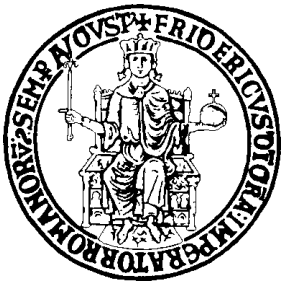
UNIVERSITY OF NAPLES "FEDERICO II"



DOCTORATE SCHOOL IN BIOLOGY

Cycle XXXIV

Omic approaches for the identification of biosynthetic pathways for bioactive compounds in the diatom *Thalassiosira rotula*

Coordinator
Prof. Sergio Esposito

Ph.D. student
Federica Di Costanzo

Tutor
Dr. Giovanna Romano

Co-tutor
Dr. Valeria Di Dato

2018 – 2021

TABLE OF CONTENTS

ABSTRACT

Diatoms are eukaryotic unicellular microorganisms widespread in all environments and responsible for about 20 % of the primary productivity on Earth. The evolutionary success of these organisms is greatly due to their metabolic plasticity that allows them to survive in different and, sometimes, challenging conditions of life through the production of a variety of bioactive molecules for defense, communication and adaptation to environmental changes.

In the last years, diatoms have gained increasing attention for their bioactive molecules that may find applications in pharmaceutical, cosmeceutical, nutraceutical and other biotechnological sectors and for their potential to become "biofactories" for a sustainable and eco-friendly production of several valuable metabolites.

My PhD thesis work was aimed at the exploitation of the biotechnological potential of the diatom *Thalassiosira rotula*, with a particular focus on lipid-derived bioactive compounds. Indeed, *T. rotula* was selected for its previously reported ability to produce unsaturated aldehydes as defence mechanism against copepods grazing that furthermore demonstrated to possess anti-proliferative and pro-apoptotic activities against different human carcinoma cell lines. Through the analysis of its transcriptome, several pathways involved in the production of bioactive molecules, i.e. secologanin, polyketides, prostaglandins and phytosterols, have been identified. Their existence has been confirmed by evaluating the gene expression of the key enzymes involved in each of the pathways listed. Moreover, their expression level was assessed along the growth or under different nutritional stresses, to evaluate their regulation and possible involvement in the response to stress conditions. For some of these pathways, the end products have been qualitatively and/or quantitatively analysed.

In addition to the transcriptome mining, the interaction of *T. rotula* with associated bacteria was studied, revealing the existence of three novel bacterial species. The analysis of their genomes revealed the presence of different biosynthetic gene clusters, among which a mixed non-ribosomal peptide synthetase/type 1 polyketide synthase (NRPS/T1PKS) pathway, potentially responsible for the production of interesting bioactive compounds. A qPCR-based approach helped in the characterization of their requirements for growing through the test of different culture media, both in presence or

absence of the diatom, and of their tolerance to different antibiotic mix and concentrations.

Altogether, the results here reported provide new stimuli in the field of marine biotechnology, as they show the presence of functionally active interesting biosynthetic pathways in the cosmopolitan species *T. rotula* and its associated bacterial community, paving the way for the utilization of these microorganisms for future biotechnological developments.

ACKNOWLEDGEMENTS

I would like to acknowledge some people without whom it would be impossible to achieve this goal. Firstly, I express my deepest gratitude to my tutor, Dr. Giovanna Romano, for her patience and kindness and for the guidance and continuous support during these three years. I am warmer grateful to my co-tutor, Dr. Valeria Di Dato, for her support in all the aspects of my PhD and her encouragement and positivity. Their precious advice and teachings helped me both personally and professionally.

I want to thank Dr. Adele Cutignano for her patience and fruitful Skypes conversations regarding GC-MS analysis. Thanks also to Dr. Lonnie Van Zyl and Prof. Marla Trinidade for welcoming me in their lab at the University of Western Cape in Cape Town.

I would like to thank all the Marine Biotechnology Department's "lab grande" of the Stazione Zoologica Anton Dohrn for being my family in the last years. Especially I thank Camilla, for showing me in the practice that kindness generates other kindness. Thanks to Laura for the support and the laughs, and to Ida for her precious advice and support every time I needed. Thanks to Massimo for his willingness in the support of my work. I thank Mariano, Gennaro and Martina for all the advice, the coffee breaks and the more and less serious conversations that enriched my lab life. I want to thank Giovanna, Elena and Flavio for making my days in the lab more cheerful, for the lunches and the coffee, the encouragement and support especially in the last phase of the PhD. Thanks to Viviana and Giorgio for being there for me whenever I needed comfort, advice and support.

I thank my family (Mamma, Papà, Valentina and Roberta, Chiara, Pierpaolo, Luca, Lucky and Max) for their immense support to all I did in my life and for being my "safe place". Thanks to Alessia, Davide, Lorena and Antonio for their friendship and the funny experiences shared together. Thanks to Armando's family, which is also mine, and to Armando itself, for being the light that guide me when I get lost.

I dedicate this milestone of my professional life to my grandparents, Nonna Anna and Nonno Renato, that always encouraged me and believed that I would realize all my dreams. Thanks Nonna for being always with me even if I can't see you anymore.

LIST OF FIGURES

LIST OF TABLES

THESIS OUTLINE

The outline of my thesis work is reported below.

The **General Introduction** contains background information about the topic addressed in this thesis, reporting the main findings about diatom secondary metabolites and the potential biotechnological applications of these microalgae. In particular, the first paragraph briefly describes the morphology of diatoms and their peculiar evolutionary history. The second and third parts are focused on the biotechnological applications of diatoms-bacteria community and of diatoms-derived bioactive compounds, especially lipid-derivatives. Finally, the model organism, i.e. *T. rotula*, is described with a focus on the available information regarding its metabolism. The description of prostaglandins biosynthetic pathway reported in the last part of the introduction is modified from the review "Di Costanzo, F., Di Dato, V., Ianora, A., Romano, G. Prostaglandins in Marine Organisms: A Review. *Marine Drugs* **17**, 428 (2019)".

Chapter 1 describes the *de-novo* transcriptome sequencing of *T. rotula* grown in control conditions compared to a nutritional stress conditions, i.e. silica depletion. Transcriptome analysis allowed the identification of pathways for the biosynthesis of interesting bioactive molecules with biotechnological applications, namely secologanin, polyketides and prostaglandins. The expression levels of the enzymes involved in the biosynthesis of these molecules have been evaluated and their belonging to diatom rather than to associated bacteria, has been confirmed using the diatom axenic genome. The results listed in this chapter are reported in the publication "Di Dato, V., Di Costanzo, F., Barbarinaldi, R., Perna, A., Ianora, A., Romano G. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Scientific Reports* **9**, *9893* (2019)". In addition, Chapter 1 also synthesize the preliminary results regarding the *T. rotula* axenic genome sequencing.

Chapter 2 represents a focus on the prostaglandin biosynthetic pathway in *T. rotula*. In particular, the transcript coding for *T. rotula* cyclooxygenase (COX) has been characterized and its identity supported by *in silico* analysis. Moreover, the gene expression levels of the key biosynthetic enzymes and the pool of prostaglandin molecules released in the medium have been analysed in different growth phases. On the basis of the results obtained, a possible signalling role for these molecules within the plankton community was hypothesized. The results listed in this chapter of the thesis are

reported in the publication "Di Dato, V., Barbarinaldi, R., Amato, A., Di Costanzo, F., Fontanarosa, C., Perna, A., Amoresano, A., Esposito, F., Cutignano, A., Ianora, A., Romano, G. Variation in prostaglandin metabolism during growth of the diatom *Thalassiosira rotula*. *Scientific Reports* **10**, 5374 (2020)".

Chapter 3 reports the reconstruction of the phytosterols biosynthetic pathway in *T. rotula* through transcriptomic analysis. The expression levels of the key enzymes have been evaluated and the sterols profile has been quantitatively and qualitatively characterized in *T. rotula* cultures grown in different nutritional stresses, i.e. nitrogen, silica and phosphate limitation, in comparison with a control condition. The goal of this study was to characterize the biosynthetic pathway in this diatom and to find appropriate conditions to enhance phytosterol production for possible biotechnological applications, giving their hypocholesterolemic, anti-diabetic, anti-inflammatory, neuro-modulatory and anti-cancer activities. These results are part of a manuscript in preparation.

Chapter 4 describes the identification, through a metagenomics analysis, of three novel bacterial species (named Cl-1, Cl-2 and Cl-8) in association with *T. rotula* and only one clone (namely FE7) of *S. marinoi*. The exploration of the three bacteria's genomes allowed the identification of different biosynthetic gene clusters, among which a mixed non-ribosomal peptide synthetase/type 1 polyketide synthase (NRPS/T1PKS) potentially responsible for the production of interesting bioactive molecules. With a qPCR-based approach, their growth requirements have been characterized by using different media, both in presence and in absence of the diatoms, and the activity of the β-lactamase/cephalosporinase enzymes identified in Cl-8 genome has been supported by testing the tolerance of the three bacterial species to different antibiotic mixes and concentrations. Moreover, the abundance of one of the species identified named Cl-1, which possesses a biosynthetic gene cluster coding for ectoine, was found to be correlated with the production of this osmolyte in xenic cultures grown at different salinities along ten days of culturing. These results are reported in the publication "Di Costanzo, F., Di Dato, V., van Zyl, L. J., Cutignano, A., Esposito, F., Trindade, M., Romano, G. Three novel bacteria associated with two centric diatom species from the Mediterranean Sea, *Thalassiosira rotula* and *Skeletonema marinoi*. *International Journal of Molecular Sciences* **22**, 13199 (2021)".

Finally, the **General Discussion** argues the main findings obtained through this doctoral work, highlighting their novelty and importance in the context of biotechnological applications of diatoms-derived bioactive compounds.

GENERAL INTRODUCTION

General characteristics of diatoms

Diatoms (class Bacillariophyceae) are unicellular photosynthetic eukaryotes that greatly contribute to the biogeochemical cycles of several key nutrients, are responsible for one fifth of the total photosynthesis on Earth[1] and are the main oxygen producer in marine ecosystems[2].

Diatoms are characterized by an incredible morphological diversity, with cell sizes that ranges from a few micrometers to a few millimeters and species-specific frustules shapes[3]. Moreover, they may live both as single cells or organized in chains[4]. Diatoms are characterized also by a peculiar external structure of amorphous silica named frustule that is composed of two valves, one bigger (epitheca) than the other (hypotheca), that fits together like a Petri dish[1]. The region overlapping the two thecae is surrounded by a grindle band, with each theca possessing micro and nanopatterns of pores with different sizes, shapes and patterns characteristic of each species[3,5]. The silica frustule offers mechanical protection from predators, is involved in gliding motility in benthonic diatoms, in increasing diatoms sinking rate and, together with osmolytes and other molecules, regulates their floating along the water column[6]. The pores allow the exchanges of gas and nutrients with the external environment[5].

On the basis of both morphological and phylogenetic analysis diatoms can be subdivided into two main groups: centric and pennate[1,7]. Centric diatoms are characterized by circular valves and are mainly planktonic while pennate diatoms have elongated valves and are mainly benthonic[8]. Centric diatoms can be further divided, on the basis of their symmetry, into radial centric diatoms (the oldest ones) and bi/multipolar centric diatoms, that evolved 150 million years ago[8]. Pennate diatoms comprises both the araphid pennate, that evolved 93 million years ago, and the raphid pennate, appeared 73 million years ago[8].

It has been estimated that there are from 12.000 - 30.000 up to ~ 200.000 different diatoms species distributed in water environments worldwide[1,9,10]. This ecological success is greatly due to their unique combination of metabolic features, acquired as genetic heritage during their peculiar evolutionary history[11,12]. Indeed, diatoms appeared more than 250 million years ago arising from a secondary endosymbiosis event characterized by the engulfment of a red alga, itself derived from a primary endosymbiosis, by a eukaryotic

host[1,13]. This red alga over the time evolved into plastids but the gene-transfer to the genome of the eukaryotic host continued during the evolution[14].

Considering that pennate and centric diatoms diverged "only" ~ 90 million years ago, their genomes structures demonstrated to be highly different, with the 40 % of their genes not shared between the two lineages[7]. Moreover, their dynamic evolutionary history is characterized by gains and losses of specific genes via horizontal-gene transfers from both bacteria and green algae, other than genes family expansions and domains recombination due to transposable elements[1,7]. All these mechanisms contributed in shaping a peculiar metabolic plasticity in diatoms, that allowed them to spread worldwide and rapidly diversify in the great variety of species that currently exist[7].

Bacteria interacting with diatoms and their biotechnological potential

Diatoms-bacteria relationships evolved in more than 200 million of years of cohabitation in the same environment[15] during which they established complex and heterogeneous networks, made of multiple cell-to-cell interactions, cohorts and consortia[16], to facilitate nutrient exchanges, communication and defence[15,16]. In these relationships, diatoms rely on bacteria for the uptake of nutrients that they are not able to produce, such as vitamins[17] and nitrogen[18], and for protection from pathogenic bacteria and competitive algal species[19]. On the other side, bacteria benefit from algal excretions of organic matter and nutrients[16].

It is known that more than 5 % of the diatoms genome is formed by bacterial genes attesting the evolutionary history of interaction between the two[19].

The majority of the exchanges at the basis of diatoms-bacteria interactions occurs in the phycosphere, a water layer not mixed with the surrounding water[15]. In this physical space that surrounds diatom's cell surface, nutrients and exudates are subjected to a diffusive transport[16]. Amino acids, carbohydrates and organic acids, that are passively excreted by diatoms during their early growth phases, can concentrate or form gradients in the phycosphere, attracting bacteria through chemotaxis[16]. In the same manner, chemo repellent molecules can inhibit bacteria encounters with diatoms[15]. Once reached the phycosphere, bacteria can establish cooperative or competitive interactions with diatoms, both mediated by the exchange of metabolites with a heterogeneous chemical structure, ranging from gaseous to complex compounds[20]. In this environment, bacteria often associate also among them forming consortia in which commensal or mutualistic

relationships occurs with the goal to degrade and utilize high molecular weight compounds released by diatoms[21].

Despite the influence that diatoms-bacteria interactions have in the ecology, evolution and physiology of diatoms[19] and in the regulation and structure of the marine food web[22], these are still not explored in depth[23]. This is mainly due to the fact that the traditional culturing approaches allow to cultivate, isolate and identify only a small fraction of the microorganisms living on Earth[23]. The great part of the studies on diatoms-bacteria interactions focus on the isolation of cultivable bacteria[24] or on the use of model systems formed by axenic diatoms co-cultured with single bacterial species[25]. However, these systems represent a simplified condition that does not explore properly the natural complexity of these associations[26]. For the same reasons, the structure of microbiota, as well as the active role of diatoms in shaping it, are not well known[26,27]. However, the development of pipelines based on metagenomics analysis allowed to explore the composition and the stability of the microbiota in different diatoms species[23,24] or its changes in dependence to sampling times and locations[28]. Less is known about the bioactive compounds produced by bacteria associated to diatoms and their biotechnological potential. Indeed, currently the main biotechnological application of bacteria associated with diatoms is in their use to control diatoms harmful algal blooms[15]. In addition, the increase of diatoms biomass caused by the establishment of nutrients exchanges with bacteria[29] can be exploited for biofuel production[29]. Finally, the cooperation between them can be useful in the removal of nutrients excess from surface waters and in the wastewater treatment through diversified pathways belonging to both the organisms[30,31].

Biotechnological applications of diatoms

Diatoms, and microalgae in general, evolved in a challenging environment characterized by the competition for nutrients, space and light[32] and the abundance of viruses, bacteria and grazers[33]. The combination of diatoms unique genomic features and the necessity to perceive and counteract external stimuli, allowed the development of peculiar physical (especially their silica frustule) and chemical features[32]. Indeed, diatoms are able to produce several compounds which have been shown to exert defensive and signalling roles, as for example toxins and unsaturated aldehydes[34,35]. Some of these compounds also showed to possess a series of different bioactivities that, together with other properties such as their fast growth rate, high photosynthetic efficiency, high lipids and

biomass production[36], made advantageous the use of this class of microalgae for several biotechnological applications[34].

Bioactive compounds can derive both by primary and secondary metabolisms and their production can be influenced by changes in environmental conditions[37]. Thus, manipulating cultivation parameters, it is possible to control the production of certain bioactive compunds[38]. Moreover, diatoms can be cultivated in all the seasons in photobioreactors, using non cultivable lands, not potable waters[39], solar energy and CO_2, that they fix with an high efficiency, allowing the large-scale production of bioactive molecules in a sustainable manner[38].

The main approach used to identify bioactive compounds is based on the cultivation of microalgae in different stressful conditions, as for example nutrient deprivation or alteration in light and temperature, with the goal to possibly enhance the production of the compounds of interest[40]. Intriguingly, the same microalgal species can produce different bioactive compounds when challenged by different stresses[40]. The biomass obtained is subjected to chemical extraction, fractionation and parallel screening for different bioactivities, a pipeline defined as bioactivity-guided fractionation[40]. In addition, the recent increasing availability of genomics, transcriptomic and metabolomics data allows also to perform *in silico* approaches for the identification of the enzymes responsible for the production of interesting bioactive compounds[40]. That approach can be useful for the simultaneous evaluation of the entire pool of molecule produced by the selected species[40] but also be preliminary for further manipulations of gene expression levels of interesting enzymes to force a pathway toward a higher production of valuable molecules[40]. As an example, new data on sterol biosynthesis in *Phaeodactylum tricornutum* were obtained using gene silencing and heterologous gene expression approaches, as well as enzyme inhibitors[41].

Diatoms biotechnological applications principally concern their use in bioremediation, in biofuel production, in nutraceutical and cosmeceutical sectors[42]. Moreover, diatoms extracts demonstrated to possess immunomodulatory[43], anti-oxidants[42], anti-cancer[44], anti-inflammatory[42] and anti-convulsant[45] properties, raising the interest of pharmaceutical industry. Among diatoms' most interesting and more abundantly produced bioactive compounds are pigments, vitamins, polysaccharides[44], sterols and polyunsaturated fatty acids (PUFAs), as well as other lipids derivatives[42].

Focusing on lipid compounds, diatoms' total lipids and triacylglycerides (TAGs) are substrates for the production of different biofuels[46]. In normal culturing conditions the

total lipid amount in diatoms represent the 22.7 % of the dry cell weight but it can be increased[46]. It is known that some stresses are able to shift the diatoms metabolism toward the accumulation of lipids, as for example nitrogen deficiency or silica limitation in the culturing medium[46,47]. Moreover, lipids accumulation can be stimulated also increasing photosynthetic efficiency, for example lifting CO_2 levels in the medium to 1-5 % as observed for *Thalassiosira pseudonana*[48] or increasing light intensity as observed in *P. tricornutum*[49]. Another strategy could be the use of an organic dye, named cy5, introduced inside the cells simply adding it into the culture medium, acting as a photosynthetic antenna which leads to increasing cell density and free-fatty acid production, as observed in *Thalassiosira weissflogii*[50]. Production of TAGs and, in general, total lipids can be increased also overexpressing the genes encoding for the key enzymes, i.e. glycerol-sn-3-phosphate acyl-transferase (GPAT), that catalyse the first esterification step and diacylglycerol acyl-transferase (DGAT), that catalyse the final step in TAGs biosynthesis[51].

Regarding possible pharmaceutical applications, it is worth mentioning PUFAs, which are very abundant in diatoms. For example, eicosapentaenoic acid (EPA) and docosapentaenoic acid (DHA) are promising compounds due to their anti-inflammatory properties, involvement in neurological and visual development during embryogenesis, help in the control of neurological effects of Parkinson and in the treatment of Alzheimer disease[42]. Moreover, PUFAs, especially omega-3, are used in the nutraceutical industry for their protective effect against cardiovascular diseases and their ability to improve metabolic rates, to regulate blood pressure and glucose levels[51]. Interesting PUFAs-derived compounds are the three polyunsaturated aldehydes (PUAs) identified by Miralto *et al.*[18], in *Thalassiosira rotula*, *Skeletonema costatum* (successively named as *Skeletonema marinoi* by Sarno *et al.*[52]) *and Pseudo-nitzschia delicatissima* (successively named as *Pseudo-nitzschia arenysensis* by Quijano *et al.*[53]), that showed anti-proliferative activity on human colon adenocarcinoma cell line (CACO-2) with one of the three, i.e. 2-trans-4-trans-decadienal, exerting the strongest effect[27].

Other interesting lipid compounds are phytosterols, that demonstrated to possess a plethora of different bioactivities among which hypocholesterolemic, anti-diabetic, anti-inflammatory, neuro-modulatory and neuroprotective ones[55,56]. In addition, campesterol, β-sitosterol, fucosterol and stigmasterol demonstrated to possess anti-cancer activities[42]. Squalene, mainly extracted from shark liver oil, has a documented activity in improving skin hydration and is currently applied in the cosmeceutical sector[56]. Squalene, together with its hydrogenated form squalane, is also a main component of parenteral emulsions

used for vaccines and drugs delivery[56]. Land plants are currently the main commercial source of phytosterols but the increasing market demand allowed to identify diatoms, and microalgae in general, as valuable and sustainable sources of these compounds[56].

Relevant bioactive compounds are also monogalactosyl diacylglicerols (MGDG). In particular, MGDG isolated from *P. tricornutum* showed to have pro-apoptotic activity in immortalized mammalian epithelial cells in a preliminary screening for a potential anti-cancer application[44]. In addition, sulfolipids, that are derivatives of both mono- and digalactosyl diacylglycerols, have been applied to produce a synthetic sulfoglycolipid patented as Sulfavant and in development as vaccine adjuvant[43].

Very recently, monoacylglicerides (MAGs) with C_{16} - C_{20} alkyl chains (among which the C_{16} ones are diatoms-specific), were purified from the centric diatom *S. marinoi* and demonstrated to possess anti-proliferative activity against haematological and colon cancer cells lines[57].

Finally, the carotenoid fucoxantin, responsible for the golden-brownish colour of diatoms, is widely used in nutraceutical and cosmeceutical industries for its anti-oxidant properties and in pharmaceutical sector for its anti-inflammatory, anti-obesity, anti-diabetic, anti-cancer, anti-osteoporotic, anti-hypertensive, and anti-microbial activities[58].

Diatoms, as well as other microalgae, can be used in the treatment of different wastewaters, as for example those deriving from aquaculture or textile industries, and in turn the biomass accumulated during that process can be used as biofuel source[59]. They are capable to adsorb heavy metals, especially copper and cadmium, from the pollutant waters detoxifying them inside the cells through a process unique of diatoms and microalgae that involve, as protecting mechanism, metal binding peptides named phytochelatins[36,59]. Moreover, they are capable to capture carbon and nitrogen from wastewaters, being able to simultaneously treat them and generate carbon-neutral fuel as sustainable alternative to fossil fuels[59,60].

In the last decade, diatoms frustules have raised interest from a technological point of view as possible alternative to synthetic materials, for their biocompatibility and mechanical strength properties[61]. In addition, it is possible to modify frustules structure, thickness and pore sizes, simply through the addition of non-essential metals into the culture medium, to fit several pharmaceutical and technological applications[62] among which their use as drug-delivery systems, as catalysts, as filtration systems and as biosensors[59,61].

The biotechnological applications of diatoms here discussed, together with the compounds involved, are summarized in Figure I.

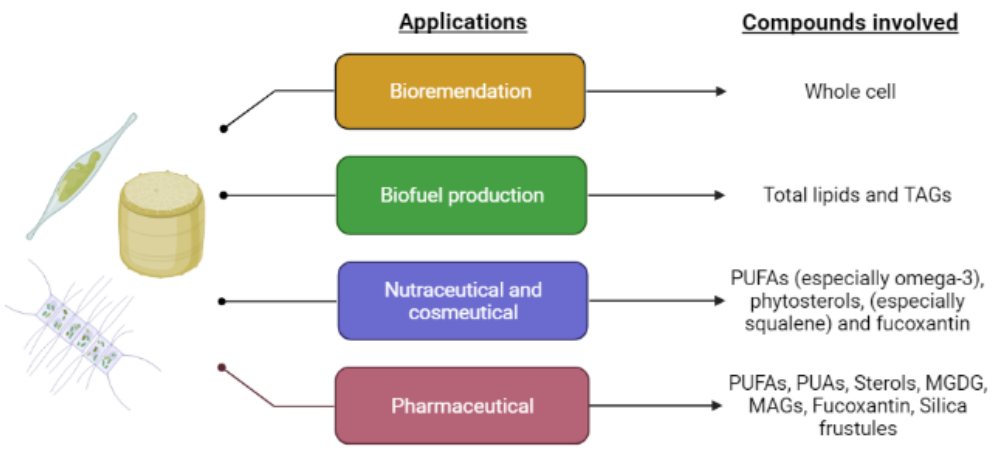


Figure I. Scheme of the different applications of diatoms in biotechnology and of the main lipid-derived compounds involved. Figure created with BioRender (https://biorender.com/).

Model organism: *Thalassiosira rotula*

T. rotula (Meunier, 1910), synonym of *Thalassiosira gravida*, is a planktonic and cosmopolitan diatom species most abundantly recorded during spring and summer in Northern European seas[63]. This planktonic diatom species belongs to the centric lineage and is characterized by the presence of a thick bundle, made of organic threads, that connect the center of adjacent cells to form chains[63] (Figure II). Each cell has a discoid shape, with a diameter ranging from 8 to 60 µm and a height of 5-20 µm. Cells valves are flat, with margins slightly rounded and covered with radial ribs[63]. The chloroplasts are numerous and distributed near the cell wall[63] (Figure II).

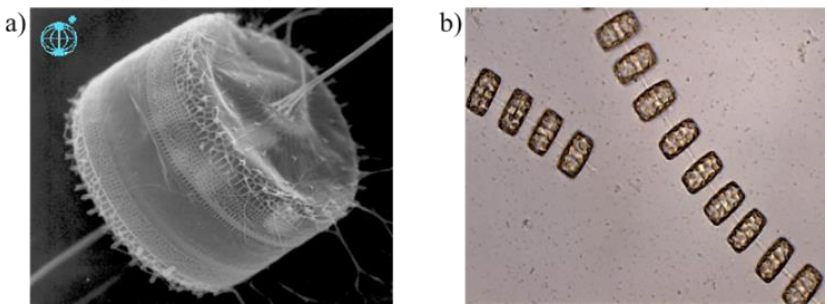


Figure II. Images of *Thalassiosira rotula* Meunier 1910.

a) SEM image of a single cell showing the valves structure and the central thick bundle (https://planktonnet.awi.de/index.php?contenttype=image_details&itemid=13809#content); **b)** Light microscope image of the *T. rotula* lab clone FE80c belonging to the Marine Biotechnology Department's

culture collection of Stazione Zoologica Anton Dohrn, Naples (Italy). The image shows the *T. rotula* cells organized in chains and connected each other by the thick bundle. Yellow-brown and round shaped chloroplasts are visible near each cell wall.

The study of the secondary metabolites in *T. rotula* focused mainly on lipid-derived compounds. In particular, *T. rotula* was the first diatom species that demonstrated to impair copepods (diatoms grazers) reproduction success[64] through the release of fatty acid-derived compounds having antimitotic activity. When this diatom species was used to feed the adult copepod *Temora stylifera*, the hatching success of the eggs laid by the female *Temora* feeding on *T. rotula* was seriously impaired, while no impairment was observed when the diet was based on the dinoflagellate *Prorocentrum minimum*[64]. It was subsequently observed that *T. rotula* is able to arrest embryonic development[65], while the few embryos that manage to hatch generated malformed nauplii[66]. Ianora *et al.*[67], demonstrated that diatom-associated bacteria are not responsible for these effects, since axenic *T. rotula* were capable to induce an even stronger impairment than xenic ones. Miralto *et al.*[35], finally isolated from *T. rotula*, *S. costatum* and *P. delicatissima* three PUAs, i.e. trans-4-cis-7-cis-decatrienal, 2-trans-4-trans-7-cis-decatrienal and 2-trans-4-trans-decadienal, that were responsible for both the reduction of fecundity and hatching success previously observed.

PUAs demonstrated to be chemical mediators both in defense and signaling mechanisms, being able to negatively affect also other phytoplankton species and some bacterial species, thus influencing the community composition of the plankton[68]. Diatoms themselves are able to perceive decadienal-like PUAs, inhibiting their growth or activating cell death mechanisms through a calcium-dependent release of nitric oxide[69].

To determine PUAs biosynthetic pathway in diatoms most of the studies have been performed using *T. rotula* as model species. A distinctive feature of diatoms is the absence of oxylipins derived from C_{18} PUFAs differently from algae and plants, while C_{20} PUFAs are preferred substrates, as occurs in animals[70]. Moreover, the C_{16} PUFAs hexadecatrienoic acid (HTrA) and hydroxytetradecanoic acid (HTA), deriving from chloroplastic galactolipids, are reported to be the main substrates for PUAs biosynthesis in *T. rotula*[71,72]. In this diatom species Ponhert[73] firstly observed that free PUFAs drastically increase in cells after wounding, hypothesizing the involvement of a wounding-activated phospholipase A_2 (PLA_2) able to release PUFAs in response to this external stimulus[52]. Following studies revealed also the presence of lipolytic acyl hydrolases (LAHs) in diatoms, acting instead or together with the phospholipase A_2[74].

PUFAs released either from phospho- and galactolipids are substrates of lipoxygenases (LOXs), which activity is stereo- and regio-specific, being responsible for the great diversity of diatoms oxylipins[74,75]. The LOXs peroxidation of PUFAs generates fatty acids hydroperoxides (FAHs)[70,74], that are further processed by a hydroperoxyde lyase (HPL) catalyzing the cleavage of a carbon-carbon bond using its cytochrome P450 system[76]. It is reported that the *T. rotula* lyase seems not to follow the canonical HPL mechanism of action, since it seems to use water from the medium to assist the cleavage of the carbon-carbon bond[76].

LOX/HPL pathways are responsible for the production of PUAs and omega-oxo-fatty acids[58] (Figure III). Other oxylipins, i.e. hydroxy-fatty acids, carbonyl-containing fatty acids and epoxyalcohol fatty acids are produced in diatoms from FAHs generated by LOXs activity[70] (Figure III).

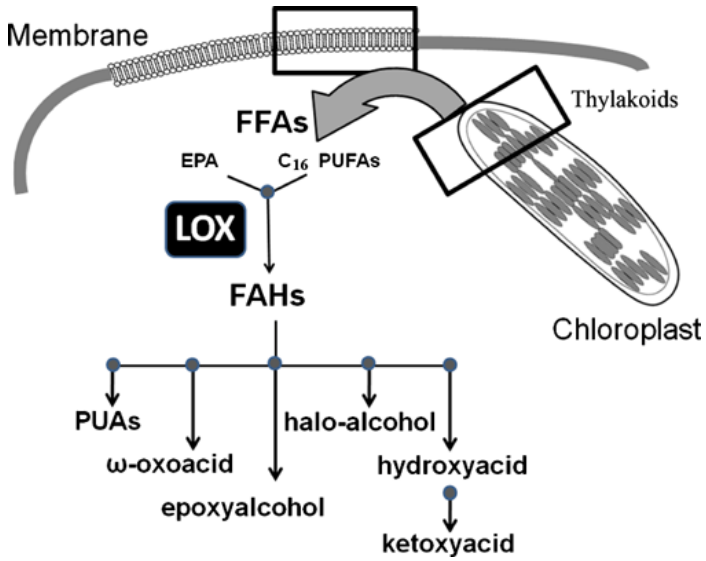


Figure III. Schematic representation of the LOX pathway in diatoms (Cutignano *et al.*, 2011)[70].

PUAs are not ubiquitous in diatoms and considerable species-specific differences occur in their production, as well as in oxylipins in general[70]. Interestingly, Pohnert *et al.*[77] reported that also different strains of the same species vary in their oxylipins profile and are able to induce different degrees of inhibition on copepod egg-hatching, as observed for two *T. rotula* strains named TR1 and TR2.

In 2017, Di Dato *et al.*[78] revealed, for the first time, the presence of an active prostaglandins (PGs) biosynthetic pathway in *S. marinoi*, extending the knowledge about PUFAs metabolism in diatoms. The key enzymes of this pathway have been identified also in *T. rotula*, as described in detail in Chapter 2 of this thesis, and the *in silico* analysis

of transcriptomes from different species in different growth conditions suggests the presence of PGs in many other diatom species[79].

Prostaglandins consist of a cyclopentanone nucleus with two side chains. Currently, three classes of prostaglandins are categorized, based on the number of double bonds present within the molecule and on the fatty acid precursor. Prostaglandins belonging to the series 1 have one double bond and derive from 8,11,14-eicosatrienoic acid (di-homolinolenic acid, ETrA), those of the series 2 have two double bonds and derive from 5,8,11,14-eicosatetraenoic acid (arachidonic acid, ARA), and those of the series 3 have three double bonds and derive from EPA

Prostaglandins derive by the sequential actions of highly specific enzymes (Figure IV). As occurs for PUAs, their synthesis is initiated by PLA_2, that hydrolyse membrane phospholipids at the sn-2 position liberating free fatty acid precursors, mainly ARA. Membrane-released ARA is then rapidly converted through the cyclization and inclusion of molecular oxygen in the precursor by the action of cyclooxygenase (COXs) enzymes into the unstable metabolite PGG_2, which is subsequently reduced to PGH_2 by the same enzyme.

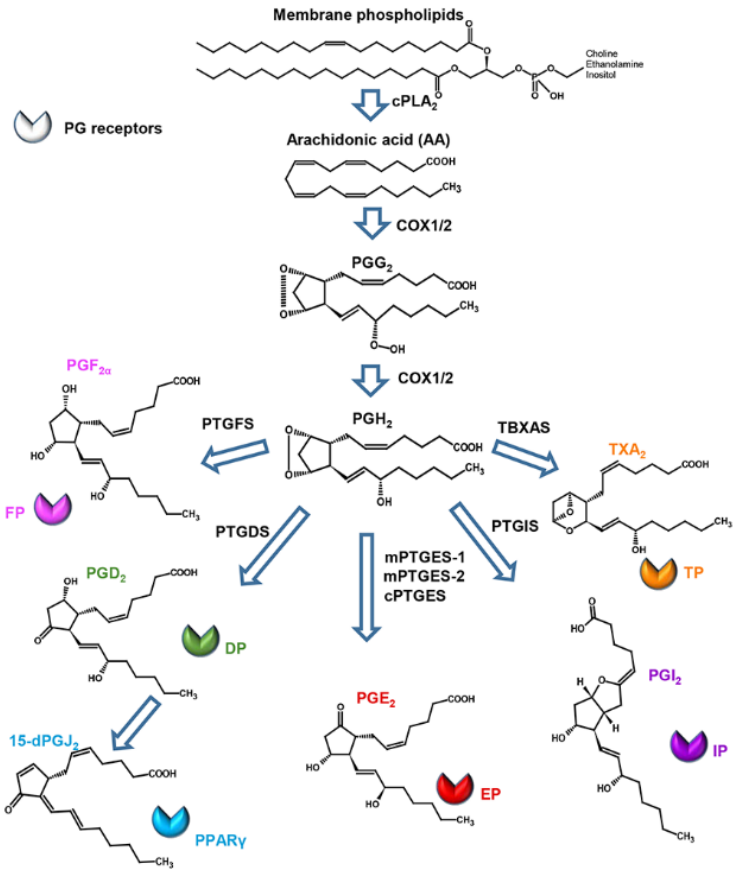


Figure IV. Prostaglandin biosynthetic pathway (Di Costanzo *et. al.*, 2019[79]). Enzymes involved in the pathway are reported next to the arrows.

The majority of the studies conducted on *T. rotula* focused on PUFAs, PUAs and prostaglandins. Beside these compounds, other lipids derivatives that have been preliminarily studied in *T. rotula* are sterols. In particular, the characterization of the sterol profile of this species revealed the production of 24-methylene cholesterol as major sterol, a characteristic shared among all the *Thalassiosirales* that could be useful for diatoms taxonomy[80,81]. On the contrary, there is no information available on the structure of sterols biosynthetic pathway in this diatom.

AIM OF THE THESIS

Diatoms are gaining increasing interest as valuable and sustainable sources of bioactive molecules with biotechnological applications[59]. In this frame, the aim of my PhD project was to exploit the biotechnological potential of the Neapolitan centric diatom *Thalassiosira rotula*, through an -omic approach. To this end, the first step pursued was the sequencing of the transcriptome and the genome of our model diatom. Particular attention was given to fatty acid-derived bioactive compounds, since in this diatom species it was previously reported the production of signalling and defence molecules, i.e. unsaturated aldehydes, that furthermore demonstrated to possess anti-proliferative and pro-apoptotic activities against different human carcinoma cell lines[35].

Through transcriptome analysis pathways codifying for secologanin, polyketides, prostaglandins and phytosterols have been identified. The expression of the key enzymes involved in each of the pathways listed have been subsequently validated by Real-time qPCR. Moreover, their expression level was assessed along the growth curve or under different nutritional stresses, to evaluate their regulation and possible involvement in the response to stress conditions. Finally, the end products of some of these pathways have been qualitatively and/or quantitatively analysed.

At the moment, the complete annotation of *T. rotula* genome is still in progress but its preliminary data obtained were used in support to the characterization of the biosynthetic pathways described in this thesis.

References

1. Armbrust, E. V. The life of diatoms in the world's oceans. *Nature* **459**, 185–192 (2009).

2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**, 237–240 (1998).

3. De Tommasi, E., Gielis, J. & Rogato, A. Diatom Frustule Morphogenesis and Function: a Multidisciplinary Survey. *Marine Genomics* **35**, 1–18 (2017).

4. Kooistra, W., Gersonde, R., Medlin, L. & Mann, D. The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence. *Evolution of Primary Producers in the Sea* 207–249 (2007) doi:10.1016/B978-012370518-1/50012-6.

5. Goessling, J. W. *et al.* Structure-based optics of centric diatom frustules: modulation of the in vivo light field for efficient diatom photosynthesis. *New Phytologist* **219**, 122–134 (2018).

6. Finkel, Z. & Kotrc, B. Silica Use Through Time: Macroevolutionary Change in the Morphology of the Diatom Fustule. *Geomicrobiol J* **27**, 596–608 (2010).

7. Bowler, C. *et al.* The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239–244 (2008).

8. Sorhannus, U. A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. *Mar Micropaleontol* **65**, 1–12 (2007).

9. Malviya, S. *et al.* Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci U S A* **113**, E1516-1525 (2016).

10. Cavalier-Smith, T. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma* **255**, 297–357 (2018).

11. Allen, A. E. *et al.* Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **473**, 203–207 (2011).

12. Athanasakoglou, A. & Kampranis, S. C. Diatom isoprenoids: Advances and biotechnological potential. *Biotechnology Advances* **37**, 107417 (2019).

13. Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science* **281**, 200–207 (1998).

14. Armbrust, E. V. *et al.* The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**, 79–86 (2004).

15. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between Diatoms and Bacteria. *Microbiol Mol Biol Rev* **76**, 667–684 (2012).

16. Seymour, J. R., Amin, S. A., Raina, J.-B. & Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton-bacteria relationships. *Nat Microbiol* **2**, 17065 (2017).

17. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B_{12} through a symbiotic relationship with bacteria. *Nature* **438**, 90–93 (2005).

18. Foster, R. A. *et al.* Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. *ISME J* **5**, 1484–1493 (2011).

19. Sison-Mangus, M. P., Jiang, S., Tran, K. N. & Kudela, R. M. Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J* **8**, 63–76 (2014).

20. Paul, C., Mausz, M. & Pohnert, G. A co-culturing/metabolomics approach to investigate chemically mediated interactions of planktonic organisms reveals influence of bacteria on diatom metabolism. *Metabolomics* **9**, 349–359 (2013).

21. Mönnich, J. *et al.* Niche-based assembly of bacterial consortia on the diatom *Thalassiosira rotula* is stable and reproducible. *ISME J* **14**, 1614–1625 (2020).

22. Cirri, E. & Pohnert, G. Algae bacteria interactions that balance the planktonic microbiome. *New Phytologist* **223**, (2019).

23. Behringer, G. *et al.* Bacterial Communities of Diatoms Display Strong Conservation Across Strains and Time. *Front Microbiol* **9**, (2018).

24. Kojadinovic-Sirinelli, M. *et al.* Exploring the microbiome of the 'star' freshwater diatom *Asterionella formosa* in a laboratory context. *Environ. Microbiol.* **20**, 3601–3615 (2018).

25. Baker, L. J. & Kemp, P. F. Bacterial inoculations can perturb the growth trajectory of diatoms with an existing microbiome. *PeerJ* **8**, (2020).

26. Shibl, A. A. *et al.* Diatom modulation of select bacteria through use of two unique secondary metabolites. *PNAS* **117**, 27445–27455 (2020).

27. Majzoub, M. E. *et al. Phaeobacter inhibens* controls bacterial community assembly on a marine diatom. *FEMS Microbiol. Ecol.* **95**, (2019).

28. Ajani, P. A. *et al.* The Microbiome of the Cosmopolitan Diatom *Leptocylindrus* Reveals Significant Spatial and Temporal Variability. *Front Microbiol* **9**, 2758 (2018).

29. Yao, S. *et al.* Microalgae-bacteria symbiosis in microalgal growth and biofuel production: a review. *J Appl Microbiol* **126**, 359–368 (2019).

30. Liu, J. *et al.* Advanced nutrient removal from surface water by a consortium of attached microalgae and bacteria: A review. *Bioresour Technol* **241**, 1127–1137 (2017).
31. Mujtaba, G. & Lee, K. Advanced Treatment of Wastewater Using Symbiotic Co-culture of Microalgae and Bacteria. *Applied Chemistry for Engineering* **27**, 1–9 (2016).
32. Ianora, A. *et al.* New trends in marine chemical ecology. *Estuaries and Coasts: J ERF* **29**, 531–551 (2006).
33. Ingebrigtsen, R. A., Hansen, E., Andersen, J. H. & Eilertsen, H. C. Light and temperature effects on bioactivity in diatoms. *J Appl Phycol* **28**, 939–950 (2016).
34. Lauritano, C. & Ianora, A. Chemical Defense in Marine Organisms. *Mar Drugs* **18**, E518 (2020).
35. Miralto, A. *et al.* The insidious effect of diatoms on copepod reproduction. *Nature* **402**, 173–176 (1999).
36. Baldisserotto, C., Sabia, A., Ferroni, L. & Pancaldi, S. Biological aspects and biotechnological potential of marine diatoms in relation to different light regimens. *World J Microbiol Biotechnol* **35**, 35 (2019).
37. Romano, G. *et al.* Marine microorganisms as a promising and sustainable source of bioactive molecules. *Mar Environ Res* **128**, 58–69 (2017).
38. de Morais, M. G., Vaz, B. da S., de Morais, E. G. & Costa, J. A. V. Biologically Active Metabolites Synthesized by Microalgae. *BioMed Research International* **2015**, e835761 (2015).
39. Bayu, A., Rachman, A., Noerdjito, D., Putra, M. & Widayatno, W. High-value chemicals from marine diatoms: a biorefinery approach. *IOP Conference Series: Earth and Environmental Science* **460**, 012012 (2020).
40. Lauritano, C., Ferrante, M. I. & Rogato, A. Marine Natural Products from Microalgae: An -Omics Overview. *Mar Drugs* **17**, E269 (2019).
41. Fabris, M. *et al.* Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *New Phytologist* **204**, 521–535 (2014).
42. Tiwari, A. *et al.* Therapeutic attributes and applied aspects of biological macromolecules (polypeptides, fucoxanthin, sterols, fatty acids, polysaccharides, and polyphenols) from diatoms - A review. *Int J Biol Macromol* **171**, 398–413 (2021).
43. Riccio, G. & Lauritano, C. Microalgae with Immunomodulatory Activities. *Mar Drugs* **18**, 2 (2019).
44. Martínez Andrade, K. A., Lauritano, C., Romano, G. & Ianora, A. Marine Microalgae with Anti-Cancer Properties. *Mar Drugs* **16**, 165 (2018).

45. Brillatz, T. *et al.* Zebrafish-based identification of the antiseizure nucleoside inosine from the marine diatom *Skeletonema marinoi*. *PLoS One* **13**, e0196195 (2018).
46. Hildebrand, M., Davis, A. K., Smith, S. R., Traller, J. C. & Abbriano, R. The place of diatoms in the biofuels industry. *Biofuels* **3**, 221–240 (2012).
47. d'Ippolito, G. *et al.* Potential of lipid metabolism in marine diatoms for biofuel production. *Biotechnol Biofuels* **8**, 28 (2015).
48. Sabia, A., Clavero, E., Pancaldi, S. & Salvadó Rovira, J. Effect of different CO_2 concentrations on biomass, pigment content, and lipid production of the marine diatom *Thalassiosira pseudonana*. *Appl Microbiol Biotechnol* **102**, 1945–1954 (2018).
49. Kurpan Nogueira, D. P., Silva, A. F., Araújo, O. Q. F. & Chaloub, R. M. Impact of temperature and light intensity on triacylglycerol accumulation in marine microalgae. *Biomass and Bioenergy* **72**, 280–287 (2015).
50. Leone, G. *et al.* Incorporating a molecular antenna in diatom microalgae cells enhances photosynthesis. *Sci Rep* **11**, 5209 (2021).
51. Santin, A. *et al.* Highly Valuable Polyunsaturated Fatty Acids from Microalgae: Strategies to Improve Their Yields and Their Potential Exploitation in Aquaculture. *Molecules* **26**, 7697 (2021).
52. Sarno, D., Kooistra, W. H. C. F., Medlin, L. K., Percopo, I. & Zingone, A. Diversity in the Genus *Skeletonema* (Bacillariophyceae). II. an Assessment of the Taxonomy of *S. Costatum*-Like Species with the Description of Four New Species. *Journal of Phycology* **41**, 151–176 (2005).
53. Sonia Isabel Quijano-Scheggia *et al.* Morphology, physiology, molecular phylogeny and sexual compatibility of the cryptic *Pseudo-nitzschia delicatissima* complex (Bacillariophyta), including the description of *P. arenysensis* sp. nov. *Phycologia* **48**, 492–509 (2009).
54. Sansone, C. *et al.* Diatom-Derived Polyunsaturated Aldehydes Activate Cell Death in Human Cancer Cell Lines but Not Normal Cells. *PLoS One* **9**, e101220 (2014).
55. Jaramillo-Madrid, A. C., Ashworth, J., Fabris, M. & Ralph, P. J. Phytosterol biosynthesis and production by diatoms (Bacillariophyceae). *Phytochemistry* **163**, 46–57 (2019).
56. Randhir, A., Laird, D. W., Maker, G., Trengove, R. & Moheimani, N. R. Microalgae: A potential sustainable commercial source of sterols. *Algal Research* **46**, 101772 (2020).
57. Miceli, M. *et al.* Monoacylglycerides from the Diatom *Skeletonema marinoi* Induce Selective Cell Death in Cancer Cells. *Marine Drugs* **17**, 625 (2019).

58. Seth, K. *et al.* Bioprospecting of fucoxanthin from diatoms - Challenges and perspectives. *Algal Research* **60**, 102475 (2021).
59. Sharma, N. *et al.* Diatoms Biotechnology: Various Industrial Applications for a Greener Tomorrow. *Frontiers in Marine Science* **8**, 106 (2021).
60. Wang, J.-K. & Seibert, M. Prospects for commercial production of diatoms. *Biotechnology for Biofuels* **10**, 16 (2017).
61. Wang, Z., Gong, D. & Cai, J. Diatom Frustule Array for Flow-Through Enhancement of Fluorescent Signal in a Microfluidic Chip. *Micromachines* **12**, 1017 (2021).
62. Soleimani, M. *et al.* Modifying the thickness, pore size, and composition of diatom frustule in *Pinnularia* sp. with Al^{3+} ions. *Sci Rep* **10**, 19498 (2020).
63. EOS - Phytoplankton Encyclopedia Project. https://www.eoas.ubc.ca/research/phytoplankton/diatoms/centric/thalassiosira/t_rotula.html.
64. Ianora, A. & Poulet, S. A. Egg viability in the copepod *Temora stylifera*. *Limnology and Oceanography* **38**, 1615–1626 (1993).
65. Poulet, S. A., Ianora, A., Miralto, A. & Meijer, L. Do diatoms arrest embryonic development in copepods? *Marine Ecology Progress Series* **111**, 79–86 (1994).
66. Poulet, S. A., Laabir, M., Ianora, A. & Miralto, A., Reproductive response of *Calanus helgolandicus*. I. Abnormal embryonic and naupliar development. *Marine Ecology Progress Series* **129**, 85–95 (1995).
67. Ianora, A., Poulet, S. A., Miralto, A. & Grottoli, R. The diatom *Thalassiosira rotula* affects reproductive success in the copepod *Acartia clausi*. *Marine Biology* **125**, 279–286 (1996).
68. Ianora, A. *et al.* The Relevance of Marine Chemical Ecology to Plankton and Ecosystem Function: An Emerging Field. *Mar Drugs* **9**, 1625–1648 (2011).
69. Vardi, A. *et al.* A Stress Surveillance System Based on Calcium and Nitric Oxide in Marine Diatoms. *PLOS Biology* **4**, e60 (2006).
70. Cutignano, A. *et al.* Lipoxygenase Products in Marine Diatoms: A Concise Analytical Method to Explore the Functional Potential of Oxylipins. *Journal of Phycology* **47**, 233–243 (2011).
71. Cutignano, A. *et al.* Chloroplastic Glycolipids Fuel Aldehyde Biosynthesis in the Marine Diatom *Thalassiosira rotula*. *ChemBioChem* **7**, 450–456 (2006).
72. d'Ippolito, G. *et al.* New C_{16} fatty-acid-based oxylipin pathway in the marine diatom *Thalassiosira rotula*. *Org Biomol Chem* **3**, 4065–4070 (2005).

73. Pohnert, G. Phospholipase A_2 activity triggers the wound-activated chemical defense in the diatom *Thalassiosira rotula*. *Plant Physiol* **129**, 103–111 (2002).
74. Fontana, A. *et al.* Chemistry of oxylipin pathways in marine diatoms. *Pure and Applied Chemistry* **79**, 481–490 (2007).
75. Lamari, N. *et al.* Specificity of Lipoxygenase Pathways Supports Species Delineation in the Marine Diatom Genus *Pseudo-nitzschia*. *PLoS One* **8**, e73281 (2013).
76. Barofsky, A. & Pohnert, G. Biosynthesis of Polyunsaturated Short Chain Aldehydes in the Diatom *Thalassiosira rotula*. *Org. Lett.* **9**, 1017–1020 (2007).
77. Pohnert, G. *et al.* Are volatile unsaturated aldehydes from diatoms the main line of chemical defence against copepods? Mar Ecol Prog Ser. *Marine Ecology - Progress Series, v.245, 33-45 (2002)* **245**, (2002).
78. Di Dato, V. *et al.* Animal-like prostaglandins in marine microalgae. *ISME J* **11**, 1722–1726 (2017).
79. Di Costanzo, F., Di Dato, V., Ianora, A. & Romano, G. Prostaglandins in Marine Organisms: A Review. *Mar Drugs* **17**, (2019).
80. Barrett, S. M., Volkman, J. K., Dunstan, G. A. & LeRoi, J.-M. Sterols of 14 species of marine diatoms (Bacillariophyta). *J Phycol* **31**, 360–369 (1995).
81. Volkman, J. K. & Hallegraeff, G. M. Lipids in marine diatoms of the genus *Thalassiosira*: Predominance of 24-methylenecholesterol. *Phytochemistry* **27**, 1389–1394 (1988).

Chapter 1

Identification of the secologanin, polyketides and prostaglandins biosynthetic pathways in the diatom *Thalassiosira rotula*

Adapted from the published article:

Di Dato, V., Di Costanzo, F., Barbarinaldi, R., Perna, A., Ianora, A., Romano G. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Scientific Reports* **9**, 9893 (2019).

1.1 Abstract

Diatoms are phytoplankton eukaryotic microalgae that are widely distributed in the world's oceans and are responsible for 20–25 % of total carbon fixation on the planet. Using transcriptome sequencing, here we show for the first time that the ubiquitous diatom *Thalassiosira rotula* expresses biosynthetic pathways that potentially lead to the synthesis of interesting secondary metabolites with pharmaceutical applications such as polyketides, prostaglandins and secologanin. We also show that these pathways are differentially expressed in conditions of silica depletion in comparison with standard growth conditions.

1.2 Introduction

Diatoms are a major group of eukaryotic microalgae in the phytoplankton widely distributed in the world's oceans, and capable through photosynthesis of fixing about 20–25 % of the total carbon on the planet[1]. Their worldwide distribution is supported by a molecular tool-kit (M-T) that allows them to adapt to different conditions through the perception of environmental cues and the control of competitors and grazers[2–5]. Some of the metabolites which are part of this M-T have already been identified as having potential applications as pharmaceuticals or cosmeceuticals[6]. Nevertheless, the full potential of marine diatoms still needs to be unlocked. One way to unlock this potential is to expose diatoms to stressful conditions to force changes in their metabolism and activate biosynthetic pathways inducing the production of secondary metabolites. One of the stressful conditions to promote such changes is nutrient depletion during microalgae growth. For example, phosphorus (P) deficiency in *Thalassiosira rotula* induces a remodelling of the transcriptome including changes in cellular P allocation patterns, enzyme activity and lipid composition, whereas nitrogen (N) deprivation reduces primary carbon metabolism leading to accumulation of lipids[7]. In addition to P and N, also carbon deprivation, independently from the light intensity (30, 300 or 1000 μmol photon m^{-2} s^{-1}), induces a rearrangement of the Carbon Concentration Mechanisms (CCM) toward the synthesis of pyruvate[8,9] reorienting the carbon metabolism toward lipid accumulation. A similar remodelling of metabolic pathways occurs in conditions of silica depletion[10]. Diatoms require silica (Si) to generate the intricate frustules that surround the cell and allow the entrance of light to the chloroplasts and the transport of gases and solutes. The silica cell wall is believed to provide an ecological advantage over other phytoplankton groups due to its mechanical strength that protects diatoms against grazers. Deprivation

of Si arrests progression of cell cycle events, such as cell division and DNA replication, attesting the importance of silica in the diatom life cycle[11].

Here we analyse the transcriptome of the diatom *Thalassiosira rotula*, isolated from the Gulf of Naples in the Mediterranean Sea, comparing cultures grown in normal condition (CTR) versus cultures grown in conditions of stress such as silica depletion (SiDepl). Our analysis shows that this diatom species has the potential to produce new metabolites, i.e. secologanin, polyketides and prostaglandins, the related pathways of which are differentially expressed under this stress condition. We experimentally confirmed the expression of these pathways identified in the newly sequenced transcriptome of our clone and also compared their expression with those from other *T. rotula* clones, grown under different stress conditions and sequenced in the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP)[12] supported by the Moore Foundation. Our study also confirms that the identified pathways are actually present in the *T. rotula* genome and do not derive from bacteria commonly associated with diatom cells.

1.3 Results

1.3.1 *De Novo* transcriptome assembly and functional annotation

Total RNA from *Thalassiosira rotula* cultured in f/2 complete medium (control condition, CTR), and in f/2 silica depleted medium, (stress condition, SiDepl), was obtained from cells collected at the beginning of the stationary phase of growth, when cells are more prone to produce secondary metabolites[13]. Table 1.1 shows the statistics for the total transcriptome assembly after filtering of unspecific sequences. The total assembly was composed by 142 Mbp in 107,289 transcripts grouped into 66,496 genes, with a N50 of 2,034 bp. The final dataset translated into proteins (minimum length 50 aa) includes 60,172 protein sequences, 28,128 (46.7 %) of which were identified as complete, i.e. having a start methionine and a stop codon.

Table 1.1. Transcriptome sequencing statistics.

