UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING,
SCIENCE AND MATHEMATICS

SCHOOL OF OCEAN AND EARTH SCIENCES

A comparative analysis of the calcification transcriptome and proteome of *Emiliania huxleyi*

by

Holger Anlauf

Thesis for the degree of Doctor of Philosophy

September 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Doctor of Philosophy

A COMPARATIVE ANALYSIS OF THE CALCIFICATION TRASNCRIPTOME AND PROTEOME OF *EMILIANIA HUXLEYI*

Holger Anlauf

Calcium carbonate precipitation by marine organisms is an acknowledged contributor to the global carbon cycle. Coccolithophores are unicellular marine phytoplankton, that by excreting calcium scales and performing photosynthesis, contribute largely to the flux of atmospheric carbon to the surface and deeper oceans. The impact of elevated future ocean temperature and ocean acidity on the mechanism and rates, by which coccolithophores secret their calcium carbonate scales internally and contribute to the global carbon cycle, has been widely studied. However, biomineralisation in coccolithophores, the expression of molecular pathways, and the timing of gene expression related to calcification are still poorly understood. To better understand the process of calcification the transcriptome and proteome of calcifying G1-phase *Emiliania huxleyi* cells were investigated in the light and dark using next generation techniques. The results showed clear differences in *both* the transcriptomic and proteomic profiles between the photosynthetically enhanced calcification and the dark calcification phase. Interestingly, the bulk of the biomineralisation genes were higher expressed in the dark calcification phase at low calcification rates, suggesting that a large proportion of the molecular calcification machinery is bound to the Golgi apparatus and endoplasmic reticulum, which are complemented in the early G1-phase following cytokinesis. Furthermore, the results suggest that a set of biomineralisation genes exhibits continuous expression in both conditions of the G1 phase, whereas other genes are more abundantly expressed in the calcification phase. The importance of the calcium binding proteins calreticulin, calnexin, and calmodulin in the calcification phase was confirmed by transcriptomic and proteomic data. Proton pumping V-type ATPases were found higher expressed in dark phase but was still highly expressed in the enhanced calcification phase in the light. Calcium transport related gene expression of members of the NCKX ($Na^+/Ca2^+-K^+$ exchanger), NCX (Na^+/Ca^{2+}exchanger), and CAX (calcium exchanger) were stronger in the low calcification phase, whereas SERCA-type calcium transporting ATPases were nearly equally expressed in both condition but originating from different genes that were expressed in either the light or the dark. Furthermore, transcriptome exploration suggested syntaxin and synaptobrevin could play an important role in calcification related vesicle fusion. The results have important implications for better understanding the timing of calcification related gene expression throughout the *E. huxleyi* cell cycle and for potential transcriptomic plasticity in response to changing environmental conditions.

Table of Contents

List of Tables

List of Figures

DECLARACTION OF AUTHORSHIP

I,   HOLGER ANLAUF

declare that this thesis and work presented in it are my own and has been generated by me as a result of my own research.

A COMPARATIVE ANALYSIS OF THE CALCIFICATION TRASNCRIPTOME AND PROTEOME OF *EMILIANIA HUXLEYI*

I confirm that:

1. This work done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have quoted from the work of others, this is always clearly attributed;

4. Where I have consulted the published work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. None of this work has been published before submission.

Sigend:

Date: 9th of September 2015

Acknowledgements

So many people from all around the world have provided support, motivation and encouragement throughout this bumpy road to the completion of my Ph.D.. These Acknowledgements claim not be complete or represent any hierarchical order in appreciation.

However, at first I need to thank my final supervisors at the National Oceanography Centre, Dr. Chris Hauton who, above all, believed in me and was committed to support me at all times and gave important reviews on my writing. Chris kept the goal in sight.

Thanks to Dr. Debora Iglesias-Rodrigues, Dr. Julie LaRoche, and Dr. David O'Connor for their support and giving me the opportunity to join NOCS and complete my Ph.D. within the Calmaro program. Special thanks to Dr. Richard Edwards, Dr. Reider Andreson, and Paul Skipp who were invaluable in finding my way through the transcriptomic and proteome jungle of my data. Thanks to Dr. John Gittins for his directions in the lab and Dr. Bethan Jones for getting me started with the *E. huxleyi* culturing and proteomics work. Dr. Alex Poulton was patient with me and the *E. hux* cultures, which for some time had an experiment-fright and were committing suicide instead of utilizing the isotopic carbon source we had prepared for them. Thanks to Dr. Sophie Richier and Dr. Luke Mackinder for the valuable science think tanks on biomineralisation and the molecular drivers. Special gratitude to Dr. Aldo-Aquino-Cruz for his companionship and help during the 24-hour over-night sampling.

My family and so many friends at NOCS and outside have stayed with me on this road. Flo, Frauke, Hanna, Alex, Lauren, Sophie, Christian, Martha, FE, Victoire, Lavinia, Sudipta, the unknown George, Basti, Aaron, Christian K, Jan, Socratis, Katrin, Kathrine and more gave me perspective, encouragement and laughs that kept me going. The community of the National Oceanographic Centre and KAUST provided the important social and logistic infrastructures.

The completion of my Ph.D. would not have been possible without the boost of support I have gotten at KAUST. Especially Dr. Susana Carvalho and Prof. Burt Jones made it possible and relentlessly encouraged me to get it done.

A strong and important impulse came from my grandma that provided a necessary kick. Personally, the PhD thesis has been the toughest challenge of my life so far. And I believe, that it wouldn't be as difficult to prepare for and complete a swim across the British Channel. It has been a huge step for me but a relatively small one for science, I guess. At last, thanks to the ocean, for providing the interest, passion, energy, and the subject of study.

.

Abbreviations

2DE	two dimensional electrophoresis
A	adenine
ACC	amorphous calcium carbonate
AS	amino acid
At	total alkalinity
ATP	Adenosin-tri-phosphate
bp	base pairs
CAP	coccolith-associated polysaccharide
CAX	calcium exchanger
CCMP	canadian centre for the culture of microorganisms
CDC	cell division control gene/protein
cDNA	complementary deoxy-ribonucleic acid
Chl a	chlorophyll a
Ci	inorganic carbon
CP	chloroplast
CRM	certified reference material
CV	coccolith vesicle
Dark	experimental culture harvested in the dark period
DIC	dissolved inorganic carbon
DNA	deoxy-ribonucleic acid
DTT	threo-1,4-Dimercapto-2,3-butanediol
ECB	early cell cycle box
ER	endoplasmic reticulum
EST	expressed sequence tag
fg	femto gram
Fm	maximum fluorescence
Fo	initial fluorescence
fpkm	frames per kilobase per million kilobases (unit for transcript abundance)
Fv	variable fluorescence
Fv/Fm	photosynthetic quantum yield/ photosynthetic efficiency PS II
G1-phase	Growth1 or Gap 1 phase of the cell cycle
G2-phase	pre-mitotic phase or final phase of the interphase
Ga	billion years
GB	Golgi body

GO	gene ontology
GPA	glutamic acid, proline and alanine
H3	Histone 3
HSP	heat shock protein
ID	identifier
JGI	Joint Genome Institute
L	Litre
LC	liquid chromatography
LC/LC	multidimensional liquid chromatography
LD	light/dark cylce
LDS	Lithium dodecyl sulphate
Light	experimental culture harvested in the light period
ln	natural logarithm basis e
M-phase	mitotic phase / cell division
Ma	million years
MADS-box	gene-family for DNA-binding proteins/transcription factors
MALDI	matrix-assisted laser desorption ionization
mE	micro Einstein
μL	micro Litre
mL	milli Litre
μM	micromolar
mM	millimolar
MPSS	massive parallel signature sequencing
mRNA	messenger ribonucleic acid
MS	mass spectrometer
MS/MS	tandem mass spectrometer
NA	not available
NCKX	Na^+/Ca^{2+}-K^+ exchanger
ncRNA	non coding ribonucleic acid
NCX	Na^+/Ca^{2+} exchanger
NGS	next generation sequencing
NZEH	New Zealand Emiliania huxleyi strain
OA	Ocean Acidification
PCL	member of the cyclin-dependent kinase family
pCO_2	partial pressure of carbon dioxide
pI	isoelectric value of protein
PIC	particulate inorganic carbon

PLY	Plymouth culture collection of marine microalgae
PM	plasma membrane
POC	particulate organic carbon
PON	particulate organic nitrogen
PS I	photosystem I
PS II	photosystem II
PSC	highly acidic polysaccharides
qPCR	quantitative polymerase chain reaction
RNA-Seq	RNA sequencing
rRNA	ribosomal ribonucleic acid
RuBisCo	Ribulose-1,5-bisphosphate carboxylase oxygenase
S-phase	Synthesis phase of the cell cycle/replication of DNA
SAGE	serial analysis of gene expression
SDS	sodium dedecyl sulfate
SERCA	sarco/endoplasmic reticulum Ca^{2+} - ATPase
SLC	solute carrier
SSH	suppressive subtractive hybridization
TPC	total particulate carbon
tRNA	transfer ribonucleic acid
v/v	volume in volume
v/w	mass in volume
VCX-type	vacuolar cation exchangers

Chapter 1.   General Introduction

In nature the counterpart of chaos is not cosmos, but evolution
(Gould and Waller, 2008)

1.1. Phytoplankton evolution –
   Coccolithophores and the supremacy of *Emiliania huxleyi*

The rise of marine photosynthetic organisms began with the evolution of prokaryotic anoxygenic photoautotrophs ~3.5 Ga (billion years) ago (Finazzi and Moreau, 2010). Over hundreds of millions of years, prokaryotic oxygenic photoautotrophs oxidized the Earth's atmosphere and the oceans by the fixation of CO_2 using energy from sunlight (Gould and Waller, 2008) and allowed the evolution of single celled eukaryotic microorganisms – *Protista* (Haeckel, 1866). Circa 1.8 Ga ago an endosymbiotic event occurred. A photoautotroph cyanobacterium was encapsulated by the early eukaryote establishing the ancestor of all eukaryotic photosynthetic (Bhattacharya and Meldin, 1995; Keeling, 2004; Tirichine and Bowler, 2011). The enslaved cyanobacterium evolved into a plastid, the organelles of plants and algae that provide photosynthetic and other biochemical pathways. In a secondary endosymbiotic event of still debated timeframe, around 1-1.9 Ga ago, the ancestral phytoplankton host engulfed a red algae establishing the "red linage" of nanoplankton, which is grouped within the chromalveolates (Yoon et al., 2004; de Vargas et al., 2007; Keeling, 2009). Three phytoplankton groups of this "red linage" namely dinoflagellates, coccolithophores and diatoms, all containing the photosynthetic pigments chlorophyll $a + c$ (Saez et al., 2004), dominate the modern ocean. The fossil record shows that the present dominance of dinoflagellates, coccolithophores and diatoms began in the Mesozoic Era - 251 to 65 Ma (million years) ago (Falkowski, 2004; Bown et al., 2004). Coccolithophores are the most abundant group of the Haptophytes and contribute significantly to the group's biodiversity (Jordan and Chamberlain, 1997). The prymnesiophyte ancestor of the coccolithophores developed the ability to form ornamented plate scales by controlled intracellular precipitation of calcium carbonate ($CaCO_3$) between 329 and 220 Ma ago (de Vargas et al., 2007; Liu et al., 2010) and have played key roles in the global cycling of carbon and the development of Earth's present climate system through their photosynthetic fixation of carbon and calcium carbonate precipitation (Thierstein et al., 1977; Robertson et al., 1994; Buitenhuis et al., 1996; Holligan et al., 1993; Ridgwell, 2005). From the Triassic until the present, over 4000 discrete morphological types of coccoliths have been confirmed in the geological record (de Vargas et

al., 2007). Today's oceans inhabit only ~280 morphological distinct coccolithophores (Young and Henriksen, 2003) of which *Emiliania huxleyi* (Lohmann) Hay et Mohler (Fig. 1) is the most abundant cosmopolitan coccolithophore (Winter et al., 1994; Bijma et al., 2001; Beaufort et al., 2007).
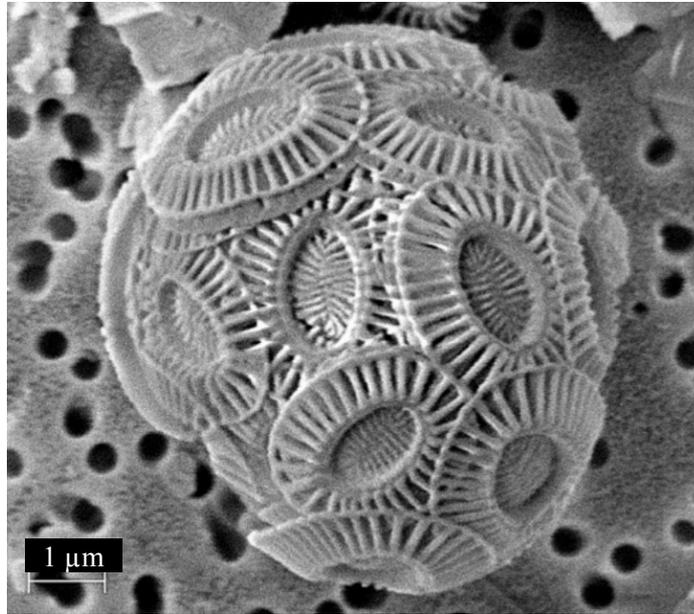
Figure 1-1: *Emiliania huxleyi* SEM. Scanning electron image (SEM) of an *Emiliania huxleyi* cell showing interlocking coccoliths creating the coccosphere. The specimen was isolated in a sample from Raunefjorden, Norway.

It is believed that *E. huxleyi* evolved around 250,000 years ago from the genus *Gephyrocapsa* (Thierstein et al., 1977; de Vargas et al., 2007). Physiological, immunological, and morphological differences separate 200 different strains within the species complex *E. huxleyi* (van Bleijswijk et al., 1991; J. R. Young et al., 1992; Young, 1994; Medlin et al., 1996; Young et al., 1999; Findlay and Giraudeau, 2000). Its cells are only 2.5 - 5.5 μm in diameter (van Bleijswijk et al., 1991). When *E. huxleyi* blooms occur they may cover over 1×10^6 km^2 exceeding cell concentrations of 10^6 cells L^{-1} (Holligan et al., 1993; Tyrell & Merico, 2004). Therefore, *E. huxleyi* blooms can affect a) regional weather patterns through the production of dimethylsulfoniopropionate, which after more reaction steps to SO_2 and sulphate acts as a cloud formation nuclei (Charlson et al., 1987; Hegg et al., 1991), b) seawater alkalinity (Holligan et al., 1993; Robertson et al., 1994), and c) the marine carbon pump (Buitenhuis et al., 1996; Rost and Riebesell, 2003) (see Section 1.2). The ease of culturing and maintaining *E. huxleyi* in the laboratory has allowed a great deal of scientific progress in terms of understanding the biology, physiology and diversification of the *E. huxleyi* species complex. Although a single species, it is

becoming increasingly evident that large genetic and physiological variations between *E. huxleyi* strains exist (Medlin et al., 1996). *Emiliania huxleyi* shows a high intraspecific physiological variability in the ability to grow in different salinities (Brand, 1984), synthesize long chain lipids (Riebesell et al., 2000; Volkman et al., 1980), and to form coccoliths under elevated CO_2 in seawater (Langer et al., 2009; Lohbeck et al., 2014). A high genetic heterogeneity has been suggested to explain the intraspecific physiological response variability in *E. huxleyi* species complex and that this may also found its overall ecological success (Lohbeck et al., 2012). High levels of genetic variability were also evident in blooms of *E. huxleyi* showing differences between geographic locations, over time, within and in between blooms and populations (Medlin et al., 1996; Iglesias-Rodriguez et al., 2006). This resulted in a discussion about the species concept in *E. huxleyi* and a proposed bacterial-like pan-genome (Medini et al., 2005). A pan-genome comprises a group of core genes (found in all strains) and dispensable genes - specific only to certain strain (Medini et al., 2005; Read et al., 2013) and could explain the physiological response variability between *E. huxleyi* strains.

1.2. The biogeochemical processes of the carbon cycle and the importance of coccolithophores

The global carbon cycle describes the fluxes (pumps) of organic and inorganic carbon between mayor reservoirs e.g. the terrestrial biosphere, the atmosphere, the lithosphere and the oceans (Fig. 2). The fluxes of carbon between the carbon reservoirs are driven and interconnected by physical, chemical and biological processes. The oceans store the largest carbon dioxide inventory on the planet containing ~ 50-60 times more carbon than the atmosphere (Siegenthaler and Sarmiento, 1993; Houghton, 2007; Riebesell et al., 2009). In the ocean the solubility pump, biological pump, and carbonate counter pump (Volk and Hofert, 1985; Falkowski, 2000) describe the uptake of CO_2 from the atmosphere, its biological fixation to organic molecules by photosynthesis or respiratory processes and the release of CO_2 from biogenic calcification (Fig.2), respectively.
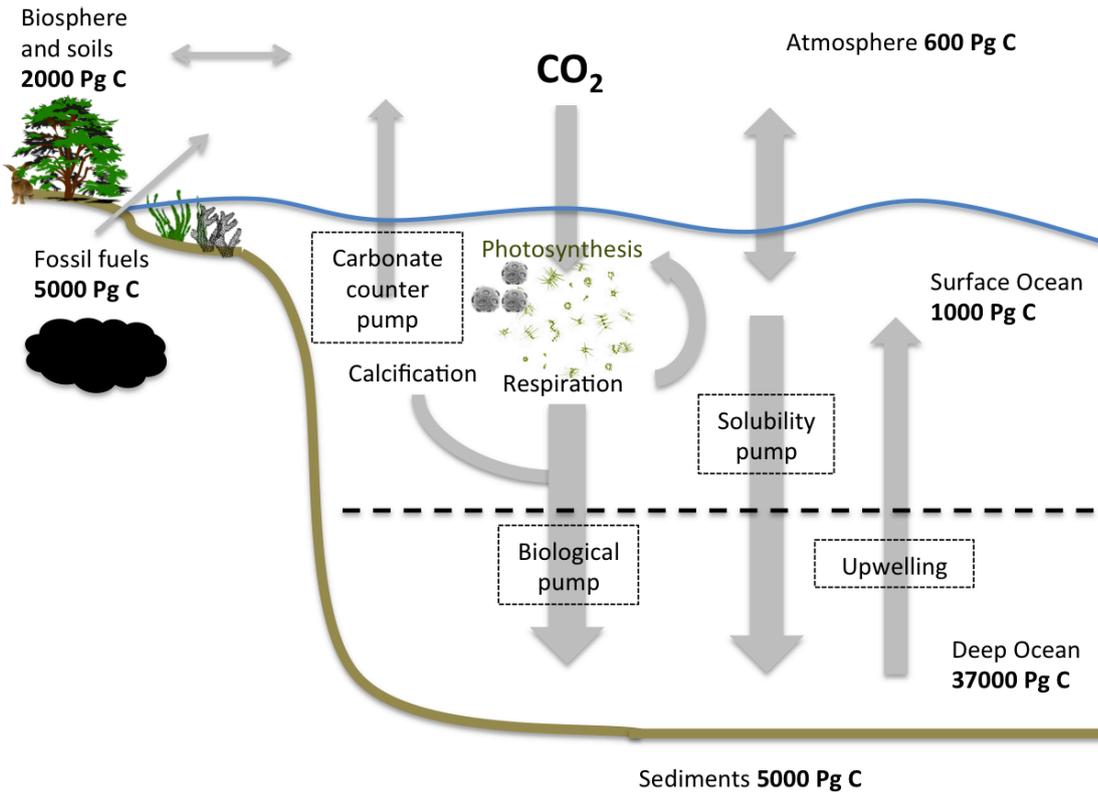
Figure 1-2: The global carbon cycle. Fluxes of carbon cycling between the biosphere, atmosphere and ocean. Dashed boxes label the major marine carbon pumps. The approximate storage capacities of the reservoirs (bold letters) is given in Pg C (10^{15} g C = gigatons C) after Zeebe and Ridgwell (2011).

In the Solubility pump (Figure 1-2) carbon dioxide dissolves in seawater and reacts with water molecules forming carbonic acid subsequently dissociating to carbonate, bicarbonate, and protons. This solubility pump is reflected by equation 1.

$$CO_2 + H_2O \rightleftarrows H_2CO_3 \rightleftarrows HCO_3^- + H^+ \rightleftarrows CO_3^{2-} + 2H^+ \qquad \textit{(Equation 1)}$$
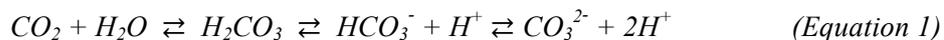
The biological pump (Figure 1-2) comprises the net carbon flux from biological processes such as photosynthetic carbon fixation and respiration/degradation of organic matter into the deep sea. Carbon fixation into 3-phosphoglycerate catalyzed by Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCo) is the main processes taking up CO_2 from the surface ocean and influencing the biological carbon pump. Regardless of their commonly microscopic size between 0.4 and 200 μm unicellular marine phytoplankton provide 45% of the global annual net primary production but represents less than 1% of the Earth's photosynthetic biomass (Field et al., 1998). Equation 2 shows the uptake of CO_2 and its photosynthetic conversion to particulate organic carbon (POC) by phytoplankton. Decomposition of organic matter and respiration are described by the reversion of equation 2 to produce CO_2.

$$CO_2 + H_2O \longrightarrow CH_2O + O_2 \hspace{3cm} \textit{(Equation 2)}$$

   Coccolithophores comprise around 10% of the phytoplankton biomass (Tyrrell and Young, 2009). While the contribution of coccolithophores to global marine photosynthetic carbon fixation is relatively small (Gregg and Casey, 2007), coccolithophores have contributed largely to the formation of deep-sea $CaCO_3$ deposits around the globe (Volk and Hoffert, 1985; Milliman and Droxler, 1996), driving the Carbonate counter pump (Fig. 2). Coccolithophores and other marine calcifiers utilize HCO_3^- ions when producing $CaCO_3$ minerals (Nimer et al., 1992; Buitenhuis et al., 1999; Herfort et al., 2002). The reaction of $CaCO_3$ precipitation requires concentration of Ca^{2+} and CO_3^{2-} ions above the saturation state (Equation 3.1). Carbonate ions are indirectly produced in a high pH milieu (Equation 3.2). During the conversion of bicarbonate to carbonate, H^+ are produced, indirectly increasing CO_2 levels due to lower pH levels (Equation 3.3). Therefore, the process of calcification becomes a net source of CO_2 when HCO_3^- is utilized (Equation 3.4.).

$$Ca^{2+} + CO_3^{2-} \longrightarrow CaCO_3 \hspace{3cm} \textit{(Equation 3.1)}$$

$$HCO_3^- \longrightarrow CO_3^{2-} + H^+ \hspace{3cm} \textit{(Equation 3.2)}$$

$$H^+ + HCO_3^- \longrightarrow H_2O + CO_2 \hspace{3cm} \textit{(Equation 3.3)}$$

$$Ca^{2+} + 2\,HCO_3^- \longrightarrow CaCO_3 + H_2O + CO_2 \hspace{2cm} \textit{(Equation 3.4)}$$

   Coccolithophores, scleractinian corals, foraminifera, molluscs and coralline algae are the most significant calcium carbonate producers in the ocean (Milliman, 1993; Iglesias-Rodriguez et al., 2002). One third to half of the present marine $CaCO_3$ is formed by coccolithophores (Milliman, 1993; Baumann et al., 2004) contributing largely to the global carbonate budget (Archer et al., 2000). Up to 70% of coccolithophore scales dissolve in the upper 1300 m of the ocean, taking up additional CO_2 from seawater (Sabine et al., 2004). Coccolithophores' calcite scales provide also ballast for the export of particulate organic carbon (Milliman, 1993; Armstrong et al., 2001; Klaas and Archer, 2002) and form large deep-sea sediments (Volk and Hoffert, 1985; Milliman and Droxler, 1996). A reduction of calcification rates in coccolithophores would therefore directly affect the marine carbon cycle (Zondervan et al., 2001). Less bicarbonate would be converted to CO_2 reducing the sea to air CO_2-flux in the carbonate counter pump (Riebesell et al. 2009). Furthermore, less organic matter would be stored in the deep ocean because ballasting by coccoliths and sinking rates would decrease while $CaCO_3$ - dissolution remains (Armstrong et al., 2001; Klaas and Archer, 2002; Engel et al., 2009).

About half of the CO_2 - emissions from fossil fuel combusted have sequestered in the ocean where it has caused a decrease in oceanic pH by 0.1 units since pre-industrial times (Haugan and Drange, 1996; Sabine et al., 2004). This phenomenon is termed ocean acidification (OA) (Raven et al., 2005) and was found to affect the calcification and growth rates in coccolithophores and other marine calcifiers (Riebesell et al., 2000; Reynaud et al., 2003; Langer et al., 2006; Orr et al., 2005; Fabry, 2008). However, *E. huxleyi's* calcification and growth responds to OA is not uniform (e.g. Riebesell et al., 2000; Iglesias-Rodriguez et al., 2008). Diverse responses in growth rates and the physiological processes involved in calcification are the basis for the observed intraspecific variability (Langer et al., 2009). For a wide range of marine metazoans the process of calcification has an important physiological role that defines the organisms' ecological niche (Cusack and Freer, 2008). Therefore, present knowledge is far from developing a generalized model of calcification and its atmospheric feedbacks if details of the processes involved in calcification remain illusive (Zondervan, 2007; Riebesell et al., 2009). New incentive to study the biomineralisation physiology *of E. huxleyi* were commenced a) to improve our understanding of the biochemical and molecular pathways of the processes of calcification, b) to understand how the genetic plasticity can explain the variable responses, and c) to develop a greater understanding of the likely feedbacks changes in coccolithophore calcification suggest for the global carbon cycle.

1.3. Coccolithophore biomineralisation – Coccolithogenesis

Biomineralisation describes the production of minerals to form shells, scales and skeletons in living organisms (Weiner et al., 2003). Biominerals nucleate in physically and chemically isolated intracellular or extracellular compartments under rigorous biological control often directly connected to an organic matrix (Lowenstam, 1981; Lowenstam and Weiner, 1989). Therefore, specific morphological features have to be present in the organism to establish, isolate and conserve the chemical gradients of pH homeostasis and ion concentrations (e.g. Ca^{2+}, carbonate ions) for the reaction towards the biogenic precipitation of the mineral.

The globally most abundant biogenic minerals are calcium carbonate minerals (Lowenstam and Weiner, 1989), of which 7 polymorphs are crystalline and one amorphous. Crystalline polymorphs of $CaCO_3$ are calcite, aragonite and vaterite. Calcite is more resistant to dissolution in seawater than aragonite and high-Mg calcite (Milliman, 1974). Magnesium calcite (Mg-calcite) and protodolomite contain magnesium. Hydrocerussite contains lead and monohydrocalcite water in combination $CaCO_3$. Amorphous calcium carbonate (ACC) is a non-crystalline precursor of calcium carbonate, which in marine species has been postulated to act as

a transport-form of calcium carbonate to the site of crystallization (Addadi et al., 2003; Weiner et al., 2003).

The intracellular nature of the calcium carbonate mineralization process in coccolithophores is a unique feature. Calcification in *E. huxleyi* is a process under intense biological control (Paasche, 1962). From scanning electron microscopy (SEM) studies in the species *E. huxleyi* (Wilbur and Watabe, 1963; Klaveness, 1976; van der Wal et al., 1983; van Emburg et al., 1986; Westbroek et al., 1989; Young et al., 1999), *Pleurochrysis carterae* (Manton and Leedale, 1969; van der Wal et al., 1987), *Coccolithus pelagicus* (Manton and Leedale, 1969; Taylor et al., 2007), *Scyphosphaera apsteinii* (Drescher et al., 2012), and *Hymenomonas coronata* and *Ochrosphaera verrucosa* (Inouye and Chihara, 1980), to name only a selection of studies, we have learned about the morphology of biomineralisation in coccolithophores. The mineralization of coccoliths occurs in an intracellular, Golgi-derived compartment, namely the coccolith vesicle (CV). The CV is positioned adjacent to the nucleus (van der Wal et al., 1983). The reticular body (RB) is a labyrinthine membrane system that connects distally to the CV throughout the calcification process in *E. huxleyi*, *Gephyrocapsa sp.* and *C. pelagicus* (Klaveness, 1976). However, a different feature is present in *P. Carterae*, where multiple vesicles show coccoliths at varying growth stages and form a trans-Golgi without a RB (van der Wal et al., 1983). The reticular body is expected to provide rapid ion transport and supersaturated conditions inside the CV (Brownlee and Taylor, 2004; Taylor et al., 2011). During coccolithogenesis, the coccolith-vesicle moves away from the nucleus, the CV membrane and plasma membrane fuse and the coccolith is then extruded through the layer of organic scales on the cell surface subsequently interlocking with adjacent coccoliths (Marsh, 2003; Taylor et al., 2007).

1.4. The molecular basis of biomineralisation in coccolithophores

Coccolithophogenesis is a complex process under rigorous genetic control involving the expression of many proteins of yet unknown functions to provide a network of interacting structural and regulatory molecules (Young and Henriksen, 2003; Henriksen et al., 2004; de Vargas et al., 2007). For the precipitation of calcium carbonate ion concentrations of calcium and carbonate above the saturation state of calcium carbonate must be reached inside the CV and H^+ ions need to be removed from the CV's lumen. Ion-transporters or exchangers that involve, transport, and co-transport Ca^{2+}, Mg^{2+}, H^+, Cl^-, Na^+ ions, highly acidic calcium binding proteins, carbonic anhydrase, and organic matrix-associated polysaccharides are generally involved in controlling coccolithophogenesis. The following paragraphs will introduce specific

important macromolecules involved in molecular pathways for cell functioning and especially calcification, such as Ca^{2+} transport, carbonate or inorganic carbon (Ci) transport, H^+ transport, and organizing and controlling the biomineralisation process. Table 1-1 summarizes the molecules involved in the process of calcification in *E. huxleyi*.

Ca^{2+} transport

In seawater – the external medium - calcium is a conservative element with concentrations of around 10 mM. Net fluxes of Ca^{2+} and inorganic carbon (Ci) from the external medium to the intracellular CV and low concentrations of free cytosolic Ca^{2+} of around $<<$ 1 μM have to be sustained to guarantee coccolith production, avoid toxic effects of high Ca^{2+}, and coccolithophore functioning (Brownlee and Taylor, 2004; Verret et al., 2010; Araki and Gonzales, 1998; Sanders et al., 2002; von Dassow, 2009). The absence of Ca^{2+} causes a decline of growths rates by hampering important cell signalling processes involved in cell division and lowers PIC production (Herfort et al., 2004; Timborn et al., 2007; Leonardos et al., 2009; Mackinder et al., 2011). At the plasma membrane (PM) of coccolithophores a variety of Ca^{2+} - channels/transporters were identified that modulate the Ca^{2+} PM passage. Voltage-gated Ca^{2+} permeable channels (CA_v) were suggested to regulate the Ca^{2+} uptake from the external medium (Mackinder et al., 2010; Dolphin, 2009). A recent model by Mackinder et al. (2011) suggests that the peripheral endoplasmic reticulum adjacent to the PM might act as a trap and storage pool for Ca^{2+} ions. The Ca^{2+} storage capabilities of the ER are well documented (Meldolesi and Pozzan, 1998). In the ER proteins such as calreticulin bind calcium ions (Jacopo and Pozzan, 1998; Wahlund et al., 2004; Quinn et al., 2006). The calcium flux through the cytosol into the CV and ER is probably mediated by SERCA-type ATPases (sarco/endoplasmic reticulum Ca^{2+} - ATPase) and Ca^{2+}/H^+ VCX-type (vacuolar cation exchangers) to enhance an Ca^{2+}-rich lumen of the ER and achieve the required supersaturated conditions for the precipitation of $CaCO_3$ inside the CV (Mackinder et al., 2010).

Inorganic carbon transport

Evidence shows that bicarbonate is the primary carbon source for the synthesis of $CaCO_3$ in coccolithophores (Frankignoulle et al., 1994; Israel and González, 1996; Raven, 2011; Bach et al., 2013). Transfer of bicarbonate along a gradient through the plasma membrane into the cytosol is thought to occur passively (Brownlee et al., 1995; Anning et al., 1996). Furthermore, recent finding also suggest the involvement of a putative HCO_3^- transporters of the SLC44 (SLC4 family) in the carbonate / bicarbonate balance in *E. huxleyi* (von Dassow et al., 2009).

The solute carrier SLC4 exchanging Cl^- / HCO_3^{2-} was successfully inhibited in *E. huxleyi* (Herfort et al., 2002) and it was suggested that it might interact with carbonic anhydrase enzymes (CA) in other eukaryotes (Vince and Reithmeier, 2000; Sterling et al., 2002).

Carbonic anhydrases are ubiquitous metalloenzymes and catalyze the reaction presented in equation 4 in both directions.

$$CO_2 + H_2O \rightleftarrows HCO_3^- + H^+ \hspace{2cm} \text{(Equation 4)}$$

The reaction is fundamental for the buffering of the bicarbonate system and therefore acid/base compensation in most organisms throughout all kingdoms. In higher plants and phytoplankton it is key to deliver CO_2 to RuBisCo for photosynthesis but isoforms of CA were suggested to be located in the cytosol and periplasmic space (Badger and Price, 1994; Mackinder et al., 2010). It was suggested that CA facilitates the conversion of HCO_3^- into CO_2 at the cytosolic face of the plasma membrane decreasing the local concentration of HCO_3^- at the cytosolic transport site (Isenberg et al., 1963). The expression of transcripts was up-regulated in calcifying versus non-calcifying phases supporting it (Quinn et al., 2006; Richier et al., 2009). Twelve different transcripts of CA have been identified in *E. huxleyi* (von Dassow et al., 2009). Specifically α-CA (prokaryotic like) and γ-CA - possibly located in or adjacent to the CV – were confirmed to be involved in the calcification process (Soto et al., 2006; Quinn et al., 2006). The CA reaction (Equation 4) also binds protons derived from the process of calcification and produces CO_2 for photosynthesis.

H^+ transport

The regulation of proton levels is required to achieve certain membrane potentials, stable conditions of cytosolic pH and to excrete protons derived from CA activity, in the chloroplast and the CV to stabilize enzyme functioning, photosynthetic efficiency, and a regime for calcium carbonate precipitation. Calcification and photosynthesis affect with cytosolic pH. Proton pumping from the CV into the cytosol is necessary to create pH values of around 7 in the *E. huxleyi* cytoplasma and around 8.2 in the CV favouring the passive conversion of HCO_3^- to CO_3^{2-} (Dixon et al., 1998; Brownlee et al., 1995). The synthesis of $CaCO_3$ using HCO_3^- produces protons and decreases the pH in the CV (refer to equation 3) requiring further proton pumping. Protons may also leak across membranes because of a conductive gradient that would be present at the cytosolic CV boundary (Suffrian et al., 2011), requiring continuous proton pumping. P-type (plasma membrane) ATPases and V- (vacuole) type ATPases are universal in eukaryotes removing protons from the cytosol (Finbow et al., 1997) and the latter may be Ca^{2+}

stimulated, as shown in *Pleurochrysis sp.* (Araki and Gonzalez, 1998; Corstjens et al., 2001).
The electrochemical gradient between seawater pH 8.2 and cytosol pH 7 creates a plasma
membrane potential greater than -60 mV where high voltage gated H^+- permeable channels
(HVCN1) could remove intercellular H^+ (Corstjens et al., 2001; Mackinder et al., 2010).

1.5. The putative function of molecules involved in the nucleation of calcium carbonate

    An organic matrix within the CV is important for the ordered nucleation of calcium
carbonate in coccolithophores. At the initiation of the coccolith formation a proto-coccolith ring
forms around an organic base plate of highly acidic polysaccharides (PSC). Coccolithosomes,
Golgi-derived polysaccharide and Ca^{2+} dense vesicles, supply polysaccharides and Ca^{2+} to the
coccolith forming vesicles (Outka and Williams, 1973; van der Wal et al., 1983). In *P. carterae*
three highly acidic polysaccharides, PSC-1, PSC-2 and PSC-3, have been identified to play a
role in calcite precipitation (Marsh et al., 1992). Immunolocalization has identified that PSC-1
and PSC-2 are synthesized in the Golgi cisternae, stay associate with the organic base plate
from the onset of crystallization until and after the coccolith is extruded to the cell surface
(Marsh, 1994). Mutants without PSC-2 produced less than 5% calcite compared to wild-type
cells (Marsh and Dickinson, 1997) and PSC-3 mutants were unable to form mature coccoliths as
seen in wild-type cells (Marsh et al., 2002).

    Emiliania huxleyi appears to have one polysaccharide termed the coccolith-associated
polysaccharide (CAP) associated with extracellular coccoliths (de Jong et al., 1976). This
complex polysaccharide is a galacturonomannan that consists of a sulphated mannan backbone
with a variety of side chains rich in galacturonic acid (Vliegenthart, 1981), showing similarities
to PSC-3 (Marsh, 2003). CAP has the ability to inhibit crystal formation in a supersaturated
solution of Ca^{2+} and CO_3^{2-} (Borman et al., 1982) but also directs crystal growth by its specific
affinity to calcite (Henriksen et al., 2004). In *E. huxleyi* a highly acidic macromolecule
containing high levels of glutamic acid, proline and alanine (GPA) has potential Ca^{2+}-binding
capacity (Corstjens et al., 1998) and has been identified in *E. huxleyi* morphotypes A and B. The
expression of the coccolith morphology motif (CMM) genomic region in the *E. huxleyi*
correlates with the morphotype of the coccoliths (Schroeder et al., 2005). Counter-intuitively, a
down regulation of GPA gene transcripts was found in calcifying cells suggesting a possible but
debatable inhibitor role of GPA at high Ca^{2+} concentrations (Mackinder et al., 2011).
Mackinder et al. (2010) suggested that an amorphous calcium carbonate (ACC) phase is being
created in smaller vesicles (precursor of CV compartments) or Golgi cisternae, transported and
merging into a CV, where the crystallization is directed by silk fibroin proteins in a non-

aqueous environment (Weiner et al., 1984). An transient precursor such as ACC is found in mollusc larval shells transforming into aragonite (Weiss et al., 2002) and in the spicules of sea urchin larvae eventually forming calcite (Beniash et al., 1997). The proposed presence of ACC in coccolithophores finds theoretical support by the limits of the CV membrane that cannot fit the required abundance of membrane associated transporters to establish the chemical environment for $CaCO_3$ precipitation and uphill Ca^{2+} -transport onto its surface (Mackinder et al., 2010). If the larger surface of an endomembrane system as the Golgi body (cis-Golgi network) could provide sufficient protein bound Ca^{2+} the coccolithosome would represent a carrying vesicles and the CV would sustain precipitation favourable conditions. A likely acidic lumen of the *E. huxleyi* Golgi body (Llopis et al., 1998) would support the directed movement of calcium ions into the Golgi body.

Table 1-1: Molecules potentially involved in calcification. A selection of organic molecules in *E. huxleyi* possibly involved in calcification processes and their predicted location, function, and reference (GB/ER: Golgi body/endoplasmatic reticulum, CV: coccolith vesicle, CP: chloroplast, PM: plasma membrane, CS: cytosol).

Process	Protein Name	Location	Protein function/origin	Related publication
Ca^{2+} transport	Vacuolar Ca^{2+}/H^+ antiporter	GB/ER	Cation-exchanger, (H^+ / Ca^{2+} - VCX_1-type antiporter)	Von Dassow et al., 2009
	Putative Na^+/Ca^{2+} exchanger	GB/ER	Efflux of cytosolic Ca^{2+} (NCKX family)	Von Dassow et al., 2009
	SERCA-type Ca^{2+} -pump	GB/ER	ATP driven, Mg^{2+}, as cofactor	Von Dassow et al., 2009
	CA_v	GB/ER	Voltage gated Ca^{2+} - channel	Richier et al., 2011; Dolphin, 2009
	Putative Ca^{2+} channels	PM	In flux of Ca^{2+}, ATP or electrochemical gradient driven	Mackinder et al., 2011
Ci – transport	Carbonic anhydrases,	CS	DIC regulation (*E. hux* - specific – δ-EhCA1, γ-EhCA2)	Von Dassow et al., 2009; Richier et al., 2011
	$Cl-/HCO3^-$ exchanger,	PM	Membrane transporter, solute carrier 4 (SLC 4-family)	Von Dassow et al., 2009
H^+ transport	Vacuolor type H^+ - ATPase	CV	H+ removal from the CV, possibly Ca^{2+} stimulated	Corstjens et al., 2001 Araki and Gonzalez, 1998
	Clathrin	CV/ER	In clathrin-coated vesicles, in association with V-ATPases	Jones et al., 2011
	P-type H^+ - ATPase	PM		Araki and Gonzalez, 1998
Ca^{2+} binding	CAP (Coccolith-associated glutamic acid)	CV	Regulates directional crystal growth, similar to acidic polysaccharides	de Jong et al., 1976; Marsh, 2003b; Henriksen et al., 2004
	Calreticulin Calnexin	ER, CV	Ca^{2+} homeostasis, cellular functions (Ca^{2+}, chaeperone – assist in non-covalent folding)	Wahlund et al., 2004; Quinn et al., 2006
	Calmodulin (CaM)	ER, CV	Calcium-binding messenger protein	Sotoj et al., 2006
Nucleation	Highly acidic polysachharides	CV	Initiation of coccolith formation. Three identified forms: PSC1, PSC2, and PSC3.	Outka and Williams, 1973; van der Wal et al., 1983; Marsh et al., 1992
	GPA	CV	$CaCO_3$ nucleation and regulating growth	Corstjens et al., 1998
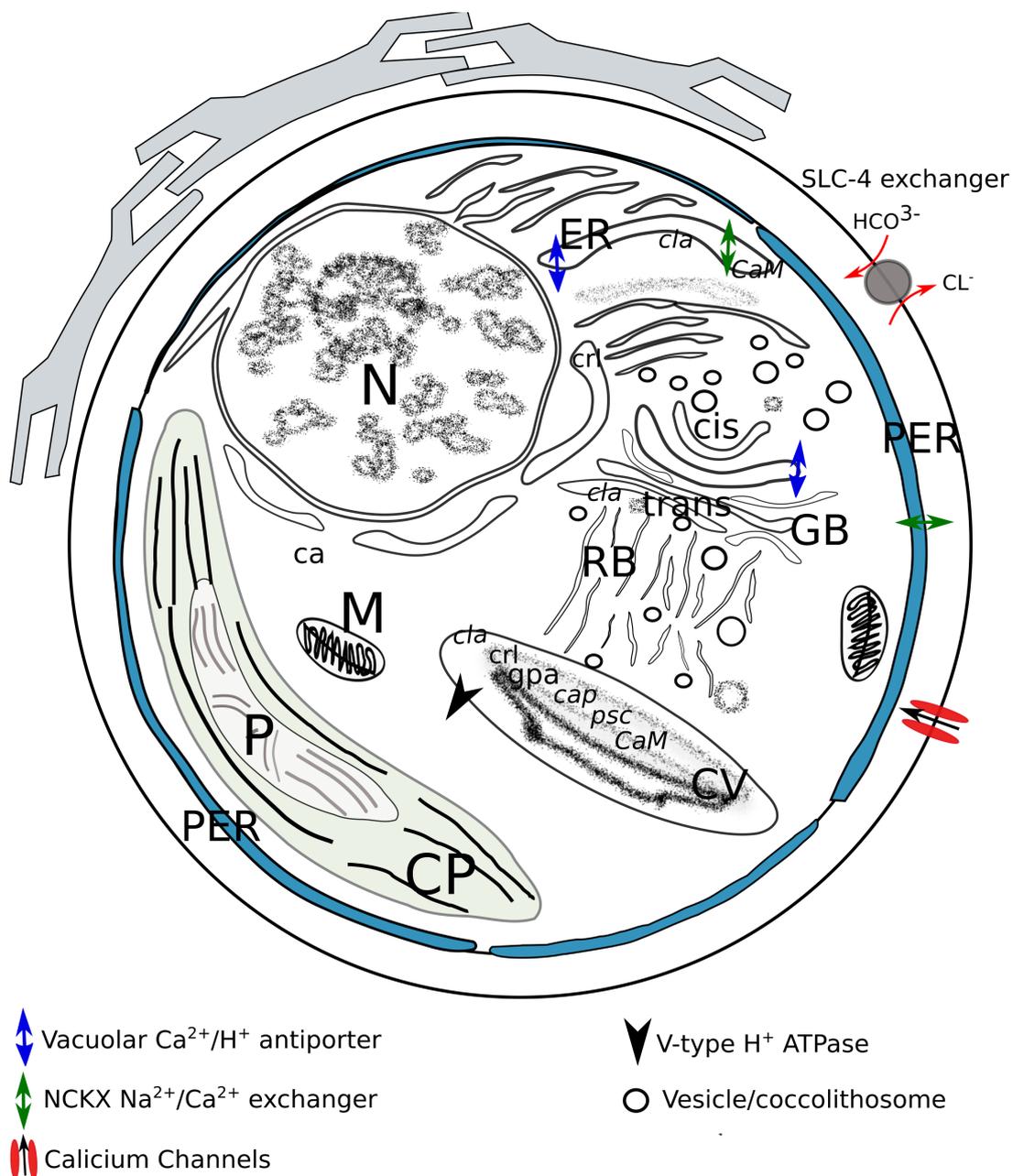
Figure 1-3: Cell structure of *Emiliania huxleyi* showing cell structure, suggested transporters and molecules involved in biomineralisation according to the reviewed literature and summary in Table 1-1. ER: endoplasmic reticulum, N: nucleus, RB: reticular body, M: mitochondria, CP: chloroplast, P: pyrenoid, PER: peripheral endoplasmic reticulum, GB (cis and trans face): Golgi body. Italic letters show the suggested position of molecules involved in calcification processes, such as: *cla*: clathrin, *CaM*: calmodulin, *crl*: calreticulin, *gpa*: glutamic acid, proline and alanine, *cap*: coccolith-associated glutamic acid, *psc*: highly acidic polysachharides, and *ca*: carbonate anhydrase.

1.6. Understanding cell functioning in the post-genome era

In the post-genome era scientists are studying the products of the genome, namely proteins and expressed RNAs, such as messenger RNA (mRNA), tRNA (transfer RNA) and rRNA (ribosomal RNA) to quantify gene expression and unravel the relationship between the genome and cell functioning. The cell's functional properties are delivered by the transcription of the coding regions of a gene (exon) into mRNA translation by ribosomes into amino acid (AS) strands and the folding of such polypeptides into globular or fibrous forms, facilitating a biological function. Cytosolic post-translational modification such as acetylation, adenylation, phosphorylation, sulphuration or addition of N- or O-linked saccharide are critically influencing the formation of the functional 3-D structure, activity, distribution, and the subcellular compartmental placement of the protein (Kalia and Gupta, 2005). Stimuli to the cell can induce transient event, where protein modification not necessarily requires the expression of a gene (Mitton and Kranias, 2003).

In transcriptomic research all RNA molecules including mRNA, rRNA, tRNA and non-coding RNA (ncRNA) (Huettenhofer and Vogel, 2006) are in the focus of interest. The transcripts represent a signature of gene-expression and activated genes. The entity of transcripts also provides the essence in understanding the functional and regulatory elements at the transcriptional and post-transcriptional level. One application of transcriptomics is the quantification of gene expression changes at different developmental stages or physiological conditions (Wang et al., 2009). Furthermore, whole cell transcriptome assessments can be used conduct *de novo* sequencing of an organism's genome (Grabherr et al., 2011). In metatranscriptomics RNA sequences from a group of interacting organisms are assessed and may point to important and novel molecules in ecosystem functioning (Warnecke and Hess, 2009).

Unlike the transcriptome, which is subject to post-transcriptional modification, the proteome gives insight into the protein inventory of a cell or organisms. Messenger RNA is often altered or not translated into proteins at all (Anderson and Seilhamer, 1997; Chen et al.; 2002; Tian et al., 2004; Choi et al., 2008). Proteins are much more stable and reflect active biochemical pathways more effectively than transcripts. In proteomic research, the quantity, diversity, structures and functions of proteins are studied (Anderson and Anderson, 1998; Blackstock and Weir, 1999). The term proteome originally describes the 'total protein complement of a genome' (Wasinger et al., 1995). Presently, most studies address the 'whole cell proteome', which should reflect the entire protein inventory of the cell at a current physiological state or point in time (Wilkins et al., 1996), but due to methodological constrains only subsets of the

'whole cell proteome' may be observed. Studying the physiological state of a cell at the protein expression level considers post-translational modifications (Olsen et al., 2006; López, 2007), non-translated mRNA (Eddy, 2001) and genetic regulatory processes. The proteome reflects a snapshot of the putative effective cellular processes at a given moment (Wilkins et al., 1996). Proteome studies in combination with transcriptome studies give comprehensive information about cell gene expression, protein translation and the physiological status of the cell.

Because of modifications at the post-transcriptional and post-translational level, whole-cell transcriptome and proteome assessments capture the complexity of cell functioning better than the genome sequence. Because, the total mRNA population of a cell does not correlate with the abundance of proteins (Anderson and Seilhamer, 1997; Chen et al. 2002; Tian et al., 2004; Choi et al., 2008) the proteome gives improved information about current physiological processes. Furthermore, the proteome is more stable than RNA transcripts (Anderson and Anderson, 1998). However, in standalone proteome or transcriptome studies limits of the inferences that can be drawn exist. For instance, the amount of ncRNA may be very large, whereas in more complex organisms only 5-10 % of the genetic code may be transcribed into mRNA (Frith et al., 2005). Alternative splicing or alternative transcription initiation and termination, and RNA editing can produce variants of mRNA that do not reflect the genetic origin. Protein's dynamic structure, their interaction with other proteins and molecules make it difficult to receive more than a snapshot of the cell's proteome (Humphrey-Smith et al., 1997). Furthermore, there are many technical challenges in extracting and characterizing proteins that cannot be easily amplified and have several post-translational modifications if they are yet unknown (Blackstock and Weir, 1999).

1.6.1. A history of transcriptome research technologies

In the mid-1970s the first quantification of mRNA using the Northern blotting method, which applies gel electrophoresis and blotting with hybridization probe-dependent detection of target RNAs (Alwine et al., 1977) became available. Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR) improved the quantification of RNA from genes of pathways of interest in 1992, because the produced complementary DNA (cDNA) was more stable, easier to sequence and to quantify. RT-qPCR was further enhanced by the development of real-time qPCR in 1996 that delivers immediate results (VanGuilder et al., 2008).

The era of transcriptomics and transcript discovery also began in the 1990s when great progress in the analysis of transcripts was made by establishing high-density microarray platforms. Microarray platforms allow parallel measurements of thousands of different transcripts, revealing a snapshot of the wider transcriptional state of an organism rather than looking at a limited number of transcripts of interest, as the case in PCR based studies (VanGuilder et al., 2008). Microarrays are based on the hybridization of transcripts from a biological sample to complementary DNA (cDNA) sequences, also called probes, located on a solid surface, such as a chip. The probes are arranged spatially in cohorts (features) on the chip for recruiting reverse transcribed sequences of the sample. The quantification of transcript abundance is achieved by fluorescent signals emitted by the reverse transcribed RNA copies recruiting to the features. For this, the extracted poly-adenylated mRNA was reverse transcribed in the presence of nucleotides that are linked to fluorescent dyes. The resulting fluorescent complementary DNA molecules bind to the sequences located on the microarray. A final washing step removes unbound cDNA, before the microarray is scanned with a laser that excites the fluorescent dye associated with the cDNA. The brightness of each feature is a measure of the relative abundance of the transcripts of each gene that is represented on the array.

An early sequencing-based technology to measure the abundance of mRNA transcripts is the Serial Analysis of Gene Expression (SAGE, Velculescu et al., 1995). In SAGE large numbers of short sequence tags excised from mRNA transcripts are reverse transcribed and the cDNA sequenced on genome sequencers. Bioinformatic tools are then used to identify the source gene of the sequenced tags. The advantage of SAGE is that it holds the possibility to discover unknown transcripts, isoforms or splicing products of transcripts (Velculescu et al., 1995).

1.6.2. Next generation sequencing (NGS)

More recently next-generation sequencing technologies have become available to assess the total transcript diversity and abundance. This approach, called RNA-seq, is based on the general procedure by which the selectively extracted RNA molecules of interest, such as eukaryotic mRNA via poly-adenylated tags, are revers transcribed to cDNA. The cDNA sequences are determined by rapid parallel sequencing (Wang et al., 2009). Next generation sequencing of transcriptomes has advantages over microarray studies due to the potential of isoform discovery and the ability to determine transcript abundances at a higher resolution than typically possible with microarrays. With RNA-seq genome-wide transcription may be analysed, thus providing additional features such as, analysis of novel transcripts, siRNA, miRNA and alternative splicing events. Furthermore, RNA-seq holds keys to analyse the regulation of gene expression by exposing transcribed but non-translated regions, such as UTR (un-translated regions) that

may act in regulating gene expression (Nookaew et al., 2012). A typical RNA-seq experiment generates large numbers of short reads that are aligned to an annotated reference genome using mapping tools (Martin and Wang, 2012). In next generation sequencing read lengths between 30 and 400 bp, depending on the sequencing technology and platform, are possible. Commonly, used sequencing platforms are SOLEXA, 454, and Illumina platforms. The continuous advances in sequencing technologies and protocols have established RNA-seq as an attractive analytical tool in transcriptomics and reduced running costs of an experiment (Nookaew, 2012).

Next generation sequencing transcriptomic experiments produce large amounts of data. Therefore, the bioinformatic toolbox is also in continuous expansion to increase the validity of the assumptions of functional properties concluded from NGS experiments. Many statistical methods applied in microarrays can be applied with a few modifications to NGS data sets (Fang et al., 2012). The approach to analyse the data comprises three steps: data preprocessing, statistical analysis, and functional interpretation. The data-preprocessing step includes artifact filtering and short read alignment/assembly. To identify differentially expressed transcripts among different samples or conditions statistical tests are performed. The results can be further analyzed to gain functional insights using gene ontology annotations (Fang et al., 2012). Statistical tools that distinguish for example the overexpression of transcripts comprised by a functional group, such as GO (gene ontology), are also available (Chen et al., 2012). More modern bioinformatic algorithms are able to conduct de-novo sequencing of a genome from the transcriptome information (Birol et al., 2009).

1.6.1. Next generation high throughput transcriptomic methods

Originally a variety of transcriptome analysis methods were developed based on sequencing previously cloned tags located in specific transcript locations (usually 3' or 5' ends) such as serial analysis of gene expression (SAGE) (Velculescu et al., 1995) and massively parallel signature sequencing (MPSS) (Brenner et al., 2000) Quantitative polymerase change reaction (qPCR) - experiments and microarrays also use pre-selected sequences from known genomes to assess transcripts at different physiological states of an organism (Quinn et al., 2006). Using hybridization-based microarrays expression levels over the entire transcriptome are accessible. However, background hybridization levels and different hybridization properties of the probes may bias the results (Denoeud et al., 2008).

The Sanger – sequencing is a method along with computational means that has increased the quality and speeds of DNA strand sequencing (Sanger et al., 1997). In expressed sequence tag (EST) analysis, randomly picked complementary DNA (cDNA) representing RNA transcripts

are partially sequenced by Sanger sequencing and redundancies removed to create a cDNA library (Nagaraj et al., 2006; Bonaldo et al., 1996). The double stranded cDNA is synthesized by reverse transcription using a specialized enzyme, the reverse transcriptase. However this approach is relatively low throughput, expensive and generally not quantitative (Wang et al., 2009). Serial analysis of gene expression (SAGE) (Velculescu et al., 1995; Harbers et al., 2005), cap analysis of gene expression (CAGE) (Kodzius et al., 2006), and massively parallel signature sequencing (MPSS) (Brenner et al., 2000) use tags to provide higher throughput and gene expression levels. However, expensive Sanger sequencing and short tags providing a limited resolution of transcripts and isoforms are disadvantageous (Wang et al., 2009).

Denoeud et al. (2008) and Wang et al. (2009) introduced RNA-Seq (RNA sequencing) that can create a library of cDNA fragments with adaptors attached to one or both ends and by this assess the total RNA or a fraction, such as poly (A)+ - extracted. High-throughput sequencing platforms such as those from SOLEXA or Illumina sequence each molecule from one end (single-end sequencing) or both ends (pair-end sequencing) in the range of 30-400bp (base pairs). The reads reflect the gene expression levels and structure for each gene after alignment to a reference genome or reference transcripts. A greater number of alternative splicing events can be revealed in RNA-seq and it promises the discovery of novel sequences and molecules that have not been annotated or sequenced previously (Denoeud et al., 2008). The on-going advances in next generation sequencing techniques and decreasing costs will continue to make RNA-seq a broadly used technique addressing biological and ecological questions.

1.6.2. HiSeq Illumina based RNA sequencing

Next generation genomics apply massive parallel sequencing technologies followed by computationally matching the readouts to known reference sequences. One of the PCR-based next generation sequencing technologies for massive parallel RNA sequencing was introduced in 2007; it utilizes the Illumina sequencing platform (Shokralla et al., 2012). The Illumina platform comprises the steps of library preparation (ligation of Illumina specific adapters), cluster generation (bridge amplification) and DNA sequencing. The general approach in most next generation sequencing technologies is very similar. The DNA of interest is randomly fragmented; the fragments are then ligated to custom linkers. The linkers establish a library on a solid surface of the cell where the DNA is amplified and then sequenced by stepwise nucleotide synthesis cycles. During each cycle of the sequencing process nucleotides attach to millions of the amplified DNA-fragments of interest and are detected in parallel over the entire surface of a flow cell lane by light sensitive sensors (Mardis, 2008). Each lane in an Illumina flow cell can hold one DNA sample. The steps of cDNA fragmentation, adapter ligation, flow cell surface

attachment, bridge amplification, building of double stranded DNA strands, and denaturation of double stranded are repeated until dNTPs are depleted and clusters of single-stranded DNA were built (Figure 1-4). After attaching primers to the DNA during each sequencing cycle one base labelled with a specific fluorescent dye is added. A camera records the position of the fluorescent signal when the flow-cell is scanned with an excitation laser.
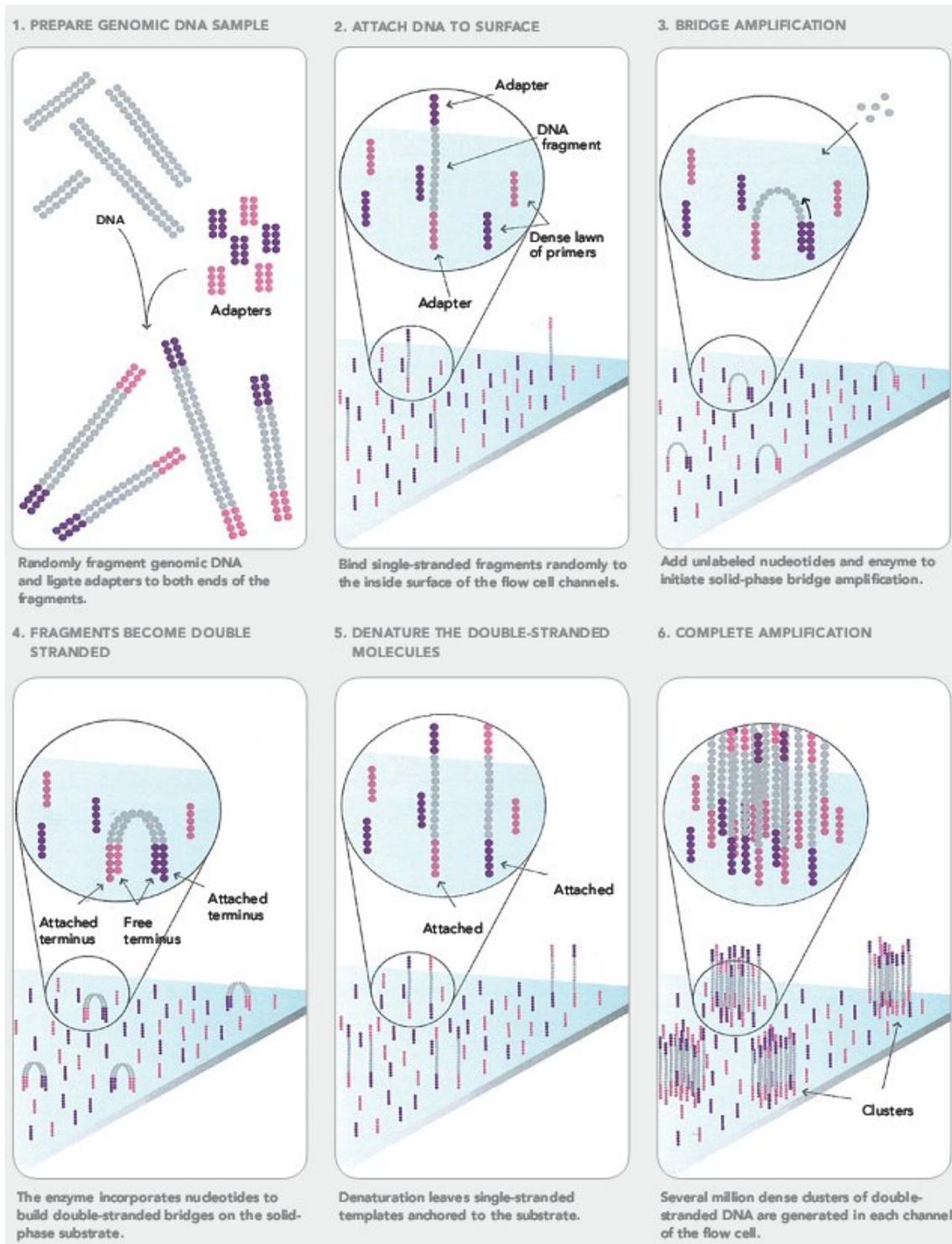
Figure 1-4: Illumina Sequencing technology. Illustration of the steps occurring in the flow cell up to DNA cluster generation using Illumina Sequencing technology. Image presented with courtesy of Illumina Incorporation.

1.6.3. Advanced proteomics methods and principles of mass spectrometry proteomics

Traditionally, proteomic studies were conducted largely by high-resolution two dimensional gel electrophoresis (2DE) that separates proteins according to their charge and size (O'Farrell, 1975). The detection of some post-translational modifications of proteins is also possible with 2DE (Anderson & Anderson, 1998). Though a promising technique at its time 2DE could not be improved to overcome issues of low sample throughput, poor reproducibility and poor resolution for proteins of extreme isoelectric (p*I*) values, and high labour intensity (O'Connor et al., 2000; Rehm, 2006). Modern quantitative high-resolution methods apply mass-spectrometry (MS) that can identify large numbers of proteins by their biological mass at their core (O'Connor et al., 2000; López, 2007).

In mass-spectrometry the mass to charge ratio of ions of an ionized samples is measured. Electrical and magnetic fields within the mass spectrometer alter the speed and direction of the injected ions. The degree by which the ion is deflected depends on its mass-to-charge ratio. The principle parts of sections of a basic mass spectrometer are the ion source, deflector and detector (see Figure 1 -5).
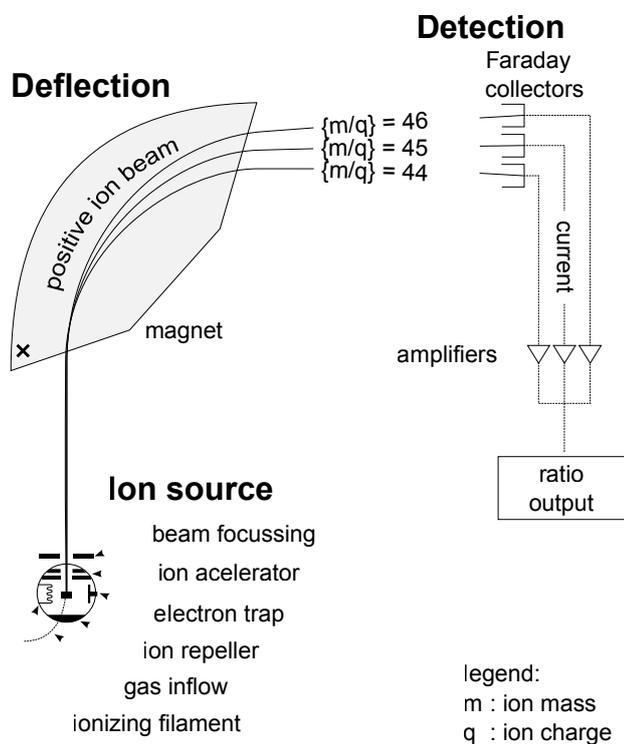
Figure 1-5: Schematics of a mass spectrometer after Devon Fyson (Wikimedia Commons), showing the key features of sample ionization, deflection and detection.

When a detector, such as discrete dynode electron multiplier, for example, receives the ions, their relative abundance is displayed as spectra. Another approach to obtain mass analysis of ions in mass spectrometry is by time of flight mass spectrometry (TOF) (Wiley and McLarren, 1955). In TOF, the mass-to-charge ratio is determined by the time it takes for ions with the same kinetic energy to reach the detector. Accelerated by an electrical field the ions arrival at the detector only depends on their mass-to-charge ratio (*m/z*). The principle design of MS and TOF was further developed. In the tandem mass spectrometry (MS/MS) two mass spectrometry steps excel the analytical capabilities of a single instrument. MS/MS involves that fragments of the precursor ions are scanned. In the first MS – step ions are selected according to their *m/z*. Fragment ions are produced by a variety of processes, for example, dissociation, collision, ion-molecule reaction, or photo-dissociation. The separation and identification of fragment ions of a particular mass-to-charge ratio occurs in the second stage of mass spectrometry. Quadrupoles have become an important and economic part in mass spectrometer, as they can be utilized as detector of *m/z* (Q), filter, focussing, or collision units (q) for ions of specific mass to charge ratios if multiple quadrupoles are aligned. Four cylindrical parallel rods, of which opposing rods receive an AC current with a DC offset for ion manipulation, are assembled in one quadrupole (Gross, 2011). By operating quadrupoles with the AC current and DC offset at radio frequency (r.f. mode) an average three-dimensional electrical force is created that focuses or traps electrically charged ions (Watson and Sparkman, 2007).

In mass spectrometry proteomics research two general strategies are followed, which are controlled by means of sample preparation. The bottom-up approach uses complete that are enzymatically digested into peptides whereas in the top-down approach intact proteins or their exact fragments are analyzed. The top-down approach requires large sample sizes and high mass accuracy instruments compared to the more versatile bottom-up approach, in which a greater variety of mass spectrometry instrument and reduced sample amount can be utilized. When dealing with complex protein samples the bottom-up approach is preferred (Yates et al., 2009). The investigation of a cell's proteome follows the bottom-up approach and has similarities to whole DNA sequencing, where small parts of the DNA sample are created and clustering overlapping DNA sequences into larger sequences reveals the entire DNA sequence. Because of the fractionation of DNA or protein digestion into peptides the techniques were given the terms "shotgun genomics", and "shotgun proteomics", respectively (Wolters et al., 2001; Aebersold and Mann, 2003). In shotgun proteomics the masses from peptides of a protein digest are matched to known or predicted masses of peptides in database searches (Perkins et al., 1999; Yates et al., 2009). Based on in silico-generated fragmentation patterns or reading frame translations together with genetic database searches the components of the proteome can

be established (e.g. Jones et al., 2013). Bottom-up proteomic studies apply tandem mass spectrometry and the associated computational pipelines as core instrumental features when investigating chemical protein modification and conduct quantitative analysis of proteins in cells or tissues. However, the data is not without bias and may show limited protein sequence coverage of the identified peptides, missing post-translational modifications, and bias related to the origin of redundant peptide sequences (Yates et al., 2009).

A liquid chromatograph (LC) or capillary electrophoresis system can separate polar peptides of low volatility according to their net charge prior to mass spectrometric analysis. Thus, the protein identification and quantification process in MS-proteomics becomes a multi-step procedure. At first the protein-concoction is digested into a complex peptide mixture and run in a LC, for example. The chromatographically focused fractions of the peptide mixture are ionized and injected into the mass spectrometer for analysis. Nano-electrospray ionization (Mann and Wilm, 1995), partial (no discharge) atmospheric pressure chemical ionization (Cristoni et al., 2002), and matrix-assisted laser desorption ionization (MALDI) (Patterson and Aebersold, 1995) are common ionization techniques in mass spectrometry. Electrospray ionization multidimensional liquid chromatography combined with tandem mass spectrometry (LC/LC-MS/MS) is a preferred method for dissolved protein samples and delivers weight signatures of peptides that require further in silico annotation by reference genomes or EST (expressed sequence tags) – libraries to identify the functional proteins (e.g. Jones et al., 2011). The MALDI expanded the application field in biological research as it could be utilized to ionize protein dots directly from 2D electrophoresis gels using automated stepper motors, for example (Chernushevich et al., 2001). One downside of MALDI mass spectrometry is the lack of sufficiently detecting peptides of low abundance from silver stained gels (Berndt et al, 1999). However, the laser ionization was not restricted to samples attached to matrix or surfaces and could also ionize elutes from continuous flow after liquid chromatography providing high sample throughput rates and high mass accuracy (Pasch and Pode, 1995; Nagra and Li, 1995). MALDI-generated ions are predominantly singly charged providing a preferable ion stream for top-down analysis of high-molecular-weight proteins. The low shot-to-shot reproducibility and labour intense sample preparation (Yates et al., 2009) evolved into a method using digested proteins as a source and an optional layered methods. If the peptide passed the MALDI-TOF MS without identification nanoelectrospray tandem mass spectrometry was applied to ensure peptide identification (Shevchenko et al., 2000). By coupling MALDI with orthogonal injection hybrid quadrupole time-of-flight mass spectrometer (MALDI QqTOF) a further substantial breakthrough in proteomics research was made. However, the Orbitrap® is the state of the art technology in bottom-up proteomic research. In Orbitrap®, ions are trapped in orbits around static electric fields. The ions spin around a central electrode and oscillate in axial direction. The

axial oscillations of the ion rings are detected by an outer electrode and amplified. The m/z ratio

is calculated from imaged orbits based on the principle of Fourier transform ion cyclotron

resonance mass spectrometry. MALDI is the ionization method of choice in Orbitrap®, because

it works better with pulses of ions rather than continuous streams of ions (Yates et al., 2009;

Gross, 2011).

1.6.4. MALDI quadrupole time-of-flight mass spectrometry

MALDI QqTOF (Figure 1-5) is a next generation proteomic research tool that has been

appreciated by the analytical community for its robustness and its powerful and unique

capabilities (Chernushevich et al., 2001). The QqTOF mass spectrometry harvests the power of

analytical (Q) and collision (q) quadrupoles (Shevchenko et al., 2000; Chernushevich et al.,

2001). As the MALDI ionization produces pulses of ions rather than continuous beams it was

incompatible with scanning mass analysers. Hence a cooling and focussing quadrupole is added

immediately after the ionization (Figure 1-6). In MS mode the Q1 quadrupole could perform

MS analysis by using the TOF section as a total ion current detector only. The advantage of

better peptide identification through TOF spectra implies that the Q1 is used for Q1 calibration

and tuning only (Chernushevich et al., 2001). TOF measurements are of multiple ions are

synchronized by Q2, which traps ions before releasing modulated bursts into the TOF.
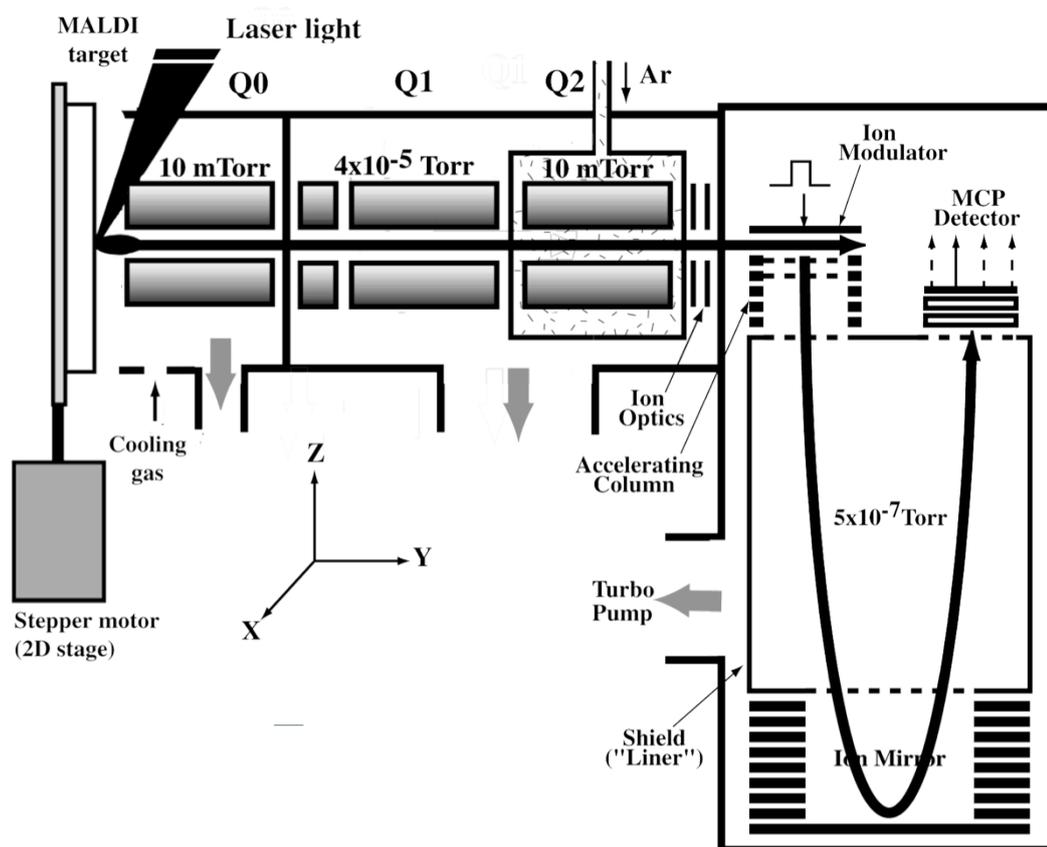
Figure 1-6: MALDI QqTOF mass spectrometer design shown with a matrix or gel source as samples (2D stage). In MALDI QpTOF, Q refers to a mass-resolving quadrupole (Q1), q refers to hexapole collision cell (Q2) (r.f.-only quadrupole) and TOF refers to a time-of-flight mass spectrometer. The Q0 is an additional r.f. quadrupole that was added to provide collisional cooling and focusing of the ions created by the laser impulse. In MS only operations the Q1 would act as the mass analyzer. Since the TOF provides improved spectra quality the Q1 is only used for calibration and tuning. The Q2 acts as a collision cell trapping ions before releasing short burst into the TOF modulator and finally to the TOF itself (after Chernushevich et al., 2001)

The resulting spectra benefit from the high resolution and mass accuracy of the TOF instruments, and also from their ability to record all ions in parallel, without scanning (Chernushevich et al., 2001). A MALDI QqTOF can identify the components of simple mixtures if the relative abundance of the components differs by the factor 10. An upstream high performance separation technique like UPLC or LC for complex protein mixtures with lower relative abundance of peptides in a trypsin digested protein sample will increase the analytical performance of MALDI QqTOF further, as the peptides entering the MS analysis have been pre-fractionated by size (Reemtsma, 2003; Farré et al., 2007). Furthermore, chromatographic fractionation in combination with QqTOF mass spectrometer provides a significantly higher mass resolution than quadrupole analyzers by themselves and the available mass resolution of TOF instruments diminish the problem of isobaric interferences allowing for analysis of complex samples using of isobaric tags (Farré et al., 2007).

1.6.5. Transcriptomics and proteomics in coccolithophore studies

The genome *E. huxleyi* (strain CCMP1516) has been sequenced and is publically available at the Joint Genome Institute (www.doe.jgi.gov). Recent studies on *E. huxleyi* have used this reference and attempted to shed light into the function of transcripts and the molecular origin of coccolith formation using expressed sequence tags (ESTs), suppressive subtractive hybridization (SSH), serial analysis of gene expression (SAGE), cDNA microarrays, and qPCR to identify numerous proteins and genes also found active in the biomineralisation process of other marine organisms (Wahlund et al., 2004; Nguyen et al., 2005; Dyhrman et al., 2006; Quinn et al., 2006; Fujiwara et al., 2007; Richier et al., 2009; Dassow et al., 2009; Richier et al., 2011). These experiments were designed to reveal differences in gene expression related to calcification rates using nutrient starvation or strains showing different calcification properties. However, the transcripts provided only low recovery rates for proteins and their isoforms putatively involved in the calcification process, encourages subsequent studies applying state of the art sequencing techniques to unravel the genome functioning of *Emiliania huxleyi* and coccolithophores (Dyhrman et al., 2006; Quinn et al., 2006).

Pioneering proteomics-research on coccolithophores by Jones et al. (2010) identified 99 proteins operating from a variety of physiological pathways in the *E. huxleyi* strain NZEH (PLY ♯ M219) (Jones et al., 2011). However, this protein recovery rate appears very low when compared for example to *Saccharomyces cerevisiae* where 4300 verified proteins were identified (King et al., 2006). To discover proteins involved in the calcification process, pCO_2-enrichment experiments, known to enhance calcification in *E. huxleyi* NZEH were conducted (Jones et al., 2013). Under high pCO_2 conditions, only 4 homologous protein groups were significantly down-regulated in the NZEH strain, namely: Histones H2A, H3, and 4 (H4) and a chloroplastic 30S ribosomal protein. Histone 4 is a structural component of the nucleosome and is a methyl donor involved in transcription and translational processes (Chiang et al., 1996). These results however are difficult to evaluate with respect to biomineralisation, because a) both molecules are unlikely to play a role in biomineralisation process and b) *E. huxleyi* PLY ♯ M219 under high pCO_2 was found to accumulate more PIC. The underlying mechanism for the increased PIC accumulation remains unclear. The identification of more key genes and proteins is required to shed light the functioning of calcification processes (Richier et al., 2009).

1.6.6. Key aims and Objectives

- Determine the growth rates, physiological parameters, and rates of photosynthesis and calcium carbonate production using stable isotope incubations over 24 hours.
- Investigate cell cycle related differences of DNA content over a 24-hour period to conclude cell cycle phases.
- Determine adequate cell harvesting times for the sampling of transcriptome and proteome at high and low calcification phases where other cell cycle phase related processes and the involved gene expression are limited.
- Compare the transcriptomes at high and low calcification phases of the cell cycle of *Emiliania huxleyi* with an emphasis on transcripts that are potentially involved in biomineralisation using next generation sequencing techniques.
- Compare the proteome at high and low calcification phases of the *Emiliania huxleyi* cycle with an emphasis on proteins that are potentially involved in biomineralisation applying shotgun-proteomics for protein identification and protein quantification.
- Investigate general differences of the transcriptome and proteome at high and low calcification phases of the *Emiliania huxleyi* cell cycle..
- Investigate coherent molecular and gene expression characteristics of the observed transcriptome and proteome patterns.

Chapter 2. Rates of photosynthesis and calcification in the cell cycle of *Emiliania huxleyi* (Lohmann), Hay et Mohler

2.1. Introduction

Coccolithophores possess varying rates of cellular calcification based on their genetic repertoire and plasticity, depending on environmental conditions, and cell or life cycle phases (Paasche, 2002). For example, high, low- or non-calcifying strains of the coccolithophore *E. huxleyi* have been observed, including the high calcifying strain NZEH (PLY# M219) (Iglesias-Rodriguez et al., 2008; Shi et al., 2009) and the low calcifying strain 279 (Scottish Marine Biological Association) (Nimer et al. 1992). Limited availability of phosphorous has been shown to result in an increase of calcium carbonate per cell (Paasche, 1998; Müller et al., 2008) as low light conditions were found to decrease net calcification rates (Bleijswijk et al., 1994). The life cycle of *E. huxleyi* is characterised by three different life-cycle phases or cell types: 1) diploid non-motile coccolith-bearing (calcified cells); 2) diploid non-motile non-calcified (naked) cells; and 3) haploid non-calcified cells of the flagella phase bearing only organic scales.

Within a natural 24 h cycle of light and dark, calcifying *E. huxleyi* cells generally undergo phases of higher and lower calcification rates and pass through the cell cycle (Müller et al., 2008; Paasche and Brubak, 1994; Paasche, 2002). These diurnal fluctuations in calcification rates correlate with the two phases, which are dominated by photosynthetic carbon assimilation and cell division. A more detailed look at the cell cycle with respect to cell division separates four phases: the G1-phase of photosynthetic carbon assimilation (Mitchison, 1971), the S-phase where the genomic information is replicated and the G2 and M-phases, where cell division is arranged, coordinated and completed (Figure 2-1). In the cell cycle of *E. huxleyi* maximal calcification rates are closely correlated with the photosynthetic carbon assimilation (G1-phase) (Paasche, 1962). Calcification rates arrest when the DNA is replicated (S-phase) to prepare for cell division (G2/M phase). As the majority of the cell population of *E. huxleyi* divided within a short period cells subsequently continued into the early G1-phase, where no light for photosynthesis existed (Müller et al., 2008). Some calcification in the S and G2/M phase were reported to occur in *E. huxleyi*, but at much decreased rates compared to the photosynthetic phase (Balch et al. 1992; Paasche and Brubak 1994; Sekino and Shiraiwa 1996). Furthermore, a population of cells undergoing the cell cycle is not fully synchronized (Hagiwara et al., 2012).
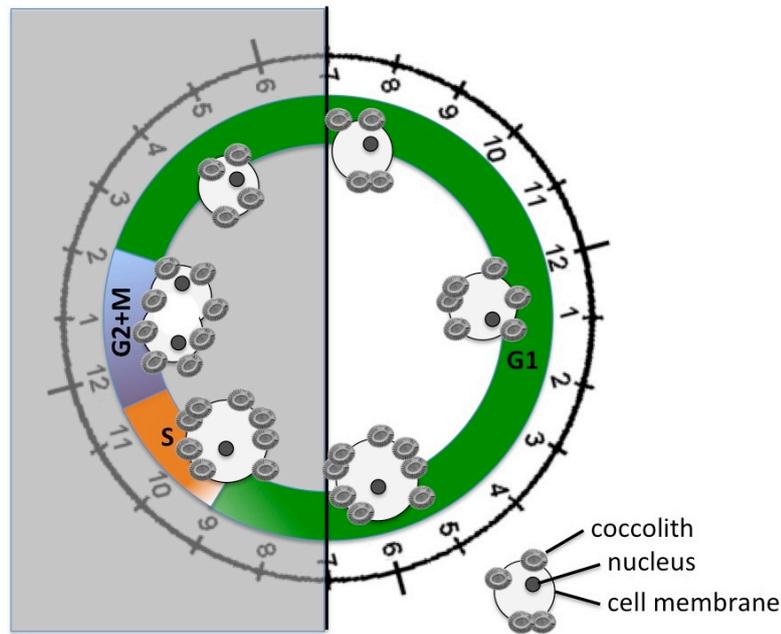
Figure 2-1: Proposed general scheme of a cell cycle of a coccolithophore based on finding of Müller et al. 2008, showing G1 –assimilation phase or gap 1, S – DNA synthesis – phase, G2 + M-cell division - mitosis-phase. Calcification and photosynthetic carbon assimilation in the G1 – phase is illustrated as a relative increase in lith-numbers and cell size. The timing and the duration of the cell cycle phases are roughly estimated from Mueller et al., 2008 Figure B, whereas in reality the phases overlap and especially the S and G2+M-phase can not be clearly separated in a population of cells. The grey shaded area indicates the dark incubation phase.

Two general models describing the mechanisms driving the cell cycle have been proposed: an internal clock (Edmunds and Adams, 1981) and a triggered oscillation, such as by the external light dark cycle that solely triggers the cell cycle phases (Spudish and Sager, 1980). Recent research has illuminated the controls of the cell cycle at the genome level (e.g. Kondorosi et al., 2000; Bisova et al., 2005; Inzé et al., 2006; Carretero-Paulet et al., 2010). A cascade of interacting molecules and gene clusters has been found to control the cell cycle oscillation. In yeast, for example, cyclins (PCL9, CLN3) and regulators for DNA replication (CDC6) initiate the cell cycle (Bastajian et al., 2013). A key role is played by the Early Cell cycle Box (ECB) comprising a gene cluster, which is regulated by further transcription factors (MADS-box's *Mcm1* or the repressors *Yox1* or *Yhp1*) describing a principal cell cycle interface through which the cell may change its fate (Bastajian et al., 2013). It is likely that a similar feature such as the ECB is present in other organisms. A change of fate leading towards a delay of cell division may occur if environmental conditions or resources within the cell are unfavourable. The survival strategy for delayed or halted reproduction is known from higher plants and animals and is a common feature in nature (e.g. Kirk, 1997). In *E. huxleyi,* a delay in cell division might also reflect a response to adverse environmental conditions. *Emiliania huxleyi's* cell division

rates were reported to reach up to 2.8 divisions per day when sufficient nutrients and light were supplied (Price et al., 1998). Moreover, a surplus of nutrients or light was found to extend the duration of the G1 phase and interrupt cell division leading to an increase of the $CaCO_3$-content per cell (Müller at al., 2008). An increase of calcium per cell was also found when *E. huxleyi* was exposed to increased levels of CO_2 (Iglesias-Rodriguez et al., 2008; Jones et al., 2013). The research summarised in this chapter investigated the diurnal rates of photosynthesis and calcification. Using ^{14}C-radionuclide incorporation, together with flow-cytometric DNA-content analysis for cell cycle phase determination, the diurnal cycle of the *E. huxleyi strain* PLY♯ M217 was described. These data were used in subsequent experiments to identify the target times to obtain samples for transcriptome and proteome analysis.

