Response of *Thalassiosira oceanica* and natural microbial communities to ocean acidification: a meta-omics comparison from unialgal cultures to mesocosms

Dissertation

in fulfilment of the requirements
for Dr. rer. nat.

of the Faculty of Mathematics and Natural Sciences
at the Christian-Albrechts University of Kiel

submitted by

Alexandra-Sophie Roy

Kiel, November 2013

First referee:	Prof. Dr. Julie LaRoche
Second referee:	Prof. Dr. Philip Rosenstiel
Date of oral examination:	16th December 2013
Approved for publication:	16th December 2013

Signed (Title, first and surnames), Dean:

Contents

1. Zusammenfassung

2. Summary

1. Zusammenfassung

Die Produktion von anthropogem Kohlenstoffdioxid (CO_2) z.B. durch die Verbrennung fossiler Energieträger steigt jährlich an. Die Konsequenzen sind aber nach wie vor weitestgehend unbekannt. Ein direkter Effekt der steigenden atmosphärischen CO_2 Konzentrationen sind die erhöhte Aufnahme durch Ozeane. CO_2 dissoziiert in Meerwasser, was schließlich zu einem Anstieg der Bikarbonat (HCO_3^-) und Protonen (H^+) Konzentrationen führt, während gleichzeitig die Konzentration des Karbonates (CO_3^{2-}) sinkt. Der Anstieg von H^+ Ionen lässt sich auch in der Abnahme des ph-Wertes im Oberflächenwasser sehen und wird Ozeanversauerung (OV) genannt. Die vorliegende Dissertation untersuchte mittels Hochdurchsatz-Sequenziermethoden die Auswirkungen der OV auf das Transkriptom von *Thalassiosira oceanica*, einer ökologisch wichtigen Kieselalge, und auf das Metatranscriptome einer natürlichen mikrobiellen Artengemeinschaft in der Arktis in Spitzbergen.

Verschiedene Gene wurden als Funktion der CO_2^- Konzentration unterschiedlich stark exprimiert. Die ursprünglich in *Emiliania huxleyi* von erhöhten pCO_2 Konzentrationen beeinflussten Gene waren jedoch nicht signifikant betroffen. Obwohl kalzifizierende Organismen wie *E. huxleyi* im Allgemeinen für ihre Empfindlichkeit gegenüber OV bekannt sind, scheinen Kieselalgen wie *T. oceanica* gar nicht oder sogar vorteilhaft von erhöhtem CO_2 Konzentrationen beeinflusst zu sein.

Hochdurchsatz-Sequenziermethoden wurden auch in zwei Experimenten genutzt, die Teil der großangelegten *in situ* Studie mit pelagischen Mesokosmen im Rahmen der EPOCA 2010 arktischen Kampagne waren. Diese Studien umfasste die erste kombinierte metagenomische und metatranskriptomische Charakterisierung einer arktischen mikrobiellen Artengemeinschaft. Die 16S rRNA Gen Ampliconsequenzierung wurde genutzt um den Effekt der OV auf die Struktur der Artengemeinschaft zu erfassen. Die Studie zeigte, dass CO_2 nicht eine der strukturierenden Hauptvariablen war, sondern vielmehr Zeit, Nährstoffzugabe und Wasserherkunft. Außerdem verdeutlichte die Studie, dass CO_2 nur auf 15 seltene Taxa einen signifikanten Einfluss hatte. Um zusätzlich die funktionelle Diversität der mikrobiellen Artengemeinschaft zu charakterisieren, wurde das Metagenome und das Metatranskriptome einer Kontrollprobe und einer Probe aus einem erhöhten CO_2-

Mesokosmus an fünf Zeitpunkten, die über die Dauer des gesamten Experiments verteilt waren, sequenziert. Die phylogenetische Analyse zeigte, wie bereits früher entdeckt, dass die Struktur der mikrobiellen Artengemeinschaft in ihrer Gesamtheit nicht stark von den CO_2 Konzentrationen beeinflusst wurde, sondern vielmehr signifikant über den Verlauf des Experiments variierte. Die Analyse der Metatranskriptome ließ erkennen, dass wieder nur eine kleine Anzahl der Arten und Gene signifikant vom CO_2 beeinflusst wurde. In ihrer Gesamtheit machte die Mesokosmenstudie deutlich, dass die Effekte der OV auf Struktur und Metabolismen der arktischen mikrobiellen Artengemeinschaft in Spitzbergen sich auf seltene Arten und spezialisierte funktionale Gene beschränkte.

Die in dieser Dissertation präsentierten Ergebnisse erweitern das Wissen des Effekts der OV auf die Kieselalge *T. oceanica*, wie auch auf die arktische mikrobielle Artengemeinschaft in Spitzbergen. Sie unterstreicht die Bedeutung der Hochdurchsatz-Sequenzierungsanalysen für den Nachweis des Effektes des Klimawandels auf unsere Ozeane.

2. Summary

The production of anthropogenic carbon dioxide (CO_2) via fossil fuels combustion increases yearly and the consequences on the environment are still to some extend unknown. The most direct effect of increased atmospheric CO_2 is its increased absorption by the oceans. The result of this absorption is the complex dissociation of CO_2 in seawater which ultimately creates an increase in bicarbonate (HCO_3^-) and in proton (H^+) concentration while it decreases carbonate ion (CO_3^{2-}) concentration. The increase of H^+ concretely decreases the oceans' surface pH and is termed ocean acidification (OA). This doctoral dissertation investigated the effect of OA on *Thalassiosira oceanica*, an ecologically relevant diatom highly resistant to iron limitation, as well as on a natural Arctic microbial community. It further compared the results of unialgal cultures to mesocosms.

High-throughput sequencing was used to investigate how the transcriptome of unialgal cultures of *T. oceanica* were affected by elevated pCO_2 and how it will affect the global gene expression in this species. Multiple genes were differentially expressed as a function of CO_2; however, the genes previously identified in *Emiliania huxleyi* as influenced by elevated pCO_2 were not significantly affected in *T. oceanica*. Calcifying organisms like *E. huxleyi* are known for their sensitivity to OA and in contrast elevated CO_2 might be neutral or beneficial for diatoms.

This type of sequencing was also used with two experiments conducted as part of a large-scale *in situ* costal pelagic mesocosms study part of the EPOCA 2010 Arctic campaign. These studies constitute the first combined metagenomic - metatranscriptomic community characterization of an Arctic fjord microbial community. The 16S rRNA amplicon sequencing was used to define the effect of OA on the community structure and indicated that CO_2 was not one of the main structuring variables whereas time, nutrient addition, water origin were. This study further established that the effect of CO_2 was significant on only 15 rare taxa. To characterize the metabolic diversity of the microbial community and the expression of functional genes, metagenome and metatranscriptome information was obtained by high-throughput sequencing of mRNA from one control and one elevated pCO_2 mesocosms at five time points throughout the duration of the experiment. Phylogenetic analysis indicated that the overall microbial community structure was, as previously established (16S analysis), not

strongly influenced by CO_2 treatments but varied significantly over time. The analysis of the metatranscriptome showed that only a small number of species and functional genes were affected significantly by pCO_2. The overall results of the mesocosms study demonstrate that the effects of OA on the Svalbard's Arctic microbial community structure and metabolisms are restricted to low-abundance species and specialized functional genes like for example DMSP breakdown or acid resistance.

The results presented in this dissertation increase the knowledge on the effect of OA on the diatom *T. oceanica* as well as the Svalbard's Arctic microbial community. It also highlights the importance of high-sequencing analyses for establishment of the effect of climate changes on our oceans.

3. Introduction

3. Introduction

3.1 The rise of phytoplankton

The earth was formed approximately 4.6 billion years ago and the first life is believed to have appeared within the first billion years. At that time only a few organisms thrived in the oceans and plankton dominated life on the planet, directing the carbon dioxide (CO_2) – oxygen (O_2) flows. The infinite availability of seawater, sunlight and CO_2 drove prokaryotic oxygenic autotroph, which appeared ≈2.4 billion years ago (Falkowski *et al.*, 2004), to take advantage and proliferated. Coincidentally, a radiation in the evolution of planktonic species happened, leading to a speciation in compound utilization (photoautotrophic versus chemoautotrophic). The appearance of phytoplankton populating today's earth and ocean was directed by primary, secondary and tertiary endosymbiosis (Timmis *et al.*, 2004; Oudot-Le Secq *et al.*, 2007; Woodson *et al.*, 2008; Kleine *et al.*, 2009). The first eukaryotic phytoplankton originated from the primary endosymbiosis of a mitochondrion-containing heterotrophic eukaryote with a cyanobacterium-like prokaryote ≈1.5-1.2 billion years ago; creating primary chloroplasts in Glaucophyta, Chlorophyta and Rhodophyta (Dyall *et al.*, 2004; Oudot-Le Secq *et al.*, 2007; Woodson *et al.*, 2008). Subsequently, secondary endosymbiosis occurred several times throughout the evolutionary history of algae, creating the commonly termed, "green" and "red lineage". The green algal lineage is attributed to the absorption of a Chlorophyta by a heterotrophic host which led to the Chlorarachniophyta and Euglenophyta. Whereas the red algal lineage resulted from the engulfment of a chloroplast-containing Rhodophyta (≈240 million years ago - Falkowski *et al.*, 2004) by a heterotrophic host and gave secondary photosynthetic chloroplasts of, for instance, the Heterokontophytes (containing *Bacillariophyceae*) (Armbrust *et al.*, 2004; Oudot-Le Secq *et al.*, 2007; Kleine *et al.*, 2009). Recent genomic study further suggests that some diatoms nuclear genes were obtained from Chlorophyta derivation (Moustafa *et al.*, 2009). A long string of evolutionary processes produced present-day organisms and much information is yet to be found but Saltzman *et al.* (2011) interestingly showed that phytoplankton were present in the ocean at least 500 million years ago when a large amount of oxygen re-entered the oceans and atmosphere.

3.2 Marine phytoplankton

Phytoplankton is the general assemblage of photoautotrophic organisms containing ≈4000-5000 classified species (Sournia *et al.*, 1991; Tett *et al.*, 1995). Phytoplankton is at the bottom of the marine food web being the main food source for zooplankton and responsible for about 98% of the net marine primary production (≈50Pg carbon per years - Longhurst *et al.*, 1995; Field *et al.*, 1998). The organisms composing phytoplankton are classified based on their phylogenetic relationships and are functionally divided in regards to their biochemical uses (Field *et al.*, 1998) as photoheterotroph, chemoheterotroph, chemoautotroph and photoautotroph (e.g. Synechococcacles, Bacillariophyta, Chlorophyta, etc. - van den Hoek *et al.*, 1997; Giovannoni *et al.*, 2000).

3.2.1 *Bacillariophyceae*

New molecular evidences and earliest fossilized siliceous *Bacillariophyceae* (diatoms) frustules place their appearance between ≈250 (Raven *et al.*, 2004) and 185 million years ago (Kooistra and Medlin, 1996). The diatom's class contain over 250 genera and with species that colonized marine, brackish and freshwater ecosystems, they are amongst the most diverse and successful photosynthetic organisms on the planet. Diatoms are divided, based on the arrangement of their frustules, into two groups: the centric and the pennate diatoms. The presence of the siliceous cell walls create an elevated demand for silica (Si); making them important players in controlling the global biogeochemical Si cycle (Falkowski *et al.*, 1998; Tréguer *et al.*, 2013) and the only living organisms included in the silicon budget of the world's ocean (Baines *et al.*, 2012; Tréguer *et al.*, 2013). The fossilized structures of diatoms, diatomite, has been used commonly in multiple commercial products (filters, deodorants, decoloring agents, insecticide, drug carriers (Scala *et al.*, 2001; Losic *et al.*, 2009; Losic *et al.*, 2010) for several decades, demonstrating the immense importance of diatoms in our daily lives. It was further indicated that the Si cycle is strongly interconnected to the atmospheric carbon dioxide cycle (Pondaven *et al.*, 2000) via the exponential connection between areas with numerous opal numbers (Si-bearing solutions precipitation form) and high level of carbon export or production. Diatoms are further suspected to play an important role in the phosphate cycle by exporting it in their cells to the sea floors (Diaz *et al.*, 2008). However, the diatoms' major role in the marine ecosystem is their contribution to primary production which accounts for ≈40% of the yearly oceanic primary production (Nelson *et al.*, 1995; Dugdale *et al.*, 1998), making them the most ecologically successful eukaryote of all time (Armbrust *et*

al., 2004; Bowler *et al.*, 2008). The exact reasons for their immense ecological success are still subject to intense investigation today.

3.2.2 *Thalassiosira* species

The genus *Thalassiosira* is the most diversified of the *Bacillariophyceae*, with more than a hundred species (Round *et al.*, 1990; Hasle *et al.*, 1996). It is, furthermore, the most investigated and thus the best known genus of its class. The *Thalassiosira* species wide distribution around the globe, their occurrence in coastal as well as in oceanic regions and their ability to be easily cultured in laboratories make them of particular interest for researchers. *Thalassiosira* species were identified as major contributors of diatoms' spring bloom, often being the dominant organisms (Riebesell, 1989; Shinada *et al.*, 1999; Popovich *et al.*, 1999). *Thalassiosira pseudonana* CCMP1335 was one of the first diatoms to have its complete genome sequenced (Armbrust *et al.*, 2004). Its genome gave the first insights of diatom's nuclear genes and revealed unexpected metabolic pathways and adaptations. For example, their ability to survive long period of darkness has been linked to high lipid content and storage (used for metabolic intermediate and ATP generation) allowing rapid growth upon the return of light (Armbrust *et al.*, 2004). In this thesis, experiments were mainly conducted on *Thalassiosira oceanica*, a close oceanic relative of the coastal model diatom *T. pseudonana*. The availability of the genome offered the opportunity to compare our sequencing results and thus to make the characterization of different genes and evaluate their expression under different OA conditions.

3.3 Marine chemistry

Background information on marine chemistry is presented in this section in order to comprehend the basic principles behinds the experimental setup presented in this thesis. The short review on the carbonate chemistry is based on a more detailed review by Zeebe and Wolf-Gladrow (2001) and the iron (Fe) chemistry on diverse peer-reviewed research.

3.3.1 Basic carbonate chemistry

Carbon chemistry is important to understand because the oceans absorb about 50% of the anthropogenic released CO_2 (Sabine *et al.*, 2004) from, for instance, fossil fuel burning and deforestation. The dissolution of CO_2 in the ocean is described in relation to temperature and salinity in Henry's law ($[CO_2] = K_0\, pCO_2$), where K_0 is the Henry's law constant and pCO_2 the partial pressure of CO_2. In seawater, the dissolved CO_2 is reacting with water (H_2O) to form true carbonic acid (H_2CO_3) which has a 0.3% smaller concentration than the original CO_2 concentration ($[CO_2]$) but is however chemically non differentiable. Its concentration ($[H_2CO_3]$) is thus by convention included in the $[CO_2]$. The dissolution of H_2CO_3 results in bicarbonate ions (HCO_3^-) and hydrogen protons (H^+) which, in turn, dissociate into carbonate ions (CO_3^{2-}) and hydrogen (H^+) ions. Theses dissociations are represented in the famous equilibrium and simplified equilibrium equations (1) (2).

$$CO_2 + H_2O \leftrightarrow H_2CO_3 \leftrightarrow HCO_3^- + H^+ \leftrightarrow CO_3^{2-} + 2H^+ \quad (1)$$

$$CO_2 + H_2O \overset{K_1^*}{\leftrightarrow} HCO_3^- + H^+ \overset{K_2^*}{\leftrightarrow} CO_3^{2-} + 2H^+ \quad (2)$$

The dissociation processes are characterized by K_1^* and K_2^*, the temperature, salinity and pressure dissociation constants expressed as stoichiometric dissociation constants (3) (4) previously determined by Lee *et al*., (2000).

$$K_1^* = \frac{[HCO_3^-]\,[H^+]}{[CO_2]} \quad (3)$$

$$K_2^* = \frac{[CO_3^{2-}]\,[H^+]}{[HCO_3^-]} \quad (4)$$

Once the values of K_1^* and K_2^* are known, it is possible to calculate the proportion of the elements in (2). For example, in a typical surface water system with temperature of 15°C, salinity of 35 and pH of 8, the calculations using the constants were done and produced the reference figure by Roy (Roy *et al.*, 1993) presented in Figure 1. It demonstrated that the

variation of one carbonate chemistry component directly leads to the variation of all others. A decrease in [CO_2] increases the [HCO_3^-] which thus increases the [CO_3^{2-}] and [H^+].

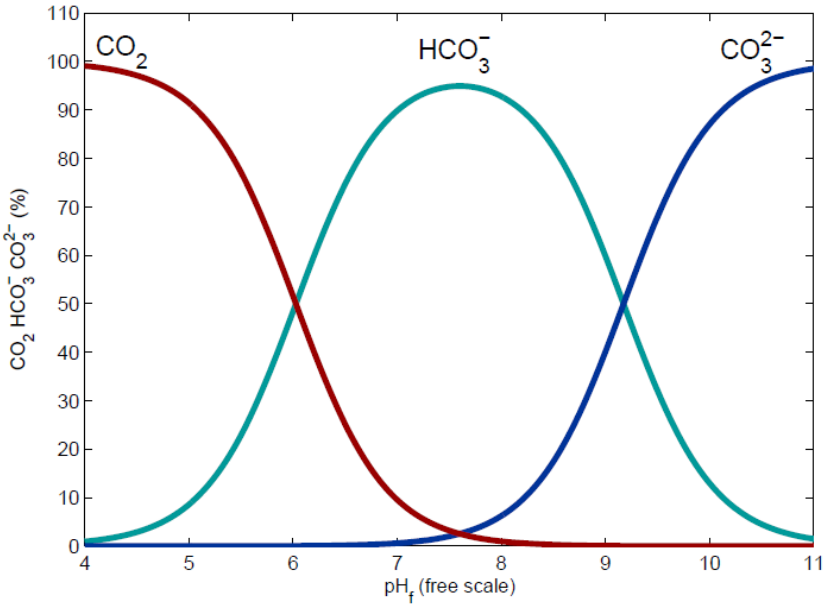


Figure 1. Proportion relative of carbonate chemistry species CO_2, HCO_3^- , and CO_3^{2-} in relation with different pH (free scale) at standard conditions as showed by Roy *et al.*, (1993).

To further fully understand the carbonate system, knowledge of the interaction between the dissolved inorganic carbon (DIC) and the total alkalinity (TA) in the carbonate system are essential. Briefly, DIC is the sum of the three dissolved inorganic carbon species (5) and TA is the number of moles of H^+ ion equivalent to the excess of proton acceptors over proton donors in one kilogram of sample as defined by Dickson (Dickson, 1981) (6).

$$DIC = [CO_2] + [HCO_3^-] + [CO_3^{2-}] \quad (5)$$

$$TA = [HCO_3^-] + 2[CO_3^{2-}] + [B(OH)_4^-] + [OH^-] + 2[PO_4^{3-}] + [H_3SiO_4^-] + [NH_3^-] + [HS^-] - [H^+] - [HSO_4^-] - [HF] - [H_3PO_4] \quad (6)$$

Other parameters, like calcium carbonate, are also important in understanding the complete carbonate system and how future environmental conditions will influence marine systems. However for diatoms, other parameters related or affected by carbon are more important; for

instance, iron (Fe) and silicate (Si) are essential for diatoms and were indicated as influenced by carbon concentration variation.

3.3.2 Iron in the ocean

Iron is an essential micronutrient for all living microorganisms. In particular, phototrophic organisms require Fe for many of the photosynthetic enzymes as it is a powerful electron carrier involved in photosynthetic electron transport and involved in the synthesis of chlorophyll (Martin *et al.*, 1988; Martin, 1990). The Fe solubility in seawater is controlled by salinity and is reduced by its increase, resulting in lower dissolved [Fe] in the ocean relative to freshwater lakes or rivers. Principally, Fe dissolves in coastal areas (Pitchford *et al.*, 1999) where it reaches high concentrations (in the micromolar range) whereas in tropical oceanic waters concentrations are in the subnanomolar range (Figueres *et al.*, 1978). Consequently, large oceanic areas depleted in Fe cover about 20% of the world's oceans (Pitchford *et al.*, 1999). These areas are called High-Nutrient Low-Chlorophyll regions (HNLC) because all essential nutrients for phytoplankton growth are present in excess with the exception of Fe which limits growth and keeps the chlorophyll concentrations low (Martin *et al.*, 1994; Pitchford *et al.*, 1999). It was clearly demonstrated that phytoplanktonic growth was directly correlated to the [Fe] in the oceans (Martin *et al.*, 1988). It was however further observed that Fe could reach some Antarctic oceanic regions through atmospheric particles of dust containing Fe (Martin *et al.*, 1988). This phenomenon was later observed in other regions of the world and guided the reasoning for the Fe hypothesis, stating that by introducing the limiting element Fe, a fast reaction of the phytoplankton could be triggered and lead to an increased capture and sequestration of carbon (Martin *et al.*, 1990; Martin, 1990). This speculation led to the accomplishment of at least 12 Ocean Iron Fertilization mesoscale experiments (for example: IronEX - Pitchford *et al.*, 1999; SOFeX - Coale *et al.*, 2004; Boyd, 2004; EIFEX Hoffmann *et al.*, 2006; etc.), which globally concluded that dissolved Fe addition in HNLC region stimulated diatoms biomass as well as small class phytoplankton. The ability of diatoms to directly respond to Fe allowed them to shortly become the dominant species. This dominance resulted in significant increased drawdown of CO_2 and ultimately, to increased production of particulate organic carbon (Coale *et al.*, 1996; Pitchford *et al.*, 1999; Jickells *et al.*, 2005; Boyd *et al.*, 2012; Smetacek *et al.*, 2012; Williamson *et al.*, 2012).

3.3.3 Iron bioavailability

As Fe is highly insoluble in seawater, its dissolution in O_2-rich water Fe (oxidized form Fe(III)), tends to settle rapidly. Indeed, Fe(III) tends to dissociate itself and form Fe hydroxide ($Fe(OH)^{2+}$, $Fe(OH)_2^+$, $Fe(OH)_3$, $Fe(OH)_4^-$), Fe oxide and various Fe(III) oxide-hydroxide (FeOOH) (Nishioka *et al.*, 2005; Cullen *et al.*, 2006; Bergquist *et al.*, 2007; Wu, 2007). The current dogma states that none of these forms are biologically available for phytoplankton as photosynthetic organisms can only assimilate Fe(II). This soluble and biologically available form of Fe is however scarce in the ocean (Hutchins *et al.*, 1999; Kuma *et al.*, 2000; Chen *et al.*, 2001). Most Fe species in the ocean are not dissolved freely but rather part of organic complexes binding to Fe(III). Therefore, to be assimilated, Fe(II) must first be extracted from its organic carriers. Photosynthetic organisms are often deficient in Fe and have developed special strategies to efficiently incorporate the Fe species needed (Castruita *et al.*, 2007; Strepek *et al.*, 2004). Consequently, several organisms like yeast, fungi and more importantly diatoms are found to be able to reduce Fe(III) into Fe(II); thus allowing the uptake of organically complexed Fe (Shaked *et al.*, 2005).

The diatoms' response to Fe limitation was characterized by the discovery of the replacement of ferredoxin, a Fe-containing protein, by flavodoxin, a larger protein that does not contain Fe observed in Fe-limited growth conditions (LaRoche *et al.*, 1996; McKay *et al.*, 1997; McKay *et al.*, 1999). It was subsequently established that *T. oceanica* was able to maintain a higher growth rate than *T. pseudonana* and *Thalassiosira weissflogii* under Fe limitation (Sunda *et al.*, 1991; Maldonado *et al.*, 1996; Strzepek *et al.*, 2004). The tolerance of this species for low Fe conditions was attributed to a modified photosynthetic architecture, with reduced amounts of the Fe-rich PSI proteins and the replacement of cytochrome C6 with plastocyanin, a protein that does not contain Fe (Peers *et al.*, 2006). In order to further characterize the Fe uptake system of diatom, it is necessary to search in other species. Van Ho (Van Ho *et al.*, 2002) described in yeast the Fe transport in between membranes as being done by a multicopper oxidase producing specific permease working on the previous reoxidation. Also, research on Chlorophyta showed Fe to be stored in cellular surface by specialized Fe-binding molecules and thereafter assimilate by acidic vacuoles (Paz *et al.*, 2007). Even if these organisms are not close relatives of diatoms, they contain some genes similar to diatoms (ferric reductases and multicopper oxidases) indicating that their Fe transport systems could be comparable to the one of diatoms. It further suggests that diatoms are probably using reductive mechanisms for

Fe uptake. It is believed that diatoms Fe uptake systems could comprise receptors for organic Fe complexes, redox enzymes, ferric reductases, multicopper ferrous oxidase and ion permeases all involved in Fe imports. The first three manuscripts presented used similar research techniques to identify candidate components of a high-affinity Fe-uptake system, cellular Fe requirements as well as the location of ferredoxin to enlighten the possible adaptation of *T. oceanica* to low-Fe.

3.3.4 The biological and solubility pumps

The biological pump is the cumulation of all biologically-driven processes using or transporting carbon to the ocean's water column. The atmospheric CO_2 is firstly dissolved in the ocean's surface and is thereafter used by photosynthetic organisms that generally constitute the staple diet of higher organisms. These organisms are part of a food web from which dead organisms or fecal matter containing carbon sink (biologically or physically) towards the ocean's floor. The particulate organic carbon is either re-circulated during their descant or trapped in sediments awaiting remineralization (Figure 2 left side). The physico-chemical counterpart of the biological pump is the solubility pump which also carries carbon from the surface (absorbed from the atmosphere) to the water column and ultimately to the deep, cold ocean water where it enters the deep water circulation for thousands of years. The physico-chemical pump is, however, controlled by the solubility of carbon dioxide mostly depending on the thermohaline circulation of the oceans (Figure 2 right side).

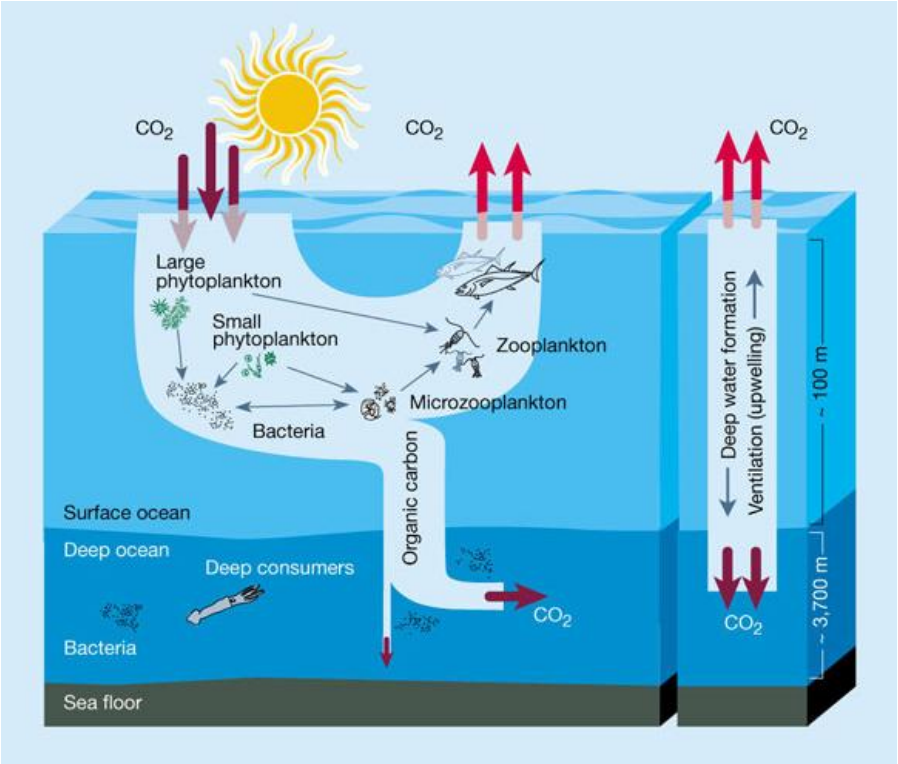


Figure 2. Schematic representation of the biological (left) and solubility (right) pumps from Chisholm (Chisholm, 2000).

3.4 Increase in anthropogenic CO_2

Since the industrial revolution, an increase in the release of anthropogenic CO_2 via deforestation, agriculture, uses of cement and fossil fuels, has been observed. The predicted atmospheric pressure of CO_2 (pCO_2) for the end of the century (2100) will be greater than 800 µatm (Friedlingstein *et al.*, 2006). The Intergovernmental Panel on Climate Change (IPCC) has been investigating human-driven climate changes and compiling all related researches to draw an overall picture of such effects. In 2007, it was concluded that this increase in pCO_2 and the lack of success in controlling greenhouse gas emissions has potentially influenced the planet's temperature (via greenhouse gas retention - IPCC, 2007). The main results of global climate changes on the oceans were identified as ocean warming and acidification.

3.4.1 Ocean warming

The warming of the oceans is indirectly caused by the increased [CO_2] in the atmosphere via greenhouses gases effects (Levitus *et al.*, 2001). This temperature increase is predicted to have critical impact on the environment such as rising sea level (Bindoff *et al.*, 2007; Marcelja, 2010; Meehl *et al.*, 2007) or the melting of Arctic ice causing a freshening of polar regions (Bethke *et al.*, 2006) and an alteration in the global ocean circulation. It will also influences the evaporation rate (intensified) leading to increase in salinity and water stratification, thus reducing nutrient exchanges between deep water masses and the euphotic zone (Yeh *et al.*, 2009) and possibly creating an amplified shortage of nutrients in the oceans.

3.4.2 Ocean acidification

The increased atmospheric CO_2 absorbed by the ocean (≈50% - Sabine *et al.*, 2004) where it dissolves as shown in (1) results in an increased [H^+] which accumulates and alters the ocean's pH. It was predicted that the ocean's pH will decrease by 0.25-0.45 units before 2100 (Orr *et al.*, 2005). The acidification of the oceans, termed Ocean Acidification (OA), decreases the availability of carbonate ions (Wolf-Gladrow *et al.*, 1999; Zeebe *et al.*, 2001; Caldeira *et al.*, 2003; Newbold *et al.*, 2012) and is suspected to reduce the Fe bioavailability (Shi *et al.*, 2010). These alteration's effects on our oceans are still not completely known but their influence on the physiology of marine organisms as well as their role in biochemical cycles and ocean biochemistry is widely recognized. The effect of OA is until now believed to be species-specific, potentially having a greater negative impact on calcifying organisms (coral, pteropodes, etc. - Orr *et al.*, 2005).

3.5 Scientific approaches for the study of ocean acidification

3.5.1 Culture and mesocosms

Multiple laboratory cultures using single species have been made to evaluate the effect of climate changes like increase in light intensities, decrease in pH, increasing pCO_2, decrease of calcium carbonate and Fe or of any of these parameters combined together (Richier *et al.*, 2011; Wu *et al.*, 2010; Lohbeck *et al.*, 2012; Shi *et al.*, 2010). In the present thesis, two types

of batch cultures, closed system and open CO_2 system, were employed. Typical batch cultures consist of culturing cells in a closed system where media volume, temperature, light (intensity and duration) as well as the quantity of nutrient are all fixed and controlled. Cells are normally harvested before all nutrients are consumed and the cells driven into a limitation state (LaRoche *et al.,* 2010). These closed-system batch cultures were used within the framework of manuscript I, II and III. Variations to typical batch culture protocol in order to optimize treatments are sometimes possible. Inspired by chemostat systems, special modifications to batch culture vessels were made to allow constant inflow of CO_2 and thus reduce pressure variation in the vessels used for manuscript IV. A new experimental method called mesocosms was developed in recent years to characterize the reaction of an entire community to OA (Riebesell *et al.,* 2013). Mesocosms studies are used to manipulate pCO_2 and control all other parameters in natural community enclosures allowing conclusions to be applied to a whole community instead of single species (manuscript V and VI). These two techniques permit researchers to compare how species actually react and interact with each other to climate changes by getting a more complete picture.

3.5.2 Sequencing techniques

The emergence of nucleic acid sequencing technique allowed scientists to gather genetic information on any type of organisms. Sequencing techniques started with the cloning of environmental DNA and functional expression screening followed by chain terminator Sanger sequencing (Sanger *et al.*, 1977; Sanger *et al.*, 1975) and was then followed by the rapid development of direct random shotgun sequencing (Roche 454, Illumina). The 454 is based on pyrosequencing technology and generates ≈400 000 reads per run. Each reads is ≈400 bases pairs long (Ronaghi *et al.*, 1998) which gives great advantages when de novo assembly of whole genomes are in order. The Illumina sequencing platform uses a new techniques called sequencing by synthesis where the, for example, DNA strands are amplified by bridge amplification resulting in cluster of clonal copies of the given DNA. The Illumina sequencing techniques can be used on four sequencing systems from which two were used in the present thesis. The HiSeq2000 generally gave 180 million reads per lane with a length of 2 x 101 reads and can run up to eight individual samples at a time but takes about two weeks to obtain results. The MiSeq sequencer is the newest technology developed and produces results more rapidly (≈27 hours). It results in 7.5 million reads per lane with a length of 2 x 150 read. Both sequencers offer the ability to run multiple samples as paired end sequencing which increases

the number of samples analysed per run and thus reducing sequencing costs. Both methods allow scientists to precisely qualify communities' assemblages (metagenome) and gene expression to the species (transcriptome) or to the community (metatranscriptome) level. Indeed, the past 10 years have seen the developments of sequencing and bioinformatics techniques that have enabled researchers of all backgrounds (not only experts) to reinvigorate the way science is done. The evolution of meta-analyses reviewed by Thomas and colleagues (Thomas *et al.*, 2012), showed that detailed descriptions of the community structure or gene expression can be done without the need to cultivate individual environmental isolates (Margulies *et al.*, 2005; Sogin *et al.*, 2006; Huber *et al.*, 2007; Huse *et al.*, 2008; Gilbert *et al.*, 2009; Agogue *et al.*, 2011). These methods further allow observing how communities are influenced by environmental factors or treatments and draw a precise picture of their impact. Multiple sequencing techniques were used in the making of the present dissertation keeping up with the technological development of sequencing instrumentation and approaches. Indeed, the sequencing for manuscript I and II was done on the 454 sequencer because the reads length facilitates *de nov*o assembly and other sequencers were not yet available or were too recent for our use. Sanger sequencing was used in manuscript III for the rapid sequencing of a limited number of distinct nucleic acid fragments (almost directly) while the Illumina HiSeq sequencing platform was used for the *T. oceanica* transcriptome of manuscript IV and for the 16S rRNA amplicon sequencing of the Svalbard's mesocosms presented in manuscript VI because it offers a little more sequencing depth which could potentially uncover rare organisms or genes. Finally, the MiSeq platform sequenced the metagenome and metatranscriptome of manuscript V because it produces longer reads. However, the quality of sequencing being the same between the HiSeq and the MiSeq, the sequencer was selected in regards to their availability.

3.6 Aim of this thesis

The experiments presented in this dissertation underlines the importance of comparing laboratory cultures to mesocosms approaches in order to understand and qualify the effect of OA on marine phytoplankton especially on diatoms and Arctic microbial communities. The present thesis intended to analyze OA effects and influences on community structure and gene expression of the previously mentioned marine phytoplankton by qualifying changes in uni-algal cultures, *T. oceanica*, and in large community; to reduce climate changes extrapolations.

In order to achieve these goals, multiple essential aims were established:

- Determine if the ferredoxin gene relocation from the chloroplast to the nucleus was solely present in the *T. oceanica* CCMP1005 strain or if the relocation was widely spread within the species or in the Thalassiosira genus by testing multiple strains of two more species said *T. pseudonana* and *T. weissflogii*.

- Establish direct effects of increased pCO_2 on the transcriptome of *T. oceanica* CCMP1005 via modified batch cultures respecting OA guidelines.

- Characterizing the effect of OA on Arctic bacterial community structure through time and pCO_2 gradient in a large *in situ* pelagic mesocosm experiment via 16S rRNA amplicon sequencing and to give some insights on isolating water in mesocosms enclosures.

- Describe further the variation of the Svalbard's Arctic microbial assemblages and the differences in their functional gene expression between one un-manipulated control mesocosm (≈185 µatm) and one manipulated mesocosm (≈1050 µatm) through time using shotgun metagenomic and metatranscriptomic sequencing.

- Demonstrate the importance of high-throughput sequencing in obtaining a better overview of OA effects.

4. Manuscripts

4. Manuscripts

4.1 Co-authorship publications

4.1.1 List of co-authorship publications, research focus and declaration of contribution

Publication I: Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation

Research focus: Sequence, complete and describe the *T. oceanica* chloroplast genome.

Contribution: Alexandra-Sophie Roy cultured the algae, carried out the extractions and RT-qPCR work and commented on the manuscript.

Publication II: Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation.

Research focus: Characterize the response of *T. oceanica* to deplete iron via genomic, transcriptomic and proteomic analysis.

Contribution: Alexandra-Sophie Roy cultured the algae, carried out the extractions and RT-qPCR work and commented on the manuscript.

4.1.2 Publication I.

Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation.

Lommer *et al.*, 2010

***BMC Genomics* 2010, 11:718**
doi:10.1186/1471-2164-11-718

Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation

Markus Lommer[1], Alexandra-Sophie Roy[1], Markus Schilhabel[2], Stefan Schreiber[2], Philip Rosenstiel[2], Julie LaRoche[1*]

Abstract

Background: Although the importance and widespread occurrence of iron limitation in the contemporary ocean is well documented, we still know relatively little about genetic adaptation of phytoplankton to these environments. Compared to its coastal relative *Thalassiosira pseudonana*, the oceanic diatom *Thalassiosira oceanica* is highly tolerant to iron limitation. The adaptation to low-iron conditions in *T. oceanica* has been attributed to a decrease in the photosynthetic components that are rich in iron. Genomic information on *T. oceanica* may shed light on the genetic basis of the physiological differences between the two species.

Results: The complete 141790 bp sequence of the *T. oceanica* chloroplast genome [GenBank: GU323224], assembled from massively parallel pyrosequencing (454) shotgun reads, revealed that the petF gene encoding for ferredoxin, which is localized in the chloroplast genome in *T. pseudonana* and other diatoms, has been transferred to the nucleus in *T. oceanica*. The iron-sulfur protein ferredoxin, a key element of the chloroplast electron transport chain, can be replaced by the iron-free flavodoxin under iron-limited growth conditions thereby contributing to a reduction in the cellular iron requirements. From a comparison to the genomic context of the *T. pseudonana* petF gene, the *T. oceanica* ortholog can be traced back to its chloroplast origin. The coding potential of the *T. oceanica* chloroplast genome is comparable to that of *T. pseudonana* and *Phaeodactylum tricornutum*, though a novel expressed ORF appears in the genomic region that has been subjected to rearrangements linked to the petF gene transfer event.

Conclusions: The transfer of the petF from the cp to the nuclear genome in *T. oceanica* represents a major difference between the two closely related species. The ability of *T. oceanica* to tolerate iron limitation suggests that the transfer of petF from the chloroplast to the nuclear genome might have contributed to the ecological success of this species.

4.1.3 Publication II.

Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation.

Lommer *et al.*, 2012

***Genome Biology* 2012, 13:R66**
doi:10.1186/gb-2012-13-7-r66

Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation

Markus Lommer[1], Michael Specht[2], Alexandra-Sophie Roy[1], Lars Kraemer[3], Reidar Andreson[4,5], Magdalena A Gutowska[6], Juliane Wolf[2], Sonja V Bergner[2], Markus B Schilhabel[3], Ulrich C Klostermeier[3], Robert G Beiko[7], Philip Rosenstiel[3], Michael Hippler[2] and Julie LaRoche[1,8,*]

Abstract

Background
Biogeochemical elemental cycling is driven by primary production of biomass via phototrophic phytoplankton growth, with 40% of marine productivity being assigned to diatoms. Phytoplankton growth is widely limited by the availability of iron, an essential component of the photosynthetic apparatus. The oceanic diatom *Thalassiosira oceanica* shows a remarkable tolerance to low-iron conditions and was chosen as a model for deciphering the cellular response upon shortage of this essential micronutrient.

Results
The combined efforts in genomics, transcriptomics and proteomics reveal an unexpected metabolic flexibility in response to iron availability for *Thalassiosira oceanica* CCMP1005. The complex response comprises cellular retrenchment as well as remodeling of bioenergetic pathways, where the abundance of iron-rich photosynthetic proteins is lowered, whereas iron-rich mitochondrial proteins are preserved. As a consequence of iron deprivation, the photosynthetic machinery undergoes a remodeling to adjust the light energy utilization with the overall decrease in photosynthetic electron transfer complexes.

Conclusions
Beneficial adaptations to low-iron environments include strategies to lower the cellular iron requirements and to enhance iron uptake. A novel contribution enhancing iron economy of phototrophic growth is observed with the iron-regulated substitution of three metal-containing fructose-bisphosphate aldolases involved in metabolic conversion of carbohydrates for enzymes that do not contain metals. Further, our data identifies candidate components of a high-affinity iron-uptake system, with several of the involved genes and domains originating from duplication events. A high genomic plasticity, as seen from the fraction of genes acquired through horizontal gene transfer, provides the platform for these complex adaptations to a low-iron world.

Keywords
Thalassiosira oceanica, diatoms, iron limitation, genomics, transcriptomics, proteomics, ISIP1, FBA

4.2 First-authorship publications

4.2.1 List of first-authorship publications, research focus and declaration of contribution

Publication III: The transfer of the ferredoxin gene from the chloroplast to the nuclear genome is confined to T*halassiosira oceanica* within the genus T*halassiosira*. Roy *et al.*, in revision for Journal of Phycology

Research focus: Determine if the PETF gene transfer, from the chloroplast to the nucleus, is unique to *T. oceanica* in the *Thalassiosira* genus.

Contribution: Idea: Alexandra-Sophie Roy and Julie LaRoche

Experimental work: Alexandra-Sophie Roy

Analysis of data: Alexandra-Sophie Roy

Preparation of manuscript: Alexandra-Sophie Roy with comments from co-authors

Publication IV: Effect of increased pCO_2 on the transcriptome of *Thalassiosira oceanica* (CCMP1005). Roy *et al.*, in prep.

Research focus: Establish the effect of elevated pCO_2 on the transcriptome of *T. oceanica*.

Contribution: Idea: Alexandra-Sophie Roy and Julie LaRoche as part of Biological Impact of Ocean ACIDification (BIOACID)

Experimental work: Alexandra-Sophie Roy

Analysis of data: Alexandra-Sophie Roy, Dhwani Desai and Matthias Barann (sequencing mapping analysis)

Preparation of manuscript: Alexandra-Sophie Roy with comments from co-authors

Publication V: Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms. Roy *et al.*, 2013, Biogeosciences

Research focus: Characterize the bacterial assemblage of an Arctic community and its variation in relation to time and pCO_2 using 16S rRNA database.

Contribution: Idea: Alexandra-Sophie Roy and Julie LaRoche as part of the European project on ocean acidification (EPOCA)

Experimental work: EPOCA Svalbard's mesocosms team (sampling) and Alexandra-Sophie Roy (samples extraction, distribution and process)

Analysis of data: Alexandra-Sophie Roy, Sean M. Gibbons

Preparation of manuscript: Alexandra-Sophie Roy with comments from co-authors

Publication VI: Ocean acidification effects on the community and gene expression of Arctic bacteria and phytoplankton during a coastal pelagic mesocosms event. Roy *et al.*, in prep. for ISME Journal

Research focus: Characterize the influence of time and pCO_2 on the transcriptome of a microbial Arctic community using next-generation sequencing.

Contribution: Idea: Alexandra-Sophie Roy and Julie LaRoche as part of the European project on ocean acidification (EPOCA)

Experimental work: EPOCA Svalbard's mesocosms team (sampling) and Alexandra-Sophie Roy (samples extraction, distribution and process)

Analysis of data: Alexandra-Sophie Roy, Harald Schunck, Sean M. Gibbons, Dhwani Desai

Preparation of manuscript: Alexandra-Sophie Roy with comments from co-authors

4.2.2 Publication III.

The transfer of the ferredoxin gene from the chloroplast to the nuclear genome is confined to *Thalassiosira oceanica* within the genus *Thalassiosira*.

Roy *et al.*,
in revision for Journal of Phycology

THE TRANSFER OF THE FERREDOXIN GENE FROM THE CHLOROPLAST TO THE NUCLEAR GENOME IS CONFINED TO *THALASSIOSIRA OCEANICA* WITHIN THE GENUS *THALASSIOSIRA*[1].

Alexandra-Sophie Roy[2], Markus Lommer and Julie LaRoche[*]

[1]GEOMAR Helmholtz Center for Ocean research Kiel, Düsternbrooker Weg 20, 24105, Kiel, Germany.

*present address: Biology Department Dalhousie University, 5850 University Avenue, B3K 6R8, Halifax, NS, Canada

[2]Author for correspondence: email sroy@geomar.de, Tel: +49 0431 4262, Fax: +49 0431 4446

Running title: FERREDOXIN IN *THALASSIOSIRA*'S GENUS

ABSTRACT

The location of the ferredoxin gene (chloroplastic or nuclear) was determined in 12 strains of *Thalassiosira* covering three species. DNA sequencing and qPCR were used to establish that the ferredoxin gene is located in the nuclear genome of all confirmed *T. oceanica* strains (CCMP0999, 1001, 1005, 1006) tested. Conversely, all *T. pseudonana* (CCMP1012, 1013, 1014, 1335) and *T. weissflogii* (CCMP1010, 1049, 1052) strains tested have retained the gene in the chloroplast genome, as is commonly found for *Bacillariophyceae*. We further demonstrated that the strain-specific location of the ferredoxin gene was reliably determined from total genomic DNA by using qPCR assay of single copy nuclear and chloroplastic genes. Sequencing of the 5.8S and ITS regions further established that the classification of *T. oceanica* strain CCMP1616 needs to be re-assessed. The ferredoxin gene relocation gives an insight into the evolutionary processes leading to cp genome reduction and suggests ecological adaptation for such genomic rearrangements.

Keywords (5-10 words): Ferredoxin, gene location, gene transfer, *Thalassiosira* genus, *T. oceanica*, *T. pseudonana*, *T. weissflogii*.

Abbreviation: bp, base pairs; cp, chloroplast; Ct, threshold cycle; e.g., for example; Fe, iron; HNLC, High-Nutrient-Low-Chlorophyll; ITS, internal transcribed spacer; PETF, ferredoxin;

$PETF_{CP}$, cp genome PETF; $PETF_{NUC}$, nuclear genome PETF; qPCR, quantitative polymerase chain reaction; rbcs, small subunit of Ribulose-1,5-bisphosphate carboxylase oxygenase or RuBisCO; spp., species; TEF1A, translation elongation factor 1 alpha.

INTRODUCTION

Diatoms are one of the most ecologically successful eukaryotes of all time. These primary producers can account for ≈40% of the yearly oceanic primary production (Dugdale *et al.*, 1998, Nelson *et al.*, 1995). Although diatoms have been widely investigated, the reason for their immense success is still poorly understood. Some correlate diatoms' success to their functional plasticity, which is manifested by their dominance in very diverse environments from nutrient-rich coastal areas to Fe-depleted HNLC oceanic regions (Strzepek *et al.*, 2004). Although diatoms are not the dominant phytoplankton in HNLC regions, their presence in Fe-limited ecosystems indicates that they are adapted to survive Fe limitation (Strzepek *et al.*, 2004). The addition of this limiting element to HNLC waters generally triggers major diatom blooms (Chisholm *et al.*, 2001, Boyd *et al.*, 2007) possibly leading to an increased anthropogenic carbon sequestration into the deep ocean.

Since the emergence of the photosynthetic eukaryote via primary endosymbiogenesis of a *Cyanobacteria* (Martin *et al.*, 2002, Kleine *et al.*, 2009), ancestral eukaryotic algae have evolved via multiple genomic modifications, creating present day microalgae. In fact, the photosynthetic apparatus involves more than 700 genes which are coordinated by the plastid and nuclear genomes. However, modern cp genomes host only about 200 of these genes (Kleffmann *et al.*, 2004), indicating that a massive transfer of cp genes to the nuclear genome has taken place after the initial endosymbiosis event. Although 200 genes seem little in comparison to the whole genome, their regulation under diverse environmental changes is unquestionably critical to ascertain as the genes in the cp genome are fundamental photosynthetic genes. Two mechanisms, deletion and relocation, have been identified as possible drivers for cp genome evolution (Oudot-Le Secq *et al.*, 2007) but cannot guarantee the disappearance of the selected gene from its original location (Oudot-Le Secq *et al.*, 2007) nor can it identify if the modification led to a successful integration and functional activation in the nuclear genome (Timmis *et al.*, 2004). It has recently been demonstrated that gene relocation from the cp to the nucleus increases with environmental stress, particularly with heat, in higher plants (Wang *et al.*, 2012). It was further presented that such transfers happen in open chromatin regions, which have an increased accessibility by temperature stress (Wang *et al.*, 2013). A similar scenario can be envisioned for unicellular algae, especially for diatoms that inhabit marine oligotrophic surface waters subject to high light intensities.

As one of the most widely diversified phylogenetic taxa, diatoms display a remarkable range in cell and genome size that has had an impact on their evolutionary trajectory (von Dassow *et al.*, 2008). In particular, their circular cp genome is subject to frequent internal rearrangement (Oudot-Le Secq *et al.*, 2007) that promotes the transfer of organelle-encoded genes to the nuclear genome (Martin *et al.*, 2002, Richly *et al.*, 2004, Bock *et al.*, 2008, Sheppard *et al.*, 2009). In *Thalassiosira oceanica* Hasle CCMP1005, an oceanic diatom adapted to chronically low Fe concentrations (Strzepek *et al.*, 2004), the PETF gene encoding for the ferredoxin protein PETF has been transferred to the nuclear genome (Lommer *et al.*, 2010). This is unusual because the PETF gene is commonly found in the cp genome of most diatoms including the closely related coastal species *Thalassiosira pseudonana* Hasle & Heimdal and *Thalassiosira weissflogii* (Grunow) G. Fryxell & Hasle (Gueneau *et al.*, 1998, Lommer *et al.*, 2012). It has previously been established that *T. oceanica* maintains a higher growth rate under Fe limitation than *T. pseudonana* and *T. weissflogii* (Sunda *et al.*, 1991, Maldonado *et al.*, 1996, Strzepek *et al.*, 2004). Its tolerance for low Fe conditions has been attributed to a modified photosynthetic architecture, with reduced amounts of the Fe-rich PSI proteins and the replacement of cytochrome C6 with plastocyanin, a protein that does not contain Fe (Peers *et al.*, 2006). Similarly, the replacement of the Fe-containing protein PETF by flavodoxin, which does not contain Fe, has often been observed in Fe-limited growth conditions (LaRoche *et al.*, 1995, McKay *et al.*, 1997). We have previously hypothesized that the transfer of the PETF gene may provide *T. oceanica* with an additional advantage to tolerate Fe limitation. To identify whether this PETF relocation is constrained to *T. oceanica* species, or whether it is scattered throughout various isolated strains of the *Thalassiosira* genus, we designed primers specific for three *Thalassiosira* spp. and used qPCR to determine the location of the PETF gene in five strains of *T. oceanica* (CCMP0999, 1001, 1005, 1006, 1616), four of *T. pseudonana* (CCMP1012, 1013, 1014, 1335) and three of *T. weissflogii* (CCMP1010, 1049, 1052). The transfer of the PETF gene was assessed by comparing the Ct values of reference single copy genes with known compartmental locations, either chloroplastic (rbcs) or nucleic (TEF1A), to the Ct values of the PETF gene ($PETF_{CP}$ or $PETF_{NUC}$). The presence of multiple cp in *Thalassiosira* spp. results in multiple copies of cp genomes and because cp genome copies are naturally enriched relative to the nuclear genome, single copy cp-encoded genes are easily separated from single copy nuclear-encoded genes. Single copy genes located in the nuclear genome of eukaryotic diploid cells are present in two copies per cell while genes located in the cp genome have several copies. This study aimed to establish whether the PETF

gene transfer was unique to *T. oceanica* CCMP1005 or more widely distributed among the *Thalassiosira* genus.

MATERIAL AND METHODS

Strains and cultures

A total of 12 strains from three *Thalassiosira* species (Table 1) were obtained from the National Center for Marine Algae and Microbiota (NCMA formerly CCMP, http://ccmp.bigelow.org) and were cultured in f/2 artificial seawater medium (Guillard *et al.*, 1962, Guillard, 1975). Each strain was cultured in duplicate at its optimal temperature and light conditions (results Table 4) and was harvested in late exponential phase within a time span of 7 to 15 days. Cells were collected by filtration on polycarbonate filters (Durapore® 47 mm, Millipore, Darmstadt, Germany) with pore sizes compatible to the algae's size. Each cell sample was resuspended in a small volume of medium and centrifuged at 11 000 rpm for 10 min at 4°C in order to concentrate the cells in pellets, which were thereafter flash frozen in liquid nitrogen and stored at –80°C until DNA extraction was performed.

Nucleic acid extraction

Total nucleic acid was extracted using the "DNeasy Plant mini" kits from Qiagen® (Qiagen, Hilden, Germany) following the standard manufacturer's protocol with minor modifications. Changes to the protocol included a 10 sec sonication of the cell pellet following the addition of lysis buffer, the washing step with AW buffer was done twice, and the rinsing step with 500 µl of 99% ethanol was done thrice as recommended to increase yield. DNA was eluted in 250 µl elution buffer and was finally sheared using a sterile hypodermic needle (30 gauge) attached to a sterile 1 ml syringe. DNA amounts were assessed as the mean of triplicate measurements by micro-volume spectrophotometer nanodrop ND-1000 (PeqLab GmbH, Erlangen, Germany) measurements and quality was tested using the automated electrophoresis system "Experion^{TM} DNA-12K" analysis kit from Bio-Rad Laboratories (Hercules, CA, USA). All samples were kept at −80°C until further analysis.

qPCR

qPCR was performed with SYBR green master mix on the ABI Prism® SDS 7000 (Applied Biosystem by Life Technologies^{TM}, Carlsbad, CA, USA) to determine the location of the gene

of interest PETF, by comparing its Ct value to single copy gene of reference. Following our previous work (Lommer *et al.*, 2010), it was established that to accurately determine the location of the PETF gene, two reference genes were necessary for comparison. The first selected gene was the TEF1A, established as a single copy gene by screening diverse phylogenetically unrelated genomes, representing the nuclear location of the gene and used for normalization of the Ct values. The second gene chosen was rbcs encoding for the small subunit of RuBisCO, which represented the single copy cp gene used to characterize the cp location. The 18S gene was also detected with general primers for all *Thalassiosira* used in the study in order to confirm that all strains belonged to the *Thalassiosira* genus (data not presented). The qPCR reactions (25 µl) were composed of 12.5 µl Platinum SYBRgreen® (Invitrogen by Life Technologies ™, Carlsbad, CA, USA), 0.5 µl of 10 µM forward and reverse primers each, 6.5 µl PCR-H_2O and 5 µl of sheared gDNA template. Triplicates of two biological samples per strain and no-template controls were run for each primer pair. The designed primers sets (Primer Express v2.0 by Applied biosystem by Life Technologies™, Carlsbad, CA, USA), were tested and optimized (Table 2; Lommer *et al.*, 2010) to detect these specific genes (amplicon of approximately 100 bp) with high amplification efficiency (>95%). The primer's efficiency was tested by comparing amplification of same DNA amount from diverse *T. oceanica*, *T. pseudonana* and *T. weissflogii* and by using the LinReg program (http://www.gene-quantification.de/) to calculate the PCR efficiency of each primer sets. These reactions were treated with cycling conditions of 2 min at 50°C (once), 2 min at 95°C (once), 40 cycles of 0:15 min at 95°C and 0:30 min at 60°C, 0:15 min at 95°C (once) followed by a dissociation stage of 0:20 min at 60°C and a final extension of 0:15 min at 95°C. Relative gene copy number was assessed using relative amplification values represented by Ct, where each Ct was equal to a 2 fold concentration difference. The mean Ct values of triplicate reactions of the two biological replicates at a threshold level of 0.2 were used to compare the values of the gene of interest to the ones of the genes of reference. The PETF copy number is assumed to be detected at Ct values matching either the rbcs or the TEF1A gene; thus identifying the potential relocation of the gene to the nuclear genome. To obtain a clear gene location, the gene copy number was calculated by normalizing the gene of interest's Ct values to the TEF1A Ct values and was thereafter used to calculate the relative gene copy numbers with the formula $2^{\Delta Ct}$; where the concentration fold difference is represented by $\Delta Ct = Ct_{(\text{gene of interest})} - Ct_{(\text{gene of reference})}$ (Pitti *et al.*, 1998, Pelham *et al.*, 2006).

ITS-5.8S PCR

The ITS1-5.8S-ITS2 regions were prepared for sequencing using standard PCR composed of 2.5 µl of 10 x dreamTaq buffer (Fermentas, Life science, Thermo Fisher Scientific, Waltham, MA, USA), 1.25 µl of dNTP's 10 mM, 2.5 µl of each 10Mm primers (Table 2; Eurofins, Ebersberg, Germany), 0.25 µl of DreamTaq™ DNA polymerase (Fermentas, Life science, Thermo Fisher Scientific, Waltham, MA, USA), 1.5 µl of DNA Template and 14.5 µl of PCR water. These reactions were treated with modified ITS cycling conditions as described in Kumar and Shukla (2005) with an initial denaturation of 10 min at 95°C, a repeated 10 times denaturation step of 1:00 min at 95°C, an annealing temperature of 1:00 min at 55°C and an elongation step of 1:30 min at 72°C. The final extension was of 10:00 min at 72°C. Different sets of primers were used to assure maximum amplification efficiency of each species. The size of the PCR products was thereafter verified by electrophoresis on a 1.3 % agarose gel (>500 bp) to assure presence and amplification of the DNA fragment. Positive samples were sequenced.

Sequencing

Standard 3730xl Sanger sequencing (Applied biosystem by Life Technologies™, Carlsbad, Ca) of the ITS1-5.8S-ITS2 regions was done and consensus sequences for each strain obtained using DNASTAR (http://www.dnastar.com/; Madison, WI, USA.). These sequences were afterward used to identify closest neighbours (data not presented) by BLASTN analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Bethesda, MD, USA) and to construct phylogenetic trees. Once different *Thalassiosira* sequences were gathered from the databases SILVA (http://www.arb-silva.de/, Bremen, Germany), NCBI (http://www.ncbi.nlm.nih.gov/, Bethesda, MD, USA) and the present sequencing, two different groups (*Thalassiosira* sp.'s 18S; *T. oceanica-T. pseudonana-T. weissflogii*'s 5.8S+ITS2 - Table S1) were created from which multiple interspecific (18S and 5.8S) and intraspecific (ITS2) alignments were done using the ClustalW alignment option in the software GENEIOUS (http://www.geneious.com/, Auckland, New Zealand); these alignments were used to generate neighbour joining phylogenetic trees with calculated bootstrap values based on 1000 random resampling using the same software.

RESULTS

Transfer of PETF gene in Thalassiosira spp.

We applied a qPCR assay to determine the location of the PETF gene in multiple strains of three *Thalassiosira* spp., *T. oceanica*, *T. pseudonana* and *T. weissflogii*. Examples of the results obtained for $PETF_{NUC}$ and $PETF_{CP}$ location are respectively presented in Fig. 1a and b, where the PETF location is illustrated by identifying which gene of reference values (cp or nucleus) correspond to the one of the gene of interest. In *T. oceanica* CCMP 1006 (Fig. 1a), the mean Ct value +/-SD (18.79 +/-0.12) of the PETF amplification curve was not statistically different from the mean Ct +/-SD of TEF1A (19.27 +/-0.34; p= 0.01; Table S2) but was from the one of rbcs (14.28 +/-0.03; p= 3.71E-8; Table S2); confirming a $PETF_{NUC}$. In contrast, the *T. weissflogii* CCMP 1049 (Fig. 1b) mean PETF Ct value +/-SD (14.88 +/-0.45) was not statistically different than the rbcs mean Ct +/-SD (14.94 +/-0.50; p= 0.36; Table S2), while the one for TEF1A was (19.56 +/-1.24; p= 1.74E-4; Table S2); confirming the presence of a PETF in the cp genome. Five strains of *T. oceanica* (CCMP0999, 1001, 1005, 1006 and 1616) were tested and in four, the PETF gene transfer to the nuclear genome was confirmed (Fig. 2 and Table 4). However, the qPCR results were not conclusive for *T. oceanica* CCMP1616 as gene copies amplification from the species-specific primers failed. *T. oceanica* CCMP1616 originates from the Mediterranean Sea whereas the four others are from the oceanic North Atlantic. The strains of *T. pseudonana* (CCMP1012, 1013, 1014 and 1335) and *T. weissflogii* (CCMP1010, 1049 and 1052) were all confirmed to have retained the PETF gene in the cp genome (Fig. 2 and Table 4).

Phylogenetic PETF tree and classification using 5.8S+ITS2

A phylogenetic tree based on the entire cp genome of diverse taxa (Fig. 3, lower section) evolved from primary endosymbiosis of the cyanobacteria *Synechocystis* (Martin *et al.*, 2002) is used here to show the PETF location in multiple taxa. The right side cluster (**A**) represents a typical $PETF_{NUC}$ genome and the left side cluster of the tree (**B**) a typical $PETF_{CP}$ genome (branches redrawn from Martin *et al.*, 2002). The *Bacillariophyceae* class, containing the Thalassiosirales, is included in the left side cluster with the typical $PETF_{CP}$ gene. As the qPCR results did not yield a clear answer for the localization of the PETF gene in *T. oceanica* CCMP1616, the *Thalassiosira* phylogenetic 18S tree was used to shed some light on the taxonomic classification of the Thalassiosirales (Fig. 4). This tree shows five major groups defining the genus which is divided into a monophyletic (upper) and a polyphyletic (lower)

cluster. Most strains are grouped with their respective species as expected but some are found within branches allocated to other spp. as seen for *T. oceanica* 1001, which clusters within the *T. pseudonana* spp., and for an unspecified *T. oceanica* within the *T. weissflogii* species. To further characterize the species used in the present study, a phylogenetic tree was created with the highly discriminating sequences of the 5.8S gene and the ITS2 region (Fig. 3 upper section). These two regions were used together in order to obtain a more detailed species identification (Moniz *et al.*, 2009, Sorhannus *et al.*, 2010). Each cluster contained in the upper section of the tree (branch annotated 90.3) demonstrated the intraspecific relationship between each species. The phylogenetic tree presented in Fig. 3 confirmed that four of the strains used in the present study were *T. oceanica* and that the *T. oceanica* CCMP1616 was more closely related to a different *Thalassiosira* spp. and may need to be reclassified. The 5.8S gene sequences are generally very similar between spp. of the same genus, thus a BLAST analysis of the 5.8S sequence from *T. oceanica* CCMP 1616 against the NCBI database suggest that this strain is more closely related to *Thalassiosira punctigera* (Castacane) Hasle (JQ217344.1-99% identity), *Thalassiosira rotula* (CCMP 1018 and 1047) Meunier (EU600777.1 and EF208798.1-99% identity), and *Thalassiosira aestivalis* Gran and Angst (EF208797.1-99% identity) than to the *T. oceanica* strains used in this study. Finally, the clusters produced for *T. pseudonana* and *T. weissflogii* also confirm the monophyletic origin of each strain in their respective spp..