Assembly (Mbp)	142
Transcripts (n=)	107,289
Genes (n=)	66,496
Mean GC%	47.53
Average contig lenght (bp)	1,332.76
Median contig lenght (bp)	970
N50 (bp)	2,034
Proteins (n=)	60,172

The sequences were also analysed for the presence of repetitive elements and more than 3 Mbp were identified as repetitive elements including, for example, DNA transposons, retro-elements, satellites, and rRNA. Functional annotation of the combined transcriptomes was performed on proteome sequences, 89 % of which showed homology with other diatoms, in particular *T. pseudonana*, *T. oceanica* and *Phaeodactylum tricornutum* (Supplementary Data 1 and 2). Out of a total 60,172 peptide sequences, 63 % had at least one blast hit, and almost half (35 %) retrieved a blast hit against conserved domains (Figure 1.1 a). The functional annotation of peptides assigned the majority of proteins and domains to the Gene Ontology (GO) terms in the Molecular Functions (MF) category (Table 1.2). Except for the Cellular Components (CC) category, for both Biological Processes (BP) and MF categories, the number of associated domains was almost half the number of associated protein sequences. The "ATP binding" GO term was

the most represented in the MF category while “transmembrane transport” was the most represented GO term in the BP category, and “integral to membrane” was the most represented GO term in the CC category (Supplementary Data 3 to 8, SI1: Supplementary Figure 1.1 a–f).

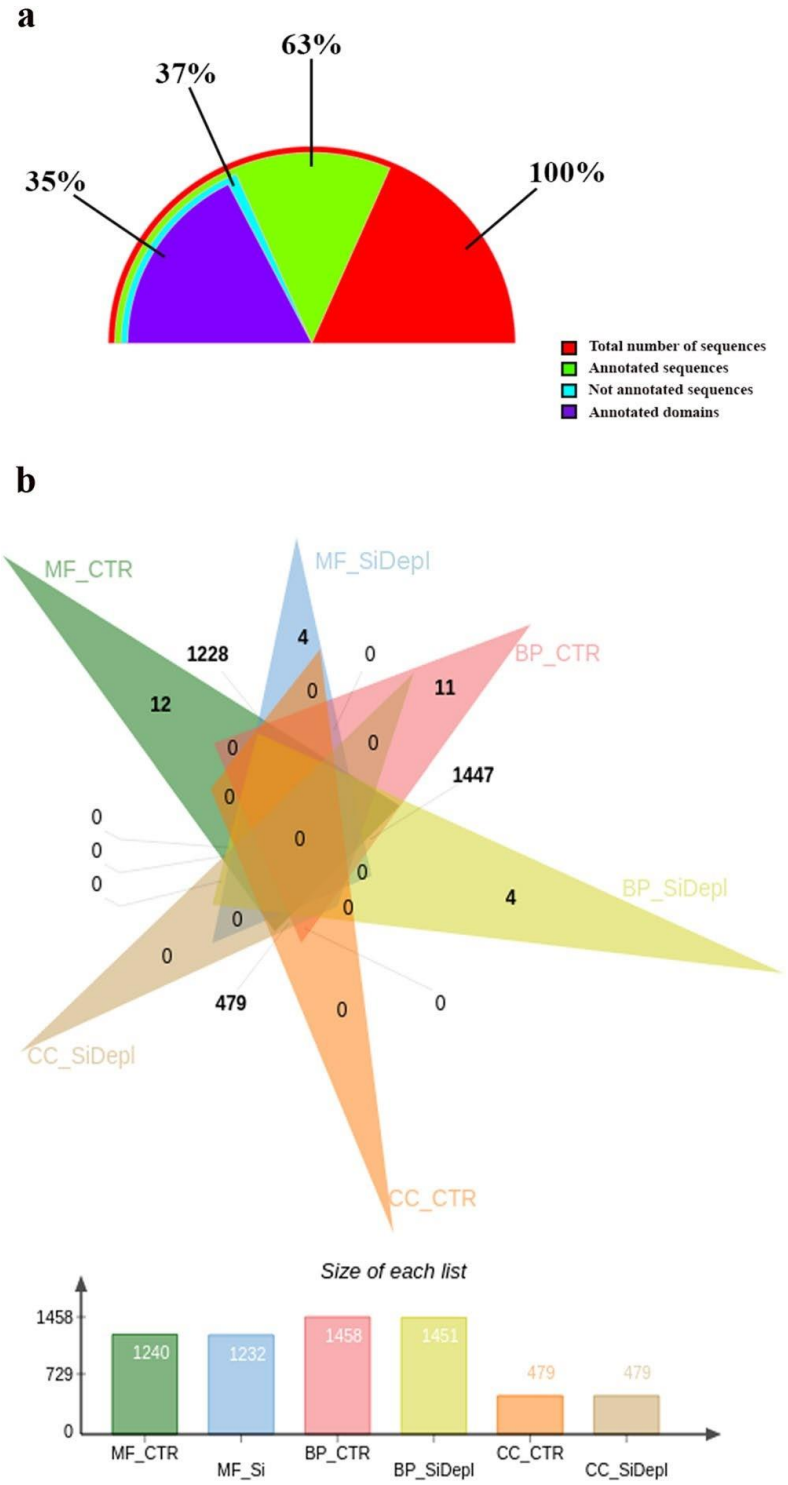


Figure 1.1. *Thalassiosira rotula* functional annotation. (**a**) Percentage of annotated sequences and domains of the total number of peptides in the diatom *Thalassiosira rotula*. (**b**) Venn diagram showing the number of common and unique GO terms between the two growth conditions for each category in the diatom *Thalassiosira rotula*. Abbreviations: MF: Molecular Functions; CTR: control condition; SiDepl: silica depleted medium; BP: Biological Processes; CC: Cellular Components.

Table 1.2. Number of sequences associated to each GO term category. Subdivision of the sequencing retrieving at least one blast hit and an association with a Gene Ontology number in the three functional categories: Molecular Functions (MF), Biological Processes (BP) and Cellular components (CC).

Number of sequences with at least one blast hit	37,999
Number of peptides sequences associated to Molecular Functions GO terms	26,255
Number of domains associated to Molecular Functions GO terms	11,622
Number of peptides sequences associated to Biological Processes GO terms	15,459
Number of domains associated to Biological Processes GO terms	7809
Number of peptides sequences associated to Cellular Components GO terms	17,794
Number of domains associated to Cellular Components GO terms	2694

1.3.2 Functional annotation analysis: SiDepl versus CTR

Principal Component Analysis showed that replicates 1 and 4, in the CTR and SiDepl group, respectively, did not behave coherently within the group, so they were excluded from subsequent analysis (Figure 1.2).

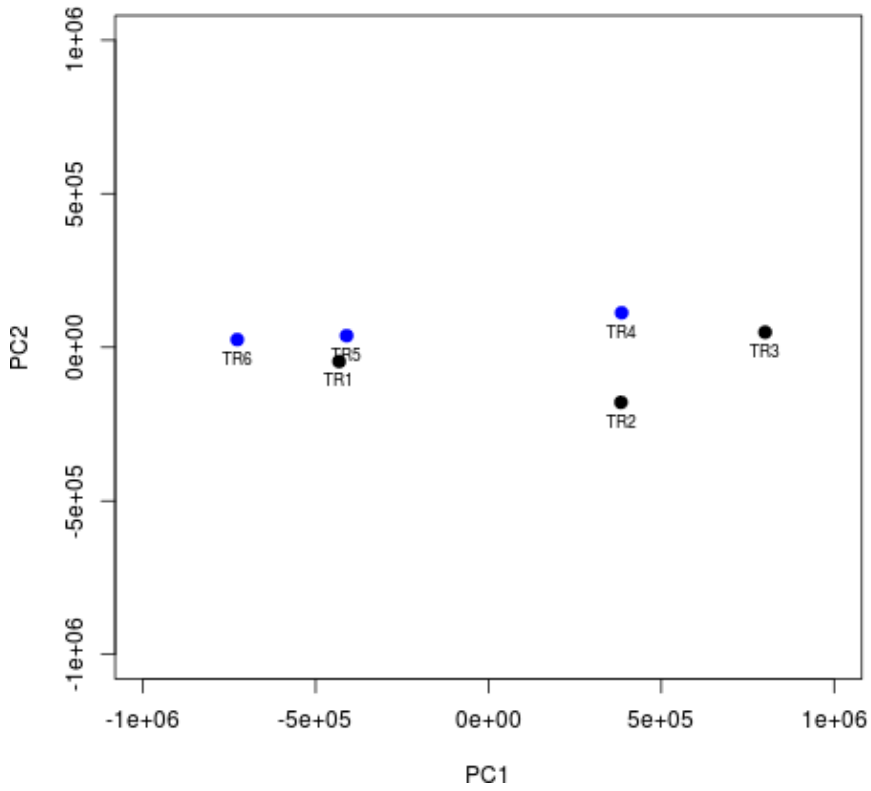


Figure 1.2. Principal component analysis (PCA) of each *Thalassiosira rotula* replicate, with PC1 plotted against PC2. In total, 6 cultures were used: three cultivated in normal condition and three cultivated in a medium depleted of silica. **TR1**, **TR2** and **TR3** correspond to the control replicates (cultures cultivated in complete medium). **TR4**, **TR5** and **TR6** correspond to stress replicates (cultures cultivated in silica depleted medium).

Fisher test on the number of GO terms associated with each peptide sequence for each treatment did not reveal any significant difference. However, we were able to identify uniquely annotated GO terms. The Venn diagram in Figure 1.1 b shows the occurrence of 12 and 4 unique MF GO terms in CTR and SiDepl, respectively. Similarly, 11 and 4 BP GO terms were uniquely annotated in CTR and SiDepl, respectively. No unique GO term annotations were present in the CC category for both treatments (Supplementary Data 9 for complete list of GO Terms). Table 1.3 lists the unique GO terms found. Except for 'Interleukin-8 biosynthetic process', 'lactate biosynthetic process', CTR specific, and '1-3-beta-D-glucan biosynthetic process', SiDepl specific in the BP category, all other terms referred to more broad categories (Table 1.3).

Table 1.3. Unique GO terms. Unique GO terms for each growth condition for each category: MF: Molecular Function; BP: Biological Processes; CC: Cellular Components.

Category	GO Number	GO Term	CTR	SiDepl
MF	0004062	ATP-dependent helicase activity		
	0008379	L-leucine transaminase activity		
	0008658	Metalloaminopeptidase activity		
	0015416	Oxidoreductase activity		
	0015430	Oxidoreductase activity, acting on a sulfur group of donors, disulfide as acceptor		
	0015604	Oxidoreductase activity, acting on iron-sulfur proteins as donors	+	–
	0005391	Hydro-lyase activity		
	0008909	NAPDH binding		
	0034189	Pyrophosphatase activity		
	0043621	RNA strand-exchange activity		
	0015439	snRNA stem-loop binding		
	0015439	Oxidoreductase activity, acting on a sulfur group of donors, NAD or NADP as acceptor		
	0003983	Anion transmembrane-transporting ATPase activity	–	+

	0008320	Lipid transporter activity		
	0008761	Methyltransferase activity		
	0009002	Nicotinate phosphoribosyltransferase activity		
BP	0001407	Body fluid secretion	+	−
	0007281	Interleukin-8 biosynthetic process		
	0008354	Lactate biosynthetic process		
	001096	Negative regulation by host of viral transcription		
	0032802	Post-chaperonin tubulin folding pathway		
	0042744	Purine ribonucleoside salvage		
	004523	Regulation of exit from mitosis		
	0044458	Regulation of protein phosphatase type 2A activity		
	0050427	RNA modification		
	0051923	Sterol biosynthetic process		
	0006011	Cellular amino acid metabolic process	−	+
	001932	Peptidyl-tyrosine dephosphorylation		
	0043953	Regulation of mRNA stability		
	0000022	(1-3)-beta_D-glucan biosynthetic process		

1.3.3 Differential expression analysis

In silico expression analysis on single transcripts identified 5,114 DET (Differential Expressed Transcripts) in SiDepl vs CTR, with 2,949 upregulated and 2,165 downregulated transcripts (Supplementary Data 10). The GO terms associated to the DETs are reported in SI1: Supplementary Table 1.2 and Supplementary Figure 1.2.

Enzymes in the chains for the sphingolipid, sterol, spermidine, secologanin, polyunsaturated fatty acids (PUFA), malonylCoA, lipotheicoic, L-ascorbate, L-cysteine biosynthesis and isoprenoid pathways were upregulated in SiDepl condition. Considering 2-fold change as the minimum value for significant results, the proportion of significant DET was 20 % for downregulated and 34 % for upregulated transcripts. Among these groups, 1,604 (54 %) of the upregulated and 1,278 (60 %) of the downregulated

transcripts were not annotated but indicated as “protein”, or “predicted protein” or “hypothetical protein” or “NA” (not annotated at all). It is interesting to note that part of these “un-annotated” groups were considerably differentially expressed. Indeed, 21 % of the “un-annotated”-upregulated transcripts had an expression fold change between +2.50 and +19.12, and 13 % of the “un-annotated”-downregulated transcripts had an expression fold change between −2.50 and −20.03 indicating the presence in these groups of still unknown functions that are very important for the homeostasis of the cells (Supplementary Data 10).

Of the downregulated transcripts, it is interesting to note that the ‘ribosomal protein s6 kinase beta 2’ showed a fold change of −18, whereas the upregulated transcripts ‘CMP-sialic acid transporter’ and ‘peroxisomal trans- 2-enoyl- reductase’ showed a fold change of +15 and +8, respectively.

1.3.4 *In silico* pathway expression analysis

Total number of general pathways, for both growth conditions, was 68 with all essential and secondary metabolic pathways present. The most represented pathways, based on the number of associated sequences, are those involved in carbohydrate, lipid and protein metabolism (Table 1.4).

Table 1.4. Top 10 most represented general pathways based on the number of associated sequences.

Pathway name	Number of associated sequences
Protein modification	456
Amino-acid biosynthesis	240
Cofactor biosynthesis	186
Lipid metabolism	175
Carbohydrate degradation	169
Amino-acid degradation	95
Porphyrin containing compounds metabolism	77
Purine metabolism	62
Sulphur metabolism	48
Carbohydrate biosynthesis	48
Carbohydrate metabolism	48

Analysis of general pathways expressed in FPKM (Fragments Per Kilobase of transcript per Million mapped reads) indicated some differences between the two conditions, CTR and SiDepl (Table 1.5). The 'protein biosynthesis' and 'flavonoid metabolism' pathways had almost three- and two-fold lower expression values in CTR with respect to SiDepl, respectively. 'Carbohydrate degradation' and 'thermoadapter biosynthesis' on the contrary were almost two-fold higher in CTR with respect to SiDepl. Despite these differences, FPKM box plot expression analysis revealed no statistically significant differences between the two conditions (Supplementary Table 1.3 (in SI1), Supplementary Files 1 and 2).

Table 1.5. Top 10 most expressed pathways in FPKM, Fragments Per Kilobase of transcript per Million mapped reads. Differentially expressed pathways are reported in bold.

Pathway Name	CTR_FPKM	SiDepl_FPKM	Fold SiDepl vs CTR
Protein biosynthesis	**88.73**	**245.92**	**+2.8**
Carbohydrate degradation	**42.97**	**26.17**	**−1.6**
Thermoadapter biosynthesis	**36,45**	**17.8**	**−1.9**
Flavonoid metabolism	**27.73**	**47**	**+1.7**
Organosulfur biosynthesis	24.26	19.77	
Carbohydrate biosynthesis	20.89	16.51	
Lipid metabolism	19.65	21.85	
Spore coat biogenesis	18.6	28.11	
Carbohydrate acid metabolism	18.07	17.59	
Nucleotide-sugar biosynthesis	15.07	21.1	

The number of pathways identified in both conditions on the second level of annotation was 258 (Supplementary Data 11). The most represented pathways were 'protein ubiquitination' (293 sequences) followed by 'protein glycosylation' (127 sequences) and 'glycolysis' (105 sequences). Also within this level of pathway annotation there were no statistically significant differences in FPKM values between the two conditions (Supplementary Files 3 and 4).

The pipeline annotated also a third level with 327 pathways for CTR and 325 for SiDepl (Supplementary Data 12). The best represented pathways were: 'Pyruvate from D-glyceraldehyde 3-phosphate step 3/5' with 26 associated sequences; 'D-ribulose 5-phosphate from D-glucose 6-phosphate (oxidative stage) step 1/3' with 25 associated

sequences, ‘glutathione from L-cysteine and L-glutamate step 1/2’ with 21 associated sequences, ‘D-glyceraldehyde 3-phosphate and glycerone phosphate from D-glucose step 4/4’ with 20 associated sequences, ‘L-cysteine from L-serine’ step 2/2’ with 20 associated sequences. Again, no differences in FPKM values were appreciable (Supplementary Files 5 and 6).

1.3.5 *In silico* expression analysis of selected pathways

Despite the fact that no significant differences were evident in the FPKM expression of the annotated pathways, we focused our attention on single transcripts associated with pathways leading to secondary metabolites that have been poorly studied or that were previously unknown for diatoms: secologanin, prostaglandin and polyketides biosynthesis (Supplementary File 4, highlighted pathways).

1.3.6 Secologanin biosynthesis

The ‘Secologanin synthase (SLS)’ annotation in the transcriptome of *T. rotula* was associated to four peptides that when aligned matched one another, with the peptide TR26037|c1_g1_i1 being the longest (SI1: Supplementary Figure 1.4). Heat map in Figure 1.3 a shows expression levels of the corresponding transcript, calculated in FPKM, which was ten times higher in SiDepl with respect to CTR (Figure 1.3). In addition, the related transcript was also present in the DET table with a logFC of +3.1 (Supplementary Data 10).

1.3.7 Prostaglandin biosynthesis

Peptide TR754|c0_g1_i1 annotated as ‘Prostaglandin G/H synthase 2 (PGHS-2)’, TR25058|_c1_g1_i2 annotated as ‘Prostaglandin E synthase 2-like (mPGES-2)’ and TR11659|c0_g1_ i1 annotated as ‘Prostaglandin F synthase (PGFS)’ were associated to the ‘Prostaglandin biosynthesis’ pathway (Supplementary Data 13). In addition, three putative peptides and two transcripts were associated to the ‘Prostaglandin F synthase’ function. Table 1.6 shows the blast results of the five PGFS associated transcripts. TR11659|c0_g1_i1 was chosen for the following analysis since this transcript blasted to other *Thalassiosira* species and to a dinoflagellate PGFS function. Moreover, the match

with *T. pseudonana* was a protein annotated with the more general function of aldo/keto oxidoreductase, which is a family of proteins that also include PGFS.

Furthermore, browsing the 'Enzyme Description' column of the annotation table (Supplementary Data 13) we found six peptides annotated as 'Carbonyl reductase (NADPH)' or 'Prostaglandin E_2 9-reductase' that reduces prostaglandin E_2 (PGE_2) to prostaglandin F2α (PGF2α) (SI1: Supplementary Table 1.4) differently from PGFS that produces PGF2α starting from prostaglandin D_2 (PGD_2). However, we did not include this enzyme in our analysis as it was not annotated in the specific pathway. Heat Map (Figure 1.3 a) shows no difference in mPGES-2 expression calculated in FPKM, between the two growth conditions, while PGFS and PGHS-2 expression were downregulated in SiDepl by two- and three-fold, respectively. PGHS-2 was also present in the DET table with a LogFC of −1.57 (Supplementary Data 10).

Table 1.6. Blast results of the prostaglandin F synthase related transcript.

Transcript	Annotation on:	First three BlastX results	Query cover (%)	Identity (%)
TR11659|c0_g1_i1	Peptide	Aldo-keto oxidoreductase *Thalassiosira pseudonana*	76	63
		Hypothetical protein *Thalassiosira ocenanica*	74	58
		Prostaglandin F synthase *Symbiodinium microadriaticum*	70	45
TR3003|c0_g1_i1	Peptide	Hypothetical protein *Thalassiosira pseudonana*	72	66
		Reductase with NAD or NADP as acceptor *Phaeodactylum tricornutum*	76	48
		Aldo/keto reductase *Fragilariopsis cylindrus*	69	48
TR17558|c0_g1_i1	Peptide	Hypothetical protein *Thalassiosira ocenanica*	42	42
		Hypothetical protein *Fistulifera solaris*	41	44

		Acetyl-CoA synthetase *Fistulifera solaris*	41	43
TR22640|c0 _g1_i1	Transcript	Predicted protein *Thalassiosira pseudonana*	72	75
		Hypothetical protein *Thalassiosira oceanica*	76	60
		Hypothetical protein *Fistulifera solaris*	58	59
TR60026|c0 _g1_i1	Transcript	Predicted protein *Thalassiosira pseudonana*	67	82
		Hypothetical protein *Thalassiosira oceanica*	71	74
		Hypothetical protein *Fistulifera solaris*	55	54

1.3.8 Polyketides biosynthesis

No specific pathway annotation exists for polyketides biosynthesis as they are included in the general lipid metabolism pathway. Searching in the complete annotation table (Supplementary Data 13), we found polyketide-related functional annotations of at least 4 putative enzymes, i.e. 'Polyketide cyclase dehydrase and lipid transport protein' (PK_Cyc, TR3754|c0_g1_i2), 'Polyketide enoylreductase' (PK_ER, TR30039|c0_g1_i1), 'Polyketide hydroxylase-like' (PK_Hxl, TR21208|c0_g1_i1), 'Phthiocerol synthesis polyketide synthase type I' (ppsC, TR40329|c0_g1_i6). Except for ppsC synthase, whose related transcript was downregulated in the SiDepl samples by two-fold (in FPKM value) and had a LogFC value equal to −0.99 in DET table (Supplementary Data 10), the other three related transcripts were almost equal in expression with low FPKM values in both growth conditions (Figure 1.3 a).

Finally, we also found annotations for single modules associated to both polyketide synthesis and fatty acid elongation, like acyltransferase, acyl carrier protein, β-ketoacyl reductase and ketosynthase, methyl transferases, thioesterases and dehydrogenase domains (Supplementary Data 13 to 20). These transcripts were not considered for the expression analysis due to their high number of redundant related transcripts.

1.3.9 Comparative analysis of selected pathways with other strains of *Thalassiosira*

To understand if selected pathway expression in the transcriptome of *T. rotula* is a prerogative of the strain from the Mediterranean, we used the public available MMETSP transcriptomic dataset to verify their presence in other *T. rotula* strains, 3096 and GSO102, isolated from different geographical areas and grown in different conditions (Table 1.7 and Figure 1.3 b).

The available conditions were: 3096LC = day portion of the day/night cycle; 3096DC = dark portion of the day/night cycle; GSO102Fe = iron depletion; GSO102P = phosphorus depletion; GSO102Fe2 = iron and phosphorus depletion; GSO102R = nutrient replete.

Secologanin synthase was present in the 3096LC and GSO102Fe2's transcriptomes but not in those of 3096DC, GSO102Fe, GSO102P and GSO102R (Table 1.7, Figure 1.3 b). Of all the conditions considered in which we found SLS annotated, silica depletion and the light portion of the day/night cycle were similar and had the highest values (41.02 FPKM and 39.6 FPKM, respectively) compared to the other conditions.

Among the polyketide biosynthesis enzymes, PKS_ER annotation has not been reported for other strains. The synthase ppsC was present only in 3096DC and the best condition of expression was found in our clone in CTR. The hydrolase-lyase PK_HxL was present in both 3096LC and 3096DC. It was present with two isoforms, here named as PK_Hlx_1 and PK_Hlx_2, in 3096LC and with one isoform, PK_Hlx_1, in 3096DC and in our clone in CTR. The highest expression for PK_Hlx_1 was in the 3096DC condition. The PK-Hxl_2, present only in 3096LC, showed a higher FPKM value than PK_Hlx_1. The GSO102 strain, in all conditions tested, did not show any polyketide related enzyme annotation. Finally, all strains had annotations for the single modules associated to both polyketide synthesis and fatty acid elongation (Supplementary Data 13 to 20).

Examining the prostaglandin pathway, we did not find any annotation for PGHS-2 in strains 3096 and GSO102. On the contrary, PGFS was present in all strains, together with a second isoform present in 3096LC, 3096DC, GSO102P and GSO102Fe2. The transcript for mPGES-2 was present only in the 3096LC transcriptome. Finally, annotation for prostaglandin D synthase (PTGDS) that was absent in our transcriptomes, was instead present in the 3096LC, 3096DC, GSO102P and GSOFe2 transcriptomes. The best condition of expression was 3096DC for PTGDS, 3096LC and CTR for mPGES-2, GSO102Fe and GSO102Fe2 for PGFS_1 and 3096DC for PGFS_2.

Table 1.7. Presence of the selected pathways in different *Thalassiosira rotula* strains. N= not present; Y= present

***T. rotula* Strain**	CCMP 3096	CCMP 3096	GSO 102	GSO 102	GSO 102	GSO 102	CCMP1647	CCMP1647
MMETSP code	0403	0404	910	911	912	913	—	—
Treatment	Light cycle	Dark Cycle	0.4um PO_4	4nM Fe, 12.5um EDTA	4nM Fe, 12.5um EDTA, 0.4uM PO_4	replete condition	Complete Medium	$SiOH_4$ Limited
Site Name	Pacific Ocean	Pacific Ocean	Pacific Ocean	Pacific Ocean	Pacific Ocean	Pacific Ocean	Mediterranean sea	Mediterranean sea
Prostaglandin G/H synthase (PGHS-2)	N	N	N	N	N	N	Y	Y
Prostaglandin E synthase (mPGES-2)	Y	N	N	N	N	N	Y	Y
Prostaglandin F synthase (PGFS)	Y	Y	Y	Y	Y	Y	Y	Y
Prostaglandin D synthase (PTGDS)	Y	Y	Y	N	Y	N	N	N
Phthiocerol synthesis polyketide synthase type I (ppsC)	N	Y	N	N	N	N	Y	Y
Polyketide cyclase dehydrase and lipid transport protein (PYL5)	N	N	N	N	N	N	Y	Y
Polyketide enoylreductase (PK_ER)	N	N	N	N	N	N	Y	Y
Polyketide hydroxylase-like (PK_HxL)	Y	Y	N	N	N	N	Y	Y
Secologanin synthase (SLS)	Y	N	N	N	Y	N	Y	Y

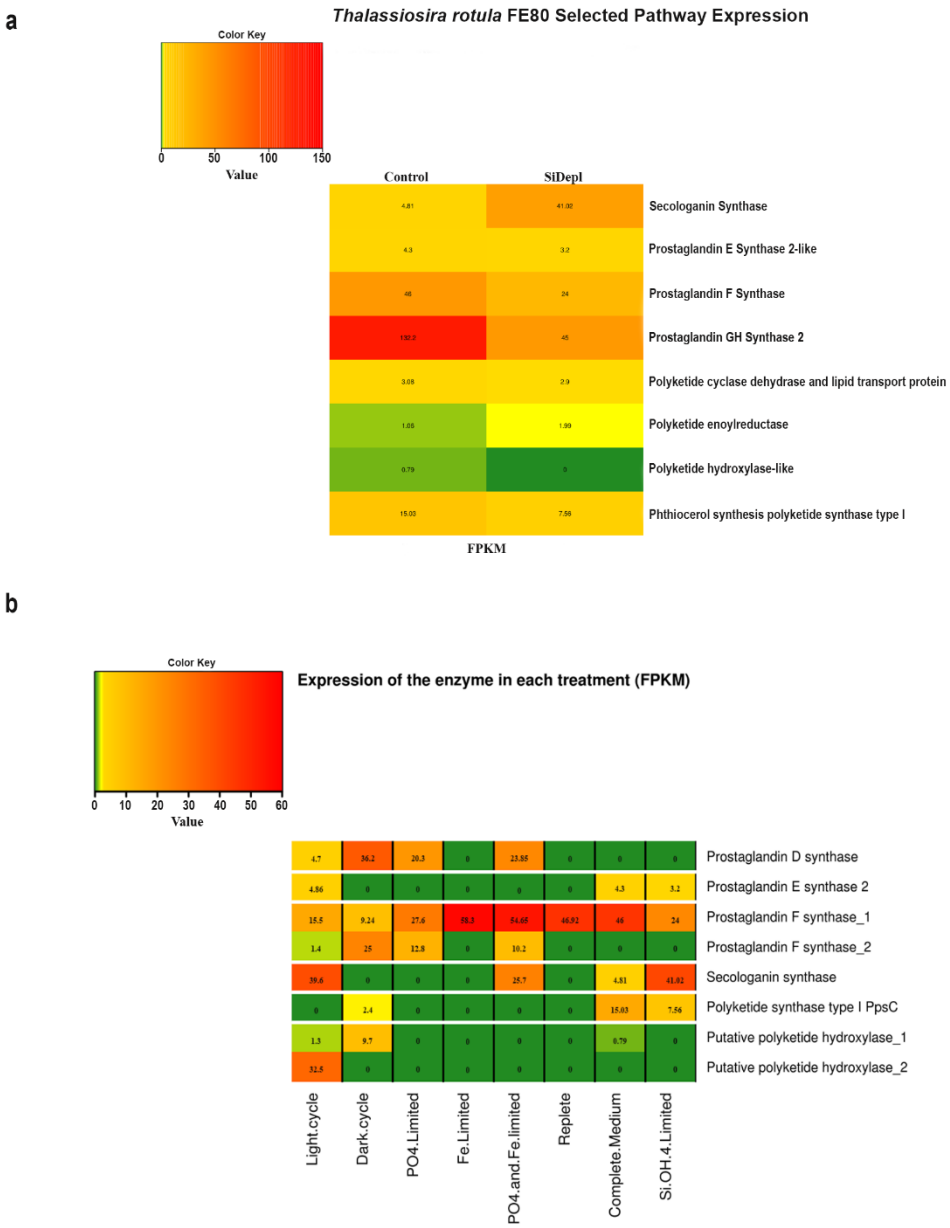


Figure 1.3. *In silico* expression analysis of selected pathways. (**a**) Heat map showing the presence and the expression levels, calculated in FPKM on the contig's reads, of transcripts, belonging to our *Thalassiosira rotula* from the Mediterranean Sea, in the selected pathways for each growth condition considered. (**b**) Heat map showing the presence and the expression levels, calculated in FPKM on the contig's reads of the transcripts in the selected pathways of the *Thalassiosira rotula* strains, grown in different conditions, from the MMETSP project in comparison with our clone from the Mediterranean Sea (Complete medium and $SiOH_4$ limited).

1.3.10 Experimental confirmation of selected pathways

In order to confirm *in silico* expression data, we performed quantitative Polymerase Chain Reaction (qPCR) amplification on coding DNA (cDNA) from the same sample used for

RNA-seq and qPCR amplification on genomic DNA (gDNA) from an axenic *T. rotula* (Ax) and a non-axenic *T. rotula* (Mx) culture (Figure 1.4) in which bacterial DNA was predominant (unpublished data).

1.3.11 Secologanin synthase

qPCR on cDNA confirmed the significant overexpression of Secologanin synthase (SLS) in SiDepl with respect to CTR with a p-value of 0.0062 (Figure 1.4 a (1)). PCR on Ax- and Mx-gDNA confirmed the presence of the corresponding gene exclusively in the algal genome excluding a co-partnership with the associated bacteria for the possible production of secologanin (Figure 1.4 a (2)). qPCR amplification on Ax-gDNA had a cycle threshold (Ct) value of 21,6 compared to 31,5 of the amplification on the Mx-gDNA, confirming the algal specific expression of SLS (Figure 1.4 a (3)). Melt curve analysis showed one peak for the Ax-gDNA overlapping with the cDNA peak, while the Mx-gDNA showed two separated peaks. Sequencing of the corresponding amplicon definitively demonstrated that the additional amplicon from the Mx culture was non-specific.

1.3.12 Prostaglandin biosynthesis

qPCR (Figure 1.4 b (1)) showed a significant downregulation of all three transcripts in SiDepl with p-value = 0.02 for PGHS-2, 0.01 for mPGES-2 and 0.07 for PGFS. PGHS-2 had the highest level of expression in both conditions with respect to PGFS and mPGES-2 that had expression levels in the order of 10^{-2} and 10^{-3}, respectively.

Amplifications on gDNAs (Figure 1.4 b (2)) and sequencing of the amplicons demonstrated the specific origin of the corresponding transcripts from the algae and not from the associated bacteria. All three transcript amplicons had a Ct value equal to about 21 on the Ax-gDNA and equal to about 33 on the Mx-gDNA, confirming the specific derivation of the prostaglandin biosynthetic pathway transcripts from the alga and not from the associated bacteria (Figure 1.4 b (3A–C)). Differently from PGHS-2 and mPGES-2, PGFS melt curve on non-axenic Mx-gDNA showed one peak that did not overlap with that of the Ax-gDNA. Sequencing of the corresponding amplicon demonstrated its lack of specificity. The absence of a peak corresponding to the one present in the Ax-gDNA is probably due to its very low abundance also considering the low presence of eukaryotic DNA in the sample (unpublished data).

1.3.13 Polyketides biosynthesis

qPCR (Figure 1.4 c (1)) confirmed the results of the FPKM calculation showing no regulation of PKS_ER, and PK_Hxl but a significant downregulation of the ppsC transcript with a p-value = 0.03. Classical PCR amplification (Figure 1.4 c (2)), amplicon sequencing and qPCR analysis on gDNA (amplification and melt curve, Figure 1.4 c (3A–D) of ppsC, PK_Hxl, PKS_ER, of the Ax-gDNA, demonstrated the specificity of this transcript to the eukaryotic alga.

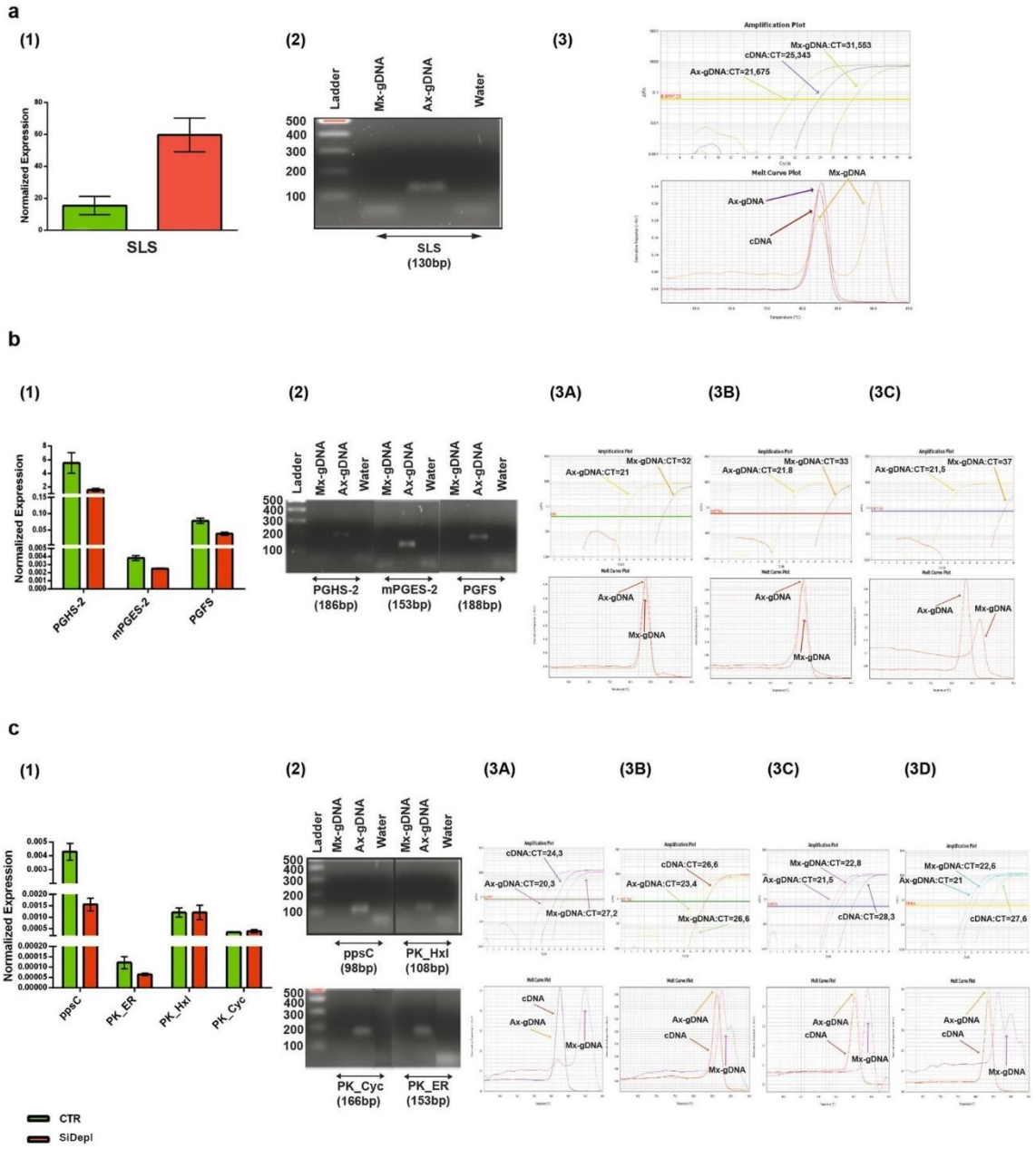


Figure 1.4. Experimental validation of the transcripts associated to the selected pathways in the *Thalassiosira rotula* from the Mediterranean Sea. (**a**–**c**) qPCR on cDNA (1), PCR on gDNA (2), qPCR on gDNA (3). (**a**) Secologanin synthase; (**b**) Prostaglandin biosynthesis; (**c**) Polyketides biosynthesis. Abbreviations: PGHS-2: Prostaglandin G/H synthase 2; mPGES-2: Prostaglandin E synthase 2-like; PGFS:

Prostaglandin F synthase; ppsC: Phthiocerol synthesis polyketide synthase type I; PK_ER: Polyketide enoylreductase; PK_Cyc: Polyketide cyclase dehydrase and lipid transport protein; PK_Hxl: Polyketide hydroxylase-like; Ax: Axenic genomic DNA; Mx: mixed Bacterial- Algae genomic DNA; CT: cycle threshold. Original gel picture is shown in Supplementary Figure 1.9.

1.3.14 Sequencing of the *T. rotula* axenic genome.

To further explore the potentiality of *T. rotula* to produce valuable bioactive molecules, we also sequenced its axenic genome. The best assembly was obtained using the Join Redundans pipeline, which results are showed in Table 1.8.

Table 1.8: Results of the assembly of *T. rotula* axenic genome performed with the Join Redundans pipeline.

Contigs (n=)	2941
N50 (Kbp)	531
Length of the smallest contigs (bp)	501
Length of the longest contigs (Mbp)	4.1
Mapped Illumina sequences (%)	99.23
Mapped PacBio sequences (%)	96.65
Completeness of the genome compared to the Stramenopiles group (%)	95
Completeness of the genome compared to the Eukaryotes group (%)	25.5
Estimate size of the genome (Mbp)	673

The completeness of the genome compared to the Eukaryotes is of 25.5 %, consistently with what has been observed for the genome of *T. pseudonana*[14].
In contrast with the other ten diatom genomes sequenced until now, whose maximum size was ~200 Mbp, our genome resulted be extremely bigger being of 673 Mbp, despite the number of genes was similar, consisting of about 35.000 genes (Figure 1.5).

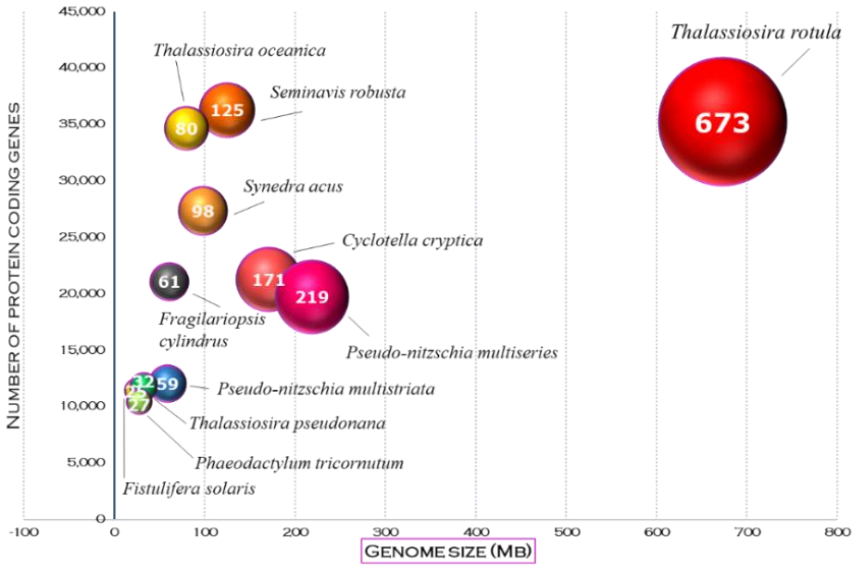


Figure 1.5. Bubble plot representing the genome sizes in relation to the number of protein coding genes for all the diatoms which genome is available in comparison with *T. rotula* genome size.

The analysis of the genomic sequences with two different pipelines specific for the identification of transposable elements highlighted their presence for about the 63 % of the total genomic size. This result is very interesting, given the growing credit that these sequences are assuming as fundamental elements for the evolution and regulation of gene expression, especially in response to environmental changes, and needs to be further investigated.

1.4 Discussion

Expression analysis to identify transcripts and differentially expressed metabolic pathways between control and silica-limited growth conditions highlighted the presence of a large group of uncharacterized proteins in the diatom *Thalassiosira rotula*, which were differentially expressed between the two growth conditions, representing, as such, a large source of potentially still undiscovered new functions and correlated metabolites. Among the differentially expressed annotated sequences, we found that Ribosomal protein s6 kinase beta 2 (S6K2) was the one with the highest downregulation (−18.03-fold change) in the silica-limited condition with respect to the control. S6K2 is an important cell growth regulator and metabolism and its downregulation is consistent with the status of starvation induced by silica limitation[15]. S6K2 is indeed activated in response to growth factors, cytokines and nutrients, together with other kinases of the same family. The activity of S6 kinases is linked to fundamental cellular processes, including protein synthesis, mRNA processing, glucose homeostasis, cell growth and survival. They also play an important role in several human pathologies, including obesity, diabetes, ageing and cancer[16].

Upregulated functions in silica-limited conditions with respect to the control include upregulation of CMP-sialic acid transporter (+15.15) and peroxisomal trans-2-enoyl-reductase (+8), indicating that cells are coping with starvation by accumulating defensive molecules and lipids confirming again the results reported by Heydarizadeh in which light stress and carbon deprivation stress redirect the CCM toward lipids accumulation[8,9]. CMP-sialic acid is one of a series of nucleotide sugar transporters, a class of transporters that have been associated with inflammation, tumours, virus infections, and other processes[17]. It transports CMP-sialic acid from the cytosol into Golgi vesicles where glycosyltransferases are activated and produce sialilated glycolipids that can act as defensive molecules. The presence of sialic acids has been ascertained only in animals, even though recently they have also been discovered in plants[18,19] and in green microalgae[20], but their role is still uncertain.

Peroxisomal trans-2-enoyl-reductase seems to be involved in the peroxisomal degradation of phytols, a product derived from chlorophyll degradation, and has also been proposed to participate in a peroxisomal fatty acid chain elongation system, indicating that cells are probably accumulating long fatty acids in response to starvation status[21].

The CTR-specific uniquely annotated GO terms, ‘Interleukin-8 biosynthetic process’ and ‘lactate biosynthetic process’, and the SiDepl specific ‘1-3-beta-D-glucan biosynthetic

process' are all involved in defensive systems and in pathogenic mechanisms in humans[22,23]. It would thus be very interesting to further investigate these processes to unveil their role in diatoms. Worthy of mention is also the upregulation of the 'flavonoid metabolism' pathway in SiDepl, which suggests that this condition could be useful to possibly increase the yield of this important class of molecules that are beneficial for human health[24].

In addition to the above-mentioned differentially expressed transcripts, also enzymes in the chains for sphingolipid, sterol, spermidine, PUFA, malonylCoA, lipotheicoic, l-ascorbate, l-cysteine biosynthesis, and isoprenoid pathways were upregulated in SiDepl confirming that the depletion of silica is a good condition to stimulate the synthesis of bioactive molecules.

Furthermore, with our study we show for the first time, that diatoms are capable of synthesising secondary metabolites such as polyketides and secologanin, the latter at increased levels under silica starvation. Moreover, we confirm the presence of genes that synthesize animal hormone-like molecules, such as prostaglandins also in this diatom species[25], *Thalassiosira rotula*, other than *Skeletonema marinoi* already published in the 2017.

Secologanin is involved in a complex metabolic pathway leading to the synthesis of Monoterpenoid Indole Alkaloids (MIAs), a large class of chemically different bioactive molecules considered very attractive due to their activity against cancer and other severe diseases in humans[26]. Some of the most active anticancer drugs such as camptothecans and vinca alkaloids[27] have this type skeleton. Secologanin is a precursor for many of these compounds that are present in low amounts in the Madagascar periwinkle *Catharanthus roseus* leaves[28,29]. The enzymes involved in its formation are target of increasing research efforts in the phytochemical area, but to date they have been identified only in a few plant species. Discovering secologanin synthase in the genome of a diatom species paves the way to a completely new research opportunity for the understanding of the synthetic pathway and the role of secologanin in eukaryotic unicellular photosynthetic organisms.

Surprisingly, we found the transcripts for enzymes involved in polyketide biosynthesis. Polyketides (PKs) are an important class of molecules with a broad range of biological and pharmacological activities[30]. They include toxins, siderophores, pigments, antibiotics, cytostatic, and immune-suppressants. PK Synthases (PKSs) are widely distributed in bacteria and sporadically in Archaea and Eukarya where they probably derive from Bacteria via horizontal gene transfer[31–33]. There has been no report until now on the existence of polyketide enzymes in a diatom species even if this is not the first

attempt to find them in diatoms. Kohli *et al*. (2016) made an evolutionary study on PKSs in marine microbial genera, including diatoms, utilizing the data in the MMETSP database. The species they considered also included the *T. rotula* strains we used in our *in silico* analysis. However, they were unable to find any PKS in their transcriptome, nor in all the Bacillariophytes they took into consideration[30] whereas we found a low expression of PKS in the strain 3096 transcriptome, both in the dark (DC) and light portion (LC) of the day/night cycle. A possible reason for this discrepancy is the low sensitivity of the annotation pipeline used for data analysis in the MMETSP project, while for our analysis, we used Annocript[34], an assembler pipeline specifically developed in our institute that differs from the ones used in the MMETSP project.

Prostaglandins are a very interesting class of PUFA-derived molecules acting as hormones that play a pivotal role in many physiological processes in animals[35]. They are present not only in mammals, but also in both marine and terrestrial invertebrates[36] and only few studies report their occurrence in terrestrial plants[37].

Some macroalgae also possess the prostaglandin pathway, but only a limited number of prostaglandin molecules have been described from this source[38,39]. Recently, we were able to demonstrate the presence of this pathway also in diatoms, showing that *Skeletonema marinoi* has the ability to synthesize the complete panel of three classes of prostaglandins[25]. With the present study, we demonstrate the expression of the prostaglandin pathway in another diatom species and the significant downregulation of the first enzyme, PGHS-2, induced by silica depletion, differently from *S. marinoi* in which we did not observe any regulation of the pathway by nutrient limitation[25].

Overall, the results obtained from the transcriptomic analysis of *T. rotula*, provide evidence for the presence of biosynthetic pathways involved in the production of secondary metabolites with potentially interesting biotechnological applications and the possibility to modulate their expression by varying culturing conditions. Although the results do not definitely provide information on the conditions to stimulate secondary metabolite production, the use of different transcriptomes used together may facilitate the understanding of the most suitable conditions for the expression of the metabolites of interest. Another criticism to our work could be that our *T. rotula* strain from the Mediterranean Sea (Gulf of Naples, Italy) and *T. rotula* strains from other geographic regions do not represent the same species. However, a recent phylogenetic study by Whittaker *et al*.[40] among different *Thalassiosira* species and strains, including a *T. rotula* strain from the Gulf of Naples, concluded that "*T. rotula* lineages should be considered as a single species". Altogether, our results provide new stimuli to investigate the role

and function of these newly discovered molecules in diatoms, especially as concerns their interactions with other phytoplankton species and with the surrounding environment.
Finally, the incredibly large dimension of the *T. rotula* axenic genome hosting a high quantity of transposable elements, allows to hypothesize a role for *T. rotula* as model species to study transposons in diatoms.

1.5 Materials and Methods

1.5.1 Strain culturing and RNA extraction

T. rotula, strain CCMP1647, lab-name FE80, was isolated in 2011 in the Gulf of Naples (40°48.5′N, 14°15′E), Mediterranean Sea. Clonal cultures were established by isolating single cells from phytoplankton net samples collected from the surface layer of the water column. Cultures were grown in sterile filtered oligotrophic seawater, with a salinity of 36 ppt, amended with f/2[41] nutrients at a temperature of 18–20 °C, at 12:12 h light:dark cycle, with a photon flux of 100 μ mol photons m^{-2} s^{-1}, under agitation true filtered air influx.

2 L cultures, containing 2500 cells/mL, were grown in complete f/2 (CTR) and in silica depleted f/2 media with 36 μM Na_2SiO_3 (SiDepl), in triplicates, until they reached the stationary growth phase (day 6 for CTR cultures and day 7 for the SiDepl cultures). Cultures were harvested by filtration onto 1.2 μm pore size filters (RAWP04700, Millipore, Burlington, Massachusetts, USA) and immediately frozen in liquid nitrogen.

10 L cultures of *T. rotula*, in triplicate, were used to follow their growth from day 3 to day 10. Every day, 250 mL of each culture was harvested by filtration onto 1.2 μm pore size filters (RAWP04700, Millipore, Burlington, Massachusetts, USA) and immediately frozen in liquid nitrogen, and 100–200 mL of culture media recovered from the cell filtration was collected and stored at −80 °C until sample processing.

Total RNA was extracted with the RNeasy Micro Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions including a step with DNase digestion. RNA concentration was determined using a Qubit® 2.0 Fluorometer (Invitrogen, Waltham, Massachusetts, USA) and a quality check was performed by gel electrophoresis (1 % agarose w/v) and an Agilent2100 bioanalyzer (Agilent Technologies, Santa Clara, California, USA).

1.5.2 Antibiotic Treatment to produce axenic cultures

1 mL of exponentially growing culture was inoculated in f/2 medium containing final concentrations of 0.1 mg/mL Streptomycin (A1852, PanReac Applichem), 0.06 mg/mL Penicillin (A1837, PanReac Applichem), 1 mg/mL Ampicillin (A0839, PanReac Applichem), 0.1 mg/mL Kanamicine (A1493, PanReac Applichem) and 0.02 mg/mL Cefotaxime (C7039, Sigma-Aldrich, St. Louis, Missouri, USA) and allowed to grow for 5–6 days under standard growth conditions. Bacterial contamination was checked in two

ways (SI1: Supplementary Figure 1.8): (i) by staining DNA with DAPI and examining cultures under the microscope to check for the presence/absence of bacterial nucleoids; (ii) by performing peptone tests. For DAPI staining, 1 µL of DAPI stock solution (4′,6-diamidino-2-phenylindole, 1 mg/mL, Roche, Basel, Switzerland) was added to 1 mL of formalin preserved culture, incubated for 10 min and observed under the epifluorescence microscope. For peptone tests, 1 mL of diatom culture was added to a tube containing a peptone solution (1 mg/mL), incubated in the dark and checked after 2–3 days and 1–2 weeks; growth of bacteria in the tubes indicated contamination. If bacterial contamination persisted, the treatment was repeated. Large volume cultures used for DNA extraction were grown with antibiotics and the contamination tests were always performed on an aliquot of the culture.

1.5.3 DNA extraction

Axenic *T. rotula* cells (strain FE80) were collected onto a 1.2 µm RAWP membrane filter (Millipore, Burlington, Massachusetts, USA). The filter was rinsed with 1.5 mL seawater and cells were further collected into Eppendorf tubes and pelleted by centrifugation at 6000 rpm at 4 °C for 5 min (Eppendorf Centrifuge 5810 R, Eppendorf, Amburg, Germany). The DNA was extracted following a Phenol-Chloroform extraction method[42] with slight modifications that included cell disruption by adding 400 mg of 0.2–0.3 mm diameter silica beads and vortex mixing at 30 hrz for 85 s (3 times), and cooling the pellet on ice between the vortex mixings. The extracted DNA was ethanol precipitated, air dried, dissolved in 50 µL of sterile water and stored at −20 °C until sequencing.

1.5.4 Axenic gDNA sequencing, assembly and annotation

Sequencing, assembling and annotation of the axenic gDNA have been performed by the Sequentia Biotech company (Barcelona, Spain). A total of two sequencing runs with the Illumina platform (Illumina, San Diego, California, USA) and three sequencing run with the Pac-Bio Sequel I system (Pacific Biosciences, Menlo Park, California, USA), using the real-time single molecule sequencing technology (SMRT), were performed in order to obtain the sequencing of long genomic fragments and an adequate coverage. The assembly of the genomic sequences with the transcriptomic ones obtained previously, was performed with the following pipelines: Canu, Canu Redundans, Flye, Flye Redundans, Join, Join Redundans, the latter giving the best results. The software Pilon

(https://github.com/broadinstitute/pilon) was used to correct the PacBio raw assemblies with the Illumina data and they were merged to create a high quality consensus assembly. Finally, the algorithms Tephra and EDTA have been applied for the search of transposons (Personal communication). The annotation of *T. rotula* axenic genome is still in progress. For the genome structural and functional annotation, the *de novo* assembly of *T. rotula* transcriptome will be used to identify and annotate the coding genes.

1.5.5 RNA sequencing

Next generation sequencing experiments, including sample quality control, were performed by Genomix4life S.R.L. (Baronissi, Salerno, Italy). Indexed libraries were prepared from 2 μg/each purified RNA with TruSeq Stranded mRNA Sample Prep Kit (Illumina, San Diego, California, USA) according to the manufacturer's instructions. Libraries were quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA) and pooled such that each index-tagged sample was present in equimolar amounts, with a final concentration of the pooled samples of 2 nM.
The pooled samples were subject to cluster generation and sequencing using an Illumina HiSeq 2500 System (Illumina, San Diego, California, USA) in a 2 × 100 paired-end format at a final concentration of 8 pmol.

1.5.6 Transcriptome assembly and annotation

Illumina paired-end 100 bp reads from 6 *T. rotula* samples were processed to produce the transcriptome assembly. Raw reads were trimmed and clipped with BBDuk (https://jgi.doe.gov/data-and-tools/bbtools/) setting a minimum Phred-like quality of 35 and a minimum length of 35 nucleotides.
The quality of the reads before and after trimming was checked with the software FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Possible human contaminant reads were removed after mapping the high quality reads against the reference human genome (GRCh38) with STAR (v 2.4.0j)[43]. High quality reads were then normalized with Trinity[44] using the options: –SS_lib_type RF–pairs_together–max_cov 50. *De novo* transcriptome assembly was then performed with Trinity using the options: –SS_lib_type RF–no_normalize_reads–min_kmer_cov 1–KMER_SIZE 32. Transcriptome redundancy was removed with CD-HIT-EST[45] using the following options: -r 0 -g 1. A filter for contaminants was performed by BLASTing the transcripts

against the NCBI nr database, discarding all the sequences having a significant hit (e-value < = 0.0001) against Bacteria or Metazoa.

The completeness of the assembly was checked against the Core Eukaryotic Genes database (http://korflab.ucdavis.edu/Datasets/genome_completeness/ and http://korflab.ucdavis.edu/datasets/cegma/, access date: January to March 2018). *In silico* translation was performed with TransDecoder[46] whereas Functional Annotation was performed with Blast2GO software[47]. Transcriptome assembly, general transcriptome and general proteome statistics are reported in Supplementary Tables 1.1 a-c.

1.5.7 Transcriptome expression quantification and differential expression analysis

Transcript expression quantification was performed using Express (v 1.5.1)[48] after mapping the reads against the assembly with STAR[43]. Posterior counts were used as input to perform transcript differential expression analysis with EBSeq[49] transcripts with a probability of being differentially expressed higher than 0.95 were considered significant. The contigs and peptide sequences for *T. rotula* FE80 (CCMP1647) can be found in Supplementary Files 7 and 8. Peptide sequences were used for the subsequent analyses. To annotate the translated transcriptome, we used the custom pipeline Annocript[34]. We used the Swiss-Prot (SP) and UniRef90[50] (version: August 2013) for the blastp against proteins with the following parameters: word_size = 4; e-value = 10−5; num_descriptions = 5; num_alignments = 5; threshold = 18. For each sequence the best hit, if any, was chosen. Rpsblast parameters, to identify domains composition of putative proteins in the Conserved Domains Database, were: e-value = 10−5; num_descriptions = 20; num_alignments = 20. The software returned GO functional classification[51], the Enzyme Commission IDs[52] and Pathways[53] descriptions associated to the resulting best matches. R scripts were used to perform further analyses and graphs (http://www.r-project.org/). Venn diagram was done using the freely available software interactivenn (http://www.interactivenn.net/). We used the t-test to compare the FPKM for groups of peptides associated to the same GO/pathway and the Fisher exact test to compare the corresponding proportions of transcripts.