2.2. Material and Methods

2.2.1. Culturing

The strain *Emiliania huxleyi* PLY# M217 [also known as CCMP 1516 strain (Provasoli-Guillard National Centre for the Culture of Marine Phytoplankton, Maine, USA)], was obtained from the Plymouth Culture Collection of Marine Microalgae. At the National Oceanography Centre, Southampton batch cultures of PLY# M217 were maintained in f/50-Si (Si-free) seawater medium (Guillard, 1975; Balch et al., 1992). The culture medium was prepared from filter-sterilised seawater obtained during a cruise offshore from Plymouth (U.K.) using a pore size of 0.22 μm (Millipore, Billerica, USA). The medium contained final nutrient concentrations of 36 μM nitrate and 1.45 μM phosphate.

To ensure that the cells were in exponential growth at the beginning of the experiment, the 1 L medium in each replicate was inoculated to achieve 500 cells per mL. The pre-experimental cultures were exposed to irradiance levels of 94 ± 4 μE m^{-1} sec^{-1} (n=20) over a 12:12 light: dark (LD) cycle at 18 °C and were mixed daily. The pre-experimental cultures reared in exponential growth for at least 20 generations prior to radioisotope incubations.

2.2.2. Experimental Design

Three stocks of 14 L culture media in 20L Nalgene$^®$ (Thermo Scientific™, Loughborough, U.K.) culture vessels were inoculated with 2000 *E. huxleyi* cells in exponential growth. The stock cultures were grown for five generations until cell densities reached 50000 to 100000 cells mL^{-1}. The stock cultures provided three aliquot cultures of 50 mL for radioisotope incubation

with ^{14}C-labelled sodium bicarbonate (NOC/RAM/730A). The sampling of aliquot cultures commenced at 10 am on the 19th April 2011 and then every other h until 10 am, of the 20th April 2011. Particulate organic carbon (POC) and particulate inorganic carbon (PIC) were measured following the micro-diffusion technique after Paasche and Brubak (1994) and Balch et al. (2000). For biogeochemical and chemical analysis including cell counts, nutrient measurements, chlorophyll *a* per cell, particulate organic carbon and particulate organic nitrogen (PON), flow cytometric DNA content analysis, chlorophyll, and photosynthetic health, a 250 ml sample was obtained at each sampling time from stock culture (see Fig. 2-2).
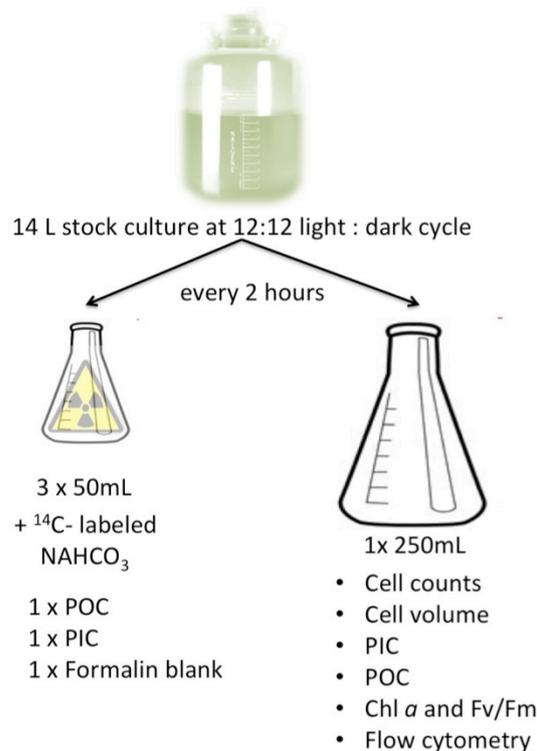
Figure 2-2: Experimental setup – cell cycle related calcification rates of *E. huxleyi*. Sampling layout for two hourly measurements for each of three stock cultures growing *Emiliania huxleyi*

2.2.3. Assessment of DNA content per cell

For the assessment of DNA content per cell one sample of 1.8 mL was taken from the well-mixed 250 mL culture aliquot and 0.2 mL of a glutaraledehyde-paraformaldehyde mixture (10% and 5% (v/v)) was added to fix the sample. The samples were allowed to incubate for two hours at room temperature and then stored at -80 $^{\circ}$C until further analysis. The DNA content was quantified by staining the defrosted samples with SYBR green I (Sigma-Aldrich, St. Louis, USA) for 45 min. SYBR green specifically binds to double-stranded DNA and the emitted fluorescent signal is proportional to the DNA content of the cells (Wittwer et al., 1997; Nunez,

2001). Multifluorescent beads (diameter 0.5 μm; Polysciences Europe GmbH, Heidelberg, Germany) were added to each sample for size calibration. The fluorescence of the cells was measured using a FACS Calibur flow cytometer (BD Biosciences, San Jose, CA, U.S.A.) with assistance from Mr. Ross Holland (National Oceanography Centre, Southampton). The ratios of cell cohorts in different cell cycle phases were then analysed using the WinMDI Vers. 2.8 software (Windows Multiple Document Interface for Flow Cytometry, Joe Trotter, San Diego, USA).

2.2.4. Cell density, cell volume, and growth rates assessment

A sample of 5 mL, taken from the 250 mL aliquot, was used to quantify cell density and cell volume. Duplicates of 1 mL from the samples were diluted 1:10 with 3 % NaCl (w/v) and analysed using a Multisizer™ 3 Coulter Counter[®] (Beckman Coulter (UK) Ltd, High Wycombe, United Kingdom) fitted with a 70 μm aperture. Averages of triplicate cell counts and cell volume measurements were assessed using the Multisizer 3 software package (Version 3.51, Beckman Coulter Inc, Brea, USA). Growth rates (μ) per h in each experimental stock culture were calculated according to Equation 1.

Equation 1: $\mu = (\ln C_1 - \ln C_0)/\Delta t$

μ: Growth rate
C_0: Initial or previous cell counts [cells/mL]
C_1: Most recent cell counts [cells/mL]

2.2.5. Particulate organic carbon (POC) and particulate organic nitrogen (PON)

To assess the contents of organic carbon and organic nitrogen per cell, duplicates of 50 mL, taken from the 250 mL culture aliquot, were filtered on pre-combusted (450°C for 12 hours) 25 mm diameter GF/F filters (Whatman, Maidstone, United Kingdom). The filters containing the samples were then stored at -20 °C until further processing. Samples and blanks were prepared for analysis by removing inorganic carbonates by sulphuric acid (H_2SO_4) fumes under vacuum for 36 hours (Verado et al., 1990). After drying for 24 hours at 50 °C, the filters were pelleted in pre-combusted aluminium foil discs (Elemental Microanalysis Ltd, Oakhampton, United Kingdom). The final quantifications of POC and PON were completed by Mr. Robert Head, Plymouth using a Thermo Finnigan Flash EA1112 elemental analyser (Thermo Scientific™, Loughborough, United Kingdom) calibrated and standardised with acetanilide. Total organic carbon and organic nitrogen production were then calculated per cell in pmol for every two-hour sampling interval.

2.2.6. Particulate inorganic carbon (PIC) analysis

To analyse the cellular PIC, duplicate samples of 50 mL from the 250 mL culture aliquot were collected on 0.2 μm Nucleopore[®] filters (Whatman, Maidstone, United Kingdom). At the end of the filtration each filter was rinsed with ~ 1 mL of alkaline ultrapure Milli-Q[TM] water (pH ~9) and then stored at -20 ^{o}C until analysis. Eventually, the samples were defrosted and treated overnight with 10 mL 0.1 M nitric acid (HNO_3) (Romil, Cambridge, United Kingdom) to dissolve the calcium carbonate containing coccoliths. Additionally, standards in the expected concentration range were prepared in the same way. Prior to PIC analyses, samples and standards were filtered through 0.22 μm Millex filters (Millipore, Billerica, MA, U.S.A) and then analysed by Mr. Darryl Green (National Oceanography Centre, Southampton) on an inductively coupled plasma-optical emission spectrometer (ICP-OES, Agilent Technologies, Wokingham, United Kingdom).

2.2.7. Nutrient measurements and dissolved inorganic carbon (DIC) analyses

A sufficient supply of nutrients is vital for phytoplankton cultures. Reduced macronutrient concentrations of N and P will alter cell physiology and in the case of *E. huxleyi* affect calcification rates (Paasche, 2000; Mueller et al., 2008). Low nutrient consumption may indicate an impaired physiological state. Nutrient concentrations of the stock cultures were measured every two hours. By filtering 30mL of stock culture through 0.22 μm polycarbonate filters cells were separated from the media. The filtrate was then transferred into a Falcon[®]-tube and stored at -20 °C until further analysis. Nutrients were measured on a Skalar San+ auto analyser (Skalar (UK) Ltd., Wheldrake, United Kingdom), calibrated for the expected range of nutrient concentrations. The detection limit of NO_3/NO_2 and PO_4 was 0.03 μmol L^{-1} and 0.01 μmol L^{-1}, respectively. The nutrient analysis was carried out by Mr. Mark Stinchcombe (NOCS, National Oceanography Centre, Southampton). Each sample was analysed in duplicates and average values were computed for each sampling point in replicate cultures. The nutrient consumption rates were calculated per cell per hour.

2.2.8.Dissolved inorganic carbon (DIC) and total alkalinity (At)

To monitor any modifications of the carbonate system within the *E. huxleyi* cultures, seawater carbonate chemistry parameters, such as dissolved inorganic carbon, total alkalinity, and pH were assessed at three time points in the course of the experiment (at 10:00 am, 8:00 pm, and 8:00 am on the 20th April 2011). For this, approximately 250 mL of stock culture was siphoned from the stock cultures into a borosilicate bottle (Duran, Fisher Scientific, Leicestershire, UK). To prevent further changes of the carbonate chemistry 50 µL of 3.5% (w/v) mercuric chloride ($HgCl_2$) were added and the DIC samples stored in the dark until further analysis. DIC and TA were then measured at the carbonate facility at the National Oceanography Centre, Southampton by Dr. Cynthia Dumousseaud using an Apollo SciTech DIC infrared analyser (AS-C3) (Apollo SciTech, Inc., Bogart, GA 30622 USA) and an Apollo SciTech AS-ALK2 Alkalinity Titrator (Apollo SciTech, Inc., Bogart, GA 30622 USA). Certified reference materials (CRM) provided by A. G. Dickson (Scripps Institution of Oceanography, University of California San Diego, U.S.A) were used as standards to calibrate the system prior each measurement. To correct the measurements for various parameters including titration acid density, nutrient concentration of the sample, temperature, salinity and CRM a MATLAB script (originally designed by Dr. Dorothee Bakker at University of East Anglia, U.K. and modified by Dr. Cynthia Dumousseaud at National Oceanography Centre, Southampton) was used.

2.2.9.Chlorophyll *a* measurements

Measurements of extracted chlorophyll *a* in algae are an established physiological parameter that can indicate N or P limitation of the cell (Geider et al., 1993; La Roche et al., 1993; Geider et al., 1998). Chlorophyll *a* (Chl *a*) concentrations were measured at 10:00 am, 4:00 pm, 8:00 pm, 0:00 am, 4:00 am + 1 day, 8:00 am + 1 day, and 10:00 am + 1 day. From the well-mixed 250 mL stock culture aliquot, a volume of 10 mL was gently vacuum filtered through a GF/F filter (25 mm diameter). The filters were transferred into glass tubes and 6 ml of 90% acetone was added. Chlorophyll pigments were extracted passively in acetone over night (12 hours) at 4° C in the dark. Quantification of chlorophyll *a* was completed on a 10AU Turner fluorometer (Turner Designs, Sunnyvale, USA) using borosilicate test tubes (Turner Designs, Sunnyvale, USA). The system was equipped with a blue mercury vapour lamp and excitation (436 nm) and emission (680 nm) filters according to the method after Welschmeyer (1994). Chlorophyll *a* concentrations in the acetone extract were calculated per mL sample after Equation 2 and are presented as Chl *a* per *E. huxleyi* cell for each sampling time.

Equation 2: $Chl\ a\ conc.\ (\mu g\ ml^{-1}) = C\ x\ (\frac{v}{V})$

> C= conc. of Chl *a* from Turner fluorometer
>
> v= volume of acetone extract
>
> V= volume of culture sample

2.2.10. Assessment of photosynthetic health

The photosynthetic efficiency or quantum yield of photosynthesis (Fv/Fm) can provide a measure of the physiological state of phytoplankton cells particularly when under nutrient stress (Kobler et al., 1983; Geider et al., 1993; Beardall et al., 2001; Moore et al., 2006). The photosynthetic quantum yield quantifies the efficiency by which the absorbed irradiance is utilised by the Photosystem II complex (PSII) (Krause et al., 1991); the maximum PSII photochemical efficiency typically decreases under nutrient starvation or extreme irradiance (Kolber et al. 1988, 1994; Geider et al. 1993b). For Fv/Fm measurements a culture aliquot of ~ 5 mL was stored in the dark for at least 25 min to relax the reaction centres of the PSII. Then Fv/Fm values were assessed by measuring F_0 and Fm, the minimum and the maximum *in vivo* chlorophyll *a* fluorescence yield after dark-adaptation, using a Fast Repetition Rate Fluorometer (FRRF, Chelsea II, Chelsea Technologies Group Ltd., West Molsey, United Kingdom). The instrument fired 100 flashes of 440 nm at 1.1 μs intervals to saturate the PSII. Values for F_0 and Fm were directly read by the FastInP-software (Vers. 1.0.0, Chelsea Technologies Group Ltd., West Molsey, United Kingdom) on the FRRF-workstation. The Fv/Fm value was then calculated based on the following equations:

Equation 3: $Fv = Fm - Fo$

Equation 4: $Fv/Fm = \frac{(Fm-Fo)}{Fm}$

> Fv = Variable fluorescence
>
> Fm = Maximum fluorescence
>
> Fo = Initial fluorescence
>
> Fv/fm = Photosynthetic efficiency of photosystem II

Values of Fv/Fm were plotted over time and compared to values reported in the literature, to estimate the photosynthetic health of *E. huxleyi* cells in this experiment. Photosynthetic quantum yield was assessed at 10:00 am, 12:00 am, 4:00 pm, 8:00 pm, 0:00 am, 4:00 am + 1 day, 8:00 am + 1 day, and 10:00 am + 1 day.

2.2.11. Radioisotope measurements of photosynthesis and calcification

To measure the rates of photosynthesis and calcification in *E. huxleyi*, the ^{14}C-incoporation into the organic and inorganic carbon fractions of the cells was measured using the general methodology described in Paasche and Brubak (1994) and Balch et al. (2000). Three aliquot cultures of 50 mL were immediately spiked with 5.25 µCi (194.25 kBq) ^{14}C-labelled sodium bicarbonate (Sigma, NOC/RAM/730A). One aliquot was poisoned with two mL of borate-buffered formaldehyde solution to create a blank that corrected for non-biological isotope exchange (Paasche, 1963). All ^{14}C-spiked aliquots were exposed to the same LD cycle as the stock cultures for 2 hours to allow ^{14}C incorporation into organic and inorganic fractions. As a result incubations starting at 8:00 pm and 8:00 am were exposed to one hour of light and one hour of darkness. The incomplete incubations of one hour in the light and one hour in the dark were excluded from the data presentation. Following the incubation for two hours, duplicate 25 mL fractions were filtered onto 25 mm diameter 0.4 µm polycarbonate filters (Whatman, Maidstone United Kingdom) for total particulate carbon (TPC) and incorporated particulate inorganic carbon. The particulate organic carbon fraction was calculated as the difference of TPC and PIC (Balch et al., 2000). For the assessment of radioisotope incorporation the TPC filter was transferred to a scintillation vial and 15 mL of scintillation cocktail (Ultima GoldTM Cocktail, Perkin Elmer, Cambridge, United Kingdom) were added. The PIC filter was transferred to a separate 20 mL glass scintillation vial, which was sealed by a septum. The septum had a plastic beaker attached to it, in which a GF/F (Whatman, Maidstone, United Kingdom) filter drenched with 200 µL phenethylamine was placed. Phenethylamine has the capacity to trap CO_2 by forming a carbonate salt. After injecting one millilitre of 1% phosphoric acid through the septum onto the PIC filter the inorganic carbon was freed as ^{14}C-labelled CO_2. The phenethylamine-drenched filter captured the developing CO_2 for radioisotope determination (Woeller, 1961). The septum closed scintillation vial containing phosphoric acid was left over night for a full dissolution of the particulate inorganic carbon fraction. Both, the PIC-filter and phenethylamine drenched filter were then transferred into separate fresh 20 mL glass vials and 15 mL of scintillation cocktail added to each. Eventually, the samples and scintillation cocktail were incubated for 24 hours at room temperature. The activity originating from the ^{14}C-labelled sodium bicarbonate incorporation was assessed on a TriCarb 2100TR (Perkin Elmer, Cambridge, United Kingdom) liquid scintillation spectrometry counter.

2.3. Results

The following sections will summarize the data collected from an *E. huxleyi* culture over a period of 24 hours using isotopic and non-isotopic methods with particular focus on calcification rates, rates of photosynthesis, and the diurnal DNA contents.

2.3.1. Dynamics of DNA content, cell density, growth rates, and cell volume of *E. huxleyi* stock cultures over a 24 hour period

To determine the cell cycle phase in *E. huxleyi* stock cultures the relative DNA content calculated by the WinMDI software was assessed every two hours over a period of 24 hours. The histograms in Figure 2-3 showed an increase of relative DNA content compared to the total DNA content in the first 16 hours (until 2 am in the Dark period) of the measurements. After 2 am the peak of lower relative DNA content increased while less cells with higher DNA content were observed. The histograms show no clear separation of cell cycle phases.
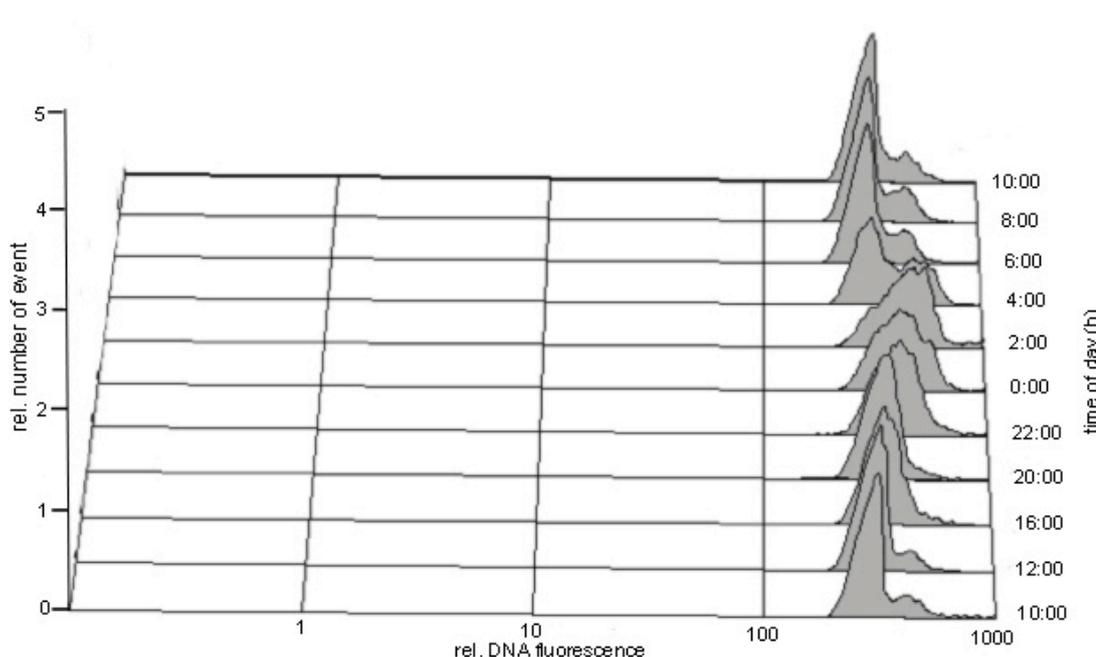
Figure 2-3: Relative DNA content in *Emiliania huxleyi* over 24-hour, showing relative DNA fluorescence histograms of *E. huxleyi* PLY# M217 at sampling times over the duration of the cell cycle experiment.

The percentages of cells in cell cycle phase were derived from the data presented in the histograms (Figure 2-3) and are presented in Figure 2-4D. A doubling of the relative DNA content of the G1-phase concluded the separation of S-Phase from the G1-phase. The G2+M phase could not be clearly separated from the S-phase, because of intermediate character and short duration of the G2+M cell phase. The 1-n cells in the early G1-phase were more apparent

allowing for a better separation of the G2+M phase from the early G1-phase. Therefore, a combined S+G2+M cell phase was presented in Figure 2-4D. During the light period, the greatest proportion of cells (> 85 %) was in the G1-phase. Relative DNA content increased from 10:00 pm to 2:00 am 20th April 2011 when the greatest proportion of cells was observed in the S+G2+M-phase. By 6:00 am the relative DNA content in the majority of had dropped to G1-phase values. Over the course of the 24-hour experiment cell densities increased more than 2 fold (in detail 2.8, 2.3, and 2.1 fold in the stock cultures PLY217-1, PLY217-2, and PLY217-3, respectively; see Figure 2-4A). The most dramatic increase in cell densities occurred between 0:00 am and 4:00 am in the Dark period. The cell population continued to grow at approximately half as fast between 4:00 am and 6:00 am. This was also reflected by the highest calculated growth rates per hour at 2:00 am in the stock cultures PLY217-1, PLY217-2, and PLY217-3 with 0.15, 0.15, and 0.14 μ h^{-1}, respectively (see Figure 2-4B). The cell volume was greatest between 4:00 pm 19th April 2011 and 2:00 pm 20th April 2011 into the dark period (see Figure 2-4C). The average cell volume in this time frame was 94 μm^3. The average cell volume had decreased to around 50 μm^3 per cell at the end of the 24 hours of measurements.
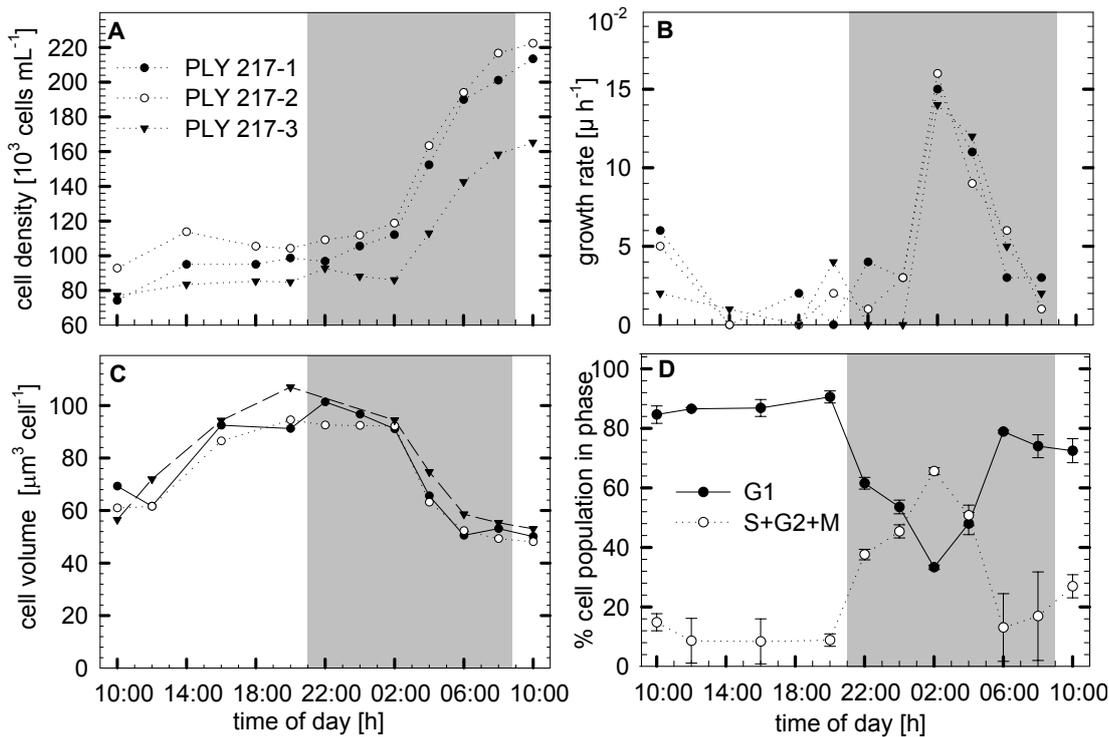
Figure 2-4: Dynamics of cell density, growth rates, cell volume, and cell cycle phase of *E. huxleyi* over a 24-hour period.. A) Cell density; B) Growth rates [μ h^{-1}], C) Cell volume [$\mu m \cdot cell^{-1}$], D) Percentages of cells in each cell cycle phase; G1 (solid line) and S+G2+M (dotted line). For A), B), and C) the measurement in the replicate cultures PLY217-1, PLY217-2, and PLY217-3 are presented (see legend Figure A). Grey areas represent the dark incubation phase of the 12:12 hours LD cycle. Error bars represent one standard deviation of three measurements from three independent experiments.

2.3.2. Production of organic carbon, organic nitrogen, the ratio of organic carbon and nitrogen and cell volume

The production of organic carbon and organic nitrogen occurred during the light incubation period of the LD-cycle and correlated with the extension of the cell volume. The highest values for particulate organic carbon were observed at 8:00 pm with 1.33 ± 0.05 µmol cell^{-1} and lowest organic carbon contents per cell were found at 8:00 am + 1day with 0.62 ± 0.07 µmol cell^{-1}, showing less than half of the maximum value (see Figure 2-5). Organic nitrogen per cell was at maximum at 12:00 am with 0.145 ± 0.01 µmol cell^{-1} and at minimum at 8:00 am 20th April 2011 with 0.078 ± 0.01 µmol cell^{-1}. The ratio of organic carbon to organic nitrogen showed a peak at 8:00 pm with 14.2 and was smallest at 8:00 am 20th April 2011 day with 7.9 (Figure 2-5).
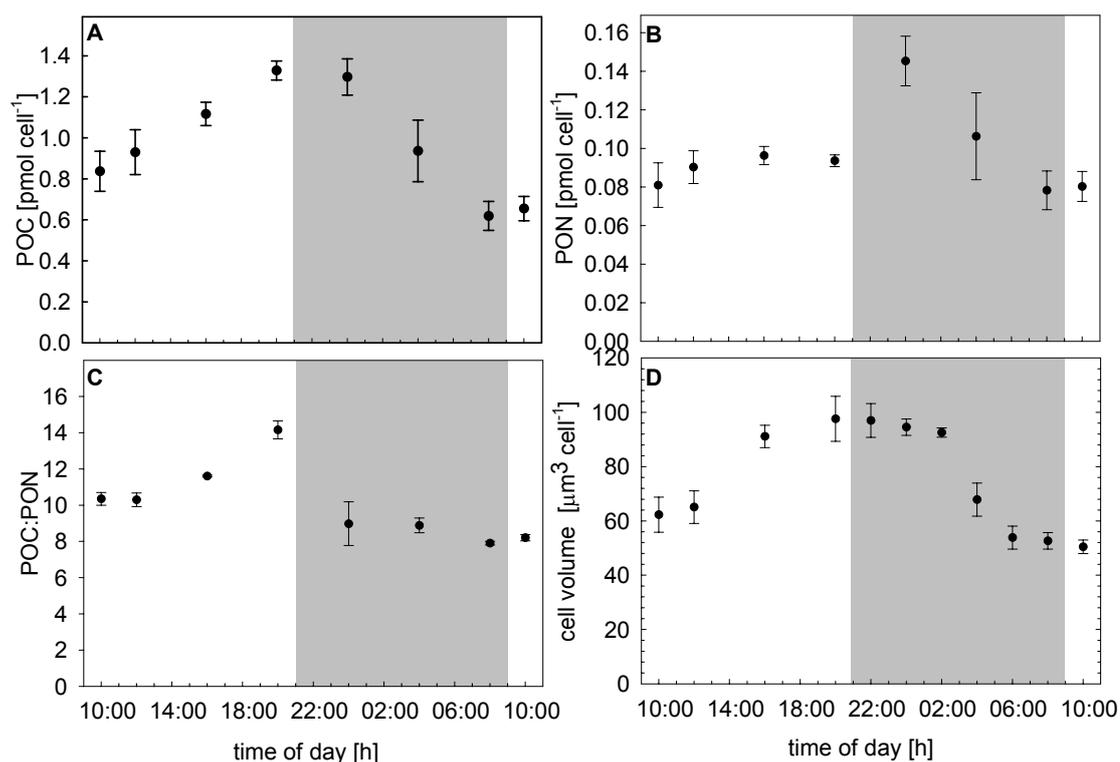
Figure 2-5: POC, PON, POC:PON dynamics of *E. huxleyi* over a 24-hour period. Averages of A) Particulate organic carbon, B) Particulate organic nitrogen, C) Ratios of particulate organic carbon to organic nitrogen, and D) Cell volume measurements of three stock cultures over a period of 24 hours are presented. Grey areas represent the dark incubation phase of the 12:12 hours light-dark incubation cycle. Error bars represent one standard deviation of three measurements from three independent experiments.

2.3.3. Production of Particulate inorganic carbon
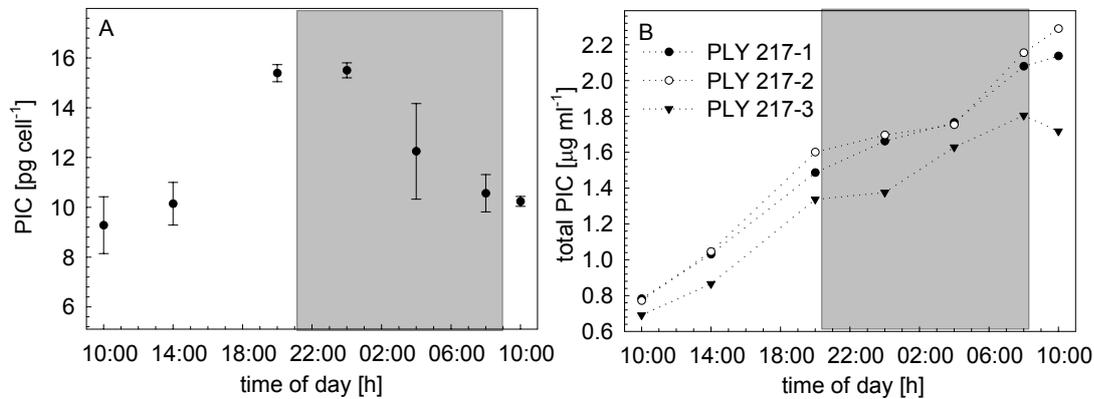
Figure 2-6: PIC dynamics of *E. huxleyi* over a 24-hour period, showing A) Averages of particulate inorganic carbon ($CaCO_3$) per cell and B) Measurement of total particulate inorganic carbon per mL of *E. huxleyi* culture in the replicate stock culture PLY217-1, PLY217-2, and PLY217-3. Grey areas represent the dark incubation phase of the 12:12 hours Light-Dark incubation cycle. Error bars represent one standard deviation of three measurements from three independent experiments.

The accumulation of PIC is equivalent to the production of calcium carbonate scales by coccolithophores. The amount of PIC per cell increased steadily during the illuminated period of the LD-cycle. The maximum of about 0.39 ± 0.008 pmol cell^{-1} was reached at 12:00 am (Figure 2-6A). After the initiation of cell division at around 2 am the PIC content per cell decreased. Between 10:00 am and 8:00 am the PIC production rate was calculated at an average of 0.0152 pmol cell^{-1} h^{-1}. Assuming that the average calcium content of a single coccolith was 6.69 fmol lith^{-1} (Fagerbakke et al., 1994) between 10:00 am and 8:00 pm coccoliths were produced at a rate of 2.24 liths h^{-1}. In the dark phase PIC per cell decreased at rate a of 0.0154 pmol cell^{-1} h^{-1}. The total PIC per ml increased from 18.7 ± 1.2 nmol mL^{-1} to 51.1 ± 7.4 nmol mL^{-1} in the course of 24 hours (see Figure 2-6B). An increase of total PIC ml^{-1} was also observed at lower rates throughout the dark period.

2.3.4. Nutrient concentrations, dissolved inorganic carbon and total alkalinity
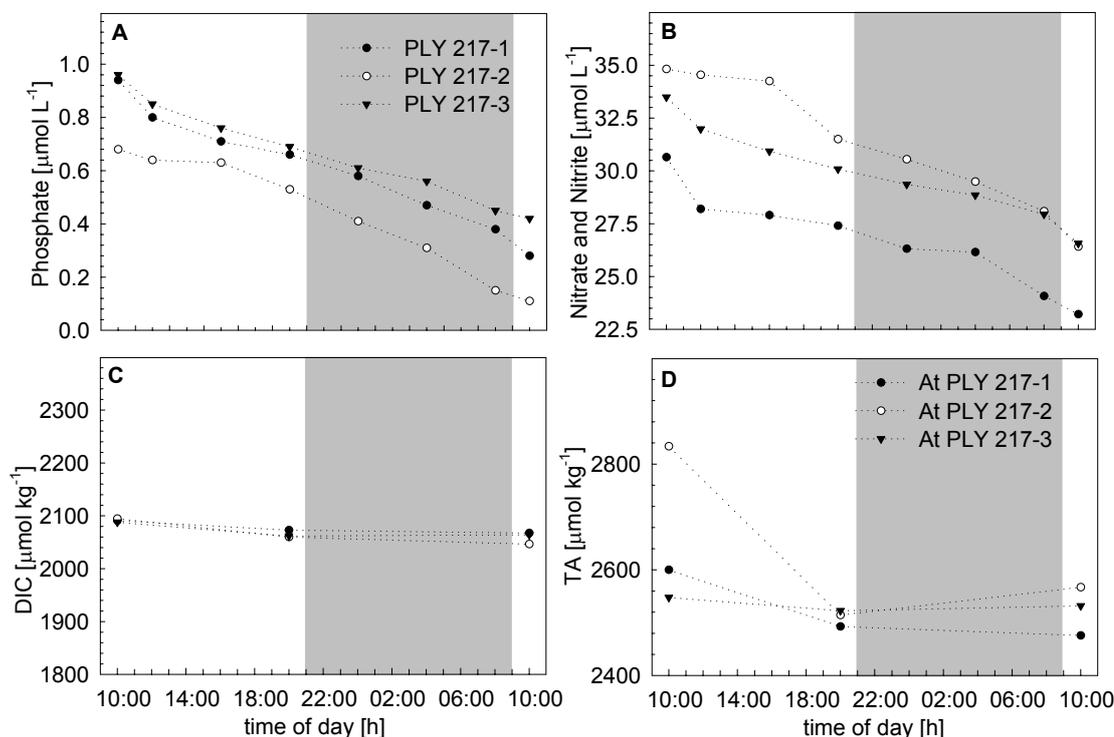
Figure 2-7: Macronutrients in *E. huxleyi* cultures over a 24-hour period. Macro nutrient concentrations and seawater carbonate chemistry parameters in the replicate stock cultures PLY217-1, PLY217-2, and PLY217-3 are presented: A) Phosphate and B) Nitrate and nitrite over a period of 24 hours. The seawater carbonate chemistry parameters dissolved inorganic carbon C) and total alkalinity D) at 10:00, 20:00 and 10:00 the following morning are presented. Grey areas represent the dark incubation phase of the 12:12 hours Light-Dark incubation cycle.

E. huxleyi cells did not exhaust the macro nutrients PO_4 and NO_3 in any of the replicate stock cultures and did not alter total alkalinity and dissolved inorganic carbon in seawater. The average nutrient consumption per cell per h in both the light and dark incubation phase is given in Figure 2-7. The consumption of nitrogen was significantly higher in the light with 14.0 ± 5.20 10^{-9} mmol cell^{-1} h^{-1} compared to 3.41 ± 1.91 10^{-9} mmol cell^{-1} h^{-1} in the dark (p < 0.001; two tailed t-test). In the dark, nitrogen uptake rates were decreased by 76% in the dark and phosphorous uptake rates decreased by 39% (Figure 2-7A and B). The calculated average consumption rate of phosphate was 4.42 ± 2.94 10^{-10} mmol cell^{-1} h^{-1} in the light and 27.0 ± 2.72 10^{-11} mmol cell^{-1} h^{-1} in the dark. The corresponding ratios of nitrogen and phosphorous uptake Light:Dark ratios were N:P $_{light}$ 33.6 ± 9.48 and N:P $_{dark}$ = 9.60 ± 1.17 (mean ± s.d., n=3). The concentrations of dissolved inorganic carbon concentrations per kilogram seawater (SW) showed an average of 26 µmol kg^{-1} SW during the day, while at night DIC decreased only by 6 µmol kg^{-1} SW. Total alkalinity decreased by 150 µmol kg^{-1} SW during the day and increased by 15 µmol kg^{-1} SW during night time (see Figure 2-7C and D).

2.3.5. Chlorophyll contents and photosynthetic health over a 24-hour period
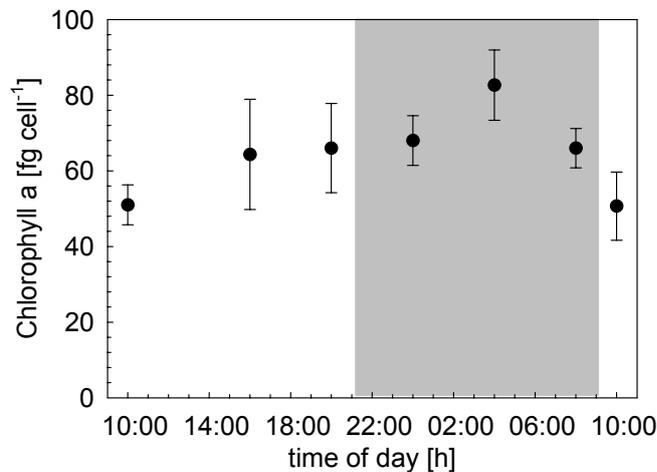
Figure 2-8: *Chl a* dynamics of *E. huxleyi* over a 24-hour period. Average of chlorophyll *a* content in *E. huxleyi* cells in fg cell^{-1} over the period of 24 hours. Grey-shaded area illustrates the dark incubation period.

To assess the photo-physiological health of the *E. huxleyi* cells in the stock cultures over the duration of the cell cycle experiment the chlorophyll *a* content per cell and the photosynthetic efficiency were assessed. Chlorophyll *a* showed steady increase from 10:00 am to 4:00 am the next morning. After 4:00 am the chlorophyll content per cell decreased (Figure 2-8). The chlorophyll *a* contents per cell were smallest at the begin of the light period at 10:00 am with 51 ± 6 fg Chl *a* cell^{-1} (Figure 2-8). The concentration then ranged between 65 ± 15 and 82 ± 9 (maximum at 4:00 am + 1 day) fg Chl *a* cell^{-1}. By 10:00 am + 1 day Chl *a* contents had reached the values from the previous day, measuring 51 ± 9 fg Chl *a* cell^{-1}.
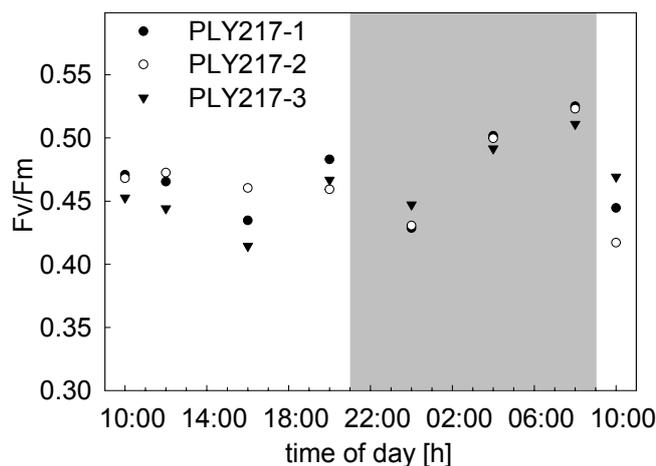
Figure 2-9: Photosynthetic efficiency dynamics of *E. huxleyi* over a 24-hour period. Photosynthetic efficiency (Fv/Fm) of *E. huxleyi* cells in the replicate stock cultures PLY217-1, PLY217-2, and PLY217-3 over a period of 24 hours. Grey-shaded area indicates the dark incubation phase.

The quantum yield of photosynthesis in the PSII was similar in all stock cultures (PLY217-1, PLY217-2, and PLY217-3; Figure 2-9). The Fv/Fm values in the light incubation period ranged between 0.436 ± 0.022 and 0.471 ± 0.01. Within the dark incubation period Fv/Fm values increased reaching the maximum of 0.52 ± 0.007 at 8:00 am the next morning. With the onset of the new light incubation period at 10:00 am 20th April 2011 the quantum yield in PSII dropped to 0.444 ± 0.026.

2.3.6. Rates of photosynthesis and calcification from ^{14}C-labelled $NaHCO_3$ incubation over a 24-hour period

To determine the rates of photosynthesis and calcification *E. huxleyi* were exposed to ^{14}C-labelled bicarbonate source. The incorporation of ^{14}C into particulate organic and particulate inorganic carbon was assessed and the rate of photosynthesis and calcification rates derived. Calcification and photosynthesis were both enhanced in the light incubation period. While photosynthesis came to a complete stop in the dark incubation period, calcification continued at strongly reduced rates. The ^{14}C labelled bicarbonate incubations suggest that calcification rates (in pmol $CaCO_3$ cell[-1] h[-1]) in *E. huxleyi* were at their maximum of 0.042 ± 0.002 pmol $CaCO_3$ cell[-1] h[-1] at 6:00 pm. Dark calcification was detected throughout the dark incubation period, but at rates of only 0.005 to 0.008 pmol $CaCO_3$ cell[-1] h[-1] (Figure 2-10A). The calculated maximum coccolith production rate was 1.56 liths h[-1] at 6 pm (Figure 2-10B). The highest rates of photosynthetic carbon assimilation were observed at 4:00 pm with 0.048 ± 0.005 pmol $CaCO_3$ cell[-1] h[-1], while photosynthesis came to a complete halt in the dark incubation period (Figure 4-10C). In the light incubation phase the ratio of calcification: photosynthesis (CF:PS) increased

gradually from 0.5 at 10 am to 1.1 at 6:00 pm (Figure 2-10D). The CF:PS - value of 1.4 at 8:00

pm is omitted from the evaluation because of the 1 h incubation light (see Section 2.2.2). In the

dark incubation period (CF:PS) increased to values of up to 22.7 due to low rates of

photosynthesis. Between 10:00 am and 8:00 pm the average CF: PS ratio was 0.9 ± 0.23.
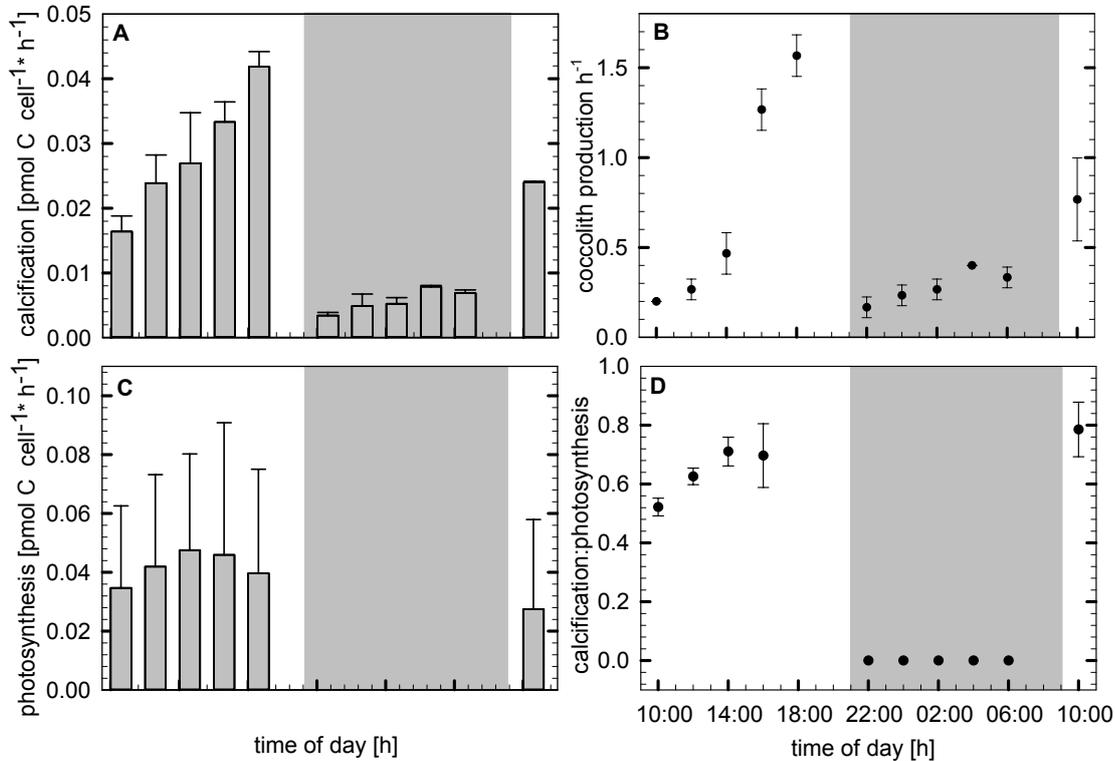
Figure 2-10: Rates of photosynthesis and calcification of *E. huxleyi* over a 24-hour period.
Photosynthesis and calcification measurements based on the incorporation of ^{14}C-labelled
bicarbonate. Incubations that started at 8:00 pm and 8:00 am were excluded from the presentation,
because of only receiving one hour of light exposure. Rates for photosynthesis were set to zero in the
dark and calcification: photosynthesis ratios not calculated for the dark incubation period. A)
Calcification rates per cell; B) coccoliths production per cell and h; C) photosynthesis rates per cell;
and D) calcification: photosynthesis. Error bars indicate one standard deviation of three independent
replicates. The dark incubation phase is shaded grey.

2.4. Discussion

Within the focus of this study, the DNA content and physiological parameters of *E. huxleyi* PLY♯M217 were followed over a period of 24 hours to describe cell cycle phases and periods of maximum and minimum calcification rates. The relative DNA content and the assessed physiological parameters namely maximum rates of calcification and photosynthesis, chlorophyll *a* content, particulate organic carbon and particulate organic nitrogen per cell, and cell volume, followed a 24-hour periodicity as previously described (e.g. Bleijswijk et al., 1994; Bleijswijk and Veldius, 1995; Paasche, 2002; Müller et al., 2008). Photosynthesis was linked to the diurnal rhythm of light and dark. The maximum rates of calcification occurred during the G1-phase of the cell cycle at the end of the light incubation phase between 6:00 and 8:00 pm when photosynthetic activity was also high. In the combined S+G2+M phase (dark incubation) calcification was observed only at rates of 12% to 19% of those measured during the late G1-phase. The following paragraphs discuss the assessed physiological parameters of *E. huxleyi* PLY♯M217 in more detail. Furthermore, the adequate cell harvesting times for investigating the transcriptome and proteome at rates of high and low calcification is concluded from the relative DNA content per cell, the calcification rates over 24-hours, and knowledge of cell cycle physiological characteristics.

Environmental factors such as elevated temperature, prolonged illumination and nutrient replete conditions are known to encourage cell division in natural coccolithophore populations and cultures (Paasche and Brubak, 1996; Paasche, 1967; Fritz and Balch, 1996; Price et al., 1998; Müller et al., 2008; Poulton et al., 2010). The observed growth rate of 0.9, which is an equivalent to 1.3 divisions per cell per day, suggests that some cell cohorts could have divided multiple times within the 24 hours of sampling. Growth rates above 0.69 (equivalent to one division per day) were also reported by Mullin et al. (1966), Paasche (1967), Balch et al. (1996), Eker-Develi et al. (2006). Cell size, temperature, light and the proposed internal clock have been reported as commonly active triggers for cell division in algae (Donnan et al., 1985; Lam et al., 2001). Observations in mamallian cell lines showed that larger cells pass into the S-phase more quickly than smaller cells (Ronning and Lindmo, 1983), with larger cells showing higher rates of DNA synthesis (Ronning and Seglen, 1982). The average cell volume presented in Figure 2-4C indicates that cell volume decreased over time but remained higher at the end of the 24 hours period compared to the start of the 24-hour monitoring period. Histograms of cell size in the cultures at the end of the 24-hour monitoring period also showed cohorts of the cell with larger cells after cell devisions had occurred (data not presented) compared to the beginning of

the monitoring period. These observations could support the assumption that a proportion of the cell population were dividing multiple times. Further support of asynchronous and multiple fissions per day is given by the DNA content of the cells. The histograms of relative DNA content and cells in cell cycle phases (Figure 2-3 and Figure 2-4D) show that cells with 2n DNA were still present in the morning of April 20th 2011 and that a proportion of the cell population remained or returned into the S+G2+M -phase. This suggests that growth rates of 0.9 were achieved by cohorts of cells passing the S+G2+M -phase multiple times (Chisholm et al. 1984). Furthermore, asynchronous and multiple fission of cell cohorts in the culture justify the conservative distinction between the G1 and S+G2+M phase because of cells with overlapping cell cycle phases S, G2, and M in the cell population. Müller et al. (2008) and Bleijswijk and Veldhuis (1995) investigated the cell cylce features of *E. huxleyi* CCMP 371 and also found higher proportions of the cell population in the S-phase rather than the G2/M-phase.

Algal culture growth, kinetics of macronutrient uptake, cellular concentrations of chlorophyll *a*, photosynthetic quantum yield (Fv/Fm), cellular POC, and cellular PON have been suggested as measures for algae physiology impairment through nutrient limitation or other stressors (Kobler et al., 1983; Geider et al., 1993; Beardall et al., 2001; Moore et al., 2006). Even though natural *E. huxleyi* populations thrive well under both high and low nutrient regimes (Zondervan, 2007) *E. huxleyi* lab cultures limited in phosphorous showed reduced growth rates (Paasche, 1998; Riegman et al., 2000; Paasche 2002), increased cellular organic carbon content (Paasche and Brubak, 1994; Paasche, 1998; Riegman et al., 2000; Shiraiwa et al., 2003), and an increase in number of liths per cell (Paasche, 1998). Nitrogen limitation in lab cultures has been shown to result in a reduction of cell size and putatively cellular organic carbon content and a small increase in calcite per cell (Riegman et al., 2000; Sciandra et al., 2003, Zondervan, 2007; Mueller et al., 2008). Therefore, it is not suggested by the observed growth rates, organic carbon content per cell, and inorganic carbon per cell that macronutrients were limiting for growth (see Section 2.3). The measurements of PIC per cell (Figure 2-6A), and particulate organic carbon (Figure 2 – 5A) were not significantly different between the start and end of the 24-hour sampling period ($p < 0.05$, two-tailed t-test, n = 6; data not shown). However, cell volume was significantly lower at the end of the 24-hour sampling period ($p = 0.04$, two-tailed t-test, n = 6; data not shown). Nevertheless, the slight significant decrease of cell volume within 24 hours does not suggest that the cells were experiencing nutrient limitation.

A sufficient supply of the macronutrients nitrogen and phosphorous is central to cell functioning, physiology, and cell division. The nutrient concentrations in this cell cylce

experiment were significantly reduced during the course of the 24-hour monitoring period. The total uptake rates of nutrients in the light and dark incubation period were different with respect to previously reported uptake rates (see Müller et al. 2008). Nitrogen uptake rates in the light were around 5 times lower and nitrogen uptake in the dark was around 10 times lower in this experiment compared to Müller et al. (2008). The phosphorous uptake rates in the light and dark incubation period in this study were both 10 times lower compared to Müller et al. (2008; Table 1). The observed differences in nutrient uptake rates might be explained by the different genetic repertoire of the strains CCMP 371 and PLY♯ M217. Despite the observaion of different uptake rates of phosphorous is was confirmed that phospohorous uptake in the light and dark are not significantly different (Müller et al., 2008). The uptake ratio of N:P in the light versus the dark was significantly reduced by about 70% (p = 0.012, two-tailed t-test, n = 6; data not shown) (compare Riegman et al., 2000; and Müller et al., 2008). Ronning and Lindmo, 1998 showed that the demand for nitrogen is lower in the dark because protein synthesis is suppressed in the dark, in particular when cells pass through the S and G2/M-phase. In this study the consumption of nitrogen was also significantly reduced in the dark (p < 0.001, two-tailed t-test, n = 15; data not shown). Therfore it can be inferred that the higher nitrogen uptake rates in the light also reflected elevated protein synthesis and biomass accumulation (compare Ronning and Lindmo, 1998; Müller et al., 2008).

The chlorophyll *a* content per cell was found to be lower or within the range of previously reported values (e.g. Muggli and Harrison, 1996, 1997; 49 to 130 fg Chl *a* cell^{-1}; Eker-Develi et al., 2006; 200 fg Chl *a* cell^{-1}; Houdan et al., 2005; 117 fg Chl *a* cell^{-1}; Suggett et al., 2007; 97 – 132 fg Chl *a* cell^{-1}; Loebl et al., 2010; 140 fg Chl *a* cell^{-1}). The lower values found in this study could reflect a strain specific feature. From the lower Chl *a* content alone an impairment of *E. huxleyi* physiology can not be inferred (Kruskopf and Flynn, 2005).

Measurements of photosynthetic health (Fv/Fm) can rapidly detect nutrient stress in phytoplankton cultures that are not adapted to nutrient limitation (Parkhill et al., 2001) by inducing only little stress on the cells (Falkowski and Kolber, 1995). Phytoplankton cells require nitrogen to synthesize proteins for PSII repair, because light frequently damages the PSII reaction centers (Nishiyama et al., 2006; Ragni et al., 2008). The observed photosythetic efficiency (Fv/Fm) – measurements do not suggest an impairment of the photosystem II in *E. huxleyi* cells and therefore nutrient starvation. Under nutrient starvation Fv/Fm was found to decrease (Sugett et al., 2009). However, Ho et al. (2003) and Suggett et al. (2007) reported generally higher values for Fv/Fm ranging between 0.557 and 0.58 for the *E. huxleyi* strains

B11, B92 and the field strain ASM 1, compared to the average range of Fv/Fm of 0.435 and 0.520 found in *E. huxleyi* strain PLY♯M217. This difference is likely to represent a 'taxonomic signature' or difference in photophysiological performance between the strains (Suggett et al., 2009). Interestingly, a peak value of photosythetic efficiency was observed at the end of the dark incubation period (Figure 2-9) when most cells of the population were about 3 hours into the G1-phase. This peak in photosythetic efficiency can be explained by the 'young cells' and 'young photosystems' showing increased values of Fv/Fm (Post et al., 1985). Furthermore, a similar pattern of lower Fv/Fm values at the onset of the G2/M was shown in diatoms (Claquin et al., 2004).

Nutrient limitation affects growth and carbon accumulation in *E. huxleyi* (reviewed by Paasche, 2002). Any significant changes of the cellular organic carbon, cellular inorganic, and organic nitrogen content were however not observed between the beginning and end of the 24-hour experiment ($p < 0.05$ for POC, PIC and PON, two-tailed t-test, n = 6; data not shown), that would indicate a limitation of P and N supply to the *E. huxleyi* cells. P limited cells in *E. huxleyi* cultures are known to show an increase of organic carbon content per cell (e.g. Paasche and Brubak, 1994; Riegman et al., 2000). Indicative of N limitation are a decrease in POC and in cell volume (Riegman et al., 2000; Sciandra et al., 2003; Müller et al., 2008) and a moderate increase or even a decrease in PIC (Paasche, 1998; Fritz, 1999). However, it should be noted that Müller et al. (2008) found higher contents of POC and PON in the *E. huxleyi* strain CCMP 371 under nutrient replete conditions and observed similar concentrations of organic nitrogen only in N-limited cultures.

One emphasis of this experiment was to measure the rates of photosynthesis and calcification over a 24-hour period. The data presented here support previous conclusions that rates of calcification and photosynthesis in *E. huxleyi* follow a light dark cycle (Müller et al. 2008; Paasche 1962). Calcification rates were found to be dependent on the irradiance levels (Paasche, 1964; Paasche, 1965; Linschooten et al., 1991; Nimer and Merrett, 1993; Bleijswijk 1994; Paasche, 1999; Zondervan et al., 2002) and our results supported the coupling of calcification on photosynthesis, suggesting that energy deriving from photosynthesis accelerates the precipitation of $CaCO_3$ (Anning et al., 1996). Dark calcification (Figure 2-10A) was observed as was previously reported (Paasche, 1966; Balch et al., 1992; Nimer an Merrett, 1992; Paasche and Brubak, 1994; Sekino and Shiraiwa, 1994, 1996). Rates as low as 10 to 15 % of the calcification rate in the light were reported for the dark (Paasche, 2002), which compare to rates of between 12 and 19 % of the maximum calcification rate in the light incubation period in this study. Mueller et al. (2008) however, reported an absence of dark calcification.

The energy required to drive dark calcification was suggested to derive from mitochondrial respiration (Sekino and Shiraiwa, 1996). In this study a lag-phase was observed before the maximum rates of calcification, rates of coccolith production, and photosynthesis were reached. The inorganic carbon assimilation increased over a period of 9 hours until 6:00 pm when maximum rates were reached. This might indicate the tight energetic coupling of photosynthesis and calcification (Anning et al, 1996) or that the degree at which biochemical energy from photosynthetic carbon fixation supplied for calcification was lower in the early hours of the light incubation phase because other cellular processes had higher energetic demands. It is known that the assimilated carbon is incorporated at different rates in molecular components such as proteins, polysaccharides, lipids, and low molecular weight molecules throughout the cell cycle (Fernandez et al., 1996). Therefore, it can be assumed that theses quotas change over time and under different physiological demands of other cell processes.

A comparison with data from the available literature showed that calcification and photosynthetic rates were within the range of previous results (see Table 2-1). The results were closest to the studies by Paasche and Brubak (1994) and Balch et al. (1996). Those two studies were also found to reflect rates similar to those found in natural subsurface calcifying populations at the Iceland Basin (Poulton et al., 2010). However, notable differences between the results presented here and those obtained by Mueller et al. (2008) were found. Nevertheless, the different kinetics of inorganic and organic carbon assimilation did not affect the characteristics of the *E. huxleyi* cell cycle.

Table 2-1: Experimental conditions and observed growth rates in previous studies and this study. Summary of experimental conditions and observed growth rates in previous studies using strains of *Emiliania huxleyi,* which investigated growth, calcification, photosynthesis properties applying ^{14}C-labelling. CF: calcification, PS: photosynthesis. Values for CF and PS are given in fmol cell^{-1} h^{-1}. Abbreviations: av = average and pk = peak value (maximum).

E. huxleyi Strain	Growth rate [m]	Medium/ condition	CF	PS	CF : PS	Reference
E88	n.a	F/2, 50 mE m^{-2}s^{-1}	1.8	2.2	0.9	Nimer and Merrett (1992)
5/90/25j	1.26	P=0.6 mM, N= 125 mM, 20°C	~23.4 -32.4	~26-36	~0.82-0.9	Paasche and Brubak (1994)
E88	0.99	K, 17±0.5°C, cont.75 mE m^{-2}s^{-1} continuously	18	42	0.43	Balch et al. (1996)
CCMP371	0.72	F\50. 21°C	8	~18.5	0.44	Mueller et al. (2008 Fig.2)
PLY ♯ M217	0.9	F\50. 19°C 94 mE m^{-2} s^{-}	av=29 pk=42	av=40 pk=48	av=0.75 pk=0.88	This study

It can be concluded that in this study, *E. huxleyi* cells were not exposed to nutrient limitations and that their growth was well mantained in the culture conditions. The results of the ^{14}C-bicarbonate incubation, the determination of calcification rates and rates of photosynthesis over 24 hours, as well as the analysis of relative DNA content per cell, of three replicate cultures of *E. huxleyi* PLY♯M217 revealed a 24-hour pattern similar to previously reported light dark periodicity (Müller et al. 2008; Paasche 1962). The bulk of calcification could be correlated with the late G1-phase of the cell cycle. Due to the asynchronous cell cycle progression and suggested multiple division of the cells in the dark it was uncertain to clearly seaprate between the S and G2+M cell cycle phases. Therefore, only two cell cylce phases, namely the G1 and S+G2+M phase, were separated. The observation that DNA replication and nuclear and cellular division may overlap in non-G1 phases (Zachleder et al., 2016) supports the functional separation in two cell cycle phases in this study in respect to subsequent molecular studies. In the S+G2+M phase demands of nitrogen and showed lower deman for nitrogen in the dark incubation period because protein synthesis is suppressed in the dark, when cells pass through the S +G2+M-phase (Ronning and Lindmo, 1998). However, Ferandez et al. (1996) suggested that the overall carbon incorporation into proteins is much higher in the dark than in the light. Prevailing cell cycle processes of microalgae in the G1-phase suggest elevated contents of RNA and proteins in the late G1-phase. In the early G1-phase, still in the dark period, cell division was completed and RNA and proteins related to cell division processes, which could provide noise in the transcriptome and preoteome analysis, were thought to be at low levels (Zachleder and Šetlík, 1988; Zachleder et al., 2016). The results of cell population growth and DNA content suggested that the cells were in the G1-phase at both harvesting time. Therefore, the time windows for harvesting cells at low and high calcification rates for transcriptome and proteome analysis were chosen between 7:00 pm in the light incubation period, when maximum calcification rates were observed and the bulk of the cells were in the late G1 cell-cycle phase. Hereafter, the samples of the late G1-phase are referred to as (high calcification – late G1-phase (hc-lG1 or just HC) samples. The harvesting time at low calcification rates was chosen at 7:00 am, the end of the dark incubation period, when cells showed low calcification rates, the bulk of the cells had completed cell division, and cells had transited into the early G1-phase. The harvesting time at 7:00 am is here after referred to as low calcification – early G1-phase (lc-eG1 or just LC).

Chapter 3.   The transcriptome of *Emiliania huxleyi* at high and low calcification rates

"Perhaps all forms of life on this planet use essentially the same genetic language."
(Nirenberg, 1968)

3.1. Introduction

In all organisms genetic information is stored in DNA that is replicated during cell division or transcribed to form proteins after gene expression. Gene expression and protein synthesis comprises two steps: (1) transcription of DNA to RNA and (2) translation of RNA to proteins. During transcription, RNA transcripts complementary to one strand of DNA are synthesised. Throughout translation, the information in the RNA sequence is converted into sequences of amino acids and polypeptides. This flow of genetic information for the purpose of DNA replication or the transcription into RNA sequences and the translation into proteins is also referred to as the central dogma of molecular biology. With the exception of RNA viruses, where the information that is stored in RNA sequences can be reverse transcribed into DNA, the direction of the flow of genetic information is from DNA to RNA to proteins (Snustad and Simmons, 2000).

There are many types of RNA, such as mRNA (messenger RNA), tRNA (transfer RNA), rRNA (ribosomal RNA), snRNA (small nuclear RNA), and small interfering RNA (siRNA) (Kanno et al., 2005; Onodera et al., 2005). The information to synthesize polypeptides is stored in the mRNA sequence. In eukaryotes pre-mRNA is an immature mRNA precursor that is edited by affiliating additional 7-methyl guanosine caps or poly (A) (a sequence of many adenosine nucleotides) to the transcript or splicing noncoding intron sequences from the transcript. In eukaryotes, most of the maturation of mRNA occurs in the nucleus, whereas polypeptide synthesis occurs in the ribosomes (macromolecular protein biosynthesis units comprised of both protein and rRNA) of the cytoplasm. Ribosomal RNA is essential for the growth of the polypeptide sequences in the ribosome, because rRNA catalyses the chemical bond between two amino acids. Furthermore, rRNAs were found to evolve and are used for the identification of taxonomic groups by separating rRNAs according to their rate of sedimentation during sucrose gradient centrifugation. Transfer RNAs adapt specifically to amino acids and the codon in the mRNA. The tRNA decodes the triplets of RNA bases (codon) into the according amino acid. The following sections focus on the expression of protein coding genes and discuss mRNA or transcripts abundance at high and low calcification rates in the coccolithophore *E. huxleyi*.

3.1.1.Molecular controls of gene expression at the transcript level

Gene expression describes the process in which a protein is produced from the template of genomic gene sequence. A regulatory macromolecular machinery selectively controls gene expression in response to environmental factors, developmental stage or cell type; thus making the processes contributing to gene expression highly dynamic. The regulation of gene expression involves the interaction of activators and repressors that operate at different and multiple layers of the transcription process and protein synthesis machinery. Activating or repressing transcription factors (Eberhardy and Farnham, 2002), the coiling of the chromosomes through nucleosome modification (Lee and Young, 2000), initiation cofactors for the RNA II polymerase (Pol II) (Jang et al., 2005), the elongation (Aso et al., 1995) or pausing of mRNA synthesis (Muse et al., 2007), and mRNA processing (Keller and Noon, 1984) are some of the levels at which the expression of protein coding genes is regulated within the cell. Multiple genes may be controlled by activating or repressing gene-regulatory processes (Lee and Young, 2000; Wijnen and Young, 2006). These regulatory processes might comprise just one of many interacting activators or even a combination of activators and repressors (Levine and Tijan, 2003). Transcript synthesis in eukaryotes starts by binding activators of gene expression to upstream activating sequences. A specific DNA-binding domain within the transcription factor protein binds to the promoter DNA sequence of the gene. Transcription factors direct the specific binding of the RNA polymerases I, II, and III to the DNA sequence and initiates the transcripts synthesis (Snustad and Simmons, 2000; Lee and Young, 2000). The transcription factors may act as an activator by attracting and stimulating the coupling of RNA polymerase to the transcription start site gene sequences or as a repressors by preventing the binding of RNA polymerase or other required proteins to the gene sequence (Lee and Young, 2000; Snustad and Simmons, 2000). In *Emiliania huxleyi* a total of 419 different transcription factors were identified, which contribute around 1% of its proteome (Rayko et al., 2010).

The degree of coiling of the DNA double helix not only establishes the condensation of chromosomes but also exposes DNA sequences that are accessible to transcription. Histones coil the DNA to form nucleosomes, which are the basic building blocks of the chromatin structure. In supercoiled DNA, as in chromosomes, the transcription machinery is unlikely to recruit to a promoter and start the transcription process. Processes that can unfold the chromatin structure for transcription are modifications of the histone complexes, such as methylation as well as acetylation and deacetylation (Margueron et al., 2005).

In eukaryotes an array of transcription factors, such as TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, and TATA-box binding proteins, complement the RNA polymerase II holo-enzyme to create the pre-initiation complex that defines the transcription start site (TSS) (Bengal et al., 1991Chiang et al., 1993; Tan et al., 1994; Näär et al., 2001; Kettenberger et al., 2004; Thomas and Chiang, 2006; Kwak and Lis, 2012). In regulated transcription additional general cofactors are often required to mediate between the gene-specific activators and the general transcription machinery (Dynlacht et al., 1991; Thomas et al., 2006). Both, stimulating and repressing co-factors have been discovered that are important in binding the promoter to the specific transcription factors regulate basal transcription (reviewed by Thomas et al., 2006).

Nuclear pre-mRNA is subject to post-transcriptional modifications. Once the modifications are conducted the mRNA is ready for translation into the amino acid sequence. Post-transcriptional modifications are the excision of non-coding introns from the transcript (Keller and Noon, 1984), the addition of 7‒methyl guanosine cap to the 5' end of the transcript (Shatkin, 1976), and polyadenylation to the 3' end of the transcript. The capping of 7-methyl guanosine caps occurs shortly after the initiation of transcription adding protection to the growing RNA strand from degradation by nucleases. The addition of poly(A) tails occurs through endonucleolytic cleavage in response of a downstream polyadenylation signal the AAUAAA sequence (Snustad and Simmons, 2000). Non-coding introns are excised on complex ribonucleoprotein structures called spliceosomes. In spliceosomes previously assembled snRNA – protein complexes and protein splicing factors cut out non-coding sequences and ligate the adjacent exons. In some cases the multiple introns are removed by splicing the pre-mRNA. If two successive introns are excised, the exon between them will be cut out. Alternative splicing, exon removal, and transcript isoforms create different polypeptides (Houk et al., 1991; Snustad and Simmons, 2000).

Modulating the transcriptional outcome through RNA polymerase II (Pol II) pausing can also control the process of transcript elongation. Hence, influencing the number of transcribed sequences directly (Kwak and Lis, 2013). The control of transcript elongation occurs at early elongation and productive elongation stages, adding a further level of gene expression control. Early transcript elongation pausing in the promoter-proximal regions is required to add the 5' caps (Rasmussen and Lis, 1993) and decrease the response time of the transcription machinery to changes in physiological demands as during heat shock (Andrulis et al., 2000). Throughout later periods of the transcription process Pol II passes through numerous obstacles, including

intrinsic pause sites and nucleosomes, at which point elongation factors are required to promote the continuation of transcript elongation (Young and Lee, 2000). For elongation pausing, passing elongation pausing or enhancements of the elongation process an array of nucleoproteins are utilized (Kwak and Lis, 2013).

3.1.2. Physiological feedback mechanism in gene expression

The patterns of gene expression in a cell are a direct response to the cell cycle stage, cell development, and environmental conditions (Snustad and Simmons, 2000). The changes in transcription may be induced, for example, in response to light or temperature (Ritossa 1962; Ballantuono et al., 2012) or internal triggers, such as hormones in multicellular eukaryotes or the intrinsic circadian clock (Johnson et al., 1993; Snustad and Simmons, 2000; Farré 2012). An induction of gene expression may involve a receptor, usually a receptor-protein that experiences an alteration in its three dimensional structure in response to changes in the physical or chemical environment of the cell (Urizar et al., 2000; Snustad and Simmons, 2000; Yamamoto, 2005). The receptor-protein may undergo redox reactions or phosphorylation in the case of DNA-binding proteins. These two examples of chemical reactions often occur at key molecules when gene expression is induced or paused.

Heat or cold, for example, are an important inducer of heat shock proteins, which occur in all organisms and assist in protein refolding under stress conditions (Feder and Hofmann, 1999; Wang et al., 2004). However, other environmental stressors may also induce the production of heat shock proteins, such as drought, oxidative stress, salinity, UV-radiation, or nutrient depletion (Vierling 1991; Rizhsky et al., 2002; Pockley, 2003; Wang et al., 2004). Heat shock proteins were also found to play an important role in cell signalling and immune response (Pockley, 2003). In *Drosophila melanogaster* the promoter of the heat shock protein genes *hsp70* is induced by the heat shock transcription factor (Hsf). Under ambient temperature Hsf is present in the nuclei as a non-promoter binding form. Upon heat stress Hsf is phosphorylated and binds to the DNA sequence upstream *hsp70*, at the heat shock response element section of the promoter (Snustad and Simmons, 2000).

Light acts as an important environmental factor inducing physiological responses and gene expression in many organisms (Wijnen and Young, 2006). In photosynthetic organisms light levels influence for example the state of the light-harvesting complexes and photomorphogenesis – the light mediated development in plants (Rochaix, 2014; Wu, 2014). For example, light was found to induce transcription in the nucleus of *Arabidopsis thaliana* via triggering the phytochrome RED/FAR RED (R/FR)-light photoreceptor, the UVB

photoreceptors and cytochrome directly (Chen, 2008; Fankhauser and Chen, 2010). The photoreceptor triggering may spark further diverse processes of gene expression regulation, such as photophysiological processes optimizing photosynthetic efficiency by readjusting the absorption cross sections of PSII and PSI, whereas the mobile LHCII (Light harvesting Complex of photosystem II) antenna is being redistributed. The redistribution of LHCII may occur as a short-term response or through transcriptional and translational regulation of LHC gene expression for prolonged modifications (Pfannschmidt et al., 2001; Rochaix, 2014). Different light qualities and harmful UV light were found to trigger differential gene expression of up to one third of genes in seedlings of *Arabidopsis thaliana* (Ma et al., 2001). Structural modifications of the photosynthetic complexes and within the thylakoid membrane of the LHCII (PSII) can induce protection mechanisms to avoid damage to the photosynthetic apparatus by high intensities of light and harmful spectral characteristics of light (ultraviolet light) (Rochaix, 2014). Light affects gene expression at the transcriptional, posttranscriptional, and translational level in plants. Hundreds to thousands of genes are differently expressed following changes in the light regime (Wu, 2014; Pfannschmidt et al., 2001; Wijnen and Young, 2006). In coccolithophores the rates of calcification are increasing in the Light period (e.g. Paasche et al., 1969, Balch et al., 1993) by the higher availability of energy of carbon fixation (Bach et al., 2013). Even though the suggestion that calcification acts as a carbon concentrating mechanism was disproved (Bach et al., 2013) transcriptional responses in calcification related genes were shown to depend on the light periodicity (Richier et al., 2009).

The periodicity of the diurnal light regime and the light/dark transition were found to act as an additional "zeitgeber" for the in circadian gene expression fluctuations that activate regulatory circuits preventing, for example, DNA damage from mutagenic ultra violet light during cell division (Nikaido and Johnson, 2000; Chen and McKnight, 2007). Hence, the diurnal light level fluctuations are closely linked with the endogenous timekeeping mechanism, the circadian clock, which enables organisms to respond as they experience environmental change and prepare for coming change (McClung, 2014). Therefore, gene expression in the advanced dark phase may prepare photosynthetic cells for optimal physiological performance during the light (Green et al., 2002; Dodd et al., 2005; Winjen et al., 2006).

3.1.3.Gene expression related to biomineralisation in coccolithophores

The genetic control of coccolithophogenesis involves the expression of many proteins of yet unknown functions to provide a network of interacting structural and regulatory molecules (Young and Henriksen, 2003; Henriksen et al., 2004; de Vargas et al., 2007). Previous coccolithophore research identified molecules and genes of interest that are involved in biomineralisation by comparing calcifying and non-calcifying strains, calcifying diploid and naked (non-calcifying) haploid cells, stimulating or altering the calcification performance of coccolithophores through culturing conditions (e.g. Nguyen et al., 2005; Quinn et al., 2006; VanDassow et al., 2009, Mackinder et al., 2011; Emery et al., 2012). An outline of the molecules involved in calcium carbonate crystallization in coccolithophores is given in chapter 1 of this monograph. Here, the results from studies comparing calcifying and non-calcifying stages of coccolithophores are summarized in the context of this chapter, which investigates the expression of genes involved in the biomineralisation process in the Dark period (low calcification rates) and Light period (high rates of calcification).

Richier et al. (2009) investigated the expression of specific genes related to biomineralisation in haploid and diploid cells in the light and dark period. Key proteins, such as calmodulin, the *E. huxleyi* type of carbonic anhydrase (γ -*EhCA2*), and the calcium binding protein (GPA) were significantly higher expressed in calcifying diploid cells in the light (Richier et al., 2009). Comparative studies of gene expression related to biomineralisation in phosphorous replete and phosphorous limited cultures showed that a significant number of genes related to stress compensation and cell repair processes, such as HSP 70, HSP 80, HSP 81, HSP 82, and HSP 90, and the co-chaperonins Dna J and Dna K were also over-expressed (Wahlund et al., 2004). Quinn et al. (2006) found 127 significantly up down regulated genes when comparing phosphorous replete and phosphorous limited cultures. The majority of the identified molecules were involved in cellular metabolism, ion channels, protein transport, vesicular trafficking, and cell signaling. In phosphorous limited calcifying cells putative gamma-carbonic anhydrase was significantly higher expressed (Quinn et al., 2006). Patterns of gene expression in isogenic haploid non-calcifying flagellated cells and diploid calcifying cells of different life cycle phases indicated diploid specific expression of genes potentially involved in calcification. In diploid cells genes potentially related to biomineralisation included Ca^{2+}, H^+, and HCO^{3-} -pump coding sequences. Specifically, transcripts of Ca^{2+}-transporting ATPase, K^+ - dependent Ca^{2+}/Na^+ exchanger NCKX1, and V-type H^+ - ATPases subunits a, d, M9.7, c/c', A, B F, and H were more abundant in diploid calcifying cells (von Dassow et al., 2009). In naked and coccoliths-bearing cells of *Pleurochrysis haptonemofera* Fujiwara et al. (2007) found 54

genes of 3-fold higher expression specific to the coccoliths-bearing phase of the life cycle, of which 32 matched to registered sequences, among them sequences referring to carbonic anhydrases, Myb1, and a mitochondrial Ca^{2+} - dependent solute carrier (for a summary, see Table 1-1).

Additional knowledge of yet unidentified genes involved in the biomineralisation process in the diurnal phases of coccolithophorgenesis is required to better understand the processes of biomineralisation in general and the timing of relevant gene expression events. Recent studies, have suggested molecules putatively involved in the biomineralisation processes in *E. huxleyi* and extended existing models of calcification suggesting a role of putative Ca^{2+}-channels in the control of Ca^{2+} flux from the peripheral endoplasmic reticulum into the maturing CV, the requirement of H^+ removal from the cytosol by voltage gated H^+ channels and an amorphous calcium carbonate pre-cursor (Mackinder et al., 2010). This study aims to identify the patterns of biomineralisation related gene expression and possibly identify new links of molecules and biomineralisation applying next generation RNA sequencing techniques on calcifying and non-calcifying *E. huxleyi* cells.

3.1.4. Strategies for gene-expression analysis

In transcriptome research the diversity, structure, and sequences of RNA molecules are investigated with the aim to connect the genome to gene functioning. Two main approaches are followed in transcriptome studies, the knowledge-driven and the data-driven approach. The choice of the approach depends on the goals of the investigator. In the knowledge driven approach the researcher knows the genes or physiological pathways and the proteins of interest from a sequenced genome or the molecular process of interest. In a data-driven approach the entire population of mRNA in a cell or tissue is sequenced. The data was analysed using known sequences and gene homologies from other organisms. When supplemented with their functions the discovered transcripts become truly meaningful and useful to increase our understanding of the molecules involved in molecular physiology.

3.1.5. Aims and Objectives

Next generation sequencing techniques open the door to a wider exploration of the transcriptomes of *E. huxleyi* showing elevated and no calcification, promising also the illumination of new genes potentially involved in the calcification processes. Previously, the molecular controls of calcification in *E. huxleyi* were studied using phosphorous limitation (Wahlund et al., 2004a; Nguyen et al., 2005; Quinn et al., 2006), comparing isogenic non-

calcifying haploid with calcifying diploid life cycle-phases (Fujiwara et al., 2007; Richier et al., 2009; von Dassow et al., 2009;), and relating isogenic non-calcifying cultures to calcifying cultures (Mackinder et al., 2011). We have previously established (Chapter 2) that the bulk of calcification in this *E. huxleyi* PLY♯ M217 took place during the light phase in the late G1-phase and is presumably energetically driven by photosynthetic carbon fixation. Furthermore, giving the assumption that higher gene expression related to biomineralisation processes is linked to the higher rates of calcification, gene expression important to coccolithogenesis could be explored by assessing the transcriptome in a high calcification late G1-phase (HC). Low levels of biomineralisation related gene expression could be assumed when no photosynthesis is occurring but cell are in the G1-phase (Paasche 2000) (see Chapter 2). In order to compare high calcification rates related gene expression to low calcification rates related gene expression the target cell harvesting time was chosen in the at low calcification rates in the early G1-phase (LC). Investigating *E. huxleyi* at calcifying and low calcifying phases of the G1-phase has the further advantage of limiting the degree of cell division derived gene expression noise. Furthermore, by using the next generation full transcriptome analysis approach other levels of molecular processes potentially involved in biomineralisation may become apparent. Therefore, cultures of *E. huxleyi* strain PLY♯ M217 were harvested at low and high rates of calcification in the G1-phase of the cell cycle. The transcriptomes at low and high calcification rates were sequenced on Hi Seq Illumina platforms and significant differences in transcript abundance between the low and high calcification phases were evaluated.

The aims of this investigation were:

- to sequence the transcriptomes of iso-genetic cultures from low and high calcification phases in the G1–phase to reduce cell division related transcript recovery,
- to determine significant differences in transcript abundance and diversity between the low and high calcification phase, and
- to investigate details of the molecular machinery that drives calcifcation.

3.1.6. Working hypothesis

The null hypothesises for these studies are:

- There are no differences in transcript abundance between the transcriptomes of *E. huxleyi* PLY♯ M217 from the high and low rate calcification period within the G1-phase;
- There are no differences in transcript diversity between the transcriptomes of *E. huxleyi* PLY♯ M217 from the high and low rate calcification period within the G1-phase;
- The general expression of genes related to calcification is not enhanced in the high calcification period when calcification rates were found to be at a maximum.

3.2. Materials and Methods

To identify the molecular processes involved in the formation of coccoliths RNA and proteins were extracted from *Emiliania huxleyi* PLY♯ M217 cultures when cells were calcifying at high and low rates. The timeframes for culture harvesting were determined from the results of Chapter 2. The physiological performance of the cultures was assessed to exclude any potential limitation of nutrients or alternations of the cell population development that would affect gene expression. The extracted transcripts of messenger RNA at high and low calcification rates were sequenced using next generation sequencing techniques. Subsequently, a thorough evaluation of the gene expression patterns in respect to the conditions high calcification (light) and low calcification (dark) was conducted and candidate genes involved in calcification were identified.

3.2.1. Experimental Design

Exponentially growing stock cultures of *Emiliania huxleyi*, strain PLY♯M217, previously adjusted to light, temperature and nutrient regimes (F/50), were used to inoculate three 16-litres stock cultures in 20L Nalgene® culture vessels. The 16L cultures were reared for 3 to 5 generations until cell densities exceeded 50000 to 100000 cells mL^{-1}, in order to supply sufficient coccolithophore biomass for RNA and protein extractions.

3.2.2. Algae Culturing

Stock cultures of *Emiliania huxleyi* PLY♯ M217 were reared for 20 generation in 2L culture flasks in 1L filter-sterilized F/50 culture medium based on oceanic natural seawater (Guillard, 1975). The seawater was collected from offshore Plymouth (UK) and filtered through 0.22 μm pore-size Millipore-filters (Millipore, Billerica, USA). All algae cultures were grown at 19 °C, irradiance levels of 103 ± 6 μE sec^{-1} m^{-1} (n=32) in a 12:12 Light:Dark (LD) cycle using Sylvania white fluorescent F36W/135-T8 bulbs (Havells Sylvania Europe, Raunheim, Germany), and macronutrient concentrations of 36 μM nitrate and 1.45 μM phosphate at the time of culture inoculation. The algae cultures were mixed daily and cell densities were monitored using a Neubauer haemocytometer (Weber Scientific International, Teddington, UK).

Several 20 L Nalgene® culture vessels were prepared with16 litres of F/50 culture medium based on sterilized oceanic seawater. Each culture medium was inoculated with 15000 cells ml^{-1} and reared for 3 to 5 generations until the cell densities exceeded 50000 to 100000 cells mL^{-1}. Throughout the experiment some cultures were aborted because the growth of the cell

population collapsed. Only *E. huxleyi* cultures that remained in the exponential growth phase were used for transcriptome analysis. In total three replicate cultures for each calcification phase (low and high) were harvested. Hereafter, the replicate cultures were referred to as HC-1 to HC-3 and LC-1 to LC-3.

To estimate the physiological state of the *E. huxleyi* cells physiological and seawater parameters, such as cell densities, cell volume, POC, PON, PIC, nutrient uptake, DIC, total alkalinity, chlorophyll *a*, and photosynthetic health were assessed one day before culture harvesting and at the day of culture harvesting. The methods used to assess the cell physiological and seawater parameters were described in detail in Chapter 2 of this monograph. Hereafter, only modifications from the methods described in Chapter 2 are elaborated.