DISCUSSION

Phylogenetic classification

The *Thalassiosira* 18S tree presents multiple major phylogenetic clusters (Fig. 4) with the two upper groups containing most spp. of *Thalassiosira,* all originating from a single common ancestor and making the upper section of the 18S tree a monophyletic *Thalassiosira* cluster. The lower sections of the tree contain the two atypical *Thalassiosira* spp., *T. pseudonana* and *T. weissflogii*, together with some other *Thalassiosira* spp. and several non-*Thalassiosira* spp. (*Cyclotella* spp.) exhibiting a classic polyphyletic cluster. This 18S clusters pattern correspond to the ones previously established with one cluster containing multiple *Thalassiosira* spp. and separated *T. pseudonana, T. weissflogii* and *Cyclotella* spp. clusters (Luddington *et al.*, 2012); thus confirming prior conclusions concerning diverse *Thalassiosira*'s ancestral situation. The different ancestral origins of the presented

Thalassiosira group can be connected to the early arbitrary classification of phytoplankton, which was for the diatoms mostly based on the physical appearance of the architecture of frustules. Most *T. oceanica* strains have not yet been sequenced and are therefore not included in databases; leaving the *T. oceanica* branch containing only one representative of the strain 1001, one of unknown strain (HM991) and the strain (1005) sequenced by Lommer *et al.* (2010). The taxonomic identity of the strain called *T. oceanica* CCMP1616 was doubted after observation of our qPCR results and to clarify the confusion about this organism, a cladogram was done using the sequences of the 5.8S+ITS2 region (Fig. 3) obtained during the present experiment. This region was selected for its high sequencing resolution and because it is known to be a good eukaryotic barcoding region (Coleman, 2003, Coleman, 2009, Sorhannus *et al.*, 2010, Caisova *et al.*, 2011, Senapin *et al.*, 2011, Whittaker *et al.*, 2012). Our results clearly demonstrate that CCMP1616 does not belong to the *T. oceanica* as it is not part of the monophyletic *T. oceanica* branch but clusters as an independent branch in the tree. To clarify the phylogenetic situation of "*T. oceanica* CCMP1616", blast searches with the 5.8S DNA sequence against the NCBI NR database placed this strain in the upper *Thalassiosira* 18S cluster as all its close relatives are contained in it; resolving some obscurity around this sp. (Fig. 3). We further compared our results to the 5.8S and ITS phylogenetic trees drawn by von Dassow *et al.* (2008), confirming that most tested strains are grouped within their respective monophyletic groups as previously established. Fortunately, the uses of new sequencing technique will resolve many of these misidentifications.

PETF gene transfer

The gene coding for PETF is typically found in the cp of the *Rhodophyta*, *Glaucophyta*, *Cryptophyta* and *Bacillariophyceae* (diatoms) whereas some other groups, *Alveolata*, *Chlorophyta* and *Streptophyta*, were shown to have this gene in their nuclear genome (Martin *et al.*, 2002, Lommer *et al.*, 2010). The groups with the nuclear PETF have a smaller cp genome ranging from ≈30-100 protein genes versus ≈≥140 for the ones which retained PETF in the cp (Kleffmann *et al.*, 2004, Lommer *et al.*, 2010). This gene content reduction is believed to result from gene loss during gene transfer causing genomic rearrangements or just by the transfer of the genes themselves (Leister, 2005, Lommer *et al.*, 2010). Indeed, when insertion of organelle DNA to the nucleus happens, a major restructuring follows from evolutionary pressures to reach the less costly and most functional possible version of the organism (Timmis *et al.*, 2004, Leister, 2005).

Stramenopiles (including *Bacillariophyceae*) typically have a $PETF_{CP}$ however, it was previously discovered that in *T. oceanica* CCMP1005, a diatom with the typically large cp genomic content of Stramenopiles (≈140 cp genes), the PETF gene was relocated in the nuclear genome and replaced by the gene orf127 in the cp genome (Lommer *et al.*, 2010). It further indicated that the transferred gene is functional. The gene transfer may have been facilitated by the intrinsic features of the $PETF_{CP}$ gene flanking regions. This functionally competent transfer might be explained by an energy economy because having a gene in the nuclear instead of cp genome could save some management energy (Allen *et al.*, 2005, Woodson *et al.*, 2008). Moreover, large scale cp genome reduction should benefit as well from an improved regulation of the transferred genes (Woodson *et al.*, 2008) as it is known that the transfer process could possibly keep the genes in both places creating a duplicate-gene; thus raising a more complicated management of gene expression regulation and the protein level control as well (Oudot-Le Secq *et al.*, 2007, Woodson *et al.*, 2008). We can further speculate that having both gene coding for flavodoxin and PETF located in the same region, typically the cps, could complicate the alternate regulation of these proteins and make it more economic to have one in the cp and one in the nucleus. It was previously demonstrated that when Fe is limited, flavodoxin is up-regulated in many species of diatoms in order to continue an efficient control of the electron transfer (LaRoche *et al.*, 1995). This type of plasmid-to-nucleus transfer is only possible in organisms containing multiple organelles (Wang *et al.*, 2012), here multiple cp, and this cp DNA lysis is considered a very important source for nuclear genome renewal or adaptation (Allen *et al.*, 2005). Woodson (2008) argued that plasmid-to-nucleus gene transfers are advantageous as they allow simpler genome regulation when successfully done as it is for *T. oceanica*; the possible logistic challenges of controlling separated genome system (organellar versus nuclear) by the cell therefore avoided. Present qPCR results showed that all *T. oceanica* strains tested (except CCMP 1616) had their PETF gene amplification values corresponding to a location in the nuclear genome ($PETF_{NUC}$ ≈ TEF1A, Table S2) whereas the PETF gene of all tested *T. pseudonana* and *T. weissflogii* strains correlated with a cp gene location, presenting a non transfer ($PETF_{CP}$ ≈ rbcs, Table S2). Organellar DNA transfers were demonstrated to be continuously in progress and more frequent than anticipated (Fujita *et al.*, 2008, Kleine *et al.*, 2009) and thus, even if only a handful of research proved such transfer (cp-to-nuclear in transgenic tobacco plant and mitochondrial-to-nucleus in yeast (Thorsness *et al.*, 1990, Huang *et al.*, 2003). New findings provided further evidences that high temperature could be the driving condition for functional gene transfer (Wang *et al.*, 2012). Indeed, temperature stress was connected to the increased

accessibility of chromatin regions which were identified as the regions where successful organelle-to-nucleus transfers take place (Wang *et al.*, 2013). In the present *Thalassiosira* spp. case, the investigated species evolve in two different environmental temperatures; *T. oceanica* grows in warmer water of 22-26°C whereas *T. pseudonana* and *T. weissflogii* in 11-16°C water. This difference but mostly the increase of temperature initially experienced by T. *oceanica* might have created the type of stress related to the improved accessibility of chromatin and thus might have led to the successful PETF transfer.

Lommer *et al.* (2010) demonstrated that looking thoroughly at the cp genome can help solving some uncertainties related to gene transfers but here we showed that the use of simple qPCR analysis with critically well designed primers can in the future be used for such observation. The reasons for successful organellar gene transfer are still unknown and need to be further investigated; however, some connections to ecological advantages through evolution are made. This capacity to adapt or even just to acclimate to a new environment will directly influence the composition of marine communities and their effect on the oceanic ecosystems (Ocean Studies Board, 2010). Diatoms are well known for their ability to adapt successfully to new environmental conditions (von Dassow *et al.*, 2008) and since one of the most successful diatom classes is *Thalassiosira* (Round, 1990, Kaczmarska *et al.*, 2006), their response to environmental trigger to form great blooms in any oceanic water around the globe must be understood to predict their reaction to major environmental changes (e.g. ocean acidification and CO_2 sequestration).

CONCLUSION

The present study showed that all cultivated and analysed algae belong to the *Thalassiosira* genus and were correctly classified except strain CCMP1616, which does not appear to be a *T. oceanica*. Within these three closely related species, only *T. oceanica* underwent a PETF gene transfer from cp to nuclear genome. It also demonstrated that qPCR is an appropriate tool for determining such gene transfer in closely related species and strains for which complete genome information is not available; therefore giving an insight into cp genome reduction evolutionary processes. Our results showed that the PETF transfer did not occur in *T. pseudonana* and *T. weissflogii* and confirms that the PETF gene transfer happened after their species differentiation. The high tolerance to severe Fe limitation and the capacity to

grow in higher temperature of *T. oceanica* is shown in multiple strains of this species and demonstrate that this transfer probably occurred once in the ancestor of all present days *T. oceanica*. The new hypothesis that temperature might affect the accessibility of chromatin and its involvement in organelle-to-nucleus gene transfer (Wang *et al.*, 2013) brings some insight into the mechanisms for such transfer. This hypothesis further supports that the functional transfer of the PETF gene to the nuclear genome in *T. oceanica* may have originated from *T. oceanica*'s preference for higher growth temperature, coupled with chronically low Fe concentrations in warm, oligotrophic surface waters.

ACKNOWLEDGEMENTS

We would like to thank Tania Klüver and Benjamin Weigel for their help with the laboratory culturing of the algae. We are grateful to the Institute of Clinical Molecular Biology in Kiel for providing Sanger sequencing. Finally, we also thank L.V. Smith, D. Lemke and several anonymous reviewers that significantly improved the manuscript. This work was supported by BMBF project BIOACID (D10/1.1.2) and DFG grant RO3128/6-1 to JLR.

REFERENCES

Allen, J. F., Puthiyaveetil, S., Strom, J. & Allen, C. A. 2005. Energy transduction anchors genes in organelles. *Bioessays* **27**:426-35.

Bock, R. & Timmis, J. N. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays* **30**:556-66.

Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. A., Buesseler, K. O., Coale, K. H., Cullen, J. J., de Baar, H. J. W., Follows, M., Harvey, M., Lancelot, C., Levasseur, M., Owens, N. P. J., Pollard, R., Rivkin, R. B., Sarmiento, J., Schoemann, V., Smetacek, V., Takeda, S., Tsuda, A., Turner, S. & Watson, A. J. 2007. Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions. *Science* **315**:612-17.

Caisova, L., Marin, B. & Melkonian, M. 2011. A close-up view on ITS2 evolution and speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). *Bmc Evolutionary Biology* **11**.

Chisholm, S. W., Falkowski, P. G. & Cullen, J. J. 2001. Oceans - Dis-crediting ocean fertilization. *Science* **294**:309-10.

Coleman, A. W. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends Genet.* **19**:370-75.

Coleman, A. W. 2009. Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution* **50**:197-203.

Dugdale, R. C. & Wilkerson, F. P. 1998. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* **391**:270-73.

Fujita, K., Ehira, S., Tanaka, K., Asai, K. & Ohta, N. 2008. Molecular phylogeny and evolution of the plastid and nuclear encoded cbbX genes in the unicellular red alga Cyanidioschyzon merolae. *Genes Genet. Syst.* **83**:127-33.

Gueneau, P., Morel, F., Laroche, J. & Erdner, D. 1998. The petF region of the chloroplast genome from the diatom Thalassiosira weissflogii: sequence, organization and phylogeny. *European Journal of Phycology* **33**:203-11.

Guillard, R. R. & Ryther, J. H. 1962. Studies of marine planktonic diatoms .1. cyclotella nana hustedt, and detonula confervacea (cleve) gran. *Can. J. Microbiol.* **8**:229-&.

Guillard, R. R. L. 1975. Culture of phytoplankton for feeding marine invertebrates. *In*: Smith, W. L. a. C., y M.H. [Ed.] *Culture of Marine Invertebrate Animals.* Plenum Press, New York, USA., pp. 26-60.

Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**:72-76.

Kaczmarska, I., Beaton, M., Benoit, A. C. & Medlin, L. K. 2006. Molecular phylogeny of selected members of the order Thalassiosirales (Bacillariophyta) and evolution of the fultoportula (vol 42, pg 121, 2006). *J. Phycol.* **42**:1373-73.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W. & Baginsky, S. 2004. The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* **14**:354-62.

Kleine, T., Maier, U. G. & Leister, D. 2009. DNA Transfer from Organelles to the Nucleus: The Idiosyncratic Genetics of Endosymbiosis. *Annual Review of Plant Biology*. pp. 115-38.

Kumar, M. & Shukla, P. K. 2005. Use of PCR targeting of internal transcribed spacer regions and single-stranded conformation polymorphism analysis of sequence variation in different regions of rRNA genes in fungi for rapid diagnosis of mycotic keratitis. *J. Clin. Microbiol.* **43**:662-68.

LaRoche, J., Murray, H., Orellana, M. & Newton, J. 1995. Flavodoxin expression as an indicator of iron limitation in marine diatoms. *J. Phycol.* **31**:520-30.

Leister, D. 2005. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* **21**:655-63.

Lommer, M., Roy, A. S., Schilhabel, M., Schreiber, S., Rosenstiel, P. & LaRoche, J. 2010. Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics* **11**.

Lommer, M., Specht, M., Roy, A.-S., Kraemer, L., Andreson, R., Gutowska, M. A., Wolf, J., Bergner, S. V., Schilhabel, M. B., Klostermeier, U. C., Beiko, R. G., Rosenstiel, P., Hippler, M. & LaRoche, J. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology* **13**.

Luddington, I. A., Kaczmarska, I. & Lovejoy, C. 2012. Distance and Character-Based Evaluation of the V4 Region of the 18S rRNA Gene for the Identification of Diatoms (Bacillariophyceae). *Plos One* **7**:e456664.

Maldonado, M. T. & Price, N. M. 1996. Influence of N substrate on Fe requirements of marine centric diatoms. *Marine Ecology Progress Series* **141**:161-72.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. & Penny, D. 2002. Evolutionary analysis of Arabidopsis,

cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **99**:12246-51.

McKay, R. M. L., Geider, R. J. & LaRoche, J. 1997. Physiological and biochemical response of the photosynthetic apparatus of two marine diatoms to Fe stress. *Plant Physiol.* **114**:615-22.

Moniz, M. B. J. & Kaczmarska, I. 2009. Barcoding diatoms: Is there a good marker? *Mol. Ecol. Resour.* **9**:65-74.

Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. 1995. Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles* **9**:359-72.

Ocean Studies Board 2010. Ocean Acidification: A National Strategy to Meet the Challenges of a Changing Ocean. *In*: Press, T. N. A. [Ed.]. National Academy of Sciences, Washington, D.C., USA.

Oudot-Le Secq, M. P., Grimwood, J., Shapiro, H., Armbrust, E. V., Bowler, C. & Green, B. R. 2007. Chloroplast genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana: comparison with other plastid genomes of the red lineage. *Mol. Genet. Genomics* **277**:427-39.

Peers, G. & Price, N. M. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441**:341-44.

Pelham, R. J., Rodgers, L., Hall, I., Lucito, R., Nguyen, K. C. Q., Navin, N., Hicks, J., Mu, D., Powers, S., Wigler, M. & Botstein, D. 2006. Identification of alterations in DNA copy number in host stromal cells during tumor progression. *Proc. Natl. Acad. Sci. U. S. A.* **103**:19848-53.

Pitti, R. M., Marsters, S. A., Lawrence, D. A., Roy, M., Kischkel, F. C., Dowd, P., Huang, A., Donahue, C. J., Sherwood, S. W., Baldwin, D. T., Godowski, P. J., Wood, W. I., Gurney, A. L., Hillan, K. J., Cohen, R. L., Goddard, A. D., Botstein, D. & Ashkenazi, A. 1998. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**:699-703.

Richly, E. & Leister, D. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Molecular Biology and Evolution* **21**:1972-80.

Round, F. E. 1990. Diatom communities - their response to changes in acidity. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **327**:243-49.

Senapin, S., Phiwsaiya, K., Kiatmetha, P. & Withyachumnarnkul, B. 2011. Development of primers and a procedure for specific identification of the diatom Thalassiosira weissflogii. *Aquaculture International* **19**:693-704.

Sheppard, A. E. & Timmis, J. N. 2009. Instability of Plastid DNA in the Nuclear Genome. *Plos Genetics* **5**.

Sorhannus, U., Ortiz, J. D., Wolf, M. & Fox, M. G. 2010. Microevolution and Speciation in Thalassiosira weissflogii (Bacillariophyta). *Protist* **161**:237-49.

Strzepek, R. F. & Harrison, P. J. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* **431**:689-92.

Sunda, W. G., Swift, D. G. & Huntsman, S. A. 1991. Low iron requirement for growth in oceanic phytoplankton. *Nature* **351**:55-57.

Thorsness, P. E. & Fox, T. D. 1990. Escape of DNA from mitochondria to the nucleus in saccharomyces-cerevisiae. *Nature* **346**:376-79.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**:123-U16.

von Dassow, P., Petersen, T. W., Chepurnov, V. A. & Armbrust, E. V. 2008. Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus Thalassiosira (Bacillariophyceae). *J. Phycol.* **44**:335-49.

Wang, D., Lloyd, A. H. & Timmis, J. N. 2012. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proc. Natl. Acad. Sci. U. S. A.* **109**:2444-48.

Wang, D. & Timmis, J. N. 2013. Cytoplasmic Organelle DNA Preferentially Inserts into Open Chromatin. *Genome Biology and Evolution* **5**:1060-64.

Whittaker, K. A., Rignanese, D. R., Olson, R. J. & Rynearson, T. A. 2012. Molecular subdivision of the marine diatom Thalassiosira rotula in relation to geographic distribution, genome size, and physiology. *Bmc Evolutionary Biology* **12**.

Woodson, J. D. & Chory, J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* **9**:383-95.

Table 1. Tested *Thalassiosira* strains environmental details (n/a = not available, Env. T = environmental temperature).

Species	Strain CCMP	Env. T °C	Origin	Environment	Taxonomic authors
T. oceanica	999	22-26	Continental slope; North Atlantic	warm core eddy	(Hustedt) Hasle & Heimdal
T. oceanica	1001	22-26	Continental slope; North Atlantic	n/a	(Husdedt) Hasle & Heimdal
T. oceanica	1005	22-26	Sargasso Sea	oceanic	(Husdedt) Hasle & Heimdal
T. oceanica	1006	22-26	Sargasso Sea	oceanic	(Husdedt) Hasle & Heimdal
T. oceanica	1616	22-26	Eastern Mediterranean Sea	n/a	(Husdedt) Hasle & Heimdal
T. pseudonana	1012	11-16	Swan River Estuary, Perth, Australia	estuarine	(Husdedt) Hasle & Heimdal
T. pseudonana	1013	11-16	Conwy, Gwynedd, Wales, United Kingdom	estuarine	(Husdedt) Hasle & Heimdal
T. pseudonana	1014	11-16	North Pacific Gyre	n/a	(Husdedt) Hasle & Heimdal
T. pseudonana	1335	11-16	Long Island, NY, USA	estuarine shallow bay	(Husdedt) Hasle & Heimdal
T. weissflogii	1010	11-16	Stream, between Bermuda and New York	oceanic	(Husdedt) Hasle & Heimdal
T. weissflogii	1049	11-16	Long Island Sound Sea, NY, USA	n/a	Fryxell & Hasle
T. weissflogii	1052	11-16	Skagerrak sea, Oslo Fjord, Norway	n/a	Fryxell & Hasle

Table 2. Oligonucleotides primers used for qPCR analysis and PCR amplification for the ITS1-5.8S-ITS2 region used for Sanger sequencing

Species	**Gene**	**Forward primers (5'→ 3')**	**Reverse primer(5'→ 3')**
Thalassiosira sp.	18S	GGATTTCTGGCAGGAGCGA	GCCCGACAACTTAATGCCAA
T. oceanica	$PETF_{NUC}$	GAGGTTGACCAGTCCGAGCA	GGATCTCGCAGTCCGACTTG
T. pseudonana/weissflogii	$PETF_{CP}$	CGTGCTGGTGCTTGTTCTACAT	CAAAACCAGCACCCATTTGAT
T. oceanica	TEF1A	GTCCGTCTCCCCATCTCCA	CAACAACATCTCCAGGGACGA
T. pseudonana	TEF1A	CCACTACACCATTGTCGATGCT	GGCAGGCACAAGAAGAAGACC
T. weissflogii	TEF1A	CTTCCGGTTCGTCTTCCCAT	CATCTCCCGGACGAAGAGTTC
T. oceanica	rbcs	TAGAGTGGACAGATGATCCGCA	AATTCAAACATTACTGTCGCAGGA
T. pseudonana	rbcs	TAGAATGGACAGATGATCCACA	AATTCGAACATTACTGTTGCAGGA
T. weissflogii	rbcs	TAGAGTGGACAGATGATCCACA	AATTCGAACATTACTGTCGCAGGA
T. oceanica	ITS1-5.8S-ITS2	TCCGTAGGTGAACCTGCGG	GCTTAAATTCGGCGGGTAGTCT
T .pseudonana/weissflogii	ITS1-5.8S-ITS2	TCCGTAGGTGAACCTGCGG	TCCTCCGCTTATTAATATG

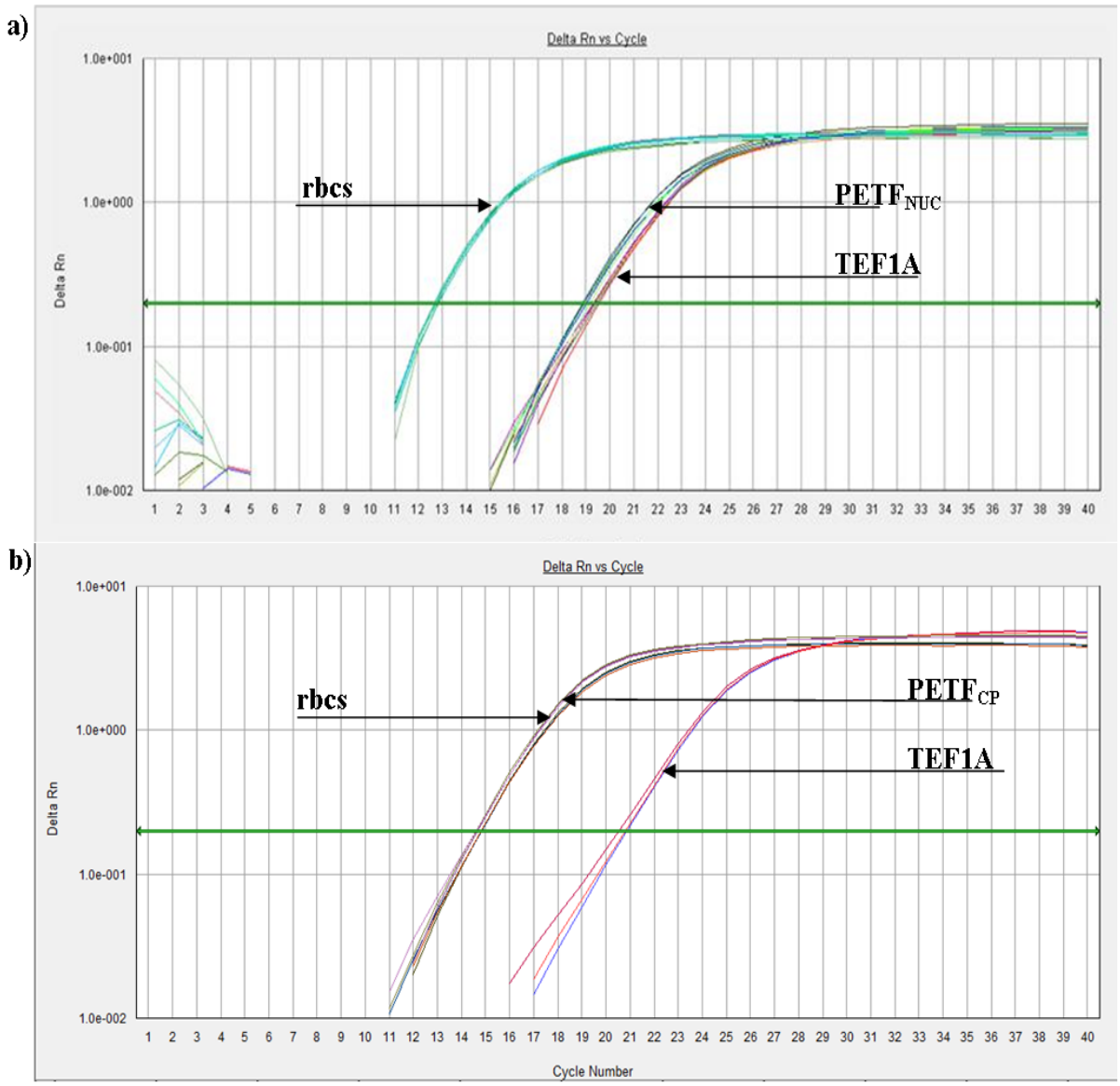


Fig. 1. qPCR amplification curves demonstrating the location of the PETF gene respectively in the nucleus a) *T. oceanica* (e.g. 1006) and in the cp *T. pseudonana* and b) *T. weissflogii* (e.g. 1049).

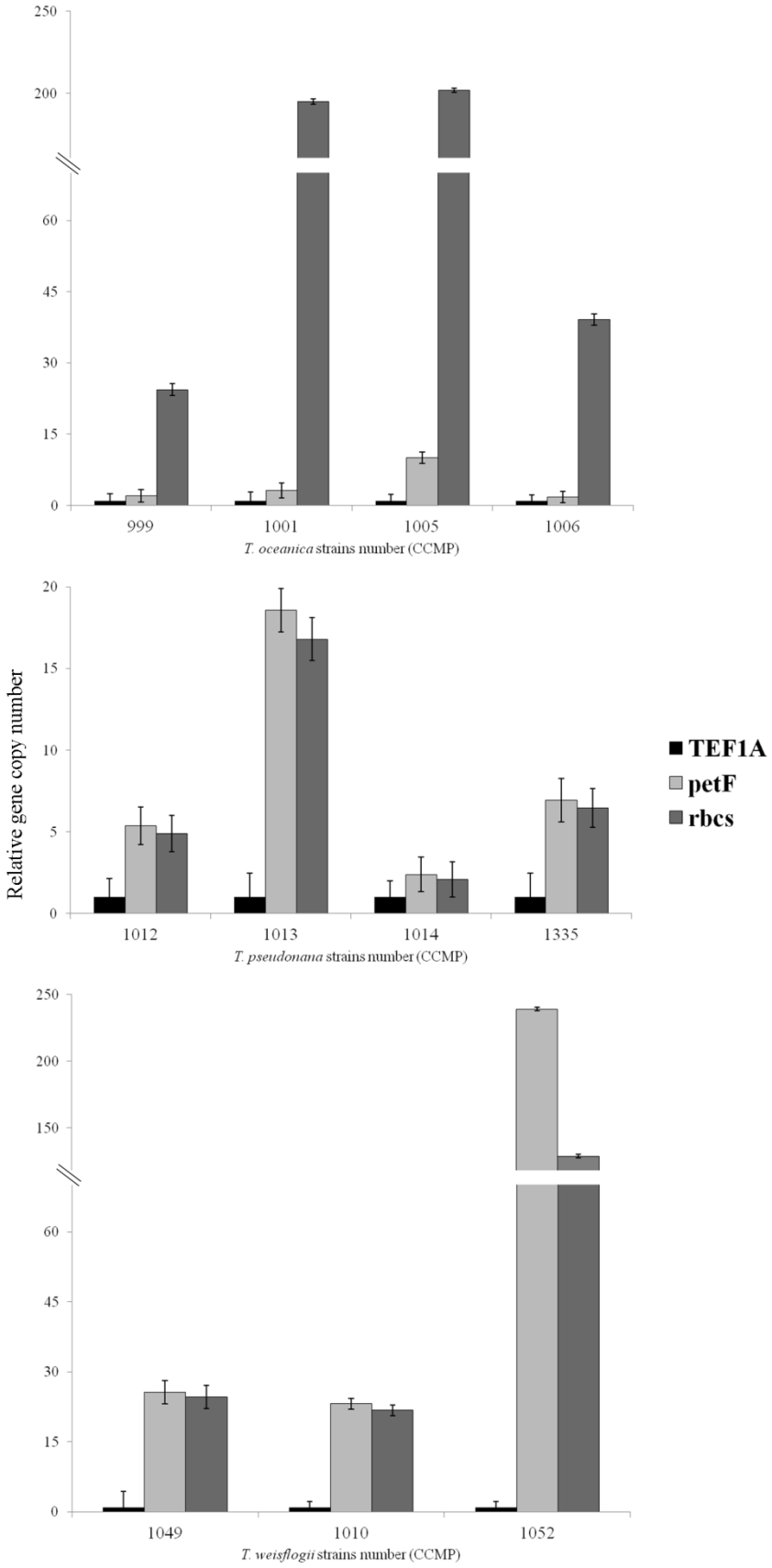


Fig. 2. Relative gene copy number obtained using $2^{\Delta Ct}$ +/-SE for three studied *Thalassiosira* with only one sp., *T. oceanica*, having a nuclear $PETF_{NUC}$ (PETF ≈ TEF1A).

Table 4. Studied gene details for each *Thalassiosira* cultured under growth preferences of 24.6°C at 70-90 µE for *T. oceanica* and of 16°C at 100 µE for *T. pseudonana* and *T. weissflogii* under14/10 hr and 12/12 hr light/dark cycles, respectively.

Species	**Strain**	**Gene**	**Copy number**	**Location**	**PETF status**
T. oceanica	0999, 1001,	PETF	single copy	nucleus	transferred
	1005, 1006	TEF1A	single copy	nucleus	
		18S	multiple copy	nucleus	
		rbcs	single copy	cp	
Unknown	1616	PETF	unclear	unclear	unclear
Thalassiosira		TEF1A	unclear	unclear	
		18S	multiple copy	nucleus	
		rbcs	single copy	cp	
T. pseudonana	1012, 1013,	PETF	multiple copy	cp	not transferred
	1014, 1335	TEF1A	single copy	nucleus	
		18S	multiple copy	nucleus	
		rbcs	single copy	cp	
T. weissflogii	1010, 1049,	PETF	multiple copy	cp	not transferred
	1052	TEF1A	single copy	nucleus	
		18S	multiple copy	nucleus	
		rbcs	single copy	cp	

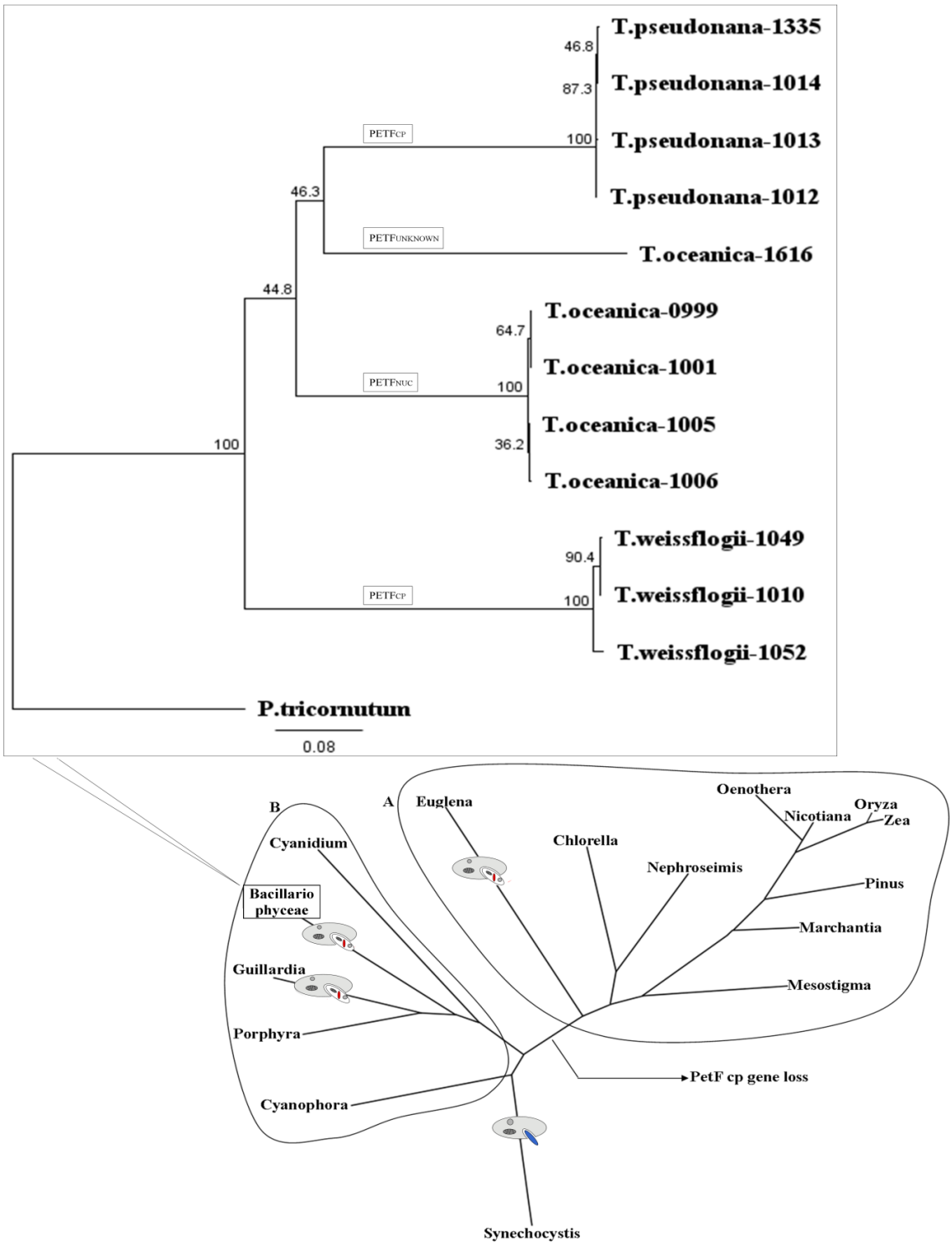


Fig. 3. Phylogenetic tree of the PETF's location redrawn from Martin *et al.* (2002), where **A** shows *Chlorophyta-Plantae* division with a $PETF_{NUC}$ and **B** the *Rhodophyta* and taxa evolved from second endosymbyosis of a *Rhodophyta* with a $PETF_{CP}$. *Bacillariophyceae* taxon leads to 5.8S+ITS2 phylogenetic tree of experimentally cultured *Thalassiosira* with *Pheaodactylum tricornutum* as outgroup.

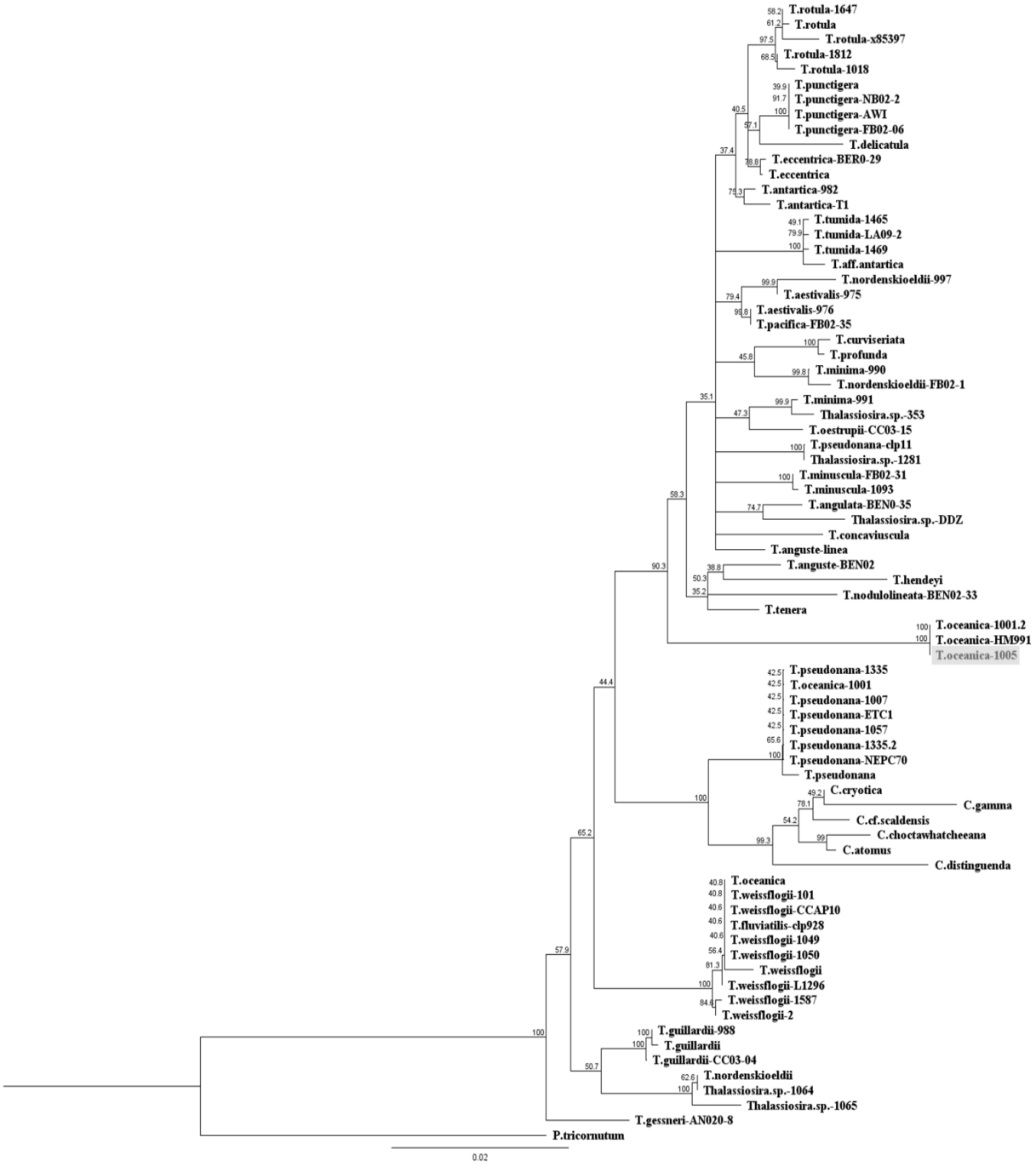


Fig. 4. Interspecific phylogenetic tree of genus *Thalassiosira* based on 18S gene sequences from databases and one experimentally acquired (grey highlight) with *Cyclotella* spp. and Phaeodactylum tricornutum as outgroup.

Table S1. CLUSTALW sequences alignments of the 5.8S+ITS2 regions.

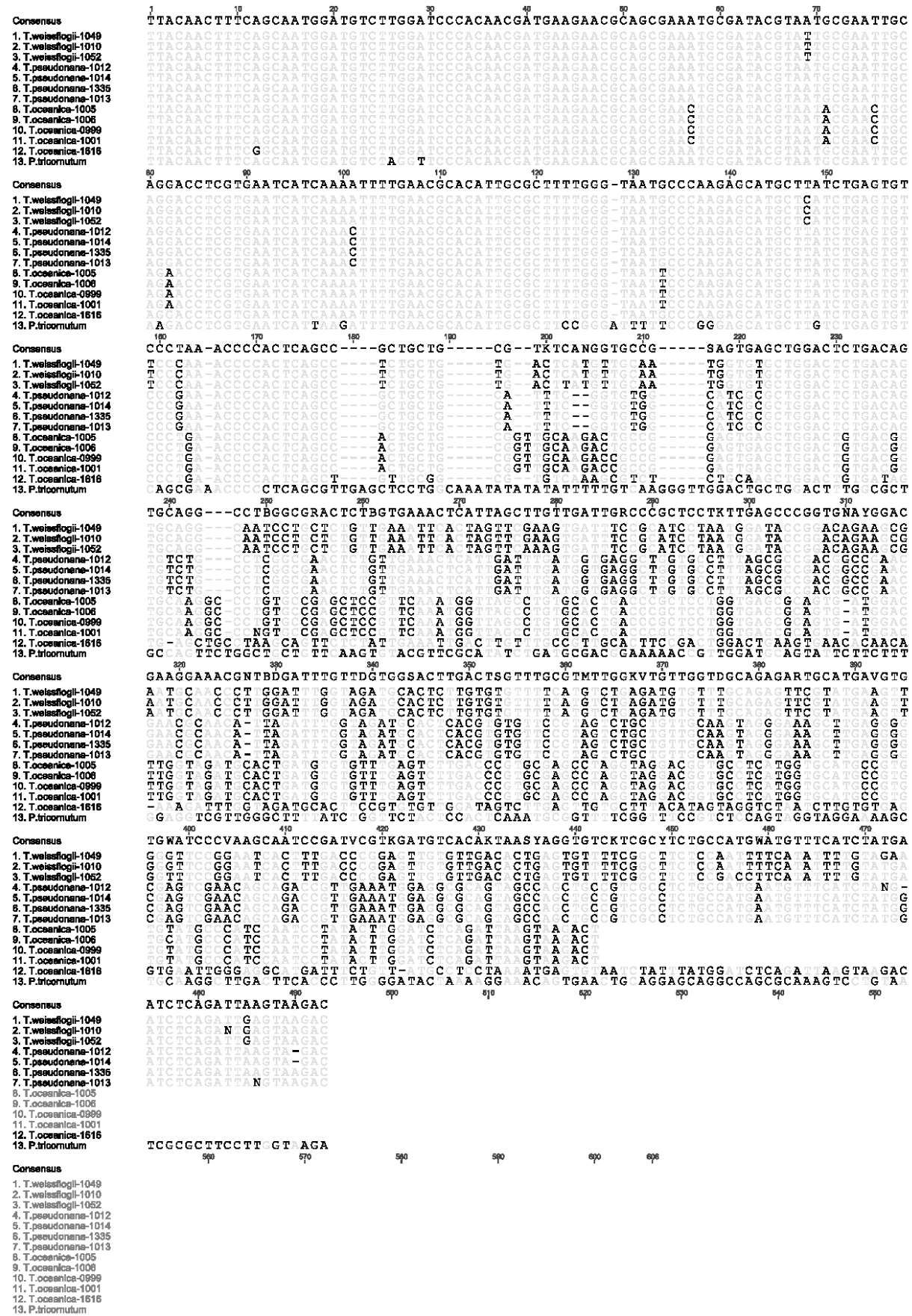


Table S2. Paired two samples for means t-test with p value between PETF and each reference gene for 2 strains presented in Fig. 1. Significant p <0.01.

Strains	**Gene**	**Mean Ct**	**SD**	**p value**
1006	PETF	18.79	0.12	
	rbcs	14.28	0.03	**3.71E-08**
	TEF1A	19.57	0.34	0.01
1049	PETF	14.88	0.45	
	rbcs	14.94	14.94	0.36
	TEF1A	19.55	19.56	**1.74E-04**

4.2.3 Publication IV.

Effect of increased pCO_2 on the transcriptome of *Thalassiosira oceanica* (CCMP1005).

Roy *et al.*,
in prep.

Effect of increased pCO_2 on the transcriptome of *Thalassiosira oceanica* (CCMP1005).

Alexandra-Sophie Roy[1], Dhwani Desai[3], Markus Lommer[1], Matthias Barann[2], M. Schillhabel[2], Philip Rosenstiel[2], Julie LaRoche[1,3]

[1]GEOMAR – Helmholtz-Zentrum für Ozeanforschung Kiel, Düsternbrooker Weg 20, 24105, Kiel, Germany. [2]Institut für Klinische Molekularbiologie der Christian-Albrechts-Universität zu Kiel, Arnold-Heller-Straße 3, 24105, Kiel, Germany. [3]Department of Biology, Dalhousie University, Halifax, N.S. Canada.

ABSTRACT

The increased intake and dissolution of anthropogenic CO_2 in our oceans result in a decrease in pH and carbonate ion concentration. The OA process creates new ecological pressures that impact the entire marine ecosystem from primary producers to large marine organisms. To further our understanding of the impact of OA on diatoms, axenic diluted batch cultures of *Thalassiosira oceanica* were exposed to 380 and 1000 µatm to determine the effect of CO_2 levels on gene expression. Messenger RNA from replicates cultures were sequenced with the Illumina HiSeq2000, producing transcriptomes for ambient and elevated CO_2 treatments. Multiple unknown or predicted proteins were up-regulated in either one of the CO_2 treatments. From the known protein-coding genes, FCP's, NADH dehydrogenase and TPR-Sel1 protein were significantly up-regulated in 1000 µatm, while stress indicating (pol protein) and cell structural proteins were up-regulated in 380 µatm. Genes involved in different carbon or transport mechanisms found to be differentially regulated in function of pCO_2 in *Emiliania huxleyi* did not change significantly, as a function of CO_2 levels, in *T. oceanica*. The fact that elevated CO_2 does not seem to induce major changes in *T. oceanica* suggest that this species may adapt or acclimate to future ocean CO_2 levels. However, further research to determine the function of unknown and predicted proteins is needed as it is essential to evaluate OA's effect on diatoms, which are among the most important primary producers in the oceans.

Keywords: Diatoms, ocean acidification, elevated CO_2, *Thalassiosira oceanica*, transcriptome, Illumina HISeq.

1 INTRODUCTION

Since the pre-industrial era, anthropogenic atmospheric carbon dioxide (CO_2) has been increasing rapidly and is expected to reach a partial pressure as high as ≈1000 µatm by the end of the 21st century (Plattner *et al.*, 2001, Friedlingstein *et al.*, 2006). The increased absorption of CO_2 in the oceans, ≈30-40% (Hoegh-Guldberg *et al.*, 2010, Hoegh-Guldberg, 2010), reacts with water molecules creating bicarbonate (HCO_3^-) and hydrogen (H^+) ions. These newly released H^+ ions accumulate, decrease the surface ocean pH (predicted to decrease by 0.25-0.45 units before the end of the century; Orr *et al.*, 2005), and thus decrease the availability of carbonate ions (CO_3^{2-}) (Wolf-Gladrow *et al.*, 1999, Caldeira *et al.*, 2003, Newbold *et al.*, 2012). These chemical alterations, coined ocean acidification (OA), may affect the physiology of marine organisms and their roles in the global biogeochemical cycles and ocean biogeochemistry (Falkowski *et al.*, 1998, Worden *et al.*, 2008).

The evaluation of the detailed effects of this increased CO_2 release and the OA results has been an active area of marine research (Feely *et al.*, 2004, Orr *et al.*, 2005, Shi *et al.*, 2010, Richardson *et al.*, 2012, Rossoll *et al.*, 2012, Brussaard *et al.*, 2013, Melzner *et al.*, 2013, Schulz *et al.*, 2013). The calcification processes in calcifying organisms like coral, foraminifera, pteropodes, urchins, etc. (Kleypas *et al.*, 2006, Ries *et al.*, 2009, Pespeni, 2013) have been found to be negatively affected. *Emiliania huxleyi* for example, increased its photosynthesis under high CO_2 pressure (pCO_2), high temperatures and reduced light intensities (Feng *et al.*, 2008) but decreased its cocoliths formation (Mackinder *et al.*, 2011, Bach *et al.*, 2013, Riebesell *et al.*, 2011). The diazotrophic cyanobacterium *Crocosphaera sp*. also increased photosynthesis and N_2 fixation rates under high CO_2 and elevated iron (Fe) (Fu *et al.*, 2008). Evolutionary studies on organisms with short generation times found that different populations (natural and laboratory) of *Chlamydomonas sp.* did not developed any adaptations to elevated CO_2 but that reaction to elevated pCO_2 was attributable to ancestral and chances events (Collins *et al.*, 2006a, Collins *et al.*, 2006b). A more extensive study on *E. huxleyi* (Lohbeck *et al.*, 2012) found that most strains could maintain their normal functionality via evolution. Marine diatoms have also been intensively investigated, however, most efforts were concentrated on the effect of elevated pCO_2 on their carbon-concentrating mechanism (CCM). Indeed, it was found in previous studies that the CCM was down-regulated by high CO_2 (Wu *et al.*, 2010, Hopkinson *et al.*, 2011, Young *et al.*, 2012, Hopkinson *et al.*, 2013) and that organisms would be limited in CO_2 without its activity

during low CO_2 concentration (Riebesell *et al.*, 1993, Hopkinson *et al.*, 2011, Allen *et al.*, 2012, Reinfelder *et al.*, 2004). Studies on diatoms have found that for both *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* exposure to high pCO_2 resulted in higher growth rates (Sobrino *et al.*, 2008, Wu *et al.*, 2010). Increase in photosynthetic rates was also observed in natural assemblages of phytoplankton subjected to high CO_2, however, this was accompanied by a shift in the community structure favouring large chain-forming species (Tortell *et al.*, 2008). Laboratory culture work is important for correctly evaluating the response of single species to elevated CO_2 as well as the comparison of unialgal culture response with that of natural communities, such as found in mesocosms studies.

The recent advances in high-throughput molecular techniques enable researchers to obtain global snapshots of the transcriptome of an organism or microbial community in response to changes in environmental conditions (reviewed by Thomas *et al.*, 2012). Here, we used high-throughput sequencing to analyse the effect of CO_2 on *T. oceanica*, an ecologically relevant diatom common in oligotrophic oceans. Diatoms are one of the most ecologically successful eukaryotes, accounting for ≈40% of the yearly oceanic primary production. Indeed, diatom species are found in a wide range of environments including High-Nutrient Low-Chlorophyll (HNLC) regions, which are Fe limited, demonstrating that they have acquired adaptation to tolerate Fe limitation (Strzepek *et al.*, 2004).

In HNLC regions, the importance of Fe for phytoplankton and diatoms is easily demonstrable by its addition which stimulates phytoplanktonic growth or more precisely, triggers major diatom blooms (Chisholm *et al.*, 2001, Boyd *et al.*, 2007, Allen *et al.*, 2008). This rapid phytoplanktonic growth can result in an increased carbon sequestration to the deep ocean by mass sinking (Timmermans *et al.*, 2001, Raven *et al.*, 2004, Allen *et al.*, 2008, Pollard *et al.*, 2009). More precisely, this increased carbon sink is created by the aggregation tendency of high density diatoms which tend to die rapidly (nutrients or space limitation - Bidle *et al.*, 2004, Bidle *et al.*, 2008), thus carrying the carbon in their cells to the ocean floors. Once on the ocean floor, the cells and carbon are sedimented (Hoffmann *et al.*, 2006) and trapped for thousands of years (Smetacek *et al.*, 2008, Buesseler *et al.*, 2004); consequently increasing the carbon ocean sinks. However, recent work suggests that the bioavailability of Fe will be reduced in an acidic ocean (Shi *et al.*, 2010), which is believed to have consequences on the formation of large phytoplankton blooms by, for example, impairing the diatoms and cocolithophores Fe uptake (grown in acidified media) without reducing their cellular

requirement (Shi *et al.*, 2010). The presented work investigated the response of the centric diatom *T. oceanica* (CCMP 1005) grown under present day, ≈380 µatm, and predicted end of the century, ≈1000 µatm, pCO_2. *T. oceanica* was chosen because it is an ecologically important species that is tolerant to Fe limitation and inhabits oligotrophic waters where conditions are stable. In addition, its genome is available (Lommer *et al.*, 2010, Lommer *et al.*, 2012) permitting detailed transcriptomic studies for this diatom species.

2 MATERIAL AND METHODS

2.1 Algae strains and cultures

Axenic clonal pre-acclimated isolates of *T. oceanica* CCMP 1005 were cultured in 5 L bottles filled with f/8 artificial sea water (Kester *et al.*, 1967) supplied with 10 µM $FeCl_3$ at 100 µE, 24.6°C and a 14/10 h light/dark cycle. Pre-mixed CO_2 (CO_2 - N45 and synthetic air - N50) with pCO_2 of 380 and 1000 µatm (Air Liquide, Düsseldorf, Germany) was constantly bubbled (very low in-flow) in the culture vessel using special axenic Fe-free airtight lids and lines. The bubbling rates were visually controlled by counting the number of bubbles released per 30 sec and were equilibrated between bottles. Three bottles for each treatment were cultured at the same time and each bottle was shaken by delicate rotation and inversion once per day throughout the experiment. Duplicate or triplicate of each treatment respecting all conditions for proper OA studies were produced (LaRoche *et al.*, 2010). The cultures were grown until a predetermined (repeated test cultures) time laps, when a cell density that draws down less than 5% of the initial Dissolved Inorganic Carbon (DIC) was reached. The cell inoculum number was pre-calculated using the growth rate under the respective treatment in order to produce multiple generations without surpassing the 5% DIC drawdown. The growth of the cultures was monitored by regular cell counts and all cultures were kept axenic by working under a clean bench to avoid bacterial contamination.

2.2 Culture sampling and measurements

Cultures were sampled during the exponential growth phase at the predetermined cell density. First, 500 ml samples for DIC and total alkalinity (TA) were taken using axenic Fe-free tubing into axenic, acid-cleaned glass bottles to avoid any contamination. The samples were immediately poisoned with 100 µl mercury chloride ($HgCl_2$) to avoid CO_2 consumption from cells present in the sample. The measurements were executed sequentially with the

coulometric SOMMA system (University of Rhode Island, RI, USA) and the **V**ersatile **IN**strument for the **D**etermination of **T**otal inorganic carbon and titration **A**lkalinity (VINDTA - University of Rhode Island, RI, USA), for DIC and TA, respectively. Subsequently, 10 ml were collected for pH measurements at growth temperature. Another few millimeters were used to measure the chlorophyll fluorescence (fv/fm) with a phytoplankton analyzer system (PHYTO-PAM, Walz GmbH, Eichenring, Germany) to evaluate the status of each culture. The corresponding pCO_2 was calculated on the CO_2sys program (Lewis *et al.*, 1998) by using the measured carbonate system parameters DIC, TA and pH, the temperature, the salinity, the calculated phosphate and silicate concentrations (mmol/KG SW). Cells were manually counted under the microscope using a Fuchs Rosenthal haemocytometer to control the number of cells actually present in each culture. The cultures were also sampled by filtering 3 x 20 ml (for each analysis) on pre-combusted GF/F filters (Whatman GmbH, Dassel, Germany) to measure chlorophyll (chl *a*), particulate organic carbon (POC) and particulate organic nitrogen (PON) following standard measurement protocols (Sharp, 1974). The final step was to harvest the cells by filtering the whole culture through 47 mm 5 µ-polycarbonate-filters (Durapore® 47 mm, Millipore, Darmstadt, Germany) for RNA and DNA. The filters were resuspended into a small amount of artificial sea water and were thereafter centrifugated at 4°C for 10 min at 11 000 rpm in order to prepare a non-denatured sample for sequencing analysis. Cell pellets were frozen in liquid nitrogen and stored at -80°C until further analysis.

2.3 DNA extraction and sample preparation

Total nucleic acid was extracted using the "Total RNA and DNA purification - NucleoSpin® RNA II RNA/DNA buffer" kits from Macherey-Nagel (Macherey-Nagel Gmbh & Co. KG, Düren, Germany) following the standard manufacturer's protocol after sonication of the cell pellets for 10 sec after the addition of lysis buffer. Both RNA and DNA were isolated during the experiment; however, DNA was not used in the present study. The quantity of products was assessed as the mean of triplicate measurements by micro-volume spectrophotometer nanodrop ND-1000 (PeqLab GmbH, Erlangen, Germany) measurements and quality of RNA was tested with the automated electrophoresis system Experion^{TM} RNA-StdAens analysis kit from Bio-Rad Laboratories (Hercules, California, USA). All samples were kept at −80°C until further analysis.

2.3 Sequencing

Metatranscriptomic paired-end libraries (2 x 100 bp) were generated using mRNA from the cultured *T. oceanica* on the Illumina TruSeq RNA Sample Preparation Kits V2. Libraries were created by following the manufacturer's protocol (Illumina Inc., CA, USA) and were then sequenced on an Illumina HiSeq2000 (Illumina Inc., CA, USA).

2.4 Annotation and Assembly

Reference gene annotation was done by Roche 454 massive parallel pyrosequencing of *T. oceanica* cDNA libraries prepared from total RNA extracted from Fe replete and limited cultures grown at present day pCO_2, as in Lommer *et al.* (2010, 2012). This genome is accessible on the *T. oceanica* genome browser (http://134.245.218.26/cgi-bin/gbrowse/Toceanica/) and gives a framework onto which the transcriptome obtained in this study can be recruited.

2.5 Mapping

Reads were mapped against a multi-fasta file containing the sequences of putative transcripts applying Burrows-Wheeler Aligner (Li *et al.*, 2009) with a seed length of 35 bp and standard parameters (default by BWA at http://bio-bwa.sourceforge.net/bwa.shtml). The total number of reported alignments for each reference sequence was counted, as was the number of unique alignments (determined by BWA's X0 flag).

2.6 Statistical analysis

A log likelihood statistic ratio denoted "R" was used to calculate the significance of any gene differentially expressed (Stekel *et al.*, 2000). The formula for the log likelihood ratio is presented in (1) and (2).

$$R_j = \sum_{i=1}^{m} x_{i,j} \log\left(\frac{x_{i,j}}{N_i f_j}\right) \quad (1)$$

In formula (1), *m* is the libraries number of five and six for ambient and elevated pCO_2, *xi,j* is the number of reads of the gene of interest, *j* in the *i*th library, *Ni* is the total number of reads in the *i*th library and *fj* is total reads number for gene of interest *j* divided by the sum of reads in the *i*th library given by formula (2)

$$f_i = \sum_{i=1}^{m} x_{i,j} \; / \sum_{i=1}^{m} \mathrm{N}_i \qquad (2)$$

The statistical limit can be fixed to different value in relation to library size but it is generally fixed to R > 8. Only the top 500 genes (including unknown and predicted proteins) with the highest R value and expression ratio (mean read number in elevated divided by mean read number in ambient) were kept throughout the analysis. Differential gene expression was also analysed using a metatranscriptomics data overview called **M**icrobial **A**ssemblage **N**ormalized **T**ranscript **A**nalysis (**MANTA**) **R**atio-**A**verage (**RA**) plots from the MANTA package (http://armbrustlab.ocean.washington.edu/download/R-packages/manta-PSA.2011-Visual.Comparative.Transcriptomics-MANTA.RA.pdf and http://armbrustlab.ocean.washington.edu/resources/meta/manta) and the statistical test "exactTest" (edger package) in order to obtain an independent assessment of differentially expressed genes (effect's direction). All manta related tasks were done on the R project software (R Development Core Team, 2011).

3 RESULTS

3.1 Cultures characteristics

As expected, significant differences between pCO_2 levels were found for the DIC, TA, pH and C (POC); indicating that the cultures grew in the initially prescribed conditions of 380 and 1000 µatm, as no significant differences were seen between the beginning and the end of the cultures. Furthermore, no significant effect of elevated pCO_2 on the cell size and growth rate of *T. oceanica* (Table 1) was found.

3.2 Sequencing statistics

Sequencing statistics are presented in Table 2 with about 30,000,000 sequences per library (except the libraries in 1000 µatm BR1c and BR2b) and a maximum of about 28% were classified as unmapped sequences.

Table 1. Overview of measured parameters during experimental culturing events. Significant p-values ($p < 0.05$) are marked bold, n/d = not defined because of irrelevance and n/a = not available because of low replication.

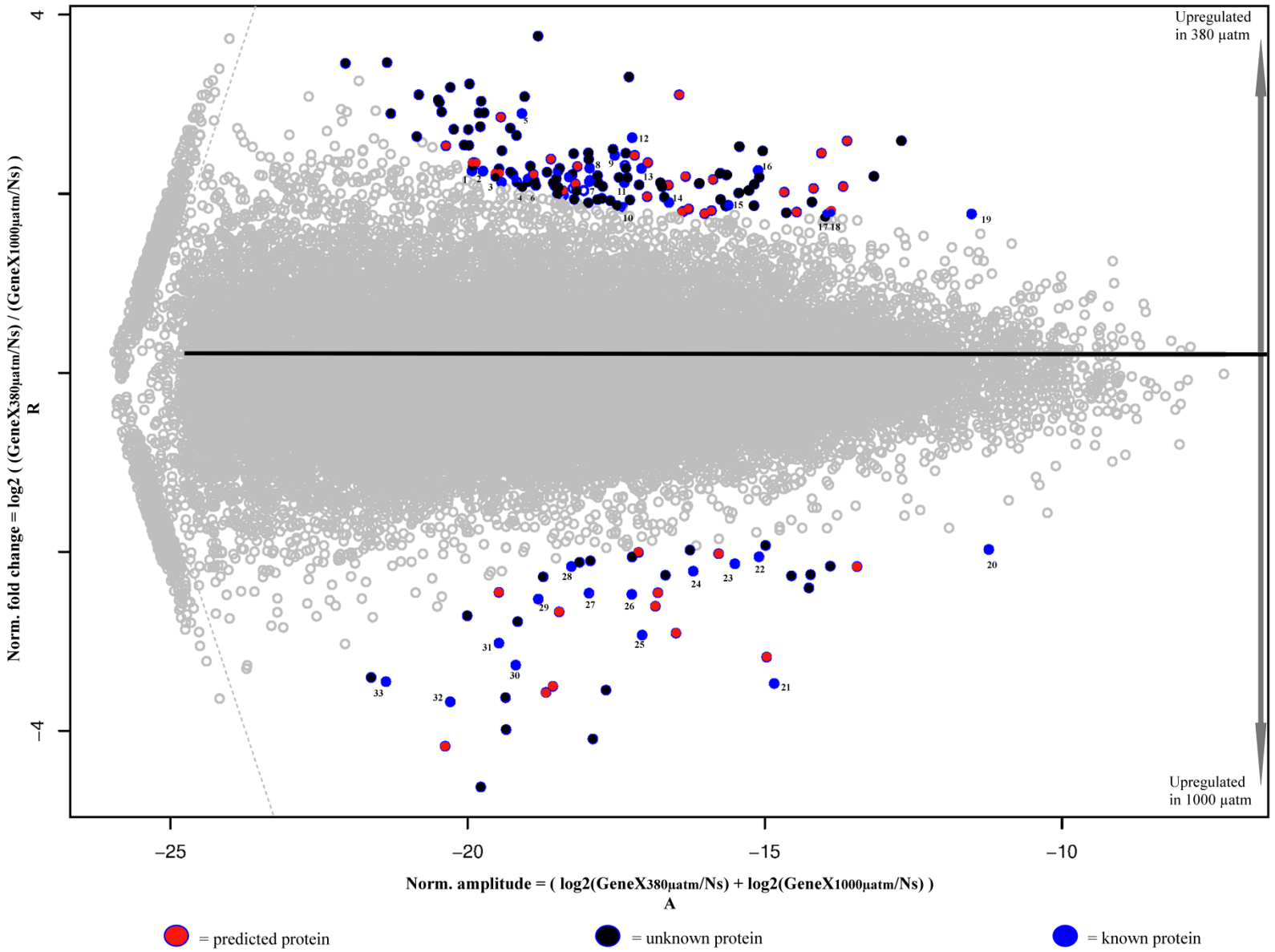
	380		1000		t-test
	Mean	**(SD)**	**Mean**	**(SD)**	**p value**
DIC beg (µmol/kg)	2010	(61)	2125	(31)	**0.00**
DIC end (µmol/kg)	1932	(34)	2123	(45)	**1.09E-05**
DIC % difference	3.9		0.1		n/d
TA beg (µmol/kg)	2310	(75)	2240	(28)	**0.05**
TA end (µmol/kg)	2333	(71)	2264	(51)	**0.05**
TA % difference	-1.0		-1.0		n/d
fv/fm	0.54	(0.23)	0.59	(0.02)	0.31
pH beg	8.11	(0.02)	7.76	(0.03)	**8.57E-09**
pH end	8.25	(0.11)	7.86	(0.04)	**3.13E-04**
% difference	-1.67		-1.32		n/d
≈ cell number/ml	96534	(48971)	139705	(43311)	0.08
Cell size					
diameter (µm)	6.17	(0.69)	6.50	(0.94)	0.11
length (µm)	7.09	(1.17)	7.07	(0.53)	0.47
Growth rate	1.54	(0.13)	1.60	(0.05)	n/a
DNA (ng/µl)	37.7	(9.6)	28.3	(5.3)	n/d
RNA (ng/µl)	353.2	(137.6)	444.0	(138.4)	n/d
Chl *a* (µg/L)	20.1	(8.4)	36.9	(26.6)	0.09
Chl *a* (µg/cell)	2.17E-07	(0.35E-07)	2.41E-07	(1.31E-07)	0.35
C (µmol/kg)	184.95	(58.63)	258.90	(74.08)	**0.048**
N (µmol/kg)	20.79	(8.11)	25.75	(6.48)	0.15
C/N (mean of C/N's per filter)	9.30	(1.95)	9.99	(3.37)	0.17

Table 2. Illumina HISeq sequencing statistics.