1.5.8 Bioinformatic identification and selection of pathways of interest

Pathways annotated as ‘prostaglandin biosynthesis’ and ‘secologanin biosynthesis’ were found among the second level pathways list generated within the Annocript pipeline

annotation of the proteome from the *T. rotula* CCMP 1647 RNA-seq. Representations of Monoterpenoid and prostaglandin biosynthetic pathways are reported in Supplementary Figures 1.3 and 1.5, respectively. Transcripts associated to the above pathways were extrapolated from the total proteome annotation table (Supplementary Data 13). Transcript and enzymes involved in polyketide biosynthesis were found by searching in the transcriptome and proteome blast description table, using the term 'polyketide' as search term. A representation of the polyketide synthase activity is reported in Supplementary Figure 1.6.

1.5.9 Comparative analysis

Fasta files and contig reads of *T. rotula* strains CCMP3096 and GSO102's RNA-seq were retrieved from the public databases CAMERA (http://camera.crbs.ucsd.edu/mmetsp/list.php, access date: September to November 2018) and iMicrobes interactive query tool for microbial data (http://data.imicrobe.us/sample/view/1867, access date: September to November 2018). Corresponding proteomes were annotated with Annocript and the output files for pathway and global annotation were used to search for the selected pathways. For comparisons among the proteomes, we used FPKM. Paired aligned reads were used to calculate the Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) of each sequence for each transcriptome as a measure of expression levels. The FPKM were calculated as follows: FPKM = [mapped reads pairs]/([length of transcript]/1000)/([total reads pairs]/10^6).

1.5.10 Reverse transcription

1 μg of total RNA used for RNA-seq was retro-transcribed in the T100 Thermal cycler (Bio-Rad Laboratories, Hercules, California, USA) following the manufacturer's instructions of the sensiFAST™ cDNA synthesis kit (Cat. No. BIO-65054, Bioline Reagents Ltd., London, UK).

1.5.11 Primer design and real time quantitative PCR

Candidate reference genes and genes of interest were selected considering the annotation of the peptides reported in the annotated transcriptome of *T. rotula* FE80 (CCMP1647)

(Supplementary Data 13). Primers for amplification of selected transcripts, designed using Primer3 program V. 0.4[54,55], are listed in Table 1.9, together with their relative sequences and characteristics.

Table 1.9: List of primers used for amplification of selected transcripts. Tm= temperature of melting; Amp. Length= amplicons length.

Gene	Primer Sequence		Length(bp)	Tm (°C)	GC content(%)	Amp. Length (bp)
	Forward	Reverse				
COX-1 or PgG/Hs2	TCATCAAGGGAGGAGAATGG	CTTCCACCAAGAGCGAAGAC	20	58.4/60.5	50/55	170
$PgEs_2$	TTCCAAACAGGGCAAGTTAC	TTGCACGAGACAGATTGGAG	20	56.4/58.4	45/50	183
PgFs	TCTCCCCTATCGAGGGTTCT	AGCTCCACTCTGCTATCC	20/18	60.5/56.3	55/56	114
RPS	GCGCCTTGACAATACCGTCGATG	GACCTCATTGGACAGGTTCTTCC	23	66.6/64.6	57/52	148
H_4	GACAACATCCAGGGCATCAC	CTCGGTGTAGGTAACGGAGTC	20/21	60.5/63.2	55/57	153
TUBα	GTATCGCCAGCTCTACCATCC	GTGGCGTGGAAGATGAGGAATC	21/22	63.2/64.2	57/55	176
TUBβ	GAAACGGCCGATATGTTCCC	CAGTGTAATGTCCCTTGGCCC	20/21	60.5/63.2	55/57	159
TBP	CCTTCTTCAACCCCTCCACCAAC	GTTCGCTCATCCCACGTTTTCG	23/22	66.6/64.2	57/55	161
GAPDH	CTGCGAAGGTCCATCCACCGTC	CCACAAGGAGTACGAGAACAGC	22	67.9/64.2	64/55	172
ACTIN	TCGGCCCTTGAGAAGAGTTTCG	GATGGTCTGGAAAGTGGAGTCC	22	64.2/64.2	55/55	147
COPα	CCCATGCACCTGACTTGATGC	GTGAACTGACAGCACCCAAG	21/20	63.2/60.5	57/55	156
UBI	GGGAGCGATGCGATAATACG	GAGGGAACAACGGAGGAAGAAC	20/22	60.5/64.2	55/55	156
EF1α	GGGAGCGATGCGATAATACG	GTTGTGCATCGAGGGGAATC	21/20	60.5/60.5	55/55	151
EctA	GATCTACACAGCGATGCGAC	GCAACCTGATCCAATGCGAT	20/20	60.5/58.4	55/50	160
EctB	CCGAAGCGACCCTGAAATC	GTCTCTGATGCGCAATGGTC	19/20	59.5/55	58/60.5	162
Secologanin	ACAGGGGTCGTATTGTGCTG	CGACTCCTCTTGCACACTCA	20/20	55/55	60.5/60.5	130
PKS_Synt	AGATGGAGGGCTTTTGGTGG	CAGAGGATATGCGTCGCTGT	20/20	55/55	60.5/60.5	100
PKS_ER	TAGTATCCAACCCCATGCCG	CGGTTTCTTGTTCGCTTCCA	20/20	55/50	60.5/58.4	153
PKS_HxL	CACAATATCGCGTGGAGGC	GTCCAAGTACGACGCGGAA	19/19	58/58	59.5/59.5	108
PKS_Cyc	CATACTTTCGGGTTTGGGCC	ACCGTATCGCAATGCCATTC	20/20	55/50	60.5/58.4	166

Each sequence was initially tested by standard PCR. Reactions were carried out in 25 µL volume with 2.5 µL of 10× PCR reaction buffer (Roche, Basel, Switzerland), 2.5 µL of 10 × 2 mM dNTP, 0.3 µL of 5 U/µL Taq (Roche, Basel, Switzerland), 1 µL 10 µM of each oligo, 1 µL of cDNA templates and nuclease-free water up to 25 µL. The PCR program consisted of a denaturation step at 95 °C for 3 min, 40 cycles at 95 °C for 30 s, 53 °C 30 s, 72 °C for 30 s, and a final extension step at 72 °C for 7 min. Amplified PCR products were analysed by agarose gel electrophoresis.

The resulting bands were excised from the gel and extracted according to the GenElute Gel Extraction Kit protocol (Sigma-Aldrich, St. Louis, Missouri, USA). Sequences were obtained by BigDye Terminator Cycle Sequencing Technology (Applied Biosystems, Waltham, Massachusetts, USA) and purified using the Agencourt CleanSEQ Dye terminator removal Kit (Agencourt Bioscience Corporation, Beverly, Massachusetts, USA) in automated robotic station Biomek FX (Beckman Coulter, Pasadena, California, USA). Products were analysed on the Automated Capillary Electrophoresis Sequencer 3730 DNA Analyser (Applied Biosystems, Waltham, Massachusetts, USA). Alignments were performed with BioEdit software V. 7.0.5.3 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) .

Reverse transcription-quantitative PCR (rt-qPCR) experiments were performed in MicroAmp Optical 384-Well reaction plate (Applied Biosystems, Waltham, Massachusetts, USA) with Optical Adhesive Covers (Applied Biosystems, Waltham, Massachusetts, USA) in a Viia7 Real Time PCR System (Applied Biosystem, Waltham, Massachusetts, USA). Five serial dilutions of mixed cDNAs were used to determine primer reaction efficiency using the formula: $E = 10^{-1/slope}$ (SI1-Supplementary Figure 1.7). The PCR volume for each sample was 10 µL, with 5 µl of SensiFAST TM SYBR® Lo-ROX Kit (BIO_94020, Bioline Reagents Ltd., London, UK), 1 µL of cDNA template (1 to 5 dilution each template) and 4 µL of 0.7 µM oligo mix (forward and reverse). Program reaction used was: 95 °C for 20 s, 40 cycles of 95 °C for 1 s and 60 °C for 20 s. The program was set to reveal the melting curve of each amplicon from 60 °C to 95 °C, and read every 0.5 °C. Single peaks for all genes confirmed gene-specific amplification and the absence of primer-dimers. All RT-qPCR reactions were carried out in triplicate to capture intra-assay variability. Each assay included three no-template negative controls for each primer pair. The normalized expression levels of each gene of interest relative to the most stable reference genes (see Supplementary Information 2 and 3 for reference genes analysis), actin and TBP, were calculated by using the Q-Gene tool[56]. Only TBP normalized values were reported in the main text and figure results. Relative expression

ratios above two fold were considered significant. Statistical analysis was performed using the unpaired t-test with Welch's correction for comparison between conditions, using GraphPad Prim statistic software V. 6.01 (GraphPad Software Inc. San Diego, California, USA). Three different algorithms were utilized to identify the best reference genes in our experimental design: BestKeeper[57]; NormFinder[58] and geNorm[59].

1.5.12 Statistics

Using GraphPadPrism6 software t-test was performed between CTR and SiDepl sample for qPCR amplification data to determine significant differences between treatments.
R software was used for bioinformatics analysis of the transcriptome performing the following tests: Fisher test to identify unique GO terms; t-test between the CTR and SiDepl condition considering the FPKM expression value of each pathway; heat maps in different conditions for the selected pathways; boxplot analysis of the expression in FPKM of each pathway in each level of annotation for each replicate and group.

1.5.13 Data Availability

The sequencing data described in this article, i.e. the transcript and peptide sequences of *T. rotula*, strain CCMP1647, lab-name FE80, are in Supplementary Files 7 and 8 respectively. The sequencing data relative to the *T. rotula* strains CCMP3096 and GSO102 used for the comparative analysis are available at https://www.imicrobe.us/#/projects/104.

1.6 References

1. Kroth, P. G. *et al*. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* **3**, e1426 (2008).

2. Nisbet, R. E. R., Kilian, O. & McFadden, G. I. Diatom Genomics: Genetic Acquisitions and Mergers. *Current Biology* **14**, R1048–R1050 (2004).

3. Armbrust, E. V. The life of diatoms in the world's oceans. *Nature* **459**, 185–192 (2009).

4. Lauritano, C., Carotenuto, Y., Miralto, A., Procaccini, G. & Ianora, A. Copepod Population-Specific Response to a Toxic Diatom Diet. *PLOS One* **7**, e47262 (2012).

5. Falciatore, A., d'Alcalà, M. R., Croot, P. & Bowler, C. Perception of environmental signals by a marine diatom. *Science* **288**, 2363–2366 (2000).

6. Romano, G. *et al*. Marine microorganisms as a promising and sustainable source of bioactive molecules. *Mar. Environ. Res.* **128**, 58–69 (2017).

7. Muhseen, Z. T., Xiong, Q., Chen, Z. & Ge, F. Proteomics studies on stress responses in diatoms. *Proteomics* **15**, 3943–3953 (2015).

8. Heydarizadeh, P. *et al*. Carbon Orientation in the Diatom *Phaeodactylum tricornutum*: The Effects of Carbon Limitation and Photon Flux Density. *Front Plant Sci* **10** (2019).

9. Heydarizadeh, P. *et al*. Response of CO_2-starved diatom *Phaeodactylum tricornutum* to light intensity transition. *Philos. Trans. R. Soc. Lond., B, Biol. Sci*. **372** (2017).

10. Shrestha, R. P. & Hildebrand, M. Evidence for a regulatory role of diatom silicon transporters in cellular silicon responses. *Eukaryotic Cell* **14**, 29–40 (2015).

11. Brzezinski, M., Olson, R. & Chisholm, S. Silicon availability and cell-cycle progression in marine diatoms. *Marine Ecology Progress Series* **67**, 83–96 (1990).

12. Keeling, P. J. *et al*. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **12**, e1001889 (2014).

13. Pan, Y., Bates, S. S. & Cembella, A. D. Environmental stress and domoic acid production by *Pseudo-nitzschia*: a physiological perspective. *Nat. Toxins* **6**, 127–135 (1998).

14. Armbrust, G. *et al*. The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution and Metabolism. *Science* **1**;306(5693):79-86 (2004).

15. Shrestha, R. P. *et al*. Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* **13**, 499 (2012).

16. Fenton, T. R. & Gout, I. T. Functions and regulation of the 70 kDa ribosomal S6 kinases. *Int. J. Biochem. Cell Biol.* **43**, 47–59 (2011).

17. Gerardy-Schahn, R., Oelmann, S. & Bakker, H. Nucleotide sugar transporters: biological and functional aspects. *Biochimie* **83**, 775–782 (2001).

18. Bakker, H. *et al*. A CMP-sialic acid transporter cloned from *Arabidopsis thaliana*. *Carbohydr. Res.* **343**, 2148–2152 (2008).

19. Takashima, S. *et al*. Analysis of CMP-sialic acid transporter-like proteins in plants. *Phytochemistry* **70**, 1973–1981 (2009).

20. Mamedov, T. & Yusibov, V. Green algae *Chlamydomonas reinhardtii* possess endogenous sialylated N-glycans. *FEBS Open Bio* **1**, 15–22 (2011).

21. Das, A. K., Uhler, M. D. & Hajra, A. K. Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme A reductase cDNAs. *J. Biol. Chem.* **275**, 24333–24340 (2000).

22. Yuan, A., Chen, J. J. W., Yao, P.-L. & Yang, P.-C. The role of interleukin-8 in cancer cells and microenvironment interaction. *Front. Biosci.* **10**, 853–865 (2005).

23. Ferguson, B. S. *et al*. Lactate metabolism: historical context, prior misinterpretations, and current understanding. *Eur. J. Appl. Physiol.* **118**, 691–728 (2018).

24. Tungmunnithum, D., Thongboonyou, A., Pholboon, A. & Yangsabai, A. Flavonoids and Other Phenolic Compounds from Medicinal Plants for Pharmaceutical and Medical Aspects: An Overview. *Medicines (Basel)* **5** (2018).

25. Di Dato, V. *et al*. Animal-like prostaglandins in marine microalgae. *ISME J* **11**, 1722–1726 (2017).

26. O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat Prod Rep* **23**, 532–547 (2006).

27. Venditto, V. J. & Simanek, E. E. Cancer Therapies Utilizing the Camptothecins: A Review of *in Vivo* Literature. *Mol Pharm* **7**, 307–349 (2010).

28. Irmler, S. *et al*. Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804 (2000).

29. English, B. J. & Williams, R. M. A divergent strategy for the synthesis of secologanin derived natural products. *J. Org. Chem.* **75**, 7869–7876 (2010).

30. Kohli, G. S., John, U., Van Dolah, F. M. & Murray, S. A. Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. *ISME J* **10**, 1877–1890 (2016).

31. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. USA* **111**, 9259–9264 (2014).

32. Stewart, C., Vickery, C. R., Burkart, M. D. & Noel, J. P. Confluence of structural and chemical biology: plant polyketide synthases as biocatalysts for a bio-based future. *Curr. Opin. Plant Biol.* **16**, 365–372 (2013).

33. Wang, H., Sivonen, K. & Fewer, D. P. Genomic insights into the distribution, genetic diversity and evolution of polyketide synthases and nonribosomal peptide synthetases. *Curr. Opin. Genet. Dev.* **35**, 79–85 (2015).

34. Musacchia, F., Basu, S., Petrosino, G., Salvemini, M. & Sanges, R. Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. *Bioinformatics* **31**, 2199–2201 (2015).

35. Funk, C. D. Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* **294**, 1871–1875 (2001).

36. Rowley, A. F., Vogan, C. L., Taylor, G. W. & Clare, A. S. Prostaglandins in non-insectan invertebrates: recent insights and unsolved problems. *J. Exp. Biol.* **208**, 3–14 (2005).

37. Groenewald, E. G. & van der Westhuizen, A. J. Prostaglandins and related substances in plants. *The Botanical Review* **63**, 199–220 (1997).

38. Ritter, A. *et al*. Copper stress induces biosynthesis of octadecanoid and eicosanoid oxygenated derivatives in the brown algal kelp *Laminaria digitata*. *New Phytol.* **180**, 809–821 (2008).

39. Imbs, A. B. *et al*. Distribution of polyunsaturated fatty acids in red algae of the genus *Gracilaria*, a promising source of prostaglandins. *Russ J Mar Biol* **38**, 339–345 (2012).

40. Whittaker, K. A., Rignanese, D. R., Olson, R. J. & Rynearson, T. A. Molecular subdivision of the marine diatom *Thalassiosira rotula* in relation to geographic distribution, genome size, and physiology. *BMC Evolutionary Biology* **12**, 209 (2012).

41. Guillard, R. R. L. Culture of Phytoplankton for Feeding Marine Invertebrates. In Culture of Marine Invertebrate Animals: Proceedings - 1st Conference on Culture of Marine Invertebrate Animals *Greenport* (eds Smith, W. L. & Chanley, M. H.) 29–60, (Springer US, 1975).

42. Sabatino, V. *et al*. Establishment of Genetic Transformation in the Sexually Reproducing Diatoms *Pseudo-nitzschia multistriata* and *Pseudo-nitzschia arenysensis* and Inheritance of the Transgene. *Mar. Biotechnol*. **17**, (4):452-62 (2015).

43. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
44. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
45. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
46. Haas, B. J. *et al*. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494–1512 (2013).
47. Conesa, A. *et al*. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
48. Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods* **10**, 71–73 (2013).
49. Leng, N. *et al*. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035–1043 (2013).
50. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
51. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
52. Bairoch, A. The ENZYME database in 2000. *Nucleic Acids Res.* **28**, 304–305 (2000).
53. Morgat, A. *et al*. UniPathway: a resource for the exploration and annotation of metabolic pathways. *Nucleic Acids Res.* **40**, D761–769 (2012).
54. Untergasser, A. *et al*. Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
55. Koressaar, T. & Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289–1291 (2007).
56. Simon, P. Q-Gene: processing quantitative real-time RT-PCR data. *Bioinformatics* **19**, 1439–1440 (2003).
57. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
58. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify

genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).

59. Vandesompele, J. *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 0034 (2002).

Chapter 2:

Variation in prostaglandin metabolism during *T. rotula* growth

Adapted from the published article:

Di Dato, V., Barbarinaldi, R., Amato, A., Di Costanzo, F., Fontanarosa, C., Perna, A., Amoresano, A., Esposito, F., Cutignano, A., Ianora, A., Romano, G. Variation in prostaglandin metabolism during growth of the diatom *Thalassiosira rotula*. *Scientific Reports* **10**, 5374 (2020).

2.1 Abstract

Prostaglandins (PGs) are hormone-like mediators in many physiological and pathological processes that are present in all vertebrates, in some terrestrial and aquatic invertebrates, and have also been identified in some macroalgae. They have recently been reported also in marine microalgae but their role as chemical mediators is largely unknown. Here we studied the expression pattern of the PG biosynthetic pathway during different growth phases of the centric diatom *Thalassiosira rotula* and assessed the release of PGs in the surrounding environment for the first time. We show that enzymes responsible for PGs formation such as cyclooxygenase, prostaglandin E synthase 2-like and prostaglandin F synthase are mainly expressed at the end of the exponential phase and that PGs are released especially during the stationary and senescent phases, suggesting a possible signalling function for these compounds. Phylogenetic analysis of the limiting enzyme, COX, indicate the presence in diatoms of more than one enzyme related to the oxidative metabolism of fatty acids belonging to the peroxidase-cyclooxygenase superfamily. These findings suggest a more complex evolution and diversity of metabolic pathways leading to the synthesis of lipid mediators in diatoms.

2.2 Introduction

Arachidonic acid (ARA), eicosapentaenoic acid (EPA), eicosatrienoic acid (ETrA) and docosahexaenoic acid (DHA) are polyunsaturated fatty acids that are physiologically important for animals at all taxonomic levels, including humans. EPA and DHA contribute to the healthy functioning of the cardiovascular system[1] and are precursors of important classes of fatty acid-derivatives playing multiple signalling roles in inflammation and immune responses, platelet aggregation and tumour growth[2,3]. Among these, the inflammation process is one of the most important mechanisms adopted by organisms in response to various external stimuli[3]. Inflammation involves a complex interplay of signalling molecules whose final goal is to restore the healthy status of a cell or tissue. Consequences of sustained inflammation are indeed the development of serious diseases such as cancer and autoimmune disorders[4].

Included in the eicosanoids are prostaglandins (PGs) synthesized principally from ARA in animals, but also from EPA and ETrA, through the enzymatic route initiated by cyclooxygenase (COXs) enzymes[5]. PGs are molecules with a hormone-like behaviour playing a prominent role in many physiological processes that have been principally studied in animals[3]. The expression of the COXs enzymes is mandatory for their

synthesis. COXs exist in two isoforms that differ for their subcellular localization and for their expression timing.

COX-1 is located in the endoplasmic reticulum and is constitutively expressed at constant levels in many tissues unless external cues, such as tumour promoting factor, cytokine and growth factor, induce an increase in its expression level. COX-2 is the inducible form, which is not detectable unless a trigger similar to those that stimulate COX-1 expression occurs. COX-2 is located in the nuclear envelope and appears to be a target for cancer therapy[6].

PGs synthesis, in mammals, is initiated by phospholipases (PLAs), a family of enzymes that hydrolyse membrane phospholipids liberating the precursors, ARA, EPA, and ETrA. These are rapidly converted, through cyclization and inclusion of molecular oxygen, into the unstable metabolite PGG_2 by the action of COXs enzymes that subsequently reduce it into PGH_2. The PGH_2 is then transformed into the ultimate prostaglandin E_2, D_2, $F_{2\alpha}$, prostacyclin or tromboxanes by the successive action of PGE, PGD, PGF, prostacyclin and tromboxanes A synthases[7], respectively. Depending on which precursor is used, three series of PGs are produced, each having its own receptor series whose binding determine the type of message transduced[8].

Plants synthesize molecules similar in structure to PGs, such as jasmonic acid, which regulates plant reproduction and fruit ripening processes[9], responses to biotic and abiotic stresses, responses to external damage due to mechanical injury, herbivore and insect attack[10]. However, the presence of PGs has also been reported in some plant species, such as onions and poplar trees[11].

In the marine environment, PGs have been isolated from the animal and plant kingdoms. In vertebrates such as sharks, carps and salmons they are involved in osmoregulation, regulation of branchial ion fluxes, ovulation and spawning[7], while in invertebrates such as crustaceans, corals, and molluscs their role is still unknown. Among marine photoautotrophs, macroalgae and microalgae (e.g. *Euglena gracilis*) have shown PGs synthesis under stress conditions, however it is unclear what role they play in the mediation of stress response in these organisms[7]. Bacterial representatives have also been discovered with some cyanobacteria species synthesizing PGs[7].

Despite the fact that only mammals have been deeply studied to understand PGs metabolism and functions, the discovery of PGs in other organisms should not be surprising. Indeed, evolutionary studies of the peroxidase-cyclooxygenase superfamily, to which COX enzymes belong, demonstrate its wide distribution in all the domains of life[12], with seven main families that are well conserved animal cyclooxygenases, with

bacterial sequences that are still putative while many of the animal non-mammalian phyla, such as Cnidaria, Mollusca, Arthropoda and Chordata, have been cloned and studied even if the role of PGs also in these organisms is still unclear[12–14]. A peculiar branch of this family that diverged early from cyclooxygenases is the plant alpha dioxygenases (α-DOX) that partly lost the peroxidase activity maintaining instead the oxygenase activity responsible for the transformation of fatty acids to hydroperoxides.

In a previous study, we explored a diatom species for PGs metabolism and demonstrated the synthesis of all three series of PGs[15]. Diatoms are a class of Stramenopiles characterized by a peculiar cell organization[16,17], cell division mode[18] and complex genomes[19]. They originated from a secondary endosymbiotic event[20] which likely led to organisms displaying intricate metabolic pathways retrieved from the different entities that coexist in diatom genomes. Marine diatoms play a crucial role in the world carbon cycle being responsible for a non-negligible part of CO_2 fixation on a yearly basis[21]. We demonstrated that two clones of *Skeletonema marinoi* differed for the expression level of the PG enzymatic pathway and for the amount of molecules produced[15]. In addition, *in-silico* analysis demonstrated that COX was present also in other diatom species and that diatom and animal COXs were remarkably similar[15]. In *Phaeodactylum tricornutum*, a model diatom species, compounds similar to PGs, called isoprostanoids, have been reported, despite the absence of COX sequence annotated in the genome. This finding suggests the existence of an alternative, non-enzymatic, origin of this class of compounds[22].

We found that the PGs pathway[23] was up-regulated in the diatom *Thalassiosira rotula* under silica deprivation stress, in comparison with control conditions (as described in Chapter 1). *T. rotula* is a major diatom species blooming in many areas of the world's oceans. This finding has fostered the current more thorough investigation on the expression and activity of the PG pathway in *T. rotula* during the different phases of its growth, measuring for the first time the release of PGs outside the cells into the surrounding growth medium. In addition, we conducted an evolutionary and *in silico* study of the *T. rotula* COX (TrotCOX) demonstrating that it actually belongs to the peroxidase-cyclooxygenase superfamily.

2.3 Results

2.3.1 *In-silico* structural reconstruction of TrotCOX

Since COX is the most relevant enzyme in the pathway, essential for the initiation of PGs synthesis, we used Phyre2 (Protein Homology/analogY Recognition Engine V 2.0)[24] to predict the possible structure of TrotCOX protein and to confirm that it belongs to peroxidase-cyclooxygenase superfamily. The multiple alignments developed by the Phyre2 program retrieved as best hits 13 structures of proteins belonging to the peroxidase-cyclooxygenase family (Table 2.1, Supplementary Figure 2.1). All of them shared 100 % confidence and 94–99 % coverage (Table 2.1). The first structure in the list shared a 19 % identity and corresponded to a *Bos taurus* lactoperoxidase (BtLPO) already reported as representative of the peroxidase- cyclooxygenase superfamily[12] (Table 2.1, Supplementary Figure 2.2). Interestingly the second to the fifth structures corresponded to an *Ovis aries* prostaglandin G/H synthase (OaCOX1) all sharing a 28 % identity, confirming the cyclooxygenase structure of the TrotCOX protein (Table 2.1, Supplementary Figure 2.3). Structures from line 6 to 13, sharing lower identity (26 % to 19 %, Table 2.1), were however representative of cyclooxygenases, myeloperoxidase or oxidoreductase protein structures from *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana* and *Oryza sativa*, all of which are representative of the peroxidase-cyclooxygenase superfamily.

Table 2.1. Results of multi alignments of TrotCOX protein sequences with available protein structures. Refer to Supplementary Figure 2.1 for a complete list.

Template	Template id/DOI/web link	Alignment Coverage	Confidence	% of identity
Buffalo lactoperoxidase	c2gjmA 10.2210/pdb2GJM/pdb	97	100	19
Sheep prostaglandin g/h synthase 1	c2oyup 10.2210/pdb2OYU/pdb	99	100	28
Sheep prostaglandin g/h synthase 1	c1pggB 10.2210/pdb1PGG/pdb	99	100	28
Sheep prostaglandin g/h synthase 1	c1ht8B 10.2210/pdb1HT8/pdb	99	100	28
Sheep prostaglandin g/h synthase 2	d1q4ga1 http://scop.berkeley.edu/search/?oldURL=1&key=%22d1q4ga1%22&key=d1q4ga1	99	100	28

Dictyostelium discoideum Peroxidase	c6ercA 10.2210/pdb6ERC/pdb	97	100	26
human promyeloperoxidase	c5mfaA 10.2210/pdb5MFA/pdb	97	100	20
Mouse Cycloxygenase 2	c1ddxA 10.2210/pdb1DDX/pdb	99	100	27
Mouse Cycloxygenase 2	c3pghD 10.2210/pdb3PGH/pdb	99	100	27
Mouse Cycloxygenase 2	d1cvua1 http://scop.berkeley.edu/search/?oldURL=1&key=%22d1cvua1%22&key=d1cvua1	99	100	27
Human Myeloperoxidase isoform C	c1d2Vd 10.2210/pdb1D2V/pdb	94	100	20
Arabidopsis fatty acid alphadioxygenase (*Arabidopsis thaliana*)	c4hhsA 10.2210/pdb4HHS/pdb	98	100	19
Oryza sativa fatty acid alphadioxygenase	c4kvjA 10.2210/pdb4KVJ/pdb	99	100	19

Figures 2.1 b and c show the results of the comparison among TrotCOX (Figure 2.1 a) predicted structure and known models of the peroxidase-cyclooxygenase family, alias the BtLPO (3BXI.pdb, 10.2210/pdb3bxi/pdb) 12,25 and OaCOX1 structures. The main alpha helices that characterize the cyclooxygenases and accommodate the heme prosthetic group are present and almost perfectly overlapped with the references BtLPO and OaCOX1.

The *in silico* predicted model constructed on the BtLPO showed conserved domain sites in the diatom sequences (Figures 2.1 d and 2.2): catalytic arginine was found (R275) in the distal heme side, embedded in the characteristic motif XRXXEX, while glutamate that is usually conserved, was not in TrotCOX but was mutated to leucine (L278). The latter seems to be a characteristic of diatoms as none of the diatom sequences (Supplementary Figure 2.4) here analysed showed a glutamate in that position. The proximal heme sides are relatively conserved in TrotCOX with the highly conserved arginine (R371), histidine (H373), isoleucine (I376) and glutamic acid (E454) unvaried in TrotCOX and in most of the other sequences. Surprisingly, the conserved asparagine and arginine that are normally bonded via a hydrogen bond were substituted with a threonine (T453) and a cysteine (C456), respectively.

The calcium binding site in TrotCOX shows the relatively conserved Isoleucine (I184) and LYG motif (L198, Y199, G200) (Figures 2.1 d and 2.2). The sequence representation by Logo (Figure 2.2) confirmed this residue conservation also among all the diatom sequences considered for the phylogenetic analysis.

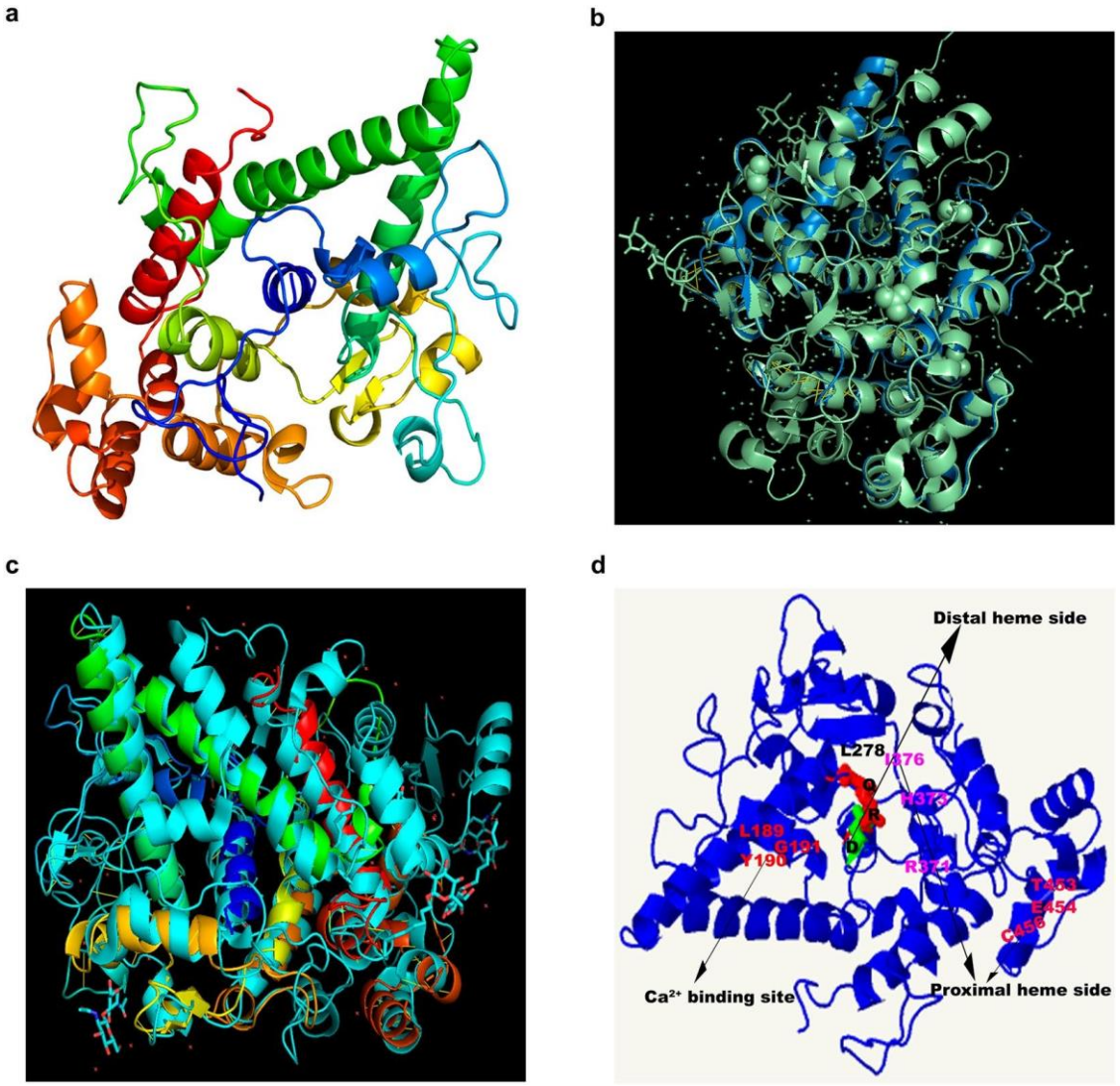


Figure 2.1. *In silico* reconstruction of the TrotCOX protein structure compared with the known reference proteins for the peroxidase-cyclooxygenase superfamily. (**a**) TrotCOX structure based on the 2.75 Å resolved x-ray diffraction of BtLPO. (**b**) Comparison of the *in silico* structure of the TrotCOX protein and the 2.75 Å resolved x-ray diffraction of BtLPO showing the perfect overlap of the alpha-helices embedding the heme sides. In blue the TrotCOX, in light green the BtLPO. (**c**) Comparison of the *in silico* structure of the TrotCOX protein and the 2.75 Å resolved x-ray diffraction of OaCOX1 showing the perfect overlap of the alpha-helices embedding the heme sides. Rainbow colored helices correspond to TrotCOX, the light-blue colored helices correspond to OaCOX1. (**d**) Illustration of the conserved catalytic sites described in the text. Q, R, D and L278 black letters refer to the conserved amino acids of the two distal heme sides with L278 being the mutated original E. Pink and light red letters indicate the conserved amino acids in the two proximal heme sides. The red letters indicate the amino acids forming the calcium binding site.

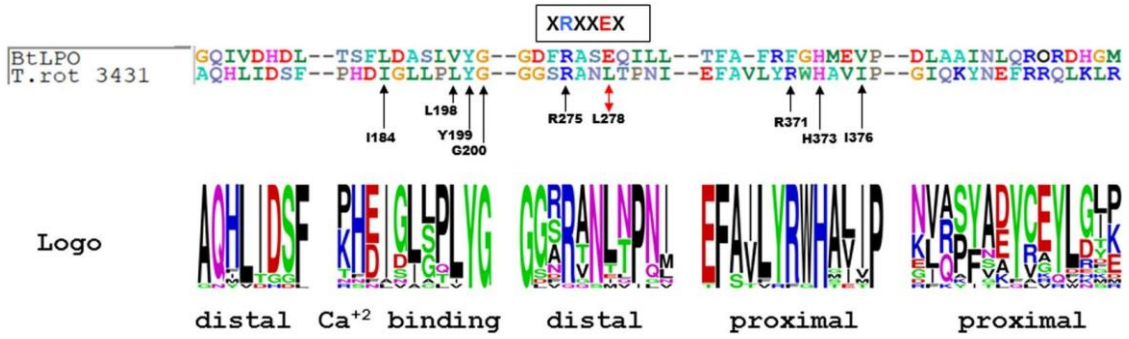


Figure 2.2. TrotCOX vs BtLPO amino acid sequence alignment at the distal and proximal heme sides and at the calcium binding sites[32]. The sequence Logo of the same portion of the alignment for the diatom specific clade is shown below. The sequence Logo is a resumed representation of an alignment among two or more sequences. The height of each letter is proportional to its frequency and the most common one is placed on the top[55]. Important amino acidic residues reported in the text are indicated on the *T. rotula* sequence. BtLPO = *Bos taurus* LPO; T. rot = *Thalassiosira rotula*.

Using a previous alignment[15], populated with other sequences retrieved from NCBI via a pBLAST, we performed a refined phylogenetic analysis comparing the COX sequence from *T. rotula* with that of other organisms (Supplementary Table 2.1). Most of the diatom sequences clustered in a large well-supported clade (Maximum Likelihood (ML) bootstrap value 100) closely related to the animal COX clade (family 4, cyclooxygenases), although the phylogenetic relationship was not well resolved (ML 35) (Figure 2.3). This large diatom clade was divided into a main well-supported clade including the vast majority of diatom sequences, among which the TrotCOX sequence (Figure 2.3), and in an end-clade (ML 94) grouping both prokaryotic (*Nostoc* sp., a cyanobacterium, the actinomycetales *Herbidospora mongoliensis* and *Rhodococcus gordoniae*) and diatom (*Grammatophora oceanica*, *Staurosira* sp., and *Chaetoceros* cf. *neogracile*) sequences. This diatom/bacterial clade can be hypothesized as being the 5a family within the peroxidase-cyclooxygenase super family[12].

Noteworthy, the pennate biraphid diatom sequence from *Nitzschia* sp. clusters basally to the animal COX clade (ML 56). In addition to that large clade, a second very well resolved (ML 100) super clade was composed of two clades. One of these forks includes only diatom sequences (*Fistulifera solaris* and *S. marinoi* 31509 and 1511), whereas the other bifurcates in two sister clades. One of these sister clades groups only two prokaryotic sequences, and the other bifurcates in two branches. The first branch groups plant α-DOX proteins (family 4 cyclooxygenase[12]) and the other groups diatoms (*Skeletonema gretae*, *Pseudo-nitzschia pungens* and *Fragilariopsis kerguelensis*) (Figure 2.3).

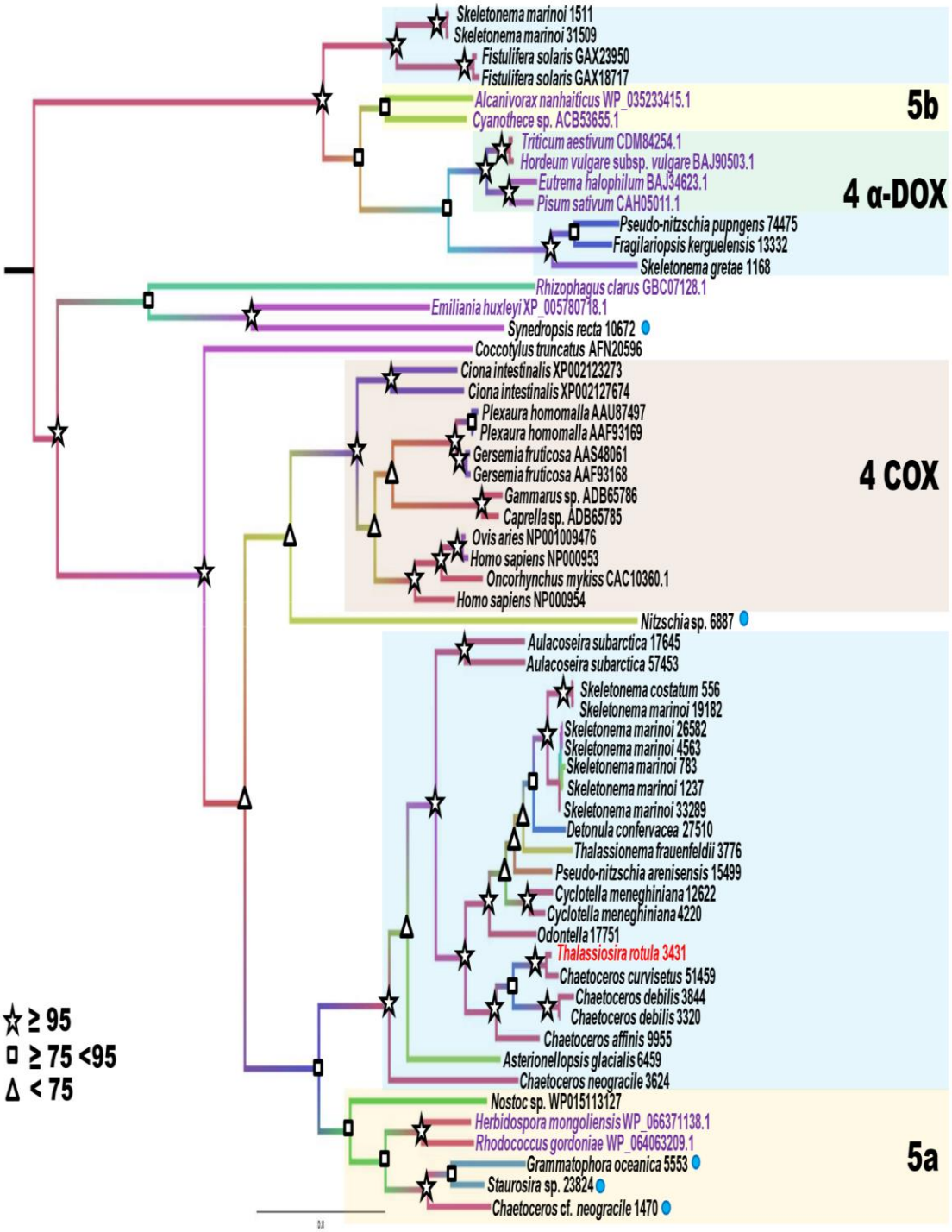


Figure 2.3. Phylogenetic analysis of the *Thalassiosira rotula* COX. Diatom sequences are scattered in the ML phylogenetic tree, either in diatom specific clades, or clustered together with sequences from other taxa. Legend: TrotCOX is indicated in red; cyan dot: diatom sequences not clustering the diatoms-specific clade; cyan-shaded clade: diatom specific clade; 5a: 5a-family in the diatom/bacterial clade; 4 COX: animal COX clade (family 4, cyclooxygenases); 4 α-DOX: plant α-DOX proteins (family 4, cyclooxygenases); 5b: 5b-family cyanobacteria. The nomenclature used for the family classification follows Zámocký *et al.*[12].

2.3.2 Gene expression analysis

Figure 2.4 illustrates differential expression analysis by qPCR of the three genes involved in the PG pathway during *T. rotula* growth (Figure 2.4 d). One-way ANOVA analysis shows, for each gene, a statistically significant difference ($0.0001 < p\text{-value} < 0.0002$) among the six time points considered along the 10-day *T. rotula* growth curve (Figure 2.4, Table 2.2).

Each gene showed a statistically significant expression peak around the late exponential/early stationary phase, i.e. day 5 for COX and at day 4 for mPTGES and PTGFS (Figure 2.4 a and Table 2.2). In addition, COX-day 6 expression persisted with significant difference versus COX-day 7 and COX-day 10 expressions. mPTGES-day 3 expression was statistically significantly higher than mPTGES-days 6, 7 and 10 expressions (Figure 2.4 b and Table 2.2) but had levels lower than mPTGES-day 4 expression. Similarly, PTGFS-day 3 expression was statistically significantly lower than day 4 but higher than PTGFS-days 7 and 10 expressions. Analogously, PTGFS-day 5 expression was still statistically significantly higher than PTGFS-day 7 and 10 expressions (Figure 2.4 c and Table 2.2).

The expression peaks were only transient because a quick decrease in the expression levels occurred soon afterwards, reaching very low levels at the end of the stationary (day 7) and senescent (day 10) phases. Moreover, the initial basal expression levels of the three enzymes were completely different. COX expression was much higher compared to mPTGES and PTGFS, and mPTGES was more expressed compared to PTGFS (Figure 2.4, Day3). However, from day 7 to 10, the expression of all three genes decreased, reaching similar levels. The reference genes[23] utilized were stable, indicating that the down regulation of the analysed genes was not due to senescence but rather to a specific pathway response.

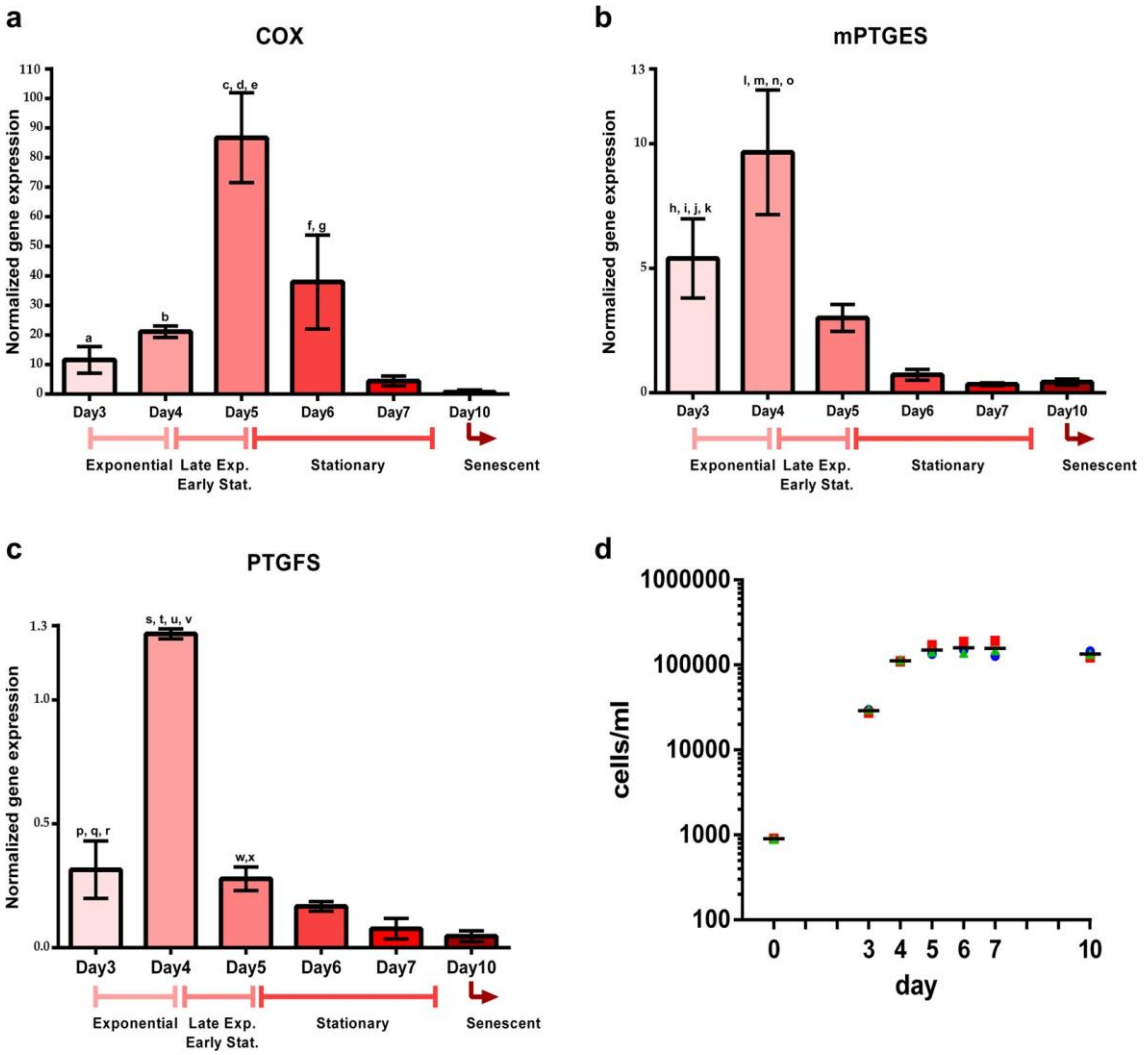


Figure 2.4. Gene expression analysis of the three genes identified in the *Thalassiosira rotula* prostaglandin pathway. Expression levels of COX (**a**), mPTGES (**b**) and PTGFS (**c**) genes during the growth of *T. rotula* from day 3 to day 10. Results are based on TBP housekeeping gene normalized value. Normalization based on Actin as housekeeping gene gave the same results. Panels a, b and c report the corresponding growth phase (Abbreviations: Exp.: Exponential phase; Stat.: Stationary phase). Y axis reports the Mean Normalized Expression (MNE) of three biological and technical replicates using TBP as reference gene. Letters on bars refer to significant daily comparisons. Refer to Table 2.2 for the levels of significance and the corresponding comparisons. (**d**) Growth curves of the 3 independent *T. rotula* cultures utilized to analyse gene expression. Horizontal lines over the symbols indicate the mean value of the three values for each time point. Y axis is in logarithmic scale.

Table 2.2. One-way ANOVA analysis over the time points along the 10 days of *T. rotula* growth curve showing day by day comparisons for each gene with the corresponding significance. Abbreviations: ns: not significant; *adjusted p-value < 0.05; **adjusted p-value < 0.001; ***0.0001 < adjusted p-value < 0.0005; ****p-value < 0.0001.

	Tukey's multiple comparisons test	**Adjusted p-value**	**Significant**	**Reference symbol in Fig. 2.4**	**Adjusted p-value**	**Significant**	**Reference symbol in Fig. 2.4**	**Adjusted p-value**	**Significant**	**Reference symbol in Fig. 2.4**
gene			COX			mPTGES			PTGFS	
ANOVA P value			0.0002			0.0002			<0.0001	
	Day3 vs Day4	0.8823	ns		0.0412	*	h	<0.0001	****	p
	Day3 vs Day5	0.0005	***	a	0.4826	ns		0.9745	ns	
	Day3 vs Day6	0.1282	ns		0.0247	*	i	0.1358	ns	
	Day3 vs Day7	0.9720	ns		0.0298	*	j	0.0079	**	q
	Day3 vs Day10	0.8710	ns		0.0328	*	k	0.0038	**	r
	Day4 vs Day5	0.0007	***	b	0.0056	**	l	<0.0001	****	s
	Day4 vs Day6	0.3705	ns		0.0003	***	m	<0.0001	****	t
	Day4 vs Day7	0.4802	ns		0.0005	***	n	<0.0001	****	u
	Day4 vs Day10	0.3069	ns		0.0005	***	o	<0.0001	****	v
	Day5 vs Day6	0.0053	**	c	0.5259	ns		0.3359	ns	
	Day5 vs Day7	0.0003	***	d	0.4658	ns		0.0207	*	w
	Day5 vs Day10	0.0002	***	e	0.4979	ns		0.0094	**	x
	Day6 vs Day7	0.0442	*	f	0.9996	ns		0.4457	ns	
	Day6 vs Day10	0.0263	*	g	0.9999	ns		0.2070	ns	
	Day7 vs Day10	0.9988	ns		>0.9999	ns		0.9740	ns	

2.3.3 LC-MS/MS analysis

In addition to gene expression levels, we also studied the pathway activity by identifying the PGs released outside the cells by *T. rotula* and measured their concentration in the culture medium (Figure 2.5 a–c). By LC-MS/MS in comparison with pure standards (Supplementary Figures 2.5–2.8), we detected three PGs in the culture medium, namely PGE_2, PGB_2 and 15-d-PGJ_2. In addition, in the same medium, we also identified PGA_2, PGD_1, PGD_3, PGE_1, PGE_2, PGE_3, PGEM, PGFM, 15-d-PGD_2 and 2,3-dinor-11b-PGF_2 that we classified as "putative" since they have been identified through comparison of experimental mass fragmentation data obtained in this study with those reported in the literature (Supplementary Figure 2.9, Supplementary Table 2.2)[26,27].

Overall, a high variability among replicates was evident as indicated by the large error bars (Figure 2.5 a–c). In addition, we observed an up-and-down trend of each measured PG concentration from days 3 to 5. Conversely, measurements from day 6, corresponding to the full stationary phase, to day 10, showed a linear trend toward increasing concentration values, except for the PGE_2, PGE_3, 15-d-PGJ_2, and PGFM (Figure 2.5 a–c). In particular, PGE_2 had a constant low level reaching zero on day 10 (Figure 2.5 a), while PGEM showed an exponential trend starting from day 4, i.e. late exponential phase of growth, reaching the highest value among all the other PGs identified (Figure 2.5 b).

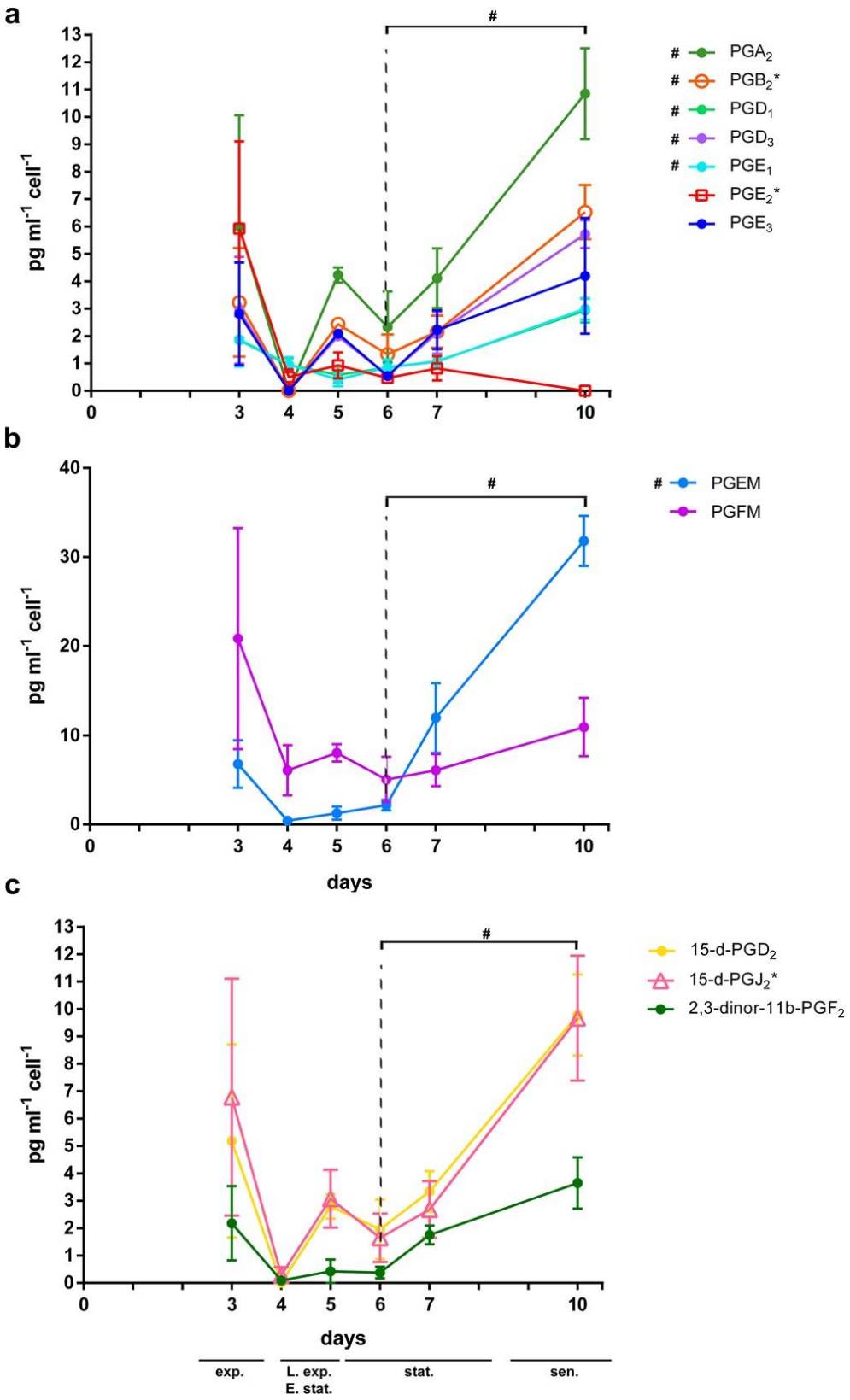


Figure 2.5. LC-MS/MS identification and quantification of prostaglandins in the culture medium during *Thalassiosira rotula* growth. Concentration trends of the prostaglandins in pg mL^{-1} $cell^{-1}$ during the 10 days were identified and measured by either comparison with pure standards (asterisks) or by comparison with fragmentation data reported in the literature. # indicates PGs for which Day 6 vs Day10 values are significantly different by one-way ANOVA analysis considering day 6 as reference time point. Dashed lines refer to day 6.