3.2.3. Culture harvesting

Cells were harvested for biomass collection at 7:30 pm and 7:00 am for the light incubation period and dark incubation period, respectively. The results of cell population growth and DNA content (Chapter 2) suggested that the cells were in the G1-phase at both harvesting time. Naturally RNA degrades rapidly (Jan, 2002; Baker et al., 2004; Deutscher, 2006). To limit RNA alteration a flow through filtration set-up as shown in Figure 3-1 was used to collect the cells. The cells were retrieved on 3 μm pore-size polycarbonate filters Ø147mm (Millipore Billerica, MA, U.S.A) fitted in a pancake-filter holder (see Figure 3-1). A Watson-Marlow Bredel Pump 620S (Falmouth, Cornwall, UK) at 50 rpm gently pressed the total coccolithophore suspension from the 20L Nalgene® culture vessels through the polycarbonate filter. To avoid sample contamination especially with RNAses or proteases all parts of the filtration set-up, (e.g. siphons, tubing, filter-holder, and filters) were washed for 24 hours in 10% hydrochloric acid prior to commencing the cell harvesting. Once the total *E. huxleyi* – suspension was siphoned through the filtration set-up the polycarbonate filter containing algae cells was transferred quickly onto a sterile custom made ice cooled plastic half pipe (10 cm diameter). The cells were washed off the filter with sterile filtered seawater (0.22 mm polycarbonate, Millepore®) and the concentrated suspension collected in a 50 mL falcon tube. The cell suspension was immediately centrifuged for 5 min at 4630 rpm and 4 °C (Rotanta 460R, Hettich Zentrifugen, Germany). The seawater supernatant was discarded, the concentrated cell pellet snap-frozen in liquid nitrogen and stored at -80 °C until RNA extractions were conducted. This harvesting procedure was completed within 25-30 minutes.
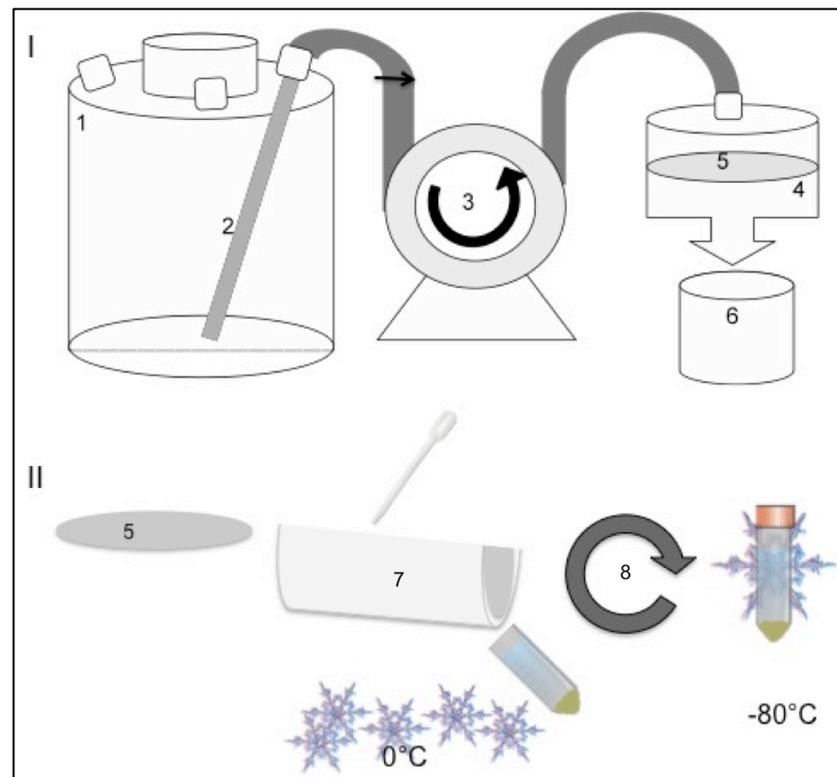
Figure 3-1: Harvesting workflow of *E. huxleyi* cultures. I – Pumping of culture medium from culture vessel (1) through siphon (2) via peristaltic pumping (3) onto 147 mm diameter polycarbonate filters (5) fitted in pancake filter membrane holder (4). The filtrate is collected in another culture vessel for sterilization (6). II – Collection of coccolithophores from filters (5) in a half-pipe (7) with sterile filtered seawater under cooling, centrifugation (8) for 5 min at 4630 rpm, snap-freezing of pellet and storage at -80°C.

3.2.4. Assessment of physiological performance of *E. huxleyi* cells

To estimate the physiological performance of *E. huxleyi* cells during the experiment the following parameters were assessed according to the methods described in Chapter 2 without modifications:

- Cell density
- Growth rate
- Photosynthetic health
- Particulate organic carbon (at one day prior and the day of harvesting)
- Particulate organic nitrogen (at one day prior and the day of harvesting)
- Particulate inorganic carbon (at one day prior and the day of harvesting)
- pH
- DIC, total alkalinity, salinity (at the time of harvesting)
- Nutrient concentrations (at the time of harvesting)

3.2.5. Statistics on non-molecular data

Differences in the physiological performance (see section 2.3) of *Emiliania huxleyi* cells in cultures harvested during the light and in the dark were graphically presented using SigmaPlot 12.5 (Systat Software, Salisbury, UK). Statistical tests, such as two-sample t-test, Mann-Whitney Rank Sum Test were applied on cell growth and fitness parameters to test for differences between conditions (HC and LC). In the case of the larger data set sizes retrieved for photosynthetic health and growth rate ANCOVA models in STATISTICA Version 10 (StatSoft Inc., Tulsa, USA) were applied to test for significant changes over time and between the conditions, using day as covariate. Shapiro-Wilk tests for normality and equal variance tests were performed in SigmaPlot 12.5 to confirm the suitability of the parametric design. Results of the normality tests and equal variance tests are not presented.

3.2.6. RNA extractions

Total RNA was extracted following a TRI Reagent® (Sigma-Aldrich, Dorset, UK) method. The TRI® total RNA extraction protocol uses the principle of RNAse inactivation by guanidinium isothiocyanate and an acidic phenol/chloroform separation of lipids and proteins from the RNA containing aqueous phase. The greatest yields of RNA quantity and quality were achieved using TRI Reagent® extraction followed by a Nucleospin® II (Machery-Nagel GmbH & Co KG, Düren, Germany) standard clean up protocol, which was supported by two subsequent DNA digestion steps. At first, a fraction of the frozen cell pellet (see section 3.2.3) was separated and transferred to a DNA/RNA free 2 mL Eppendorf tube. Then, 1.5 mL of cooled TRI Reagent® were added and the pellet resuspended by pipetting. Cell lysis was achieved by vortexing the TRI Reagent® cell suspension for 5 minutes with 20 acid – washed glass beads (0.5 mm diameter) added to the 2 mL Eppendorf tube. The suspension of disrupted cells was allowed 15 minutes at 37 °C for a complete inactivation of RNAses and dissociation of nucleoprotein complexes. The cell debris was separated from the suspension by centrifugation at 14000 g, 15 min, 4 °C (Mikro 22R centrifuge, Hettich Zentrifugen, Germany). Subsequently, the RNA was extracted and separated from DNA and proteins by adding 300 μL of chloroform (Sigma-Aldrich, Dorset, UK) and vigorous vortexing for 15 seconds. After allowing 2 min at room temperature centrifugation for 15 min at 12000 g and 4 °C (Mikro 22R centrifuge, Hettich Zentrifugen, Germany) completed the formation of the aqueous phase. The aqueous phase was pipetted into a fresh micro-centrifuge tube (1.5 mL) and another chloroform wash step conducted to remove remaining lipids and proteins. From the final collected aqueous

phase the RNA was precipitated with - 20 °C isopropanol (Sigma-Aldrich, Dorset, UK), incubated at - 20 °C for 20 min and the RNA pelleted by centrifugation for 15 min (12000g, 4 °C). After removal of the supernatant the RNA pellet was washed with 75% ethanol (Sigma-Aldrich, Dorset, UK) once and dried on ice under sterile conditions (laminar flow system Foster BHG 2006; Foster-air, Cornaredo, Italy). Subsequently, a first DNA digestion step using Ambion®-Turbo DNAse-free (Life Technologies Co.) was performed following the manufacturer's instruction. Preliminary trials, revealed that residual DNA on acrylamide gels was still present after the Ambion®-Turbo DNAse digestion step. Therefore, the resulting sample was submitted to a second on column DNA-digestion step using standard Nucleospin® II DNAse followed by a Nucleospin® II column clean up for RNA samples from reaction mixtures. The purified RNA sample was diluted in 100 μL DNAse/RNAse - free water (HyClone^TM, Thermo Scientific, Loughborough, UK) and stored on ice for shorter periods or at - 80 °C until further processing.

3.2.7. RNA Quality and Quantity

The spectrometric absorbance ratio at wavelength 280 and 260 nm A_{260}/A_{280} was analysed as a measure for RNA purity and quantity (Glasel 1995) using a ND-1000 spectrophotometer (NanoDrop®, Thermo Scientific, Wilmington, USA). The absorbance at 260 nm (A260) is a measure of nucleic acid concentration, whereas the absorbance at 280 nm shows amounts of contaminants, such as proteins or phenols, which may inhibit enzymatic reactions in downstream protocols. An A_{260}/A_{280} ratio of 1.8 to 2.0 is preferable for RNA sequencing or complementary DNA library construction. RNA integrity describes the contribution of RNA subsets of different lengths to a population of RNA and it is of greatest importance when applications involve RNA quantitation for gene expression studies because it reflects the degradation status of the isolated RNA (Denisov et al., 2008). Naturally, cells contain only small amounts of messenger RNA but greater amounts of ribosomal RNA 18S and 28S fractions (5 kb and 2 kb in size, respectively). Therefore, a 28S/18S rRNA ratio close to 2 should be indicative for intact mRNA (Sambrook et al., 1987). The RNA integrity in the described experiments was measured on an automated capillary micro fluid electrophoresis system - Experion^TM (Experion^TM Bio-Rad Laboratories, Inc., Hemel Hempstead, UK) (Imbeaud et al., 2005) using StdSens chips and reagents. The Experion^TM - system delivers results in form of a virtual gel (electropherogram) and the RNA Quality Indicator (RQI-value). Three regions of the electropherogram are taken into account when calculating the RQI value: the 28S region, the 18S region, and the pre - 18S regions. The Experion^TM Bio-Rad system compares the sample

to a standard and returns RQI values between 10 (intact RNA) and 1 (highly degraded RNA) for each eukaryotic RNA sample (for details see Denisov et al., 2008). An electropherogram also gives details about the presence of rRNA and tRNA segments – in the range of 5S and 5.8S - in the sample. For the subsequent steps of RNASeq Illumina sequencing a RQI threshold of 7.5 was used as a minimum requirement. The Experion^{TM} Bio-Rad StdSens protocol and recommendations were followed in detail to assess the integrity of the RNA samples. In the case of poor RQI values the extraction from the frozen cell pellet (section 3.2.6) was repeated until RNA quality specifications were affirmed. For sequencing 12 ng RNA diluted in TE-buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) were aliquoted stored at - 80 °C.

3.2.8. Independent approach of promoter expression analysis

Following RNA extraction and RNA quality validation (section 3.2.7) another aliquot of 15 ng total RNA per sample was sent to DNAFORM Inc. (Leading Venture Plaza-275-1, Ono-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0046, Japan) for cap analysis of gene expression (CAGE) (Shiraki et al., 2003; Kodzius et al., 2006). The CAGE identifies the exact transcription start site by sequencing short 5'-sequences of the mRNA population. In CAGE the beginning of the 5'mRNA is reverse transcribed after selecting the biotin residue at the cap sites using streptavidin-coated magnetic beads (Corninci et al., 1996). CAGE raw sequences have ben independently analysed by Dr. Reidar Andreson (University of Tartu, Estonia) to investigate differential patterns of promoter expression in *E. huxleyi* PLY♯M217 at rates of high and low calcification.

3.2.9. Illumina sequencing

For the investigation of the transcriptomes from the at times of high and low calcification in the cell cycle of *E. huxleyi* the extracted RNA was sent to the Institute of Clinical Molecular Biology (IKMB) at the Christian-Albrechts-University Kiel, Germany for sequencing. In total 6 samples (3 x HC, 3 x LC condition) were shipped on dry ice to the IKMB for Illumina based RNA sequencing. The IKMB followed the Illumina TrueSeq sample prepation guide. RNA was reverse transcribed into cDNA using Sera Mag$^{®}$ oligo(dT)$_{14}$ tags, magnetic Oligo(dT) microparticles, which covalently bind specifically to polyadenylated mRNA and therefore establishing a population of cDNA from purely eukaryotic mRNA (see Sera-Mag manual for details). The prepared cDNA from low and high calcification periods of *E. huxleyi* was sequenced using a raw Illumina 2000 system sequencing at a depths of 100 bp paired reads.

Raw sequence reads were provided by the IKMB and in silico processed by Dr. Reidar Andreson, University of Tartu, Estonia.

3.2.10. Processing of paired end Illumina output data

The raw sequences from the Illumina sequencing were quality checked using FastQC Version 0.11.4 (www.bioinformatics.babraham.ac.uk/projects/). The sequencing data was then converted into sequence counts and gene expression analysis compatible data by Dr. Reidar Andreson, University of Tartu, Estonia. The software package Tophat (Trapnell et al., 2009; http://tophat.cbcb.umd.edu) was applied for clustering and mapping the transcript sequences to the reference genome of *E. huxleyi* CCMP 1516 version from 25.4.2008 available at http://genome.jgi.doe.gov/Emihu1/Emihu1.home.html. Then the Cufflinks package (Trapnell et al. 2010, http://bio.math.berkeley.edu/cufflinks) was utilized to test for significant differences in the abundance of transcripts between the light and dark condition.

Tophat uses a two-step read-mapping algorithm suitable for the alignment of reads from RNA-Seq experiments (Trapnell et al., 2009). First, the sequences are mapped to the reference genome using Bowtie, a short DNA sequence aligner (Langmead et al., 2009) that also accounts for 'initially unmapped reads'. The initial consensus of mapped regions is computed using the subroutine Maq (Li et al., 2008). In the analysis Dr. Reidar Andreson applied the following Tophat-parameters: number of threads 30 bases, inner-distance 300 bases, minimal intron length 20 bases, maximum intron-length 1000 bases, minimal anchor-length 10 bases, splice-mismatches 1, minimum-closure-intron 20 bases, maximum-closure-intron 1000 bases, minimal coverage-intron 20 bases, maximal coverage-intron 1000 bases, minimal segment-intron 20 bases, and maximal segment-intron 1000 bases. In the case of paired end reads, as produced by the Illumina 2000® sequencer in this study, the counting of duplicate reads from the paired end approach was avoided.

The Cufflinks package was applied to estimate the abundance of transcripts from the paired end reads of the Illumina 2000® - data. Furthermore, Tophat-annotation was conducted to compare the differences in gene expression levels. The algorithm aligns overlapping 'bundles' of fragments, identifying spliced mRNA isoforms as pairs of 'incompatible' fragments by taking into account the distribution of fragment lengths for the identification of isoforms. Cufflinks counts the transcript abundance in fpkm (fragments per kilobase per million) and utilizes a beta-negative binomial model to estimate the variance of the RNA-Seq data and to

distinguish significant differences between treatment conditions such as the HC and LC running a t-test-like statistic. The significance is based on q-values. The q-value is a correction of the p-value that accounts for the likelihood of *a type 1 error* using a significance level threshold of 0.05 (see Benjamini and Hochberg, 1995). Cufflinks 2.0 produces multiple result tables that contain transcript abundance, isoform abundance, promoter area usage, transcription start sites counts, the coding sequence, and the significantly different transcripts. Subsequently, the R - package CummeRbund (available http://compbio.mit.edu/cummeRbund/) (Trapnell et al., 2012) was applied to build a database from the cufflinks output files. This database was used as the core for the analysis of differential expression between the conditions HC and LC applying procedures described in the CummeRbund manual and to create further tables isolating specific results of interest and connecting transcripts with annotation data (see details in section 3.2.11). The pipeline of transcript sequence mapping and differential gene expression analysis is summarized in Figure 3.2.

3.2.11. Assessment of data quality and of transcript abundance

The cufflinks output files were connected using the R-package CummeRbund to create a SQL – related database that linked transcripts with parameters such as expression data, isoforms, and transcription start site. In order to investigate the general data for over-dispersion, the quality of the sequencing data and the performance of the Tophat-Cufflinks pipeline graphical visualization, Box-Plots, and hierarchical clustering techniques were applied.

3.2.12. Assessment of transcript abundance in the high rates of calcification (HC) and low rates of calcification (LC) condition and overall identification success

Overrepresented transcripts in the light were assigned to the HC condition and transcripts overrepresented in the dark were assigned to the LC condition for all subsequent evaluations. Histograms of the transcript abundances and the fragment-sizes were produced through the R - package CummRbund to investigate general differences between the transcriptomes from the HC and LC condition. The overall identification success of isoform, transcription start sites (TSS), coding sequences (CDS), promoters, splicing regions, and regulated CDS was called from the CummeRbund database. For details of the commands used refer to the CummeRbund manual (http://compbio.mit.edu/cummeRbund/manual_2_0.html).

3.2.13. Modification of the Cuffdiff output

The cuffdiff output file "gene_exp_func.diff" was the source of differential gene expression analysis between the HC and HC condition, whereas the significance level for differential gene expression between the conditions was $q < 0.05$. In the original output file some transcripts were present as duplicates and the clustering by the Tophat-Cufflinks pipeline was apparently omitted. To correct duplicate transcripts, the cufflinks output table "gene_exp_func.diff " was modified for downstream analysis. Multiple expression values for the same transcript sequence were simply summarized. According to Trapnell at al. (2012), adding transcript abundances referring to the same gene or transcript is permitted because the fpkm-values are directly proportional to the abundance of transcripts.

3.2.14. Gene recruitment and annotation success

The numbers of transcripts that could be recruited to genes of the JGI *E. huxleyi* reference genome (http://genome.jgi.doe.gov/Emihu1/Emihu1.home.html) were counted. Additional annotation information, such as KOG (euKaryotic Orthologous groups), KEGG (Kyoto Encyclopedia of Genes and Genomes), and protein information based on non-redundant protein NR, SwissProt and UniProt databases was downloaded from JGI sources and connected to the gene expression results. Subsequently, the transcript annotation success was assessed at the levels of gene recruitment, KOG annotations, KEGG annotations, and protein annotations and differences between the conditions evaluated.

3.2.15. Assessment of functional differences between the high calcification (HC) and low calcification (LC) condition using KOG-class annotations

The largest amount of functional annotations was provided by KOG information. Therefore, the KOG annotations were used to investigate functional differences between the conditions. The numbers of transcripts in relation to their predominance in the HC or LC condition were used to investigate differences in functional expression within KOG-groups and KOG-classes. The arbitrary significance level of $p < 0.01$ for Chi-Square tests was applied to classify significant differences in gene expression between KOG-classes. The data provided low degrees of freedom. Therefore, the Chi-square test with Yates-correction, was applied to distinguish significant deviation of the ratio of the number of significantly overexpressed genes LC:HC in each KOG-class to the overall ratio LC:HC of the gene expression by means of transcripts (19093:15310). Furthermore, an arbitrary threshold for the ratio of gene expression of all genes

expressed in each KOG-class and the ratio of gene expression for significantly higher expressedgenes was set at 2 and 0.5 to indicate difference in expression patterns in the functional KOG-classes. This ratio is considered to be comparable to the Chi-square approach. However, in those KOG-classes were omitted from consideration where the thresholds were created by an absence of transcripts in either condition.

3.2.16. Assessment of gene expression differences between the high calcification (HC) and low calcification (LC) condition

Gene expression differences between the HC and the LC condition were investigated using a number of approaches. First, it was of interest which transcripts might be absent from the other calcification condition in the G1-phase (HC or LC). Therefore, the genes unique to either condition were investigated. Genes with a potential role in the calcification process in *E. huxleyi* were taken in focus. Secondly, the functional groups according to KOG-annotations were applied to subdivide HC and LC transcriptomes. Following this KOG-class grouping, the frequency of transcripts in the HC condition was plotted against the frequency of transcripts in the LC condition, marking non-significantly and significantly more abundant transcripts. This graphical presentation of the transcripts in each KOG-class was used to identify interesting clusters of significant transcripts. The genes and the functions of the members of the selected transcripts were investigated in detail. Thirdly, a selection of molecules, previously reported in the literature to play a role in the processes surrounding calcification and coccolith formation in *E. huxleyi,* was made (Marsh et al. 1992; Wahlund et al., 2004; Nguyen et al. 2005; Quinn et al., 2006; Richier et al., 2009; von Dassow, 2009; Macckinder et al., 2009, 2011, 2012; Emery et al., 2012). The JGI gene ID were retrieved for the molecules of interest, such as V -type proton ATPases, clathrin, P - type proton ATPase, proton exchanger, Ca^{2+} - transporters, calmodulin, Ca^{2+} - binding proteins, GPA, bicarbonate transporter, and carbonate anhydrase (CA) and their gene expression levels plotted.
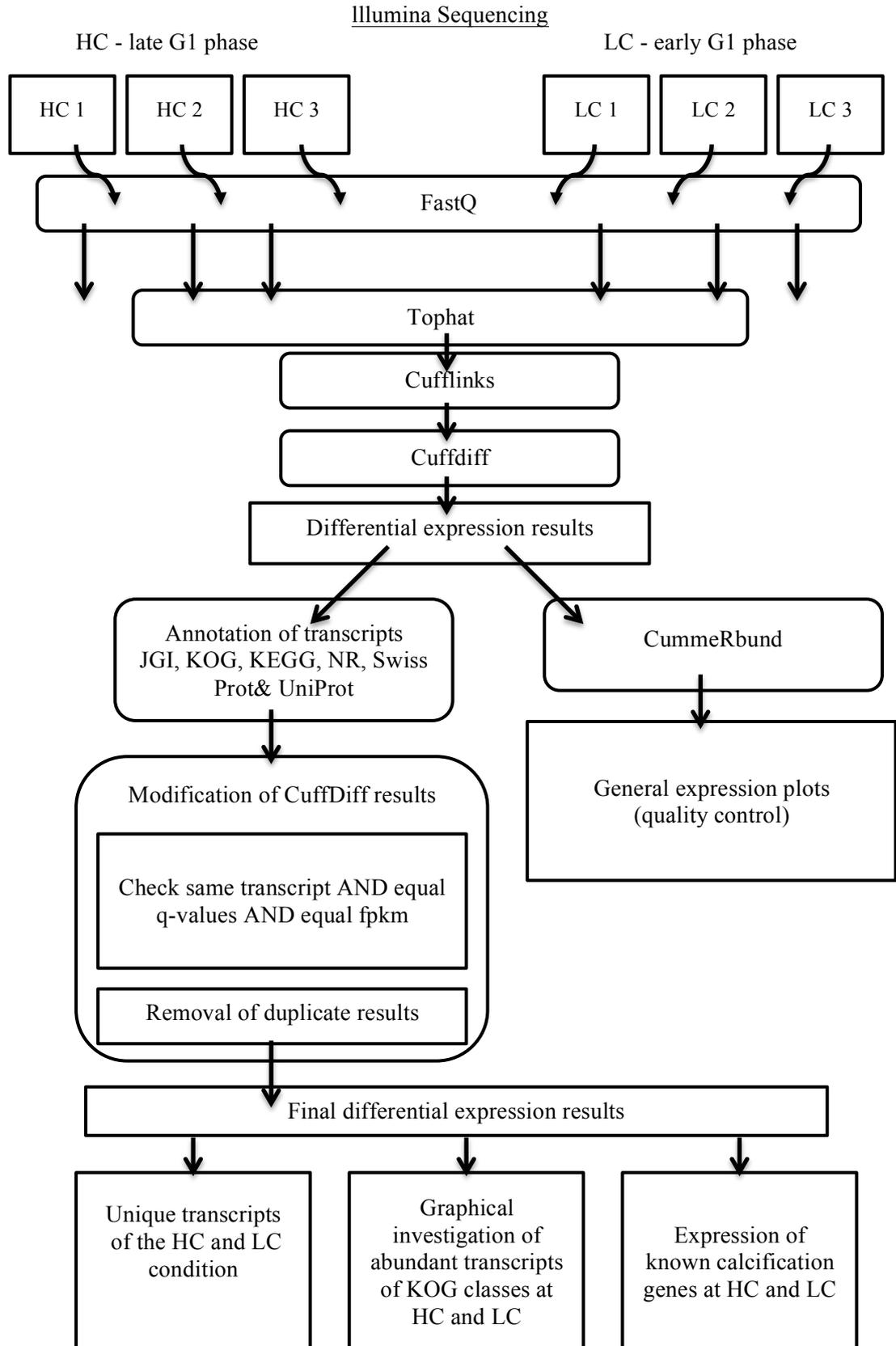
Figure 3-2: Flowchart showing the bioinformatics pipeline and data analysis from raw Illumina results to differential expression results comparing gene expression at the transcript level in the high calcification rate (late G1-phase) and low calcification rate phase (early G1-phase) of *Emiliania huxleyi* PLY♯ M217.

3.3. Results

The physiological parameters between the day before cell harvesting and the day of harvesting showed significant differences for PON and PIC. No significant differences were found between the high calcification and the low calcification condition. This suggests that the physiology of the *E. huxleyi* cell population was not altered. The results are graphically summarized in Figure 3.3.

3.3.1. Cell fitness parameters in experimental *Emiliania huxleyi* cultures

Growth rates per day were calculated from daily cell density measurements to assess the development of the *E. huxleyi* PLY♯ M217 cultures. Furthermore, physiological growth parameters such as particulate organic carbon per cell, particulate organic nitrogen, particulate inorganic carbon, and the photosynthetic quantum yield (Fv/Fm) were quantified to judge the fitness of the cells. The ratios of POC:PON and PIC:POC were calculated.

3.3.1.1. Cell culture growth

E. huxleyi PLY♯ M217 grew exponentially in all F/50 natural seawater cultures. Figure 3-3 A and B illustrate cell densities and growth rates observed in the *E. huxleyi* cultures until cell harvesting. Five out of 6 cultures showed a lag phase of one to two days before growth rates increased to above 0.5 d^{-1} (Figure 3-3 B). In one cultures (HC 3) this lag phase extended over two days. The maximum growth rate of 1.41 d^{-1} was observed in culture HC 3 between day 3 and day 4. Average growth rates of HC and LC cultures were 0.69 ± 0.02 (n=3; SD) and 0.68 ± 0.04 (n=3; SD), respectively. Despite the observed lag phase the growth rates in the culture vessels were not significantly different (ANOVA; df: 4,27; F = 0.147, p = (0.978).

3.3.1.2. Particulate organic carbon (POC) per cell

One day before harvesting POC was in average 1.36 ± 0.1 pmol $cell^{-1}$ in the light and 1.40 ± 0.27 pmol $cell^{-1}$ in the LC cultrues. At the day of harvest lower POC values were observed, being 1.04 ± 0.09 pmol $cell^{-1}$ and 1.19 ± 0.26 pmol $cell^{-1}$, for the HC and LC cultures respectively (see Figure 3-3 C). The sampling time was scheduled earlier than on the day before the harvesting due the extended work plan at the day of harvesting. No significant differences between conditions and sampling days became apparent (Kruskal-Wallis ANOVA; df: 3, H= 3.667, p = 0.381).
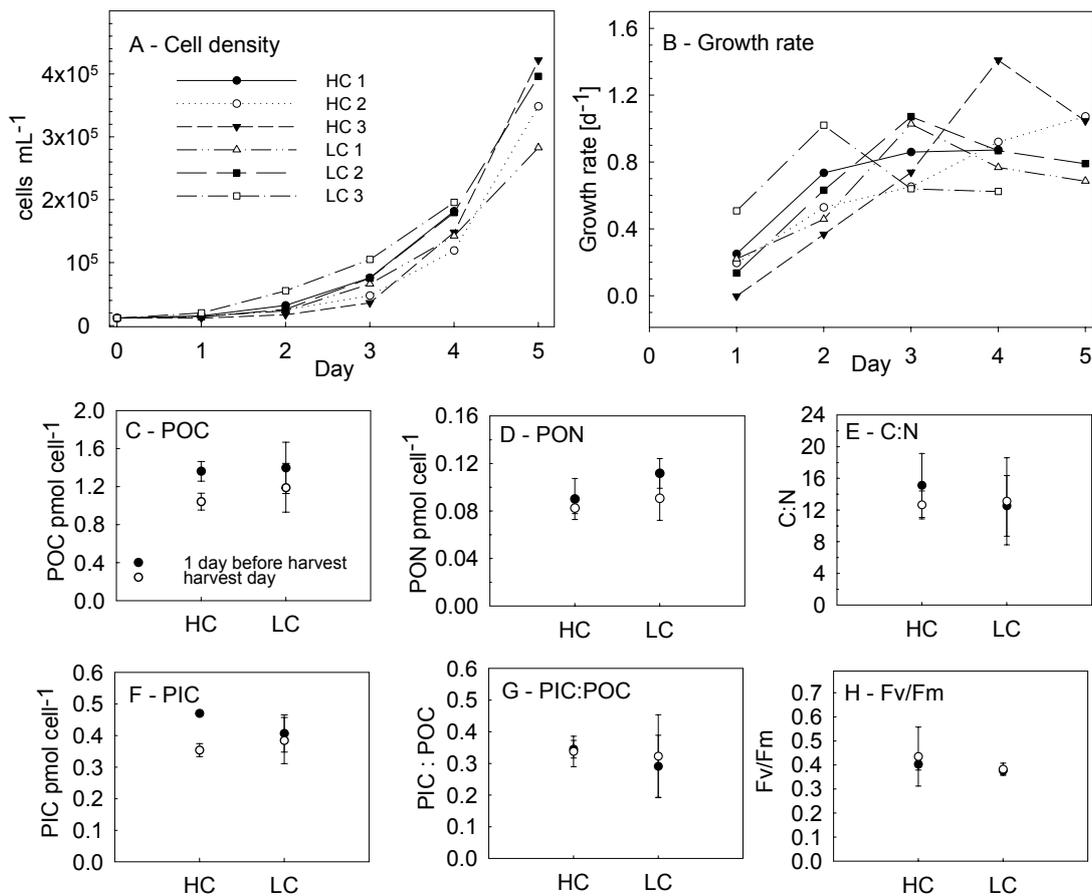
Figure 3-3: Cell fitness parameters of *E. huxleyi* cells in culture for transcriptome and proteome analysis. A - Cell densities in *E. huxleyi* cultures, B - Growth rates [μ] per day of *E. huxleyi* cultures; HC 1 to HC 3 replicate cultures for harvesting at high calcification rates and LC 1 to LC 3 for cultures harvested at low calcification rates. C – POC (Particulate Organic Carbon) content per cell, D – PON (Particulate Organic Nitrogen) content per cell, E – C:N (Ratios of particulate inorganic carbon to organic carbon), F – PIC (Particulate Inorganic Carbon) per cell, G – PIC:POC (Particulate Inorganic Carbon to Particulat Organic Carbon ratio), and H – Fv/Fm (average values of quantum yield of PSII based on FRRF, Chelsea II, measurements of *E. huxleyi* cells in cultures 1 day before and at the day of harvesting. For C, F, and H n = 3. For D, E, and G n = 2. Error bars show standard deviation from the mean.

3.3.1.3.   Particulate Organic Nitrogen (PON) per cell

Averages of PON at the time of harvesting were 0.082 ± 0.007 pmol cell^{-1} and 0.091 ± 0.018 pmol cell^{-1} for the HC and LC cultures, respectively (compare Figure 3-3 D). No significant differences between conditions became apparent No significant difference was found in respect to PON between the different sampling times (Kruskal-Wallis ANOVA; df: 3, H= 2.833, p = 0.543).

3.3.1.4.   Particulate Carbon to Nitrogen ratio (C:N) per cell

The ratios of organic carbon and nitrogen per cell at the day of sampling were 12.7 ± 1.8 and 13.1 ± 5.5 in the HC and LC cultures, respectively (see Figure 3 – 3 E). At the day before cell harvest C:N values were slightly higher with 15.1 ± 4.0 and 12.5 ± 3.8 for the HC and LC condition, respectively. No significant differences were found in the C:N values between the HC and LC cultures and the different sampling times (Kruskal-Wallis ANOVA; df: 3, H = 4.167, p = 0.324).

3.3.1.5.   Particulate Inorganic Carbon (PIC) per cell

One day before the cell harvesting the PIC in the HC and LC cultures was 0.47 ± 0.01 and 0.41 ± 0.06, respectively. Just prior cell harvest the PIC content was 0.34 ± 0.05 pmol cell^{-1} and 0.32 ± 0.13 pmol cell^{-1} in the HC and LC condition, respectively (see Figure 3-3 F). No significant difference was found for PIC between the conditions and the different sampling times (ANOVA; df: 3,8, F = 3.212, p = 0.083).

3.3.1.6.   Particulate Inorganic Carbon : Particulate Organic Carbon ratio (PIC:POC)

The ratio of PIC to POC was conserved over the two days of measurement (see Figure 3-3 G). It was between 0.35 ± 0.03 and 0.34 ± 0.05 for the HC cultures and between 0.29 ± 0.10 and 0.32 ± 0.13 in the LC cultures, for the day before harvest and the day of harvest respectively. No significant difference was found in respect to PON between the HC cultures and the different sampling times. There was no difference in PIC:POC between one day before harvesting and the day of harvesting and the HC and the LC conditions (Kruskal-Wallis ANOVA; df: 3, H = 5.500, p = 0.139).

3.3.1.7.   Photosynthetic health

The average photosynthetic efficiency was observed at 0.403 ± 0.024 (n=3; SD) in the HC cultures and 0.377 ± 0.016 (n=3; SD) in the LC cultures one day before cell harvesting. At the day of cell harvesting Fv/Fm values were 0.435 ± 0.122 (n=3; SD) at HC condition and 0.382 ± 0.025 (n=3; SD) at LC condition (see Figure Figure 3-3 H). No significant difference was found in respect to Fv/Fm values between the HC and LC conditions and the different sampling times (ANOVA; df: 3,8 , F = 2.153, p = 0.172).

3.3.1.8. Nutrient consumption

Nutrient consumption of the cell populations in the cultures is shown in Table 3-1. Cells consumed in average 14.3 ± 2.2 mmol $L^{-1}d^{-1}$ of nitrogen and 0.530 ± 0.08 mmol $L^{-1}d^{-1}$ phosphorous per day. No significant differences of nitrogen or phosphorous uptake within all *E. huxleyi* cultures were found (ANOVA [N uptake]; df: 1,5, F = 0.126, p = 0.741; (ANOVA [P uptake]; df: 1,5, F = 1.926, p = 0.237).

Table 3-1: Nutrient consumption in *E. huxelyi* cultures within the last day before culture harvesting.

Culture	N uptake [μmol $L^{-1}d^{-1}$]	P uptake [μmol $L^{-1}d^{-1}$]
HC 1	15.57	0.64
HC 2	16.97	0.60
HC 3	11.30	0.46
LC 1	12.00	0.45
LC 2	14.89	0.51
LC 3	14.87	0.50

3.3.2. Dissolved inorganic carbon, total alkalinity and pH

Results for dissolved inorganic carbon, total alkalinity and pH are given in Table 3-2. In average DIC was in average 1789 ± 159 mmol kg^{-1} SW in the HC cultures and 1945 ± 66 mmol kg^{-1} SW in the LC cultures. Total alkalinity (At) means were 2102 ± 110 mmol kg^{-1} SW in the HC condition samples and 2228 ± 48 mmol kg^{-1} SW in the LC condition samples. Values of pH were 8.04 ± 0.02 in average in the HC cultures and 8.02 ± 0.03 in average in the LC cultures.

Table 3-2: Seawater parameters in the *E. huxelyi* culture medium before the harvesting of cultures. Dissolved inorganic carbon (DIC) and total alkalinity (At) are presented. Averaged values for pH and standard deviation over the growth period of 5 days for *E. huxleyi* cultures.

Culture	DIC [mmol kg^{-1} SW]	At [mmol kg^{-1} SW]	pH (n = 5) (Average ± SD)
HC 1	1967.5	2222.7	8.02 ± 0.04
HC 2	1736.1	2076.9	8.05 ± 0.06
HC 3	1663.3	2007.4	8.04 ± 0.04
LC 1	1929.0	2223.2	7.99 ± 0.08
LC 2	1888.3	2183.0	8.03 ± 0.03
LC 3	2018.3	2278.3	8.05 ± 0.05

3.3.3. Transcriptomics

3.3.3.1.   RNA quality and integrity

The RNA purity achieved the recommended values for the ratios 260/230 and 260/280 (compare Table 3-3).

Table 3-3: NanoDrop parameters of RNA samples assigned for Illumina transcriptome sequencing

Sample - Label	260/230	260/280
HC 1 – L1	2.04	2.33
HC 2 – L2	2.05	2.01
HC 3 – L3	2.02	2.41
LC 1 – D1	2.04	2.32
LC 2 – D2	2.04	1.95
LC 3 – D3	2.04	2.38

Sufficient integrity of the RNA samples extracted from cells of the HC and LC conditions is confirmed by the RQI values. The 28S:18S ribosomal RNA ratio is below 2 in all samples, which is a common finding in algae and other marine organisms (Dr. John Gittins pers. communication) (see Table 3-4 and Figure 3-11).

Table 3-4: RNA concentrations and RQI values from chip-based electrophoresis RNA samples for Illumina based transcriptome analysis.

Sample - Label	RNA conc. [ng mL^{-1}]	Ratio [28S:18S]	RQI
HC 1 – L1	417.0	1.67	8.8
HC 2 – L2	379.8	1.72	8.6
HC 3 – L3	468.4	1.57	8.4
LC 1 – D1	350.1	1.59	8.6
LC 2 – D2	468.2	1.68	9.0
LC 3 – D3	262.5	1.81	8.8
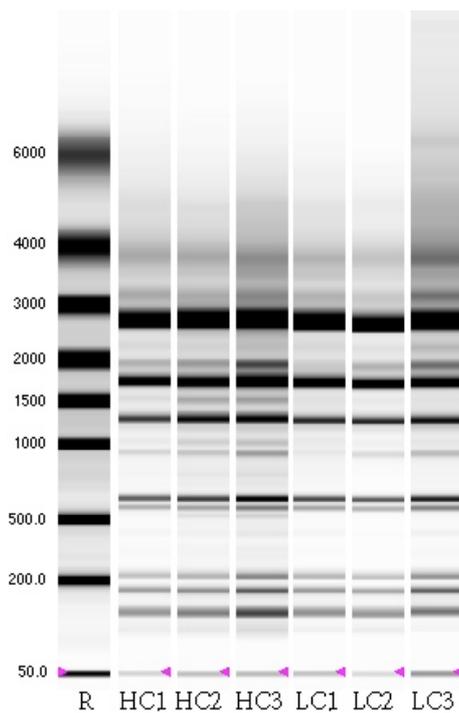
Figure 3-4: Virtual gel of RNA-samples from chip-based electrophoresis. R: Ladder, HC1 – HC3: Samples taken at the end of the Light period – HC-condition, LC1 – LC3: Samples taken at the end of the Dark period – LC-condition.

3.3.3.2.   Illumina raw sequences quality check and results of sequencing

The raw sequences from the Illumina sequencing were quality checked using FastQC Version 0.11.4 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). All sequences were 101 bp long. No sequences of poor quality were observed (data not shown). No sequence was flagged of poor quality. However, the per base sequence quality showed dropping reliability of base identification after around 80 bp with quality score values below 25 in all replicate samples. This is not an uncommon observation in Illumina sequencing and due to photobleaching of the dyes from laser scanning. The per-sequence quality score was good in all sequenced samples. No concerns resulted from using the Illumina raw sequences in the Tophat – Cufflinks pipeline. Basic statistics of Illumina reads including total reads in each sample and unique hits to the reference genome are summarized in Table 3-5.

Table 3 – 5: Basic results of the Illumina experiment.

Sample	Total reads	GC content	Reads removed	Unique hits to reference genome
HC 1	14864840	66 %	0	73.36
HC 2	18687928	66 %	0	74.08
HC 3	14537310	66 %	0	73.62
LC 1	17147800	67 %	0	73.62
LC 2	15536898	67 %	0	73.72
LC 3	14313635	66 %	0	74.38

3.3.3.3. General results and quality control of the Tophat – Cufflinks pipeline

An over-dispersion of transcript frequencies or differences in the abundances of transcripts of different length between the replicate samples was not observed. Over-dispersion of transcripts in the RNASeq replicates was visualized using Box-plot (see Figure 3-5). The boxes indicate that transcripts with higher fpkm were found in replicates of the LC condition. Outliers with a high fpkm-value were represented more abundant in the HC condition (one fpkm equals 10000 transcripts per 100 ng RNA; Trapnell et al., 2010). In the dark (LC condition) outliers of low fpkm-value appear to be more frequent. Replicate HC1 shows a greater abundance of transcripts within the first to third quartile of the data. This represents a slight dissimilarity to the replicates HC2 and HC3 and is reflected also in the hierarchical cluster analysis (see Figure 3-6). However, the dissimilarity is small and only present in one replicate. In subsequent analyses of gene expression differences this circumstance is likely not to cause a significant effect.
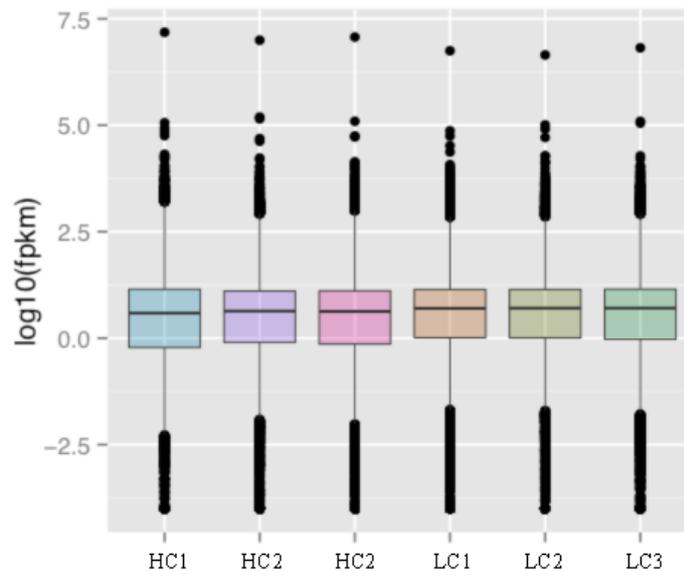
Figure 3-5: Transcript frequencies of trascriptomic samples. CummeRbund-Box-Plot showing the presence of transcripts in each replicate of the condition HC and LC in the G1-phase of *Emiliania huxleyi* (x-axis) presented as fragments per kilobase of transcript per million mapped (fpkm), including outliers. One fpkm equals 10000 transcripts per 100 ng RNA (Trapnell et al., 2010).
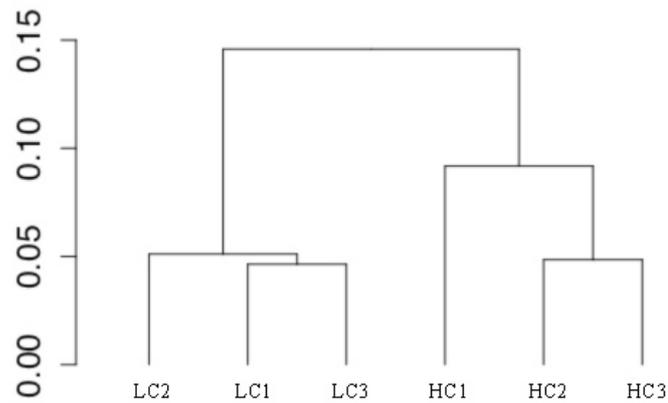
Figure 3-6: Dissimilarity-analysis of transcriptome samples. CummeRbund-Hierachical clustering analysis showing dissimilarities of all transcripts in the replciates of the conditions HC and LC.

3.3.3.4.   Transcript abundances relative to HC and LC condition

The frequency of transcripts relative to their length in fpkm is presented in Figure 3-7 to visualize differences in the population of transcripts of different sequence length in the HC and LC condition. It was obvious that transcripts of around $10^{0.6}$ to $10^{1.2}$ fpkm were more abundant in the dark (low calcification rates condition). In the light (high calcification rates condition) transcripts in the ranges of around 10^{-1} to $10^{0.4}$ fpkm and around 10^{2} to $10^{3.2}$ fpkm were more frequent.
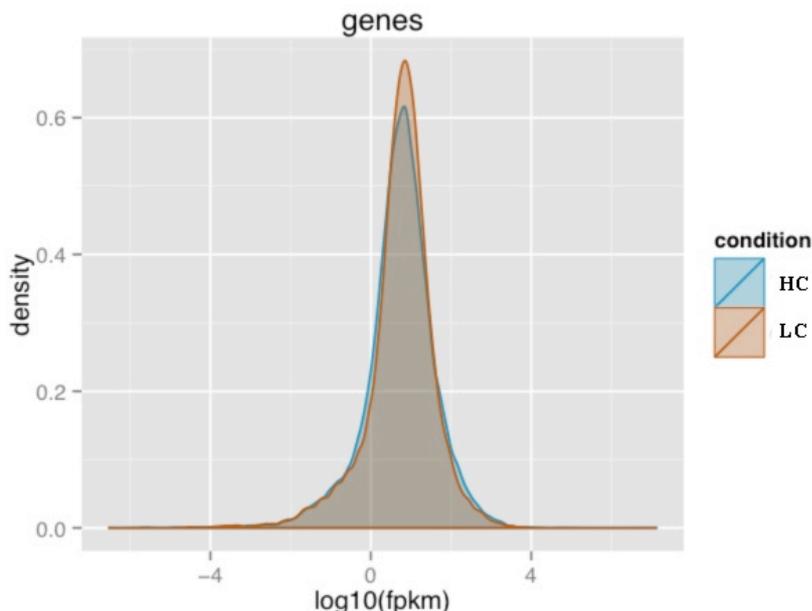
Figure 3-7: Transcript length versus transcript abundance in HC and LC condition. Transcript abundance (density – y axis) in the HC and LC condition relative to transcript length (x-axis) is presented. In this presentation 5269 cases in the HC condition and 5263 cases in the LC condition were automatically removed by the CummeRbund package.

General gene recruitment success with Cufflinks:

Cufflinks conducted counts of reference genome matches in the combined transcriptomes of *E. huxleyi* PLY♯ M217 in HC and LC conditions. A total of 43009 genes were counted in both conditions. The original results from the Cufflinks differential expression analysis algorithm were corrected and general statistics of sequence annotation were performed. Low quality or inconclusive data (see Section 3.2.13) was omitted from the analysis and multiple gene hits were merged by transcript abundances. Then the gene expression data set comprised of 34403 transcripts of which 24352 transcripts found matching sequences on the JGI *E. huxleyi* reference genome and 9999 could be mapped to know genes of the JGI *E. huxleyi* reference genome (see Table 3-6).

Table 3-6: Annotation success of transcripts. The table summarizes the transcripts found in *E. huxleyi* in the HC and LC condition, showing recruitment success to the JGI- reference genome, protein knowledge base matches of transcript clusters, and the number of retrieved KOG annotations. The number of transcripts found significantly more abundant in the HC or the LC condition are given in the columns: HC sig. (significantly more abundant transcripts in HC condition – high calcification and light incubation period) and LC sig. (significantly more abundant transcripts in LC condition – low calcification and dark incubation period).

	Total	HC total	HC sig.	LC total	LC sig.
JGI reference gene hits	9999	6446	1846	3553	515
Protein Sequences hits	15150	7099	1332	8051	1299
Transcripts with KOG annotation	9974	4564	420	5410	1065
Total number of transcripts	34403	15310	3216	19093	3762

Annotation success:

A total of 15150 transcript-clusters, 44% of the transcriptomes, matched to known proteins. Functional annotation based on euKaryotic Orthologous groups (KOG) could be provided for 22 % (7689) of the transcriptome (see Figure 3-8). However, 228 transcript-sequences had only KOG annotation but no matches in protein databases. The total number of transcripts having KOG annotations was 9974.
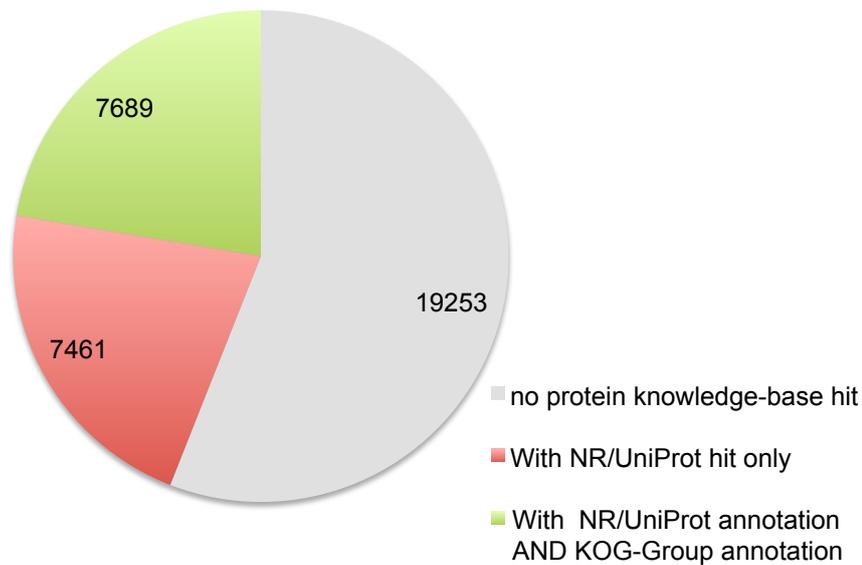
Figure 3-8: Transcript annotation success. The degree of non-redundant and UniProt protein and euKaryotic Orthologous groups (KOG) annotation in the combined transcriptomes of HC and LC conditions is presented.

KEGG-Pathway annotations:

Annotations of KEGG-pathways only provided annotations for 2748 recruited transcripts. A figure presenting the transcripts abundance for each KEGG pathway class can be found in the Appendix A of this monograph.

Significantly different expressed genes:

Between the HC and LC condition 6978 transcripts (20 %) out of the total 34403 transcripts were significantly different expressed.
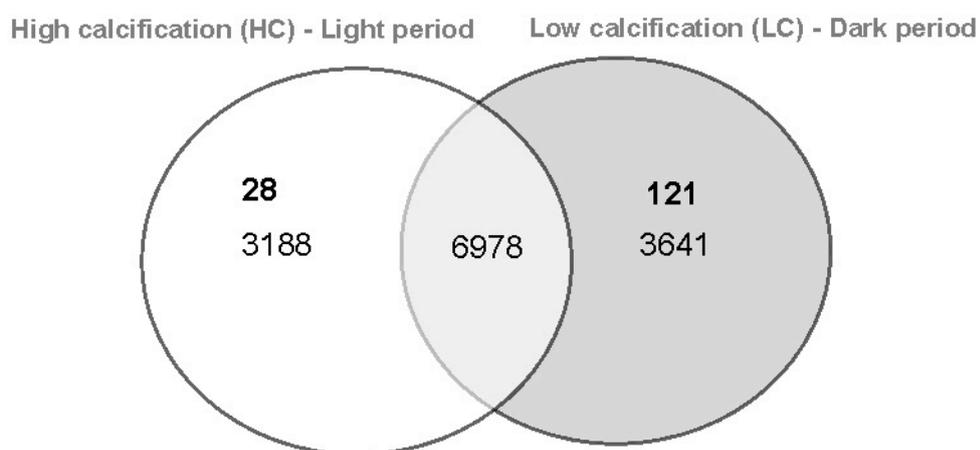
Figure 3-9: Significantly different and unique transcripts. Number of significantly different transcripts found in the transcriptomes of *E. huxleyi* in the HC and LC conditions are shown together with transcripts that were unique (bold letters) for the HC and LC conditions.

Table 3-7: Annotation success of significantly different transcripts. Number of transcripts found in the transcriptomes of *E. huxleyi* in the HC and LC conditions are presented in respect to counts of significantly different transcripts at the q-value threshold < 0.05 and numbers of transcripts with annotation information.

Category	Total	HC	LC	Significant
Total number	34403	15310	19093	6978
Significantly different (q =< 0.05)	6978	3216	3762	
Unique transcripts (q =< 0.05)		28	121	
All unique transcripts		153	177	
NR/UniProt/SwissProt hit	15150	7099	8051	2631
Sig. with NR/UniProt/SwissProt hit	2631	1332	1299	

A total of 3216 transcripts were significantly more frequent in the light (high calcification condition), of which 28 transcripts were unique for the HC condition. In the dark (low calcification condition) 121 transcripts were unique and 3641 transcripts significantly more abundant (compare Figure 3-9 and Table 3-6). Out of the total of 15150 transcripts with protein annotation 47% were higher expressed in the HC and 53% were higher expressed in the LC condition. Only 2631 of the protein annotated transcripts showed significant differences, of which 1332 were significantly higher expressed in the HC condition and 1299 significantly higher expressed in the LC condition (compare Table 3-7).

3.3.3.5.   Functional differences of the *Emiliania huxleyi* transcriptome in the HC and LC condition based on euKaryotic Orthologous groups (KOG) annotations

The euKaryotic Orthologous groups (KOG) annotations provided the highest number of functional annotations represented with 9974 transcripts. Therefore, KOG-groups and KOG-classes were used to investigate general functional features of the transcriptomes at the HC and LC condition. To evaluate differences in functional features between the HC and the LC condition the number of transcripts for all featured KOG-groups and KOG-classes were counted in respect to predominance in the HC or LC condition of the G1-phase of the cell cycle and their significant different overexpression in the HC or LC condition.

KOG-Group level:

In this study it was observed that higher numbers of transcripts were more frequent in the dark period (compare Table 3-6). This was also reflected by the counts of transcripts with KOG-groups and KOG-class annotations. At the functional level of KOG-group no group showed pronounced presence of genes relative to the condition and total ratio of transcript predominance (see Table 3-8).

Table 3-8: KOG-group association of transcriptomes. Numbers of transcripts annotated with KOG-groups, for the total population, transcripts overexpressed in the Light or Dark condition, and significantly different transcripts in the HC or LC condition.

KOG-Group	Total	Overall at		Significant at	
		HC	LC	HC	LC
Cellular processes and signalling	3357	1543	1814	277	343
Information storage and processing	2190	1081	1109	157	198
Metabolism	2242	934	1308	319	208
Poorly characterized	2185	1006	1179	178	205

The highest number of transcripts (3357) was found in the "Cellular processes and signalling" group and provided 34 % of the transcripts, combined in the HC and LC condition. Significantly more frequent transcripts in the LC and HC condition of the KOG-group 'cellular processes and signalling' transcripts contributed 18 % and 17 % (343 and 277), respectively (compare Figure 3-10). Twenty-two per cent of the transcripts in the HC and LC condition were related to 'Information storage and processing'. However, higher rates of significant gene expression were found in the Dark condition. A portion of 22 % of the gene transcripts attributed to the KOG-group 'Metabolism', which were slightly more frequent in the dark (58 % of 2242). Only 25 % of the 'Metabolism' gene transcripts were significantly more abundant in the light (HC condition) compared to 30 % significantly more abundant in the LC condition. Another 22 % of the gene transcripts were 'Poorly characterized'.
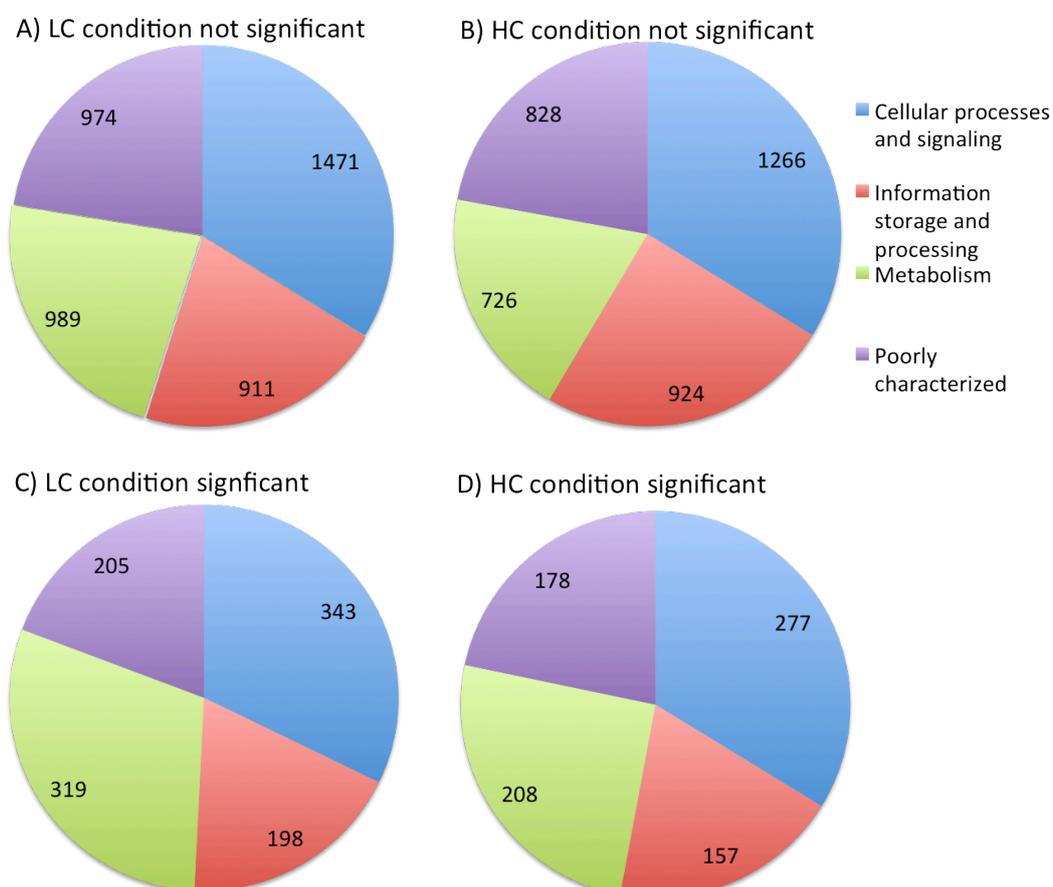
Figure 3-10: KOG-group association of transcriptomes. Number of transcript hits assigned to KOG-groups being singificantly or not significantly over-presented in the HC (high calcifcation) or LC (low calcifcation) condition. Significance level was set at q < 0.05.

KOG-Class level:

Functional differences in gene expression between the HC and LC condition were assessed at KOG-class level. The ratio LC:HC of the number of all genes was used as the expected ratio to which the ratio of significantly more abundant transcripts for each KOG-class was compared by means of Chi-square statistics (see Table 3-9 for a summary of the statistical results). Furthermore, the LC:HC ratio being equal or exceeding the arbitrary thresholds of 0.5 or 2 was used to identify relevant KOG-classes for cell functional evaluation (see Table 3-10). The LC:HC ratios of the KOG-classes 'Translation, ribosomal structure and biogenesis', 'Posttranslational modification, protein turnover, and chaperones', and 'Lipid transport and metabolism' showed significant deviation from the expected ratio (Chi-square test, $p < 0.01$, see Table 3-8).

The abundance ratios of the expressed genes from the entire gene population exceeded the thresholds of 2 (higher abundance in the dark incubation period) only in the KOG-class 'Nuclear structure'. However, 'Nuclear structure' expression was even higher for significantly different expressed KOG-class annotated genes, showing a ratio of 4. For significantly different expressed KOG-class annotated genes the ratios exceeded 2 also in the KOG-classes 'Translation, ribosomal structure and biogenesis', 'Nucleotide transport and metabolism', 'Inorganic ion transport and metabolism', 'Extracellular structures', 'Cell cycle control, cell division, chromosome partitioning', 'Carbohydrate transport and metabolism', and 'Amino acid transport and metabolism', showing higher numbers of KOG-class annotated genes to be expressed in the LC condition. The results for significantly more abundant transcripts for each KOG-class by means of Chi-square statistics and the abundance ratios of expressed KOG-class annotated genes for the entire gene population and the significantly higher expressed genes are summarized in Figure 3-11.

Table 3-9: Differences in transcriptomes by KOG-classes. Results of the analysis for significant differences between the expected LC:HC ratio from the whole transcriptome population and the observed LC:HC ratio of significantly more abundant transcript in each KOG-class from Chi-Square tests with Yates-correction (p<0.01) testing. KOG-classes that showed significantly different LC:HC ratios were marked bold. The transcript abundances in each KOG-Class can be found in Table 3-10.

KOG class	All genes vs. sig. KOG-Class
Translation, ribosomal structure and biogenesis	**<0.001**
Transcription	0.812
Signal transduction mechanisms	0.224
Secondary metabolites biosynthesis, transport and	0.521
RNA processing and modification	0.045
Replication, recombination and repair	0.023
Posttranslational modification, protein turnover	**0.001**
Nucleotide transport and metabolism	0.313
Nuclear structure	0.099
Lipid transport and metabolism	**0.009**
Intracellular trafficking, secretion, and vesicular transport	0.101
Inorganic ion transport and metabolism	0.060
General function prediction only	0.608
Function unknown	0.016
Extracellular structures	0.022
Energy production and conversion	0.949
Defence mechanisms	0.176
Cytoskeleton	0.195
Coenzyme transport and metabolism	0.926
Chromatin structure and dynamics	0.873
Cell wall/membrane/envelope biogenesis	0.755
Cell motility	0.386
Cell cycle control, cell division, chromosome partitioning	0.012
Carbohydrate transport and metabolism	0.023
Amino acid transport and metabolism	0.032

Table 3-10: KOG-class association of transcriptomes. Total numbers of transcripts and significantly higher abundant transcripts in the light and dark incubation period (HC and LC condition), which were KOG-class annotated, are presented. The calculated ratios of transcript numbers LC: HC for the total transcripts and significantly higher expressed transcripts are given. The KOG-classes are marked bold upon exceeding either the arbitrary expression ratio thresholds of 2 or 0.5. Asterisks at the KOG-class name indicate the significant difference of gene expression ratios (p < 0.01) according to Table 3-9.

KOG class	LC condition Not sig.	sig.	Unique	HC condition Not sig.	sig.	unique	Total ratio LC:HC	Ratio Sig. LC:HC
Translation, ribosomal structure and biogenesis *	**155**	**58**	**1**	**262**	**16**	**0**	**0.8**	**3.6**
Transcription	292	74	0	240	56	0	1.2	1.3
Signal transduction mechanisms	387	112	1	271	73	4	1.4	1.5
Secondary metabolites biosynthesis, transport and catabolism	71	33	2	56	32	0	1.2	1.0
RNA processing and modification	274	32	1	269	42	0	1.0	0.8
Replication, recombination and repair	100	15	0	80	26	0	1.1	0.6
Posttranslational modification, protein turnover chaperones *	480	75	0	491	101	0	0.9	0.7
Nucleotide transport and metabolism *	**37**	**7**	**0**	**50**	**2**	**0**	**0.8**	**3.5**
Nuclear structure	**16**	**12**	**0**	**11**	**3**	**0**	**2.0**	**4.0**
Lipid transport and metabolism *	130	27	0	132	42	3	0.9	0.6
Intracellular trafficking, secretion, and vesicular transport	167	19	0	141	26	0	1.1	0.7
Inorganic ion transport and metabolism	**111**	**49**	**0**	**60**	**24**	**0**	**1.9**	**2.0**
General function prediction only	711	163	2	546	122	1	1.3	1.3
Function unknown	263	42	1	282	56	0	0.9	0.8
Extracellular structures	**143**	**42**	**1**	**106**	**17**	**1**	**1.5**	**2.5**
Energy production and conversion	122	33	2	127	26	1	1.0	1.3
Defence mechanisms	16	0	1	20	3	0	0.7	0.0
Cytoskeleton	207	73	2	183	45	3	1.2	1.6
Coenzyme transport and metabolism	67	18	0	46	14	0	1.4	1.3
Chromatin structure and dynamics	90	19	0	73	17	0	1.2	1.1
Cell wall/membrane/envelope biogenesis	51	10	0	39	6	0	1.4	1.7
Cell motility	4	0	0	4	2	0	0.7	0.0
Cell cycle control, cell division, chromosome partitioning	**63**	**26**	**1**	**39**	**7**	**0**	**1.9**	**3.7**
Carbohydrate transport and metabolism	**213**	**63**	**0**	**117**	**31**	**0**	**1.9**	**2.0**
Amino acid transport and metabolism	**175**	**63**	**1**	**99**	**30**	**1**	**1.8**	**2.1**
Not annotated *	10896	2697	1	8250	2396	0	1.3	1.1
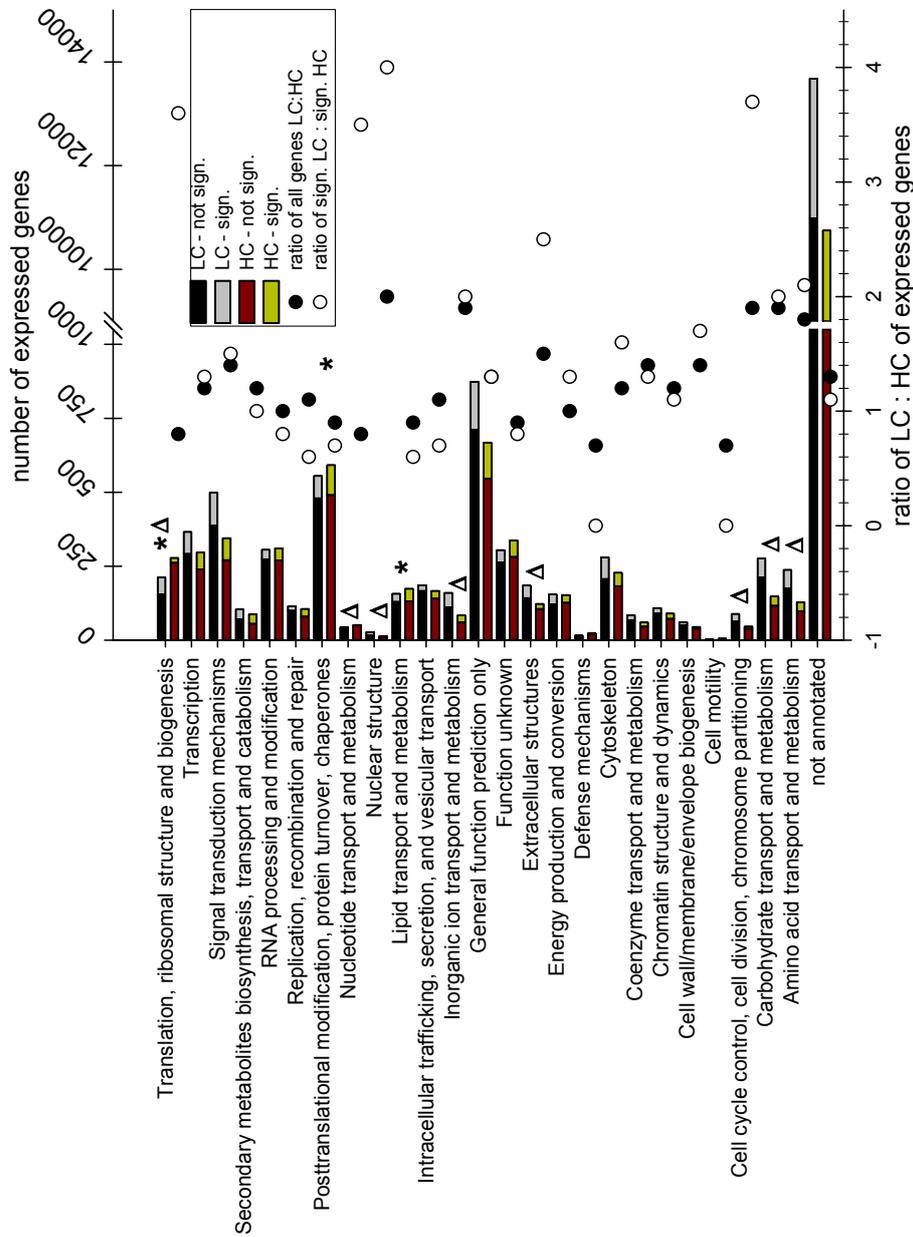
Figure 3-11: Differences of transcriptomes by KOG-classes. Number of transcripts (top x-axis) per KOG-class vindicating significant and not significant differences in LC and HC conditions (see legend) are presented. Significant differences of the observed transcript abundance ratio LC:HC of KOG-class against entire transcriptomes (p < 0.01, Chi-Square test with Yates-correction as in Table 3-8) is indicated by asterisks. (Δ) Indicating expression ratios LC : HC exceeding the threshold of 0.5 or 2 (see Table 3- 9).

Comparison of unique and significantly higher expressed transcripts in the high calcification (HC) and low calcification (LC) conditions of the G1 cell-cycle phase

The number of total unique transcripts and unique significant transcripts varied between the conditions. In total 153 transcripts were unique in the HC condition and 177 transcripts were unique in the LC condition. For a large number of unique transcripts no annotation data was available. Only 28 unique transcripts were significantly increased at HC and 121 unique transcripts were significantly increased in the LC condition. For the HC condition only 9 of the significant transcripts matched known proteins. In total, gene annotations for 54 significantly overexpressed and unique transcripts in the LC condition were retrieved. Further details of the annotation success of unique transcripts are given in Table 3-11.

Table 3-11: Unique transcripts in the light and dark and their annotation success. Number of the light and dark condition with and without annotation (KOG-desc.: KOG-description, Sig.: significant, Not sig.: not significant)

	HC conditions			LC condition			Grand total
	Total	Not sig.	Sig	Total	Not sig.	Sig	
Sequenced transcripts	153	125	28	177	56	121	330
Recruited genes	82	73	9	97	46	51	179
With KOG-desc.	14	12	2	16	6	10	30
With protein annotation	40	34	6	54	27	27	94

Unique significant transcripts in the high calcification in the G1-phase condition

The three most frequent significantly unique transcripts in the HC condition with known annotation were similar to a proton-coupled amino acid transporter 3 (JGI♯ 225551, UniProt ID ID: Q4V8B1, q-value: 0.0096), a pentatricopeptide repeat-containing protein At2g18940 (JGI♯ 205977, UniProt ID: Q64624, q-value: 0.1284), and a putative homeobox protein R749 (JGI♯ 250018, UniProt ID: Q5UP03, q-value: 0.01298). The pentatricopeptide repeat-containing protein At2g18940 was also unique and significantly transcribed by the JGI♯ 249953. For further details of unique significant transcripts in HC condition refer to Table 3-12. The most abundant transcript (JGI♯ 352298) only found at HC had an unknown function. The full list of unique transcripts in the HC conditon is given in the Appendix A Table 2, including transcripts not being significantly different, without annotation data or matching genes in the *E. huxleyi* reference genome.

Unique significant transcripts in the low calcification in the G1-phase condition

In the LC condition the three most significantly and uniquely abundant transcripts were a probable fatty acid methyltransferase (JGI♯ 454253, UniProt ID: P31049, q-value: 1.18 10-16), a transcript recruiting to gene JGI♯ 446057, q-value: 2.54 10-15 without a known protein, and a putative uncharacterized protein (JGI♯ 205338, UniProt ID: D8LH42, q-value: 2.88 10-15) with a suggested signal transducer activity inferred from electronic annotation. For further details of unique significant transcripts at LC refer to Table 3-13. The full list of unique transcripts found in the LC condition treatment is given in the Appendix A Table 3, including transcripts without being significantly different, having annotation data provided or recruiting to known genes in the *E. huxleyi* reference genome.

Table 3-12: Significantly higher abundant transcripts unique in the high calcification (HC) condition of the G1-phase of the *E. huxleyi* cell cycle, with match of sequence to the reference genes (Sig.: if significant).

Gene	Transcripts at HC [fpkm]	q_value	Sig.	UniProt ID	Function
225551	8.63	0.0096	yes	Q4V8B1	Proton-coupled amino acid transporter 3
205977	2.48	0.0128	yes	O64624	Pentatricopeptide repeat-containing protein At2g18940
250018	14.50	0.0130	yes	Q5UP03	Putative homeobox protein R749
249953	1.67	0.0153	yes	O64624	Pentatricopeptide repeat-containing protein At2g18940
201176	5.97	0.0204	yes	-	-
435200	1.12	0.0241	yes	Q6LM37	Flavohemoprotein
352298	44.25	0.0293	yes	-	-
98132	1.83	0.0388	yes	A0PLM4	Transcriptional regulatory protein
369373	1.71	0.0390	yes		*Serine/threonine protein kinase

* based on KOG definition

Table 3-13: Significantly higher abundant transcripts unique in the low calcification (LC) condition of the G1-phase of the *E. huxleyi* cell cycle with match of sequence to the reference genes (Sig.: if significant).

Gene	Transcripts at LC [fpkm]	q_value	Sig.	UniProt ID	Function
454253	24.87	1.18E-16	yes	P31049	Probable fatty acid methyltransferase
446057	24.85	2.54E-15	yes	-	-
205338	271.96	2.88E-15	yes	D8LH42	Putative uncharacterized protein
439818	13.28	1.54E-14	yes	-	*FOG: Zn-finger
435479	12.24	5.90E-13	yes	-	-
309174	27.64	4.28E-12	yes	Q8MTI2	Putative surface protein bspA-like
196275	6.57	8.81E-11	yes	Q9N3Z3	Serine/threonine-protein kinase chk-1
450646	42.00	1.12E-10	yes	-	-
455889	7.65	3.08E-09	yes	-	*Collagens (type IV and type XIII)
208723	10.01	7.65E-09	yes	-	-
459932	5.50	1.07E-07	yes	F0Y858	Putative uncharacterized protein
208470	4.09	2.10E-07	yes	Q54Y32	MAP kinase phosphatase with leucine-rich repeats protein 3
456341	5.24	2.94E-07	yes	F0Y858	Putative uncharacterized protein
456791	4.31	4.70E-07	yes	-	-
450077	4.62	7.99E-07	yes	-	-
435146	5.27	1.18E-06	yes	-	-
456340	4.40	1.27E-06	yes	P55176	UPF0012 hydrolase in pqqF 5'region
240013	4.19	1.42E-06	yes	-	-
258092	16.09	3.01E-06	yes	B7G3J7	Formamidase-like protein
221272	3.30	3.08E-06	yes	A9A494	Porphyromonas-type peptidyl-arginine deiminase
217995	2.94	2.40E-05	yes	C1EIL5	Predicted protein
233951	2.78	6.33E-05	yes	F2TWL3	Putative uncharacterized protein

96

Table 3 – 13 continued

Gene	Transcripts at LC [fpkm]	q_value	Sig.	UniProt ID	Function
207272	4.43	7.15E-05	yes	F0Y048	Putative uncharacterized protein
454351	3.58	9.21E-05	yes	B0G143	Mitochondrial substrate carrier family protein ucpB
247181	3.97	0.00015	yes	B7G168	Predicted protein
352052	6.62	0.00017	yes	-	-
249502	100.07	0.00028	yes	-	-
218752	1.60	0.00031	yes	A9A494	Porphyromonas-type peptidyl-arginine deiminase
243061	8.94	0.00041	yes	-	-
109294	2.55	0.00107	yes	-	-
222516	3.89	0.00111	yes	Q6DGV7	Mpv17-like protein 2
237168	2.44	0.00123	yes	-	-
251402	2.18	0.00143	yes	-	-
461189	2.58	0.00150	yes	Q9HYH5	Putative aldolase class 2 protein PA3430
117279	0.96	0.00169	yes	-	*LRR-containing protein
113812	0.90	0.00227	yes	Q99NF1	Beta,beta-carotene 9',10'-oxygenase

Table 3 – 13 continued

Gene	Transcripts at LC [fpkm]	q_value	Sig.	UniProt ID	Function
424057	4.73	0.00258	yes	-	-
232962	2.56	0.00335	yes	Q5TZ51	Protein Mpv17
248076	5.33	0.00455	yes	Q8NUI5	Lipase 1
211836	3.71	0.00663	yes	Q10977	Phthiocerol synthesis polyketide synthase type I PpsA
113553	2.20	0.00820	yes	Q99NF1	Beta,beta-carotene 9',10'-oxygenase
248875	0.88	0.01044	yes	P55176	UPF0012 hydrolase in pqqF 5'region
211047	1.92	0.01163	yes	Q5F371	Protein strawberry notch homolog 1
359963	1.63	0.01301	yes	-	-
103163	1.15	0.01364	yes	-	-
232178	1.77	0.01490	yes	C1DZB7	Predicted protein
249876	1.54	0.01581	yes	-	-
244412	0.97	0.01944	yes	A9G2Z3	Putative membrane protein
255132	0.88	0.02808	yes	-	-
229436	1.88	0.03236	yes	-	-
369188	1.96	0.04740	yes	-	-

* based on KOG-definition

Transcripts from each KOG-class selected by means of graphical distinctiveness

A graphical presentation approach was used to identify distinct significantly more abundant transcripts in the high calcification (HC) and low calcification (LC) condition of the G1-phase of the *E. huxleyi* cell cycle. Hence, the frequency of the transcripts of each KOG - class was plotted for the light (x-axis; HC condition) and dark (y-axis; LC condition) and are presented (see Figure 3-12 to Figure 3-16). From the figures significantly more abundant transcripts were extracted and presented in the Appendix A Table 4. The focus was to extract transcripts from genes, which are potentially involved in the calcification processes in *E. huxleyi*. Fifteen transcripts potentially related to calcification in *E. huxleyi* were identified (see Table 3-14). Ten of the identified transcripts were significantly more frequent in the HC condition and only 5 were significantly more frequent in the LC condition.