CO_2	Biological replicate	Sample replicate	Total aligned fwReads	Total aligned revReads	Total alignments	Total reads	Unmapped reads	Unmapped %
Ambient (≈380 µatm)	BR1	a	15,802,981	15,644,424	31,447,963	43,341,072	11,893,109	27.44
		b	16,274,705	16,121,391	32,395,810	44,388,858	11,993,048	27.02
	BR2	a	14,891,692	14,743,300	29,635,554	39,819,742	10,184,188	25.58
		b	14,507,264	14,368,737	28,875,497	38,739,540	9,864,043	25.46
		c	15,744,779	15,597,363	31,342,765	42,866,322	11,523,557	26.88
Elevated (≈1000 µatm)	BR1	a	15,666,193	15,519,739	31,185,259	42,746,372	11,561,113	27.05
		b	14,768,478	14,641,468	29,410,179	40,843,020	11,432,841	27.99
		c	11,908,972	11,775,221	23,683,809	32,004,876	8,321,067	26.00
	BR2	a	15,610,359	15,425,302	31,035,127	41,290,862	10,255,735	24.84
		b	9,127,785	9,022,561	18,150,199	24,492,714	6,342,515	25.90
		c	18,014,296	17,831,673	35,846,133	47,953,430	12,107,297	25.25

Table 3. Multiple cellular ribosomal proteins expression ratios. Gene number corresponds to the *T. oceanica* browser annotation. Significant p-values ($p < 0.05$) are marked bold.

Ribosomal Compartment	Gene number	Mean reads (380 µatm)	Mean reads (1000 µatm)	High/low ratio	Mean ratio +/- SD
MT	g27363	10.40	18.33	1.76	
MT	g30227	98.60	65.33	0.66	
MT	Tass1192	0.00	0.00	---	
MT	g27361	2.40	3.50	1.46	
MT	Tass2936	0.00	0.00	---	
MT	Tass7983	0.00	0.33	---	
MT	Tass1077	0.00	0.00	---	
MT	g17721	3.40	4.17	1.23	
MT	Tass2773	0.00	0.33	---	
MT	Tass373	3.00	9.67	3.22	
MT	Tass1096	1.80	2.50	1.39	
MT	g27369	6.40	12.83	2.01	1.68 +/- 0.8
			t-test p-value MT versus CP		**0.01**
			t-test p-value MT versus NC		**0.03**
CP	g1142	26457.80	21923.00	0.83	
CP	g30197	23.40	19.00	0.81	
CP	g10714	30.80	18.17	0.59	
CP	g26201	6.20	4.00	0.65	
CP	g30232	85.60	46.50	0.54	
CP	g30219	51.20	44.67	0.87	
CP	g30227	98.60	65.33	0.66	
CP	Tass11157	4.40	0.83	0.19	
CP	g26181	393.00	486.33	1.24	
CP	g30207	19.20	23.17	1.21	
CP	g17915	899.60	863.50	0.96	0.78 +/- 0.3
			t-test p-value CP versus NC		**0.03**
NC	g18532	8251.60	7890.00	0.96	
NC	Tass2718	12.80	12.83	1.00	
NC	g8069	5469.80	5593.50	1.02	
NC	g34623	31851.60	21497.67	0.67	
NC	g19689	4560.80	4494.67	0.99	
NC	g1960	6902.00	7167.17	1.04	
NC	g17987	13028.80	11842.00	0.91	
NC	g6278	3159.20	3322.50	1.05	
NC	g15098	1719.20	2014.17	1.17	0.98 +/- 0.1


Upregulated
in 380 µatm
Upregulated
in 1000 µatm
Norm. fold change = log2 ((GeneX380µatm/Ns) / (GeneX1000µatm/Ns))
R
Norm. amplitude = (log2(GeneX380µatm/Ns) + log2(GeneX1000µatm/Ns))
A
4
−4
−25
−20
−15
−10
= predicted protein
= unknown protein
= known protein

1 g28627-pol protein
2 g10592-hypothetical protein BRAFLDRAFT 131122 gb|EEN53864.1|
3 g33335-hypothetical protein Aasi 0854 gb|ACE06224.1|
4 g4360-PREDICTED: similar to conserved hypothetical protein, partial
5 g13790-sel1 repeat protein gb|EAR88461.1|
6 g35192-cytoplasmic dynein heavy chain-like protein gb|ACI64606.1|
7 g25299-putative lipoprotein gb|EDM77756.1|
8 g29247-putative lipoprotein gb|EDM77756.1|
9 g14374-hypothetical protein CBG18550 emb|CAP35977.1|
10 g7866-Serine/threonine protein kinase
11 g34366-pol protein
12 g27435-hypothetical protein Psyc 0360 gb|AAZ18229.1|
13 g3276-myosin heavy chain-like protein gb|EED88779.1|
14 g20707-pectate lyase gb|ADE01341.1|
15 g23789-epsilon frustilin
16 g31520-myosin heavy chain-like protein gb|EED88779.1|
17 g7638-ATP-dependent DNA helicase (predicted) emb|CAA18870.1|
18 g22991-hypothetical protein gb|EDQ93048.1|
19 g37510-tyrocidine synthase 3 gb|AAC45930.1|
20 g1207-fucoxanthin chl. a/c binding protein emb|CAA80677.1|
21 g29769-hypothetical protein NATL1 08871 gb|ABM75445.1|
22 g2177-expressed unknown protein
23 g25221-hypothetical protein THAPSDRAFT 33389 gb|EED93256.1|
24 g2906-NNT NAD transhydrogenase gb|EED92787.1|
25 g4999-FOG: TPR repeat, SEL1 subfamily
26 g6811-fucoxanthin chl a/c light-harvesting protein gb|EED86590.1|
27 g26452-conserved hypothetical protein gb|EEY57439.1|
28 g778-tripartite motif-containing 65, isoform CRA b
29 g31889-hypothetical protein THAPSDRAFT 261827 gb|EED93810.1|
30 g25222-shikimate kinase gb|EBA00232.1|
31 g13157-hypothetical protein THAPSDDRAFT 264671 gb|EED87914.1|
32 g26335-PKD domain containing protein gb|EEG20097.1|
33 g1579-enhanced entry protein enhc, tetratricopeptide repeat family

Fig. 1. MANTA-RA plot showing the expression of all (grey points) and significantly differentially expressed (blue borders) genes obtained from the transcriptome of *T. oceanica* with the 33 known protein-coding genes (#1-33) listed below figure.

3.3 Cellular ribosomal protein

Ribosomal proteins indicate the basic metabolism of the nucleus (NC), chloroplast (CP) and mitochondria (MT) and can be used as metabolic reference by looking at the expression of the genes located in these compartments. Genes within these categories were randomly selected and the corresponding mean read abundances (mean 380 µatm versus mean 1000 µatm) were used to calculate the mean expression ratios (mean 1000 µatm / mean 380 µatm) (Table 3). The MT, CP and NC ratios were 1.68, 0.78 and 0.98 respectively indicating an increased activity of the MT genes while the CP decreased slightly in elevated CO_2.

3.4 Gene expression overview

All predicted genes on which experimental reads were mapped are presented in the gene expression overview MANTA-RA plot (Fig. 1). The black horizontal line across the plot represents the trimmed mean of fold changes and the genes above this line are upregulated in ambient pCO_2 (380 µatm) or downregulated in 1000 µatm. Conversely, the genes presented under the line are upregulated in 1000 µatm or downregulated in 380 µatm. All points situated to the left of the grey dotted lines represent the genes that have zero read under either one or both conditions. A total of 184 genes of the total 38 993 unique genes (Augustus predicted and manually assigned genes) entries were statistically differentially expressed and of these, we concentrated our analysis to the 33 genes which were associated with known proteins coding genes. Fig. 1 also includes all statistically differentially expressed unknown and predicted proteins. A complete list of these statistically differentially expressed 184 genes including their p-values is presented in Table S1.

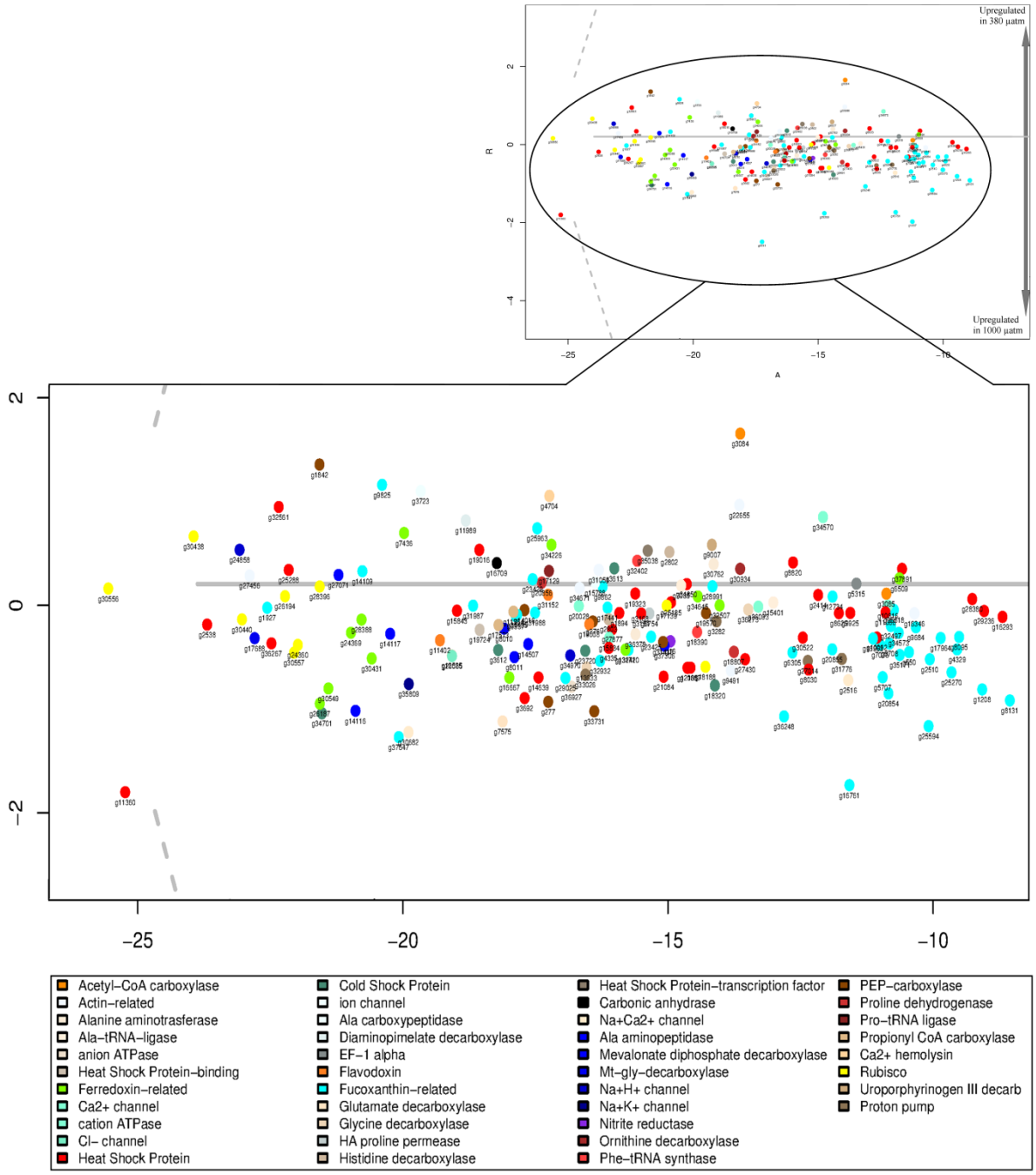


Fig. 2. MANTA-RA plot showing color-coded genes previously established in *E. huxleyi* as being significantly differentially expressed under different CO_2 pressures, gene numbers taken from the annotation provided on the *T. oceanica* browser. Normalised amplitude (average read counts) called A is the x-axis while normalized fold change called R is represented by the y-axis (axis formulas presented in Fig. 1.).

3.5 Genes differentially regulated with CO_2 levels

The same MANTA-RA plot annotated with genes identified in *E. huxleyi* as being affected in any way by changes in CO_2 and pH is presented in Fig. 2. The genes differentially expressed in *E. huxleyi* as a function of CO_2 were not significantly affected by elevated pCO_2 in *T. oceanica*, although most followed a trend similar to what has been found in *E. huxleyi*. For example, vacuolar proton pump proteins (all related proteins) are found in the lower (up-regulated in 1000 µatm) middle part of the figure, ribulose-1.5-biphosphate carboxylase/oxygenic (RuBisCO) is found in the lower-left side, heat shock and fucoxanthin to the lower-right side of the MANTA-RA. However, none were identified in the top-ranked, significantly differentially expressed genes by either of the statistical analyses performed on our data.

3.6 Log likelihood ratio statistic

As in the previous study by Stekel *et al.* (2012), the log likelihood ration was fixed to be significant when above 8 ($R > 8$) and thus gave a total of 26,520 of the 38,993 total Augustus unique predicted genes (Augustus predicted and manually assigned genes) significantly differentially expressed under elevated or ambient pCO_2. Of this total, 8532 genes had a $R > 100$ which contained 1298 with a $R > 1000$ and finally, of these, 93 genes had a $R > 10\ 000$. Of all these genes, we concentrated our analyses on the 500 genes with the highest expression in either the ambient or elevated pCO_2, resulting in 250 genes in each condition. The complete list of these genes is presented in Table S2 and Table S3, whereas a short version of the most important genes is listed in Table 4.

Table 4. Differentially expressed genes obtained from the log likelihood R calculations; where ratio represents the mean read abundance in elevated pCO_2 (1000 µatm) / mean read abundance in ambient pCO_2 (380 µatm). Upreg. States in which pCO_2 (µatm) the upregulation is seen. Gene numbers correspond to unique genes annotated for the *T. oceanica* browser. The genes found in both present and known proteins of the MANTA-RA tests are marked with *.

Gene	Gene description	R	ratio	Upreg.
g19004	alkaline phosphatase [*T. pseudonana* CCMP1335] gb|EED95733.1|	14.57	6.67	1000
g34583	alkaline phosphatase-like protein [Teredinibacter turnerae T7901] gb|ACR13720.1|	16.85	10.00	1000
g7758	carboxylesterase [Mycobacterium parascrofulaceum ATCC BAA-614] gb|EFG79051.1|	13.49	4.17	1000
g2513	cell surface protein [Methanosarcina acetivorans C2A] gb|AAM07636.1|	351.08	3.12	1000
g35803	COG0790: FOG: TPR repeat, SEL1 subfamily [Haemophilus influenzae R2846]	367.55	3.60	1000
g23643	COG0790: FOG: TPR repeat, SEL1 subfamily [Haemophilus influenzae R2846]	25.29	3.13	1000
Tass373	cytochrome oxidase subunit II [*T. pseudonana*] gb|AAZ99398.1|	14.14	3.22	1000
g1579	enhanced entry protein, tetratricopeptide repeat family [Cox.burnetii Dugway 5J108-111] gb|ABS76720.1|	130.87	15.23	1000
g2532	exonuclease [*T. pseudonana* CCMP1335] gb|EED90855.1|	8.57	4.29	1000
g15465	expressed protein [Oryza sativa (japonica cultivar-group)]	35.13	3.13	1000
g2177	expressed unknown protein [Ectocarpus siliculosus]	1910.60	3.82	1000
g24161	flap endonuclease-1 [Candidatus Methanoregula boonei 6A8] gb|ABS56620.1|	14.42	4.32	1000
g4999	FOG: TPR repeat, SEL1 subfamily [Butyrivibrio fibrisolvens 16/4]	1494.77	7.05	1000
g6811*	fucoxanthin chl *a*/*c* light-harvesting protein [*T. pseudonana* CCMP1335] gb|EED86590.1|	1040.18	5.17	1000
g25269	Fucoxanthin chlorophyll a/c protein [*P. tricornutum* CCAP 1055/1] gb|EEC44018.1|	3605.46	3.11	1000
g1207*	Fucoxanthin-chlorophyll a-c binding protein F, chloroplastic emb|CAA80677.1|	51918.00	3.61	1000
g10419	mu subunit of AP4-like protein [*T. pseudonana* CCMP1335] gb|EED96067.1|	9.97	3.33	1000
g30046	myosin heavy chain [Dictyostelium discoideum]	14.33	4.06	1000
g27367	NADH dehydrogenase subunit 7 [*T. pseudonana*] gb|AAZ99426.1|	22.13	5.52	1000
Tass9655	NADH dehydrogenase subunit 9 [*T. pseudonana*] gb|AAZ99425.1|	8.68	6.67	1000
g2906*	NNT NAD transhydrogenase [*T. pseudonana* CCMP1335] gb|EED92787.1|	1901.60	4.28	1000

g4387	oxidoreductase, short chain dehydrogenase/reductase family gb|EDX85241.1|	518.46	3.44	1000
g26335	PKD domain containing protein [Opitutaceae bacterium TAV2] gb|EEG20097.1|	290.39	13.93	1000
g28285	replication protein A2 [*T. pseudonana* CCMP1335] gb|EED92004.1|	22.76	3.55	1000
g27362	ribosomal protein S11 [*T. pseudonana*] gb|AAZ99415.1|	16.72	5.19	1000
g20895	Sel1 domain protein repeat-containing [alpha proteobacterium BAL199] gb|EDP64864.1|	19.00	5.07	1000
g3129	Sel1 domain protein repeat-containing [Haemophilus influenzae 7P49H1] gb|EEP45669.1|	9.13	3.17	1000
g35956	Sel1 domain-containing protein [Magnetococcus sp. MC-1] gb|ABK44504.1|	513.86	3.27	1000
g32947	Sel1 repeat-containing protein [Chlorobium chlorochromatii CaD3] gb|ABB29186.1|	12.78	6.17	1000
g25222	shikimate kinase [Marinobacter sp. ELB17] gb|EBA00232.1|	327.52	9.31	1000
g21380	subtilisin-like serine protease [*T. pseudonana* CCMP1335] gb|EED89728.1|	35.93	6.58	1000
g1691	thioredoxin [*T. pseudonana* CCMP1335] gb|EED93987.1|	626.08	3.47	1000
g7664	TPR repeat-containing SEL1 subfamily protein [Acinetobacter baumannii AB900]	12.09	3.19	1000
g3807	TPR repeat-containing SEL1 subfamily protein [Acinetobacter baumannii AB900]	11.31	3.38	1000
g778	tripartite motif-containing 65, isoform CRA_b [Homo sapiens]	460.75	4.22	1000
g9740	zinc finger family protein [Arabidopsis lyrata subsp. lyrata] gb|EFH58975.1|	31.30	9.17	1000
g36264	CBL-interacting serine/threonine-protein kinase [Phytophthora infestans T30-4] gb|EEY62050.1|	25.33	0.22	380
g20706	collagen, type XVIII, alpha 1 [Xenopus laevis] gb|AAL14257.1|	165.75	0.25	380
g36714	conserved hypothetical protein [Phytophthora infestans T30-4] gb|EEY54214.1|	190.35	0.24	380
g7104	conserved hypothetical protein [Phytophthora infestans T30-4] gb|EEY68651.1|	184.58	0.24	380
g13132	conserved unknown protein [Ectocarpus siliculosus]	27.23	0.17	380
g9703	conserved unknown protein [Ectocarpus siliculosus]	26.54	0.24	380
g35192*	cytoplasmic dynein heavy, chain-like protein [*T. pseudonana* CCMP1335] gb|ACI64606.1|	130.48	0.21	380
g35193	cytoplasmic dynein heavy, chain-like protein [*T. pseudonana* CCMP1335] gb|ACI64606.1|	76.36	0.21	380
g23789*	epsilon frustilin [Navicula pelliculosa]	938.07	0.25	380
g17677	flagellar associated protein putative: flagellar associated protein [Ectocarpus siliculosus]	45.19	0.24	380
g31520*	myosin heavy chain-like protein [*T. pseudonana* CCMP1335] gb|EED88779.1|	1795.31	0.19	380
g3276*	myosin heavy chain-like protein [*T. pseudonana* CCMP1335] gb|EED88779.1|	457.72	0.19	380
g20707	pectate lyase [Haloferax volcanii DS2] gb|ADE01341.1|	450.53	0.24	380
g21711	pol protein [*Phaeodactylum tricornutum*]	40.05	0.21	380
g34366*	pol protein [*T. pseudonana*]	338.42	0.21	380

g26721	pol protein [*T. pseudonana*]	205.26	0.25	380
g27186	pol protein [*T. pseudonana*]	62.88	0.20	380
g28627	pol protein [*T. pseudonana*]	61.88	0.18	380
g13210	pol protein [*T. pseudonana*]	54.34	0.18	380
g30640	pol protein [*T. pseudonana*]	38.43	0.22	380
g25850	pol protein [*T. pseudonana*]	26.79	0.20	380
g29247*	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	275.15	0.18	380
g25299*	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	235.41	0.21	380
g28690	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	134.41	0.24	380
g24620	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	57.38	0.14	380
g23160	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	46.42	0.22	380
g13790	sel1 repeat protein [Tetrahymena thermophila] gb|EAR88461.1|	191.55	0.12	380
g7866	Serine/threonine protein kinase [Ectocarpus siliculosus]	288.85	0.25	380
g9409	Serine/threonine-protein kinase PLK4 gb|EAL29859.1|	25.36	0.14	380

4 DISCUSSION

Our experiment consisted of characterizing the transcriptome of *T. oceanica* grown at elevated and ambient CO_2 while maintaining defined pCO_2, pH, total DIC and TA. This was achieved by maintaining constant light aeration of the cultures with premixed CO_2 in a closed system, thereby avoiding any gas exchanges with the atmosphere. The risk of causing minimal "bubbling stress" (Richier *et al.*, 2011) was favored over addition of hydrochloric acid (HCl) and sodium carbonate ($Na_2CO_3^-$) as multiple pre-cultures identified this technique as more disturbing (produced large cells aggregate) for *T. oceanica*. Cultures grown at the same time with and without constant aeration to evaluate the extension of the "bubbling stress" were done and demonstrated no difference in cell growth or size (results not shown).

All carbonate system parameters (Table 3) were significantly different between CO_2 pressures but not measurements (beginning versus end) making every culture used in this analysis conforms to conditions outlined by the OA guidelines (Riebesell *et al.*, 2010). The growth, cell yield and chlorophyll content of the cultured *T. oceanica*, as assessed, were increased (even if minimally) by elevated CO_2 as found in a similar studies on *P. tricornutum* and *T. pseudonana* cultured under high CO_2 and different light intensities (Sobrino *et al.*, 2008, Wu *et al.*, 2010). This all indicate that *T. oceanica*'s photosynthetic rate (fv/fm not significantly different) seems to be affected by the increase of CO_2 as previously suggested. Finally, the C:N ratios indicated that as expected from similar experiment, the C and N mechanisms increased a little but are not influenced by the CO_2 pressure (Sugie *et al.*, 2013).

4.1 Cellular activity indication

Evaluation of the influence of elevated pCO_2 on the general cell metabolism by observation of cellular ribosomal proteins indicated that the MT activity increased while the CP decreased under elevated CO_2. It was previously observed (Lommer *et al.*, 2012), that under stress like reduced Fe concentration, the CP genes expression was impaired whereas the MT genes expression increased; possibly as a stress compensation mechanism. In the present case, the elevated CO_2 concentrations provided increased carbon resource (photosynthesis) to the cell and is considered as optimal growth condition (Wu *et al.*, 2010, Crawfurd *et al.*, 2011). Wu and colleagues (2010) found that *P. tricornutum* had an increased photosynthetic rate when photosynthetic resource was increased, and that the cell's general systemic functions were believed to be more active. Consequently the MT, known as the central energy-converting

compartment, had to increase its activity as the results of increased photosynthesis. It is similarly hypothesized here for the elevated pCO_2 treatment, comparatively to the results of stress as seen in our previous experiment with different Fe concentrations.

4.2 Transcriptomic profile of *T. oceanica* in elevated CO_2

All genes expressed under ambient or elevated CO_2, are presented in Fig. 1, which shows that the transcriptome of *T. oceanica* does not appear, at first glance, to be very sensitive to varying pCO_2 and pH levels. However, it is important to note that some of the outliers, representing the significantly differentially expressed genes, might be essential genes for the survival or the maintenance of the organism. Here, we looked at 184 statistically differentially expressed genes (MANTA-RA) with most of them representing either unknown (107 out of 184) or predicted protein (44 out of 184). All known proteins of these 184 genes (33 genes) are listed under the figure with most of them also established significantly differentially expressed by the log likelihood R (genes present in both statistical analyses are marked with *).

If, assuming the possibility that elevated pCO_2 or in our case the predicted pCO_2 for the end of the century, does not necessarily create stress conditions for certain phytoplankton but might instead be beneficial by creating an increased photosynthesis resource is true; we compared our top significantly differently expressed genes (R values/MANTA-RA) to the ones obtained in our previous transcriptomic research on stress induced Fe-deplete cultures of the same diatom (Lommer *et al.*, 2012). Only a few of the genes upregulated in either one treatment, 380 or 1000 µatm, are discussed below in making a first characterization of the effect of elevated CO_2 on *T. oceanica* transcriptome.

The gene called NNT producing NADH dehydrogenase (g2906*) and NADH dehydrogenase subunit 7 & 9 gene (g27367, Tass9655) are genes localised in the MT. These genes were upregulated in 1000 µatm, in accordance with the general increased MT gene expression observed under elevated pCO_2. NADH dehydrogenase is an enzyme responsible for catalysing NADH into NAD but its exact role is currently not completely known. It has been linked to the CCM which increases or replenishes the CO_2 concentration at the fixation sites surrounding the enzyme RuBisCO in order to ensure a proper fixation for vital demands (Burkhardt *et al.*, 2001, Hopkinson *et al.*, 2011, Alterio *et al.*, 2012, Zabaleta *et al.*, 2012, Hopkinson *et al.*, 2013). Throughout evolution, diverse RuBisCO and CCM's have been

produced to create the optimal version of these enzymes (Young *et al.*, 2012). One of these CCMs, called basal CCM (bCCM), was described in higher plants and Cyanobacteria by Zabaleta *et al.* (2012) as being composed of mitochondrial carbonic anhydrase (CA- both β-CA and γ-CA) from the mitochondrial NADH dehydrogenase complex. Therefore, we hypothesize that the upregulation of the NADH dehydrogenase enzyme in 1000 µatm contributes to the CCMs known to reduce their activity in presence of elevated concentration of CO_2 (Burkhardt *et al.*, 2001, Hopkinson *et al.*, 2011, Alterio *et al.*, 2012, Young *et al.*, 2012, Zabaleta *et al.*, 2012, Hopkinson *et al.*, 2013). Furthermore, the Tass9655 was also identified as upregulated in the Fe replete condition (Lommer *et al.*, 2012), strongly underlining the fact that this enzyme seems to be up-regulated under optimal growth conditions. More research is needed to define the importance of NADH genes but its involvement with different carbon-related functions definitely marks its importance in OA research.

Other proteins up-regulated in elevated CO_2 are the fucoxanthin chl *a*/*c* light-harvesting protein (FCP) (g6811*, g1207*, g25269) localized in the CP. These proteins are part of the light-harvesting antenna systems present in multiple organisms (Chromophyta, Bacillariophyta, Cyanobacteria) but are also important pigment binding protein of diatoms believed to have been acquired during second endosymbiosis of a red algae (Pan *et al.*, 2013, Ikeda *et al.*, 2013). The function of this protein superfamily (*T. pseudonana* was demonstrated to have 30 and *P. tricornutum* 40 FCP candidates - Armbrust *et al.*, 2004, Nymark *et al.*, 2009, respectively) can be mainly divided in two types: as harvesting light necessary for oxygenic photosynthesis or as energy-dissipating proteins (Neilson *et al.*, 2010, Ikeda *et al.*, 2013, Sturm *et al.*, 2013). The differentiation of these proteins and their distribution between photosystems (I and II) is difficult because of the diatoms' hard silica shells making unaltered protein's extraction difficult (Ikeda *et al.*, 2013). However with the present increase of photosynthetic yield (indicated by fv/fm), the upregulated FCPs, here under 1000 µatm, are the light-harvesting types as carbon substrate is largely available for photosynthesis. The activity of the FCPs is therefore probably increased in order to complete photosynthesis.

Another protein-coding gene up-regulated under elevated CO_2 is the group "tetratricopeptide repeat (TPR) - Sel1 subfamily domain" proteins (g35803, g23643, g4999, g7664, g3807, g20895, g3129, g35956, g32947-MANTA ID g13790). This group of proteins represent the genes which appeared the most often in the top 500 significantly differentially expressed

genes (Table S2 and S3). This group of proteins can be divided in two different proteins types. The first part is the TPR, a repeat of 34 amino acids found in a large variety of proteins, which form scaffolds that mediate protein to protein interactions or participate in multiple protein assembly (Das *et al.*, 1998, Lima *et al.*, 2010, Whitfield *et al.*, 2010). These repeats are believed to have the general task of regulating the complexes of several important "unknown" processes (Krachler *et al.*, 2010). More elaborate, specialised studies demonstrated that these repeats were present in several important systems like bacterial cellulose and poly-b-DGGIcNAc development or biofilm formation (Keiski *et al.*, 2010, Whitfield *et al.*, 2010). The second part of the group is the Sel1 protein subfamily, known for containing repeats (here TPR) and for being found in bacteria. They are identified as proteins important in pathogens-host cell interactions which down-regulates host-signaling-surface cells called Notch receptors (Periz *et al.*, 1999, Singh *et al.*, 2009, Schmitz-Esser *et al.*, 2010). These proteins are known for being components of the endoplasmic reticulum (Kaneko *et al.*, 2007, Ninagawa *et al.*, 2011) involved in transport for removing unfolded proteins and for maintaining a high Ca^{2+} concentration in the lumen. The direct function of these repeat-containing proteins is still widely unknown, however, they may be involved in stress response, especially in pathogen-host interaction. Further studies (Kaneko *et al.*, 2007, Sharifmoghadam *et al.*, 2009, Singh *et al.*, 2009, Schmitz-Esser *et al.*, 2010, Ninagawa *et al.*, 2011) established that Sel1 proteins are present when the cells are undergoing stress (for example: osmotic variation, nutrient limitation, symbiont or parasite). TPR-Sel1's presence in 1000 µatm is unexpected in regards to the elevated CO_2 beneficial scenario but it was also present in the replete Fe experiment (g35803) presented in Lommer *et al.*, (2010, 2012). At this time, it is not possible to speculate what stressed the cells in both experiments because the cells were not reacting to the presumed treatments (bottle effect, density, light, etc.). However, the fact that multiple genes are up-regulated in both experiments (this study and Lommer *et al.* 2012), shows that there are similarities between elevated CO_2 concentration and high dissolved Fe.

Another gene upregulated in both replete conditions (present elevated CO_2 and previous replete Fe) is the cytochrome oxidase subunit II (Tass373) often abbreviated to cox2. It is a component of the respiratory chain; more precisely, it is the terminal oxidase of the cell respiration that catalyzes electron transfer from cytochrome C to oxygen and, protons translocation reactions across mitochondrial membranes (Cooper *et al.*, 1991, Tsukihara *et al.*,

1996). This gene is localised in the MT, again supporting that elevated CO_2 stimulated MT gene expression.

The following sections present some of the upregulated in the ambient (≈380 µatm) pCO_2 condition. The first coding genes were the RNA polymerase proteins or pol proteins (g21711, g34366, g26721*, g27186, g28627, g13210, 30640, g25850) which were well represented by 8 genes coding for these proteins found in the top 500 genes. These proteins are from viral origin (retroviruses) and encode multiple enzymatic functions ultimately leading to the integration of modified cDNA to the organism's genome (Maumus *et al.*, 2009). These proteins are possible stress indicators, at which time pol proteins tend to increase mutations via genomic rearrangements as seen in *P. tricornutum* during sexual reproduction (Maumus *et al.*, 2009). These rearrangements ultimately lead to newly recombined genome and these recombinations have the potential to help the organisms to cope better with new stressors or changing environments (Maumus *et al.*, 2009). In our case, these proteins are strongly expressed under present-day pCO_2, thus reinforcing that higher CO_2 levels may provide better growth conditions for diatoms.

The putative lipoprotein genes (g29247*, g25299*, g28690, g24620, g23160, g13790) were also found to be up-regulated in the ambient pCO_2 and in the low Fe condition of our previous study. The reason for such an expression pattern is unknown, but since a lipoprotein is a combination of proteins and lipids, allowing the latter to be transported between the intracellular and extracellular membranes of a cell, is it possible that such lipids are needed in stressful conditions like low pCO_2 or depleted Fe. Furthermore, in the event of difficult or stressful times, diatoms tend to switch from asexual to sexual reproduction (Waite *et al.*, 1992, Rijstenbil *et al.*, 1994, Armbrust, 1999, Drebes, 1977) as, even if this is energetically more "costly", it is a good investment to pool genes for recombination and thus increases the chances to get advantageous genes for the experienced treat. Sexual reproduction was established as the driving source of genetic variation in the marine environment (Waite *et al.*, 1992) with stressed cells evolving in direction of survival of the fittest via either mutation or genomic adaptation. Despite that the presence of sexual reproduction has not yet been established in *T. oceanica*, it was speculated that multiple genes associated with restructuring proteins could be up-regulated in the ambient CO_2 (which in our elevated-CO_2 better scenario would be considered a limiting environment). In fact, the previously presented pol proteins are associated with mutation and play a certain restructuring role. Another gene up-regulated

in ambient CO_2 was the flagellar protein (g17677) and is associated to cellular mobility. However since diatoms are non motile organisms, this gene could hypothetically be involved in spermatozoid motility during sexual reproduction (Drebes, 1977). Another gene up-regulated in ambient CO_2 in relation to motility, is the cytoplasmic dynein heavy chain-like protein (g35192*, g35193). This protein is a converter responsible for the transformation of ATP into mechanical energy available for movements (Moore *et al.*, 2013). It is responsible for long-range transfer in between organelles (Barlan *et al.*, 2013) whereas myosin is responsible for short-range transfer. Concisely, myosin (g31520*, g3276*) was also up-regulated genes in this pCO_2, fitting nicely into the sexual reproduction context.

Stress has also generally been associated to have an effect on cell wall (Davis *et al.*, 2011) and could be here hypothetically related to collagen protein (type XVIII, alpha 1 - g20706), another gene up-regulated at ambient pCO_2. Indeed, there seems to be an unknown connection between silica and collagen (collagen/silicate in chimeras in sponge spicules – van der Walle, 2012) and could here mean that some collagen proteins in relation to diatom cell wall are altered by lower CO_2 level. Finally, epsilon frustilins (g23789*) were up-regulated in ambient pCO_2 and are generally found associated to the outer layers of frustules of diatoms (Kroeger *et al.*, 1996, van de Poll *et al.*, 1999 Santos *et al.*, 2013). They are further suspected to be involved in the actual building of the cell walls (Kroeger *et al.*, 1996, van de Poll *et al.*, 1999, Kroeger *et al.*, 2008) but not to their silicification. These last proteins indicate that diatoms' cell wall proteins are to some extend downregulated by the increase of CO_2.

4.3 Carbon influenced genes

Multiple genes identified by previous studies on the calcifying Haptophyta *E. huxleyi* (Mackinder *et al.*, 2011, Richier *et al.*, 2011, Lohbeck *et al.*, 2012, Bach *et al.*, 2013) as being significantly affected by increases of pCO_2 or by reduction of pH were not part of the 184 significantly differentially expressed genes (MANTA-RA) or part of the top 500 genes from the log likelihood ratio analysis of *T. oceanica*. Calcifying organisms like *E. huxleyi* have been identified as more susceptive as other organisms to acidification changes because of the decrease in calcification rate and the increase of calcareous material dissolution (Gattuso *et al.*, 1998, Riebesell *et al.*, 2000, Langdon *et al.*, 2003, Orr *et al.*, 2005). These genes are presented in Fig. 2 in order to visualise their expression levels but were not significantly affected by increase of pCO_2 in *T. oceanica*. All genes were found closely above or under the trimmed mean of fold-changes values; representing a small variation in the gene regulation. It

is important to note even lacking of marked difference in expression levels, these genes were widely distributed on the x-axis, representing a significant difference in the expression abundance. For example, multiple genes coding for fucoxanthin chl *a*/*c* proteins (g17964, g25594, g8095, g4329, g25270, g1208, g8131) and heat shock proteins (g16293, g28380, g292360) are found at the end of the tail and represent some of the most abundant genes. RuBisCO (g30556, g30438, g30440), PSI-ferredoxin-binding protein (g26187, g30549, g24369, g28388) and the exchangers/transporters (NA+/H+, vacuolar H+, anion and cation transporting, Na/Ca, etc. see list) expected to be influenced, are found in the upregulated, under 1000 µatm section but are not significantly differentially expressed. All genes coding for any type of carboxylase (g33731, g277, g276, g1842, g14016, g19570, g17573, g11364) are also found in the middle-left of the graph and are thus also sparsely expressed. These findings suggest that elevated pCO_2 does not affect *T. oceanica*'s carbon related system greatly. Our results agree with the reaction of *T. pseudonana* to elevated pCO_2 as no significant effect of this increase was found (Crawfurd *et al.*, 2011); except for the C/N ratios which were found to decrease with the increase of CO_2 concentration whereas these ratios were found to increase in the present study. Crawfurd *et al.*, (2011) found only one putative protein, d-carbonic anhydrase (CA-4), to have a slight decreased transcription after the length on the experiment (three months continuous culture). Furthermore, after completion of their long experiment, no adaptation of *T. pseudonana* could be demonstrated. Both the present study and the study from Crawfurd and colleagues (2011) confirm the possible benefit of elevated pCO_2 on diatoms compare to Haptophyta. Indeed, a previous study (Rost *et al.*, 2003) established that the reaction to CO_2 augmentation was species-specific and that the diatom *Skeletonema costatum* had a highly efficient and regulated carbon acquisition system compared to *E. huxleyi* which had an inefficient carbon uptake system and low carbonic anhydrase activities. They attributed these differences to the functional group to which the species belong. Indeed calcifiers are renowned for their sensitivity to acidification. On the basis of previous and present studies, increasing pCO_2 are suspected to have minimal effect on diatoms while others parameters included in global warming (temperature, light intensities, surface water stratification, etc.) might have the predominant effect on the diatoms' efficient ability to make photosynthesis as previously demonstrated (Sobrino *et al.*, 2008, Wu *et al.*, 2010). However, some effect might be masked by the fact that some CO_2 affected genes are not yet known and therefore not yet investigated. In order to counteract this un-characterization, unknown or predicted protein found to be abundantly or highly significantly differentially expressed need to be investigated intensively (undergoing investigation).

5 CONCLUSION

The transcriptomics results on *T. oceanica* shown in our study, present new elements on the effect of elevated pCO_2 on this species and possibly confirm results on other diatoms. A significant effect of increasing CO_2 on gene expression was demonstrated with gene such as FCP's or NADH dehydrogenase being up-regulated in 1000 µatm. However, the genes previously reported as influenced by increasing CO_2 in *E. huxleyi* were not affected here. We speculate that the genes coding for unknown and predicted proteins presented as statistically differentially expressed in either pCO_2 could belong to one of the mechanisms indicted as vulnerable to OA and need to be further investigated.

T. oceanica was no affected negatively but possibly positively by the acidification as the enrichment of CO_2 permits increased photosynthesis and could increase growth rates. It seems that the scenario considering the effect of elevated CO_2 as reducing photosynthetic stress might be possible. However, other conditions created by global warming , like increased light intensities or surface water stratification, might be the parameters influencing the diatoms' primary production as similar results on *P. tricornutum* and *T. pseudonana* were previously established (Sobrino *et al.*, 2008, Wu *et al.*, 2010). The global importance of diatoms as principal player in primary production makes it more than necessary to continue to analyse and quantify the effects of OA on their remarkable photosynthetic efficiency. However, the lack of changes in gene expression in regards to the carbon related mechanisms could suggest that *T. oceanica* might be able to adapt or acclimate to OA. Experiments involving transcriptome profiling can help to elucidate the function of all gene expression affected by changing environmental factors. The combination of the log likelihood ratio and MANTA-RA analyses allowed the identification of genes significantly affected by CO_2 and the characterization of the effect's direction; resulting in an overview of elevated CO_2 on the transcriptome of *T. oceanica*. It is, furthermore, essential to compare current results from unialgal cultures to metatranscriptomic results of environmental samples, as cultures may not represent the response of a community or a mixed natural population of phytoplankton.

6 ACKNOWLEDGEMENT

We would like to thank T. Klüver and B. Weigel for their participation in culturing and sampling *T. oceanica*. We also thank M. S. Parker and S. Bender for their help with the MANTA package. We are grateful to the Institute of Clinical Molecular Biology in Kiel for providing Illumina sequencing. This work was supported by BMBF project BIOACID (D10/1.1.2) and the European Project on Ocean Acidification (EPOCA), which received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 211384.

7 REFERENCES

Allen, A. E., LaRoche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., Finazzi, G., Fernie, A. R. & Bowler, C. 2008. Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. *Proc. Natl. Acad. Sci. U. S. A.* **105**:10438-43.

Allen, A. E., Moustafa, A., Montsant, A., Eckert, A., Kroth, P. G. & Bowler, C. 2012. Evolution and Functional Diversification of Fructose Bisphosphate Aldolase Genes in Photosynthetic Marine Diatoms. *Molecular Biology and Evolution* **29**:367-79.

Alterio, V., Langella, E., Viparelli, F., Vullo, D., Ascione, G., Dathan, N. A., Morel, F. M. M., Supuran, C. T., De Simone, G. & Monti, S. M. 2012. Structural and inhibition insights into carbonic anhydrase CDCA1 from the marine diatom Thalassiosira weissflogii. *Biochimie* **94**:1232-41.

Armbrust, E. V. 1999. Identification of a new gene family expressed during the onset of sexual reproduction in the centric diatom Thalassiosira weissflogii. *Appl. Environ. Microbiol.* **65**:3121-28.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S. G., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kroger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P. & Rokhsar, D. S. 2004. The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. *Science* **306**:79-86.

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. *New Phytologist* **199**:121-34.

Barlan, K., Rossow, M. J. & Gelfand, V. I. 2013. The journey of the organelle: teamwork and regulation in intracellular transport. *Current Opinion in Cell Biology* **25**:483-88.

Bidle, K. D. & Falkowski, P. G. 2004. Cell death in planktonic, photosynthetic microorganisms. *Nature Reviews Microbiology* **2**:643-55.

Bidle, K. D. & Bender, S. J. 2008. Iron starvation and culture age activate metacaspases and programmed cell death in the marine diatom Thalassiosira pseudonana. *Eukaryotic Cell* **7**:223-36.

Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. A., Buesseler, K. O., Coale, K. H., Cullen, J. J., de Baar, H. J. W., Follows, M., Harvey, M., Lancelot, C., Levasseur, M., Owens, N. P. J., Pollard, R., Rivkin, R. B., Sarmiento, J., Schoemann, V., Smetacek, V., Takeda, S., Tsuda, A., Turner, S. & Watson, A. J. 2007. Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions. *Science* **315**:612-17.

Brussaard, C. P. D., Noordeloos, A. A. M., Witte, H., Collenteur, M. C. J., Schulz, K., Ludwig, A. & Riebesell, U. 2013. Arctic microbial community dynamics influenced by elevated CO_2 levels. *Biogeosciences* **10**:719-31.

Buesseler, K. O., Andrews, J. E., Pike, S. M. & Charette, M. A. 2004. The effects of iron fertilization on carbon sequestration in the Southern Ocean. *Science* **304**:414-17.

Burkhardt, S., Amoroso, G., Riebesell, U. & Sultemeyer, D. 2001. CO_2 and HCO_3^- uptake in marine diatoms acclimated to different CO_2 concentrations. *Limnol. Oceanogr.* **46**:1378-91.

Caldeira, K. & Wickett, M. E. 2003. Anthropogenic carbon and ocean pH. *Nature* **425**:365-65.

Chisholm, S. W., Falkowski, P. G. & Cullen, J. J. 2001. Oceans - Dis-crediting ocean fertilization. *Science* **294**:309-10.

Collins, S. & Bell, G. 2006a. Evolution of natural algal populations at elevated CO_2. *Ecology Letters* **9**:129-35.

Collins, S. & Bell, G. 2006b. Rewinding the tape: Selection of algae adapted to high CO_2 at current and pleistocene levels of CO_2. *Evolution* **60**:1392-401.

Cooper, C. E., Nicholls, P. & Freedman, J. A. 1991. Cytochrome-c-oxidase - structure, function, and membrane topology of the polypeptide subunits. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **69**:586-607.

Crawfurd, K. J., Raven, J. A., Wheeler, G. L., Baxter, E. J. & Joint, I. 2011. The Response of Thalassiosira pseudonana to Long-Term Exposure to Increased CO_2 and Decreased pH. *Plos One* **6**.

Das, A. K., Cohen, P. T. W. & Barford, D. 1998. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *Embo Journal* **17**:1192-99.

Davis, S. J., Peters, G. P. & Caldeira, K. 2011. The supply chain of CO_2 emissions. *Proc. Natl. Acad. Sci. U. S. A.* **108**:18554-59.

Drebes, G. 1977. Sexuality. *In*: Werner, D. [Ed.] *The biology of diatoms.* University of California Press, Berkeley (u.a), pp. 250-83.

Falkowski, P. G., Barber, R. T. & Smetacek, V. 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**:200-06.

Feely, R. A., Sabine, C. L., Lee, K., Berelson, W., Kleypas, J., Fabry, V. J. & Millero, F. J. 2004. Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans. *Science* **305**:362-66.

Feng, Y., Warner, M. E., Zhang, Y., Sun, J., Fu, F. X., Rose, J. M. & Hutchins, D. A. 2008. Interactive effects of increased pCO(2), temperature and irradiance on the marine coccolithophore Emiliania huxleyi (Prymnesiophyceae). *European Journal of Phycology* **43**:87-98.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C. & Zeng, N. 2006. Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison. *Journal of Climate* **19**:3337-53.

Fu, F. X., Mulholland, M. R., Garcia, N. S., Beck, A., Bernhardt, P. W., Warner, M. E., Sanudo-Wilhelmy, S. A. & Hutchins, D. A. 2008. Interactions between changing pCO(2), N-2 fixation, and Fe limitation in the marine unicellular cyanobacterium Crocosphaera. *Limnol. Oceanogr.* **53**:2472-84.

Gattuso, J. P., Frankignoulle, M., Bourge, I., Romaine, S. & Buddemeier, R. W. 1998. Effect of calcium carbonate saturation of seawater on coral calcification. *Global and Planetary Change* **18**:37-46.

Hoegh-Guldberg, O. 2010. Dangerous shifts in ocean ecosystem function? *Isme Journal* **4**:1090-92.

Hoegh-Guldberg, O. & Bruno, J. F. 2010. The Impact of Climate Change on the World's Marine Ecosystems. *Science* **328**:1523-28.

Hoffmann, L. J., Peeken, I., Lochte, K., Assmy, P. & Veldhuis, M. 2006. Different reactions of Southern Ocean phytoplankton size classes to iron fertilization. *Limnol. Oceanogr.* **51**:1217-29.

Hopkinson, B. M., Dupont, C. L., Allen, A. E. & Morel, F. M. M. 2011. Efficiency of the CO_2-concentrating mechanism of diatoms. *Proc. Natl. Acad. Sci. U. S. A.* **108**:3830-37.

Hopkinson, B. M., Meile, C. & Shen, C. 2013. Quantification of Extracellular Carbonic Anhydrase Activity in Two Marine Diatoms and Investigation of Its Role. *Plant Physiol.* **162**:1142-52.

Ikeda, Y., Yamagishi, A., Komura, M., Suzuki, T., Dohmae, N., Shibata, Y., Itoh, S., Koike, H. & Satoh, K. 2013. Two types of fucoxanthin-chlorophyll-binding proteins I tightly bound to the photosystem I core complex in marine centric diatoms. *Biochimica Et Biophysica Acta-Bioenergetics* **1827**:529-39.

Kaneko, M., Yasui, S., Niinuma, Y., Arai, K., Omura, T., Okuma, Y. & Nomura, Y. 2007. A different pathway in the endoplasmic reticulum stress-induced expression of human HRD1 and SEL1 genes. *Febs Letters* **581**:5355-60.

Keiski, C. L., Harwich, M., Jain, S., Neculai, A. M., Yip, P., Robinson, H., Whitney, J. C., Riley, L., Burrows, L. L., Ohman, D. E. & Howell, P. L. 2010. AlgK Is a TPR-Containing Protein and the Periplasmic Component of a Novel Exopolysaccharide Secretin. *Structure* **18**:265-73.

Kester, D. R., Duedall, I. W., Connors, D. N. & Pytkowic.Rm 1967. preparation of artificial seawater. *Limnol. Oceanogr.* **12**:176-&.

Kleypas, J. A., Feely, R. A., Fabry, V. J., Langdon, C., Sabine, C. L. & Robbins, L. L. 2006. Impacts of Ocean Acidification on Coral Reefs and Other Marine Calcifiers: A Guide for Future Researchs: A Guide for Future Research, report of a workshop held 18–20 April 2005. NSF, NOAA, and the U.S. Geological Survey, pp. 88.

Krachler, A. M., Sharma, A. & Kleanthous, C. 2010. Self-association of TPR domains: Lessons learned from a designed, consensus-based TPR oligomer. *Proteins-Structure Function and Bioinformatics* **78**:2131-43.

Kroeger, N., Bergsdorf, C. & Sumper, M. 1996. Frustilins: Domain conservation in a protein family associated with diatom cell walls. *European Journal of Biochemistry* **239**:259-64.

Kroeger, N. & Poulsen, N. 2008. Diatoms-From Cell Wall Biogenesis to Nanotechnology. *Annual Review of Genetics.* pp. 83-107.

Langdon, C., Broecker, W. S., Hammond, D. E., Glenn, E., Fitzsimmons, K., Nelson, S. G., Peng, T. H., Hajdas, I. & Bonani, G. 2003. Effect of elevated CO_2 on the community metabolism of an experimental coral reef. *Global Biogeochemical Cycles* **17**.

LaRoche, J., Rost, B. & Engel, A. 2010. Bioassays, batch culture and chemostat experimentation. *In*: Riebesell, U., Fabry, V. J., Hansson, L. & Gattuso, J. P. [Eds.] *Guide to best practices for ocean acidification research and data reporting.* Publications Office of the European Union, Luxembourg, pp. 81-94.

Lewis, E. & Wallace, D. W. R. 1998. Program developed for CO_2 system calculations. *Carbon Dioxide Information Analysis Center.* U.S. Department of Energy, OakRidge, Tennessee.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754-60.

Lima, M. D., Eloy, N. B., Pegoraro, C., Sagit, R., Rojas, C., Bretz, T., Vargas, L., Elofsson, A., de Oliveira, A. C., Hemerly, A. S. & Ferreira, P. C. G. 2010. Genomic evolution and complexity of the Anaphase-promoting Complex (APC) in land plants. *Bmc Plant Biology* **10**.

Lohbeck, K. T., Riebesell, U. & Reusch, T. B. H. 2012. Adaptive evolution of a key phytoplankton species to ocean acidification (vol 5, pg 346, 2012). *Nature Geoscience* **5**:917-17.

Lommer, M., Specht, M., Roy, A.-S., Kraemer, L., Andreson, R., Gutowska, M. A., Wolf, J., Bergner, S. V., Schilhabel, M. B., Klostermeier, U. C., Beiko, R. G., Rosenstiel, P., Hippler, M. & LaRoche, J. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology* **13**.

Lommer, M., Roy, A. S., Schilhabel, M., Schreiber, S., Rosenstiel, P. & LaRoche, J. 2010. Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics* **11**.

Mackinder, L., Bach, L., Schulz, K., Wheeler, G., Schroeder, D., Riebesell, U. & Brownlee, C. 2011. The molecular basis of inorganic carbon uptake mechanisms in the coccolithophore emiliania huxleyi. *European Journal of Phycology* **46**:142-43.

Maumus, F., Allen, A. E., Mhiri, C., Hu, H. H., Jabbari, K., Vardi, A., Grandbastien, M. A. & Bowler, C. 2009. Potential impact of stress activated retrotransposons on genome evolution in a marine diatom. *BMC Genomics* **10**.

Melzner, F., Thomsen, J., Koeve, W., Oschlies, A., Gutowska, M. A., Bange, H. W., Hansen, H. P. & Kortzinger, A. 2013. Future ocean acidification will be amplified by hypoxia in coastal habitats. *Marine Biology* **160**:1875-88.

Moore, D. J., Onoufriadis, A., Shoemark, A., Simpson, M. A., zur Lage, P. I., de Castro, S. C., Bartoloni, L., Gallone, G., Petridi, S., Woollard, W. J., Antony, D., Schmidts, M.,

Didonna, T., Makrythanasis, P., Bevillard, J., Mongan, N. P., Djakow, J., Pals, G., Lucas, J. S., Marthin, J. K., Nielsen, K. G., Santoni, F., Guipponi, M., Hogg, C., Antonarakis, S. E., Emes, R. D., Chung, E. M. K., Greene, N. D. E., Blouin, J. L., Jarman, A. P. & Mitchison, H. M. 2013. Mutations in ZMYND10, a Gene Essential for Proper Axonemal Assembly of Inner and Outer Dynein Arms in Humans and Flies, Cause Primary Ciliary Dyskinesia. *American Journal of Human Genetics* **93**:346-56.

Neilson, J. A. D. & Durnford, D. G. 2010. Evolutionary distribution of light-harvesting complex-like proteins in photosynthetic eukaryotes. *Genome* **53**:68-78.

Newbold, L. K., Oliver, A. E., Booth, T., Tiwari, B., DeSantis, T., Maguire, M., Andersen, G., van der Gast, C. J. & Whiteley, A. S. 2012. The response of marine picoplankton to ocean acidification. *Environmental Microbiology* **14**:2293-307.

Ninagawa, S., Okada, T., Takeda, S. & Mori, K. 2011. SEL1L Is Required for Endoplasmic Reticulum-associated Degradation of Misfolded Luminal Proteins but not Transmembrane Proteins in Chicken DT40 Cell Line. *Cell Structure and Function* **36**:187-95.

Nymark, M., Valle, K. C., Brembu, T., Hancke, K., Winge, P., Andresen, K., Johnsen, G. & Bones, A. M. 2009. An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom Phaeodactylum tricornutum. *Plos One* **4**.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M. F., Yamanaka, Y. & Yool, A. 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**:681-86.

Pan, X. J., Wong, G. T. F., Ho, T. Y., Shiah, F. K. & Liu, H. B. 2013. Remote sensing of picophytoplankton distribution in the northern South China Sea. *Remote Sensing of Environment* **128**:162-75.

Periz, G. & Fortini, M. E. 1999. Ca2+-ATPase function is required for intracellular trafficking of the Notch receptor in Drosophila. *Embo Journal* **18**:5983-93.

Pespeni, M. H. 2013. Evolutionary and ecological genomics in a changing world: integrating Next-Gen data with environmental variation to reveal local adaptation. *Integrative and Comparative Biology* **53**:E165-E65.

Plattner, G. K., Joos, F., Stocker, T. F. & Marchal, O. 2001. Feedback mechanisms and sensitivities of ocean carbon uptake under global warming. *Tellus Series B-Chemical and Physical Meteorology* **53**:564-92.

Pollard, R. T., Salter, I., Sanders, R. J., Lucas, M. I., Moore, C. M., Mills, R. A., Statham, P. J., Allen, J. T., Baker, A. R., Bakker, D. C. E., Charette, M. A., Fielding, S., Fones, G. R., French, M., Hickman, A. E., Holland, R. J., Hughes, J. A., Jickells, T. D., Lampitt, R. S., Morris, P. J., Nedelec, F. H., Nielsdottir, M., Planquette, H., Popova, E. E., Poulton, A. J., Read, J. F., Seeyave, S., Smith, T., Stinchcombe, M., Taylor, S., Thomalla, S., Venables, H. J., Williamson, R. & Zubkov, M. V. 2009. Southern Ocean deep-water carbon export enhanced by natural iron fertilization. *Nature* **457**:577-U81.

R Development Core Team 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienne, Austria.

Raven, J. A. & Waite, A. M. 2004. The evolution of silicification in diatoms: inescapable sinking and sinking as escape? *New Phytologist* **162**:45-61.

Reinfelder, J. R., Milligan, A. J. & Morel, F. M. M. 2004. The role of the C-4 pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiol.* **135**:2106-11.

Richardson, A. J., Brown, C. J., Brander, K., Bruno, J. F., Buckley, L., Burrows, M. T., Duarte, C. M., Halpern, B. S., Hoegh-Guldberg, O., Holding, J., Kappel, C. V., Kiessling, W., Moore, P. J., O'Connor, M. I., Pandolfi, J. M., Parmesan, C., Schoeman, D. S., Schwing, F., Sydeman, W. J. & Poloczanska, E. S. 2012. Climate change and marine life. *Biology Letters* **8**:907-09.

Richier, S., Fiorini, S., Kerros, M. E., von Dassow, P. & Gattuso, J. P. 2011. Response of the calcifying coccolithophore Emiliania huxleyi to low pH/high pCO(2): from physiology to molecular level. *Marine Biology* **158**:551-60.

Riebesell, U., Wolfgladrow, D. A. & Smetacek, V. 1993. Carbon-dioxide limitation of marine-phytoplankton growth-rates. *Nature* **361**:249-51.

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 CO_2. *Nature* **407**:364-67.

Riebesell, U., Fabry, V. J., Hansson, L. & Gattuso, J.-P. 2010. *Guide to best practices for ocean acidification research and data reporting.* Publications Office of the European Union, Luxembourg, 260.

Riebesell, U. & Tortell, P. D. 2011. Effects of ocean acidification on pelagic organisms and ecosystems. *In*: Gattuso, J.-P. & Hansson, L. [Eds.] *Ocean acidification.* Oxford Univ. Press.

Ries, J. B., Cohen, A. L. & McCorkle, D. C. 2009. Marine calcifiers exhibit mixed responses to CO_2-induced ocean acidification. *Geology* **37**:1131-34.

Rijstenbil, J. W., Derksen, J. W. M., Gerringa, L. J. A., Poortvliet, T. C. W., Sandee, A., Vandenberg, M., Vandrie, J. & Wijnholds, J. A. 1994. Oxidative stress-induced by copper - defense and damage in the marine planktonic diatom ditylum-brightwellii, grown in continuous cultures with high and low zinc levels. *Marine Biology* **119**:583-90.

Rossoll, D., Bermudez, R., Hauss, H., Schulz, K. G., Riebesell, U., Sommer, U. & Winder, M. 2012. Ocean Acidification-Induced Food Quality Deterioration Constrains Trophic Transfer. *Plos One* **7**.

Rost, B., Riebesell, U., Burkhardt, S. & Sultemeyer, D. 2003. Carbon acquisition of bloom-forming marine phytoplankton. *Limnol. Oceanogr.* **48**:55-67.

Santos, J., Almeida, S. F. P. & Figueira, E. 2013. Cadmium chelation by frustilins: a novel metal tolerance mechanism in Nitzschia palea (Kutzing) W. Smith. *Ecotoxicology* **22**:166-73.

Schmitz-Esser, S., Tischler, P., Arnold, R., Montanaro, J., Wagner, M., Rattei, T. & Horn, M. 2010. The Genome of the Amoeba Symbiont "Candidatus Amoebophilus asiaticus" Reveals Common Mechanisms for Host Cell Interaction among Amoeba-Associated Bacteria. *Journal of Bacteriology* **192**:1045-57.

Schulz, K. G., Bellerby, R. G. J., Brussaard, C. P. D., Büdenbender, J., Czerny, J., Engel, A., Fischer, M., Koch-Klavsen, S., Krug, S. A., Lischka, S., Ludwig, A., Meyerhöfer, M., Nondal, G., Silyakova, A., Stuhr, A. & Riebesell, U. 2013. Temporal biomass dynamics of an Arctic plankton bloom in response to increasing levels of atmospheric carbon dioxide. *Biogeosciences* **10**:161-80.

Sharifmoghadam, M. R. & Valdivieso, M. H. 2009. The Fission Yeast SEL1 Domain Protein Cfh3p a novel regulator of the glucan synthase bgs1p whose function is more relevant under stress conditions. *Journal of Biological Chemistry* **284**:11070-79.

Sharp, J. H. 1974. Improved analysis for particulate organic carbon and nitrogen from seawater. *Limnol. Oceanogr.* **19**:984-89.

Shi, D. L., Xu, Y., Hopkinson, B. M. & Morel, F. M. M. 2010. Effect of Ocean Acidification on Iron Availability to Marine Phytoplankton. *Science* **327**:676-79.

Singh, M., Mukherjee, P., Narayanasamy, K., Arora, R., Sen, S. D., Gupta, S., Natarajan, K. & Malhotra, P. 2009. Proteome Analysis of Plasmodium falciparum Extracellular Secretory Antigens at Asexual Blood Stages Reveals a Cohort of Proteins with Possible Roles in Immune Modulation and Signaling. *Molecular & Cellular Proteomics* **8**:2102-18.

Smetacek, V. & Naqvi, S. W. A. 2008. The next generation of iron fertilization experiments in the Southern Ocean. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* **366**:3947-67.

Sobrino, C., Ward, M. L. & Neale, P. J. 2008. Acclimation to elevated carbon dioxide and ultraviolet radiation in the diatom Thalassiosira pseudonana: Effects on growth, photosynthesis, and spectral sensitivity of photoinhibition. *Limnol. Oceanogr.* **53**:494-505.

Stekel, D. J., Git, Y. & Falciani, F. 2000. The comparison of gene expression from multiple cDNA libraries. *Genome Research* **10**:2055-61.

Strzepek, R. F. & Harrison, P. J. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* **431**:689-92.

Sturm, S., Engelken, J., Gruber, A., Vugrinec, S., Kroth, P. G., Adamska, I. & Lavaud, J. 2013. A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids. *Bmc Evolutionary Biology* **13**.

Sugie, K. & Yoshimura, T. 2013. Effects of pCO_2 and iron on the elemental composition and cell geometry of the marine diatom Pseudo-nitzschia pseudodelicatissima (Bacillariophyceae). *J. Phycol.* **49**:475-88.

Thomas, T., Gilbert, J. & Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2**.

Timmermans, K. R., Gerringa, L. J. A., de Baar, H. J. W., van der Wagt, B., Veldhuis, M. J. W., de Jong, J. T. M., Croot, P. L. & Boye, M. 2001. Growth rates of large and small Southern Ocean diatoms in relation to availability of iron in natural seawater. *Limnol. Oceanogr.* **46**:260-66.

Tortell, P. D., Payne, C. D., Li, Y. Y., Trimborn, S., Rost, B., Smith, W. O., Riesselman, C., Dunbar, R. B., Sedwick, P. & DiTullio, G. R. 2008. CO(2) sensitivity of Southern Ocean phytoplankton. *Geophysical Research Letters* **35**.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., ShinzawaItoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. 1996. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 angstrom. *Science* **272**:1136-44.

van de Poll, W. H., Vrieling, E. G. & Gieskes, W. W. C. 1999. Location and expression of frustilins in the pennate diatoms Cylindrotheca fusiformis, Navicula pelliculosa, and Navicula salinarum (Bacillariophyceae). *J. Phycol.* **35**:1044-53.

van der Walle, C. F. 2012. Towards a Bottom-up Approach for Mimicking Marine Sponge Spicules. *Silicon* **4**:23-31.

Waite, A. & Harrison, P. J. 1992. Role of sinking and ascent during sexual reproduction in the marine diatom ditylum-brightwellii. *Marine Ecology Progress Series* **87**:113-22.