One-way ANOVA analysis (Table 2.3) shows a statistically significant difference over the considered time points for PGA_2, PGB_2, PGD_1, PGD_3, PGE_1, PGEM, 15-d-PGD_2 and 2,3-dinor-11b-PGF_2. Comparison of day 4 versus day 10 values were statistically significant for the majority of the compounds, due to concentrations dropping to zero on day 4 (Figure 2.5). PGEM concentrations, varied significantly from day 4 to 10, even if some differences were detected, i.e. Day5 vs Day6, Day5 vs Day7 and Day6 vs Day7 were not significant because of the large error bars (Figure 2.5 c, Table 2.3).

One-way ANOVA, calculated considering day 6 as reference point, shows the same significant differences over time points confirming that differences in concentrations at day 6 vs day 10 are statistically significant for all the PGs, except for PGE_2, PGE_3, 15-d-PGJ_2, and PGFM (Table 2.4).

Table 2.3. One-way ANOVA analysis over the time points considered during the 10 days of the *T. rotula* growth curve.

PG	Tukey's Multiple comparisons test	Adjusted p-value/Significance											
		PGA_2	PGB_2	PGD_1	PGD_3	PGE_1	PGE_2	PGE_3	PGEM	PGFM	15-d-PGJ_2	15-d-PGD_2	2,3-dinor-11b-PGF_2
ANOVA P value		**0.0306***	**0.001****	**0.042***	**0.008****	**0.028***	0.068ns	0.221ns	**<0.0001******	0.385ns	0.0701ns	**0.0207***	**0.0289***
	Day3 vs Day4	0.3042ns	0.255ns	0.759ns	0.241ns	0.754ns	0.110ns	0.5728ns	0.4080ns	0.4514ns	0.3189ns	0.3011ns	0.3651ns
	Day3 vs Day5	0.9845ns	0.991ns	0.463ns	0.969ns	0.317ns	0.158ns	0.9976ns	0.5487ns	0.5900ns	0.8117ns	0.9034ns	0.5375ns
	Day3 vs Day6	0.7557ns	0.743ns	0.683ns	0.439ns	0.670ns	0.108ns	0.7541ns	0.7086ns	0.3843ns	0.5486ns	0.7405ns	0.5104ns
	Day3 vs Day7	0.9794ns	0.968ns	0.856ns	0.982ns	0.832ns	0.144ns	0.9991ns	0.6146ns	0.4521ns	0.7442ns	0.9648ns	0.9979ns
	Day3 vs Day10	0.5089ns	0.238ns	0.624ns	0.289ns	0.588ns	0.071ns	0.9574ns	**<0.0001********	0.7949ns	0.9210ns	0.4168ns	0.6952ns
	Day4 vs Day5	0.6404ns	0.525ns	0.994ns	0.610ns	0.9581ns	0.999ns	0.8104ns	0.9998ns	0.9998ns	0.9309ns	0.8326ns	0.9993ns
	Day4 vs Day6	0.9504ns	0.922ns	0.999ns	0.997ns	0.9999ns	0.999ns	0.9994ns	0.9931ns	>0.9999ns	0.9970ns	0.9550ns	0.9997ns
	Day4 vs Day7	0.6655ns	0.637ns	0.999ns	0.557ns	0.9999ns	0.999ns	0.7685ns	**0.0349***	>0.9999ns	0.9624ns	0.7139ns	0.5886ns

	Day4 vs Day10	**0.0174***	**0.005****	0.103ns	**0.006****	0.0916ns	0.999ns	0.2003ns	**<0.0001****	0.9873ns	0.0732ns	**0.0128***	**0.0381***
	Day5 vs Day6	0.9786ns	0.963ns	0.998ns	0.851ns	0.9823ns	0.999ns	0.9345ns	0.9997ns	0.9986ns	0.9961ns	0.9990ns	>0.9999ns
	Day5 vs Day7	0.9999ns	0.999ns	0.974ns	0.999ns	0.9169ns	0.999ns	0.9999ns	0.0542ns	0.9998ns	>0.9999 ns	0.9999ns	0.7733ns
	Day5 vs Day10	0.219ns	0.098ns	**0.043***	0.090ns	**0.0233***	0.995ns	0.8025ns	**<0.0001****	0.9988ns	0.3050ns	0.0944ns	0.0663ns
	Day6 vs Day7	0.9839ns	0.989ns	0.999ns	0.808ns	0.9995ns	0.999ns	0.9091ns	0.0861ns	>0.9999ns	0.9992ns	0.9899ns	0.7478ns
	Day6 vs Day10	0.0731ns	**0.026***	0.082ns	**0.013***	0.0714ns	0.999ns	0.3164ns	**<0.0001****	0.9706ns	0.1516ns	>0.0523 ns	0.0612ns
	Day7 vs Day10	0.2056ns	0.071ns	0.142ns	0.105ns	0.118ns	0.997ns	0.8417ns	**0.0005*****	0.9874ns	0.2539ns	0.1367ns	0.4585ns

Table 2.4. One-way ANOVA analysis over the time points day 6 to day 10 of the 10 days of *T. rotula* growth curve, considering Day 6 as reference time point.

PG	**Tukey's Multiple comparisons test**	**Adjusted p-value/Significance**											
		PGA_2	PGB_2	PGD_1	PGD_3	PGE_1	PGE_2	PGE_3	PGEM	PGFM	15-d-PGJ_2	15-d-PGD_2	2,3-dinor-11b-PGF_2
ANOVA P value		**0.0306***	**0.001****	**0.042***	**0.008****	**0.028***	0.068ns	0.221ns	**<0.0001****	0.385ns	0.0701ns	**0.0207***	**0.0289***
	Day6 vs Day3	0.5521	0.537	0.472	0.257	0.458	**0.051**	0.550	0.498	0.218	0.345	0.534	0.313
	Day6 vs Day4	0.8574	0.8	0.999	0.987	0.99	>0.999	0.997	0.973	0.999	0.987	0.868	0.998
	Day6 vs Day5	0.9278	0.9	0.993	0.679	0.938	0.999	0.823	0.998	0.994	0.984	0.995	>0.999
	Day6 vs Day7	0.9430	0.1	0.996	0.618	0.997	0.999	0.774	**0.04**	0.999	0.996	0.962	0.543
	Day6 vs Day10	**0.0334**	**0.011**	**0.038**	**0.006**	**0.032**	0.999	0.172	**<0.0001**	0.906	0.074	**0.023**	**0.028**

2.4 Discussion

Diatoms are a very important group of microalgae populating all aquatic niches and able to fix about 20 % of global carbon production[28]. Their elevated adaptability is due to a rich set of metabolic pathways coded by their genome[29]. Their crucial ecological role[30,31] is coupled to an emerging biotechnological interest due to their ability to produce high added value molecules. One of the metabolisms recently identified in these microalgae is the pathway for the enzymatically-mediated synthesis of PGs, principally studied in mammals but less in plants or microorganisms[15,23].

The key step in PG biosynthesis is catalysed by COX enzymes belonging to the heme-peroxidase protein superfamily. Evolutionary studies of this superfamily demonstrated an independent evolution of four superfamilies, including peroxidase–cyclooxygenase[12], each possessing a peculiar folding of the heme peroxidase domain. In particular, the peroxidase-cyclooxygenase superfamily has a wide distribution in all living kingdoms[12]. Phylogenetic analysis of the *Thalassiosira rotula* COX presented here widens the analysis reported previously in Di Dato *et al.*[15] with the addition of a few sequences that have better resolved some of the clades identified in that study. In particular, the sequences from *Grammatophora oceanica* 5553, *Staurosira* 23824, *Chaetoceros cf. neogracile* 1470 clustered with *Nostoc* WP015113127 thereby questioning the validity of the analysis. It was not clear enough whether the clade grouping the three diatom and *Nostoc* sequences could be identified as a diatom specific clade derived from cyanobacteria. In the present analysis, new bacterial sequences were added and clustered together with the above-mentioned diatom and *Nostoc* sequences showing that this clade is indeed a bacterial-like clade. Moreover, our phylogenetic analysis highlights the existence of more than one peroxidase-cyclooxygenase enzyme related to the oxidative metabolism of fatty acids in diatoms. The four sub-families in the peroxidase-cyclooxygenase superfamily (family 4: cyclooxygenase and α-DOX; 5a and 5b, bacterial peroxidase-cyclooxygenase families) did not evolved from one single protein, but rather appeared independently during evolution[32]. COX, were formerly known as ‘animal heme-dependent peroxidases’. However, as demonstrated here and in our previous work[15,23], this enzyme is present and active in diatoms as well, thus the denomination ‘animal heme-dependent peroxidases’ would deserve a revision in order to avoid confusion.

α-DOX belongs to the same peroxidase–cyclooxygenase superfamily as COX and some diatoms (e.g. *S. marinoi*) possess this protein along with COX. This questions the origin and evolution of this class of enzymes in diatoms and shows that this group of microalgae

is capable of manipulating fatty acids in different ways: diatom α-DOX enzymes may act on medium chain saturated fatty acids, such as 16:033, while COX on long chain-PUFAs (e.g. AA and EPA)[34]. Also, the close proximity of bacterial and diatom sequences and the topology of the tree we present, questions whether diatoms acquired the genes coding for COX via lateral gene transfer[35], which occurred during diatom evolution[36–38] or whether these genes were acquired during endosymbiotic events, at least for species like *S. marinoi* possessing both α-DOX-like and COX-like enzymes. In the case of *T. rotula*, α-DOX-like is absent, while the pennate model diatom *Phaeodactylum tricornutum* completely lacks the enzymatic pathway leading to prostaglandins though it can produce isoprostanoids non-enzymatically[22]. Indeed, the ability to synthesize eicosanoids, including lipoxygenase-derived oxylipins and aldehydes, is not robustly conserved in diatoms[39,40] where genes involved in these metabolisms might have been lost during evolution.

In our *in silico* study of TrotCOX protein sequence and structure we found an overall homology and the conservation of motifs characterizing the heme-peroxidase protein superfamily, corroborating also the assignment of this sequence to the peroxidase-cyclooxygenase superfamily[12]. The catalytic sites, the motifs and their order along the protein sequence, are very conserved compared to the representative sequences of the peroxidase-cyclooxygenase superfamily, even if with some differences. The glutamate embedded in the characteristic motif XRXXEX, usually conserved, in TrotCOX is mutated into a leucine. This mutation seems to be characteristic of diatoms, since not only TrotCOX but also the other diatom sequences considered for the phylogenetic analysis show this mutation. Considering the different cellular environments, the leucine residue may still play in diatoms the same role as glutamic acid in other organisms although the glutamic acid is a charged amino acid while leucine is a hydrophobic uncharged amino acid. In addition, in the proximal heme sides, the conserved asparagine and arginine normally linked via a hydrogen bond, playing crucial roles in alpha helix structures[41], are substituted with a threonine and a cysteine, respectively. This change should not compromise the alpha helix stability present in this region.

Our experimental data demonstrate that TrotCOX is active and works like the animal COX, being able to synthesize the same PGs produced by animals. The presence of PGE_3, a specific derivative of EPA, is in line with the abundance of this PUFA precursor, which represents a hallmark of diatoms[42]. Interestingly, the general increase of PGs released outside the cells started from day 5 onwards, i.e., when cells were entering the stationary phase that coincides with the maximum expression of COX. This stage is generally

associated with an increase in the naturally associated bacterial population. During the transition between the stationary and senescent phases, nutrients decrease while dying algal cells generate organic matter that can feed associated bacteria. Bacteria are able to assimilate phosphate better than algae, especially at low concentrations[43], and compete with algae for inorganic nutrients[44]. This increased pressure may stimulate PGs synthesis in diatoms as already shown in animals, where COX expression is stimulated in the presence of bacteria[45] and in humans where PGE_2 have been correlated to viral load and infection severity in influenza[46–49]. In line with this hypothesis, similar compounds, such as hydroxylated eicosapentaenoic acid 15-HEPE, were shown to be up-regulated and released from cells when the diatom *S. marinoi* was exposed to pathogenic bacteria[50].

PGE_2 is the most abundant prostanoid in the human body, but it is also very unstable since it is rapidly converted into PGE-M[46–49] that is considered to mirror the systemic levels of PGE_2 formation. In our study, PGE_2 was present at low levels in the medium during the exponential growth phase and was almost absent in the late stationary phase (from day 7) while PGE-M levels increased exponentially up to the senescent phase (Figure 2.5). The peak in the expression of the related enzyme, PTGES, at the onset of the stationary phase and the sustained release of PGs afterwards suggest a possible role of prostaglandins in communication and cell signalling. Their release outside the cells, in addition to their sustained presence in a saline environment, is quite striking and has never been observed before in diatoms. PGs are known to exert a wide range of effects in different organisms, including the induction of inflammatory processes, injury and pain in humans where they have been best studied. The fact that they have been identified in organisms ranging from unicellular diatoms, corals and jellyfish to arthropods, molluscs and mammals denotes that they regulate important physiological processes that have been conserved through evolution.

In conclusion, this study confirms that diatoms possess a molecular toolbox generally believed to be unique to higher organisms such as mammals. The production and release of PGs by some diatoms and the variation in the expression levels of PG biosynthetic pathway during different growth stages, strongly suggest a relevant and possible signalling role for these molecules within the plankton community that needs to be further investigated. This characteristic may have contributed to render diatoms one of the most successful groups of organisms in the world's oceans.

2.5 Materials and Methods

2.5.1 Strain and cell cultures

Thalassiosira rotula, strain FE80C, was isolated in 2011 in the Gulf of Naples (40°48.5'N, 14°15'E), Mediterranean Sea. Clonal cultures were established by isolating single cells or short chains from phytoplankton net samples collected from the surface layer of the water column. Cultures were grown in sterile filtered oligotrophic seawater at 36 ppt amended with f/2[51] nutrients and maintained at a temperature of 20 °C, at 12:12 h light:dark cycle, and with a photon flux of 100 μmol photons $m^{-2}s^{-1}$.

10 L cultures of *T. rotula*, in triplicate, were used to follow their growth from day 3 to day 10. Every day, 250 mL of each culture was harvested by filtration onto 1.2 μm pore size filters (RAWP04700, Millipore, Burlington, Massachusetts, USA) and immediately frozen in liquid nitrogen. 100–200 mL of culture media recovered from the cell filtration was collected and stored at −80 °C until sample processing. Initial cell concentrations were about 5000 cells/mL upon inoculation. Culture growth was monitored daily from samples fixed with one drop of Lugol (final concentration of about 2 %) and counted in a Bürker counting chamber under an Axioskop 2 microscope (20×) (Carl Zeiss GmbH, Jena, Germany).

2.5.2 RNA extraction and reverse transcription

To proceed to total RNA extraction, filters were covered with 1.5 mL TRIsure (BIO-38033, Bioline Reagents Ltd., London, UK) to which glass beads (G1277, Sigma-Aldrich, St. Louis, Missouri, USA) were added. Cells were disrupted on a thermo-shaker (Eppendorf ThermoMixer C, Eppendorf, Amburg, Germany) at 60 °C for 10 min at 1200 rpm. Filter and glass beads were discarded, and the extraction was continued according to the TRIsure instructions. DNase treatment was carried out using DNase I recombinant, RNase-free (Roche, Basel, Switzerland) according to manufacturer's protocol to eliminate potential genomic DNA contamination. The efficiency of the DNase digestion was checked by testing DNA primer capability to amplify an amplicon with traditional PCR. PCR reactions were carried out in 25 μL volume with 2.5 μL of 10× PCR reaction buffer (Roche, Basel, Switzerland), 2.5 μL of 2 mM dNTP, 0.3 μL of 5 U/μL Taq (Roche, Basel, Switzerland), 1 μL of 20 pmol/μL for each oligo, 1 μL of RNA templates and nuclease-free water up to 25 μL. The PCR program consisted in a denaturation step at 95 °C for 5 min, 35 cycles at 95 °C for 30 s, 53 °C for 30 s and 72 °C for 45 s, and a final

extension step at 72 °C for 7 min. Amplified PCR products were analysed by agarose gel electrophoresis. Total RNAs were purified and concentrated using RNeasy MinElute Cleanup Kit (Cat. No./ID: 74204, Qiagen, Hilden, Germany) and eluted in 30 µL of RNase-free water. Concentrations of the resulting RNA samples were assessed by absorbance at 260 nm (ND-1000 Spectrophotometer; NanoDrop Technologies, Wilmington, Delaware, USA). The integrity of total RNA was checked by agarose gel electrophoresis. One microgram of each RNA sample was retro-transcribed in cDNA following the manufacturer's instructions (5X All-In_One RT Master Mix, abm, Applied Biological Materials Inc., Richmond, Canada) using the T100 Thermal cycler (Bio-Rad Laboratories, Hercules, California, USA). Retro-transcribed samples were again checked for DNA contamination by traditional PCR, as above, using intron-spanning primer pairs.

2.5.3 Phylogenetic analyses

For cyclooxygenase phylogeny, the alignment from Di Dato *et al*.[15] was retrieved and all the sequences were blasted against the NCBI database. The best hits from each pBLAST were added to the analysis. Sequences were visualized and aligned in BioEdit Sequence Alignment Editor[52] using the ClustalW alignment algorithm implemented in BioEdit. The alignment was performed using MUSCLE online software and compared to the ClustalW output. The two alignments were identical. For ML phylogenetic analysis both MEGAX[53] and RAxML were used. Model test implemented in MEGAX identified the LG + G + I as the best substitution model on the basis of the Bayesian Information Criterion (BIC). 1000 bootstrap replicates were performed. Two different analyses were run in RAxML, one with the boot stopping option and one with 100 bootstrap replicates. All the analyses involved 58 amino acid sequences and 612 positions (including gaps). The two RAxML and the MEGA X trees were identical in topology and support. Phylogenetic trees were visualized and edited in the FigTree (Tree Figure Drawing Tool Version 1.4.3) software (http://tree.bio.ed.ac.uk/). The complete list of species and sequences used in the present work for phylogenetic analyses are listed in Supplementary Table 2.1. The corresponding sequences are reported at the end of the Supplementary File, both the aligned sequences (with gaps) and the ungapped sequences. The sequences from strain SkelB that were annotated as *Skeletonema dohrnii* in MMETSP database were changed to *S. marinoi* after ribosomal sequence identification. Sequence ID were left unchanged. The phylogenetic tree was built on scale in the branch length (scale bar reported on Figure 2.3). Sequences

are identified by the species name, followed by their MMETSP ID (deprived of the taxon).

2.5.4 *In-silico* TrotCOX protein structure modelling and comparison

We used Phyre2 (Protein Homology/analogY Recognition Engine V 2.0)[24] to predict the possible structure of TrotCOX protein. In order to compare the TrotCOX predicted structure to known models of the peroxidase-cyclooxygenase family, the BtLPO (3BXI. pdb, 10.2210/pdb3bxi/pdb)[12,25] and OaCOX1 structures were retrieved from the RCSB database (https://www.rcsb.org) and compared to TrotCOX using PyMOL (Molecular Graphics System, Version 2.0 Schrödinger, LLC).

2.5.5 Bioinformatic identification of the prostaglandin pathway

Pathway annotated as 'prostaglandin biosynthesis' was found among the second level pathways list generated within the Annocript pipeline annotation of the proteome from the *T. rotula* CCMP 1647 RNA-seq. Transcripts associated to the pathway were extrapolated from the total proteome annotation table[23].

2.5.6 Primer design and real time quantitative PCR

Candidate reference genes and genes of interest were selected considering the annotation of the peptides reported in the annotated transcriptome of *T. rotula* FE80 (CCMP1647)[23]. Oligo sequences utilized[23] are listed in Table 2.5.

Table 2.5. List of oligo sequences. Abbreviations: Tm= Temperature of melting.

Gene	Primer Sequence Forward	Primer Sequence Reverse	Length (bp)	Tm °C	GC content (%)	Amplicon Length (bp)
COX-1 - PgG/Hs2	TCATCAAGGGA GGAGAATGG	CTTCCACCAAG AGCGAAGAC	20	58.4/60.5	50/55	170
PgEs2	TTCCAAACAGG GCAAGTTAC	TTGCACGAGAC AGATTGGAG	20	56.4/58.4	45/50	183
PgFs	TCTCCCCTATCG AGGGTTCT	AGCTCCACTCTG CTATCC	20/18	60.5/56.3	55/56	114
TBP	CCTTCTTCAACC CCTCCACCAAC	GTTCGCTCATCC CACGTTTTCG	23/22	66.6/64.2	57/55	161

ACTIN	TCGGCCCTTGAG AAGAGTTTCG	GATGGTCTGGA AAGTGGAGTCC	22	64.2/64.2	55/55	147

Each sequence was initially tested by standard PCR in a 25 µL final volume with 2.5 µL of 10× PCR reaction buffer (Roche, Basel, Switzerland), 2.5 µL of 10 × 2 mM dNTP, 0.3 µL of 5 U/µL Taq (Roche, Basel, Switzerland), 1 µL 10 µM of each oligo, 1 µL of cDNA templates and nuclease-free water up to 25 µL. Standard PCR amplification program was used, i.e. 95 °C for 3 min, 40 cycles at 95 °C for 30 s, 53 °C 30 s, 72 °C for 30 s, and a final extension step at 72 °C for 7 min. Amplified PCR products were analysed by agarose gel electrophoresis and verified by sanger sequencing. Reverse transcription-quantitative PCR (rt-qPCR) experiments were performed in MicroAmp Optical 384-Well reaction plate (Applied Biosystems, Waltham, Massachusetts, USA) with Optical Adhesive Covers (Applied Biosystems, Waltham, Massachusetts, USA) in a Viia7 Real Time PCR System (Applied Biosystem, Waltham, Massachusetts, USA). Five serial dilutions of mixed cDNAs were used to determine primer reaction efficiency[23] using the formula: $E = 10^{-1/slope}$. The PCR volume for each sample was 10 µL, with 5 µl of SensiFAST TM SYBR® Lo-ROX Kit (BIO_94020, Bioline Reagents Ltd., London, UK), 1 µL of cDNA template (1 to 5 dilution each template) and 4 µL of 0.7 µM oligo mix (forward and reverse). Program reaction used was: 95 °C for 20 s, 40 cycles of 95 °C for 1 s and 60 °C for 20 s. The program was set to reveal the melting curve of each amplicon from 60 °C to 95 °C, and read every 0.5 °C. Single peaks for all genes confirmed gene-specific amplification and the absence of primer-dimers. All RT-qPCR reactions were carried out in triplicate to capture intra-assay variability. Each assay included three no-template negative controls for each primer pair. The normalized expression levels of each gene of interest relative to the most stable reference genes[23], actin and TBP, were calculated by using the Q-Gene tool[54]. Only TBP normalized values were reported in the main text and figure results. Relative expression ratios above two fold were considered significant[23].

2.5.7 Prostaglandin extraction

1 µg of Prostaglandin E_2-d4 (CAYMAN CHEMICAL, Michigan, USA) was added as internal standard to the media (100–200 mL) recovered from the cell culture, from day 3 to day 10, through a filtration step on 1.2 µm nitrocellulose filters (RAWP04700, Millipore, Burlington, Massachusetts, USA). Culture media were loaded onto pre-packed CHROMABOND® HR-X cartridges (500 mg/6 mL) previously activated with methanol

(12 mL) and milliQ water (12 mL). After a preliminary desalting step with 12 mL of milliQ water, collection of the organic components was achieved by elution with 16 mL of methanol followed by 16 mL of methanol/dichloromethane (1:1). The two organic fractions were combined and dried under reduced pressure for LC-MS/MS analysis.

2.5.8 LC-MS/MS analysis

Thalassiosira rotula water samples extracted as above, were re-suspended in 0.5 mL methanol and analyzed by tandem mass spectrometry in MRM (Multiple Reaction Monitoring) mode on a 4000 QTRAP® LC/MS/MS System (Applied Biosystems, Waltham, Massachusetts, USA), working in negative ion mode and coupled to a 1100 nanoHPLC system (Agilent Technologies, Santa Clara, California, USA). Prostaglandins were separated by using a micro C18 column (10 cm ×1.0 mm, 5 μm). The mobile phase was generated by mixing eluent A (water, 0.1 % Acetic Acid) and eluent B (acetonitrile/isopropanol 50/50) and the flow rate was 30 nL/min. Elution started at 20 % B up to 95 % B in 15 min. Tandem mass spectrometry was performed using a turbo ion spray source operating in negative mode, and the multiple reaction monitoring (MRM) mode was used for the selected analytes. Mass parameters (4000 qtrap AB science) were as follows: curtain gas 20 psi, GS 1/2 50/50 psi, ion spray voltage −5500 V, DP, −60 V; Dwell Time 25 ms and Temperature of 550 °C. Quantitative analysis was performed by monitoring a unique production[26] arising from collision-induced fragmentation of the deprotonated selected parent compound after proper optimization of mass spectral parameters. After testing the scan mode including all the transitions, only the most intense transitions were chosen for each molecule.

2.5.9 Statistical analysis

One-way ANOVA ($\alpha = 0.05$) with Tukey's post-hoc test was performed using GraphPad Prism6.0 (GraphPad Software Inc. San Diego, California, USA). The analysis was performed to determine significant differences among the time points (days) considered during the *T. rotula* growth. Both qPCR amplification and LC-MS/MS data have been analysed.

2.6 References

1. Gladyshev, M. I., Sushchik, N. N. & Makhutova, O. N. Production of EPA and DHA in aquatic ecosystems and their transfer to the land. *Prostaglandins Other Lipid Mediat.* **107**, 117–126 (2013).

2. Serhan, C. N., Chiang, N., Dalli, J. & Levy, B. D. Lipid mediators in the resolution of inflammation. *Cold Spring Harb. Perspect. Biol.* **7**, a016311 (2014).

3. Wiktorowska-Owczarek, A., Berezińska, M. & Nowak, J. Z. PUFAs: Structures, Metabolism and Functions. *Adv. Clin. Exp. Med.* **24**, 931–941 (2015).

4. Fullerton, J. N. & Gilroy, D. W. Resolution of inflammation: a new therapeutic frontier. *Nat. Rev. Drug. Discov.* **15**, 551–567 (2016).

5. Korbecki, J., Baranowska-Bosiacka, I., Gutowska, I. & Chlubek, D. Cyclooxygenase pathways. *Acta Biochim. Pol.* **61**, 639–649 (2014).

6. Versteeg, H. H., van Bergen en Henegouwen, P. M., van Deventer, S. J. & Peppelenbosch, M. P. Cyclooxygenase-dependent signalling: molecular events and consequences. *FEBS Lett.* **445**, 1–5 (1999).

7. Di Costanzo, F., Di Dato, V., Ianora, A. & Romano, G. Prostaglandins in Marine Organisms: A Review. *Mar Drugs* **17**, (2019).

8. Clària, J. Cyclooxygenase-2 biology. *Curr. Pharm. Des.* **9**, 2177–2190 (2003).

9. Schultz, J. C. & Appel, H. M. Cross-Kingdom Cross-Talk: hormones shared by plants and their insect herbivores. *Ecol.* **85**, 70–77 (2004).

10. Ruan, J. *et al*. Jasmonic Acid Signaling Pathway in Plants. *Int. J. Mol. Sci.* **20**, (2019).

11. Groenewald, E. G. & van der Westhuizen, A. J. Prostaglandins and related substances in plants. *Botanical Rev.* **63**, 199–220 (1997).

12. Zámocký, M. *et al*. Independent evolution of four heme peroxidase superfamilies. *Arch. Biochem. Biophysics* **574**, 108–119 (2015).

13. Varvas, K. *et al*. Direct evidence of the cyclooxygenase pathway of prostaglandin synthesis in arthropods: genetic and biochemical characterization of two crustacean cyclooxygenases. *Insect Biochem. Mol. Biol.* **39**, 851–860 (2009).

14. Koljak, R. *et al*. The basis of prostaglandin synthesis in coral: molecular cloning and expression of a cyclooxygenase from the Arctic soft coral *Gersemia fruticosa*. *J. Biol. Chem.* **276**, 7033–7040 (2001).

15. Di Dato, V. *et al*. Animal-like prostaglandins in marine microalgae. *ISME J.* **11**, 1722–1726 (2017).

16. Flori, S. *et al*. Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nature Communications* **8** (2017).

17. Tanaka, A. *et al*. Ultrastructure and membrane traffic during cell division in the marine pennate diatom *Phaeodactylum tricornutum*. *Protist* **166**, 506–521 (2015).

18. De Martino, A., Amato, A. & Bowler, C. Mitosis in diatoms: rediscovering an old model for cell division. *BioEssays* **31**, 874–884 (2009).

19. Bowler, C. *et al*. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nat.* **456**, 239–244 (2008).

20. Moustafa, A. *et al*. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Sci.* **324**, 1724–1726 (2009).

21. Matsuda, Y., Hopkinson, B. M., Nakajima, K., Dupont, C. L. & Tsuji, Y. Mechanisms of carbon dioxide acquisition and CO 2 sensing in marine diatoms: a gateway to carbon metabolism. *Phil. Trans. R. Soc. B* **372**, 20160403 (2017).

22. Lupette, J. *et al*. Non-enzymatic synthesis of bioactive isoprostanoids in the diatom *Phaeodactylum* following oxidative stress. *Plant. Physiol.* **178**, 1344–1357 (2018).

23. Di Dato, V. *et al*. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Sci. Rep.* **9**, 9893 (2019).

24. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

25. Singh, A. K. *et al*. Inhibition of lactoperoxidase by its own catalytic product: crystal structure of the hypothiocyanate-inhibited bovine lactoperoxidase at 2.3-Å resolution. *Biophysical J.* **96**, 646–654 (2009).

26. Wang, Y., Armando, A. M., Quehenberger, O., Yan, C. & Dennis, E. A. Comprehensive ultra-performance liquid chromatographic separation and mass spectrometric analysis of eicosanoid metabolites in human samples. *J. Chromatogr. A* **1359**, 60–69 (2014).

27. Yang, J., Schmelzer, K., Georgi, K. & Hammock, B. D. Quantitative profiling method for oxylipin metabolome by liquid chromatography electrospray ionization tandem mass spectrometry. *Anal. Chem.* **81**, 8085–8093 (2009).

28. Kroth, P. G. *et al*. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* **3**, e1426 (2008).

29. Busseni, G. *et al*. Meta-omics reveals genetic flexibility of diatom nitrogen transporters in response to environmental changes. *Mol. Biol. Evolution* **36**(11), 2522–2535 (2019).

30. Vardi, A., Thamatrakoln, K., Bidle, K. D. & Falkowski, P. G. Diatom genomes come of age. *Genome Biol.* **9**, 245 (2008).

31. Bozarth, A., Maier, U.-G. & Zauner, S. Diatoms in biotechnology: modern tools and applications. *Appl. Microbiol. Biotechnol.* **82**, 195–201 (2008).

32. Zamocky, M., Jakopitsch, C., Furtmüller, P. G., Dunand, C. & Obinger, C. The peroxidase-cyclooxygenase superfamily: reconstructed evolution of critical enzymes of the innate immune system. *Proteins* **72**, 589–605 (2008).

33. Hamberg, M., Ponce de Leon, I., Rodriguez, M. J. & Castresana, C. Alpha-dioxygenases. *Biochem. Biophys. Res. Commun.* **338**, 169–174 (2005).

34. Smith, W. L., Garavito, R. M. & DeWitt, D. L. Prostaglandin endoperoxide H synthases (cyclooxygenases)-1 and -2. *J. Biol. Chem.* **271**, 33157–33160 (1996).

35. Sieber, K. B., Bromley, R. E. & Dunning Hotopp, J. C. Lateral gene transfer between prokaryotes and eukaryotes. *Exp. Cell Res.* **358**, 421–426 (2017).

36. Sorhannus, U. Evolution of antifreeze protein genes in the diatom genus *Fragilariopsis*: evidence for horizontal gene transfer, gene duplication and episodic diversifying selection. *Evol. Bioinform. Online* **7**, 279–289 (2011).

37. Raymond, J. A. & Kim, H. J. Possible role of horizontal gene transfer in the colonization of sea ice by algae. *PLoS ONE* **7**, e35968 (2012).

38. Diner, R. E. *et al*. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc. Natl. Acad. Sci. USA* **114**, E6015–E6024 (2017).

39. Teng, L. *et al*. Evolution and expansion of the prokaryote-like lipoxygenase family in the brown alga *Saccharina japonica*. *Front. Plant. Sci.* **8**, 2018 (2017).

40. Orefice, I. Structural and functional analysis of lipoxygenase enzymes in marine diatoms. (The Open University of London, 2013).

41. Gray, T. M. & Matthews, B. W. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α-helices and its relevance to membrane-bound proteins. *J. Mol. Biol.* **175**, 75–81 (1984).

42. Stonik, V. & Stonik, I. Low-molecular-weight metabolites from diatoms: structures, biological roles and biosynthesis. *Mar. Drugs* **13**, 3672–3709 (2015).

43. Amin, S. A. *et al*. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nat.* **522**, 98–101 (2015).

44. Grossart, H.-P., Czub, G. & Simon, M. Algae-bacteria interactions and their effects on aggregation and organic matter flux in the sea. *Environ. Microbiol.* **8**, 1074–1084 (2006).

45. Martínez-Colón, G. J. & Moore, B. B. Prostaglandin E_2 as a regulator of immunity to pathogens. *Pharmacology Therapeutics* **185**, 135–146 (2018).

46. Medina, S. *et al*. Assessment of oxidative stress markers and prostaglandins after chronic training of triathletes. *Prostaglandins Other Lipid Mediat.* **99**, 79–86 (2012).

47. Idborg, H. *et al*. Evaluation of urinary prostaglandin E_2 metabolite as a biomarker in infants with fever due to viral infection. *Prostaglandins Leukot. Essent. Fat. Acids* **91**, 269–275 (2014).

48. Kekatpure, V. D. *et al*. Elevated levels of urinary PGE-M are found in tobacco users and indicate a poor prognosis for oral squamous cell carcinoma patients. *Cancer Prev. Res.* **9**, 428–436 (2016).

49. Aoki, T. & Narumiya, S. Prostaglandins and chronic inflammation. *Trends Pharmacol. Sci.* **33**, 304–311 (2012).

50. Meyer, N., Rettner, J., Werner, M., Werz, O. & Pohnert, G. Algal oxylipins mediate the resistance of diatoms against algicidal bacteria. *Mar. Drugs* **16**, 486 (2018).

51. Guillard, R. R. L. Culture of phytoplankton for feeding marine invertebrates. In *Culture of Marine Invertebrate Animals* (eds. Smith, W. L. & Chanley, M. H.) 29–60 (Springer US, 1975).

52. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Ser.* **41**, 95–98 (1999).

53. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

54. Simon, P. Q-Gene: processing quantitative real-time RT-PCR data. *Bioinforma.* **19**, 1439–1440 (2003).

55. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18**(20), 6097–6100 (1990).

Chapter 3

Study of the sterols biosynthetic pathway and evaluation of sterol content under different nutritional conditions in *T. rotula*

Paper in preparation for publication:

Di Costanzo F., Di Dato V., Cutignano A., Romano G.

3.1 Abstract

Sterols are a class of amphipathic lipids with a plethora of different bioactivities, such as hypocholesterolemic, anti-diabetic, anti-inflammatory, anti-cancer, neuro-modulatory and neuroprotective ones, which makes them attracting for the market. Diatoms demonstrated to produce a great variety of sterols, among which the 24-methlylene cholesterol (24-MC) is one of the most abundant. Sterol biosynthetic pathway has been studied only in few species in recent years, highlighting several differences among their most specialized steps that lead to the production of different sterols.

The preliminary results here reported highlighted the structure of the sterol biosynthetic pathway in the diatom *Thalassiosira rotula*. With the goal to identify the optimal condition to enhance sterols production in this diatom, we evaluated the gene expression and the quantity of end-product in cultures grown in standard conditions and under nitrogen, silica and phosphate starvation, harvested at the stationary growth phase. Overall, our results highlighted the production of 24-MC as major sterol and a lower production of this compound in all the stress conditions tested. The better understanding of the structure of *T. rotula* sterol pathway obtained through this study could be useful to associate genetic and metabolic manipulation methods to the canonical cultivation methods to elicit sterols production in this species.

3.2 Introduction

Sterols are triterpenoids of fundamental importance in all the eukaryotes, being involved in the regulation of membrane fluidity and dynamics, in signal transduction, in binding of different molecules to the cell surface and in the production of secondary metabolites and hormones [1–3]. In particular plant sterols i.e. phytosterols, increase membrane cohesion allowing plants to be less sensitive to temperature shocks[4]. Phytosterols are also involved in morphogenesis[5] and embryonic growth[4]. In microalgae, they influence photosynthesis, fatty acid metabolism and growth[5] and some specific sterols, namely sterols sulfates, demonstrated to be involved in triggering cell death in the diatom *Skeletonema marinoi*, likely as regulator of blooms dynamics[6].

Precursors of phytosterols are isoprenoids, coming out from the cytosolic mevalonate (MVA) pathway, whilst the non-mevalonate or 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, located in the plastid, is the precursor of carotenoids, pigments and other isoprenoids[1,5]. Green algae and some red algae lost the MVA pathway during the

evolution and are able to use the MEP one to produce also sterols building blocks[1,5]. In diatoms there are contrasting information about the biosynthesis of sterols precursors, since in some species, as *Nitzschia ovalis*, *Rhizosolenia setigera*, *Cyclotella criptica* and *S. marinoi*, only the MVA pathway is active while in others, like *Haslea ostrearia* and *Chaetoceros muelleri* only the MEP one[5,7]. Additionally, some other diatoms, like *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, possess both MEP and MVA pathways[7,8]. The distribution of the MVA and MEP pathways in diatoms according to *in silico* transcriptome analysis is reported in Di Dato *et al.*,[8]. Figure 3.1 schematically reports the enzymes involved in both the MEP and MVA pathways, both leading to the biosynthesis of isopentenyl pyrophosphate (IPP).

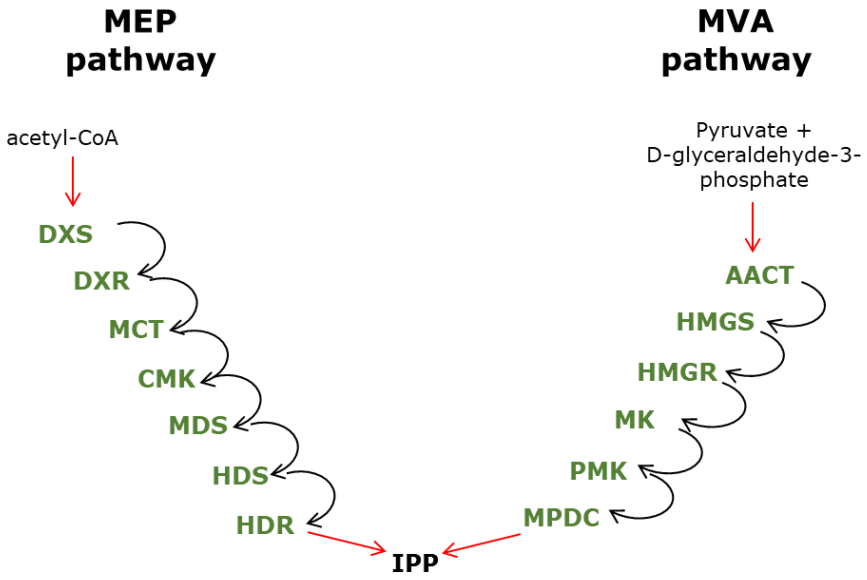


Figure 3.1. Scheme reporting the enzymes involved in the MEP and MVA pathways with the indication of substrates and final products. Abbreviations corresponding to each enzyme are reported in Tables 3.1 and 3.2 in the Results paragraphs 3.3.1 and 3.3.2.

The dimerization of farnesyl pyrophosphate (FPP), obtained from the isopentenyl pyrophosphate (IPP) derived from the MEP/MVA pathways, allows the formation of squalene that, both in plants, fungi and animals, is converted into 2,3 epoxysqualene by a flavoprotein monooxygenase named squalene epoxidase (SQE)[1,5,7]. The activity of SQE represents a checkpoint in animal cholesterol synthesis since it is regulated both by sterol levels and by a post-translational regulation[7]. On the contrary, the mechanisms by which microalgae sense sterols levels and regulate their biosynthesis are not well understood[9]. In diatoms, 2,3 epoxysqualene formation is catalyzed by an alternative SQE (AltSQE) that, differently from the canonical SQE, belongs to the fatty acid hydroxylase

superfamily[10]. AltSQE is mutually exclusive with the canonical SQE since no organisms harboring both the enzymes have been identified until now[10]. In addition, AltSQE homologous sequences have been identified also in other not-diatom species demonstrating that, as for SQE, it is widespread in the tree of life[10]. In the successive biosynthetic step, 2,3 epoxysqualene is cyclized by lanosterol synthase to form lanosterol in animals and fungi and by cycloartenol synthase to form cycloartenol in plants[1]. The possible occurrence of animal/fungal and plant reactions also in diatoms has been inferred in *P. tricornutum*, *C. muelleri* and *T. pseudonana*, by transcriptomic analysis and the use of chemical inhibitors[11], as well as in *S. marinoi* and *C. cryptica* by transcriptome analysis, labelling methods and GC-MS profiling[5]. All these species demonstrated to possess only the cycloartenol synthase enzyme, from which they are also able to synthetize cholesterol[5,11]. Other diatoms, as the pennate diatom *Halsea sp.*, seem to produce lanosterol and its derived sterols, while *Stauroneis simulans* produce sterols both from lanosterol and cycloartenol[12]. Overall, given the fact that sterol biosynthetic pathway has been extensively studied only in few diatoms species until now, it is not yet completely clear which is the route, via cycloartenol or lanosterol, by which diatom synthetize sterols[2]. All these core reactions, that allow the formation of 24-methylfecosterol, are conserved in diatoms while the downstream reactions become particularly complex and specialized[1,7].

Contrary to the scarcity of information regarding the biosynthetic pathway, the sterol profile of these microalgae have been extensively studied[13–15]. While animals and fungi possess a sterol composition mainly dominated by cholesterol and ergosterol, respectively, and plants possess a variegated sterols pool comprising cholesterol, stigmasterol, sitosterol, campesterol, a 24-methyl sterol and two 24-ethyl sterols as principal constituents[16], in diatoms more than 25 different sterols have been reported until now[2,12]. All diatoms possess sterols with 27-29 carbon atoms with the exception of *S. simulans* that possess also C_{30} and C_{31} sterols[2,12]. An extensive study of sterol profile in several diatoms revealed a high variability and complexity[12] with the 24-methylcholesta-5,24(24^1)-dien-3β-ol, diatomsterol (a C-24 epimer of brassicasterol) and 24-methylene cholesterol being the predominant sterols[2,3,5]. Even if they are not exclusive of a certain taxonomic group, in some cases sterol distribution can be useful for diatoms taxonomy as, for example, for *Thalassiosirales* that have a characteristic sterol profile with a higher abundance of 24-methylene cholesterol[3].

Since sterols functions in microalgae are related to the interaction with the environment, several studies have been conducted to evaluate sterol content in different growth conditions[9]. In some microalgae, sterols distribution can be influenced by parameters as salinity, temperature and nutrients availability, including nitrogen, phosphate and silica[9,12]. In some cases, the environmental stresses applied changed the profile of only minor sterols. As an example, when *P. tricornutum* is grown at 4 °C, the amount of campesterol increased and cholesterol decreased while brassicasterol, the most abundant in this diatom, remains unaltered[9]. In the same way, salinity changes influenced only the minor sterols of *P. tricornutum*, *C. muelleri* and *T. pseudonana* where the amount of the principal sterol did not change[9]. Other studies have been focused on the identification of the best growth condition to stimulate phytosterols accumulation. Indeed, as reviewed by Randhir *et al.*[17], in general a higher sterol content can be achieved: (i) cultivating diatoms until stationary growth phase (*T. pseudonana* represent an exception since it has equal amount of sterols also during exponential growth); (ii) modulating nutrients concentration; (iii) increasing salinity, temperature and light; (iv) or applying combinations of these conditions.

Plant phytosterols have been commercialized in nutraceutical industry for their hypocholesterolemic properties, being incorporated into various foods, and possess also anti-diabetic, anti-inflammatory neuro-modulatory and neuroprotective activities[2,17]. In addition, squalene, mainly derived from shark liver oil for its commercial uses, has a documented activity in improving skin hydration, finding application in the cosmeceutical sector[17]. Squalene and its hydrogenated form squalane are also principal components of parenteral emulsions used for vaccines and drug delivery[17]. Campesterol and β-sitosterol possess anti-proliferative activities against breast, colon and prostate cancer cells, fucosterol against HeLa cells (human cervical cell lines)[18] while the stigmasterol produced by the diatom *Navicula incerta* possess apoptotic activity against hepatocarcinoma cells[19].

Nowadays, vegetables and tall oils derived from land plants represent the main source of sterols but they are not able to satisfy the increasing demand of these molecules by the market[17]. Microalgae can fulfill market needs due to their growth behavior, environmentally sustainable cultivation and the capability of modulating sterol content acting on environmental conditions[2,17]. Moreover, molecular approaches, as for example the use of RNA interference (RNAi), can be used to manipulate the diatoms whose

genomes or transcriptomes are available toward an augmented sterol production by blocking competitive pathways or over-expressing specific genes[17].

In this chapter of the thesis, I describe the reconstruction of the sterol biosynthetic pathway in the diatom *Thalassiosira rotula*. I have firstly evaluated the occurrence of MVA and/or MEP pathways in this diatom and their differential expression in transcriptomes obtained from cultures grown in control versus silica limited conditions (details regarding transcriptome sequencing are reported in Chapter 1). Moreover, the relative gene expressions of the key enzymes of the sterols biosynthetic pathway and the end-product profile and quantities have been investigated in cultures challenged with different nutritional stresses. The main goal of this study was to add knowledge to the phytosterols synthesis in diatoms and to obtain preliminary information regarding the potentiality of our *T. rotula* as source of valuable sterols for biotechnological applications.

3.3 Results

3.3.1 MVA/MEP biosynthetic pathways in *T. rotula*

The enzymes involved in both the 2-C-methyl-D-erythritol 4-phosphate (MEP) and the mevalonate (MVA) biosynthetic pathways in *T. rotula* were firstly searched in the annotation table of its transcriptome. The search was then refined by a blastp in the *T. rotula* transcriptome of the enzymes of *Catharanthus roseus*, a plant species in which these pathways have been extensively characterized[20]. The entire set of the enzymes in the MVA pathway have been retrieved, as reported in Table 3.1.

Table 3.1. List of transcripts involved in the MVA pathway found annotated in *T. rotula*. In bold underlined are reported the enzymes not annotated but identified through blastp search. Abbreviations: Transcr. ID= transcript identifier; Enzyme ID= enzymes EC numbers; Enzyme desc= enzymes description; Pw Lev1/2/3= Pathways of first/second/third level associated to the transcripts as they are found in UniPathway.

Transcr. ID	Enzyme ID	Enzyme desc. and Gene name	Pw Lev1	Pw Lev2	Pw Lev3
TR29102 c1_g1_i1	**2.3.1.9**	**Acetoacetyl-CoA thiolase (AACT)**	-	-	
TR16704 c0_g1_i2	2.3.3.10	Hydroxy-methylglutaryl-CoA synthase (HMGS)	Metabolic intermediate biosynthesis	(R)-mevalonate biosynthesis	(R)-mevalonate from acetyl-CoA: step 2/3
TR1112 c0_g1_i3	1.1.1.34	Hydroxyl-methyglutaryl-CoA reductase (HMGR)	Metabolic intermediate biosynthesis	(R)-mevalonate biosynthesis	(R)-mevalonate from acetyl-CoA: step 3/3
TR8551 c0_g1_i1	2.7.1.36	Mevalonate kinase (MK)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via mevalonate pathway	isopentenyl diphosphate from (R)-mevalonate: step 1/3
TR55782 c0_g1_i1	2.7.4.2	Phosphomevalonate kinase (PMK)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via mevalonate pathway	isopentenyl diphosphate from (R)-mevalonate: step 2/3

TR6289 c0_g1_i1	4.1.1.33	Diphosphomevalonate decarboxylase (MPDC)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via mevalonate pathway	isopentenyl diphosphate from (R)-mevalonate: step 3/3

Figure 3.2 shows the *in silico* expression, calculated as Reads Per Kilobase Million (RPKM) (Supplementary Table 3.1 a), of the MVA pathway in the two growth conditions, i.e. complete medium (CTR) and silica depleted medium (SiDepl). Hydroxyl-methyglutaryl-CoA reductase (HMGR), the rate-limiting enzyme in the pathway[7], resulted to be differentially expressed, being down-regulated in the SiDepl condition compared to the CTR. However, the high standard deviation made the difference not significant, suggesting also a high variability and instability of expression among the replicates in these conditions.

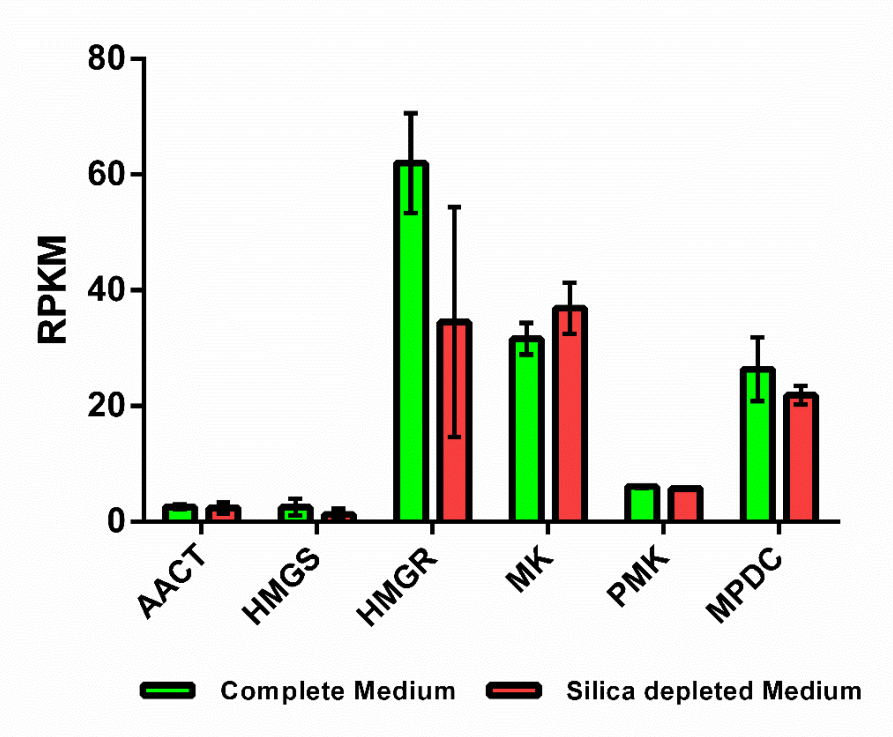


Figure 3.2. MVA pathway in *T. rotula*. Expression in RPKM of the enzymes in the mevalonate pathway identified in the *T. rotula* transcriptome. The comparison was made between transcriptome from cells grown in complete medium (green bars) and in silica depleted medium (red bars). The abbreviations corresponding to the enzymes names are listed in Table 3.1.

In the same manner, the search of the enzymes constituting the MEP pathway in *T. rotula* transcriptome retrieved the complete set, listed in Table 3.2.

Table 3.2. List of transcripts involved in the MEP pathway found annotated in *T. rotula*. In bold underlined are reported the enzymes not annotated but identified through blastp search. Abbreviations: Transcr. ID= transcript identifier; Enzyme ID= enzyme's EC number; Enzyme desc= enzyme's description; Pw Lev1/2/3= Pathways of first/second/third level associated to the transcripts as they are found in UniPathway.

Transcr. ID	Enzyme ID	Enzyme desc. and Gene name	Pw Lev1	Pw Lev2	Pw Lev3
TR37205 c0_g1_i1	2.2.1.7	1-deoxy-D-xylulose-5-phosphate synthase (DXS)	Metabolic intermediate biosynthesis	1-deoxy-D-xylulose 5-phosphate biosynthesis	1-deoxy-D-xylulose 5-phosphate from D-glyceraldehyde 3-phosphate and pyruvate: step 1/1
TR2820 c0_g1_i1	1.1.1.267	1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via DXP pathway	isopentenyl diphosphate from 1-deoxy-D-xylulose 5-phosphate: step 1/6
TR46467 c0_g1_i1	2.7.7.60	2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase. (MCT)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via DXP pathway	isopentenyl diphosphate from 1-deoxy-D-xylulose 5-phosphate: step 2/6
TR2889 c0_g1_i1	**2.7.1.148**	**4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK)**	-	-	-
TR30 c0_g2_i1	4.6.1.12	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECS)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via DXP pathway	isopentenyl diphosphate from 1-deoxy-D-xylulose 5-phosphate: step 4/6
TR12957 c0_g1_i2	1.17.7.1	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via DXP pathway	isopentenyl diphosphate from 1-deoxy-D-xylulose 5-phosphate: step 5/6
TR12704 c0_g1_i1	1.17.7.4	4-hydroxy-3-methylbut-2-enyl diphosphate reductase. (HDR)	Isoprenoid biosynthesis	isopentenyl diphosphate biosynthesis via DXP pathway	isopentenyl diphosphate from 1-deoxy-D-xylulose 5-phosphate: step 6/6

Figure 3.3 shows the differential expression, in RPKM (Supplementary Table 3.1 b), of the MEP pathway's enzymes between the two-growth conditions in *T. rotula*. The expression was similar for all the enzymes in the two conditions, except for 2-C-methyl-D-erythritol 4-phosphate cytidylyl transferase (MCT) and 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECS), that resulted to be significantly up-regulated in SiDepl compared to CTR (with p-values of 0.005 and 0.03, respectively).

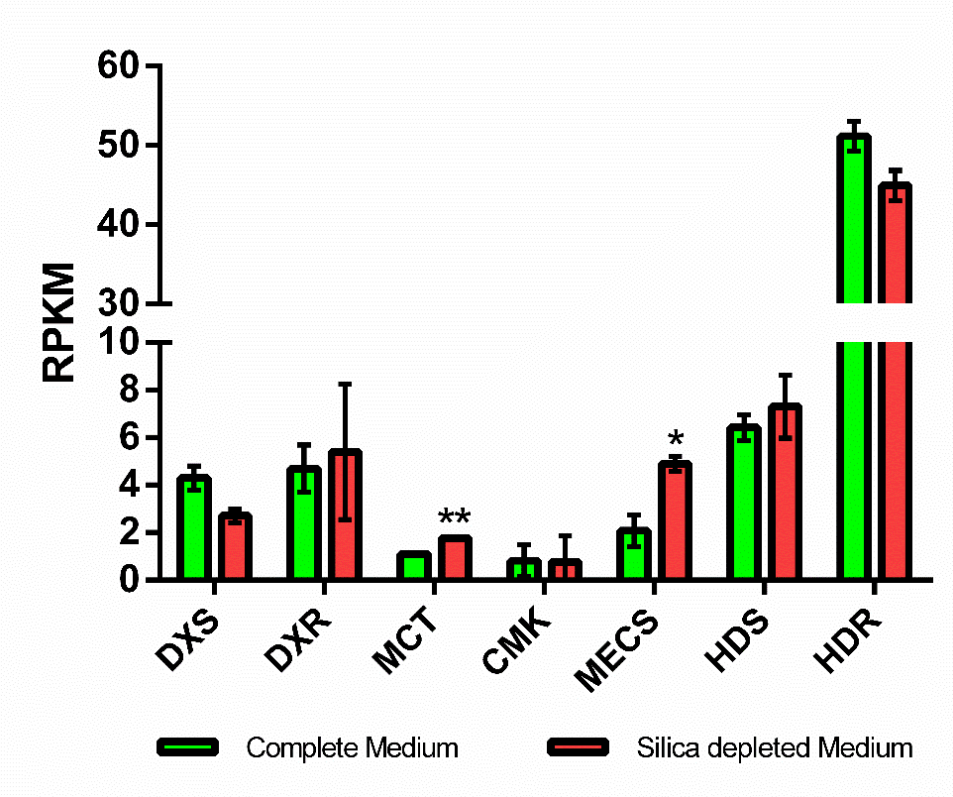


Figure 3.3. MEP pathway in *T. rotula*. Expression in RPKM of the enzymes in the MEP pathway identified in *T. rotula* transcriptome from cells grown in complete medium (green bars) compared to silica depleted medium (red bars). Significance was calculated by performing a multiple t-test, each row individually without correction for multiple comparisons. The abbreviations corresponding to the enzymes names are listed in Table 3.2.