The three most abundant transcripts in the HC condition were from the clusters recruiting to calreticulin (JGI♯ 426711, UniProt ID: P15253, q - value: 0.031, 1206 fpkm in the HC condition, log2fold-value: -3.1), GPA - calcium binding protein (JGI♯ 431830, UniProt ID: Q0MYW8, q - value: $1.29 \ 10^{-5}$, 1026 fpkm in the HC condition, log2fold - value: -6.16), and a synaptobrevin-B (JGI♯ 444996, UniProt ID: Q54GB3, q - value: 0.031, 561 fpkm in the HC condition, log2fold-value: -2.84). The three most abundant transcripts in the LC condition were from clusters recruiting to a sodium/potassium/calcium exchanger 1 (NCKX1) (JGI♯ 447939, UniProt ID: Q9QZM6, q - value: $1.77 \ 10^{-5}$, 897 fpkm in LC condition, log2fold-value: 4.03), a probable sodium/potassium/calcium exchanger CG1090 (JGI♯ 354606, UniProt ID: Q9VN12, q-value: 0.008, 553 fpkm in LC condition, log2fold -value: 2.62), and a vacuolar cation / proton exchanger 5 (CAX5) (JGI♯ 416800, UniProt ID: Q8L783, q - value: $7.1 \ 10^{-5}$, 549 fpkm in LC condition, log2fold - value: 2.7). Further transcripts of the HC and LC condition identified by graphical means and potentially involved in the coccoliths production processes are listed in Table 3-43 and Appendix A Table 4.
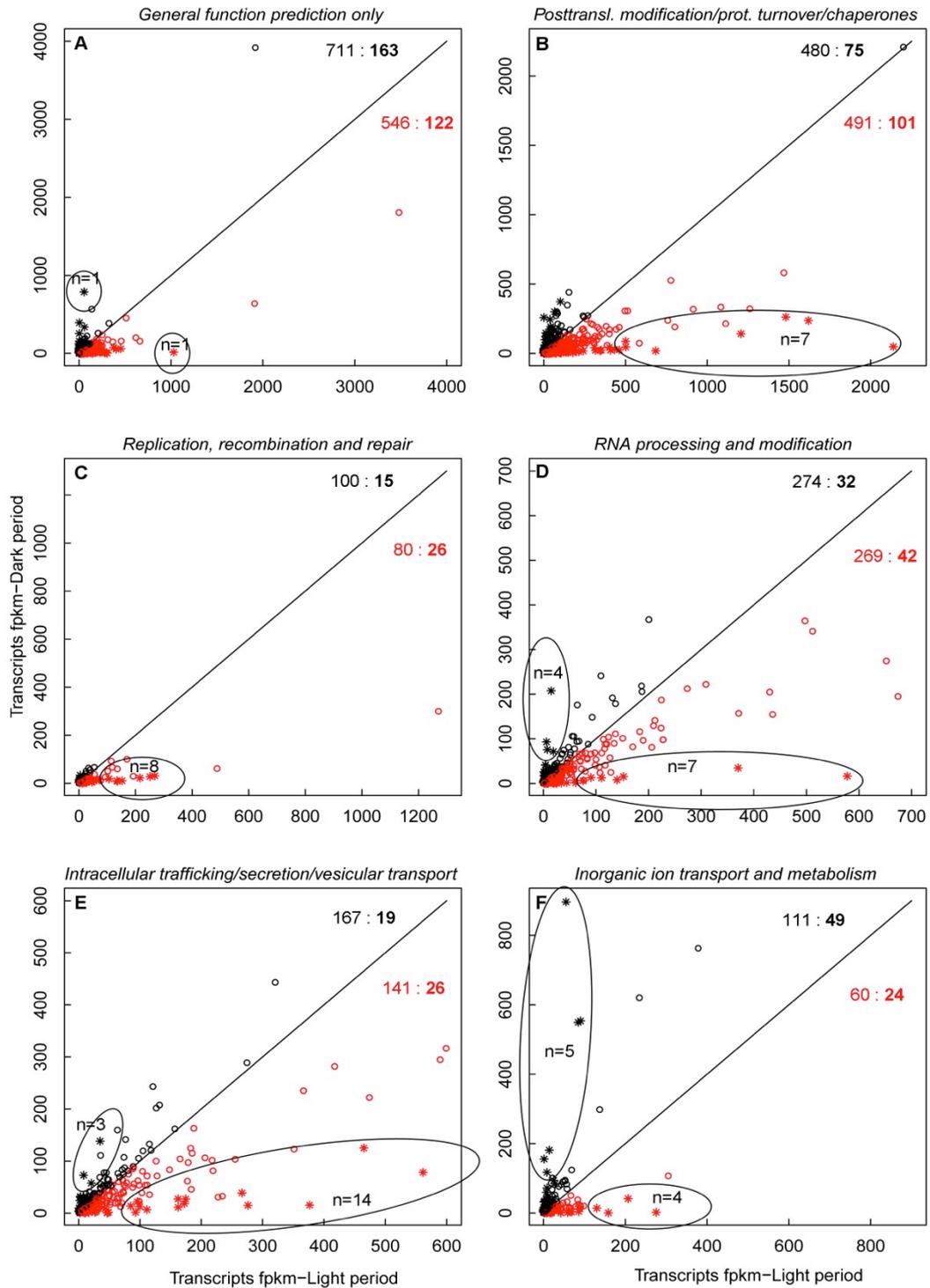
Figure 3-12: Transcript abundance by KOG-class (1-6) in the light (HC condition) (x-axis) against transcript abundance in the dark (LC condition) (y-axis) for the KOG-classes A) General function prediction only, B) Posttranslational modification, protein turnover and chaperones, C) Replication, recombination and repair, D) RNA processing and modification, E) Intracellular trafficking, secretion, and vesicular transport, and F) Inorganic ion transport and metabolism. Small circles show not significant transcripts and small asterisks show significant transcripts. In red: transcripts overrepresented the HC condition and in black transcripts overrepresented in the LC condition. The ratios in red and black show the total number of non-significant and significant transcripts (bold) for the HC and LC condition, respectively. Clusters of significant transcripts of interest are selected by ellipses.
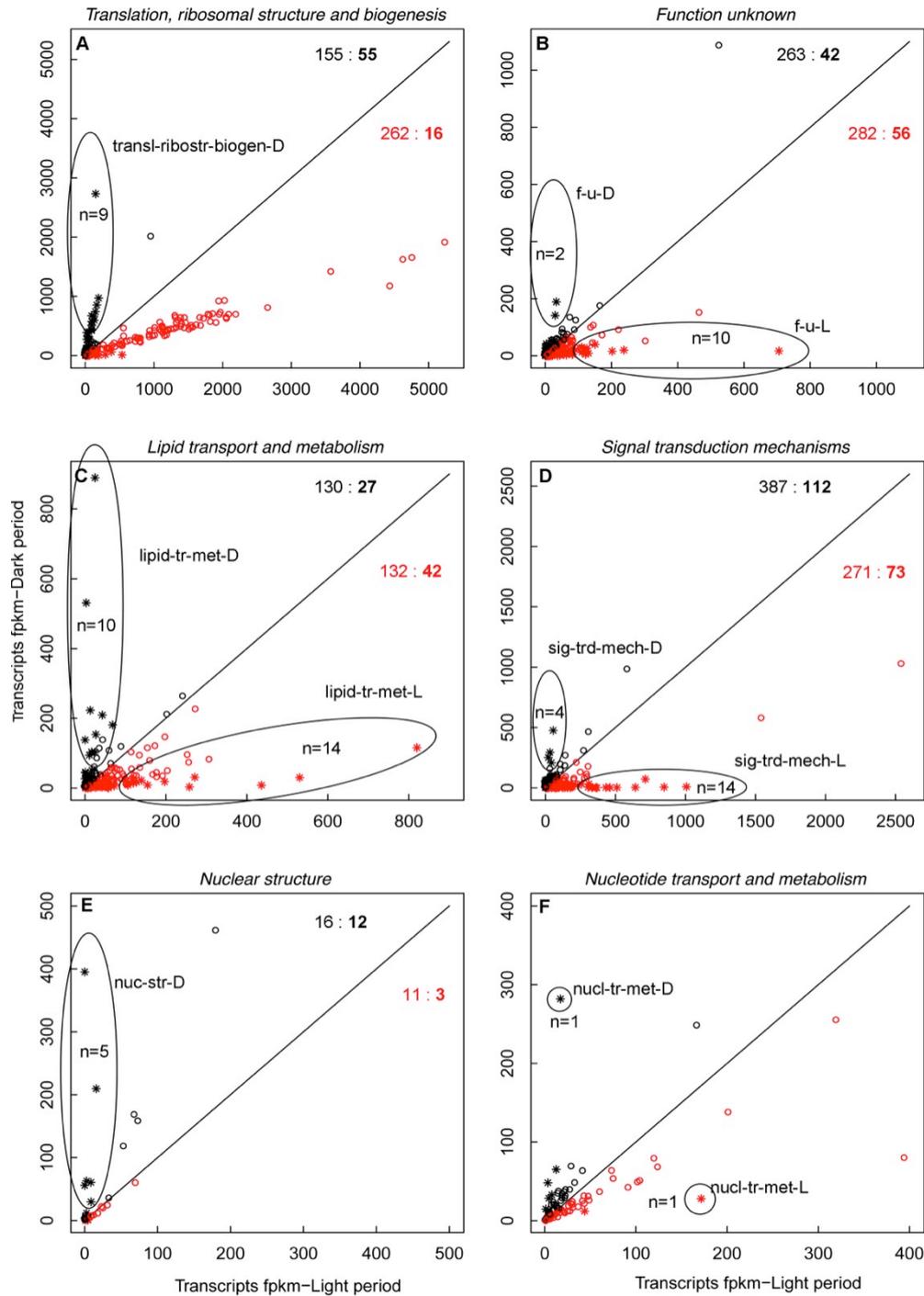
Figure 3-13: Transcript abundance by KOG class (7-12) in the light (HC condition) (x-axis) against transcript abundance in the dark (LC condition) (y-axis) for the KOG-classes A) Translation, ribosomal structure and biogenesis, B) unknown Function, C) Lipid transport and metabolism, D) Signal transduction mechanism, E) Nuclear structure, and F) Nucleotide transport and metabolism. Small circles show not significant transcripts and small asterisks show significant transcripts. In red: transcripts overrepresented the HC condition and in black transcripts overrepresented in the LC condition. The ratios in red and black show the total number of non-significant and significant transcripts (bold) for the HC and LC condition, respectively. Clusters of significant transcripts of interest are selected by ellipses.
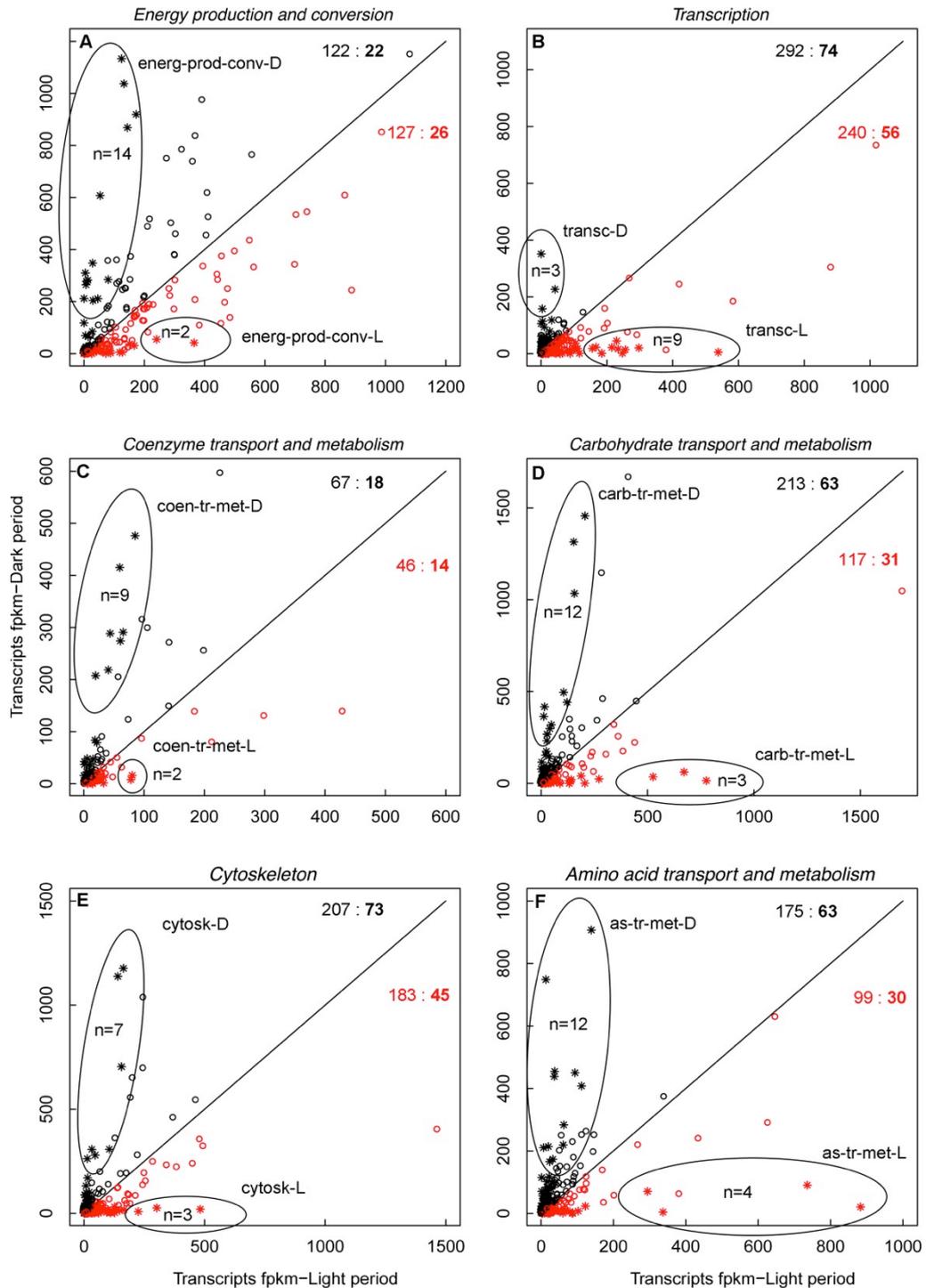
Figure 3-14: Transcript abundance by KOG-class (13-18) in the light (HC condition) (x-axis) against transcript abundance in the dark (LC condition) (y-axis) for the KOG-classes A) Energy production and conversion, B) Transcription, C) Coenzyme transport and metabolism, D) Carbohydrate transport and metabolism, E) Cytoskeleton, and F) Amino acid transport and metabolism. In red: transcripts with abundance ratios towards the HC condition and in black with ratios towards the LC condition. Ratios in red and black show the total number of non-significant and significant transcripts (bold) for the HC and LC condition, respectively. Clusters of significant transcripts of interest are pooled by ellipses.
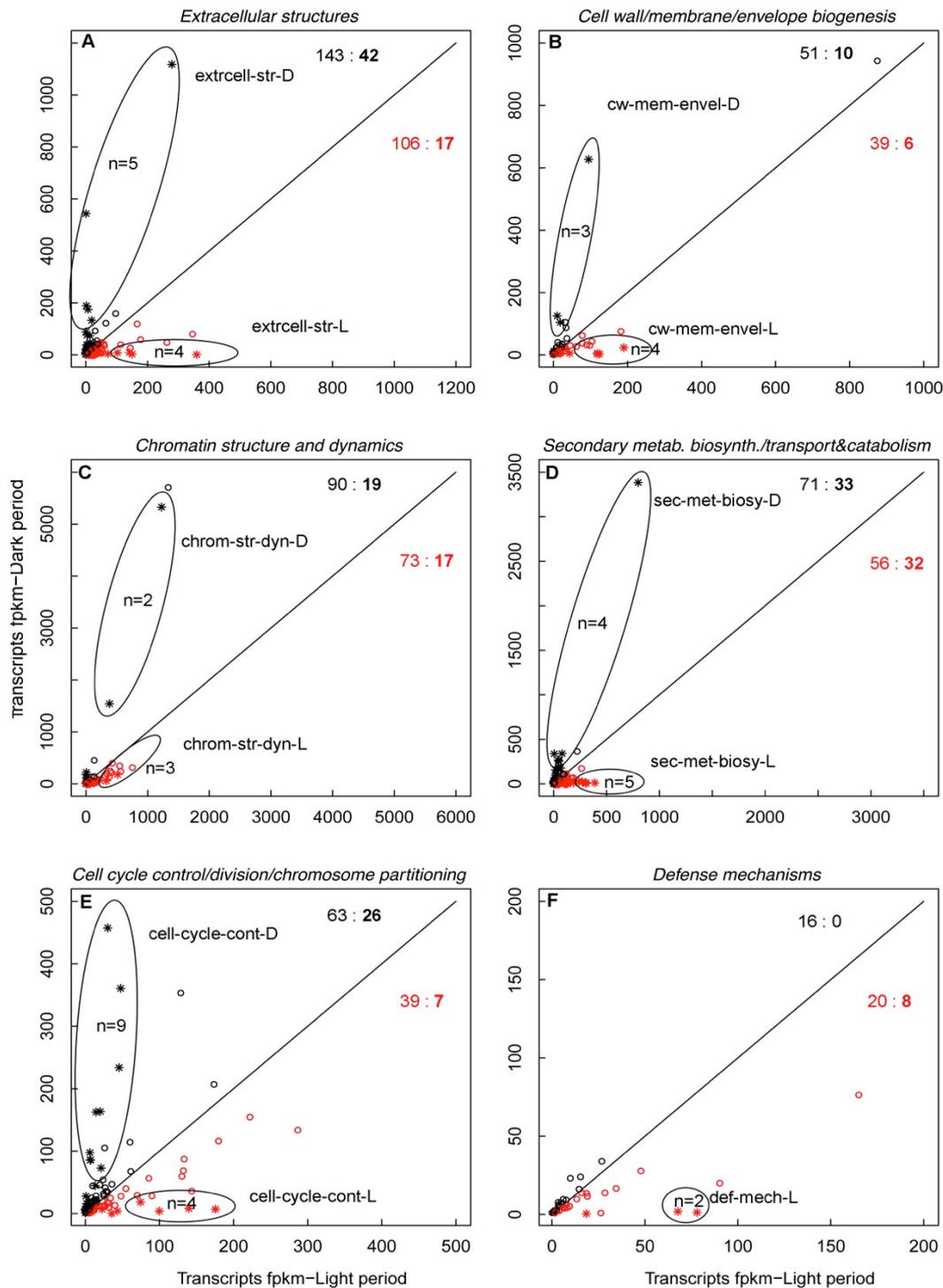
Figure 3-15: Transcript abundance by KOG-class (19-24) in the light (HC condition) (x-axis) against transcript abundance in the dark (LC condition) (y-axis) for the KOG-classes A) Extracellular structures, B) Cell wall, membrane and envelope biogenesis, C) Chromatin structure and dynamics, D) Secondary metabolism, biosynthesis, transport and catabolism, E) Cell cycle control, division, and chromosome partitioning, and F) Defence mechanism. Small circles show not significant transcripts and small asterisks show significant transcripts. In red: transcripts overrepresented the HC condition and in black transcripts overrepresented in the LC condition. The ratios in red and black show the total number of non-significant and significant transcripts (bold) for the HC and LC condition, respectively. Clusters of significant transcripts of interest are selected by ellipses.
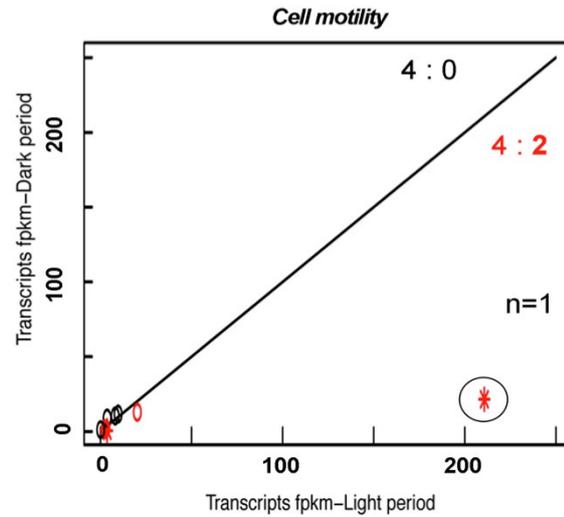
Figure 3-16: Transcript abundance by KOG-class (25) in the light (HC condition) (x-axis) against transcript abundance in the dark (LC condition) (y-axis) for the KOG-classes cell motility. Small circles show not significant transcripts and small asterisks show significant transcripts. In red: transcripts overrepresented the HC condition and in black transcripts overrepresented in the LC condition. The ratios in red and black show the total number of non-significant and significant transcripts (bold) for the HC and LC condition, respectively. Clusters of significant transcripts of interest are selected by ellipses.

Table 3-14: Significantly more abundant transcripts potentially involved in the calcification process of *E. huxleyi* by KOG-Class. Genes were selected after clustering transcripts based on KOG –annotation as presented in Figure 3-12 to Figure 3-16. All genes selected per KOG – class are given in the Appendix A, Table 4.

Condition	JGI gene	fpkm at HC	fpkm at LC	UniProt ID	Function or KOG-description (KOG)
HC	431830	1029	14	Q0MYW8	GPA - Putative calcium binding protein
HC	426711	1206	141	P15253	Calreticulin
HC	444996	561	78	Q54GB3	Synaptobrevin-B
HC	437571	266	39	P47193	Vesicle-associated membrane protein 2,
HC	351006	162	27	O70480	Vesicle-associated membrane protein 4,
HC	463384	100	12	D3Z5L6	MFS-type transporter C6orf192 homolog,
HC	460762	96	22	Q9ZRD6	VAMP-like protein YKT61, Synaptobrevin
HC	434324	85	14	P93654	Syntaxin-22
HC	466232	276	2	Q9HGM6	Putative transporter C543.05c, HCO_3 transport
HC	314659	207	41	Q9HGM6	Putative transporter C543.05c, HCO_3 transport
LC	99733	21	58	Q84WW5	Vesicle-associated protein 1-3
LC	447939	55	897	Q9QZM6	Sodium/potassium/calcium exchanger 1
LC	354606	90	553	Q9VN12	Probable sodium/potassium/calcium
LC	416800	84	549	Q8L783	Vacuolar cation/proton exchanger 5
LC	463095	1	156	Q9XES1	Calcium-transporting ATPase 4, ER-type

3.3.3.7. Transcript abundance in the high and low calcification rate conditions in the G1-phase of the *E. huxleyi* cell cycle of genes with a suggested role in the calcification processes

A selection of molecules, previously reported in the literature to play a role in the processes surrounding calcification and coccolith formation in *E. huxleyi* (Marsh et al. 1992; Wahlund et al., 2004; Nguyen et al., 2005; Quinn et al., 2006; Richier et al., 2009; von Dassow, 2009; Mackinder et al., 2009, 2011, 2012; Emery et al., 2012) are presented in Figure 3-17 to Figure 3-21. The individual figures represent clusters of molecules according to their function and characteristics.
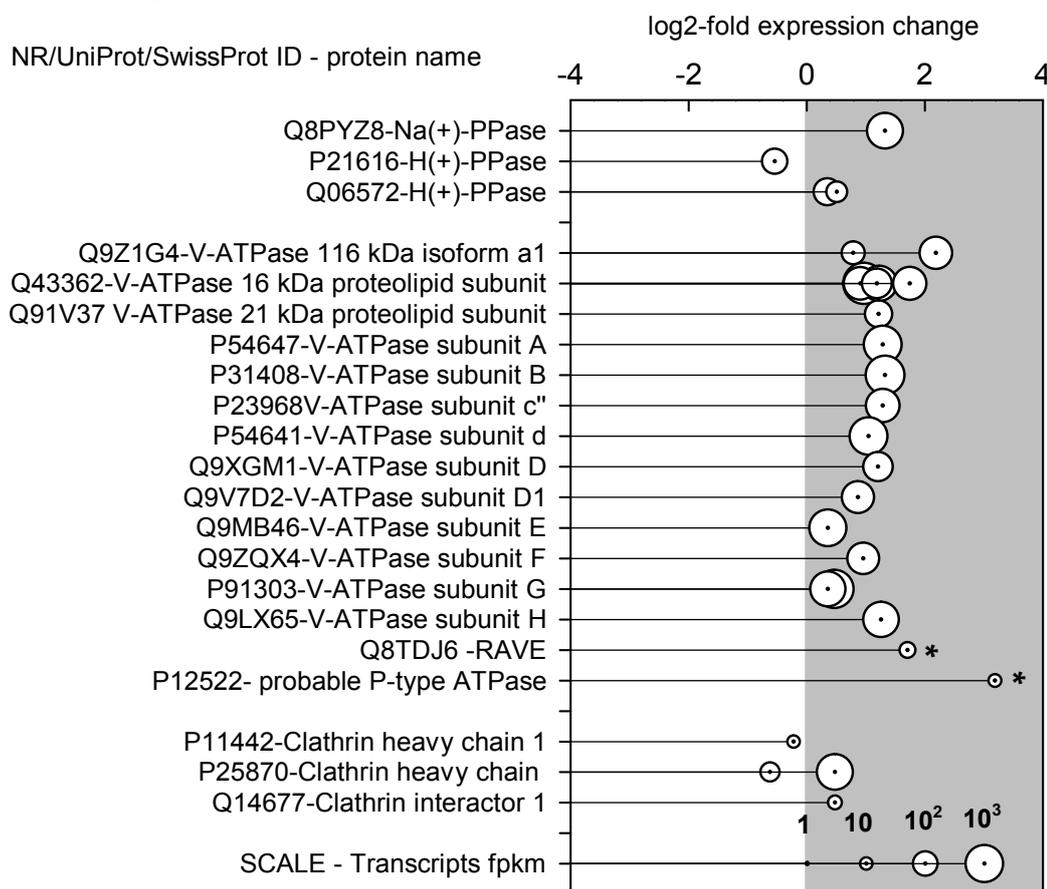
Figure 3-17: Transcript abundance and expression ratios of vacuolar type membrane proton pump ATPases (V-ATPase) in light and dark incubation period (shaded grey), a regulator of V-ATPase assembly (RAVE), and clathrin. Positive 2^X-fold expression values indicate elevated expression in the low calcification condition. The sizes of the circles indicate the total abundance of transcripts found in high and low calcification conditions. The results table can be found in Appendix A Table 5.

Proton pumping pyrophosphatases (PPase), V-type (vacuolar) proton ATPases, P-type proton ATPase, and clathrin expression

V-type proton ATPases and PPases were suggested to paly an important role in the control of the pH in organelles (Mackinder et al., 2010), where calcium carbonate is precipitated and to stabilize the electrochemical gradient across membranes (Suffrian et al., 2011). Transcripts of V-type proton ATPases and PPases with suggested vacuolar location were present in transcriptomes of high and low calcificaion condition (compare Figure 3-17). Higher abundances of transcripts of the V-type proton ATPases and PPases cluster were observed in the low calcification (LC) condition. Three transcripts for pyrophosphate-energized vacuolar membrane proton pumps (Na^+ and H^+-PPases) were found more abundant in low calcification condition. Two genes expressed the H^+-PPase with the UniProt ID: Q06572. Furthermore, transcripts for 13 peptides of the V-type proton ATPase were present, whereas the 16 kDa proteolipid subunit of the V-type ATPase formed a cluster by 7 genes, the V-type proton ATPase 116 kDa subunit a isoform 1 and probable V-type proton ATPase subunit G formed clusters of two genes (for details of the expression levels of each transcript refer to Appendix A Table 5). Only the P21616 peptide of the H^+-PPase of the described V-ATPase cluster was higher expressed in the high calcification (HC) condition. None of the V-ATPase related transcripts were significantly higher expressed in HC or LC condition. However, the predicted RAVE (regulator of V-ATPase assembly) a complex subunit RAV1/DMX protein of the WD-repeat superfamily, JGI♯ 461741 was significantly higher expressed in the LC condition, showing 3-fold more transcripts in the dark and a total of 21 fkpm. Only the transcript of P12522 being a P-type ATPase 1B was 9-fold significantly higher expressed in the LC condition. The total abundance of transcripts in both periods was 11.15 fpkm.

Clathrin was suggested to play an indirect role in the calcification process by being present in clathrin-coated vesicles and occurring together with V-type proton ATPases. The observed expression of clathrin did show any significant increase in the HC or LC condition. Four genes were expressed and the bulk of transcripts (801 fpkm) matched the clathrin heavy chain P25870, whereas 25 % more were present in the LC condition (see Appendix A, Table 5 for details).

Expression of proton antiporter and proton exchanger and voltage gated proton channels

Further proton transport across membranes may involve sodium proton antiporters, chloride proton exchanger transporters, proton sugar co-transporters, and voltage-gated H^+ channel proteins. The abundance of transcripts of the above-mentioned transporters is presented in Figure 3-18.

Proton exchanger transcript clusters
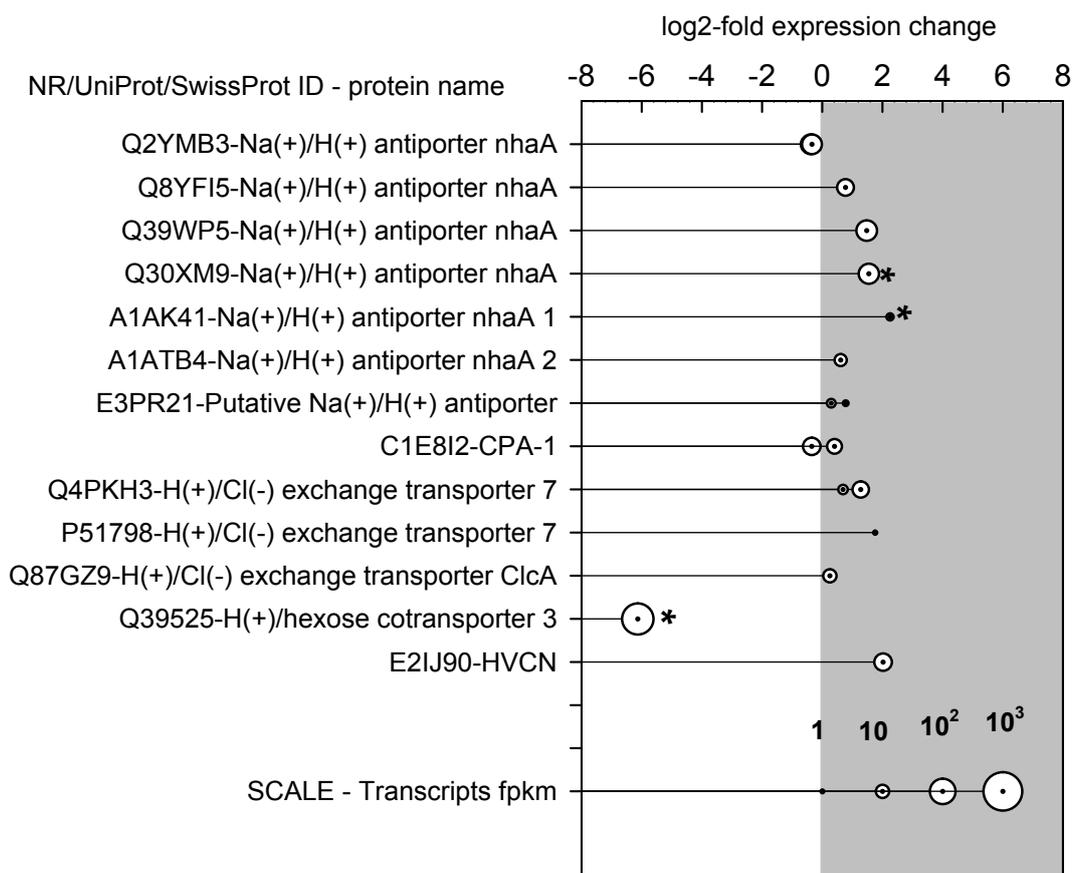
Figure 3-18: Transcript abundance and expression ratios of proton exchanger found at light and dark incubation period (shaded grey). CPA1: monovalent cation: proton antiporter-1 family. HVCN: Voltage-gated H^+ channel protein. Positive 2^X-fold expression values indicate elevated expression in the low calcification condition. The sizes of the circles indicate the total abundance of transcripts found in high and low calcification conditions. The results table can be found in Appendix A Table 6.

Sodium proton antiporters, chloride proton exchanger transporters, proton sugar co-transporters, and voltage-gated H^+ channel proteins were present in the high and low calcification rate conditions of the G1-phase of the *E. huxleyi* cell cycle. Significant differences

for the expression levels between the HC and LC condition were found for transcripts of the Na$^+$/H$^+$ antiporter nhaA protein (UniProt ID Q30XM9) and Na$^+$/H$^+$ antiporter nhaA1 Protein (UniProt ID ID A1AK41) in the LC condition representing 27 fpkm and 2 fpkm transcripts, respectively (compare Appendix A Table 6). In the light incubation period (HC condition) the H$^+$/hexose co-transporter 3 was significantly higher expressed (280 fpkm transcripts).

Observed Ca^{2+} - transporter expression:

Calcium transporters are an important part of cell signalling pathways and fundamental for reaching the saturation state for the precipitations of calcium carbonate. Calcium transporters of the K$^+$ dependent Ca^{2+}/Na$^+$ exchanger (NCKX), vacuolar cation/proton exchanger (VCX), cation/proton exchanger (CAX), calcium transporting ATPase – endoplasmic reticulum (SERCA) type, together with plasma membrane calcium-transporting ATPases (PMCA) and voltage dependent calcium channels were found in the transcriptomes at high and low calcification rate conditions of the G1-phase of the *E. huxleyi* cell cycle (compare Figure 3-19). The clusters of transcripts for the NCKX1 protein (UniProt ID: Q9QZM), the probable sodium / potassium / calcium exchanger CG1090 protein (UniProt ID Q9VN12), vacuolar cation/proton exchanger 5 (UniProt ID: Q8L783).), the SERCA- type 4 ATPase (UniProt ID: Q9XES1), the plasma membrane calcium-transporting ATPase 4 (UniProt ID: Q64542), and the voltage-dependent calcium channel subunit alpha-2/delta-2 (UniProt ID Q9NY47) were significantly more expressed in the LC condition. Furthermore, the long transient receptor potential cation channel proteins (UniProt IDs: Q91YD4, P48994, and O94759) were expressed in the LC condition. The *E. huxleyi* calcification specific cation/proton exchanger CAX 3 (von Dassow et al., 2009) was apparently not found in the transcriptomes of the HC and LC conditions. For details about transcript abundance refer to Appendix A Table 7.

Ca^{2+}- transporter transcript clusters
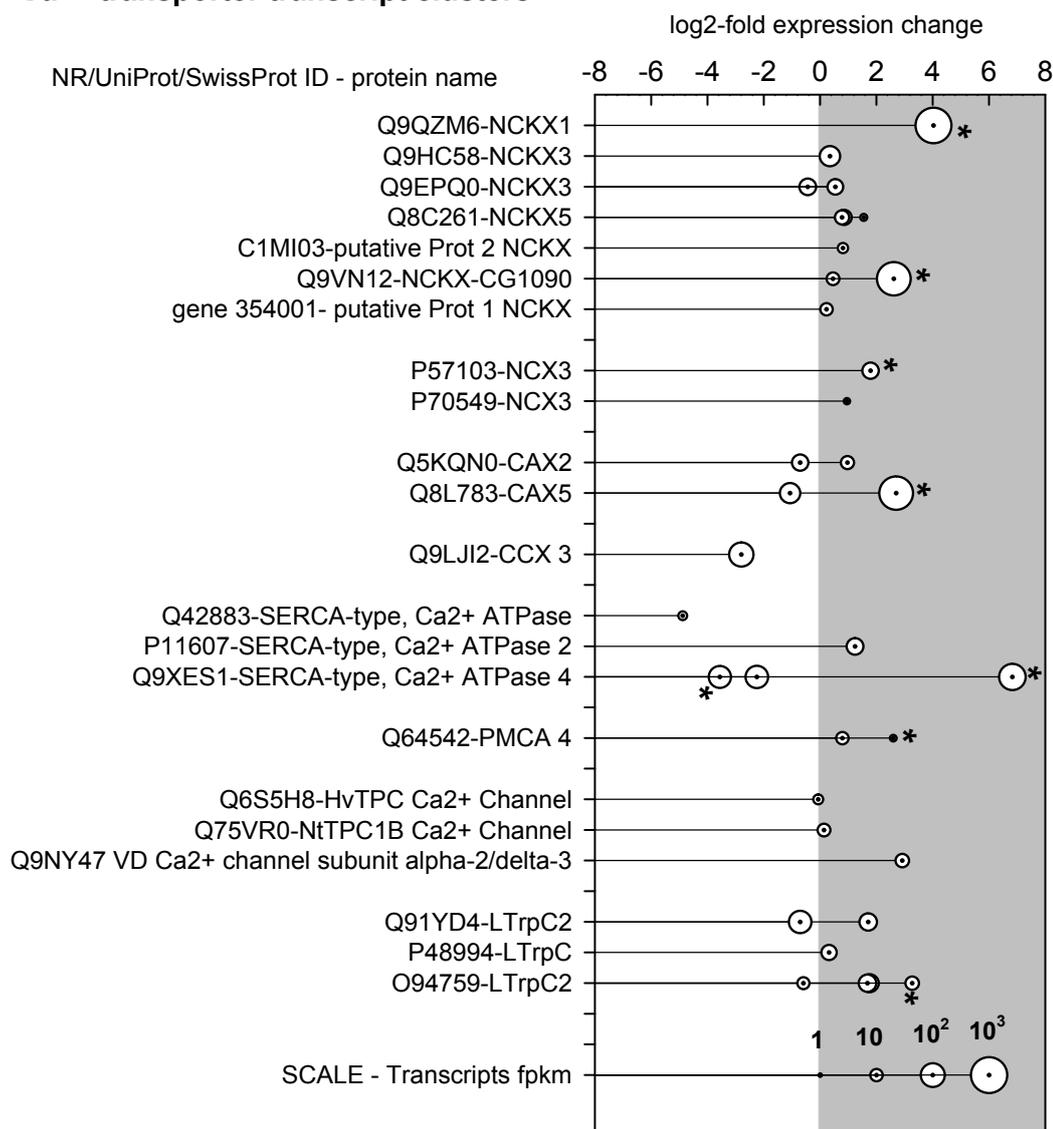
Figure 3-19: Transcript abundance and expression ratios of Ca^{2+} - transport related proteins found at light and dark incubation period (shaded grey). NCKX: K$^+$ dependent Ca^{2+}/Na$^+$ exchanger, VCX: vacuolar cation/proton exchanger, CAX: cation/proton exchanger, SERCA: calcium transporting ATPase – endoplasmic reticulum type, PMCA: plasma membrane calcium-transporting ATPase, VD: voltage dependent, Ca^{2+}/Mg^{2+}-permeable cation channels LTrpC: Long transient receptor potential cation channel. Positive 2^X-fold expression values indicate elevated expression in the low calcification condition. The sizes of the circles indicate the total abundance of transcripts found in high and low calcification conditions. The results table can be found in Appendix A Table 7.

Observed calmodulin and Ca^{2+} - binding proteins expression

Calcium binding proteins play an important part in reaching and maintaining saturated concentrations of calcium ions prior to the precipitation calcium carbonate provide a pool of calcium ions within the cell (Schroeder et al., 2005). The results of the observed transcripts for calcium binding proteins are presented in Figure 3-20.
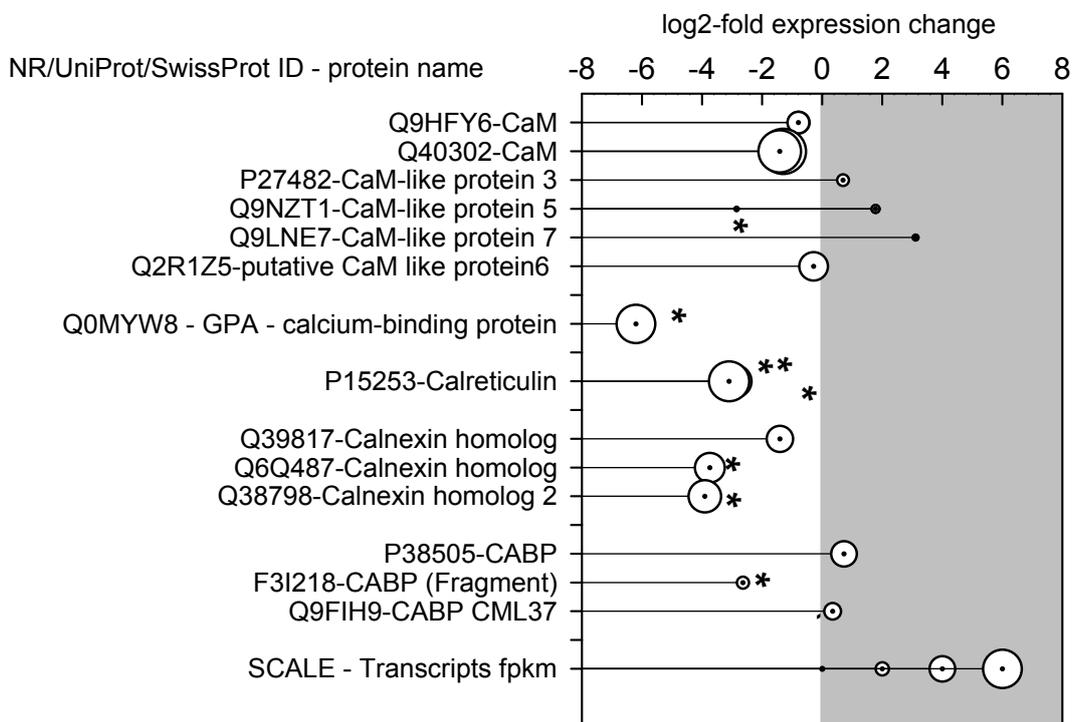
Figure 3-20: Transcript abundance and expression ratios of calmodulin (CaM) and Ca^{2+} - binding proteins found at light and dark incubation period (shaded grey). CABP: calcium binding protein. Positive 2^X-fold expression values indicate elevated expression in the low calcification condition. The sizes of the circles indicate the total abundance of transcripts found in high and low calcification conditions. The results table can be found in Appendix A Table 8.

Calmodulin transcripts were more frequent in the HC condition, namely Q9HFY6, Q40302, and the calmodulin like protein 6 (Q2R1Z5) (compare Figure 3-20 and Appendix A Table 8). Transcripts of calmodulin like proteins were also observed in the low calcification rate condition (LC) of the G1-phase of the *E. huxleyi* cell cycle but at much lower rates. However, the only significant difference in expression rates between the HC and LC conditions was found for clamodulin CaM-like protein 5 (Q9NZT1) showing only 1 fpkm transcripts. In total 4265 fpkm transcripts of the different calmodulin isoforms were found in the HC condition compared

to 1736 fpkm transcripts in the LC condition (see Appendix A Table 8). The known GPA calcium binding protein (Q0MYW8) was significantly high expressed in the cells of the G1-phase showing higher calcification rates in the light incubation period. In the HC 1029 fpkm compared to 14 fpkm in the LC condition of GPA transcripts were observed. Two genes of calreticulin (UniProt ID: P15253) expressed significantly higher transcripts in the HC condition (1516 fpkm) compared to 183 fpkm transcripts in the LC condition. Furthermore, genes of calnexin homologs (UniProt ID: Q39817, Q6Q487, and Q38798) produced significantly more transcripts in the HC condition (596 fpkm) compared to 67 fpkm to the LC condition. Transcripts of the predicted calcium binding protein (UniProt ID: F31218) were also significantly higher expressed in the HC condition.

Observed Bicarbonate transport and carbonate anhydrase expression

The transport of bicarbonate and activity of carbonate anhydrase provides CO_2 and CO_3^{2-} - ions for photosynthesis and calcium carbonate synthesis in *E. huxleyi* (von Dassow et al., 2009; Richier et al., 2011). A total of 20 genes of the *E. huxleyi* genome related to bicarbonate transport were tested for differential expression between the high (HC) and low calcification rate (LC) conditions of the G1-phase of the *E. huxleyi* cell cycle and results are presented in Figure 3-21. Transcripts relating to 9 genes showed significantly different abundance in 4 gene clusters. Gene expression related to bicarbonate transport was significantly higher for 5 out of 9 genes in the LC condition and for 4 out of 9 genes in the HC condition. The genes JGI♯ 99943, 426735, 200137, 198643, and 120259 code for the isoforms of the anion exchanger 2 protein (UniProt ID: P23347) and were higher expressed in the LC condition comprising a total of 726 fpkm, whereas transcripts of the genes JGI♯ 99943, 200137, 198643, and 120259 were significantly more abundant in the LC condition. The putative Na^+-independent Cl^-/HCO^{3+} exchanger AE1 of the SLC4 family (UniProt. ID: Q9HGM6) was more frequent in the HC condition and transcribed by 3 genes. Transcripts recruited to the *E. huxleyi* gene JGI♯ 466232 showed 276 fpkm in the HC condition, which was significantly higher than in the LC condition (2 fpkm). Transcripts recruited to the *E. huxleyi* gene JGI♯ 314659 were also significantly more abundant in the HC condition showing 207 fpkm transcripts in the light incubation period compared to 4 fpkm in the dark incubation period. The sodium bicarbonate co-transporter 3 of the SLC 7 family (JGI gene ♯ 469783, UniProt. ID: Q8BTY2) was significantly higher expressed in the LC condition with 15 fpkm compared to one fpkm transcripts in the HC condition. Furthermore, the uncharacterized vacuolar membrane protein of the SLC 26 family (UniProt. ID: YGR125W) involved in sulphate/bicarbonate/oxalate exchanger was expressed by three genes, of which two were significantly more expressed. Genes relating to the sulphate/bicarbonate/oxalate exchanger were significantly more expressed in both the HC and LC condition, whereas JGI gene ♯ 98125 expression was significant for the HC condition and JGI gene ♯ 460215 expression was significantly increased in the LC condition. The total transcript

abundance of the YGR125W homologs was 34 fpkm in the HC and LC condition. For details of transcript abundance related to bicarbonate transport refer to Appendix A Table 9.

The gene expression differences in the HC and LC condition of seven genes related to carbonic anhydrase were investigated. In general, carbonic anhydrase expression was elevated in the LC (low calcification) condition compared to the HC (high calcification) condition (compare Figure 3-21). Increasingly expressed genes for carbonic anhydrases were found significant in the LC condition. Three homologs of the carbonic anhydrase (UniProt ID: O52535) were present, of which only transcripts recruiting to JGI gene ♯ 233460 were significant in the LC condition. Out of two homologs of carbonic anhydrase UniProt ID Q50940 in the LC condition only transcripts mapping to JGI gene ♯ 62679 were significant. Interestingly, the sequences of most transcripts were mapped to the *E. huxleyi* specific delta-carbonic anhydrase (UniProt ID: Q0ZB86, JGI gene♯ 436031), which were only present in the LC condition with 74 fpkm transcripts. Details for all transcripts related to carbonic anhydrases are given in Appendix A Table 9.

HCO$_3^-$ transporter & carboante anhydrase transcript clusters
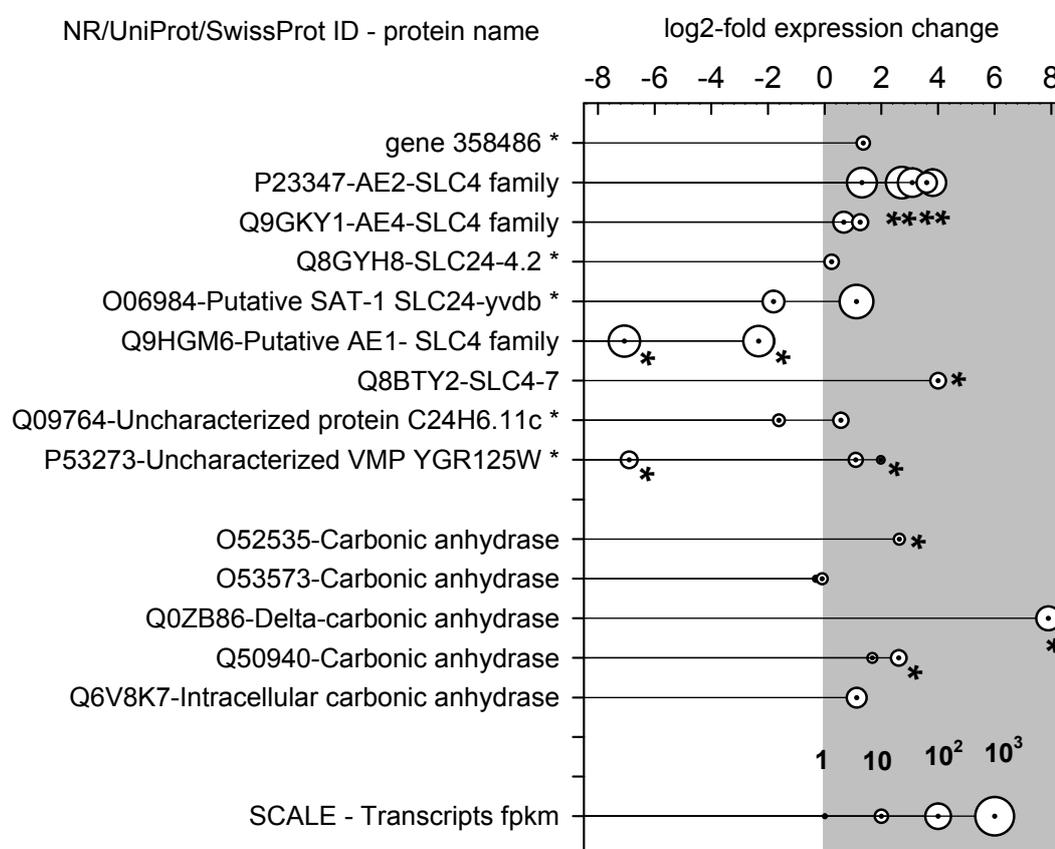
Figure 3-21: Transcript abundance and expression ratios of bicarbonate transport and carbonic anhydrase found at light and dark incubation period (shaded grey). AE: Anion exchanger. SLC: Solute carrier family. VMP: vacuolar membrane protein. Asterisks at the protein name indicate EuKaryotic Orthologous Groups (KOG) definition: Sulphate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family). Positive 2^X-fold expression values indicate elevated expression in the low calcification condition. The sizes of the circles indicate the total abundance of transcripts found in high and low calcification conditions. The results table can be found in Appendix A Table 9.

3.4. Discussion

In this study, experimental cultures of the unicellular coccolithophore *E. huxleyi* PLY♯ M217 were reared in F/50 medium to investigate the transcriptomes at high and low calcification rates of the G1-phase of the cell cycle. Previous studies observed different gene expression in *Emiliania huxleyi* related to life cycle phases, different cell cycle phases, under calcification stimulating by phosphorous limitation, and different carbonate chemistry regimes (e.g. Quinn et al., 2006; von Dassow et al., 2009; Richier et al., 2009; Mackinder et al., 2011; Bach et al., 2013). This study showed that gene expression related to biomineralisation differs significantly within the G1 cell-cycle phase at high and low calcification rates, the light and the dark incubation period, respectively. Interestingly, the expression of proteins, which were previously reported to be involved in coccolithophogenesis, was more frequent in the low calcification phase of the G1-phase. This suggests patterns of gene expression, in which a significant amount and molecules involved in biomineralisation processes, namely ion transporting pathways, are expressed before the calcification machinery maximizes under light exposure. For example, the bulk of V-type ATPases related transcripts (see Appendix A Table 5 for details), around 75% of the cation proton exchanger CAX5 transcripts, the majority of K^+ dependent Ca^{2+} / Na^+ exchanger (NCKX) transcripts, and higher numbers of clathrin transcripts were observed when cells were actually calcifying at low rates. This might suggest that large proportion of the potential biomineralisation proteins is synthesized during or just after mitosis and that a smaller degree of those proteins needs continuous synthesis during periods of highest calcification rates. Furthermore, the observed gene expression patterns also support the general importance of calcium binding proteins in the process of coccolithophogenesis. Calcium binding proteins were increasingly expressed in the high calcification condition of the G1-phase, which could also indicate that many of those calcium-binding proteins are lost during coccolithogenesis. In particular GPA showed 70-fold increase in the light enhanced calcification phase.

The following paragraphs will lead to putting the gene expression results from the transcriptomes of early and late G1-phase *E. huxleyi* cells, in a wider context. First, the experimental methodology, culturing conditions, transcriptome analysis using functional groups and a graphical approach are reassessed and discussed. Eventually existing models of calcification in coccolithophores are discussed with the results from this study taking into account models of the algae cell cycle and cell division.

Methodological considerations of the *E. huxleyi* culture conditions

The culturing conditions, depletion of nutrients and the time point of cell harvesting may have affected the physiological performance and therefore calcification of *E. huxleyi* in the presented study. Using cultures of *E. huxleyi* grown in F/50 medium might have affected the general patterns of gene expression but should not have affected the investigation of gene expression relevant to coccolithophogenesis. The F/50 medium is considered to limit the supply of phosphate to coccolithophores and thereby increase the accumulation of calcium carbonate per cell and cause physiological stress to the cells (Paasche, 2002; Wahlund et al., 2004). Under phosphate limited conditions general gene expression was reduced in *E. huxleyi* compared to gene expression in phosphate-replete conditions (Wahlund et al., 2004b; Nguyen et al., 2005). Phosphate limitation was found to cause gene expression linked to stress and cell defence, among them genes coding for homologs of HSP70 (Wahlund et al., 2004b). Testing the expression of HSP 70 homologs in the transcriptomes of the high calcification period and low calcification period showed that 4 homologs were significantly more frequent in the light period (high calcification condition) (see Appendix A Figure A-2 and Appendix A Table 11), which could indicate phosphate depletion at the time of cell harvesting for transcriptome analysis. The cells require phosphate for a diversity of processes such as the synthesis of DNA and RNA. Under phosphate limited conditions cells might not be able to pass into cell division phase in the dark because at earlier cell cycle check points the cell had not attained its critical size and energy reserves (Zachleder et al., 2016). As a result of prolonged cell cycle and cell growth, more calcium carbonate per cell is accumulated until the cellular reservoirs of phosphate have been filled. Only when phosphate pools were restored the cells may pass through cell division. However, an accumulation of calcium carbonate per cell (PIC increase) and a halting of the cell cycle cannot be concluded from the increase in cell densities and growth rates (Figure 3-3 A and B). The PIC per cell (Figure 3-3 F) remained stable at the termination of the experiment. Furthermore, the POC and PON results did not show any differences between the day before cell harvest and the day of harvest. It was therefore thought that the cells were not altered in their physiology during the experiment and a phosphate limitation was not given. If phosphate was depleted, even though the data suggest otherwise, phosphate limitation should enhance gene expression related to biomineralisation in *E. huxleyi* (Wahlund et al., 2004b; Nguyen et al., 2005; Quinn et al., 2006).

Methodological considerations of the transcriptome analysis

Matching the raw transcript sequences of late and early G1-phase to the *Emiliania huxleyi* genome v.1.0 (April 25, 2008; http://genome.jgi.doe.gov/Emihu1/Emihu1.home.html) identified 34403 transcripts using NGS Illumina sequencing of mRNA. However, only 9999 gene transcripts matched to known proteins suggesting the relatively poor annotation of the *Emiliania huxleyi* genome v.1.0, a large degree of regulatory transcripts, and post-transcriptional modifications. Nevertheless, the present study revealed nearly twice the number of transcripts as reported in the EST analysis by von Dassow et al. (2009) (19188 ESTs; strain RCC 1217). However, von Dassow et al. (2009) reported nearly twice as much transcripts using the strain CCMP 1516, B-morphotype, which showed 72,513 ESTs. From the most recent genome of *E. huxleyi* CCMP1516 it is thought that the *E. huxleyi* genome comprises of 142Mb containing 30,569 protein-coding genes of which 28429 were matched to known transcripts in databases (Read et al., 2013). The results of this study found 9999 transcripts matching the 2008 version of the reference genome but 15150 transcripts matching known protein sequences. Again the comprehensiveness of the *Emiliania huxleyi* genome v.1.0 (April 25, 2008) might have influenced the annotation success of the observed transcripts. At the time of analysis the newly sequenced genome http://www.ebi.ac.uk/ena/data/view/ GCA_000372725.1 (Read et al., 2013) was not available at the time.

The Illumina sequencing and Tophat-cufflinks pipeline found large numbers of differentially expressed transcripts. However, the Tophat-cufflinks pipeline has been criticized for its return of increased number of false positives (Rapaport et al., 2013). The accuracy issues with significant differential gene expression in the Tophat-cufflinks pipeline were addressed by the introduction of the $q-value$, which provides higher correctness (Trapnell et al., 2012) and reduces false positive identifications (Rapaport et al., 2013). Microarray based gene expression analysis applied thresholds of $p < 0.01$ and 2-fold-expression greater than two to identify significantly different transcripts abundance in *E. huxleyi* cultures grown in nutrient replete and nutrient deplete conditions (Quinn et al., 2006). In a simulation, the threshold values of Quinn et al. (2006) were applied to the differential expression results of the Tophat-cufflinks pipeline. Thus, the numbers of significantly different transcripts were reduced by 43% from 6978 significantly different expressed genes ($q < 0.5$) to 3942 significantly different expressed genes ($p < 0.01$; 2-fold expression greater than 2). Quinn et al. (2006) showed that of a total of 2298 ESTs 127 transcripts (5%) showed significantly different up or down regulation. In this study 20% of the transcripts were significantly up or down regulated following the Tophat-cufflinks pipeline and even after applying the additional thresholds ($q < 0.01$ and 2-fold expression greater than 2) 11% of the transcripts were different in cells of the early G1-phase low-calcifying and late G1-phase high calcifying conditions. Further transcriptome analysis was conducted without the implementation of arbitrary thresholds because other means (e.g. graphical selection by transcript frequency and previously reported genes) were used to select the differentially expressed genes of interest.

General expression patterns using functional physiological groups

The affiliations of the proteins transcribed by the observed transcripts to functional groups such as the euKaryotic Orthologous groups (KOG) were used to investigate general functional differences in the transcriptomes of the high calcification (HC) and low calcification (LC) condition of *E. huxleyi* cells in the late and early G1phase, respectively. Functional KOG annotation groups showed the highest number of KOG annotations and were therefore applied for the general inferences of functional characteristics of the cells in the HC and LC condition. 'Cell signalling' was found to be the most important functional characteristic of the expressed genes according to the retrieved transcripts with KOG-group annotation. Genes involved in 'Metabolism' and 'Information storage and processing' were almost equally expressed in the Light and Dark period. In all KOG-groups more transcripts were abundant in the LC condition (see Table 3-7). Wahlund et al. (2004) reported that most transcripts in nutrient replete and nutrient deplete cultures referred to the functional group of 'Metabolism', and could not discover a pattern towards over-representation relating to culture conditions, which increase calcification or alter calcification.

At the KOG-class the general expression patterns showed that transcripts belonging to 'Translation, ribosomal structure and biogenesis ', 'Posttranslational modification, protein turnover, chaperones', and 'Lipid transport and metabolism' were significantly different from the overall expression ratios in the HC and LC condition. 'Translation, ribosomal structure and biogenesis ' showed more significantly expressed transcripts in the LC condition, whereas more significantly expressed genes of 'Posttranslational modification, protein turnover, chaperones' and 'Lipid transport and metabolism' were observed in the HC condition (late G1-phase). Increased numbers of significantly expressed genes of the KOG-class 'Translation, ribosomal structure and biogenesis' in the LC condition could indicate that a higher general activity of cell structure synthesis, due to internal 'zeitgeber', in preparation for the onset of the HC condition (McClung, 2014) or for complementation of cell assemblies divided during cytokinesis. Elevated gene expression related to the KOG-class 'Translation, ribosomal structure and biogenesis'in the early G1-cell-cycle-phase enables photosynthetic organisms to efficiently produce carbohydrates at an early stage of the photosynthetic period already (Green et al., 2002; Dodd et al., 2005; Winjen et al., 2006).

The ratio of genes showing significantly different expression was increased in the KOG-classes 'Translation, ribosomal structure and biogenesis', 'Nucleotide transport and metabolism', 'Inorganic ion transport and metabolism', 'Extracellular structures', 'Cell cycle control, cell division, chromosome partitioning', 'Carbohydrate transport and metabolism', and 'Amino acid transport and metabolism' towards the Dark period. Von Dassow et al. (2009) showed that transcripts belonging to the KOG-Classes 'Signal transduction mechanisms' and 'Cytoskeleton' were significantly more abundant in non-calcifying 1N *E. huxleyi* cells. In calcifying 2N *E. huxleyi* cells von Dassow et al. (2009) showed that transcripts of the KOG-

Class 'Translation, ribosomal structure and biogenesis', 'Signal transduction mechanisms', and 'Cytoskeleton' were overrepresented. Only the total ratios of genes expressed in the KOG-classes 'Translation, ribosomal structure and biogenesis' and 'Nucleotide transport and metabolism' were more abundant in calcifying cells in the light phase of this study.

It was unexpected to find transcripts of the functional class summarizing 'inorganic ion transport' and 'carbohydrate transport and metabolism mechanisms' more abundant in the dark period (low calcification condition), as genes of these groups should be more active during biomineralisation and photosynthesis. However, the observed patterns further support that the readiness for biomineralisation and photosynthetic machinery is established before the start of the light phase as part of the synthesis of cell structures (e.g. Golgi apparatus), which were reduced in their extend during cell division (Nebenfuehr et al., 2000; Puhka et al., 2007). This readiness could be accomplished by inducing gene expression in a periodicity that is slightly uncoupled from the physiological activity and more related to the cell organelle construction.

It seems that care must be taken when estimations of the functional machinery are made from interpreting functional annotations. The functional groups comprise transcripts, which are active in many physiological pathways including calcification and photosynthesis. Furthermore, the evaluation does not consider transcript abundance and the potential amount of translated proteins. The expression patterns by comparison of KOG-class representatives, or by other functional annotations (e.g. GO annotations), could mask the actual general gene expression patterns that equally account for transcript abundance.

Unique transcripts in the early and late G1-phase involved in biomineralisation

Previously, the observations of unique transcripts in conditions of increased or altered biomineralisation of *E. huxleyi* suggested that the unique transcripts relate to genes, which could be involved in calcification (Quinn et al., 2006). The gene expression analysis of transcripts in the high calcification (HC) and low calcification (LC) conditions, when cells were in the late and early G1-phase, respectively, revealed 153 unique transcripts in the HC condition and 177 transcripts in the LC condition. The numbers of unique transcripts in the HC and LC condition follow the general pattern of higher transcript abundance in the early G1-phase (LC condition). However, only 18% of the unique transcripts in the late G1-phase and 68% of the unique transcripts in the early G1-phase were found differently expressed (q < 0.05). In particular, in the late G1-phase (high calcification condition) the percentage of significant transcripts was low.

Most of the transcripts unique in the HC conditions were not annotated and thus attributing functions to the most abundant transcripts is lacking. However, the second most abundant unique transcript In the HC condition is referred to Q5UP03 having a potential role in regulation of transcription. The third most abundant unique to the HC condition transcript translated to the proton-coupled amino acid transporter 3

(UniProt ID: Q4V8B1), which could be part of processes involved in the precipitation of calcium carbonate in *E. huxleyi*. Coccoliths are complex structures made generally of subsequent layers of organic matrix and calcium carbonate. The organic matrix is important for the directed precipitation of calcium carbonate forming the characteristic 3-D structure of the calcite scales (Marsh, 1986; Arias and Fernandez, 2008). The intercellular location of Q4V8B1 is suggested in the endoplasmatic reticulum using Euk-mPLoc 2.0 (Chou and Shen, 2010). Calcium carbonate and calcium binding proteins were found to be abundant in the ER and the CV (Corstjens et al., 1998; Mackinder et al., 2010). Furthermore, single amino acids were found to modulate the precipitation of calcium carbonate crystals in situ (Briegel et al. 2012) and amino acids were also found stabilizing amorphous calcium carbonate (ACC) in vitro in the red claw crayfish *Cherax quadricarinatus* (Bentov et al., 2010). In coccolithophores it was suggested that an ACC phase might play a role in coccolithophore-genesis (Mackinder et al., 2010) additionally requiring transporters Mg^{2+} and protons to sustain the pH and aqueous phase of ACC in the maturating CV before a directed precipitation of calcium carbonate is mediated by an increase in pH in the coccolithosome. Hence, the proton-coupled amino acid transporter 3 (UniProt ID: Q4V8B1) could be an important molecule under photosynthetically enhanced calcification that in conserving the Ca^{2+} in coccolithosomes by ACC conformation and thereby reducing the transport of protons from the ER, coccolithosomes, and CV into the cytosol. Four additional membrane-associated proton-coupled amino acid transporters were found in the transcriptome, showing 224 fpkm. The potential function in regard to ACC stabilization of these transporters in coccolithophores requires further investigation.

In the non-calcifying state of *E. huxleyi* (LC condition) the abundance of unique transcripts was greater than in the HC condition (see Table 3-10). In comparison Quinn et al. (2006) found more significantly up-regulated genes by using micro-arrays, which can be explained by targeting ESTs that were previously identified by Wahlund et al. (2004a). The restricted selection of target genes might have missed many transcripts as the knowledge of the *E. huxleyi* genome was still at an early stage. The unique genes of the Dark period of this study seem, as expected, to have no relevance for processes involved in calcification.

Exploration for genes involved in calcification from the graphical evaluation of significantly different transcript abundance

The expressed transcripts were clustered according to KOG-Class annotations; their expression levels at high and low calcification rates graphically illustrated, and thus obvious patterns of gene regulation explored. Out of a total of 271 graphically selected transcripts fifteen genes with a putative involvement in biomineralisation related processes were identified. Ten transcripts potentially involved in biomineralisation were significantly overrepresented in the calcification phase (late G1-phase) and five transcripts showed elevated frequencies in the low calcification phase (early G1-phase).

The most abundant transcript of those selected in the calcification phase (late G1-phase) was calreticulin (UniProt ID: P15253) showing nearly nine times higher expression in the HC condition. Calreticulin binds Ca^{2+} - ions and was found in other calcifying organisms such as corals (Tambutté et al., 2011) as well as *E. huxleyi* (Wahlund et al., 2004b; Quinn et al., 2006; Mackinder et al., 2010). Calreticulin was reported to be abundant in the endoplasmatic reticulum where it binds Ca^{2+} and buffering concentrations between 100-500 µM (Berridge, 2002) or it acts as a stress response protein (Pockley, 2003). Furthermore, calreticulin regulates the expression of SERCA-type 2 Ca - ATP pumps (John et al., 1998). SERCA-type Ca transporters were previously found in *E. huxleyi* (von Dassow et al., 2009) but the limited Ca - transport potential of the SERCA-type Ca - ATP pump suggested to be insufficient to establish the required Ca – concentrations in the CV alone (Mackinder et al., 2010). However, in combination with calsequestrin calreticulin may enable the ER to modulate impulses of calcium flux (Berridge, 2002) that could initiate coccolithogenesis, if the properties of the ER persist in the CV.

The second most abundant transcripts in the HC condition recruited to gene 431830 coding for a further putative calcium binding protein GPA (Corstjens et al., 1998). GPA was found up-regulated in calcifying cells (Quinn et al., 2006) and was present in calcifying 2N and non-calcifying 1N cells (Richier et al., 2009; Mackinder et al., 2010). In this study, JGI gene ♯ 431830 was 70 fold (1029 fpkm in HC versus 14 fpkm in LC) higher expressed in the calcification phase of *E. huxleyi*, supporting its direct or indirect involvement in calcification. An elevation of GPA-expression in calcifying *E. huxleyi* cells in nutrient replete cultures was previously reported (Richier et al., 2009), thus supporting the key importance of GPA in coccolithophogenesis. Furthermore, Mackinder et al. (2010) suggested that the silk fibrion protein similarities of GPA might make it important in the formation of amorphous calcium carbonate (ACC), as an intermediate form in advance of calcium carbonate crystallisation. However, the existence of ACC in coccolithophores is still debated.

Another potentially relevant process for coccolithogenesis might be membrane fusion and exocytosis, in which syntaxin and synaptobrevin are part of the SNARE complex (Soluble NSF Attachment Protein Receptor). Syntaxin and synaptobrevin were expressed through JGI gene ♯ 434324 and JGI gene ♯ 444996, 437571, 351006, and JGI gene ♯ 460762, respectively (compare Table 3-13). Both proteins are part of the SNARE complex, which is involved in vesicle membrane fusion. During the calcification process it is thought that coccoliths are formed in the CV. When the coccolith has formed the CV fuses with the cell membrane to expose the coccoliths to the outside of the cell (Taylor, 2007). Furthermore, from the Golgi apparatus polysaccharide and Ca^{2+} - rich vesicles, termed coccolithosomes, constantly supply polysaccharide and Ca^{2+} to the coccolith-forming vesicles (Outka and Williams, 1971; van der Wal et al., 1983); a process that also requires vesicle membranes to fuse.

Interestingly, significantly elevated transcript abundance of the MFS transporter homolog was also found in the high calcification condition. MFS-type transporters are capable of transporting small solutes in response to chemiosmotic ion gradients and are characterized as single-polypeptide secondary carriers. However, it is not clear how MFS – transporter activity could assist the biomineralisation and what solutes the MFS-transporters could process within the coccolithophores. In the low calcification condition, three genes (JGI gene ♯ 354606, 447939 and 416800) were found to express the greatest number of sequences similar to members of the $Na^+ / Ca^{2+} - K^+$ exchanger (NCKX) family. NCKX cation / Ca^{2+} - exchangers play an important role in Ca^{2+} - homeostasis. Proteins of the NCKX – family were suggested to contribute significantly to the Ca^{2+} transport into endomembrane precursor CV compartments (Mackinder et al., 2011). Here however, the putative NCKX family member's gene was expressed significantly higher in the low calcification phase (early G1-phase), when no calcification occurred (refer to section 2, Figure 2.10). Nevertheless, multiple genes in the *E. huxleyi* genome for NCKX – like proteins and their expression levels will be discussed in more detail below to evaluate the potential activity of NCKX transporters during the calcification phase. The calcium transporters that were expressed in non-calcifying early G1-phase cells (compare Table 3 – 13) may still play a role in calcification during the light phase. In the early G1-phase after cell division was completed general protein synthesis is increased (Snustad and Simmons, 2000). Calcium transporters are most likely expressed as well to construct the functionality of organelles, membranes and compartments to enable the establishment of Ca^{2+} gradients; for example between the cytosol and the ER lumen before the onset of the Light period.

Discovering expressed biomineralisation gene using transcript abundance

An analysis of the top 1000 most abundant transcripts of the light phase demonstrated transcript abundances ranging from $13 \ 10^6$ fpkm to 225 fpkm. The selection of the 1000 most frequent transcripts coding for proteins potentially involved in calcification is presented in Appendix A Table 12. The most abundant transcript in the light (JGI *E. huxleyi* scaffold 781:3342-3438) is newly described as an un-reviewed hypothetical protein partial mRNA (NCBI XM_005771519.1; Read et al., 2013). One protein-homologous group previously not discussed in this monograph was discovered in the 1000 most abundant transcripts that indicates a potential role in the calcification processes. A GTP-binding protein SAR (UniProt ID: Q01476) a subunit of the COPII vesicle coat complex showed 255 fpkm and 103 fpkm transcripts in cells showing elevated and reduced calcification rates, respectively. Clathrin was previously identified as another highly abundant transcript in the both conditions. Clathrin peptides were also identified to have significantly higher relative abundance in calcifying *E. huxleyi* cells by the proteome analysis. Jones et al. (2011) suggested clathrin to be important in the biomineralisation process because it occurs together with V-type ATPase proton transporter in clathrin-coated vesicles (Forgac, 2000). Also vacuolar-type proton

pumping ATPase was identified in the coccolithophore vesicle membranes (Corstjens et al., 2001), suggesting that these two molecules exist together in coccolith vesicles. Clathrin's significance for vesicle-mediated transport in the silica-precipitating diatom *T. pseudonana* was previously suggested from proteome analysis (Nunn et al., 2009). Vesicle-mediated transport in coccolithophores with clathrin and/or GTP-binding protein SAR could deliver minerals from the Golgi network to the CV via coccolithosomes. The V-type ATPase could control the luminal pH of the coccolithosome and trigger forced alkalinisation. The presence and location of clathrin and GTP-binding proteins using molecular staining should be investigated. Such a study might reveal that clathrin and the COP vesicle coat complex should be placed in a more general model for vesicle transport in marine phytoplankton if it is active in diatoms and coccolithophores alike. Vesicle fusion assisted by the SNARE complex is another related process that was indicated by the significant higher abundant transcripts of syntaxin and synaptobrevin, which are part of the SNARE complex, in the light period (see Table 3-13).

Expression of genes with a previously suggested role in coccolithogenesis in *E. huxleyi*

The expressions of known genes coding for vacuolar membrane proton pump ATPases (V-ATPase), proton exchanger, Ca^{2+} - transport proteins, calmodulin (CaM) and Ca^{2+} - binding proteins, and bicarbonate transport related protein were investigated in periods of calcification and no calcification over the cell cycle of *E. huxleyi*. The general expression of these genes is higher in the early G1-phase (no calcification), with the exception of genes coding for calmodulin (CaM) and Ca^{2+} - binding proteins. The genes (JGI♯ 373343, 442625, and 443126) coding for CaM (UniProt. ID: Q40302) were the highest expressed of the Ca^{2+} - binding proteins. Previously, Richier et al. (2009) also found that CaM and GPA expression was elevated in the Light period of diploid *E. huxleyi* RCC1217. As reported above calreticum and calnexin expression was highly up-regulated in the calcification phase (Light period). The constant demand for calcium binding proteins may result from the continuous construction of coccolithosomes to supply calcium to the CV. The high abundance of transcripts of calcium binding proteins could also suggest that the molecules are instable or are lost after the CV merges with the cell membrane and the coccolith is expelled to the outside of the cell. Calnexin and calreticulin were assigned to the ER (Brodski and Skach, 2011), whereas no preferred location for the calmodulin is known, as it may be found at different subcellular locations such as the cytoplasm, organelles or plasma organelle membranes (InterPro Protein Archive). Hence, calmodulin in *E. huxleyi* could also be a messenger protein for calcium transport to receiver proteins (Chou et al., 2001), such as other calcium binding proteins. Continuous synthesis of calreticulin, calnexin, and CaM appeared to be present in calcifying cells for the delivery of calcium or stabilization of calcium concentrations within coccolithosomes and the CV.

The relatively under-represented transcripts of NCKX, CAX, and, calcium channels in the calcification phase could indicate a lower significance of calcium transporter expression for the maintenance of coccolithophogensis. In fact the low expression of NCKX, CAX, and calcium channels might also indicate that the calcium transporters are more stable within the membrane of the ER and that there is a transverse flow of calcium in the ER from areas with high abundance of NCKX, CAX, and calcium channels to areas of coccolithosome-formation. Only the transporters that are lost from the ER by coccolithosome-formation need to be resynthesized. This could explain why expression levels of NCKX, CAX, and calcium channels were lower in the calcification period than in the low calcification period.

The control of the pH is vital for calcium carbonate precipitation and important for coccolith production. Hence, proton levels need to be controlled by active proton pumping to establish low proton concentrations inside the CV. Previously suggested proton pumps in *E. huxleyi* include Ca^{2+} -stimulated vacuolar V-type-ATPase and P-type ATPase (inorganic pyrophosphate driven H^+ pumps) (Corstjens et al., 2001). Only one Vacuolar H^+-pyrophosphatase (P21616) showed higher expression in the Light period. Nevertheless, transcripts of V-type ATPase and P-type ATPase were also present in the calcification phase but at roughly $\frac{1}{2}$ to $\frac{1}{3}$ of the densities as in the Dark period, but still reaching frequencies of approximately 1000 and 130 fpkm per gene. The observed expression patterns suggest that V-type proton ATPase and P-type proton ATPase are mainly translated during the Dark period and maybe at a very early stage of the Light period. However, expression levels for V-type H^+ ATPase and P-type H^+ ATPase are also considerable in the calcification period. Vacuolar type proton ATPases were suggested to be pivotal for proton pumping in coccolithophores and were located at the coccolith producing membrane (Corstjens et al., 2001). Mackinder et al. (2010) discussed the possibility of a unique pumping direction of V-type proton ATPases and V-PPases, in which case proton pumping would occur from the organelle (CV) into the cytosol. Potentially, the considerable expression of V-type H^+ ATPase and P-type H^+ ATPase during the Light period is required to replace molecules that are lost during coccolithophogenesis, whereas the higher expression of V-type ATPases and P-type ATPases in the Dark period implies that the molecules are required at many locations of the cell for transporting protons and are incorporated into cell endosomal membranes, such as those of the ER. Further proton exchangers were also more abundant in the Dark period. The significantly expressed Na^+/H^+ antiporters (UniProt ID: Q39WP5 and Q30XM9) are both considered integral components of membranes and are involved in pH regulation (Mackinder et al., 2010). However, both genes were significantly higher expressed in the dark, showing only very little transcripts or none in the Light period (see Appendix A Table 6 for details). Only the proton transporting protein H^+/hexose cotransporter 3 (UniProt ID: Q39525) was more highly expressed in the calcification phase and is specifically involved in co-transporting sugar across the plasmalemma (Stadler et al., 1995). H^+/hexose cotransporter 3 was not suggested to play a role in calcium carbonate precipitation in coccolithophores.

Interestingly, the voltage gated proton channel (UniProt ID: E2IJ90) that was suggested to play an important role in pH homeostasis in calcifying cells of *Coccolithus pelagicus ssp braarudii*, as it mediates rapid H^+ efflux (Taylor et al., 2011), was not over expressed in calcifying cells of *Emiliania huxleyi* in the Light period.

The intracellular calcification of coccolithophores requires constant fluxes of inorganic carbon and calcium into the cell and the CV. A variety of potential Ca^{2+} -transporters, such as CAX, NCKX, NCX, and SERCA-type Ca^{2+} -ATPases have been identified as important candidates to raise Ca^{2+} concentrations in the CV or its precursors (Mackinder et al., 2010). The transcriptome analysis did ascertain the significantly over expression of SERCA type calcium-transporting ATPase 4 (JGI gene ♯: 251608, UniProt ID: Q9XES1) in the calcifying period and by JGI gene ♯: 463095 in the Dark period. But Q9XES1 was also expressed by JGI gene ♯: 429294 in the HC condition, leading to almost balanced levels of Q9XES1 expression in the HC and LC condition (compare Appendix A, Table 7). The SERCA – type calcium-transporting ATPase was found to have limited calcium transport capacities to provide all the necessary calcium transport. Therefore, NCKX/NCX, Ca^{2+}/H^+ exchangers, and Ca^{2+}/Na^+ exchangers were suggested to provide additional calcium flux (Mackinder et al., 2010). Transcripts of greater abundance than 25 fpkm in the light were given by a NCKX1 (JGI gene ♯: 447939), a probable sodium/potassium/calcium exchanger 1 of the NCKX1-type (JGI gene ♯: 354606), CAX5 (JGI gene ♯: 416800 and 415715), and a transient receptor potential cation channel subfamily M member (2 LTrpC-2, JGI gene ♯: 460292), which could help to recover the calcium fluxes across the ER - membrane and potentially across the coccolithosome membrane.

The primary source of carbonate for the formation of calcium carbonate is bicarbonate (Nimer et al., 1997; Paasche, 2001; Bach et al., 2012). In this study, the bulk of carbonic anhydrase was expressed in the early G1-phase. Under low DIC conditions carbonic anhydrase was found more highly expressed (Mackinder et al., 2011; Bach et al., 2012). However, the DIC characteristics of the cultures (concentrations of HCO_3^-, CO_3^{2-}, and CO_2; data not shown) do not suggest that carbonic anhydrase expression was stimulated. Hence, the CO_2 levels in the cultures were sufficient for efficient photosynthesis (Ross et al., 2003; Bach et al., 2012). Richier et al. (2009) reported elevated expression of *E. huxleyi* carbonic anhydrase in the light in K2 medium. The location of CA is assumed in cellular organelles and the plasmalemma (Kitao et al., 2008; Soto et al, 2006). Recent data suggested an absence of cytosolic CA in *E. huxleyi* (Suffrian et al., 2012; Bach et al., 2012). Under low CO_2 concentration Bach et al. (2012) observed an increased expression of an anion exchanger like protein 1 (AEL1) of the solute carrier 4 family (SLC4) with a suggested location at the plasma membrane and induction at low CO_2 concentrations.

Transcripts of AE1 like proteins (UniProt ID: Q9HGM6) were more abundant in the calcification period and JGI gene ♯ 466232 and ♯ 314659 significantly more highly expressed, supporting the importance of the AEL1 in coccolithophore calcification and photosynthesis. However, another AE1 related protein (UniProt ID: P23347) was significantly higher expressed in the LC condition.

Biomineralisation gene expression in the G1 cell-cycle phase of *E. huxleyi*

The differences in the gene expression profiles between the high and low calcification conditions of the G1-phase *E. huxleyi* cells became apparent at many levels of the analysis. The observed patterns of gene expression presented here, however, were not related to sampling at different cell-cycle phases (G1, S, or G2+M) but exclusively related to different calcification rates in dark and photosynthetically enhanced phases of the G1-phase. Thus, the results presented here, are novel insights into the synthesis of compartments and biological pathways in *E. huxleyi* and constitute that a status quo at the onset of the photosynthesis phase in a diurnal cycle is reached, which facilitates immediately increasing calcification rates in the light phase. Furthermore, the observed patterns show that a large proportion of the molecular machinery driving calcification is bound to the general cell structure or cell compartments. Significantly higher gene expression and transcript diversity was observed in non-calcifying early G1-phase *E. huxleyi* cells (see section 3.3.3) compared to highly calcifying late G1-phase cells. The null hypothesis that there are no differences in gene expression between non-calcifying early G1-phase cells and calcifying late G1-phase cell related to genes involved in calcification processes in respect to transcript abundance and diversity, must be rejected. The expression of many genes previously related to processes involved in calcification was found much higher in the low calcifying early G1-phase compared to the highly calcifying late G1-phase cells. For example, all investigated ion-transporter transcript clusters showed more transcripts in the Dark period. However, this observation was sometimes offset by the higher expression of genes in the late G1-phase relating to the same protein (e.g. SERCA type calcium-transporting ATPase 4, UniProt ID: Q9XES1), which could indicate differences in gene expression triggering and control as well as that the transporter genes, which were highly expressed in the late G1-phase period, are involved in coccolithophogenesis. On the other, this pattern may also reflect the re-construction of molecular machinery that was separated during cytokinesis. Golgi stacks, endoplasmic reticulum, ribosomes, mitochondria and chloroplasts are known to be equally partitioned between daughter cells during cytokinesis (Nebenfuhr et al., 2000; Puhka et al., 2007). Following cytokinesis the ER and Golgi membranes are synthesized, as the ribosomal-ER-Golgi cell physiological network is pivotal for protein translation and posttranslational protein modifications. These networks ensure that at dawn efficient photosynthetic carbon assimilation can render cell growth and accumulation of energy reserves for passing cell division check points of the next cell division cycle (Green et al., 2002; Dodd et al., 2005; Winjen et al., 2006b; Zachleder et al., 2016). The synchronisation gene expression and the

environmental light - dark rhythm improves the physiological performance of photosynthetic organisms (Green et al., 2002; Dodd et al., 2005; Winjen et al., 2006b). Hence, serving the need for maximized cell growth and ecological prevailing. It is an energetic advantage especially for photosynthetic cells to establish the physiological networks before the next illumination period. Because calcification and photosynthesis in coccolithophores is tightly coupled (Balch et al., 1992), the genes relating to the physiological networks for biomineralisation, Golgi apparatus and ER for example, and photosynthesis were all expressed in the early G1-phase. Therefore, the timing of the sampling can explain many biases of the expression levels observed for molecules with known function in coccolith calcification. The model that gene expression of transporter proteins potentially involved in coccolith production (e.g. NCKX, VCX) was increased in the early G1-phase to complement the reduced cytoplasmic compartments of ER and Golgi apparatus after cytokinesis is supported by elevated levels of gene expression of proteins with known involvement in either ribosomes, mitochondria, or chloroplasts in the early G1-phase. Significantly higher expressed proteins in the early G1-phase showed ratios of 35:39, 38:46, and 121:147 of significantly expressed in the dark versus total significantly expressed protein for ribosomal, mitochondrial, and chloroplastic proteins, respectively. The expression of chloroplastic, ribosomal and mitochondrial genes was found to be reduced in the G1-phase for *Cyanidioschyzon merolae.* Periodicity of gene expression directly related to cell cycle phases or plastid gene expression in the unicellular red alga *Cyanidioschyzon merolae*. 158 genes were only induced during the S or G2/M-phase, of which 93 were known and contained genes related to mitochondrial division (Fujiwara et a., 2009). Overall the observed gene expression patterns are likely following an endogenous clock set by periodic environmental cues such as the diurnal cycle of light and dark in *E. huxleyi* cultures (Wijnen and Young, 2006; McClung, 2014).

Considerations for *Emiliania huxleyi* calcification models

Genes that were previously reported to be involved in *E. huxleyi* calcification process are not necessarily found at higher expression rates in the photosynthesizing late G1-phase cell. In fact, the transcriptome patterns suggest that suggest that more and less expendable proteins interact during coccolithogenesis. Nevertheless, null-hypothesis that the general expression of genes related to calcification is not enhanced in the high calcification period when calcification rates were found to be at a maximum, can only be partly rejected. Future models of biomineralisation for *E. huxleyi* should distinguish between proteins that are more decisive for coccolithogenesis and such that are very relevant for general cell functioning but partly expended in the process of coccolith production. The level at which proteins are conserved is partly due to the membrane bound location of proteins involved in ion-transport, such as NCKX transporters, V-type Ca^{2+}/H^+ antiporter, and V-type ATPases. However, in the scenario where coccolithosomes are expelled from the membrane of the Golgi body some of the transporters are lost in the process; hence requiring expression in the calcification phase of *E. huxleyi*. The expression of proteins with a high disposal rate such as calcium

binding proteins calmodulin, calreticulin, and calnexin was concentrated in the highly calcifying late G1-phase. These expression patterns confirmed the importance of continuous calcium binding protein synthesis in the calcification period to maintain coccolithophogenesis. The highly acidic and calcium binding macromolecule, containing high levels of glutamic acid, proline and alanine, (GPA) was expressed 70 fold higher in the sampled calcification period. Furthermore, the observations of genes uniquely expressed in the HC and LC condition has suggested new genes with potential roles in the stabilisation of amorphous calcium carbonate and processes of membrane fusion. However, the presence of ACC has yet not been confirmed and the proton-coupled amino acid transporter 3 (UniProt ID: Q4V8B1) gene was not showing as high rates of expression as the ion transporter genes in early G1-phase cells. The SNARE complex that combines the revealed syntaxin and synaptobrevin proteins could be relevant for the fusion of vesicle membranes in the growth of the CV. The functions and locations of the proton-coupled amino acid transporter 3 and the SNARE related proteins within *E. huxleyi* should be further investigated to reveal or not to reveal their involvement in processes related to calcification, ACC stabilisation, and vesicle membrane fusion.

Outlook

The results presented here are of high relevance for the design of future experiments and confirm the great importance of Ca^{2+} binding proteins in coccolithophogenesis. The expression patterns clearly suggest that a large proportion of biomineralisation genes are expressed in the early G1-phase of the *E. huxleyi* cell cycle; in particular those genes transcribing proteins that were required to complement the reduced size of cell compartment as a result of cytokinesis, but which also play an important role in the process of coccolith production. More conserved proteins involved in biomineralisation of *E. huxleyi*, which were found to be highly expressed in the non-coccolith producing early G1-phase, may provide the transport rates required to create the high calcium concentrations in coccolithosomes. The ion transport proteins expressed in the calcification phase are mainly required to maintain the ion concentrations in the coccolithosomes and CV and replace proteins that were lost in the process of coccolithogenesis.

Chapter 4.   The proteome of *Emiliania huxleyi* at high and low calcification rates

"Cellular reality is more elaborate than the dreams of even the nucleus itself."

(Anderson and Anderson 1998)

4.1. Introduction

Studying the proteomes of marine organisms has developed into a new and ambitious field of marine molecular biology that aims to provide deeper understanding of the physiological functioning of species in the marine realm (López, 2007). At the organism as well as the population level new insights gained by studying the proteome and therefore the drivers for the organism's physiological activity have increased our understanding of key ecological processes. For example, a variety of quantitative expression proteomics studies have been conducted on different marine taxa, such as on the bivalve *Pinctada margaritifera* (Joubert et al., 2010) – focusing on the expression of shell matrix proteins using EST analysis of cells of the calcifying mantle and verifying shell matrix protein presence using proteome analysis of the shell, microalga such as the diatom *Thalassiosira pseudonana* - investigating the general physiology and acclimation in response to nutrient availability (Dyhrman et al., 2012), on the bivalve *Mytilus galloprovincialis* - investigating proteomic responses to different temperature regimes (Tomanek & Zuzow, 2010), and *Emiliania huxleyi* - investigating changes due to ocean acidification and light intensity (Jones et al., 2013; McKew et al., 2013). Meta-population studies applying proteomics identified how the utilization of nutrients and energetic flows shift over an environmental gradient of macronutrients in the South Atlantic. A significant fraction of the microbes showed a high presence of TonB-dependent transporters, which are proteins able to utilize light for generating a proton motive force that increases the efficiency of ATP synthases and ATP independent nutrient uptake (Morris et al., 2010).

The importance of biomineralisation at the organismal level for the control of optimal physiological conditions and on an ecosystem scale for the transport of inorganic carbon was elaborated in Chapter 1. Proteomic studies have aimed to reveal the proteins involved in constructing the underlying molecular and cellular mechanisms involved in calcification (Joubert et al., 2010; Tomanek et al., 2011; Jones et al., 2013). In the coral *Acropora millepora* Ramos-Silva et al. (2013) confirmed the importance of skeletal organic matrix proteins (SOMPs) and inferred on the evolutionary origin of the transcoding SOMP genes. In close relationship with polysaccharides, SOMPs and amino acids form the organic matrix directing and controlling precipitation of calcium carbonate in the sub-calicoblastic space (Tambutté et

al., 2011; Venn et al., 2011; Ramos-Silva et al., 2013). A total of 36 SOMPs were identified in a study of the
proteome of *Acropora millepora*. Interestingly the evolutionary origin of many of the identified SOMPs
pointed towards non-calcifying cnidarians and only two homologs were restricted to Scleractinia (Ramos-
Silva et al., 2013).

    The number of next generation proteomics studies on coccolithophores is still limited. Jones et al. (2011)
have published pioneering work on the proteome of *E. huxleyi* , identifying a set of 99 proteins in the strains
NZEH (PLY♯ M219), CCMP1516, and CCMP371. Some of the identified proteins, such as carbonic
anhydrase, vacuolar ATPase units, and clathrin were discussed to be involved in *E. huxleyi* calcification
processes (Jones et al., 2011). The importance of V-type ATPases in the establishment of proton gradients and
high pH values in the CV (coccolith vesicle) promoting calcium carbonate precipitation was supported
(Corstjens et al., 2001; Mackinder et al., 2010). Also, clathrin - coated vesicle membranes contain V- ATPases
in eukaryote cells (Kirchhausen, 2000), was found by Jones et al., (2011). Hence, Jones et al. (2011)
hypothesized that clathrin may constitute the membranes of the CV or ER in coccolithophores. Clathrin -
coated vesicles were found to provide membrane trafficking pathways within Golgi networks (Lewin and
Mellman, 1998; Kirchhausen, 2000).