Whitfield, C. & Mainprize, I. L. 2010. TPR Motifs: Hallmarks of a New Polysaccharide Export Scaffold. *Structure* **18**:151-53.

Wolf-Gladrow, D. A., Riebesell, U., Burkhardt, S. & Bijma, J. 1999. Direct effects of CO_2 concentration on growth and isotopic composition of marine plankton. *Tellus Series B-Chemical and Physical Meteorology* **51**:461-76.

Worden, A. Z. & Not, F. 2008. Ecology and Diversity of Picoeukaryotes. *In*: Kirchman, D. [Ed.] *Microbial Ecology of the Ocean, 2nd Edition.* Wiley.

Wu, Y., Gao, K. & Riebesell, U. 2010. CO_2-induced seawater acidification affects physiological performance of the marine diatom Phaeodactylum tricornutum. *Biogeosciences* **7**:2915-23.

Young, J. N., Rickaby, R. E. M., Kapralov, M. V. & Filatov, D. A. 2012. Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367**:483-92.

Zabaleta, E., Martin, M. V. & Braun, H. P. 2012. A basal carbon concentrating mechanism in plants? *Plant Science* **187**:97-104.

SUPPLEMENTARY MATERIAL

Table S1. List of the 184 outliers significantly differentially expressed identified by the MANTA-RA analysis. up = upregulated at 380 µatm and dn = downregulated at 380 µatm.

Position	Gene number	Gene description	p-value
up	g7638	ATP-dependent_DNA_helicase_(predicted)_sp|O60177|YG42_SCHPO_RecName:_Full=Uncharacterized_ATP-dependent_helicase_C23E6.02_emb|CAA18870.1|	0.0002285
dn	g26452	conserved_hypothetical_protein_gb|EEY57439.1|	1.33E-05
up	g35192	cytoplasmic_dynein_heavy_chain-like_protein_gb|ACI64606.1|	0.0001293
dn	g1579	enhanced_entry_protein_enhC,_tetratricopeptide_repeat_family_gb|ABS76720.1|	2.61E-05
up	g23789	epsilon_frustilin	0.0001198
dn	g2177	expressed_unknown_protein	9.44E-05
dn	g4999	FOG:_TPR_repeat,_SEL1_subfamily	1.15E-07
dn	g6811	fucoxanthin_chl_a/c_light-harvesting_protein_gb|EED86590.1|	6.40E-06
up	g33335	hypothetical_protein_Aasi_0854_gb|ACE06224.1|	1.00E-04
up	g10592	hypothetical_protein_BRAFLDRAFT_131122_gb|EEN53864.1|	0.0001736
up	g14374	Hypothetical_protein_CBG18550_emb|CAP35977.1|	2.86E-06
up	g22991	hypothetical_protein_gb|EDQ93048.1|	0.0002137
dn	g29769	hypothetical_protein_NATL1_08871_gb|ABM75445.1|	2.24E-10
up	g27435	hypothetical_protein_Psyc_0360_gb|AAZ18229.1|	4.69E-07
dn	g31889	hypothetical_protein_THAPSDRAFT_261827_gb|EED93810.1|	2.37E-05
dn	g13157	hypothetical_protein_THAPSDRAFT_264671_gb|EED87914.1|	2.55E-06
dn	g25221	hypothetical_protein_THAPSDRAFT_33389_gb|EED93256.1|	5.32E-05
up	g3276	myosin_heavy_chain-like_protein_gb|EED88779.1|	7.35E-06
up	g31520	myosin_heavy_chain-like_protein_gb|EED88779.1|	4.71E-06
dn	g2906	NNT_NAD_transhydrogenase_gb|EED92787.1|	3.48E-05
up	g20707	pectate_lyase_gb|ADE01341.1|	0.0001175
dn	g26335	PKD_domain_containing_protein_gb|EEG20097.1|	1.29E-07
up	g28627	pol_protein	0.0002126

up	g34366	pol_protein	3.29E-05
up	g29247	putative_lipoprotein_gb|EDM77756.1|	1.22E-05
up	g25299	putative_lipoprotein_gb|EDM77756.1|	4.25E-05
dn	g1207	RecName:_Full=Fucoxanthin-chlorophyll_a-c_binding_protein_F,_chloroplastic;_Flags:_Precursor_emb|CAA80677.1|	0.000151
up	g37510	RecName:_Full=Tyrocidine_synthase_3;_AltName:_Full=Tyrocidine_synthase_III;_Includes gb|AAC45930.1|	0.000198
up	g13790	sel1_repeat_protein_gb|EAR88461.1|	6.69E-07
up	g7866	Serine/threonine_protein_kinase	0.0002464
dn	g25222	shikimate_kinase_gb|EBA00232.1|	1.45E-07
dn	g778	tripartite_motif-containing_65,_isoform_CRA_b	0.0001139
up	g4360	PREDICTED:_similar_to_conserved_hypothetical_protein,_partial	9.65E-05
up	g23632	predicted_protein_gb|ACI64100.1|	0.0002455
up	g19858	predicted_protein_gb|ACI64335.1|	0.0001604
dn	g26312	predicted_protein_gb|ACI64480.1|	5.67E-06
up	g15673	predicted_protein_gb|ACI64504.1|	1.85E-05
up	g14969	predicted_protein_gb|ACO66050.1|	1.31E-05
dn	g27127	predicted_protein_gb|EEC44967.1|	3.86E-06
up	g25760	predicted_protein_gb|EEC45730.1|	1.18E-05
up	g37838	predicted_protein_gb|EEC47605.1|	0.0001336
up	g21839	predicted_protein_gb|EEC49229.1|	0.0001934
dn	g34984	predicted_protein_gb|EEC49407.1|	9.06E-12
up	g2724	predicted_protein_gb|EED86247.1|	3.33E-05
up	g15505	predicted_protein_gb|EED86392.1|	0.000181
up	g16279	predicted_protein_gb|EED87346.1|	0.0001453
up	g20146	predicted_protein_gb|EED87699.1|	0.0001842
up	g22649	predicted_protein_gb|EED88551.1|	2.31E-05
up	g19358	predicted_protein_gb|EED88827.1|	0.0001202
up	g29613	predicted_protein_gb|EED89370.1|	0.0001697
up	g5654	predicted_protein_gb|EED89555.1|	1.86E-06
dn	g4258	predicted_protein_gb|EED89900.1|	0.0001021
up	g24104	predicted_protein_gb|EED90138.1|	2.09E-07

dn	g17733	predicted_protein_gb|EED90323.1|	0.0001884
up	g21470	predicted_protein_gb|EED90337.1|	1.32E-05
up	g25617	predicted_protein_gb|EED90427.1|	1.45E-05
up	g17424	predicted_protein_gb|EED90483.1|	4.61E-06
up	g9305	predicted_protein_gb|EED90483.1|	8.95E-06
dn	g31184	predicted_protein_gb|EED90970.1|	1.80E-06
up	g13797	predicted_protein_gb|EED91646.1|	0.0001384
dn	g36876	predicted_protein_gb|EED91931.1|	1.00E-08
up	g509	predicted_protein_gb|EED92045.1|	0.0001834
up	g36024	predicted_protein_gb|EED92555.1|	3.35E-09
dn	g2080	predicted_protein_gb|EED93488.1|	8.21E-09
up	g17395	predicted_protein_gb|EED93652.1|	7.37E-05
up	g4085	predicted_protein_gb|EED94366.1|	8.84E-05
dn	g2412	predicted_protein_gb|EED94446.1|	4.44E-09
dn	g15967	predicted_protein_gb|EED94525.1|	0.0001354
dn	g16223	predicted_protein_gb|EED94684.1|	1.04E-07
up	g35783	predicted_protein_gb|EED94913.1|	0.0002487
dn	g19676	predicted_protein_gb|EED95041.1|	4.37E-09
dn	g6566	predicted_protein_gb|EED95318.1|	0.0002212
up	g28470	predicted_protein_gb|EED95364.1|	2.48E-06
up	g8305	predicted_protein_gb|EED95371.1|	7.85E-07
up	g4663	predicted_protein_gb|EED96119.1|	2.61E-05
up	g25904	predicted_protein_gb|EED96376.1|	0.0002017
dn	g29771	predicted_protein_gb|EED96591.1|	3.52E-05
dn	g30492	unknown	3.72E-11
dn	g20586	unknown	6.30E-10
dn	Tass2435	unknown	0.0001454
dn	g33243	unknown	1.77E-08
dn	g3347	unknown	9.23E-10
dn	g18672	unknown	6.11E-06

up	g33941	unknown	3.44E-05
up	g13296	unknown	1.88E-06
up	g16685	unknown	3.02E-06
up	g2790	unknown	7.20E-07
up	g10551	unknown	4.68E-06
up	g22791	unknown	2.25E-06
up	g12011	unknown	0.0001596
up	g26731	unknown	6.14E-06
up	g34140	unknown	3.87E-05
up	g22601	unknown	6.19E-06
up	g14582	unknown	2.13E-05
up	g14588	unknown	0.0001142
up	g28556	unknown	0.0001021
up	g11899	unknown	2.90E-05
up	g9212	unknown	4.27E-05
up	g4896	unknown	0.0001093
up	g31299	unknown	7.32E-06
up	g31500	unknown	8.12E-05
up	g10331	unknown	3.79E-05
up	g17478	unknown	3.10E-05
up	g37066	unknown	2.26E-05
up	g18960	unknown	8.33E-06
up	g9780	unknown	5.42E-05
up	g37686	unknown	0.0001951
up	g30406	unknown	9.65E-05
up	g25618	unknown	2.10E-05
up	g19525	unknown	6.99E-06
up	g7853	unknown	2.00E-05
up	g14174	unknown	9.20E-06
up	g28262	unknown	0.0001195

dn	g4949	unknown	4.86E-05
dn	g19604	unknown	5.21E-06
dn	g16582	unknown	0.0001072
dn	g10378	unknown	1.64E-05
dn	Tass5711	unknown	2.80E-05
dn	g4049	unknown	1.81E-05
dn	g8518	unknown	3.75E-05
dn	g10652	unknown	0.0001689
dn	g299	unknown	0.0001655
dn	g27029	unknown	0.0001943
dn	g2691	unknown	0.0002382
up	g30151	unknown	0.0001592
up	g24699	unknown	1.16E-05
up	g4559	unknown	3.07E-10
up	g31662	unknown	1.29E-06
up	g10920	unknown	1.40E-05
up	g16788	unknown	9.88E-10
up	g19274	unknown	0.0002148
up	g8722	unknown	1.10E-06
up	g10184	unknown	9.93E-06
up	g8304	unknown	1.18E-07
up	g32908	unknown	1.08E-06
up	g28934	unknown	9.93E-06
up	g35300	unknown	1.19E-05
up	g33713	unknown	2.30E-06
up	g28662	unknown	2.18E-06
up	g17609	unknown	0.000222
up	g33814	unknown	7.61E-06
up	g26415	unknown	2.64E-05
up	g6918	unknown	4.16E-06
up	g30636	unknown	3.72E-05
up	g4561	unknown	6.55E-05
up	g653	unknown	0.0002037
up	g23148	unknown	4.09E-05
up	g10797	unknown	2.30E-05
up	g6275	unknown	8.50E-05
up	g30141	unknown	0.0001977
up	g32426	unknown	2.07E-05
up	g29942	unknown	3.02E-05
up	g33657	unknown	1.76E-05
up	g12610	unknown	0.0001083
up	g16710	unknown	5.09E-05
up	g31120	unknown	3.27E-05
up	g18983	unknown	9.62E-05
up	g36721	unknown	0.0001
up	g36613	unknown	3.03E-05
up	g33668	unknown	0.0001066
up	g27710	unknown	0.0001066
up	g31345	unknown	0.0001764
up	g19074	unknown	0.0001768
up	g23821	unknown	3.95E-05
up	g7677	unknown	7.63E-05
up	g1506	unknown	0.0001444
up	g11422	unknown	0.0001498
up	g19566	unknown	0.0001213
up	g10807	unknown	0.0002385
up	g34796	unknown	0.00015
up	g31908	unknown	7.69E-05
up	g29743	unknown	7.82E-05
up	g37640	unknown	0.0002399

up	g33564	unknown	2.02E-05	up	g32757	unknown	0.0001154
up	g9364	unknown	6.11E-06	up	g22828	unknown	0.0002208
up	g37486	unknown	6.57E-05	up	g16763	unknown	0.0001254
up	g25917	unknown	6.76E-05	up	g27683	unknown	0.0002201
up	g31355	unknown	1.99E-07	up	g28686	unknown	0.0002003
up	g23287	unknown	5.19E-07				

Table S2. List of reference gene number presented in the MANTA-RA plot (Fig.2) showing the corresponding gene number and the gene description.

Gene number	Gene description	Gene number	Gene description
g3085	acetyl-coa carboxylase	g19724	heat shock protein binding / unfolded protein binding
g3084	acetyl-coa carboxylase	g15843	heat shock protein DnaJ
g3168	actin- actin related protein	g14639	heat shock protein DnaJ domain protein
g22655	actin like protein	g19016	heat shock protein DnaJ family protein
g16346	actin-like protein	g11360	heat shock protein DnaJ-like
g31054	actin-related protein centractin-like protein	g3181	heat shock protein HslVU- ATPase
g30762	alanine aminotransferase	g3692	heat shock protein- HSP70- hsc70
g36378	alanine aminotransferase- aminotransferase class I	g27877	heat shock protein- HSP70- hsc70
g7085	alanine-trna ligase	g9925	heat shock protein Hsp90
g15401	alanine-trna ligase	g31850	heat shock protein/chaperone
g36927	anion transporting ATPase	g16293	heat shock protein/chaperone
g20028	calcium channel	g1894	heat shock protein-like protein
g15834	calcium/proton exchanger- calcium antiporter	g6754	high affinity proline permease
g34570	cation channel	g2802	histidine decarboxylase
g36093	cation transport ATPase	g2803	histidine decarboxylase
g10816	cation transporting ATPase	g35038	HSF32- heat shock transcription factor
g20065	channel voltage activated chloride channel	g16709	intracellular carbonic anhydrase
g8820	chaperone- heat shock protein 70	g30682	K-independent Na+/Ca2+ exchanger JSX
g18320	cold shock protein	g11088	low molecular weight heat shock protein
g34701	cold shock protein CspE	g31710	fucoxanthin chlorophyll a/c family- lhca clade
g3613	cold-shock DNA-binding domain-containing protein	g27071	membrane alanine aminopeptidase
g3612	cold-shock DNA-binding domain-containing protein	g14507	membrane alanine aminopeptidase
g3723	cyclic nucleotide and voltage-activated ion channel	g37306	membrane alanine aminopeptidase

g15788	D-alanyl-D-alanine carboxypeptidease- putative
g11989	diaminopimelate decarboxylase
g32561	DNAJ heat shock N-terminal domain-containing protein
g2538	DNAJ heat shock N-terminal domain-containing protein
g36267	DNAJ heat shock N-terminal domain-containing protein
g5315	elongation factor 1-alpha
g27456	F-actin-capping protein subunit beta- putative
g13242	ferredoxin
g32507	ferredoxin oxidoreductase
g7436	ferredoxin-dependent glutamate synthase- fusion of large and small subunits
g6509	ferredoxin-nadp reductase
g11402	flavodoxin
g31152	flavodoxin
g19008	flavodoxin
g9825	fucoxanthin chl a/c light-harvesting protein
g5707	fucoxanthin chl a/c light-harvesting protein
g16963	fucoxanthin chl a/c light-harvesting protein
g6811	fucoxanthin chl a/c light-harvesting protein
g35171	fucoxanthin chl a/c light-harvesting protein- lhcr type
g29025	fucoxanthin chl a/c light-harvesting protein- lhcr type
g34573	fucoxanthin chl a/c protein- lhca clade
g32932	fucoxanthin chlorophyll a/c binding protein
g8131	fucoxanthin chlorophyll a/c binding protein
g32497	fucoxanthin chlorophyll a/c binding protein
g35518	fucoxanthin chlorophyll a/c light-harvesting protein
g9862	fucoxanthin chlorophyll a/c light-harvesting protein
g5708	fucoxanthin chlorophyll a/c light-harvesting protein-
g25963	fucoxanthin chlorophyll a/c protein
g8010	membrane alanine aminopeptidase
g8011	membrane alanine aminopeptidase
g14117	mevalonate disphosphate decarboxylase-like protein
g17688	mitochondrial glycine decarboxylase T-protein
g17371	mitochondrial glycine decarboxylase T-protein
g14116	MPDC mevalonate diphosphate decarboxylase
g34970	Na+/H+ exchanger
g35809	Na+/K+ ATPase
g2516	Na-Ca exchanger/integrin-beta4
g34226	NADP-dependent ferredoxin reductase and adrenodoxin reductase- fprA-like protein
g34645	NADP-dependent ferredoxin reductase and adrenodoxin reductase- fprA-like protein
g16	nitrite reductase-ferredoxin dependent
g30934	ornithine decarboxylase
g32402	phenylalanine tRNA synthtetase-like protein
g18390	phenylalanine-trna synthetase-like protein
g33731	phosphoenolpyruvate carboxylase
g276	phospoenolpyruvate carboxylase
g277	phospoenolpyruvate carboxylase
g1842	precursor of carboxylase pyruvate carboxylase
g19570	precursor of carboxylase pyruvate carboxylase
g14016	precursor of carboxylase pyruvate carboxylase
g32265	precursor of carboxylase pyruvate carboxylase
g24858	sodium/hydrogen exchanger 2-like
g18807	proline dehydrogenase
g20856	proline-specific peptidase
g17129	Proline--tRNA ligase
g17573	propionyl-coa carboxylase
g11364	propionyl-coa carboxylase

g1927	fucoxanthin chlorophyll a/c protein
g37647	fucoxanthin chlorophyll a/c protein
g36248	fucoxanthin chlorophyll a/c protein
g17964	fucoxanthin chlorophyll a/c protein
g4329	fucoxanthin chlorophyll a/c protein
g550	fucoxanthin chlorophyll a/c protein
g2510	fucoxanthin chlorophyll a/c protein 4
g25270	fucoxanthin chlorophyll a/c protein 4
g8095	fucoxanthin chlorophyll a/c protein -LI818 clade
g14109	fucoxanthin chlorophyll a/c protein- LI818 clade
g23429	fucoxanthin chlorophyll a/c protein- LI818 clade
g28991	fucoxanthin chlorophyll a/c protein- LI818 clade
g12734	fucoxanthin chlorophyll a/c protein- LI818 clade
g31987	fucoxanthin chlorophyll a/c protein- LI818 clade
g31988	fucoxanthin chlorophyll a/c protein- LI818 clade
g31495	fucoxanthin chlorophyll a/c protein- LI818 clade
g9937	fucoxanthin chlorophyll a/c protein- LI818 clade
g9684	fucoxanthin-chlorophyll a-c binding protein- plastid precursor
g7575	glutamate decarboxylase
g13633	glycine decarboxylase h-protein
g36273	glycine decarboxylase p-protein
g12525	glycine decarboxylase t-protein
g25288	heat shock protein
g25185	heat shock protein
g37891	heat shock protein
g21084	heat shock protein
g12165	heat shock protein
g27430	heat shock protein
g17441	protein fucoxanthin chlorophyll a/c protein
g23425	protein fucoxanthin chlorophyll a/c protein
g25269	protein fucoxanthin chlorophyll a/c protein
g6305	protein fucoxanthin chlorophyll a/c protein
g20855	protein fucoxanthin chlorophyll a/c protein
g20854	protein fucoxanthin chlorophyll a/c protein
g1208	protein fucoxanthin chlorophyll a/c protein
g1207	protein fucoxanthin chlorophyll a/c protein
g16761	protein fucoxanthin chlorophyll a/c protein
g25594	protein fucoxanthin chlorophyll a/c protein
g7036	protein fucoxanthin chlorophyll a/c protein
g26187	PSI ferredoxin-binding protein II
g30549	PSI ferredoxin-binding protein II
g24369	PSI ferredoxin-binding protein II
g28388	PSI ferredoxin-binding protein II
g30431	PSI ferredoxin-binding protein II
g4704	putative Ca2+-binding hemolysin
g23720	putative cold-shock protein
g31413	pyruvate carboxylase-like protein
g16667	reductase of ferredoxin thioredoxin reductase
g18188	ribulose-1-5-bisphosphate carboxylase/oxygenase large subun
g17139	ribulose-1-5-bisphosphate carboxylase/oxygenase small subun
g30556	Rubisco expression protein
g30438	Rubisco expression protein
g28396	Rubisco expression protein
g30440	Rubisco expression protein
g26194	Rubisco expression protein
g24360	Rubisco expression protein

g8626	heat shock protein	g30557	Rubisco expression protein
g8030	heat shock protein	g19323	Similarities Heat shock protein 70-probable alternative splicing
g10062	heat shock protein	g34671	suppressor of actin mutations protein-like protein
g20807	heat shock protein	g9491	actin binding protein cofilin-like protein
g4335	Heat shock protein 40 like protein DnaJ domain containing protein	g9007	uroporphyrinogen III decarboxylase
g28380	heat shock protein 70	g31776	vacuolar membrane proton pump- inorganic pyrophosphatase
g30522	heat shock protein 70	g27014	vacuolar proton pump alpha subunit
g29236	heat shock protein 70	g3282	vacuolar proton pump D subunit
g2414	heat shock protein belonging to the HSP70 family	g33026	v-type H+ transporting ATPase

Table S3. Detailed significantly differentially expressed genes in elevated pCO_2 where ratio represents the mean read abundance in elevated pCO_2 (1000 µatm) divided by the mean read abundance in ambient pCO_2 (380 µatm).

Gene	Gene description	R	Ratio
g1207	Fucoxanthin-chlorophyll a-c binding protein F, chloroplastic emb|CAA80677.1|	51918.00	3.61
g29771	predicted protein [T. pseudonana CCMP1335] gb|EED96591.1|	10326.80	4.12
g8518	unknown	9636.32	4.10
g29769	hypothetical protein NATL1_08871 [Prochlorococcus marinus str. NATL1A] gb|ABM75445.1|	8548.04	10.22
g19604	unknown	6813.28	4.86
g2412	predicted protein [T. pseudonana CCMP1335] gb|EED94446.1|	6782.36	8.32
g4049	unknown	5810.71	4.38
g10378	unknown	4300.46	4.43
g25269	Fucoxanthin chlorophyll a/c protein [P. tricornutum CCAP 1055/1] gb|EEC44018.1|	3605.46	3.11
g2691	unknown	3261.69	3.50
g16223	predicted protein [T. pseudonana CCMP1335] gb|EED94684.1|	2154.41	6.91
g25221	hypothetical protein THAPSDRAFT_33389 [T. pseudonana CCMP1335] gb|EED93256.1|	2151.78	4.04
g2177	expressed unknown protein [Ectocarpus siliculosus]	1910.60	3.82
g2906	NNT NAD transhydrogenase [T. pseudonana CCMP1335] gb|EED92787.1|	1901.60	4.28
g15967	predicted protein [T. pseudonana CCMP1335] gb|EED94525.1|	1743.92	3.73
g4999	FOG: TPR repeat, SEL1 subfamily [Butyrivibrio fibrisolvens 16/4]	1494.77	7.05
g26312	predicted protein [T. pseudonana CCMP1335] gb|ACI64480.1|	1490.32	5.08
g19255	unknown	1416.30	3.16
g34984	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC49407.1|	1385.07	15.84
g2213	predicted protein [T. pseudonana CCMP1335] gb|EED93577.1|	1339.24	3.33
g3347	unknown	1337.67	10.87
g31184	predicted protein [T. pseudonana CCMP1335] gb|EED90970.1|	1330.10	5.63
Tass5711	unknown	1169.02	4.42
g6811	fucoxanthin chl a/c light-harvesting protein [T. pseudonana CCMP1335] gb|EED86590.1|	1040.18	5.17
g27029	unknown	1007.84	3.64

g20820	PREDICTED: hypothetical protein XP_843354 [Canis familiaris]	1005.60	3.51
g299	unknown	959.56	3.85
g6566	predicted protein [T. pseudonana CCMP1335] gb|EED95318.1|	833.96	3.69
g2040	predicted protein [T. pseudonana CCMP1335] gb|EED94719.1|	751.16	3.37
g19676	predicted protein [T. pseudonana CCMP1335] gb|EED95041.1|	699.04	11.52
g2080	predicted protein [T. pseudonana CCMP1335] gb|EED93488.1|	634.08	10.52
g1691	thioredoxin [T. pseudonana CCMP1335] gb|EED93987.1|	626.08	3.47
g14812	hypothetical protein R2601_14365 [Roseovarius sp. HTCC2601] gb|EAU48157.1|	598.35	3.45
g17733	predicted protein [T. pseudonana CCMP1335] gb|EED90323.1|	518.76	3.96
g4387	oxidoreductase, short chain dehydrogenase/reductase family gb|EDX85241.1|	518.46	3.44
g20586	unknown	515.29	15.79
g35956	Sel1 domain-containing protein [Magnetococcus sp. MC-1] gb|ABK44504.1|	513.86	3.27
g30492	unknown	488.94	24.30
g26452	conserved hypothetical protein [Phytophthora infestans T30-4] gb|EEY57439.1|	485.90	5.10
g1079	predicted protein [T. pseudonana CCMP1335] gb|ACI64660.1|	485.86	3.16
g778	tripartite motif-containing 65, isoform CRA_b [Homo sapiens]	460.75	4.22
g33243	unknown	436.16	11.64
g16860	unknown	429.78	3.36
g33755	predicted protein [T. pseudonana CCMP1335] gb|EED96180.1|	407.35	3.17
g27127	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC44967.1|	404.53	6.04
g35803	COG0790: FOG: TPR repeat, SEL1 subfamily [Haemophilus influenzae R2846]	367.55	3.60
g37778	predicted protein [T. pseudonana CCMP1335] gb|EED90826.1|	353.65	3.23
g2513	cell surface protein [Methanosarcina acetivorans C2A] gb|AAM07636.1|	351.08	3.12
g31889	hypothetical protein THAPSDRAFT_261827 [T. pseudonana CCMP1335] gb|EED93810.1|	329.43	5.44
g25222	shikimate kinase [Marinobacter sp. ELB17] gb|EBA00232.1|	327.52	9.31
g18672	unknown	317.21	6.49
g17672	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC45419.1|	311.18	3.29
g36876	predicted protein [T. pseudonana CCMP1335] gb|EED91931.1|	310.57	19.85
g10738	unknown	304.15	3.93
g13157	hypothetical protein THAPSDRAFT_264671 [T. pseudonana CCMP1335] gb|EED87914.1|	295.45	7.73

g30401	unknown	291.39	3.47
g26335	PKD domain containing protein [Opitutaceae bacterium TAV2] gb|EEG20097.1|	290.39	13.93
g12787	unknown	281.37	3.46
g1340	unknown	269.18	3.53
g37400	unknown	266.66	3.21
g16582	unknown	233.17	4.54
g19196	unknown	223.78	3.40
g4258	predicted protein [T. pseudonana CCMP1335] gb|EED89900.1|	207.49	5.33
g20948	unknown	192.16	3.36
g32826	predicted protein [T. pseudonana CCMP1335] gb|EED90770.1|	188.07	3.99
g31504	predicted protein [Micromonas sp. RCC299] gb|ACO67020.1|	178.87	3.59
g31805	predicted protein [T. pseudonana CCMP1335] gb|EED86462.1|	176.53	3.72
g10652	unknown	174.98	4.05
g4949	unknown	162.53	6.49
g15044	hypothetical protein MAB_2456 [Mycobacterium abscessus ATCC 19977] emb|CAM62537.1|	156.95	3.59
g6064	predicted protein [T. pseudonana CCMP1335] gb|EED95041.1|	153.02	4.50
g22699	hypothetical protein all7332 [Nostoc sp. PCC 7120] dbj|BAB77090.1|	144.98	3.22
g1579	enhanced protein,, tetratricopeptide repeat family [C. burnetii Dugway 5J108-111] gb|ABS76720.1|	130.87	15.23
g25705	unknown	126.21	3.64
g32267	unknown	125.81	3.17
g15306	unknown	109.36	5.37
g846	predicted protein [T. pseudonana CCMP1335] gb|EED90428.1|	104.42	4.35
g20729	unknown	101.13	5.85
g6790	predicted protein [T. pseudonana CCMP1335] gb|EED86325.1|	96.35	4.52
g34751	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC43682.1|	93.65	3.31
g5655	predicted protein [T. pseudonana CCMP1335] gb|EED95317.1|	91.14	4.65
Tass2435	unknown	87.54	15.37
g5055	predicted protein [T. pseudonana CCMP1335] gb|EED87214.1|	82.55	3.60
g8699	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC49724.1|	80.95	3.18
Tass604	unknown	77.04	3.58

g36297	unknown	74.40	4.05
g1214	hypothetical protein BRAFLDRAFT_94613 [Branchiostoma floridae] gb|EEN43721.1|	69.56	3.29
g33281	unknown	68.89	3.47
g34983	predicted protein [Nematostella vectensis] gb|EDO43196.1|	61.99	5.40
g27292	predicted protein [T. pseudonana CCMP1335] gb|ACI64808.1|	60.85	3.33
g21445	unknown	57.53	3.91
g30941	predicted protein [T. pseudonana CCMP1335] gb|EED94104.1|	54.76	7.33
g36270	unknown	51.77	10.83
g23803	unknown	51.28	3.75
Tass5079	unknown	50.92	3.63
g21126	unknown	49.93	3.20
g28612	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC44875.1|	44.78	3.59
g25732	unknown	43.81	5.29
g33147	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC48580.1|	42.66	3.40
Tass11045	unknown	41.88	3.20
g33396	conserved hypothetical protein [uncultured Prochlorococcus marinus clone ASNC1092]	41.13	6.79
g8153	predicted protein [T. pseudonana CCMP1335] gb|EED92122.1|	38.41	4.27
g21381	unknown	38.05	4.46
g1571	unknown	37.97	4.51
g28780	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC47111.1|	37.76	3.27
g27502	hypothetical protein PE36_12787 [Moritella sp. PE36] gb|EDM67322.1|	36.31	4.08
g21380	subtilisin-like serine protease [T. pseudonana CCMP1335] gb|EED89728.1|	35.93	6.58
Tass11054	unknown	35.82	3.44
g15465	expressed protein [Oryza sativa (japonica cultivar-group)]	35.13	3.13
g31818	unknown	32.58	3.52
g34930	unknown	32.29	3.64
g17703	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC43604.1|	32.06	3.57
g14105	predicted protein [T. pseudonana CCMP1335] gb|ACI64742.1|	32.02	3.21
g9740	zinc finger family protein [Arabidopsis lyrata subsp. lyrata] gb|EFH58975.1|	31.30	9.17
g8677	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC44195.1|	30.97	4.04

g22475	unknown	30.85	3.97
Tass9110	unknown	29.55	4.23
g10391	predicted protein [T. pseudonana CCMP1335] gb|EED88426.1|	29.50	3.37
g3720	unknown	29.26	3.81
g24747	predicted protein [T. pseudonana CCMP1335] gb|EED96320.1|	27.52	3.23
g4232	predicted protein [T. pseudonana CCMP1335] gb|ACI64176.1|	27.31	5.21
g9027	hypothetical protein THAPSDRAFT_264671 [T. pseudonana CCMP1335] gb|EED87914.1|	27.13	5.63
g23643	COG0790: FOG: TPR repeat, SEL1 subfamily [Haemophilus influenzae R2846]	25.29	3.13
g8952	unknown	25.20	19.17
g596	unknown	24.60	3.36
g1222	unknown	24.41	3.40
g4922	unknown	24.24	5.06
g35810	unknown	24.08	3.30
g23542	predicted protein [T. pseudonana CCMP1335] gb|EED88309.1|	23.71	3.88
g3533	unknown	23.26	3.60
g3821	unknown	22.94	3.61
g28285	replication protein A2 [T. pseudonana CCMP1335] gb|EED92004.1|	22.76	3.55
g36516	unknown	22.65	3.62
g37613	unknown	22.44	3.38
g27367	NADH dehydrogenase subunit 7 [T. pseudonana] gb|AAZ99426.1|	22.13	5.52
Tass2293	unknown	21.59	5.30
g32077	unknown	21.50	3.83
g11655	unknown	21.00	3.91
g25991	unknown	20.91	6.30
g22457	predicted protein [T. pseudonana CCMP1335] gb|ACI64757.1|	20.68	3.18
g8260	unknown	20.56	4.17
g34402	unknown	20.22	3.59
g20895	Sel1 domain protein repeat-containing prot. [alpha proteobacterium BAL199] gb|EDP64864.1|	19.00	5.07
g14223	unknown	18.82	3.30
g25993	unknown	18.79	7.92

g11876	predicted protein [T. pseudonana CCMP1335] gb|EED91211.1|	18.61	4.22
g8036	hypothetical protein BRAFLDRAFT_117088 [Branchiostoma floridae] gb|EEN67695.1|	18.20	4.58
Tass4713	unknown	17.99	4.17
g22602	unknown	17.97	3.45
g13672	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC50029.1|	17.59	3.89
g34008	unknown	17.29	6.17
g35379	unknown	17.12	3.21
g20188	unknown	17.04	17.08
g30038	unknown	16.93	4.26
g34583	alkaline phosphatase-like protein [Teredinibacter turnerae T7901] gb|ACR13720.1|	16.85	10.00
g27362	ribosomal protein S11 [T. pseudonana] gb|AAZ99415.1|	16.72	5.19
g36877	unknown	16.52	3.57
g5434	unknown	16.21	3.99
g36210	unknown	16.04	3.92
g7186	unknown	16.01	3.59
g13314	predicted protein [T. pseudonana CCMP1335] gb|EED86905.1|	15.82	3.13
g29740	unknown	15.79	6.46
g23928	unknown	15.15	3.33
g14946	predicted protein [T. pseudonana CCMP1335] gb|EED92674.1|	15.03	3.59
g36895	predicted protein [T. pseudonana CCMP1335] gb|ACI65031.1|	15.01	4.44
g30007	unknown	14.89	3.82
g6473	unknown	14.66	3.72
g19004	alkaline phosphatase [T. pseudonana CCMP1335] gb|EED95733.1|	14.57	6.67
g6043	predicted protein [T. pseudonana CCMP1335] gb|EED88523.1|	14.55	3.82
g31167	unknown	14.52	5.00
g24161	flap endonuclease-1 [Candidatus Methanoregula boonei 6A8] gb|ABS56620.1|	14.42	4.32
g3646	unknown	14.36	4.65
g30046	myosin heavy chain [Dictyostelium discoideum]	14.33	4.06
g2540	unknown	14.30	4.38
g24625	unknown	14.27	3.14

Tass373	cytochrome oxidase subunit II [T. pseudonana] gb|AAZ99398.1|	14.14	3.22
g4694	unknown	14.12	5.00
g36596	unknown	13.73	15.00
g7758	carboxylesterase [Mycobacterium parascrofulaceum ATCC BAA-614] gb|EFG79051.1|	13.49	4.17
g28530	unknown	13.48	6.15
g4699	predicted protein [T. pseudonana CCMP1335] gb|EED93532.1|	13.42	3.65
g2083	unknown	13.40	7.08
g15269	unknown	13.26	3.87
g11420	predicted protein [T. pseudonana CCMP1335] gb|EED88641.1|	13.20	3.33
g34801	unknown	13.11	7.71
g17999	unknown	12.89	3.66
g32947	Sel1 repeat-containing protein [Chlorobium chlorochromatii CaD3] gb|ABB29186.1|	12.78	6.17
g32946	conserved hypothetical protein [Oxalobacter formigenes HOxBLS] gb|EEO27702.1|	12.68	3.65
g18881	unknown	12.61	4.00
g16246	unknown	12.16	4.72
g32597	unknown	12.15	5.00
g7664	TPR repeat-containing SEL1 subfamily protein [Acinetobacter baumannii AB900]	12.09	3.19
g8766	PREDICTED: similar to predicted protein [Hydra magnipapillata]	12.02	3.47
g32155	unknown	11.91	5.36
g21988	unknown	11.65	13.33
Tass5238	unknown	11.63	3.85
g30589	unknown	11.52	4.24
g8333	unknown	11.39	3.40
g3807	TPR repeat-containing SEL1 subfamily protein [Acinetobacter baumannii AB900]	11.31	3.38
g16708	predicted protein [T. pseudonana CCMP1335] gb|EED89884.1|	11.05	19.17
g35323	predicted protein [T. pseudonana CCMP1335] gb|EED86996.1|	11.00	3.25
g21221	unknown	10.80	6.67
g22721	predicted protein [T. pseudonana CCMP1335] gb|EED88043.1|	10.77	3.29
g20728	unknown	10.52	3.33
g26547	unknown	10.52	4.79

g34396	predicted protein [T. pseudonana CCMP1335] gb|EED91771.1|	10.48	8.75
g28806	unknown	10.47	3.69
g8283	unknown	10.45	3.13
Tass4124	unknown	10.40	4.58
g34696	unknown	10.37	3.33
g4817	unknown	10.36	4.31
g37627	unknown	10.21	3.13
g3050	unknown	10.17	6.25
Tass2605	unknown	10.06	5.00
g27788	unknown	10.00	3.79
g25486	hypothetical protein Aasi_1217 [Candidatus Amoebophilus asiaticus 5a2] gb|ACE06544.1|	10.00	4.33
g677	unknown	9.99	6.94
g10419	mu subunit of AP4-like protein [T. pseudonana CCMP1335] gb|EED96067.1|	9.97	3.33
g21407	unknown	9.89	6.39
g24319	unknown	9.83	3.43
g18993	unknown	9.61	3.96
g28818	unknown	9.59	4.72
g2922	predicted protein [T. pseudonana CCMP1335] gb|EED94104.1|	9.53	3.89
g137	hypothetical protein Aasi_0854 [Candidatus Amoebophilus asiaticus 5a2] gb|ACE06224.1|	9.35	10.83
g32838	unknown	9.33	3.15
g32836	unknown	9.19	6.00
g25500	unknown	9.18	3.33
g32268	unknown	9.16	5.33
g3129	Sel1 domain protein repeat-containing prot. [Haemophilus influenzae 7P49H1] gb|EEP45669.1|	9.13	3.17
g19404	unknown	9.13	4.06
g30336	unknown	8.94	3.53
g22233	unknown	8.84	3.13
g15126	unknown	8.83	3.56
g31134	unknown	8.79	6.94
g18216	unknown	8.71	5.83

Tass9655	NADH dehydrogenase subunit 9 [T. pseudonana] gb|AAZ99425.1|	8.68	6.67
g37604	unknown	8.67	3.15
g3875	hypothetical protein THAPSDRAFT_15858 [T. pseudonana CCMP1335] gb|EED95351.1|	8.66	3.47
g37420	unknown	8.63	3.50
g5761	unknown	8.62	5.83
g2532	exonuclease [T. pseudonana CCMP1335] gb|EED90855.1|	8.57	4.29
g11413	unknown	8.44	3.39
g24664	unknown	8.41	5.12
g27911	unknown	8.30	3.33
g20405	unknown	8.26	3.50
g11326	unknown	8.25	6.67
g23115	unknown	8.21	3.52
g35378	unknown	8.11	10.83
g17804	unknown	8.04	3.15
g9188	unknown	8.01	4.38

Table S4. Detailed significantly differentially expressed genes in ambient pCO_2 where ratio represent the mean read abundance in elevated pCO_2 (1000 µatm) divided by the mean read abundance in ambient pCO_2 (380 µatm).

Gene	Gene description	R	Ratio
g31355	unknown	15922.45	0.15
g19525	unknown	7325.98	0.20
g24104	predicted protein [T. pseudonana CCMP1335] gb|EED90138.1|	6671.93	0.15
g8305	predicted protein [T. pseudonana CCMP1335] gb|EED95371.1|	3793.33	0.17
g15673	predicted protein [T. pseudonana CCMP1335] gb|ACI64504.1|	3470.88	0.22
g22649	predicted protein [T. pseudonana CCMP1335] gb|EED88551.1|	2749.91	0.22
g29743	unknown	2500.54	0.24
g2790	unknown	2498.58	0.16
g31520	myosin heavy chain-like protein [T. pseudonana CCMP1335] gb|EED88779.1|	1795.31	0.19
g2724	predicted protein [T. pseudonana CCMP1335] gb|EED86247.1|	1794.52	0.23
g36613	unknown	1719.30	0.22
g36024	predicted protein [T. pseudonana CCMP1335] gb|EED92555.1|	1666.85	0.11
g14174	unknown	1494.43	0.20
g23287	unknown	1470.95	0.16
g33657	unknown	1447.62	0.21
g32757	unknown	1414.29	0.25
g23821	unknown	1248.55	0.23
g31299	unknown	1226.85	0.20
g18960	unknown	1123.85	0.20
g21470	predicted protein [T. pseudonana CCMP1335] gb|EED90337.1|	943.87	0.20
g23789	epsilon frustilin [Navicula pelliculosa]	938.07	0.25
g25760	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC45730.1|	880.52	0.20
g16788	unknown	836.55	0.09
g32426	unknown	828.29	0.21

g16763	unknown	807.12	0.25
g31908	unknown	755.98	0.24
g4663	predicted protein [T. pseudonana CCMP1335] gb|EED96119.1|	606.33	0.21
g27435	hypothetical protein Psyc_0360 [Psychrobacter arcticus 273-4] gb|AAZ18229.1|	536.28	0.15
g17424	predicted protein [T. pseudonana CCMP1335] gb|EED90483.1|	525.15	0.18
g7853	unknown	513.53	0.20
g10797	unknown	493.00	0.21
g22601	unknown	489.28	0.18
g7677	unknown	487.97	0.23
g31120	unknown	459.00	0.22
g3276	myosin heavy chain-like protein [T. pseudonana CCMP1335] gb|EED88779.1|	457.72	0.19
g20707	pectate lyase [Haloferax volcanii DS2] gb|ADE01341.1|	450.53	0.24
g13296	unknown	427.44	0.16
g28470	predicted protein [T. pseudonana CCMP1335] gb|EED95364.1|	422.84	0.17
g4085	predicted protein [T. pseudonana CCMP1335] gb|EED94366.1|	415.90	0.23
g29942	unknown	372.90	0.21
g22791	unknown	370.14	0.17
g14374	Hypothetical protein CBG18550 [Caenorhabditis briggsae] emb|CAP35977.1|	362.01	0.17
g9305	predicted protein [T. pseudonana CCMP1335] gb|EED90483.1|	345.63	0.19
g34366	pol protein [T. pseudonana]	338.42	0.21
g4559	unknown	336.68	0.06
g34796	unknown	333.36	0.24
g37066	unknown	329.64	0.20
g23148	unknown	316.16	0.21
g19566	unknown	312.06	0.24
g25618	unknown	294.09	0.20
g26731	unknown	289.46	0.17
g7866	Serine/threonine protein kinase [Ectocarpus siliculosus]	288.85	0.25
g10551	unknown	281.03	0.16
g29247	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	275.15	0.18

g33668	unknown	269.27	0.22
g14969	predicted protein [Micromonas sp. RCC299] gb|ACO66050.1|	254.96	0.18
g16710	unknown	247.06	0.21
g18983	unknown	240.33	0.22
g16685	unknown	236.09	0.16
g25299	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	235.41	0.21
g17478	unknown	233.96	0.20
g22828	unknown	221.34	0.25
g1506	unknown	215.52	0.24
g37640	unknown	212.33	0.25
g26721	pol protein [T. pseudonana]	205.26	0.25
g8304	unknown	204.41	0.10
g14582	unknown	203.10	0.18
g30636	unknown	201.44	0.20
g13790	sel1 repeat protein [Tetrahymena thermophila] gb|EAR88461.1|	191.55	0.12
g36714	conserved hypothetical protein [Phytophthora infestans T30-4] gb|EEY54214.1|	190.35	0.24
g10331	unknown	187.45	0.20
g11422	unknown	185.65	0.24
g9780	unknown	184.60	0.20
g7104	conserved hypothetical protein [Phytophthora infestans T30-4] gb|EEY68651.1|	184.58	0.24
g27710	unknown	178.61	0.22
g36721	unknown	177.10	0.22
g20706	collagen, type XVIII, alpha 1 [Xenopus laevis] gb|AAL14257.1|	165.75	0.25
g9212	unknown	164.69	0.19
g8722	unknown	161.95	0.09
g11899	unknown	159.62	0.19
g17395	predicted protein [T. pseudonana CCMP1335] gb|EED93652.1|	151.55	0.21
g19074	unknown	150.47	0.23
g9364	unknown	148.41	0.14
g4360	PREDICTED: similar to conserved hypothetical protein, partial [Ciona intestinalis]	145.58	0.20

g29613	predicted protein [T. pseudonana CCMP1335] gb|EED89370.1|	141.45	0.22
g25617	predicted protein [T. pseudonana CCMP1335] gb|EED90427.1|	140.47	0.17
g11773	hypothetical protein [Monosiga brevicollis MX1] gb|EDQ88709.1|	140.19	0.25
g30280	unknown	136.42	0.25
g15441	unknown	136.03	0.25
g15398	predicted protein [Micromonas sp. RCC299] gb|ACO62854.1|	135.35	0.24
g28690	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	134.41	0.24
g37838	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC47605.1|	134.26	0.22
g4561	unknown	133.04	0.20
g12315	unknown	133.02	0.23
g7225	unknown	132.97	0.24
g35192	cytoplasmic dynein heavy chain-like protein [T. pseudonana CCMP1335] gb|ACI64606.1|	130.48	0.21
g5654	predicted protein [T. pseudonana CCMP1335] gb|EED89555.1|	126.75	0.12
g19858	predicted protein [T. pseudonana CCMP1335] gb|ACI64335.1|	126.63	0.19
g33713	unknown	125.06	0.11
g10807	unknown	123.30	0.24
g12610	unknown	122.89	0.21
g17608	unknown	122.31	0.23
g34140	unknown	119.22	0.18
g30141	unknown	119.14	0.21
g6275	unknown	119.09	0.21
g6918	unknown	118.53	0.13
g31345	unknown	114.74	0.23
g28556	unknown	112.26	0.18
g31662	unknown	110.98	0.08
g37686	unknown	110.72	0.20
g30406	unknown	106.94	0.20
g32908	unknown	106.00	0.10
g28262	unknown	105.94	0.20
g31500	unknown	104.76	0.20

g33814	unknown	104.30	0.13
g37486	unknown	102.70	0.14
g17574	unknown	102.66	0.22
g33564	unknown	102.46	0.13
g25904	predicted protein [T. pseudonana CCMP1335] gb|EED96376.1|	100.48	0.20
g33941	unknown	98.54	0.16
g22068	unknown	98.37	0.24
g28662	unknown	98.02	0.11
g9749	unknown	96.77	0.23
g19191	unknown	96.17	0.25
g36577	predicted protein [T. pseudonana CCMP1335] gb|ACI65036.1|	94.42	0.24
g28934	unknown	93.90	0.10
g10184	unknown	93.82	0.10
g4896	unknown	93.05	0.19
g653	unknown	90.51	0.20
g33335	hypothetical protein Aasi_0854 [Candidatus Amoebophilus asiaticus 5a2] gb|ACE06224.1|	88.91	0.19
g6114	unknown	88.02	0.23
g14588	unknown	84.72	0.18
g35300	unknown	82.89	0.11
g33059	predicted protein [T. pseudonana CCMP1335] gb|EED86272.1|	79.16	0.18
g7238	unknown	78.36	0.23
g25116	unknown	77.27	0.23
g12286	predicted protein [T. pseudonana CCMP1335] gb|EED91579.1|	76.91	0.24
g35193	cytoplasmic dynein heavy chain-like protein [T. pseudonana CCMP1335] gb|ACI64606.1|	76.36	0.21
g10592	hypothetical protein BRAFLDRAFT_131122 [Branchiostoma floridae] gb|EEN53864.1|	75.18	0.18
g25917	unknown	75.06	0.15
g22424	unknown	74.71	0.21
g13797	predicted protein [T. pseudonana CCMP1335] gb|EED91646.1|	74.69	0.17
g19821	unknown	72.99	0.24
g24353	unknown	72.61	0.25

g26415	unknown	72.32	0.13
g34904	unknown	71.75	0.23
g35665	predicted protein [T. pseudonana CCMP1335] gb|EED93640.1|	67.90	0.24
g19358	predicted protein [T. pseudonana CCMP1335] gb|EED88827.1|	67.78	0.17
g19375	unknown	67.06	0.18
g17609	unknown	66.38	0.12
g37892	unknown	64.89	0.24
g13991	unknown	64.78	0.24
g12011	unknown	64.27	0.17
g18566	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC43895.1|	64.01	0.22
g27186	pol protein [T. pseudonana]	62.88	0.20
g9066	unknown	62.66	0.20
g28627	pol protein [T. pseudonana]	61.88	0.18
g36186	unknown	60.48	0.22
g33895	predicted protein [T. pseudonana CCMP1335] gb|EED87337.1|	60.18	0.22
g31811	unknown	58.59	0.22
g10920	unknown	57.92	0.08
g24620	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	57.38	0.14
g14869	unknown	56.80	0.17
g23666	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] gb|EEC46601.1|	56.67	0.25
g3974	unknown	56.45	0.25
g20724	predicted protein [T. pseudonana CCMP1335] gb|EED86411.1|	54.98	0.25
g13210	pol protein [T. pseudonana]	54.34	0.18
g13828	unknown	54.14	0.17
g26048	unknown	52.15	0.25
g6917	unknown	51.82	0.19
g16279	predicted protein [T. pseudonana CCMP1335] gb|EED87346.1|	51.37	0.15
g2096	unknown	51.26	0.24
g9171	predicted protein [T. pseudonana CCMP1335] gb|EED86536.1|	50.86	0.15
g5389	unknown	50.10	0.18

g29131	predicted protein [T. pseudonana CCMP1335] gb|EED89253.1|	49.63	0.21
g24699	unknown	49.40	0.05
g36959	predicted protein [Phaeodactylum tricornutum CCAP 1055/1] ref|XP_002185602.1|	49.23	0.23
g27762	unknown	48.67	0.24
g4783	unknown	48.58	0.20
g27690	unknown	47.78	0.25
g340	unknown	47.77	0.18
g30691	unknown	46.76	0.19
g30151	unknown	46.68	0.03
g912	unknown	46.63	0.08
g23160	putative lipoprotein [Plesiocystis pacifica SIR-1] gb|EDM77756.1|	46.42	0.22
g16956	unknown	46.34	0.25
g17677	flagellar associated protein putative: flagellar associated protein [Ectocarpus siliculosus]	45.19	0.24
g5879	predicted protein [T. pseudonana CCMP1335] gb|EED93667.1|	45.00	0.25
g29495	unknown	44.38	0.15
Tass7588	unknown	44.21	0.25
g10309	unknown	43.28	0.23
g11495	unknown	42.58	0.24
g17542	unknown	42.47	0.21
g19274	unknown	42.22	0.09
g21686	unknown	42.06	0.24
g30755	unknown	40.83	0.11
g10724	unknown	40.59	0.20
g36539	predicted protein [T. pseudonana CCMP1335] gb|EED88050.1|	40.36	0.22
g21711	pol protein [Phaeodactylum tricornutum]	40.05	0.21
g17637	unknown	39.90	0.19
g25189	unknown	39.15	0.24
g7425	unknown	39.02	0.19
g7971	predicted protein [T. pseudonana CCMP1335] gb|EED90321.1|	38.66	0.22
g30640	pol protein [T. pseudonana]	38.43	0.22

g7911	unknown	38.37	0.19
g37002	unknown	37.42	0.10
g18956	predicted protein [T. pseudonana CCMP1335] gb|EED86982.1|	37.36	0.23
g24244	hypothetical protein LgasM_08201 [Lactobacillus gasseri MV-22]	37.34	0.23
g31630	unknown	37.12	0.24
g17898	unknown	36.67	0.18
g25846	unknown	36.50	0.18
g24917	unknown	35.47	0.21
g6095	unknown	35.11	0.24
g37001	unknown	35.03	0.18
g18364	unknown	34.67	0.11
g1175	unknown	34.56	0.08
g23954	unknown	33.90	0.18
g16711	unknown	33.62	0.20
g7869	unknown	33.04	0.19
g29982	unknown	32.94	0.19
g25243	unknown	32.15	0.13
Tass2979	unknown	32.11	0.21
g35975	unknown	32.09	0.25
g35670	unknown	32.05	0.18
g32699	unknown	31.77	0.24
g31631	unknown	31.45	0.14
g29219	unknown	31.28	0.15
g33777	unknown	31.08	0.22
g16841	unknown	30.76	0.25
g15628	unknown	29.76	0.13
g34416	unknown	29.67	0.24
g19029	unknown	29.49	0.19
g28417	unknown	29.12	0.20
g29922	unknown	29.07	0.17

g29364	predicted protein [T. pseudonana CCMP1335] gb|EED95057.1|	28.47	0.15
g35064	unknown	27.93	0.08
g34144	unknown	27.93	0.22
g25662	unknown	27.68	0.20
g917	unknown	27.62	0.17
g17208	predicted protein [T. pseudonana CCMP1335] gb|EED89950.1|	27.47	0.22
g13132	conserved unknown protein [Ectocarpus siliculosus]	27.23	0.17
g25850	pol protein [T. pseudonana]	26.79	0.20
g16132	unknown	26.78	0.23
g9703	conserved unknown protein [Ectocarpus siliculosus]	26.54	0.24
g23210	unknown	26.32	0.17
g19254	hypothetical protein DORFOR_00347 [Dorea formicigenerans ATCC 27755] gb|EDR48223.1|	25.74	0.21
g25600	predicted protein [T. pseudonana CCMP1335] gb|EED86325.1|	25.46	0.13
g9409	Serine/threonine-protein kinase PLK4 gb|EAL29859.1|	25.36	0.14
g36264	CBL-interacting serine/threonine-protein kinase [Phytophthora infestans T30-4] gb|EEY62050.1|	25.33	0.22
g21476	unknown	24.87	0.25

4.2.4 Publication V.

Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms.

Roy *et al.*, 2013
Biogeosciences 2013, 10, 555:566
doi:10.5194/bg-10-555-2013

Biogeosciences, 10, 555–566, 2013
www.biogeosciences.net/10/555/2013/
doi:10.5194/bg-10-555-2013

Biogeosciences

Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms

A.-S. Roy[1], S. M. Gibbons[2,3], H. Schunck[1], S. Owens[2,8], J. G. Caporaso[2,4], M. Sperling[1,9], J. I. Nissimov[5], S. Romac[6], L. Bittner[6], M. Mühling[10], U. Riebesell[1], J. LaRoche[1,*], and J. A. Gilbert[2,7]

[1]GEOMAR Helmholtz Center for Ocean Research Kiel, Düsternbrooker Weg 20, 24105, Kiel, Germany
[2]Argonne National Laboratory, Institute for Genomic and Systems Biology, 9700, S. Cass Ave. Lemont, IL 60439, USA
[3]Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637, USA
[4]Department of Computer Science, Northern Arizona University, Flagstaff, AZ 86001, USA
[5]Plymouth Marine Laboratory, Prospect Place, Plymouth, Devon, PL1 3DH, UK
[6]CNRS, UMR7144 (Paris 6), Université Pierre et Marie Curie, Station Biologique de Roscoff, Place Georges Teissier, 29682, Roscoff, France
[7]Department of Ecology and Evolution, University of Chicago, 5801 South Ellis Avenue, Chicago, IL 60637, USA
[8]Computation Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[9]Alfred-Wegener-Institute, Am Handelshafen 12, 27570, Bremerhaven, Germany
[10]Institute of Biological Sciences, TU Bergakademie Freiberg, Leipziger Str. 29, 09599, Freiberg, Germany
[*]present address: Dalhousie University, 5850 University Avenue, Halifax, NS, B3K 6R8, Canada

Correspondence to: A.-S. Roy (sroy@geomar.de)

Received: 31 August 2012 – Published in Biogeosciences Discuss.: 26 September 2012
Revised: 19 December 2012 – Accepted: 2 January 2013 – Published: 29 January 2013

Abstract. The impact of ocean acidification and carbonation on microbial community structure was assessed during a large-scale in situ costal pelagic mesocosm study, included as part of the EPOCA 2010 Arctic campaign. The mesocosm experiment included ambient conditions (fjord) and nine mesocosms with pCO_2 levels ranging from $\sim$145 to $\sim$1420 µatm. Samples for the present study were collected at ten time points ($t-1$, $t1$, $t5$, $t7$, $t12$, $t14$, $t18$, $t22$, $t26$ to $t28$) in seven treatments (ambient fjord ($\sim$145), $2\times\sim$185, $\sim$270, $\sim$685, $\sim$820, $\sim$1050 µatm) and were analysed for "small" and "large" size fraction microbial community composition using 16S rRNA (ribosomal ribonucleic acid) amplicon sequencing. This high-throughput sequencing analysis produced $\sim$20 000 000 16S rRNA V4 reads, which comprised 7000 OTUs. The main variables structuring these communities were sample origins (fjord or mesocosms) and the community size fraction (small or large size fraction). The community was significantly different between the unenclosed fjord water and enclosed mesocosms (both control and elevated CO_2 treatments) after nutrients were added to the mesocosms, suggesting that the addition of nutrients is the primary driver of the change in mesocosm community structure. The relative importance of each structuring variable depended greatly on the time at which the community was sampled in relation to the phytoplankton bloom. The sampling strategy of separating the small and large size fraction was the second most important factor for community structure. When the small and large size fraction bacteria were analysed separately at different time points, the only taxon pCO_2 was found to significantly affect were the Gammaproteobacteria after nutrient addition. Finally, pCO_2 treatment was found to be significantly correlated (non-linear) with 15 rare taxa, most of which increased in abundance with higher CO_2.

1 Introduction

The acidification of our oceans, caused predominantly by dissolution of anthropogenic carbon dioxide (CO_2) in seawater, has the potential to affect the physiology of marine microbes. Therefore, because marine microbes play a major role in global biogeochemical cycles, this increase may

have unforeseen consequences on ocean biogeochemistry (Falkowski et al., 2008; Worden and Not, 2008). Experimental manipulation of the partial pressure of carbon dioxide (pCO_2) in marine mesocosms has demonstrated species-specific physiological responses to elevated dissolved CO_2 concentrations. For example, delayed or decreased coccolithophore calcification (Delille et al., 2005), a significant increase in photosynthetic capacity (Fu et al., 2008), higher CO_2 and N_2 fixation (Hutchins et al., 2007), and a decreased abundance of picoeukaryotes (Newbold et al., 2012) have been observed. However, the response of bacterial communities to elevated pCO_2 concentrations is less defined, with mixed reports of both significant increases in bacterial protein production (Grossart et al., 2006), and no significant changes in microbial community structure (Tanaka et al., 2008; Allgaier et al., 2008; Newbold et al., 2012). For example, during the 2008 PeECE III mesocosms study, elevated pCO_2 had no significant impact on bacterial abundance, diversity, or activity; however, the community structure of the small size fraction bacteria was significantly altered by the induced phytoplankton bloom (Allgaier et al., 2008; Arnosti et al., 2011; Riebesell et al., 2008).

While these existing studies have observed little impact of elevated pCO_2 on microbial community structure, they were all performed with molecular techniques that offered limited taxonomic resolution (e.g. High-Performance Liquid Chromatography, Denaturing Gradient Gel Electrophoresis, Terminal Restriction Fragment Length Polymorphism). To improve that resolution, this study employed high-throughput amplicon sequencing of 16S rRNA to characterize microbial taxonomic community dynamics. High-throughput amplicon sequencing provides an efficient method to obtain a deep molecular overview of microbial community structure, without having to cultivate environmental isolates (Agogué et al., 2011; Gilbert et al., 2009; Hubert et al., 2007; Huse et al., 2008; Margulies et al., 2005; Sogin et al., 2006). In this study, the variation of microbial assemblages was characterised through time, across a gradient of pCO_2, in a large-scale in situ pelagic mesocosm experiment in the coastal Arctic Ocean. In addition to characterizing the detailed response of the microbial community structure to elevated pCO_2, the analysis of the 16S rRNA database provided insight on the effect of isolating the water column in a mesocosm, and to investigate the community structure response to elevated pCO_2.

2 Methods

2.1 Location and carbonate system manipulation

The European Project on Ocean Acidification (EPOCA) supported a large mesocosm experiment in the Arctic which was conducted in the water of Kongsfjorden, Svalbard, Norway (78°56.2′ N, 11°53.6′ E) during the months of June and July 2010. Throughout the experiment, diverse environmental parameters were measured to explore the effect of ocean acidification (OA) on multiple biological processes. Briefly, nine mesocosms containing about 45 m^3 of seawater and reaching down to 15 m depth were deployed from Ny-Ålesund and pCO_2 was manipulated by the addition of CO_2-saturated seawater in seven mesocosms, resulting in initial pCO_2 ranging from ∼ 286 to ∼ 1420 µatm. Two of the mesocosms were not manipulated and served as controls with starting pCO_2 of ∼ 185 µatm. Additionally, samples were taken directly from the fjord (initial pCO_2 ∼ 145 µatm) in which the mesocosms were suspended and from which the mesocosm water originated. These samples were used to monitor any natural changes in pCO_2 that may occur in the ambient water during the course of the experiment and were also important for detecting deviations in pCO_2 between the fjord and the untreated mesocosms with time. To promote phytoplankton growth, all nine mesocosms were subjected to nutrient additions (nitrate (NO_3), phosphate (PO_4) and silicate (Si)) on day (t) 13, creating pre-nutrient ($t-1$ to $t12$) and post-nutrient ($t13$ to $t30$) periods (Fig. 1). Detailed information about the experimental set-up, the mesocosms deployment, the carbonate chemistry, and the nutrients additions can be found in this issue in Riebesell et al. (2013), Czerny et al. (2013a, b), Bellerby et al. (2013), and Schulz et al. (2013), respectively.

2.2 Sampling, filtration and sample selection

A total of 10 L of water was collected using integrated water sampler (Hydrobios, Kiel, Germany) between 0 and 12 m water depth, from the fjord (∼ 145 µatm), and six mesocosms (starting $pCO_2 = 2\times \sim 185$, ∼ 270, ∼ 685, ∼ 820, ∼ 1050 µatm) on $t-1$, $t1$, $t5$, $t7$, $t12$, $t14$, $t18$, $t22$, $t26$ and $t28$ (Fig. 1). Only six of the mesocosms were chosen for this study due to time, personnel and equipment constraints. The collected water was first pre-filtered on a 20 µm sieve, and sequentially filtered through a 10 µm, a 3 µm filter to isolate associate-particle bacterial fraction (large size fraction) and through a 0.2 µm filter to isolate the small size fraction (Durapore® 47 mm, Millipore). To avoid nucleic acid degradation, processing of the samples from filtration to flash-freezing (in liquid nitrogen) was performed within 30 min of the sampling event. Samples were then stored at −80 °C until DNA/RNA extraction.