3.3.2 Sterols pathway in *T. rotula*

To guide the search for the sterols biosynthetic pathway enzymes in *T. rotula* transcriptome, the enzymes annotated in the *S. marinoi* and *C. criptica* pathways[5] and in the pathways of *T. pseudonana*, *C. muelleri* and *P. tricornutum*[11] have been taken as query. The enzymes identified with this approach are listed in Table 3.3.

Table 3.3. List of transcripts involved in the sterols biosynthesis found annotated in *T. rotula* transcriptome. Abbreviations: Transcr. ID= transcript identifier; Enzyme ID= enzyme's EC number; Desc. SP= description of the best annotation result according to Uniprot_KB; Enzyme desc= enzyme's description; Pw Lev1/2/3= Pathways of first/second/third level associated to the transcripts as they are found in UniPathway.

Transcr. ID	Enzyme ID	Desc. SP	Enzyme desc. and Gene name	Pw Lev1	Pw Lev2	Pw Lev3
Central steps of isoprenoids biosynthesis						
TR59462 c0_g2_i1	2.5.1.1	Geranyl pyrophosphate synthase	Geranyl pyrophosphate synthase (GDPS)	Isoprenoid biosynthesis	geranyl diphosphate biosynthesis	geranyl diphosphate from dimethylallyl diphosphate and isopentenyl diphosphate: step 1/1
TR38140 c0_g1_i1	2.5.1.10	Farnesyl pyrophosphate synthase	Farnesyl pyrophosphate synthase (FPPS)	Isoprenoid biosynthesis	farnesyl diphosphate biosynthesis	farnesyl diphosphate from geranyl diphosphate and isopentenyl diphosphate: step 1/1
TR9895 c0_g1_i1	2.5.1.10	Farnesyl pyrophosphate synthase 1	Farnesyl pyrophosphate synthase (FPPS)	Isoprenoid biosynthesis	farnesyl diphosphate biosynthesis	farnesyl diphosphate from geranyl diphosphate and isopentenyl diphosphate: step 1/1
Sterol biosynthesis						
TR34822 c0_g1_i1	2.5.1.21	Squalene synthase	Squalene synthase (SQS)	Terpene metabolism	lanosterol biosynthesis	lanosterol from farnesyl diphosphate: step 1/3
TR39301 c2_g2_i2	5.4.99.8	Cycloartenol synthase	Cycloartenol synthase (CAS)	-	-	-
TR39301 c2_g2_i3	5.4.99.7	Lanosterol synthase	Lanosterol synthase (LAS)	Terpene metabolism	lanosterol biosynthesis	lanosterol from farnesyl diphosphate: step 3/3
TR16425 c0_g1_i1	2.1.1.41	Sterol 24-C-methyltransferase erg6	Sterol 24-C-methyl-transferase (SMT1_A)	Steroid metabolism	ergosterol biosynthesis	ergosterol from zymosterol: step 1/5
TR26964 c0_g1_i1	2.1.1.41	Sterol 24-C-methyltransferase erg6	Sterol 24-C-methyl-transferase (SMT1)	Steroid metabolism	ergosterol biosynthesis	ergosterol from zymosterol: step 1/5
TR28292 c0_g1_i1	2.1.1.41	Sterol 24-C-methyltransferase erg6	Sterol 24-C-methyl-transferase (SMT1_B)	Steroid metabolism	ergosterol biosynthesis	ergosterol from zymosterol: step 1/5

TR9224 c0_g1_i1	2.1.1.41	Cycloartenol-C-24-methyltransferase	Sterol 24-C-methyl-transferase (SMT1_C)	Steroid biosynthesis	sterol biosynthesis	-
TR21471 c0_g1_i1	2.1.1.41	Cycloartenol-C-24-methyltransferase	Sterol 24-C-methyl-transferase (SMT1_D)	Steroid biosynthesis	sterol biosynthesis	-
TR5301 c0_g4_i1	1.14.18.9	Methylsterol monoxigenase 1	Methylsterol monoxigenase (MSMO)	Steroid biosynthesis	zymosterol biosynthesis	zymosterol from lanosterol: step 3/6
TR5301 c0_g4_i2	1.14.18.9	Methylsterol monoxigenase 1	Methylsterol monoxigenase (MSMO)	Steroid biosynthesis	zymosterol biosynthesis	zymosterol from lanosterol: step 3/6
TR18756 c0_g1_i1	1.1.1.270	3-keto-steroid reductase	3-beta-hydroxysteroid 3-dehydrogenase (3-KSR)	Steroid biosynthesis	zymosterol biosynthesis	zymosterol from lanosterol: step 5/6
TR1055 c0_g1_i1	5.5.1.9	Cycloeucalenol cycloisomerase	Cycloeucalenol cycloisomerase (CPI1)	-	-	-
TR1055 c0_g1_i2	5.5.1.9	Cycloeucalenol cycloisomerase	Cycloeucalenol cycloisomerase (CPI1)	-	-	-
TR9931 c0_g1_i1	1.14.14.154	Sterol 14-demethylase	Sterol 14-demethylase (CYP51G1)	-	-	-
TR9931 c0_g1_i2	1.14.14.154	Sterol 14-demethylase	Sterol 14-demethylase (CYP51G1)	-	-	-
TR9931 c0_g1_i3	1.14.14.154	Sterol 14-demethylase	Sterol 14-demethylase (CYP51G1)	-	-	-
TR31276 c1_g1_i2	1.3.1.70	Delta(14)-sterol reductase	Delta(14)-sterol reductase (TM7SF2)	Steroid biosynthesis	zymosterol biosynthesis	zymosterol from lanosterol: step 2/6
TR84 c1_g1_i3	1.1.1.170	Sterol-4-alpha-carboxylate 3-dehydrogenase decarboxylating	3-beta-hydroxysteroid-4-alpha-carboxylate 3-dehydrogenase (decarboxylating) (NSDHL)	Steroid biosynthesis	zymosterol biosynthesis	zymosterol from lanosterol: step 4/6
TR2119 c0_g3_i2	1.14.19.20	Delta(7)-sterol-C5(6)-desaturase	Delta(7)-sterol 5(6)-desaturase (SC5DL)	-	-	-
TR49413 c0_g1_i1	1.3.1.21	7-dehydrocholesterol reductase	7-dehydrocholesterol reductase (DHCR7)	Lipid metabolism	steroid biosynthesis	-
TR19296 c0_g1_i1	1.3.1.71	Delta(24(24(1)))-sterol reductase	Delta(24(24(1)))-sterol reductase (ERG4)	Steroid biosynthesis	ergosterol biosynthesis	ergosterol from zymosterol: step 5/5

The identity of all the transcripts retrieved from the *in silico* analysis and reported in Table 3.3 have been confirmed through their blast in the NCBI database and domains analysis. Among these transcripts, some needed further steps of confirmation, as described below.

No SQE sequences have been found annotated in *T. rotula* transcriptome, in accordance with what already described for other diatoms species[10]. Indeed, in diatoms, the first and rate limiting oxygenation reaction of the sterol biosynthetic pathway is catalysed by another enzyme, named AltSQE[10]. The search of AltSQE in *T. rotula* transcriptome have been performed through the blast of all the known diatoms AltSQE sequences reported in Pollier *et al.*[10]. We retrieved as results two transcripts, 'TR2119|c0_g3_i2' and 'TR48825|c0_g1_i1', matching all the queries with high bit scores and low e-values. The clusterization analysis of these two transcripts with the online tool CD-HIT, confirmed that they are two different genes. The first (TR2119|c0_g3_i2) was annotated in *T. rotula* transcriptome as delta(7)-sterol-C5(6)-desaturase (EC 1.14.19.20, in Table 3.3), an enzyme belonging to the fatty acids hydroxylase superfamily (that is the same superfamily to which AltSQE belongs), and was included in the genes expression analysis reported below. On the contrary, the second transcript retrieved (TR48825|c0_g1_i1) wasn't associated to any functional annotation. Further analysis is needed to assess whether this last one could be the *T. rotula* AltSQE and for this reason its gene expression has not been evaluated in this study.

In order to verify the possible function of the transcripts 'TR39301 c2_g2_i2' and 'TR39301 c2_g2_i3' annotated as 'Cycloartenol synthase' and 'Lanosterol synthase' respectively (Table 3.3), we firstly compared their nucleotide sequences with CD-HIT obtaining as results only one cluster, with the first transcript being the referring protein. The domains analysis finally confirmed their identity as 'Cycloartenol synthase'. In addition, a further search in *T. rotula* transcriptome retrieved no others Lanosterol synthase annotated transcripts, confirming the absence of this branch of the pathway.

The *in silico* analysis performed allowed to identify other transcripts that are included in the sterol biosynthetic pathway. In particular, five different transcripts were found annotated as 'Sterol-24-C-methyl transferase' or 'Cycloartenol-C-24-methyltransferase' (EC 2.1.1.41 in Table 3.3), both corresponding to the isoform 1 of the plant enzyme (SMT1), catalysing the conversion of cycloartenol into 24-methylene cycloartenol[21]. On the contrary, no enzymes have been found annotated as the plant isoform 2 (SMT2, EC 2.1.1.143) that catalyse the formation of 24-ethylenelophenol from 24-methylenelophenol[5].

Finally, we also found a transcript annotated as '3-beta-hydroxysteroid 3-dehydroganase' synonym of '3-keto steroid reductase' (EC 1.1.1.270, in Table 3.3), an enzyme that participates in the biosynthesis of cholesterol from lanosterol[22]. However, this enzyme has not been detected in the cholesterol biosynthetic pathway of other diatoms[1,11] and its presence in our diatom species needs further investigation. For these reasons, it was not included in the genes expression analysis described below.

3.3.2.1 qPCR gene expression analysis

For the genes expression analysis, we focused our attention on the key genes involved in the sterols biosynthetic pathway and identified in *T. rotula* transcriptome. Their differential expressions have been evaluated in cultures grown in nitrogen (NDepl), silica (SiDepl) and phosphate (PDepl) limited media. *T. rotula* cultures grown in standard medium were used as control (CTR) for RT-qPCR analysis. Figure 3.4 shows the expression ratio of the identified genes, as fold change, in each of the stressing growing condition applied to the algae with respect to the CTR. In the NDepl condition, a general down-regulation trend is observable (except for SMT_B-C) with a significant variation only the SMT_D enzyme (p-value=0.001) (Cycloartenol-C-24-methyltransferase, transcript TR21471 c0_g1_i2). The SiDepl condition showed the same trend of gene expression observed in the NDepl (except for SMT_A-C), with a significant down-regulation of SMT_D (p-value=0.001) and of DHCR7 (7-dehydrocholesterol reductase) (p-value=0.002) with respect to the CTR. On the contrary, in PDepl the general trend of the tested genes is an up-regulation, except for SQS (squalene synthase), SMT_A and SMT_C, with only DHCR7 being significantly up-regulated (p-value=0.048) (Figure 3.4). Interestingly, SQS, that is the limiting enzyme for the sterols biosynthesis[23], is down-regulated, even if not significantly, in all the conditions tested (Figure 3.4).

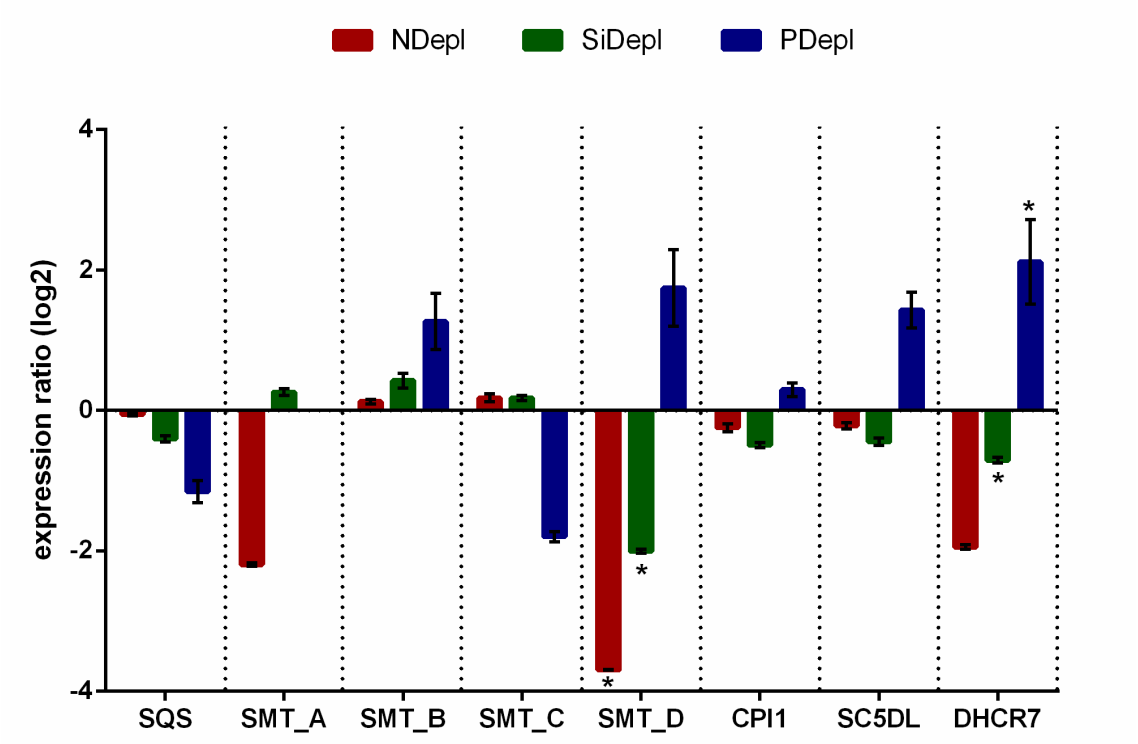


Figure 3.4. Expression levels (y-axis, Mean ± SD) of the key enzymes in *T. rotula* sterols biosynthetic pathway. Cultures were grown in limited media and named as follow: NDepl = nitrogen starved cultures (red bars), SiDepl= silica starved cultures (green bars) and PDepl= phosphate starved cultures (blue bars). Cells cultured in normal medium were used as control condition. **SQS**= squalene synthase; **SMT_A**= Sterol-24-C-methyltransferase erg6, transcript TR16425|c0_g1_i1; **SMT_B**= Sterol-24-C-methyltransferase erg6, transcript TR28292|c0_g1_i1; **SMT_C**= Cycloartenol-C-24-methyltransferase, transcript TR9224|c0_g1_i2; **SMT_D**= Cycloartenol-C-24-methyltransferase, transcript TR21471 c0_g1_i2; **CPI1**= Cycloleucalenol cycloisomerase; **SC5DL=** Delta(7)-sterol 5(6)-desaturase; **DHCR7** = 7-dehydrocholesterol reductase; Asterisks indicate statistical significance of the genes expression levels in each nutritional stress compared to the control according to the Pair Wise Fixed Reallocation Randomisation Test implemented in REST.

3.3.2.2 Qualitative and quantitative analysis of *T. rotula* sterols

The phytosterols pool in *T. rotula*, grown in the different experimental conditions tested, was evaluated at a qualitative and quantitative level by GC-MS analysis. The resulting chromatograms highlighted the production of 24-methylene cholesterol (24-MC) as major sterol in all the conditions tested (Figure 3.5, peak 2) along with a minor (<5 %) sterol putatively identified as 24-methylcholest-5-en-3β-ol (campesterol) not considered in the quantification analysis (red arrows in Figure 3.5).

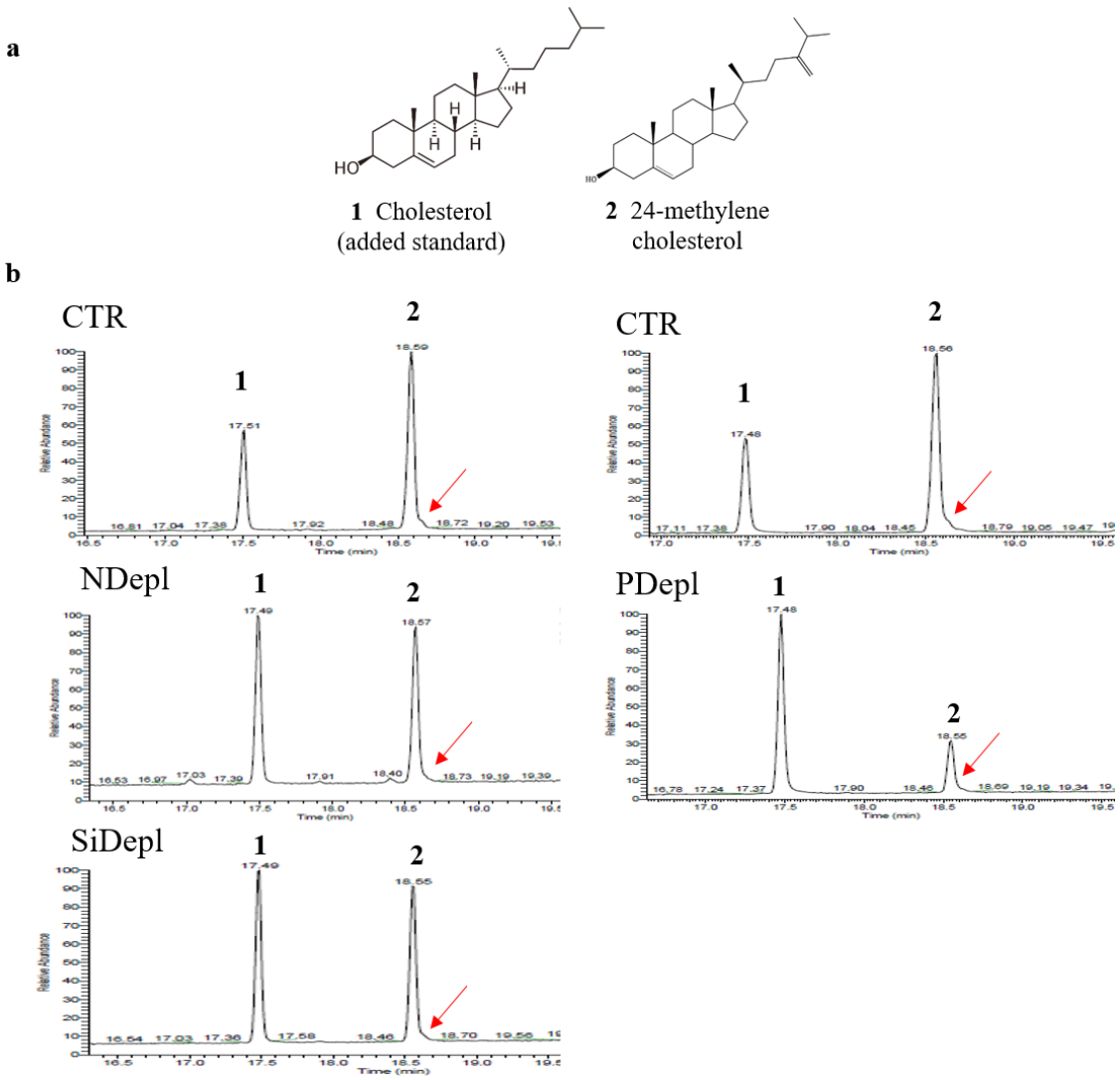


Figure 3.5. *T. rotula* sterols profile. **a.** molecular formula of cholesterol (added as internal standard) and 24-methylene cholesterol (24-MC), the major sterol produced by *T. rotula*; **b.** GC-MS chromatograms of *T. rotula* cultures grown in different nutritional starvation conditions (CTR=control, NDepl= nitrogen starved cultures, SiDepl= silica starved cultures, PDepl= phosphate starved cultures). **1**= cholesterol peak; **2**= 24-MC peak. The red arrows indicate the minor sterol detected along the analysis.

Figure 3.6 shows the GC-MS quantization of the main *T. rotula* sterol, i.e. 24-MC, in all the experimental growth conditions tested. As reported in the figure, in Si-starv the quantity of 24-MC was comparable to the control (Figure 3.6 a-c), whilst both in NDepl and PDepl there was a significant reduction of 24-MC quantity in comparison to the control when normalized per cell (Figure 3.6 a, p-values of 0.002 and 0.03, respectively) and per dry weight (Figure 3.6 b, p-values of 0.0004 and 0.02, respectively). A decreasing trend, not significant due to the high error bars, is evident when 24-MC absolute quantity is normalized on the mg of lipid extract (Figure 3.6 c). As shown in Figure 3.6 d, none of the treatments tested impacted on *T. rotula* total lipid metabolism, since in NDepl and

PDepl the change in the total lipids quantity was not significant, while in SiDepl the quantity was equal to the CTR.

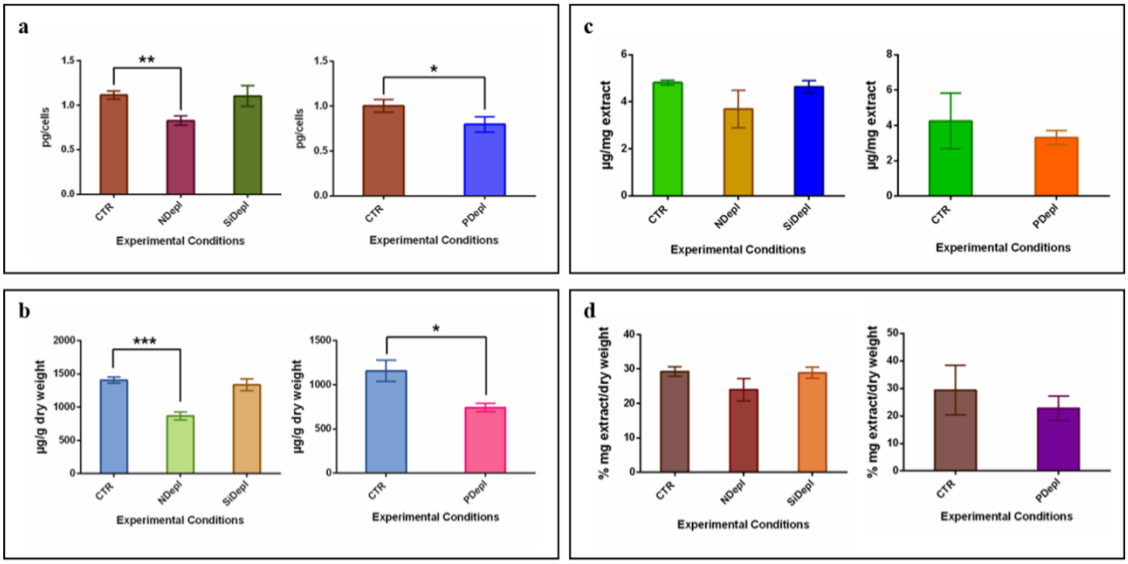


Figure 3.6. 24-MC quantification with GC-MS. The quantity was expressed as: **a.** pg of 24-MC per cell (pg/cell); **b.** µg of 24-MC per g of pellet's dry weight (µg/g dry weight); **c.** µg of 24-MC per mg of lipid extract (µg/mg lipid extract). **d.** % of mg of extract per mg of dry weight (% mg extract/dry weight), that shows the % of total lipids in relation to the biomass. Asterisks indicates samples with statistical significant differences between treatment and control according to unpaired t-test. *: $p<0.05$, **: $p<0.005$, ***: $p<0.0005$.

3.3.2.3 Proposed structure of *T. rotula* sterols biosynthetic pathway

On the basis of the results obtained from the *T. rotula* transcriptome *in silico* analysis and on the GC-MS analysis of the end-products, together with the already available information in the literature[1,5,11], the *T. rotula* sterol biosynthetic pathway have been reconstructed as shown in Figure 3.7.

Differently from what is reported for the diatoms which pathway have been already studied[1,5,11], all the enzymes involved in the cholesterol biosynthesis have not been detected in *T. rotula* transcriptome (Figure 3.7).

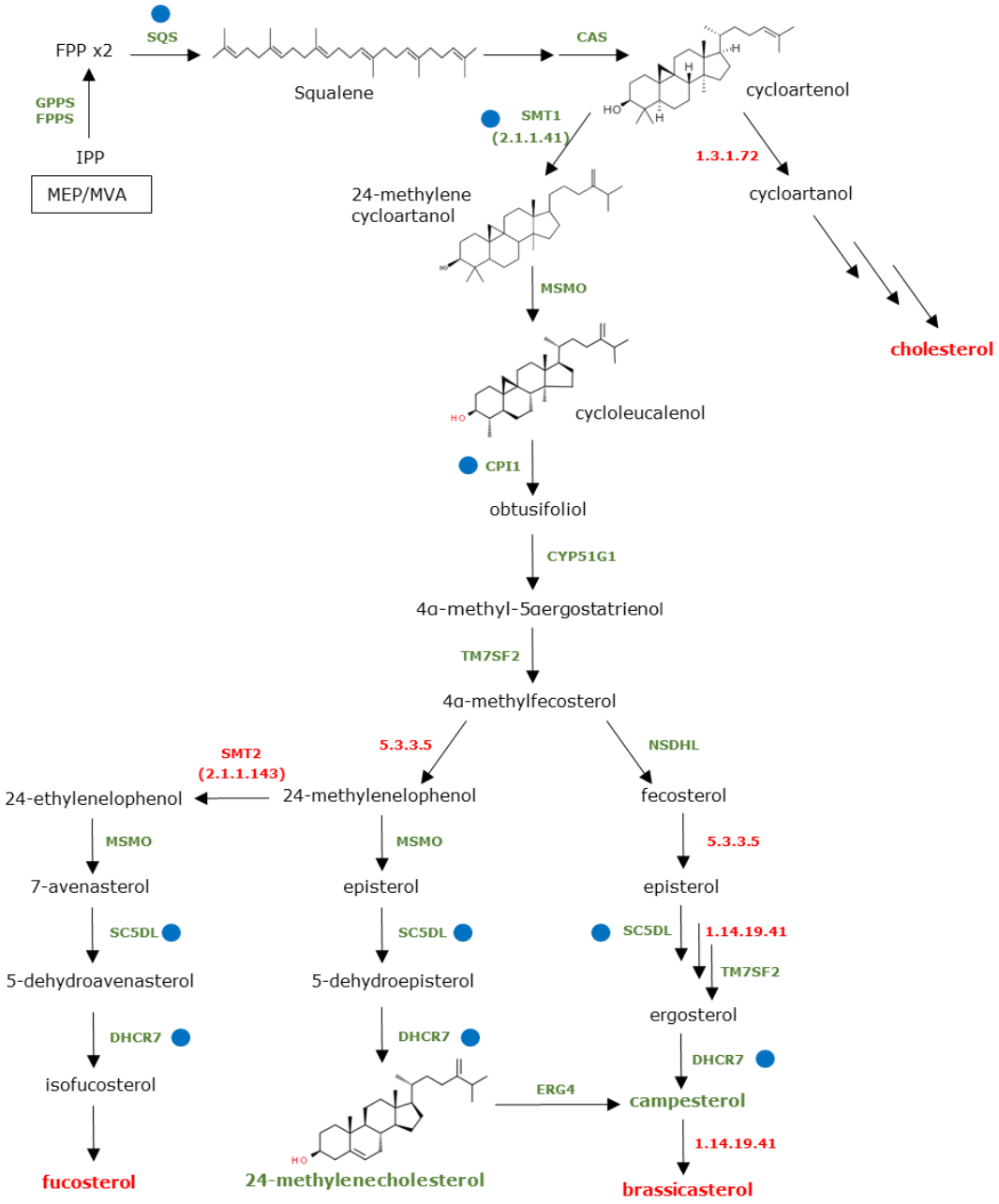


Figure 3.7. Proposed sterols biosynthetic pathway in *T. rotula*, based on *S. costatum*, *C. criptica*[5], *P. tricornutum*, *T. pseudonana* and *C. muelleri*[11] data.

Legend: green names: Enzymes identified in *T. rotula*; red EC numbers: Enzymes not detected in *T. rotula*; blue circles: enzymes which expression have been evaluated through qPCR in the different experimental conditions tested; green end-products: products detected in the *T. rotula* through GC-MS analysis of its sterol profile; red end-products: products not detected in *T. rotula* during the GC-MS sterol profile analysis. Abbreviations: IPP=isopentenyl-pyrophosphate; FPP=farnesyl-pyrophosphate.

3.3.3 The Iridoid pathway in *T. rotula*

Besides sterols, also carotenoids, Monoterpenoid indole alkaloids (MIAs) and other terpenoids originate from the IPP derived especially from the MEP pathway. In Chapter

1 we reported the identification, for the first time, of secologanin synthase (SLS) in *T. rotula* as significantly up-regulated in SiDepl with respect to CTR. For these reasons, we searched all the Iridoid pathway enzymes, to which SLS belongs, in order to evaluate, *in silico,* if the SiDepl condition could shift the lipid metabolism toward the MIAs synthesis instead of sterols. This search did not retrieve the complete set of necessary enzymes. Indeed, geraniol synthase (GES), iridoid synthase (IS) and 7-deoxyloganetic acid synthase (7-DLGT) were not found. On the contrary, we identified the enzymes 'Isopentenyl-diphosphate Delta-isomerase' (IDI), 'Geranyl pyrophosphate synthase' (GDPS), 'Cytochrome P450-dependent monooxygenase' (G10H), 'Loganic acid methyltransferase' (LAMT), 'Secologanin synthase' (SLS), 'tryptophan decarboxylas' (TDC) and 'Strictosidine synthase' (STR) (Figure 3.8).

Figure 3.8 shows the different expression, reported as RPKM (Supplementary Table 3.1 c), of the Iridoid pathway enzymes between the two-growth conditions compared with the transcriptome analysis. Unless for IDI and G10H, all the enzymes showed an up-regulation trend in SiD but, by multiple t-test analysis, only 7-DLGT and SLS were significantly up-regulated (p-values of 0.03 and 0.006, respectively).

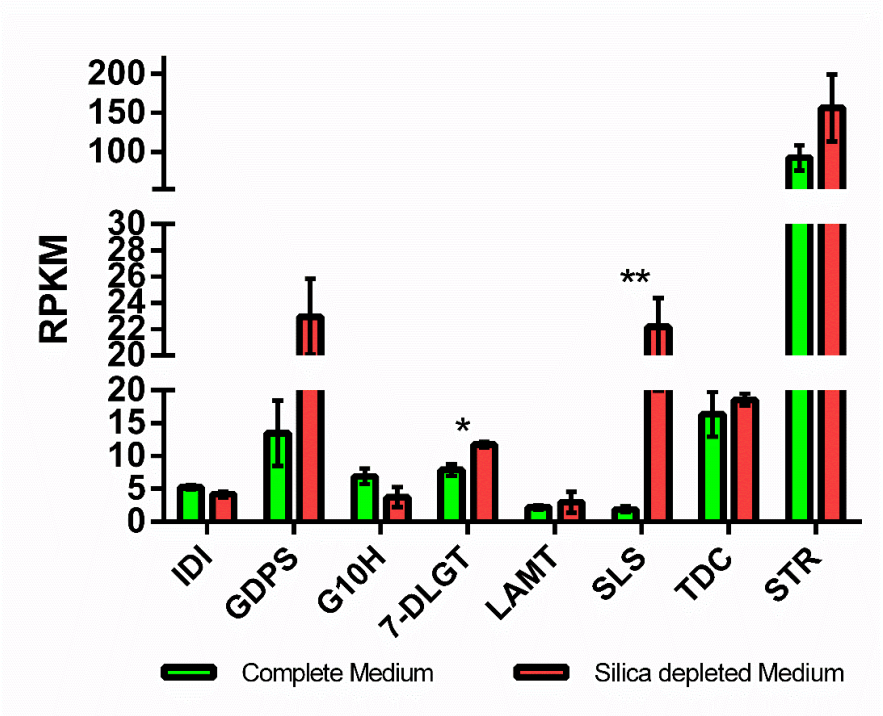


Figure 3.8. Iridoid pathway in *T. rotula*. Expression, calculated in RPKM, of the enzymes in the Iridoid pathway identified in *T. rotula* transcriptome from cells grown in complete medium (green bars) compared to silica depleted medium (red bars). Significance was calculated by performing a multiple t-test, each row individually without correction for multiple comparisons.

3.4 Discussion

The results obtained with this preliminary study, allowed to propose the structure of the phytosterol biosynthetic pathway of the centric diatom species *T. rotula*, collected in the Gulf of Naples. The main enzymes involved were identified through *in silico* transcriptome analysis of *T. rotula* cultures challenged with silica limitation (SiDepl) in comparison to a control (CTR) condition (details regarding the transcriptome sequencing are reported in Chapter 1 of this thesis). The first step in the biosynthesis of phytosterols is the production of Isopentenyl pyrophosphate (IPP) through the MVA or the MEP pathways[5,7]. Our results, clearly highlighted the presence of the complete sets of enzymes belonging to both routes in this diatom, as was previously confirmed *in silico* for other *Thalassiosira* species[8]. However, it is known that in organisms that harbour both pathways, such as plants, the MVA provides the precursors for sterol biosynthesis[11], while the MEP generates carotenoids, MIAs and other terpenoids[22]. The differential expression of the MVA pathway's enzymes, evaluated in RPKM, evidenced the down-regulation, in SiDepl compared to CTR, of the HMGR enzyme, catalyzing the rate-limiting conversion of β-Hydroxy β-methylglutaryl-CoA (HMG-CoA) into mevalonate[20], even if the difference was not significant due to the high standard deviation. Due to its involvement as key enzyme in both sterols and isoprenoids biosynthesis, in mammals this enzyme is highly regulated at the transcriptional, translational and post-translational levels[25], while not much is known about its regulation in diatoms[7]. We also reported that the MCT and MECS enzymes, belonging to the MEP pathway, were significantly up-regulated in SiDepl with respect to the CTR. These findings, even if need experimental confirmation, suggested that in SiDepl the precursors "flux" seems to go through the MEP pathway instead of MVA. In Chapter 1 we reported the up-regulation, in SiDepl compared to CTR, of SLS, the key enzyme in the Iridoid pathway. In this chapter of the thesis, we identified also other enzymes of the Iridoid pathway and evaluated their *in silico* differential expression, highlighting the general up-regulation of the pathway in SiDepl. Altogether, these results allowed to hypothesize that SiDepl is not the preferable condition to stimulate sterol pathway expression while it activates the MEP and the Iridoid pathway, resulting to be an intriguing condition in which evaluate MIAs production.

With a similar *in silico* approach, we also reported the identification of the main enzymes involved in the biosynthesis of phytosterols. We were not able to confirm unambiguously the presence of the AltSQE enzyme in *T. rotula*. Its absence in the annotation table can

be due to the fact that the annotation of our transcriptome precedes the first characterization of this enzyme performed by Pollier et al[10]. Moreover, the high similarity (above 60 %) between delta(7)-sterol-C5(6)-desaturase and AltSQE, already reported for other diatom species[5], make necessary the use, as future perspective for this work, of chemical inhibitors[11] or labelled precursors[5] to confirm the identity of the transcript without functional annotation that we retrieved as result of our analysis.

We confirmed that the pathway goes through 2,3 epoxysqualene cyclization by the cycloartenol synthase while we did not detected the lanosterol synthase, in accordance with literature data on diatoms[5,11], further supporting the theory that the cycloartenol-based sterol biosynthesis is predominant in these organisms[11]. All the enzymes detected in the diatom species in which the pathway has been studied through transcriptome analysis, i.e. *S. costatum*, *C. criptica*[5], *P. tricornutum*, *T. pseudonana* and *C. muelleri*[11], were also identified in *T. rotula*. However, no functional annotation was found in the *T. rotula* transcriptome for the enzyme 'cholestenol Delta-isomerase' (EC 5.3.3.5, in red in Figure 3.7) that is responsible for the conversion of 4α-methylfecosterol in 24-methylenelophenol and of fecosterol into episterol in the *P. tricornutum*, *T. pseudonana* and *C. muelleri* pathways proposed by Jaramillo-Madrid *et al.*,[11]. Nevertheless, evaluating *T. rotula* sterol profile through GC-MS, we identified 24-MC as major sterol, confirming previous studies on this diatom species[15,26], and campesterol as minor sterol. Both these compounds derive from the branch of the pathway that starts from cholestenol Delta-isomerase activity[11], thus indirectly confirming its presence in *T. rotula* and also demonstrating that only this branch of the pathway is active. Moreover, while five different SMT1 were found annotated in this species, no SMT2, that catalyze the conversion of 24-methylenelophenol into 24-ethylenelophenol[11], was found, suggesting the inactivity of the branch of the pathway leading to fucosterol production[11] (Figure 3.7), also confirmed by its absence in *T. rotula* sterol profile. Another inactive branch in *T. rotula* is the one leading to cholesterol biosynthesis, since nor the enzymes and the end-product were detected through both *in silico* and GC-MS analysis.

Noteworthy, there is a general lack of information that relates culture level parameters to biochemical and genetic changes within cells and this represents a limiting factor in the improving of phytosterol production[17]. In addition, a wide variability in the effect of culture conditions on phytosterol production in microalgae is reported in literature[17]. Regarding our model diatom, i.e. *T. rotula*, no studies have been conducted before on the effect of nutrients variations on sterol production.

We evaluated the gene expression of the key enzymes of the pathway in cultures challenged with different nutritional stresses, i.e. nitrogen, silica and phosphate limitation, and in a standard condition (CTR) constituted by cultures grown in f/2 media[27], that represents an enrichment of nutrients in comparison to the natural environment. Moreover, we harvested the cells in stationary phase, that is reported to be the growth phase in which an higher sterol content is present in diatoms[17].

We found a very low genes expression in almost all the conditions tested, being the threshold cycle (Ct) of each gene above or equal to 30 with the exception of DHCR7 that has a Ct mean of ~28 in all the conditions. A general down-regulation trend is evident in NDepl and SiDepl and up-regulation trend in PDepl. Although gene expression is not indicative of the production of the end-product, we noticed that SQS, the first enzyme of the pathway committed specifically to sterols biosynthesis[23], had a down-regulation trend, even if not significantly, in all the stress conditions tested with respect to the CTR.

We also evaluated the amount of 24-MC in *T. rotula* grown in the selected stress conditions, revealing the decrease in its production in NDepl and PDepl. In addition, the % of total lipids produced, normalized on the dry weight of the pellet, did not significantly vary in all the condition tested, suggesting that the decrease of sterol content in nitrogen and phosphate limitation is compensated by the production of other classes of lipids. Overall, the CTR condition was the one in which the sterol production was higher, confirming that nutrient supplement is necessary for sterol production.

Further deepening investigation is necessary to understand whether a post-transcriptional regulation occurs when the pathway activity does not correlate with the molecule production.

Altogether, our results led to a better understanding of the putative structure of the sterol biosynthetic pathway in *T. rotula*. This improved understanding of the pathway, together with the genome and transcriptome sequencing we performed for this species (Chapter 1), could be helpful in the application of molecular approaches, such as RNAi or CRISPR/Cas methods[17], to perturb the pathway allowing the enhancement of the sterol content in this diatom as future perspective of this study. Unfortunately, the stress conditions preliminarily tested in this work were not able to increase sterol production in our model species, thus the effect of different culturing parameters should be tested in future studies, as temperature and light.

The interest in the identification of stress conditions able to stimulate 24-MC production in this diatom species is due to the fact that it is the necessary substrate of sterols with

anti-cancer activities, but its total chemical synthesis is difficult and costly[28]. Moreover, despite no information is reported in literature regarding a putative biotechnological application of this compound, a study in progress in our group is revealing a cytotoxic activity of 24-MC toward breast and lung cancer cells lines (Personal communication).

3.5 Materials and Methods

3.5.1 Strain culturing

Thalassiosira rotula, strain CCMP1647, lab-name FE80, was isolated in 2011 in the Gulf of Naples (40°48.5′N, 14°15′E), Mediterranean Sea. Clonal cultures were established by isolating single cells from phytoplankton net samples collected from the surface layer of the water column. Cultures were grown in sterile filtered oligotrophic seawater with a salinity of 36 ppt amended with f/2 nutrients[27] at a temperature of 18 °C, at 12:12 h light:dark cycle, with a photon flux of 100 μmol photons m^{-2} s^{-1}.

2 L cultures containing 5000 cells/mL were grown in complete f/2 (CTR), in silica depleted f/2 media with 36 μM Na_2SiO_3 (SiDepl), in nitrogen depleted f/2 media with 30 mM $NaNO_3$ (NDepl) and phosphate depleted f/2 media with 0.5 μM NaH_2PO_4 (PDepl)[29,30], all under agitation through filtered air influx. Since phosphate starved cultures have not been set up contemporary to the other nutritional starvations, specific CTR cultures have been simultaneously cultured to compare gene expression and sterols quantification. All the experimental conditions tested were set up in biological triplicates. 1 mL of each culture was fixed with one drop of Lugol (final concentration of about 2 %) and counted in a gridded Sedgewick-Rafter counting cell under an Axioskop 2 microscope (20 ×) (Carl Zeiss GmbH, Jena, Germany). The growth curves obtained for each experimental condition are reported in Supplementary Figure 3.1. Cultures were harvested at their stationary growth phases (day 6 for CTR and NDepl conditions and day 5 for SiDepl and PDepl conditions). 100 mL of each culture and triplicates have been used for RNA extraction and gene expression analysis, while 1.5 L were collected for sterols quantification. All the samples were obtained by centrifugation at 3200 rpm. for 10 min at 4 °C (Eppendorf Centrifuge 5810 R, Eppendorf, Amburg, Germany). The 100 mL pellets were covered with 1 mL TRIsure (BIO-38033, Bioline Reagents Ltd., London, UK) and vortexed. All the pellets were immediately frozen in liquid nitrogen and stored at -80°C until use.

3.5.2 Total lipids extraction and fractionation

Cells pellets from 1.5 L of the cultures were lyophilized over-night and 1/3 of each dry pellet was used for lipid extraction. Only for the PDepl condition all the dry pellet was used and 1/3 of the resulting 24-MC absolute quantity was normalized on cells and dry weight corresponding to 500 mL of culture (see “GC-MS analysis” for the determination

of the absolute quantity). 60 µg of cholesterol (Sigma-Aldrich, St. Louis, Missouri, USA) were added as internal standard to the lyophilized pellet. The total lipids were extracted in methyl tert-butyl ether (MTBE) and the fatty acids were methylated with diazomethane (CH_2N_2).

The total lipids extracts were resuspended in petroleum ether 100 % and loaded onto pre-packed CHROMABOND® SiOH cartridges (500 mg/6 mL) previously activated with petroleum ether 100 % (9 mL). The chromatographic elution was accomplished with the following solvent gradient:

- 9 mL of petroleum ether (100 %);
- 18 mL of petroleum ether/diethyl ether (95:5);
- 18 mL of petroleum ether/diethyl ether (9:1);
- 18 mL of petroleum ether/diethyl ether (85:15);
- 18 mL of petroleum ether/diethyl ether (8:2);
- 9 mL of petroleum ether/diethyl ether (6:4);
- 9 mL of petroleum ether/diethyl ether (2:8);
- 9 mL of dichloromethane/methanol (9:1);

Fractions of 9 mL volume were collected, pooled according to their chromatographic (TLC) profile (reported in Supplementary Figure 3.2) and dried under vacuum. Fractions containing sterols were silylated with *N*,*O*-Bis(trimethylsilyl) trifluoroacetamide (BSTFA) (CAS Number 25561-30-2, Sigma-Aldrich, St. Louis, Missouri, USA) immediately before GC-MS analysis.

3.5.3 GC-MS analysis and quantitative determination of 24-MC

GC-MS analyses were performed on an ion trap mass spectrometer equipped with EI source (70 eV) and autosampler (Polaris Q; ThermoScientific, Waltham, Massachusetts, USA) coupled with a gas chromatographic system (GCQ; ThermoScientific, Waltham, Massachusetts, USA) with a 5 % phenyl-methyl polysiloxane column (VF-5ms, Agilent, 30 m × 0.25 mm × 0.25 µm) using high-purity helium as the gas carrier. Elution of silyl sterols required a temperature program starting at 180 °C for 2 min, followed by a 10 °C min^{-1} gradient up to 280 °C, then 4 °C min^{-1} up to 320 °C, holding for 6 min. Microalgal samples were dissolved in dichloromethane at a final internal standard (IS) concentration of 60 µg/ml and directly injected in split (1:10) mode. Inlet temperature was 270 °C,

transfer line was set at 280 °C, and ion source temperature was 250 °C. Mass spectra were acquired in the range 50-700 *m/z*.

Absolute amount of 24-MC was determined by using an external calibration curve (Supplementary Figure 3.3) of a pure commercial standard of 24-MC in the range (150 - 2.5 µg/ml) and by using cholesterol as IS at a concentration of 60 µg/ml. Analyte peak areas at the 5 calibration points were integrated and normalized by IS area, affording a linear calibration curve with R^2= 0.998 and equation y= 0.0203x – 0.0421. The chromatograms obtained for each replicate in all the experimental conditions tested are reported in Supplementary Figure 3.4.

3.5.4 Bioinformatics identification of the pathway

For the search of the enzymes of the pathway, the *T. rotula* transcriptome and genome FASTA files were used to create databases with the standalone setup BLAST+ v 2.11.0 (https://www.ncbi.nlm.nih.gov/books/NBK279690/). Default parameters were used to perform tBLASTn, BLASTn and BLASTp searches of different sequences in the *T. rotula* databases. The identity of all the enzymes was confirmed both with a search in NCBI and InterPro Scan (https://www.ebi.ac.uk/interpro/) databases, where the HMMER (Web version 2.41.2) search tool was used, with default parameters, to refine the analysis through the study of protein's domain organization.

For transcripts present in multiple copies, redundant sequences were identified with CD-HIT-EST using the default parameters (http://weizhong-lab.ucsd.edu/cdhit-web-server/cgi-bin/index.cgi?cmd=cd-hit-est). Alignments have been performed with the NCBI blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with default parameters.

3.5.5 RNA extraction and cleanup

To proceed to total RNA extraction, glass beads (G1277, Sigma-Aldrich, St. Louis, Missouri, USA) were added to the 100 mL pellets with TRIsure. Cells were disrupted on a thermo-shaker (Eppendorf ThermoMixer C, Eppendorf, Amburg, Germany) at 60 °C for 10 min at 1200 rpm. Cells debris and glass beads were discarded, and the extraction was continued according to the TRIsure instructions. DNase treatment was carried out using DNase I recombinant, RNase-free (Roche, Basel, Switzerland) according to manufacturer's protocol to eliminate potential genomic DNA contamination. Total RNAs were purified and concentrated using RNeasy MinElute Cleanup Kit (Cat. No./ID: 74204,

Qiagen, Hilden, Germany) and eluted in 30 µL of RNase-free water. Concentrations of the resulting RNA samples were assessed by absorbance at 260 nm (ND-1000 Spectrophotometer; NanoDrop Technologies, Wilmington, Delaware, USA). The integrity of total RNA was checked by agarose gel electrophoresis (1 % agarose w/v).

3.5.6 Reverse transcription and evaluation of gDNA contamination

1 µg of total RNA for CTR, SiDepl and NDepl and 0.5 µg of PDepl and its CTR were retro-transcribed in complementary DNA (cDNA) following the manufacturer's instructions (5X All-In_One RT Master Mix, abm, Applied Biological Materials Inc., Richmond, Canada) using the T100 Thermal cycler (Bio-Rad Laboratories, Hercules, California, USA). This kit includes a gDNA removing step. Retro-transcribed samples were checked for DNA contamination by traditional PCR using intron-spanning primer pairs designed on the housekeeping gene COP-A (Table 3.4, Supplementary Figure 3.5). PCR reactions were carried out in 25 µL volume with 2.5 µL of 10× PCR reaction buffer (Roche, Basel, Switzerland), 2.5 µL of 2 mM dNTP, 0.3 µL of 5 U/µL Taq (Roche, Basel, Switzerland), 1 µLof 20 pmol/µL for each primer, 1 µL of RNA templates (2 µL of PDepl and its CTR RNAs) and nuclease-free water up to 25 µL. The PCR program consisted in a denaturation step at 95 °C for 5 min, 35 cycles at 95 °C for 30 s, 60 °C for 30 s and 72 °C for 45 s, and a final extension step at 72 °C for 7 min.

Table 3.4: Primer used for DNA contamination check.

Primer name		Primer sequence	Tm (°C)	cDNA amplicon size	gDNA amplicon size
COP-A	**F**	GTGTTCATTGCGTGCCATCT	66.5	120	206
	R	GCTAGTTTTGCCAGGCGTCT	65.8		

Amplified PCR products were analysed by 1.5 % agarose gel electrophoresis and the resulting bands were excised from the gel and extracted according to the GenElute Gel Extraction Kit protocol (Sigma-Aldrich, St. Louis, Missouri, USA). Sequences were obtained by BigDye Terminator Cycle Sequencing Technology (Applied Biosystems, Waltham, Massachusetts, USA) and purified using the Agencourt CleanSEQ Dye terminator removal Kit (Agencourt Bioscience Corporation, Beverly, Massachusetts, USA) in automated robotic station Biomek FX (Beckman Coulter, Pasadena, California,

USA). Products were analysed on the Automated Capillary Electrophoresis Sequencer 3730 DNA Analyser (Applied Biosystems, Waltham, Massachusetts, USA). Obtained sequences were analysed with Chromas v 2.6.6 (http://technelysium.com.au/wp/chromas). Identity of the retrieved sequences was verified by blasting the amplicon sequences in *T. rotula* transcriptome with the standalone setup BLAST+ v 2.11.0.

3.5.7 Best reference analysis

To study the best genes to be used for data normalization in RT-qPCR experiments, a set of six genes already identified as internal controls for *T. rotula* gene expression analysis in Di Dato *et al.*, 2019[31] (and reported in "APPENDIX A – Supplementary Information Chapter 1") were selected: Actin, TBP, COP-A, RPS, EIF1α, H4. The tool RefFinder (https://www.heartcure.com.au/reffinder/), that integrates the computational programs BestKeeper, NormFinder, geNorm and Delta-Ct-method, was used to identify the best reference genes in all our experimental conditions (Table 3.5).

Table 3.5: List of the analysed reference genes reporting the best reference according to the Ranking Order (Better--Good--Average).

method	1	2	3	4	5	6
Delta CT	Actin	TBP	RPS	H4	COP-A	EIF1α
BestKeeper	RPS	TBP	Actin	H4	COP-A	EIF1α
NormFinder	Actin	TBP	RPS	H4	COP-A	EIF1α
geNorm	Actin|TBP		RPS	H4	COP-A	EIF1α
Recommended comprehensive ranking	Actin	TBP	RPS	H4	COP-A	EIF1α

The graphs showing the best reference genes according to all the four computational programs are reported in Supplementary Figure 3.6 a-e. On the basis of the best reference analysis results, Actin, TBP and RPS were selected as reference genes for the gene expression analysis.

3.5.8 Quantitative PCR analysis

Quantitative PCR (qPCR) experiments were performed in MicroAmp Optical 384-Well reaction plate (Applied Biosystems, Waltham, Massachusetts, USA) with Optical

Adhesive Covers (Applied Biosystems, Waltham, Massachusetts, USA) in a Viia7 Real Time PCR System (Applied Biosystem, Waltham, Massachusetts, USA). Five serial dilutions of mixed RNAs were used to obtain the standard curve. The PCR volume for each sample was 10 µL, with 5 µl of SensiFASTTM SYBR® Lo-ROX Kit (BIO_94020, Bioline Reagents Ltd., London, UK), 1 µL of cDNA template (1:10 dilution of CTR, SiDepl and NDepl templates and 1:15 dilutions of PDepl with its CTR templates) and 4 µL of 0.7 µM oligo mix (forward and reverse). The oligo used for qPCR are listed in Table 3.6.

Table 3.6 List of oligo used for qPCR analysis. Abbreviations: Tm=Temperature of melting;

Primer name	**Forward sequence**	**Reverse sequence**	**Tm (°C)**	**Ampl. Length (bp)**
Actin	TCGGCCCTTGAGAAG AGTTTCG	GATGGTCTGGAAAG TGGAGTCC	64.2/64.2	147
TBP	CCTTCTTCAACCCCT CCACCAAC	GTTCGCTCATCCCA CGTTTTCG	66.6/64.2	161
RPS	GCGCCTTGACAATAC CGTCGATG	GACCTCATTGGACA GGTTCTTCC	66.6/64.6	148
SQS	TCTCTAGCGGGTGAA CTACA	CCTAAATCGCCACT ACCACT	60.6/61	152
SMT_A	TTTTAAGGGAGGGA ACGTGG	TTCCTTCGGCTGAT TCAAGT	65.2/64	181
SMT_B	AAATCTAGGTGGAG TTCGCA	AATCCTAGCATGTG ACCACG	61.7/63	195
SMT_C	AATGATGAGTACGG AGGGTG	GAGTACTTCGTGAG CCTGAG	61.9/59.8	176
SMT_D	AGTCAAAGCGTCAA TTTCGG	CCAGCCCCATTCAT AGAAGT	64.4/63.2	163
CPI1	GCACTCCTTCCATAA CCGTA	CCAGTAATTGCCGA TGAAGG	62/64.6	167
SC5DL	GTATGATGGCTGTTG GAGGT	GGTAGATCCCACAC GATGTC	61.9/62.4	189

DHCR7	TTAGCATGACGTTTT GGCAG	CAATCGAACGCATG TATCCC	64.1/65.5	162

Program reaction used was: 95 °C for 20 s, 40 cycles of 95 °C for 1 s and 60 °C for 20 s. The program was set to reveal the melting curve of each amplicon from 60 °C to 95 °C, and read every 0.5 °C. Single peaks for all genes confirmed gene-specific amplification and the absence of primer-dimers. All qPCR reactions were carried out in triplicate to capture intra-assay variability. Each assay included three no-template negative controls for each primer pair. The quantification of the samples was conducted using the standard curve with the qPCR software QuantStudio Real-Time PCR (Applied Biosystems, Waltham, Massachusetts, USA). qPCR results have been analysed with the Relative expression software tool (REST) (https://www.gene-quantification.info/) that test for significance by a Pair Wise Fixed Reallocation Randomisation Test and plot the results using standard error (SE) estimation via a complex Taylor algorithm. qPCR standard curves of the primer pairs used are reported in Supplementary Figures 3.7 and 3.8, while qPCR slope, efficiency (%) and R_2 of the primers are listed in Tables 3.2 and 3.3.

3.5.9 Statistical analysis

Statistical analysis was performed using GraphPad Prism 6.0 (GraphPad Software Inc. San Diego, California, USA). A multiple t-test, each row individually without correction for multiple comparisons, have been performed to assess the significance in MEP and MVA enzymes' differential expression in control and silica-depleted transcriptomes. An unpaired t-test have been performed in the comparison between the 24-MC quantity in control and the experimental conditions (i.e. nitrogen, silica and phosphate depletion) tested. Statistical analysis of the qPCR results have been performed with REST using 2000 randomisations of the Pair Wise Fixed Reallocation Randomization Test implemented in the software.

3.6 References

1. Fabris, M. *et al.* Tracking the sterol biosynthesis pathway of the diatom *Phaeodactylum tricornutum*. *New Phytologist* **204**, 521–535 (2014).

2. Jaramillo-Madrid, A. C., Ashworth, J., Fabris, M. & Ralph, P. J. Phytosterol biosynthesis and production by diatoms (Bacillariophyceae). *Phytochemistry* **163**, 46–57 (2019).

3. Stonik, V. & Stonik, I. Low-Molecular-Weight Metabolites from Diatoms: Structures, Biological Roles and Biosynthesis. *Mar Drugs* **13**, 3672–3709 (2015).

4. Dufourc, E. J. Sterols and membrane dynamics. *J Chem Biol* **1**, 63–77 (2008).

5. Gallo, C. *et al.* Diatoms synthesize sterols by inclusion of animal and fungal genes in the plant pathway. *Sci Rep* **10**, 4204 (2020).