    More recently a quantitative *E. huxleyi* proteome study explored differences in the proteomes at ambient and
elevated CO_2 concentrations (Jones et al., 2013). The study yielded 115 homologous protein groups of these,
46 could be explored in respect to their differential expression in the experimental treatments (Jones et al.,
2013). However, no proteins known or with a suggested role in coccolithogenesis were significantly affected
by ocean acidification. Only the histones H2A, H3 and H4 and a chloroplastic 30S ribosomal protein S7 were
down-regulated under elevated CO_2 concentrations. But trends of membrane related processes, e.g. an up-
regulation of an acyl-carrier protein, could indicate increased lipid production and increased vesicle membrane
synthesis under elevated CO_2. Furthermore, an up-regulation of vacuolar membrane bound V – type proton
ATPases could indicate stronger regulation of cellular and vesicle pH in an ocean acidification scenario. The
proteome studies on *Emiliania huxleyi* have been proven as very challenging due to protein extract quality
issues, possibly resulting from high levels of lipid, polysaccharide or salt contamination in the protein extracts
(Jones et al., 2009; Jones et al., 2013). Furthermore, the genomic data on coccolithophores has only recently
been improved by the publication of the pan genome of *Emiliania huxleyi* (Read et al., 2013). The protein
identification yield for *Emiliania huxleyi* was improved to 500 highly abundant and confident proteins from a
total of 1835 proteins in the study of McKew et al., 2013. The protein abundance in different light incubations
was investigated and showed that proteins of the photosystem I, photosystem II, and proteins involved in
biosynthesis and oxidative stresses were more abundant in the high light treatment (McKew et al., 2013).
Calcification rates also increased by two fold in the high light treatment (1000 μmol photons m^{-2} s^{-1}). A
detailed study of the proteome harvested from cells showing higher calcification rates indicated that V-type

proton ATPases and the calcium-binding protein calreticulin were more abundant (supplementary data from McKew et al., 2013). The study by McKew and colleagues (2013) greatly increased recovery yields of proteins predicted by the *E. huxleyi* 1516 'Filtered ("best") Models' gene set (http://genome.jgi-psf.org/Emihu1/Emihu1.home.html; Read et al., 2013). However, the number of recovered proteins from the *E. huxleyi* proteomes out of a 39126 predicted by gene models (Jones et al., 2013; McKew et al., 2013, Read et al., 2013) is still relatively small in comparison to recovery rates in *Saccharomyces cerevisiae* where 4300 confirmed proteins were recovered from a genome of 5,770 genes (King et al., 2006).

In this study the proteomes from *E. huxleyi* cells from the late G1-phase (high calcification) and early G1-phase (low calcification) (as described in Chapter 3) are compared using a quantitative shotgun proteomic approach similar to Jones et al. (2011), to identify proteins of the calcification period that might be involved in the process of biomineralisation in coccolithophores. The consensus of recovered proteins and observed transcripts is evaluated.

4.1.1. Aims and Objectives

The proteome directly reflects the given physiological state of the cells and in the case of *Emiliania huxleyi* promises to show molecules utilized in calcification processes. By investigating differences in the proteomes of cells harvested from cultures at time points where low and high rates of calcification were previously confirmed (see Chapter 2), proteins involved in the calcification processes were identified. Therefore, proteins were extracted from cells harvested in the late and early G1-phase (light and dark incubation period) and a quantitative comparison of the proteomes of *E. huxleyi* by means of multidimensional liquid chromatography tandem mass spectrometry (LC/LC-MS/MS) and iTRAQ labelling was conducted.

The aims of this study were:

- to assess the proteomes of iso-genetic cultures of *Emiliania huxleyi* at negligible (early G1-phase) and high rates of calcification (late G1-phase) of the cell cycle,
- to compare the proteomes at negligible and high calcification rates for differences in protein abundance,
- to elucidate proteins potentially involved in biomineralisation, and
- to evaluate the proteome with the transcriptome (see Chapter 3) for matching gene expression patterns.

4.1.2. Null hypothesis

The working hypothesises for these studies are:

- There are no differences in proteomes of the highly calcifying and low calcifying *E. huxleyi* cells;
- Proteins potentially involved in the process of calcification are not expressed in the calcification period (late G1-phase).

4.2. Material and Method

4.2.1. Experimental Design

From the same exponentially growing cultures harvested as described in Chapter 3 proteins were extracted and subjected to proteome analysis. Six batches of 16 L cultures of *E. huxleyi* PLY♯M217 were grown with starting cell densities of 5000 cell mL^{-1}. Cell harvesting was conducted when cell densities exceeded 50,000 – 100,000 cells mL^{-1}. Cells were harvested 9 hours into the light period and 10 hours into the dark period at 6:00 pm and 7:00 am using inline filtration (see Figure 3.1, Chapter 3). Samples for nutrient analysis, cell count, cell volume, DIC, PIC, and POC were taken immediately before cell harvesting. The assessed biogeochemical parameters were determined following the described methods in Chapter 3. Cells were collected from a 147 mm filter with sterile filtered seawater and concentrated by centrifugation for 5 min at 46000 g (Rotanta 460R, Hettich Zentrifugen, Germany).

4.2.2. Extraction of proteins

Around 50% of the frozen cell pellet from 4.2.1 were transferred to a 50mL Falcon® tubes and ten millilitres of 100 mM triethylammonium bicarbonate (TEAB, Sigma Aldrich, Poole, UK), 0.1% sodium dodecyl sulphate (SDS, Sigma-Aldrich, WGK Germany) buffer were added. Probe-sonication on ice was conducted to resuspend and break the cells by shearing forces using a VC300 Vibracell sonicator (Sonics and Materials, USA, 20 kHz frequency, 10% duty cycle). Twenty-five, ten seconds bursts on output 3, with each burst separated by a 30 second pause for cooling were most efficient to destroy the cells and solubilize the proteins. The cell debris was separated from the protein extract by centrifugation (46000 g, 30min, 4°C, Rotana 460R). The supernatant was transferred into a new 50mL Falcon tube. Four volumes of 100% Acetone (HPLC-grade, Thermo-Fisher, Loughborough, UK) cooled to -20 °C were added and gently mixed with the TEAB-buffer. The proteins were allowed 36 hours at -20 °C to precipitate. The precipitated proteins were collected by centrifugation (20000 g, 30min, 4 °C). The acetone-supernatant was discarded and the protein pellet resuspended in cold (-20°C) 80% acetone for further purification steps. After a resting time of around 5 minutes at -20 °C the protein was pelleted by centrifugation (20000 g, 30min, 4 °C). This washing procedure with -20 °C 80% acetone was conducted twice. The final protein pellet was dried on ice in a laminar flow environment and the protein sample re-dissolved in TEAB-buffer (100mM Triethylammonium bicarbonate buffer 1.0 M, pH 8.5±0.1, 0.1% (w/v) SDS). The protein samples were stored at -20 °C until proteome analysis.

4.2.3. Protein quality assessment and troubleshooting

The thawed protein samples from 4.2.2 were resuspended by repeated cycles of vortexing and sonication at 37 °C to increase the solubility of proteins. The protein solution was centrifuged at 10000 g for 10 minutes to pellet insoluble molecules. One-dimensional gradient SDS-polyacrylamide gel electrophoresis was performed to assess protein quality. For SDS gel electrophoresis the samples were prepared by adding 30 µL of the sample supernatant, 10 µL 4 x LDS-loading buffer (Invitrogen, Paisley, UK) and 2 µL of 0.5 M DTT (*threo*-1,4-Dimercapto-2,3-butanediol, Sigma-Aldirch, UK). Samples were heated for 10 min at 75°C for protein denaturation before loading the gel pockets. The separation of proteins into fractions was performed on a 4-12% gradient SDS-polyacrylamide gel (NuPAGE®, dimensions: 8.0 cm x 8.0 cm, Invitrogen, Paisley, UK) in NuPAGE® MOPS SDS running buffer (Thermo Fisher Scientific, UK) at 200 V for 55 min. Proteins were stained with Colloidal Coomassie Brilliant Blue 0.2% (w/v) dye (Colloidal CBB, Sigma-Aldrich, Dorset, UK) incubation for 30 minutes. Subsequently, de-staining using 30% methanol and 2.5% glacial acetic acid was conducted over night. The de-stained gels were imaged on a VersaDoc imaging system and the PDQuest software (Bio- Rad, Hercules, CA, USA).
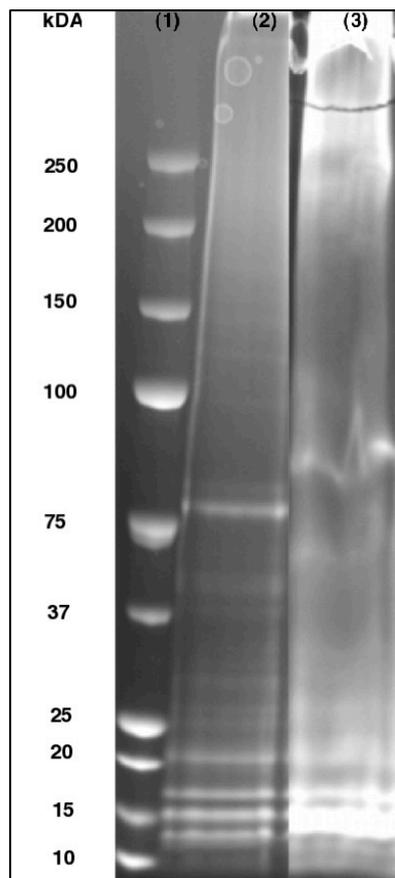
Figure 4-1:Example of results of 1D-SDS NuPage® gels illustrating the encountered issue with smearing. (1) Protein weight ladder, (2) sample after additional ultra-sonication and acetone washing and (3) sample showing smearing of protein material.

Early 1D-SDS gels showed smearing and low levels of protein fractioning (Figure 4-1). Lipids, DNA, polysaccharides, and salts were observed to alter the quality of protein samples. Nagai et al. (2008) suggested that viscous polysaccharides cause smearing on SDS-PAGE gels produced from protein extracts of kelp (*Ecklonia kurome*). Furthermore, small amounts of lipids may deform the protein-fraction-bands on a gel (Pokhariyal et al., 2014). Therefore, the cleaning steps were varied and tested to reduce the concentration of contaminants that caused smearing. The attempts to improve protein quality included desalting columns, additional washes with acetone after four short one-second bursts from the sonicator to resuspend the pellet, and stronger as well as longer centrifugation at 10000g. The additional acetone washes in combination with re-suspending the protein pellets with short sonication bursts showed the best fractionation of proteins (compare Figure 4-1).

4.2.4. Protein concentrations assessment

Protein sample concentrations in 100 mM TEAB-buffer were assessed prior to the LC/LC-MS/MS runs applying a bicinchoninic acid (BCA) protein assay (Pierce Biotechnology, Rockford, USA). The working reagent was prepared for the 96-well microplate procedure following the manufacturer's instructions. Standards of BSA concentrations from 0.1 mg mL^{-1} to 1.0 mg mL^{-1} and sample aliquots were prepared using 20 µL of samples or standards in 200 µL of working reagent. Protein–stain was allowed to develop for 30 min in the dark at room temperature. Plates were read at 570 nm using a Dynex plate reader and analysed with the Revelation 3.2 software (Dynex Technologies Limited, Worthing, UK). For the iTRAQ® labelling and subsequent tandem mass spectrometry quantification and identification of proteins the sample concentrations were adjusted to be equal.

4.2.5. iTRAQ®– labelling

The iTRAQ® technique describes a non-gel proteomic approach simultaneously identifying and measuring abundance levels of proteins in a mixture containing peptides from up to 8 different sample extracts (Applied Biosystems, Foster City, U.S.A). The key is to assign the peptides to the different samples within the protein cocktail by attaching probes – 'tags' - that differ in isotope composition to the N-terminus of the amino acids. The isotopic signature of the iTRAQ® tags (reporter ions) are detected by the mass spectrometer without interacting chemically or technically with the detection of the amino acids. By associating the relative quantities of the reporter ions with the identified proteins the protein abundance levels in the different samples in the mixture can be compared (Ross et al. 2004; Choe et al., 2007).

The procedure was kindly conducted by Dr. Paul Skipp, Institute for Life Sciences, and is briefly described here (for details see Skipp et al., 2005; Jones et al., 2013). The manufacturer's instructions (Applied Biosystems, Foster City, U.S.A.) were followed to perform the iTRAQ®-labeling of samples. Accordingly,

each sample containing around 100 μg protein was reduced, cysteine blocked, digested with 10 μL trypsin at 37 °C overnight, vacuum-dried for 1 hour at 45 °C, to which 30 μL iTRAQ® solution buffer were added. The iTRAQ® labels 113 to 118 for Light and Dark period samples (three replicate samples per condition) were attached to the N-terminus of peptides in isopropanol at pH above 7.5 within 2 hours of incubation. The labelled samples were stored at -20 °C prior to liquid chromatography (LC) and tandem mass spectrometry analysis.

4.2.6. Liquid chromatography mass spectrometric analysis of peptides

Dr. Paul Skipp at the Institute for Life Sciences (University of Southampton, UK) also conducted the separation of peptides by liquid chromatography and identification by mass spectrometry as previously described (Skipp et al., 2005). Hence, in the first step, the digested and iTRAQ® - labeled peptides were injected onto a nano-capillary two dimensional reverse phase (2D-RP) LC system, using a Waters C18 RP, 3 μm, 100 Å (150 mm x 75 μm, internal diameter) column for separation of peptides based on their hydrophobicity (Waters, Elstree, UK). The LC column reduced the complexity of the peptide-mixture and before the separated peptides were ionized in an in-line Matrix-Assisted Laser Desorption/Ionization (MALDI) phase of the hot plume of an ablated gas plasma. The ionized molecules were accelerated to the quadrupole time-of flight tandem mass spectrometer (Q-Tof Global Ultima, Waters, Milford, USA). According to the mass-to-charge-ratio (m/z) of the eluting peptide the collision energy was acquired during a survey scan at m/z 300 to 1700 switching the criteria for MS to MS/MS, the ion intensity and charge state. The spectra of the peptides were compared to fragmentation data of the peptides in the database to identify the proteins. During MALDI the iTRAQ®-tag is also separated from the peptide releasing a unique signal in the mass spectrometer. A relative quantification of the number of peptides or their fragments of one sample (tag) is achieved by comparing the number of reporter ions (tags) released in coherence with the peptide being reported in other samples, for example between the peptide ratios of the same peptide labelled with the tags 113 and 114.

4.2.7. In silico identification of proteins

The MS/MS generated spectra of the peptides or fragments showed peaks that were compared against a library of known mass fingerprints of amino acids and peptides within the MASCOT Distiller Software v.2.2.1 (Matrix Science, London, UK), applying a probabilistic scoring algorithm for statistical confirmation of matches between observed and projected peptide fragments (Koenig et al., 2008). The peptide mass fingerprinting is based on the assumption that a protein will produce a specific set of peptides when it is digested with a site-specific protease such as trypsin, which represent a virtual set of masses to compare with

the MS results. The matching of the experimental protein mass to the predicted masses of proteins is conducted within the MASCOT Distiller Software referring to a large computer database containing predicted proteins from various reference genomes. The MASCOT software compared the experimental spectra of peptides against the predicted proteins of the *Emiliania huxleyi* CCMP1516 draft genome produced by the US Department of Energy Joint Genome Institute (JGI; http://www.jgi.doe.gov, downloaded 22.2.2012, file: "Emihu1_all_proteins.fasta"). The reference sequences consisted of 115,995 non-redundant sequences from all models generated by the JGI annotation pipeline. Furthermore, the experimental scans were also searched against a taxonomically restricted database, which contained 196,774 proteins (93,921,998 residues) constructed from all Alveolata, Cryptophyta, Haptophyceae, Rhodophyta, and Stramenopile sequences available at UniProtKB Release 15.1 (UniProt 2008). The mass tolerance for all searches was set at 150 ppm and fragment mass tolerance at 0.25 Da. Additional settings were carbamidomethylation set to fixed modification, oxidation of methionine to variable modification, and the maximum missing cleavage was set to one. Protein matches were only accepted if supported by 2 or more peptides and under the significance threshold of $p < 0.05$. Non - *E. huxleyi* proteins from the taxonomically restricted database (TAXDB) were defined and removed as redundant if supported by less than 2 peptides. Protein matches to TAXDB were validated using homology in a phylogenetic context by HAQESAC (Homologue Alignment Quality, Establishment of Subfamilies and Ancestor Construction) (Edwards et al., 2007). The automated analysis pipeline HAQESAC generated multiple sequence alignments and phylogenetic trees of protein families. Proteins were searched using BLAST (e < 10-4) against the previously created taxonomically restricted protein database (TAXDB) and the ≤ 50 closest homologues were aligned with MAFFT (Katoh and Toh, 2008). The homologues were used to construct a phylogenetic tree with 1000 bootstraps using FastTree (Price et al., 2009). The findings of the protein identification by phylogenetic information were validated manually to remove false positives and prefer proteins taxonomically closer to *E. huxleyi*. The database constructing and bioinformatic processing was conducted by Dr. Richard Edwards (former Institute for Life Sciences, University of Southampton, UK)

4.2.8. Data processing

Redundancy and duplicate matches of proteins to the *E. huxleyi* JGI reference genome data and the TAXDB were manually removed by similarities of protein identifier, sequence length and quantitative iTRAQ data. The redundancy filtered results of the MASCOT identification and iTRAQ relative protein quantification for matches against the *Emiliania huxleyi* genome and the TAXDB were combined and annotated. Functional annotations, such as gene ontology. (GO) biological process, molecular function, cellular component, and GO IDs were retrieved from the UniProtKB ID mapping service (http://www.uniprot.org; Huntley et al., 2015). KOG annotations were anticipated to allow functional comparison to the transcriptome. KOG annotations from *E. huxleyi* JGI reference genome annotations yielded insufficient results for any functional conclusions.

Only 5 out of 403 total proteins had KOG annotations provided by JGI. Hence, GO annotations were the main feature to compare functional properties in the Light and Dark condition. For the functional annotation GO biological process only the first GO term was used to summarize the results. The entire proteome was also searched for proteins with a potential role in calcification based on protein description and GO annotations. The significantly expressed proteins in the Light and Dark condition were filtered and evaluated for dominant physiological processes. Furthermore, to find consensus of the proteomics and transcriptomics data (Chapter 3) the MASCOT proteins identified by the JGI *E. huxleyi* reference genome were investigated for consensus hits using the JGI gene IDs of 34403 transcripts (Chapter 3) and the UniProt IDs of transcripts using R.

4.3. Results

The experimental design was the same as described in Chapter 3 of this monograph. The same physiological parameters such as, PIC, POC, PON, growth rates, and nutrient consumption were relevant for the evaluation of the cell cultures. In the following section, only additional proteome related results of *E. huxleyi* in the highly calcifying late G1-phase and low calcifying early G1-phase, the light and dark incubation period, respectively, are presented. The results for cell physiological parameters, calcification, nutrient consumption, dissolved inorganic carbon, and irradiance levels are presented in subchapter 3.3.

4.3.1. Protein yield and integrity of protein extract

The extraction of proteins using TEAB-buffer yielded protein concentrations of 2.6 to 4.2 mg ml^{-1} (see Table 4-1) sufficient for iTRAQ labelling and MS/MS based peptide sequence assessment.

Table 4-1: Protein concentrations in extracts from the collected cell pellets from experimental cultures at the late G1-phase (highly calcifying cell, HC) and early G1-phase (low calcifying cells, LC)

Culture	Protein conc. in mg ml^{-1}
HC-1	3.1
HC-2	3.3
HC-3	3.4
LC-1	2.6
LC-2	4.2
LC-3	3.2

4.3.2. Peptide identification and annotation success

The following sections describe the general and more specific protein identification and annotation success, in respect to functional categories, significantly more relatively higher abundant proteins, and proteins potentially involved in the biomineralisation process in *E. huxleyi.*

4.3.2.1. General identification and annotation success

The MASCOT-identified peptides were mapped against the *E. huxleyi* genome (dataset JGI) and taxonomically restricted database (dataset TAXDB) (see section 4.2.7). In the identification pipeline a total of 496 hits to both reference databases were achieved. However, 55 of the hits to the JGI reference database did not return a BLAST hit ($e < 10^{-4}$). The remaining 441 proteins showing BLAST annotated findings recruited to a total of 329 protein clusters (compare Table 4-2). Differences in the total yield of identified proteins occurred in the iTRAQ experiment. Quantitative protein information was returned for 409 hits to JGI and TAXDB reference databases in total. The iTRAQ pipeline returned a total of 403 non-redundant hits, equivalent to 339 peptide matches, and recruiting to 287 protein clusters (for details see Table 4-2). In the following sections the evaluation of the quantitative iTRAQ proteomics data is central for the presentation.

Table 4-2: Protein identification success in the identification pipeline and iTRAQ experiment (nr : non redundant; JGI: matches against the JGI reference genome of 22.2.2012; TAX DB: matches against the taxonomically restricted data base created be Dr. Richard Edwards, see Subchapter 4.2.8)

	Identification pipeline			iTRAQ experiment			
Dataset	Total peptide hits	No Blast hits removed	Protein clusters	iTRAQ peptide hits	nr iTRAQ peptide hits	nr protein	nr protein clusters
JGI	406	351	241	330	327	263	235
TAXDB	90	90	88	79	76	76	65
Total	496	441	329	409	403	339	287

In the HC condition 77 peptides matching to known proteins out of 90 total hits were up-regulated. Higher abundance ratios of proteins in the LC condition were found in 26 cases, of which 23 returned protein matches. The majority of the total hits and annotated protein showed no significantly increased relative abundance in the HC or LC condition. This means that 26% of the identified proteins showed up or down regulation in both experimental conditions (for details see Table 4-3).

Table 4-3: Counts of significantly more abundant peptides in the HC and LC condition. (E.hux.G.: JGI *E. huxleyi* reference genome, Sig. higher rel. abund.: significant higher relative abundant proteins based on iTRAQ peptide ratios).

	Total hits	Hits with UniProt ID	Hits without UniProt ID	Hits in E. hux.G.	Hits in TaxaDB
HC	90	77	13	75	15
LC	26	23	3	19	7
Not sign.	287	239	48	233	54
Total	403	339	66	327	76

4.3.2.2.   Functional and localisation prediction annotation of proteins

Only 5 KOG annotations were returned from the JGI reference genome mapping results. Therefore, functional information based on gene ontology (GO) annotation was more abundantly available. The UniProt knowledge base GO annotation is evaluated at the level of biological process and cellular component. For the summary of the GO annotation success refer to Table 4-4.

Table 4-4: Success of gene ontology (GO) annotation of peptides in the HC and LC condition at the levels biological process and cellular component (Sig. higher rel. abund.: significant higher relative abundant proteins based on iTRAQ peptide ratios).

Sig. higher rel. abund.	Gene Ontology (GO)	Gene Ontology (biological process)	Gene Ontology (cellular component)
HC	70	60	44
LC	18	12	13
Not sign.	210	211	116
Total	298	211	173

The number of proteins attributed to the dominant biological processes based on GO annotations in the HC and LC condition are given in Table 4-5 and Table 4-6, respectively. Proteins involved in energy turnover processes such as photosynthesis, ATP hydrolysis glycolysis, carbohydrate metabolism, cell homeostasis, protein transport, and metabolism were activated in the HC condition (see Table 4-5 for details). In the dark incubation period biological processes involved in mitosis (e.g. microtubule-based process [GO:0007017], protein polymerization [GO:0051258], nucleosome assembly [GO:0006334]), protein synthesis (e.g. protein folding [GO:0006457], translation [GO:0006412]), amino acid synthesis (e.g. gluconeogenesis [GO:0006094], pyruvate metabolic process [GO:0006090], carboxylic acid metabolic process [GO:0019752]), and cell wall synthesis (terpenoid biosynthetic process [GO:0016114]) are indicated.

Table 4-5: Significantly overrepresented peptides in the HC condition by GO: biological processes. The number of significantly overrepresented peptides in the HC condition by Gene Ontology biological processes annotation is presented.

Gene Ontology – biological process [GO IDs]	Number of Peptides
Photosynthesis; light harvesting [GO:0009765]; protein-chromophore linkage [GO:0018298]	14
ATP hydrolysis coupled proton transport [GO:0015991]; ATP synthesis coupled proton transport [GO:0015986]	8
Glycolytic process [GO:0006096]	5
Protein folding [GO:0006457]	3
Carbohydrate metabolic process [GO:0005975]; glycerol-3-phosphate catabolic process [GO:0046168]	3
Glucose metabolic process [GO:0006006]; glycolytic process [GO:0006096]	2
ATP synthesis coupled proton transport [GO:0015986]	2
Cell redox homeostasis [GO:0045454]; cellular response to oxidative stress [GO:0034599]; glycerol ether metabolic process [GO:0006662]; oxidation-reduction process [GO:0055114]; protein folding [GO:0006457]; protein folding [GO:0006457]; sulphate assimilation [GO:0000103];	1
Photosynthesis; light harvesting [GO:0009765]	1
Protein transport [GO:0015031]; small GTPase mediated signal transduction [GO:0007264]	1
Mitochondrial respiratory chain complex III assembly [GO:0034551]	1
Carbohydrate metabolic process [GO:0005975]	1
Oxygen transport [GO:0015671]	1
Sulphate assimilation [GO:0000103]	1
One-carbon metabolic process [GO:0006730]	1
Glucose metabolic process [GO:0006006]	1
ATP hydrolysis coupled proton transport [GO:0015991]	1
Ubiquitin-dependent protein catabolic process [GO:0006511]	1
Gluconeogenesis [GO:0006094]; glycolytic process [GO:0006096]	1
Protein import [GO:0017038]; protein targeting [GO:0006605]	1
Arginyl-tRNA aminoacylation [GO:0006420]	1
ATP hydrolysis coupled proton transport [GO:0015991]; ATP metabolic process [GO:0046034]	1
Pyridoxal phosphate biosynthetic process [GO:0042823]; vitamin B6 biosynthetic process [GO:0042819]	1
Translation [GO:0006412]	1
Cell redox homeostasis [GO:0045454]; glutathione metabolic process [GO:0006749]	1
Peptidyl-pyrromethane cofactor linkage [GO:0018160]; tetrapyrrole biosynthetic process [GO:0033014]	1
Protein folding [GO:0006457]; response to stress [GO:0006950]	1
Reductive pentose-phosphate cycle [GO:0019253]	1
Photosynthetic electron transport in photosystem II [GO:0009772]; Protein-chromophore linkage [GO:0018298]	1
Carotenoid biosynthetic process [GO:0016117]	1

Table 4-6: Significantly overrepresented peptides in the LC condition by GO: biological processes. The number of significantly overrepresented peptides in the LC condition by Gene Ontology biological processes annotation is presented.

Gene Ontology – biological process [GO IDs]	Number of Peptides
Microtubule-based process [GO:0007017]; protein polymerization [GO:0051258]	3
Nucleosome assembly [GO:0006334]	2
Protein folding [GO:0006457]	1
Translation [GO:0006412]	1
Transmembrane transport [GO:0055085]	1
Protein catabolic process [GO:0030163]	1
Gluconeogenesis [GO:0006094]; pyruvate metabolic process [GO:0006090]	1
Terpenoid biosynthetic process [GO:0016114]	1
Carboxylic acid metabolic process [GO:0019752]	1

In general, the majority of up regulated proteins in highly calcifying cells were tagged to the chloroplast [GO:0009507] (n = 21). Fewer proteins related to the cytoplasm [GO:0005737] (n = 3), Golgi apparatus [GO:0005794] (n = 1), membrane [GO:0016020] (n = 1), mitochondrion [GO:0005739] (n = 1), and nucleosome [GO:0000786] (n = 1) were also identified (see Table 4-6). Interestingly, two proteins related to proton-transporting V-type ATPases [GO:0033179 and GO:003318] were identified in highly calcifying cells. For more details of gene ontology cellular component annotation in the HC condition refer to Table 4-7. In the dark incubation period (LC condition) the largest cohort of proteins recruited to the nucleosome [GO:0000786] (n = 6), mitosis in the form of microtubule [GO:0005874] (n = 2), Cytoplasm [GO:0005737] (n = 2). For further details of gene ontology cellular component annotation in the LC condition refer to Table 4-8.

Table 4-7: Significantly overrepresented peptides in the light incubation period by GO: cellular component. The number of significantly overrepresented peptides in the HC condition by Gene Ontology: cellular component annotation is presented.

Gene Ontology – cellular component [GO IDs] LIGHT	Number of Peptides
Chloroplast [GO:0009507]; integral component of membrane [GO:0016021]; photosystem II [GO:0009523]	14
Proton-transporting ATP synthase complex, catalytic core F(1) [GO:0045261]	6
Chloroplast thylakoid membrane [GO:0009535]; proton-transporting ATP synthase complex, catalytic core F(1) [GO:0045261]	4
Glycerol-3-phosphate dehydrogenase complex [GO:0009331]	3
Cytoplasm [GO:0005737]	2
Cell [GO:0005623]	1
Chloroplast [GO:0009507]	1
Chloroplast [GO:0009507]; integral component of membrane [GO:0016021]; photosystem II [GO:0009523]; thylakoid membrane [GO:0042651]	1
Chloroplast stroma [GO:0009570]; chloroplast thylakoid membrane [GO:0009535]	1
Chloroplast thylakoid membrane [GO:0009535]; integral component of membrane [GO:0016021]; light-harvesting complex [GO:0030076]; photosystem II [GO:0009523]	1
Clathrin coat of coated pit [GO:0030132]; clathrin coat of trans-Golgi network vesicle [GO:0030130]	1
Cytoplasm [GO:0005737]; nucleus [GO:0005634]; proteasome core complex, alpha-subunit complex [GO:0019773]	1
Golgi apparatus [GO:0005794]	1
Membrane [GO:0016020]	1
Mitochondrion [GO:0005739]	1
Nucleosome [GO:0000786]; nucleus [GO:0005634]	1
Phosphopyruvate hydratase complex [GO:0000015]	1
Plasma membrane [GO:0005886]	1
Proton-transporting V-type ATPase, V0 domain, across membrane transport [GO:0033179]	1
Proton-transporting V-type ATPase, V1 domain, across membrane transport [GO:0033180]	1
Ribosome [GO:0005840]	1

Table 4-8: Significantly overrepresented peptides in the dark incubation period by GO: cellular component. The number of significantly overrepresented peptides in the LC condition by Gene Ontology: cellular component annotation is presented.

Gene Ontology – cellular component [GO IDs] DARK	Number of Peptides
Nucleosome [GO:0000786]; nucleus [GO:0005634]	6
Microtubule [GO:0005874]	2
Cytoplasm [GO:0005737]	1
Cytoplasm [GO:0005737]; microtubule [GO:0005874]	1
Integral component of membrane [GO:0016021]	1
Nucleus [GO:0005634]	1
Small ribosomal subunit [GO:0015935]	1

The gene ontology tables considering all protein hits can be found in Appendix B Table 2 and Table 3.

4.3.2.3. Significantly expressed proteins in the highly and low calcifying *E. huxleyi* cells

The quantitative proteomics experiment using iTRAQ showed that 110 peptides with known UniProt IDs showed significant higher relative abundance in highly calcifying *E. huxleyi* cells and cells calcifying at low rates (see Table 4-3 and Table 4-9). Protein function and gene ontology frequencies show a similar general trend that distinct functional protein groups were up-regulation in either the light or dark incubation period (Section 4.3.2.2.). Proteins involved in photosynthesis and ATP synthesis were up regulated in the HC condition. Proteins suggested to be involved in the calcification process, namely clathrin (Dark:Light = 0.664; UniProt ID: C1E1W7; JGI407983) and subunits of V-type proton ATPases (e.g. UniProt ID: D8TKK9, JGI43077; Dark:Light = 0.813) were more abundant in the HC condition. In *E. huxleyi* cells calcifying at lowest rates, if at all, histones and tubulins were significantly upregulated.

Table 4-9: Significantly more abundant proteins in the HC and LC condition. Proteins showing significant relative up- and down-regulation of identified peptides at high calcification (HC condition) or in low calcifying cell (LC condition). If LC:HC < 1 (proteins in low calcifying cells versus proteins in high calcifying cells abundance ratio) proteins were more abundant in highly calcifying cells. The abundance ratio of all peptides is given in Appendix B. LC: HC sig. dir.: indicating significant more abundant peptides in the HC or LC condition. (DB ID: identifier for matches to TaxDB and JGI reference genome, Consensus: reference ID for match in HAQESAC. Shading of table rows indicates proteins with abundance ratio towards the LC condition).

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
JGI65061	Q21A28	Chloroplast light harvesting protein isoform 12	HC	0.594	0.433	0.751	HAJGI222
TAXDB13	Q4G3F4	Ribulose bisphosphate carboxylase large chain	HC	0.598	0.544	0.660	TAXDB13
JGI281147	Q5UU97	Enolase	HC	0.600	0.447	0.729	HAJGI283
JGI206028	Q21A28	Chloroplast light harvesting protein isoform 12	HC	0.632	0.487	0.763	HAJGI261
JGI66310	F0XZQ3	Putative uncharacterized protein	HC	0.641	0.416	0.911	HAJGI043
TAXDB73	A3FQF5	Uncharacterized protein	HC	0.647	0.528	0.799	TAXDB73
JGI269056	Q39709	Fucoxanthin-chlorophyll a-c binding protein	HC	0.654	0.565	0.845	HAJGI382
TAXDB14	Q4G396	Photosystem II CP43 reaction center protein	HC	0.681	0.581	0.801	TAXDB14
TAXDB22	B8C4K2	Uncharacterized protein	HC	0.691	0.443	0.950	TAXDB22
JGI62246	Q21A73	Chloroplast light harvesting protein isoform 6	HC	0.693	0.544	0.787	HAJGI246
JGI109974	NA	NA	HC	0.699	0.479	0.998	NA
JGI220283	NA	NA	HC	0.700	0.653	0.772	NA
JGI103600	D8LB65	Glucokinase	HC	0.703	0.522	0.884	HAJGI275
JGI67584	Q21A28	Chloroplast light harvesting protein isoform 12	HC	0.704	0.489	0.926	HAJGI264
JGI218041	P51821	ADP-ribosylation factor 1	HC	0.704	0.568	0.924	HAJGI036
JGI45869	Q21A76	Chloroplast light harvesting protein isoform 3	HC	0.705	0.538	0.937	HAJGI029
JGI104438	C1DZ32	Predicted protein	HC	0.710	0.540	0.843	HAJGI260
JGI281510	F0YM05	Putative uncharacterized protein	HC	0.710	0.589	0.853	HAJGI085
JGI96082	NA	NA	HC	0.718	0.594	0.805	NA
TAXDB75	Q5ENR8	Phosphoglycerate kinase	HC	0.720	0.647	0.818	TAXDB75
JGI90250	F0YFF1	Putative uncharacterized protein	HC	0.726	0.630	0.804	HAJGI405
JGI237243	NA	NA	HC	0.727	0.498	0.955	HAJGI191
JGI275754	Q5ENS1	Chloroplast phosphoribulokinase	HC	0.727	0.604	0.883	HAJGI167
JGI67108	Q21A72	Chloroplast light harvesting protein isoform 7	HC	0.728	0.596	0.974	HAJGI299

Table 4-9: Significantly more abundant proteins on the Light and Dark phase, continued.

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
JGI407983	C1E1W7	Clathrin heavy chain	HC	0.664	0.387	1.053	HAJGI365
JGI283485	B8CFP9	Predicted protein	HC	0.728	0.573	0.877	HAJGI071
JGI88932	NA	NA	HC	0.732	0.679	0.821	NA
TAXDB81	Q40612	ATP synthase subunit beta, chloroplastic	HC	0.735	0.622	0.864	TAXDB81
JGI99131	F0Y6E3	Putative uncharacterized protein	HC	0.744	0.589	0.858	HAJGI125
JGI406321	NA	NA	HC	0.751	0.696	0.870	HAJGI381
JGI120262	F0Y2W6	Putative uncharacterized protein	HC	0.756	0.613	0.881	HAJGI186
JGI120829	Q21A75	Chloroplast light harvesting protein isoform 4	HC	0.757	0.461	0.998	HAJGI025
JGI122970	Q01AH9	Heat shock protein 70 / HSP70 (ISS)	HC	0.758	0.664	0.861	HAJGI280
TAXDB85	D8TRA2	ATP synthase subunit beta	HC	0.760	0.672	0.836	TAXDB85
JGI216402	G4ZP25	Putative uncharacterized protein	HC	0.762	0.587	0.975	HAJGI372
TAXDB61	Q4G3C8	ATP synthase subunit beta, chloroplastic	HC	0.763	0.659	0.892	TAXDB61
JGI273928	G5A9A0	Putative uncharacterized protein	HC	0.765	0.673	0.870	HAJGI259
JGI103637	Q21A76	Chloroplast light harvesting protein isoform 3	HC	0.766	0.612	0.945	HAJGI035
JGI102701	Q21A76	Chloroplast light harvesting protein isoform 3	HC	0.767	0.614	0.953	HAJGI209
JGI271331	D0NEK2	Adenosylhomocysteinase	HC	0.774	0.760	0.792	HAJGI136
JGI407305	Q21A76	Chloroplast light harvesting protein isoform 3	HC	0.774	0.688	0.845	HAJGI008
JGI106019	A8I1X3	Putative uncharacterized protein	HC	0.777	0.536	0.996	HAJGI140
JGI112702	NA	NA	HC	0.780	0.617	0.961	NA
JGI358141	Q84LQ0	Cytosolic glyceraldehyde 3-phosphate dehydrogenase	HC	0.781	0.661	0.950	HAJGI335
JGI230835	Q21A55	Chloroplast glyceraldehyde-3-phosphate dehydrogenase	HC	0.782	0.706	0.850	HAJGI027
JGI102901	Q5QD45	Plastid C1 class II fructose bisphosphate aldolase	HC	0.786	0.684	0.855	HAJGI079
TAXDB41	Q4G3C9	ATP synthase epsilon chain, chloroplastic	HC	0.797	0.687	0.949	TAXDB41

Table 4-9: Significantly more abundant proteins on the Light and Dark phase, continued.

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
TAXDB05	Q4G397	ATP synthase subunit alpha, chloroplastic	HC	0.798	0.720	0.909	TAXDB05
JGI88203	A6YAZ8	Putative plastid transketolase	HC	0.799	0.750	0.880	HAJGI1397
JGI212512	G5A9A0	Putative uncharacterized protein	HC	0.800	0.712	0.887	HAJGI1121
TAXDB88	Q21A54	Glyceraldehyde-3-phosphate dehydrogenase	HC	0.808	0.758	0.878	TAXDB88
JGI199538	NA	NA	HC	0.810	0.717	0.976	HAJGI237
JGI43077	D8TKK9	Vacuolar H+ ATPase V0 sector, subunit D	HC	0.813	0.751	0.887	HAJGI1048
JGI120592	Q0MYX3	Proteasome subunit alpha type	HC	0.817	0.670	0.994	HAJGI1098
JGI61340	Q21A74	Chloroplast light harvesting protein isoform 5	HC	0.818	0.724	0.934	HAJGI380
JGI87079	B9PNV8	Heat shock protein, putative	HC	0.819	0.723	0.935	HAJGI1162
JGI270399	Q15GC7	Triosephosphate isomerase	HC	0.823	0.725	0.995	HAJGI1387
JGI88977	A4RW83	Protein translocase subunit SecA, chloroplastic	HC	0.824	0.697	0.938	HAJGI1402
JGI366337	A8J387	Predicted protein	HC	0.828	0.724	0.919	HAJGI266
JGI235106	NA	NA	HC	0.831	0.739	0.965	HAJGI138
JGI222386	P48414	V-type proton ATPase catalytic subunit A	HC	0.833	0.681	0.971	HAJGI1081
JGI272624	A4S2C6	Predicted protein	HC	0.833	0.703	0.994	HAJGI1037
JGI96943	C1EIH7	Putative uncharacterized protein	HC	0.836	0.726	0.931	HAJGI1267
JGI198360	C5KNV3	Molecular chaperone, putative	HC	0.836	0.726	0.931	HAJGI1007
JGI206438	D0NBZ5	Nucleolar protein NOP5	HC	0.839	0.695	0.961	HAJGI1097
JGI265597	G5A9A0	Putative uncharacterized protein	HC	0.839	0.773	0.940	HAJGI1205
TAXANA	C5K6Q8	Heat shock protein, putative	HC	0.841	0.662	0.987	NA
JGI102800	G0QQ50	Putative uncharacterized protein	HC	0.843	0.671	0.981	HAJGI1019
JGI254140	NA	NA	HC	0.843	0.696	0.996	HAJGI1026
JGI100088	NA	NA	HC	0.845	0.751	0.905	HAJGI1349
JGI218081	Q21A28	Chloroplast light harvesting protein isoform 12	HC	0.846	0.787	0.948	HAJGI1242
JGI350570	NA	NA	HC	0.847	0.709	0.949	NA
JGI202968	F0YD68	Putative uncharacterized protein	HC	0.848	0.749	0.969	HAJGI291
JGI252355	F0XWG3	Putative plastid light harvesting protein isoform 40	HC	0.851	0.806	0.910	HAJGI1030

Table 4-9: Significantly more abundant proteins on the Light and Dark phase, continued.

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
JGI209460	G4YJL5	Putative uncharacterized protein	HC	0.854	0.577	0.994	HAJGI220
JGI402650	C1EHC0	ATP synthase subunit alpha	HC	0.855	0.754	0.951	HAJGI377
JGI360985	G4Z7L2	Putative uncharacterized protein	HC	0.865	0.713	0.969	HAJGI157
TAXDB90	G0QRF6	ATP synthase subunit alpha	HC	0.865	0.741	0.998	TAXDB90
TAXDB48	Q49JI1	Histone H2A	HC	0.867	0.789	0.927	TAXDB48
JGI99724	D8LHC1	Glutathione reductase	HC	0.871	0.730	0.979	HAJGI386
JGI360891	NA	NA	HC	0.871	0.770	0.970	NA
JGI272626	F0YAX4	Putative uncharacterized protein	HC	0.874	0.774	0.962	HAJGI038
JGI89571	Q5ENR5	Phosphoglycerate kinase	HC	0.877	0.768	0.973	HAJGI1147
JGI270342	D8U3I9	Putative uncharacterized protein	HC	0.877	0.752	0.987	HAJGI103
JGI266768	B8C830	Thioredoxin	HC	0.882	0.766	0.978	HAJGI200
JGI276141	D0NY29	60S ribosomal protein L3	HC	0.883	0.765	0.968	HAJGI073
JGI105420	Q5DK81	Heat shock protein 90	HC	0.887	0.811	0.964	HAJGI092
JGI197716	Q21A11	ATP synthase gamma chain	HC	0.906	0.870	0.964	HAJGI367
TAXDB77	D7FXG1	ATP synthase subunit beta	HC	0.924	0.876	0.978	TAXDB77
JGI412390	C1FE16	ATP synthase subunit beta	HC	0.938	0.835	0.998	HAJGI1058
JGI97545	B6KVV6	Proliferation-associated protein 2G4, putative	HC	0.957	0.929	0.999	HAJGI1017
JGI266878	F0YE54	Putative uncharacterized protein	LC	1.081	1.007	1.205	HAJGI319
TAXDB02	C1FFK0	Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)	LC	1.107	1.010	1.215	TAXDB02
JGI105146	C1E0B7	Elongation factor tu gtp-binding domain protein 2	LC	1.130	1.049	1.209	HAJGI1104
JGI64083	F0YST3	Putative uncharacterized protein	LC	1.133	1.027	1.267	HAJGI289
JGI231324	D0NEH7	26S protease regulatory subunit 7	LC	1.137	1.002	1.312	HAJGI235
JGI365196	B6DX96	Putative chloroplast 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	LC	1.141	1.022	1.259	HAJGI268
JGI71008	NA	NA	LC	1.161	1.007	1.372	NA
JGI95565	Q49JJ2	Histone H2A	LC	1.161	1.017	1.368	HAJGI144

Table 4-9: Significantly more abundant proteins on the Light and Dark phase, continued.

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
JGI267796	Q1WLZ2	40S ribosomal protein S3	LC	1.165	1.082	1.264	HAJGI296
JGI279740	C1FFE2	Predicted protein	LC	1.161	1.007	1.372	HAJGI074
JGI256830	NA	NA	LC	1.168	1.100	1.280	HAJGI208
JGI271455	F0YFV2	Putative uncharacterized protein	LC	1.170	1.034	1.441	HAJGI185
JGI96192	Q013K3	Histone H2B	LC	1.207	1.002	1.403	HAJGI401
TAXIDB65	F0Y8U4	Putative uncharacterized protein (Fragment)	LC	1.213	1.052	1.397	TAXIDB65
JGI76377	B7G0V7	Predicted protein	LC	1.244	1.030	1.389	HAJGI1143
JGI201707	A4RS62	Histone H4	LC	1.277	1.094	1.524	HAJGI379
JGI255477	A7AT84	Histone H3, putative	LC	1.281	1.063	1.502	HAJGI111
JGI72235	C1MVY5	Histone H2A	LC	1.305	1.156	1.477	HAJGI082
TAXIDB19	Q00Y53	Histone H2B	LC	1.333	1.076	1.631	TAXIDB19
TAXIDB71	D0MWJ7	Histone H2A	LC	1.362	1.181	1.524	TAXIDB71
TAXIDB28	C1E6T8	Histone H4	LC	1.373	1.280	1.580	TAXIDB28
JGI277202	Q5G920	Alpha-tubulin	LC	1.476	1.285	1.821	HAJGI1156
TAXIDB53	Q5G917	Alpha-tubulin (Fragment)	LC	1.504	1.242	1.709	TAXIDB53
JGI374572	NA	NA	LC	1.550	1.147	1.916	HAJGI1158
JGI87697	F0Y9S1	Putative uncharacterized protein	LC	1.577	1.383	1.764	HAJGI307
TAXIDB72	F0XW25	Tubulin alpha-2 chain	LC	2.388	1.484	3.448	TAXIDB72

4.3.2.4. Proteins of interest involved in biomineralisation in highly calcifying and low calcifying *E. huxleyi* cells

The abundance ratios for all identified proteins can be found in Appendix B Table 1. The entire population of identified proteins was investigated for proteins involved in biomineralisation at the gene ontology level and by protein name. The results show that four proteins potentially involved in *E. huxleyi* biomineralisation were relatively more abundant in calcifying cells of the late G1 cell cycle phase (see Table 4-10 and Table 4-11). The peptides related to biomineralisation matched three proton V-ATPases and one clathrin (see section 4.3.2.3). Gene ontology suggested three proteins involved in the calcification process, whereas two are predicted and uncharacterized. Furthermore, a putative alpha-actinin-1 and oxygen evolving enhancer 1 (JGI265795, UniProt ID: Q8L878; Dark: Light = 0.944) was suggested by gene ontology to play a role in calcification. In cells showing low rates of calcification (LC condition) (JGI271455, UniProt ID: F0YFV2; LC: HC = 1.1170) was indicated to be involved in calcification trusting GO annotations. However, WolfsPSort (Horton et al., 2007) predicts the location of the uncharacterised protein in the chloroplast.

Table 4-10: Proteins with potential role in calcification in the HC and LC condition. Known proteins involved in biomineralisation observed in the proteomes of the HC and LC conditions. LC: HC sig. dir.: indicating significantly more abundant peptides in the HC or LC condition. (DB ID: identifier for matches to TaxDB and JGI reference genome, Consensus: reference ID for match in HAQESAC).

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
JGI43077	D8TKK9	Vacuolar H+ ATPase V0 sector, subunit D	HC	0.813	0.751	0.887	HAJGI048
JGI222386	P48414	V-type proton ATPase catalytic subunit A	HC	0.833	0.681	0.971	HAJGI081
JGI276075	D8U1L3	Vacuolar ATP synthase subunit E		0.894	0.772	1.047	HAJGI202
JGI407983	C1E1W7	Clathrin heavy chain	HC	0.664	0.387	1.053	HAJGI365

Table 4-11: GO: calcification proteins in the HC and LC condition. Proteins with "calcification" as a keyword in gene ontology annotation in the proteomes of the HC and LC condition. LC: HC sig. dir.: indicating significantly more abundant peptides in the HC or LC condition. (DB ID: identifier for matches to TaxDB and JGI reference genome, Consensus: reference ID for match in HAQESAC). Shading of table rows indicates proteins with abundance ratio towards the LC condition.

DB ID	UniProt ID	Protein description	LC:HC sig. dir	LC:HC	LC:HC Min	LC:HC Max	Consensus
JGI408848	D0NLH6	Alpha-actinin-1, putative		0.918	0.759	1.122	HAJGI110
JGI113436	A8IJG5	Predicted protein		0.942	0.789	1.101	HAJGI262
JGI265795	Q8L878	Oxygen evolving enhancer 1		0.944	0.765	1.224	HAJGI298
JGI271455	F0YFV2	Putative uncharacterized protein	LC	1.170	1.034	1.441	HAJGI185

4.3.2.5.   Consensus of transcriptome and proteome

To find consensus of the proteomics and transcriptomics data (Chapter 3) the MASCOT peptides identified by the JGI *E. huxleyi* reference genome and TAXDB were compared as described in section 4.2.8. Out of 496 protein hits only 51 were found to show a consensus transcript based on the JGI gene ID. A total of 20 matches based on protein identifies (UniProt ID) were found in the proteome and transcriptome. After removing duplicates in the combined results based on JGI gene ID 35 matches remained, of which 31 had relative expression / abundance data. Of those 31 matches 12 showed the same direction towards the HC or LC condition, whereas only 2 showed higher abundance in cells with low calcification rates. The expression direction was found significantly different only in one case; the chloroplastic fucoxanthin-*chl a-c* binding protein (JGI ID: 269056; UniProt ID: Q39709) was more abundant in the HC condition. The protein matches in the proteome and transcriptome are presented in Table 4-12. None of the proteins match by the proteome against the transcriptome appears to be involved in biomineralisation.

Table 4-12: Proteome – transcriptome consensus. Matching hits (proteins) in the proteome and transcriptome datasets are presented. Abundance ratios for the proteins and transcripts are presented. dir. prot: direction of abundance ratio of proteins. LC:HC: peptide ratio average, LC:HC min.: peptide ratio minimum, LC:HC max.: peptide ratio maximum, log 2x: log 2-fold ratio of transcripts, dir transcr.: direction of transcript abundance, sig.: significantly different transcript abundance. Italic and bold JGI ID mark matches where the direction of abundance ratio of proteins and transcripts is the same.

JGI ID	UniProt ID	Description	dir prot.	LC: HC	LC:HC min.	LC:HC max.	log2 x	dir transcr.	sig.
67108	Q2IA72	Chloroplast light harvesting protein isoform 7	HC	0.728	0.596	0.974	0.7	LC	no
243741	D8TS49	Component of cytosolic 80S ribosome and 60S large subunit		0.872	0.596	1.353	-1.28	HC	no
240795	G4ZPW0	Elongation factor Tu		NA	NA	NA	-0.14	HC	no
269056	Q39709	Fucoxanthin-chlorophyll a-c binding protein, chloroplastic	HC	0.654	0.565	0.845	-2.25	HC	yes
198360	C5KNV3	Molecular chaperone, putative		0.836	0.726	0.931	-0.5	HC	no
206438	D0NBZ5	Nucleolar protein NOP5		0.839	0.695	0.961	-1.14	HC	no
193927	B8BU34	Predicted protein		0.88	0.618	1.669	1.78	LC	yes
102657	A4RTV8	Predicted protein		0.895	0.799	1.055	-1.26	HC	no
96453	B8LEJ5	Predicted protein		0.951	0.615	2.008	5.22	LC	yes
200165	B7G9A2	Predicted protein		0.98	0.808	1.164	-0.51	HC	no
108180	C1N4F5	Predicted protein	LC	1.058	0.871	1.346	-1.27	HC	no
279740	C1FFE2	Predicted protein		1.161	1.007	1.372	-1.07	HC	no
233632	B8CAB4	Predicted protein		1.795	0.588	6.25	-3.22	HC	yes
351398	B5Y3N7	Predicted protein		NA	NA	NA	1.87	LC	yes
108073	Q50KB1	Protein disulfide-isomerase-like protein EhSep2		1.087	0.781	1.546	-2.4	HC	no
88977	A4RW83	Protein translocase subunit SecA, chloroplastic	HC	0.824	0.697	0.938	1.84	LC	yes
281510	F0YM05	Putative uncharacterized protein	HC	0.71	0.589	0.853	4.01	LC	yes
106019	A8I1X3	Putative uncharacterized protein	HC	0.777	0.536	0.996	-2.48	HC	no
202968	F0YD68	Putative uncharacterized protein	HC	0.848	0.749	0.969	0.39	LC	no
212481	D7FJQ9	Putative uncharacterized protein		0.874	0.735	1.167	1.92	LC	yes
365947	E1ZDS1	Putative uncharacterized protein		0.885	0.690	1.117	-1.21	HC	no

Table 4 12: Proteome – transcriptome consensus continued:

JGI ID	UniProt ID	Description	dir prot.	LC: HC	LC:HC min.	LC:HC max.	log2 x	dir transcr.	sig.
258285	E1Z343	Putative uncharacterized protein		0.900	0.706	1.116	-0.83	HC	no
273308	B4ZG44	Putative uncharacterized protein		0.91	0.781	1.054	4.67	LC	yes
269305	F0Y0B9	Putative uncharacterized protein		0.973	0.829	1.23	3.84	LC	yes
216106	D8TLK9	Putative uncharacterized protein		1.024	0.943	1.073	0.78	LC	no
284352	D7FSF0	Putative uncharacterized protein		1.122	0.847	1.776	2.67	LC	yes
87697	F0Y9S1	Putative uncharacterized protein	LC	1.577	1.383	1.764	-3.15	HC	yes
61467	F0Y333	Putative uncharacterized protein		NA	NA	NA	0.79	LC	no
215132	D7G200	Putative uncharacterized protein		NA	NA	NA	-1.91	HC	no
237243	no match	NA		0.727	0.498	0.955	0.79	LC	no
199538	no match	NA		0.810	0.717	0.976	1.63	LC	yes
105039	no match	NA		0.875	0.750	1.015	-0.27	HC	no
227764	no match	NA		0.972	0.509	1.555	1.5	LC	yes
100546	no match	NA		0.978	0.791	1.153	0.2	LC	no
220968	no match	NA		1.199	0.805	1.862	1.56	LC	yes

4.4. Discussion

To date five studies were investigating the proteome of the globally important marine calcifying coccolithophore *Emiliania huxleyi*. Pioneering work on the *E. huxleyi* proteome by Jones et al. (2011) identified 99 proteins operating from a variety of physiological pathways applying a similar protocol for protein extraction and identification. Protein identification was improved to 115 (Jones et al., 2013) by studying different *E. huxleyi* strains and the proteome response to elevated CO_2 conditions. In response to different light regimes 500 highly abundant and confident proteins were identified (McKew et al., 2013). Furthermore, Rose et al. (2014) explored microdomains in the host virus interaction of *Emiliania huxleyi* virus infections using shotgun proteomics and observed proteins specifically associated with cellular stress, host defence, programmed cell death and immunity pathways. Alcolombri et al. (2015) were interested identifying the genetic origin of the enzymes for responsible forming dimethyl sulphide (DMS) from dimethylsulfoniopropionate (DMSP) in *Emiliania huxleyi*. DMS is an important signal in marine food webs and can influence local weather by affecting cloud formation. Shotgun proteomics and transcriptomics were combined and identified the *Alma1* gene in *E. huxleyi*, which yielded high DMSP-lyase activity.

In this study, the quantitative proteome characteristics of the unicellular coccolithophore *E. huxleyi* PLY♯ M217 at high and low calcification rates of the G1-phase of the cell cycle were investigated to determine abundance patterns of protein potentially involved in biomineralisation. The proteome study also sought to find consensus with the transcriptome (Chapter 3).

A total of 403 different peptides were identified during the iTRAQ experiment. Of those 339 were matching known non-redundant proteins in the UniProt database. The proteins comprised of 287 protein clusters, which represent a significant improvement in protein identification yield in comparison to Jones et al., 2013. The results showed that the proteomes of the late G1-phase (high calcification) and early G1-phase (low calcification) showed significant differences in relative protein abundance; however, only five proteins involved in calcification processes were identified and quantified. Both null-hypotheses that there are no differences in proteomes of the highly calcifying and low calcifying *E. huxleyi* cells and proteins potentially involved in the process of calcification are not expressed in the calcification period (late G1-phase) must be rejected. Significant up-regulation of calcification related proteins was accounted for three proteins of clathrin and V-type proton ATPases in the calcification phase. The V-type ATPases were suggested for proton pumping at the coccolith producing membrane (Corstjens et al.,

2001) to establish a flux of protons into the cytosol to establish high pH conditions in the CV (Mackinder et al., 2010). Furthermore, the identified clathrin (UniProt ID: C1E1W7; JGI407983) is known to play a role in vesicle trafficking protein of trans-Golgi vesicle transport (Kirchhausen, 2000) and occurs together with clathrin-coated vesicles (Forgac, 2000). Jones et al. (2011) suggested an importance of clathrin in the biomineralisation process because it occurs together with V-type ATPase proton transporter in clathrin-coated vesicles (Forgac, 2000). Furthermore, clathrin was abundant in the proteome of the silica-precipitating diatom *T. pseudonana* (Nunn et al., 2009). Hence, vesicle mediated transport of minerals from the Golgi network and V-type ATPase forced alkanilisation could be a more general model for biomineralisation if it is active in diatoms and coccolithophores alike. Furthermore, clathrin was found in higher abundance in phosphate starved cultures of the brown alga *Aureococcus anophagefferens*, which could suggest that clathrin coated vesicles may act in nutrient scavenging from the environment (Wurch et al., 2011). The growth of the experimental *E. huxleyi* cultures and observed nutrient concentrations, as documented in Chapter 3, did not imply nutrient limitation. Therefore, an expression of clathrin in response to nutrient limitation is not suggested and the bulk expression of clathrin is most likely related to clathrin-coated vesicle construction.

Further exploration of the proteome using the keyword "calcification" in GO annotations revealed additional proteins. Any role of these proteins selected by their gene ontology annotation in coccolithogenesis has not been previously suggested. The α– actinin (JGI40884, D0NLH6) has a known affection for Ca^{2+} – ions and has an important role in cytogenesis and muscle function (Jayadev et al., 2014). The abundance ratio between the cell of the early and late G1-phase showed no significant difference and so a role for α– actinin in the biomineralisation processes could not be inferred. The oxygen evolving enhancer 1 (JGI265795, UniProt ID: Q8L878, protein name: *psbO*) is part of a complex, comprised of a cluster of manganese, calcium and chloride ions binding extrinsic proteins (Bricker et al., 2012), which catalyse the splitting of water to O_2 and $4 H^+$ (GO functional annotations, source www.uniprot.org; Huntley et al., 2015). The location of *psbO* is at the thylakoid membrane facing the thylakoid lumen (Ferreira et al., 2004). The binding of calcium ions within the oxygen-evolving centre is fundamental for the mechanism of water oxidation (Siegbahn, 2002). Therefore, it can be concluded that Q8L878 plays no role in biomineralisation and is closely involved in photosynthetic processes.

In contrast to Jones (2010) Histone 4 (H4) and other histones were up regulated at low calcification rates. Histone 4 was down regulated in *E. huxleyi* PLY♯M219 under high pCO_2

levels (Jones, 2013). However, both observations are not in disagreement if under elevated CO_2 the cell cycle and culture growth rates are slowed and the cells undergo mitosis less frequently. Histone 4 is a structural component of the nucleosome and is a methyl donor involved in transcription and translational processes (Chiang et al., 1996). Therefore, *E. huxleyi* cells could require less newly synthesized histones for nucleosome construction under high CO_2 (Jones, 2013).

Heat shock proteins play an important role in the proteome response to environmental stressors, immune response, and cell signalling (Pockley, 2003; Wang et al., 2004). Here, heat shock proteins are discussed because in previous proteome studies of *E. huxleyi* heat shock proteins were observed (Jones et al., 2011; Jones et al., 2013, McKew et al., 2013) and nutrient depletion stress can be inferred from their expression values, whereas higher expression of HSP 70 was found in phosphate limited cultures (Wahlund et al., 2004b). Heat shock protein (HSP) expression responded when *E. huxleyi* cells were exposed to higher light intensities (McKew et al., 2013). An elevation in HSP expression was also observed in the light incubation period in this study, as 3 out of 10 observed HSPs were significantly upregulated in the HC condition. Periodic expression patterns of HSPs in *E. huxleyi* have not been reported before. Human HSPs expression was regulated and variable during the cell cycle, whereas a 10-to15-fold expression increase of HSP 7O mRNA was found upon entry into S-phase; levels of HSP 70 expression decreased by the late S and G2 phase (Milarski & Morimoto, 1986). Therefore, the observed higher presence of heat shock proteins in the late G1-phase could be a result of the chosen sampling times. The *E. huxleyi* cultures in this experiment were harvested in the late G1 and early G1 phase, and the observed abundance of HSP could reflect the relative heat shock protein abundance at these cell cycle phases. Furthermore, heat shock proteins are also involved in non-stress related protein folding and housekeeping processes (Picard, 2002). The very balanced expression of 8 peptides matching HSPs in the late and early G1-phase supports the housekeeping function of HSP in *E. huxleyi* (see Appendix A Table 11). The maturation of specific protein complexes also requires the presence of HSPs. For example, during the insertion of the precursor of the major light-harvesting complex of photosystem II into the thylakoid chloroplastic HSP 70 was found to play a key role (Yalosky et al, 1992). Because photosynthesis-related proteins were most abundant in the Light condition it can be speculated that a bulk of the expressed heat shock proteins was involved in protein folding processes of photosynthesis related proteins. However, with the available data the dominant cause of higher HSP expression in the late G1-phase remains unclear. It would require additional sampling points throughout the cell cycle to improve the understanding of HSP expression in the cell cycle of *E. huxleyi*, which was not the objective of this study.

The observed overexpression of proteins related to pathways such as photosynthesis, glycolysis, vesicle transport, and cell repair in the Light period was conclusive. The majority of the identified proteins in the HC condition were associated with the chloroplast and inner cell membranes, and for a minority of proteins a function in cytoplasm or the Golgi-network was suggested. Considering that a large proportion of the biomineralisation proteins would be active within endomembranes (e.g. V-type proton ATPases), finding only four proteins potentially involved in biomineralisation could indicate insufficient extraction of proteins, issues with salt contamination of the sample, or insufficient dynamic range of the mass spectrometer. Previously, protein extraction of *E. huxleyi* using TEAB buffer and subsequent proteome analysis revealed a maximum of 115 homologous protein groups when searched against JGI *E. huxleyi* genome, the taxonomically restricted UniProt KB database, and an EST database were combined. Of those 115 homologues protein groups 46 showed differential expression using iTRAQ (Jones et al., 2013). This study, using only JGI *E. huxleyi* genome (version from 22.2.2012) and the taxonomically restricted database showed 105 homologous proteins with 35 homologous proteins being differentially expressed. However, the JGI-genome search by Jones et al. (2013) yielded only 40 hits compared to 327 proteins (287 protein clusters) in this study, using the same version of the reference genome. Previously, limitations in the protein extraction using TEAB buffer were discussed (Jones, 2011); because of low protein recovery rates and observed smearing in gel electrophoresis. In this study an optimized protocol to achieve a 4 times faster cell harvesting time in comparison to Jones et al. (2011) was implemented, which might explain higher rates of protein matches and identification. However, *E. huxleyi* TEAB buffer protein extraction appears to alter protein recovery when compared with SDS/TRIS-buffer (Qiagen) extraction (McKew et al., 2013). The SDS/TRIS-buffer would allow for simultaneous extraction of DNA, RNA, and proteins and appears to improve the protein extraction greatly. McKew and colleagues identified 1835 proteins and 500 with high confidence against the newly published JGI *E. huxleyi* genome using label free protein quantification (McKew et al., 2013; Read et al., 2013). The iTRAQ – technique is not compatible with reagents containing primary amines and the RLT Buffer was not considered as a suitable extraction buffer for the iTRAQ experiment, as not recommended by the iTRAQ manufacturer (Applied Biosystems, Foster City, U.S.A). Furthermore, the matrix-assisted laser desorption ionization technique produced higher noise to sample ratios if salty elutes are utilized (Yates et al., 2009). Salt were removed from the protein sample using desalination column. However, non-covalently binding salt in the protein sample could have carried salts over to the matrix-assisted laser desorption ionization and mass spectrometric analysis. The utilized LC/LC-MS/MS (MALDI QqTOF, Global Ultima 2, Waters) has a lower dynamic rate, lower protein discovery rates, and mass accuracy than an Orbitrap mass spectrometer (Yates et

al., 2009). A hybrid high-resolution LTQ/Orbitrap Velos instrument (Thermo Scientific) was used by McKew et al., 2009, which may explain the higher protein discovery rates in the McKew study.

   To exclude the possibility that the different protein discovery rates relate to the different versions of the reference genome, MS/MS data were matched against the newly published JGI *E. huxleyi* gene set (JGI; http://www.jgi.doe.gov, downloaded 15.6.2015, file: "Emihu1_best_proteins.fasta"). Preliminary results suggested that the more recent reference genome yielded 507 genome hits (see Appendix B Table 4), which is only around 28% of the identification success achieved by McKew et al. (2013) but an improvement of 25 % more identification compared with the older version of the reference genome. Therefore, using the newly annotated *E. huxleyi* genome improves the yield of protein identification but may only explain the differences in protein identification success between this study and McKew et al. (2013) to some degree. Preliminary evaluation suggests that the yield of peptide identification on the data presented in this study could be further improved by employing the BUDAPEST pipeline (Bioinformatics Utility for Data Analysis of Proteomics using ESTs provided by Dr. Richard Edwards and referenced in Jones et al., 2011; Jones et al., 2013) using an extended set of reference genome databases. The extended in silico identification suggests 1372 peptide matches (data not presented in this thesis). The identified 1372 proteins from searches against the JGI genome, taxonomically restricted databases, metatranscriptomic databases, *E. huxleyi* EST databases, and non-redundant protein sequences (Uniprot database) are believed to contain redundancies that were not removed at the time of writing. Preliminary, results suggest that 350 homologous protein groups were identified, which would improve the protein clustering by 12 %. Clustering the protein matches from the matches against the JGI and TAXDB references returned 287 protein clusters. Potentially, extended sequence database searches and improvement of the protein extraction protocol, or a label free MS/MS protein quantification approach could further improve yield of the *E. huxleyi* proteome. In the future, modifications of protein extract preparation, selection of lysis-buffer, and the utilized MS protein identification and quantification technology (e.g. Orbitrap) should be considered to achieve greater proteome coverage and reveal additional proteins relevant to biomineralisation in *E. huxleyi*.

Consensus of transcriptome and proteome

   The transcriptome and proteome of *Emiliania huxleyi* were evaluated with a focus on genes significantly overexpressed and potentially involved in biomineralisation. Based on the protein ID, a total of 35 genes were represented in both the transcriptome and proteome reflecting a 7 %

success in identifying matches of proteins and transcript. Frith et al. (2005) note that translation of only 5-10 % of mRNA occurs. Unfortunately, none of the matched genes in the transcriptome and proteome indicated proteins potentially involved in calcification. The indicated low rate of finding matches based on the JGI gene identifier (genetic sequence) could imply high levels of post-translational modification. The annotated transcripts presented in Chapter 3 of this monograph were related to the EuKaryotic Orthologous group class (KOG-class) posttranslational modification, protein turnover chaperones in *both* cell showing high rates of calcification and low rates of calcification, HC and LC condition, respectively. In the Haptophyta *Prymnesium parvum* most of the discovered transcripts also belonged to the KOG-class posttranslational modification, protein turnover chaperones (Claire, 2006). Therefore, higher rates of posttranslational modifications could indicate a more general characteristic of Haptophyta. However, it must be noted that the protein sequences were not directly matched against the transcripts of the Illumina experiment described in Chapter 3. Here, the consensus finding was based solely on gene identifier (JGI ID) and protein identifier (UniProt ID). A sequence based matching of proteome and transcriptome data is future work.

A closer investigation of matching transcripts and proteins potentially involved in calcification in *E. huxelyi* based on the protein function showed that V-type proton ATPases and clathrin were represented in the transcriptomes and proteomes of the HC and LC condition. Vacuolar proton ATPases and clathrin were previously observed in the proteome of *E. huxleyi* (Jones et al., 2011; Jones et al., 2013). Transcripts and peptides matching V-type proton ATPase were more present in the calcification phase. The proteomes showed an increased relative abundance of clathrin in the calcification phase. However, the same gene or protein for clathrin was not matched in the transcriptome and proteome. Furthermore, clathrin transcripts were less abundant in the HC condition, 355 fpkm compared to 471 fpkm transcripts in the LC conditon. In the case of clathrin protein and transcript abundance patterns in the HC and LC condition did not agree. The high presence of clathrin would support the model of Golgi-derived vesicle mediated growth of the coccolithosome (Jones et al., 2011). The transcriptome showed further transcripts with high abundance in the calcification phase relating to the vesicle coating complexes COP I and COP II (e.g. JGI ID: 432215, 57758). The COPI involve clathrin to transport molecules between the Golgi and the endoplasmic reticulum (McMahon and Mills, 2004). The cell division cell cycle homolog 48 was observed by Jones (2010) in the proteomes of *E. huxleyi* and suggested that it could be operating in vesicular transport from the transitional ER to the Golgi. Cell cycle homolog 48 was also observed in the proteomes of the HC and LC condition, the late and early G1-phase of the cell cycle. The transcriptome showed 68 transcripts

transcribing for 51 different proteins that are involved in vesicle transport (data not presented in this thesis). The majority of the transcripts related to proteins such as clathrin, vesicle complex COP I and COP II, and alpha SNAP, a protein required for vesicle fusion. Vesicle transport is important for coccolithogenesis and it appears that many proteins may be present in *E. huxleyi*, which could be involved in vesicle formation, vesicular transport, vesicle fusion, and coccolithosome merging. However, clathrin and other vesicle transport associated vesicle play a role in numerous cellular processes.

Conclusions and outlook

  This study presents the first results of proteomes of *Emiliania huxleyi* at different calcification rates during the G1-phase of the cell cycle. Relative protein quantification shows clear differences between the highly calcifying and the poorly calcifying phase. Photosynthesis and calcification related proteins were present in the calcification phase. V-type proton ATPase proteins and the calcium binding proteins clathrin relevant for biomineralisation were identified in the calcification phase. The number of matches of proteins and transcripts was low; however within the range of the suggested 5-10 % of transcripts being translated to proteins (Frith, et al., 2005). Transcripts in the cells were represented by 7 % of the identified proteins. The direction of transcript and protein abundance agreed in 34 % of the cases. Albeit the yield of protein identification using TEAB-buffer for protein extraction was improved compare to previous studies (Jones et al., 2011; Jones et al., 2013), the observed proteome is likely to present only a fraction of the actual proteome of *Emiliania huxleyi*. Future work should address matching the peptide information directly to the observed transcriptome and aim to apply the most state of the art mass spectrometry proteomics technique for protein identification. The utilization of the Orbitrap technology would greatly improve the protein discovery. Furthermore, label-free quantification methods may provide a more elaborate choice of protein extraction buffer to increase the knowledge of diurnal proteomes in coccolithophores. These initial novel results of diurnal proteome characteristics in *E. huxleyi* should encourage to conduct further experiments to extend the knowledge of cell cycle proteomes using additional sampling time-points and improved protein extraction, identification, and quantification protocols.

Chapter 5.   Synopsis and General Discussion

Project summary:

Identifying how marine organisms respond to changes in their environment at the physiological level is fundamental to understand processes in nature, the ecology of organisms, and their interactions at larger environmental scales. The motivation for this study was to investigate the calcification processes of the coccolithophore *Emiliania huxleyi* because of its importance for global biogeochemical cycles and to increase the understanding of the fundamental molecular processes driving calcification in the species and in broader phytoplankton communities. The high abundance and global distribution and the significant contribution of coccolithophores to the global carbon cycle have brought them to the focus of scientific endeavours. In recent decades the potential susceptibility of the carbon pumping by coccolithophores' photosynthesis and calcification due to ocean acidification was addressed by many studies (e.g. Riebesell et al., 2000; Reynaud et al., 2003; Langer et al., 2006; Orr et al., 2005; Fabry, 2008; Jones et al., 2013). Because biomineralisation in *Emiliania huxleyi* occurs internally and cells rarely exceed 5 nm in diameter it has been difficult to study the biology of coccolithogenesis. Molecular tools have already proven successful in improving our understanding of biomineralisation in *Emiliania huxleyi* (e.g. Wahlund et al., 2004a; Nguyen et al., 2005; van Dassow et al., 2009; Jones et al., 2011; McKew et al., 2013) and the response to ocean acidification (Richier et al., 2011; Lohbeck et al., 2012; Bach et al., 2013; Jones et al., 2013).

To better understand the fundamental processes driving calcification, unaltered biomineralisation properties throughout the cell cycle were investigates by assessing physiological parameters and a novel combination of whole cell transcriptome and proteome analysis. First, the cell-cycle phases, calcification rates and rates of photosynthesis in *Emiliania huxleyi* PLY♯ M217 were studied over a 24-hour period using flow cytometry and stable carbon isotope incorporation to resolve appropriate sampling time points at high and low calcification rates for whole cell transcriptome and whole cell proteome sampling. The cell-cyclic analysis of DNA content per cell allowed for the separation of the G1-phase from a combined S+G2/M-phase. The patterns of carbon assimilation showed a clear pattern of maximum calcification occurring at 9 hours into the photosynthetic light period when coccolith production was found to reach 1.6 per hour. At 9 hours into the dark phase calcification rates of 19 % of the maximum calcification rates in the light were observed (see chapter 2). According to the analysis of cell population growth and DNA content analysis cells were found to be G1-phase of the cell cycle

at both sampling times. Dark calcifications rates of 10-15 % were reported in previous studies (Paasche, 1966; Balch et al., 1992; Nimer an Merrett, 1992; Paasche and Brubak, 1994; Sekino and Shiraiwa, 1994, 1996). The energy driving the dark calcification was suggested to derive from mitochondrial respiration (Sekino and Shiraiwa, 1996). Calcification is significantly enhanced in photosynthesizing cells (Paasche, 2002).

The transcriptome and proteome of *Emiliania huxleyi* were evaluated with a focus on genes potentially involved in biomineralisation, which showed significant differential expression between low and high calcification states of the G1 cell-cycle phase. Three general approaches were undertaken to explore differential expression relevant to the calcification process in the two conditions at transcriptome and proteome level: a) using functional KOG-class (euKaryotic Orthologous groups) annotation groups or GO (gene ontology) groups to retrieve general expression patterns, b) using a graphical approach to mine for clusters of significant differences in higher abundant transcripts of biomineralisation genes in each KOG-class, and c) exploring the differential expression of known biomineralisation relevant genes at peptide and transcript level. The results have been discussed in respect to calcification physiology in coccolithophores using the exclusive transcriptome and proteome data and the consensus of transcriptome and proteome. Here methodological considerations are discussed and the results are put into a larger context of coccolithophore biology, cell cycle regulated gene expression in *Emiliania huxleyi*, and future research implications.

Evaluation of transcriptome analysis approach

The transcripts from the Illumina experiment described in Chapter 3 were matched against the older JGI *Emiliania huxleyi* genome version from 23.8.2012 using the Tophat – cufflink package (Trapnell et al., 2012). Subsequent to this work, a new version of the *E. hux* genome has been released and, in future work, the transcripts could be matched against the most recent version of the *E. huxleyi* genome (Read et al., 2013) to improve transcript annotations. Furthermore, it should be thoroughly researched beforehand what features the transcriptome analysis packages have and what kind of difficulties users have experienced. There are many publically available molecular expression data analysis tools available through www.bioconductor.org (Huber et al., 2014). A good choice of the analytical software package should consider subsequent analysis. In this study it was difficult to apply the data from the cufflinks output to other analysis tools due to compatibility issues that could not be resolved. An enrichment analysis functional GO annotation using, for example the topGo R script (Alexa & Rahnenfuhrer, 2010), was difficult to apply to the cufflinks output. Therefore, the analysis was

based on a basic summary of KOG annotations. A functional or pathway analysis might reveal insights into activated physiological pathways and interacting components. A combined RNA and protein extraction protocol such as the AllPrep DNA/ RNA/Protein Mini Kit (Qiagen) may have improved the overall quality of the proteome and transcriptome assessment.

Evaluation of cell harvesting, RNA, and protein extraction approach

RNA and proteins are vulnerable to degrading. The rapid sampling of live cells and RNA extraction was pivotal for obtaining high quality data. Previous proteomics studies on *E. huxleyi* at the National Oceanography Centre, Southampton in collaboration with the Institute for Life Sciences, University of Southampton showed poorer yields in protein quality, whereas the assessed proteomes were thought to represent only a part of the actual proteome of *E. huxleyi*. Therefore, it was anticipated to accelerate the cell harvesting process and protein extraction. The MASCOT software successfully fingerprinted Mass spectra of 409 different peptides. The peptides were successfully matched to 287 different proteins, which is an increase in the *Emiliania huxleyi* protein identification yield in an iTRAQ experiment of 250 % compared to Jones et al. (2013). The accelerated cell harvesting method and the resulting shorter period until snap freezing may have caused less protein degradation. The cell harvesting method described in Chapter 3 took 25 to 30 minutes compared to 2 hours in previous studies (Jones et al. 2011; Jones et al., 2013). Additional acetone wash steps as described in Chapter 3 may have further improved the protein sample quality. The proteome discovery rate would be greatly improved using a more advanced mass spectrometry system, such as the Orbitrab®.

Evaluation of proteome analysis approach

The quantitative proteomics approach for *Emiliania huxleyi* using isobaric tags (iTRAQ) has been previously established at the Institute for Life Sciences, University of Southampton. The iTRAQ method returns relative peptide abundance ratios between samples per peptide, from which relative protein abundance between conditions can be concluded. The relative abundance lacks information of absolute protein abundance in the cell. Information on absolute or relative to the absolute proteome peptide abundance would be important to understand the kinetics of biomineralisation in *E. huxleyi*. For example, it was estimated that 250,000 SERCA-type Ca^{2+} ATPase molecules would be required to provide the Ca^{2+} transport for one-hour's coccolith production (Mackinder et al., 2011). However, the Ca^{2+} transport by SERCA type Ca^{2+} ATPase cannot be achieved alone and other Ca^{2+}/H^+ exchanger were suggested at other locations, preferably the endoplasmic reticulum, to achieve the transport kinetics. A more accurate quantification of protein abundance and diversity and compartmental location in *E. huxleyi*

would be pivotal to better understand the ion transport kinetics. The technical implementation of truly quantitative proteomics remains challenging. Improved MS/MS proteomics techniques (e.g. Orbitrap®) are now available, which apply a combination of iTRAQ and unlabelled proteomics techniques (SWATH-MS) (Basak et al., 2015) or novel isobaric-labelled techniques using internal heavily labelled standards (Curran et al., 2015). Furthermore, peptide sequences should be matched directly to the transcript sequences and subsequently matched to the most recent *E. huxleyi* reference consensus analysis of proteome and transcriptome. The combination of a simultaneous DNA/RNA/protein extraction and a label free quantitative proteomics approach, such as by using normalized counts of MS/MS spectra assigned to a specific peptide, protein abundances could be estimated and methodological bias from indices application could be minimized (Schulze and Usadel, 2010; McKew et al., 2013). However, the protein identification and relative quantification presented here was more successful than previously reported for an iTRAQ experiment on *Emiliania huxleyi* (Jones et al., 2010). Out of 403 identified proteins using only the JGI *E. huxleyi* genome and taxonomically restricted reference databases 116 peptides showed significantly different abundances between the early and late G1-phase (dark and light incubation period). Interestingly, more proteins with relative abundances were significantly higher in the high calcification rate phase (n = 90) compared to the low calcification rate phase (n = 26).

Dark calcification in the early G1-phase of the cell-cycle

The assessment of calcification rates over 24 hours (Chapter 2) showed detectable calcification in the dark phase in the range of those previously reported (Paasche, 2002). It was argued that calcification is restricted to the G1-phase of the cell cycle. However, the existence of dark calcification in coccolithophores has been discussed. During nuclear division electron microscopy did not detect coccolith vesicles (van Emburg, 1989). Linschooten et al. (1991) showed that dark calcification was not detectable in cultures of synchronous cell division using ^{45}Ca as a tracer of biomineralisation. No existence of dark calcification is further supported by dark calcification only being accountable to cells not undergoing cell division in the dark. As such phosphorous-limited coccolithophore cultures, with inhibited nuclear division, were showing dark calcification (Paasche and Brubak, 1994; van Bleijswijk et al., 1994). Dark calcification requires energy for ion transport and synthesis of cellular structures, which is supplied from respiration (Sekino and Shiraiwa, 1996). Transcription and protein synthesis also require energy through hydrolysis of ATP (Snustad and Simmons, 2000; Robinson and Oijen, 2013). The observed transcript abundance shows that high rates of transcription occurred in the early-G1 (experimental low calcification) phase in the *E. huxleyi* cultures, which suggests

sufficient ATP availability. Therefore, it can be hypothesized that cohorts of old and new generation of the cells performed the observed dark calcification, which were in the early G1-phase. The results from the DNA content analysis support that the cell cycle of the cell population was not synchronised. If we assume, that a fraction of the cell population was always in the G1-phase and that proteins relevant to the calcification process were present the observation of dark calcification throughout the dark incubation period is explained. A high abundance of calcification specific transcripts in the early G1-phase and the nature of cytokinesis, by which ER and Golgi networks are distributed equally between daughter cells, further support that the proteins and calcification pathways were present at all times. The available proteome data also shows that, for example the V-type ATPase proton transporter and clathrin were present in the early G1-phase (compare Table 4.10). The raw MASCOT data from the iTRAQ experiment suggests that a total of 23 V-type proton ATPases matching peptides and 10 clathrin peptides were counted in both conditions (data not shown). The average protein abundance ratios were 0.847 for type proton ATPases and 0.664 for clathrin relative to the early G1-phase, which would suggest that some of these proteins were present a could provide low rates of calcium carbonate production. If cell structures and proteins were present and synthesized in the early G1-phase a molecular biomineralisation machinery of smaller size could be present in the cells, contributing to rates of dark calcification. It could be argued that the observed biomineralisation peptides in the dark incubation period originated from those cells, which did not pass cell division. However, the observed growth rates in this study indicated that the cell population more than duplicated (average 2.4 fold, see section 2.3.1) over the course of 24 hours, making it unlikely that cells were not passing through the S + G2/M phase. Furthermore, during mitosis the proteins and cell structures are divided between the daughter cells. Paasche (2002) noted that respiratory CO_2 production would oppose calcification in the absence of photosynthesis. However, *E. huxleyi* has a high internal buffering capacity for CO_2 and takes up external CO_2 as its primary source for photosynthesis to reach maximum POC production rates. Furthermore, bicarbonate was found to be the primary source for PIC production and elevated CO_2 did not affect calcification (Bach et al., 2013). The additional CO_2 formed during dark calcification and respiration might be transported to the chloroplast or buffered in the cytosol. This argument would support an effective dark calcification, which is driven by the remaining and newly developing molecular biomineralisation machinery in the targeted cell phase. To verify the observation of dark calcification the redevelopment of the CV could be observed using cross-polarized light microscopy and electron microscopy alongside isotope incorporation. Additional sampling point at which proteomes and transcriptomes are analyzed could greatly improve our understanding of the natural gene expression and existing molecular pathway in the cell cycle of *E. huxleyi*.

Transcriptome and proteome patterns related to calcification in different stages of *Emiliania huxleyi* G1-phase showing high calcification and low calcification rates

Cell cycle-specific transcription, targeted proteolytic degradation, and protein modification are regulators of cell cycle progression (Cardone and Sassone-Corsi, 2003). Chauton et al. (2013) showed high periodicity in gene regulation of carbon fixation, storage, and utilization in the diurnal cycle related genes of the diatom *Phaeodactylum tricornutum*. Hence, a cell-cycle regulation of gene expression related to calcification in *Emiliania huxleyi* was expected. The transcriptome profiles from the light and dark incubation period from this study showed significant differential expression. Interestingly, the transcriptome in the dark was 20 % larger showing 19093 transcripts compared to 15310 transcripts in the light. This overall pattern of higher transcript abundance in the dark is also reflected in, for example, the unique transcripts observed at the dark and light sampling point, the higher abundance of transcripts belonging to V-type-proton ATPases, SERCA-type calcium ATPases, and Na^+ - dependent Ca^{2+} exchanger (NCKX and NCX) (see Chapter 3 for details) in cells of the early G1-phase cells (dark period). The finding of higher transcript abundance and significant differential expression of genes potentially involved in calcification process in the dark is intriguing at first, knowing that the bulk of calcification in coccolithophores is linked to the light period. Statistical analysis showed that, 11 and 22 genes, with link to processes involved in calcification, were significantly higher expressed in highly calcifying and low calcifying cells, respectively. The majority of significantly expressed genes in the calcification phase indicated that calcium binding proteins such as calreticulin, calnexin, and calmodulin are continuously highly expressed in the light. Calreticulin transcripts, for example, were 8-fold more abundant in the light, at1516 fpkm. Calcium binding protein seem to be in high demand in the light phase, which suggests that continuous synthesis is required because the molecules are lost during coccolithophogenesis.