2.3 DNA extraction, PCR, and Sequencing

Total nucleic acid was extracted from the 0.2 and 3 µm filters using the "Total RNA and DNA purification – NucleoSpin® RNA II RNA/DNA buffer" kits from Macherey-Nagel (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Standard protocol with minor modifications was followed. Changes to the protocol included making the filters brittle by immersing the samples in liquid nitrogen while still

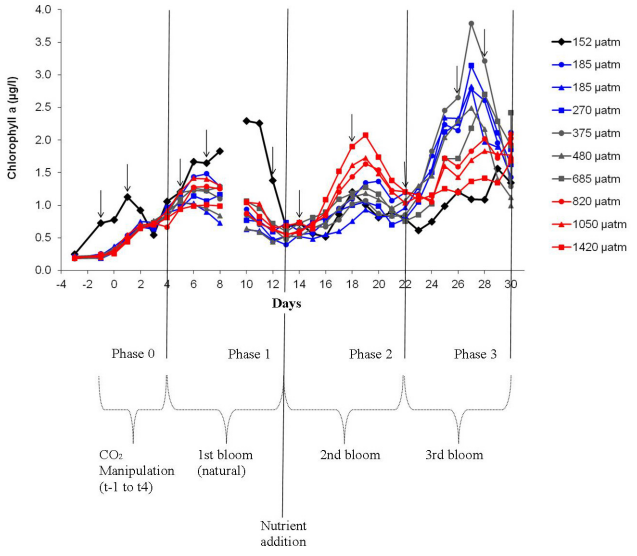


Fig. 1. Chlorophyll *a* ($\mu g\,L^{-1}$) concentration measurements plotted against days, where arrows mark time points analysed in the present study. Figure derived from Schulz et al. (2013).

in the cryovials to facilitate disruption and homogenization. The filters were crushed with RNase-free plastic pestles and lysozyme was directly added to the broken filter pieces while still in the cryovial. Both the RNA and DNA were isolated during the experiment. However, the RNA was kept for further purposes. DNA quality and quantity were assessed by micro-volume spectrophotometer nanodrop ND-1000 (PeqLab GmbH, Erlangen, Germany) measurements. All samples were kept at −80 °C until further analysis.

Polymerase chain reaction (PCR) and sequencing were performed following the Illumina HiSeq2000 and MiSeq V4-16S rRNA protocol (Caporaso et al., 2012). Briefly, the V4 region of the 16S rRNA gene was amplified with region-specific primers that included the Illumina paired-end flow-cell adapter sequences (Illumina Inc., CA, USA). The barcode was read using the custom index sequencing primer in an additional cycle (12 bp). Each sample was amplified in triplicate, and was pooled afterwards. Each 25 µL PCR reaction contained 12 µL of MoBio PCR Water (certified DNA-free), 10 µL of 5 Prime HotMasterMix, 1 µL of Forward Primer (5 µM initial concentration), 1 µL Golay Barcode Tagged Reverse Primer (5 µM initial concentration), and 1 µL of template DNA. The reactions were heated to 94 °C for 3 min for their initial denaturation, followed by 35 cycles in series of 94 °C for 45 s, 50 °C for 60 s, and 72 °C for 90 s. The amplicons were quantified using Quant-it™ Picogreen® (Invitrogen by Life Technologies™, CA, USA), and pooled in equal amounts (ng) into a 1.5 mL tube. Once pooled, the entire amplicon pool was cleaned up with the MO-BIO UltraClean® PCR Clean-Up Kit (MO-BIO Laboratories, Inc., CA, USA). Finally, the pooled samples were quantified using a Qubit® fluorometer (Invitrogen by Life Technologies™, CA, USA), and the molarity was estimated based on amplicon length. From this estimate, dilutions were made down to 2 µM and the standard Illumina sample preparation for sequencing was followed. Pooled amplicons were sequenced using custom sequencing primers, Read 1, Read 2, and Index. These sequencing primers were designed to be complementary to the V4 amplification primers to avoid sequencing of the primers. Amplicons were sequenced in a paired-end, 100 bp × 100 bp cycle run on the Illumina HiSeq2000, at a concentration of 4 pM with a 10 % PhiX spike. An entire control lane devoted to PhiX is also useful when sequencing low base diversity samples, like amplicons, and was included in the present analysis.

2.4 Sequence data analysis

All sequence analyses were performed using Quantitative Insights Into Microbial Ecology v. 1.5.0 (QIIME; Caporaso et al., 2010). QIIME defaults were used for quality filtering of raw Illumina data (including chimeras). Unique operational taxonomic units (OTUs) were picked against the Greengenes (McDonald et al., 2012) database and pre-clustered at 97 % identity; sequences that did not hit the reference collection were discarded. Representative sequences were aligned to the Greengenes core set with PyNAST (Caporaso et al., 2010). All sequences that failed to align were discarded. A phylogenetic tree was built from the alignment, and taxonomy was assigned to each sequence using the Ribosomal Database Project (RDP) classifier (Wang et al., 2007) retrained on Greengenes. Samples were rarefied to an even depth of 81 181 sequences and only the OTUs that appeared at least twice in the dataset were included in the further analyses; 106 singleton OTUs were not included in this analysis.

2.5 Statistical analysis

Multivariate analysis of microbial community structure was carried out in CANOCO 4.54 (ter Braak and Šmilauer, 2002), where the count of each OTU (97 % similarity) was used as a measure of abundance. All analyses had samples as scaling focus, and all species data were Hellinger-transformed using the program PrCoord 1.0 (Legendre and Gallagher, 2001; ter Braak and Šmilauer, 2002). Analysis of variance (ANOVA) followed by a Tukey test was done to test for significant differences between treatments (i.e. control vs. fjord, fjord vs. mesocosm, control vs. mesocosm) within each abundant phylum. Detrended correspondence analysis of the transformed OTU abundance data showed axis lengths < 3.0, suggesting a linear treatment of the data (Ramette, 2007). Redundancy analysis (RDA), with manual forward selection and Monte Carlo permutation tests (999 permutations), was used to evaluate effects of environmental variables (salinity, temperature, pH, chlorophyll *a*, etc.) on the microbial community composition. An indirect gradient analysis (PCoA)

was used to plot the distribution of samples in ordination space, with important environmental variables (as indicated by forward selection) overlaid as supplementary data. Microbial community composition differences were assessed by UniFrac (Lozupone and Knight, 2005) distance using QIIME (Caporaso et al., 2010).

In order to assess whether or not particular taxa were significantly influenced by pCO_2, a Bonferroni-corrected g-test was done using QIIME to remove significance due to chance. All analyses were considered to have a significant difference if $p < 0.05$ after Bonferroni correction.

Contour plots presenting mean abundance count plotted against pCO_2 and time (days) of the three most abundant genus of the OTUs significantly correlated to pCO_2 were created using Ocean data view (Bremen, Germany).

3 Results

The 256 sequenced samples generated ∼20 000 000 16S rRNA V4 reads (∼2 510 000 sequences per treatment); which clustered at 97 % sequence identity into 6821 OTUs.

3.1 Experimental timeline

Phytoplanktonic bloom evolution was identified using the daily measured chlorophyll *a* (chl *a*) concentration (µg L^{-1}) (Fig. 1). The chl *a* protocol and patterns are presented in Schulz et al. (2013). Briefly, all treatments (fjord included) underwent a natural bloom between $t0$ and $t11$, with highest chl *a* concentrations on $t6$. Subsequently, a second and third strong phytoplankton bloom happened only in the mesocosms following nutrient addition on $t13$. The second bloom had its highest chl *a* concentration on $t19$ and the third one, which varied greatly between mesocosms, reached its highest concentration on $t27$. These 3 blooms were represented as four general phases in phytoplankton chlorophyll phases defined by Schulz et al. (2013): phase 0 occurred from the start of the experiment on $t-4$ until adjustment of CO_2 was completed on $t4$; phase 1 started with the end of CO_2 addition on $t4$ until the nutrient additions on $t13$; phase 2 included the end of the first bloom on $t13$ to the end of the second bloom on $t22$; and phase 3 started from the end of the second bloom on $t22$ and lasted until the end of the experiment, on $t30$ (the chl *a* minimum of the third bloom was not recorded) (Fig. 1). Detailed fluctuations of chl *a*, nutrient concentrations, pH and pCO_2 are presented in this issue in Schulz et al. (2013) and Bellerby et al. (2013).

3.2 Community-structuring variables

The significant structuring variables for the total community during the post-nutrient addition period ($t13$–$t30$) of the experiment were (in order of explanatory importance) "fjord vs. mesocosm origin" (i.e. whether the sample was from water contained in a mesocosm or from the open fjord), sampling strategy (i.e. physical fractionation into small and large particle sizes), Si concentration, PO_4 concentration, mean primary production 14C-POC (PP), temperature (T), and pH (Fig. S1 and Table 1). The microbial community in the small size fraction (0.2–3 µm) from the fjord and all the analysed mesocosms was dominated by Proteobacteria (in order of abundance: Gamma (γ)- , Alpha (α)- and Beta (β)- proteobacteria) throughout the experiment. However, Proteobacteria began dropping in abundance gradually after $t7$, coincidentally with the increase in the abundance of Bacteroidetes (Fig. 2). In the large size fraction (3–12 µm) Bacteroidetes dominated consistently, while a fourth group comprised of the "Cyanobacteria and eukaryotic chloroplasts" (which included Chlorophyta, Haptophyceae, Rhodophyta and Stramenopiles) were also abundant (Fig. 2). The group classified as "others" in the small size fraction was composed predominately of Cyanobacteria at the beginning of the experiment, and of Actinobacteria towards the end (Fig. S2). In the large size fraction, the "others" group was extremely variable until $t7$. For example, at $t-1$ the fjord "others" group was dominated by the Verrucomicrobia while the mesocosms "others" groups was dominated by Actinobacteria; by $t5$ Firmicutes dominated in most mesocosms, while being almost absent from the fjord. At $t7$, the Actinobacteria was the dominant taxa in the "others" group in all treatments for the remainder of the experiment. At the end ($t28$), some Verrucomicrobia increased in the control, ∼270, and ∼685 µatm mesocosms (Fig. S2).

Once the community was analysed with regard to filter size fraction (small vs. large size fraction), the structuring community variables varied. The fjord had a significantly different assemblage from the mesocosms in the small and large size fraction before (origin 3 %–4 %) and after (origin 48 %–12 %) mesocosm nutrient addition (Table 2); however, the fjord and mesocosm communities were not significantly different until after $t5$. The microbial community in the fjord small size fraction was not significantly different from the mesocosms communities in the pre-nutrient addition phase and only the γ-proteobacterial abundance was significantly different ($p < 0.05$) between fjord and mesocosm in the post-nutrient addition phase. The fjord large size fraction microbial community was significantly different from the mesocosms during both the pre- and post-nutrient addition phases. In particular, the "Cyanobacteria and eukaryotic chloroplasts" group was significantly different between fjord and mesocosms pre- and post-nutrient addition; while the Bacteroidetes, α-proteobacteria and "others" were only significantly different post-nutrient addition (Fig. 3 and Table 3). Furthermore, the significant variables that correlated with community structure changes in the small size fraction were dimethyl sulphide (DMS-16 %), bacterial production (bp-15 %), density (d-12 %) for the pre-nutrient period ($t-4$ to $t12$), and origin (48 %), pCO_2 (10 %), day (10 %) for the post-nutrient period ($t13$–$t30$; Table 2). For the large size fraction, these variables were oxygen (O_2-7 %), DMS (7 %),

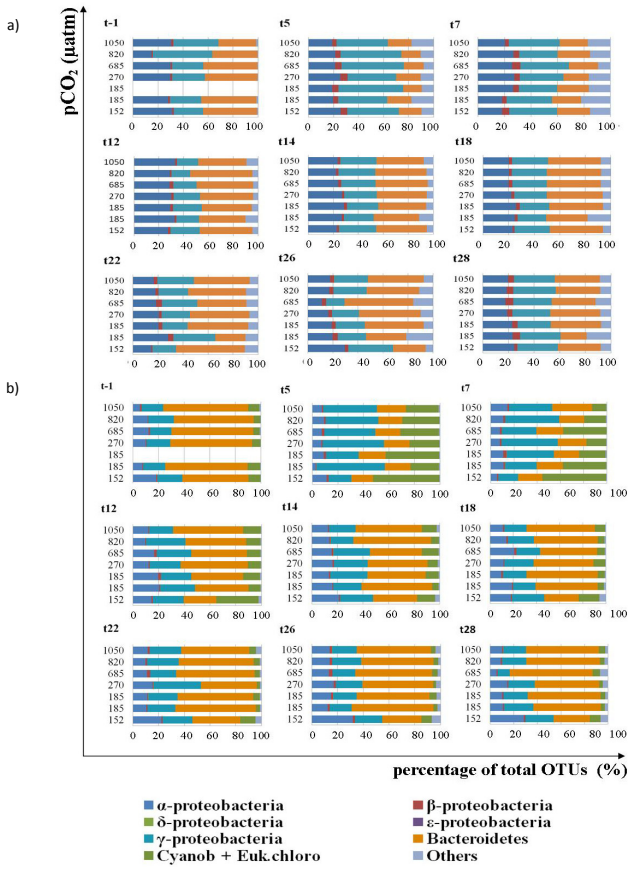


Fig. 2. Microbial community overview of the most abundant phyla in **(a)** the small (0.2–3 µm) and **(b)** the large (3–12 µm) size fraction during $t-1$, $t5$, $t7$, $t12$, $t14$, $t18$, $t22$, $t26$ and $t28$; x-axis represents percentage of total OTUs and y-axis represents pCO_2 in µatm

nitrate (NO_3-5 %) and origin (4 %) for the pre-nutrient period ($t-4$ to $t12$), and Si (27 %) and origin (12 %) for the post-nutrient addition period ($t13$–$t30$; Table 2). Therefore, the differences in the microbial community structure between the fjord and mesocosms were primarily due to the addition of nutrients to the mesocosms, and not to pCO_2 manipulation, as the control mesocosms were not significantly different from the elevated CO_2 mesocosms post-nutrient addition.

3.3 pCO_2 effect on microbial community

Although the pCO_2 treatment was not identified as a major community structuring variable, the relative abundances of 15 rare taxa (% abundance across time and treatment was <0.22%; Table 4) were significantly correlated to pCO_2 levels. From these 15 rare taxa in both small and the large size fractions, 12 showed a significant but slight increase with pCO_2, having their maximum abundances in either the medium (∼ 685 and ∼ 820 µatm) or the high (∼ 1050 µatm) pCO_2 mesocosms. The remaining three decreased, with their highest abundances in the lowest (∼ 185 µatm) pCO_2 mesocosm, or before manipulation started (Fig. 4, Figs. S3 and S4). The three most abundant of these 15 taxa were Methylotenera (β-proteobacteria), Colwellia (γ-proteobacteria) and Fluviicola (Bacteroidetes). Methylotenera and Colwellia abundances were at their highest in, respectively, the ∼ 686

Table 1. Redundancy analysis showing the significant structuring variables for the whole bacterial community during the post-nutrient addition period ($t13$–$t30$). Significant values are $p < 0.05$.

Variables	%	p	F
Origin	25	0.001	24.84
Fraction	14	0.001	17.77
Si	8	0.001	11.32
PO_4	2	0.01	2.83
Primary production	2	0.026	2.31
Temperature	2	0.042	2.26
pH	1	0.029	2.36

Table 2. Results from RDA forward selection (with Monte Carlo permutation tests) showing only the significant ($p < 0.05$) structuring variables for the small (0.2–3 µm) and the large (3–12 µm) size fraction during the pre-nutrient period from $t1$ to $t12$ (a and c, respectively) and post-nutrient period from $t13$ to $t30$ (b and d, respectively).

Variable		%	p	F
Small size fraction				
(a)	Dimethyl Sulphide	16	0.001	9.79
	Bacterial production	15	0.001	10.73
	Density	12	0.001	9.65
	NO_2	5	0.001	4.54
	Day	2	0.024	2.14
	Origin	3	0.014	2.53
(b)	Origin	48	0.001	35.75
	pCO_2	10	0.001	8.77
	Day	10	0.001	11.43
	CO_2	4	0.001	4
	Mesocosm	2	0.002	3.27
	Turbidity	3	0.001	3.91
	Primary production 14C	1	0.007	2.3
	NH_4	2	0.019	2.06
	Density	1	0.032	1.99
	Temperature	1	0.044	1.72
	PO_4	1	0.033	1.92
Large size fraction				
(c)	O_2	7	0.002	3.81
	Dimethyl sulphide	7	0.005	3.81
	Origin	4	0.016	2.59
	NO_3	5	0.014	2.95
(d)	Si	27	0.001	13.36
	Origin	12	0.001	7.11
	PO_4	4	0.039	2.24

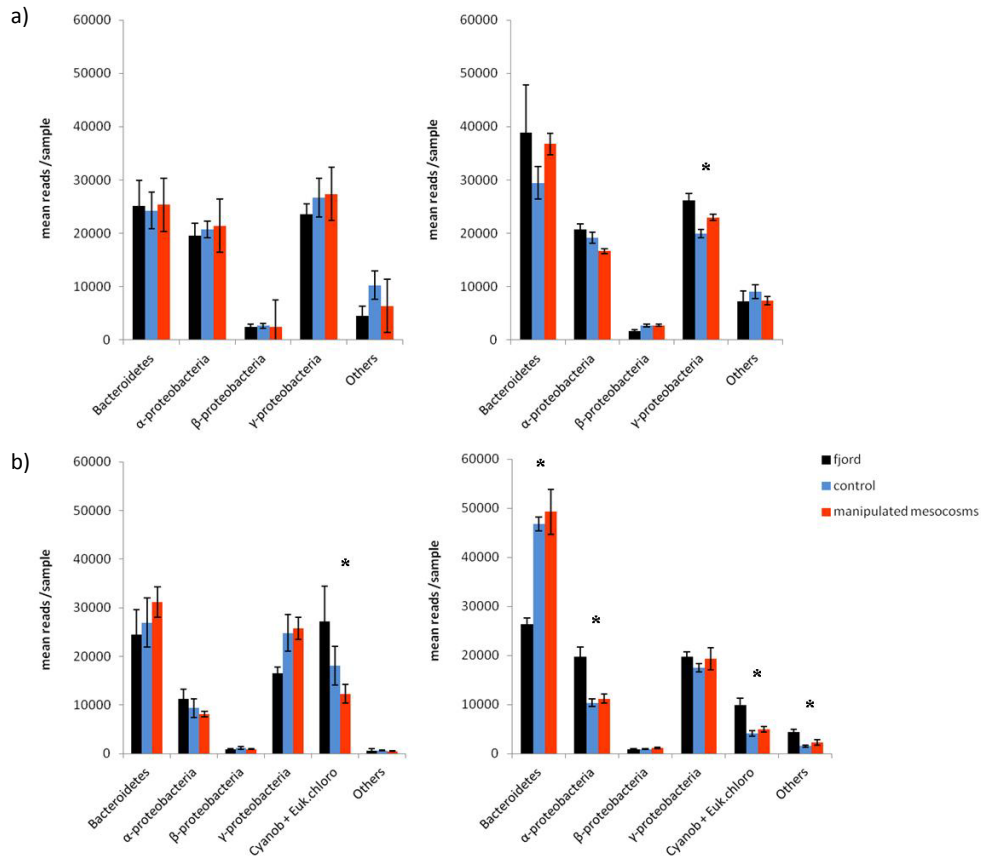


Fig. 3. Mean abundance (±SE) of the main phyla of the bacterial community for the fjord (∼ 145 µatm), the control (2 × ∼ 185 µatm) and the manipulated mesocosms (∼ 270, ∼ 685, ∼ 820, ∼ 1050 µatm) of the small **(a)** and large **(b)** size fraction pre- (left) and post- (right) nutrient addition. Phyla with significantly different p values (<0.05) as a function of samples origin are marked with an *.

and ∼ 824 µatm mesocosms toward the end of the experiment ($t22$). Fluviicola was present from the beginning of the experiment, but decreased precipitously after CO_2 was added and then recovered in abundance after $t10$, reaching its highest abundance in the 1050 µatm mesocosm between $t12$ and $t22$ (Fig. 4).

4 Discussion

4.1 Mesocosms and structuring effects

In this study, a large-scale mesocosm experiment was used to investigate the impacts of OA on the microbial community structure in a coastal, high latitude marine pelagic ecosystem. The experimental design provided the opportunity to test for the effects of four different pCO_2 concentrations (∼ 270, ∼ 685, ∼ 820, ∼ 1050 µatm) against two negative controls (∼ 185 µatm) over a six-week period. In addition, mesocosm-specific experimental artefacts were monitored by sampling the fjord microbial community throughout the course of the experiment. The microbial community structure post-nutrient-addition ($t13$) was significantly correlated with seven variables, the most influential of which was sample origin (fjord or mesocosm). The overall community structure was not significantly different between mesocosms (including control versus elevated pCO_2) over the course of the experiment. The significant effect of the mesocosm enclosures on microbial community structure could be due to the mesocosms themselves (isolating a microbial community from the surrounding fjord community) or since the effect was not significant before nutrient addition, more likely due to the addition of nutrients into the mesocosms at $t13$.

The sampling strategy separating the community into size fractions was the second most important variable in explaining differences in community structure. Before nutrient addition, the communities in the small size fraction were not significantly different between the fjord (ambient), control mesocosms, and the elevated pCO_2 mesocosms. However, after the addition of nutrients, γ-proteobacterial abundances were significantly different between fjord and mesocosms, and probably reflected the utilization of metabolites released by decaying phytoplankton in the post-bloom system. In particular, the overall abundance of Bacteroidetes in the small and large size fractions increased in post-blooms conditions,

Table 3. Analysis of variance (ANOVA) showing the relationship in between each treatment pre- and post-bloom condition for (a) small and (b) large size fraction bacteria of phyla with significant differences. Significant values are $p < 0.05$.

	Time	Phylum	Treatment	p
(a)	Post-nutrient addition	Gamma-proteobacteria	fjord-control	0.001
			mesocosm-control	0.140
			mesocosm-fjord	0.038
	Pre-nutrient addition	"Cyanobacteria + euk.chloro"	fjord-control	0.317
			mesocosm-control	0.289
			mesocosm-fjord	0.020
(b)	Post-nutrient addition	Bacteroidetes	fjord-control	0.001
			mesocosm-control	0.864
			mesocosm-fjord	0.002
		Alpha-proteobacteria	fjord-control	0.002
			mesocosm-control	0.787
			mesocosm-fjord	0.006
		"Cyanobacteria + euk.chloro"	fjord-control	0.000
			mesocosm-control	0.839
			mesocosm-fjord	0.001
		"Others"	fjord-control	0.000
			mesocosm-control	0.320
			mesocosm-fjord	0.001

Table 4. Bonferroni-corrected g-test of significance ($p < 0.05$) listing 15 taxa significantly correlated with CO_2, for both small and large size fraction; bold highlights mark the taxa presented in Fig. 4. Greengenes OTU identifiers refer to prokMSA ids in the Greengenes database.

Greengenes OTU Identifier	Taxa	Abundance	% total sequences (20 863 517)	Response to elevated pCO_2	p
114 612	Methylotenera (genus)	2907	0.014	Highest in middle pCO_2	0.000
144 699	Oceanospirillaceae (family)	1182	0.006	Increased with pCO_2	0.000
105 727	**Methylotenera (genus)**	**45 915**	**0.220**	**Highest in middle pCO_2**	**0.000**
151 803	Flavobacteriaceae (family)	1841	0.009	Increased with pCO_2	0.000
522 744	Leucothrix (genus)	130	0.001	Decreased with pCO_2	0.000
419 525	Sphingobacteriales (order)	171	0.001	Increased with pCO_2	0.000
94 238	Oxalobacteraceae (family)	322	0.002	Highest in middle pCO_2	0.000
402 252	**Fluviicola (genus)**	**20 950**	**0.100**	**Increased with pCO_2**	**0.001**
592 739	Oleibacter (genus)	2976	0.014	Highest in middle pCO_2/Increased	0.001
262 549	HTCC-1288 (genus)	25	0.001	Mixed, highest in high-middle pCO_2	0.001
140 859	*FlavobacteriumSuccinicans* (species)	344	0.0001	Decrease with pCO_2	0.004
235 556	**Colwellia (genus)**	**32 153**	**0.154**	**Highest in high-middle pCO_2**	**0.008**
591 187	Flavobacteria (class)	231	0.001	Decrease with pCO_2	0.010
243 032	Thioclava (genus)	59	0.0003	Mixed, highest in high pCO_2	0.011
554 148	SC3-41 (family)	571	0.003	Minimum increase	0.027

possibly also as a result of the dissolved organic carbon (DOC) released by a decaying algal bloom and aggregation of dying phytoplankton, respectively. The γ-proteobacteria and Bacteroidetes generally include many phytodetritus-assimilating organisms (Teske et al., 2011; Abell and Bowman, 2005; Pinhassi et al., 2004) and this would explain their increase in abundance during the demise of the bloom. Despite the observation that Bacteroidetes showed bloom-related dynamics, and contradictory to the findings of Zhang et al. (2012), no significant difference in the Bacteroidetes abundance (in either fraction) was found between the control and elevated pCO_2 mesocosms, suggesting that elevated pCO_2 did not impact the relative abundance of Bacteroidetes. However, their abundance in the fjord was significantly lower than in the mesocosms, suggesting that the nutrient addition or influence of the mesocosm enclosure did have an impact.

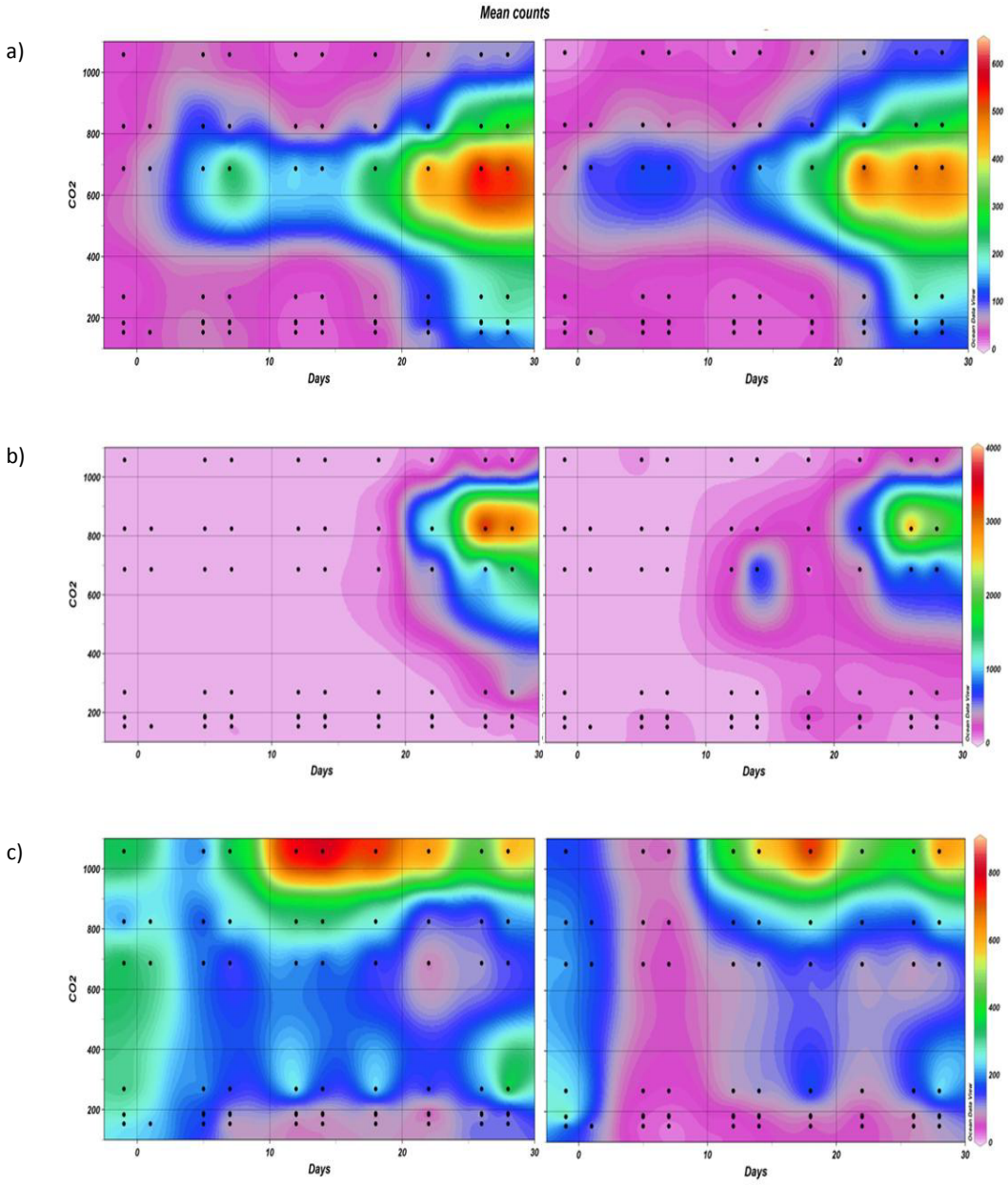


Fig. 4. Contour plots presenting the continuous interpolated mean abundance count of the three most abundant taxa that are significantly affected by pCO_2 levels **(a)** Methylotenera **(b)** Colwellia and **(c)** Fluviicola plotted against pCO_2 (µatm, y-axis) and time (days, x-axis). Left and right panel represent, respectively, the small (0.2–3 µm) and large (3–12 µm) size fraction.

The large size fraction in the mesocosms also showed differences in the relative abundance of dominant phyla following nutrient addition (t13). It has previously been established that particle-associated assemblages were predominantly connected to phytoplankton development (Riemann et al., 2000; Allgaier et al., 2008). Furthermore, differences in the "Cyanobacteria and eukaryotic chloroplasts" group were measurable before nutrient addition. However these differences appear to be related to the natural phytoplankton bloom (which occurred in the fjord and mesocosms) that reached its maximum on t7. The "post-nutrient addition" differences were significant between the fjord and mesocosms for almost every abundant phyla throughout the different phytoplankton phases; suggesting that nutrient addition influenced autotrophic and heterotrophic microbial community structure. However, no significant differences were found between the control and the elevated mesocosms, which suggests that high pCO_2 level was not an important community-structuring variable for the large size fraction in this experiment. Silica was the third most important structuring variable

and is potentially related to diatom abundance (de Kluijver et al., 2010). The recycling of Si from decaying diatoms, after a phytoplankton bloom, is carried out by a diverse fast growing bacteria related to cytophagales (from Flavobacteria; Riemann et al., 2000). Indeed, an increase in the abundance of Bacteroidetes, which contains the Flavobacteria, was observed in the post-nutrient addition phase.

However, no single environmental variable could account for the microbial community composition of the large and small size fractions for all of the phases of the mesocosm experiment (Fig. 1). Rather a shift was observed between pre- and post-nutrient addition with DMS concentration as the most influential variable for the small size fraction under pre-nutrient addition, while origin (Fjord vs. mesocosm) was most influential under post-nutrient addition conditions. Oxygen and Si were the most significant structuring variables for the large size fraction for the pre- and post-nutrient addition, respectively. Variables associated with phytoplankton bloom dynamics were most important for structuring the community, especially when looking at the taxonomic shifts between fjord, control mesocosms and elevated pCO_2 mesocosms. The differences were greater after $t13$ because of the two subsequent phytoplankton blooms that were triggered by the nutrient addition. The differences were most evident in the large size fraction, probably due to the association of the bacterial community with phytoplankton aggregates. Therefore, it is possible to state that nutrients, and the phytoplankton blooms, were the main drivers of microbial community structure in this experiment, which is in agreement with previous (Allgaier et al., 2008; de Kluijver et al., 2010) and present studies (Sperling et al., 2013).

4.2 Elevated pCO_2 effect

The effect of elevated pCO_2 on microbial community structure has also been investigated in previous (Newbold et al., 2012) or present mesocosms (Zhang et al., 2012), where no evidence of a major pCO_2 effect on the general bacterial community was found. However, other work suggests that only the community structure of the small size fraction bacteria is significantly affected by elevated pCO_2 (Allgaier et al., 2008). The extensive database of 16S rRNA sequence obtained in this study provided the high resolution necessary to study subtle but significant changes in community structure hinted at in prior studies. In agreement with Allgaier et al. (2008), the effect of elevated pCO_2 in this experiment was slight and only impacted the small size fraction bacteria after nutrient addition, which corresponded to post-nutrient addition and post-bloom conditions (after $t13$) in this study. This increased post-bloom CO_2 effect was previously observed in other mesocosms experiments (Arnosti et al., 2011; de Kluijver et al., 2010), confirming a possible increased CO_2 effect under nutrient (N, P, Si) limitation.

While pH was shown to be a weak driver of microbial community structure in our experiment, the direct impact of pCO_2 was found to be non-significant, except for 15 rare taxa, which did show a response to elevated CO_2. Therefore, the level of taxonomic resolution afforded by this study suggests that, in this ecosystem, rare organisms may be disproportionately affected by acidification. The most abundant of these 15 rare taxa was Methylotenera (genus) and had its highest mean abundance in the medium pCO_2 mesocosms (~ 685 µatm). Species from this genus are generally aerobic, ubiquitous bacteria found in a wide range of O_2, salinity, temperature and pH. Methylotenera can colonize multiple pH range (5 to 8.5) but grows optimally at pH 7.5 (Kalyuzhnaya et al., 2006; Bosch et al., 2009), suggesting that pH may strongly influence for distribution of this taxa. Indeed, the pH close to this value from $t5$ until the end of the experiment in the mesocosms with a pCO_2 over ~ 685 µatm. The highest abundance was found from $t22$ until $t28$ where the pH was 7.9 and 7.94. A lower pH was found (pHT 7.57–7.80) in the ~ 1050 µatm mesocosm but this was not accompanied by an increase in Methylonera abundance, potentially because the pCO_2 concentration itself was toxic to this species at this stage or this could represent mesocosm variability, suggesting a need for improved replication. Functionally, the species included in this genus have been described as bacteria that require organic compounds containing no carbon–carbon bonds (C_1 compounds) like methylamine and/or methanol as energy sources (Lidstrom, 2006; Kalyuzhnaya et al., 2006, 2010). These organic compounds play an important role in the global carbon cycle because they are greenhouse gases whose emissions are on a scale similar to methane (Chistoserdova et al., 2009). Further investigation of the behaviour of these C_1-compound-degraders in response to elevated CO_2 are, therefore, important for understanding biotic influences on climate dynamics. The second most abundant group of the 15 pCO_2-correlated rare taxa was Colwellia, which is a versatile group with broad temperature range tolerance. For example, the psychrophilic Arctic marine strain *Colwellia psychrerythraea* grows at a range of temperature from -1 to 10 °C (optimal growth 8 °C), *Colwellia chuckchiensis* at a range from 0 to 30 °C and *Colwellia asteriadis* spp. at a range from 4 to 25 °C. These organisms are also capable of colonising a wide range of pH from 4 to 10 (Yu et al., 2011; Choi et al., 2010; Methé et al., 2005). *C. psychrerythrea* is considered a model organism for psychrophiles and shows multiple molecular adaptations to the cold, like enzymes for cryoprotection, for dissolving high-molecular-weight organic compounds (ex. carbon), for stability in extreme environments (extracellular polymeric substances) and for cold-active processes (Methé et al., 2005; Huston et al., 2004). These features make Colwellia spp. key participants in carbon and nutrient cycling in the cold marine environments. Since some methanogenic enzymes were previously found in Colwellia spp. (Methé, et al., 2005) one can speculate that these compounds were found in greater abundance toward the end of the experiment. This would also support the presence of the Methylotenera, which increased

in abundance towards the end of the experiment. Finally the genus Fluviicola, the third most abundant OTU correlated with pCO_2, was dominant in the elevated CO_2 mesocosms ($\sim 1058\,\mu atm$). Interestingly, Fluviicola was present at the beginning of the experiment but decreased shortly after CO_2 treatment started. The abundance increased under elevated pCO_2, but stayed low in medium pCO_2 mesocosms and absent in the controls, for both size fractions. Little is known about this genus, making speculations about its ecological role difficult.

5 Conclusions

In summary, multiple parameters were found to significantly influence the structure of the bacterial community in Svalbard mesocosms. The most influential factors were the origin of the sample (fjord or mesocosms) and nutrient addition. Furthermore, the relative importance of sampling strategy (small versus large size fraction), Si, PO_4, primary production, temperature, and pH in structuring the community depended greatly on the time at which the community was sampled in relation to the phytoplankton blooms. The direct impact of pCO_2 was found to be significant for only 15 rare taxa and should be further investigated as analysis of low abundance community members is known to be problematic in 16S surveys (Bokulich et al., 2013). If confirmed, this limited pCO_2 effect could have evolutionary consequences creating a shift in the taxa dominance and/or diversity, profoundly affecting the structure of entire community in a high CO_2 world. However, it should be noted that the pCO_2 conditions in which these organisms dominated were super-elevated compared to predicted outcomes for the surface ocean under current climate change scenarios. Furthermore, the evolutionary response of the unicellular eukaryote Emiliania huxleyi to elevated CO_2 was studied by Lohbeck et al. (2012) and showed that only 500 asexual-generations were necessary to permit evolution either via adaptive changes from diverse genotype selection or via new mutations. It would be interesting to investigate how the bacterial communities from the present mesocosms experiment would evolve faced to extended elevated CO_2 exposure, allowing a longer population growth.

Future work should focus on exploring the functional responses of the community (metagenomics/metatranscriptomics) to evaluate how elevated pCO_2 or OA influence these processes over a longer time period.

Supplementary material related to this article is available online at: http://www.biogeosciences.net/10/555/2013/bg-10-555-2013-supplement.pdf.

Acknowledgements. This work is a contribution to the "European Project on Ocean Acidification" (EPOCA), which received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 211384. We gratefully acknowledge the logistical support of Greenpeace International for its assistance with the transport of the mesocosm facility from Kiel to Ny-Ålesund and back to Kiel. We also thank the captains and crews of M/V *ESPERANZA* of Greenpeace and R/V *Viking Explorer* of the University Centre in Svalbard (UNIS) for assistance during mesocosm transport and during deployment and recovery in Kongsfjorden. We thank the staff of the French-German Arctic Research Base at Ny-Ålesund, in particular Marcus Schumacher, for on-site logistical support. We would like to thank Mesoaqua for supporting J. I. N. in the field trip time to Svalbard on grant METAEPOCA. We also want to thank Tina Baustian for her help during the RNA/DNA extraction. Finally, the funding for S. M. G. was provided by NIH Training Grant 5T-32EB-009412.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: T. F. Thingstad

References

Abell, G. C. J. and Bowman, J. P.: Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean, FEMS Microbiol. Ecol., 51, 265–277, 2005.

Agogué, H., Lamy, D., Neal, P. R., Sogin, M. L., and Herndl, G. J.: Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing, Mol. Ecol., 20, 258–274, 2011.

Allgaier, M., Riebesell, U., Vogt, M., Thyrhaug, R., and Grossart, H.-P.: Coupling of heterotrophic bacteria to phytoplankton bloom development at different pCO_2 levels: a mesocosm study, Biogeosciences, 5, 1007–1022, doi:10.5194/bg-5-1007-2008, 2008.

Arnosti, C., Grossart, H. P., Mühling, M., Joint, I., and Passow, U.: Dynamics of extracellular enzyme activities in seawater under changed atmospheric pCO_2: a mesocosm investigation, Aquat. Microb. Ecol., 64, 285–298, 2011.

Bellerby, R. G. J., Silyakova, A., Nondal, G., Slagstad, D., Czerny, J., De Lange, T., and Ludwig, A.: Marine carbonate system evolution during the EPOCA Arctic pelagic ecosystem experiment in the context of simulated future Arctic ocean acidification, Biogeosciences, in review, 2013.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., and Caporaso, J. G.: Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing, Nat. Methods, 10, 57–59, doi:10.1038/nmeth.2276, 2013.

Bosch, G., Wang, T., Latypova, E., Kalyuzhnaya, M. G., Hackett, M., and Chistoserdova, L.: Insights into the physiology of Methylotenera mobilis as revealed by metagenome-based shotgun proteomic analysis, Microbiology, 155, 1103–1110, 2009.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights,

D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D, Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R.: QIIME allows analysis of high-throughput community sequencing data, Nat. Methods, 7, 335–336, 2010.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., and Knight, R.: Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, ISME J., 6, 1621–1624, doi:10.1038/ismej.2012.8, 2012.

Chistoserdova, L., Kalyuzhnaya, M. G., and Lapidus, A.: The Expanding World of Methylotrophic Metabolism, Annu. Rev. Microbiol., 63, 477–499, doi:10.1146/annurev.micro.091208.073600, 2009.

Choi, E. J., Kwon, H. C., Koh, H. Y., Kim, Y. S., and Yang, H. O.: Colwellia asteriadis sp. nov., a marine bacterium isolated from the starfish Asterias amurensis, Int. J. Syst. Evol. Micr., 60, 1952–1957, doi:10.1099/ijs.0.016055-0, 2010.

Czerny, J., Schulz, K. G., Boxhammer, T., Bellerby, R. G. J., Büdenbender, J., Engel, A., Krug, S. A., Ludwig, A., Nachtigall, K., Nondal, G., Niehoff, B., Siljakova, A., and Riebesell, U.: Element budgets in an Arctic mesocosm CO_2 perturbation study, Biogeosciences, in review, 2013a.

Czerny, J., Schulz, K. G., Ludwig, A., and Riebesell, U.: A simple method for gas exchange measurements in mesocosms and its application for carbon budgeting, Biogeosciences, accepted, 2013b.

de Kluijver, A., Soetaert, K., Schulz, K. G., Riebesell, U., Bellerby, R. G. J., and Middelburg, J. J.: Phytoplankton-bacteria coupling under elevated CO_2 levels: a stable isotope labelling study, Biogeosciences, 7, 3783–3797, doi:10.5194/bg-7-3783-2010, 2010.

Delille, B., Harlay, J., Zondervan, I., Jacquet, S., Chou, L., Wollast, R., Bellerby, R. G. J., Frankignoulle, M., Borges, A. V., Riebesell, U., and Gattuso, J. P.: Response of primary production and calcification to changes of pCO_2 during experimental blooms of the coccolithophorid *Emiliania huxleyi*, Global Biogeochem. Cy., 19, GB2023, doi:10.1029/2004GB002318, 2005.

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

Fu, F.-X., Mulholland, M. R., Garcia, N. S., Beck, A., Bernhardt, P. W., Warner, M. E., Sañudo-Wilhelmy, S. A., and Hutchins, D. A.: Interactions between changing pCO_2, N_2 fixation, and Fe limitation in the marine unicellular cyanobacterium Crocosphaera, Limnol. Oceanogr., 53, 2472–2484, 2008.

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver A., Smyth, T., Somerfield, P. J., Huse, S., and Joint, I.: The seasonal structure of microbial communities in the Western English Channel, Environ. Microbiol., 11, 3132–3139, 2009.

Grossart, H.-P., Allgaier, M., Passow, U., and Riebesell, U.: Testing the effect of CO_2 concentration on dynamics of marine heterotrophic bacterioplankton, Limnol. Oceanogr., 51, 1–11, 2006.

Hubert, J. A., Welch, D. M., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A., and Sogin, M. L.: Microbial population structures in the deep marine biosphere, Science, 318, 97–100, doi:10.1126/science.1146689, 2007.

Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A., and Sogin, M. L.: Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing, PLoS Genet, 4, e1000255, doi:10.1371/journal.pgen.1000255, 2008.

Huston, A. L., Methé, B., and Deming J. W.: Purification, Characterization, and Sequencing of an Extracellular Cold-Active Aminopeptidase Produced by Marine Psychrophile Colwellia psychrerythraea Strain 34H, Appl. Environ. Microbiol., 70, 3321–3328, doi:10.1128/AEM.70.6.3321-3328.2004, 2004.

Hutchins, D. A., Fu, F.-X., Zhang, Y., Warner, M. E., Feng, Y., Portune, K., Bernhardt, P. W., and Mulholland, M. R.: CO_2 control of Trichodesmium N_2 fixation, photosynthesis, growth rates, and elemental ratios: Implications for past, present, and future ocean biogeochemistry, Limnol. Oceanogr., 52, 1293–1304, 2007.

Kalyuzhnaya, M. G., Bowerman, S., Lara, J. C., Lidstrom, M. E., and Chistoserdova, L.: Methylotenera mobilis gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family Methylophilaceae, Int. J. Syst. Evol. Microbiol., 56, 2819–2823, 2006.

Kalyuzhnaya, M. G., Beck, D. A. C., Suciu, D., Pozhitkov, A., Lidstrom, M. E. and Chistoserdova, L.: Functioning in situ: gene expression in Methylotenera mobilis in its native environment as assessed through transcriptomics, ISME J., 4, 388–398, 2010.

Legendre, P. and Gallagher, E. D.: Ecologically meaningful transformations for ordination of species data, Oecologia, 129, 271–280, 2001.

Lidstrom, M. E.: Aerobic methylotrophic prokaryotes, in: The Prokaryotes, edited by: Balows, A., Truper, H. G., Dworkin, M., Harder, W., and Schleifer, K.-H., Springer, New York, 618–634, 2006.

Lohbeck, K. T., Riebesell, U., and Reusch, T. B. H.: Adaptive evolution of a key phytoplankton species to ocean acidification, Nat. Geosci., 5, 346–351, doi:10.1038/NGEO1441, 2012.

Lozupone, C. and Knight, R.: UniFrac: a new phylogenetic method for comparing microbial communities, Appl. Environ. Microbiol., 71, 8228–8235, 2005.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-C., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., He Ho, C., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M.: Genome sequencing in microfabricated high-density picolitre reactors, Nature, 437, 376–380, 2005.

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R., and Hugenholtz, P.: An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea, ISME J., 6, 610–618, 2012.

Methé, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X., Moult, J., Madupu, R., Nelson, W. C., Dodson, R. J., Brinkac, L. M., Daugherty, S. C., Durkin, A. S., Deboy, R. T., Kolonay, J. F., Sullivan, S. A., Zhou, L., Davidsen, T. M., Wu, M.,

Huston, A. L., Lewis, M., Weaver, B., Weidman, J. F., Khouri, H., Utterback, T. R., Feldblyum, T. V., and Fraser, C. M.: The psychrophilic lifestyle as revealed by the genome sequence of Colwellia psychrerythraea 34H through genomic and proteomic analyses, P. Natl. Acad. Sci., 102, 10913–10918, 2005.

Newbold, L. K., Oliver, A. E., Booth, T., Tiwari, B., DeSantis, T., Maguire, M., Andersen, G., van der Gast, C. J., and Whiteley, A. S.: The response of picoplankton to ocean acidification, Environ. Microbiol., 14, 2293–2307, doi:10.1111/j.1462-2920.2012.02762.x, 2012.

Pinhassi, J., Sala, M. M., Havskum, H., Peters, F., Guadayol, O., Malits, A., and Marrase, C.: Changes in bacterioplankton composition under different phytoplankton regimens, Appl. Environ. Microb., 70, 6753–6766, 2004.

Ramette, A.: Multivariate analyses in microbial ecology, FEMS Microbiol. Ecol., 62, 142–160, 2007.

Riebesell, U., Bellerby, R. G. J., Grossart, H.-P., and Thingstad, F.: Mesocosm CO_2 perturbation studies: from organism to community level, Biogeosciences, 5, 1157–1164, doi:10.5194/bg-5-1157-2008, 2008.

Riebesell, U., Czerny, J., von Bröckel, K., Boxhammer, T., Büdenbender, J., Deckelnick, M., Fisher, M., Hoffmann, D., Krug, S.A., Lenz, U., Ludwig, A., and Schulz, K. G.: Technical Note: A mobile sea-going mesocosm system: New opportunities for ocean change research, Biogeosciences, in review, 2013.

Riemann, L., Steward, G. F., and Azam, F.: Dynamics of Bacterial Community Composition and Activity during a Mesocosm Diatom Bloom, Appl. Environ. Microb., 66, 578–587, 2000.

Schulz, K. G., Bellerby, R. G. J., Brussaard, C. P. D., Büdenbender, J., Czerny, J., Fischer, M., Koch-Klavsen, S., Krug, S., Lischka, S., Ludwig, A. M. M., Nondal, G., Silyakova, A., Stuhr, A., and Riebesell, U.: Temporal biomass dynamics of an Arctic plankton bloom in response to increasing levels of atmospheric carbon dioxide, Biogeosciences, 10, 161–180, doi:10.5194/bg-10-161-2013, 2013.

Sogin, M .L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J.: Microbial diversity in the deep sea and the underexplored "rare biosphere", P. Natl. Acad. Sci., 103, 12115–12120, doi:10.1073/pnas.0605127103, 2006.

Sperling, M., Piontek, J., Gerdts, G., Wichels, A., Schunck, H., Roy, A.-S., La Roche, J., Gilbert, J., Nissimov, J. I., Bittner, L., Romac, S., Riebesell, U., and Engel, A.: Effect of elevated CO_2 on the dynamics of particle-attached and free-living bacterioplankton communities in an Arctic fjord, Biogeosciences, 10, 181–191, doi:10.5194/bg-10-181-2013, 2013.

Tanaka, T., Thingstad, T. F., Lovdal, T., Grossart, H.-P., Larsen, A., Allgaier, M., Meyeröfer, M., Schulz, K. G., Wohlers, J., Zöllner, E., and Riebesell, U.: Availability of phosphate for phytoplankton and bacteria and of glucose for bacteria at different pCO_2 levels in a mesocosm study, Biogeosciences, 5, 669–678, doi:10.5194/bg-5-669-2008, 2008.

ter Braak, C. J. F. and Šmilauer, P.: CANOCO Reference Manual and Cano-Draw for Windows User's Guide, in: Software for Canonical Community Ordination Version 4.5 Microcomputer Power, Centrum voor biometrie, Plant research international Wageningen UR, Ithaca, NY, USA, p. 500, 2002.

Teske, A., Durbin, A., Ziervogel, K., Cox, C., and Arnosti, C.: Microbial community composition and function in permanently cold seawater and sediment from an Arctic fjord of Svalbard, Appl. Environ. Microbiol., 77, 2008–2018, doi:10.1128/AEM.01507-10, 2011.

Wang Q., Garrity, G. M., Tiedje, J. M. and Cole, J. R.: Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Appl. Environ. Microbiol., 2007, 5261–5267, 2007.

Worden, A. Z., and Not, F.: Ecology and diversity of picoeukaryotes, in: Microbial Ecology of the Oceans, 2nd Edn., edited by: Kirchman, D. L., Hoboken, NJ, USA, John Wiley & Sons, 159–205, 2008.

Yu, Y., Li, H.-R., and Zeng, Y.-X.: Colwellia chukchiensis sp. nov., a psychrotolerant bacterium isolated from the Arctic Ocean, Int. J. Syst. Evol. Micr., 61, 850–853, 2011.

Zhang, R., Xia, X., Lau, S. C. K., Motegi, C., Weinbauer, M. G., and Jiao, N.: Response of bacterioplankton community structure to an artificial gradient of pCO_2 in the Arctic Ocean, Biogeosciences Discuss., 9, 10645–10668, doi:10.5194/bgd-9-10645-2012, 2012.

4.2.5 Publication VI.

Ocean acidification effects on the community and gene expression of Arctic bacteria and phytoplankton during a coastal pelagic mesocosms event.

Roy *et al.*,
in prep. to be submitted to The ISME Journal

Ocean acidification effects on the community and gene expression of Arctic bacteria and phytoplankton during a coastal pelagic mesocosms event.

A.-S. Roy[1], Harald Schunck[1], S.M. Gibbons[2,3], D. Desai[1*], S. Owens[2,4], U. Riebesell[1], J.A. Gilbert[2,5]and J. LaRoche[1,*].

[1]GEOMAR Helmholtz Center for Ocean Research Kiel, Düsternbrooker Weg 20, 24105, Kiel, Germany.

[2]Argonne National Laboratory, Institute for Genomic and Systems Biology, 9700, S. Cass Ave. Lemont, IL 60439, USA

[3]Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637, USA

[4]Computation Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

[5]Department of Ecology and Evolution, University of Chicago, 5801 South Ellis Avenue, Chicago, IL 60637, USA

*Present address: Department of Biology, Dalhousie University, Halifax, N.S. Canada

Subject Category: Integrated genomics and post-genomics approaches in microbial ecology

Keywords: metagenome, metatranscriptome, mesocosms, Arctic, ocean acidification, CO_2

Abstract

The rise of atmospheric CO_2 results in a decrease in surface ocean pH, which has a more pronounced effect in polar regions. Next-generation metagenomic and metatranscriptomic sequencing was used to evaluate the effects of such anticipated pH changes on the Arctic microbial community structure and function as part of the 2010 EPOCA Svalbard mesocosm study. Samples were collected at 5 time points, covering all major phytoplanktonic events in two mesocosms under different pCO_2 treatments (≈185 and ≈1050 µatm). A total of ≈14,000,000 DNA and ≈18,000,000 RNA (cDNA) sequences were analyzed, which were eventually classified into 10,866 and 11,595 OTUs (metagenome and metatranscriptome, respectively) and 41,414 and 25,125 distinct metabolic functions (metagenome and metatranscriptome, respectively). Phylogenetic analysis revealed that the overall microbial community structure varied significantly with time, but was not influenced by pCO_2 treatments. Only a small number of species and functional genes were significantly affected. A detailed analysis was performed on two functions significantly impacted by pCO_2 (DMSP breakdown and acid resistance) and another function essential for photosynthesis (RuBisCO). Transcript abundances for DMSP breakdown and acid resistance enzymes increased with time and large differences in the abundances of these same transcripts between pCO_2 treatments became apparent at the later time points. Acid resistance had the highest abundance on d26 while RuBisCO had varying transcript abundances with the highest abundances on d16. Our results demonstrate that the effects of OA on Arctic microbial community structure and metabolisms are restricted to low-abundance species and specialised functional genes.

Introduction

Since the beginning of the industrial era, humans have used vast amounts of fossil resources, resulting in significant carbon dioxide (CO_2) release into the atmosphere. Oceans, rivers and lakes absorbed about 30-40% of anthropogenic CO_2, but recently, the oceans have not been able to compensate, creating an imbalance between the atmosphere and the ocean (Sabine *et al.*, 2004; Orr *et al.*, 2005; Feely *et al.*, 2009; Krause *et al.*, 2012). The absorbed atmospheric CO_2 reacts with water molecules creating bicarbonate (HCO_3^-) and hydrogen (H^+) ions. This process is predicted to drive down the ocean surface pH by 0.3-0.4 units before the end of the century; thus acidifying the ocean (Zeebe and Wolf-Gladrow, 2001; Orr *et al.*, 2005; Newbold *et al.*, 2012). Ocean acidification (OA) has the potential to influence the physiology of marine microbes (bacteria and phytoplankton), which would in turn impact ocean biogeochemistry (Falkowski *et al.*, 2008; Worden and Not, 2008).

Until recently, researchers have focused their interests in characterizing the effect of OA on calcifying organisms and microbes in laboratory experiments (Shi *et al.*, 2010; Hopkinson *et al.*, 2011; Lohbeck *et al.*, 2012; Hopkinson *et al.*, 2013). For example, Fu and colleagues (Fu *et al.*, 2008) found a significant increase in photosynthetic capacity of cyanobacterium *Crocosphaera* with increasing levels of CO_2 and Hutchins *et al.* (2007) discovered a positive correlation between pressure of carbon dioxide (pCO_2) and N_2 fixation rates in *Trichodesmium* species. More recently, several mesocosm experiments have studied the effect of CO_2 enrichment on whole plankton communities giving a better representation of natural conditions. The partial pCO_2 can be manipulated in these large mesocosm enclosures, placed in natural environments and seeded with natural communities. Such experiments allow researchers to evaluate how organisms react to OA in a natural setting. Some mesocosms studies have demonstrated species-specific physiological responses to elevated dissolved CO_2 concentrations. For example, Delille (Delille *et al.*, 2005) found a delayed or decreased calcification of the calcareous plates of *Coccolithophyceae* and Newbold (Newbold *et al.*, 2012) found a decreased abundance of picoeukaryotes at high dissolved CO_2. Despite this prior work, the response of bacterial communities to elevated pCO_2 concentrations is poorly characterized. The initial studies investigating the effects of OA were published less than 10 years ago, and showed significant

increases in bacterial protein production (Grossart *et al.*, 2006) but no significant changes in microbial community structure (Allgaier *et al.*, 2008; Tanaka *et al.*, 2008; Newbold *et al.*, 2012). One example of a major mesocosm experiment is the 2008 PeECE study, which showed no significant impact of elevated pCO_2 on bacterial abundance, diversity, or activity but found that the small-sized bacteria community structure was significantly altered by nutrient additions and the subsequent phytoplankton blooms (Allgaier *et al.*, 2008; Riebesell *et al.*, 2008; Arnosti *et al.*, 2011). To deepen and clarify our knowledge on OA, the European Project on Ocean Acidification (EPOCA) designed and carried out a large mesocosms study off the coast of Svalbard, in an Arctic fjord. This region was chosen because Arctic ecosystems are particularly vulnerable to OA due to the relatively high CO_2 solubility and low carbonate saturation states of cold surface waters (Hoegh-Guldberg and Bruno, 2010). The campaign concluded that bacterial protein production in the post induced (after nutrient addition) bloom phases was enhanced by increased pCO_2 (Sperling *et al.*, 2013), as well as primary production, and suggested that efficient heterotrophic carbon utilisation had the potential to counteract enhanced autotrophic CO_2 fixation (Piontek *et al.*, 2013). It further established that elevated pCO_2 transiently influenced net community production and the stoichiometry of carbon and nutrient coupling of the plankton community in this high-latitude fjord (Tanaka *et al.*, 2013). Brussaard *et al.*, (2013) found that OA distinctly affected the composition and growth (shift to smaller organisms and reduced growth) of the Arctic picoeukaryotic photoautotroph and nanophytoplankton communities. Finally, the experiment showed that pCO_2 augmented dissolved organic carbon production, stimulating heterotrophic microorganisms, which as a consequence increased gross primary production and reduced net community production (Engel *et al.*, 2013). While extensive, these results cannot resolve many important functional questions about the effects of OA on pelagic microbes. Therefore, the present study employs metagenomic and metatranscriptomic sequencing to investigate the detailed structural and functional differences between microbial communities experiencing different levels of pCO_2.

Since the advent of next-geneneration sequencing techniques, multiple studies have been carried out ranging from single-cell cultures (Pommier *et al.*, 2007; Jeffries *et al.*, 2011; Yilmaz *et al.*, 2011; Thomas *et al.*, 2012) to diverse natural environments, such as deep sea vent systems (Xie *et al.*, 2011; Xie *et al.*, 2013) or terrestrial soils (Urich *et al.*, 2008; Keshri *et al.*, 2013) and demonstrated the power of this approach to characterize both the bacteria community structures

and their function. Arctic microbial community composition attracted much interest because of its vulnerability to climate changes and the possible loss of microbial biodiversity. Several studies identified variation between winter and summer communities as well as species and phylotypes found in this environment (Galand *et al.*, 2009; Lovejoy *et al.*, 2006; Lovejoy *et al.*, 2011). Two recent studies started to characterize the whole community structure using similar sequencing (16S rRNA sequencing and tag pyrosequencing) as our first community analysis of Svalbard. The researchers established that the Arctic surface water was dominated by α-proteobacteria, γ-proteobacteria, Flavobacteria and a few other taxa with each relative abundance varying in regards to sampling location (Teske *et al.*, 2011; Kirchman *et al.*, 2010). The first bacterial community analysis of the Svalbard mesocosms used 16S rRNA amplicon sequencing and showed the same general community structure with sample origin (fjord or mesocosms) and nutrient addition as the most influential factors for bacterial community composition. While overall bacterial community structure was not significantly impacted by pCO_2, there were 15 rare taxa that varied significantly in abundance at elevated pCO_2 levels (Roy *et al.*, 2013).

High-throughput 16S amplicon sequencing methods can accurately characterize microbial community structure, but the degree of functional information that can be gleaned from these studies is limited (Margulies *et al.*, 2005; Sogin *et al.*, 2006; Huber *et al.*, 2007; Huse *et al.*, 2008; Gilbert *et al.*, 2009; Agogue *et al.*, 2011). Thus, our initial study was complemented by shotgun metagenomic and metatranscriptomic sequencing in order to investigate the differences in functional potential and gene expression profiles that exist between communities under varying CO_2 levels. In this study, the variation of microbial assemblages and how the expression of their genes is influenced by pCO_2 was observed. Samples from one un-manipulated control mesocosm (*in situ* ≈185 μatm) and one manipulated mesocosm (≈1050 μatm) were characterized through time using the Illumina MiSeq platform. These analyses characterized the detailed response of the microbial community structure and functions to elevated pCO_2.

Materials and methods

Study area and carbonate system manipulation

A large mesocosms experiment was conducted in summer 2010 in the water of Kongsfjorden, Ny-Ålesund settlement, Svalbard, Norway (78°56.2' N; 11°53.6' E) under EPOCA. The CO_2 pressure was manipulated by addition of CO_2-saturated seawater in seven mesocosms (initial pCO_2 from ≈286 to ≈1420 µatm) and was not manipulated in two mesocosms, which served as controls (initial pCO_2 ≈185 µatm); these mesocosms contained about 45 m^3 of seawater and reached down to 15 m depth. The fjord in which the mesocosms were suspended (initial pCO_2 ≈145 µatm) was also sampled to account for the ambient characteristics of the environment. Nutrient (nitrate via $NaNO_3$, phosphate via $NaHPO_4$ and silicate via Na_2SiO_3) were added on day (d) 13 to all nine mesocosms in order to promote phytoplankton growth, thus creating two periods: pre- (d-1 to d12) and post- (d13 to d30) nutrient addition. To explore the effect of OA on multiple biological processes, diverse environmental parameters were measured in all ten treatments. Detailed information about the mesocosms experimental set-up, deployment, carbonate chemistry and nutrient additions are presented in the "Arctic ocean acidification: pelagic ecosystem and biogeochemical responses during a mesocosms study" special Biogeosciences issue (Czerny *et al.*, 2013a; Czerny *et al.*, 2013b; Riebesell *et al.*, 2013; Schulz *et al.*, 2013; Silyakova *et al.*, 2013).

Sampling, filtration and sample selection

Water was collected between 0 and 12 m water depth in each mesocosm and in the Fjord (≈10 L) using an integrated water sampler (Hydrobios, Kiel, Germany). The collected water was first pre-filtered on a 20 µm sieve, and sequentially filtered through a 12 µm, a 3 µm filters to isolate the particle-associated microbial fraction (large size fraction) and through a 0.2 µm filter to isolate the soluble microbial fraction (Durapore® 47 mm, Millipore). To avoid nucleic acid degradation, processing of the samples from filtration to flash-freezing in liquid nitrogen was performed within 30 min of the sampling event. The filtered samples were stored thereafter at -80ºC, until DNA/RNA extraction. From all collected samples, the ones from two mesocosms (mesocosms 3 for control ≈185 and mesocosms 5 for manipulated ≈1050 µatm) on d3, d7, d16, d20 and d26 were used for this meta-omics (metagenome and metatranscriptome) study. These samples were

selected based on chlorophyll *a* (Roy *et al.*, 2013; Schulz *et al.*, 2013) to assure that every stage of the mesocosms' phytoplankton blooms were analysed.

DNA/RNA extraction and cDNA synthesis

Total nucleic acids was extracted from the 3-12 μm filters using the 'Total RNA and DNA purification - NucleoSpin® RNA II RNA/DNA buffer' kits from Macherey-Nagel (Macherey-Nagel GmbH & Co. KG, Düren, Germany). A standard protocol with minor modifications was followed. Changes to the protocol included the cryo-vials containing filters being disrupted with RNase-free plastic pestles after immersion in liquid nitrogen and addition of lysozyme (damages bacterial cell walls) to facilitate homogenization of the samples and lysis of the cells. Both RNA and DNA were isolated during the experiment; DNA was directly used for whole genome shotgun sequencing on the Illumina MiSeq platform, while RNA was transcribed into cDNA using the Invitrogen superscript III cDNA synthesis kit with random hexameric primers (Qiagen®, Hilden, Germany). The RNA preparation was done following the detailed protocol from Schunck *et al.* (2013) and only shortly describe here. The synthesis was performed after first removing contaminating remainder of genomic DNA with the Turbo DNA-*free*™ kit from Ambion® (Invitrogen by Life Technologies™, CA, USA) and removing prokaryotic rRNAs with the mRNA-only™ Prokaryotic mRNA Isolation Kit (Epicentre®, WI, USA). Further depletion of bacterial rRNA was done using the Ambion® MicrobExpress™ kit (Invitrogen by Life Technologies™, CA, USA) and finally, the cleaned rRNA-depleted mRNA was amplified using Ambion® MessageAmp™ II aRNA amplification Kit (Invitrogen by Life Technologies™, CA, USA). DNA and RNA were quantified by micro-volume spectrophotometer nanodrop ND-1000 (PeqLab GmbH, Erlangen, Germany) measurements and quality-checked with the "Experion™" analysis kits (DNA and RNA) from Bio-Rad Laboratories (Hercules, CA, USA). Finally, all leftover reactants and reagents were removed using the PCR Mini Elute Kit (Qiagen®, Hilden, Germany) and samples were kept at -80 °C until further analysis.