6. Gallo, C., d'Ippolito, G., Nuzzo, G., Sardo, A. & Fontana, A. Autoinhibitory sterol sulfates mediate programmed cell death in a bloom-forming marine diatom. *Nat Commun* **8**, 1292 (2017).

7. Jaramillo-Madrid, A. C. *et al.* Overexpression of Key Sterol Pathway Enzymes in Two Model Marine Diatoms Alters Sterol Profiles in *Phaeodactylum tricornutum*. *Pharmaceuticals (Basel)* **13**, 481 (2020).

8. Di Dato, V. *et al.* Transcriptome sequencing of three *Pseudo-nitzschia* species reveals comparable gene sets and the presence of Nitric Oxide Synthase genes in diatoms. *Sci Rep* **5**, 12329 (2015).

9. Jaramillo-Madrid, A. C., Ashworth, J. & Ralph, P. J. Levels of Diatom Minor Sterols Respond to Changes in Temperature and Salinity. *Journal of Marine Science and Engineering* **8**, 85 (2020).

10. Pollier, J. *et al.* A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nat Microbiol* **4**, 226–233 (2019).

11. Jaramillo-Madrid, A. C., Ashworth, J., Fabris, M. & Ralph, P. J. The unique sterol biosynthesis pathway of three model diatoms consists of a conserved core and diversified endpoints. *Algal Research* **48**, 101902 (2020).

12. Rampen, S., Abbas, B., Schouten, S. & Sinninghe-Damste, J. A comprehensive study of sterols in marine diatoms (Bacillariophyta): Implication of their use as tracers for diatom productivity. *Limnology and Oceanography* **55**, (2010).

13. Orcutt, D. M. & Patterson, G. W. Sterol, fatty acid and elemental composition of diatoms grown in chemically defined media. *Comp Biochem Physiol B* **50**, 579–583 (1975).

14. Nuzzo, G. *et al.* UPLC⁻MS/MS Identification of Sterol Sulfates in Marine Diatoms. *Mar Drugs* **17**, E10 (2018).
15. Barrett, S. M., Volkman, J. K., Dunstan, G. A. & LeRoi, J.-M. Sterols of 14 species of marine diatoms (Bacillariophyta). *J Phycol* **31**, 360–369 (1995).
16. Dufourc, E. J. The role of phytosterols in plant adaptation to temperature. *Plant Signal Behav* **3**, 133–134 (2008).
17. Randhir, A., Laird, D. W., Maker, G., Trengove, R. & Moheimani, N. R. Microalgae: A potential sustainable commercial source of sterols. *Algal Research* **46**, 101772 (2020).
18. Tiwari, A. *et al.* Therapeutic attributes and applied aspects of biological macromolecules (polypeptides, fucoxanthin, sterols, fatty acids, polysaccharides, and polyphenols) from diatoms - A review. *Int J Biol Macromol* **171**, 398–413 (2021).
19. Kim, Y.-S., Li, X.-F., Kang, K.-H., Ryu, B. & Kim, S. K. Stigmasterol isolated from marine microalgae *Navicula incerta* induces apoptosis in human hepatoma HepG2 cells. *BMB Rep* **47**, 433–438 (2014).
20. Burlat, V., Oudin, A., Courtois, M., Rideau, M. & St-Pierre, B. Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J* **38**, 131–141 (2004).
21. Carland, F., Fujioka, S. & Nelson, T. The Sterol Methyltransferases SMT1, SMT2, and SMT3 Influence *Arabidopsis* Development through Nonbrassinosteroid Products. *Plant Physiology* **153**, 741–756 (2010).
22. Billheimer, J. T., Alcorn, M. & Gaylor, J. L. Solubilization and partial purification of a microsomal 3-ketosteroid reductase of cholesterol biosynthesis. *Arch Biochem Biophys* **211**, 430–438 (1981).
23. Tansey, T. & Shechter, I. Squalene synthase: Structure and regulation. *Progress in nucleic acid research and molecular biology* **65**, 157–95 (2001).
24. Maréchal, E. & Lupette, J. Relationship between acyl-lipid and sterol metabolisms in diatoms. *Biochimie* **169**, 3–11 (2020).
25. Burg, J. S. & Espenshade, P. J. Regulation of HMG-CoA reductase in mammals and yeast. *Progress in Lipid Research* **50**, 403–410 (2011).
26. Volkman, J. K. & Hallegraeff, G. M. Lipids in marine diatoms of the genus *Thalassiosira*: Predominance of 24-methylenecholesterol. *Phytochemistry* **27**, 1389–1394 (1988).

27. Guillard, R. R. L. Culture of Phytoplankton for Feeding Marine Invertebrates. in *Culture of Marine Invertebrate Animals: Proceedings - 1st Conference on Culture of Marine Invertebrate Animals Greenport* (eds. Smith, W. L. & Chanley, M. H.) 29–60 (Springer US, 1975).

28. Yang, J., Li, C. & Zhang, Y. Engineering of *Saccharomyces cerevisiae* for 24-Methylene-Cholesterol Production. *Biomolecules* **11**, 1710 (2021).

29. Lauritano, C., Orefice, I., Procaccini, G., Romano, G. & Ianora, A. Key genes as stress indicators in the ubiquitous diatom *Skeletonema marinoi*. *BMC Genomics* **16**, 411 (2015).

30. Lauritano, C. *et al.* New molecular insights on the response of the green alga *Tetraselmis suecica* to nitrogen starvation. *Scientific Reports* **9**, (2019).

31. Di Dato, V. *et al.* Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Sci Rep* **9**, 9893 (2019).

Chapter 4

Characterization of three novel bacteria associated with the diatoms *T. rotula* and *Skeletonema marinoi*

Adapted from the published article:

Di Costanzo, F., Di Dato, V., van Zyl, L. J., Cutignano, A., Esposito, F., Trindade, M., Romano, G. Three novel bacteria associated with two centric diatom species from the Mediterranean Sea, *Thalassiosira rotula* and *Skeletonema marinoi*. *International Journal of Molecular Sciences* **22**, 13199 (2021).

4.1 Abstract

Diatoms are a successful group of microalgae at the base of the marine food web. For hundreds of millions of years, they have shared common habitats with bacteria, which favoured the onset of interactions at different levels, potentially driving the synthesis of biologically active molecules. To unveil their presence, we sequenced the genomes of bacteria associated with the centric diatom *Thalassiosira rotula* from the Gulf of Naples. Annotation of the metagenome and its analysis allowed the reconstruction of three bacterial genomes that belong to currently undescribed species. Their investigation showed the existence of novel gene clusters coding for new polyketide molecules, antibiotics, antibiotic-resistance genes and an ectoine production pathway. Real-time PCR was used to investigate the association of these bacteria with three different diatom clones and revealed their preference for *T. rotula* FE80 and *Skeletonema marinoi* FE7, but not *S. marinoi* FE60 from the North Adriatic Sea. Additionally, we demonstrate that although all three bacteria could be detected in the culture supernatant (free-living), their number is up to 45 times higher in the cell associated fraction, suggesting a close association between these bacteria and their host. We demonstrate that axenic cultures of *T. rotula* are unable to grow in medium with low salinity (<28 ppt NaCl) whereas xenic cultures can tolerate up to 40 ppt NaCl with concomitant ectoine production, likely by the associated bacteria.

4.2 Introduction

The role of bacteria in the equilibrium of marine ecosystems is increasingly being recognized, based on the growing number of papers demonstrating their role in organic matter decomposition and their strong interaction with other marine organisms, including diatoms[1–6]. Since diatoms are at the base of the marine food web, they are fundamental to the health of the planet, playing a crucial role in carbon fixation[7] and contribute to the formation of oil deposits sinking to the bottom of the oceans due to their heavy frustule[8]. They are represented by more than 200,000 different species able to adapt to every marine niche and bloom in both coastal and oceanic areas, or wherever sufficient nutrients can sustain their growth[8,9].

Studies focusing on diatom–bacteria (D–B) interactions have reported communities of different species of bacteria living strictly on diatoms (strictly associated, SA) or in their close surroundings (free living, FL)[3,6,10,11]. The exact composition of these communities at the species level is not yet well defined and appears to be dependent on different factors,

which include the algal species and its growth stages as well as the light and temperature conditions under which they are growing[10,12–14]. The D–B interaction is predominantly based on the exchange of metabolites and nutrients (e.g., vitamin B_{12})[11], including the synthesis of species-specific nutrients, and the secretion of defence molecules against other algal or bacterial species[12,15,16]. An interesting example comes from the bacterial species *Phaeobacter inhibens* that can protect diatoms from harmful prokaryotic species by influencing and shaping the bacterial community developing around them[15,17]. In addition, bacteria seem to establish a stable long-term association with diatoms[14], and certain bacterial orders, in particular Rhizobiales and Rhodobacterales, tend to co-occur with *Symbiodiniaceae* dinoflagellates when establishing symbiosis with other organisms[18]. Moreover, the surface properties of the algal cells can be influenced by SA species that can help in the formation of aggregates[5,6]. Indeed, depending on the type of interaction established, bacteria can contribute to algal aggregation at levels depending on their community composition and the type of algal exopolymer released[5,6,10], which can also serve as a mechanism of escape from predators[5,10].

The high abundance of diverse and novel bacterial species that compose those microbial communities, both SA and FL, could harbour enormous metabolic diversity in which there is hidden potential for the discovery of yet unknown metabolites that could be valuable for exploitation. This is in part due to the space and nutrient limitations that create a competitive environment, requiring bacteria to develop an adaptive response and strategies that enable them to survive over others, including the utilization of secondary metabolites that function as repellents against competing species. Identification and characterization of these metabolites should allow for the discovery of new drugs with application in many fields[19,20]. It has been demonstrated in several cases that the bacterial symbiont, rather than the host organism, is responsible for the production of secondary metabolites beneficial to itself as well as to its host[21]. Some of these compounds have also successfully made it to market[22].

Thalassiosira rotula is a cosmopolitan coastal bloom-forming diatom species, present in the Gulf of Naples (Italy) often dominating the phytoplankton community due to its large size and elevated concentration[23]. This species has been explored for the presence of secondary metabolites produced under nutrient stress conditions, i.e., silica depletion, revealing its ability to express genes related to the synthesis of prostaglandins, iridoid type alkaloids and polyketides (as described in Chapter 1 of this thesis)[24]. *T. rotula* has also been shown to produce compounds toxic for the progeny of predators grazing on it[25] as well as antibacterial compounds[26].

While several studies have been conducted to identify the bacterial species associated with *T. rotula*[14,27,28], to our knowledge, none have explored the biotechnological potential of the bacteria associated with this diatom.

Here, we present the genomes of three novel diatom-associated bacteria reconstructed from a metagenome dataset and investigate their association with three diatom strains. The analysis of their secondary metabolite gene clusters revealed the presence of new hybrid non-ribosomal peptide synthase (NRPS)/Type 1 polyketide synthase (T1PKS) biosynthetic pathways, as well as clusters coding for ectoine, terpenes and bacteriocins.

The availability of these genomes and our preliminary analysis of the association between these bacteria and diatoms will help open the way for an improved understanding of the relationship between these organisms as well as provide the basis for a genomics-based exploration of the secondary metabolite production capability of these bacteria.

4.3 Results

4.3.1 Sequencing, Assembly, Binning and Classification of MAGs

Under antibiotic selection, the sequencing of the associated bacteria from *T. rotula* FE80 genomic DNA (gDNA) resulted in a 22 Mb bacterial metagenome dataset consisting of 416 contigs greater than 5 kb (Table 4.1).

Table 4.1. Statistics of raw sequence data from which metagenome-assembled genomes (MAGs) were recovered through de novo assembly.

Parameter Type	**Base Pairs**
Total number of reads before trimming	20,505,104
Total number of reads after trimming	19,519,539
Average read length	239.5
Number of reads in contigs	17,286,414
Number of contigs >5 kb	416
N50 (bp)	146,992
Assembly size (Mb)	22.15
Maximum contig length (bp)	952,220

The binning of the metagenomic contigs resulted in eight clusters (Figure 4.1), three of which (Cl-1, Cl-2 and Cl-8), were well defined and of sufficient completeness and quality to warrant further analysis (Table 4.2). These were deposited in the Genbank database as high quality draft genomes.

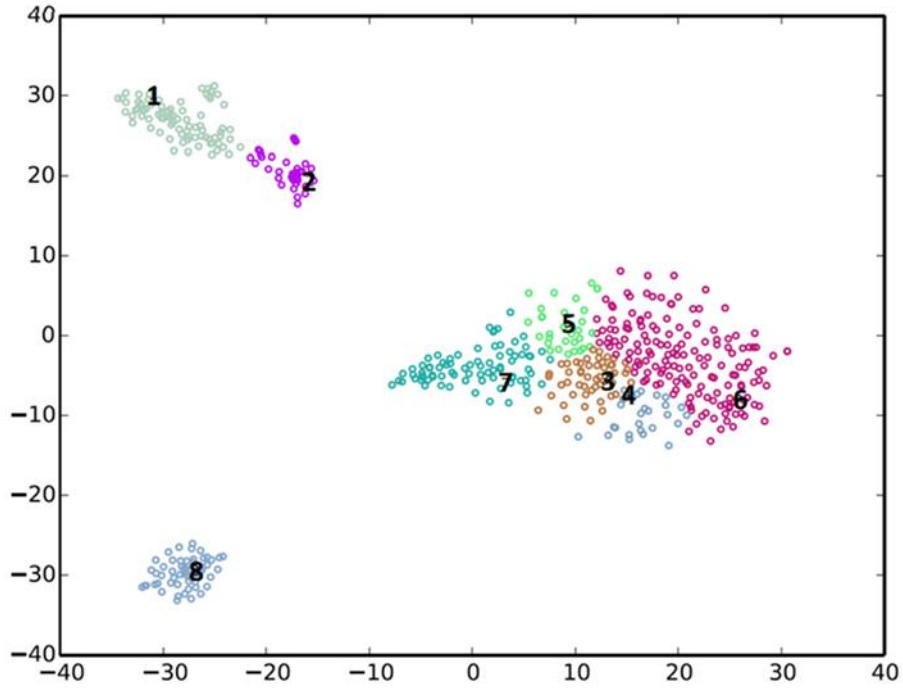


Figure 4.1. Metagenomic binning of contigs using the MyCC analysis tool[29].

Table 4.2. Statistics for selected cluster genomes.

Cluster	**Cl-1**	**Cl-2**	**Cl-8**
Completeness %	97.66	99.57	99.05
Contamination %	0.38	1.3	1.64
Number of contigs	81	20	80
N50 (bp)	98,035	327,203	346,266
Assembly size (Mb)	4.78	3.74	8.57
Maximum contig length (bp)	239,842	952,220	622,837
Average coverage	633.5	66.6	31.6
Number of CDS	4680	4101	7071
G + C%	62	56	45
Number of tRNAs	37	40	48

With an average amino acid identity (AAI) of 77.56 %, 90.05 % and 81.41 % respectively, taxonomic classification identified Cl-1 as a new species in the genus *Aestuariivita*[30],

family *Rhodobacteraceae* (therefore belonging to the larger Roseobacter group); Cl-2 as a new species in the placeholder genus *Mf105b01*[31], family *Parvibaculaceae*; and Cl-8 as a new species in the placeholder genus *Roseivirga_A*, most closely related to the MAG *Roseivirga_A sp002427745*, family *Cyclobacteriaceae*[32] (Figure 4.2; Table 4.3, Supplementary Tables 4.1-4.3).

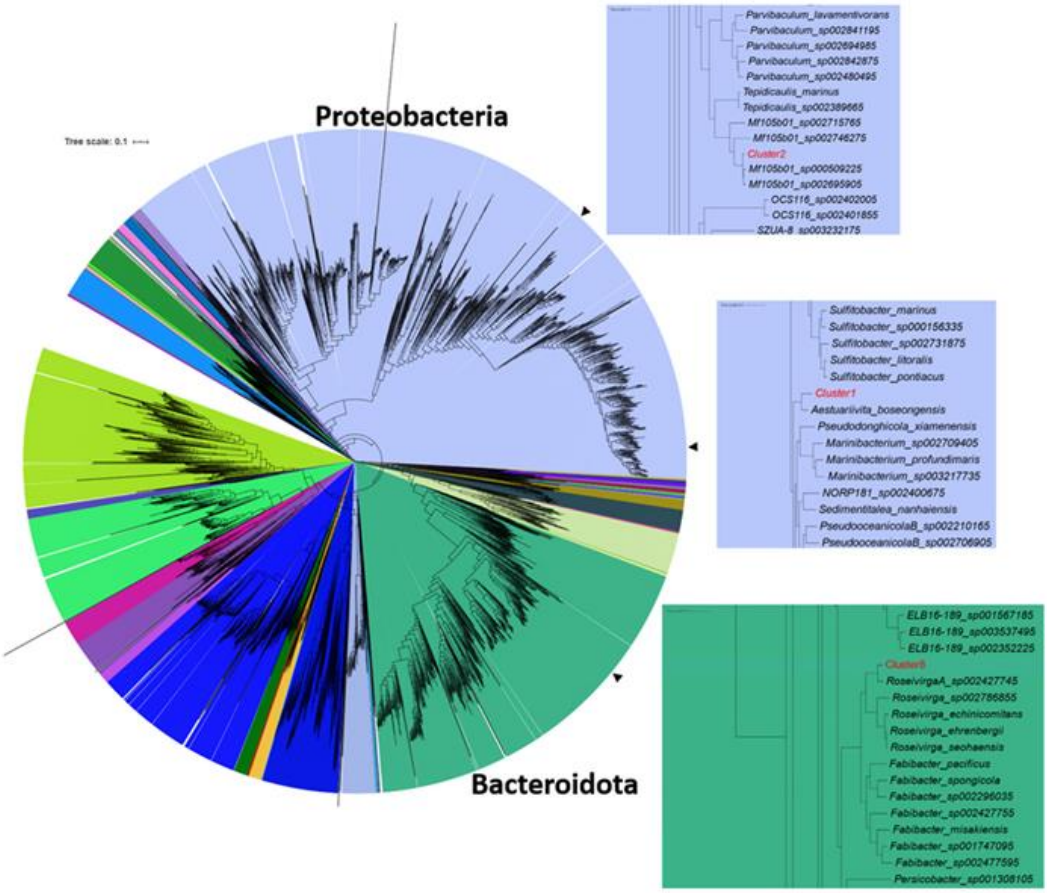


Figure 4.2. Phylogenetic placement of selected MAGs following analysis performed with GTDB-Tk[33]. The tree is coloured by phylum. Arrows indicate the positions of the clusters in the tree and the insets show enlarged views of the region of the tree, including the selected clusters.

Table 4.3. Selected MAG phylogeny.

Cluster	Cl-1	Cl-2	Cl-8
Phylum	Proteobacteria	Proteobacteria	Bacteroidota
Family	*Rhodobacteraceae*	*Parvibaculaceae*	*Cyclobacteriaceae*
Genus	*Aestuariivita*	*Mf105b01*	*Roseivirga_A*

Average Amino acid Identity (AAI)	77.56 %	90.05 %	81.41 %

Based on RAST analysis[34], we identified in Cl-1 the presence of genes encoding a photosystem II-type photosynthetic reaction center and a carotenoid pigment that may interact with the light harvesting system.

Cl-2 was identified as a new species in the genus *Mf105b01* of the *family Parvibaculaceae* that was first described as a bacterial contaminant of the genome sequence of the dinoflagellate *Symbiodinium minutum*[35].

Cl-8 is related to *Roseivirga_A sp002427745* assembled from a coral mucus metagenome[36]. RAST annotation could only place 26 % of CDS into a subsystem category for both Cl-8 and *Roseivirga_A sp002427745*, supporting the novelty in the amino acid sequences coded for by these genomes. In addition, RAST analysis showed that Cl-8 encodes 34 putative β-lactamase/cephalosporinases, similarly to *Roseivirga_A sp002427745* that also encodes 30 of these enzymes suggesting that they naturally possess multiple β- lactam resistance mechanisms.

4.3.2 Analysis of MAG Secondary Metabolite Gene Clusters

The potential secondary metabolite production from these organisms was estimated using PRISM[37] and antiSMASH[38] (Table 4.4, Supplementary Table 4.4). Both pipelines identified similar PKS and NRPS biosynthetic clusters, while antiSMASH provided a broader and more detailed cluster annotation (Table 4.4).

Table 4.4. Biosynthetic gene clusters identified by PRISM and antiSMASH.

	Cluster	**Cl-1**	**Cl-2**	**Cl-8**
PRISM	**n° of Biosynthetic clusters**	7	1	3
	Type of Biosynthetic clusters	PKS, NRPS, Acyl homoserine lactone, Ectoine	PKS	Lasso peptide, PKS, NRPS
antiSMASH	**n° of Biosynthetic clusters**	9	1	8

	Type of Biosynthetic clusters	NRPS, hserlactone, ectoine, terpenes, T1PKS, Bacteriocin	Bacteriocin	NRPS, T1PKS, T3PKS, Bacteriocin, terpenes, beta lactone

Cl-2 does not appear to encode any biosynthetic gene clusters, except for one putative bacteriocin, whereas Cl-1 and Cl-8 likely produce new NRPS/PKS hybrid compounds based on the high sequence novelty of the pathways (Supplementary Figure 4.1).

4.3.3 Ectoine Quantification in Xenic *T. rotula* Cultures

Considering the presence of a biosynthetic cluster for the synthesis of the osmolyte ectoine on the Cl-1 genome, we sought to investigate the production and contribution of ectoine by xenic *T. rotula* cultures to osmotic stress tolerance in salinities ranging from 20 ppt to 40 ppt. The ectoine concentration was measured at exponential/early stationary (day 3) and late stationary phases (day 8). Axenic cultures could not tolerate salinities below 28 ppt, while xenic cultures could grow in salinities ranging from 20 ppt to 40 ppt (personal communication).

The xenic cultures in higher salinity media (from 32 ppt to 40 ppt) had a similar growth trend, with exponential growth until day 4, while those cultured at 20 ppt, and to a lesser extent those at 24 ppt and 28 ppt, showed a significant reduction in cell numbers during the first day of culturing followed by a recovery period thereafter (Figure 4.3 a). The variability observed for the 20 ppt cultures suggests that the diatoms had difficulty in growing at this low salinity. However, after 8 days, cell density reached similar values at all salinities (Figure 4.3 a).

The ectoine amount was significantly different among different salinities and days, being higher at all salinities at day 3 compared with day 8 (Figure 4.3 b). Interestingly, at day 3, the ectoine expressed in relation to diatom cell number was higher at 20, 24 and 40 ppt, which corresponded to a lower diatom growth rate (Figure 4.3 b). At day 8 of growth in which all the cultures reached the same cell concentrations, ectoine was highest at 40 ppt, stable at 36–32 ppt and decreased at lower salinities (Figure 4.3 b).

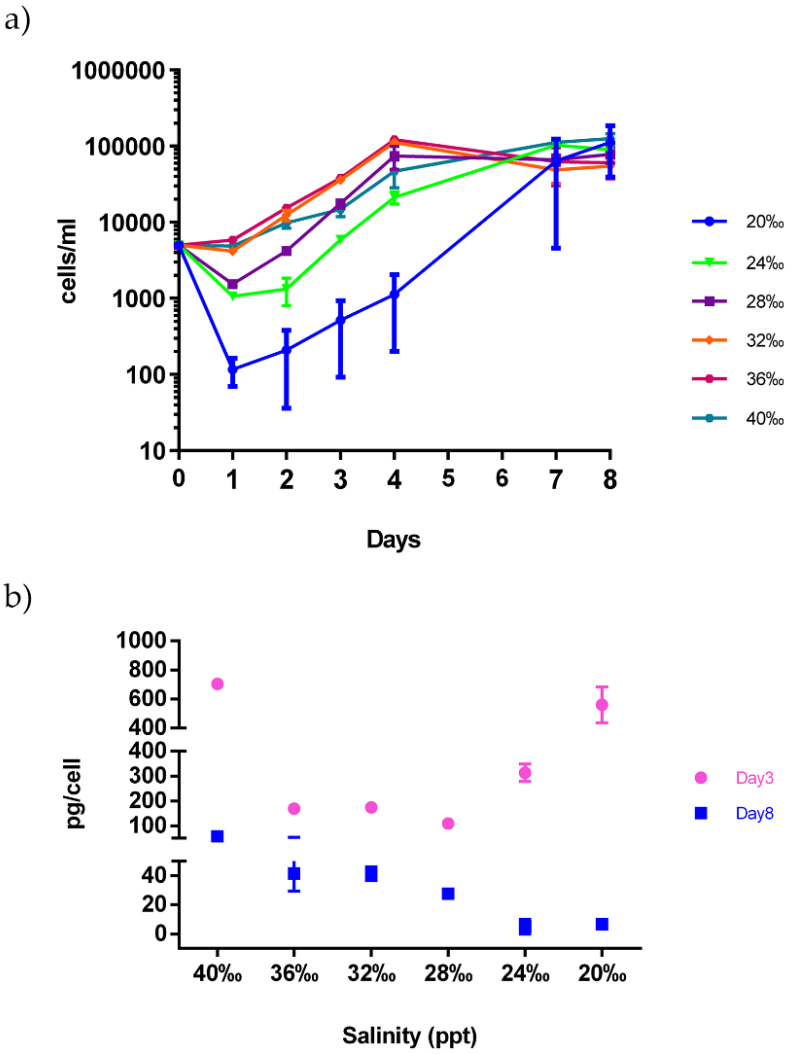


Figure 4.3. Ectoine quantification in xenic *T. rotula* culture. (**a**) Growth curves of *T. rotula* in media with different salinities; (**b**) ectoine amount, reported as pg per *T. rotula* cell. Day 3 = third day of growth; day 8 = eighth day of growth.

Cl-1 abundance evaluated in these culture fractions by qPCR is shown in Figure 4.4, where growth was only detected at day 3 and only at the 20 ppt and 24 ppt salinities and at equal amounts (Figure 4.4, Supplementary Table 4.5). On the contrary, at day 8, at all salinities, Cl-1 was clearly detectable (Figure 4.4, Supplementary Table 4.5). The comparison of Figures 4.3 b and 4.4 shows that, at day 8, the ectoine trend is concomitant with the Cl-1 abundance.

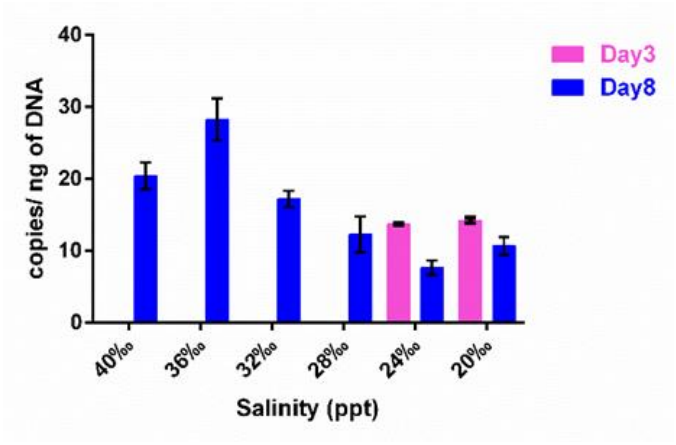


Figure 4.4. Cl-1 qPCR amplification on SA gDNA extracted from *T. rotula* xenic cultures grown at different salinities. The cultures were sampled at exponential-early stationary phase (Day 3) and at late stationary phase (Day 8) during their growth.

4.3.4 Verification of MAG Association with Diatoms

Following many unsuccessful attempts to bring the bacteria representing the MAGs into culture, and considering that they could be a co-dependent consortium, we employed PCR and qPCR as a means of tracking their growth and association with the diatom. To confirm the association of the three bacterial strains (Cl-1, Cl-2 and Cl-8) with *T. rotula* FE80, we amplified specific sequences unique to each MAG from gDNA extracted from the associated FL bacterial fraction (Figure 4.5). In addition, to verify their specific association to different diatom species or their clones, we also amplified from two clones, named FE7 and FE60, of another model centric diatom, *Skeletonema marinoi* (Figure 4.5). Amplification resulted in specific amplicons from *T. rotula* FE80 and *S. marinoi* FE7 but not FE60, demonstrating a specific clonal preference. Amplification using the Cl-2 specific primer set resulted in two amplicons, both confirmed through sequencing to be specific for Cl-2 (Supplementary Figure 4.2 b).

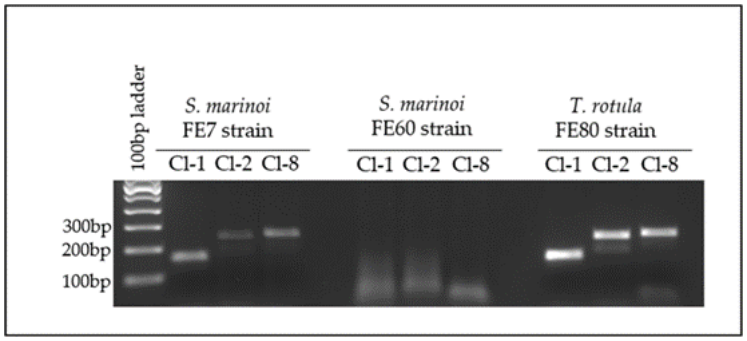


Figure 4.5. Cl-1, Cl-2 and Cl-8 PCR amplification from FL gDNA. FL fractions of each diatom strain were tested on the 7th day of growth.

Since the *T. rotula* FE80 clone we were working with lost viability, the CNR-ICB (Pozzuoli, NA, Italy) laboratories kindly provided us with a different preparation of the same clone. However, their clone was periodically treated with antibiotics that presumably led to the loss of the three MAGs under investigation. We therefore attempted to reintroduce the bacterial assemblage collected from the FL fraction of the *S. marinoi* FE7 cultures to T. rotula FE80. The reintroduction appeared successful, and the association was stable as shown by the positive and specific amplification of all three MAG sequences at each time point tested, from 1 month to 8 months after the reintroduction (Figure 4.6). The absence of diatom species cross-contamination was confirmed through microscopic observation of the culture. Indeed, *S. marinoi* and *T. rotula* are easily differentiated due to the different morphologies and dimension of the two species.

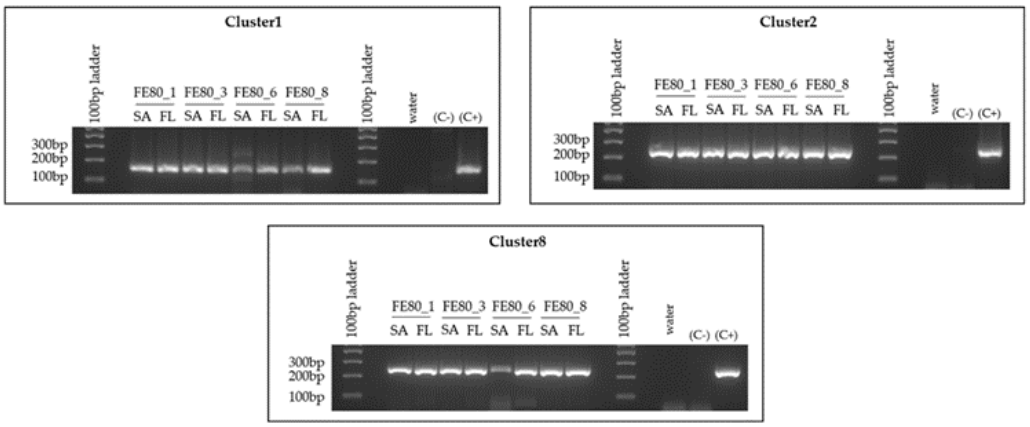


Figure 4.6. Reintroduction of the bacterial species to *T. rotula* FE80. Cl-1, Cl-2 and Cl-8 PCR amplifications were performed on gDNA extracted separately from the FL and SA fractions associated with the *T. rotula* FE80 cultures after the reintroduction of the *S. marinoi* FE7 associated - FL bacterial fraction. Abbreviations: C+: positive control = *S. marinoi* FE7 xenic total culture gDNA; C-: negative control = *T. rotula* FE80 axenic total culture-gDNA (before reintroduction of bacteria to the culture); FE80_1 = 1 month after the reintroduction; FE80_3 = 3 months after the reintroduction; 6 = 6 months after the reintroduction; FE80_8 = 8 months after the reintroduction of the bacteria. FE80 = *T. rotula* clone FE80.

4.3.5 MAG-Diatom Association Type and Temporal Variation

qPCR amplification of the MAG sequences from the FL and SA fractions of a *S. marinoi* FE7 xenic culture highlighted a preferential distribution of each MAG in the SA fraction, especially for Cl-2 and Cl-8 (Figure 4.7). Based on copy number at day 4, Cl-2 and Cl-8 were 11 and 45 times more abundant in the SA than in the FL fraction, respectively

(Supplementary Table 4.7). No appreciable difference between SA and FL could be detected for Cl-1 on day 4. After 7 days, the abundance of all three MAGs in the SA fraction increased, confirming growth and preferential association with the algae (Figure 4.7). Conversely, the FL fraction abundance of Cl-1 and Cl-2 did not change with diatom growth. On the contrary Cl-8 abundance also increased in the FL fraction with a copy number ~9 times higher at day 7 compared to day 4 (Figure 4.7). Among the three MAGs, Cl-2 showed the strongest diatom dependence (SA 67 times higher than FL at day 7) (Figure 4.7, Supplementary Table 4.7).

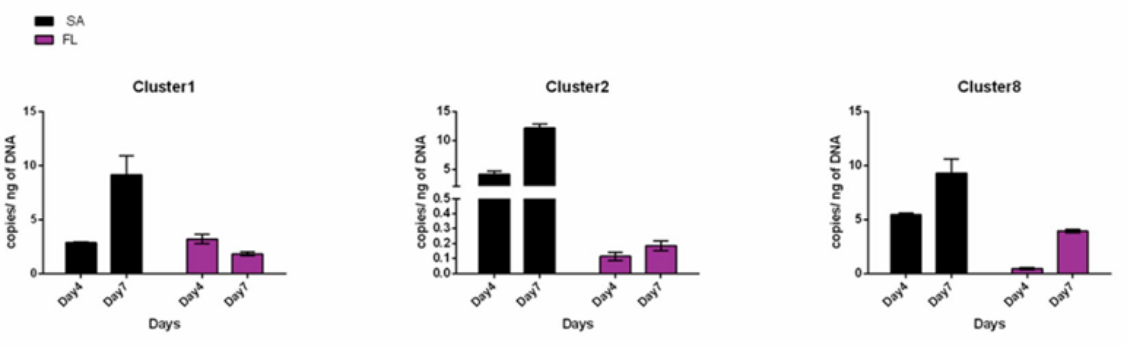


Figure 4.7. MAG distribution between the SA and FL diatom associated fractions. qPCR using Cl-1, Cl-2 and Cl-8 specific primers performed on SA and FL bacterial fractions collected from *S. marinoi* FE7 sampled on the 4th and 7th days of growth.

To further confirm the strict association and assess the possible co-dependence of the MAGs and diatoms, the S. marinoi FE7-FL fraction was cultured in three different medium types: marine broth supplemented with *S. marinoi* FE7 spent media (MB), f/2[39] media supplemented with two different concentrations of sonicated *S. marinoi* FE7 cells (f/2¼ and f/2¾).

Cultivation in MB medium resulted in a turbid culture, compared to the f/2¼ and f/2¾ cultures when bacterial density was measured spectrophotometrically (Supplementary Figure 4.7). However, the relative abundance of the MAGs did not reflect the total bacterial abundance. Cl-1, Cl-2 and Cl-8 detection by gDNA copy number determination revealed their presence when cultured in f/2¼ or f/2¾ medium but not in MB (Figure 4.8 a, Supplementary Table 4.9). Indeed, the copy number values from the MB cultures were close to zero at each time point tested, indicating no growth of the three MAGs of interest (Figure 4.8 a). On the contrary, Cl-1, Cl-2 and Cl-8 copy numbers could be detected when the liquid culture was either f/2¼ or f/2¾. The f/2¾ medium was preferred as indicated by the gDNA copy number being very low on day 3 (Figure 4.8 a), indicating that with

this media, at least 6 days are needed to detect the growth. This observation was confirmed when quantifying the MAG presence in a *S. marinoi* FE7-derived FL fraction grown in f/2¼ media treated with DNAse (to exclude possible false positives due to gDNA traces in the medium). Indeed, based on gDNA copy number the MAG detection was not due to gDNA traces, but due to actively growing bacteria (Figure 4.8 b, Supplementary Table 4.10). The MAG growth curve was thus refined and a different growth rate among the MAGs was highlighted. Cl-1 and Cl-8 had a short growing time, being detectable from day 4 and reaching a maximum at day 7, whereas Cl-2 was detectable from day 7 (Figure 4.8 b). The most abundant was Cl-2 being 4× and 17× more abundant than Cl-1 and Cl-8, respectively.

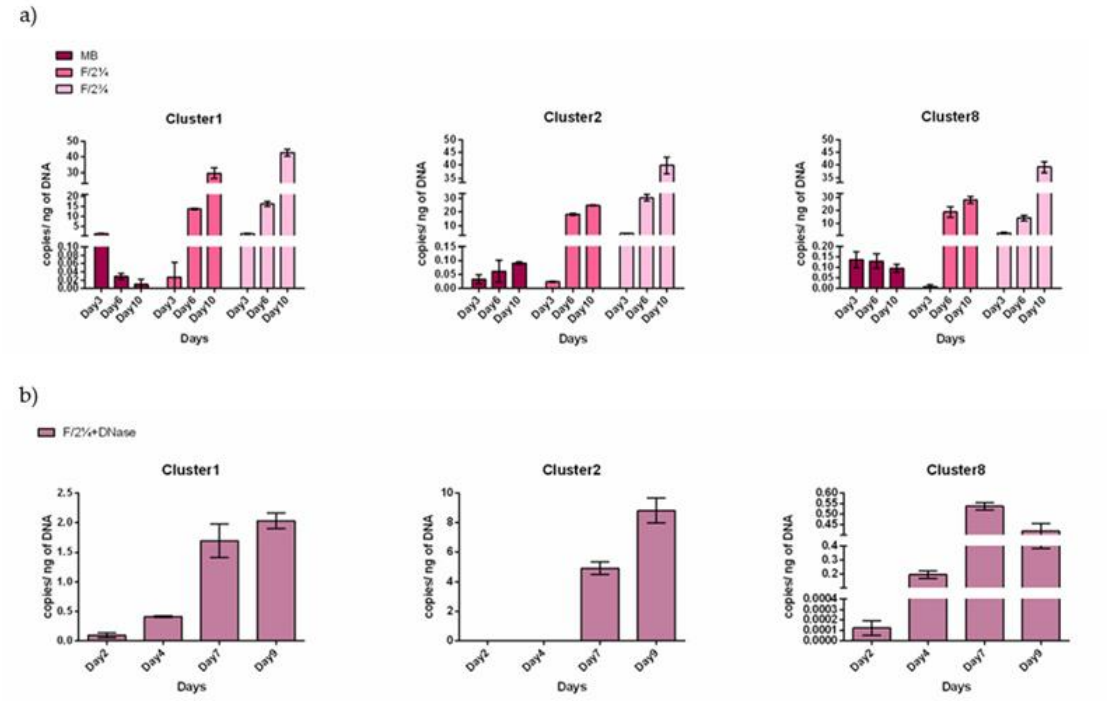


Figure 4.8. Diatom-free MAG growth. (**a**) qPCR amplification of Cl-1, Cl-2 and Cl-8 specific sequences from gDNA extracted from FL bacteria separated from *S. marinoi* FE7 cultures grown in marine broth supplemented with spent medium (MB), or in f/2 medium supplemented with different concentrations of sonicated *S. marinoi*-FE7 cells (f/2¼ and f/2¾) (see M&M for details). Sampling time points: 3rd (Day3), 6th (Day6) and 10th (Day10). (**b**) qPCR amplification of Cl-1, Cl-2 and Cl-8 on gDNA extracted from FL bacteria grown in f/2¼ media previously treated with DNase. Sampling time points: 2nd (Day2), 4th (Day4), 7th (Day7) and 9th (Day9).

4.3.6 MAG Antibiotic Tolerance

RAST analysis of the Cl-8 genome revealed 34 putative β-lactamases. Since all three

MAGs were recovered from cultures grown in a mixture of antibiotics, including Ampicillin, we evaluated their tolerance to two different β-lactam antibiotics (Ampicillin (Amp), and Cefotaxime (Ctx)) concentrations, when grown in association with algae. In addition, we also tested these antibiotics in combination with the aminoglycoside Streptomycin (Str). The tolerance was evaluated on both SA and FL fractions. Two different mixtures of antibiotics were tested: a mixture containing all three antibiotics (ACS mix), and a mix that did not contain Str (AC mix); and both mixtures were tested at two different concentrations (2× and 1×).

Cl-1 was not tolerant to the ACS mix, at either concentration, while it was weakly tolerant to the AC-2× mix (Figure 4.9 a, lanes 2, barely visible band). Similarly, Cl-2 was not tolerant to the ACS mix, weakly tolerant to AC-2× and tolerant to the AC-1×, even if the specific PCR amplification was more visible in the FL than the SA fraction (Figure 4.9 b, SA- FL lane 4).

Cl-8 displayed tolerance to the ACS-1× mix but not to the ACS-2× mix in both fractions (Figure 4.9 c). When Str was not included in the mixture, the antibiotic treatment was tolerated at both concentrations, suggesting that the putative β-lactamases identified in the Cl-8 genome may be responsible for the higher resistance observed for this bacterium (Figure 4.9 c).

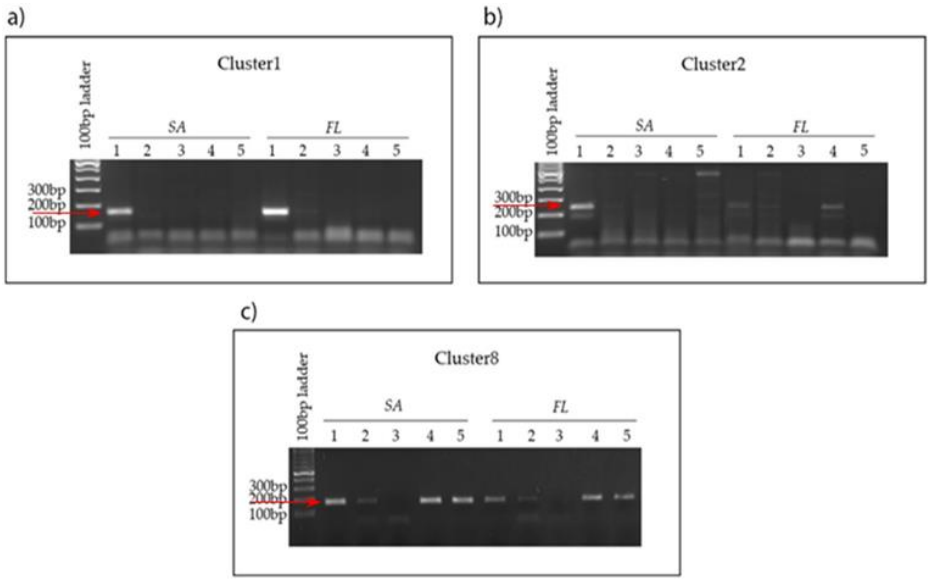


Figure 4.9. MAG antibiotic tolerance in strictly associated (SA) and free-living (FL) fractions. (**a**) Cl-1. (**b**) C-2 (**c**) Cl-8 detection in the presence of different antibiotic (Abt) mixtures and concentrations. Lanes 1, no Abt; lanes 2, 500 mg/mL Amp, 10 mg/mL Ctx (ACx2); lanes 3, 500 mg/mL Amp, 50 mg/mL Str, 10 mg/mL Ctx (ASCx2); lanes 4, 250 mg/mL Amp, 5 mg/mL Ctx (ACx1); lanes 5, 250 mg/mL Amp, 25 mg/mL Str, 5 mg/mL Ctx (ASCx1). Red arrows indicate expected amplicon size.

4.4 Discussion

Diatoms and bacteria have shared the same environment for more than 200 million years[1] and more than 5 % of the diatom genome is composed of bacterial genes[4], whose contribution has been fundamental in the diversification and the evolutionary success of diatoms[1]. Bacteria can live strictly associated to diatom cell walls or in their surrounding water space constituting very different communities[13]. However, in general, the bacterial phyla mainly associated with diatoms are Proteobacteria and Bacteroidetes[40] with few genera representative members, such as *Sulfitobacter*, *Roseobacter* and *Flavobacterium*[1]. Despite the importance of these associations in the evolutionary history of diatoms, diatom-bacteria interactions remain poorly studied[41]. Recently however, this field has garnered more attention, and studies are performed based on the analysis of the entire microbiome associated with diatoms, reporting a global perspective on their relationships[42].

Within this work, based on metagenome analysis of bacteria associated with *T. rotula* we identified three new bacterial species, named Clusters 1, 2 and 8 (Cl-1, Cl-2, Cl-8). The three selected MAGs are members of the families *Rhodobacteraceae, Parvibaculaceae* and *Cyclobacteriaceae*, known to be present in the marine environment, which have been shown to have a strong association with *Symbiodiniaceae*[18,35], while their association with the Bacillariophyceae has never been described before. These findings add novelty to what is currently understood about the bacterial species associated with diatoms. Supporting these species' novelty, for example, Cl-8, defined as a new species closely related to the *Roseivirga* genus, shared only 26 % of its CDS with the MAG *Roseivirga_A sp002427745* (81.41 % AAI) belonging to the *Cyclobacteriaceae* family[43].

In their 2020 study, Mönnich and co-workers identified *T. rotula* associated bacterial communities to be dominated by Rhodobacteraceae (30.5 %), Alteromonadaceae (27.7 %), and *Oceanospirillales* (18.5 %)[14]. The species identified in our work would not have been detected in that study, as their methodology could, at best, classify bacteria at family level and was based on an older database (SILVA v128 September 2016). This database did include the 16S rRNA sequence for *A. boseongensis*, but not the closest relatives of Clusters 2 and 8. The coarse grained-approach of identifying the major groups of bacteria associated with any environment using single marker genes has utility for environmental monitoring, but shows its limitations here as little can be said about the interactions between symbionts at this level (family or genus) and shows the advantage of performing whole genome sequencing and employing the latest databases available as opposed to targeting just the 16S rRNA marker.

Based on taxonomic classification Cl-1 was assigned to the *Aestuariivita* genus that currently consists of only two described members, namely *A. boseongensis*, isolated from a Korean tidal flat sediment in 2014[30] and *Aestuariivita atlantica* isolated from the deep sediment of the Atlantic Ocean in the 2014[44]. A phylogenomic analysis by Wirth and Whitman in 2018 reassigned *A. atlantica* to the novel genus *Pseudaestuariivita* with the type species *Pseudaestuariivita atlantica*, separating it from *A. boseongensis*[45]. This recent reassignment of genera and species highlights a gap in the classification of these genera which could include a number of novel and as yet unexplored species, such as Cl-1. The presence of genes encoding a photosystem II-type photosynthetic reaction center in Cl-1 suggested a photoheterotrophic behavior for it, although no PufX (component of the reaction center-light-harvesting 1-PufX (RC-LH1-PufX) complex) homolog could be identified. Although essential to enable photosynthesis in some organisms, such as *Rhodobacter sphaeroides*, in others, such as *Thermochromatium tepidum*, it is not required[46,47]. Its closest relative, A. boseongensis carries the same complement of photosynthesis reaction center genes, with the exception of the light-harvesting protein B-800/850 alpha and beta subunits. However, enzyme homologs involved in CO_2 fixation or carboxysome formation were not found, similar to many other aerobic anoxygenic photoheterotrophic bacteria, including the model organism *Dinoroseobacter shibae* which has been shown to use the 3- hydroxypropionate cycle for CO_2 fixation[48,49]. The key enzymes needed for the 3-hydroxypropionate cycle, (crotonyl-CoA carboxylase and propionyl-CoA carboxylase), together with the rest of the enzymes in the pathway are present in both Cl-1 and *A. boseongensis*.

The prediction for carotenoid pigment production in Cl-1 also supports its hypothesized photoheterotrophic nature and may interact with the light harvesting system. The phytoene synthase and dehydrogenase, responsible for carotenoid production, are divergently transcribed from two of the subunit genes (*chlI* and *chlD*) that compose the magnesium chelatase, essential for installing the Mg^{2+} ion in the bacteriochlorophyll protoporphyrin IX ring. This suggests that the regulation of carotenoid expression is linked to production of the chelatase responsible for bacteriochlorophyll formation and could be oxygen dependent, as observed for *R. sphaeroides*[50]. The Rhodobacteraceae are known to have phototrophic members that can be very flexible in their energy generation using chemotrophic paths if needed[50,51].

Evidence of the association of a Cl-2-like organism with *T. pseudonana* exists in the genome GCA_002380265.1 deposited on the Genbank database. Bacteria identified through 16S rRNA sequencing as belonging to the genus *Mf105b01* were found to be

associated with non-axenic *T. pseudonana* cultures (strain CCMP1335, National Center for Marine Algae and Microbiota (NCMA)), an isolate collected in 1958 from Moriches Bay (Long Island, New York) and maintained in continuous culture[31]. The genome was originally recovered from the Tara Oceans expedition data, SRA ERR599368, collected at station TARA_111 in the South Pacific Ocean[36,52]. In 2019, members of the *Mf105b01* genus were identified in the bacterial fraction associated with *T. pseudonana*[31], growing both under standard culture conditions and in the presence of oil, representing between 14 % and 25 % of the bacterial fraction strictly associated to the algae[31]. The placement of the Cl-2 species in the *Mf105b01* genus provides further support to consider that members belonging to this genus could all have close associations with phytoplankton. This is not yet well explored and warrants further investigation.

Physiological characterization of the MAGs evidenced differences in their growth rates, distribution in the diatom culture and antibiotic tolerance. An interesting observation is that the presence of all three bacteria in both *S. marinoi* FE7 and *T. rotula* FE80, and absence in *S. marinoi* FE60, cultures suggests that these may form a cohort that co-occur with a specific host. The microbial consortium showed a clear dependence on diatom organic matter when cultured in a liquid medium. In f/2 medium supplemented with sonicated diatoms, Cl-1 and Cl-8 needed at least four days, while Cl-2 needed at least seven to be detectable by qPCR, whereas little to no growth was observed in supplemented marine broth. This differs to what is present in the literature regarding the genus closest to Cl-1[30] and Cl-8[53–57], which were isolated and maintained on marine agar and marine broth, while all our attempts to culture Cl-8 on marine agar medium (including supplementation with diatom organic matter) were unsuccessful. This preference for diatom organic matter and their major distribution in the SA fraction of a classical diatom culture supports their dependence on molecules produced by the diatoms or the existence of a mutualistic relationship. Relationships based on nutrient exchange are extensively documented in literature. Diatoms release or passively exudate amino acids, carbohydrates and organic acids that are attractants for bacteria[58], while bacteria produce and release vitamins, ammonium or organosulfur molecules in exchange for organic matter released by diatoms[2,11,59]. Another difference among the three selected MAGs comes from the tolerance to β- lactam and aminoglycoside antibiotics. Cl-1 and Cl-2 were weakly tolerant to Ampicillin and Cefotaxime (β-lactam type antibiotics) and sensitive to streptomycin, whilst Cl-8 was tolerant to both β-lactam type antibiotics and Streptomycin (Str) in a dose-dependent manner. These results provide functional validation of the *in silico* prediction by RAST (β-lactamase) which identified the presence of 34 putative

genes for β-lactamase/cephalosporinases in the Cl-8 genome. This could be a common feature of the *Cyclobacteriaceae*[32], excluding the possible acquisition through horizontal transfer during culturing. However, with such a significant β-lactamase capacity, it could be considered surprising that C1-8 was not tolerant under all antibiotic mixtures tested. β-lactamases hydrolyse the β-lactam ring giving resistance to β-lactam antibiotics. However, Str can inhibit the β-lactamase enzymes, overcoming bacterial resistance[60]. The interplay between these antibiotics on the bacterial resistant machineries could explain the dose-dependent tolerance of Cl-8 to the antibiotic mix containing Str.

Functional annotation by PRISM[37] and antiSMASH[38] of pathways coding for secondary metabolites revealed the presence of biosynthetic clusters belonging to polyketides (PKS), non-ribosomal peptides (NRPs), terpenes (TP) and antibiotic classes of molecules. Moreover, the uncharacterized (novel) hybrid NRPS/PKS pathways annotated in the Cl-8 genome make it interesting for the search of molecules useful in pharmacological applications.

Among the other functional annotations, we focused our attention on the ectoine biosynthetic cluster identified in Cl-1. Ectoine is one of the compatible solutes able to accumulate in cells in response to external increases in osmotic pressure without compromising cell physiology. Other than serving as osmotic balancing agents, compatible solutes also show direct protective effects against freezing, high UV irradiation, high temperature and macromolecule denaturation under adverse conditions[61–64]. Not all of them are present in all organisms; in fact, their occurrence seems be spread out and sometimes species specific, with some functional specialization. As an example, neutral zwitterionic compatible solutes such as, among others, ectoine are frequent in many mesophilic bacteria, whilst negatively charged organic solutes such as dimyoinositolphosphate and mannosylglycerate (MG) have often been identified in hyper/thermophilic bacteria and archaea[65–68]. Based on these observations, we decided to verify the functionality of the ectoine biosynthetic cluster by measuring the ectoine produced in xenic *T. rotula* cultures under salinity stress. Our results indicated that ectoine is produced in the culture and increases under both hypo- and hyperosmotic shock. Interestingly, while axenic *T. rotula* cultures did not survive at salinities below 28 ppt, xenic cultures were able to grow at all the salinities tested and those exposed to 20 ppt salinity were able to overcome the stressful conditions and recovered their growth. The fact that axenic cultures were able to grow at 40 ppt salinity indicates that other mechanisms inside the diatom can compensate for high salinity stress. In xenic cultures, there is increased ectoine levels at both very low and high salinities (20 ppt and 40 ppt).

The capacity of Cl-1 to produce ectoine could resemble the ectoine production in *Halomonas elongata* which, when exposed to both low and high salinities, produces increased quantities of compatible solutes[69]. Thus, C1-1 could be complementing ectoine production under both hypo- and hyperosmotic conditions (20 ppt and 40 ppt). This phenomenon is actually used for the commercial production of ectoine[69] in a process called "bacterial milking". Looking at the ectoine levels reported in Figure 4.3, it looks like when fewer diatom cells are in the culture, more ectoine is produced. It would be of interest to determine if the activation of ectoine production in Cl-1 is regulated independently of the diatom host or through signalling from the diatom. Ectoine production in diatom culture was reported recently by Fenizia and co-workers who reported ectoine in *Thalassosira weissflogii* xenic cultures as being produced by the associated bacteria with a very small contribution by the algae[70]. In addition, Vallet and co-workers used the presence of ectoine and absence of choline as metabolic markers for the presence of bacteria in association with algae (*Ulva*)[71].

Cl-1 abundance in the *T. rotula* xenic cultures in the SA fraction on the 8th day of growth correlated with ectoine production. The highest copy number corresponded with higher ectoine concentrations (36 ppt and 40 ppt), decreasing with decreasing salinity, similarly to ectoine levels. On the other hand, at the 3rd day of growth, Cl-1 was undetectable, except at 20 ppt and 24 ppt salinities, which coincides with the higher ectoine concentrations measured. As another ectoine biosynthetic cluster was found in the total metagenome (Supplementary Table 4.12), we hypothesize that other bacterial species can contribute to the ectoine production in the culture that also supports microalgae growth at 40 ppt salinity. Alternatively, since Cl-1 was undetectable early on in the growth cycle, it could suggest that, if it is responsible for most of the ectoine production, the few cells that are present are expending most of their resources to produce ectoine.

In conclusion, with our work, we identified three new bacterial species associated to at least two diatom species, i.e., *Skeletonema marinoi* and *Thalassiosira rotula*. These bacteria showed diatom clonal preferences and the need for a strict association with them. They appear to contribute to the diatom adaptive response with the supply of osmolyte and antibiotic resistance and have shown potential for the production of biotechnologically interesting secondary metabolites. This work serves as a preliminary characterization of these bacteria that opens the door to further exploration of the diatom–bacteria relationship and to the biotechnological development of new and possibly pharmaceutically relevant molecules.