Cell cycle phase specific gene expression is widely supported in the scientific literature (e.g. Rustici et al, 2004; Kanesaki et al., 2012; Bastajian et al., 2013), and is triggered by external 'zeitgeber', such as the light/dark cycle (Lorenzen, 1957). The periodicity of gene expression was also found to be cell cycle-phase specific for mitochondria, chloroplast, and cell nucleus in the in the Unicellular Red Alga *Cyanidioschyzon merolae*; whereas, for example, the *cmdnm2* gene involved in plastid division was higher expressed in the G2/M phase and the *ftsZ2-1* and *ftsZ2-2* gene, involved in mitochondrial division, in the S phase (Fujiwara et al., 2009). In the diatom *Phaeodactylum tricornutum* microarray analysis from 8 sampling times showed differential gene expression for 44 % of the whole transcriptome, with highest transcriptional activity during the day. Six clusters of genes with differential expression and dominant

physiological pathways were identified. Chloroplast and nucleus-encoded ribosomal genes were upregulated in the light, heat shock proteins, chaeperones, and proteins involved in protein modification/degradation were upregulated at midday, and proteins of cellular processes such as carotenoid, chlorophyll, and fatty acid synthesis were down-regulated during night time, for example (Chauton et al., 2013). The synchronisation of the circadian clock and the environmental light - dark rhythm improves the physiological performance of photosynthetic organisms (Green et al., 2002; Dodd et al., 2005; Winjen et al., 2006b). Therefore, it is an energetic advantage, especially for photosynthetic cells, to establish the physiological networks before the next illumination period.

The diurnal periodicity of gene expression in *E. huxleyi* is hardly known. Gene expression studies using the dark phase as sampling point were focussing on three biomineralisation genes (Richier et al, 2009). In this study, it was concluded that the majority of the cell populations at the dark and light sampling points were in stages of the G1-phase (see Figure 2-4D), the late and early G1-phase, respectively. As calcification in coccolithophores is restricted to the G1-phase but increases significantly in photosynthesizing cells (Paasche 2000), the presence of gene expression related to calcification was expected. However, the results presented in this study are a novelty as they show that the larger number of known biomineralisation genes are more highly expressed in the early G1-phase (dark period). The significantly higher expression of, for example, gene 354606 (probable K^+ dependent Ca^{2+}/Na^+ exchanger), gene 463095 (SERCA - calcium-transporting ATPase 4), and 109061 (calmodulin-like protein 5 / CLP 5) in the dark would suggest that the proteins are not involved in calcification processes (see Appendix A Table 7 and 8). WolfPSort (Horton et al., 2007) identifies the ER as the location of the probable K^+ dependent Ca^{2+}/Na^+ exchanger (gene 354606, UniProt ID: Q9VN12) and SERCA - calcium-transporting ATPase 4 (gene 463095, UniProt ID: Q9XES1). The calmodulin-like protein 5 / CLP 5 (gene 109061, UniProt ID: Q9NZT1) is of mitochondrial location by WolfPSort. Therefore, the observed high general gene expression and high expression of calcification related genes reflected the re-construction of molecular machinery that was separated during cytokinesis. Golgi stacks, endosplasmic reticulum, ribosomes, mitochondria and chloroplasts are known to be equally partitioned between daughter cells during cytokinesis (Nebenfuhr et al., 2000; Puhka et al., 2007). Following cytokinesis the ER and Golgi membranes are synthesized, as the ribosomal-ER-Golgi cell physiological network is pivotal for protein translation and posttranslational protein modifications. In *E. huxleyi* calcification the ER and Golgi-apparatus are important in Ca^{2+} rich coccolithosome production that is pivotal for the precipitation of the coccolith in the coccolith vesicle. These networks also ensure that at dawn efficient photosynthetic carbon assimilation can render cell growth and

accumulation of energy reserves for passing cell division check points of the next cell division cycle (Green et al., 2002; Dodd et al., 2005; Winjen et al., 2006b; Zachleder et al., 2016). The ER and mitochondria related gene expression showed periodicity in cell-cycle of *C. merolae* (Fujiwara et al., 2009) and it can be assumed that cell organelle related gene expression occurs also in periodic waves throughout the cell cycle of *E. huxleyi.* The elevated gene expression of proteins related to biomineralisation is coherent with the post cytokinesis G1-phase characteristics of gene expression. Even the relatively high levels of calmodulin transcript abundance in the dark period (~1650 fpkm), support the synthesis of the ER networks in the early G1-phase, as calmodulin is an important component in maintaining high Ca^{2+} concentrations of the ER lumen.

Interestingly, more proteins with relative abundances were significantly higher in the late G1-phase (n = 90) compared to the early G1-phase (n = 26). This observation stands in contrast to the observed pattern of transcript abundance. The identified significantly more abundant proteins in the light were mainly involved in processes related to photosynthesis (photosystem II) and ATPase proton transport as indicated by counts of proteins' function, subcellular location, and GO cellular component annotations. The predicted locations of the proteins identified to be relatively more abundant in the light were the chloroplast, cytoplasma, Golgi apparatus, and plasma membrane. In the dark period proteins with higher relative abundance of nucleosome and microtubule origin were more common. The proteome of the light reflected the general catalytic activity in the photosynthesis phase and given by the higher relative abundance proton transporting ATPases and clathrin calcification and coccolithogenesis. Vacuolar proton ATPases and clathrin were previously observed in the proteome of *E. huxleyi* (Jones et al., 2011; Jones et al., 2013). Both proteins may be directly associated in clathrin coated vesicles (Forgac, 2000). However, both proteins have functions in a variety of cellular processes and locations. Clathrin is generally the primary component of endocytic transport systems (Takei and Hauke, 2001) and was also found to be highly expressed in the diatom *Thalassiosira pseudonana* (Nunn et al., 2009). Clathrin has a versatile function in processes such as endocytosis events that induce nutrient availability, growth of cell structures, protein transport, stabilization of forming vesicles, and the reformation of synaptic vesicles (Kirchhausen 2000; Takei and Hauke, 2001; Nunn et al., 2009). Synaptic vesicle recycling is also mediated by clathrin (Marks and McMahon, 1998). The precise function of clathrin in coccolithophores remains speculation. However, it can be assumed that clathrin coated vesicles may play a significant role in the vesicle transport by coccolithosomes to the coccolith vesicle. Jones (2010) also observed clathrin and COP in the proteomes of *E. huxleyi* and suggested that it could be operating in vesicular transport from the transitional ER to the Golgi. Furthermore, V-type

ATPases have been shown to be important for the acidification of cellular compartments, such as neuron (Moriyama et al., 1992) and were proposed to contribute to acidification of the trans-Golgi network in *Arabidopsis* (Dettmer et al., 2006). If clathrin-coated vesicles are associated with V-type ATPase in *Emiliania huxleyi* they could be involved in the acidification of the Golgi-derived coccolithosomes.

New insights into the calcification mechanisms and gene expression patterns in *E. huxleyi* from the synchronous study of the transcriptome and proteome at a different calcification levels (low and high) of the G1 phase

The proteome and transcriptome results of the calcification phase have so far accentuated the importance of four functional proteins categories for coccolithophore calcification: calcium binding proteins, clathrin-coated vesicle transport, membrane fusion, and proton-transport, which can be considered to propose a model for vesicular transport of Ca^{2+} to the coccolith vesicle. Calcium binding proteins, such as calmodulin and calreticulin have a higher affinity to Ca ions in acidic conditions (Yoo and Lewis, 1992; Nash et al., 1994) and can bind up to 25 moles of Ca^{2+} per mol calreticulin (Baksch and Michalak, 1991). Calreticulin changes into a tetramer in low pH with increased mass, with increasing Ca^{2+} binding capacity (Yoo and Lewis, 1992) and has high capacity low affinity Ca^{2+} binding characteristics (Corbett and Michalak, 2000). Calreticulin and other calcium binding proteins are associated and abundant in the Golgi network and endoplasmic reticulum, which act as a Ca-ion buffering structure in cells and modulates the uptake and release of Ca^{2+} in the endoplasmic reticulum (Corbett and Michalak, 2000; Arnaudeau 2002). Calcium loaded calreticulin could be included in coccolithosomes that are released from the trans-Golgi. The vesicle-like coccolithosome membrane could be clathrin coated (Takei and Hauke, 2001) and therefore contain numerous V-type proton ATPases. Clathrin coated vesicles originating from the Golgi were previously identified (Orci et al., 1986). Mackinder et al. (2011) suggested that a down-regulation of V-type proton ATPase could promote calcium carbonate precipitation under supersaturated conditions of Ca^{2+} and HCO_3^- and at pH 7.2. Calreticulin was found to change to a monomer under more alkaline conditions and suggested to release calcium (Yoo and Lewis, 1992). Active V-type proton ATPase could stabilize calreticulin and other calcium binding proteins at high calcium binding capacities during trans-cytosolic migration to the CV. Clathrin and COP I, COP II, and SNARE complexes may assist in the fusion of coccolithosome and coccolith vesicle membranes. The CV might already show higher pH during this fusion. The V-type ATPases are at some point down-regulated in the CV membranes to develop a more alkaline environment for calcium carbonate precipitation. V-type proton ATPases were found to have peripheral complex

V_1 (Arai, 1989). The integral membrane complex membrane V_0 and peripheral V_1 complex are easily dissociated reducing proton pumping across the membrane to a minimum (Kane, 1995; Beyenbach and Wiezorek, 2006), which could regulate the pH of the CV in favour for calcium carbonate precipitation if the CV maintained a cytoplasmic-side negative membrane potential (Mackinder et al., 2011). The results presented here strongly suggest the requirement of V-type proton ATPase, calcium binding proteins, and membrane fusion complexes, which would be required for the above scenario. Future research on *E. huxleyi* calcification should attempt to elucidate the location of calreticulin in the cell. A remaining questions, which may be very hard to address experimentally, should address the differences of pH and composition of the coccolithosomes to the CV. A lower pH value in the coccolithosomes than the CV would support the above-proposed model that coccolithosomes carry higher loads of Ca^{2+} due to calreticulin in acidified conditions.

Furthermore, the molecular results indicate some interesting characteristic of cell-cyclic gene expression in *E. huxleyi*. The transcriptome analysis identified a few genes that transcribed for the same protein but, interestingly, showed preferred expression in the early-G1-phase or late G1-phase. Thus suggesting that different promoter sites and different gene regulation pathways activated gene expression regulation. For example, the endoplasmic reticulum-type calcium transporting ATPase 4 Q9XES1 was expressed by three genes (251608, 429294, 463095). Genes 251608 and 429294 were dominantly expressed in the late G1-phase in highly calcifying cells, whereas gene 463095 was expressed almost exclusively in the early G1-phase (dark incubation period) when cell compartments like the endoplasmic reticulum were complemented after cytokinesis (compare Appendix A Table 7). A more detailed investigation of the transcriptome might reveal more such cases. Gene expression transcribing the same protein also showed genes that with higher expression than others. For example, the expression of six genes transcribing for V-type proton ATPase 16 kDa proteolipid subunit Q43362 were observed to show higher transcript abundance in the early G1-phase. Among those, gene 359783 was at least 3 times higher expressed than any other gene transcribing for Q43362. It is unclear if an elevated physiological or environmental key could trigger higher expression of the other genes and if all genes transcribing for Q43362, for example, could provide the same maximum transcription rate. The outstanding cap analysis of gene expression will provide additional insights in differences of gene expression patterns in the early and late G1-phase. A higher plasticity of the transcriptome in respect to elevating the expression of genes that are not transcribed at maximum rates, such as the V-type proton ATPases, suggests that *E. huxleyi* has a potential to compensate for calcification altering effects of ocean acidification. In fact, after 500 generations of ocean acidification exposure *E. huxelyi* cells were found to show an adaptive

response, which included a putative vacuolar type two-sector proton pump (Lohbeck et al., 2014).

Concluding remarks and future work

The original outline of the project anticipated sampling a natural coccolithophore bloom, to improve existing protocols of *Emiliania huxleyi* proteomics, and to find transcriptome and proteome matches. The data presented and discussed in this monograph represents only a fraction of the information that the whole cell transcriptomic and whole cell proteomic results hold. Future, investigations should address functional network and predict the location of the identified proteins. Furthermore, a great deal of data that was assessed during this PhD has not been used in the framework of this PhD thesis. Due to time constraints, the metatranscriptomic assessment of a natural *Emiliania huxleyi* bloom in Raunefjorden, Norway has not been fully analysed and was not presented here; this work, among others, is on going.

The investigation of the transcriptome and proteome of unaltered diurnal biomineralisation patterns has shown that multiple genes of homologous groups are expressed at different times during the cell cycle. The observations imply that a bulk of proteins involved in coccolithophore calcification, such as V-type-proton ATPases, SERCA-type calcium ATPases, and Na^+-dependent Ca^{2+} exchanger (NCKX and NCX) are expressed in the early G1-phase of the cell cycle and that ribosomal synthesis of these proteins occurs shortly after. Furthermore, the early expression of the biomineralisation genes implies that they are part of compartmental systems and organelles that are maturing in the early G1 phase. To decipher the molecular machinery at the core of biomineralisation it is critical to further engage studies using modern genomic, transcriptomic, proteomic and metabolomic techniques that target explore the expression and its control (e.g. promoter expression analysis) additional time points throughout the cell cycle. Chauton et al. (2013) have provided an example for cyclic diurnal gene expression study that suggested clusters of genes that are expressed at different times throughout the cell cycle. A similar study and molecular analysis could be conducted on coccolithophores. Furthermore, knockout, knockdown and overexpression of genes, which have become apparent in this study, should help to confirm the role of the genes putatively identified herein as being important for biomineralisation. An analysis of gene expression, proteome patterns and calcification under conditions of specifically knocked out ion transporter genes should help to differentiate between a core set of ion transporter genes involved in calcification (core set) and secondary pathways. Knockout or transformation techniques are now more widely available and have been more frequently applied in marine species, such as *Nannochloropsis gaditana* to enhance their

potential use in biofuel production (Radakovits et al., 2012). Such transformation techniques are needed in coccolithophores to understand the fundamental mechanisms driving calcification. In respect to the response of coccolithophores to ocean acidification a CO_2 exposure experiment could illuminate, if the early expression of biomineralisation proteins increases under CO_2 or if the responses occur with increasing rates of calcification during the light period. Such an experiment may also help to decipher which genes that are involved in calcification are activated as a response to elevated CO_2 and which genes are more likely expressed because they function in cellular compartments. The results presented in this study – the cell cycle phase of sampling greatly matters when drawing conclusions from gene expression studies that address biomineralisation - should be considered in the design of future experiments, in particular those addressing the response to ocean acidification and adaptation to global change. Further fundamental assessments of the calcification physiology at different cell cycle phases will improve the understanding of the level at which adaptive processes to ocean acidification in coccolithophores occur. Lohbeck et al. (2014) showed that after 500 generations of CO_2 exposure *Emiliania huxleyi* showed an adaptive response and was able to reestablish partly the calcification and growth rates. At a molecular level the adaptive response was observed as up-regulation of genes of interest involved in pH regulation and carbon transport. As this study has shown there are multiple genes that are active in processes, which could be important for calcification and those genes may be expressed at an early stage of the cell cycle. It would be most interesting to investigate how the adaptive processes are reflected in selected promoter enhancement and transcriptome plasticity (transcribing at maximum gene expression rates) of genes that are not fully expressed under ambient CO_2 and whether or not the adaptive changes are reflected in early G1-phase cells or in photosynthesizing cells of the late G1-phase. Early G1-phase adaptive changes would suggest an "evolutionary" response in *E. huxleyi* to altering ocean acidity because the extend of the ER and Golgi apparatus are improved, for example. Changes of gene expression related to calcification in photosynthesizing cells of the late G1-phase would indicate an improvement of the physiological pathways under photosynthetic enhancement of the calcification process. Such insights would increase our detailed knowledge of the molecular adaptation of *E. huxleyi* to the future high CO_2 world and furthermore, help to improve estimating the changes in future calcite production and the global carbon cycle.

Appendices

Appendix A - Supplementary results - Chapter 3 - The transcriptome of *Emiliania huxleyi* at high and low calcification rates in the G1-phase

Physiological data:

Appendix A Table - 1: Statistical results comparing the physiological data from the day before cell harvesting with the day of harvesting and the late and the early G1-phase. (KWA: Kruskal-Wallis ANOVA).

Parameter	Comparison	test	n	d.f.	F-statistics H-statistics	p-value
Fv/Fm	full	ANOVA	12	3,8	F= 2.153	0.172
POC	full	KWA	8	3	H= 3.667	0.381
PON	full	KWA	8	3	H= 2.833	0.543
PIC	full	ANOVA	12	3,8	F= 3.212	0.083
C:N	full	KWA	8	3	H= 4.167	0.324
PIC:POC	full	KWA	8	3	H= 5.500	0.139

Transcriptomics data:
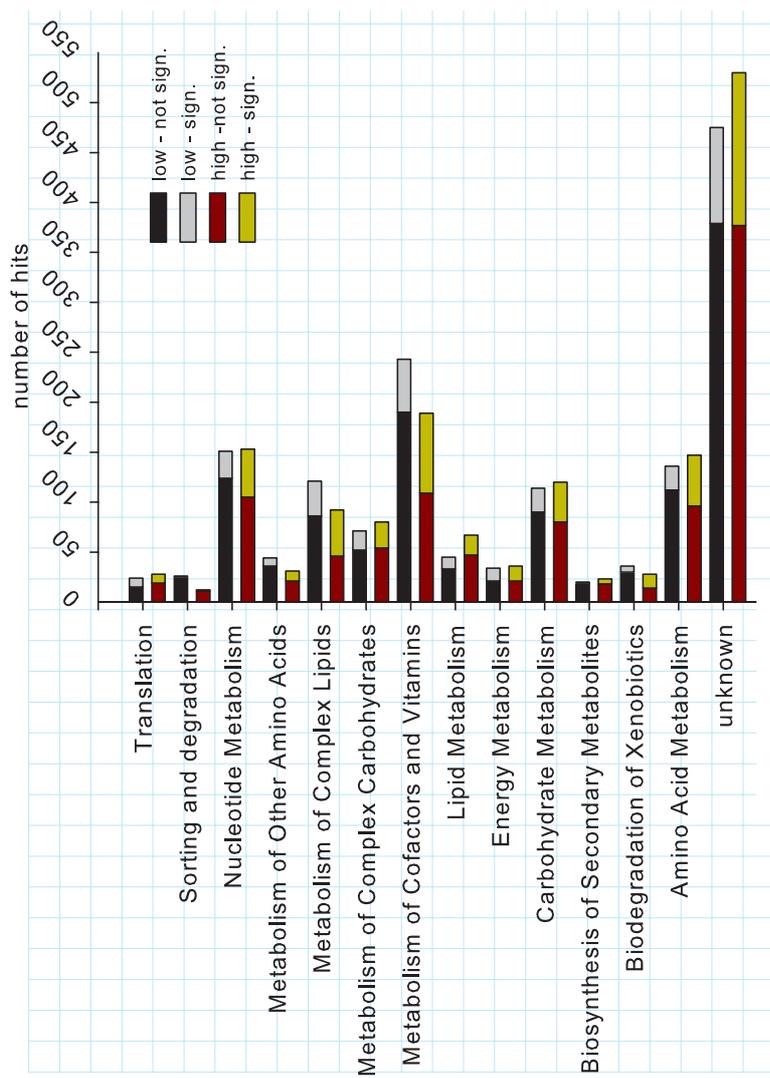
Appendix A Figure 1: Transcript abundance by KEGG-pathway classes for *Emiliania huxleyi* cells showing high and low calcifications in the G1-phase.

Appendix A Table - 2: Unique transcripts in highly calcifying cells of the late G1-phase .

Unique transcripts of the late

JGI/gene-ID	fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
- / XLOC_020772	2228.59	0.2242	no	-	-
- / XLOC_019818	592.35	0.0917	no	-	-
- / XLOC_029101	365.22	0.2379	no	-	-
- / XLOC_006956	325.07	0.1550	no	-	-
- / XLOC_039075	154.93	0.0847	no	-	-
- / XLOC_024334	68.95	0.1283	no	-	-
- / XLOC_039067	56.49	0.0715	no	-	-
- / XLOC_022511	46.63	0.0258	yes	-	-
- / XLOC_018847	45.76	0.0135	yes	-	-
352298 / XLOC_008894	44.25	0.0293	yes	-	-
- / XLOC_024335	43.29	0.0452	yes	-	-
- / XLOC_001410	42.94	0.0547	no	-	-
- / XLOC_015139	33.11	0.0568	no	-	-
- / XLOC_007296	30.79	0.1009	no	-	-
- / XLOC_000845	24.16	0.1336	no	-	-
- / XLOC_034717	24.16	0.1336	no	-	-
- / XLOC_021913	24.01	0.0114	yes	-	-
- / XLOC_031851	23.89	0.0446	yes	-	-
- / XLOC_015199	23.42	0.1128	no	-	-
- / XLOC_003721	20.55	0.0612	no	-	-
- / XLOC_002711	20.42	0.0351	yes	-	-
- / XLOC_015880	19.35	0.1170	no	-	-
100679 / XLOC_018351	18.61	0.0788	no	F0Y854	Expressed protein
- / XLOC_026071	17.39	0.0135	yes	-	-
- / XLOC_007295	16.91	0.1202	no	-	-

Unique transcripts of the late

JGI/gene-ID	G1-phase fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
- / XLOC_014742	15.20	0.0679	no	-	-
250018 / XLOC_040170	14.50	0.0130	yes	Q5UP03	Putative homeobox protein R749
44463 / XLOC_018334	13.97	0.0865	no	Q8LIG4	CBL-interacting protein kinase 3
255559 / XLOC_024876	13.01	0.1927	no	-	-
- / XLOC_002710	12.42	0.0454	yes	-	-
- / XLOC_010612	12.06	0.0167	yes	-	-
- / XLOC_012328	11.96	0.1070	no	-	-
- / XLOC_030958	10.34	0.0162	yes	-	-
- / XLOC_010900	10.28	0.0287	yes	-	-
- / XLOC_008618	9.88	0.0271	yes	-	-
- / XLOC_042065	9.88	0.1700	no	-	-
- / XLOC_007297	8.86	0.1472	no	-	-
- / XLOC_019883	8.79	0.0409	yes	-	-
225551 / XLOC_005371	8.63	0.0096	yes	Q4V8B1	Proton-coupled amino acid transporter 3
- / XLOC_004089	8.54	0.1875	no	-	-
- / XLOC_010896	8.46	0.1523	no	-	-
44813 / XLOC_006798	8.27	0.0841	no	P52046	3-hydroxybutyryl-CoA dehydratase
235819 / XLOC_016605	7.70	0.0781	no	E1ZM53	Putative uncharacterized protein
358042 / XLOC_034611	7.68	0.1763	no	-	-
- / XLOC_022048	7.28	0.0736	no	-	-
317194 / XLOC_038198	6.98	0.2272	no	-	-
- / XLOC_004722	6.76	0.1234	no	-	-
- / XLOC_016788	6.44	0.0279	yes	-	-
- / XLOC_035469	6.33	0.1568	no	-	-
- / XLOC_017020	6.16	0.1266	no	-	-
201176 / XLOC_005902	5.97	0.0204	yes	-	-

Unique transcripts of the late JGI/gene-ID	G1-phase fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
225228 / XLOC_005039	5.95	0.0668	no	-	-
- / XLOC_017535	5.72	0.1477	no	-	-
- / XLOC_031076	5.63	0.0249	yes	-	-
- / XLOC_039416	5.62	0.0236	yes	-	-
208991 / XLOC_025228	5.59	0.0615	no	-	-
- / XLOC_037632	5.31	0.0874	no	-	-
213562 / XLOC_032951	5.08	0.0783	no	-	-
- / XLOC_018766	4.98	0.0715	no	-	-
- / XLOC_007057	4.83	0.0876	no	-	-
44442 / XLOC_013259	4.71	0.0820	no	Q93VD3	CBL-interacting serine/threonine-protein kinase 23
59486 / XLOC_033585	4.60	0.7657	no	P23253	Sialidase
- / XLOC_032599	4.43	0.0388	yes	-	-
- / XLOC_010356	4.38	0.1077	no	-	-
- / XLOC_008965	4.29	0.0828	no	-	-
243770 / XLOC_026235	3.95	0.1367	no	-	-
219951 / XLOC_040503	3.88	0.0513	no	-	-
242647 / XLOC_025007	3.83	0.0805	no	-	-
99134 / XLOC_011273	3.79	0.0704	no	P52409	Glucan endo-1,3-beta-glucosidase
442881 / XLOC_014908	3.73	0.1115	no	B8GAY2	tRNA (guanine-N(7)-)-methyltransferase
432890 / XLOC_036316	3.70	0.1640	no	Q99P15	Phosphatidate phosphatase LPIN2
- / XLOC_017283	3.68	0.0323	yes	-	-
260314 / XLOC_002425	3.67	0.1296	no	Q0AS57	Peptidase S1 and S6, chymotrypsin/Hap (Precursor)
- / XLOC_006119	3.64	0.1404	no	-	-
- / XLOC_033682	3.61	0.0770	no	-	-
- / XLOC_039731	3.60	0.0736	no	-	-

Unique transcripts of the late

JGI/gene-ID	G1-phase fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
- / XLOC_008573	3.59	0.1672	no	-	-
- / XLOC_007061	3.38	0.0998	no	-	-
- / XLOC_007294	3.37	0.0979	no	-	-
- / XLOC_027620	3.14	0.0585	no	-	-
- / XLOC_023509	3.10	0.0546	no	-	-
309164 / XLOC_002899	2.98	0.0548	no	-	-
212382 / XLOC_031465	2.89	0.0782	no	P47980	Protein TIS11
357004 / XLOC_031810	2.86	0.1074	no	-	-
236785 / XLOC_017870	2.68	0.0769	no	P18126	Endoglucanase B
260625 / XLOC_003743	2.64	0.5047	no	-	-
- / XLOC_029150	2.56	0.1853	no	-	-
236638 / XLOC_017485	2.53	0.5858	no	Q7ZVE6	KDEL motif-containing protein 1
205977 / XLOC_018405	2.48	0.0128	yes	O64624	Pentatricopeptide repeat-containing protein At2g18940
70381 / XLOC_004911	2.46	0.4439	no	Q8GZ24	Putative uncharacterized protein At1g67060/F1O19_7
- / XLOC_032352	2.45	0.0694	no	-	-
368720 / XLOC_023158	2.44	0.0599	no	-	-
114681 / XLOC_016638	2.24	0.0972	no	-	-
- / XLOC_033459	2.15	0.0990	no	-	-
101089 / XLOC_019781	2.14	0.2364	no	F0YCE4	Putative uncharacterized protein
460267 / XLOC_042278	2.13	0.7376	no	-	-
- / XLOC_016541	2.10	0.1596	no	-	-
- / XLOC_020107	2.09	0.0971	no	-	-
66080 / XLOC_028085	2.06	0.7310	no	Q8WXR4	Myosin-IIIb
225509 / XLOC_005361	2.01	0.2170	no	-	-
238919 / XLOC_020182	2.00	0.5598	no	Q7XUR2	Lysine-specific histone demethylase 1 homolog 3

Unique transcripts of the late

JGI/gene-ID	G1-phase fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
70033 / XLOC_003329	1.98	0.1059	no	Q5UR67	Uncharacterized protein R617
- / XLOC_027592	1.93	0.1710	no	-	-
98132 / XLOC_005959	1.83	0.0388	yes	A0PLM4	Transcriptional regulatory protein
208652 / XLOC_024737	1.79	0.2511	no	C1N072	Predicted protein
238888 / XLOC_020156	1.78	0.4115	no	Q7VE24	Adenylyl-sulfate kinase
353266 / XLOC_014417	1.75	0.5695	no	-	-
107915 / XLOC_042479	1.73	0.0689	no	Q9Z1J8	SEC14-like protein 3
369373 / XLOC_024951	1.71	0.0390	yes	-	-
- / XLOC_017838	1.70	0.1686	no	-	-
249953 / XLOC_039941	1.67	0.0153	yes	O64624	Pentatricopeptide repeat-containing protein At2g18940
234824 / XLOC_015193	1.65	0.1199	no	-	-
224967 / XLOC_004763	1.63	0.1550	no	-	-
439062 / XLOC_041432	1.57	0.5486	no	P59968	Uncharacterized tRNA/rRNA methyltransferase Mb0905
195328 / XLOC_021657	1.56	0.2262	no	-	-
439135 / XLOC_041832	1.56	0.3794	no	-	-
55544 / XLOC_038225	1.48	0.1340	no	Q9ZUI0	Putative leucine-rich repeat receptor-like serine/threonine-protein kinase At2g24130
209316 / XLOC_025921	1.46	0.6886	no	-	-
228876 / XLOC_009104	1.45	0.5235	no	-	-
359059 / XLOC_037819	1.42	0.1057	no	-	-
194425 / XLOC_000234	1.41	0.7671	no	E9C982	Sjogren syndrome antigen B
99618 / XLOC_013707	1.39	0.1231	no	Q6AZT7	Nephrocystin-3
- / XLOC_004217	1.36	0.0629	no	-	-
238542 / XLOC_019899	1.34	0.0890	no	-	-
259646 / XLOC_042244	1.33	0.3827	no	-	-

Unique transcripts of the late					
JGI/gene-ID	G1-phase / fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
53381 / XLOC_028567	1.31	0.0751	no	Q9FYA6	Branched-chain-amino-acid aminotransferase 5, chloroplastic
118001 / XLOC_025412	1.29	0.0743	no	-	-
314118 / XLOC_010560	1.27	0.2436	no	P52409	Glucan endo-1,3-beta-glucosidase
- / XLOC_023365	1.26	0.1724	no	-	-
- / XLOC_022709	1.25	0.0353	yes	-	-
441995 / XLOC_011055	1.23	0.2190	no	-	-
196464 / XLOC_026780	1.17	0.1961	no	-	-
252131 / XLOC_005405	1.15	0.0816	no	-	-
239576 / XLOC_020921	1.15	0.0816	no	-	-
228379 / XLOC_008391	1.15	0.3387	no	-	-
435200 / XLOC_018392	1.12	0.0241	yes	Q6LM37	Flavohemoprotein
435722 / XLOC_022879	1.11	0.6750	no	Q91FH7	Putative FAD-linked sulfhydryl oxidase 347L
- / XLOC_023366	1.08	0.1419	no	-	-
194548 / XLOC_000281	1.08	0.1692	no	-	-
256592 / XLOC_030906	1.07	0.3583	no	-	-
251825 / XLOC_004247	1.03	0.1003	no	-	-
222247 / XLOC_001541	1.01	0.6332	no	-	-
233729 / XLOC_014109	1.00	0.3195	no	A0AUS0	WD repeat, SAM and U-box domain-containing protein 1
227000 / XLOC_006870	0.97	0.2076	no	-	-
310662 / XLOC_025270	0.97	0.1365	no	P55495	Uncharacterized protein y4iL
- / XLOC_027364	0.97	0.1407	no	-	-
117487 / XLOC_023944	0.97	0.1037	no	D6U4E0	Acyltransferase 3
236421 / XLOC_017284	0.95	0.1996	no	-	-
207447 / XLOC_022297	0.95	0.4150	no	O59843	1,4-beta-D-glucan cellobiohydrolase B

Unique transcripts of the late	G1-phase fpkm in late G1-phase	q_value	Sig.	UniProt ID	Function
JGI/gene-ID					
201173 / XLOC_006055	0.93	0.3271	no	Q63424	Solute carrier family 15 member 2
113645 / XLOC_014089	0.92	0.0758	no	-	-
- / XLOC_023129	0.91	0.1737	no	-	-
218775 / XLOC_039349	0.91	0.1046	no	P14328	Spore coat protein SP96

Appendix A Table - 3: Transcripts unique in *E. huxleyi* early G1-phase (low calcification rates).

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
- / XLOC_035007	2873.86	0.0479	yes	-	-
- / XLOC_007334	461.17	0.1399	no	-	-
205338 / XLOC_017557	271.96	0.0000	yes	D8LH42	Putative uncharacterized protein
- / XLOC_039432	207.08	0.0258	yes	-	-
- / XLOC_006749	169.45	0.0170	yes	-	-
- / XLOC_007641	166.95	0.0050	yes	-	-
- / XLOC_039916	122.87	0.0000	yes	-	-
- / XLOC_023432	114.46	0.0034	yes	-	-
249502 / XLOC_038428	100.07	0.0003	yes	-	-
- / XLOC_025529	94.24	0.0000	yes	-	-
- / XLOC_023232	74.53	0.0000	yes	-	-
- / XLOC_005269	66.51	0.0000	yes	-	-
- / XLOC_006849	50.67	0.0005	yes	-	-
- / XLOC_035748	42.20	0.0026	yes	-	-
450646 / XLOC_020305	42.00	0.0000	yes	-	-
- / XLOC_024344	34.16	0.0062	yes	-	-
- / XLOC_009645	32.75	0.0000	yes	-	-
- / XLOC_036400	32.54	0.0004	yes	-	-
309174 / XLOC_003093	27.64	0.0000	yes	Q8MTI2	Putative surface protein bspA-like
- / XLOC_001858	26.40	0.0000	yes	-	-
- / XLOC_031978	25.80	0.0091	yes	-	-
454253 / XLOC_002024	24.87	0.0000	yes	P31049	Probable fatty acid methyltransferase
446057 / XLOC_033730	24.85	0.0000	yes	-	-
- / XLOC_026248	23.19	0.0011	yes	-	-
- / XLOC_039074	22.64	0.0000	yes	-	-

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
- / XLOC_036606	20.72	0.0000	yes	-	-
- / XLOC_028775	19.77	0.0189	yes	-	-
- / XLOC_004785	19.23	0.0000	yes	-	-
- / XLOC_040372	19.04	0.0146	yes	-	-
- / XLOC_008781	18.27	0.0133	yes	-	-
257724 / XLOC_035222	18.10	0.1504	no	Q869P0	Probable ATP-dependent RNA helicase ddx31
- / XLOC_021020	16.90	0.0002	yes	-	-
258092 / XLOC_036904	16.09	0.0000	yes	B7G3J7	Formamidase-like protein
- / XLOC_013724	15.67	0.0005	yes	-	-
- / XLOC_018631	14.52	0.0632	no	-	-
- / XLOC_013144	13.46	0.0371	yes	-	-
439818 / XLOC_002631	13.28	0.0000	yes	-	-
- / XLOC_017538	12.96	0.0000	yes	-	-
435479 / XLOC_020267	12.24	0.0000	yes	-	-
- / XLOC_042388	11.54	0.0007	yes	-	-
- / XLOC_005332	11.30	0.0000	yes	-	-
- / XLOC_019260	10.99	0.0057	yes	-	-
208723 / XLOC_024764	10.01	0.0000	yes	-	-
- / XLOC_011815	9.90	0.0088	yes	-	-
- / XLOC_014788	9.80	0.0000	yes	-	-
- / XLOC_005601	9.00	0.0000	yes	-	-
243061 / XLOC_025503	8.94	0.0004	yes	-	-
49394 / XLOC_027735	8.77	0.1721	no	P10979	Glycine-rich RNA-binding, abscisic acid-inducible protein
- / XLOC_038400	8.51	0.0000	yes	-	-
- / XLOC_018176	7.91	0.0001	yes	-	-
- / XLOC_004381	7.87	0.0308	yes	-	-

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
455889 / XLOC_010285	7.65	0.0000	yes	-	-
- / XLOC_029998	7.52	0.0027	yes	-	-
- / XLOC_024364	7.18	0.0030	yes	-	-
- / XLOC_041374	7.18	0.0009	yes	-	-
- / XLOC_013640	7.16	0.0003	yes	-	-
- / XLOC_023541	6.87	0.0040	yes	-	-
352052 / XLOC_007394	6.62	0.0002	yes	-	-
- / XLOC_024331	6.57	0.0299	yes	-	-
196275 / XLOC_026420	6.57	0.0000	yes	Q9N3Z3	Serine/threonine-protein kinase chk-1
- / XLOC_040772	6.25	0.0231	yes	-	-
- / XLOC_013190	6.22	0.0030	yes	-	-
452468 / XLOC_034205	6.14	0.1689	no	E9C645	Putative uncharacterized protein
- / XLOC_033991	6.05	0.0131	yes	-	-
- / XLOC_018256	5.97	0.0695	no	-	-
- / XLOC_027238	5.89	0.0039	yes	-	-
- / XLOC_024385	5.88	0.0009	yes	-	-
459932 / XLOC_036896	5.50	0.0000	yes	F0Y858	Putative uncharacterized protein
- / XLOC_023121	5.40	0.0000	yes	-	-
248076 / XLOC_034319	5.33	0.0045	yes	Q8NUI5	Lipase 1
435146 / XLOC_018339	5.27	0.0000	yes	-	-
456341 / XLOC_012114	5.24	0.0000	yes	F0Y858	Putative uncharacterized protein
229350 / XLOC_009553	5.14	0.0926	no	Q8T7K0	ALG-2 interacting protein X
- / XLOC_033848	4.97	0.0000	yes	-	-
- / XLOC_001985	4.91	0.0035	yes	-	-
216305 / XLOC_035795	4.86	0.1038	no	F0Y502	Putative uncharacterized protein (Fragment)
424057 / XLOC_014299	4.73	0.0026	yes	-	-

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
450077 / XLOC_016201	4.62	0.0000	yes	-	-
70738 / XLOC_006435	4.59	0.1036	no	B7NLB5	Peroxyureidoacrylate/ureidoacrylate amidohydrolase RutB
207272 / XLOC_022237	4.43	0.0001	yes	F0Y048	Putative uncharacterized protein
456340 / XLOC_012113	4.40	0.0000	yes	P55176	UPF0012 hydrolase in pqqF 5'region
456791 / XLOC_014244	4.31	0.0000	yes	-	-
- / XLOC_031251	4.28	0.0390	yes	-	-
- / XLOC_022102	4.22	0.0008	yes	-	-
240013 / XLOC_021890	4.19	0.0000	yes	-	-
208470 / XLOC_024289	4.09	0.0000	yes	Q54Y32	MAP kinase phosphatase with leucine-rich repeats protein 3
- / XLOC_020111	4.08	0.0019	yes	-	-
247181 / XLOC_032305	3.97	0.0001	yes	B7G168	Predicted protein
222516 / XLOC_001666	3.89	0.0011	yes	Q6DGV7	Mpv17-like protein 2
308883 / XLOC_038740	3.85	0.6729	no	-	-
- / XLOC_034289	3.83	0.0000	yes	-	-
- / XLOC_018175	3.83	0.0004	yes	-	-
- / XLOC_021462	3.79	0.0004	yes	-	-
- / XLOC_009785	3.78	0.0941	no	-	-
211836 / XLOC_030827	3.71	0.0066	yes	Q10977	Phthiocerol synthesis polyketide synthase type I PpsA
223170 / XLOC_002452	3.69	0.1888	no	E9CG20	Embryogenesis-associated protein EMB8
454351 / XLOC_002508	3.58	0.0001	yes	B0G143	Mitochondrial substrate carrier family protein ucpB
- / XLOC_004833	3.45	0.0002	yes	-	-
- / XLOC_005270	3.34	0.0101	yes	-	-
221272 / XLOC_042335	3.30	0.0000	yes	A9A494	Porphyromonas-type peptidyl-

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
- / XLOC_021870	3.27	0.0267	yes	-	arginine deiminase
- / XLOC_033689	3.12	0.0597	no	-	-
195290 / XLOC_021645	3.08	0.3922	no	O31404	Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha
217995 / XLOC_037996	2.94	0.0000	yes	C1EIL5	Predicted protein
233951 / XLOC_014272	2.78	0.0001	yes	F2TWL3	Putative uncharacterized protein
426599 / XLOC_010056	2.72	0.5138	no	Q9FL69	Probable ADP-ribosylation factor GTPase-activating protein AGD5
461189 / XLOC_021379	2.58	0.0015	yes	Q9HYH5	Putative aldolase class 2 protein PA3430
232962 / XLOC_013178	2.56	0.0033	yes	Q5TZ51	Protein Mpv17
- / XLOC_008610	2.56	0.0627	no	-	-
109294 / XLOC_004057	2.55	0.0011	yes	-	-
255893 / XLOC_027244	2.54	0.2228	no	-	-
- / XLOC_017664	2.51	0.0308	yes	-	-
- / XLOC_026932	2.50	0.0003	yes	-	-
237168 / XLOC_018165	2.44	0.0012	yes	-	-
- / XLOC_008707	2.41	0.0094	yes	-	-
254752 / XLOC_019564	2.37	0.4152	no	-	-
- / XLOC_003248	2.34	0.0232	yes	-	-
- / XLOC_007186	2.32	0.0461	yes	-	-
- / XLOC_009931	2.32	0.0005	yes	-	-
113553 / XLOC_013918	2.20	0.0082	yes	Q99NF1	Beta,beta-carotene 9',10'-oxygenase
196894 / XLOC_030373	2.19	0.2222	no	-	-
251402 / XLOC_002738	2.18	0.0014	yes	-	-

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
369186 / XLOC_024471	2.04	0.0501	no	-	-
- / XLOC_008682	2.03	0.0017	yes	-	-
244577 / XLOC_027869	1.98	0.1308	no	Q9ST27	Phototropin-2
244576 / XLOC_027868	1.98	0.1308	no	Q9ST27	Phototropin-2
208026 / XLOC_023608	1.98	0.1308	no	Q9ST27	Phototropin-2
369188 / XLOC_024460	1.96	0.0474	yes	-	-
- / XLOC_033477	1.96	0.0127	yes	-	-
211047 / XLOC_029003	1.92	0.0116	yes	Q5F371	Protein strawberry notch homolog 1
- / XLOC_033938	1.90	0.0895	no	-	-
229436 / XLOC_009577	1.88	0.0324	yes	-	-
455708 / XLOC_009450	1.87	0.2293	no	-	-
452714 / XLOC_035411	1.79	0.1901	no	Q7PC76	Glucomannan 4-beta-mannosyltransferase 1
232178 / XLOC_012271	1.77	0.0149	yes	C1DZB7	Predicted protein
208111 / XLOC_023743	1.74	0.6037	no	B8BXH2	D-lactate dehydrogenase
- / XLOC_015733	1.70	0.2607	no	-	-
- / XLOC_004041	1.67	0.0064	yes	-	-
211666 / XLOC_029727	1.65	0.4960	no	-	-
359963 / XLOC_041264	1.63	0.0130	yes	-	-
245483 / XLOC_029078	1.62	0.2712	no	F0XX34	DNA polymerase
218752 / XLOC_039208	1.60	0.0003	yes	A9A494	Porphyromonas-type peptidyl-arginine deiminase
249876 / XLOC_039717	1.54	0.0158	yes	-	-
214286 / XLOC_033579	1.39	0.1144	no	-	-
- / XLOC_023263	1.38	0.0035	yes	-	-
115968 / XLOC_019986	1.34	0.0794	no	-	-
211783 / XLOC_030802	1.28	0.3546	no	-	-

Unique transcripts of the early

JGI/gene-ID	G1-phase fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
- / XLOC_015760	1.28	0.0604	no	-	-
217295 / XLOC_037315	1.15	0.6789	no	Q6ZQP7	Uncharacterized protein LOC284861
103163 / XLOC_029304	1.15	0.0136	yes	-	-
216181 / XLOC_035607	1.15	0.0612	no	B5Y3V8	Predicted protein
455975 / XLOC_010574	1.14	0.5251	no	Q2KHZ2	HBS1-like protein
230353 / XLOC_010590	1.14	0.4936	no	-	-
202497 / XLOC_011174	1.13	0.3932	no	-	-
449016 / XLOC_004485	1.11	0.2110	no	Q9EX73	Monoterpene epsilon-lactone hydrolase
234141 / XLOC_014712	1.10	0.7185	no	-	-
253036 / XLOC_009393	1.10	0.1436	no	-	-
212940 / XLOC_032136	1.08	0.1255	no	Q397B0	Transcriptional regulator, AraC family
205336 / XLOC_017556	1.08	0.5700	no	Q8BZ25	Ankyrin repeat and protein kinase domain-containing protein 1
219179 / XLOC_039750	1.04	0.1194	no	-	-
229875 / XLOC_009942	1.04	0.4709	no	-	-
- / XLOC_028155	1.03	0.0021	yes	-	-
215526 / XLOC_034875	1.03	0.4934	no	-	-
236751 / XLOC_017884	1.03	0.5248	no	F0VJG8	Putative uncharacterized protein
238038 / XLOC_019306	1.02	0.4301	no	-	-
206847 / XLOC_019800	1.01	0.5487	no	Q0WVK7	Pentatricopeptide repeat-containing protein At1g05670, mitochondrial
244412 / XLOC_027759	0.97	0.0194	yes	A9G2Z3	Putative membrane protein
117279 / XLOC_023358	0.96	0.0017	yes	-	-
426727 / XLOC_012634	0.93	0.5608	no	Q8T135	Kinesin-related protein 5
236239 / XLOC_017181	0.90	0.3933	no	A4RY54	Predicted protein

Unique transcripts of the early G1-phase

JGI/gene-ID	fpkm in early G1-phase	q_value	sig.	UniProt ID	Function
113812 / XLOC_014649	0.90	0.0023	yes	Q99NF1	Beta,beta-carotene 9',10'-oxygenase
212962 / XLOC_032145	0.89	0.5149	no	A0AUS0	WD repeat, SAM and U-box domain-containing protein 1
248875 / XLOC_036897	0.88	0.0104	yes	P55176	UPF0012 hydrolase in pqqF 5'region
- / XLOC_034399	0.88	0.0670	no	-	-
255132 / XLOC_022217	0.88	0.0281	yes	-	-
241150 / XLOC_023145	0.85	0.7507	no	C1EIX2	Predicted protein
122137 / XLOC_003630	0.84	0.1509	no	-	-

Appendix A Table - 4: Transcript of the graphically comprised cohorts from KOG-classes and their annotation.

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q_value	Sig.	Organism	Function
General function								
HC	431830	1029	14	Q0MYW8	1.29E-05	yes	*Emiliania huxleyi*	Putative calcium binding protein
HC	366938	453	60	Q8CIP4	1.34E-02	yes	*Mus musculus*	MAP/microtubule affinity-regul. kinase 4
HC	443010	403	53	-	4.57E-02	yes	-	-
HC	439733	377	74	-	3.58E-02	yes	-	-
HC	354307	307	3	-	8.12E-11	yes	-	
HC	463672	295	12	Q8H485	1.92E-04	yes	*Oryza sativa subsp. japonica*	Tubby-like F-box protein 11
HC	432500	280	4	Q39525	5.47E-07	yes	*Parachlorella kessleri*	H(+)/hexose cotransporter 3
LC	462385	54	788	P80030	1.25E-06	yes	*Brassica napus*	Enoyl-[acyl-carrier-protein] reductase [NADH], chloroplastic
LC	465364	3	394	P38230	0.00E+00	yes	*Saccharomyces cerevisiae* (strain ATCC 204508 / S288c)	Probable quinone oxidoreductase
LC	433442	58	336	P29618	2.27E-03	yes	*Oryza sativa subsp. japonica*	Cyclin-dependent kinase A-1
LC	442561	2	254	Q54FR4	0.00E+00	yes	*Dictyostelium discoideum*	PXMP2/4 family protein 4
Posttranslational modification, protein turnover, chaperones								
HC	362359	2139	48	P27323	7.30E-09	yes	*Arabidopsis thaliana*	Heat shock protein 90-1
HC	420962	1619	237	Q9U639	1.14E-02	yes	*Manduca sexta*	Heat shock 70 kDa protein cognate 4
HC	363448	1481	263	Q9SKQ0	4.37E-02	yes	*Arabidopsis thaliana*	Peptidyl-prolyl cis-trans isomerase CYP19-2
HC	426711	1206	141	P15253	4.33E-03	yes	*Oryctolagus cuniculus*	Calreticulin
HC	440477	685	18	-	3.14E-03	yes	-	-
HC	432765	501	46	Q6YYB0	3.02E-04	yes	*Oryza sativa subsp. japonica*	Uncharacterized protein Os08g0359500
HC	435425	501	88	Q70YI1	1.41E-02	yes	*Legionella pneumophila*	Outer membrane protein MIP

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
								Replication, recombination and repair
HC	40909	268	31	P33991	4.00E-02	yes	*Homo sapiens*	DNA replication licensing factor MCM4
HC	51597	252	28	P30664	1.36E-02	yes	*Xenopus laevis*	DNA replication licensing factor mcm4-B
HC	358490	215	22	Q9WUK4	4.64E-03	yes	*Mus musculus*	Replication factor C subunit 2
HC	447864	152	10	Q6DFV7	3.13E-04	yes	*Mus musculus*	Nuclear receptor coactivator 7
HC	75254	137	8	P43299	1.21E-06	yes	*Arabidopsis thaliana*	Protein PROLIFERA
HC	316205	135	10	P43299	2.69E-05	yes	*Arabidopsis thaliana*	Protein PROLIFERA
HC	310913	135	10	P43299	2.72E-05	yes	*Arabidopsis thaliana*	Protein PROLIFERA
HC	467665	107	18	O60182	9.75E-03	yes	*Schizosaccharomyces pombe* (strain 972 / ATCC 24843)	Replication factor C subunit 1
								RNA processing and modification
HC	443184	578	16	-	2.03E-04	yes	-	-
HC	438654	370	34	-	6.10E-04	yes	-	-
HC	439979	152	15	-	3.46E-04	yes	-	-
HC	373806	140	8	-	8.37E-08	yes	-	-
HC	98041	111	12	-	6.95E-04	yes	-	-
HC	434023	90	13	Q0CER3	6.21E-03	yes	*Aspergillus terreus* (strain NIH 2624 / FGSC A1156)	Putative uncharacterized protein
HC	443600	75	6	Q5DU56	1.06E-03	yes	*Mus musculus*	Protein NLRC3
LC	460685	15	207	-	9.03E-11	yes	-	-
LC	432507	6	93	Q0DM51	3.24E-13	yes	*Oryza sativa subsp. japonica*	DEAD-box ATP-dependent RNA helicase 3, chloroplastic
LC	467848	7	76	Q7ZWM3	8.44E-09	yes	*Xenopus laevis*	CUGBP Elav-like family member 3-B
LC	437070	19	72	D7FJQ9	1.26E-02	yes	*Ectocarpus siliculosus*	Putative uncharacterized protein

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q_value	Sig.	Organism	Function
								Intracellular trafficking, secretion, and vesicular transport
HC	444996	561	78	Q54GB3	3.12E-02	yes	*Dictyostelium discoideum*	Synaptobrevin-B
HC	316256	465	125	Q9FNC9	8.55E-03	yes	*Arabidopsis thaliana*	Mitochondrial import receptor subunit TOM22 homolog 2
HC	443186	377	15	P0CM16	2.49E-04	yes	*Cryptococcus neoformans var. neoformans serotype* D (strain JEC21 / ATCC MYA-565)	ADP-ribosylation factor
HC	436140	276	15	Q6AXQ8	4.97E-05	yes	*Rattus norvegicus*	Meiosis-specific nuclear structural protein 1
HC	437571	266	39	P47193	9.53E-03	yes	*Xenopus laevis*	Vesicle-associated membrane protein 2
HC	66952	174	27	F1SRI0	3.49E-03	yes	*Sus scrofa*	Trafficking protein particle complex subunit 2
HC	440861	172	18	Q9FMJ0	1.04E-02	yes	*Arabidopsis thaliana*	Phosphatidylinositol 4-kinase beta 1
HC	432331	162	11	Q01968	2.26E-03	yes	*Homo sapiens*	Inositol polyphosphate 5-phosphatase OCRL-1
HC	351006	162	27	O70480	3.31E-02	yes	*Mus musculus*	Vesicle-associated membrane protein 4
HC	434042	112	7	-	6.38E-03	yes	-	-
HC	463384	100	12	D3Z5L6	3.05E-02	yes	*Mus musculus*	MFS-type transporter C6orf192 homolog
HC	460762	96	22	Q9ZRD6	1.01E-02	yes	*Arabidopsis thaliana*	VAMP-like protein YKT61
HC	440111	93	2	A7YY49	7.56E-07	yes	*Bos taurus*	Surfeit locus protein 4
HC	434324	85	14	P93654	1.64E-02	yes	*Arabidopsis thaliana*	Syntaxin-22
LC	465064	35	138	B3RPE1	3.44E-03	yes	*Trichoplax adhaerens*	Putative uncharacterized protein
LC	51911	8	73	Q96321	5.68E-03	yes	*Arabidopsis thaliana*	Importin subunit alpha-1
LC	99733	21	58	Q84WW5	3.65E-02	yes	*Arabidopsis thaliana*	Vesicle-associated protein 1-3

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
							Inorganic ion transport and metabolism	
HC	466232	276	2	Q9HGM6	4.02E-06	yes	*Schizosaccharomyces pombe* (strain 972 / ATCC 24843)	Putative transporter C543.05c
HC	314659	207	41	Q9HGM6	2.61E-02	yes	*Schizosaccharomyces pombe* (strain 972 / ATCC 24843)	Putative transporter C543.05c
HC	103004	158	0	-	9.63E-08	yes	-	-
HC	437762	130	14	Q14973	5.43E-03	yes	*Homo sapiens*	Sodium/bile acid cotransporter
LC	447939	55	897	Q9QZM6	1.77E-05	yes	*Rattus norvegicus*	Sodium/potassium/calcium exchanger 1
LC	354606	90	553	Q9VN12	7.82E-03	yes	*Drosophila melanogaster*	Probable sodium/potassium/calcium exchanger CG1090
LC	416800	84	549	Q8L783	9.64E-04	yes	*Arabidopsis thaliana*	Vacuolar cation/proton exchanger 5
LC	428968	14	181	P17847	2.71E-07	yes	*Zea mays*	Ferredoxin--nitrite reductase, chloroplastic (Fragment)
LC	463095	1	156	Q9XES1	0.00E+00	yes	*Arabidopsis thaliana*	Calcium-transporting ATPase 4, endoplasmic reticulum-type
							LC	
LC	415031	150	2730	Q8DM25	1.16E-08	yes	*Thermosynechococcus elongatus* (strain BP-1)	50S ribosomal protein L7/L12
LC	77930	193	970	B0K5S7	6.58E-05	yes	*Thermoanaerobacter sp.* (strain X514)	50S ribosomal protein L13
LC	72821	167	855	B0K5S7	4.50E-05	yes	*Thermoanaerobacter sp.* (strain X514)	50S ribosomal protein L13
LC	416208	147	743	B2TTN0	2.11E-03	yes	*Nostoc punctiforme* (strain ATCC 29133 / PCC 73102)	50S ribosomal protein L17
LC	69198	93	674	P46228	1.13E-02	yes	*Synechococcus sp.* (strain ATCC 27144))	30S ribosomal protein S1
LC	49364	124	666	P27484	3.54E-04	yes	*Nicotiana sylvestris*	Glycine-rich protein 2

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q_value	Sig.	Organism	Function
							Translation, ribosomal structure and biogenesis	
LC	468608	131	595	P36236	4.52E-03	yes	Synechocystis sp. (strain ATCC 27184 / PCC 6803 / N-1)	50S ribosomal protein L1
LC	439611	95	576	B0CAC9	1.59E-03	yes	Acaryochloris marina (strain MBIC 11017)	50S ribosomal protein L11
LC	441818	93	514	Q3MFC0	9.58E-03	yes	Anabaena variabilis (strain ATCC 29413 / PCC 7937)	50S ribosomal protein L4
							Function unknown	
HC	437226	706	17	A7S4N4	3.94E-05	yes	Nematostella vectensis	Probable serine incorporator
HC	361672	238	20	Q54RS7	3.29E-03	yes	Dictyostelium discoideum	ELMO domain-containing protein C
HC	440226	202	16	-	3.68E-04	yes	-	-
HC	441326	150	40	Q9URW6	3.26E-02	yes	Schizosaccharomyces pombe (strain 972 / ATCC 24843)	SH3 domain-containing protein PJ696.02
HC	446071	130	9	Q9H841	1.16E-03	yes	Homo sapiens	NIPA-like protein 2
HC	437398	125	26	Q2KJ22	3.23E-02	yes	Bos taurus	Protein FAM63A
HC	361462	122	5	P10775	1.12E-04	yes	Sus scrofa	Ribonuclease inhibitor
HC	358812	114	18	Q8BMW7	3.51E-02	yes	Mus musculus	Magnesium transporter NIPA3
HC	433894	111	13	-	2.59E-02	yes	-	-
HC	436574	110	30	Q8BK64	4.33E-02	yes	Mus musculus	Activator of 90 kDa heat shock protein ATPase homolog 1
LC	469213	33	189	B7G571	9.35E-03	yes	Phaeodactylum tricornutum (strain CCAP 1055/1)	Predicted protein
LC	433126	30	141	-	1.58E-04	yes	-	-

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
							Lipid transport and metabolism	
HC	360281	820	116	P79274	4.72E-02	yes	Sus scrofa	Long-chain specific acyl-CoA dehydrogenase, mitochondrial
HC	432191	530	30	B8BTZ6	8.15E-04	yes	Thalassiosira pseudonana	Predicted protein
HC	435731	436	8	P46250	9.08E-05	yes	Candida albicans (strain SC5314 / ATCC MYA-2876)	SEC14 cytosolic factor
HC	437926	271	31	Q3ZBF6	1.01E-02	yes	Bos taurus	Short-chain specific acyl-CoA dehydrogenase, mitochondrial
HC	443385	258	3	-	2.47E-05	yes	-	-
HC	445286	197	18	Q8K214	7.24E-03	yes	Mus musculus	Polycomb protein SCMH1
HC	444795	155	9	Q9LVZ3	8.35E-04	yes	Arabidopsis thaliana	Probable lipid desaturase ADS3.2, chloroplastic
HC	442614	139	22	Q7SBB6	2.61E-02	yes	Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987)	Probable C-5 sterol desaturase
HC	438299	127	17	P04634	3.25E-02	yes	Rattus norvegicus	Gastric triacylglycerol lipase
HC	456684	120	15	Q5ZKR7	1.85E-02	yes	Gallus gallus	Long-chain-fatty-acid--CoA lig. ACSBG2
HC	464665	117	26	P45954	4.74E-02	yes	Homo sapiens	Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial
HC	442519	106	8	Q9R008	1.58E-03	yes	Mus musculus	Mevalonate kinase
HC	442508	104	12	P19967	1.03E-02	yes	Drosophila melanogaster	Cytochrome b5-related protein
HC	435778	101	20	Q9FS87	2.14E-02	yes	Solanum tuberosum	Isovaleryl-CoA dehydrogenase 2, mitochondrial (Fragment)
LC	455280	25	889	P11029	7.18E-09	yes	Gallus gallus	Acetyl-CoA carboxylase
LC	429248	3	531	Q00955	0.00E+00	yes	Saccharomyces cerevisiae (strain ATCC 204508 / S288c)	Acetyl-CoA carboxylase

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
								Lipid transport and metabolism
LC	449545	13	223	P11029	2.98E-05	yes	*Gallus gallus*	Acetyl-CoA carboxylase
LC	444084	43	209	Q92038	2.67E-03	yes	*Cyprinus carpio*	Acyl-CoA desaturase
LC	54629	68	181	Q8R3F5	2.40E-02	yes	*Mus musculus*	Malonyl-CoA-acyl carrier protein transacylase, mitochondrial
LC	447535	27	153	Q92038	2.59E-05	yes	*Cyprinus carpio*	Acyl-CoA desaturase
LC	351492	0	138	Q8BHI7	0.00E+00	yes	*Mus musculus*	Elongation of very long chain fatty acids protein 5
LC	438795	16	103	Q9DEX7	6.64E-06	yes	*Danio rerio*	Fatty acid desaturase 2
LC	417285	25	103	A4FV48	2.57E-03	yes	*Bos taurus*	Fatty acid desaturase 2
LC	454147	11	94	Q9DEX7	8.33E-07	yes	*Danio rerio*	Fatty acid desaturase 2
								Signal transduction mechanisms
HC	444252	1007	9	-	4.44E-07	yes	-	-
HC	435065	847	6	-	6.04E-08	yes	-	-
HC	443525	713	72	Q869W6	2.30E-02	yes	*Dictyostelium discoideum*	Probable myosin light chain kinase
HC	435910	642	4	-	3.19E-08	yes	-	-
HC	439698	510	2	-	3.93E-09	yes	-	-
HC	443934	461	3	-	3.21E-08	yes	-	-
HC	462780	431	3	Q9JMA9	3.52E-08	yes	*Mus musculus*	Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)
HC	435909	368	1	-	9.14E-10	yes	-	-
HC	440198	360	2	-	7.23E-10	yes	-	-
HC	462351	334	2	E9AEM9	1.53E-04	yes	*Leishmania major*	Proteophosphoglycan 5
HC	431876	332	3	Q9SN43	4.74E-07	yes	*Arabidopsis thaliana*	CBL-interacting serine/threonine-protein kinase 12

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q_value	Sig.	Organism	Function
Signal transduction mechanisms								
HC	440893	310	33	B8CAB4	1.99E-02	yes	Thalassiosira pseudonana	Predicted protein
HC	462359	305	6	-	2.33E-05	yes	-	-
HC	365420	258	32	Q3E9C0	9.12E-03	yes	Arabidopsis thaliana	Calcium-dependent protein kinase 34
LC	125588	55	475	A7BQ37	4.60E-05	yes	Beggiatoa sp. PS	Receptor protein kinase
LC	420555	32	294	O22437	3.16E-04	yes	Pisum sativum	Magnesium-chelatase subunit chlD, chloroplastic
LC	442554	21	250	P39442	3.10E-06	yes	Natronomonas pharaonis	Halocyanin
LC	68547	42	208	Q8YP49	1.09E-02	yes	Nostoc sp. (strain PCC 7120 / UTEX 2576)	1-deoxy-D-xylulose 5-phosphate reductoisomerase
LC								
LC	441363	0	395	-	5.37E-07	yes	-	-
LC	455397	16	209	C1N6V7	8.01E-06	yes	Micromonas pusilla (strain CCMP1545)	Predicted protein
LC	100060	3	63	E7F3N4	7.66E-11	yes	Danio rerio	Uncharacterized protein
LC	432966	9	61	-	2.82E-05	yes	-	-
LC	432237	0	56	-	2.71E-11	yes	-	-
Nucleotide transport and metabolism								
HC	365588	172	28	Q56E62	4.60E-03	yes	Nicotiana tabacum	Nucleoside diphosphate kinase 1
LC	60978	17	282	Q7Z8P9	5.28E-13	yes	Neosartorya fumigata (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100)	Nucleoside diphosphate kinase

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q_value	Sig.	Organism	Function
								Energy production and conversion
HC	451022	365	42	Q42686	1.18E-03	yes	*Chlamydomonas reinhardtii*	Malate dehydrogenase, mitochondrial
HC	349815	241	55	P39616	4.85E-02	yes	*Bacillus subtilis*	Probable aldehyde dehydrogenase ywdH
LC	361737	125	1133	P41345	7.37E-04	yes	*Oryza sativa subsp. japonica*	Ferredoxin--NADP reductase, root isozyme, chloroplastic
LC	461699	133	1037	Q06908	5.11E-03	yes	*Odontella sinensis*	ATP synthase gamma chain, chloroplastic
LC	437335	173	919	Q9LXC9	7.28E-03	yes	*Arabidopsis thaliana*	Soluble inorganic pyrophosphatase 1, chloroplastic
LC	432385	144	868	P41345	6.08E-03	yes	*Oryza sativa subsp. japonica*	Ferredoxin--NADP reductase, root isozyme, chloroplastic
LC	415940	54	607	Q5XIZ6	4.93E-07	yes	*Danio rerio*	Glycerol-3-phosphate dehydrogenase 1-like protein
LC	460094	29	348	Q27928	2.40E-09	yes	*Drosophila pseudoobscura pseudoobscura*	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic
LC	427769	4	310	O35077	0.00E+00	yes	*Rattus norvegicus*	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic
LC	45312	80	285	Q02585	1.24E-02	yes	*Nicotiana tabacum*	Cytochrome b6-f complex iron-sulfur subunit 2, chloroplastic
LC	417078	13	283	Q1XDM1	8.33E-07	yes	*Porphyra yezoensis*	Pyruvate dehydrogenase E1 component subunit beta
LC	448908	8	278	Q1RJT3	8.64E-14	yes	*Rickettsia bellii (strain RML369-C)*	Dihydrolipoyllysine-residue acetyltransferase comp. of pyruvate dehydrogenase complex
LC	437063	7	265	O95847	0.00E+00	yes	*Homo sapiens*	Mitochondrial uncoupling protein 4
LC	436192	0	212	Q93Y52	0.00E+00	yes	*Chlamydomonas reinhardtii*	Soluble inorganic pyrophosphatase 1, chloroplastic
LC	455448	46	211	B0G143	2.77E-03	yes	*Dictyostelium discoideum*	Mitochondrial SLC protein ucpB

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
Energy production and conversion								
Dark	420435	30	204	P72740	1.24E-04	yes	Synechocystis sp. (strain ATCC 27184 / PCC 6803 / N-1)	Dihydrolipoyl dehydrogenase
Transcription								
HC	433598	539	5	P10243	3.58E-06	yes	Homo sapiens	Myb-related protein A
HC	436010	297	21	-	5.76E-04	yes	-	-
HC	452735	256	14	Q9HOI3	2.04E-02	yes	Homo sapiens	Coiled-coil domain-containing protein 113
HC	440310	247	2	-	1.56E-08	yes	-	-
HC	454961	234	17	Q0JIC2	4.56E-03	yes	Oryza sativa subsp. japonica	Transcription factor GAMYB
HC	437110	230	45	-	4.81E-02	yes	-	-
HC	436479	215	22	Q4R8Y1	4.25E-04	yes	Macaca fascicularis	Bromodomain testis-specific protein
HC	439494	184	1	Q9S7L2	1.23E-09	yes	Arabidopsis thaliana	Transcription factor MYB98
HC	441236	184	1	-	4.13E-08	yes	-	-
LC	449660	0	352	P38529	8.64E-14	yes	Gallus gallus	Heat shock factor protein 1
LC	314592	42	227	P49263	1.77E-03	yes	Xenopus laevis	Pentraxin fusion protein
LC	354747	4	158	O97670	9.22E-11	yes	Oryctolagus cuniculus	Homeobox expressed in ES cells 1 (Fragment)
Coenzyme transport and metabolism								
HC	426057	80	16	Q8VHT6	7.30E-03	yes	Rattus norvegicus	Arsenite methyltransferase
HC	351930	78	8	-	2.10E-05	yes	-	-
LC	423968	85	476	Q6H6D2	2.74E-03	yes	Oryza sativa subsp. japonica	Porphobilinogen deaminase, chloroplastic
LC	443289	60	415	Q42698	1.25E-04	yes	Catharanthus roseus	Geranylgeranyl pyrophosphate synthase, chloroplastic
LC	432924	65	291	P35055	6.65E-03	yes	Glycine max	Coproporphyrinogen-III oxidase, chloroplastic

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
							Coenzyme transport and metabolism	
LC	427895	44	289	Q55467	3.82E-05	yes	Synechocystis sp. (strain ATCC 27184 / PCC 6803 / N-1)	Magnesium-protoporphyrin O-methyltransferase
LC	436908	61	274	Q8S4R4	1.35E-02	yes	Solanum lycopersicum	Prolycopene isomerase, chloroplastic
LC	361296	41	219	A5GJR6	6.18E-04	yes	Synechococcus sp. (strain WH7803)	Uroporphyrinogen decarboxylase
LC	56011	20	207	A7HSV9	2.99E-04	yes	Parvibaculum lavamentivorans (strain DS-1 / DSM 13023 / NCIMB 13966)	Bifunctional protein FolD
LC	465530	19	84	B1XIF6	7.52E-04	yes	Synechococcus sp. (strain ATCC 27264 / PCC 7002 / PR-6)	Uroporphyrinogen decarboxylase
LC	435069	22	79	Q2S1W0	5.10E-03	yes	Salinibacter ruber (strain DSM 13855 / M31)	Uroporphyrinogen decarboxylase
							Carbohydrate transport and metabolism	
HC	436725	776	15	-	7.39E-06	yes	-	-
HC	438851	671	62	P29495	6.86E-04	yes	Propionibacterium freudenreichii subsp. shermanii	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase
HC	439674	526	36	Q9UT63	1.64E-03	yes	Schizosaccharomyces pombe (strain 972 / ATCC 24843)	Probable phosphoglycerate mutase C513.02
HC	70323	272	24	P14618	2.50E-04	yes	Homo sapiens	Pyruvate kinase isozymes M1/M2
HC	439593	205	1	Q3UP75	2.64E-09	yes	Mus musculus	UDP-glucuronosyltransferase 3A1
HC	317489	187	41	B4ETL7	2.89E-02	yes	Proteus mirabilis (strain HI4320)	Bifunctional polymyxin resistance protein ArnA
LC	437959	205	1456	Q556J0	1.13E-02	yes	Dictyostelium discoideum	Transketolase
LC	436550	153	1315	Q0PAS0	1.09E-03	yes	Campylobacter jejuni	Fructose-bisphosphate aldolase
LC	417537	156	1035	Q9SBN4	2.57E-03	yes	Volvox carteri	Phosphoglycerate kinase, chloroplastic
LC	100158	105	496	Q42971	6.31E-03	yes	Oryza sativa subsp. japonica	Enolase

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
Carbohydrate transport and metabolism								
LC	351398	121	442	Q9ZU38	2.58E-02	yes	*Arabidopsis thaliana*	Probable ribose-5-phosphate isomerase
LC	440786	16	417	B6IRB5	8.64E-14	yes	*Rhodospirillum centenum* (strain ATCC 51521 / SW)	1-deoxy-D-xylulose-5-phosphate synthase
LC	433869	12	365	Q93VR3	4.10E-10	yes	*Arabidopsis thaliana*	GDP-mannose 3,5-epimerase
LC	418341	47	319	O52402	2.98E-03	yes	*Edwardsiella ictaluri* (strain 93-146)	Fructose-bisphosphate aldolase
LC	444895	40	298	P51181	3.18E-04	yes	*Bacillus licheniformis*	Pyruvate kinase
LC	438704	28	268	P0AG10	3.16E-07	yes	*Shigella flexneri*	Ribulose-phosphate 3-epimerase
LC	438492	28	267	P0AG10	4.44E-07	yes	*Shigella flexneri*	Ribulose-phosphate 3-epimerase
LC	439215	26	232	P51181	2.25E-05	yes	*Bacillus licheniformis*	Pyruvate kinase
Cytoskeleton								
HC	439339	482	19	-	4.78E-04	yes	-	-
HC	444747	301	26	-	1.22E-03	yes	-	-
HC	432151	225	8	-	2.27E-04	yes	-	-
LC	432636	163	1177	P26302	5.11E-03	yes	*Triticum aestivum*	Phosphoribulokinase, chloroplastic
LC	445562	140	1138	Q40832	1.08E-03	yes	*Pelvetia fastigiata*	Tubulin alpha-2 chain
LC	416424	155	704	Q40832	2.61E-02	yes	*Pelvetia fastigiata*	Tubulin alpha-2 chain
LC	78244	105	308	Q08114	4.98E-02	yes	*Euplotes octocarinatus*	Tubulin alpha chain
LC	459906	32	307	Q63425	4.35E-05	yes	*Rattus norvegicus*	Periaxin
	441006	46	279	D8TPL1	1.77E-03	yes	*Volvox carteri*	Metalloproteinase, extracellular matrix glycoprotein VMP10 (Fragment)
LC	457781	13	262	-	7.44E-08	yes	-	-

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q_value	Sig.	Organism	Function
							Amino acid transport and metabolism	
HC	438642	883	21	O93796	4.55E-04	yes	Candida glabrata (strain ATCC 2001 / CBS 138 / JCM 3761 / NBRC 0622 / NRRL Y-65)	Elongation factor 3
HC	452597	736	90	Q9I2V5	1.87E-02	yes	Pseudomonas aeruginosa (strain ATCC 15692 / PAO1 / 1C / PRS 101 / LMG 12228)	Aconitate hydratase 2
HC	433953	337	4	P29786	3.39E-07	yes	Aedes aegypti	Trypsin 3A1
HC	437487	294	70	P46643	4.63E-02	yes	Arabidopsis thaliana	Aspartate aminotransferase, mitochondrial
LC	455311	139	907	P56859	2.73E-02	yes	Emericella nidulans	Phosphoadenosine phosphosulfate reduc.
LC	437052	13	748	Q31QJ2	0.00E+00	yes	Synechococcus elongatus	Glutamate-1-semialdehyde 2,1-aminomutase
LC	443753	37	455	Q5VQL3	1.72E-07	yes	Oryza sativa subsp. japonica	Phosphoenolpyruvate/phosphate translocator 3, chloroplastic
LC	352050	94	450	Q9M5A9	9.59E-03	yes	Arabidopsis thaliana	Glucose-6-phosphate/phosphate translocator 1, chloroplastic
LC	432785	36	439	Q5VQL3	1.71E-07	yes	Oryza sativa subsp. japonica	Phosphoenolpyruvate/phosphate translocator 3, chloroplastic
LC	441299	112	408	A1SJJ3	4.46E-02	yes	Nocardioides sp. (strain BAA-499 / JS614)	Argininosuccinate synthase
LC	453525	63	283	P58338	6.38E-03	yes	Rhizobium meliloti (strain 1021)	Ornithine cyclodeaminase 1
LC	427526	61	219	P00931	4.69E-02	yes	Saccharomyces cerevisiae (strain ATCC 204508 / S288c)	Tryptophan synthase
LC	420632	20	214	Q54UH8	1.43E-04	yes	Dictyostelium discoideum	D-3-phosphoglycerate dehydrogenase
LC	418314	7	210	Q8RXN3	0.00E+00	yes	Arabidopsis thaliana	Phosphoenolpyruvate/phosphate translocator 1, chloroplastic
LC	416626	32	173	Q9LF61	7.86E-04	yes	Arabidopsis thaliana	Xylulose 5-phosphate/phosphate translocator, chloroplastic

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
							Amino acid transport and metabolism	
LC	471108	24	166	Q9SE94	3.25E-04	yes	Zea mays	Methylenetetrahydrofolate reductase 1
							Extracellular structures	
HC	431808	359	1	-	5.34E-10	yes	-	-
HC	436756	150	3	-	4.20E-07	yes	-	-
HC	442609	142	9	-	6.67E-04	yes	-	-
HC	437049	103	8	A4S1T5	3.85E-02	yes	Ostreococcus lucimarinus (strain CCE9901)	Predicted protein
LC	435083	279	1118	-	4.01E-02	yes	-	-
LC	455541	1	544	-	0.00E+00	yes	-	-
LC	466583	2	189	Q8LB31	0.00E+00	yes	Arabidopsis thaliana	Putative uncharacterized protein
LC	463630	7	175	P39442	8.20E-07	yes	Natronomonas pharaonis	Halocyanin
LC	432130	19	133	P74615	2.21E-06	yes	Synechocystis sp. (strain ATCC 27184 / PCC 6803 / N-1)	Uncharacterized protein sll1483
							Cell wall/membrane/envelope biogenesis	
HC	461984	190	24	Q9UVH3	9.38E-03	yes	Mortierella alpina	Palmitoyltransferase AKR1 (Fragment)
HC	358244	124	3	Q9SYM1	1.15E-04	yes	Arabidopsis thaliana	Uncharacterized mscS family protein At1g78610
HC	433117	119	6	Q6GQX6	1.55E-07	yes	Mus musculus	Ankyrin repeat and SAM domain-containing protein 6
HC	358242	118	3	Q9SYM1	7.68E-05	yes	Arabidopsis thaliana	Uncharacterized mscS family protein At1g78610
LC	466230	95	627	Q84KI6	4.59E-03	yes	Spinacia oleracea	UDP-sulfoquinovose synthase, chloroplastic
LC	448185	10	125	Q9LXT9	1.03E-05	yes	Arabidopsis thaliana	Callose synthase 3
LC	454930	18	105	Q5W915	1.48E-03	yes	Pisum sativum	UDP-sugar pyrophospharylase

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
							Chromatin structure and dynamics	
HC	50586	513	179	P82888	4.22E-02	yes	*Olisthodiscus luteus*	Histone H4
HC	350324	331	57	Q9I8G9	1.21E-02	yes	*Gallus gallus*	Histone-binding protein RBBP7
HC	371985	221	69	P04735	3.75E-02	yes	*Psammechinus miliaris*	Late histone H2A.1
LC	460675	1226	5329	-	4.74E-02	yes	-	-
LC	442150	382	1543	Q6YNC8	4.71E-02	yes	*Ovis aries*	Histone H2A.Z
							Secondary metabolites biosynthesis, transport and catabolism	
HC	359930	388	12	P25970	2.41E-04	yes	*Myxococcus xanthus* (strain DK 1622)	Uncharacterized oxidoreductase MXAN_5909
HC	53289	306	3	O23024	1.90E-03	yes	*Arabidopsis thaliana*	Flavin-containing monooxygenase YUCCA3
HC	438972	302	16	Q99J47	1.16E-03	yes	*Mus musculus*	Dehydrogenase/reductase SDR family member 7B
HC	454764	270	23	Q9ZU35	1.56E-02	yes	*Arabidopsis thaliana*	ABC transporter G family member 7
LC	435829	798	3384	Q1IHZ5	2.45E-02	yes	*Koribacter versatilis* (strain Ellin345)	Acyl carrier protein
LC	457138	78	340	Q1IHZ5	2.11E-03	yes	*Koribacter versatilis* (strain Ellin345)	Acyl carrier protein
LC	415523	5	338	O34340	0.00E+00	yes	*Bacillus subtilis*	3-oxoacyl-[acyl-carrier-protein] synthase 2
LC	461962	44	267	P43710	2.48E-03	yes	*Haemophilus influenzae* (strain ATCC 51907 / DSM 11121 / KW20 / Rd)	3-oxoacyl-[acyl-carrier-protein] synthase 1
LC	438034	49	253	O34340	1.70E-03	yes	*Bacillus subtilis*	3-oxoacyl-[acyl-carrier-protein] synthase 2

Appendix A - Table - 4: Transcript of the graphically comprised cohorts and their annotation continued:

HC or LC condition	JGI#	fpkm in HC condition	fpkm in LC condition	NR Prot.ID	q value	Sig.	Organism	Function
Cell cycle control, cell division, chromosome partitioning								
HC	113594	176	7	-	1.89E-04	yes	-	-
HC	366332	140	8	O94580	2.98E-03	yes	Schizosaccharomyces pombe (strain 972 / ATCC 24843)	Ubiquitin and WLM domain-containing protein C1442.07c
HC	102888	100	4	-	6.88E-05	yes	-	-
HC	374001	74	18	Q00737	1.16E-02	yes	Emericella nidulans	Chromosome segregation protein sudA
LC	466575	30	457	P46277	5.47E-09	yes	Medicago sativa subsp. varia	G2/mitotic-specific cyclin-1
LC	125202	48	361	P59241	8.79E-05	yes	Rattus norvegicus	Aurora kinase A
LC	359888	46	234	P97477	3.22E-03	yes	Mus musculus	Aurora kinase A
LC	72589	20	163	Q9FJL0	1.70E-06	yes	Arabidopsis thaliana	Structural maintenance of chromosomes protein 4
LC	465321	14	163	Q54PK4	7.69E-04	yes	Dictyostelium discoideum	Structural maintenance of chromosomes protein 2
LC	456538	6	98	Q9NTJ3	3.24E-13	yes	Homo sapiens	Structural maintenance of chromosomes protein 4
LC	468124	7	86	Q9BPX3	3.26E-04	yes	Homo sapiens	Condensin complex subunit 3
LC	448323	6	85	Q15003	3.02E-05	yes	Homo sapiens	Condensin complex subunit 2
LC	444453	21	73	P0CP40	1.14E-02	yes	Cryptococcus neoformans var. neoformans serotype D (strain JEC21 / ATCC MYA-565)	Probable kinetochore protein NUF2
Defence mechanisms								
HC	58171	78	1	Q54T76	2.62E-05	yes	Dictyostelium discoideum	Probable dual specificity protein phosphatase
HC	446038	68	2	-	1.70E-11	yes	-	-
Cell motility								
HC	122222	211	21	E9AEM8	2.79E-03	yes	Leishmania major	Proteophosphoglycan ppg4

Appendix A Table - 5: Transcripts matching proteins of the V-type proton ATPase cluster and clathrin cluster.

Protein JGI#	UniProt ID	Function	Short name	2^X-fold change.	fpkm HC at condition	fpkm LC at condition	Sig.	KOG definition
		V-type proton ATPase and clathrin						
439740	Q8PYZ8	K(+)-stimulated pyrophosphate-energized sodium pump	H PPase	1.32	235	586	no	-
415047	P21616	Pyrophosphate-energized vacuolar membrane proton pump	V-type PPase	-0.55	70	48	no	-
51239	Q06572	Pyrophosphate-energized vacuolar membrane proton pump	V-type PPase	0.34	67	85	no	-
75032	Q06572	Pyrophosphate-energized vacuolar membrane proton pump	V-type PPase	0.5	20	28	no	-
464767	Q9Z1G4	V-type proton ATPase 116 kDa subunit a isoform 1	V-ATPase	0.78	28	49	no	Vacuolar H^+-ATPase V0 sector, subunit a
61253	Q9Z1G4	V-type proton ATPase 116 kDa subunit a isoform 1	V-ATPase	2.18	79	358	no	Vacuolar H^+-ATPase V0 sector, subunit a
313422	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	V-ATPase	1.25	116	277	no	Vacuolar H^+-ATPase V0 sector, subunits c/c'
359783	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	V-ATPase	0.96	748	1460	no	Vacuolar H^+-ATPase V0 sector, subunits c/c'
362459	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	V-ATPase	1.22	211	489	no	Vacuolar H^+-ATPase V0 sector, subunits c/c'
364707	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	V-ATPase	0.9	153	285	no	Vacuolar H^+-ATPase V0 sector, subunits c/c'
366512	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	V-ATPase	1.18	81	184	no	Vacuolar H^+-ATPase V0 sector, subunits c/c'
457332	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	V-ATPase	1.74	108	361	no	Vacuolar H^+-ATPase V0 sector, subunits c/c'
451883	Q91V37	V-type proton ATPase 21 kDa proteolipid subunit	V-ATPase	1.21	48	110	no	Vacuolar H^+-ATPase V0 sector, subunit c''

Appendix A - Table - 5: Transcripts matching proteins of the V-type proton ATPase cluster and clathrin cluster continued:

Protein JGI#	UniProt ID	Function	Short name	2^X-fold change.	fpkm HC condition.	fpkm LC condition	Sig.	KOG definition
		V-type proton ATPase and clathrin						
439538	P54647	V-type proton ATPase catalytic subunit A	V-ATPase	1.28	323	786	no	Vacuolar H^+-ATPase V1 sector, subunit A
435128	P31408	V-type proton ATPase subunit B, brain isoform	V-ATPase	1.32	391	976	no	Vacuolar H^+-ATPase V1 sector, subunit B
313800	P23968	V-type proton ATPase subunit c"	V-ATPase	1.28	154	374	no	Vacuolar H^+-ATPase V0 sector, subunit c"
413949	P54641	V-type proton ATPase subunit d	V-ATPase	1.04	360	739	no	Vacuolar H^+-ATPase V0 sector, subunit d
420005	Q9XGM1	V-type proton ATPase subunit D	V-ATPase	1.2	76	175	no	Vacuolar H^+-ATPase V1 sector, subunit D
425833	Q9V7D2 1	V-type proton ATPase subunit D	V-ATPase	0.86	138	251	no	Vacuolar H^+-ATPase V1 sector, subunit D
433060	Q9MB46	V-type proton ATPase subunit E	V-ATPase	0.35	411	526	no	Vacuolar H^+-ATPase V1 sector, subunit E
352209	Q9ZQX4	V-type proton ATPase subunit F	V-ATPase	0.95	128	246	no	Vacuolar H^+-ATPase V1 sector, subunit F
355949	P91303	Probable V-type proton ATPase subunit G	V-ATPase	0.46	556	765	no	Vacuolar H^+-ATPase V1 sector, subunit G
369392	P91303	Probable V-type proton ATPase subunit G	V-ATPase	0.35	299	382	no	Vacuolar H^+-ATPase V1 sector, subunit G
95543	Q9LX65	V-type proton ATPase subunit H	V-ATPase	1.25	217	518	no	Vacuolar H^+-ATPase V1 sector, subunit H
461741	Q8TDJ6	RAVE (regulator of V-ATPase assembly) complex subunit	DmX	1.7	5	16	yes	RAVE, RAV1/DMX protein, WD repeat superfamily
67081	P12522	Probable proton ATPase 1B	H Ppase	3.18	1	10	yes	Plasma membrane H^+-transporting ATPase

214

Appendix A - Table - 5: Transcripts matching proteins of the V-type proton ATPase cluster and clathrin cluster continued:

Protein JGI#	UniProt ID	Function	Short name	2^x-fold change	fpkm HC condition	fpkm LC condition	Sig.	KOG definition
		V-type proton ATPase and clathrin						
456439	P11442	Clathrin heavy chain 1		-0.23	6	5	no	Vesicle coat protein clathrin, heavy chain
455942	P25870	Clathrin heavy chain		-0.63	23	15	no	Vesicle coat protein clathrin, heavy chain
352114	P25870	Clathrin heavy chain		0.47	320	443	no	Vesicle coat protein clathrin, heavy chain
209558	Q14677	Clathrin interactor 1		0.47	6	8	no	Clathrin coated vesicle

Appendix A Table - 6: Transcripts matching proteins of the proton exchanger cluster.