Library generation and Sequencing technology

Libraries were prepared for the metagenomes from eight genomic DNA samples and for the metatranscriptomes from nine synthesized cDNA samples by following the manufacturer's protocol (Illumina Inc., CA, USA), using the Illumina TruSeq DNA Sample Preparation Kits. These libraries were then gel-extracted to obtain a uniform size distribution before sequencing.

All metagenome and metatranscriptome samples were sequenced on the Illumina MiSeq Platform (Illumina Inc., CA, USA) with six samples per 2 x 150bp MiSeq run.

Binning and storage of data

Sequences processing was carried out on the MG-RAST$_{3.2.5}$ (Meyer *et al.*, 2008) server, using their standard annotation workflow. In this study, M5NR annotations (MG-RAST's non-redundant protein database) were used.

Statistical analysis

Metagenome and metatranscriptome data were normalized to their respective sequence numbers; further statistical analysed and figures were conducted with the R software package (www.r-project.org; R Development Core Team, 2011) and the pathway mapping and visualization tool FROMP was used for visualization (Desai *et al.*, 2013).

Preliminary association of bacterial taxa to metabolic functions

A preliminary association of bacterial taxa with metabolic functions was done by MG-RAST$_{3.2.5}$ manual download of sequences of interests (acid resistance, DMSP breakdown, RuBisCO) using the workbench option. The obtained sequences were subjected to a BLASTX search with the NCBI non-redundant (NR) database. The top-hits (evalue < $1e^{-5}$) were parsed for the bacterial taxon and the complete taxonomic lineage for each hit was extracted from the NCBI Taxonomy database. The taxa associated with selected functions were then sorted according to the sample affiliations and the proportional count (%) for each taxon in each sample for the given function was represented as pie-charts using the R software package.

Results

The 17 sequenced samples resulted in a total of ≈14,000,000 metagenomic and ≈18,000,000 metatranscriptomic reads (total of 32,232,802) consisting of on average ≈2,000,000 sequences per sample (from 1,170,721 to 2,435,034); these sequences clustered into 10,866 and 11,595

genera and 41,413 and 25,125 metabolic functions for the metagenome and metatranscriptome, respectively. The two d7 metagenomes (≈185 and ≈1050 µatm) and the ≈185 µatm d20 metatranscriptome failed to sequence properly, which was probably due to the lower amount of nucleic acid in these samples.

Experimental timeline

Chlorophyll a (chl *a*) concentration (µg/L) was measured daily and was used to plot the phytoplankton bloom evolution described in Schulz *et al.* (2013). To summarize briefly, three blooms were observed throughout the experiment. The first bloom, produced by natural causes, happened from d0 to d11 reaching the highest chl *a* concentration on d6 in all treatments. The other two blooms were observed after nutrient addition on d13 with the second bloom's highest concentration on d19. The third bloom varied greatly between treatments but crested at d27, Information about nutrient concentrations, pH, pCO_2 and more detailed chl *a* patterns are presented in Schulz *et al.* (2013) and Silyakova *et al.* (2013).

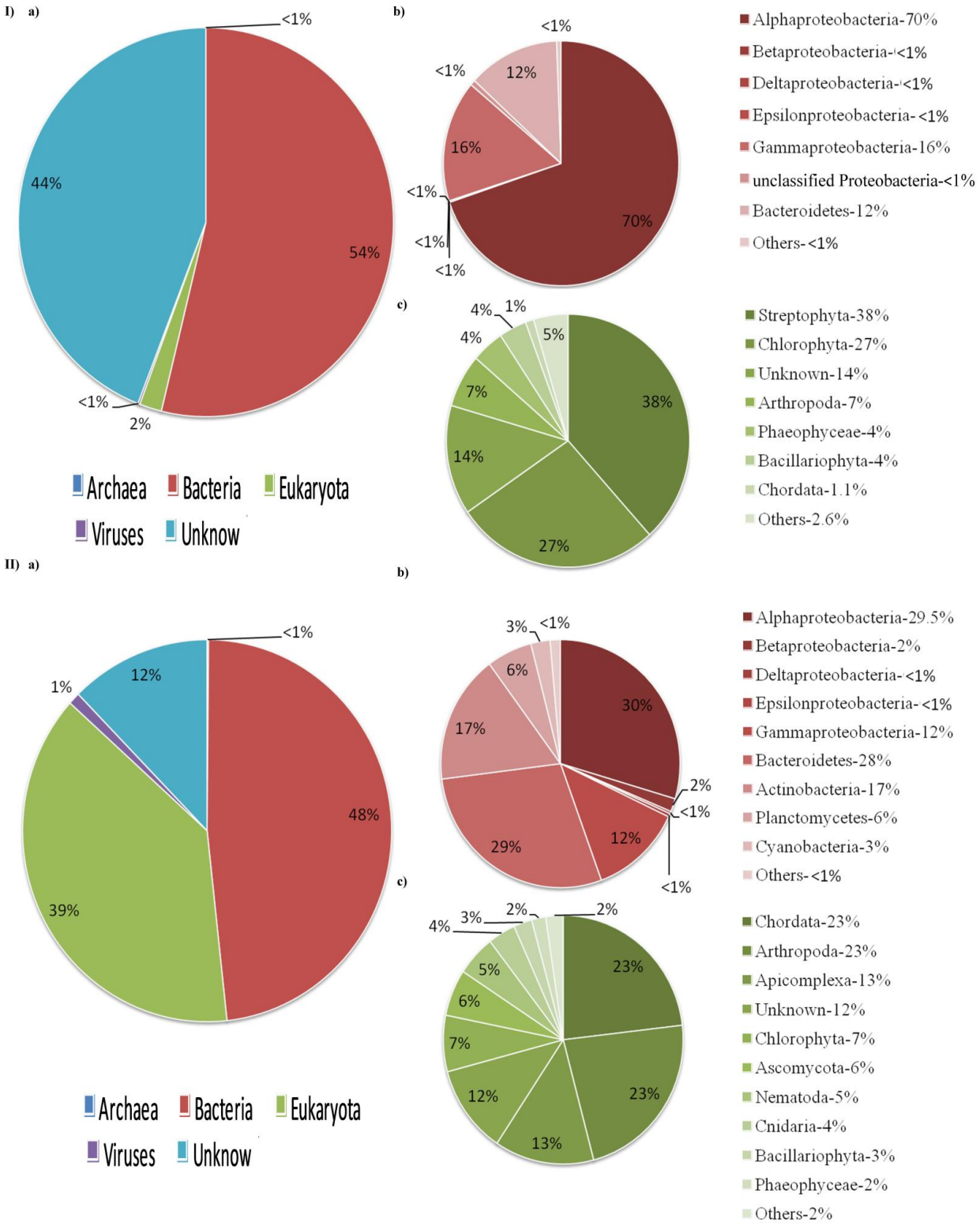


Figure 1 Overview of normalized relative abundance of taxa in the combined I) metagenomes and II) metatranscriptomes showing the a) kingdoms, b) bacterial phyla and c) eukaryotic phyla composition.

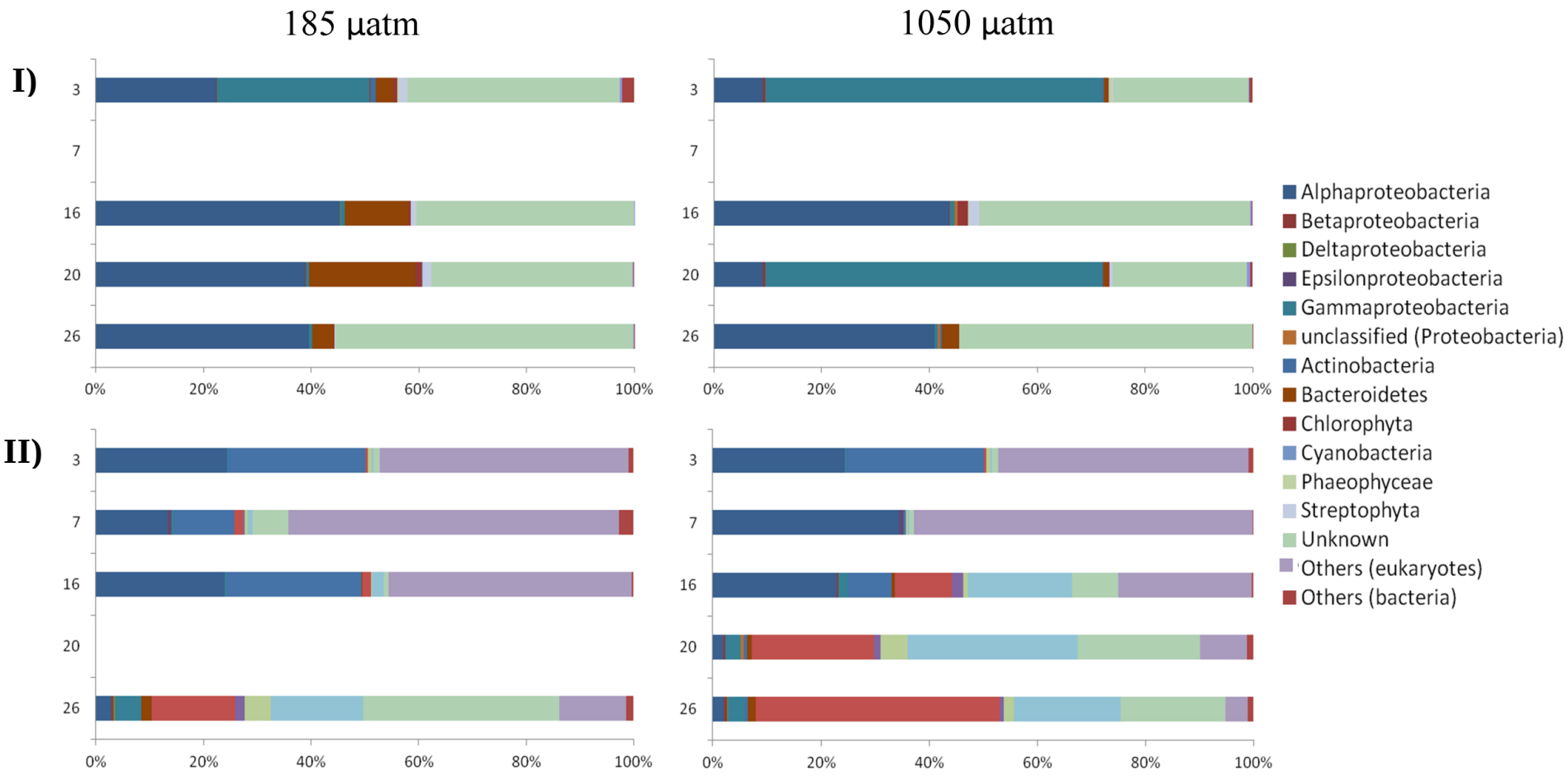


Figure 2 Normalized relative abundance overview of the 15 most abundant taxa from the combined I) metagenomes and II) metatranscriptomes grown under two different CO_2 pressures.

Community structure and function overview

The microbial community composition at kingdom, phylum, and class levels (Figure 1) for both DNA (metagenome) and cDNA (metatranscriptome) datasets shows differences across pCO_2 treatments. The majority of sequences in both libraries were annotated as bacteria - 54% (metagenome) and 48% (metatranscriptome); however the eukaryotic fraction was much larger in the metatranscriptomic than in the metagenomic database (39% vs. 2%); the eukaryotic fraction of the metatranscriptome (39%) corresponds, almost in totality, to the unknown sequences (44%) of the metagenome.

Archaea were represented in the metagenome and in the metatranscriptome by a small number of reads (<1%) without difference between the metagenome and metatranscriptome. Bacterial DNA sequences were mainly composed of Alpha (α)-proteobacteria (70%), followed by Gamma (γ)-proteobacteria (16%), Bacteroidetes (12%) and a small percentage (<1%) of "other" taxa. However, in the metatranscriptome assemblage, the α-proteobacteria (29.5%) and the Bacteroidetes (28%) were dominant, followed by the γ-proteobacteria (12%) and the Actinobacteria (17%). This assemblage also contained a significant number of Planctomycetes (6%) and Cyanobacteria (3%), thus the metatranscriptome was more diverse than the metagenome. The eukaryotic sequences from the metagenome were mainly assigned to Streptophyta (38%), Chlorophyta (27%) and other unknown taxa (14%), while the eukaryotic metatranscriptome contained mainly Chordata (23%), Arthropoda (23%), Apicomplexa (13%) and Chlorophyta (7%).

Analysing the sequence datasets separately (time vs. CO_2), the unknown sequences represented a large fraction of sequences in both databases with a much higher relative abundance in the metagenome (Figure 2). The upper part of the figure, I), shows the most abundant phyla or classes from the metagenomes, and taxa with different percentages between the CO_2 pressures (mostly at d3 and d20). Indeed, the unknown fraction was about 50% but varied throughout the experiment. Figure 2 clearly shows variation in α-, γ-proteobacteria, and Bacteroidetes (only in low CO_2) abundances through time, as well as the disappearance of the other taxa. The 1050 µatm mesocosm metagenome was alternatively dominated by the α- and γ-proteobacteria while the 185 µatm mesocosm was, after nutrient addition, dominated by the α-proteobacteria and the Bacteroidetes. Multiple taxa in the lower section metatranscriptome, Figure 2 II), were seen after nutrient addition at t13, where Chlorophyta and Streptophyta displaced γ-proteobacteria.

The metatranscriptome presented more taxa with elevated relative abundance and a very high percentage of "others" eukaryotes (Apicomplexa, Arthropoda, Ascomycota, Chordata, Cnidaria, Echinoderma, Euglenida, *Eustigmatophyceae*, Mollusca, Porifera, *Pheaophyceae*, *Xanthophyceae*) and "others" bacteria (Bacillariophyta and Firmicutes). These "others" included the more complex multi-cellular organisms like Arthropoda and Chordata, which were not essential to qualify the effect of CO_2 on bacteria or phytoplankton, and were not included in the analysis. α-proteobacteria and Actinobacteria were dominant, alternatively, in both low and high CO_2 treatments from the beginning of the experiment until d20, when Chlorophyta and Streptophyta appeared and dominated throughout the rest of experiment. The Epsilon (ε)-proteobacteria disappeared in both mesocosms after nutrient addition, and a clear separation between the two pCO_2 treatments was not observed.

Table 1 Analysis of similarity (ANOSIM) showing beta-diversity differences between treatments groups ($p < 0.05$ marked in bold; community structure calculated based on abundance of reads grouped into phylogenetic classes).

Database	**Treatment rarefication depth**	**R statistic**	**p-value**
Metatranscriptomic	CO_2_2360_16S	-0.2062	0.993
	CO_2_6300	-0.2625	1.000
	CO_2_17100	-0.2687	1.000
	day_2360_16S	0.7969	**0.004**
	day_6300	0.7188	**0.006**
	day_17100	0.7500	**0.003**
Metagenomic	CO_2_17100	-0.1354	0.921
	CO_2_18960_16S	-0.1458	0.878
	CO_2_256200	-0.1354	0.832
	day_17100	0.5625	**0.009**
	day_18960_16S	0.6875	**0.008**
	day_256200	0.625	**0.010**

An analysis of similarity (ANOSIM, Table 1) was performed to identify the taxa (to the class level) with significant differences between the CO_2 treatments. The ANOSIM showed that the CO_2 effect on the combined samples (excluding time) was not significant whereas the effect of time (excluding CO_2) was. In conclusion, CO_2 was not a significant driver of general community assemblage in either of the metagenome or metatranscriptome; however it is important to test for the significance of these variables on individual taxa. A statistical analysis (ANOVA) indentified 414 and 7 species in the metagenome and metatranscriptome (Table S1 and S2), respectively, which show significant changes in abundance between pCO_2 treatments. The detailed analysis of the significantly correlated taxa to CO_2 is not included in the present analysis but in a complementary analysis.

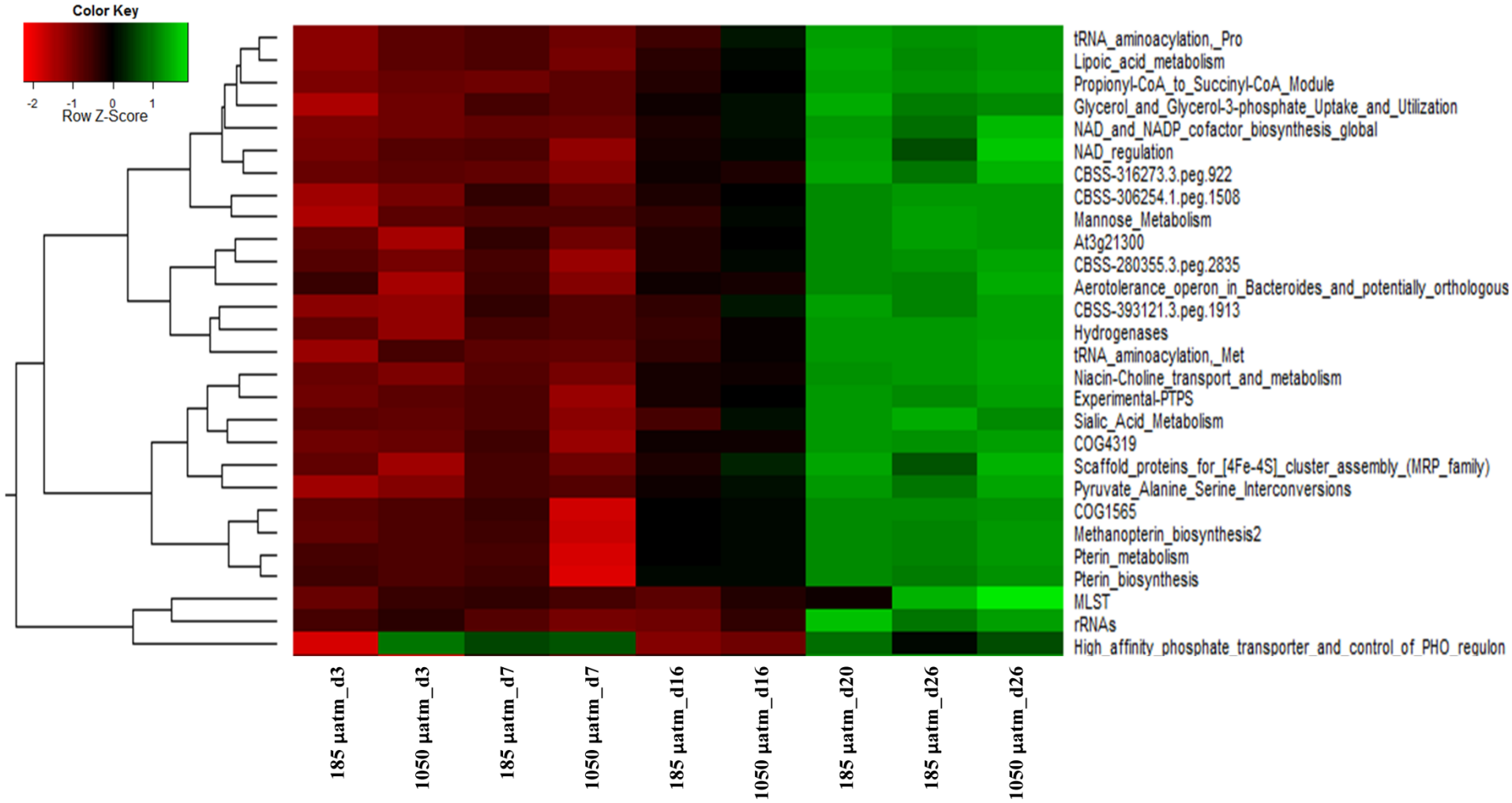


Figure 3 Heat map showing the first cluster of the metatranscriptome metabolic functions significantly ($p < 0.05$) correlated to CO_2; represents cluster **A** of supplementary Figure S1.

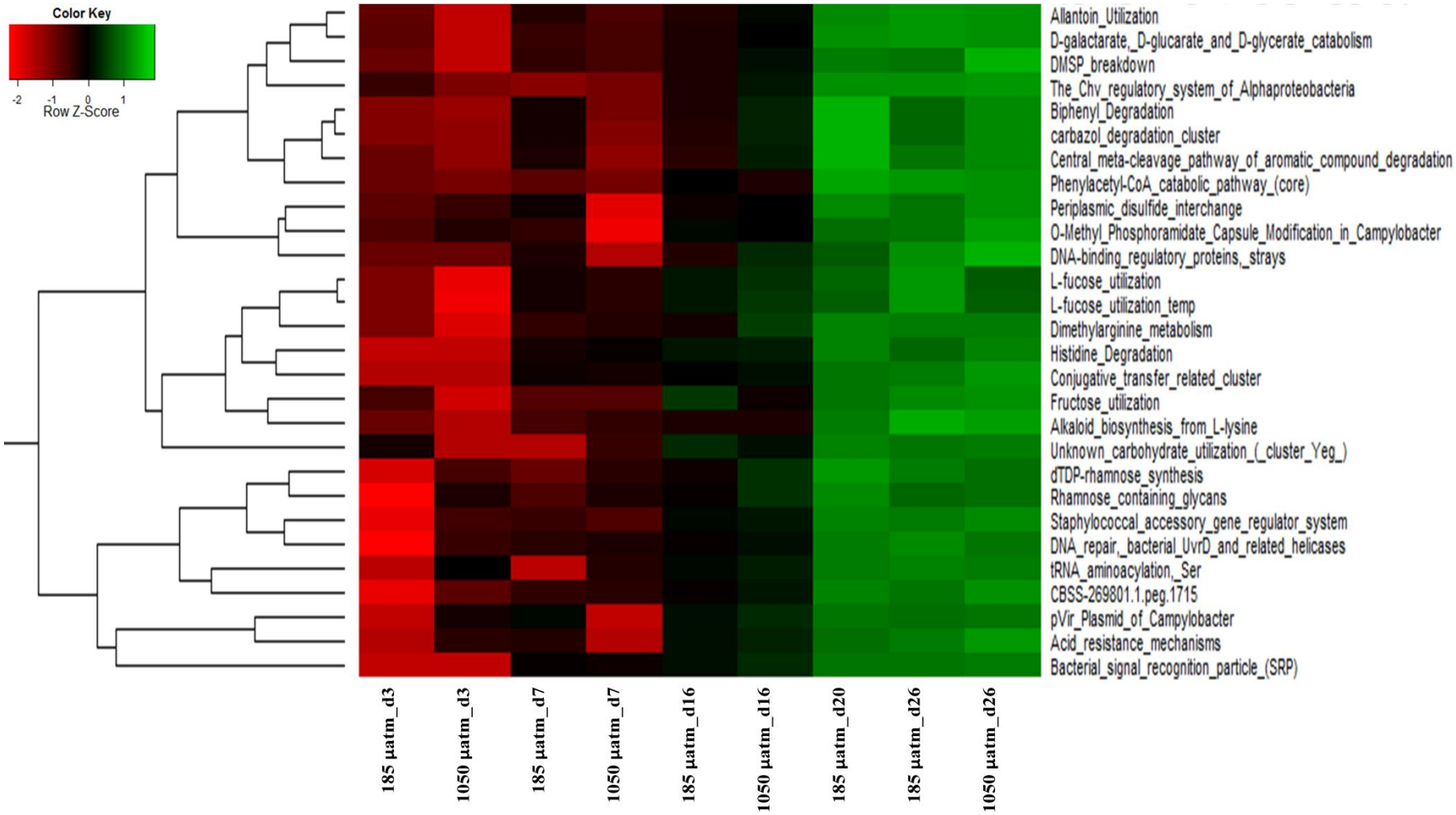


Figure 4 Heat map showing the second cluster of the metatranscriptome metabolic functions significantly ($p < 0.05$) correlated to CO_2; represents cluster **B** of supplementary Figure S1.

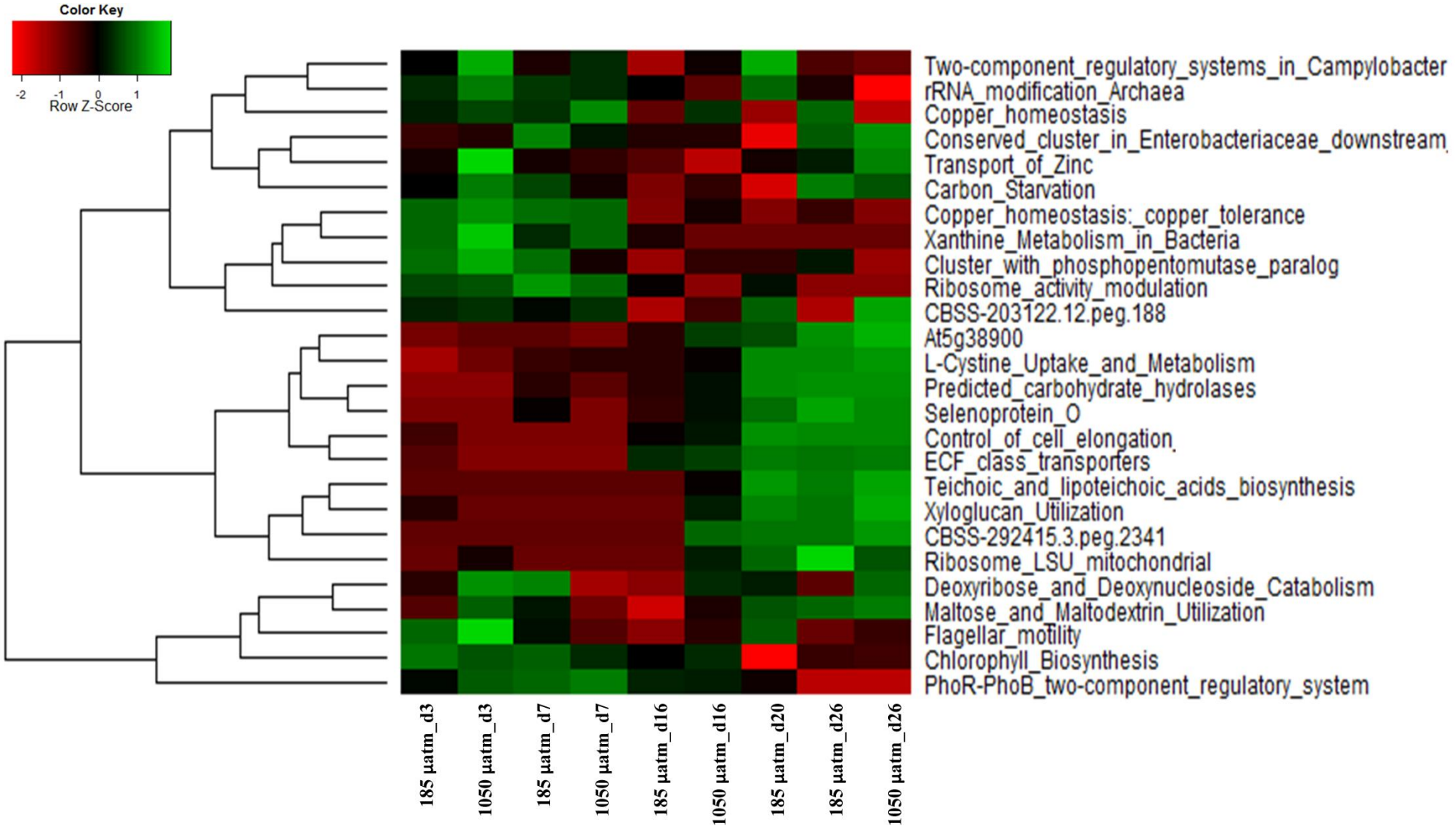


Figure 5 Heat map showing the third cluster of the metatranscriptome metabolic functions significantly ($p < 0.05$) correlated to CO_2; represents cluster **C** of supplementary Figure S1.

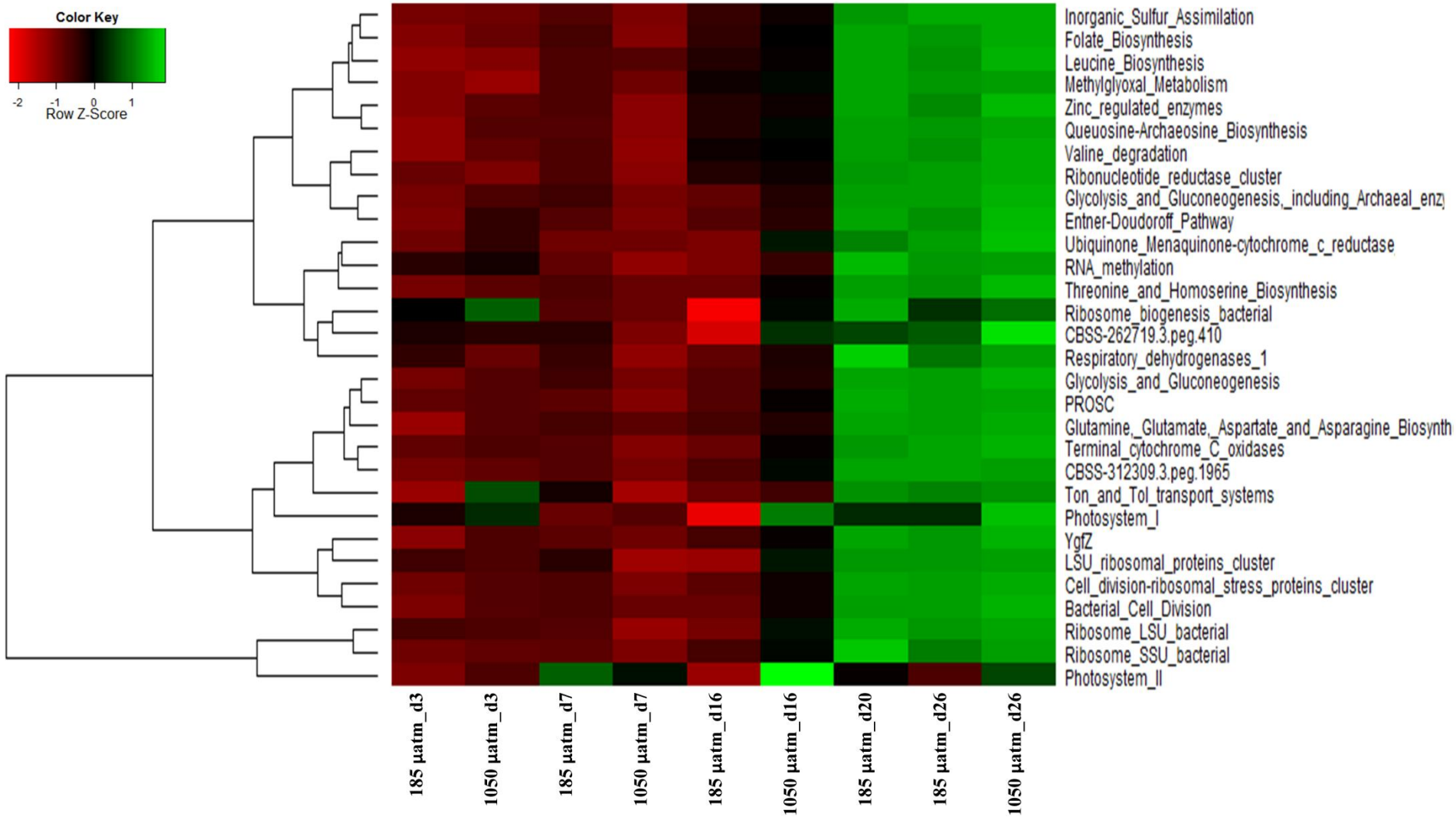


Figure 6 Heat map showing the forth cluster of the metatranscriptome metabolic functions significantly ($p < 0.05$) correlated to CO_2; represents cluster **D** of supplementary Figure S1.

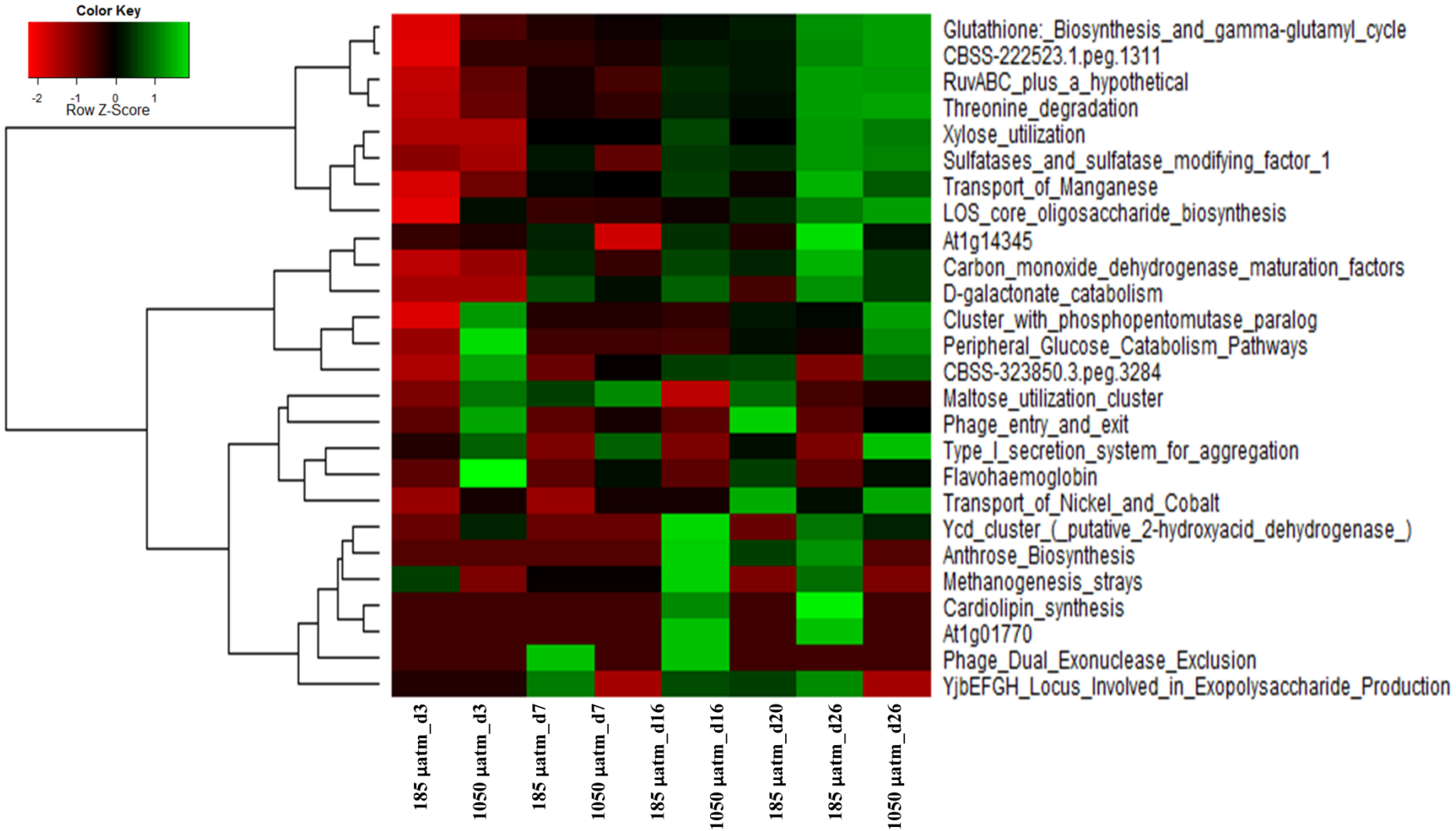


Figure 7 Heat map showing metagenome metabolic functions significantly ($p < 0.05$) correlated to CO_2.

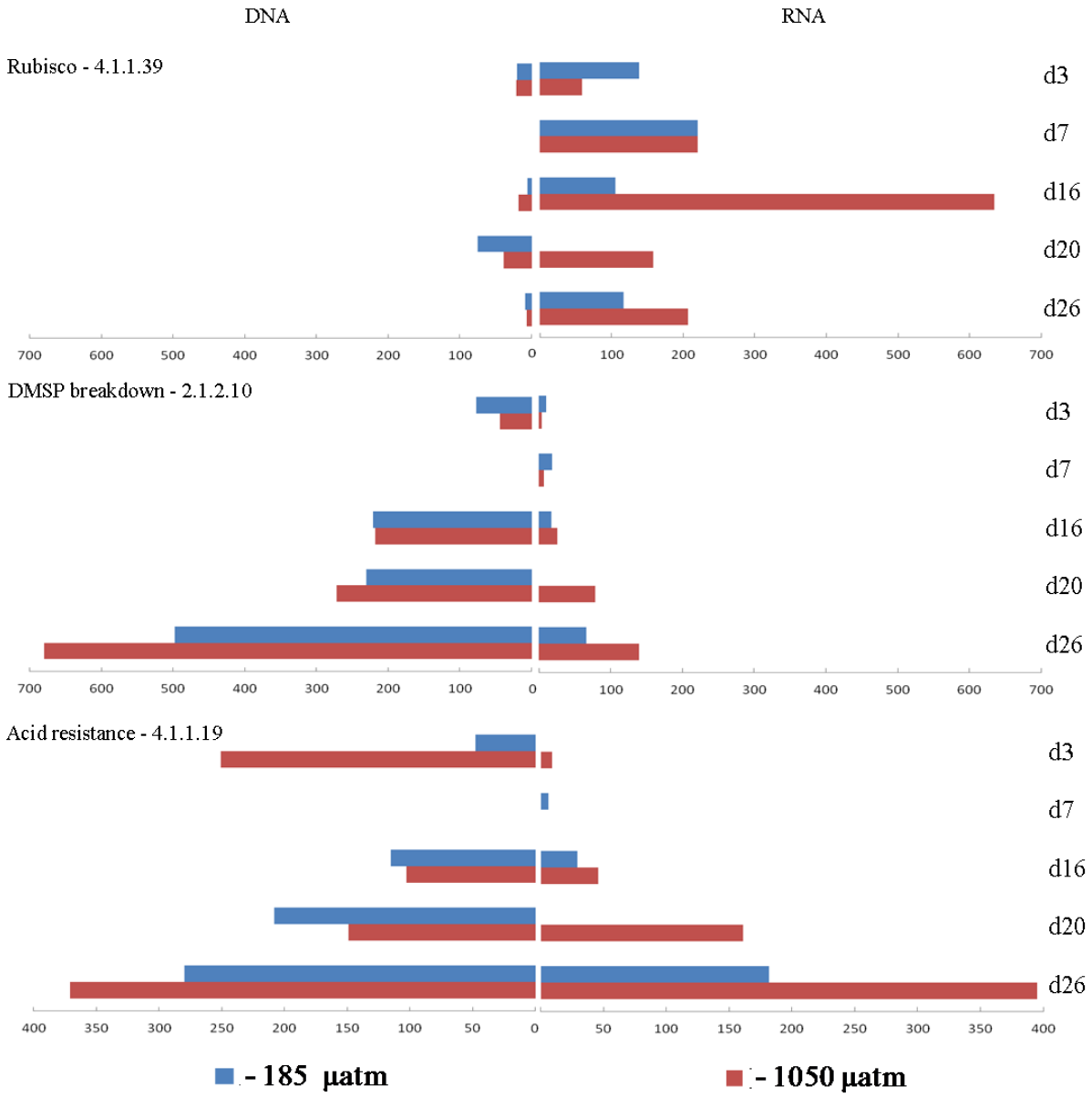


Figure 8 DNA (left) and RNA (right) relative MG-RAST sequences (≈reads) abundance for 1 essential photosynthetic metabolic function and 2 significantly differentially expressed (185 and 1050 µatm) metabolic functions; EC# included.

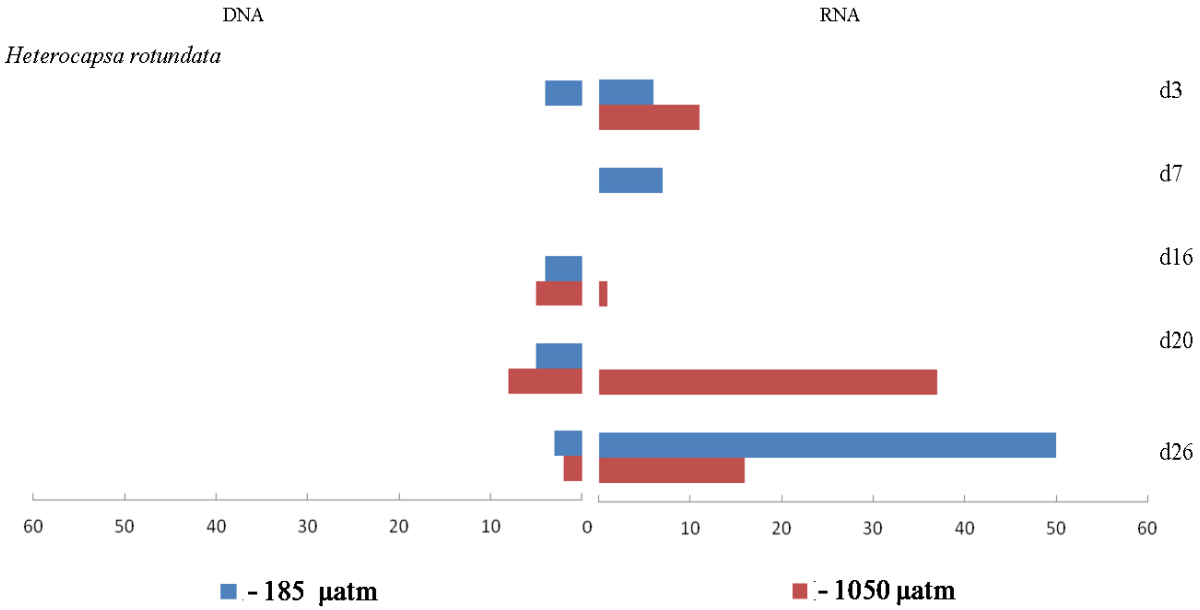


Figure 9 DNA (left) and RNA (right) relative MG-RAST sequences (≈reads) abundance for *Heterocapsa rotundata* between 185 and 1050 µatm.

pCO_2 *effect on microbial metabolic functions*

To describe the community gene expression profiles in detail, samples were tested separately in order to demonstrate the significant (non parametric Kruskal-Wallis ANOVA with R) effect of time and CO_2. The first four heat maps present chronologically arranged metatranscriptome functions, which showed significant differences in expression levels in each mesocosm (total 112 sub-groups) and were divided into 4 clusters to allow better readability of the genes (Figure 3-6 and Figure S1). It showed that the upper and lower clusters seemed to be more strongly affected by *p*CO_2 at the beginning of the experiment or before the addition of nutrient. The middle cluster indicates a strong effect throughout the experiment. Three targeted functions, two significantly affected by CO_2 levels, dimethylsulfoniopropionate (DMSP) breakdown, acid resistance (Figure 4) and Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO, EC 4.1.1.39) were further investigated to better characterize their response to CO_2, and are presented in Figure 8. This figure presents the difference between the relative abundance of genes and transcripts. The DMSP breakdown annotations (enzyme aminotransferase, DMSP breakdown hydrolase protein, DMSP demethylase [EC 2.1.2.10], DMSP demethylase transcriptional regulator) presented to opposite sequences' and transcripts abundance than RuBisCO. DMSP breakdown continuously increased with time. However, the difference between the two *p*CO_2 treatments is small, up until the end of the experiment. The sequences' abundances and transcripts of the Dinoflagellata species (*Heterocapsa rotundata)* identified by Archer *et al.* (2013) as possibly responsible for the synthesis of DMSP in the mesocosms is presented in Figure 9. It shows that the relative sequence abundances in the DMSP pathway is lower than 10% throughout the experiment, but the transcript levels rose a short time after nutrient addition. The acid resistance functional cluster, containing the enzyme Arginine decarboxylase [EC 4.1.1.1], had a large difference between mesocosms at the beginning of the experiment, with almost no transcripts in the 1050 μatm mesocosm until nutrient addition when the sequences' abundances and transcripts slowly increased to reach high level in both *p*CO_2 reaching largest differences between pressures on d26. RuBisCO was not abundant throughout the experiment however its expression is much higher, reaching its maximum at high CO_2 on d16, with no significant effect of CO_2. RuBisCO was included in the analysis as it is an essential photosynthetic enzyme involved with carbon fixation and is established as being affected by elevated *p*CO_2; significantly affected functions (total 26 sub-groups) were also identified in the DNA dataset (Figure 7) but the detailed analysis of these functions was not included in the present analysis (undergoing).

Preliminary association of taxa with metabolic functions

Preliminary results of the organism's identification for the expression of the significantly affected metabolic functions (acid resistance, DMSP breakdown and non significant function RuBisCO) are presented in the supplementary material (Figure S2 to S10). The SAR11 clade appears to be the most abundant taxa expressing acid resistance in the metagenome under both pCO_2 during the length of the experiment; except for d3 of the 1050 µatm which was dominated by *Pseudoalteromonaceae*. The organisms found in the SAR11 clade were *Candidatus pelagibacter* and an organism classified as α-proteobacteria HIMB5 while the most abundant of the *Pseudoalteromonaceae* were species of *Pseudoalteromonas*. Similar results were found in the metatranscriptome for the taxa associated to acid resistance with, however, less abundant (no sequences found in d3-185 µatm and d7-1050 µatm) and less diversified taxa found. The diversity of taxa expressing acid resistance increased with time. The comparison of both taxomic lists on d26 (Figure S3) shows that the same taxa are expressing acid resistance enzymes but with different abundance. The DMSP breakdown enzyme were associated to SAR11 (*Candidatus pelagibacter* and α-proteobacteria HIMB5) which dominated the metagenome and metatranscriptome with only one other abundant taxon, the *Rhodobacteraceae*. The comparison on d26 (Figure S6) indicated that more taxa were present in the low pCO_2 metagenome as in the high pCO_2 one whereas the opposite is demonstrated in the metatranscriptome. Rubisco is by far the function expressed by the most diversified taxa with the dominating organisms varying greatly between datasets, time and pCO_2; the comparison on d26 demonstrates this fact (Figure S9). The metagenome is however, dominated by multiple abundant taxa at d3 but quickly decrease to be dominated by Mamiellales (*Micromonas sp.*), *Bacillariophyceae*, unknown, *Xanthomonadacea*, and unclassified γ-proteobacteria under the low CO_2 while Guillardia (*Guillardia theta*), unknown and unclassified γ-proteobacteria dominated under high CO_2. The metatranscriptome was, mainly dominated by Mamiellales (*Micromonas sp* and *Ostreococcus* sp.) and Embryophyta with also a large percentage represented by *Phaeocystis* sp. throughout the experiment. The statistical analyses are presently undergoing but this is a first glance into characterizing organisms actually expressing metabolic functions significantly influenced by changing CO_2.

Discussion

The microbial community structure and the corresponding gene expression under ≈185 and ≈1050 µatm pCO_2 metatranscriptome confirmed the previous 16S results (CO_2 not significantly structuring bacterial communities, but is altering the abundances of a few rare taxa).

Community and functional structure

The general microbial overview (Figure 1) showed that Archaea constitute only a small part of the microbial community found in our metagenomes and metatranscriptomes, indicating that they are not dominating in Svalbard mesocosms' communities. The sequences matched to viral DNA/RNA appeared to be in even lower abundance, which is likely related to the choice of collection filters that are unsuitable for sampling of viruses. Only viruses that are attached or inside cells or other particles can be detected in this experiment and resulted in low viruses' abundance in our results. However, their abundance should have increased at the end of each bloom, as previously established by similar study on high CO_2 (Gilbert *et al.*, 2008) and because the dominant pico- and nano-plankton were found to be susceptible to viral lysis by Brussaard and colleagues (Brussaard *et al.*, 2013). The large proportion of "unknown" reads in the metagenomes was comparable to other similar studies. An average of 50% remained unidentified using comparable high-throughput sequencing in marine environments (Frias-Lopez *et al.*, 2008; Shi *et al.*, 2011; Stewart *et al.*, 2012; Schunck *et al.*, 2013). The "unknown" section of the metatranscriptome is smaller, although still of important magnitude. Many of the unknown RNA sequences in the metatranscriptome are potentially small RNAs-like non-coding-proteins or/and non-ribosomal RNA (Shi *et al.*, 2011) as eukaryotic and prokaryotic ribosomal RNA were removed (see material and methods)

The metatranscriptome presented a larger proportion of eukaryotic taxa, probably because eukaryotic cells contain a higher concentration of RNA than DNA compared to bacteria. A small number of eukaryotic sequences were detected in the metagenome whereas a higher proportion of eukaryotic sequences are found in the metatranscriptome.

The most abundant taxa (Figure 2) showed that Bacteroidetes do not dominate the metagenome nor the metatranscriptome assemblages, in contrast to the results obtained with the 16S amplicon sequencing (Roy *et al.*, 2013). The Bacteroidetes were found to contribute

to 50% of the 16S amplicon sequences throughout most of the experiment whereas with the approach taken here, the Bacteroidetes were almost absent in the high CO_2 and only present in the low CO_2 metagenome; suggesting that the Bacteroidetes were either not really active according to the metatranscriptomes or that 16S rRNA amplification and sequencing biases underestimated their abundance. Alternatively, it could also be due to the sequencing results being annotated against a low number of representative Bacteroidetes genomes resulting in the gene sequences to be annotated to another class. Bacteroidetes were also previously established as the dominant marine pelagic community members in diversified environment from surface water sampled at the Hawaii Ocean Time-series (station ALOHA) and the Baltic sea, but their activity was not quantified (Frias-Lopez *et al.*, 2008; Herlemann *et al.*, 2011, respectively). In the present experiment, the α-proteobacteria were dominant in the low CO_2 treatment and the α- and γ-proteobacteria were dominant in the high CO_2 treatment of the metagenome. Conversely, the abundance of α-proteobacteria was lower in the metatranscriptome corresponding to a high presence but low activity of these bacteria. Their abundance was again reported as much higher in the assemblage describe by the 16S analysis (Roy *et al.*, 2013) with the γ-proteobacterial sequences accounting for about 20% while here only present in the metagenome on d3 and d20 of the high CO_2. The Beta (β)-proteobacterial sequences were present in the 16S amplicon dataset (Roy *et al.*, 2013) in low, but consistent numbers; in this study, both the metagenome and metatranscriptome showed only negligible numbers of β-proteobacterial sequences, indicating a minor functional role of them in this pelagic community. These differences could however be due to experimental amplification biases or differences in genes annotation (Greengenes versus M5NR databases).

The sequenced samples taken on d3 demonstrate that the metagenomic microbial community in the low and high CO_2 were already showing major composition differences, possibly caused by the CO_2 manipulations in one of the enclosure. For example, the abundances of γ-proteobacteria in both metagenomes were respectively at about 30% (low) versus 60% (high) at starting conditions. Furthermore, the γ-proteobacteria were almost absent after nutrient addition but quickly came back under high CO_2 conditions. It is possible that some organisms were more sensitive to the CO_2 manipulation and drastically decreased during the handling, these organisms could possibly have recovered due to the cessation of the manipulations and the addition of nutrient making the mesocosms return to favorable environmental conditions. The composition of the diverse proteobacterial sequences and their changes over time are quite representative of the community previously reported as α- and γ-proteobacteria are

normally quite abundant in the marine ecosystem. However, the β-proteobacteria are mostly found in freshwater systems. It was previously determined that proteobacteria were generally very abundant in the marine microbial community of surface waters (Frias-Lopez *et al.*, 2008; Hewson *et al.*, 2009; Herlemann *et al.*, 2011), in marine low oxygen deeper environments (Canfield *et al.*, 2010; Stewart *et al.*, 2012; Schunck *et al.*, 2013) and in high arctic communities like Svalbard (Teske *et al.*, 2011; Roy *et al.*, 2013).

In the metatranscriptome, the beginning of the experiment is more or less similar between the different pCO_2 treatments (α-proteobacterial, Actinobacterial and many eukaryotic sequences). However, the phylogenetic affiliation based on the expressed gene pool started changing with the addition of nutrient. This effect is stronger in the high CO_2 mesocosm, as the uptake of CO_2 required for photosynthetic carbon-fixation might be facilitated by higher pCO_2 levels. The result of such an increase in abundance is represented by a high transcriptional activity, as indicated by the greater number of sequences in the metatranscriptome. The most abundant eukaryotic microorganisms were the Chlorophyta and Streptophyta which are green algae. These two divisions were presented in Figure 2 because the bacteria abundance decreases with their increase. In fact, it was previously speculated that the abundance of Chlorophyta species were increasing with latitude (Burdett *et al.*, 2013). The other abundant taxa were summarized in the "Others (eukaryotes)" as diverse taxa that were present in high abundance in both metatranscriptome (≈185 versus ≈1050 µatm), dominant between d3-d16. These other eukaryotic sequences were assigned to Apicomplexa, Arthropoda, Ascomycota, Chordata, Cnidaria, Echinoderma, Euglenida, *Eustigmatophyceae*, Mollusca, Porifera, *Pheaophyceae*, *Xanthophyceae* and fell in abundance with time, which could be caused by the nutrient being used more efficiently by the Chlorophyta and Streptophyta, and because bacteria generally have much faster cell division rates than green algae. A wide association between the taxa and their expressed transcripts in undergoing in order to facilitate the characterization of taxonomic patterns under elevated CO_2 (preliminary results presented in Figure S2 to S10b). Indeed, a shift in community structure towards the smallest phytoplankton as a result of OA was established in the present mesocosms experiment by Brussaard (Brussaard *et al.*, 2013) and would directly affect the food web structure and in turn have an impact on biogeochemical cycling.

Metabolic functions

Observation of the metabolic functional clusters (MG-RAST SEED classification level 3) significantly affected by different pCO_2 demonstrated that, even if the general impact of OA on community structure is not statistically significant, detailed investigation reveals targeted effects on certain fractions of the microbial community. This mirrors prior 16S amplicon results, which found that the abundant taxa of the community were not affected by pCO_2 whereas some rare taxa were (Roy *et al.*, 2013). Two groups of functional genes significantly affected by pCO_2 (and one important for any photosynthetic organisms) in the transcriptome data were investigated in detail.

Acid resistance enzymes

The acid resistance cluster contained arginine decarboxylase (EC 4.1.1.19), an enzyme involved in the regulation of the arginine metabolism (Wu and Morris, 1998), and widely studied in higher plants, bacteria, and mammals (Lin *et al.*, 1996; Wu and Morris, 1998; Bouchereau *et al.*, 1999; Castanie-Cornet *et al.*, 1999; Cotter and Hill, 2003). The acid resistance mechanism was described in *Escherichia coli* as the mechanisms responsible for tolerance to low pH including three characterizing systems (Lin *et al.*, 1996; Castanie-Cornet *et al.*, 1999). Two of them involve decarboxylase enzymes suspected to take up protons during the decarboxylation of arginine (or glutamate). The decarboxylation process results in the consumption of the protons contained in the cells; thus equilibrating the cell's pH avoiding lethal pH levels during acid stress events (Castanie-Cornet *et al.*, 1999). In our study, the mesocosm with the elevated CO_2 partial pressure of ≈1050 µatm had a lower pH varying from 7.3 (d3) to 7.9 (d26) compared to the pH of the ≈185 µatm mesocosm, which ranged from 8.4 (d3) to 8.5 (d26). The highest abundance of acid resistance arginine decarboxylase was found in the low pH/high CO_2 mesocosms. The microbes present in the higher pCO_2 environment were thus exposed to lower pH as the organisms exposed in the 185 µatm treatment which were evolving in an environment representative of their natural condition as the pH in 185 µatm equalled the one in the untreated Kongsfjorden (all pH data are presented in Schulz *et al.*, 2013). One interesting fact is that arginine decarboxylase process produces CO_2. This means that the increase of atmospheric CO_2 and its increase absorption by the ocean, decreases pH which in turn increases the acid resistance (arginine decarboxylase) in diverse organisms and thus produce more CO_2 (positive feedback) (Wu and Morris, 1998).

DMSP breakdown

Dimethyl sulfide (DMS) is an important reactant of the aerosol formation and growth in the (Arctic) troposphere and was included in a suite of metadata measurements in the previous 16S amplicon analysis (Roy *et al.*, 2013) of the same mesocosms study. DMS was previously established as significantly structuring the community during the mesocosm experiment (Roy *et al.*, 2013). In the present metatranscriptome, DMSP was identified as one of the multiple functional clusters significantly influenced by pCO_2 treatment level. DMSP is established as being produced by marine phytoplankton with DMS as its breakdown product (Archer *et al.*, 2011; Nevitt, 2011; Archer *et al.*, 2013; Burdett *et al.*, 2013). Briefly, DMS is involved in the global climate as a key biogenic source of aerosol creating condensation clouds. These clouds are responsible for reducing the sunlight entry in the ocean and are thus decreasing the effect of solar ultraviolet radiation a proven stressor for phytoplankton (Nevitt, 2011; Archer *et al.*, 2013; Burdett *et al.*, 2013). In addition, DMS has often been related to environmental stress and has been shown to be an antioxidant (Sunda *et al.*, 2002; Burdett *et al.*, 2013). Conversely to what was expected from the previous analysis (Sunda *et al.*, 2002) in regards to increase of DMS in relation to limited pCO_2, DMSP transcripts were twice as abundant under elevated CO_2. Prior work has suggested that intracellular DMSP concentrations in a red coralline alga were augmented by a reduced pH (Burdett *et al.*, 2013); which corresponds to the same pH level measured in the elevated pCO_2 mesocosm. It was also found that the intracellular DMSP concentration was reduced at northern latitudes in microalgae, like Chlorophyta (Burdett *et al.*, 2013). In our metatranscriptomes, elevated amounts of Chlorophyta-like sequences in the elevated pCO_2 mesocosm were found from the moment nutrient were added, which could be related to the significant differential expression of DMSP between pCO_2 treatments.

DMSP and DMS are strongly affected by low pH as their concentrations increased by 50% and decreased by 60%, respectively, right after nutrient addition in the elevated pCO_2 mesocosms (findings from Archer *et al.*, 2013). Archer and colleagues found that the synthesis rate of DMSP by phytoplankton rose with increased $[H^+]$ and was connected to inorganic carbon fixation rates (increased primary production). The increased production of DMSP was related to the increased biomass of dinoflagellates, but more precisely to the biomass of *Heterocapsa rotundata* (Archer *et al.*, 2013). The metagenome and metatranscriptome were searched for this particular species and found a low level of activity for this taxon, which almost disappeared by d16. However, *H. rotundata* transcripts increased and were quite active after nutrient addition; the 16S amplicon dataset will shortly be

screened for this species. A further search of the transcriptome is undergoing to find exactly what species are involved in the DMSP breakdown and further qualify how they are affected by OA (undergoing-preliminary Figure S5-S7b).

An extra function attributed to DMSP and thus to DMS, is its property as bio-chemical-indicator attracting predators of phytoplankton-grazing organisms. Indeed, it is believed that a large assemblage of phytoplankton would attract grazers but would also increase the production of DMS. This variation in DMS would alert predators of the presence of grazers, which would then lead to their consumption and reduce the grazing pressure on phytoplankton. It could also allow for the renewal of the phytoplankton nutrient pool via the grazer waste output (Nevitt, 2011).

RuBisCO

RuBisCO is one of the most important enzymes found in any photosynthetic organisms as it is involved in the first carboxylation step of photosynthetic carbon acquisition. It is a carbon-fixing enzyme closely interacting with the carbon concentrating mechanisms (CCM) responsible for the increase or replenishment of CO_2 concentrations at the fixation sites surrounding RuBisCO and support the vital CO_2 fixation process (Burkhardt *et al.*, 2001; Hopkinson *et al.*, 2011; Alterio *et al.*, 2012; Zabaleta *et al.*, 2012; Hopkinson *et al.*, 2013). It was further established that the CCMs are down-regulated by high CO_2 concentrations (Wu *et al.*, 2010; Hopkinson *et al.*, 2011; Young *et al.*, 2012; Hopkinson *et al.*, 2013) as they become superfluous when large amounts of CO_2 are available. Therefore, the abundance of RuBisCO transcripts were expected to be lower under elevated CO_2, which was however not the case in the present experiment. Indeed, RuBisCO was not significantly affected by the changes in pCO_2 even on d16 of the transcriptome results where it seemed to be activated (≈1050 µatm). This non-significant effect could be the result of the addition of nutrient as multiple studies on terrestrial organisms showed an increased photosynthetic rate under elevated CO_2 (Campbell *et al.*, 1988; Lesser, 1996a, b) that would only last as long as enough other nutrient, e.g. nitrogen, were available in the environment (Tissue *et al.*, 1993). It could be that the photosynthetic organisms took advantage of the new increase in nutrient as long as the competition for resources was still low.

The RuBisCO total genes abundance (MG-RAST abundance) was also fairly low and the lack of significant differences in the transcript abundance of RuBisCO was previously observed in

soybeans (Campbell *et al.*, 1988) where it was demonstrated that the range of CO_2 pressure for RuBisCO's activity to remain unresponsive was quite large (≈110-880 µatm). In our experiment, the transcripts of RuBisCO are a little more abundant in the low CO_2 however not significantly different indicating that the CCM might be a little stimulated in the low ≈185 µatm mesocosms but still offered enough photosynthetic substrate as in the ≈1050 µatm mesocosms.

Conclusion

In conclusion, this study confirms previous findings from 16S rRNA amplicon sequencing, which established that microbial community composition at Svalbard Kongsfjorden was not structured by CO_2 but did change significantly with time/nutrient addition. It also confirmed that the taxa significantly affected by CO_2 are not the most abundant genera but rather the rare ones. We further established that changing pCO_2 significantly influences metabolic functions. Two of these significantly affected functions were here analyzed, demonstrating how OA can affect genes expression. The essential photosynthetic enzyme RuBisCO was also observed in detail and showed that even if a function is not statistically significantly differentially expressed, some modifications from its production at ambient CO_2 or pH are nevertheless happening. Our findings did not establish that OA had a major impact on the microbial community of Kongsfjorden as found by Zhang *et al.* (2013). We are thus presently determining the taxa expressing the functions identified as affected by CO_2 and their phylogenetic relationships to see if they correspond to the 150 species (summed metagenome and metatranscriptome) found to be affected by OA in the present study.

Future work should be concentrated on dissecting the metagenome and metatranscriptome in order to evaluate the effect of OA on separated taxa. It should further include a specie-specific research for organisms identified by Svalbard's colleagues as having a major role in parameters analyzed; comparatively to the *H. rotundata* (genome not yet sequenced) analysis started here for its role in DMSP and DMS cycles. Also, microbial community are often facing extreme environmental changes over a short period of time (days, hours) and their ability to adapt should be included in further investigation. Lohbeck *et al.*, (2012) showed the importance of integrating adaptation in OA research. The investigations of connections (flow cytometry, microscopy,...) between community structure, genes expression/alteration and

processes affected by OA will help draw a concrete picture of the long term effects of OA on our world.

Conflicts of interest

The authors declare no conflict of interest

Acknowledgement

We would like to thank the "European Project on Ocean Acidification" (EPOCA) which received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 211384. We are grateful to every organization, vessel's crews and everyone involved in the Svalbard mesocosm experiment. We also want to thank Tina Baustian for her help during the RNA/DNA extraction. Finally, the funding for S.M.G. was provided by an EPA STAR Graduate Fellowship.

References

Agogue H, Lamy D, Neal PR, Sogin ML, Herndl GJ (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol Ecol* **20:** 258-274.