4.5 Materials and Methods

4.5.1 Strain Information and Culturing

Thalassiosira rotula strain FE80 was isolated in 2011 in the Gulf of Naples (40°48.5′ N, 14°15′ E), Mediterranean Sea (Italy) and the *Skeletonema marinoi* strain FE7 was isolated from phytoplankton samples collected during diatom blooms in the northern Adriatic Sea in 1997[72]. Clonal cultures were established by isolating single cells or short chains from phytoplankton net samples collected from the surface layer of the water column. Cultures were grown in sterile filtered oligotrophic seawater at 36 ppt amended with f/2[39] nutrients and maintained at a temperature of 20 °C, at 12:12 h light:dark cycle, and with a photon flux of 100 μmol photons m^{-2} s^{-1}. To remove excess bacteria from *T. rotula* FE80 diatom cultures, 250 mL of exponentially growing culture was inoculated in f/2 medium containing final concentrations of 2 mg/L Streptomycin (A1852, PanReac AppliChem), 2 mg/L Penicillin (A1837, PanReac AppliChem), 10 mg/L Ampicillin (A0839, PanReac AppliChem) and allowed to grow for six days under standard growth conditions. After six days, the culture was supplied with fresh f/2 media, containing antibiotics, to a final volume of 1 L and cultivation continued for seven more days. Cells were collected on a 1.2 μm RAWP membrane filter (Millipore, Burlington, Massachusetts, USA). The filter was rinsed with 1.5 mL seawater to detach the cells that were further collected into Eppendorf tubes by centrifugation at 6000 rpm at 4 °C for 5 min (Eppendorf Centrifuge 5810 R, Eppendorf, Amburg, Germany). The pellet free of supernatant was immediately frozen in liquid nitrogen and stored at -80 °C.

For ectoine and Cl-1 quantification, 5000 cells/mL were inoculated in 2 L of f/2 medium at different salinities, each in triplicate. These media were made as follow: 2 L of f/2 medium prepared with sterile seawater at 38 ppt was supplemented with 4 g of NaCl to obtain a final salinity of 40 ppt. Starting from this medium, the other salinity points were obtained by dilution with autoclaved Milli-Q water. All the salinities were checked before and after sterilization using a portable refractometer. After three days, 500 mL of each culture were collected by filtration on 3 μm polycarbonate filters (Millipore, Burlington, Massachusetts, USA). After eight days of growth, cells from 500 mL of culture were collected by centrifugation at 3000 rpm for 10 min at 4 °C. All the samples collected at the two time points were immediately frozen in liquid nitrogen and stored at −80 °C until gDNA extraction (see “DNA extraction, primer design and PCR” paragraph for details) or ectoine quantification (see “Ectoine quantification by UPLC-MS/MS” paragraph for details).

4.5.2 Ectoine Quantification by UPLC-MS/MS

Ectoine was measured applying a newly developed quantitative UPLC-mass spectrometry-based methodology by using an internal standard calibration approach. LC-MS analyses were performed on an Acquity UPLC System (Waters, Milford, Massachusetts, USA) coupled to a 3200 API Triple Quadrupole mass spectrometer (SCIEX, Foster City, California, USA) with a Turbo VTM interface equipped with a turbo ion spray probe used in positive ion mode. The chromatographic analysis was developed on a Luna Omega Polar (100 × 2.1 mm, i.d. 1.6 μm, Phenomenex, Bologna, Italy) by using as eluent A water 0.1 % FA and eluent B ACN 0.1 % FA; a gradient elution was applied from A 100 % to 70 % in 3 min, returning back at 100 % A in 0.1 min. A re-equilibration step of 1 min was included before successive runs. The flow rate was set at 0.45 mL/min. Separations were performed at a temperature of 35 °C, the autosampler was maintained at 10 °C and injections were of 1 μL. A multiple reaction monitoring (MRM) experiment was used to collect data, by setting the following source parameters: curtain gas (N_2), 20 psi; ion source gas (GS1), 55 psi; turbogas (GS2), 70 psi; desolvation temperature, 550 °C; collision activated dissociation gas (CAD), 4 a.u.; and ion spray voltage, 5500 V. For ectoine, the following transitions were monitored: 143 > 68 (quantifier) and 143 > 97 *m/z* (qualifier). A calibration curve was constructed in the range 2.5–250 ng/mL. Phenylalanine-d5 (Phe-d5) was used as internal standard (IS, 1 μg/mL) and monitored at 171 > 125 *m/z*. Analyst software (version 1.6.2; SCIEX) was used for data acquisition. Multiquant software (version 2.1.1, SCIEX) was used for quantitative analysis.

Cell pellets or filters from 500 mL cultures, in triplicate for each experimental condition (see "Strain and cell cultures" paragraph for details), including a control batch, were extracted with MeOH (3 × 5 mL). Phe-d5 was added as IS to obtain a concentration of 1 μg/mL in the final sample. Methanolic extracts were reconstituted in 90:10 (water/MeOH) before LC-MS analysis.

4.5.3 DNA Extraction and Sequencing

The collected *T. rotula* FE80 was used to extract mixed genomic DNA (mgDNA) from both algae and the associated bacteria according to the DNeasy Plant Maxi Kit (Cat. No. 68163, Qiagen, Hilden, Germany) handbook. One nanogram of total genomic DNA was used to prepare one NexteraXT library (FC-131-1024; Illumina San Diego, California, USA). The resultant library was sequenced on the Illumina MiSeq platform using a MiSeq

Reagent kit V2 (MS-102-2003; 500 cycle) to generate paired end reads (2 × 250 bp), including a 10 % PhiX Control v3 (FC-110-3001) as per the manufacturer's recommendation.

4.5.4 Metagenome Sequencing, Assembly, Binning and Annotation

Reads were processed by first mapping to the phi-X174 genome and removing reads that were not correctly demultiplexed. Only reads in pairs were taken further for assembly and singletons discarded. Next, reads containing any ambiguous nucleotides were re- moved. The remaining read pairs were merged. The merged and un-merged pairs, as well as singletons from the ambiguous nucleotide trim were co-assembled.

Assembly was performed using CLC Genomics Workbench version v7.5.1 (http://www.clcbio.com, Qiagen, Hilden, Germany) with the length fraction set at 0.8 and similarity fraction at 0.9. Mismatch insertion and deletion costs were left at their default values. The "global alignment" and "update contigs" settings were enabled, while scaffolding was turned off. The word and bubble sizes were set to automatic mode as well as paired distance detection.

The binning of metagenomic contigs was performed using MyCC (https://sourceforge.net/projects/sb2nhri/files/MyCC/, version 2017)[29] with 5 mer and meta settings, while completeness and contamination of metagenome-assembled genomes (MAGs) as well as genome quality were determined using CheckM using the lineage-specific workflow and default parameters[73].

Annotation of MAGs was performed using Prokka v1.1.2[74] through the KBase[75] online analysis platform as well as Rapid Annotations using Subsystems Technology (RAST; https://rast.nmpdr.org/). The quality of the assembled MAGs was determined using CheckM as required by the MIMAG (Minimum Information about a Metagenome-Assembled Genome) guidelines[76].

Whole genome phylogeny was determined using the Genome Taxonomy Database tool kit (GTDB-Tk) release 89 implemented in KBase[48] and the output maximum likelihood tree visualized using iTOL (https://itol.embl.de/). tRNAs were detected using the ARAGORN webserver[77]. MAGs were analysed for the presence of biosynthetic gene clusters using PRISM[37] and antiSMASH v5.1.1[38] (access date: from April to June 2020).

4.5.5 SA and FL Bacterial Fractions Separation

To obtain the two bacterial fractions, xenic S. marinoi FE7 cultures were subjected to two subsequent centrifugation steps: 3000 rpm for 10 min at 4 °C to collect the strictly associated fraction (SA) and 3000 rpm for 30 min at 4 °C in falcon followed by 10.000 rpm for 30 min at 4 °C in Eppendorf tubes (Beckman Avanti™ 30 Centrifuge, Beckman Coulter Inc., Brea, California, USA) to collect the free living (FL) bacteria. The collected samples were immediately frozen in liquid nitrogen and stored at −80 °C until use.

4.5.6 Reintroduction of Bacteria in *T. rotula* FE80 Cultures

The axenic *T. rotula* FE80 culture was provided by CNR-ICB (Pozzuoli, NA, Italy) laboratory. A total of 100 µL of FL bacteria prepared from a xenic *S. marinoi* FE7 culture and stored in a cryovial with 40 % glycerol were inoculated in a 30 mL axenic *T. rotula* FE80 culture at exponential growth phase. The culture was maintained at a temperature of 20 °C, at 12:12 h light:dark cycle, and with a photon flux of 100 µmol photons m^{-2} s^{-1} for 3 days, which was then refreshed by a 1:30 dilution with fresh f/2 medium. When stabilized, the culture was refreshed weekly in the same manner and the presence of the bacteria was periodically checked with PCR on the SA-FL fraction gDNA preparations (see "DNA extraction, primer design and PCR" paragraph for details).

4.5.7 Cl-1, Cl-2 and Cl-8 Growth Characterization

Xenic cultures were harvested after 0, 4, and 7 days of growth and bacterial fractions were separated as described before. For the antibiotic tolerance test, exponentially growing xenic S. marinoi FE7 cultures were diluted 1:10, in a final volume of 100 mL, with f/2 medium containing different combinations of the following antibiotics, Ampicillin (Amp) (A0839, PanReac AppliChem), Cefotaxime (Ctx) (C7039, Sigma-Aldrich, St. Louis, Missouri, USA) and Streptomycin (Str) (A1852, PanReac AppliChem), and allowed to grow for 72 h. At the end of the 72 h, cultures were again diluted 1:10 into f/2 medium, antibiotic treated as above and allowed to grow for 7 days.
Two antibiotic combinations and concentrations were used:

- ACx2: 500 mg/mL Amp; 10 mg/mL Ctx;
- ACSx2: 500 mg/mL Amp; 10 mg/mL Ctx; 50 mg/mL Str;
- ACx1: 250 mg/mL of Amp; 5 mg/mL Ctx;

- ACSx1: 250 mg/mL of Amp; 5 mg/mL Ctx; 25 mg/mL Str.

At the end of the treatment, cultures were harvested as described before.

4.5.8 Liquid Bacterial Cultures

FL bacteria (collected as described in the paragraph "SA and FL bacterial fractions separation") were diluted 1:10 in a final volume of 150 mL of two different media: f/2 medium supplemented with sonicated diatoms or Marine Broth (2216 DIFCO Laboratories, Detroit, Michigan, USA) supplemented with a 1:1 ratio of spent medium. The spent medium was collected from a *S. marinoi* FE7 culture at the 7th day of growth and then was filtered onto 0.22 μm pore size filters (SLGS033SB, Millipore, Burlington, Massachusetts, USA). A total of 150 mL of f/2 medium supplemented with sonicated diatoms was obtained as follow. After 7 days of growth, 150 mL of cultures were harvested by centrifugation at 3000 rpm, 10 min, 4 °C. The resulting supernatant was used to collect the FL bacterial fraction. The pellet was resuspended in 2 mL of f/2 and sonicated on ice for 3 pulses of 30 s at 40 hrz, then 30 s on ice. Then 0.5 mL of the sonicated material was resuspended in 149.5 mL of f/2, resulting in a ¼ of the final diatom cell culture concentration. The remaining sonicate material, 1.5 mL, was also resuspended in 148.5 mL of f/2, obtaining a second concentration corresponding to ¾ of the final diatom cell culture concentration. The ¼ sonicate concentration was also used for cultures in which 100 μL of the undiluted FL bacteria was inoculated into 8 mL of f/2 supplemented with sonicate treated with DNAse (1 μL for each 30 μL of sonicate) and incubated at 37 °C for 30 min before addition to the f/2 medium. All the cultures were maintained at a temperature of 20 °C, at 12:12 h light:dark cycle, and with a photon flux of 100 μmol photons m^{-2} s^{-1}.

4.5.9 Solid Media Culturing

FL bacteria, isolated as described before, were serially 1:10 diluted in f/2 medium and spread on different types of solid media: Marine Broth with 1.4 % agar (European Bacteriological Agar, Condalab, Madrid, Spain) and a second medium prepared with f/2 containing 1.4 % agar and once solidified was subsequently covered with sonicated diatoms (¼ concentration) treated with DNAse as described above. These media were also prepared by replacing the agar with 0.8 % Gellan-Gum (Gelrite, SERVA Electrophoresis GmbH, Heidelberg, Germany) as the gelling agent. Gellan-Gum was

dissolved in Milliq water to a 2× concentration (1.6 gr in 100 mL corresponding to 0.8 %) and autoclaved. Soon after, one volume of Gellan-Gum was mixed with one volume of Marine Broth or f/2 (2×, previously autoclaved) under stirring at 80 °C, and poured into petri dishes (25 mL for each plate). FL bacteria were spread on top with a metallic T- shaped spreader or streaked with a wire loop and allowed to grow at temperatures of 18 °C and 28 °C, with a 12:12 h light:dark cycle and in complete darkness, from 48 h to 10 days. Then, 10 to 50 colonies from Marine Broth plates were screened by PCR to search for Cl-8 positive colonies (see "DNA extraction, primer design and PCR" paragraph). Colonies on algae-supplemented f/2 (sonicated diatoms) plates were collected by scraping since they never reached a size to be easily collected. The material collected in this way was treated as a colony for the following PCR analysis.

4.5.10 DNA Extraction, Primer Design and PCR

Genomic DNA (gDNA) was extracted by vortexing diatom/bacterial pellets with 0.4 g of 0.2–0.3 mm glass beads (G1277-Sigma-Aldrich, St. Louis, Missouri, USA), 0.5 mL of phenol (A1153,0100, PanReac AppliChem) and 0.5 mL of TE buffer (10 mM Tris HCl pH 7.6 and 1 mM EDTA pH 8.0). DNA extraction was performed as reported in[78]. SA fraction gDNA was extracted from xenic culture pellets of mixed diatom and bacterial cells. FL fraction gDNA was extracted from the pellet deriving from centrifugation of the xenic culture supernatant. Primer3 software V. 0.4.0[79,80] was used to design PCR primers specific for sequences on the Cl-1, Cl-2 and Cl-8 contigs (Table 4.5), while 16S rRNA universal oligo primers (E9F/U1510R) were taken from[81,82]. Oligo sequences are listed below (Table 4.5):

Table 4.5. Oligo sequences. Abbreviations: Tm=Temperature of melting;

Primer Name	Forward Sequence	Reverse Sequence	Tm (°C)	Amplicon length (bp)
C8_c608	GCTCCAGTGTTTTAACCGG	CCATCTATTCTGCCGACC	62.1/60.7	251
C8_c450	TCGCCAATACTGATTATGCT	GTCGTAGTTCCTAAGGTCAC	59.7/55.3	169
C1_c182	CTGATCTGTTATATGATGCGGA	GACATGACAGTGATGCATTG	61.3/60.2	161
C2_c82	GTATCAATATCGGGCAGTGT	CGATATTCCAAATGTGAGCG	58.9/63.0	243

E9F/U1510R(16s)	GAGTTTGATCCTGGC TCAG	GGCTTACCTTGTT ACGACTT	60/53.1	1500

Each sequence was tested by PCR in a 25 µL final reaction volume with 2.5 µL of 10× PCR reaction buffer (Roche, Basel, Switzerland), 2.5 µL of 2 mM dNTP, 0.3 µL of 5 U/µL Taq (Roche, Basel, Switzerland), 0.25 µL of 1 % DMSO, 1 µL 10 µM of each oligo, 1 µL of 50 ng/µL DNA template and nuclease-free water up to 25 µL. The PCR was conducted using the T100 Thermal cycler (Bio-Rad Laboratories, Hercules, California, USA) and the PCR program consisted of a denaturation step at 95 °C for 3 min, 35 cycles at 95 °C for 30 s, 55 °C 30 s, 72 °C for 30 s, and a final extension step at 72 °C for 7 min. For PCR on colonies, the template was obtained as follows: each colony or scraped colony was resuspended in 20 µL of TE buffer (10 mM TrisCl pH 7.6 and 1 mM EDTA pH 8.0), heated at 95 °C for 3 min and centrifuged at 12.000 rpm for 2 min (Eppendorf Mini Spin Centrifuge, Eppendorf, Amburg, Germany). Then, 1 µL of the supernatant obtained was tested by PCR using the same temperature cycle and reaction mix described above. Amplified PCR products were analysed by 1.5 % agarose gel electrophoresis and the resulting bands were excised from the gel and extracted according to the GenElute Gel Extraction Kit protocol (Sigma-Aldrich, St. Louis, Missouri, USA). See Supplementary Information for uncropped gel images in Supplementary Figures 4.2 a, 4.3, 4.6 and 4.8.
Sequences were obtained by BigDye Terminator Cycle Sequencing Technology (Applied Biosystems, Waltham, Massachusetts, USA) and purified using the Agencourt CleanSEQ Dye terminator removal Kit (Agencourt Bioscience Corporation, Beverly, Massachusetts, USA) in an automated robotic station Biomek FX (Beckman Coulter Inc., Brea, California, USA). Products were analysed on the Automated Capillary Electrophoresis Sequencer 3730 DNA Analyzer (Applied Biosystems, Waltham, Massachusetts, USA). Sequences were analysed with Chromas v 2.6.6 (http://technelysium.com.au/wp/chromas).
Identity of the retrieved sequences was verified with the standalone setup BLAST+ v 2.11.0 (https://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/). See Supplementary Information for the sequences of the cluster-specific genes and 16S rRNA gene.

4.5.11 Quantitative PCR Analysis

Quantitative PCR (qPCR) experiments were performed in a MicroAmp Optical 384- Well reaction plate (Applied Biosystems, Waltham, Massachusetts, USA) with Optical

Adhesive Covers (Applied Biosystems, Waltham, Massachusetts, USA) in a Viia7 Real Time PCR System (Applied Biosystem, Waltham, Massachusetts, USA). Five serial dilutions of mixed DNAs were used to obtain the standard curve. The PCR volume for each sample was 10 μL, with 5 μL of SensiFASTTM SYBR® Lo-ROX Kit (BIO_94020, Bioline Reagents Ltd., London, UK), 1 μL of DNA template (4 ng dilution of each template) and 4 μL of 0.7 μM oligo mix (forward and reverse primers). The program reaction used was: 95 °C for 20s, 40 cycles of 95 °C for 1s and 60 °C for 20s. The program was set to reveal the melting curve of each amplicon from 60 °C to 95 °C and read every 0.5 °C. Single peaks for all genes confirmed gene-specific amplification and the absence of primer-dimers. All RT-qPCR reactions were carried out in triplicate to capture intra-assay variability. Each assay included three no-template negative controls for each primer pair. The quantification of the samples was conducted using the standard curve with the qPCR software QuantStudio Real-Time PCR (Applied Biosystems, Waltham, Massachusetts, USA). See Supplementary Information for standard and melt curves in Supplementary Figures 4.4, 4.5 and 4.9 and primer efficiency in Supplementary Tables 4.6, 4.8 and 4.11.

4.6 References

1. Amin, S.A.; Parker, M.S.; Armbrust, E.V. Interactions between Diatoms and Bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 667–684, (2012).

2. Amin, S.; Hmelo, L.R.; Van Tol, H.M.; Durham, B.; P. Carlson, L.T.; Heal, K.R.; Morales, R.L.; Berthiaume, C.T.; Parker, M.S.; Djunaedi, B.; et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* **522**, 98–101, (2015).

3. Liu, J.; Lewitus, A.J.; Brown, P.; Wilde, S.B. Growth-promoting effects of a bacterium on raphidophytes and other phytoplankton. *Harmful Algae* **7**, 1–10, (2007).

4. Sison-Mangus, M.P.; Jiang, S.; Tran, K.N.; Kudela, R.M. Host-specific adaptation governs the interaction of the marine diatom, *Pseudo-nitzschia* and their microbiota. *ISME J.* **8**, 63–76, (2013).

5. Grossart, H.P.; Kiørboe, T.; Tang, K.W.; Allgaier, M.; Yam, E.M.; Ploug, H. Interactions between marine snow and heterotrophic bacteria: aggregate formation and microbial dynamics. *Aquat. Microb. Ecol.* **42**, 19–26, (2006).

6. Gärdes, A.; Iversen, M.H.; Grossart, H.-P.; Passow, U.; Ullrich, M.S. Diatom-associated bacteria are required for aggregation of *Thalassiosira weissflogii*. *ISME J.* **5**, 436–445, (2010).

7. Field, C.B.; Behrenfeld, M.J.; Randerson, J.T.; Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science* **281**, 237–240, (1998).

8. Armbrust, E.V. The life of diatoms in the world′s oceans. *Nature* **459**, 185–192, (2009).

9. Malviya, S.; Scalco, E.; Audic, S.; Vincent, F.; Veluchamy, A.; Poulain, J.; Wincker, P.; Iudicone, D.; de Vargas, C.; Bittner, L.; et al. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. USA* **113**, E1516–E1525, (2016).

10. Grossart, H.-P.; Czub, G.; Simon, M. Algae-bacteria interactions and their effects on aggregation and organic matter flux in the sea. *Environ. Microbiol.* **8**, 1074–1084, (2006).

11. Croft, M.T.; Lawrence, A.D.; Raux-Deery, E.; Warren, M.J.; Smith, A.G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90–93, (2005).

12. Amin, S.A.; Green, D.H.; Hart, M.C.; Kupper, F.C.; Sunda, W.G.; Carrano, C.J. Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc. Natl. Acad. Sci. USA* **106**, 17071–17076 (2009).

13. Grossart, H.-P.; Levold, F.; Allgaier, M.; Simon, M.; Brinkhoff, T. Marine diatom

species harbour distinct bacterial communities. *Environ. Microbiol.*, **7**, 860–873, (2005).
14. Mönnich, J.; Tebben, J.; Bergemann, J.; Case, R.; Wohlrab, S.; Harder, T. Niche-based assembly of bacterial consortia on the diatom *Thalassiosira rotula* is stable and reproducible. *ISME J.* **14**, 1614–1625 (2020).
15. Majzoub, M.E.; Beyersmann, P.G.; Simon, M.; Thomas, T.; Brinkhoff, T.; Egan, S. *Phaeobacter inhibens* controls bacterial community assembly on a marine diatom. *FEMS Microbiol. Ecol.* **95**, (2019).
16. Teeling, H.; Fuchs, B.M.; Becher, D.; Klockow, C.; Gardebrecht, A.; Bennke, C.M.; Kassabgy, M.; Huang, S.; Mann, A.J.; Waldmann, J.; et al. Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science* **336**, 608–611, (2012).
17. Dittmann, K.K.; Sonnenschein, E.C.; Egan, S.; Gram, L.; Bentzon-Tilia, M. Impact of *Phaeobacter inhibens* on marine eukaryote-associated microbial communities. *Environ. Microbiol. Rep.* **11**, 401–413, (2018).
18. Quigley, K.M.; Roa, C.A.; Torda, G.; Bourne, D.G.; Willis, B.L. Co-dynamics of Symbiodiniaceae and bacterial populations during the first year of symbiosis with *Acropora tenuis* juveniles. *MicrobiologyOpen* **9**, e959, (2019).
19. Egan, S.; Thomas, T.; Kjelleberg, S. Unlocking the diversity and biotechnological potential of marine surface associated microbial communities. *Curr. Opin. Microbiol.* **11**, 219–225, (2008).
20. Guedes, A.C.; Amaro, H.M.; Malcata, F.X. Microalgae as sources of high added-value compounds-a brief review of recent work. *Biotechnol. Prog.* **27**, 597–613, (2011).
21. Trindade, M.; van Zyl, L.J.; Navarro-Fernández, J.; AbdElrazak, A. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Front. Microbiol.* **6**, 890, (2015).
22. Lindequist, U. Marine-Derived Pharmaceuticals: Challenges and Opportunities. *Biomol. Ther.* **24**, 561–571, (2016).
23. Krawiec, R.W. Autecology and clonal variability of the marine centric diatom *Thalassiosira rotula* (Bacillariophyceae) in response to light, temperature and salinity. *Mar. Biol.* **69**, 79–89 (1982).
24. Di Dato, V.; Di Costanzo, F.; Barbarinaldi, R.; Perna, A.; Ianora, A.; Romano, G. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Sci. Rep.* **9**, 9893, (2019).
25. Ianora, A.; Poulet, S.A.; Miralto, A.; Grottoli, R. The diatom *Thalassiosira rotula* affects reproductive success in the copepod *Acartia clausi*. *Mar. Biol.* **125**, 279–286,

(1996).

26. Qin, J.G.; D′Antignana, T.; Zhang, W.; Franco, C. Discovery of antimicrobial activities of a marine diatom *Thalassiosira rotula*. *Afr. J. Microbiol. Res.* **7**, 5687–5696, (2013).

27. Garcia, N.S.; Yung, C.-M.; Davis, K.M.; Rynearson, T.; Hunt, D.E. Draft genome sequences of three bacterial isolates from cultures of the marine diatom *Thalassiosira rotula*. *Genome Announc.* **5**, e00316-17, (2017).

28. Grossart, H. P.; Simon, M. Interactions of planktonic algae and bacteria: effects on algal growth and organic matter dynamics. *Aquat. Microb. Ecol.* **47**, 163–176, (2007).

29. Lin, H. H.; Liao, Y.-C. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci. Rep.* **6**, 24175, (2016).

30. Park, S.; Won, S. M.; Kim, H.; Park, D. S.; Yoon, J. H. *Aestuariivita boseongensis* gen. nov., sp. nov., isolated from a tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* **64**, 2969–2974, (2014).

31. Kamalanathan, M.; Chiu, M. H.; Bacosa, H.; Schwehr, K.; Tsai, S. M.; Doyle, S.; Yard, A.; Mapes, S.; Vasequez, C.; Bretherton, L.; et al. Role of polysaccharides in diatom *Thalassiosira pseudonana* and its associated bacteria in hydrocarbon presence. *Plant Physiol.* **180**, 1898–1911 (2019).

32. Pinnaka, A.K.; Tanuku, N.R.S. The family Cyclobacteriaceae. In *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, pp. 551–575, (2014) ISBN 978-3-642-38954-2.

33. Chaumeil, P.A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927, (2019).

34. Aziz, R.K.; Bartels, D.; Best, A.A.; DeJongh, M.; Disz, T.; Edwards, R.A.; Formsma, K.; Gerdes, S.; Glass, E.M.; Kubal, M.; et al. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genom.* **9**, 75, (2008).

35. Shoguchi, E.; Shinzato, C.; Kawashima, T.; Gyoja, F.; Mungpakdee, S.; Koyanagi, R.; Takeuchi, T.; Hisata, K.; Tanaka, M.; Fujiwara, M.; et al. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr. Biol.* **23**, 1399–1408, (2013).

36. Parks, D.H.; Rinke, C.; Chuvochina, M.; Chaumeil, P.-A.; Woodcroft, B.J.; Evans, P.N.; Hugenholtz, P.; Tyson, G.W. Recovery of nearly 8,000 metagenome-assembled

genomes substantially expands the tree of life. *Nat. Microbiol.* **2017**, *2*, 1533–1542, doi:10.1038/s41564-017-0012-7.

37. Skinnider, M.A.; Merwin, N.J.; Johnston, C.W.; Magarvey, N.A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **45**, W49–W54, (2017).

38. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

39. Guillard, R.R.L. Culture of phytoplankton for feeding marine invertebrates. In *Culture of Marine Invertebrate Animals: Proceedings-1st Conference on Culture of Marine Invertebrate Animals Greenport*; Smith, W.L., Chanley, M.H., Eds.; Springer US: Boston, MA, USA, pp. 29–60 (1975) ISBN 978-1-4615-8714-9.

40. Windler, M.; Tsymbal, D.; Kryvenda, A.; Straile, D.; Gruber, A.; Kroth, P. Influence of bacteria on cell size development and morphology of cultivated diatoms. *Phycol. Res.* **62**, 269–281 (2014).

41. Cirri, E.; Vyverman, W.; Pohnert, G. Biofilm interactions-bacteria modulate sexual reproduction success of the diatom *Seminavis robusta*. *FEMS Microbiol. Ecol.* **94**, (2018).

42. Behringer, G.; Ochsenkühn, M.A.; Fei, C.; Fanning, J.; Koester, J.A.; Amin, S.A. Bacterial communities of diatoms display strong conservation across strains and time. *Front. Microbiol.* **9**, 659, (2018).

43. García-López, M.; Meier-Kolthoff, J.P.; Tindall, B.J.; Gronow, S.; Woyke, T.; Kyrpides, N.C.; Hahnke, R.L.; Göker, M. Analysis of 1,000 type-strain genomes improves taxonomic classification of Bacteroidetes. *Front. Microbiol. 10*, 2083, (2019).

44. Li, G.; Lai, Q.; Du, Y.; Liu, X.; Sun, F.; Shao, Z. *Aestuariivita atlantica* sp. nov., isolated from deep-sea sediment. *Int. J. Syst. Evol. Microbiol.* **65**, 3281–3285, (2015).

45. Wirth, J.S.; Whitman, W.B. Phylogenomic analyses of a clade within the Roseobacter group suggest taxonomic reassignments of species of the genera *Aestuariivita*, *Citreicella*, *Loktanella*, *Nautella*, *Pelagibaca*, *Ruegeria*, *Thalassobius*, *Thiobacimonas* and *Tropici- bacter*, and the proposal of six novel genera. *Int. J. Syst. Evol. Microbiol.* **68**, 2393–2411, (2018).

46. Olsen, J.D.; Martin, E.C.; Hunter, C.N. The PufX quinone channel enables the light-harvesting 1 antenna to bind more carotenoids for light collection and photoprotection. *FEBS Lett.* **591**, 573–580 (2017).

47. Qian, P.; Martin, E.C.; Ng, I.W.; Hunter, C.N. The C-terminus of PufX plays a key role in dimerisation and assembly of the reaction center light-harvesting 1 complex from

Rhodobacter sphaeroides. *Biochim. Biophys. Acta Bioenerg.* **1858**, 795–803, (2017).
48. Branduardi, P.; Sauer, M. Microbial carbon dioxide fixation: new tricks for an old game. *FEMS Microbiol. Lett.* **365**, (2017).
49. Tomasch, J.; Gohl, R.; Bunk, B.; Diez, M.S.; Wagner-Döbler, I. Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *ISME J.* **5**, 1957–1968 (2011).
50. Imhoff, J.F.; Rahn, T.; Künzel, S.; Neulinger, S.C. Photosynthesis is widely distributed among Proteobacteria as demonstrated by the phylogeny of PufLM reaction center proteins. *Front. Microbiol.* **8**, 2679 (2018).
51. Simon, M.; Scheuner, C.; Meier-Kolthoff, J.P.; Brinkhoff, T.; Wagner-Döbler, I.; Ulbrich, M.; Klenk, H.-P.; Schomburg, D.; Petersen, J.; Göker, M. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J.* **11**, 1483–1499 (2017).
52. Sunagawa, S.; Coelho, L.P.; Chaffron, S.; Kultima, J.R.; Labadie, K.; Salazar, G.; Djahanschiri, B.; Zeller, G.; Mende, D.R.; Alberti, A.; et al. Ocean Plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359, (2015).
53. Nedashkovskaya, O.I.; Kim, S.B.; Lysenko, A.M.; Park, M.S.; Mikhailov, V.V.; Bae, K.S.; Park, H.Y. *Roseivirga echinicomitans* sp. nov., a novel marine bacterium isolated from the sea urchin *Strongylocentrotus intermedius*, and emended description of the genus *Roseivirga*. *Int. J. Syst. Evol. Microbiol.* **55**, 1797–1800, (2005).
54. Nedashkovskaya, O.I.; Kim, S.B.; Lee, D.H.; Lysenko, A.M.; Shevchenko, L.S.; Frolova, G.M.; Mikhailov, V.V.; Lee, K.H.; Bae, K.S. *Roseivirga ehrenbergii* gen. nov., sp. nov., a novel marine bacterium of the phylum 'Bacteroidetes', isolated from the green alga *Ulva fenestrata*. *Int. J. Syst. Evol. Microbiol.* **55**, 231–234 (2005).
55. Lau, S.C.K.; Tsoi, M.M.Y.; Li, X.; Plakhotnikova, I.; Dobretsov, S.; Wu, M.; Wong, P.K.; Pawlik, J.R.; Qian, P.-Y. Description of F*abibacter halotolerans* gen. nov., sp. nov. and *Roseivirga spongicola* sp. nov., and reclassification of *[Marinicola] seohaensis* as *Ro- seivirga seohaensis* comb. nov. *Int. J. Syst. Evol. Microbiol.* **56**, 1059–1065, (2006).
56. Jung, Y.-T.; Park, S.; Lee, J.-S.; Yoon, J.-H. *Roseivirga maritima* sp. nov., isolated from seawater. *Int. J. Syst. Evol. Microbiol.* **66**, 2664–2670 (2016).
57. Selvaratnam, C.; Thevarajoo, S.; Goh, K.M.; Chan, K.-G.; Chong, C.S. Proposal to reclassify *Roseivirga ehrenbergii* (Nedash- kovskaya et al., 2008) as *Roseivirga seohaensis* comb. nov., description of *Roseivirga seohaensis* subsp. *aquiponti* subsp. nov. and emendation of the genus *Roseivirga*. *Int. J. Syst. Evol. Microbiol.* **66**, 5537–5543 (2016).

58. Seymour, J.R.; Amin, S.A.; Raina, J.-B.; Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton– bacteria relationships. *Nat. Microbiol.* **2**, 17065 (2017).

59. Suleiman, M.; Zecher, K.; Yücel, O.; Jagmann, N.; Philipp, B. Interkingdom cross-feeding of ammonium from marine methylamine-degrading bacteria to the diatom *Phaeodactylum tricornutum*. *Appl. Environ. Microbiol.* **82**, 7113–7122 (2016).

60. Maryam, L.; Khan, A.U. A Mechanism of synergistic effect of streptomycin and cefotaxime on CTX-M-15 type β-lactamase producing strain of *E. cloacae:* A first report. *Front. Microbiol.* **7**, 2007 (2016).

61. Graf, R.; Anzali, S.; Buenger, J.; Pfluecker, F.; Driller, H. The multifunctional role of ectoine as a natural cell protectant. *Clin. Dermatol.* **26**, 326–333, (2008).

62. Pastor, J.M.; Salvador, M.; Argandoña, M.; Bernal, V.; Reina-Bueno, M.; Csonka, L.N.; Iborra, J.L.; Vargas, C.; Nieto, J.J.; Cánovas, M. Ectoines in cell stress protection: Uses and biotechnological production. *Biotechnol. Adv.* **28**, 782–801, (2010).

63. Moghaieb, R.E.; Nakamura, A.; Saneoka, H.; Fujita, K. Evaluation of salt tolerance in ectoine-transgenic tomato plants (*Lycopersicon esculentum*) in terms of photosynthesis, osmotic adjustment, and carbon partitioning. *GM Crop.* **2**, 58–65, (2011).

64. Bownik, A.; Stępniewska, Z. Ectoine as a promising protective agent in humans and animals. *Arh. Hig. Rada Toksikol.* **67**, 260–265 (2016).

65. Brown, A.D. Compatible solutes and extreme water stress in Eukaryotic micro-organisms. *Adv. Microb. Physiol.* **17**, 181– 242 (1978).

66. Kirst, G.O. Salinity tolerance of Eukaryotic marine algae. *Annu. Rev. Plant Biol.* **41**, 21–53 (1990).

67. Pade, N.; Hagemann, M. Salt acclimation of Cyanobacteria and their application in biotechnology. *Life* **5**, 25–49 (2015).

68. Empadinhas, N.; da Costa, M.S. Diversity, biological roles and biosynthetic pathways for sugar-glycerate containing compatible solutes in bacteria and archaea. *Environ. Microbiol.* **13**, 2056–2077 (2011).

69. Sauer, T.; Galinski, E.A. Bacterial milking: a novel bioprocess for production of compatible solutes. *Biotechnol. Bioeng*. **57**, 306–313 (1998).

70. Fenizia, S.; Thume, K.; Wirgenings, M.; Pohnert, G. Ectoine from bacterial and algal origin is a compatible solute in microalgae. *Mar. Drugs* **18**, 42 (2020).

71. Vallet, M.; Kaftan, F.; Grabe, V.; Ghaderiardakani, F.; Fenizia, S.; Svatoš, A.; Pohnert, G.; Wichard, T. A new glance at the chemosphere of macroalgal–bacterial interactions: in situ profiling of metabolites in symbiosis by mass spectrometry. *Beilstein*

J. Org. Chem. **17**, 1313–1322 (2021).

72. Gerecht, A.; Romano, G.; Ianora, A.; D'Ippolito, G.; Cutignano, A.; Fontana, A. Plasticity of oxylipin metabolism among clones of the marine diatom *Skeletonema marinoi* (Bacillariophyceae). *J. Phycol.* **47**, 1050–1056 (2011).

73. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

74. Seemann, T. Prokka: Rapid Prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

75. Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; Dehal, P.; Ware, D.; Perez, F.; Canon, S.; et al. KBase: The United States Department of Energy systems biology knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).

76. Bowers, R.M.; Kyrpides, N.C.; Stepanauskas, R.; Harmon-Smith, M.; Doud, D.; Reddy, T.B.K.; Schulz, F.; Jarett, J.; Rivers, A.R.; Eloe-Fadrosh, E.A.; et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).

77. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).

78. Sabatino, V.; Russo, M.T.; Patil, S.; D'Ippolito, G.; Fontana, A.; Ferrante, M.I. Establishment of genetic transformation in the sexually reproducing diatoms *Pseudo-nitzschia multistriata* and *Pseudo-nitzschia arenysensis* and inheritance of the transgene. *Mar. Biotechnol.* **17**, 452–462 (2015).

79. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3-New capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

80. Koressaar, T.; Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23**, 1289– 1291 (2007).

81. Hansen, M.C.; Tolker-Nielsen, T.; Givskov, M.; Molin, S. Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol. Ecol.* **26**, 141–149 (1998).

82. Reysenbach, A.L.; Pace, N.R. *Reliable Amplification of Hyperthermophilic Archaeal 16S rRNA Genes by the Polymerase Chain Reaction*; Robb F T Place R Archaea Cold; Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA pp. 101–105 (1995).

GENERAL DISCUSSION

The present PhD thesis is aimed at the exploration of the biotechnological potential of the diatom *Thalassiosira rotula* and its associated bacteria through an -omic approach. The expression of the identified pathways, leading to the production of different bioactive compounds, have been further assessed along the diatoms growth or under different nutritional stresses to evaluate also their involvement in diatoms eco-physiology and stress response.

In **Chapter 1** we explored the *T. rotula* transcriptome, obtained challenging this species with silica limitation as stress. The results highlighted firstly the presence of a large group of differentially expressed uncharacterized proteins that could represent a still unexplored potential source of bioactive compounds. Furthermore, we provided evidences for the presence of biosynthetic pathways leading to the production of bioactive compounds never detected before in diatoms, i.e. secologanin and polyketides. Secologanin is a precursor of monoterpenoid indole alkaloids (MIAs), a large class of compounds with known utilization as anti-cancer, anti-malarial and anti-arrhythmic[1]. Polyketides comprises a great variety of natural products, working as toxins, siderophores, pigments, antibiotics, cytostatics, and immunosuppressants[2].

We showed also the presence of genes involved in the synthesis of prostaglandins in *T. rotula*, as previously revealed for *S. marinoi*[3]. The expression levels of the key genes of each pathway have been evaluated and their belonging to the alga rather than to the associated bacteria have been confirmed using the *T. rotula* axenic gDNA. The experimental data highlighted that secologanin synthase (SLS), involved in the MIAs biosynthesis, was significantly up-regulated in silica limitation while in the same condition, cyclooxygenase (COX), the key enzyme involved in prostaglandin biosynthesis, was significantly down-regulated. On the contrary, all the enzymes involved in the polyketides biosynthetic pathway were not significantly affected by this stress demonstrating that each pathway expression levels can be modulated by varying culturing conditions. In this sense, the use of different transcriptomes, as the ones grouped by the MMETSP project, may help in the detection of the most suitable condition for the expression of genes of interest, even if further investigations are necessary to identify the conditions needed to also stimulate the compounds production.

Finally, the preliminary results regarding the genome structure of *T. rotula* revealed an incredibly big dimension of its genome due to a quantity of transposable elements higher in comparison to all the diatoms sequenced until now.

Prostaglandins, deriving from COX activities, are PUFAs-derived lipid mediator that play a pivotal role in many physiological and pathological processes, mainly studied in animals[4]. The identification of the prostaglandin biosynthetic pathway in *T. rotula*, as was for *S. marinoi*[3], represents a further enrichment in the comprehension of PUFAs metabolism in diatoms. For this reason, in **Chapter 2**, we studied more deeply the *T. rotula* prostaglandin pathway, confirming its activity through the assessment, for the first time, of prostaglandins release outside the cells, in the medium, along the diatom growth. Phylogenetic analysis of the *T. rotula* COX protein sequence, highlighted the general conservation of the catalytic sites and the motifs characterizing the heme-peroxidase protein superfamily to which it belongs.

Gene expression studies confirmed the pathway presence and expression in *T. rotula*, showing a peak of expression of prostaglandin E-synthase-2-like and prostaglandin F-synthase at the end of the exponential growth phase, one day before the peak of expression of the rate limiting enzyme COX. This last in fact showed a peak after 5 days of growth, when diatoms enter in stationary phase. This stage of the diatoms growth curve is characterized by a decrease in nutrients and an increase of organic matter released from dying algal cells, that can be substrate for bacterial growth. On the basis of the results obtained, and considering that COX expression in animals can be stimulated also by the presence of bacteria[5], we hypothesize a signalling role of prostaglandins within the plankton community, that however needs to be further investigated.

The exploration of *T. rotula* transcriptome allowed also the identification of enzymes involved in the biosynthesis of phytosterols. These compounds already demonstrated to possess a plethora of different bioactivities among which hypocholesterolemic, anti-diabetic, anti-inflammatory, anti-cancer, neuro-modulatory and neuroprotective ones[6,7]. In **Chapter 3** we report the preliminary results describing the structure of the sterols biosynthetic pathway in *T. rotula*. The goal of our investigation was also to assess whether *T. rotula* could be considered a valuable source of sterols and to identify the most suitable stress condition able to stimulate sterol production in this diatom.

The results obtained showed firstly that *T. rotula* possess both the MEP and MVA pathways leading to the biosynthesis of both sterols and terpenoids precursors. However, the *in silico* genes expression analysis highlighted the general down-regulation of the

MVA pathway and up-regulation of the MEP pathway in silica limitation. Interestingly, in organisms that possess both these routes, the MEP is mainly committed to the production of carotenoids, MIAs and other terpenoids. *In silico* analysis revealed also the general up-regulation of the Iridoid pathway in silica limitation, enriching the data reported in Chapter 1 regarding the presence of the SLS in *T. rotula*. Moreover, these data suggest that silica limitation could be an intriguing condition in which evaluate MIAs production, although it is not the preferable condition to stimulate sterol pathway expression.

We evaluated also the amount of the major sterol produced, i.e. 24-methylene cholesterol (24-MC) and the expression of the key genes involved in its synthesis in cultures challenged with different stresses, i.e. nitrogen, silica and phosphate limitation. The results obtained revealed that 24-MC amount is significantly reduced in cultures grown in nitrogen and phosphate limitation and not influenced by silica limitation. A general down-regulation trend is evident in the genes expression of the nitrogen and silica starved cultures and up-regulation trend in phosphate starved cultures. Overall, the gene expression of the key enzymes of the biosynthetic pathway seems to be very low in all the conditions tested. Further deepening investigation is necessary to understand whether a post-transcriptional regulation occurs when the pathway activity does not correlate with the molecule production.

The preliminary results here obtained allowed an improved understanding of the pathway that, together with the genome and transcriptome sequencing we performed for this species (Chapter 1), could be helpful in the application of molecular approaches, such as RNAi or CRISPR/Cas methods[7], to perturb the pathway allowing the enhancement of the sterol content in this diatom as future perspective of this study. In addition, a study in progress in our group is revealing a cytotoxic activity of 24-MC toward breast and lung cancer cells lines (Personal communication).

Diatoms have shared the same environment with bacteria for more than 200 million of years and the horizontal gene transfer occurred has been fundamental in their diversification and evolutionary success[8]. This cohabitation favoured also the onset of cooperative and competitive interactions generating complex microbial communities[9] constituted by an high diversity of bacterial species, often unidentified, that represent an underexplored source of metabolites potentially valuable for further exploitation. In **Chapter 4** we explored the metagenome of the bacterial population associated to *T. rotula*, identifying three new bacterial species that we named Clusters 1, 2 and 8 (Cl-1,

Cl-2, Cl-8). The taxonomic classification allowed to identify them as new species in the families *Rhodobacteraceae*, *Parvibaculaceae* and *Cylcobacteriaceae* respectively. The physiological characterization of the three bacteria evidenced their dependence on diatom organic matter, due to their major abundance in the strictly-associated fraction of the culture, rather than in the free-living one. Interestingly, all the three bacteria have been found in association with another diatom, i.e. *S. marinoi* clone FE7 but not in the clone FE60, suggesting the putative establishment of a cohort between the three bacteria. In addition, a different tolerance to β- lactam and aminoglycoside antibiotics have been detected, with Cl-8, the one possessing 34 putative genes for β-lactamase/cephalosporinases in its genome, being tolerant to all the mixtures in a dose-dependent manner. The functional annotation revealed the presence of biosynthetic clusters belonging to polyketides (PKS), non-ribosomal peptides (NRPs), terpenes (TP) and antibiotic classes of molecules in the genomes of the three bacteria and the presence of an uncharacterized hybrid NRPS/PKS pathways in Cl-8, which further exploration is interesting for putative pharmacological applications. Among the identified biosynthetic gene clusters, we focused our attention on the ectoine gene cluster identified in Cl-1 genome, since this compound has an important physiological role as compatible solute accumulating in diatoms cells in response to the external increase of osmotic pressure. Measuring the ectoine produced in xenic *T. rotula* cultures under salinity stress, we confirmed the functionality of Cl-1 ectoine biosynthetic gene cluster, showing also that this bacterium could complement the ectoine production by the diatom under both hypo- and hyperosmotic conditions. This characteristic of Cl-1 is interesting also from a biotechnological point of view, since it is at the base of the bacterial milking applied in the commercial production of ectoine[10]. The results presented in this chapter constitute a preliminary characterization of these bacteria that pave the way for further developments in their isolation and/or additional exploration of their bioactive compounds.

Altogether, the results obtained in this thesis work provide new stimuli in the search of bioactive compounds from diatoms, by exploring the hidden biotechnological potential of the cosmopolitan diatom *T. rotula*. In addition, our studies enrich the previous knowledge regarding diatoms lipid metabolism using *T. rotula* as model species and open new questions to answer regarding the eco-physiological role of the compounds we identified. Moreover, the study of its microbiome has provided also preliminary indications about the biotechnological potential of the bacteria associated to this diatom, never studied before from this point of view.

In conclusion, given the high number of transposable elements identified in *T. rotula* genome and the fundamental role that transposons have in diatoms evolution and in the regulation of genes expression, we purpose our *T. rotula* species as model species to study transposons in diatoms.

References

1. O'Connor, S. E. & Maresh, J. J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **23**, 532–547 (2006).

2. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *PNAS* **111**, 9259–9264 (2014).

3. Di Dato, V. *et al.* Animal-like prostaglandins in marine microalgae. *ISME J* **11**, 1722–1726 (2017).

4. Wiktorowska-Owczarek, A., Berezińska, M. & Nowak, J. Z. PUFAs: Structures, Metabolism and Functions. *Adv Clin Exp Med* **24**, 931–941 (2015).

5. Martínez-Colón, G. J. & Moore, B. B. Prostaglandin E2 as a Regulator of Immunity to Pathogens. *Pharmacol Ther* **185**, 135–146 (2018).

6. Jaramillo-Madrid, A. C., Ashworth, J., Fabris, M. & Ralph, P. J. Phytosterol biosynthesis and production by diatoms (Bacillariophyceae). *Phytochemistry* **163**, 46–57 (2019).

7. Randhir, A., Laird, D. W., Maker, G., Trengove, R. & Moheimani, N. R. Microalgae: A potential sustainable commercial source of sterols. *Algal Research* **46**, 101772 (2020).

8. Amin, S. A., Parker, M. S. & Armbrust, E. V. Interactions between Diatoms and Bacteria. *Microbiol Mol Biol Rev* **76**, 667–684 (2012).

9. Seymour, J. R., Amin, S. A., Raina, J.-B. & Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton-bacteria relationships. *Nat Microbiol* **2**, 17065 (2017).

10. Sauer, T. & Galinski, E. A. Bacterial milking: A novel bioprocess for production of compatible solutes. *Biotechnol Bioeng* **57**, 306–313 (1998).

APPENDIX A

Supplementary Information

PART 1

Supplementary Information

Chapter 1

Identification of the secologanin, polyketides and prostaglandins biosynthetic pathways in the diatom *Thalassiosira rotula*

Adapted from the published article:
Di Dato, V., Di Costanzo, F., Barbarinaldi, R., Perna, A., Ianora, A., Romano G. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Scientific Reports* (2019) **9**, 9893.

Supplementary Information 1 (SI1)

Supplementary Figures.

Supplementary Figure 1.1: Representation of the GO term categories.
Supplementary Figure 1.2: Representation of the GO terms categories associated to the Differential ExpressedTranscripts (DET).
Supplementary Figure 1.3: Representation of the Monoterpenoid biosynthetic pathway.
Supplementary Figure 1.4: Alignment of the five peptides annotated as Secologanin synthase.
Supplementary Figure 1.5: Representation of the Prostaglandin biosynthetic pathway.
Supplementary Figure 1.6: Representation of the Polyketides biosynthetic pathway.
Supplementary Figure 1.7: Oligo pairs standard curve of the transcripts in the selected pathways analysed by qPCR.
Supplementary Figure 1.8: Dapi staining and peptone assay for genomic DNA axenicity verification.
Supplementary Figure 1.9: Agarose gel of PCR amplification of *Thalassiosira rotula* FE80 clone transcripts in figure 1.4, in main text.

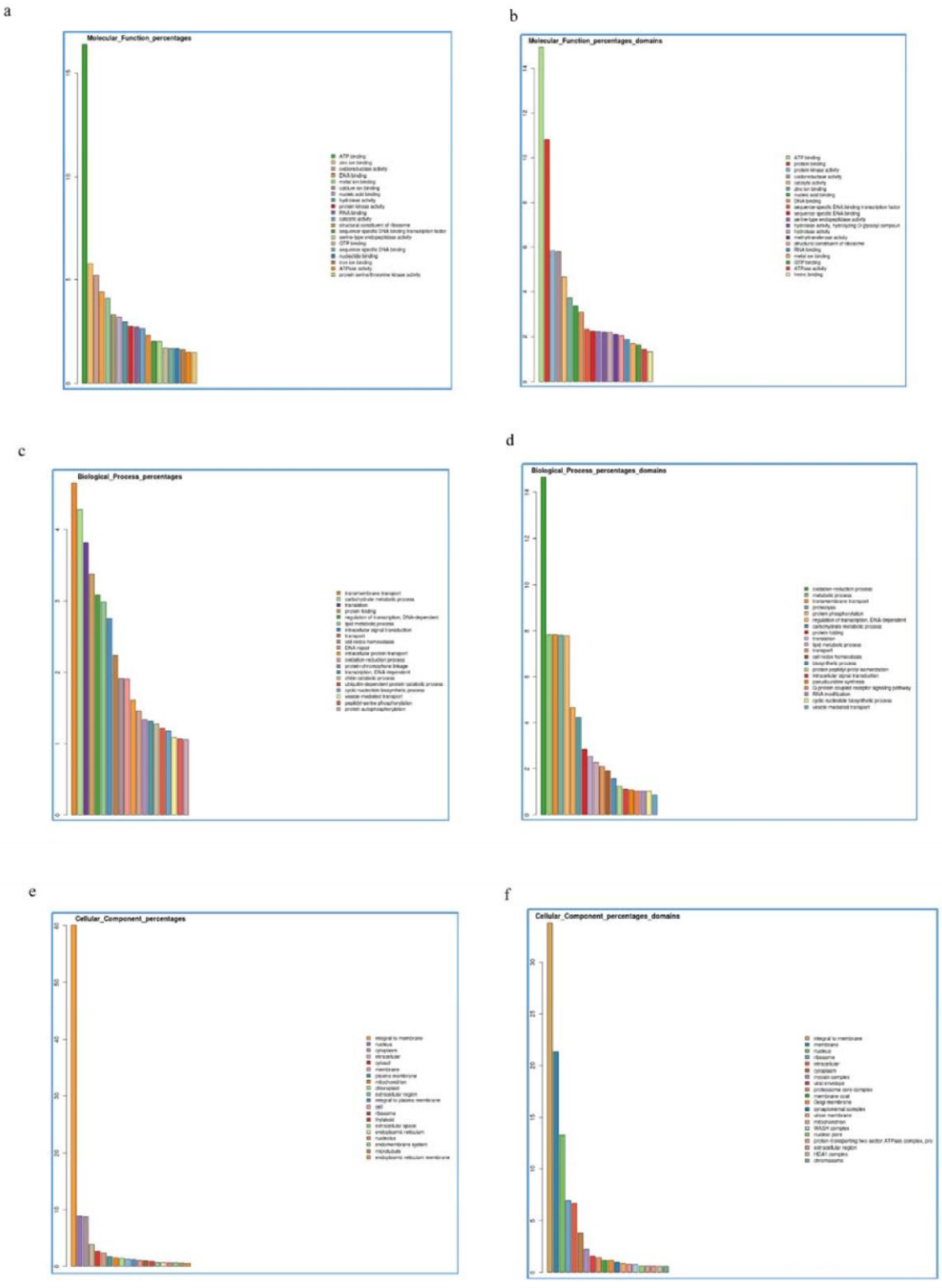
Supplementary Tables.

Supplementary Table 1.1: 1a. Transcriptome assembly statistics. 1b. General transcriptome statistics. 1c. General proteome statistics.
Supplementary Table 1.2: Ten DET-associated GO terms.
Supplementary Table 1.3: Level 1 Pathway Annotation.
Supplementary Table 1.4: Prostaglandin annotation in *Thalassiosira rotula* FE80 (CCMP1647).

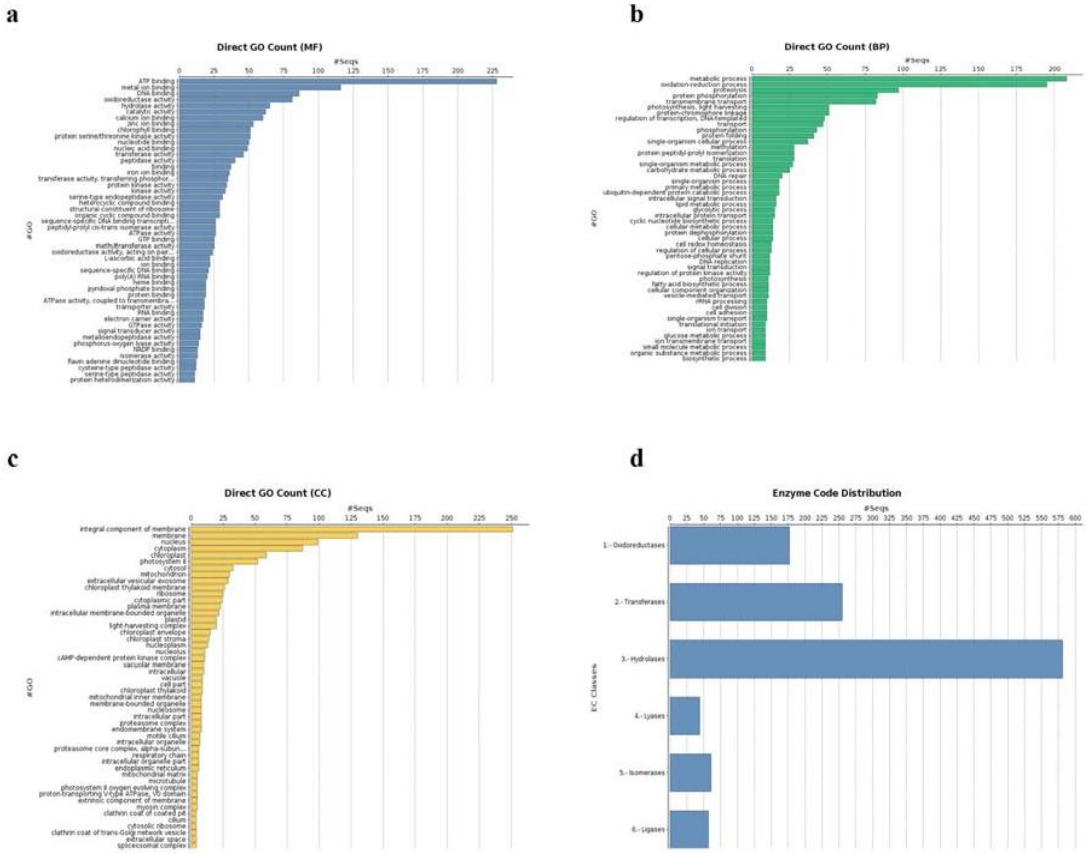
For their high number of pages, all the **Supplementary Files** are available (from page 45 onwards) at the following link:
https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-019-46276-8/MediaObjects/41598_2019_46276_MOESM1_ESM.pdf,
while all the **Supplementary Data** are available at:
https://drive.google.com/drive/folders/1UszIbyzRaejvgKpOocan44it6MtKUS-n?usp=sharing.