Protein JGI#	UniProt ID	Function	Short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG definition
		Proton Exchanger						
469557	Q2YMB3	Na(+)/H(+) antiporter nhaA	NhaA	3.1	19	14	no	-
423009	Q2YMB3	Na(+)/H(+) antiporter nhaA	NhaA	3.1	19	15	no	-
468382	Q8YFI5	Na(+)/H(+) antiporter nhaA	NhaA	2.6	7	12	no	Carbonic anhydrase
447659	Q39WP5	Na(+)/H(+) antiporter nhaA	NhaA	3.2	11	30	no	-
198981	Q30XM9	Na(+)/H(+) antiporter nhaA	NhaA	3.2	9	27	yes	-
219535	A1AK41	Na(+)/H(+) antiporter nhaA 1	NhaA	1.1	0	2	yes	-
105293	A1ATB4	Na(+)/H(+) antiporter nhaA 2	NhaA	1.9	3	5	no	Pantothenate kinase and related prot.
365993	E3PR21	Putative Na(+)/H(+) antiporter	NhaA	0.9	1	1	no	-
358003	E3PR21	Putative Na(+)/H(+) antiporter	NhaA	1.4	2	2	no	-
359372	C1E8I2	Monovalent Cation:Proton antiporter-1 family		2.7	12	9		-
120657	C1E8I2	Monovalent Cation:Proton antiporter-1 family		2.3	6	7		-
459148	Q4PKH3	H(+)/Cl(-) exchange transporter 7	ClC-7	2.6	5	13	no	Cl- channel CLC-7 and related proteins
452730	Q4PKH3	H(+)/Cl(-) exchange transporter 7	ClC-7	1.5	2	3	no	Cl- channel CLC-7 and related proteins
221620	P51798	H(+)/Cl(-) exchange transporter 7	ClC-7	0.6	0	1	no	Cl- channel CLC-7 and related proteins
450698	Q87GZ9	H(+)/Cl(-) exch. transporter ClcA	ClC-7	2.1	5	6	no	Cl- channel CLC-7 and related proteins
432500	Q39525	H(+)/hexose cotransporter 3		-6.1	280	4	yes	Predicted transporter
461000	E2IJ90	Voltage-gated H+ channel protein		2	5	19	no	Beta-catenin-binding protein APC

Appendix A Table - 7: Transcripts matching proteins of the Ca^{2+} - transport cluster.

Presence of Ca^{2+} - transport transcripts

JGI#	UniProt ID	Function	short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG-Definition
		K+ dependent Ca2+/Na+ exchanger (NCKX)						
447939	Q9QZM6	Sodium/potassium/calcium exchanger 1	NCKX1	4.0	55	897	yes	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins
461099	Q9HC58	Sodium/potassium/calcium exchanger 3	KCKX3	0.4	23	29	no	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins
447153	Q9EPQ0	Sodium/potassium/calcium exchanger 3 (Fragment)	NCKX3'	-0.4	14	11	no	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins
437634	Q9EPQ0	Sodium/potassium/calcium exchanger 3 (Fragment)	NCKX3'	0.5	4	6	no	
368842	Q9EPQ0	Sodium/potassium/calcium exchanger 3 (Fragment)	NCKX3'	0.5	8	11	no	
450681	Q8C261	Sodium/potassium/calcium exchanger 5	NCKX5	0.9	7	12	no	Splicing coactivator SRm160/300, subunit SRm300
258305	Q8C261	Sodium/potassium/calcium exchanger 5	NCKX5	1.6	1	2	no	
212884	Q8C261	Sodium/potassium/calcium exchanger 5	NCKX5	0.8	5	9	no	
469467	C1MI03	Predicted protein		0.8	2	4	no	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins
354606	Q9VN12	Probable sodium/potassium/calcium exchanger CG1090		2.6	90	553	yes	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins
219401	Q9VN12	Probable sodium/potassium/calcium exchanger CG1090		0.5	5	7	no	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins
354001	-	NCKX1 and related proteins		0.2	4	5	no	K^+-dependent Ca^{2+}/Na^+ exchanger NCKX1 and related proteins

Presence of Ca²⁺ - transport transcripts

JGI#	UniProt ID	Function	short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG-Definition
K⁺ independent Ca2⁺/Na⁺ exchanger (NCX)								
454623	P57103	Sodium/calcium exchanger 3	NCX3	1.79	5	18	yes	Ca²⁺/Na⁺ exchanger NCX1 and related proteins
369786	P70549	Sodium/calcium exchanger 3	NCX3	0.951	2		no	Ca²⁺/Na⁺ exchanger NCX1 and related proteins
Cation/proton exchanger (CAX)								
72273	Q5KQN0	Vacuolar cation/proton exchanger 2	CAX2	0.97	4	8	no	Ca²⁺/H⁺ antiporter VCX1 and related proteins
223499	Q5KQN0	Vacuolar cation/proton exchanger 2	CAX2	-0.71	14	9	no	Ca²⁺/H⁺ antiporter VCX1 and related proteins
416800	Q8L783	Vacuolar cation/proton exchanger 5	CAX5	2.7	84	549	yes	Ca²⁺/H⁺ antiporter VCX1 and related proteins
415715	Q8L783	Vacuolar cation/proton exchanger 5	CAX5	-1.07	33	16	no	-
Cation/Ca2+ exchanger (CCX)								
449053	Q9LJI2	Cation/calcium exchanger 3	AtCCX3	-2.80	89	13	no	K⁺-dependent Na⁺:Ca²⁺ antiporter
Calcium transporting ATPase – endoplasmatic reticulum type (SERCA-type)								
258138	Q42883	Calcium-transporting ATPase, endoplasmic reticulum-type	SERCA	-4.88	4	0	no	
62350	P11607	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	SERCA 2	1.24	7	17	no	
251608	Q9XES1	Calcium-transporting ATPase 4, endoplasmic reticulum-type	SERCA 4	-3.56	64	5	yes	
429294	Q9XES1	Calcium-transporting ATPase 4, endoplasmic reticulum-type	SERCA 4	-2.25	62	13	no	Ca²⁺ transporting ATPase
463095	Q9XES1	Calcium-transporting ATPase 4, endoplasmic reticulum-type	SERCA 4	6.83	1	156	yes	Ca²⁺ transporting ATPase

Presence of Ca^{2+} - transport transcripts

JGI#	UniProt ID	Function	short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG-Definition
		Plasma membrane calcium-transporting ATPase (PMCA)						
466567	Q64542	Plasma membrane calcium-transporting ATPase 4	PMCA4	0.79	4	6	no	Calcium transporting ATPase
460140	Q64542	Plasma membrane calcium-transporting ATPase 4	PMCA4	2.6	0	2	yes	Calcium transporting ATPase
		Putative proteins						
78746	P35315	Probable calcium-transporting ATPase		3.28	0	5	yes	Ca^{2+} transporting ATPase
107905	B7FRE6	Predicted protein		0.78	10	17	no	F0F1-type ATP synthase, **α** subunit
		Calcium Channels						
101180	Q6S5H8	Two pore calcium channel protein 1	HvTPC1	-0.07	3	3	no	Voltage-gated Ca^{2+} channels, alpha1 subunits
452741	Q75VR0	Two pore calcium channel protein 1B	HvTPC1	0.14	5	6	no	Voltage-gated Ca^{2+} channels, alpha1 subunits
452291	Q9NY47	Voltage-dependent calcium channel subunit alpha-2/delta-2		2.92	1	11	yes	L-type voltage-dependent Ca^{2+} channel, alpha2/delta subunit
		Transient receptor potential cation channel						
468587	Q91YD4	Transient receptor potential cation channel subfamily M member 2 (1)	LTrpC-2	1.71	7	22	no	Ca^{2+}/Mg^{2+}-permeable cation channels (LTRPC family)
462171	P48994	Transient-receptor-potential-like protein (1)		0.32	8	10	no	Receptor-activated Ca^{2+}-permeable cation channels (STRPC family)
460292	Q91YD4	Transient receptor potential cation channel subfamily M member 2 (1)	LTrpC-2	-0.71	48	29	no	Ca^{2+}/Mg^{2+}-permeable cation channels (LTRPC family)
455760	O94759	Transient receptor potential cation channel subfamily M member 2 (1)	LTrpC-2	1.76	7	25	no	Ca2^{2+}/Mg^{2+}-permeable cation channels (LTRPC family)
449985	O94759	Transient receptor potential cation channel subfamily M member 2 (1)	LTrpC-2	1.68	7	22	no	Ca^{2+}/Mg^{2+}-permeable cation channels (LTRPC family)
250817	O94759	Transient receptor potential cation channel subfamily M member 2 (1)	LTrpC-2	3.27	1	12	yes	-

Presence of Ca^{2+} - transport transcripts

JGI#	UniProt ID	Function	short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG-Definition
107737	O94759	Transient receptor potential cation channel subfamily M member 2 (1)	LTrpC-2	-0.59	6	4	no	Ca^{2+}/Mg^{2+}-permeable cation channels (LTRPC family)

(1) Predicted protein referring to KOG definition annotation

Appendix A Table - 8: Transcripts matching calmodulin and calcium binding proteins.

JGI#	UniProt ID	Function	Short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG-Definition
		Transcripts related to Calmodulin (CaM) and Ca^{2+} - binding protein						
48130	Q9HFY6	Calmodulin	CaM	-0.8	31	18	no	Calmodulin and related proteins
58485	Q9HFY6	Calmodulin	CaM	-0.8	31	18	no	Calmodulin and related proteins
373343	Q40302	Calmodulin	CaM	-1.4	2	1	no	Calmodulin and related proteins
442625	Q40302	Calmodulin	CaM	-1.3	2539	1031	no	Calmodulin and related proteins
443126	Q40302	Calmodulin	CaM	-1.4	1540	579	no	Calmodulin and related proteins
60184	P27482	Calmodulin-like protein 3	CLP3	0.7	3	5	no	Calmodulin and related proteins
47652	Q9NZT1	Calmodulin-like protein 5	CLP5	1.8	1	2	no	Calmodulin and related proteins
50530	Q9NZT1	Calmodulin-like protein 5	CLP5	1.8	1	2	no	Nucleolar GTPase/ATPase p130
109061	Q9NZT1	Calmodulin-like protein 5	CLP5	-2.9	1	0	yes	Calmodulin and related proteins
118982	Q9LNE7	Calmodulin-like protein 7	CLP7	3.1	0	2	no	Calmodulin and related proteins
61107	Q2R1Z5	Putative calmodulin-like protein 6		-0.3	96	78	no	Calmodulin and related proteins
431830	Q0MYW8	Polysaccharide associated protein GPA	GPA	-6.2	1029	14	yes	Putative calcium binding protein
446255	P15253	Calreticulin	ERp60	-2.9	310	42	yes	Calreticulin
426711	P15253	Calreticulin	ERp60	-3.1	1206	141	yes	Calreticulin
435301	Q39817	Calnexin homolog		-1.4	82	31	no	Calnexin
422633	Q6Q487	Calnexin homolog		-3.7	199	15	yes	Calnexin
49942	Q38798	Calnexin homolog 2		-3.9	316	21	yes	Calnexin
95284	P38505	Calcium-binding protein	CABP	0.7	37	62	no	Calmodulin and related proteins
434920	F3I218	Calcium-binding protein (Fragment)	CABP	-2.6	7	1	yes	-
100360	Q9FIH9	Calcium-binding protein CML37	CML37	0.4	8	11	no	Actin filament-coating protein tropomyosin

Appendix A Table - 9: Bicarbonate transport protein cluster

JGI#	UniProt ID	Function	short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG Definition (source: JGI annotation)
					Bicarbonate transport			
358486	-	-	-	1.36	3	7	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
99943	P23347	Anion exchange protein 2	AE 2	3.81	8	117	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
426735	P23347	Anion exchange protein 2	AE 2	1.31	60	150	no	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
200137	P23347	Anion exchange protein 2	AE 2	2.72	39	259	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
198643	P23347	Anion exchange protein 2	AE 2	3.09	19	158	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
120259	P23347	Anion exchange protein 2	AE 2	3.60	3	35	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
450694	Q9GKY1	Anion exchange protein 4	AE 4	0.67	16	26	no	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
196761	Q9GKY1	Anion exchange protein 4	AE 4	1.25	5	12	no	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
453061	Q8GYH8	Probable sulfate transporter 4.2	-	0.24	6	7	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
441761	O06984	Putative sulfate transporter yvdB	-	1.12	137	298	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
363809	O06984	Putative sulfate transporter yvdB	-	-1.81	40	11	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
466232	Q9HGM6	Putative transporter C543.05c	-	-7.07	276	2	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
436956	Q9HGM6	Putative transporter C543.05c	-	-2.25	90	19	no	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)
314659	Q9HGM6	Putative transporter C543.05c	-	-2.33	207	41	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC4 family)

JGI#	UniProt ID	Function	short name	2^x-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG Definition (source: JGI annotation)
								and related transporters (SLC4 family)
469783	Q8BTY2	Sodium bicarbonate cotransporter 3	SLC7	4.00	1	15	yes	Na$^+$-independent Cl/HCO3 exchanger AE1 and related transporters (SLC7 family)
448706	Q09764	Uncharacterized protein C24H6.11c	SLC26	0.57	6	9	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
121358	Q09764	Uncharacterized protein C24H6.11c	SLC26	-1.62	6	2	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
98125	P53273	Uncharacterized vacuolar membr. protein YGR125W	SLC26	-6.90	19	0	yes	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
460215	P53273	Uncharacterized vacuolar membr. protein YGR125W	SLC26	1.09	4	8	no	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
455488	P53273	Uncharacterized vacuolar membr. protein YGR125W	SLC26	1.98	1	2	yes	Sulfate/bicarbonate/oxalate exchanger SAT-1 and related transporters (SLC26 family)
			Carbonic anhydrase					
233460	O52535	Beta-Carbonic anhydrase	CA	2.63	1	6	yes	Carbonic anhydrase
239690	O53573	Beta-Carbonic anhydrase	CA	-0.31	1	1	no	Carbonic anhydrase
115240	O53573	Beta-Carbonic anhydrase	CA	-0.09	3	3	no	Carbonic anhydrase
436031	Q0ZB86	Delta-carbonic anhydrase	δ-CA	7.89	0	74	yes	Carbonic anhydrase
62679	Q50940	Carbonic anhydrase	CA	1.68	1	4	no	Carbonic anhydrase
456048	Q50940	Carbonic anhydrase	CA	2.61	2	14	yes	Carbonic anhydrase
195575	Q6V8K7	Intracellular carbonic anhydrase	CA	1.13	11	25	no	Carbonic anhydrase

Appendix A Table - 10: Aquaporin protein cluster.

JGI#	UniProt ID	Function	short name	2^X-fold change	fpkm HC condition	fpkm LC condition	sig.	KOG definition
		Aquaporins						
75635	P42767	Aquaporin PIP-type	AP	0.84	4	8	no	Aquaporin (major intrinsic protein family)
458101	Q94CS9	Probable aquaporin TIP1-2	OsTIP1:2	0.05	45	46	no	Aquaporin (major intrinsic protein family)
454880	Q94CS9	Probable aquaporin TIP1-2	OsTIP1;2	0.01	5	5	no	Aquaporin (major intrinsic protein family)
111379	O82316	Aquaporin TIP4-1	OsTIP1;4	1.55	10	29	no	Nuclear receptor coregulator SMRT/SMRTER, contains Myb-like domains
462297	Q1J3H7	Glycerol uptake facilitator GlpF, MIP/aquaporin family (Precursor)	GlpF	1.51	1	3	no	Aquaporin (major intrinsic protein family)
424870	Q1J3H7	Glycerol uptake facilitator GlpF, MIP/aquaporin family (Precursor)	GlpF	1.27	3	7	no	Aquaporin (major intrinsic protein family)
432761	P31140	Glycerol uptake facilitator protein	GlpF	-0.81	10	6	no	Aquaporin (major intrinsic protein family)

Heat Shock Protein 70 homolog expression
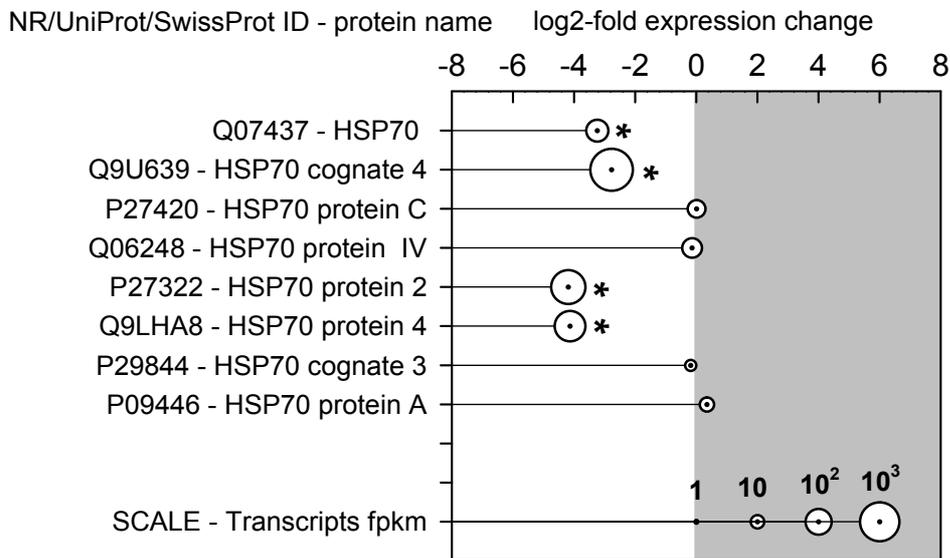
Appendix A Figure 2: Transcripts related to Heat Shock Protein 70 expression found in *E. huxleyi* cells showing high and low rates of calcification (shaded grey). Positive 2^X-fold expression values indicate elevated expression in the Dark period. The sizes of the circles indicate the total abundance of transcripts found at Light and Dark period.

Appendix A Table - 11: Details of transcript abundance of heat shock protein in *E. huxleyi* cells showing high (HC) and low rates of calcification (LC).

JGI#	UniProt ID	Short name	2^X-fold change	fpkm at HC	fpkm at LC	sig.
311681	Q07437	HSP 70 kDa	-3.24	42.8	4.5	yes
420962	Q9U639	HSP 70 kDa 4 cognate	-2.77	1619.2	236.8	yes
421892	P27420	HSP 70 kDa C	0.01	10.9	10.9	no
423268	Q06248	HSP 70 kDa IV	-0.14	16.2	14.7	no
441734	P27322	HSP 70 kDa 2 cognate	-4.19	408.7	22.5	yes
47896	Q9LHA8	HSP 70 kDa 4 cognate	-4.13	212.9	12.2	yes
59809	P29844	HSP 70 kDa 3 cognat	-0.18	3.4	3.0	no
62931	P09446	HSP 70 kDa A	0.35	5.0	6.4	no

APPENDIX A – SUPPLEMENTARY RESULTS – CHAPTER 3 – THE TRANSCRIPTOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

225

Appendix A Table - 12: Selected transcripts with potential role in biomineralisation of the 1000 most abundant transcripts in the light incubation period (high calcification rate) of the cell cycle.

JGI#	UniProt ID	Function	fpkm in Light	fpkm in Dark	2^X-fold change	q_value	Sig.
442625	Q40302	Calmodulin	2539.2	1030.6	-1.3	0.5585	no
443126	Q40302	Calmodulin	1540.1	578.9	-1.41	0.3289	no
426711	P15253	Calreticulin	1206.0	141.0	-3.1	0.0043	yes
431830	Q0MYW8	GPA	1029.1	14.4	-6.16	0.0000	yes
359783	Q43362	V-type proton ATPase 16 kDa proteolipid subunit	748.2	1460.4	0.96	0.8089	no
443525	Q869W6	Ca^{2+}/calmodulin-dependent protein kinase, EF-Hand	713.0	72.0	-3.31	0.0230	yes
355949	P91303	Probable V-type proton ATPase subunit G	556.2	765.2	0.46	0.7748	no
433060	Q9MB46	V-type proton ATPase subunit E	411.4	525.9	0.35	0.8586	no
435128	P31408	V-type proton ATPase subunit B, brain isoform	390.7	976.2	1.32	0.2916	no
413949	P54641	V-type proton ATPase subunit d	359.5	739.4	1.04	0.4089	no
439538	P54647	V-type proton ATPase catalytic subunit A	323.4	785.5	1.28	0.3852	no
352114	P25870	Clathrin heavy chain	320.5	443.1	0.47	0.7282	no
49942	Q38798	Calnexin homolog 2	316.2	21.1	-3.91	0.0000	yes
446255	P15253	Calreticulin	310.3	41.7	-2.9	0.0103	yes
432823	Q7T385	V-type proton ATPase subunit C 1-A	302.8	460.8	0.61	0.7537	no
369392	P91303	Probable V-type proton ATPase subunit G	299.3	381.5	0.35	0.8057	no
466232	Q9HGM6	Putative transporter C543.05c	275.6	2.0	-7.07	0.0000	yes
365420	Q3E9C0	Calcium-dependent protein kinase 34	257.5	31.5	-3.03	0.0091	yes
432215	Q01474	GTP-binding protein SAR1B	255.2	103.5	-1.3	0.3831	no
436781	P60881	Synaptosomal-associated protein 25	234.4	32.7	-2.84	0.0509	no

Appendix B – Supplementary results for Chapter 4 - The proteome of *Emiliania huxleyi* at high and low rates of calcification of the G1-phase

228

Appendix B Table - 1: Proteins identified in in *E. huxleyi* cells showing high (HC) and low rates of calcification (LC) with significant expression direction ratios LC: HC, and consensus sequences in phylogenetic homology analysis.

Appendix B Table 1 DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI65061	Q2IA28	Chloroplast light harvesting protein isoform 12	HC	0.594	0.433	0.751	HAJGI222
JGI281147	Q5UU97	Enolase	HC	0.600	0.447	0.729	HAJGI283
JGI206028	Q2IA28	Chloroplast light harvesting protein isoform 12	HC	0.632	0.487	0.763	HAJGI261
JGI66310	F0XZQ3	Putative uncharacterized protein	HC	0.641	0.416	0.911	HAJGI043
JGI269056	Q39709	Fucoxanthin-chlorophyll a-c binding protein, chloroplastic	HC	0.654	0.565	0.845	HAJGI382
JGI407983	C1E1W7	Clathrin heavy chain		0.664	0.387	1.053	HAJGI365
JGI411074				0.672	0.470	1.131	HAJGI055
JGI62246	Q2IA73	Chloroplast light harvesting protein isoform 6	HC	0.693	0.544	0.787	HAJGI246
JGI313236	A4S0B8	Predicted protein		0.694	0.430	1.124	HAJGI177
JGI109974		NA	HC	0.699	0.479	0.998	NA
JGI220283		NA	HC	0.700	0.653	0.772	NA
JGI103600	D8LB65	Glucokinase	HC	0.703	0.522	0.884	HAJGI275
JGI218041	P51821	ADP-ribosylation factor 1	HC	0.704	0.568	0.924	HAJGI036
JGI67584	Q2IA28	Chloroplast light harvesting protein isoform 12	HC	0.704	0.489	0.926	HAJGI264
JGI45869	Q2IA76	Chloroplast light harvesting protein isoform 3	HC	0.705	0.538	0.937	HAJGI029
JGI104438	C1DZ32	Predicted protein	HC	0.710	0.540	0.843	HAJGI260
JGI281510	F0YM05	Putative uncharacterized protein	HC	0.710	0.589	0.853	HAJGI085
JGI62634	A8J7F6	Dihydrolipoamide acetyltransferase		0.715	0.295	1.168	HAJGI021
JGI96082		NA	HC	0.718	0.594	0.805	NA
JGI61814	D8UI70	Dual function alcohol dehydrogenase		0.723	0.577	1.017	HAJGI180
JGI90250	F0YFF1	Putative uncharacterized protein	HC	0.726	0.630	0.804	HAJGI405
JGI275754	Q5ENS1	Chloroplast phosphoribulokinase	HC	0.727	0.604	0.883	HAJGI167
JGI237243			HC	0.727	0.498	0.955	HAJGI191
JGI67108	Q2IA72	Chloroplast light harvesting protein isoform 7	HC	0.728	0.596	0.974	HAJGI299
JGI283485	B8CFP9	Predicted protein	HC	0.728	0.573	0.877	HAJGI071

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI88932	NA	NA	HC	0.732	0.679	0.821	NA
JGI99131	F0Y6E3	Putative uncharacterized protein	HC	0.744	0.589	0.858	HAJGI125
JGI195196	C1N6Z5	Glutathione peroxidase		0.748	0.378	1.255	HAJGI145
JGI406321			HC	0.751	0.696	0.870	HAJGI381
JGI120262	F0Y2W6	Putative uncharacterized protein	HC	0.756	0.613	0.881	HAJGI186
JGI120829	Q2IA75	Chloroplast light harvesting protein isoform 4	HC	0.757	0.461	0.998	HAJGI025
JGI122970	Q01AH9	Heat shock protein 70 / HSP70 (ISS)	HC	0.758	0.664	0.861	HAJGI280
JGI216402	G4ZP25	Putative uncharacterized protein	HC	0.762	0.587	0.975	HAJGI372
JGI273928	G5A9A0	Putative uncharacterized protein	HC	0.765	0.673	0.870	HAJGI259
JGI103637	Q2IA76	Chloroplast light harvesting protein isoform 3	HC	0.766	0.612	0.945	HAJGI035
JGI102701	Q2IA76	Chloroplast light harvesting protein isoform 3	HC	0.767	0.614	0.953	HAJGI209
JGI68029		NA		0.770	0.619	1.012	NA
JGI407305	Q2IA76	Chloroplast light harvesting protein isoform 3	HC	0.774	0.688	0.845	HAJGI008
JGI271331	D0NEK2	Adenosylhomocysteinase	HC	0.774	0.760	0.792	HAJGI136
JGI106019	A8I1X3	Putative uncharacterized protein	HC	0.777	0.536	0.996	HAJGI140
JGI112702		NA	HC	0.780	0.617	0.961	NA
JGI358141	Q84LQ0	Cytosolic glyceraldehyde 3-phosphate dehydrogenase	HC	0.781	0.661	0.950	HAJGI335
JGI230835	Q2IA55	Chloroplast glyceraldehyde-3-phosphate dehydrogenase	HC	0.782	0.706	0.850	HAJGI027
JGI63770	Q5ENR6	Phosphoglycerate kinase		0.785	0.613	1.006	HAJGI366
JGI102901	Q5QD45	Plastid C1 class II fructose bisphosphate aldolase	HC	0.786	0.684	0.855	HAJGI079
JGI111303	A8J506	Argininosuccinate synthase	HC	0.795	0.629	1.013	HAJGI091
JGI88203	A6YAZ8	Putative plastid transketolase	HC	0.799	0.750	0.880	HAJGI397
JGI212512	G5A9A0	Putative uncharacterized protein	HC	0.800	0.712	0.887	HAJGI121
JGI209610	A8IDN1	Chaperonin 10		0.802	0.364	1.745	HAJGI369
JGI199538			HC	0.810	0.717	0.976	HAJGI237

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI62235	B8BY55	S-adenosylmethionine synthase		0.813	0.566	1.164	HAJGI086
JGI43077	D8TKK9	Vacuolar H+ ATPase V0 sector, subunit D	HC	0.813	0.751	0.887	HAJGI048
JGI120592	Q0MYX3	Proteasome subunit alpha type	HC	0.817	0.670	0.994	HAJGI098
JGI61340	Q2IA74	Chloroplast light harvesting protein isoform 5	HC	0.818	0.724	0.934	HAJGI380
JGI87079	B9PNV8	Heat shock protein, putative	HC	0.819	0.723	0.935	HAJGI162
JGI103121				0.819	0.598	1.209	HAJGI174
JGI270399	Q15GC7	Triosephosphate isomerase	HC	0.823	0.725	0.995	HAJGI387
JGI88977	A4RW83	Protein translocase subunit SecA, chloroplastic	HC	0.824	0.697	0.938	HAJGI402
JGI366337	A8J387	Predicted protein	HC	0.828	0.724	0.919	HAJGI266
JGI235106			HC	0.831	0.739	0.965	HAJGI138
JGI97031	D0NEX4	D-3-phosphoglycerate dehydrogenase		0.832	0.722	1.008	HAJGI042
JGI272624	A4S2C6	Predicted protein	HC	0.833	0.703	0.994	HAJGI037
JGI222386	P48414	V-type proton ATPase catalytic subunit A, V-ATPase subunit A	HC	0.833	0.681	0.971	HAJGI081
JGI62038	F0YLJ3	Peptidyl-prolyl cis-trans isomerase		0.833	0.513	1.025	HAJGI040
JGI70830	Q2IA73	Chloroplast light harvesting protein isoform 6		0.834	0.677	1.018	HAJGI197
JGI96943	C1EIH7	Putative uncharacterized protein	HC	0.836	0.726	0.931	HAJGI267
JGI198360	C5KNV3	Molecular chaperone, putative	HC	0.837	0.726	0.931	HAJGI007
JGI267646	B8BY55	S-adenosylmethionine synthase	HC	0.837	0.617	1.145	HAJGI061
JGI227391	B5YMK1	Predicted protein		0.839	0.562	1.314	HAJGI054
JGI265597	G5A9A0	Putative uncharacterized protein	HC	0.839	0.773	0.940	HAJGI205
JGI206438	D0NBZ5	Nucleolar protein NOP5	HC	0.839	0.695	0.961	HAJGI097
JGI97087	B8BY55	S-adenosylmethionine synthase		0.843	0.604	1.178	HAJGI013
JGI102800	G0QQ50	Putative uncharacterized protein	HC	0.843	0.671	0.981	HAJGI019
JGI254140			HC	0.845	0.696	0.996	HAJGI026
JGI100088			HC	0.846	0.751	0.905	HAJGI349
JGI308724	Q84KQ0	Elongation factor-2		0.846	0.665	1.079	HAJGI006
JGI218081	Q2IA28	Chloroplast light harvesting protein isoform 12	HC	0.846	0.787	0.948	HAJGI242

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI350570	NA	NA	HC	0.847	0.709	0.949	NA
JGI110945	B7FUR6	Violaxanthin deepoxidase		0.847	0.732	1.041	HAJGI350
JGI202968	F0YD68	Putative uncharacterized protein	HC	0.848	0.749	0.969	HAJGI291
JGI276712	Q2IA78	Chloroplast light harvesting protein isoform 1		0.848	0.709	1.053	HAJGI338
JGI252355	F0XWG3	Putative plastid light harvesting protein isoform 40	HC	0.851	0.806	0.910	HAJGI030
JGI209460	G4YJL5	Putative uncharacterized protein	HC	0.854	0.577	0.994	HAJGI220
JGI101140	Q2IA28	Chloroplast light harvesting protein isoform 12		0.854	0.714	1.004	HAJGI328
JGI402650	C1EHC0	ATP synthase subunit alpha	HC	0.855	0.754	0.951	HAJGI377
JGI231101	F0YJF2	Putative uncharacterized protein		0.855	0.699	1.000	HAJGI293
JGI99331		NA		0.855	0.673	1.082	NA
JGI207464	Q7XY98	40S ribosomal protein S6		0.857	0.667	1.037	HAJGI345
JGI65080	B8BVQ7	Predicted protein		0.857	0.688	1.021	HAJGI166
JGI234774	D0N196	3-isopropylmalate dehydrogenase		0.860	0.633	1.006	HAJGI243
JGI86588	C1E1M5	Predicted protein		0.861	0.703	1.031	HAJGI129
JGI360985	G4Z7L2	Putative uncharacterized protein	HC	0.865	0.713	0.969	HAJGI157
JGI68128	G4ZAY2	Putative uncharacterized protein		0.867	0.682	1.015	HAJGI355
JGI360891		NA	HC	0.871	0.770	0.970	NA
JGI99724	D8LHC1	Glutathione reductase	HC	0.871	0.730	0.979	HAJGI386
JGI108944	G4YS12	Putative uncharacterized protein		0.871	0.611	1.215	HAJGI016
JGI87166	A8JBC6	Pyruvate dehydrogenase E1 beta subunit		0.871	0.796	1.015	HAJGI347
JGI243741	D8TS49	Component of cytosolic 80S ribosome & 60S large subunit		0.872	0.596	1.353	HAJGI154
JGI272626	F0YAX4	Putative uncharacterized protein	HC	0.874	0.774	0.962	HAJGI038
JGI212481	D7FJQ9	Putative uncharacterized protein		0.874	0.735	1.167	HAJGI192
JGI105039				0.875	0.750	1.015	HAJGI378
JGI267363	D8TRA0	Putative uncharacterized protein		0.876	0.770	1.014	HAJGI308
JGI273597	C5L4E6	Nad dependent epimerase/dehydratase, putative		0.877	0.712	1.152	HAJGI403

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI270342	D8U3I9	Putative uncharacterized protein	HC	0.877	0.752	0.987	HAJGI103
JGI89571	Q5ENR5	Phosphoglycerate kinase	HC	0.877	0.768	0.973	HAJGI147
JGI193927	B8BU34	Predicted protein		0.880	0.618	1.669	HAJGI106
JGI89087	D8TLB0	Putative uncharacterized protein		0.881	0.776	1.028	HAJGI093
JGI266768	B8C830	Thioredoxin	HC	0.882	0.766	0.978	HAJGI200
JGI89311	Q2IA66	Chloroplast light harvesting protein isoform 13		0.883	0.753	1.080	HAJGI305
JGI276141	D0NY29	60S ribosomal protein L3	HC	0.883	0.765	0.968	HAJGI073
JGI365947	E1ZDS1	Putative uncharacterized protein		0.885	0.690	1.117	HAJGI300
JGI108665	A8J768	Ribosomal protein S14		0.886	0.762	1.004	HAJGI375
JGI105420	Q5DK81	Heat shock protein 90	HC	0.887	0.811	0.964	HAJGI092
JGI119469	Q5ENP2	Chloroplast photosystem II 12 kDa extrinsic protein		0.887	0.822	1.006	HAJGI255
JGI68146	Q2I7V4	ATP-dependent Clp protease proteolytic subunit		0.893	0.740	1.111	HAJGI225
JGI95447	B6KPP5	Glycine-rich protein 2, putative		0.894	0.763	1.114	HAJGI241
JGI276075	D8U1L3	Vacuolar ATP synthase subunit E		0.894	0.772	1.047	HAJGI202
JGI102657	A4RTV8	Predicted protein		0.895	0.799	1.055	HAJGI179
JGI96608	C1FG28	Predicted protein		0.895	0.732	1.072	HAJGI315
JGI101101	B7FVA8	Predicted protein		0.897	0.652	1.330	HAJGI130
JGI72905	Q2IA22	Ribosomal protein		0.898	0.763	1.005	HAJGI184
JGI258285	E1Z343	Putative uncharacterized protein		0.900	0.706	1.116	HAJGI370
JGI218536	E1Z823	40S ribosomal protein S18		0.901	0.778	1.014	HAJGI146
JGI207960	E1Z8Z5	Putative uncharacterized protein		0.903	0.721	1.045	HAJGI218
JGI197716	Q2IA11	ATP synthase gamma chain	HC	0.906	0.870	0.964	HAJGI367
JGI90254	D8LP24	Pyruvate kinase		0.906	0.824	1.018	HAJGI400
JGI87271	B8BPW0	Putative uncharacterized protein		0.910	0.788	1.094	HAJGI199
JGI222636		NA		0.910	0.781	1.054	NA
JGI273308	B4ZG44	Putative uncharacterized protein		0.910	0.781	1.054	HAJGI323
JGI61563	Q9AR87	Delta-aminolevulinic acid dehydratase		0.910	0.711	1.186	HAJGI277

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI62864	F0Y053	GTP-binding nuclear protein Ran		0.911	0.792	1.047	HAJGI165
JGI363444	G4ZRS7	Adenosylhomocysteinase		0.911	0.812	1.029	HAJGI056
JGI217140	F0YC48	Putative uncharacterized protein		0.911	0.804	1.149	HAJGI155
JGI235076	C1N337	Predicted protein		0.914	0.797	1.068	HAJGI134
JGI268670	E1ZAW9	Putative uncharacterized protein		0.914	0.735	1.280	HAJGI272
JGI361737		NA		0.914	0.822	1.075	NA
JGI267344	E1Z8N4	Putative uncharacterized protein		0.915	0.835	1.001	HAJGI279
JGI103652				0.915	0.592	1.399	HAJGI083
JGI1112185	D8UEM8	Plastid/chloroplast ribosomal protein L4		0.916	0.776	1.065	HAJGI122
JGI270085	G4YF15	Putative uncharacterized protein		0.916	0.664	1.437	HAJGI392
JGI86826	A4S2V6	Ribosomal protein L18		0.916	0.791	1.080	HAJGI256
JGI240824	A8J9T0	40S ribosomal protein S12		0.917	0.806	1.171	HAJGI139
JGI67049	D8TJL0	Component of cytosolic 80S ribosome and 40S small subunit		0.917	0.704	1.144	HAJGI348
JGI271767	C1ML75	Acetyl-coa carboxylase		0.918	0.864	1.010	HAJGI131
JGI408848	D0NLH6	Alpha-actinin-1, putative		0.918	0.759	1.122	HAJGI110
JGI104032	B8C070	Predicted protein		0.919	0.781	1.073	HAJGI117
JGI202270	D7FUX7	Putative uncharacterized protein		0.919	0.756	1.126	HAJGI339
JGI218259				0.920	0.751	1.261	HAJGI383
JGI57264	D8UHT7	Component of cytosolic 80S ribosome and 40S small subunit		0.922	0.593	1.311	HAJGI238
JGI272213	C1N8B3	Heat shock protein 70kDa		0.924	0.796	1.032	HAJGI404
JGI51375	C1MKE9	Luminal binding heat shock protein 70		0.927	0.758	1.049	HAJGI095
JGI1107785	B8C7K5	Sulfolipid biosynthesis protein		0.928	0.812	1.138	HAJGI229
JGI201657	A4S2V6	Ribosomal protein L18		0.930	0.822	1.067	HAJGI353
JGI269606	C1N183	Predicted protein		0.930	0.741	1.142	HAJGI033
JGI201477	D8TNL8	Prohibitin		0.932	0.720	1.189	HAJGI014
JGI64811		NA		0.933	0.835	1.015	NA

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI276434	D8LJS6	Cystathionine gamma-lyase		0.936	0.727	1.189	HAJGI075
JGI61178	F2WQ20	Malate dehydrogenase		0.936	0.850	1.014	HAJGI330
JGI101346	D0MZZ6	Callose synthase, putative		0.936	0.710	1.111	HAJGI274
JGI96515	C1E1B9	Predicted protein		0.937	0.795	1.111	HAJGI105
JGI412390	C1FE16	ATP synthase subunit beta	HC	0.938	0.835	0.998	HAJGI058
JGI63123	A7ATH4	Ribosomal protein L13, putative		0.939	0.765	1.062	HAJGI250
JGI267180	A4RQS6	Enolase		0.941	0.853	1.040	HAJGI178
JGI271119	D8LHD4	Putative leucine rich repeat protein		0.941	0.816	1.112	HAJGI257
JGI113436	A8IJG5	Predicted protein		0.942	0.789	1.101	HAJGI262
JGI105359	C1E5V7	Predicted protein		0.943	0.607	1.776	HAJGI133
JGI105526	B7FTU0	Predicted protein		0.944	0.683	1.312	HAJGI324
JGI265795	Q8L878	Oxygen evolving enhancer 1		0.944	0.765	1.224	HAJGI298
JGI41410	E1ZQ58	Putative uncharacterized protein		0.947	0.880	1.024	HAJGI087
JGI268510		NA		0.948	0.881	1.071	NA
JGI70481	C1FE16	ATP synthase subunit beta		0.948	0.848	1.007	HAJGI160
JGI109822		NA		0.949	0.846	1.009	NA
JGI64976	Q2IA28	Chloroplast light harvesting protein isoform 12		0.950	0.812	1.136	HAJGI253
JGI96453	B8LEJ5	Predicted protein		0.951	0.615	2.008	HAJGI089
JGI270629	F0XX48	Putative uncharacterized protein		0.953	0.876	1.070	HAJGI150
JGI274190	E1ZHB0	Putative uncharacterized protein		0.954	0.567	1.385	HAJGI286
JGI275203	B7S3L1	Predicted protein		0.954	0.845	1.196	HAJGI215
JGI96420		NA		0.954	0.819	1.131	NA
JGI99777	D0NVC3	26S proteasome non-ATPase regulatory subunit 3		0.954	0.754	1.230	HAJGI213
JGI76562	F0YH74	Putative uncharacterized protein		0.955	0.804	1.170	HAJGI047
JGI284516	D0P2Z9	Adenylate kinase, putative		0.955	0.772	1.071	HAJGI356
JGI97545	B6KVV6	Proliferation-associated protein 2G4, putative	HC	0.957	0.929	0.999	HAJGI017
JGI268371	Q2IA78	Chloroplast light harvesting protein isoform 1		0.957	0.726	1.217	HAJGI211

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_ Min	LC:HC_ Max	Consensus
JGI285186	A4S3B4	Chaperonin 60, mitochondrial		0.958	0.821	1.098	HAJGI169
JGI66519	C1MZG8	Predicted protein		0.959	0.805	1.078	HAJGI195
JGI97618		NA		0.959	0.585	1.462	NA
JGI282672	Q8IDD9	Ubiquitin conjugating enzyme, putative		0.960	0.765	1.175	HAJGI068
JGI201465				0.960	0.728	1.493	HAJGI090
JGI212424	B4ZFX6			0.960	0.826	1.112	HAJGI135
JGI231985	A1YQX3	Nucleic acid binding protein		0.962	0.623	1.842	HAJGI234
JGI57286	G5AFK3	60S ribosomal protein L13		0.962	0.760	1.149	HAJGI080
JGI281076	A6MVX0	ELongation factor Ts, chloroplastic		0.962	0.762	1.142	HAJGI112
JGI98483	Q5UHI6	EF-1 alpha-like protein		0.963	0.824	1.098	HAJGI368
JGI203980	G5AFK3	60S ribosomal protein L13		0.963	0.723	1.214	HAJGI010
JGI1107520	C1EEB4	Elongation factor tu		0.964	0.905	1.041	HAJGI306
JGI69320	D2D4K3	Cell division cycle protein 48		0.965	0.867	1.048	HAJGI053
JGI271735	A7YXU6	RNA recognition motif protein		0.965	0.685	1.340	HAJGI344
JGI66249	Q017S6	Stress-induced protein sti1-like protein (ISS)		0.966	0.786	1.119	HAJGI077
JGI267440	B8BXA1	Enoyl-reductase [NADH]		0.968	0.901	1.063	HAJGI376
JGI88834	D8LKS5	LRR-GTPase of the ROCO family		0.969	0.847	1.166	HAJGI049
JGI100207	Q9SWQ1	Type 1 actin		0.970	0.708	1.361	HAJGI327
JGI207878				0.971	0.712	1.245	HAJGI204
JGI271242	C1EAX2	Predicted protein		0.971	0.789	1.143	HAJGI302
JGI227764				0.972	0.509	1.555	HAJGI052
JGI203470		NA		0.973	0.847	1.144	NA
JGI225524	B8BZI8	Predicted protein		0.973	0.859	1.117	HAJGI332
JGI269305	F0Y0B9	Putative uncharacterized protein		0.973	0.829	1.230	HAJGI318
JGI272248		NA		0.974	0.850	1.133	NA
JGI269306	B7FPZ8	Predicted protein		0.975	0.762	1.218	HAJGI317
JGI217316	B8C1N2	Predicted protein		0.976	0.809	1.133	HAJGI269

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI100546				0.978	0.791	1.153	HAJGI251
JGI274007	G4YEF7	Putative uncharacterized protein		0.978	0.820	1.253	HAJGI363
JGI200165	B7G9A2	Predicted protein		0.980	0.808	1.164	HAJGI022
JGI201257				0.980	0.732	1.337	HAJGI390
JGI205571				0.981	0.791	1.200	HAJGI325
JGI358143	O04839	Glycolytic glyceraldehyde-3-phosphate dehydrogenase		0.981	0.923	1.042	HAJGI394
JGI201317	D7FW04	Magnesium chelatase subunit H, putative chloroplast		0.982	0.894	1.147	HAJGI295
JGI285672	Q50KB1	Protein disulfide-isomerase-like protein EhSep2		0.982	0.598	1.890	HAJGI393
JGI55967	D8UA34	Proteasome subunit alpha type		0.983	0.886	1.142	HAJGI183
JGI271180	F0YH80	Uroporphyrinogen decarboxylase		0.983	0.647	1.631	HAJGI132
JGI268401	F0Y5H5	Putative uncharacterized protein		0.984	0.821	1.174	HAJGI070
JGI250666				0.986	0.894	1.100	HAJGI359
JGI1106545	G5AFG3	Putative uncharacterized protein		0.988	0.838	1.199	HAJGI294
JGI405060	E7D6U3	Glutamine synthtease III		0.991	0.836	1.116	HAJGI127
JGI261410	F0Y3Y1	Putative uncharacterized protein		0.991	0.823	1.289	HAJGI287
JGI230942				0.992	0.833	1.215	HAJGI148
JGI272233				0.992	0.847	1.161	HAJGI159
JGI210921	A7AS38	Polyubiquitin, putative		0.993	0.817	1.277	HAJGI292
JGI215323	G4YSQ9	Putative uncharacterized protein		0.993	0.817	1.277	HAJGI009
JGI408844				0.999	0.809	1.252	HAJGI109
JGI198101				1.000	0.847	1.233	HAJGI248
JGI267345	F0VF26	Putative uncharacterized protein		1.001	0.684	1.319	HAJGI173
JGI68569	C1EI97	Predicted protein		1.001	0.880	1.156	HAJGI313
JGI267273	E6Y2N8	Putative mitochondrial ADP/ATP translocase		1.004	0.864	1.167	HAJGI371
JGI265835		NA		1.006	0.681	1.332	NA
JGI116551	C1E4H7	Putative uncharacterized protein		1.007	0.853	1.134	HAJGI064

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI277072	C1N2P4	Cold-shock DNA binding protein		1.007	0.861	1.282	HAJGI099
JGI268104	Q1WLZ5	14-3-3-like protein-related protein		1.016	0.914	1.125	HAJGI231
JGI118455	E1ZCS2	Putative uncharacterized protein		1.018	0.860	1.120	HAJGI078
JGI100746				1.023	0.833	1.359	HAJGI107
JGI216106	D8TLK9	Putative uncharacterized protein		1.024	0.943	1.073	HAJGI224
JGI116581				1.024	0.771	1.456	HAJGI050
JGI72237	D0NVM9	60S acidic ribosomal protein		1.024	0.913	1.143	HAJGI290
JGI204747		NA		1.025	0.945	1.166	NA
JGI282213				1.032	0.920	1.199	HAJGI163
JGI279225	Q2IA80	Chloroplast O-acetyl-serine lyase		1.040	0.752	1.475	HAJGI309
JGI58448	D8LGN4	Putative uncharacterized protein		1.047	0.849	1.353	HAJGI374
JGI282667	C1E7E6	SNF2 super family		1.047	0.920	1.214	HAJGI020
JGI275346	B8BTG2	60S ACIDIC ribosomal protein P2		1.051	0.777	1.748	HAJGI258
JGI278306	Q1WLZ5	14-3-3-like protein-related protein		1.058	0.970	1.136	HAJGI060
JGI108180	C1N4F5	Predicted protein		1.058	0.871	1.346	HAJGI343
JGI75300	F0Y2L8	Peptidyl-prolyl cis-trans isomerase		1.060	0.981	1.147	HAJGI245
JGI111254	F0Y4J8	Peptidyl-prolyl cis-trans isomerase		1.060	0.981	1.147	HAJGI176
JGI211611	B7FSL9	Predicted protein		1.063	0.747	1.497	HAJGI263
JGI357374	Q2IA80	Chloroplast O-acetyl-serine lyase		1.064	0.766	1.475	HAJGI326
JGI109201	A8J3W3	Predicted protein		1.066	0.978	1.160	HAJGI114
JGI78778	Q5UHI6	EF-1 alpha-like protein		1.066	0.906	1.274	HAJGI100
JGI282626	D8LBF4	Dihydrolipoyl dehydrogenase		1.068	0.912	1.383	HAJGI126
JGI269934				1.069	0.951	1.208	HAJGI206
JGI88448	D7FYP5	Putative uncharacterized protein		1.071	0.804	1.368	HAJGI142
JGI270992				1.075	0.833	1.406	HAJGI230
JGI402323	C5KTL6	HSP90 co-chaperone, putative		1.077	0.894	1.294	HAJGI310
JGI68819	B5Y4R6	Predicted protein		1.078	0.709	1.395	HAJGI171
JGI95574				1.080	0.751	1.534	HAJGI065

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI266878	F0YE54	Putative uncharacterized protein	LC	1.081	1.007	1.205	HAJGI319
JGI65400	D7FRM9	40S ribosomal protein S28		1.083	0.888	1.256	HAJGI187
JGI109593	D0P3G4	Putative uncharacterized protein		1.086	0.872	1.362	HAJGI240
JGI108073	Q50KB1	Protein disulfide-isomerase-like protein EhSep2		1.087	0.781	1.546	HAJGI357
JGI86748	Q2IA86	Chloroplast 50S ribosomal protein L12		1.088	0.835	1.613	HAJGI044
JGI71010	B8C6L5	Putative uncharacterized protein		1.089	0.977	1.183	HAJGI084
JGI72427	D0NXH1	Ubiquitin, putative		1.091	0.847	1.560	HAJGI364
JGI195071				1.093	0.880	1.318	HAJGI190
JGI122468				1.095	0.857	1.404	HAJGI223
JGI278991	C1EBQ0	Predicted protein		1.098	0.908	1.379	HAJGI001
JGI214274		NA		1.099	0.711	1.656	NA
JGI267274	E1Z6A8	Putative uncharacterized protein		1.113	0.604	1.894	HAJGI281
JGI97560	C1MK47	Predicted protein		1.116	0.864	1.383	HAJGI398
JGI210563	Q49JJ1	Histone H2A		1.119	0.885	1.422	HAJGI265
JGI284352	D7FSF0	Putative uncharacterized protein		1.122	0.847	1.776	HAJGI244
JGI257539	Q49JJ1	Histone H2A		1.123	0.891	1.422	HAJGI252
JGI69183	D7FRA4	Plastid ribosomal protein S1		1.123	0.967	1.370	HAJGI254
JGI105146	C1E0B7	Elongation factor tu gtp-binding domain protein 2	LC	1.130	1.049	1.209	HAJGI104
JGI64083	F0YST3	Putative uncharacterized protein	LC	1.133	1.027	1.267	HAJGI289
JGI109561	C5L3N3	3-hydroxyacyl-CoA dehyrogenase, putative		1.135	0.995	1.305	HAJGI354
JGI231324	D0NEH7	26S protease regulatory subunit 7	LC	1.137	1.002	1.312	HAJGI235
JGI110032	Q2IA85	Chloroplast 50S ribosomal protein L17		1.140	0.640	2.278	HAJGI059
JGI365196	B6DX96	Putative chloroplast 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	LC	1.141	1.022	1.259	HAJGI268
JGI116827	B7FZI9	Predicted protein		1.147	0.770	1.957	HAJGI032
JGI242836	C1E0T7	Predicted protein		1.147	0.911	1.439	HAJGI226
JGI274290	C1EGC2	Predicted protein		1.148	0.868	1.866	HAJGI214

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI65898	E1ZDT5	Proteasome subunit beta type		1.148	0.861	1.357	HAJGI311
JGI230687	Q1MWL3	Fucoxanthin chlorophyll a/c binding protein		1.157	0.974	1.321	HAJGI212
JGI95565	Q49JJ2	Histone H2A	LC	1.161	1.017	1.368	HAJGI144
JGI71008		NA	LC	1.161	1.007	1.372	NA
JGI279740	C1FFE2	Predicted protein	LC	1.161	1.007	1.372	HAJGI074
JGI267796	Q1WLZ2	40S ribosomal protein S3	LC	1.165	1.082	1.264	HAJGI296
JGI271880		NA		1.167	0.935	1.706	NA
JGI215648				1.168	0.754	1.789	HAJGI115
JGI256830			LC	1.168	1.100	1.280	HAJGI208
JGI271455	F0YFV2	Putative uncharacterized protein	LC	1.170	1.034	1.441	HAJGI185
JGI65368				1.173	0.769	1.712	HAJGI072
JGI105819	A8IZW5	RAN binding protein, RANBP1		1.175	0.906	1.645	HAJGI346
JGI270764	F0Y620	Putative uncharacterized protein		1.179	0.941	1.379	HAJGI210
JGI108396	D8TQ04	ER lumen protein retaining receptor		1.185	0.867	1.634	HAJGI137
JGI220968				1.199	0.805	1.862	HAJGI312
JGI96192	Q013K3	Histone H2B	LC	1.207	1.002	1.403	HAJGI401
JGI205593	E1Z2Y1	Putative uncharacterized protein		1.222	0.848	1.730	HAJGI384
JGI198481				1.227	0.815	2.028	HAJGI342
JGI94767				1.239	0.637	2.433	HAJGI124
JGI114035	D0NWQ9	Protein kinase, putative		1.240	0.845	2.024	HAJGI232
JGI76377	B7G0V7	Predicted protein	LC	1.244	1.030	1.389	HAJGI143
JGI364974	D7FPP1	Trifunctional enzyme subunit alpha		1.246	0.987	1.475	HAJGI284
JGI372283		NA		1.267	0.940	1.486	NA
JGI112664		NA		1.277	0.940	1.466	NA
JGI201707	A4RS62	Histone H4	LC	1.277	1.094	1.524	HAJGI379
JGI255477	A7AT84	Histone H3, putative	LC	1.281	1.063	1.502	HAJGI111
JGI369217				1.303	0.969	1.585	HAJGI322
JGI72235	C1MVY5	Histone H2A	LC	1.305	1.156	1.477	HAJGI082

240

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
JGI272682	F0YAR4	Putative uncharacterized protein		1.327	0.902	2.304	HAJGI193
JGI270658	B8BTN5	Putative uncharacterized protein		1.333	0.808	2.433	HAJGI316
JGI203946	B7G4V0	Predicted protein		1.456	0.980	2.315	HAJGI321
JGI349877				1.475	0.966	2.132	HAJGI116
JGI277202	Q5G920	Alpha-tubulin	LC	1.476	1.285	1.821	HAJGI156
JGI374572			LC	1.550	1.147	1.916	HAJGI158
JGI87697	F0Y9S1	Putative uncharacterized protein	LC	1.577	1.383	1.764	HAJGI307
JGI44488	A8JDV9	ATP synthase gamma chain		1.700	0.776	6.250	HAJGI271
JGI233632	B8CAB4	Predicted protein		1.795	0.588	6.250	HAJGI360
TAXDB13	Q4G3F4	Ribulose bisphosphate carboxylase large chain (RuBisCO large subunit) (EC 4.1.1.39)	HC	0.598	0.544	0.660	NA
TAXDB73	A3FQF5	Uncharacterized protein	HC	0.647	0.528	0.799	TAXDB73
TAXDB68	Q4G3F3	Ribulose 1,5-bisphosphate carboxylase/oxygenase small subunit	HC	0.673	0.416	1.290	TAXDB68
TAXDB14	Q4G396	Photosystem II CP43 reaction center protein (PSII 43 kDa protein) (Protein CP-43)	HC	0.681	0.581	0.801	TAXDB14
TAXDB22	B8C4K2	Uncharacterized protein	HC	0.691	0.443	0.950	TAXDB22
TAXDB12	A8IVJ3	Phospholipid-transporting ATPase (EC 3.6.3.1) (Fragment)		0.707	0.488	1.189	TAXDB12
TAXDB75	Q5ENR8	Phosphoglycerate kinase (EC 2.7.2.3)	HC	0.720	0.647	0.818	TAXDB75
TAXDB81	Q40612	ATP synthase subunit beta, chloroplastic (EC 3.6.3.14) (ATP synthase F1 sector subunit beta) (F-ATPase subunit beta)	HC	0.735	0.622	0.864	NA
TAXDB85	D8TRA2	ATP synthase subunit beta (EC 3.6.3.14)	HC	0.760	0.672	0.836	TAXDB85
TAXDB61	Q4G3C8	ATP synthase subunit beta, chloroplastic (EC 3.6.3.14) (ATP synthase F1 sector subunit beta) (F-ATPase subunit beta)	HC	0.763	0.659	0.892	NA
TAXDB57	C6KIJ6	ATP synthase subunit alpha, chloroplastic (EC 3.6.3.14) (ATP synthase F1 sector subunit alpha)		0.774	0.633	1.034	TAXDB57

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
		(F-ATPase subunit alpha)					
TAXDB31	Q9TAH9	ATP synthase subunit alpha	HC	0.774	0.633	1.034	TAXDB31
TAXDB41	Q4G3C9	ATP synthase epsilon chain, chloroplastic (ATP synthase F1 sector epsilon subunit) (F-ATPase epsilon subunit)	HC	0.797	0.687	0.949	NA
TAXDB05	Q4G397	ATP synthase subunit alpha, chloroplastic (EC 3.6.3.14) (ATP synthase F1 sector subunit alpha) (F-ATPase subunit alpha)		0.798	0.720	0.909	NA
TAXDB01	C0JWI5	Photosystem II D2 protein (PSII D2 protein) (EC 1.10.3.9) Photosystem Q(A) protein		0.805	0.629	1.011	TAXDB01
TAXDB88	Q2IA54	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	HC	0.808	0.758	0.878	TAXDB88
TAXDB56	F0YKF4	S-adenosylmethionine synthase (EC 2.5.1.6)		0.810	0.497	1.239	TAXDB56
TAXDB67	F0Y394	Phosphoglycerate kinase (EC 2.7.2.3)		0.820	0.657	1.024	TAXDB67
TAXNA	C5K6Q8	Heat shock protein, putative	HC	0.841	0.662	0.987	NA
TAXDB50	Q4G398	ATP synthase subunit delta, chloroplastic		0.848	0.654	1.103	NA
TAXDB66	A1E8Z5	ATP synthase subunit beta (EC 3.6.3.14) (Fragment)		0.853	0.737	1.014	TAXDB66
TAXDB90	G0QRF6	ATP synthase subunit alpha	HC	0.865	0.741	0.998	TAXDB90
TAXDB48	Q49JJ1	Histone H2A	HC	0.867	0.789	0.927	TAXDB48
TAXDB42	Q4G347	DNA-directed RNA polymerase subunit alpha		0.876	0.656	1.122	NA
TAXDB78	Q4G368	Cytochrome c-550 (Cytochrome c550)		0.877	0.783	1.008	NA
TAXDB84	Q309Z7	ATP synthase subunit beta (EC 3.6.3.14)		0.879	0.773	1.034	TAXDB84
TAXDB25	Q4G3C5	Photosystem II CP47 reaction center protein		0.881	0.766	1.073	TAXDB25
TAXDB23	Q4G399	ATP synthase subunit b, chloroplastic		0.888	0.746	1.034	NA
TAXDB44	A9BL97	Hsp90		0.893	0.782	1.085	TAXDB44
TAXDB89	F0YFK1	ATP synthase subunit beta (EC 3.6.3.14)		0.911	0.795	1.031	TAXDB89
TAXDB45	Q4G3A0	ATP synthase subunit b', chloroplastic		0.914	0.819	1.012	NA
TAXDB08	B8C637	HSP90 family member		0.917	0.808	1.099	TAXDB08

242

Appendix B

Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
TAXDB11	Q6TP37	Heat shock protein 90 (Fragment)		0.917	0.808	1.099	TAXDB11
TAXDB20	Q5DK79	Heat shock protein 90 (Fragment)		0.923	0.867	1.041	TAXDB20
TAXDB77	D7FXG1	ATP synthase subunit beta (EC 3.6.3.14)	HC	0.924	0.876	0.978	TAXDB77
TAXDB29	P81643	Glutamine synthetase 2 isozyme (Chloroplast GS2)		0.937	0.829	1.048	NA
TAXDB64	Q5DK81	Heat shock protein 90 (Fragment)		0.938	0.843	1.065	TAXDB64
TAXDB33	D0NGK0	Heat shock protein 70		0.939	0.881	1.045	TAXDB33
TAXDB76	G4YW40	Putative uncharacterized protein		0.939	0.881	1.045	TAXDB76
TAXDB10	Q4G342	Elongation factor Tu, chloroplastic		0.942	0.803	1.157	NA
TAXNA	F0XYY5	Putative uncharacterized protein		0.944	0.812	1.093	NA
TAXDB27	Q5UHI6	EF-1 alpha-like protein		0.952	0.841	1.070	TAXDB27
TAXDB80	Q5ENS8	Ferredoxin--NADP reductase		0.955	0.869	1.104	TAXDB80
TAXDB79	Q5CTI3	Putative biotin-(Acetyl-CoA carboxylase) ligase		0.957	0.575	1.456	TAXDB79
TAXDB87	Q4G374	Ribosomal protein S6		0.969	0.840	1.070	TAXDB87
TAXDB58	Q4G388	Mg-protoporphyrin IX chelatase		0.973	0.918	1.041	TAXDB58
TAXDB37	Q4G3D5	60 kDa chaperonin, chloroplastic		0.975	0.849	1.075	NA
TAXNA	A8HX38	Eukaryotic translation elongation factor 1 alpha 1		0.977	0.818	1.121	NA
TAXDB82	A8IFL6	Flagellar associated protein		0.977	0.761	1.188	TAXDB82
TAXDB38	Q4G3D7	Cytochrome f		0.993	0.833	1.153	NA
TAXDB39	F0XX48	Putative uncharacterized protein		0.994	0.917	1.096	TAXDB39
TAXDB04	C1NAA3	Predicted protein		0.998	0.862	1.133	TAXDB04
TAXDB59	B4ZFX6	Glutamate semialdehyde synthase		0.999	0.880	1.119	TAXDB59
TAXDB07	Q4G3D0	Clp protease ATP binding subunit		1.001	0.876	1.126	TAXDB07
TAXDB74	Q4G366	Chaperone protein dnaK (HSP70)		1.005	0.901	1.135	NA
TAXNA	P30722	Chaperone protein dnaK (HSP70)		1.007	0.890	1.148	NA
TAXDB55	Q5ENU6	ATP synthase subunit gamma		1.012	0.850	1.253	TAXDB55
TAXDB32	C1E1B9	Uncharacterized protein		1.013	0.894	1.247	TAXDB32
TAXDB43	F0YRE3	40S ribosomal protein S14		1.015	0.885	1.181	TAXDB43

Appendix B
Table 1

DB ID	UniProt	Description	Expression direction significant	LC:HC	LC:HC_Min	LC:HC_Max	Consensus
TAXDB51	Q9SWP8	Type 4 actin		1.017	0.918	1.189	TAXDB51
TAXDB06	A6XDC6	Actin		1.037	0.914	1.176	TAXDB06
TAXNA	F0Y169	Putative uncharacterized protein		1.049	0.814	1.429	NA
TAXDB83	Q2IA80	Chloroplast O-acetyl-serine lyase		1.086	0.806	1.577	TAXDB83
TAXDB46	Q4G354	30S ribosomal protein S8, chloroplastic		1.097	0.727	1.570	NA
TAXDB52	B4ZFS2	Actin (Fragment)		1.101	0.984	1.205	TAXDB52
TAXDB34	D8UMU9	Putative uncharacterized protein		1.105	0.995	1.252	TAXDB34
TAXNA	A6YGC8	50S ribosomal protein L14, chloroplastic		1.105	0.995	1.252	NA
TAXDB02	C1FFK0	Peptidyl-prolyl cis-trans isomerase (	LC	1.107	1.010	1.215	TAXDB02
TAXDB54	Q4G343	30S ribosomal protein S7, chloroplastic		1.135	0.974	1.339	NA
TAXDB26	Q4G379	CbbX (Putative rubisco expression protein)		1.147	0.878	1.757	TAXDB26
TAXDB65	F0Y8U4	Putative uncharacterized protein (Fragment)	LC	1.213	1.052	1.397	TAXDB65
TAXDB19	A4S1C9	Histone H2B	LC	1.333	1.076	1.631	TAXDB19
TAXDB71	D0MWJ7	Histone H2A	LC	1.362	1.181	1.524	TAXDB71
TAXDB28	C1E6T8	Histone H4	LC	1.373	1.280	1.580	TAXDB28
TAXDB53	Q5G917	Alpha-tubulin (Fragment)	LC	1.504	1.242	1.709	TAXDB53
TAXDB72	F0XW25	Tubulin alpha-2 chain	LC	2.388	1.484	3.448	TAXDB72

Appendix B Table - 2: Gene ontology annotations of all proteins identified in *E. huxleyi* showing highly calcifying cells of the late G1-phase.

Appendix B Table -2 Gene Ontology – biological process [GO IDs] (non significantly higher expressed)	Number of Peptides
Translation [GO:0006412]	29
Photosynthesis; light harvesting [GO:0009765]; protein-chromophore linkage [GO:0018298]	7
Protein folding [GO:0006457]	6
ATP hydrolysis coupled proton transport [GO:0015991]; ATP synthesis coupled proton transport [GO:0015986]	6
Protein folding [GO:0006457]; response to stress [GO:0006950]	6
Glycolytic process [GO:0006096]	4
one-carbon metabolic process [GO:0006730]; S-adenosylmethionine biosynthetic process [GO:0006556]	4
ATP synthesis coupled proton transport [GO:0015986]	4
Protein refolding [GO:0042026]	3
Ubiquitin-dependent protein catabolic process [GO:0006511]	2
L-serine biosynthetic process [GO:0006564]	2
Pyridoxal phosphate biosynthetic process [GO:0042823]; vitamin B6 biosynthetic process [GO:0042819]	2
tRNA aminoacylation for protein translation [GO:0006418]	2
Leucine biosynthetic process [GO:0009098]	2
Photosynthesis [GO:0015979]; photosystem II stabilization [GO:0042549]	2
Tetrapyrrole biosynthetic process [GO:0033014]	2
Regulation of transcription; DNA-templated [GO:0006355]	2
Cell redox homeostasis [GO:0045454]; protein folding [GO:0006457]	2
Fatty acid metabolic process [GO:0006631]	2
Photosynthetic electron transport in photosystem II [GO:0009772]; protein-chromophore linkage [GO:0018298]	2
ATP synthesis coupled proton transport [GO:0015986]; methylglyoxal catabolic process to D-lactate [GO:0019243]; photorespiration [GO:0009853]; proteasome core complex assembly [GO:0080129]; response to misfolded protein [GO:0051788]; response to misfolded protein [GO:0051788]; ubiquitin-dependent protein catabolic process [GO:0006511];	1
Cortical microtubule organization [GO:0043622]; gluconeogenesis [GO:0006094]; glycolytic process [GO:0006096]; protein folding [GO:0006457]; response to cadmium ion [GO:0046686]; response to cadmium ion [GO:0046686]; response to salt stress [GO:0009651];	1
Intracellular protein transport [GO:0006886]; vesicle-mediated transport [GO:0016192]	1
Glycyl-tRNA aminoacylation [GO:0006426]	1
Response to oxidative stress [GO:0006979]	1
One-carbon metabolic process [GO:0006730]	1
Glucose metabolic process [GO:0006006]	1
Arginine biosynthetic process [GO:0006526]; nucleotide biosynthetic process [GO:0009165]	1
ATP hydrolysis coupled proton transport [GO:0015991]	1
Phospholipid metabolic process [GO:0006644]	1
Branched-chain amino acid metabolic process [GO:0009081]	1

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

245

Appendix B Table -2 Gene Ontology – biological process [GO IDs] (non significantly higher expressed)	Number of Peptides
GMP biosynthetic process [GO:0006177]	1
Cytoskeleton organization [GO:0007010]	1
Acetyl-CoA biosynthetic process from pyruvate [GO:0006086]	1
Folic acid-containing compound biosynthetic process [GO:0009396]	1
rRNA processing [GO:0006364]	1
Tricarboxylic acid cycle [GO:0006099]	1
Formation of translation preinitiation complex [GO:0001731]; regulation of translational initiation [GO:0006446]	1
Porphyrin-containing compound biosynthetic process [GO:0006779]	1
Intracellular protein transport [GO:0006886]; nucleocytoplasmic transport [GO:0006913]	1
Proteolysis [GO:0006508]	1
Fatty acid biosynthetic process [GO:0006633]	1
carbohydrate metabolic process [GO:0005975]; malate metabolic process [GO:0006108]	1
(1->3)-beta-D-glucan biosynthetic process [GO:0006075]	1
Regulation of protein catabolic process [GO:0042176]	1
Protein ubiquitination [GO:0016567]; ubiquitin-dependent protein catabolic process [GO:0006511]	1
Cell division [GO:0051301]	1
Small GTPase mediated signal transduction [GO:0007264]	1
Pteridine-containing compound metabolic process [GO:0042558]	1
'de novo' pyrimidine nucleobase biosynthetic process [GO:0006207]	1
Chlorophyll biosynthetic process [GO:0015995]	1
Protoporphyrinogen IX biosynthetic process [GO:0006782]	1
Nitrogen compound metabolic process [GO:0006807]	1
Transmembrane transport [GO:0055085]	1
Ribosome biogenesis [GO:0042254]	1
Translational elongation [GO:0006414]	1
Protein catabolic process [GO:0030163]	1
Cell redox homeostasis [GO:0045454]	1
Actin filament depolymerization [GO:0030042]	1
Threonyl-tRNA aminoacylation [GO:0006435]	1
Proteolysis involved in cellular protein catabolic process [GO:0051603]	1
Intracellular transport [GO:0046907]	1
Protein retention in ER lumen [GO:0006621]; protein transport [GO:0015031]	1
Transcription; DNA-templated [GO:0006351]	1
Cytochrome c-heme linkage [GO:0018063]; photosynthesis, light reaction [GO:0019684]	1
Nitrogen fixation [GO:0009399]	1
Cellular protein modification process [GO:0006464]	1
Chlorophyll biosynthetic process [GO:0015995]; photosynthesis [GO:0015979]	1
Oxidation-reduction process [GO:0055114]; photosynthesis [GO:0015979]	1

Appendix B Table - 3: Gene ontology annotations of all proteins identified in *E. huxleyi* showing low calcification rates in the early G1-phase

Appendix B Table 3 Gene Ontology – cellular component [GO IDs] (non significantly higher expressed)	Number of Peptides
Ribosome [GO:0005840]	21
Cytoplasm [GO:0005737]	14
Chloroplast [GO:0009507]	11
Chloroplast [GO:0009507]; integral component of membrane [GO:0016021]; photosystem II [GO:0009523]	7
Intracellular [GO:0005622]	5
Large ribosomal subunit [GO:0015934]	5
Proton-transporting ATP synthase complex, catalytic core F(1) [GO:0045261]	4
Chloroplast thylakoid membrane [GO:0009535]; proton-transporting ATP synthase complex, catalytic core F(1) [GO:0045261]	3
Cytosol [GO:0005829]	3
Chloroplast [GO:0009507]; ribosome [GO:0005840]	2
Chloroplast thylakoid membrane [GO:0009535]; integral component of membrane [GO:0016021]; proton-transporting ATP synthase complex, coupling factor F(o) [GO:0045263]	2
Endoplasmic reticulum [GO:0005783]; endoplasmic reticulum lumen [GO:0005788]	2
Nucleosome [GO:0000786]; nucleus [GO:0005634]	2
Nucleus [GO:0005634]	2
Small ribosomal subunit [GO:0015935]	2
1,3-beta-D-glucan synthase complex [GO:0000148]	1
3-isopropylmalate dehydratase complex [GO:0009316]	1
Actin cytoskeleton [GO:0015629]	1
Cell [GO:0005623]	1
Cell wall [GO:0005618]; chloroplast [GO:0009507]; mitochondrial proton-transporting ATP synthase complex [GO:0005753]; nucleolus [GO:0005730]; proton-transporting ATP synthase complex, catalytic core F(1) [GO:0045261]	1
Chloroplast [GO:0009507]; integral component of membrane [GO:0016021]; photosystem II [GO:0009523]; thylakoid membrane [GO:0042651]	1
Chloroplast [GO:0009507]; large ribosomal subunit [GO:0015934]	1
Chloroplast [GO:0009507]; mitochondrion [GO:0005739]	1
Chloroplast [GO:0009507]; small ribosomal subunit [GO:0015935]	1
Chloroplast envelope [GO:0009941]; mitochondrion [GO:0005739]	1
Chloroplast stroma [GO:0009570]	1
Chloroplast thylakoid membrane [GO:0009535]; integral component of membrane [GO:0016021]; photosystem II [GO:0009523]	1
Chloroplast thylakoid membrane [GO:0009535]; integral component of thylakoid membrane [GO:0031361]	1
Chloroplast thylakoid membrane [GO:0009535]; photosystem II [GO:0009523]	1
Clathrin coat of coated pit [GO:0030132]; clathrin coat of trans-Golgi network vesicle [GO:0030130]	1
Cullin-RING ubiquitin ligase complex [GO:0031461]	1

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

247

Appendix B Table 3 Gene Ontology – cellular component [GO IDs] (non significantly higher expressed)	Number of Peptides
Cytoplasm [GO:0005737]; nucleus [GO:0005634]; proteasome core complex [GO:0005839]	1
Cytoplasm [GO:0005737]; nucleus [GO:0005634]; proteasome core complex, alpha-subunit complex [GO:0019773]	1
Cytosol [GO:0005829]; prefoldin complex [GO:0016272]	1
Endoplasmic reticulum membrane [GO:0005789]; integral component of membrane [GO:0016021]	1
Eukaryotic 43S preinitiation complex [GO:0016282]; eukaryotic 48S preinitiation complex [GO:0033290]; eukaryotic translation initiation factor 3 complex [GO:0005852]	1
Extrinsic component of membrane [GO:0019898]; integral component of membrane [GO:0016021]; photosystem II oxygen evolving complex [GO:0009654]	1
Extrinsic component of membrane [GO:0019898]; photosystem II oxygen evolving complex [GO:0009654]	1
Integral component of membrane [GO:0016021]	1
Integral component of membrane [GO:0016021]; mitochondrial inner membrane [GO:0005743]	1
Membrane [GO:0016020]	1
Mitochondrion [GO:0005739]; proton-transporting ATP synthase complex, catalytic core F(1) [GO:0045261]	1
Motile cilium [GO:0031514]	1
Phosphopyruvate hydratase complex [GO:0000015]	1
Proteasome complex [GO:0000502]	1
Proton-transporting two-sector ATPase complex, catalytic domain [GO:0033178]	1

Appendix B Table - 4: Peptides matching the most recent *E. huxleyi* genome (Read et al., 2013)

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
100546	0	1667	100
102999	2.00E-129	365	100
105039	0	598	100
106019	4.00E-124	350	100
107385	4.00E-134	377	100
107385	4.00E-134	377	100
107385	5.00E-108	310	87
108180	0	572	100
109848	0	2969	100
109848	0	2969	100
113427	0	2357	100
114964	4.00E-80	234	100
116827	0	607	66
118474	0	4175	100
124604	4.00E-97	279	100
196702	4.00E-139	391	99
198024	0	2243	100
198360	0	764	100
199538	0	925	100
200700	0	535	100
200700	0	535	100
206438	0	713	100
216021	0	841	100
216021	0	835	100
220968	7.00E-69	205	100
227764	3.00E-152	425	100
237243	4.00E-95	274	100
243741	3.00E-163	454	100
245546	1.00E-108	310	100
246130	3.00E-69	206	100
246253	0	869	82
251065	2.00E-71	211	100
251065	2.00E-71	211	100
257439	2.00E-91	265	100
258285	3.00E-108	308	100
312577	0	1599	100
312577	0	1599	100
317426	0	2635	100
317426	0	2635	100
317426	0	2629	100
317615	5.00E-153	427	100
317615	7.00E-153	427	100
317615	1.00E-148	416	100
318009	0	701	100
318009	0	701	100
349981	0	558	100

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

249

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
351077	0	562	100
351077	0	505	91
351077	0	505	91
351730	2.00E-134	378	100
351730	2.00E-134	378	100
351730	2.00E-127	360	96
352634	0	532	100
352634	0	507	96
354741	8.00E-173	480	100
354741	8.00E-173	480	100
354741	9.00E-173	480	100
354741	9.00E-173	480	100
356661	0	941	100
356661	0	941	100
361737	0	608	75
361737	0	793	100
361737	0	793	100
362112	0	900	100
362112	0	900	100
362534	3.00E-180	504	68
362722	1.00E-63	191	100
363772	1.00E-151	424	100
363772	1.00E-151	424	100
363772	3.00E-144	405	98
364889	0	515	94
365175	0	833	100
365175	0	833	100
365175	0	833	100
365947	3.00E-108	309	100
369392	3.00E-70	209	100
369392	3.00E-70	209	100
369392	3.00E-70	209	100
369425	0	803	85
369425	0	944	100
369425	0	944	100
369425	0	944	100
372394	0	671	91
372663	0	533	100
373699	3.00E-66	197	100
374282	7.00E-82	239	100
374282	7.00E-82	239	100
374282	7.00E-82	239	100
414178	2.00E-143	402	100
414178	2.00E-143	402	100
414772	8.00E-153	427	100
414772	7.00E-150	420	100
415031	3.00E-109	312	100
415405	0	913	100
415433	4.00E-122	346	91

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
415729	0	516	100
415729	0	516	100
415729	0	516	100
415940	2.00E-151	423	100
415940	2.00E-151	423	100
416424	0	918	100
416424	0	918	100
416950	0	1315	100
416950	0	1315	100
416955	1.00E-95	275	100
417078	0	1602	100
417078	0	1602	100
417078	0	1602	100
417351	5.00E-106	314	36
417468	0	539	31
417537	0	525	100
417537	0	525	100
417537	0	525	100
417816	1.00E-88	257	97
418142	9.00E-100	286	100
418229	0	779	63
418599	0	548	47
418599	0	555	47
419966	2.00E-102	294	100
419966	2.00E-102	294	100
419966	4.00E-92	267	90
420527	0	1715	100
420527	0	1672	98
421023	0	2216	95
421349	0	1038	98
421356	0	1711	99
421356	0	1660	97
421960	0	1581	100
422643	0	560	100
422643	0	560	100
422643	3.00E-104	303	53
423073	0	2655	100
423073	0	2655	100
423357	4.00E-166	488	27
423968	0	771	100
423968	0	771	100
423968	4.00E-166	468	60
424731	0	739	100
424731	0	739	100
424731	6.00E-135	387	51
426056	0	525	86
426056	0	614	100
426056	0	614	100

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

251

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
426711	0	846	100
426711	0	846	100
427020	0	520	56
427625	0	1307	100
427625	0	719	56
427625	0	1266	98
427625	0	1132	90
428226	0	1041	54
429671	7.00E-142	397	100
429767	0	624	86
430738	2.00E-174	487	74
431703	6.00E-150	419	100
431703	6.00E-150	419	100
431703	3.00E-119	339	80
431800	8.00E-103	295	92
431974	0	583	81
431974	0	723	100
431974	0	723	100
432120	8.00E-77	226	100
432120	1.00E-76	226	100
432120	1.00E-76	226	100
432385	0	733	100
432385	0	733	100
432606	0	978	100
432636	0	521	69
432636	0	759	100
432636	0	759	100
432636	0	759	100
432765	2.00E-165	459	100
432765	3.00E-165	459	100
432765	3.00E-165	459	100
432765	3.00E-165	459	100
432823	0	748	100
432823	0	748	100
432864	0	552	69
432864	0	795	100
432864	0	795	100
432864	0	784	99
432900	5.00E-107	308	73
433142	0	1037	99
433142	0	1050	100
433142	0	1050	100
433142	0	1050	100
433309	0	744	100
433309	0	750	100
433309	0	750	100
433456	4.00E-61	187	66
433847	0	562	100
433847	0	562	100

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
433847	4.00E-140	396	69
433854	2.00E-132	373	100
434108	2.00E-149	418	100
434108	2.00E-149	418	100
434108	2.00E-124	353	83
434239	0	2522	100
434239	0	2522	100
434239	0	1522	90
434513	0	538	61
434648	0	517	83
434648	0	625	100
434648	0	625	100
434751	0	740	100
434751	0	719	98
435222	2.00E-175	487	100
435222	2.00E-175	487	100
435222	3.00E-161	449	91
435472	0	521	100
435472	0	521	100
435472	0	521	100
436026	0	565	39
436026	0	1337	100
436026	0	1337	100
436026	0	1337	100
436026	1.00E-10	54	3
436143	0	548	100
436409	3.00E-167	465	100
436495	0	573	72
436495	0	794	100
436495	0	794	100
436528	0	1214	100
436615	0	504	100
436903	6.00E-179	496	100
436903	6.00E-179	496	100
436903	2.00E-167	466	95
437052	0	920	100
437052	0	920	100
437052	0	879	97
437374	6.00E-141	395	100
437374	7.00E-141	395	100
437557	4.00E-137	388	88
437959	0	786	57
437959	0	1395	100
437959	0	1395	100
437959	0	1395	100
438040	0	928	100
438040	0	928	100
438194	0	1037	100

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

253

Appendix B Table 4 E. hux JGI gene ID	E Value	Score	Hit Sim. in %
438194	0	1037	100
438194	0	858	88
438265	0	541	100
438462	0	541	100
438462	0	541	100
438462	0	541	100
438462	1.00E-179	498	94
439359	0	600	38
439359	0	1582	100
439359	0	1582	100
439359	0	1573	100
439359	8.00E-163	474	30
439486	5.00E-102	292	100
439486	5.00E-102	292	100
439486	5.00E-102	292	100
439486	5.00E-88	256	86
439538	0	1209	100
439538	2.00E-121	359	29
439544	0	1544	100
439544	2.00E-180	522	34
439617	0	816	100
439617	0	816	100
439617	0	701	88
439617	8.00E-174	489	60
439725	2.00E-174	483	100
439725	2.00E-174	483	100
439906	0	1098	82
439906	0	631	45
440119	0	729	74
440119	0	986	100
440119	0	986	100
440242	0	1400	100
440242	0	1400	100
440362	0	536	96
440362	0	560	100
440362	0	560	100
440362	0	560	100
440893	1.00E-89	281	17
440922	0	612	100
440922	0	612	100
440922	1.00E-111	323	52
441453	4.00E-145	406	100
441818	3.00E-122	350	83
441919	0	717	100
441919	0	717	100
441919	0	675	93
441942	0	756	100
442074	0	986	100
442074	0	986	100

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
442074	0	986	100
442074	2.00E-168	478	48
442092	0	1285	100
442092	0	1285	100
442209	0	1371	100
442209	0	1371	100
442340	0	1291	100
442340	0	1291	100
442340	4.00E-169	486	37
442429	9.00E-157	437	100
442429	1.00E-151	424	98
442429	2.00E-151	424	98
442547	0	514	54
442547	0	953	100
442547	0	953	100
442547	0	946	100
442739	2.00E-168	475	57
443305	0	529	100
443320	0	890	100
443778	0	1395	100
443778	0	1395	100
443778	0	1400	100
443778	1.00E-176	507	37
443828	2.00E-109	312	100
443828	2.00E-109	312	100
443828	2.00E-109	312	100
443878	3.00E-160	446	100
443878	3.00E-160	446	100
443878	3.00E-160	446	100
443878	3.00E-160	446	100
444144	5.00E-99	301	23
444382	0	976	100
444382	0	976	100
444382	2.00E-174	493	50
444779	8.00E-154	429	100
444876	2.00E-174	484	100
444895	0	954	100
444895	0	954	100
444895	0	928	99
444895	3.00E-138	399	42
444955	1.00E-173	482	100
444955	1.00E-171	477	99
445002	4.00E-101	294	68
445073	6.00E-85	247	98
445255	5.00E-83	243	88
445439	0	891	62
445439	0	891	62
445566	6.00E-129	363	97

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

255

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
445703	0	1545	100
445703	0	1545	100
445703	0	1404	93
445703	1.00E-167	487	32
446067	8.00E-175	484	99
446067	9.00E-167	464	98
446099	6.00E-175	485	100
446206	0	522	100
446206	4.00E-178	495	97
446237	1.00E-131	371	94
446255	0	511	61
446310	1.00E-154	431	100
446406	1.00E-166	464	100
446406	2.00E-111	321	68
446942	0	586	63
447219	6.00E-160	445	100
447219	7.00E-160	445	100
447219	7.00E-160	445	100
447219	3.00E-159	443	100
447372	2.00E-136	394	64
447695	0	566	60
448062	0	807	100
448062	0	807	100
448062	0	719	89
448193	0	1662	100
448193	0	1662	100
448193	3.00E-162	474	29
448569	0	688	100
448569	9.00E-157	442	65
448960	0	707	100
448960	0	707	100
448960	0	707	100
449047	2.00E-174	483	100
449047	2.00E-174	483	100
449865	0	1793	100
450336	0	1043	100
450336	0	1043	100
450594	2.00E-180	520	35
450615	0	1987	100
450615	0	1987	100
450615	0	1972	100
452198	0	569	74
452198	0	761	100
452198	0	761	100
452198	0	761	100
452284	1.00E-166	475	47
453068	6.00E-143	402	90
453373	0	672	100
453373	0	672	100

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
454897	0	1457	100
454897	0	1457	100
455078	0	1662	100
455078	0	1662	100
455280	0	3069	100
455280	5.00E-176	529	17
455300	1.00E-178	496	100
455406	0	803	100
455406	0	803	100
456187	3.00E-125	353	100
456239	0	1311	100
456254	0	602	98
457572	1.00E-125	355	100
457856	8.00E-170	472	100
458081	5.00E-108	310	95
458081	1.00E-81	241	72
458122	0	798	100
458122	0	750	95
459601	5.00E-159	443	100
459601	6.00E-159	443	100
459601	6.00E-159	443	100
459601	9.00E-144	404	94
459734	0	837	100
459734	0	837	100
459736	0	546	29
459736	0	1874	100
459736	0	1874	100
459873	0	966	100
460004	1.00E-129	367	74
460492	0	885	100
460492	0	885	100
461181	0	829	100
461699	0	657	97
461699	0	678	100
461699	0	678	100
462385	0	507	74
462385	0	681	99
462385	0	681	99
462385	0	689	100
462433	4.00E-134	377	100
462452	0	2905	100
462452	0	2905	100
462452	0	2905	100
462457	0	973	100
462457	0	975	100
462457	0	975	100
462645	0	593	100
462645	0	593	100

APPENDIX B– SUPPLEMENTARY RESULTS – CHAPTER 4 – THE PROTEOME
OF EMILIANIA HUXLEYI AT HIGH AND LOW CALCIFCATION
RATES OF THE G1-PHASE

257

Appendix B Table 4 E. hux JGI gene ID	E Value	Score	Hit Sim. in %
463036	0	666	92
463429	0	935	90
463837	0	2722	90
463837	0	3181	100
464103	0	778	99
464103	0	783	100
464103	0	783	100
464103	0	702	92
464599	0	1691	100
464599	0	1691	100
464599	0	1619	96
464740	0	4036	100
464740	0	4036	100
464972	0	514	89
464972	0	588	100
464972	0	588	100
465288	0	727	100
465288	0	727	100
465364	0	1316	100
465364	0	1279	99
465373	0	927	100
465373	5.00E-177	499	55
465509	0	1215	100
465509	0	1215	100
465509	0	1174	97
465547	0	628	100
465547	0	628	100
465547	1.00E-131	375	64
465676	0	1167	100
465926	8.00E-139	407	75
465926	3.00E-94	283	29
466230	0	783	87
466469	0	2164	58
466469	0	3430	93
466778	0	591	89
466893	0	1835	100
466893	0	1818	100
466893	2.00E-163	482	27
468608	6.00E-143	401	100
468608	6.00E-143	401	100
468689	0	671	96
470712	0	541	100
470712	0	541	100
470997	0	1500	100
49364	6.00E-52	159	100
54552	0	608	100
61467	6.00E-172	478	100
61745	8.00E-137	384	100
61745	8.00E-137	384	100

Appendix B Table 4 *E. hux* JGI gene ID	E Value	Score	Hit Sim. in %
63832	0	830	100
64238	0	513	100
66170	1.00E-82	241	100
68547	0	835	93
68856	0	1446	100
68856	0	1446	100
68856	0	1446	100
69198	0	891	100
69198	0	891	100
69198	0	891	100
69773	5.00E-104	298	100
69773	5.00E-104	298	100
69773	5.00E-104	298	100
74095	0	501	100
74095	0	501	100
74335	4.00E-93	268	100
77411	0	658	100
77411	0	658	100
77411	0	658	100
77826	1.00E-152	426	96
78752	2.00E-175	487	100
78752	2.00E-175	487	100
78806	6.00E-58	176	100
79310	9.00E-131	368	100

References

Alabadí, D.; Oyama, T.; Yanovsky, M. J.; Harmon, F. G.; Más, P.; Kay, S. A. (2001) Reciprocal Regulation Between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock. *Science*, 293 (5531), 880–883.