Allgaier M, Riebesell U, Vogt M, Thyrhaug R, Grossart HP (2008). Coupling of heterotrophic bacteria to phytoplankton bloom development at different pCO(2) levels: a mesocosm study. *Biogeosciences* **5:** 1007-1022.

Alterio V, Langella E, Viparelli F, Vullo D, Ascione G, Dathan NA *et al* (2012). Structural and inhibition insights into carbonic anhydrase CDCA1 from the marine diatom Thalassiosira weissflogii. *Biochimie* **94:** 1232-1241.

Archer SD, Safi K, Hall A, Cummings DG, Harvey M (2011). Grazing suppression of dimethylsulphoniopropionate (DMSP) accumulation in iron-fertilised, sub-Antarctic waters. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **58:** 839-850.

Archer SD, Kimmance SA, Stephens JA, Hopkins FE, Bellerby RGJ, Schulz KG *et al* (2013). Contrasting responses of DMS and DMSP to ocean acidification in Arctic waters. *Biogeosciences* **10:** 1893-1908.

Arnosti C, Grossart HP, Muhling M, Joint I, Passow U (2011). Dynamics of extracellular enzyme activities in seawater under changed atmospheric pCO(2): a mesocosm investigation. *Aquatic Microbial Ecology* **64:** 285-298.

Bouchereau A, Aziz A, Larher F, Martin-Tanguy J (1999). Polyamines and environmental challenges: recent development. *Plant Science* **140:** 103-125.

Brussaard CPD, Noordeloos AAM, Witte H, Collenteur MCJ, Schulz K, Ludwig A *et al* (2013). Arctic microbial community dynamics influenced by elevated CO_2 levels. *Biogeosciences* **10:** 719-731.

Burdett HL, Donohue PJC, Hatton AD, Alwany MA, Kamenos NA (2013). Spatiotemporal Variability of Dimethylsulphoniopropionate on a Fringing Coral Reef: The Role of Reefal Carbonate Chemistry and Environmental Variability. *Plos One* **8**.

Burkhardt S, Amoroso G, Riebesell U, Sultemeyer D (2001). CO_2 and HCO3- uptake in marine diatoms acclimated to different CO_2 concentrations. *Limnol Oceanogr* **46:** 1378-1391.

Campbell WJ, Allen LH, Bowes G (1988). Effects of CO_2 concentration on rubisco activity, amount, and photosynthesis in soybean leaves. *Plant Physiol* **88:** 1310-1316.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science* **330:** 1375-1378.

Castanie-Cornet MP, Penfound TA, Smith D, Elliott JF, Foster JW (1999). Control of acid resistance in Escherichia coli. *Journal of Bacteriology* **181:** 3525-3535.

Cotter PD, Hill C (2003). Surviving the acid test: Responses of gram-positive bacteria to low pH. *Microbiology and Molecular Biology Reviews* **67:** 429-+.

Czerny J, Schulz KG, Boxhammer T, Bellerby RGJ, Büdenbender J, Engel A *et al* (2013a). Implications of elevated CO_2 on pelagic carbon fluxes in an Arctic mesocosm study – an elemental mass balance approach. *Biogeosciences* **10:** 3109-3125.

Czerny J, Schulz KG, Ludwig A, Riebesell U (2013b). Technical Note: A simple method for air–sea gas exchange measurements in mesocosms and its application in carbon budgeting. *Biogeosciences* **10:** 1379-1390.

Delille B, Harlay J, Zondervan I, Jacquet S, Chou L, Wollast R *et al* (2005). Response of primary production and calcification to changes of pCO(2) during experimental blooms of the coccolithophorid Emiliania huxleyi. *Global Biogeochemical Cycles* **19**.

Desai DK, Schunck H, Loser JW, LaRoche J (2013). Fragment recruitment on metabolic pathways: comparative metabolic profiling of metagenomes and metatranscriptomes. *Bioinformatics* **29**: 790-791.

Engel A, Borchard C, Piontek J, Schulz KG, Riebesell U, Bellerby R (2013). CO_2 increases 14C primary production in an Arctic plankton community. *Biogeosciences* **10:** 1291-1308.

Falkowski PG, Fenchel T, Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320:** 1034-1039.

Feely RA, Doney SC, Cooley SR (2009). Ocean Acidification: Present Conditions and Future Changes in a High-CO_2 World. *Oceanography* **22:** 36-47.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105:** 3805-3810.

Fu FX, Mulholland MR, Garcia NS, Beck A, Bernhardt PW, Warner ME *et al* (2008). Interactions between changing pCO(2), N-2 fixation, and Fe limitation in the marine unicellular cyanobacterium Crocosphaera. *Limnol Oceanogr* **53:** 2472-2484.

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc Natl Acad Sci U S A* **106:** 22427-22432.

Gilbert JA, Field D, Huang Y, Edwards R, Li WZ, Gilna P *et al* (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *Plos One* **3**.

Gilbert JA, Field D, Swift P, Newbold L, Oliver A, Smyth T *et al* (2009). The seasonal structure of microbial communities in the Western English Channel. *Environmental Microbiology* **11:** 3132-3139.

Grossart HP, Allgaier M, Passow U, Riebesell U (2006). Testing the effect of CO_2 concentration on the dynamics of marine heterotrophic bacterioplankton. *Limnol Oceanogr* **51:** 1-11.

Herlemann DPR, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *Isme Journal* **5:** 1571-1579.

Hewson I, Paerl RW, Tripp HJ, Zehr JP, Karl DM (2009). Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. *Limnol Oceanogr* **54:** 1981-1994.

Hoegh-Guldberg O, Bruno JF (2010). The Impact of Climate Change on the World's Marine Ecosystems. *Science* **328:** 1523-1528.

Hopkinson BM, Dupont CL, Allen AE, Morel FMM (2011). Efficiency of the CO_2-concentrating mechanism of diatoms. *Proc Natl Acad Sci U S A* **108:** 3830-3837.

Hopkinson BM, Meile C, Shen C (2013). Quantification of Extracellular Carbonic Anhydrase Activity in Two Marine Diatoms and Investigation of Its Role. *Plant Physiol* **162:** 1142-1152.

Huber JA, Mark Welch D, Morrison HG, Huse SM, Neal PR, Butterfield DA *et al* (2007). Microbial population structures in the deep marine biosphere. *Science* **318:** 97-100.

Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, Sogin ML (2008). Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *Plos Genetics* **4**.

Hutchins DA, Fu FX, Zhang Y, Warner ME, Feng Y, Portune K *et al* (2007). CO_2 control of Trichodesmium N-2 fixation, photosynthesis, growth rates, and elemental ratios: Implications for past, present, and future ocean biogeochemistry. *Limnol Oceanogr* **52:** 1293-1304.

Jeffries TC, Seymour JR, Gilbert JA, Dinsdale EA, Newton K, Leterme SSC *et al* (2011). Substrate Type Determines Metagenomic Profiles from Diverse Chemical Habitats. *Plos One* **6**.

Keshri J, Mishra ., Jha B (2013). Microbial population index and community structure in saline-alkaline soil using gene targeted metagenomics. *Microbiological Research* **168**: 165-173.

Kirchman DL, Cottrell MT, Lovejoy C (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology* **12:** 1132-1143.

Krause E, Wichels A, Gimenez L, Lunau M, Schilhabel MB, Gerdts G (2012). Small Changes in pH Have Direct Effects on Marine Bacterial Community Composition: A Microcosm Approach. *Plos One* **7**.

Lovejoy C, Massana R, Pedros-Alio C (2006). Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl Environ Microbiol* **72:** 3085-3095.

Lovejoy C, Potvin M (2011). Microbial eukaryotic distribution in a dynamic Beaufort Sea and the Arctic Ocean. *J Plankton Res* **33:** 431-444.

Lesser MP (1996a). Elevated temperatures and ultraviolet radiation cause oxidative stress and inhibit photosynthesis in symbiotic dinoflagellates. *Limnol Oceanogr* **41:** 271-283.

Lesser MP (1996b). Acclimation of phytoplankton to UV-B radiation: Oxidative stress and photoinhibition of photosynthesis are not prevented by UV-absorbing compounds in the dinoflagellate Prorocentrum micans. *Marine Ecology Progress Series* **132:** 287-297.

Lin JS, Smith MP, Chapin KC, Baik HS, Bennett GN, Foster JW (1996). Mechanisms of acid resistance in enterohemorrhagic Escherichia coli. *Appl Environ Microbiol* **62:** 3094-3100.

Lohbeck KT, Riebesell U, Reusch TBH (2012). Adaptive evolution of a key phytoplankton species to ocean acidification (vol 5, pg 346, 2012). *Nature Geoscience* **5:** 917-917.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376-380.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* **9**.

Nevitt GA (2011). The Neuroecology of Dimethyl Sulfide: A Global-Climate Regulator Turned Marine Infochemical. *Integrative and Comparative Biology* **51:** 819-825.

Newbold LK, Oliver AE, Booth T, Tiwari B, DeSantis T, Maguire M *et al* (2012). The response of marine picoplankton to ocean acidification. *Environmental Microbiology* **14:** 2293-2307.

Orr JC, Fabry VJ, Aumont O, Bopp L, Doney SC, Feely RA *et al* (2005). Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437:** 681-686.

Piontek J, Borchard C, Sperling M, Schulz KG, Riebesell U, Engel A (2013). Response of bacterioplankton activity in an Arctic fjord system to elevated pCO_2: results from a mesocosm perturbation study. *Biogeosciences* **10:** 297-314.

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P *et al* (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16:** 867-880.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing: Vienne, Austria.

Riebesell U, Bellerby RGJ, Engel A, Fabry VJ, Hutchins DA, Reusch TBH *et al* (2008). Comment on "Phytoplankton Calcification in a High-CO(2) World". *Science* **322**.

Riebesell U, Czerny J, von Bröckel K, Boxhammer T, Büdenbender J, Deckelnick M *et al* (2013). Technical Note: A mobile sea-going mesocosm system – new opportunities for ocean change research. *Biogeosciences* **10:** 1835-1847.

Roy AS, Gibbons SM, Schunck H, Owens S, Caporaso JG, Sperling M *et al* (2013). Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms. *Biogeosciences* **10:** 555-566.

Sabine CL, Feely RA, Gruber N, Key RM, Lee K, Bullister JL *et al* (2004). The oceanic sink for anthropogenic CO_2. *Science* **305:** 367-371.

Schulz KG, Bellerby RGJ, Brussaard CPD, Büdenbender J, Czerny J, Engel A *et al* (2013). Temporal biomass dynamics of an Arctic plankton bloom in response to increasing levels of atmospheric carbon dioxide. *Biogeosciences* **10:** 161-180.

Schunck H, Lavik G, Desai DK, Großkopf T, Kalvelage T, Löscher CR *et al* (2013). Giant Hydrogen Sulfide Plume in the Oxygen Minimum Zone off Peru Supports Chemolithoautotrophy. *PLoS ONE* **8**.

Shi DL, Xu Y, Hopkinson BM, Morel FMM (2010). Effect of Ocean Acidification on Iron Availability to Marine Phytoplankton. *Science* **327:** 676-679.

Shi YM, Tyson GW, Eppley JM, DeLong EF (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *Isme Journal* **5:** 999-1013.

Silyakova A, Bellerby RGJ, Schulz KG, Czerny J, Tanaka T, Nondal G *et al* (2013). Pelagic community production and carbon-nutrient stoichiometry under variable ocean acidification in an Arctic fjord. *Biogeosciences* **10:** 4847-4859.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* **103:** 12115-12120.

Sperling M, Piontek J, Gerdts G, Wichels A, Schunck H, Roy AS *et al* (2013). Effect of elevated CO_2 on the dynamics of particle-attached and free-living bacterioplankton communities in an Arctic fjord. *Biogeosciences* **10:** 181-191.

Stewart FJ, Ulloa O, DeLong EF (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **14:** 23-40.

Sunda W, Kieber DJ, Kiene RP, Huntsman S (2002). An antioxidant function for DMSP and DMS in marine algae. *Nature* **418:** 317-320.

Tanaka T, Thingstad TF, Lovdal T, Grossart HP, Larsen A, Allgaier M *et al* (2008). Availability of phosphate for phytoplankton and bacteria and of glucose for bacteria at different pCO(2) levels in a mesocosm study. *Biogeosciences* **5:** 669-678.

Tanaka T, Alliouane S, Bellerby RGB, Czerny J, de Kluijver A, Riebesell U *et al* (2013). Effect of increased pCO_2 on the planktonic metabolic balance during a mesocosm experiment in an Arctic fjord. *Biogeosciences* **10:** 315-325.

Teske A, Durbin A, Ziervogel K, Cox C, Arnosti C (2011). Microbial Community Composition and Function in Permanently Cold Seawater and Sediments from an Arctic Fjord of Svalbard. *Appl Environ Microbiol* **77:** 2008-2018.

Thomas T, Gilbert J, Meyer F (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2**.

Tissue DT, Thomas RB, Strain BR (1993). Long-term effects of elevated CO_2 and nutrients on photosynthesis and rubisco in loblolly-pine seedlings. *Plant Cell and Environment* **16:** 859-865.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome. *Plos One* **3**.

Worden AZ, Not F (2008). Ecology and Diversity of Picoeukaryotes. In: Kirchman D (ed). *Microbial Ecology of the Ocean, 2nd Edition*. Wiley.

Wu GY, Morris SM (1998). Arginine metabolism: nitric oxide and beyond. *Biochemical Journal* **336:** 1-17.

Wu Y, Gao K, Riebesell U (2010). CO_2-induced seawater acidification affects physiological performance of the marine diatom Phaeodactylum tricornutum. *Biogeosciences* **7:** 2915-2923.

Xie W, Wang FP, Guo L, Chen ZL, Sievert SM, Meng J *et al* (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *Isme Journal* **5:** 414-426.

Xie ZR, Xu B, Ding JM, Liu LY, Zhang XL, Li JJ *et al* (2013). Heterologous expression and characterization of a malathion-hydrolyzing carboxylesterase from a thermophilic bacterium, Alicyclobacillus tengchongensis. *Biotechnology Letters* **35:** 1283-1289.

Yilmaz P, Gilbert JA, Knight R, Amaral-Zettler L, Karsch-Mizrachi I, Cochrane G *et al* (2011). The genomic standards consortium: bringing standards to life for microbial ecology. *Isme Journal* **5:** 1565-1567.

Young JN, Rickaby REM, Kapralov MV, Filatov DA (2012). Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society B-Biological Sciences* **367:** 483-492.

Zabaleta E, Martin MV, Braun HP (2012). A basal carbon concentrating mechanism in plants? *Plant Science* **187:** 97-104.

Zeebe RE, Wolf-Gladrow D (2001). CO_2 *in Seawater: Equilibrium, Kinetics, Isotopes*, 1st edition edn, vol. 65. Elsevier: Amsterdam.

Zhang R, Xia X, Lau SCK, Motegi C, Weinbauer MG, Jiao N (2013). Response of bacterioplankton community structure to an artificial gradient of pCO_2 in the Arctic Ocean. *Biogeosciences* **10**:3679-89.

SUPPLEMENTARY MATERIAL

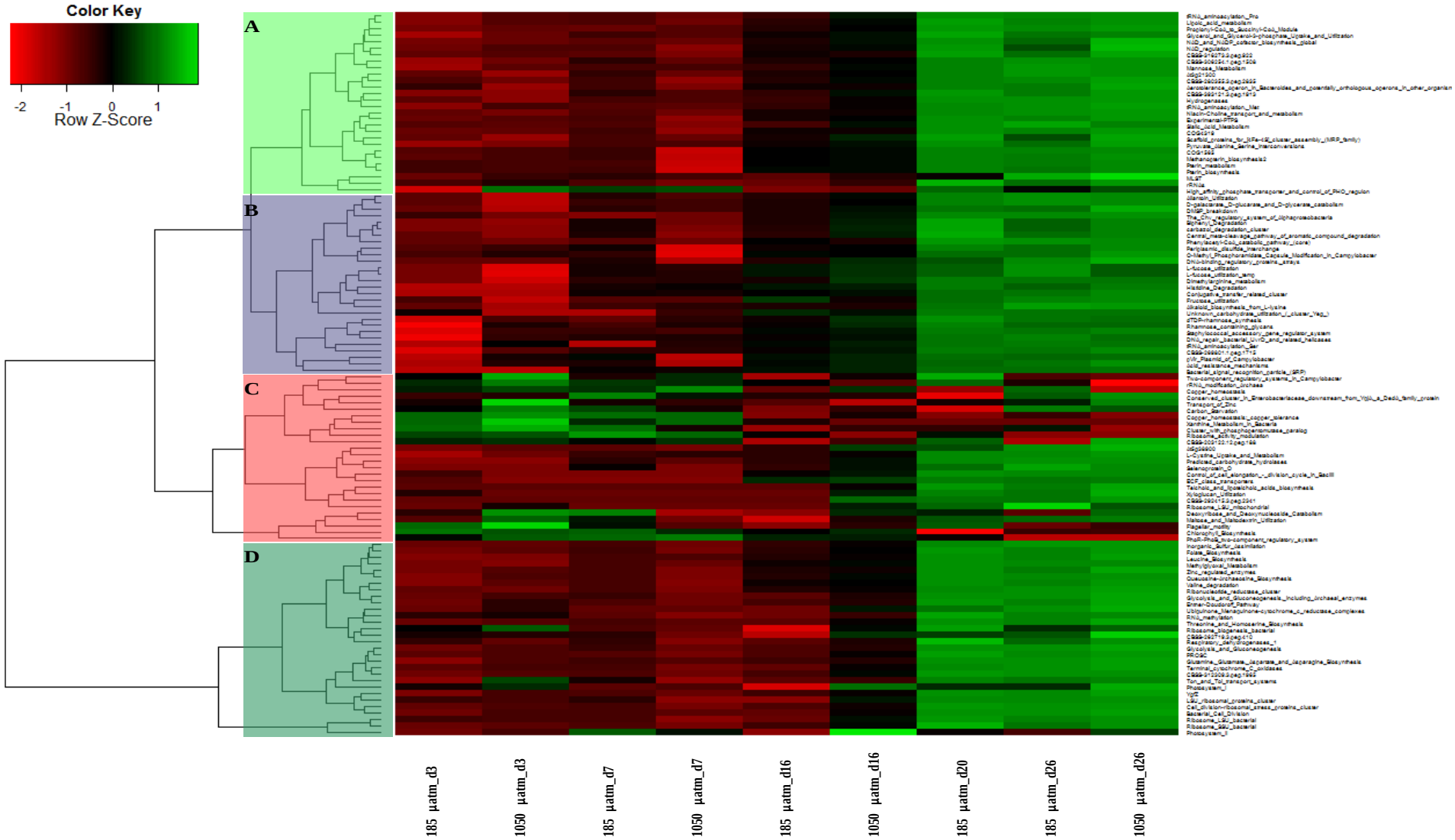


Figure S1 Heat map showing the four clusters of the metatranscriptome metabolic functions significantly ($p < 0.05$) correlated to CO_2; cluster **A**, **B**, **C** and **D** are presented in Figure 3, 4, 5 and 6, respectively, for better readability.

Table S1 Species (414) significantly differentially expressed under varying pCO_2 combined to the genus for table size reduction. The statistical variables $p < 0.05$ could not be presented in the figure as genera were combined in order to reduce table size. All genera presented are unique and all affected with a $p < 0.05$.

Domain	**Phylum**	**Class**	**Family**	**Genus**
Archaea	Euryarchaeota	Methanomicrobia	Methanospirillaceae	Methanospirillum
Bacteriaria	Actinobacteria	Actinobacteria	Actinomycetaceae	Arcanobacterium
Bacteriaria	Actinobacteria	Actinobacteria	Bifidobacteriaceae	Bifidobacterium
Bacteriaria	Actinobacteria	Actinobacteria	Bifidobacteriaceae	Gardnerella
Bacteriaria	Actinobacteria	Actinobacteria	Geodermatophilaceae	Geodermatophilus
Bacteriaria	Actinobacteria	Actinobacteria	Nocardiopsaceae	Thermobifida
Bacteriaria	Actinobacteria	Actinobacteria	Propionibacteriaceae	Propionibacterium
Bacteriaria	Actinobacteria	Actinobacteria	Sanguibacteraceae	Sanguibacter
Bacteriaria	Actinobacteria	Actinobacteria	Streptomycetaceae	Streptomyces
Bacteriaria	Bacteroidetes	Bacteroidia	Bacteroidaceae	Bacteroides
Bacteriaria	Bacteroidetes	Bacteroidia	Porphyromonadaceae	Porphyromonas
Bacteriaria	Bacteroidetes	Flavobacteriia	Flavobacteriaceae	Capnocytophaga
Bacteriaria	Bacteroidetes	Flavobacteriia	Flavobacteriaceae	Cellulophaga
Bacteriaria	Bacteroidetes	Flavobacteriia	Flavobacteriaceae	Flavobacterium
Bacteriaria	Chlamydiae	Chlamydiia	Chlamydiaceae	Chlamydia
Bacteriaria	Chlorobi	Chlorobia	Chlorobiaceae	Chlorobium
Bacteriaria	Chlorobi	Chlorobia	Chlorobiaceae	Prosthecochloris
Bacteriaria	Cyanobacteria	unclassified	Prochlorococcaceae	Prochlorococcus
Bacteriaria	Cyanobacteria	unclassified	unclassified	Synechococcus
Bacteriaria	Cyanobacteria	unclassified	unclassified	Synechocystis
Bacteriaria	Deferribacteres	Deferribacteres	Deferribacteraceae	Denitrovibrio
Bacteriaria	Deinococcus-Thermus	Deinococci	Trueperaceae	Truepera
Bacteriaria	Firmicutes	Bacilli	Alicyclobacillaceae	Alicyclobacillus

Bacteriaria	Firmicutes	Bacilli	Bacillaceae	Bacillus
Bacteriaria	Firmicutes	Bacilli	Bacillaceae	Geobacillus
Bacteriaria	Firmicutes	Bacilli	Bacillaceae	Halobacillus
Bacteriaria	Firmicutes	Bacilli	Bacillaceae	Lysinibacillus
Bacteriaria	Firmicutes	Bacilli	Enterococcaceae	Enterococcus
Bacteriaria	Firmicutes	Bacilli	Lactobacillaceae	Lactobacillus
Bacteriaria	Firmicutes	Bacilli	Leuconostocaceae	Oenococcus
Bacteriaria	Firmicutes	Bacilli	Listeriaceae	Listeria
Bacteriaria	Firmicutes	Bacilli	Paenibacillaceae	Paenibacillus
Bacteriaria	Firmicutes	Bacilli	Staphylococcaceae	Staphylococcus
Bacteriaria	Firmicutes	Bacilli	Streptococcaceae	Streptococcus
Bacteriaria	Firmicutes	Clostridia	Clostridiaceae	Clostridium
Bacteriaria	Firmicutes	Clostridia	Clostridiales Family XI. Incertae Sedis	Anaerococcus
Bacteriaria	Firmicutes	Clostridia	Clostridiales Family XI. Incertae Sedis	Finegoldia
Bacteriaria	Firmicutes	Clostridia	Eubacteriaceae	Eubacterium
Bacteriaria	Firmicutes	Clostridia	Lachnospiraceae	Roseburia
Bacteriaria	Firmicutes	Clostridia	Syntrophomonadaceae	Syntrophothermus
Bacteriaria	Firmicutes	Clostridia	Thermoanaerobacteraceae	Thermoanaerobacter
Bacteriaria	Firmicutes	Clostridia	Thermoanaerobacterales Family III	Thermoanaerobacterium
Bacteriaria	Fusobacteria	Fusobacteriia	Fusobacteriaceae	Fusobacterium
Bacteriaria	Planctomycetes	Planctomycetia	Planctomycetaceae	Rhodopirellula
Bacteriaria	Proteobacteria	Alphaproteobacteria	Acetobacteraceae	Acetobacter
Bacteriaria	Proteobacteria	Alphaproteobacteria	Anaplasmataceae	Ehrlichia
Bacteriaria	Proteobacteria	Alphaproteobacteria	Anaplasmataceae	Neorickettsia
Bacteriaria	Proteobacteria	Alphaproteobacteria	Anaplasmataceae	Wolbachia
Bacteriaria	Proteobacteria	Alphaproteobacteria	Aurantimonadaceae	Fulvimarina
Bacteriaria	Proteobacteria	Alphaproteobacteria	Bartonellaceae	Bartonella
Bacteriaria	Proteobacteria	Alphaproteobacteria	Beijerinckiaceae	Methylocapsa
Bacteriaria	Proteobacteria	Alphaproteobacteria	Bradyrhizobiaceae	Afipia
Bacteriaria	Proteobacteria	Alphaproteobacteria	Bradyrhizobiaceae	Nitrobacter

Bacteriaria	Proteobacteria	Alphaproteobacteria	Brucellaceae	Brucella
Bacteriaria	Proteobacteria	Alphaproteobacteria	Brucellaceae	Ochrobactrum
Bacteriaria	Proteobacteria	Alphaproteobacteria	Caulobacteraceae	Brevundimonas
Bacteriaria	Proteobacteria	Alphaproteobacteria	Methylobacteriaceae	Methylobacterium
Bacteriaria	Proteobacteria	Alphaproteobacteria	Rhizobiaceae	Agrobacterium
Bacteriaria	Proteobacteria	Alphaproteobacteria	Rhodobacteraceae	Paracoccus
Bacteriaria	Proteobacteria	Alphaproteobacteria	Rhodospirillaceae	Azospirillum
Bacteriaria	Proteobacteria	Alphaproteobacteria	Rickettsiaceae	Rickettsia
Bacteriaria	Proteobacteria	Betaproteobacteria	Burkholderiaceae	Burkholderia
Bacteriaria	Proteobacteria	Betaproteobacteria	Burkholderiaceae	Cupriavidus
Bacteriaria	Proteobacteria	Betaproteobacteria	Burkholderiaceae	Polynucleobacter
Bacteriaria	Proteobacteria	Betaproteobacteria	Burkholderiaceae	Ralstonia
Bacteriaria	Proteobacteria	Betaproteobacteria	Comamonadaceae	Acidovorax
Bacteriaria	Proteobacteria	Betaproteobacteria	Comamonadaceae	Comamonas
Bacteriaria	Proteobacteria	Betaproteobacteria	Comamonadaceae	Delftia
Bacteriaria	Proteobacteria	Betaproteobacteria	Comamonadaceae	Hydrogenophaga
Bacteriaria	Proteobacteria	Betaproteobacteria	Neisseriaceae	Chromobacterium
Bacteriaria	Proteobacteria	Betaproteobacteria	Neisseriaceae	Lutiella
Bacteriaria	Proteobacteria	Betaproteobacteria	Neisseriaceae	Neisseria
Bacteriaria	Proteobacteria	Betaproteobacteria	Nitrosomonadaceae	Nitrosospira
Bacteriaria	Proteobacteria	Betaproteobacteria	Oxalobacteraceae	Herbaspirillum
Bacteriaria	Proteobacteria	Betaproteobacteria	Rhodocyclaceae	Aromatoleum
Bacteriaria	Proteobacteria	Betaproteobacteria	unclassified	Methylibium
Bacteriaria	Proteobacteria	Deltaproteobacteria	Cystobacteraceae	Stigmatella
Bacteriaria	Proteobacteria	Deltaproteobacteria	Desulfohalobiaceae	Desulfohalobium
Bacteriaria	Proteobacteria	Deltaproteobacteria	Desulfuromonadaceae	Desulfuromonas
Bacteriaria	Proteobacteria	Deltaproteobacteria	Geobacteraceae	Geobacter
Bacteriaria	Proteobacteria	Deltaproteobacteria	Pelobacteraceae	Pelobacter
Bacteriaria	Proteobacteria	Epsilonproteobacteria	Campylobacteraceae	Arcobacter
Bacteriaria	Proteobacteria	Epsilonproteobacteria	Campylobacteraceae	Campylobacter

Bacteriaria	Proteobacteria	Epsilonproteobacteria	Helicobacteraceae	Helicobacter
Bacteriaria	Proteobacteria	Gammaproteobacteria	Acidithiobacillaceae	Acidithiobacillus
Bacteriaria	Proteobacteria	Gammaproteobacteria	Aeromonadaceae	Aeromonas
Bacteriaria	Proteobacteria	Gammaproteobacteria	Aeromonadaceae	Tolumonas
Bacteriaria	Proteobacteria	Gammaproteobacteria	Alteromonadaceae	Marinobacter
Bacteriaria	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Buchnera
Bacteriaria	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Erwinia
Bacteriaria	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Escherichia
Bacteriaria	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Salmonella
Bacteriaria	Proteobacteria	Gammaproteobacteria	Enterobacteriaceae	Shigella
Bacteriaria	Proteobacteria	Gammaproteobacteria	Francisellaceae	Francisella
Bacteriaria	Proteobacteria	Gammaproteobacteria	Halothiobacillaceae	Halothiobacillus
Bacteriaria	Proteobacteria	Gammaproteobacteria	Legionellaceae	Legionella
Bacteriaria	Proteobacteria	Gammaproteobacteria	Moraxellaceae	Acinetobacter
Bacteriaria	Proteobacteria	Gammaproteobacteria	Moraxellaceae	Psychrobacter
Bacteriaria	Proteobacteria	Gammaproteobacteria	Moritellaceae	Moritella
Bacteriaria	Proteobacteria	Gammaproteobacteria	Oceanospirillaceae	Oceanospirillum
Bacteriaria	Proteobacteria	Gammaproteobacteria	Pasteurellaceae	Actinobacillus
Bacteriaria	Proteobacteria	Gammaproteobacteria	Pasteurellaceae	Bibersteinia
Bacteriaria	Proteobacteria	Gammaproteobacteria	Pasteurellaceae	Haemophilus
Bacteriaria	Proteobacteria	Gammaproteobacteria	Piscirickettsiaceae	Methylophaga
Bacteriaria	Proteobacteria	Gammaproteobacteria	Piscirickettsiaceae	Thiomicrospira
Bacteriaria	Proteobacteria	Gammaproteobacteria	Pseudoalteromonadaceae	Pseudoalteromonas
Bacteriaria	Proteobacteria	Gammaproteobacteria	Pseudomonadaceae	Cellvibrio
Bacteriaria	Proteobacteria	Gammaproteobacteria	Pseudomonadaceae	Pseudomonas
Bacteriaria	Proteobacteria	Gammaproteobacteria	Shewanellaceae	Shewanella
Bacteriaria	Proteobacteria	Gammaproteobacteria	unclassified	Teredinibacter
Bacteriaria	Proteobacteria	Gammaproteobacteria	Vibrionaceae	Photobacterium
Bacteriaria	Proteobacteria	Gammaproteobacteria	Vibrionaceae	Vibrio
Bacteriaria	Proteobacteria	Gammaproteobacteria	Xanthomonadaceae	Xanthomonas

Bacteriaria	Proteobacteria	Gammaproteobacteria	Xanthomonadaceae	Xylella
Bacteriaria	Proteobacteria	unclassified	unclassified	unclassified
Bacteriaria	Spirochaetes	Spirochaetia	Spirochaetaceae	Borrelia
Bacteriaria	Spirochaetes	Spirochaetia	Spirochaetaceae	Treponema
Bacteriaria	Tenericutes	Mollicutes	Mycoplasmataceae	Mycoplasma
Bacteriaria	Tenericutes	Mollicutes	Mycoplasmataceae	Ureaplasma
Bacteriaria	Verrucomicrobia	unclassified	Methylacidiphilaceae	Methylacidiphilum
Bacteriaria	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiaceae	Akkermansia
Eukaryota	Apicomplexa	Aconoidasida	unclassified	Plasmodium
Eukaryota	Arthropoda	Insecta	Cecidomyiidae	Mayetiola
Eukaryota	Arthropoda	Insecta	Noctuidae	Spodoptera
Eukaryota	Ascomycota	Sordariomycetes	Sordariaceae	Neurospora
Eukaryota	Bacillariophyta	Mediophyceae	Eupodiscaceae	Odontella
Eukaryota	Chlorophyta	Prasinophyceae	unclassified	Pyramimonas
Eukaryota	Chlorophyta	Ulvophyceae	Ulvellaceae	Ochlochaete
Eukaryota	Chordata	Actinopterygii	Serranidae	Epinephelus
Eukaryota	Chytridiomycota	Chytridiomycetes	Spizellomycetaceae	Spizellomyces
Eukaryota	Nemertea	Anopla	Lineidae	Lineus
Eukaryota	Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigaceae	Piromyces
Eukaryota	Phaeophyceae	unclassified	Laminariaceae	Laminaria
Eukaryota	Streptophyta	Coniferopsida	Pinaceae	Keteleeria
Eukaryota	Streptophyta	Coniferopsida	Pinaceae	Pinus
Eukaryota	Streptophyta	Equisetopsida	Equisetaceae	Equisetum
Eukaryota	Streptophyta	Ginkgoopsida	Ginkgoaceae	Ginkgo
Eukaryota	Streptophyta	Jungermanniopsida	Frullaniaceae	Frullania
Eukaryota	Streptophyta	Jungermanniopsida	Metzgeriaceae	Metzgeria
Eukaryota	Streptophyta	Liliopsida	Arecaceae	Elaeis
Eukaryota	Streptophyta	Liliopsida	Dioscoreaceae	Dioscorea
Eukaryota	Streptophyta	Liliopsida	Musaceae	Musa
Eukaryota	Streptophyta	Liliopsida	Poaceae	Avena

Eukaryota	Streptophyta	Liliopsida	Poaceae	Oryza
Eukaryota	Streptophyta	Liliopsida	Smilacaceae	Smilax
Eukaryota	Streptophyta	Sphagnopsida	Sphagnaceae	Sphagnum
Eukaryota	Streptophyta	unclassified	Amaranthaceae	Atriplex
Eukaryota	Streptophyta	unclassified	Amaranthaceae	Beta
Eukaryota	Streptophyta	unclassified	Apiaceae	Anthriscus
Eukaryota	Streptophyta	unclassified	Apiaceae	Daucus
Eukaryota	Streptophyta	unclassified	Aquifoliaceae	Ilex
Eukaryota	Streptophyta	unclassified	Asteraceae	Conyza
Eukaryota	Streptophyta	unclassified	Convolvulaceae	Ipomoea
Eukaryota	Streptophyta	unclassified	Garryaceae	Aucuba
Eukaryota	Streptophyta	unclassified	Malvaceae	Gossypium
Eukaryota	Streptophyta	unclassified	Nymphaeaceae	Nuphar
Eukaryota	Streptophyta	unclassified	Rutaceae	Citrus
Eukaryota	Streptophyta	Zygnemophyceae	Desmidiaceae	Cosmocladium
Eukaryota	unclassified	Bangiophyceae	Bangiaceae	Porphyra
Eukaryota	unclassified	Dinophyceae	Gonyaulacaceae	Alexandrium
Eukaryota	unclassified	Dinophyceae	Gymnodiniaceae	Karenia
Eukaryota	unclassified	Dinophyceae	Pfiesteriaceae	Pfiesteria
Eukaryota	unclassified	Florideophyceae	Palmariaceae	Palmaria
Eukaryota	unclassified	unclassified	Mortierellaceae	Mortierella
Eukaryota	unclassified	unclassified	Saprolegniaceae	Achlya
Eukaryota	unclassified	unclassified	Trypanosomatidae	Leishmania
Eukaryota	unclassified	unclassified	unclassified	Phytophthora
Unclassified	unclassified	unclassified	unclassified	unclassified
Viruses	unclassified	unclassified	Microviridae	Microvirus
Viruses	unclassified	unclassified	Podoviridae	Bpp-1-like viruses
Viruses	unclassified	unclassified	Podoviridae	T7-like viruses
Viruses	unclassified	unclassified	Poxviridae	Capripoxvirus

Table S2 Seven significantly differentially expressed ($p < 0.05$) species under varying pCO_2

Phylum	Class	Family	Genus	Species	Prob	Bonferroni corrected	FDR corrected
Bacteroidetes	Flavobacteriia	Flavobacteriaceae	Ornithobacterium	Ornithobacterium rhinotracheale	0.004	1.887	1.887
Proteobacteria	Gammaproteo-bacteria	Pasteurellaceae	Pasteurella	Pasteurella multocida	0.011	4.996	2.498
Proteobacteria	Gammaproteo-bacteria	Pseudoalteromo-nadaceae	Pseudoalteromonas	Pseudoalteromonas sp. KMM 701	0.011	4.996	1.665
not ID	not ID	not ID	not ID	not ID	0.011	4.996	1.249
unclassified	unclassified	unclassified	unclassified	uncultured marine microorganism	0.012	5.419	1.084
Proteobacteria	Gammaproteo-bacteria	Moritellaceae	Moritella	Moritella profunda	0.027	12.021	2.003
Bacteroidetes	Cytophagia	Cytophagaceae	Cytophaga	Cytophaga sp. T-588	0.046	20.514	2.931

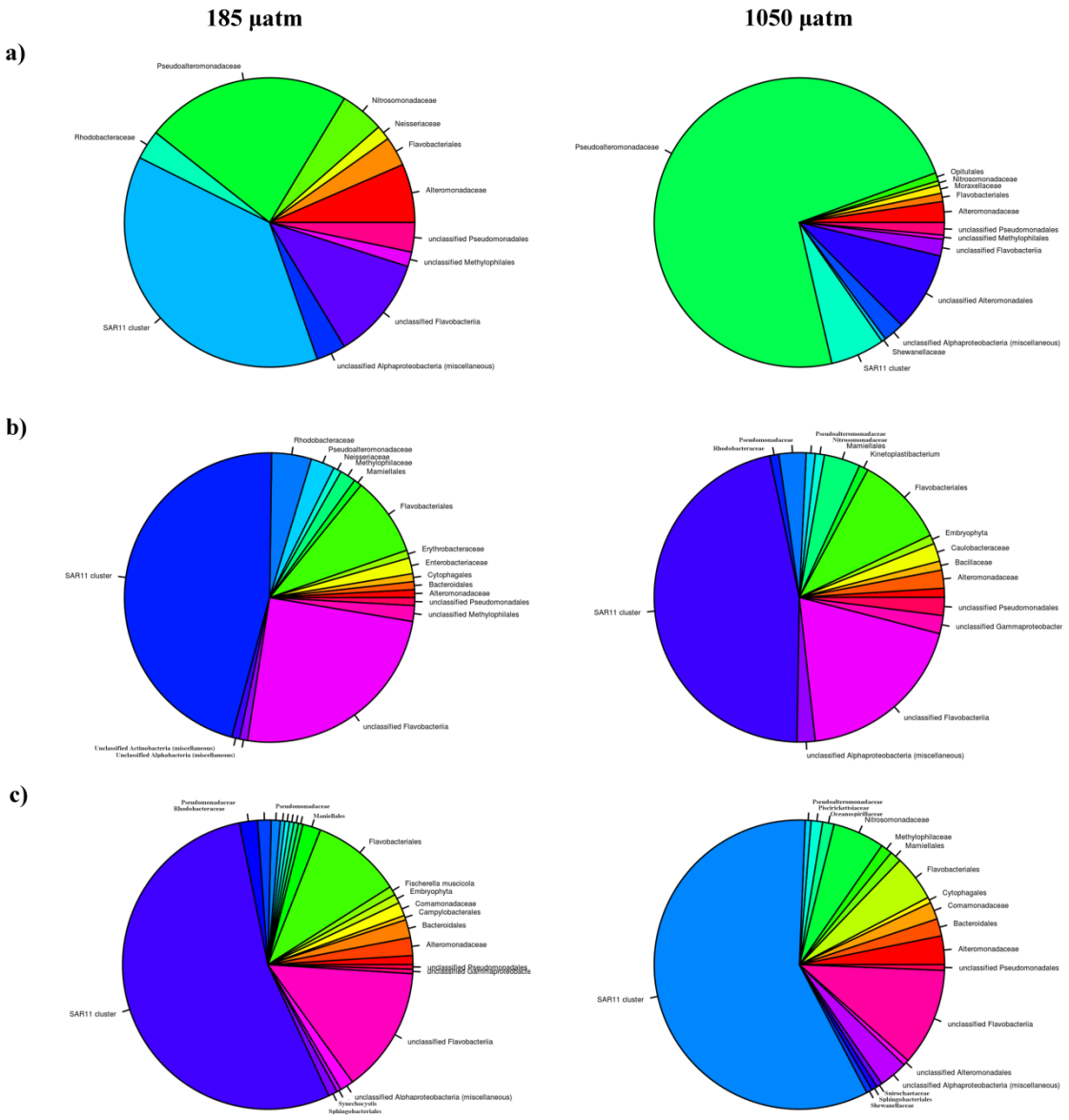


Figure S2 Taxa (represented in %) expressing significantly differentially expressed metabolic function acid resistance from the metagenome on a) d3, b) d16 and c) d20 under two pCO_2. Metagenomic libraries for d7 are not presented as sequencing failed. Colors are not taxa-specific.

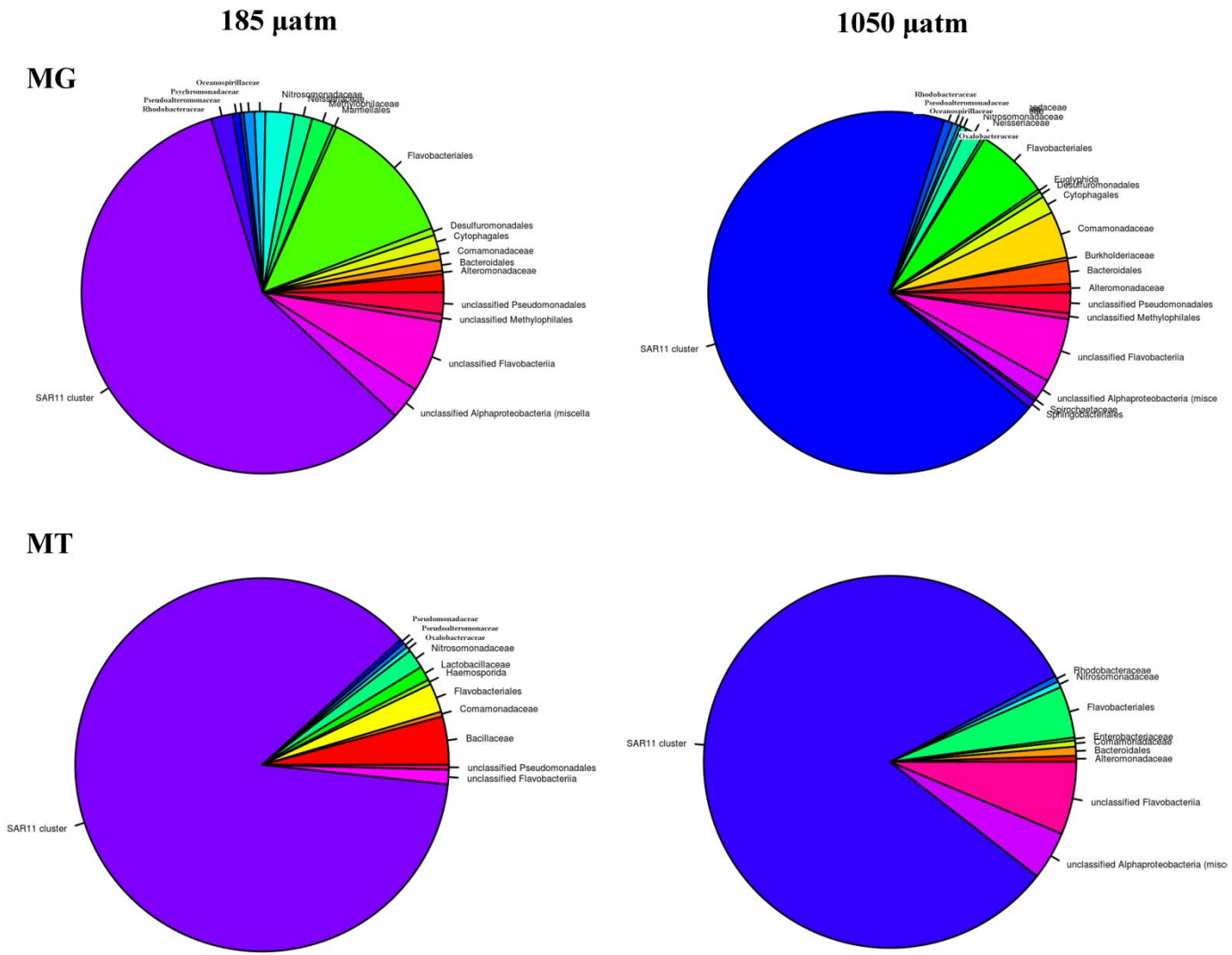


Figure S3 Taxa (represented in %) expressing significantly differentially expressed metabolic function acid resistance from the metagenome (MG) versus the metatranscriptome (MT) on d26 on under two pCO_2. Colors are not taxa-specific.

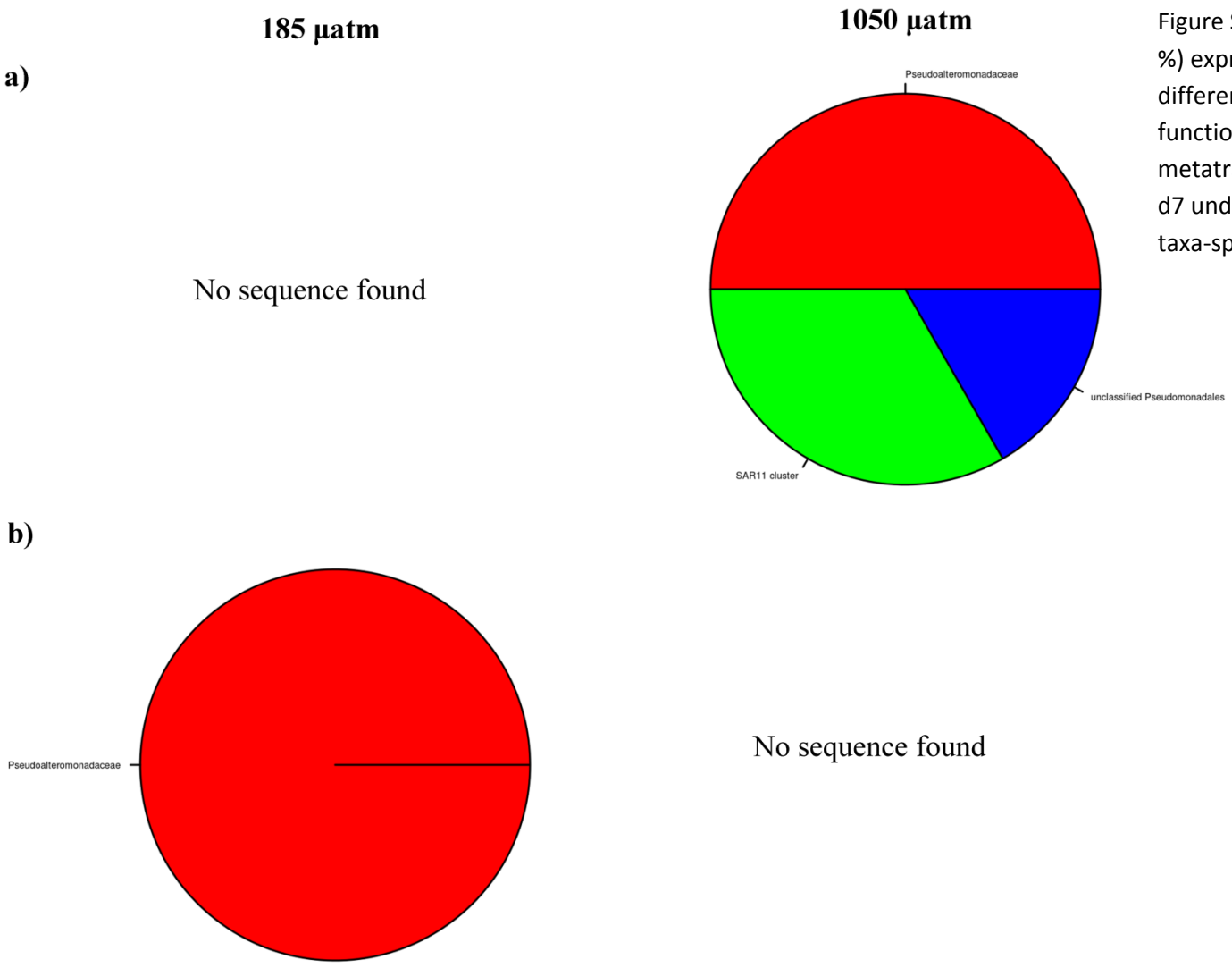
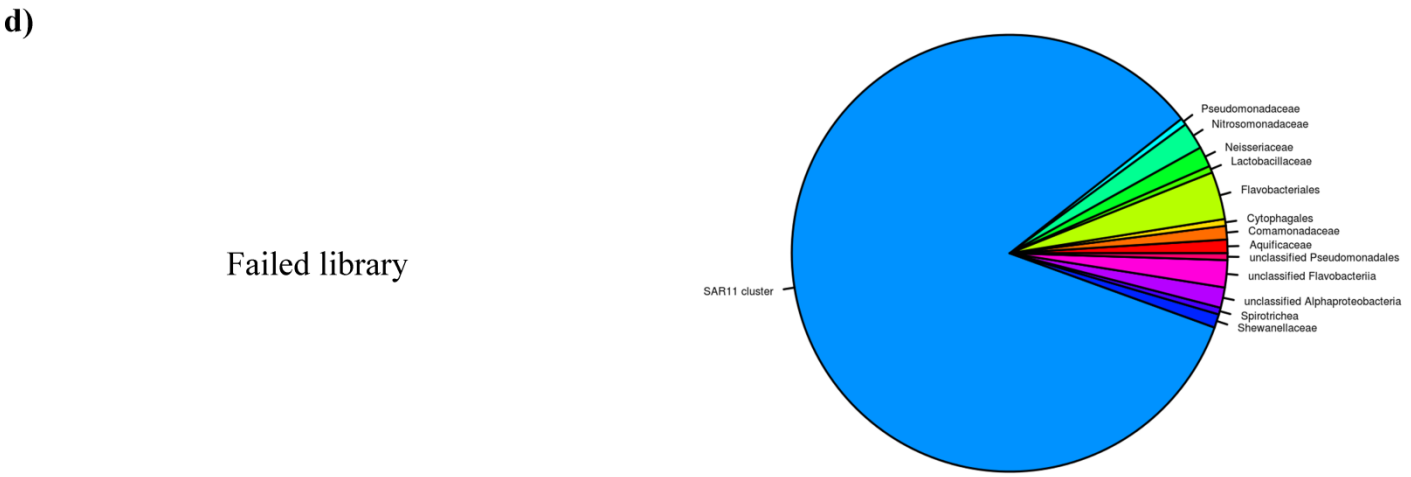


Figure S4a Taxa (represented in %) expressing significantly differentially expressed metabolic function acid resistance from the metatranscriptome on a) d3 and b) d7 under two pCO_2. Colors are not taxa-specific.

Figure S4b Taxa (represented in %) expressing significantly differentially expressed metabolic function acid resistance from the metatranscriptome on c) d16 and b) d20 under two pCO_2. Metatranscriptomic libraries for d20-185 µatm is not presented as sequencing failed. Colors are not taxa-specific.



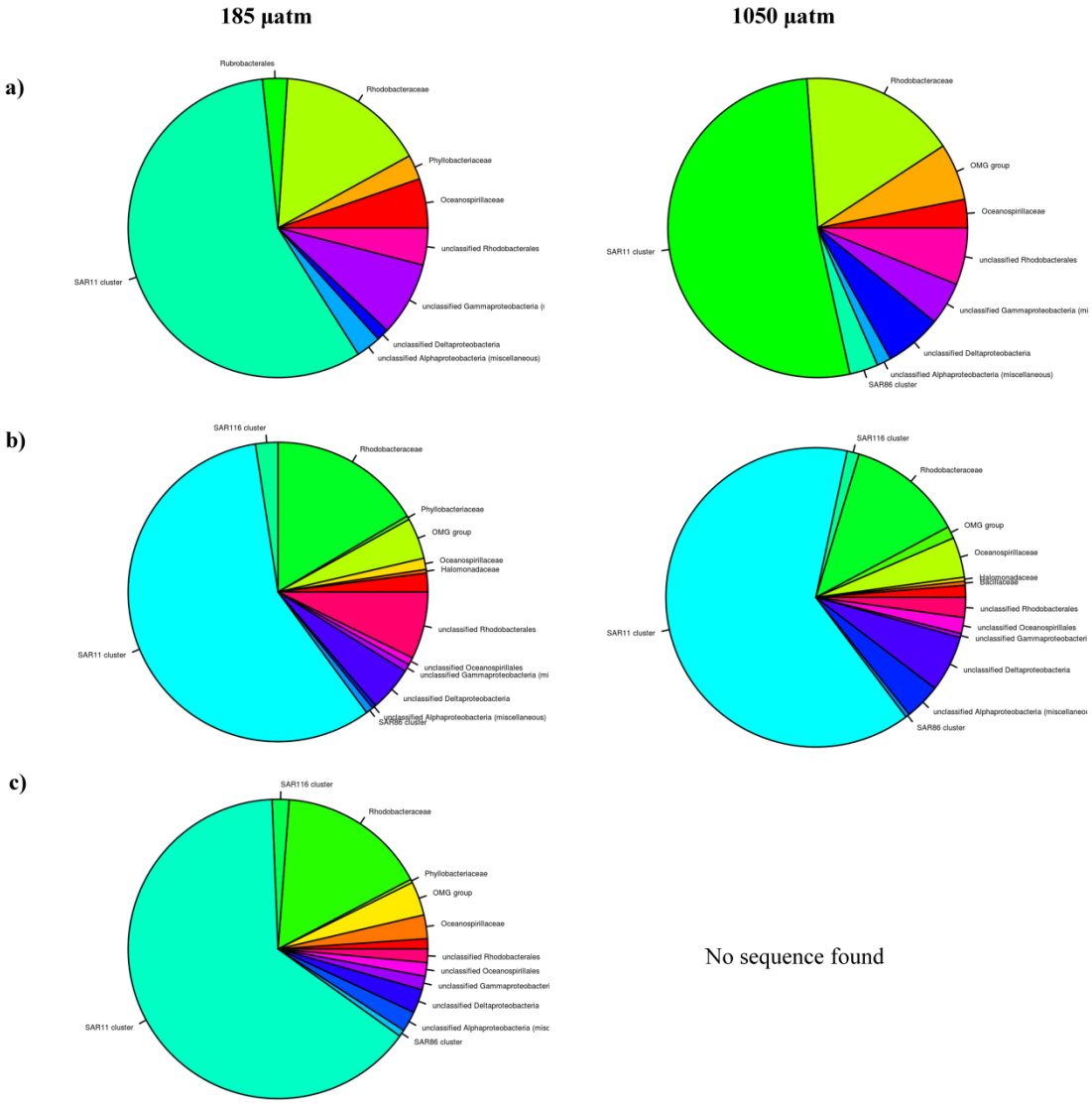


Figure S5 Taxa (represented in %) expressing significantly differentially expressed metabolic function DMSP breakdown from the metagenome on a) d3, b) d16 and c) d20 under two pCO_2. Colors are not taxa-specific.

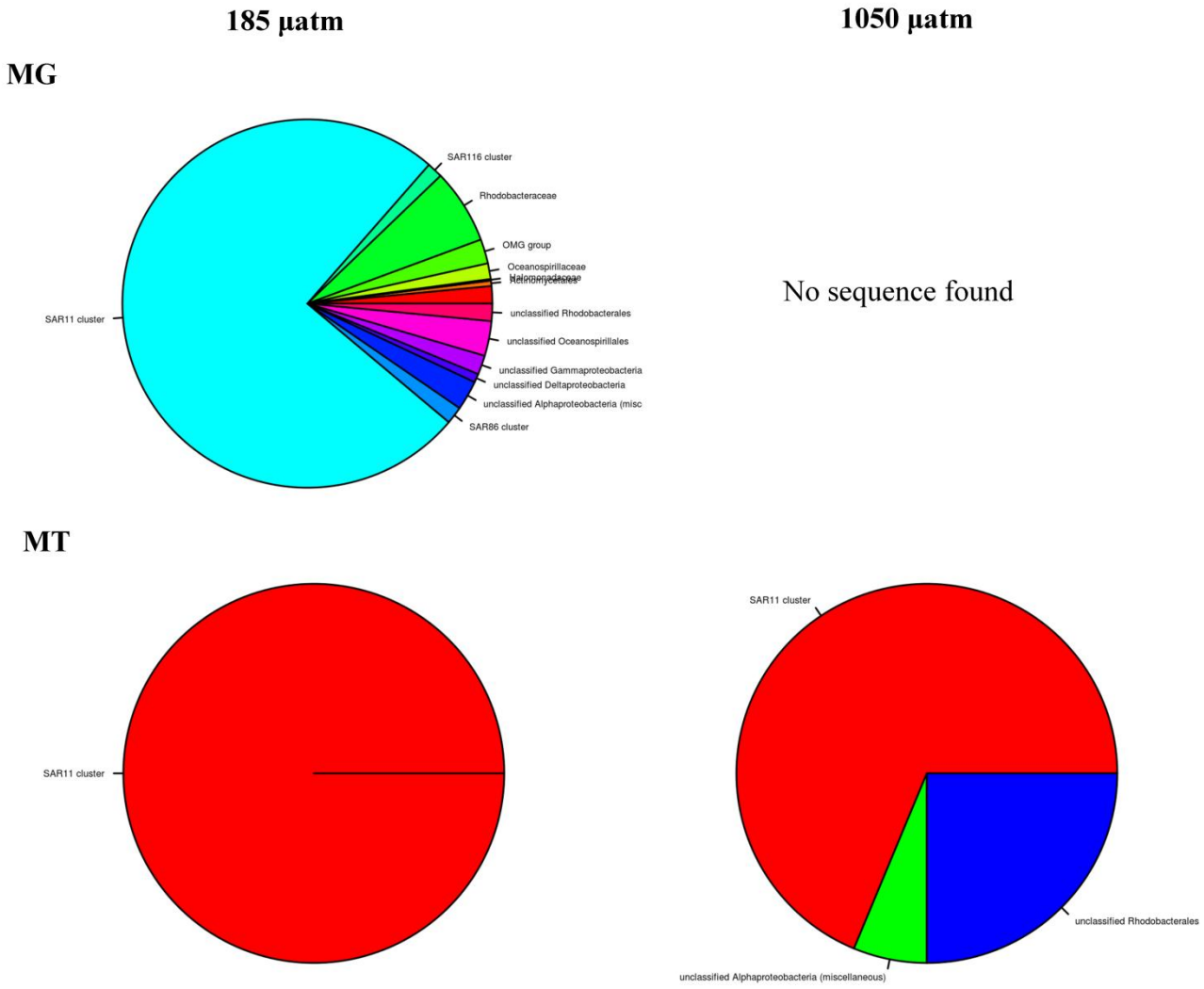


Figure S6 Taxa (represented in %) expressing significantly differentially expressed metabolic function DMSP breakdown from the metagenome (MG) versus the metatranscriptome (MT) on d26 under two pCO_2. Colors are not taxa-specific.

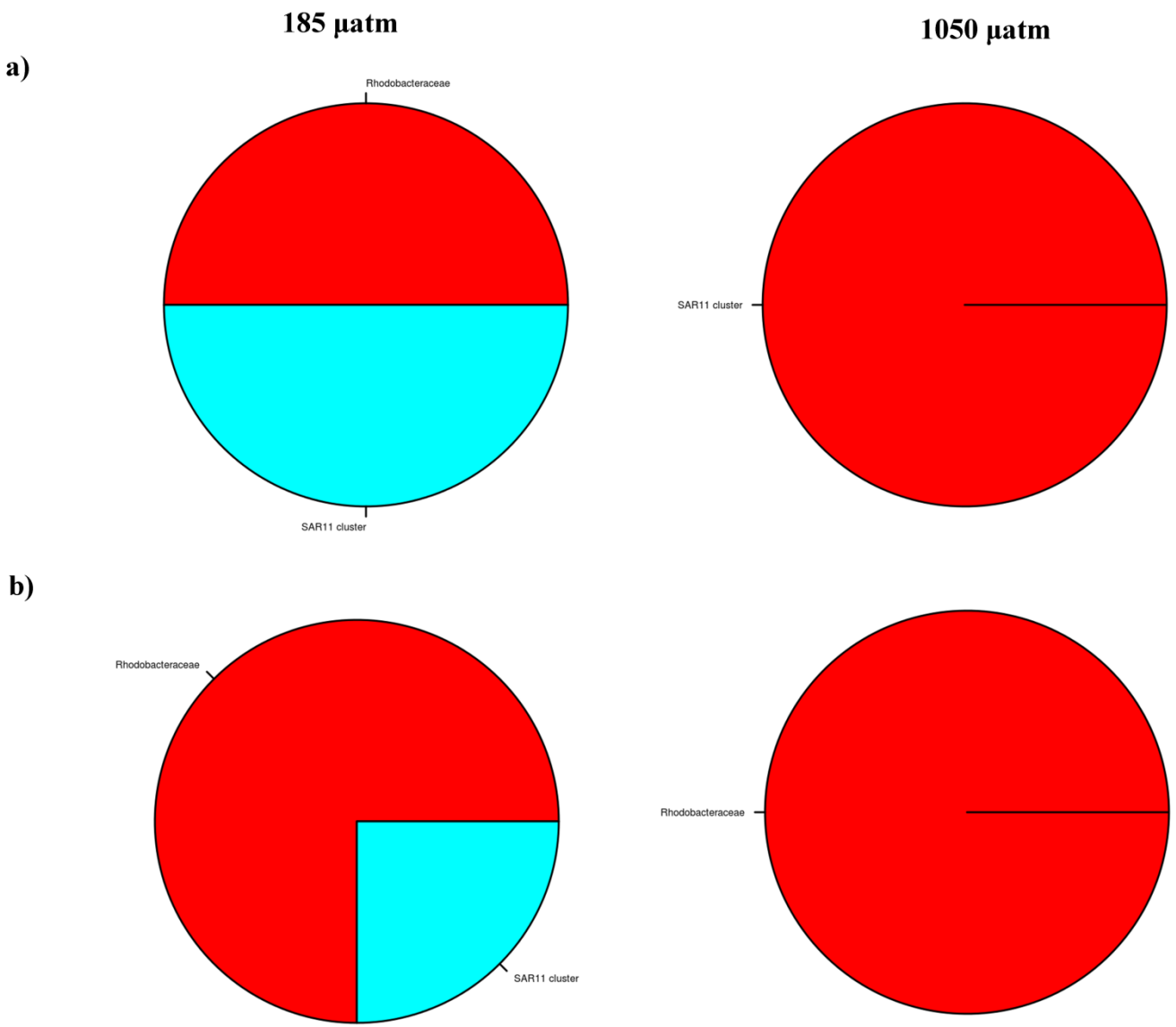


Figure S7a Taxa (represented in %) expressing significantly differentially expressed metabolic function DMSP breakdown from the metatranscriptome on a) d3 and b) d7 under two pCO_2. Colors are not taxa-specific.