Addadi, L.; Raz, S.; Weiner, S. (2003) Taking Advantage of Disorder: Amorphous Calcium Carbonate and Its Roles in Biomineralization. *Adv. Mater.*, 15 (12), 959–970.

Alcolombri, U.; Ben-Dor, S.; Feldmesser, E.; Levin, Y.; Tawfik, D. S.; Vardi, A. (2015) Identification of the algal dimethyl sulfide-releasing enzyme: A missing link in the marine sulfur cycle. *Science*, 348 (6242), 1466–1469.

Alwine, J. C.; Kemp, D. J.; Stark, G. R. (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U.S.A.*, 74 (12), 5350–5354.

Anders, S.; Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, 11 (10)

Anderson, L.; Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, 18 (3-4), 533–537.

Anderson, N. L.; Anderson, N. G. (1998) Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis*, 19 (11), 1853–1861.

Andrulis, E. D.; Guzman, E.; Doering, P.; Werner, J.; Lis, J. T. (2000) High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes & Development*, 14 (20), 2635–2649.

Anning, T.; Nimer, N.; Merrett, M. J.; Brownlee, C. (1996) Costs and benefits of calcification in coccolithophorids. *Journal of Marine Systems*, 9 (1-2), 45–56.

Arai, H.; Pink, S.; Forgac, M. (1989) Interaction of anions and ATP with the coated vesicle proton pump. *Biochemistry*, 28 (7), 3075–3082.

Araki, Y.; Gonzalez, E. L. (1998) V- and P-type Ca^{2+}-stimulated ATPases in a calcifying strain of *Pleurochrysis sp.* (Haptophyceae). *J Phycol*, 34 (1), 79–88.

Archer, D.; Winguth, A.; Lea, D.; Mahowald, N. (2000) What caused the glacial/interglacial atmospheric pCO_2 cycles? *Rev. Geophys.*, 38 (2), 159–189.

Arias, J. L.; Fernández, M. S. (2008) Polysaccharides and Proteoglycans in Calcium Carbonate-based Biomineralization. *Chem. Rev.*, 108 (11), 4475–4482.

Armstrong, R. A.; Lee, C.; Hedges, J. I.; Honjo, S.; Wakeham, S. G. (2001) A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, 49(1), 219–236.

Arnaudeau, S. (2002) Calreticulin Differentially Modulates Calcium Uptake and Release in the Endoplasmic Reticulum and Mitochondria. *J Biol Chem*, 277 (48), 46696–46705.

Ateweberhan, M.; Feary, D. A.; Keshavmurthy, S.; Chen, A.; Schleyer, M. H.; Sheppard, C. R. C. (2013) Climate change impacts on coral reefs: Synergies with local effects, possibilities for acclimation, and management implications. *Palaeogeography, Palaeoclimatology, Palaeoecology*.

Bach, L. T.; Mackinder, L. C. M.; Schulz, K. G.; Wheeler, G.; Schroeder, D. C.; Brownlee, C.; Riebesell, U. (2013) Dissecting the impact of CO_2 and pH on the mechanisms of photosynthesis and calcification in the coccolithophore *Emiliania huxleyi*. *J Phycol*, 199 (1), 121–134.

Badger, M. R.; Price, G. D. (1994) The Role of Carbonic Anhydrase in Photosynthesis. *Annu. Rev. Ecol. Evol. Syst.*, 45 (1), 369–392.

Baker, K. E.; Condon, C. (2004) Under the Tucson sun: a meeting in the desert on mRNA decay. *RNA*, Vol. 10, 1680–1691.

Balch, W. M.; Drapeau, D. T.; Fritz, J. J. (2000) Monsoonal forcing of calcification in the Arabian Sea. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, 47 (7-8), 1301–1337.

Balch, W. M.; Holligan, P. M.; Kilpatrick, K. A. (1992a) Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliania huxleyi. Continental Shelf Research*, 12 (12), 1353–1374.

Balch, W. M.; Holligan, P. M.; Kilpatrick, K. A. (1992b) Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliania huxleyi. Continental Shelf Research*, 12 (12), 1353–1374

Baumann, K. H.; Böckel, B.; Frenz, M. (2004) Coccolith contribution to South Atlantic carbonate sedimentation. In: *Coccolithophores: from molecular processes to global impact* (H. R. Thierstein & J. Young, eds.); Berlin: Springer.

Beardall, J.; Young, E.; Roberts, S. (2001) Approaches for determining phytoplankton nutrient limitation. *Aquat. sci.*, 63 (1), 44–69.

Bellantuono, A. J.; Granados-Cifuentes, C.; Miller, D. J.; Hoegh-Guldberg, O.; Rodriguez-Lanetty, M. (2012) Coral thermal tolerance: tuning gene expression to resist thermal stress. *PLoS ONE*, 7 (11).

Bengal, E.; Flores, O.; Krauskopf, A.; Reinberg, D.; Aloni, Y. (1991) Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Mol. Cell. Biol.*, 11 (3), 1195–1206.

Beniash, E.; Aizenberg, J.; Addadi, L.; Weiner, S. (1997) Amorphous calcium carbonate transforms into calcite during sea urchin larval spicule growth. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264 (1380), 461–465.

Benjamini, Y.; Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*, 57 (1), 289–300.

Bentov, S.; Weil, S.; Glazer, L.; Sagi, A.; Berman, A. (2010) Stabilization of amorphous calcium carbonate by phosphate rich organic matrix proteins and by single phosphoamino acids. *Journal of Structural Biology*, 171 (2), 207–215.

Berridge, M. J. (2002) The endoplasmic reticulum: a multifunctional signaling organelle. *Cell Calcium*, 32 (5), 235–249.

Beyenbach, K. W.; Wieczorek, H. (2006) The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. *J. Exp. Biol.*, 209 (Pt 4), 577–589.

Bhattacharya, D.; Medlin, L. (1995) Phylogeny of plastids A review based on comparisons of small subunit ribosomal RNA coding regions. *J Phycol*.

Birol, I.; Jackman, S. D.; Nielsen, C. B.; Qian, J. Q.; Varhol, R.; Stazyk, G.; Morin, R. D.; et al. (2009) De novo transcriptome assembly with ABySS. *Bioinformatics*, 25 (21), 2872–2877.

Bisova, K.; Krylov, D. M.; Umen, J. G. (2005) Genome-wide annotation and expression profiling of cell cycle regulatory genes in *Chlamydomonas reinhardtii. Plant Physiol.*, 137 (2), 475–491.

Blackstock, W. P.; Weir, M. P. (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends in Biotechnology*, 17 (3), 121–127.

Bonaldo, M. F.; Lennon, G.; Soares, M. B. (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.*, 6 (9), 791–806.

Borman, A. H.; Jong, E. W.; Huinzinga, M.; Kok, D.; Westbroek, P.; BOSCH, L. (1982) The role in $CaCO_3$ crystallization of an acid Ca^{2+}-Binding polysaccharide associated with coccoliths of *Emiliania huxleyi. Eur J Biochem*, 129 (1), 179–183.

Bown, P. R.; Lees, J. A.; Young, J. R. (2004) Calcareous nannoplankton evolution and diversity through time. *Coccolithophores-from molecular processes to global impact.-Berlin (Springer)*, 481–508.

Brenner, S.; Johnson, M.; Bridgham, J.; Golda, G.; Lloyd, D. H.; Johnson, D.; Luo, S.; et al. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, 18 (6), 630–634.

Bricker, T. M.; Roose, J. L.; Fagerlund, R. D.; Frankel, L. K.; Eaton-Rye, J. J. (2012) Biochimica et Biophysica Acta. *BBA - Bioenergetics*, 1817 (1), 121–142.

Briegel, C.; Seto, J. (2012) Single Amino Acids as Additives Modulating $CaCO_3$ Mineralisation. In: *Advanced Topics in Biomineralization* (J. Seto, ed.), 1–16.

Brodsky, J. L.; Skach, W. R. (2011) Protein folding and quality control in the endoplasmic reticulum: Recent lessons from yeast and mammalian cell systems. *Curr. Opin. Cell Biol.*, 23 (4), 464–475.

Brownlee, C.; Davies, M.; Nimer, N.; Dong, L.; Merrett, M. (1995) Calcification, photosynthesis and intracellular regulation in *Emiliania huxleyi. Bulletin de l'Institut Oceanographique (Monaco)*, 14, 19–35.

Brownlee, C.; Taylor, A. (2004) Calcification in coccolithophores: A cellular perspective. In: *Coccolithophores: from molecular processes to global impact* (H. R. Thierstein & J. Young, eds.); Berlin: Springer.

Buitenhuis, E. T.; De Baar, H. J. W.; Veldhuis, M. J. W. (1999) Photosynthesis and calcification by *Emiliania huxleyi* (Prymnesiophyceae) as a function of inorganic carbon species. *J Phycol*, 35 (5), 949–959.

Buitenhuis, E. T.; Van Bleijswijk, J.; Bakker, D.; Veldhuis, M. J. W. (1996) Trends in inorganic and organic carbon in a bloom of *Emiliania huxleyi. Mar Ecol-Prog Ser*, 143, 271–282.

Cardone, L.; Sassone-Corsi, P. (2003) Timing the cell cycle. *Nature Cell Biology* 5(10),859–861.

Carretero-Paulet, L.; Galstyan, A.; Roig-Villanova, I.; Martínez-García, J. F.; Bilbao-Castro, J. R.; Robertson, D. L. (2010) Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiology*, 153 (3), 1398–1412.

Charlson, R. J.; Lovelock, J. E.; Andreae, M. O.; Warren, S. G. (1987) Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature*, 326 (6114), 655–661.

Chen, C.-M.; Shih, T.-H.; Pai, T.-W.; Liu, Z.-L.; Chang, M. D.-T. (2012) RNA-seq coverage effects on biological pathways and GO tag clouds, 240–245; IEEE.

Chen, G.; Gharib, T. G.; Huang, C.-C.; Taylor, J. M. G.; Misek, D. E.; Kardia, S. L. R.; Giordano, T. J.; et al. (2002) Discordant Protein and mRNA Expression in Lung Adenocarcinomas. *Molecular & Cellular Proteomics*, 1 (4), 304–313.

Chen, M. (2008) Phytochrome nuclear body: an emerging model to study interphase nuclear dynamics and signaling. *Current Opinion in Plant Biology*, 11 (5), 503–508.

Chen, Z.; McKnight, S. L. (2007) A conserved DNA damage response pathway responsible for coupling the cell division cycle to the circadian and metabolic cycles. *Cell cycle*, 6(23), 2906–2912.

Chiang, P.; Gordon, R.; Tal, J.; Zeng, G.; Doctor, B.; Pardhasaradhi, K.; McCann, P. (1996) S-Adenosylmethionine and methylation. *The FASEB Journal*, 10 (4), 471–480.

Chisholm, S. W.; Vaulot, D.; Olson, R. J. (1984) Cell cycle control in phytoplankton: comparative physiology and ecology. In: *Cell cycle clocks* (L. N. J. Edmunds, ed.), 365–394; New York (USA).

Choe, L.; D'Ascenzo, M.; Relkin, N. R.; Pappin, D.; Ross, P.; Williamson, B.; Guertin, S.; et al. (2007) 8-Plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*, 7 (20), 3651–3660.

Choi, Y. W.; Park, S. A.; Lee, H. W.; Kim, D. S.; Lee, N. G. (2008) Analysis of growth phase-dependent proteome profiles reveals differential regulation of mRNA and protein in *Helicobacter pylori*. *Proteomics*, 8 (13), 2665–2675; WILEY-VCH Verlag.

Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. (2001) Solution structure of Ca^{2+} - calmodulin reveals flexible hand-like properties of its domains. *Nature*, 8 (11), 990–997;

Chou, K.-C.; Shen, H.-B. (2010) A New Method for Predicting the Subcellular Localization of Eukaryotic Proteins with Both Single and Multiple Sites: Euk-mPLoc 2.0. (D. P. Martin, Ed.)*PLoS ONE*, 5 (4), e9931.

Claquin, P.; Kromkamp, J. C.; Martin-Jezequel, V. (2004) Relationship between photosynthetic metabolism and cell cycle in a synchronized culture of the marine alga *Cylindrotheca fusiformis* (Bacillariophyceae). *European Journal of Phycology*, 39 (1), 33–41.

Corstjens, P. L. A. M.; Araki, Y.; Gonzalez, E. L. (2001) A coccolithophorid calcifying vesicle with a vacuolar-type ATPase proton pump: cloning and immunolocalization of the V0 subunit C. *J Phycol*, 37 (1), 71–78.

Corstjens, P. L. A. M.; van der Kooij, A.; Linschooten, C.; Brouwers, G.-J.; Westbroek, P.; de Vrind-de Jong, E. W. (1998) GPA, a calcium-binding protein in the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae). *J Phycol*, 34 (4), 622–630.

Cusack, M.; Freer, A. (2008) Biomineralization: Elemental and Organic Influence in Carbonate Systems. *Chem. Rev.*, 108 (11), 4433–4454.

Dassow, von, P.; Ogata, H.; Probert, I.; Wincker, P.; Da Silva, C.; Audic, S.; Claverie, J.-M.; et al. (2009) Transcriptome analysis of functional differentiation between haploid and diploid cells of *Emiliania huxleyi*, a globally significant photosynthetic calcifying cell. *Genome Biol.*, 10 (10), R114.

de Jong, E. W.; Bosch, L.; Westbroek, P. (1976) Isolation and characterization of a Ca^{2+} - binding polysaccharide associated with coccoliths of *Emiliania huxleyi* (Lohmann) Kamptner. *Eur J Biochem*, 70, 611–621.

de Vargas, C.; Aubry, M.; Probert, I.; Young, J. (2007) *Origin and evolution of coccolithophores: From coastal hunters to Oceanic farmers*. (P. Falkowski & A. Knoll, Eds.), 251–285.

DeCoursey, P. J.; Walker, J. K.; Smith, S. A. (2000) A circadian pacemaker in free-living chipmunks: essential for survival? *J. Comp. Physiol. A*, 186 (2), 169–180.

Denoeud, F.; Aury, J.-M.; Da Silva, C.; Noel, B.; Rogier, O.; Delledonne, M.; Morgante, M.; et al. (2008) Annotating genomes with massive-scale RNA sequencing. *Genome Biol.*, 9 (12), R175.

Deutscher, M. P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Research*, 34 (2), 659–666.

Dixon, G. K.; Brownlee, C.; Merrett, M. J. (1989) Measurement of internal pH in the coccolithophore *Emiliania huxleyi* using 2′,7′-bis-(2-carboxyethyl)-5(and-6) carboxyfluorescein acetoxymethylester and digital imaging microscopy. *Planta*, 178 (4), 443–449.

Dodd, A. N.; Salathia, N.; Hall, A.; Kévei, E.; Tóth, R.; Nagy, F.; Hibberd, J. M.; et al. (2005) Plant Circadian Clocks Increase Photosynthesis, Growth, Survival, and Competitive Advantage. *Science*, 309 (5734), 630–633.

Dolphin, A. C. (2009) Calcium channel diversity: multiple roles of calcium channel subunits. *Curr. Opin. Neurobiol.*, 19 (3), 237–244.

Donnan, L.; Carvill, E. P.; Gilliland, T. J.; John, P. C. L. (1985) The cell cycles of *Chlamydomonas* and *Chlorella*. *New Phytol*, 99 (1), 1–40.

Drescher, B.; Dillaman, R. M.; Taylor, A. R. (2012) Coccolithogenesis in *Scyphosphaera apsteinii* (Prymnesiophyceae). *J Phycol*,48 (6) 1343-1361.

Dyczkowski, J.; Vingron, M. (2005) Comparative analysis of cell cycle regulated genes in eukaryotes. *Genome Informatics Series*.

Dyhrman, S. T.; Haley, S. T.; Birkeland, S. R.; Wurch, L. L.; Cipriano, M. J.; McArthur, A. G. (2006) Long serial analysis of gene expression for gene discovery and transcriptome profiling in the widespread marine coccolithophore *Emiliania huxleyi. Appl. Environ. Microbiol.*, 72 (1), 252–260.

Dyhrman, S. T.; Jenkins, B. D.; Rynearson, T. A.; Saito, M. A.; Mercier, M. L.; Alexander, H.; Whitney, L. P.; et al. (2012) The Transcriptome and Proteome of the Diatom *Thalassiosira pseudonana* Reveal a Diverse Phosphorus Stress Response. (P. Santos, Ed.)*PLoS ONE*, 7 (3), e33768.

Eberhardy, S. R.; Farnham, P. J. (2002) Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J. Biol. Chem.*, 277 (42), 40156–40162.

Eddy, S. R. (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet*, 2 (12), 919–929.

Edmunds, L. N. J.; Adams, K. J. (1981) Clocked cell cycle clocks. *Science*, 211, 1002–1012.

Edwards, R. J.; Moran, N.; Devocelle, M.; Kiernan, A.; Meade, G.; Signac, W.; Foy, M.; et al. (2007) Bioinformatic discovery of novel bioactive peptides. *Nat Chem Biol*, 3 (2), 108–112.

Emery, L.; Whelan, S.; Hirschi, K. D.; Pittman, J. K. (2012) Protein Phylogenetic Analysis of Ca^{2+}/cation Antiporters and Insights into their Evolution in Plants. *Front Plant Sci*, 3, 1.

Engel, A.; Szlosek, J.; Abramson, L.; Liu, Z. F.; Lee, C. (2009) Investigating the effect of ballasting by $CaCO_3$ in *Emiliania huxleyi*: I. Formation, settling velocities and physical properties of aggregates. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, 56 (18), 1396–1407.

Fabry, V. J. (2008) Marine calcifiers in a high-CO_2 ocean. *Science*, 320 (5879), 1020–1022.

Fagerbakke, K. M.; Heldal, M.; Norland, S.; Heimdal, B. R.; Båtvik, H. (1994) *Emiliania huxleyi*. Chemical composition and size of coccoliths from enclosure experiments and a Norwegian fjord. *British Phycological Journal*; Taylor & Francis Group.

Falkowski, P. (2000) The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science*, 290 (5490), 291–296.

Falkowski, P. G. (2004) The Evolution of Modern Eukaryotic Phytoplankton. *Science*, 305 (5682), 354–360.

Falkowski, P. G.; Kolber, Z. (1995) Variations in Chlorophyll Fluorescence Yields in Phytoplankton in the World Oceans. *Functional Plant Biol.*, 22 (2), 341–355.

Fankhauser, C.; Chen, M. (2008) Transposing phytochrome into the nucleus. *Trends in Plant Science*, 13 (11), 596–601.

Farre, E. M. (2012) The regulation of plant growth by the circadian clock. *PLANT BIOLOGY*, {14} ({3}), {401–410}.

Feder, M. E.; Hofmann, G. E. (1999) Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annu. Rev. Physiol.*, 61, 243–282.

Field, C. B.; Behrenfeld, M. J.; Randerson, J. T. (1998) Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*.

Finazzi, G.; Moreau, H. (2010) Genomic insights into photosynthesis in eukaryotic phytoplankton. *Trends in Plant Science*.

Finbow, M. E.; Harrison, M. A. (1997) The vacuolar H+-ATPase: a universal proton pump of eukaryotes. *The Royal Society*, 324 (Pt 3), 697–712.

Findlay, C. S.; Giraudeau, J. (2000) Extant calcareous nannoplankton in the Australian Sector of the Southern Ocean (austral summers 1994 and 1995). *Marine Micropaleontology*, 40 (4), 417–439.

Forgac, M. (2000) Structure, mechanism and regulation of the clathrin-coated vesicle and yeast vacuolar H+-ATPases. *J. Exp. Biol.*, 203 (1), 71–80.

Frankignoulle, M.; Canon, C.; Gattuso, J.-P. (1994) Marine calcification as a source of carbon dioxide: positive feedback of increasing atmospheric CO_2. *Limnol. Oceanogr.*, 39 (2), 458–462.

Frith, M. C.; Pheasant, M.; Mattick, J. S. (2005) Genomics: The amazing complexity of the human transcriptome. *Nature*, 13 (8), 894–897.

Fritz, J. J. (1999) Carbon fixation and coccolith detachment in the coccolithophore *Emiliania huxleyi* in nitrate-limited cyclostats. *Mar Biol*, 133 (3), 509–518.

Fritz, J. J.; Balch, W. M. (1996) A light-limited continuous culture study of *Emiliania huxleyi*: determination of coccolith detachment and its relevance to cell sinking. *Journal of Experimental Marine Biology and Ecology*, 207 (1), 127–147.

Fujiwara, S.; Hirokawa, Y.; Takatsuka, Y.; Suda, K.; Asamizu, E.; Takayanagi, T.; Shibata, D.; et al. (2007) Gene Expression Profiling of Coccolith-Bearing Cells and Naked Cells in Haptophyte *Pleurochrysis haptonemofera* with a cDNA Macroarray System. *Mar Biotechnol*, 9 (5), 550–560.

Fujiwara, T.; Misumi, O.; Tashiro, K.; Yoshida, Y.; Nishida, K.; Yagisawa, F.; Imamura, S.; et al. (2009) Periodic Gene Expression Patterns during the Highly Synchronized Cell Nucleus and Organelle Division Cycles in the Unicellular Red Alga Cyanidioschyzon merolae. *DNA Res.*, 16 (1), 59–72.

Gehlenborg, N.; O'Donoghue, S. I.; Baliga, N. S.; Goesmann, A.; Hibbs, M. A.; Kitano, H.; Kohlbacher, O.; et al. (2010) Visualization of omics data for systems biology. *Nature Publishing Group*, 7 (3s), S56–S68.

Geider, R. J.; Roche, J.; Greene, R. M.; Olaizola, M. (1993) Response of the photosynthetic apparatus of *Phaeodactylum tricornutum* (Bacillariophyceae) to nitrate, phosphate, or iron starvation1. *J Phycol*, 29 (6), 755–766.

Geider, R.; Macintyre; Graziano, L.; McKay, R. M. (1998) Responses of the photosynthetic apparatus of *Dunaliella tertiolecta* (Chlorophyceae) to nitrogen and phosphorus limitation. *European Journal of Phycology*, 33 (4), 315–332.

Gould, S. B.; Waller, R. F. (2008) Plastid evolution. *Annu Rev Plant Biol.*

Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 29 (7), 644–652.

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

Green, R. M.; Tingay, S.; Wang, Z.-Y.; Tobin, E. M. (2002) Circadian rhythms confer a higher level of fitness to Arabidopsis plants. *Plant Physiol.*, 129 (2), 576–584.

Gregg, W. W.; Casey, N. W. (2007) Modeling coccolithophores in the global oceans. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, 54 (5), 447–477.

Guillard, R. R. L. (1975) Culture of phytoplankton for feeding marine invertebrates. In: *Culture of marine invertebrates* (W. Smith & M. Chanley, eds.), 26–60.

Hagiwara, S.; Bolige, A.; Zhang, Y. L.; Takahashi, M.; Yamagishi, A.; Goto, K. (2002) Circadian gating of photoinduction of commitment to cell-cycle transitions in relation to photoperiodic control of cell reproduction in Euglena. *Photchemistry and Photobiology*, (76/1), (105–115).

Harbers, M. M.; Carninci, P. P. (2005) Tag-based approaches for transcriptome research and genome annotation. *Nat Methods*, 2 (7), 495–502.

Haugan, P. M.; Drange, H. (1996) Effects of CO_2 on the ocean environment. *Energy Conversion and Management*, 37 (6-8), 1019–1022.

Hegg, D. A.; Ferek, R. J.; Hobbs, P. V.; Radke, L. F. (1991) Dimethyl sulfide and cloud condensation nucleus correlations in the northeast Pacific Ocean. *J. Geophys. Res.*, 96, 13189–13191.

Henriksen, K.; Stipp, S. L. S.; Young, J. R.; Marsh, M. E. (2004) Biological control on calcite crystallization: AFM investigation of coccolith polysaccharide function. *ammin*, 89 (11-12), 1709–1716.

Herfort, L.; Loste, E.; Meldrum, F.; Thake, B. (2004) Structural and physiological effects of calcium and magnesium in *Emiliania huxleyi* (Lohmann) Hay and Mohler. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 148 (3), 307–314.

Herfort, L.; Thake, B.; Roberts, J. (2002) Acquisition and use of bicarbonate by *Emiliania huxleyi*. *New Phytol*. 156, 427-456

Ho, T.-Y.; Quigg, A.; Finkel, Z. V.; Milligan, A. J.; Wyman, K.; Falkowski, P. G.; Morel, F. M. M. (2003) The elemental composition of some marine phytoplankton. *J Phycol*, 39 (6), 1145–1159.

Holligan, P. M.; Fernandez, E.; Aiken, J.; Balch, W. M.; Boyd, P.; Burkill, P. H.; Finch, M.; et al. (1993) A biogeochemical study of the coccolithophore, *Emiliania huxleyi*, in the North Atlantic. *Global Biogeochem Cy*, 7 (4), 879–900.

Horton, P.; Park, K.-J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C. J.; Nakai, K. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, 35 (Web Server issue), W585–7.

Houck, K. A.; Ferrara, N.; Winer, J.; Cachianes, G.; Li, B.; Leung, D. W. (1991) the vascular endothelial growth factor family: identification of a fourth molecular species and characterization of alternative splicing of rna. *Molecular Endocrinology*, 5 (12), 1806–1814.

Houghton, R. A. (2007) Balancing the Global Carbon Budget. *Annu. Rev. Earth Planet. Sci.*, 35 (1), 313–347.

Humphery-Smith, I.; Cordwell, S. J.; Blackstock, W. P. (1997) Proteome research: Complementarity and limitations with respect to the RNA and DNA worlds. *J Phycol*, 18 (8), 1217–1242.

Huntley, R. P.; Sawford, T.; Mutowo-Meullenet, P.; Shypitsyna, A.; Bonilla, C.; Martin, M. J.; O'Donovan, C. (2015) The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Research*, 43 (D1), D1057–D1063.

Hüttenhofer, A.; Vogel, J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Research*, 34 (2), 635–646.

Iglesias-Rodriguez, M. D.; Brown, C. W.; Doney, S. C.; Kleypas, J.; Kolber, D.; Kolber, Z.; Hayes, P. K.; et al. (2002) Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids. *Global Biogeochem Cy*, 16 (4), –.

Iglesias-Rodriguez, M. D.; Halloran, P. R.; Rickaby, R. E. M.; Hall, I. R.; Colmenero-Hidalgo, E.; Gittins, J. R.; Green, D. R. H.; et al. (2008) Phytoplankton Calcification in a High-CO_2 World. *Science*, 320 (5874), 336–340.

Iglesias-Rodriguez, M. D.; Schofield, O. M.; Batley, J.; Medlin, L. K.; Hayes, P. K. (2006) Intraspecific genetic diversity in the marine coccolithophore *Emiliania huxleyi* (Prymnesiophyceae): The use of microsatellite analysis in marine phytoplankton population studies. *J Phycol*, 42 (3), 526–536.

Inouye, I.; Chihara, M. (1980) Laboratory culture and taxonomy of *Hymenomonas coronata* and *Ochrosphaera verrucosa* (Class Prymnesiophyceae) from the Northwest Pacific. *Bot Mag Tokyo*, 93 (3), 195–208.

Inzé, D.; De Veylder, L. (2006) Cell cycle regulation in plant development 1. *Annu. Rev. Genet.*, 40, 77–105.

Isenberg, H. D.; Lavine, L. S.; Weissfellner, H. (1963) The suppression of mineralization in a coccolithophorid by an inhibitor of carbonic anhydrase. *J Eukaryotic Microbiology*, 10 (4), 477–479.

Izban, M. G.; Luse, D. S. (1991) Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes & Development*, 5 (4), 683–696.

Jackowiak, P.; Nowacka, M.; Strozycki, P. M.; Figlerowicz, M. (2011) RNA degradome--its biogenesis and functions. *Nucleic Acids Research*, 39 (17), 7361–7370.

Jain, C. (2002) Degradation of mRNA in *Escherichia coli*. *J Phycol*, 54 (6), 315–321.

Jang, M. K.; Mochizuki, K.; Zhou, M.; Jeong, H.-S.; Brady, J. N.; Ozato, K. (2005) The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell*, 19 (4), 523–534.

Jayadev, R.; Kuk, C. Y.; Low, S. H.; Murata-Hori, M. (2014) Calcium sensitivity of α-actinin is required for equatorial actin assembly during cytokinesis. *Cell Cycle*, 11 (10), 1929–1937.

John, L. M.; Lechleiter, J. D.; Camacho, P. (1998) Differential modulation of SERCA2 isoforms by calreticulin. *J. Cell Biol.*, 142 (4), 963–973.

Johnson, C. H.; Golden, S. S.; Ishiura, M.; Kondo, T. (1996) Circadian clocks in prokaryotes. *Molecular Microbiology*, 21 (1), 5–11.

Jones, B. M.; Edwards, R. J.; Skipp, P. J.; O'Connor, C. D.; Iglesias-Rodriguez, M. D. (2011) Shotgun Proteomic Analysis of *Emiliania huxleyi*, a Marine Phytoplankton Species of Major Biogeochemical Importance. *Mar Biotechnol*, 13 (3), 496–504.

Jones, B. M.; Iglesias-Rodriguez, M. D.; Skipp, P. J.; Edwards, R. J.; Greaves, M. J.; Young, J. R.; Elderfield, H.; et al. (2013) Responses of the *Emiliania huxleyi* Proteome to Ocean Acidification. *PLoS ONE*, 8 (4), e61868.

Jordan, R. W.; Chamberlain, A. H. L. (1997) Biodiversity among haptophyte algae. *Biodiversity and Conservation*, 6 (1), 131–152.

Joubert, C.; Piquemal, D.; Marie, B.; Manchon, L.; Pierrat, F.; Zanella-Cléon, I.; Cochennec-Laureau, N.; et al. (2010) Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. *BMC Genomics*, 11 (1), 613–.

Kalia, A.; Gupta, R. P. (2005) Proteomics: A Paradigm Shift. *Critical Reviews in Biotechnology*, 25 (4), 173–198.

Kane, P. M. (1995) Disassembly and reassembly of the yeast vacuolar H^+-ATPase in vivo. *J Biol Chem.*, 270 (28), 17025–17032.

Kanesaki, Y.; Imamura, S.; Minoda, A.; Tanaka, K. (2012) external light conditions and internal cell cycle phases coordinate accumulation of chloroplast and mitochondrial transcripts in the red alga cyanidioschyzon merolae. *DNA Res.*, 19 (3), 289–303.

Kaniewska, P.; Anthony, K.; Sampayo, E. M. (2013) Implications of geometric plasticity for maximizing photosynthesis in branching corals. *Mar Biol*. DOI 10.1007/s00227-013-2336-z

Kanno, T.; Huettel, B.; Mette, M. F.; Aufsatz, W.; Jaligot, E.; Daxinger, L.; Kreil, D. P.; et al. (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nature*, 37 (7), 761–765.

Karako-Lampert, S.; Katcoff, D. J.; Achituv, Y.; Dubinsky, Z.; Stambler, N. (2004) Do clades of symbiotic dinoflagellates in scleractinian corals of the Gulf of Eilat (Red Sea) differ from those of other coral reefs? *Journal of Experimental Marine Biology and Ecology*, 311 (2), 301–314.

Katoh, K.; Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, 9 (4), 286–298.

Keeling, P. J. (2004) Diversity and evolutionary history of plastids and their hosts. *Am. J. Bot.*, 91 (10), 1481–1493.

Keeling, P. J. (2009) Chromalveolates and the Evolution of Plastids by Secondary Endosymbiosis. *Journal of Eukaryotic Microbiology*, 56 (1), 1–8.

Keller, E. B.; Noon, W. A. (1984) Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci. U.S.A.*, 81 (23), 7417–7420.

Kettenberger, H.; Armache, K.-J.; Cramer, P. (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell*, 16 (6), 955–965.

King, N. L.; Deutsch, E. W.; Ranish, J. A.; Nesvizhskii, A. I.; Eddes, J. S.; Mallick, P.; Eng, J.; et al. (2006) Analysis of the *Saccharomyces cerevisiae* proteome with PeptideAtlas. *Genome Biol.*, 7 (11), R106.

Kirchhausen, T. (2000) Clathrin. *Annu. Rev. Biochem.*, (69), 699–727.

Kirk, K. L. (1997) life-history responses to variable environments: starvation and reproduction in planktonic rotifers. *Ecology*, 78 (2), 434–441.

Kitao, Y.; Harada, H.; Matsuda, Y. (2008) Localization and targeting mechanisms of two chloroplastic beta-carbonic anhydrases in the marine diatom *Phaeodactylum tricornutum*. *Physiol Plant*, 133 (1), 68–77.

Klaas, C.; Archer, D. E. (2002) Association of sinking organic matter with various types of mineral ballast in the deep sea: Implications for the rain ratio. *Global Biogeochem Cy*, 16 (4), 1116; AGU.

Klaveness, D. (1972) *Coccolithus huxleyi* (Lohmann) Kamptner. 1. Morphological investigations on the vegetative cell and the process of coccolith formation. *Protistologica (Paris. 1965)*, 8 (3).

Klaveness, D. (1976) Emiliania huxleyi (Lohmann) Hay and Mohler 3. Mineral deposition and origin of matrix during coccolith formation. *Protistologica*, 12, 217–224.

Kodzius, R.; Kojima, M.; Nishiyori, H.; Nakamura, M.; Fukuda, S.; Tagami, M.; Sasaki, D.; et al. (2006) CAGE: cap analysis of gene expression. *Nat Methods*, 3 (3), 211–222.

Koenig, T.; Menze, B. H.; Kirchner, M.; Monigatti, F.; Parker, K. C.; Patterson, T.; Steen, J. J.; et al. (2008) Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J Proteome Res*, 7 (9), 3708–3717.

Kolber, Z. S.; Barber, R. T.; Coale, K. H.; Fitzwateri, S. E.; Greene, R. M.; Johnson, K. S.; Lindley, S.; et al. (1994) Iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature*, 371 (6493), 145–149.

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

Kondorosi, E.; Roudier, F.; Gendreau, E. (2000) Plant cell-size control: growing by ploidy? *Palaeogeography, Palaeoclimatology, Palaeoecology*, 3 (6), 488–492.

Krause, G. H.; Weis, E. (1991) Chlorophyll fluorescence and photosynthesis: the basics. *Annu Rev Plant Biol*, 42 (1), 313–349.

Krause, G. H.; Winter, K.; Krause, B.; Virgo, A. (2015) Light-stimulated heat tolerance in leaves of two neotropical tree species, Ficus insipida and *Calophyllum longifolium*. *Functional Plant Biol.*, 42 (1), 42.

Kruskopf, M.; Flynn, K. J. (2005) Chlorophyll content and fluorescence responses cannot be used to gauge reliably phytoplankton biomass, nutrient status or growth rate. *New Phytol*, 169 (3), 525–536.

La Claire, J. W. (2006) Analysis of Expressed Sequence Tags from the Harmful Alga, *Prymnesium parvum* (Prymnesiophyceae, Haptophyta). *Mar Biotechnol*, 8 (5), 534–546.

La Roche, J.; Geider, R. J.; Graziano, L. M.; Murray, H.; Lewis, K. (1993) Induction of specific proteins in eukaryotic algae grown under iron-, phosphorus-, or nitrogen-deficient conditions. *J Phycol*, 29 (6), 767–777.

Lam, C. M. C.; Chong, C.; Wong, J. T. Y. (2001) A dinoflagelalte mutant with higher frequency of multiple fission. *Prtoplasma*, 216, 75–79.

Langer, G.; Geisen, M.; Baumann, K.-H.; KlÃ s, J.; Riebesell, U.; Thoms, S.; Young, J. R. (2006) Species-specific responses of calcifying algae to changing seawater carbonate chemistry. *Geochem. Geophys. Geosyst.*, 7, 1-12.

Langer, G.; Nehrke, G.; Probert, I.; Ly, J.; Ziveri, P. (2009) Strain-specific responses of *Emiliania huxleyi* to changing seawater carbonate chemistry. *Biogeosciences*, 6 (11), 2637–2646.

Leonardos, N.; Read, B.; Thake, B.; Young, J. R. (2009) No mechanistic dependence of photosynthesis on calcification in the coccolithophorid *Emiliania huxleyi* (Haptophyta). *J Phycol*, 45 (5), 1046–1051.

Levine, M.; Tjian, R. (2003) Transcription regulation and animal diversity. *Nature*, 424 (6945), 147–151.

Lewin, D. A.; Mellman, I. (1998) Sorting out adaptors. *Biochim. Biophys. Acta*, 1401 (2), 129–145.

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 (1), 161–176.

Llopis, J.; McCaffery, J. M.; Miyawaki, A.; Farquhar, M. G.; Tsien, R. Y. (1998) Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc Natl Acad Sci U S A*, 95 (12), 6803–6808.

Lohbeck, K. T.; Riebesell, U.; Reusch, T. B. H. (2012) Adaptive evolution of a key phytoplankton species to ocean acidification. *Nature Geosci*, advance online publication.

Lohbeck, K. T.; Riebesell, U.; Reusch, T. B. H. (2014) Gene expression changes in the coccolithophore *Emiliania huxleyi* after 500 generations of selection to ocean acidification. *Proceedings of the Royal Society B: Biological Sciences*, 281 (1786), 20140003–20140003.

Lorenzen, H. (1957) Synchrone Zellteilungen von Chlorella bei verschiedenen Licht-Dunkel-Wechseln. *Flora (Jena)*, 144, 473–496.

Lowenstam, H. (1981) Minerals formed by organisms. *Science*, 211 (4487), 1126–1131.

Lowenstam, H. A.; Weiner, S. (1989) *On Biomineralization*; Oxford University Press.

López, J. L. (2007) Applications of proteomics in marine ecology. *Mar Ecol-Prog Ser*, 332, 275–279.

Mackinder, L.; Wheeler, G.; Schroeder, D.; Dassow, von, P.; Riebesell, U.; Brownlee, C. (2011) Expression of biomineralization-related ion transport genes in Emiliania huxleyi. *Environ. Microbiol.*, 13 (12), 3250–3265.

Mackinder, L.; Wheeler, G.; Schroeder, D.; Riebesell, U.; Brownlee, C. (2010) Molecular Mechanisms Underlying Calcification in Coccolithophores. *Geomicrobiology Journal*, 27 (6), 585–595.

Mann, M.; Wilm, M. (1995) Electrospray mass spectrometry for protein characterization. *Trends in Biochemical Sciences*, 20 (6), 219–224.

Manton, I.; Leedale, G. F. (1969) Observations on the Microanatomy of *Coccolithus Pelagicus* and *Cricosphaera Carterae*, With Special Reference to the Origin and Nature of Coccoliths and Scales. *J. Mar. Biol. Ass.*, 49 (01), 1–16.

Marsh, M. E. (1986) Biomineralization in the presence of calcium-binding phosphoprotein particles. *J. Exp. Zool.*, 239 (2), 207–220.

Marsh, M. E. (1994) Polyanion-mediated mineralization assembly and reorganization of acidic polysaccharides in the Golgi system of a coccolithophorid alga during mineral deposition. *Protoplasma*.

Marsh, M. E. (2003) Regulation of CaCO3 formation in coccolithophores. *Comp Biochem Physiol B Biochem Mol Biol*, 136 (4), 743–754.

Marsh, M. E.; Chang, D. K.; King, G. C. (1992) Isolation and characterization of a novel acidic polysaccharide containing tartrate and glyoxylate residues from the mineralized scales of a unicellular coccolithophorid alga *Pleurochrysis carterae*. *J Biol Chem*, 267 (28), 20507–20512.

Marsh, M. E.; Dickinson, D. P. (1997) Polyanion-mediated mineralization — mineralization in coccolithophore (*Pleurochrysis carterae*) variants which do not express PS2, the most abundant and acidic mineral-associated polyanion in wild-type cells. *Protoplasma*, 199 (1-2), 9–17.

Marsh, M. E.; Ridall, A. L.; Azadi, P.; Duke, P. J. (2002) Galacturonomannan and Golgi-derived membrane linked to growth and shaping of biogenic calcite.

Martin, J. A.; Wang, Z. (2011) Next-generation transcriptome assembly. *Nature*, 12 (10), 671–682.

Matsuda, Y.; Nakajima, K.; Tachibana, M. (2011) Recent progresses on the genetic basis of the regulation of CO_2 acquisition systems in response to CO2 concentration. *Photosynthesis Research*, 109 (1-3), 191–203.

Mattick, J. S.; Makunin, I. V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, 15 Spec No 1, R17–29.

Más, P. (2008) Circadian clock function in *Arabidopsis thaliana*: time beyond transcription. *Trends Cell Biol.*, 18 (6), 273–281.

McKew, B. A.; Lefebvre, S. C.; Achterberg, E. P.; Metodieva, G.; Raines, C. A.; Metodiev, M. V.; Geider, R. J. (2013) Plasticity in the proteome of *Emiliania huxleyi* CCMP 1516 to extremes of light is highly targeted. *New Phytol*, 200 (1), 61–73.

McMahon, H. T.; Mills, I. G. (2004) COP and clathrin-coated vesicle budding: different pathways, common approaches. *Curr. Opin. Cell Biol.*, 16 (4), 379–391.

Medini, D.; Donati, C.; Tettelin, H.; Masignani, V.; Rappuoli, R. (2005) The microbial pan-genome. *Current Opinion in Genetics & Development*, 15 (6), 589–594.

Medlin, L. K.; Barker, G. L. A.; Campbell, L.; Green, J. C.; Hayes, P. K.; Marie, D.; Wrieden, S.; et al. (1996) Genetic characterisation of *Emiliania huxleyi* (Haptophyta). *Journal of Marine Systems*, 9 (1,Äì2), 13–31.

Meldolesi, J.; Pozzan, T. (1998) The endoplasmic reticulum Ca^{2+} store: a view from the lumen. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 23 (1), 10–14.

Menges, M.; Hennig, L.; Gruissem, W.; Murray, J. A. H. (2002) Cell cycle-regulated gene expression in Arabidopsis. *J. Biol. Chem.*, 277 (44), 41987–42002.

Milliman, J. D. (1974) *Marine carbonates. Recent sedimentary carbonates*; Springer.

Milliman, J. D. (1993) Production and accumulation of calcium carbonate in the ocean: Budget of a nonsteady state. *Global Biogeochem Cy*, 7 (4), 927–957; AGU.

Milliman, J.; Droxler, A. (1996) Neritic and pelagic carbonate sedimentation in the marine environment: ignorance is not bliss. *Geologische Rundschau*, 85 (3), 496–504.

Mitchison, J. M. (1971) *The biology of the cell cycle*; CUP Archive.

Mitton, B.; Kranias, E. G. (2003) Proteomics: Its potential in the post-genome era. *Hellenic J Cardiol*.

Montsant, A.; Allen, A. E.; Coesel, S.; Martino, A. D.; Falciatore, A.; Mangogna, M.; Siaut, M.; et al. (2007) Identification and comparative genomic analysis of signaling and regulatory components in the diatom *Thalassiosira pseudonana*. *J Phycol*, 43 (3), 585–604.

Moore, C. M.; Suggett, D. J.; Hickman, A. E.; Kim, Y.-N.; Tweddle, J. F.; Sharples, J.; Geider, R. J.; et al. (2006) Phytoplankton Photoacclimation and Photoadaptation in Response to Environmental Gradients in a Shelf Sea. *Limnol. Oceanogr.*, 51 (2), 936–949.

Moriyama, Y.; Maeda, M.; Futai, M. (1992) The role of V-ATPase in neuronal and endocrine systems. *J. Exp. Biol.*, 172, 171–178.

Morris, R. M.; Nunn, B. L.; Frazar, C.; Goodlett, D. R.; Ting, Y. S.; Rocap, G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *Isme J*, 4 (5), 673–685.

Muggli, D. L.; Harrison, P. J. (1997) Effects of iron on two oceanic phytoplankters grown in natural NE subarctic pacific seawater with no artificial chelators present. *Journal of Experimental Marine Biology and Ecology*, 212 (2), 225–237.

Muggli, D. L.; Lecourt, M.; Harrison, P. J. (1996) Effects of iron and nitrogen source on the sinking rate, physiology and metal composition of an oceanic diatom from the subarctic Pacific. *Mar Ecol-Prog Ser*, 132, 215–227.

Mullin, M. M.; Sloan, P. R.; Eppley, R. W. (1966) Relationship Between Carbon Content, Cell Volume, and Area in Phytoplankton. *Limnol. Oceanogr.*, 11 (2), 307–311.

Muse, G. W.; Gilchrist, D. A.; Nechaev, S.; Shah, R.; Parker, J. S.; Grissom, S. F.; Zeitlinger, J.; et al. (2007) RNA polymerase is poised for activation across the genome. *Nature*, 39 (12), 1507–1511.

Müller, M. N.; Antia, A. N.; LaRoche, J. (2008) Influence of cell cycle phase on calcification in the coccolithophore *Emiliania huxleyii*. *Limnol. Oceanogr.*, 53 (2), 7.

Nagai, K.; Yotsukura, N.; Ikegami, H.; Kimura, H.; Morimoto, K. (2008) Protein extraction for 2-DE from the lamina of *Ecklonia kurome* (Laminariales): recalcitrant tissue containing high levels of viscous polysaccharides. *Electrophoresis*, 29 (3), 672–681.

Nagaraj, S. H.; Gasser, R. B.; Ranganathan, S. (2006) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics*, 8 (1), 6–21.

Nash, P. D.; Opas, M.; Michalak, M. (1994) Calreticulin: not just another calcium-binding protein. *Mol Cell Biochem*, 135 (1), 71–78.

Nazareth Bastajian, H. F. B. J. A. (2013) Bck2 Acts through the MADS Box Protein Mcm1 to Activate Cell-Cycle-Regulated Genes in Budding Yeast. *PLoS Genet*, 9 (5)

Näär, A. M.; Lemon, B. D.; Tjian, R. (2001) Transcriptional cactivator complexes. *Annu. Rev. Ecol. Evol. Syst.*, 70 (1), 475–501.

Nguyen, B.; Bowers, R. M.; Wahlund, T. M.; Read, B. A. (2005) Suppressive Subtractive Hybridization of and Differences in Gene Expression Content of Calcifying and Noncalcifying Cultures of *Emiliania huxleyi* Strain 1516. *Appl. Environ. Microbiol.*, 71 (5), 2564–2575.

Nieduszynski, C. A.; Liti, G. (2011) From sequence to function: Insights from natural variation in budding yeasts. *BBA - General Subjects*, 1–8.

Nimer, N. A.; Dixon, G. K.; Merrett, M. J. (1992) Utilization of inorganic carbon by the coccolithophorid *Emiliania huxleyi* (Lohmann) Kamptner. *New Phytol*, 120 (1), 153–158.

Nimer, N. A.; Iglesias-Rodriguez, M. D.; Merrett, M. J. (1997) Bicarbonate utilization by marine phytoplankton species. *J Phycol*, 33 (4), 625–631.

Nimer, N. A.; Merrett, M. J. (1991) Calcification and utilization of inorganic carbon by the coccolithophorid *Emiliania huxleyi* Lohmann. *New Phytol*, 121 (2), 173–177.

Nimer, N. A.; Merrett, M. J. (1993) Calcification rate in *Emiliania huxleyi* Lohmann in response to light, nitrate and availability of inorganic carbon. *New Phytol*, 123 (4), 673–677.

Nishiyama, Y.; Allakhverdiev, S. I.; Murata, N. (2006) A new paradigm for the action of reactive oxygen species in the photoinhibition of photosystem II. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 1757 (7), 742–749.

Nookaew, I.; Papini, M.; Pornputtapong, N.; Scalcinati, G.; Fagerberg, L.; Uhlen, M.; Nielsen, J. (2012) A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 40 (20), 10084–10097.

Nunez, R. (2001) DNA measurement and cell cycle analysis by flow cytometry. *Current issues in molecular biology*, 3, 67–70.

Nunn, B. L.; Aker, J. R.; Shaffer, S. A.; Tsai, S.; Strzepek, R. F.; Boyd, P. W.; Freeman, T. L.; et al. (2009) Deciphering diatom biochemical pathways via whole-cell proteomics. *Aquat. Microb. Ecol.*, 55 (3), 241–253.

O'Connor, C. D. C.; Adams, P. P.; Alefounder, P. P.; Farris, M. M.; Kinsella, N. N.; Li, Y. Y.; Payot, S. S.; et al. (2000) The analysis of microbial proteomes: strategies and data exploitation. *Electrophoresis*, 21 (6), 1178–1186.

O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem. 1975, 250:4007-4021.*

Oehlen, L. J.; McKinney, J. D.; Cross, F. R. (1996) Ste12 and Mcm1 regulate cell cycle-dependent transcription of FAR1. *Mol. Cell. Biol.*, 16 (6), 2830–2837.

Oliva, A.; Rosebrock, A.; Ferrezuelo, F.; Pyne, S.; Chen, H.; Skiena, S.; Futcher, B.; et al.

276

(2005) The Cell Cycle–Regulated Genes of *Schizosaccharomyces pombe*. *PLoS Biol*, 3 (7), e225.

Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell*, 127 (3), 635–648.

Olson, R. J.; Vaulot, D.; Chisholm, S. W. (n.d.) Effects of Environmental Stresses on the Cell Cycle of Two Marine Phytoplankton Species.

Onodera, Y.; Haag, J. R.; Ream, T.; Costa Nunes, P.; Pontes, O.; Pikaard, C. S. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell*, 120 (5), 613–622.

Orci, L.; Malhotra, V.; Amherdt; Serafini, T.; Rothman, J. E. (1986) Dissection of a single round of vesicular transport: sequential intermediates for intercisternal movement in the Golgi stack. *Cell*, 56 (3), 357–368.

Orr, J. C.; Fabry, V. J.; Aumont, O.; Bopp, L.; Doney, S. C.; Feely, R. A.; Gnanadesikan, A.; et al. (2005) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature*, 437 (7059), 681–686.

Outka, D. E.; Williams, D. C. (1971) Sequential Coccolith Morphogenesis in Hymenomonas carterae. *J Eukaryotic Microbiology*, 18 (2), 285–297.

Outka, D. E.; Williams, D. C. (1973) Green Algae - Carole A. Lembi - Google Books. *Green algae*.

Paasche, E. (1962) Coccolith Formation. *Nature*, 193 (4820), 1094–1095.

Paasche, E. (1963) The Adaptation of the Carbon-14 Method for the Measurement of Coccolith Production in *Coccolithus huxleyi*. *Physiol Plant*, 16 (1), 186–200; Blackwell Publishing Ltd.

Paasche, E. (1965) The Effect of 3-(p-Chlorophenyl)-1,1-Dimethylurea (CMU) on Photosynthesis and Light-Dependent Cocoolith Formation in *Coccolithus huxleyi*. *Physiol Plant*, 18 (1), 138–145.

Paasche, E. (1966) Adjustment to Light and Dark Rates of Coccolith Formation. *Physiol Plant*, 19 (2), 271–278.

Paasche, E. (1967) Marine Plankton Algae Grown with Light-Dark Cycles. I. Coccolithus huxleyi. *Physiol Plant*, 20 (4), 946–956

Paasche, E. (1998) Roles of nitrogen and phosphorus in coccolith formation in *Emiliania huxleyi* (Prymnesiophyceae). *European Journal of Phycology*, 33 (1), 33–42.

Paasche, E. (2002) A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions. *Phycologia*, 40 (6), 503–529

Paasche, E.; Brubak, S. (1994) Enhanced calcification in the coccolithophorid *Emiliania huxleyi* (Haptophyceae) under phosphorus limitation. *Phycologia*, 33 (5), 324–330.

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

Patterson, S. D.; Aebersold, R. (1995) Mass spectrometric approaches for the identification of gel-separated proteins. *Electrophoresis*, 16 (10), 1791–1814.

Perales, M.; Más, P. (2007) A functional link between rhythmic changes in chromatin structure and the Arabidopsis biological clock. *Plant Cell*, 19 (7), 2111–2123.

Pockley, A. G. A. (2003) Heat shock proteins as regulators of the immune response. *Lancet*, 362 (9382), 469–476.

Pokhariyal, R.; Basnet, B.; Gupta, A. K.; Sehrawat, U.; Jain, S. K.; Hariprasad, G. (2014) Effect of Lipid on Gel Based Electrophoretic Proteomic Experiments. *Journal of Proteins & Proteomics*, 5 (2).

Post, A. F.; Eijgenraam, F.; Mur, L. R. (1985) Influence of light period length on photosynthesis and synchronous growth of the green alga *Scenedesmus protuberans*. *British Phycological Journal*, 20 (4), 391–397.

Poulton, A. J.; Charalampopoulou, A.; Young, J. R.; Tarran, G. A.; Lucas, M. I.; Quartly, G. D. (2010) Coccolithophore dynamics in non-bloom conditions during late summer in the central Iceland Basin (July-August 2007). *Limnol. Oceanogr.*, 55 (4), 1601–1613.

Price, L. L.; Yin, K.; Harrison, P. J. (1998) Influence of continuous light and L:D cycles on the growth and chemical composition of Prymnesiophyceae including coccolithophores. *Journal of Experimental Marine Biology and Ecology*, 223 (2), 223–234.

Price, M. N.; Dehal, P. S.; Arkin, A. P. (2009) FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Mol. Biol. Evol.*, 26 (7), 1641–1650.

Quinn, P.; Bowers, R. M.; Zhang, X.; Wahlund, T. M.; Fanelli, M. A.; Olszova, D.; Read, B. A. (2006) cDNA Microarrays as a Tool for Identification of Biomineralization Proteins in the Coccolithophorid *Emiliania huxleyi* (Haptophyta). *Appl. Environ. Microbiol.*, 72 (8), 5512–5526.

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

Ramos-Silva, P.; Kaandorp, J.; Huisman, L.; Marie, B.; Zanella-Cléon, I.; Guichard, N.; Miller, D. J.; et al. (2013) The Skeletal Proteome of the Coral *Acropora millepora*: The Evolution of Calcification by Co-Option and Domain Shuffling. *Mol. Biol. Evol. 30 (9) pp. 2099-2112*

Rasmussen, E. B.; Lis, J. T. (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc. Natl. Acad. Sci. U.S.A.*, 90 (17), 7923–7927.

Raven, J.; Caldeira, K.; Elderfield, H.; Hoegh-Guldberg, O.; Liss, P.; Riebesell, U.; Shepherd, J.; et al. (2005) Ocean acidification due to increasing atmospheric carbon dioxide. *The Royal Society*.

Rayko, E.; Maumus, F.; Maheswari, U.; Jabbari, K.; Bowler, C. (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *J Phycol*, 188 (1), 52–66.

Read, B. A.; Kegel, J.; Klute, M. J.; Kuo, A.; Lefebvre, S. C.; Maumus, F.; Mayer, C.; et al. (2013) Pan genome of the phytoplankton Emiliania underpins its global distribution. *Nature*, 499 (7457), 209–213; Nature Publishing Group.

Reynaud, S.; Leclercq, N.; Romaine-Lioud, S.; Ferrier-PagËs, C.; Jaubert, J.; Gattuso, J.-P. (2003) Interacting effects of CO2 partial pressure and temperature on photosynthesis and calcification in a scleractinian coral. *Global Change Biol*, 9 (11), 1660–1668.

Richier, S.; Kerros, M. E.; de Vargas, C.; Haramaty, L.; Falkowski, P. G.; Gattuso, J. P. (2009) Light-dependent transcriptional regulation of genes of biogeochemical interest in the diploid and haploid life cycle stages of *Emiliania huxleyi*. *Appl. Environ. Microbiol.*, 75 (10), 3366–3369.

Riebesell, U.; Koertzinger, A.; Oschlies, A. (2009) Sensitivities of marine carbon fluxes to ocean change. *Proc Natl Acad Sci U S A*, 106 (49), 20602–20609.

Riebesell, U.; Zondervan, I.; Rost, B.; Tortell, P. D.; Zeebe, R. E.; Morel, F. M. M. (2000) Reduced calcification of marine plankton in response to increased atmospheric CO2. *Nature*, 407 (6802), 364–367.

Riegman, R.; Stolte, W.; Noordeloos, A. A. M.; Slezak, D. (2000) Nutrient Uptake and Alkaline Phosphatase (Ec 3:1:3:1) Activity of Emiliania Huxleyi (PRYMNESIOPHYCEAE) During Growth Under N and P Limitation in Continuous Cultures. *J Phycol*, 36 (1), 87–96.

Robbens, S.; Khadaroo, B.; Camasses, A.; Derelle, E.; Ferraz, C.; Inzé, D.; Van de Peer, Y.; et al. (2005) Genome-wide analysis of core cell cycle genes in the unicellular green alga *Ostreococcus tauri*. *Mol. Biol. Evol.*, 22 (3), 589–597.

Robertson, J. E.; Robinson, C.; Turner, D. R.; Holligan, P.; Watson, A. J.; Boyd, P.; Fernandez, E.; et al. (1994) The impact of a coccolithophore bloom on oceanic carbon uptake in the northeast Atlantic during summer 1991. *Deep-Sea Res Pt I*, 41 (2), 297–314.

Robinson, A.; van Oijen, A. M. (2013) Bacterial replication, transcription and translation: mechanistic insights from single-molecule biochemical studies. *Nat. Rev. Microbiol.*, 11 (5), 303–315.

Ronning, O. W.; Lindmo, T. (1983) Progress through G1 and S in relation to net protein accumulation in human NHIK 3025 cells. *Experimental Cell Research*, 144 (1), 171–179.

Rose, S. L.; Fulton, J. M.; Brown, C. M.; Natale, F.; Van Mooy, B. A. S.; Bidle, K. D. (2014) Isolation and characterization of lipid rafts in *Emiliania huxleyi*: a role for membrane microdomains in host-virus interactions. *Environ. Microbiol.*, 16 (4), 1150–1166.

Rosenwasser, S.; Graff van Creveld, S.; Schatz, D.; Malitsky, S.; Tzfadia, O.; Aharoni, A.; Levin, Y.; et al. (2014) Mapping the diatom redox-sensitive proteome provides insight into response to nitrogen stress in the marine environment. *Proc. Natl. Acad. Sci. U.S.A.*, 111 (7), 2740–2745.

Ross, P. L. (2004) Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics*, 3 (12), 1154–1169.

Rost, B.; Riebesell, U.; Burkhardt, S.; Sültemeyer, D. (2003) Carbon Acquisition of Bloom-Forming Marine Phytoplankton. *Limnol. Oceanogr.*, 48 (1), 55–67.

Rustici, G.; Mata, J.; Kivinen, K.; Lió, P.; Penkett, C. J.; Burns, G.; Hayles, J.; et al. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat Genet*, 36 (8), 809–817.

Rymarquis, L. A.; Kastenmayer, J. P.; Hüttenhofer, A. G.; Green, P. J. (2008) Diamonds in the rough: mRNA-like non-coding RNAs. *Trends in Plant Science*, 13 (7), 329–334.

Sabine, C. L.; Feely, R. A.; Gruber, N.; Key, R. M.; Lee, K.; Bullister, J. L.; Wanninkhof, R.; et al. (2004) The Oceanic Sink for Anthropogenic CO2. *Science*, 305 (5682), 367–371.

Saez, A. G.; Probert, I.; Young, J. R.; Edvardsen, B.; Eikrem, W.; Medlin, L. K. (2004) A review of the phylogeny of the Haptophyta. In "Coccolithophores: from Molecular Processes to Global Impact". *Coccolithophores-from molecular processes to global impact . -Berlin (Springer)*.

Sanders, D.; Pelloux, J.; Brownlee, C.; Harper, J. F. (2002) Calcium at the crossroads of signaling. *Plant Cell*, 14, 401–417.

Sanger, F.; Nicklen, S.; Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, 74 (12), 5463–5467.

Schroeder, D. C.; Biggi, G. F.; Hall, M.; Davy, J.; Martínez, J. M.; Richardson, A. J.; Malin, G.; et al. (2005) A genetic marker to separate *Emiliania huxleyi* (Prymnesiophyceae) morphotypes. *J Phycol*, 41 (4), 874–879.

Schulze, W. X.; Usadel, B. (2010) Quantitation in Mass-Spectrometry-Based Proteomics. *Annu Rev Plant Biol*, 61 (1), 491–516.

Sciandra, A.; Harlay, J.; Lefèvre, D.; Lemée, R.; Rimmelin, P.; Denis, M.; Gattuso, J.-P. (2003) Response of coccolithophorid *Emiliania huxleyi* to elevated partial pressure of CO2 under nitrogen limitation. *Mar Ecol-Prog Ser*, 261.

Sekino, K.; Shiraiwa, Y. (1994) Accumulation and Utilization of Dissolved Inorganic Carbon by a Marine Unicellular Coccolithophorid, *Emiliania huxleyi*. *Plant and Cell Physiology*, 35 (3), 353–361.

Sekino, K.; Shiraiwa, Y. (1996) Evidence for the Involvement of Mitochondrial Respiration in Calcification in a Marine Coccolithophorid, *Emiliania huxleyi*. *Plant and Cell Physiology*, 37 (7), 1030–1033.

Shatkin, A. J. (1976) Capping of eucaryotic mRNAs. *Cell*, 9 (4), 645–653; Cell Press.

Shi, D.; Xu, Y.; Morel, F. M. M. (2009) Effects of the pH/pCO2 control method on medium chemistry and phytoplankton growth. *Biogeosciences*, 6 (7), 1199–1207.

Shiraiwa, Y. (2003) Physiological regulation of carbon fixation in the photosynthesis and calcification of coccolithophorids. *Comp Biochem Physiol B Biochem Mol Biol*, 136 (4), 775–783.

Shiraki, T.; Kondo, S.; Katayama, S.; Waki, K.; Kasukawa, T.; Kawaji, H.; Kodzius, R.; et al. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*, 100 (26), 15776–15781.

Siegenthaler, U.; Sarmiento, J. L. (1993) Atmospheric carbon dioxide and the ocean. *Nature*, 119–125.

Skipp, P.; Robinson, J.; O'Connor, C. D.; Clarke, I. N. (2005) Shotgun proteomic analysis of *Chlamydia trachomatis*. *Proteomics*, 5 (6), 1558–1573.

Snustad, D. P.; Simmons, M. J. (2000) *Principles of Genetics*; John Wiley and Sons, Inc.

Soto, A. R.; Zheng, H.; Shoemaker, D.; Rodriguez, J.; Read, B. A.; Wahlund, T. M. (2006) Identification and Preliminary Characterization of Two cDNAs Encoding Unique Carbonic Anhydrases from the Marine Alga *Emiliania huxleyi*. *Appl Environ Microbiol*, 72 (8), 5500–5511.

Spudish, J. L.; Sager, R. (1980) Regulation of the Chlamydomonas cell cycle by light and dark. *J. Cell Biol.*, 85, 136–145.

Stadler, R.; Wolf, K.; Hilgarth, C.; Tanner, W.; Sauer, N. (1995) Subcellular localization of the inducible Chlorella HUP1 monosaccharide-H+ symporter and cloning of a co-induced galactose-H+ symporter. *Plant Physiol.*, 107 (1), 33–41; Am Soc Plant Biol.

Sterling, D.; Alvarez, B. V.; Casey, J. R. (2002) The extracellular component of a transport metabolon. Extracellular loop 4 of the human AE1 Cl-/HCO3- exchanger binds carbonic anhydrase IV. *J Biol Chem*, 277 (28), 25239–25246.

Suffrian, K.; Schulz, K. G.; Gutowska, M. A.; Riebesell, U.; Bleich, M. (2011) Cellular pH measurements in Emiliania huxleyi reveal pronounced membrane proton permeability. *New Phytol*, 190 (3), 595–608.

Suggett, D. J.; Le Floc'H, E.; Harris, G. N.; Leonardos, N.; Geider, R. J. (2007) Different strategies of photoacclimation by two strains of *Emiliania huxleyi* (Haptophyta) 1. *J Phycol*, 43 (6), 1209–1222.

Suggett, D. J.; Moore, C. M.; Hickman, A. E.; Geider, R. J. (2009) Interpretation of fast repetition rate (FRR) fluorescence: signatures of phytoplankton community structure versus physiological state. *Mar Ecol-Prog Ser*, 376, 1–19.

Takei, K.; Haucke, V. (2001) Clathrin-mediated endocytosis: membrane factors pull the trigger. *Trends Cell Biol.*, 11 (9), 385–391.

Tambutte, E.; Tambutte, S.; Segonds, N.; Zoccola, D.; Venn, A.; Erez, J.; Allemand, D. (2011) Calcein labelling and electrophysiology: insights on coral tissue permeability and calcification. *Proceedings of the Royal Society B: Biological Sciences*, 279 (1726), 19–27.

Tambutté, S.; Holcomb, M.; Ferrier-PagËs, C.; Reynaud, S.; Tambutté, É.; Zoccola, D.; Allemand, D. (2011) Coral biomineralization: from the gene to the environment. *Journal of Experimental Marine Biology and Ecology*, 408 (1), 58–78.

Tan, S.; Conaway, R. C.; Conaway, J. W. (1995) Dissection of transcription factor TFIIF functional domains required for initiation and elongation. *Proc. Natl. Acad. Sci. U.S.A.*, 92 (13), 6042–6046.

Taylor, A. R.; Chrachri, A.; Wheeler, G.; Goddard, H.; Brownlee, C. (2011) A voltage-gated H+ channel underlying pH homeostasis in calcifying coccolithophores. *PLoS Biol*, 9 (6), e1001085.

Taylor, A. R.; Russell, M. A.; Harper, G. M.; Collins, T. F. T.; Brownlee, C. (2007) Dynamics of formation and secretion of heterococcoliths by *Coccolithus pelagicus ssp. braarudii*. *European Journal of Phycology*, 42 (2), 125–136.

Thierstein, H. R.; Geitzenauer, K. R.; Molfino, B.; Shackleton, N. J. (1977) Global synchroneity of late Quaternary coccolith datum levels Validation by oxygen isotopes. *Geology*, 5 (7), 400–404.

Thomas, M. C.; Chiang, C.-M. (2006) The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.*, 41 (3), 105–178.

Tian, Q.; Stepaniants, S. B.; Mao, M.; Weng, L.; Feetham, M. C.; Doyle, M. J.; Yi, E. C.; et al. (2004) Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Molecular & Cellular Proteomics*, 3 (10), 960–969.

Tirichine, L.; Bowler, C. (2011) Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant J*, 66 (1), 45–57.

Tomanek, L.; Zuzow, M. J. (2010) The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress. *Journal of Experimental Biology*, 213 (20), 3559–3574.

Tomanek, L.; Zuzow, M. J.; Ivanina, A. V.; Beniash, E.; Sokolova, I. M. (2011) Proteomic response to elevated pCO$_2$ levles in eastern oysters,*Crassostrea virginica*: eveidence for oxidative stress. *Journal of Experimental Biology*, 214, 1836–1844.

Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D. R.; Pimentel, H.; et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature*, 7 (3), 562–578.

Trimborn, S.; Langer, G.; Rost, B. (2007) Effect of varying calcium concentrations and light intensities on calcification and photosynthesis in *Emiliania huxleyi. Limnol. Oceanogr.*, 52 (5), 2285–2293.

Tyrell, T.; Merico, A. (2004) *Emiliania huxleyi*: bloom observations and the conditions that induce them. In: *Coccolithophores: from molecular processes to global impact.* (H. R. Thierstein & J. R. Young, eds.).

Tyrrell, T.; Young, J. R. (2009) *COCCOLITHOPHORES*. (J. H. Steele & K. K. Turekian, Eds.), 2nd ed., 606–614; Elsevier.

Urizar, N. L.; Dowhan, D. H.; Moore, D. D. (2000) The farnesoid X-activated receptor mediates bile acid activation of phospholipid transfer protein gene expression. *J. Biol. Chem.*, 275 (50), 39313–39317.

Van Bleijswijk, J. D. L.; Kempers, R. S.; Veldhuis, M. J.; 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 (2), 230–241.

van Bleijswijk, JDL; Veldhuis, M. (1995) In situ gross growth rates of *Emiliania huxleyi* in enclosures with different phosphate loadings revealed by diel changes in DNA content. *Mar Ecol-Prog Ser*, 121, 271–277.

van Bleijswijk, Judith; Wal, P.; Kempers, R.; Veldhuis, M.; Young, J. R.; Muyzer, G.; Vrind-de Jong, E.; et al. (1991) Distribution of two types of *Emiliania huxleyii* (Prymnesiophyzeae) in the northeast Atlantic region as determined by immunofluorescence and coccolith morphology. *J Phycol*, 27 (5), 566–570.

van den Heuvel, S.; Harlow, E. (1993) Distinct roles for cyclin-dependent kinases in cell cycle control. *Science*, 262 (5142), 2050–2054.

van der Wal, P.; de Jong, E. W.; Westbroek, P.; de Bruijn, W. C.; Mulder-Stapel, A. A. (1983) Polysaccharide localization, coccolith formation, and golgi dynamics in the coccolithophorid *Hymenomonas carterae. Journal of Ultrastructure Research*, 85 (2), 139–158;

van der Wal, P.; de Vrind, J. P. M.; de Vrind-de Jong, E. W.; Borman, A. H. (1987) Incompleteness of the coccosphere as a possible stimulus for coccolith formation in *Pleurochrysis carterae* (Prymnesiophyceae). *J Phycol*, 23 (1), 218–221.

van der Wal, P.; Jong, E. W.; Westbroek, P.; Bruijn, W. C.; Mulder-Stapel, A. A. (1983) Ultrastructural polysaccharide localization in calcifying and naked cells of the coccolithophorid *Emiliania huxleyi*, 118 (2), 157–168.

Van Emburg, P. R. (1989, September 8) Coccolith formation in *Emiliania huxleyi*.

Van Emburg, P. R.; de Jong, E. W.; Daems, W. T. (1986) Immunochemical localization of a polysaccharide from biomineral structures (coccoliths) of *Emiliania huxleyi*. *Journal of Ultrastructure and Molecular Structure Research*, 94 (3), 246–259.

Vandenbussche, F.; Verbelen, J. P.; Van Der Straeten, D. (2005) Of light and length: regulation of hypocotyl growth in Arabidopsis. *Bioessays*, 27 (3), 275–284; Wiley Online Library.

VanGuilder, H. D.; Vrana, K. E.; Freeman, W. M. (2008) Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, 44 (5), 619–626.

Velculescu, V. E.; Zhang, L.; Vogelstein, B.; Kinzler, K. W. (1995) Serial analysis of gene expression. *Science*, 270 (5235), 484–487.

Venn, A.; Quinn, J.; Jones, R.; Bodnar, A. (2009) P-glycoprotein (multi-xenobiotic resistance) and heat shock protein gene expression in the reef coral *Montastraea franksi* in response to environmental toxicants. *Aquat. Toxicol.*, 93 (4), 188–195.

Venn, A.; Tambutté, É.; Holcomb, M.; Allemand, D.; Tambutté, S. (2011) Live Tissue Imaging Shows Reef Corals Elevate pH under Their Calcifying Tissue Relative to Seawater. *PLoS ONE*, 6 (5), e2001.

Verardo, D. J.; Froelich, P. N.; McIntyre, A. (1990) Determination of organic carbon and nitrogen in marine sediments using the Carlo Erba NA-1500 analyzer. *Deep Sea Research Part A. Oceanographic Research Papers*, 37 (1), 157–165.

Verret, F.; Wheeler, G.; Taylor, A. R.; Farnham, G.; Brownlee, C. (2010) Calcium channels in photosynthetic eukaryotes: implications for evolution of calcium-based signalling. *New Phytol*, 187 (1), 23–43.

Vierling, E. (1991) The roles of heat shock proteins in plants. *Annu Rev Plant Biol*, 42 (1), 579–620.

Vince, J. W.; Reithmeier, R. A. (2000) Identification of the carbonic anhydrase II binding site in the Cl(-)/HCO(3)(-) anion exchanger AE1. *Biochemistry*, 39 (18), 5527–5533.

Vliegenthart, J.; van Halbeek, H.; Dorland, L. (1981) The aplicability of 50-Mhz high-resolution 1h-NMR spectroscopyfor the structure determination of carbohydrates derived from glycoproteins. *Pure and Applied Chemistry*, 53, 45–77.

Volk, T.; Hoffert, M. I. (1985) Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO_2 changes. In: *The Carbon Cycle and Atmospheric CO2: Natural Variations Archean to Present*, Vol. 32, 99–110.

Volkman, J. K.; Eglinton, G.; Corner, E. D. S.; Sargent, J. R. (1980) Novel unsaturated straight-chain C37 C39 methyl and ethyl ketones in marine sediments and a coccolithophore *Emiliania huxleyi. Palaeogeography, Palaeoclimatology, Palaeoecology*, 12, 219–227.

Wahlund, T. M.; Hadaegh, A. R.; Clark, R.; Nguyen, B.; Fanelli, M.; Read, B. A. (2004) Analysis of Expressed Sequence Tags from Calcifying Cells of Marine Coccolithophorid (*Emiliania huxleyi*). *Mar Biotechnol*, 6 (3), 278–290.

Wahlund, T. M.; Zhang, X.; Read, B. A. (2004) Expressed sequence tag profiles from calcifying and non-calcifying cultures of *Emiliania huxleyi. micropalaeontology*, 50 (Suppl 1), 145–155.

Wang, W.; Vinocur, B.; Shoseyov, O.; Altman, A. (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, 9 (5), 244–252; Elsevier.

Wang, Z.; Gerstein, M.; Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10 (1), 57–63.

Warnecke, F.; Hess, M. (2009) A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts. *Journal of Biotechnology*, 142 (1), 91–95.

Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; et al. (1995) Progress with gene-product mapping of the Mollicutes:Mycoplasma genitalium. *J Phycol*, 16 (1), 1090–1094.

Weiner, S.; Dove, P. M.; De Yoreo, J. J. (2003) *Biomineralization*. (J. J. Rosso, Ed.)*Reviews in Mineralogy and Geochemistry* Reviews in Mineralogy and Geochemistry, Vol. 54.

Weiner, S.; Levi-Kalisman, Y.; Raz, S.; Addadi, L. (2003) Biologically Formed Amorphous Calcium Carbonate. *Connective Tissue Research*, 44 (1), 214–218.

Weiner, S.; Traub, W.; Parker, S. B. (1984) Macromolecules in mollusc shells and their functions in biomineralization. *Philosophical Transactions of the Royal Society London B Biological Science*, 304, 425–434.

Weiss, I. M.; Tuross, N.; Addadi, L.; Weiner, S. (2002) Mollusc larval shell formation: amorphous calcium carbonate is a precursor phase for aragonite. *Journal of Experimental Zoology*, 293 (5), 478–491.

Welschmeyer, N. A. (1994) Fluorometric Analysis of Chlorophyll a in the Presence of Chlorophyll b and Pheopigments. *Limnol. Oceanogr.*, 39 (8), 1985–1992.

Westbroek, P.; Young, J. R.; Linschooten, K. (1989) Coccolith Production (Biomineralization) in the Marine Alga *Emiliania huxleyi. Journal of Eukaryotic Microbiology*, 36 (4), 368–373.

Wijnen, H.; Young, M. W. (2006) Interplay of circadian clocks and metabolic rhythms. Annual Review of Genetics, 40, 409–448.

Wilbur, K. M.; Watabe, N. (1963) Experimental Studies on Calcification in Molluscs and the Alga Coccolithus Huxleyi. *J Phycol*, 109 (1), 82–112.

Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J.-C.; Yan, J. X.; et al. (1996) From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Arnino Acid Analysis. *Nat. Biotechnol.*, 14 (1), 61–65.

Wittwer, C. T.; Herrmann, M. G.; Moss, A. A.; Rasmussen, R. P. (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques*, 22 (1), 130–1–134–8.

Woeller, F. H. (1961) Liquid scintillation counting of C-14-labelled CO2 with phenethylamine. *Anal Biochem*, 2, 508–511.

Wurch, L. L.; Bertrand, E. M.; Saito, M. A.; Van Mooy, B. A. S.; Dyhrman, S. T. (2011) Proteome Changes Driven by Phosphorus Deficiency and Recovery in the Brown Tide-Forming Alga *Aureococcus anophagefferens*. *PLoS ONE*, 6 (12), e28949;

Yalovsky, S.; Paulsen, H.; Michaeli, D.; Chitnis, P. R.; Nechushtai, R. (1992) Involvement of a chloroplast HSP70 heat shock protein in the integration of a protein (light-harvesting complex protein precursor) into the thylakoid membrane. *Proc. Natl. Acad. Sci. U.S.A.*, 89 (12), 5616–5619.

Yoo, S. H.; Lewis, M. S. (1992) Effects of pH and Ca2+ on monomer-dimer and monomer-tetramer equilibria of chromogranin A. *J Biol Chem*.

Yoon, H. S.; Hackett, J. D.; Ciniglia, C.; Pinto, G.; Bhattacharya, D. (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.*, 21 (5), 809–818.

Young, J. N. (1994) Functions of coccoliths. In: *Coccolithophores* (A. Winter & W. Siesser, eds.), 63–82; Cambridge University Press.

Young, J. R.; Andruleit, H.; Probert, I. (2009) Coccolith function and morphogenesis: insights from appendage-bearing coccolithophores of the family Syracosphaeraceae (Haptophyta). *J Phycol*, 45 (1), 213–226.

Young, J. R.; Davis, S. A.; Bown, P. R.; Mann, S. (1999) Coccolith Ultrastructure and Biomineralisation, 126 (3), 195–215.

Young, J. R.; Didymus, J. M.; Brown, P. R.; Prins, B.; Mann, S. (1992) Crystal assembly and phylogenetic evolution in heterococcoliths. *Nature*, 356 (6369), 516–518.

Young, J. R.; Henriksen, K. (2003) Biomineralization Within Vesicles: The Calcite of Coccoliths. *Reviews in Mineralogy and Geochemistry*, 54 (1), 189–215.

Zeebe, R.; Ridgwell, A. (2011) Past changes in ocean carbonate chemistry. In: *Ocean Acidification* (J. P. Gattuso, ed.), 21–40; Oxford University Press.

Zondervan, I. (2007) The effects of light, macronutrients, trace metals and CO2 on the production of calcium carbonate and organic carbon in coccolithophores, A review. *Deep-Sea Res. Part II-Top. Stud. Oceanogr.*, 54 (5), 521–537.

Zondervan, I.; Rost, B.; Riebesell, U. (2002) Effect of CO_2 concentration on the PIC/POC ratio in the coccolithophore Emiliania huxleyi grown under light-limiting conditions and different daylengths. *Journal of Experimental Marine Biology and Ecology*, 272 (1), 55–70.

Zondervan, I.; Zeebe, R. E.; Rost, B.; Riebesell, U. (2001) Decreasing marine biogenic calcification: a negative feedback on rising atmospheric pCO2. *Global Biogeochem Cy*, 15 (2), 507–516.

Bibliography

Jones, B., 2010. The proteome of *Emiliania huxleyi* and its responses to CO_2 induced ocean acidification. University of Southampton, August 2010

Rehm, H. (2006) *Protein biochemistry and proteomics*; Elsevier.

Snustad, D.P. & Simmons, M.J., 2000. *Principles of Genetics*, John Wiley and Sons, Inc.

Tyrrell, T. & Young, J.R., 2009. *Coccolithophores* 2nd ed. J. H. Steele & K. K. Turekian, eds., Elsevier.

Weiner, S., Dove, P.M. & De Yoreo, J.J., 2003a. *Biomineralization* J. J. Rosso, ed., Mineralogical Society of America.

Zachleder, V., Bisova, K., Vitova, M., 2016. The Cell Cycle of Microalgae. In: *The Physiology of Microalgae*. M.A. Borowitzka, J. Beardall, J.A. Raven, eds., Springer. pp. 3–46.