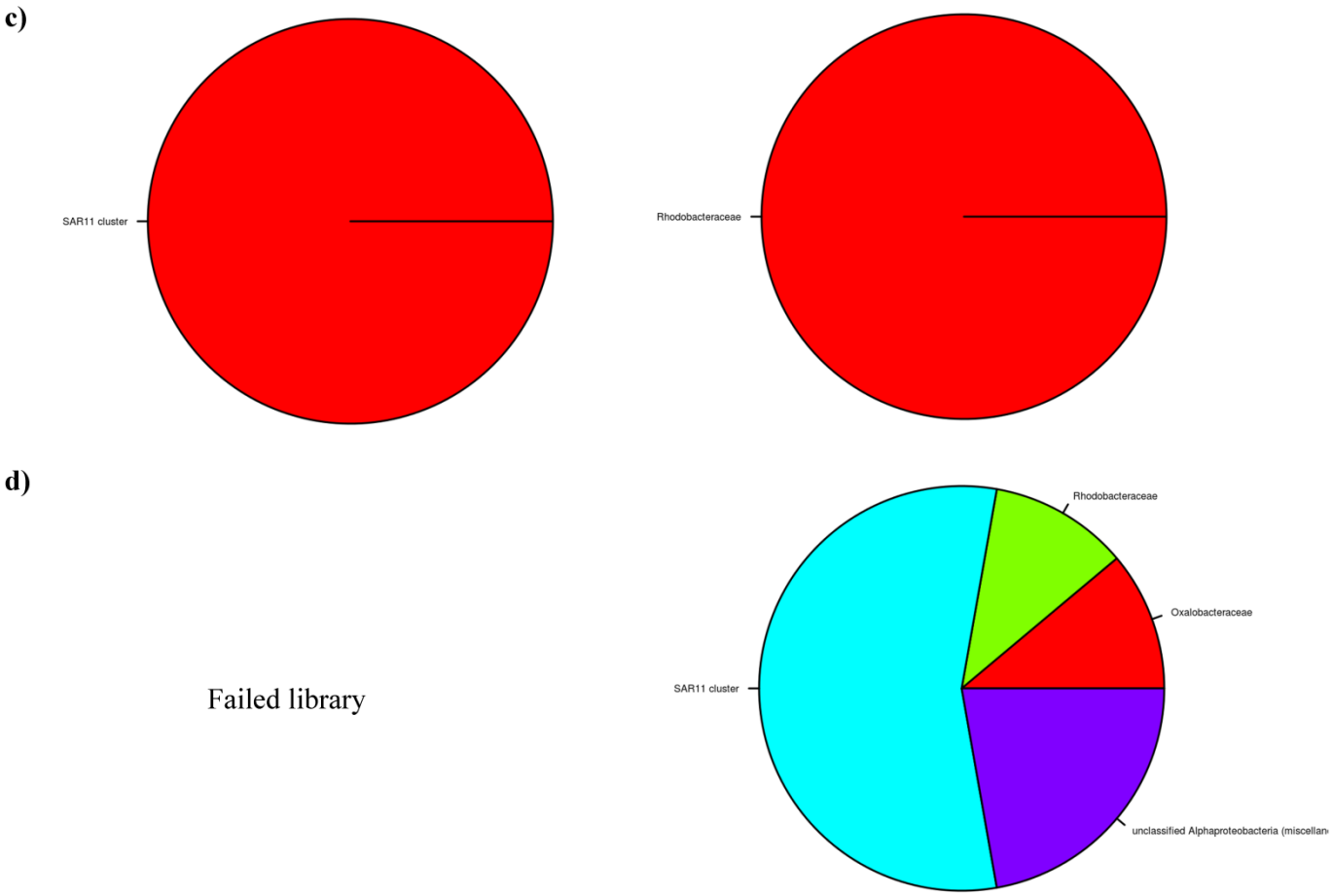


Figure S7b Taxa (represented in %) expressing significantly differentially expressed metabolic function DMSP breakdown from the metatranscriptome on c) d16 and b) d20 under two pCO_2. Metatranscriptomic libraries for d20-185 µatm is not presented as sequencing failed. Colors are not taxa-specific.

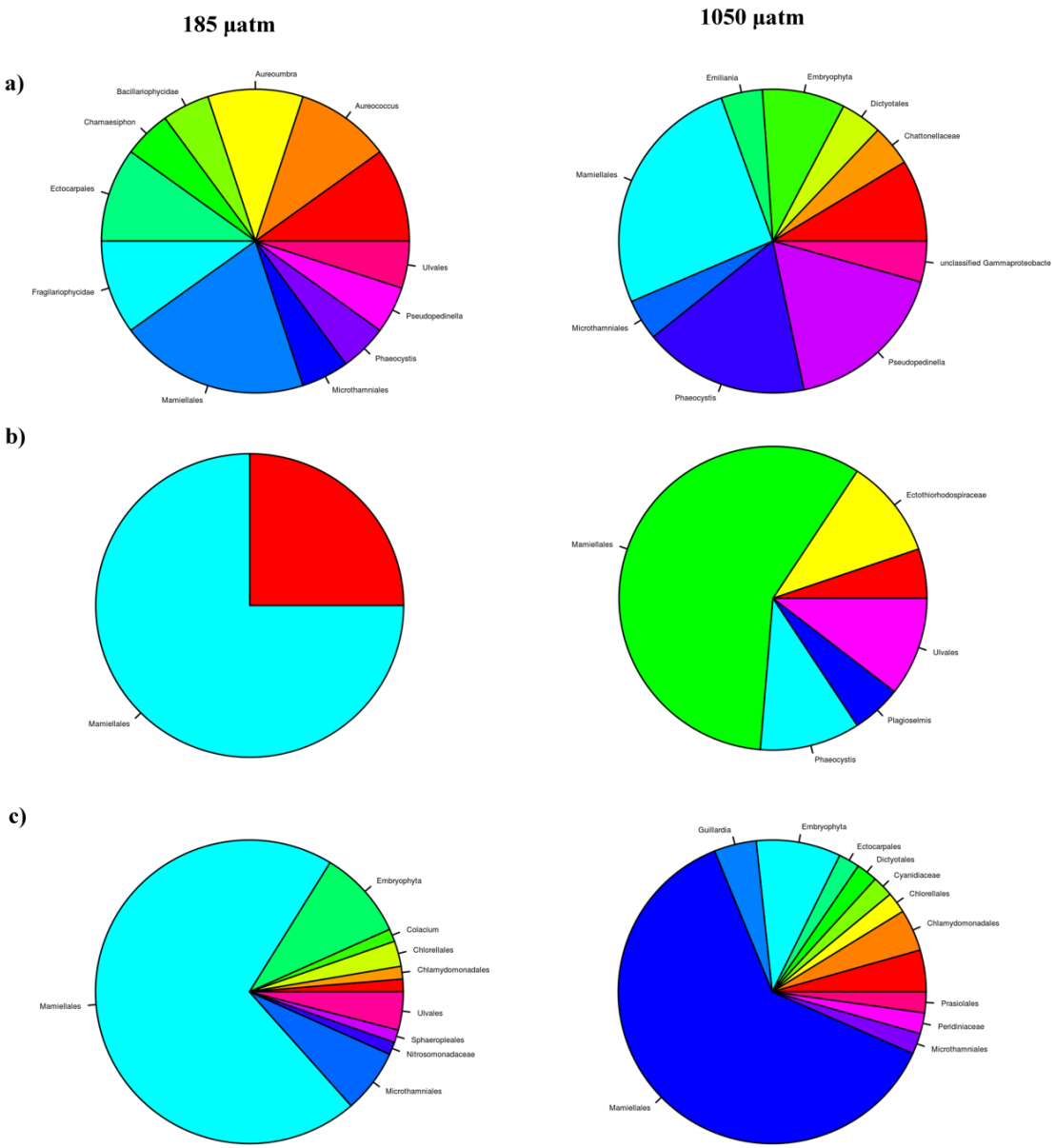


Figure S8 Taxa (represented in %) expressing significantly differentially expressed metabolic function RuBisCO from the metagenome on a) d3, b) d16 and c) d20 under two pCO_2. Metagenomic libraries for d7 are not presented as sequencing failed. Colors are not taxa-specific.

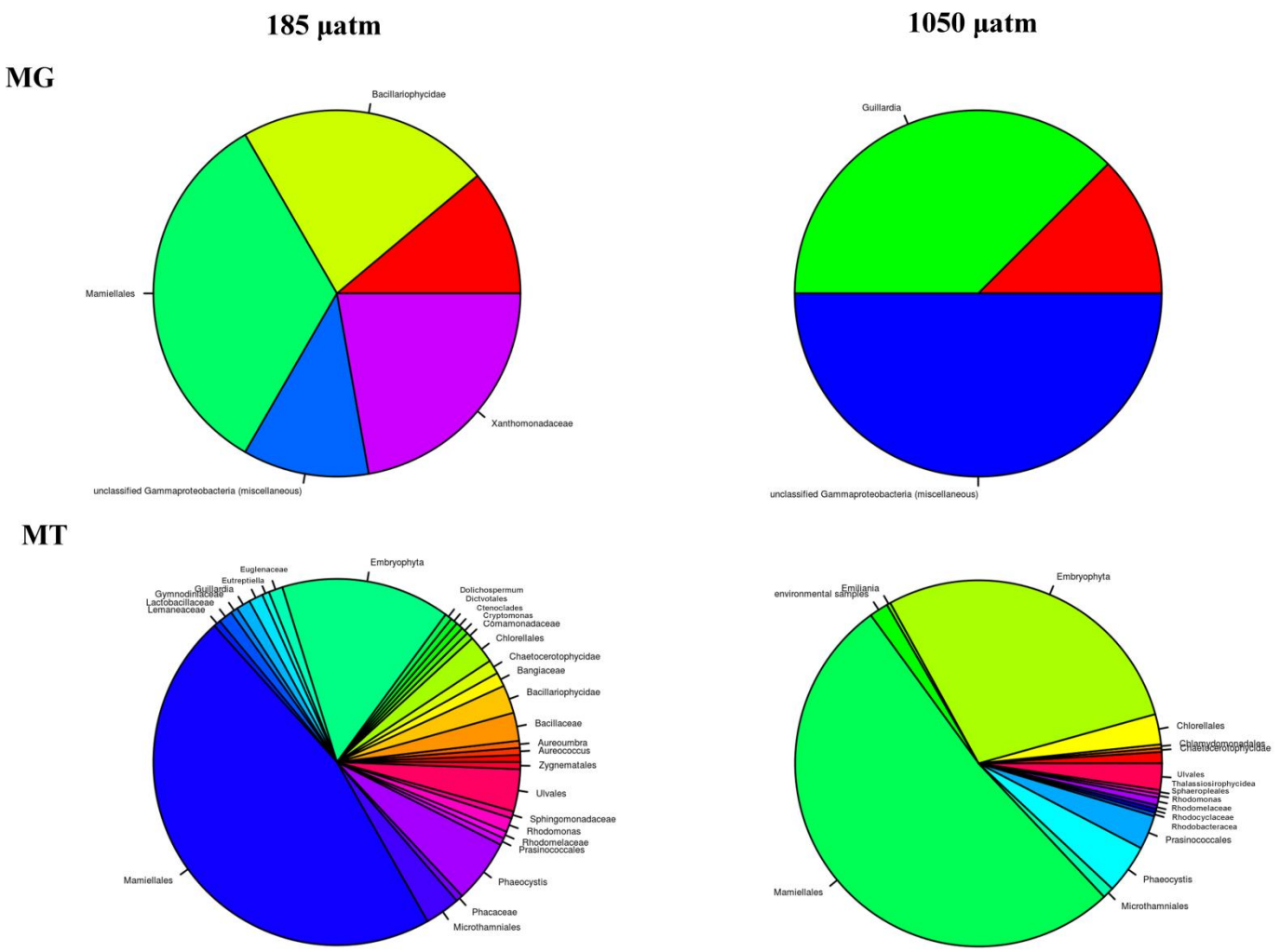


Figure S9 Taxa (represented in %) expressing significantly differentially expressed metabolic function RuBisCO from the metagenome (MG) versus the metatranscriptome (MT) on d26 under two pCO_2. Colors are not taxa-specific.

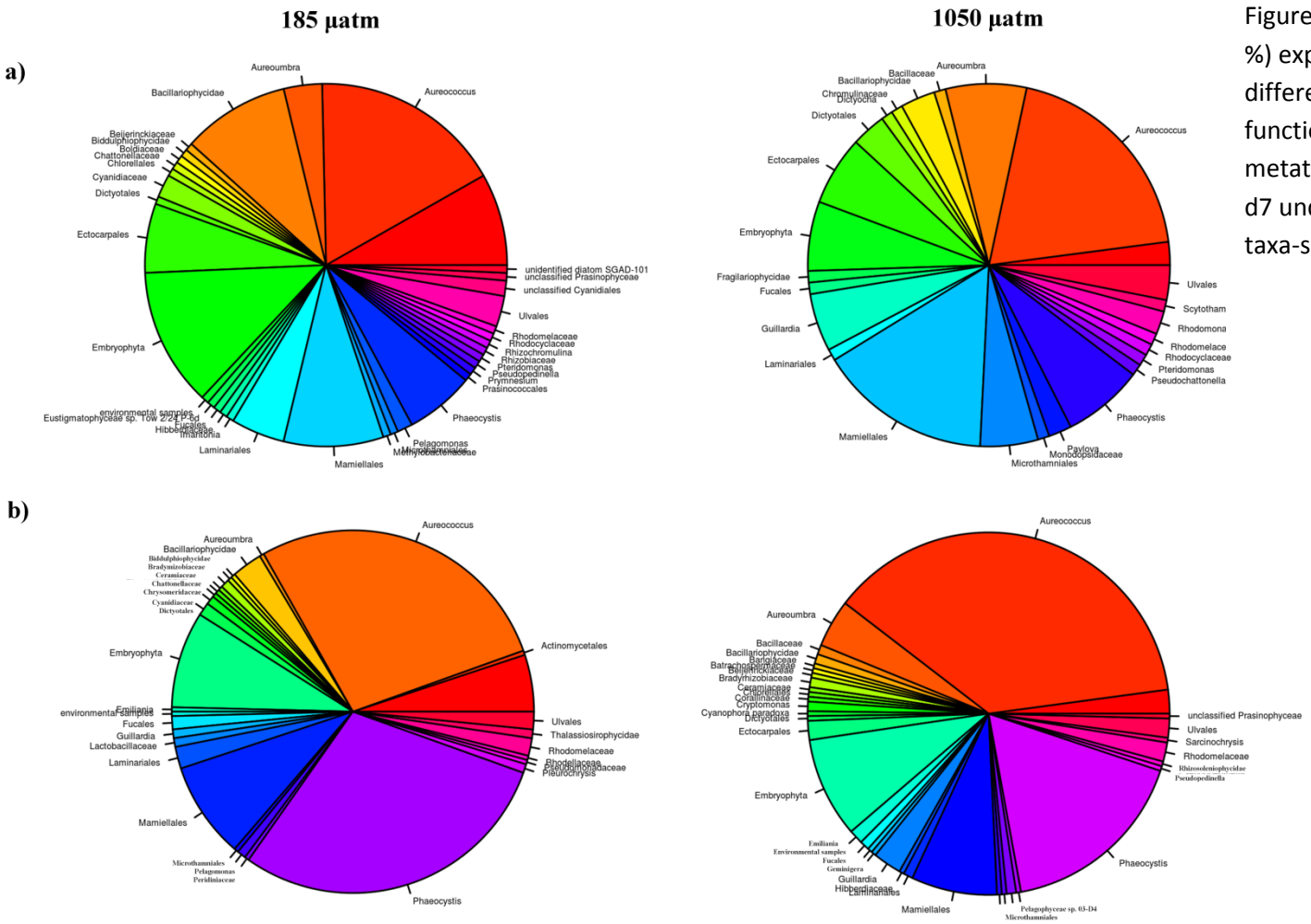


Figure S10a Taxa (represented in %) expressing significantly differentially expressed metabolic function RuBisCO from the metatranscriptome on a) d3 and b) d7 under two pCO_2. Colors are not taxa-specific.

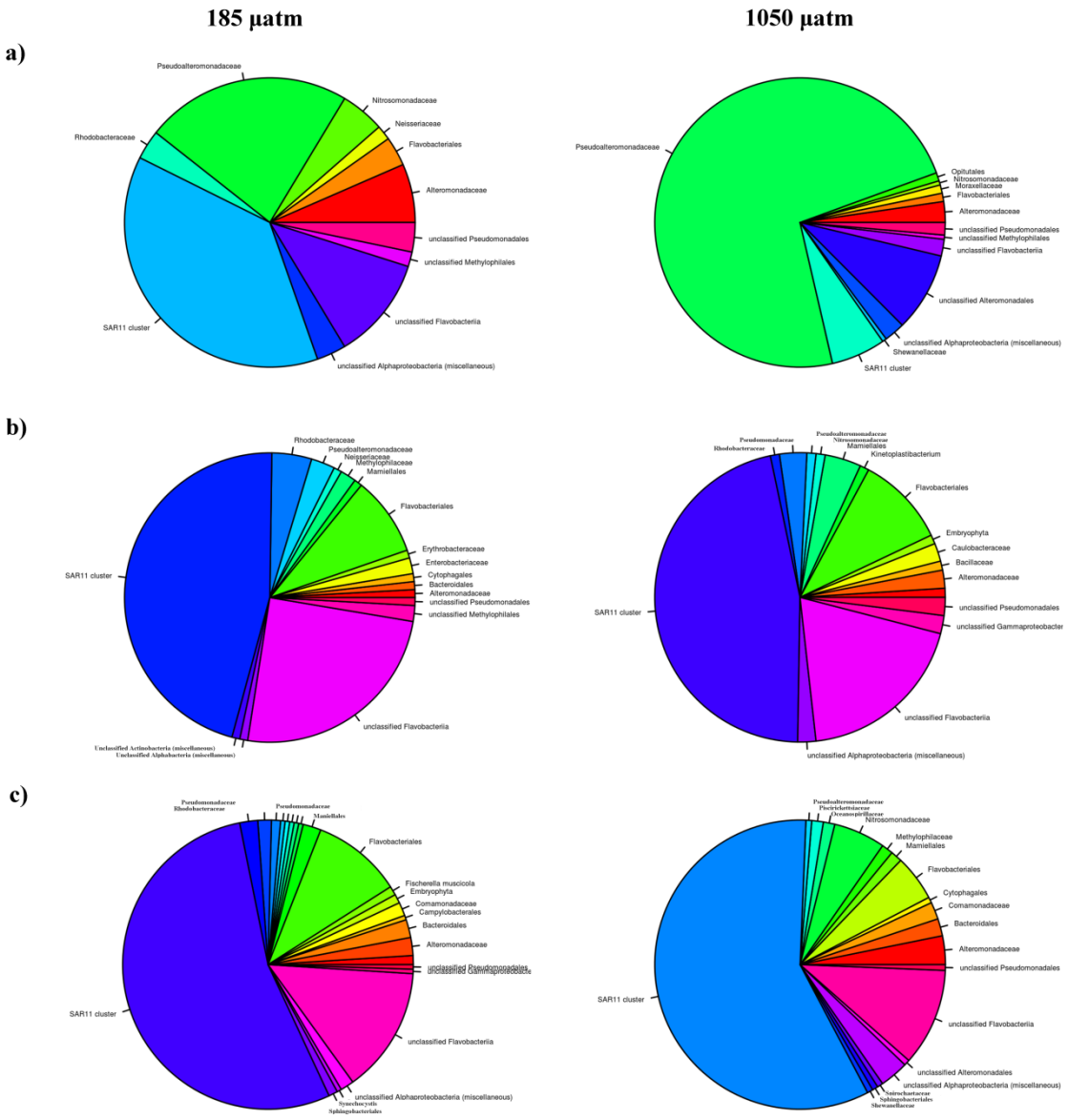


Figure S10a Taxa (represented in %) expressing significantly differentially expressed metabolic function RuBisCO from the metatranscriptome on a) d16 and b) d20 under two pCO_2. Colors are not taxa-specific.

5. Synthesis and future perspectives

5. Synthesis and future perspective

The results presented in this dissertation provide a significant step toward understanding the effect of OA on cultured *T. oceanica* and on whole natural Arctic microbial communities using multiple molecular biological and sequencing techniques (qPCR, Sanger, 454 and Illumina sequencing).

High-throughput sequencing technologies have been established as powerful descriptive and analytical tools for the characterization of community structures and their function in their respective environments (Frias-Lopez *et al.*, 2008; Canfield *et al.*, 2010; Gilbert *et al.*, 2011; Stewart *et al.*, 2011; Thomas *et al.*, 2012; Schunck *et al.*, 2013). The first part of the work presented in this dissertation (co-authorship manuscript I - Lommer *et al.*, 2010 and II - , Lommer *et al.*, 2012) was done as a collaborative effort to obtain genome information on one ecologically relevant diatom in oligotrophic oceans. The importance of diatoms in the carbon cycle (primary producers) justifies the acquisition of diatom genome information. In particular, *T. oceanica* is highly tolerant to limiting Fe conditions. The discovery that ferredoxin was relocated in one strain of *T. oceanica* (CCMP1005) but not in *T. pseudonana* (CCMP1335) nor in *T. weissflogii* (CCMP1052) was used to speculate that the transfer of the ferredoxin gene to the nuclear genome may in fact be an adaptation to low Fe conditions in *T. oceanica*. The relocation of the ferredoxin gene was confirmed in all tested *T. oceanica* strains but not in other strains of other closely related *Thalassiosira* (via gDNA qPCR assay of single copy nuclear and chloroplastic genes) (manuscript III). The experiment determined that the relocation happened in all tested strains of *T. oceanica* (CCMP0999, 1001, 1005, 1006) while none of the tested strains of *T. pseudonana* (CCMP1012, 1013, 1014, 1335) nor of *T. weissflogii* (CCMP1010, 1049, 1052) underwent such an event; confirming that this relocation happened after species differentiation.

The newly obtained *T. oceanica* genome and its annotation through the replete/deplete Fe transcriptomes were considered as first step into observing OA effect on *Thalassiosira* species as the bioavailability of Fe decreases with acidification of the oceans (Shi *et al.*, 2010). The Fe studies (Lommer *et al.*, 2010; Lommer *et al.*, 2012) set the basic laboratory skills needed for the manipulation experiments of CO_2 in *T. oceanica*'s cultures respecting the guidelines on ocean acidification (Riebesell *et al.*, 2010). The Illumina sequencing resulted in more detailed transcriptomes than the Fe-manipulated ones as the Illumina technology produces an

increased sequence number, resulting in a more comprehensive overview of *T. oceanica*. Multiple protein-coding genes were significantly affected by CO_2 (manuscript IV) with the most interesting up-regulated-genes under 380 µatm annotated pol proteins, cytoplasmic dynein heavy proteisn and myosin proteins while fucoxanthin proteins, TPR-SEL1 proteins and NNT NAD proteins under 1000 µatm. The transcriptome gives detailed information on the functional diversity of *T. oceanica*, and although many of the genes with functions and roles identified as affected by CO_2 in the calcifiers *Emiliania huxleyi* (Mackinder *et al.*, 2011; Lohbeck *et al.*, 2012; Bach *et al.*, 2013) they were not significantly influenced here. Calcifiers are susceptible to OA because of their external calcium carbonate structures however diatoms do not appear to be strongly affected by the increase of pCO_2 (Wu *et al.*, 2010; Crawfurd *et al.*, 2011). Nevertheless, it is essential to establish how OA will affect diatoms as they are one of the most important groups of primary producers. The work presented here provides an important list of genes that are responsive to CO_2 levels but for which the function is still unknown. These genes and their functional characterization should be the focus of further studies because they might provide important adaptation or change in metabolism for diatoms as OA continues to increase. In the future, we intend to look at the most significantly affected unknown genes and search the neighbouring genes for indication of their roles in the genome and their genomic context. Another future analysis planned, is the reassembly of the *T. oceanica* genome as Illumina sequencing gives a higher number of shorter reads (compared to 454) and could possibly produce a better assembly resolution.

Cultures experiment such as those described in manuscript I, II, III and IV are useful but not always representative of the behaviour of a species in its natural environmental conditions. Mesocosms experiments, like the EPOCA 2010 Svalbard campaign, were designed to complement cultures findings. The Svalbard mesocosms experiment represents the first study on an Arctic community where growth and procession were observed over an extended period of time and under multiple pCO_2. The samples dedicated to sequencing were prepared according to the established RNA sequencing with filtration and conservation of the samples within a short period of time. Long filtration time of environmental samples has been directly linked to small fragile RNA degradation, which biases the determination of the microbial community structure (Gilbert *et al.*, 2010). The sample filtration time during the Svalbard experiment was kept to the shortest feasible given that the location of some mesocosms in relation to the laboratory made it difficult to respect the accepted 20-minute filtration time for the analysis of transcriptomic. This time period was sometimes increased to 30 minutes due to

difficult logistics. In the future, designing portable filtration systems allowing filtration directly at the water collection site or preservation techniques at the sampling site may help resolve these issues.

The first metatranscriptomic study focussing on environmental microbial functions was carried out by Poretsky and colleagues in 2005 on the Sanger sequencer (Poretsky *et al.*, 2005). They established the first RNA samples preparation protocol for such studies and were the first to directly target mRNA isolated from a natural environment. Our studies protocol was based on the improved Poretsky protocol (Poretsky *et al.*, 2005; Poretsky *et al.*, 2009; Schunck *et al.*, 2013). Following the updated protocol, prokaryotic rRNA was removed as in Schunck *et al.* (2013) and the removal of eukaryotic rRNA was tested prior to the analysis of the mesocosm samples and resulted in removal of approximately 55-60% of the ribosomal sequences. The removal of the ribosomal RNA was successful and the resulting amplification of the mRNA produced successfully 17 high quality libraries from the 20 samples selected from the 3-12 µm fraction.

The Svalbard mesocosms experiment was conducted according to the detailed guidelines on experimental work on OA (Riebesell *et al.*, 2010) and accurately mimicked OA conditions. It involved multiple parameters believed to be necessary in understanding the Arctic ecological responses to future climate changes (Biogeosciences special issue "Arctic ocean acidification: pelagic ecosystem and biochemical responses during a measocosm study"). The first sequencing analysis of the Svalbard mesocosms was done using 16S rRNA V4 amplicon (publication V). The 16S rRNA gene is known for its high degree of accuracy in classifying bacteria, even from rare or uncharacterized clades (Song *et al.*, 2005). A large number of 16S rRNA gene sequences from the environment and from culture collections are available on public databases like SILVA, Greengenes, and NCBI. The taxonomic annotation of unknown sequences from the environment is limited by the quality of our databases. In the present experiment, the Greengenes 16S rRNA database was used as a reference database for alignment construction. Taxonomic annotations were determined using the RDP classifier, trained on Greengenes reference OTUs (at 97% sequence identity). This study highlighted the fact that CO_2 was not significantly structuring the dominant members of the microbial community but had a significant effect on 15 rare taxa. The bacterial community was structured mainly by time, origin of the sample (Fjord versus mesocosms) and nutrient addition. Others parameters, sampling strategies (0.2-3 µm versus 3-12 µm filters' fraction),

silicate, phosphate, primary production, temperature and pH also participated in sculpting the community. The overall dominating taxa were the alpha-, gamma-proteobacteria and Bacteroidetes in both fractions; additionally, the Cyanobacteria and Eukaryotes were found to be a large part of the 0.3-12 µm fraction before nutrient addition. In conclusion, the effect of CO_2 on the microbial community composition was limited but could still have some evolutionary consequences by creating a shift in the taxa dominance and/or diversity affecting the entire community structure by for example favouring some rare taxa. At the time of publication of our study, similar results were revealed by Zhang for the same mesocosms (Zhang *et al.*, 2013). Following the 16S rRNA analysis, the metagenome and metatranscriptome of two mesocosms were used to compare and deepen the prior results on OA (manuscript VI) by extending the analysis to biochemical function rather than limiting it to phylogenetic composition. The analyses of total metagenome confirmed the 16S rRNA results, also indicating that no major effects of OA were found on the microbial community structure except for a few rare or low-abundance taxa. Some genes coding for different metabolic functions were also significantly differentially expressed by CO_2 and the impact on the community is presently being evaluated via the analysis of the taxa significantly expressing these metabolic functions. It will further be useful to observe whether the anticipated shift in community happened or not. The importance of compiling the different parameters measured during the Svalbard's study was accentuated by Archer *et al.*'s (2013) suggestion that *Heterocapsa rotundata*'s played an important role in the DMSP cycle and its metagenomic/metatranscriptomic profile (manuscript VI). Archer *et al.* (2013) found, using flow cytometry, that there was a strong dependence between the abundance of *H. rotundata* and the DMSP and concluded that dinoflagellates were responsible for the DMSP increase. Our results show that this species was in very low abundance throughout the experiment but had an increasing abundance in transcripts level after nutrient addition possibly representing Archer et al. (2013) findings. The metagenome and metatranscriptome are further being used to establish if dinoflagellates were indeed the main taxa containing enzymes involved in DMSP production. The roles of the few taxa significantly affected by pCO_2 are important to evaluate as they appear to be the most prone to the effect of OA (Roy *et al.*, 2013). The taxa actually expressing the functions and enzymes significantly differentially expressed under different CO_2 are presently being investigated in order to highlight more connection between community and OA affected parameters.

Finally, to connect the *T. oceanica* transcriptome and the meta-omics work on the Svalbard's community, a first search for diatoms in the Arctic community was done. Only 2% of the total eukaryotic sequences were annotated as belonging to Bacillariophyta, corresponding to 0.11% of the total sequences. These sequences were matched to the *Bacillariophyceae* (subclass *Bacillariophycidae*), *Coscinodiccophyceae*, *Fragilariophyceae* and *Mediophyceae*, the four Bacillariophyta classes presently known (data not presented). These sequences were further analysed statistically (two way ANOVA) for effect of time and pCO_2, where no significance (p value = 0.08 and 0.06; respectively) was found. The metabolic functions corresponding to these Bacillariophyta sequences obtained via MG-RAST best hit search were also statistically evaluated for differential expression and were matched to 84 functions which were significantly affected by time (Figure 3 in Appendix) and none by pCO_2. These significantly time-affected functions are involved in ATP synthase, photosystem I and II, cytochrome complexes (electron carriers) and other ribosomal functions. Further investigation is needed as these functions are all indirectly connected to survival and photosynthesis. The possibility that OA's effect on organisms exposed for a relative short time (30 days) could suggest the occurrence of a certain adaptation. A longer exposure (years) might be necessary for the observation of significant effects as bacteria are often exposed to extreme variation in environmental conditions and might have the ability to compensate for negative environment.

The experiments presented showed that the analysis of metagenomes and metatranscriptomes of complex natural microbial communities subjected to different treatments can produce compelling information on OA. Metagenome and metatranscriptome analyses create enormous datasets and lead to the discovery of large numbers of novel sequences and metabolic functions. The undergoing analysis connecting taxa to significantly affected metabolic function will further improve our knowledge of OA. The present research included the top best blast hits, however, only a few organisms have had their whole genome sequenced which sometimes can lead to some biased annotation needing reassessment. The size of the output also creates some difficulties in getting an overview of the data and manual analysis of large sequences is always strenuous if not impossible. The urgent need for an user-friendly efficient computational platform for analysing metadata is acknowledged by the community as every user is faced with the challenge of viewing sequencing results. The importance of public portals is also critical in order to share results and improvement of sequencing protocol. One option to lighten next-generation sequencing analysis could be its combination with single-cell sequencing or single molecule real time sequencing (Shapiro *et*

al., 2013); which in the case of metadata could help sort what organisms directly express one particular gene of interest.

The present dissertation provides a detailed view of OA consequences on diatoms and Arctic microbial community. The unexpected evolution of sequencing technologies during the past decade and studies combining 16s rRNA gene amplicon, metagenome and metatranscriptome provide evidences that unknown and rare taxa/functions play a more important role in natural communities as previously expected. The possibility that sequencing techniques and bioinformatics tools could continue to be developed so efficiently, expand the hope that combination of such methods can elucidate processes until now thought to be uncharacterizable.

6. References

6. References

Agogue, H., Lamy, D., Neal, P. R., Sogin, M. L. & Herndl, G. J. 2011. Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol. Ecol.* **20**:258-74.

Archer, S. D., Kimmance, S. A., Stephens, J. A., Hopkins, F. E., Bellerby, R. G. J., Schulz, K. G., Piontek, J. & Engel, A. 2013. Contrasting responses of DMS and DMSP to ocean acidification in Arctic waters. *Biogeosciences* **10**:1893-908.

Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S. G., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kroger, N., Lau, W. W. Y., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A., Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P. & Rokhsar, D. S. 2004. The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. *Science* **306**:79-86.

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. *New Phytologist* **199**:121-34.

Baines, S. B., Twining, B. S., Brzezinski, M. A., Krause, J., Vogt, S., Assael, D. & McDaniel, H. 2012. Significant silicon accumulation by marine picocyanobacteria. *Nature Geoscience* **5**:886-91.

Bergquist, B. A., Wu, J. & Boyle, E. A. 2007. Variability in oceanic dissolved iron is dominated by the colloidal fraction. *Geochimica Et Cosmochimica Acta* **71**:2960-74.

Bethke, I., Furevik, T. & Drange, H. 2006. Towards a more saline North Atlantic and a fresher Arctic under global warming. *Geophysical Research Letters* **33**.

Bindoff, N. L., Willebrand, J., Artale, V., Cazenave, A., Gregory, J., Gulev, S., Hanawa, K., Le Quéré, C., Levitus, S., Nojiri, Y., Shum, C. K., Talley, L. D. & Unnikrishnan, A. 2007. Observations: oceanic climate change and sea level. *In*: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L. [Eds.] *Climate Change 2007: The physical science basis. Contribution of working group I to*

the fourth assessment report of the Intergovernmental Panel on Climate Change'. Cambridge University Press, Cambridge, United Kingdom and New York.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van De Peer, Y. & Grigoriev, I. V. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**:239-44.

Boyd, P. 2004. Ironing out algal issues in the southern ocean. *Science* **304**:396-97.

Boyd, P. W., Bakker, D. C. E. & Chandler, C. 2012. A New Database to Explore the Findings from Large-Scale Ocean Iron Enrichments Experiments. *Oceanography* **25**:64-71.

Caldeira, K. & Wickett, M. E. 2003. Anthropogenic carbon and ocean pH. *Nature* **425**:365-65.

Canfield, D. E., Stewart, F. J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E. F., Revsbech, N. P. & Ulloa, O. 2010. A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science* **330**:1375-78.

Castruita, M., Elmegreen, L. A., Shaked, Y., Stlefel, E. & Morel, F. 2007. Comparison of the kinetics of iron release from a marine (Trichodesmium erythraeum) Dps protein and mammalian ferritin in the presence and absence of ligands. *Journal of Inorganic Biochemistry* **101**:1686-91.

Chen, M. & Wang, W. X. 2001. Bioavailability of natural colloid-bound iron to marine plankton: Influences of colloidal size and aging. *Limnol. Oceanogr.* **46**:1956-67.

Chisholm, S. W. 2000. Oceanography - Stirring times in the Southern Ocean. *Nature* **407**:685-87.

Coale, K. H., Johnson, K. S., Fitzwater, S. E., Gordon, R. M., Tanner, S., Chavez, F. P., Ferioli, L., Sakamoto, C., Rogers, P., Millero, F., Steinberg, P., Nightingale, P., Cooper, D., Cochlan, W. P., Landry, M. R., Constantinou, J., Rollwagen, G., Trasvina, A. & Kudela, R. 1996. A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**:495-501.

Coale, K. H., Johnson, K. S., Chavez, F. P., Buesseler, K. O., Barber, R. T., Brzezinski, M. A., Cochlan, W. P., Millero, F. J., Falkowski, P. G., Bauer, J. E., Wanninkhof, R. H., Kudela, R. M., Altabet, M. A., Hales, B. E., Takahashi, T., Landry, M. R., Bidigare, R. R., Wang, X. J., Chase, Z., Strutton, P. G., Friederich, G. E., Gorbunov, M. Y., Lance, V. P., Hilting, A. K., Hiscock, M. R., Demarest, M., Hiscock, W. T., Sullivan, K. F., Tanner, S. J., Gordon, R. M., Hunter, C. N., Elrod, V. A., Fitzwater, S. E., Jones, J. L., Tozzi, S., Koblizek, M., Roberts, A. E., Herndon, J., Brewster, J., Ladizinsky, N., Smith, G., Cooper, D., Timothy, D., Brown, S. L., Selph, K. E., Sheridan, C. C., Twining, B. S. & Johnson, Z. I. 2004. Southern ocean iron enrichment experiment: Carbon cycling in high- and low-Si waters. *Science* **304**:408-14.

Crawfurd, K. J., Raven, J. A., Wheeler, G. L., Baxter, E. J. & Joint, I. 2011. The Response of Thalassiosira pseudonana to Long-Term Exposure to Increased CO_2 and Decreased pH. *Plos One* **6**.

Cullen, J. T., Bergquist, B. A. & Moffett, J. W. 2006. Thermodynamic characterization of the partitioning of iron between soluble and colloidal species in the Atlantic Ocean. *Marine Chemistry* **98**:295-303.

Diaz, J., Ingall, E., Benitez-Nelson, C., Paterson, D., de Jonge, M. D., McNulty, I. & Brandes, J. A. 2008. Marine polyphosphate: A key player in geologic phosphorus sequestration. *Science* **320**:652-55.

Dickson, A. G. 1981. An exact definition of total alkalinity and a procedure for the estimation of alkalinity and total inorganic carbon from titration data. *Deep-Sea Research Part a-Oceanographic Research Papers* **28**:609-23.

Dugdale, R. C. & Wilkerson, F. P. 1998. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* **391**:270-73.

Dyall, S. D., Brown, M. T. & Johnson, P. J. 2004. Ancient invasions: From endosymbionts to organelles. *Science* **304**:253-57.

Falkowski, P. G., Barber, R. T. & Smetacek, V. 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**:200-06.

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O. & Taylor, F. J. R. 2004. The evolution of modern eukaryotic phytoplankton. *Science* **305**:354-60.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**:237-40.

Figueres, G., Martin, J. M. & Meybeck, M. 1978. Iron behavior in the zaire estuary. *Netherlands Journal of Sea Research* **12**:329-37.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W. & DeLong, E. F. 2008. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. U. S. A.* **105**:3805-10.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C. & Zeng, N. 2006. Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison. *Journal of Climate* **19**:3337-53.

Gilbert, J. A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T., Somerfield, P. J., Huse, S. & Joint, I. 2009. The seasonal structure of microbial communities in the Western English Channel. *Environmental Microbiology* **11**:3132-39.

Gilbert, J. A., Meyer, F., Schriml, L., Joint, I. R., Muhling, M. & Field, D. 2010. Metagenomes and metatranscriptomes from the L4 long-term coastal monitoring station in the Western English Channel. *Standards in Genomic Sciences* **3**:183-93.

Gilbert, J. A. & Dupont, C. L. 2011. Microbial Metagenomics: Beyond the Genome. *Annual Review of Marine Science, Vol 3.* pp. 347-71.

Giovannoni, S. & Rappe, M. 2000. Evolution, diversity and molecular ecology of marine prokaryotes. *In*: L., K. D. [Ed.] *Microbial Ecology of the Oceans.* Wiley- Liss, New York, USA, pp. 47-83.

Hasle, G. R. & Syvertsen, E. E. 1996. Marine diatoms. *In*: Thomas, C. R. [Ed.] *Identifying Marine Phytoplankton.* Academic Press, San Diego, pp. 5-385.

Hoffmann, L. J., Peeken, I., Lochte, K., Assmy, P. & Veldhuis, M. 2006. Different reactions of Southern Ocean phytoplankton size classes to iron fertilization. *Limnol. Oceanogr.* **51**:1217-29.

Huber, J. A., Mark Welch, D., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. A. & Sogin, M. L. 2007. Microbial population structures in the deep marine biosphere. *Science* **318**:97-100.

Huse, S. M., Dethlefsen, L., Huber, J. A., Welch, D. M., Relman, D. A. & Sogin, M. L. 2008. Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. *Plos Genetics* **4**.

Hutchins, D. A., Witter, A. E., Butler, A. & Luther, G. W. 1999. Competition among marine phytoplankton for different chelated iron species. *Nature* **400**:858-61.

IPCC. 2007. *'Working group I fourth assessment report "The Physical Science Basis".'* Cambridge University Press Edn., Cambridge, United Kingdom and New Your, NY, USA,

Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Bergametti, G., Brooks, N., Cao, J. J., Boyd, P. W., Duce, R. A., Hunter, K. A., Kawahata, H., Kubilay, N., laRoche, J., Liss, P. S., Mahowald, N., Prospero, J. M., Ridgwell, A. J., Tegen, I. & Torres, R. 2005. Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science* **308**:67-71.

Kleine, T., Maier, U. G. & Leister, D. 2009. DNA Transfer from Organelles to the Nucleus: The Idiosyncratic Genetics of Endosymbiosis. *Annual Review of Plant Biology*. pp. 115-38.

Kooistra, W. & Medlin, L. K. 1996. Evolution of the diatoms (bacillariophyta) .4. Reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Molecular Phylogenetics and Evolution* **6**:391-407.

Kuma, K., Katsumoto, A., Shiga, N., Sawabe, T. & Matsunaga, K. 2000. Variation of size-fractionated Fe concentrations and Fe(III) hydroxide solubilities during a spring phytoplankton bloom in Funka Bay (Japan). *Marine Chemistry* **71**:111-23.

LaRoche, J., Boyd, P. W., McKay, R. M. L. & Geider, R. J. 1996. Flavodoxin as an in situ marker for iron stress in phytoplankton. *Nature* **382**:802-05.

LaRoche, J., Rost, B. & Engel, A. 2010. Bioassays, batch culture and chemostat experimentation. *In*: Riebesell, U., Fabry, V. J., Hansson, L. & Gattuso, J. P. [Eds.] *Guide to best practices for ocean acidification research and data reporting.* Publications Office of the European Union, Luxembourg, pp. 81-94.

Lee, K., Millero, F. J., Byrne, R. H., Feely, R. A. & Wanninkhof, R. 2000. The recommended dissociation constants for carbonic acid in seawater. *Geophysical Research Letters* **27**:229-32.

Levitus, S., Antonov, J. I., Wang, J. L., Delworth, T. L., Dixon, K. W. & Broccoli, A. J. 2001. Anthropogenic warming of Earth's climate system. *Science* **292**:267-70.

Lohbeck, K. T., Riebesell, U. & Reusch, T. B. H. 2012. Adaptive evolution of a key phytoplankton species to ocean acidification (vol 5, pg 346, 2012). *Nature Geoscience* **5**:917-17.

Lommer, M., Roy, A. S., Schilhabel, M., Schreiber, S., Rosenstiel, P. & LaRoche, J. 2010. Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics* **11**.

Lommer, M., Specht, M., Roy, A.-S., Kraemer, L., Andreson, R., Gutowska, M. A., Wolf, J., Bergner, S. V., Schilhabel, M. B., Klostermeier, U. C., Beiko, R. G., Rosenstiel, P., Hippler, M. & LaRoche, J. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology* **13**.

Longhurst, A., Sathyendranath, S., Platt, T. & Caverhill, C. 1995. An estimate of global primary production in the ocean from satellite radiometer data. *J. Plankton Res.* **17**:1245-71.

Losic, D., Mitchell, J. G. & Voelcker, N. H. 2009. Diatomaceous Lessons in Nanotechnology and Advanced Materials. *Advanced Materials* **21**:2947-58.

Losic, D., Yu, Y., Aw, M. S., Simovic, S., Thierry, B. & Addai-Mensah, J. 2010. Surface functionalisation of diatoms with dopamine modified iron-oxide nanoparticles: toward magnetically guided drug microcarriers with biologically derived morphologies. *Chemical Communications* **46**:6323-25.

Mackinder, L., Bach, L., Schulz, K., Wheeler, G., Schroeder, D., Riebesell, U. & Brownlee, C. 2011. The molecular basis of inorganic carbon uptake mechanisms in the coccolithophore emiliania huxleyi. *European Journal of Phycology* **46**:142-43.Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A. & Hollibaugh, J. T. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* **71**:4121-26.

Maldonado, M. T. & Price, N. M. 1996. Influence of N substrate on Fe requirements of marine centric diatoms. *Marine Ecology Progress Series* **141**:161-72.

Maldonado, M. T., Allen, A. E., Chong, J. S., Lin, K., Leus, D., Karpenko, N. & Harris, S. L. 2006. Copper-dependent iron transport in coastal and oceanic diatoms. *Limnol. Oceanogr.* **51**:1729-43.

Marcelja, S. 2010. The timescale and extent of thermal expansion of the global ocean due to climate change. *Ocean Science* **6**:179-84.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. T., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P. G., Begley, R. F. & Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376-80.

Martin, J. H. & Fitzwater, S. E. 1988. Iron-deficiency limits phytoplankton growth in the northeast pacific subarctic. *Nature* **331**:341-43.

Martin, J. H. 1990. Glacial-interglacial CO_2 change: The iron hypothesis. *Paleoceanography* **5**:1-13.

Martin, J. H., Gordon, R. M. & Fitzwater, S. E. 1990. Iron in antarctic waters. *Nature* **345**:156-58.

Martin, J. H., Coale, K. H., Johnson, K. S., Fitzwater, S. E., Gordon, R. M., Tanner, S. J., Hunter, C. N., Elrod, V. A., Nowicki, J. L., Coley, T. L., Barber, R. T., Lindley, S., Watson, A. J., Vanscoy, K., Law, C. S., Liddicoat, M. I., Ling, R., Stanton, T., Stockel, J., Collins, C., Anderson, A., Bidigare, R., Ondrusek, M., Latasa, M., Millero, F. J., Lee, K., Yao, W., Zhang, J. Z., Friederich, G., Sakamoto, C., Chavez, F., Buck, K., Kolber, Z., Greene, R., Falkowski, P., Chisholm, S. W., Hoge, F., Swift, R., Yungel, J., Turner, S., Nightingale, P., Hatton, A., Liss, P. & Tindale, N. W. 1994. Testing the iron hypothesis in ecosystems of the equatorial pacific-ocean. *Nature* **371**:123-29.

McKay, R. M. L., Geider, R. J. & LaRoche, J. 1997. Physiological and biochemical response of the photosynthetic apparatus of two marine diatoms to Fe stress. *Plant Physiol.* **114**:615-22.

McKay, R. M. L., La Roche, J., Yakunin, A. F., Durnford, D. G. & Geider, R. J. 1999. Accumulation of ferredoxin and flavodoxin in a marine diatom in response to Fe. *J. Phycol.* **35**:510-19.

Meehl, G. A., Stocker, T. F., Friedlingstein, P., Collins, W. D., Gaye, A. T., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C. B., Watterson, I. G.,

Weaver, A. J. & Zhao, Z.-C. 2007. Global climate projection. *In*: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. & Miller, H. L. [Eds.] *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Moustafa, A., Beszteri, B., Maier, U. G., Bowler, C., Valentin, K. & Bhattacharya, D. 2009. Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. *Science* **324**:1724-26.

Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. 1995. Production and dissolution of biogenic silica in the ocean - revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles* **9**:359-72.

Newbold, L. K., Oliver, A. E., Booth, T., Tiwari, B., DeSantis, T., Maguire, M., Andersen, G., van der Gast, C. J. & Whiteley, A. S. 2012. The response of marine picoplankton to ocean acidification. *Environmental Microbiology* **14**:2293-307.

Nishioka, J., Takeda, S., de Baar, H. J. W., Croot, P. L., Boye, M., Laan, P. & Timmermans, K. R. 2005. Changes in the concentration of iron in different size fractions during an iron enrichment experiment in the open Southern Ocean. *Marine Chemistry* **95**:51-63.

Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G. K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M. F., Yamanaka, Y. & Yool, A. 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**:681-86.

Oudot-Le Secq, M. P., Grimwood, J., Shapiro, H., Armbrust, E. V., Bowler, C. & Green, B. R. 2007. Chloroplast genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana: comparison with other plastid genomes of the red lineage. *Mol. Genet. Genomics* **277**:427-39.

Paz, Y., Shimoni, E., Weiss, M. & Pick, U. 2007. Effects of iron deficiency on iron binding and internalization into acidic vacuoles in Dunaliella salina. *Plant Physiol.* **144**:1407-15.

Peers, G. & Price, N. M. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441**:341-44.

Pitchford, J. W. & Brindley, J. 1999. Iron limitation, grazing pressure and oceanic high nutrient-low chlorophyll (HNLC) regions. *J. Plankton Res.* **21**:525-47.

Pondaven, P., Ragueneau, O., Treguer, P., Hauvespre, A., Dezileau, L. & Reyss, J. L. 2000. Resolving the 'opal paradox' in the Southern Ocean. *Nature* **405**:168-72.

Popovich, C. A. & Gayoso, A. M. 1999. Effect of irradiance and temperature on the growth rate of Thalassiosira curviseriata Takano (Bacillariophyceae), a bloom diatom in Bahia Blanca estuary (Argentina). *J. Plankton Res.* **21**:1101-10.

Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W. M., Moran, M. A. & Hollibaugh, J. T. 2005. Analysis of microbial gene transcripts in environmental samples. *Appl. Environ. Microbiol.* **71**:4121-26.

Poretsky, R. S., Hewson, I., Sun, S. L., Allen, A. E., Zehr, J. P. & Moran, M. A. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology* **11**:1358-75.

Raven, J. A. & Waite, A. M. 2004. The evolution of silicification in diatoms: inescapable sinking and sinking as escape? *New Phytologist* **162**:45-61.

Richier, S., Fiorini, S., Kerros, M. E., von Dassow, P. & Gattuso, J. P. 2011. Response of the calcifying coccolithophore Emiliania huxleyi to low pH/high pCO(2): from physiology to molecular level. *Marine Biology* **158**:551-60.

Riebesell, U. 1989. Comparison of sinking and sedimentation-rate measurements in a diatom winter spring bloom. *Marine Ecology Progress Series* **54**:109-19.

Riebesell, U., Fabry, V. J., Hansson, L. & Gattuso, J.-P. 2010. *Guide to best practices for ocean acidification research and data reporting.* Publications Office of the European Union, Luxembourg, 260.

Riebesell, U., Czerny, J., von Bröckel, K., Boxhammer, T., Büdenbender, J., Deckelnick, M., Fischer, M., Hoffmann, D., Krug, S. A., Lentz, U., Ludwig, A., Muche, R. & Schulz, K. G. 2013. Technical Note: A mobile sea-going mesocosm system – new opportunities for ocean change research. *Biogeosciences* **10**:1835-47.

Ronaghi, M., Uhlen, M. & Nyren, P. 1998. A sequencing method based on real-time pyrophosphate. *Science* **281**:363-+.

Round, F. E., Crawford, R. M. & Mann, D. G. 1990. *The Diatoms:Biology and Morphology of the Genera.* Cambridge University Press, Cambridge, U.K.,

Roy, R. N., Roy, L. N., Vogel, K. M., Portermoore, C., Pearson, T., Good, C. E., Millero, F. J. & Campbell, D. M. 1993. The dissociation-constants of carbonic-acid in seawater at

salinities 5 to 45 and temperatures 0-degrees-c to 45-degrees-c. *Marine Chemistry* **44**:249-67.

Roy, A. S., Gibbons, S. M., Schunck, H., Owens, S., Caporaso, J. G., Sperling, M., Nissimov, J. I., Romac, S., Bittner, L., Mühling, M., Riebesell, U., LaRoche, J. & Gilbert, J. A. 2013. Ocean acidification shows negligible impacts on high-latitude bacterial community structure in coastal pelagic mesocosms. *Biogeosciences* **10**:555-66.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., Wanninkhof, R., Wong, C. S., Wallace, D. W. R., Tilbrook, B., Millero, F. J., Peng, T. H., Kozyr, A., Ono, T. & Rios, A. F. 2004. The oceanic sink for anthropogenic CO_2. *Science* **305**:367-71.

Saltzman, M. R., Young, S. A., Kump, L. R., Gill, B. C., Lyons, T. W. & Runnegar, B. 2011. Pulse of atmospheric oxygen during the late Cambrian. *Proc. Natl. Acad. Sci. U. S. A.* **108**:3876-81.

Sanger, F. & Coulson, A. R. 1975. Rapid method for determining sequences in DNA by primed synthesis with DNA-polymerase. *J. Mol. Biol.* **94**:441-&.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**:5463-67.

Scala, S. & Bowler, C. 2001. Molecular insights into the novel aspects of diatom biology. *Cellular and Molecular Life Sciences* **58**:1666-73.

Schunck, H., Lavik, G., Desai, D. K., Großkopf, T., Kalvelage, T., Löscher, C. R., Paulmier, A., Contreras, S., Siegel, H., Holtappels, M., Rosenstiel, P., Schilhabel, M. B., Graco, M., Schmitz, R. A., Kuypers, M. M. M. & LaRoche, J. 2013. Giant Hydrogen Sulfide Plume in the Oxygen Minimum Zone off Peru Supports Chemolithoautotrophy. *PLoS ONE* **8**.

Shaked, Y., Kustka, A. B. & Morel, F. M. M. 2005. A general kinetic model for iron acquisition by eukaryotic phytoplankton. *Limnol. Oceanogr.* **50**:872-82.

Shapiro, E., Biezuner, T. & Linnarsson, S. 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**:618-30.

Shi, D. L., Xu, Y., Hopkinson, B. M. & Morel, F. M. M. 2010. Effect of Ocean Acidification on Iron Availability to Marine Phytoplankton. *Science* **327**:676-79.

Shinada, A., Shiga, N. & Ban, S. 1999. Origin of Thalassiosira diatoms that cause the spring phytoplankton bloom in Funka Bay, southwestern Hokkaido, Japan. *Plankton Biol. Ecol.* **46**:89-93.

Smetacek, V., Klaas, C., Strass, V. H., Assmy, P., Montresor, M., Cisewski, B., Savoye, N., Webb, A., d'Ovidio, F., Arrieta, J. M., Bathmann, U., Bellerby, R., Berg, G. M., Croot, P., Gonzalez, S., Henjes, J., Herndl, G. J., Hoffmann, L. J., Leach, H., Losch, M., Mills, M. M., Neill, C., Peeken, I., Rottgers, R., Sachs, O., Sauter, E., Schmidt, M. M., Schwarz, J., Terbruggen, A. & Wolf-Gladrow, D. 2012. Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature* **487**:313-19.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* **103**:12115-20.

Song, Y. L., Liu, C. X., Bolanos, M., Lee, J., McTeague, M. & Finegold, S. M. 2005. Evaluation of 16S rRNA sequencing and reevaluation of a short biochemical scheme for identification of clinically significant Bacteroides species. *J. Clin. Microbiol.* **43**:1531-37.

Sournia, A., Chretiennotdinet, M. J. & Ricard, M. 1991. Marine-phytoplankton - how many species in the world ocean. *J. Plankton Res.* **13**:1093-99.

Stewart, F. J., Sharma, A. K., Bryant, J. A., Eppley, J. M. & DeLong, E. F. 2011. Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities. *Genome Biology* **12**.

Strzepek, R. F. & Harrison, P. J. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* **431**:689-92.

Sunda, W. G., Swift, D. G. & Huntsman, S. A. 1991. Low iron requirement for growth in oceanic phytoplankton. *Nature* **351**:55-57.

Tett, P. & Barton, E. D. 1995. Why are there about 5000 species of phytoplankton in the sea. *J. Plankton Res.* **17**:1693-704.

Thomas, T., Gilbert, J. & Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation* **2**.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**:123-U16.

Tréguer, P. J. & De La Rocha, C. L. 2013. The World Ocean Silica Cycle. *In*: Carlson, C. A. & Giovannoni, S. J. [Eds.] *Annual Review of Marine Science, Vol 5.* pp. 477-501.

van den Hoek, C. M., D. & Jahns, H. M. 1997. *Algae: An introduction to phycology*. Cambridge University Press, London, UK,

Williamson, P., Wallace, D. W. R., Law, C. S., Boyd, P. W., Collos, Y., Croot, P., Denman, K., Riebesell, U., Takeda, S. & Vivian, C. 2012. Ocean fertilization for

geoengineering: A review of effectiveness, environmental impacts and emerging governance. *Process Safety and Environmental Protection* **90**:475-88.

Wolf-Gladrow, D. A., Riebesell, U., Burkhardt, S. & Bijma, J. 1999. Direct effects of CO_2 concentration on growth and isotopic composition of marine plankton. *Tellus Series B-Chemical and Physical Meteorology* **51**:461-76.

Woodson, J. D. & Chory, J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* **9**:383-95.

Wu, J. F. 2007. Determination of picomolar iron in seawater by double Mg(OH)(2) precipitation isotope dilution high-resolution ICPMS. *Marine Chemistry* **103**:370-81.

Wu, Y., Gao, K. & Riebesell, U. 2010. CO_2-induced seawater acidification affects physiological performance of the marine diatom Phaeodactylum tricornutum. *Biogeosciences* **7**:2915-23.

Yeh, S. W., Yim, B. Y., Noh, Y. & Dewitte, B. 2009. Changes in mixed layer depth under climate change projections in two CGCMs. *Climate Dynamics* **33**:199-213.

Zeebe, R. E. & Wolf-Gladrow, D. 2001. CO_2 *in Seawater: Equilibrium, Kinetics, Isotopes.* Elsevier, Amsterdam.

Zhang, R., Xia, X., Lau, S. C. K., Motegi, C., Weinbauer, M. G. & Jiao, N. 2013. Response of bacterioplankton community structure to an artificial gradient of pCO_2 in the Arctic Ocean. *Biogeosciences* **10**:3679-89.

7. Appendix

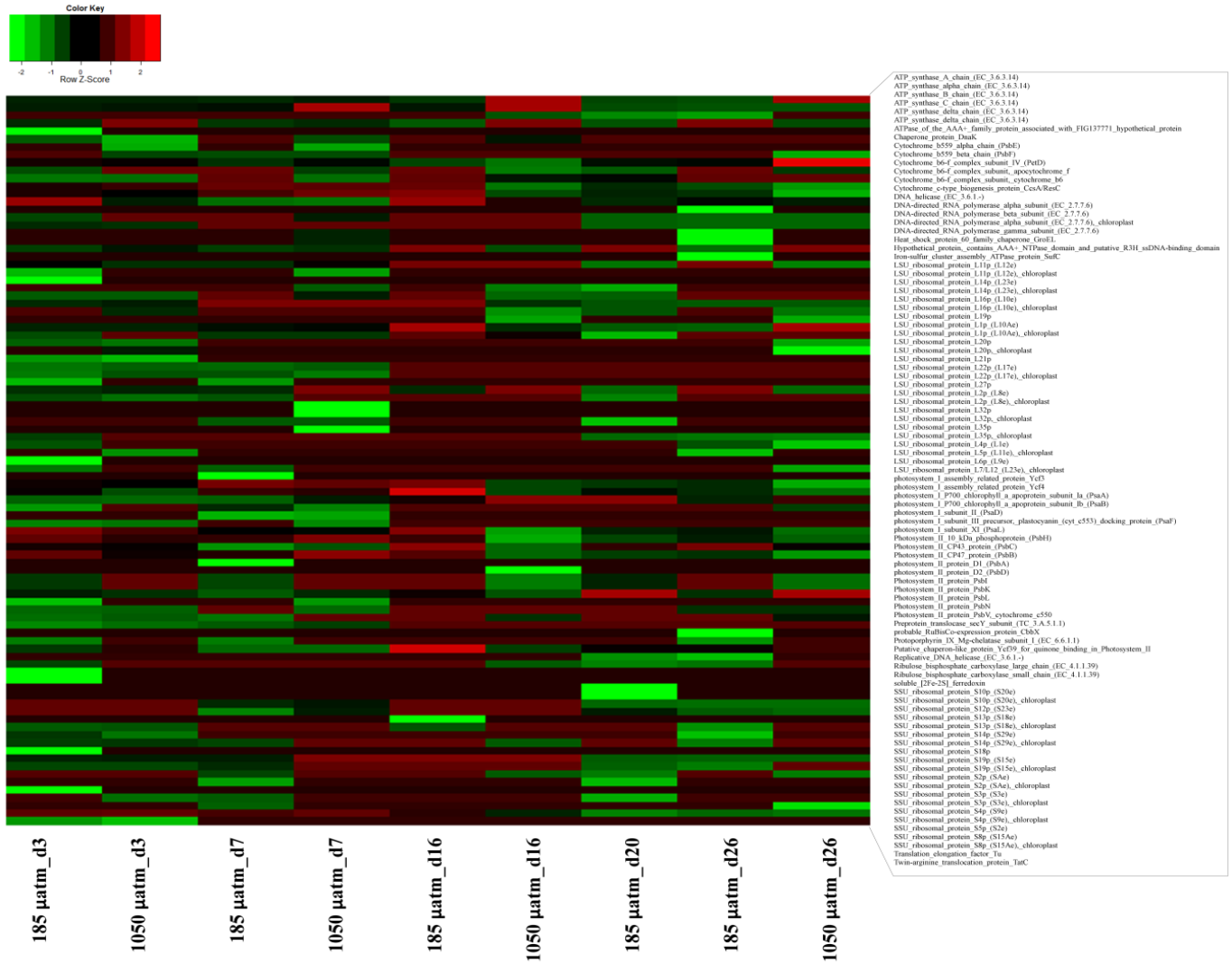


Figure 3. Heat map showing 84 metabolic functions significantly influenced by time.

8. Acknowledgements

8. Acknowledgements

I would firstly like to thank my supervisor Prof. Dr. Julie LaRoche for giving me the opportunity to make my PhD in a completely new field, letting me discover the wonderful world of phytoplankton and molecular biology, and also for always being available when needed.

Many thanks to Prof. Dr. Philip Rosenstiel and Prof. Dr. Arne Körtzinger for always making themselves available for questions and meetings.

Herzlichen Dank to Tania for her continuous help, support and for often sharing a brain. Herzlichen Dank to Diana for her "masterness" of the Mollab and moral support. Without the both of you my PhD and the daily life in Germany would have just not been the same.

DankeSchunck Liebe 'arald for all the times I was beached and you came to my rescue with a bucket or a hose ;) no seriously, thank you for always having time to help.

I also have to thank Tobi G., my office mate, for all the hours drinking coffee sharing anecdotes and delirium. Taking care of "your" plants, "your" telephone and sometimes sorting out your desk was always a good distraction.

Special thanks to Markus for his help with the analyses of *T. oceanica* as well as for countless discussions about diatoms, genomes, life, PhD, molecular biology, food, lab work, DNA degradation, etc...

Thank you to Dhwani for his help with R, MG-RAST and the transcriptomes analyses.

Thank you to Ben and Tina for always increasing the fun factor of culture work in A-M-Y and extraction work in the lab.

I also want to thank all my colleagues, the ones who left to discover the world and the ones who stayed at GEOMAR: Wiebke, Joerg, Katrin, Kai S., Kai L., Lennart, Allanah, Scarlett, Kerstin, Saskia, Silke, Christophe, David, Sebastian F., Bjorn, Helmke, Matthias and so many others. Thank you for all the discussion, the coffee and the many pleasant years. I also need to thank Ulrike and Scarlett for often giving me an oasis of peace after long hours of work in Kiel.

I am grateful for Jack A. Gilbert and Sean M. Gibbons, my American colleagues, for sharing their expertise in meta-omics analyses and for being efficient at everything even from afar.

Thank you Tobi S. for being a true friend in good, bad and "too busy" times.

Special thanks to all of my friends (Mireille, Marie-Ève, Rachel, Sara, Julie, Laura, and all the others) for their love and support but mostly for always understanding my constant absence.

I can't start to express my gratitude to my family for always being there, loving me and believing in me; "thank you" just doesn't seem enough...France, Gilles, Jean-Mathieu, Véronique, Malorie and P, un énorme merci à vous tous pour votre compréhension, amour et support.

Most of all I have to thank my partner, Dirk, for going through this with me and always being there when I needed him the most, even during bubupufferfish-time. Dirk, without you this would have not been possible. Thank you.

9. Declaration of work

9. Declaration of work

Hereby I assure that the content and design for the thesis titled:

Response of *Thalassiosira oceanica* and natural microbial communities to ocean acidification: a meta-omics comparison from unialgal cultures to mesocosms

is the result of work done apart from my supervisor's guidance or where clearly stated, completed by myself. The thesis has not been submitted to another examining body, either partially or wholly, as part of a doctoral degree unless clearly stated (publication I and II). Publication past or future of manuscripts was stated when necessary. Finally, I state that this thesis was prepared subject to the Rules of Good Scientific Practice of the German Research Foundation (DFG).

Kiel, 07th of November 2013 Alexandra-Sophie Roy