Supplementary Figures.

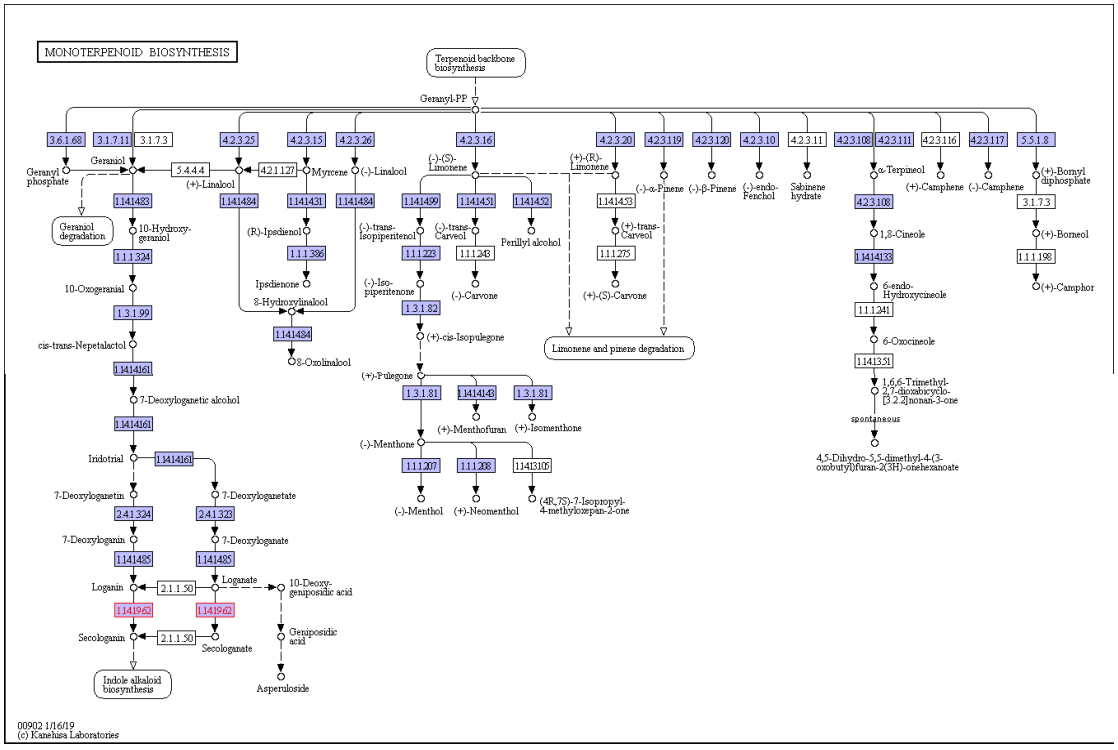
Supplementary Figure 1.1. Histograms representing the top 20 peptide sequences and domains associated to the GO term categories: a and b, Molecular Function category; c and d, Biological Processes category; e and f, Cellular Components category.



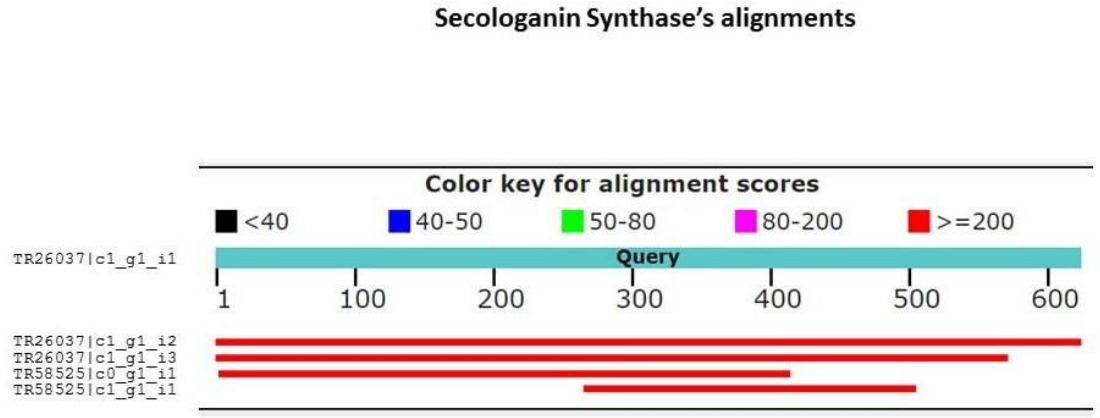
Supplementary Figure 1.2. Histograms representing the distribution and number, in the different categories, of the GO terms associated to the DET (Differential Expressed Transcripts). a) Molecular Function category; b) Biological Processes category; c) Cellular Components category, d) Enzyme classes.



Supplementary Figure 1.3. pattern of monoterpenoid biosynthesis of which secologanin synthase, indicated in red, is part. Font: KEGG: Kyoto Encyclopedia of Genes and Genomes.

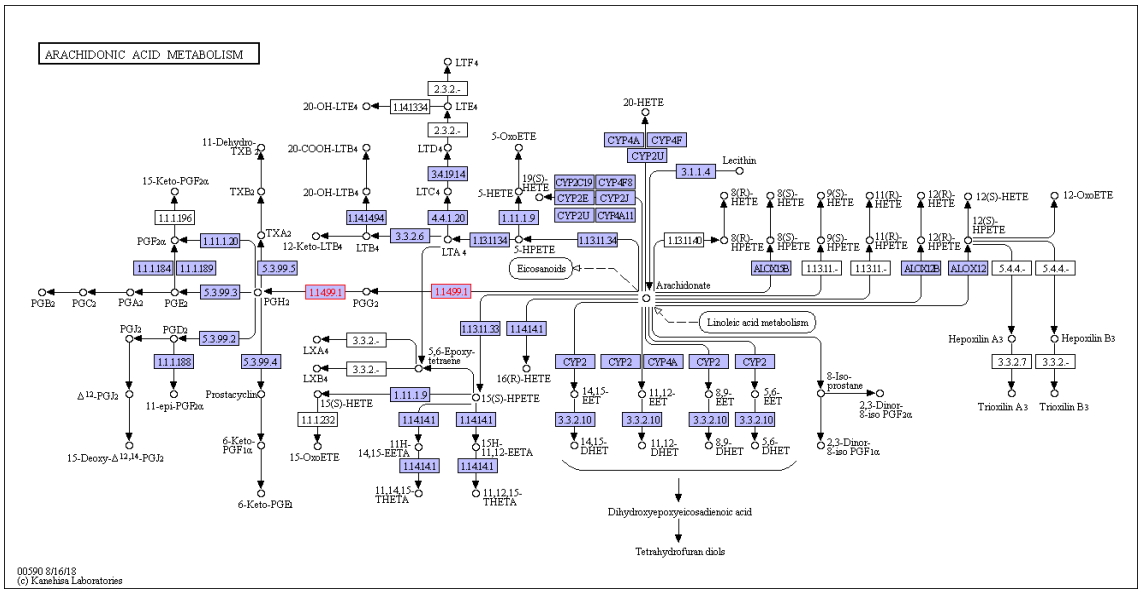


Supplementary Figure 1.4. Alignment of the five peptides annotated as Secologanin synthase.

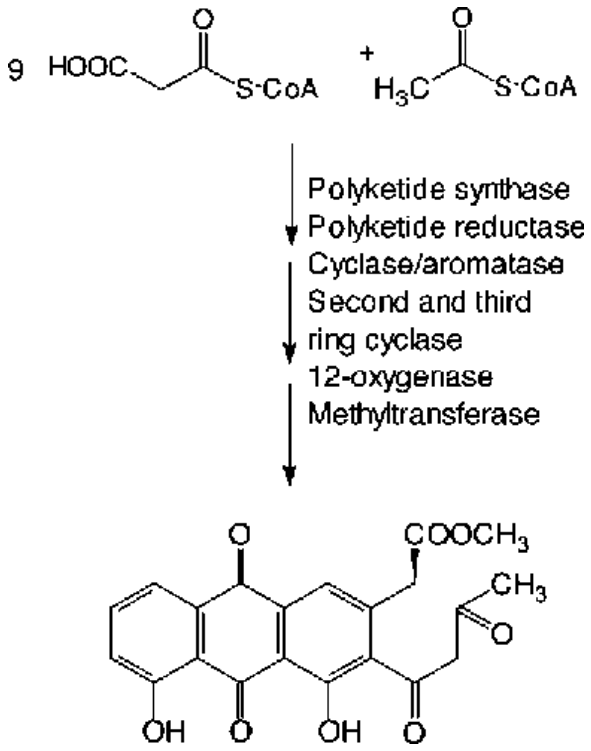


Description	Max score	Total score	Query cover	E value	Ident	Accession
TR26037|c1_g1_i2|m.127109	1288	1288	100%	0.0	99%	Query_24922
TR26037|c1_g1_i3|m.127112	1024	1024	91%	0.0	87%	Query_24923
TR58525|c0_g1_i1|m.284670	455	455	66%	2e-159	54%	Query_24924
TR58525|c1_g1_i1|m.284673	351	351	38%	6e-122	66%	Query_24925

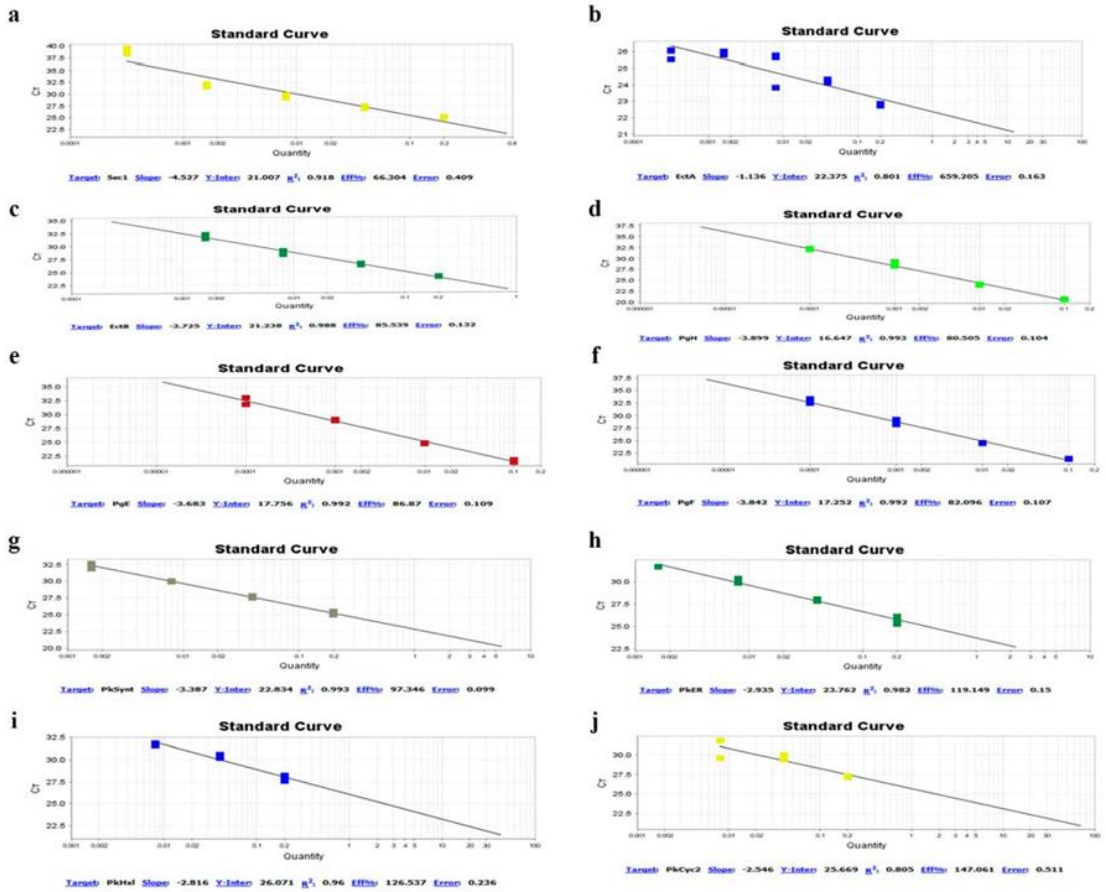
Supplementary Figure 1.5. Pattern of arachidonic acid biosynthesis of which prostaglandin enzymes are part. In red the prostaglandin G/H synthase or cyclooxygenase, from which prostaglandins derive. 5.3.99.2: microsomal Prostaglandin-E synthase 1; 1.11.1.20: prostaglandin F2α synthase. Font: KEGG: Kyoto Encyclopedia of Genes and Genomes.



Supplementary Figure 1.6. Example of iterative reactions for the synthesys of polyketides. Modified from: Sultana A. et al., 2004. Structure of the polyketide cyclase SnoaL reveals a novel mechanism for enzymatic aldol condensation. The EMBO Journal, volume 23, 9, pages 1911-21.



Supplementary Figure 1.7. Oligo pairs standard curve of the transcripts in the selected pathways analysed by qPCR. a. Secologanin synthase; b. Ecotine A; c. Ectoine B; d. PgG/Hs; e. $PgES_2$; f. PgFs; g. PKS_Synt; h. PKS_ER; i. PKS_Hxl; j. PKS_Cyc.

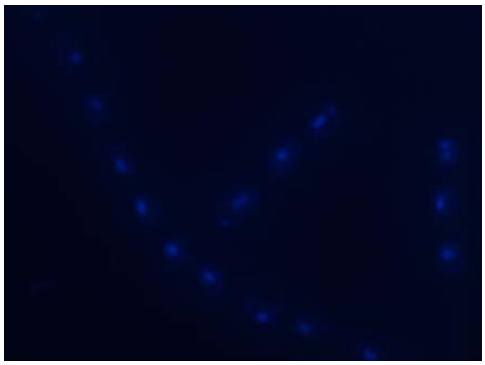


Supplementary Figure 1.8. Dapi staining and peptone assay to verify if genomic DNA was axenic.

a) Spontaneous Cholorophyll autofluorescence to assess normal cell shape;
b) Dapi staining to mark the nucleus to see the presence of bacteria associated to the cell. Only the cell nuclei are visible;
c) Peptone assay on three replicates to verify bacterial growth. The solutions are transparent indicating no bacterial growth.

a)

b)

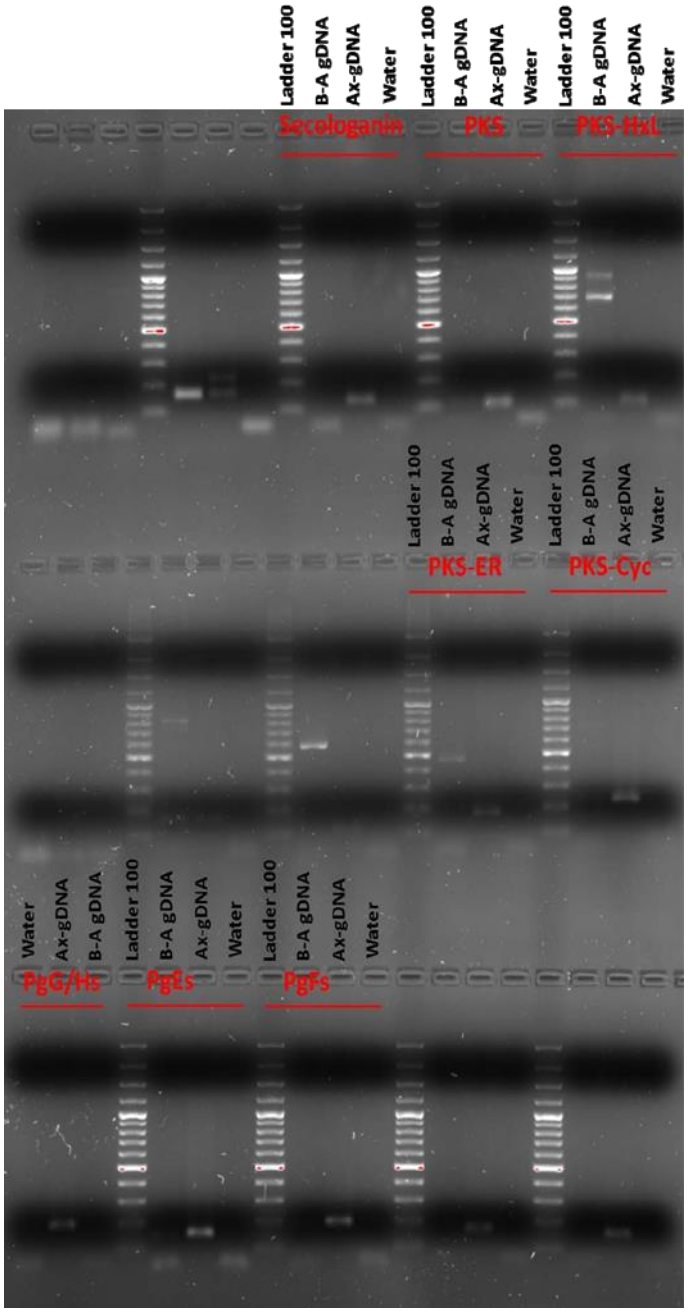


c)

Supplementary Figure 1.9. Agarose Gel of PCR amplification of the *Thalassiosira rotula*, clone FE80, transcripts showed in the figure 1.4 of the main text.

- B-A gDNA: genomic DNA of a mixed culture *Thalassiosira rotula*/associated bacteria.
- Ax-gDNA: genomic DNA of an axenic *Thalassiosira rotula* culture.

Ladder100: from bottom to top: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 3000 bps.



Supplementary Tables.

Supplementary Table 1.1 a. Transcriptome assembly statistics.

SAMPLE NAME	TREATMENT	READS BEFORE TRIMMING	READS AFTER TRIMMING
TR1	COMPLETE MEDIA	45735814	44763298
TR2	COMPLETE MEDIA	55689770	54219334
TR3	COMPLETE MEDIA	54231050	52909044
TR4	SILICA DEPLETED MEDIA	50028032	48504904
TR5	SILICA DEPLETED MEDIA	41982052	41063084
TR6	SILICA DEPLETED MEDIA	35165342	34181574

Supplementary Table 1.1 b. General transcriptome statistics.

	BEFORE FILTRATION FOR CONTAMINATION	AFTER FILTRATION FOR CONTAMINATION
MBP	152	142
TRANSCRIPT NUMBER	125135	107289
GENES	79330	66496
% GC	47.55	47.53
AVERAGE CONTIG LENGTH (BP)	1217	1332.76
MEDIAN CONTIG LENGTH (BP)	810	970
N50 (BP)	1940	2034

Supplementary Table 1.1 c. General proteome statistics.

		BEFORE FILTRATION FOR CONTAMINATION	AFTER FILTRATION FOR CONTAMINATION
PROTEIN SEQUENCES		67222	60172
COMPLETE (STARTING METHIONINE AND STOP CODON)		29304	28128
STARTING METHIONINE, NO STOP CODON		8669	7616
NO STARTING METHIONINE, YES STOP CODON		14628	13291
NO STARTING METHIONINE, NO STOP CODON		14621	11137
MEAN PEPTIDE LENGTH		-	281
MINIMUM PEPTIDE LENGTH		-	49
MAXIMUM PEPTIDE LENGTH		-	3952
N° of blast results	SWISS-PROT	-	15601
	UNIREF	-	37717
	DOMAINS	-	21011
	PEPTIDES WITH AT LEAST ONE BLAST RESULT	-	37999

Supplementary Table 1.2: Top Ten DET-associated GO terms.

Biological Processes	Cellular Components	Molecular Functions
Metabolic Process	Integral component of membrane	ATP binding
Oxidation-Reduction Process	Nucleus	Metal Ion Binding
Proteolysis	Cytoplasm	DNA Binding
Protein Phosphorylation	Chloroplast	Oxidoreductase activity
Transmembrane transport	Photosystem II	Hydrolase activity
Photosynthesis, light harvesting	Cytosol	Catalytic activity
Protein-chromophore linkage	Mitochondrion	Calcium Ion Binding
Regulation of transcription, DNA-templated	Extracellular vesicular exosome	Zinc Ion Binding
Transport	Chloroplast thylakoid membrane	Chlorophyll Binding
Phosphorylation	Ribosome	Protein Serine/Threonine Kinase Activity

Supplementary Table 1.3. Level 1 Pathway Annotation. Mean expression value, in FPKM, and statistics data of the pathways on first level of annotation on the total proteome.

#: pathways numeration; **# of ass. Seqs:** number of peptide sequences associated to each pathway; **Pathways on Level 1:** pathway's name; **CTR MEAN:** mean FPKM value among all the corresponding transcript of the complete medium cultures associated to the pathway; **CTR SD:** standard deviation of the CTR mean; **SiDepl FPKM MEAN:** mean FPKM value among all the corresponding transcript of the silica depleted medium cultures associated to the pathway; **SiDepl SD:** standard deviation of the SiDepl mean; **p-value CTR vs SiDepl:** test on the FPKM mean of the CTR versus SiDepl for each pathway; **p-adj value CTR vs SiDepl:** pvalue adjusted on the total number of associated sequences.

#	# of ass. seqs	Pathways onLevel 1	CTR FPKM MEAN	CTR SD	SiDepl FPKM MEAN	SiDepl SD	p-value CTR vs SiDepl	p-adj value CTR vs SiDepl
1	7	Alkaloid biosynthesis	1.93819087357143	1.93100775440389	12.3622564214286	14.9041172497516	0.114564625127527	1
2	3	Alkaloid degradation	1.9967997	1.48837117146259	2.60919641666667	2.14566014688563	0.707797304611788	1
3	7	Alkene biosynthesis	3.0170001	3.57221627966652	2.01524596428571	2.02850639235049	0.534072564311056	1
4	46	Amine and polyamine biosynthesis	10.7085171301143	15.0839619677234	9.77146942633582	15.6913691264389	0.759141265109627	1
5	13	Amine and polyamine degradation	1.57156235322269	1.64110401961755	1.61593673070014	1.68325362842774	0.946304022362477	1
6	2	Amine andpolyamine metabolism	4.02284685	3.99030772306456	5.44424825	0.232285638279952	0.702901173816272	1
7	240	Amino-acidbiosynthesis	9.6266325170581	32.4637595818077	11.2866950575017	52.8568136964496	0.656232622029268	1
8	95	Amino-acid degradation	5.6716736557636	11.218993623446	4.87511562220517	8.61322524297467	0.562695290408778	1
9	4	Amino-acid metabolism	13.243841875	19.6965739004505	8.3848496375	12.3864094255887	0.565898107703054	1
10	4	Aminoacyl-tRNA biosynthesis	5.444138975	6.31044116618844	6.26338385	7.83919532348209	0.876245938475542	1
11	32	Antibiotic biosynthesis	10.9575223058901	17.1491798977856	11.3091498875625	17.8616027224739	0.936233460295997	1
12	4	Aromatic compound metabolism	2.27103873763088	4.07718996467948	1.64405210003583	2.79758487981645	0.809329477753681	1
13	12	Bacterial outer membrane biogenesis	9.07707028302721	12.9355798436961	11.3307976828702	10.7835107069336	0.647645874467103	1
14	4	Capsule biogenesis	3.473348	4.33567149842042	6.3905280125	7.96247669281662	0.550361698290555	1
15	6	Carbohydrate acid metabolism	18.0743664166667	23.9999559141408	17.5919465	19.1559628760125	0.97009745943686	1
16	48	Carbohydrate biosynthesis	20.8858513833617	43.6280500889786	16.5137272622563	36.3145331426027	0.594898683302364	1
17	169	Carbohydrate degradation	42.9728390495802	226.42491928468	26.1735629970953	101.447625525475	0.378245239853317	1
18	48	Carbohydrate metabolism	11.5114837273302	16.3308996710298	11.3171064160177	18.1685114337505	0.956156619827766	1
19	19	Carotenoidbiosynthesis	3.9948779495122	4.83838511854079	3.36496709079836	3.25374585209978	0.615374133775705	1
20	2	Catecholamine biosynthesis	2.01856175	0.533213313790199	1.808283	0.402181123935473	0.702558791913895	1
21	8	Cell wall biogenesis	5.26537841875	6.4327968739877	5.0421862625	5.98166237031681	0.943730291791584	1
22	186	Cofactor biosynthesis	8.4263563238636	25.8026840511522	7.61973078395479	19.4338939703801	0.727517781932475	1
23	6	Cofactor degradation	6.91392311666667	12.1431704362354	8.13644774166667	13.3013763416456	0.871287996676324	1

24	**3**	**Cofactor metabolism**	1.94363667125	1.71070792005389	1.4350754625	0.873441316787731	0.621685095970365	1
25	**26**	**Energy metabolism**	11.5748235052116	18.0531535068561	11.0145811675983	16.8634257577519	0.90840615899133	1
26	**16**	**Fermentation**	0.397935310579678	0.704412451264113	0.885105205625	1.91650883318659	0.351897372773782	1
27	**1**	**Flavonoid metabolism**	27.731895	NA	46.99605	NA	10	1
28	**10**	**Genetic information processing**	8.309927715	7.5362047500817	5.48287793	3.48224541669386	0.301611232918807	1
29	**1**	**Glycan biosynthesis**	2.2356625	NA	3.128675	NA	10	1
30	**20**	**Glycan degradation**	7.2114768525	10.728644651084	4.8022975257675	8.80311121422541	0.442530192577737	1
31	**18**	**Glycan metabolism**	2.21121046901364	2.822066444023	2.16918962655682	2.63783765463647	0.959550834346818	1
32	**19**	**Glycerolipid metabolism**	5.58599072026316	8.20476744902123	6.04173952631579	8.62839832602791	0.868427745579636	1
33	**28**	**Glycolipid biosynthesis**	3.42656811932143	6.77484974711564	3.94197677283591	8.46621660637824	0.802412639261149	1
34	**3**	**Hormone biosynthesis**	1.58883798166667	1.87897594804455	1.53535429333333	1.24958862937039	0.969499137381665	1
35	**21**	**Isoprenoid biosynthesis**	8.08461845735294	13.0375577599331	9.09789294558824	12.37164096636	0.743402187863152	1
36	**7**	**Ketone metabolism**	2.75663957106605	2.90248527588007	2.27429664674354	2.29461346544345	0.736443822783611	1
37	**175**	**Lipid metabolism**	19.653383597443	77.0040678340764	21.8519716131275	105.980158295988	0.824443293554184	1
38	**4**	**Membrane lipid metabolism**	3.3040355	2.06912854027273	5.116475125	3.10595097012977	0.374198350328475	1
39	**27**	**Metabolic intermediate biosynthesis**	7.20515494715714	11.953221005346	6.9181116620466	10.3579764811837	0.914821263838417	1
40	**10**	**Metabolic intermediate metabolism**	6.75222531815	7.99860877282214	5.5872351644951	6.89036276021084	0.731257767515254	1
41	**16**	**Nitrogen metabolism**	2.93282775240625	4.29816271070079	3.7027467718125	6.97339241219183	0.710126837069234	1
42	**2**	**Nucleoside biosynthesis**	2.77259525	1.54572799905259	2.76126105	3.35109988059195	0.997067883355439	1
43	**35**	**Nucleotide-sugar biosynthesis**	15.0664536510462	20.2615260954377	21.0973595339467	28.0400371748146	0.299517010219956	1
44	**34**	**One-carbon metabolism**	4.3652897082669	8.97176633435821	4.22715016741417	8.49042459715027	0.948205158148916	1
45	**5**	**Organic acid metabolism**	2.939366604	2.57301072759137	2.7137586459581	3.12923916010463	0.904085053178901	1
46	**7**	**Organosulfur biosynthesis**	24.2572826571429	43.6444289237238	19.7665988571429	28.8121068192813	0.824683760954756	1
47	**43**	**Phospholipid metabolism**	4.72428288846705	5.69588511867248	5.53593137490795	7.60717240030549	0.572627363100781	1
48	**10**	**Photosynthesis**	1.979514526	1.77363261163519	1.4482921781964	1.15504667948054	0.439390337888341	1
49	**2**	**Phytoalexin biosynthesis**	12.59049875	11.3292241895309	8.8297445	6.7948796813086	0.73360541508224	1
50	**7**	**Plant hormone biosynthesis**	2.44234762142857	3.40077574119786	1.24640577142857	1.07377775759095	0.40366725583701	1
51	**24**	**Polyol metabolism**	7.6744324291875	8.95201620215457	9.00329173261382	8.31275298910688	0.596685711930388	1

52	**77**	**Porphyrin-containing compound metabolism**	5.51203371967617	10.138413377452	6.15737972043181	12.0970801479157	0.703416690572848	1
53	**7**	**Protein biosynthesis**	88.731714595	228.281078664317	245.919347917536	643.421291668945	0.560446907179927	1
54	**3**	**Protein degradation**	11.4585388333333	8.5407517994378	11.7043961666667	10.0797161999054	0.975870621104043	1
55	**456**	**Protein modification**	9.15065204228311	17.5006443204081	9.68016677262585	19.1038948893162	0.662621983834879	1
56	**62**	**Purine metabolism**	9.45315776836404	17.5755754183667	8.16373798801145	10.6739617156226	0.606128817994198	1
57	**34**	**Pyrimidine metabolism**	6.42380451179622	12.8555357231255	5.72509374014706	10.2303485577813	0.762010544473349	1
58	**14**	**Quinol/quinone metabolism**	6.13504992176471	4.77106920981642	7.08846534676471	6.07587124849835	0.614539376892254	1
59	**1**	**Secondary metabolite biosynthesis**	1.887811	NA	2.0184605	NA	10	1
60	**8**	**Secondary metabolite metabolism**	8.01349330484776	13.0120863166746	15.1421185014724	31.5947931505467	0.569180660718774	1
61	**1**	**Signal transduction**	10.61557	NA	8.9021045	NA	10	1
62	**1**	**Spore coat biogenesis**	18.60315	NA	28.106485	NA	10	1
63	**18**	**Steroid biosynthesis**	5.57185156421053	7.5054417087592	6.80928485352205	10.9452695542686	0.687146803764157	1
64	**4**	**Steroid metabolism**	12.77444025	14.7488280129856	14.2216669125	18.5241955574504	0.906895024120955	1
65	**48**	**Sulfur metabolism**	8.05854475010417	38.5316113141322	7.9326679588125	31.4647990229458	0.986051748274181	1
66	**3**	**Terpene metabolism**	3.68994991666667	5.82481319632118	3.69068256666667	5.06165386582286	0.999876811908281	1
67	**1**	**Thermoadapter biosynthesis**	36.453995	NA	17.79681	NA	10	1
68	**22**	**tRNA modification**	1.73801389812273	2.17021583703122	2.01134279954547	1.72551912375392	0.646306823900338	1

Supplementary Table 1.4: Prostaglandin annotation in *Thalassiosira rotula* FE80 (CCMP1647).

Transcript Name	DescriptionSP	EnzymeDescs	PwLev1	PwLev2	PwLev3	DescriptionUf
TR11659|c0_g1_i1	Prostaglandin F synthase	Prostaglandin-F synthase.	Lipid metabolism	prostaglandin biosynthesis	-	Aldo-keto oxidoreductase
TR13003|c0_g1_i1	Prostaglandin F synthase	Prostaglandin-F synthase.	Lipid metabolism	prostaglandin biosynthesis	-	Putative uncharacterized protein
TR17558|c0_g1_i1	Prostaglandin F synthase	Prostaglandin-F synthase.	Lipid metabolism	prostaglandin biosynthesis	-	Uncharacterized protein

TR19490|c1_g4_i3	Carbonyl reductase NADPH 1	Carbonyl reductase(NADPH).]---[Prostaglandin-E(2) 9- reductase.]---[15-hydroxyprostaglandin dehydrogenase (NADP()).	-	-	-	Uncharacterizedprotein
TR19490|c1_g4_i4	Carbonyl reductase NADPH 1	Carbonyl reductase(NADPH).]---[Prostaglandin-E(2) 9- reductase.]---[15-hydroxyprostaglandin dehydrogenase (NADP()).	-	-	-	Uncharacterized protein
TR19490|c1_g4_i5	Carbonyl reductase NADPH 1	Carbonyl reductase(NADPH).]---[Prostaglandin-E(2) 9- reductase.]---[15-hydroxyprostaglandin dehydrogenase (NADP()).	-	-	-	Uncharacterized protein
TR25058|c1_g1_i2	Prostaglandin Esynthase 2	Prostaglandin-E synthase.	Lipid metabolism	prostaglandin biosynthesis	-	Predicted protein
TR28149|c0_g1_i1	Carbonyl reductase NADPH 1	Carbonyl reductase(NADPH).]---[Prostaglandin-E(2) 9- reductase.]---[15-hydroxyprostaglandin dehydrogenase (NADP()).	-	-	-	Uncharacterized protein
TR31063|c0_g1_i1	Carbonyl reductase NADPH 1	Carbonyl reductase(NADPH).]---[Prostaglandin-E(2) 9- reductase.]---[15-hydroxyprostaglandin dehydrogenase (NADP()).	-	-	-	Uncharacterized protein
TR31063|c0_g1_i2	Carbonyl reductase NADPH 1	Carbonyl reductase(NADPH).]---[Prostaglandin-E(2) 9- reductase.]---[15-hydroxyprostaglandin dehydrogenase (NADP()).	-	-	-	Uncharacterizedprotein
TR754|c0_g1_i1	Prostaglandin G/H synthase 2	Prostaglandin- endoperoxide synthase.	Lipid metabolism	prostaglandin biosynthesis	-	Heme peroxidase

Supplementary Information 2 (SI2):

1. Analysis of the best reference gene.
2. qPCR efficiency and stability in condition of silica starvation.

1. Analysis of the best reference gene.

Putative reference genes have never been identified before in *Thalassiosira rotula*.

To study the best genes to be used for data normalization in RT-qPCR experiments, a set of ten genes commonly used as internal controls for diatom gene expression analysis[1] were selected: Actin, Ubiquitin, TBP, COPα, TUBα, TUBα, RPS, GAPDH, H_4.

Three different software, BestKeeper[2]; NormFinder[3] and geNorm[4] were used to identify the best reference genes in condition of silica starvation, taken as reference condition to analyse all the genes set. Considering that each individual algorithm has some drawbacks, we decided to use all three algorithms and then compare the results.

In Supplementary Table 1.6 is reported the qPCR analysis of each reference gene in terms of standard curve efficiency and melting curve consistency.

Supplementary Table 1.6: qPCR analysis

Gene	Standard Curve			Tm Peak
	Slope	Efficiency	R^2	
Actin	-3.553	1.9118	0.999	Single
Ubiquitin	-3.256	2.0283	0.99	Multiple
TBP	-3.694	1.8651	0.984	Single
COPα	-3.125	2.0893	0.997	Single
TUBα	-3.33	1.9966	0.992	Single
TUBα	-3.227	2.0412	0.992	Multiple
RPS	-3.09	2.1068	0.991	Single
GAPDH	-2.928	2.1955	0.996	Single
H_4	-0.908	12.6276	0.798	Multiple
EF1α	-2.85	2.2432	0.964	Single

Couple of oligo for ubiquitin, tubulin beta and histon4 were not able to amplify well as indicated by multiple temperature of melting that are synonymous of multiple amplicons. These genes were excluded from the analysis.

Supplementary Table 1.7 shows that each algorithm utilized gave GAPDH as the third better gene putting it at the second position in the final ranking of better genes. However, even if several papers have reported GAPDH as an appropriate reference gene because of its constitutively expression and its involvement in basic functions required for cell maintenance, we decided to not use it in our analysis. In fact, GAPDH has been shown to be sometimes unstable in different experimental conditions and in diatoms also seems be regulated by light, chemical stresses and along the growth curve.

Supplementary Table 1.7: Ranking Order (Better--Good--Average)_7 genes

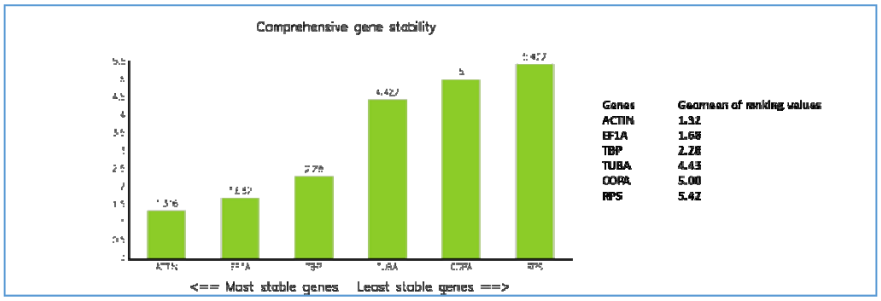
Method	***1***	***2***	***3***	***4***	***5***	***6***	***7***
Delta CT	Actin	EF1α	GAPDH	TUBα	TBP	RPS	COP α
BestKeeper	Actin	EF1α	GAPDH	TBP	RPS	COPα	TUB α
Normfinder	Actin	EF1α	GAPDH	TUBα	TBP	RPS	COP α
Genorm	GAPDH | RPS		ACTIN	TBP	EF1α	TUBα	COP α
Recommended comprehensive ranking	**Actin**	**GAPDH**	**EF1α**	**RPS**	**TBP**	**TUBα**	**COP α**

The genes finally analysed were six: actin, tubulin α, TBP (Tata Binding Protein), RPS (Ribosomal Protein Small subunit 30S), EF1α elongation Factor), COPα (Coatomer subunit alpha).
Supplementary Table 1.8 shows the best ranking with the different methods, and Supplementary Figures 1.11 to 1.15 shows the results for each methods. All algorithms indicated Actin as the most stable gene followed by EF1α, TBP, TUBα, COPα and finally RPS.

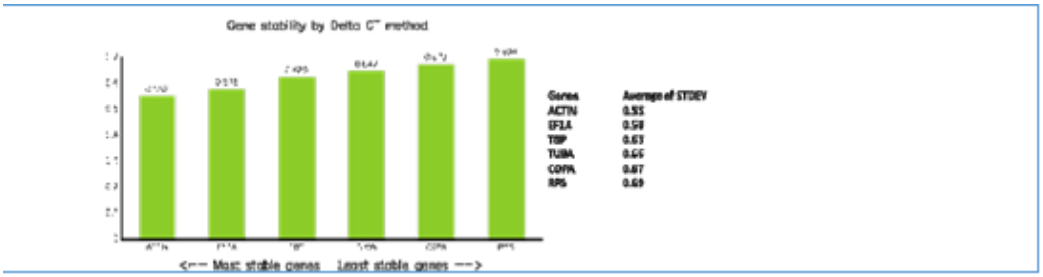
Supplementary Table 1.8: Ranking Order (Better--Good--Average)_6 genes

Method	***1***	***2***	***3***	***4***	***5***	***6***
Delta CT	ACTIN	EF1α	TBP	TUBα	COPα	RPS
BestKeeper	ACTIN	EF1α	TBP	RPS	COPα	TUBα
Normfinder	ACTIN	EF1α	TBP	TUBα	COPα	RPS
Genorm	EF1α | TBP		ACTIN	TUBα	COPα	RPS
Recommended comprehensive ranking	**ACTIN**	**EF1α**	**TBP**	**TUBα**	**COPα**	**RPS**

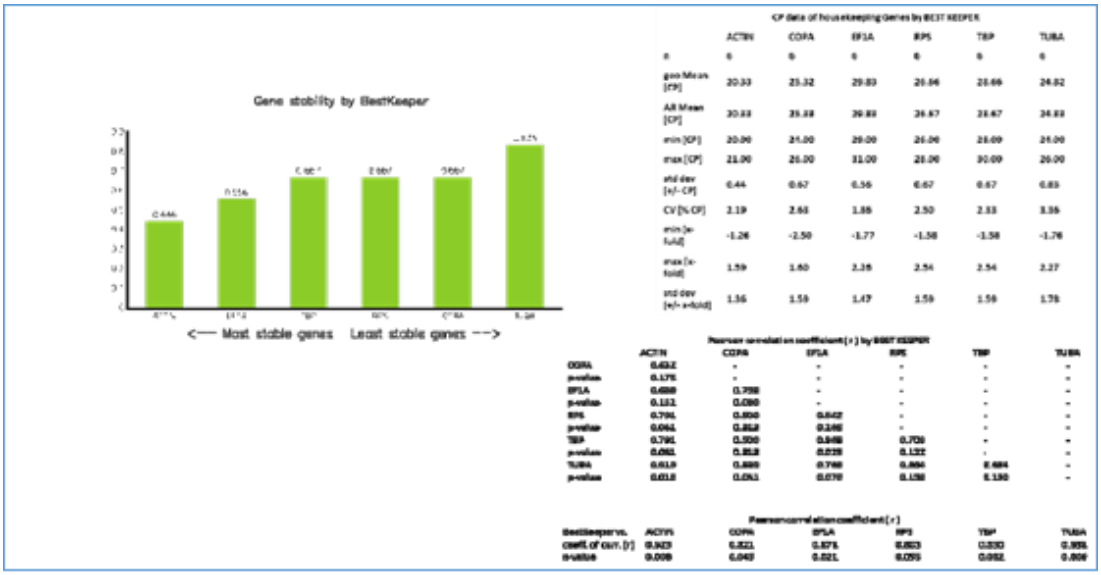
Supplementary Figure 1.11: Comprehensive gene stability.



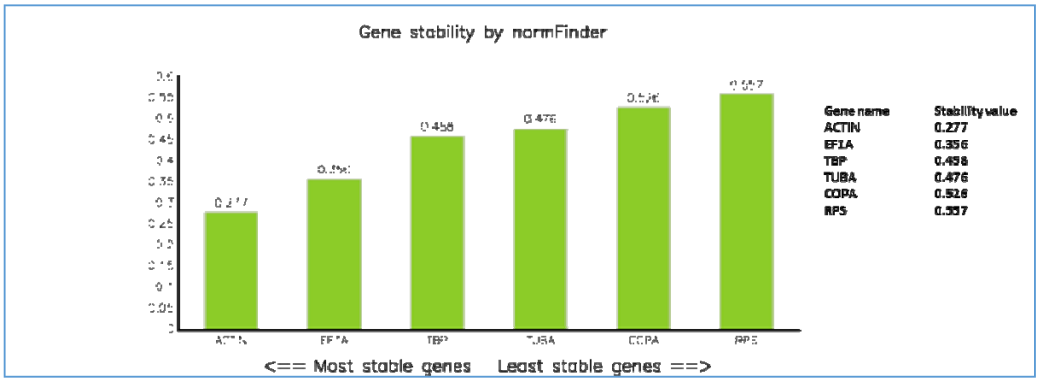
Supplementary Figure 1.12: Delta CT method.



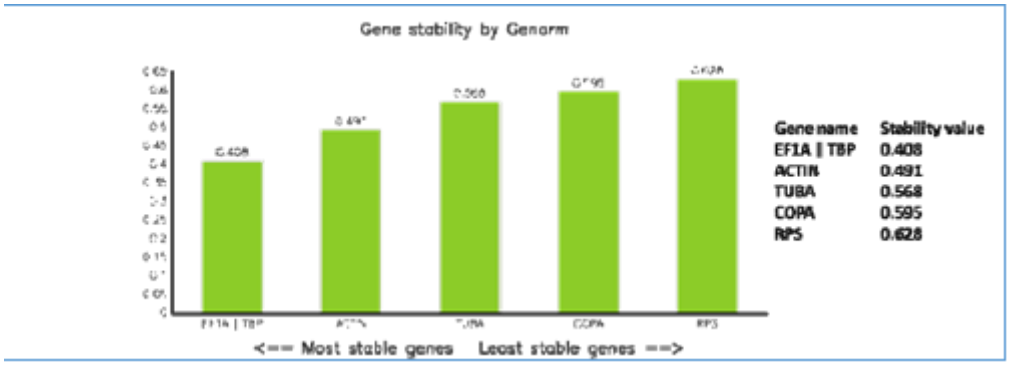
Supplementary Figure 1.13: BestKeeper.



Supplementary Figure 1.14: Normfinder.



Supplementary Figure 1.15: Genorm.



For our qPCR normalization, we decide to use Actin, EF1α and TBP. However, even designing different oligo's couple, EF1α in our subsequent experimental conditions was always instable, so we decide to use only Actin and TBP as reference genes.

2. qPCR efficiency and stability in condition of silica starvation.

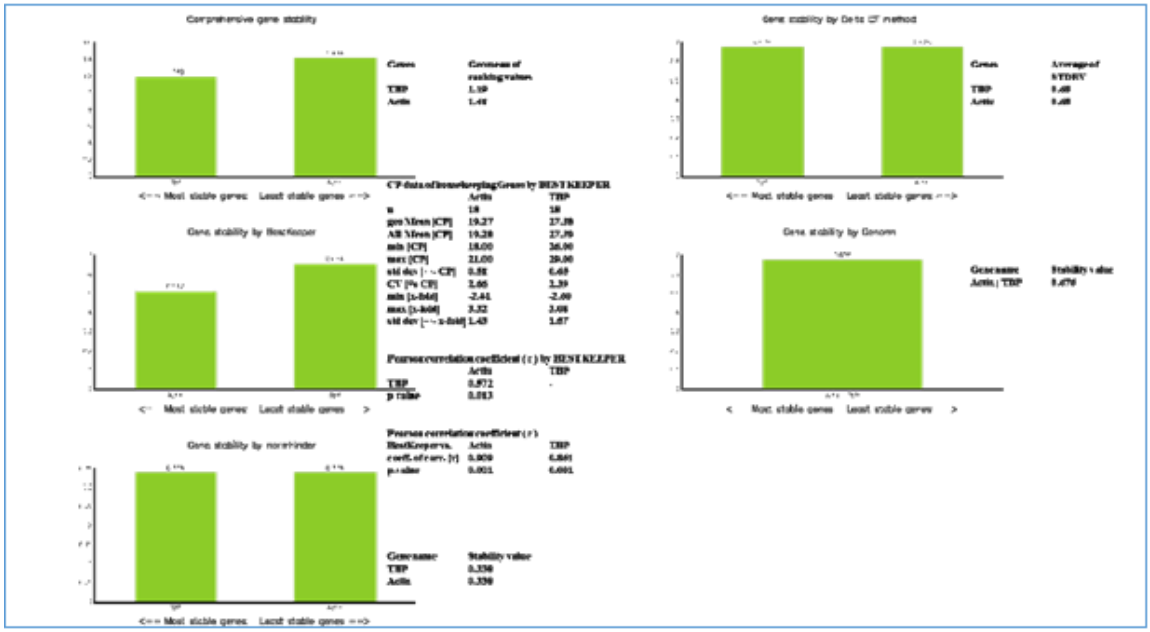
qPCR standard curves:

Gene	Standard Curve			Ct range
	Slope	Efficiency	R^2	
Actin	-3.424	1.9591	0.999	19-21.5
TBP	-3.381	1.9759	0.976	26.8-29.1
PgG/Hs_2	-3.271	2.0217	0.979	23.9-27.6
$PgEs_2$	-2.578	2.4429	0.944	26.6-28.8
PgFs	-2.796	2.2785	0.974	25.2-28.5

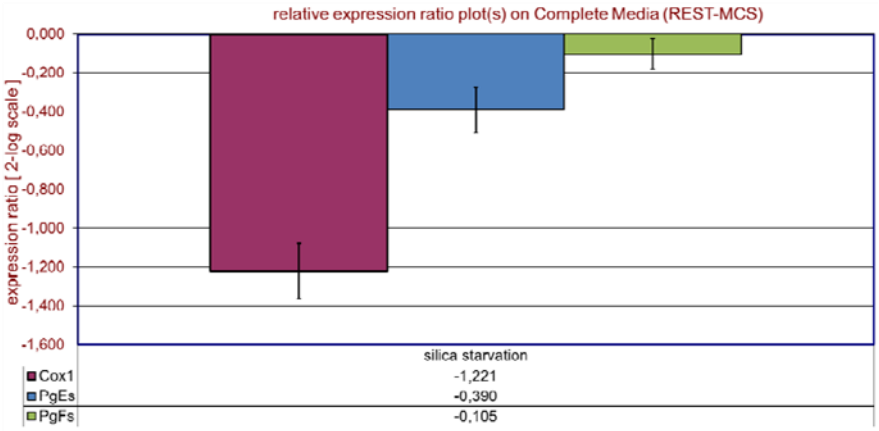
Reference genes stability:

Ranking Order (Better--Good--Average)

Method	**1**	**2**
Delta CT	**TBP**	**Actin**
BestKeeper	**Actin**	**TBP**
Normfinder	**TBP**	**Actin**
Genorm	**Actin | TBP**	
Recommended comprehensive ranking	**TBP**	**Actin**



qPCR Relative Expression Analysis by REST tool on Complete Media:



Supplementary Information 3 (SI3):

Sequences of the Reference genes and of the Transcripts in the Selected Pathways

S ribosomal protein-RPS

>TR58414|c0_g1_i2
TATCTAGAGCAGGGTCCTAAACTAGTTCGTAGTTCGAGCCTATGGTAATGATCGCATGGACG
TGATGTTTTACGACACATTATTATTATTTTCAATTTCCGTCCCATTTACAATTCCAGTGCTTCT
TCAGCTCAGGAGTGAATCAAGATTATCACTGGCATCATCCGAACTCTCTTTATCGCCACTATC
CATGCCGACGCTCAAATCATCCAATGCATCTCCCAATCCCAAAACAATATCCTTCGCGCTAG
CCTCTCTCTTCTTGCGTTCCGCCTCTTGTTTCTCCAGGTACACGCGAGCCGTCTCGTCAGCCAT
CTCATAAACGCGTTCCGGATCCTTGATCATGTCACCAGGCTCCGGTTCCAACGTCTTGGTGCT
GAGTGCTATGCGGCCATTCACCTTATCATGATCGATAATCATACACTTGATCTTCTGCCCCGG
TTGGAGGACGCGCTCAAGATCGTTGATGCGATCCATGGAAATTTGAGAAATGTGCAACAGA
CCCGACATGCCACCAACCTCGACGAAGGCTCCGTAAGGCTTGAGCGCCTTGACAATACCGTC
GATGACATCGCCCCGCGATAGGTCCTCCATCTGAGCCTCAACAACAGCACGGCGGTTACTCA
CAACAAGCTTGTTATCCTCCTCATTGACCTCCAAAAACTTCAGCGGAAGAACCTGTCCAATG
AGGTCCTCGTCAGGCAGGCGTCCCGTTAGGTGGGATCCGGGGAGGAATGCACGAAGCCCCT
CCACGAGGAAGATGGCTCCTCCGCGGTTGACGCTGACGCACTCGGCCTCAAAGGCCTCGTCC
GTATCCTGCTTGGCCACAACCTTCTCCCATGCCTCGCGGTACTGAATCCTTCGCACGGATACC
ATGAGCTGGCCGTTCTCATCCTCCTCGGAGATGATCTGGAAGTTTCGCTCTTGGTCGATTTCG
ATGAGGGATTCGATATTCTCACCCTGGACCTGGATGAGAGCAGCCTCCTGGAGTGGGAGGA
AGGCGCTGGCCTTGGCTCCAATGTCGACGATGCAGCCACCCTTGTCGTACTGGACGACGGTT
CCCTTAACGACATCATTGCGGTTGAAGGAGTAGTCGGTGGCGGCGACGGCCTCGTCGAAGG
AGGCGTAGGTGAAGCCAACTTCGTCTTCAATTTCGGGGGAGATTTCGAGTTGGGGGCCCTTA
TATTTGGACTTGATGCGGTCAAATCGGGTCATTTGTTGTACCTTAGGCTTTTCCGCCACTTCC
GAGAAGAGGGCTGTGGAATGGCGAGTTTGTCCAATGGTAGAGAACGCCGACACGGCACTCC
CTGTGAGAGCTACTAATGCGAATCCAAGCTTCATCATTCAATAATAAAACCGTATGAACTGG
GGGATTTATCGTGCCTTAATTGCGATATCAGACTATTAATAGTGGGGATATGGCGCCGTGTC
GTTGTATTGATACGGTGTATGAAGGCGGACGGTGGGGGAGATGGTAAGTCGATTATGTT
GGGAGGGCGTTGGTTGATGCGTGAG

HISTONE H4

>TR7889|c0_g1_i1
TAACATACTGTCTCTAGTGTATAAATGACAAGTGGAATCTCAAATTAGGAGAACTGACATAA
ATTAGAAATCCCATTGACACTGATCAAATCGGCCTGGTTGGATCACCATACTTGTTACCCTAT
TTTTTTTTTCCACAAACATGAGGTAGTACATCTCTCGGGTGAAAAATCATGATCAACTGGCC
GTTCAAGCAAGTTCTTACATGACACACAGTGTATGTAATTTGGTATCAGATCTAATCGAAGT
ACGAGCATCGGCTGTGAACAAGGATACAACAGTCCGTCACTGCTGAGATGCACATTTGCTTG
GTGATATTGATTGTCGTCACCTTTATTGTAATCATGGAGAGTTACCCTGTCTTCAGCACACAC
ATTCTCACACAAAATCGCAATCGTCATAAACAAAAGAAGTTAATCATGTCGTGGGTTCGGAT
TGCGTTCGAGAGAGATAGTTCGCACCAACTAAGGCGCAAAAATACATGCGTACAAATGTTTC
AGGGGTTTCGATTTTTGGTCTTATTGGATTTTCAGTTTCTGTCCTGGTTGCTAACCACAACTCT
ATTCATATTTCCATATCCCATAGTCTCCAACGAAACCCTAGTCTCCCTGTCAATTCCAATCGT
CTCCTGCTCCAAGGCACTGACCTCAGGAACAAAGT
CCAACCTGCGCTCACATTCCTCCTCCCGCAACTTCAACGAAATGTTACGCACGAGTCCCTGA
GTAATACGCTTCATCAAATGGGTAGTGAATCCGGCAATCTTATTACGCATGCGCTTCGAGGG
GGTGATGGCAATTTCATCGATGATACGCTTGTTGGTGTAGAAGTCGAGGGTGAGACGAGGGT
AGTACTTCTCGACGATGGTACGAGCGGACTTCTTAACAGTCTTGTTCGACATGTCTGCTGGAT
AAAACATTAAAAAAGCGCGCGAGTGTTTCTTTGGCAGTTGGGTGGATTTCAGCAAATCCGGT
TGCAGGGGGGTAGAGGCATCATGTCGCCATAATATCTCAATTTTATCTCTCTGTCACGGCAG
GTAAATTATCCACCGAAGCCATAAAGAGTCTTTCCCTGGCGCTTGAGGGCATAAACAACGTC
CATAGCGGTGACAGTCTTGCGCCTGGCGTGCTCCGTGTAGGTGACGGAGTCACGGATGACAT
TCTCAAGGAAGACCTTGAGCACGCCACGCGTTTCCTCATAAATGAGTCCGGAGATGCGCTTG

ACACCGCCACGACGGGCGAGGCGACGGATGGCGGGCTTGGTGATGCCTTGGATGTTATCGC
GAAGGACCTTGCGATGACGCTTGGCGCCTCCCTTTCCTAGACCTTTGCCTCCTTTTCCGCGAC
CAGACATGATGAAGTGATGTTGTTGTTTGGGTTACTGTGACGATGACAATCGTTTTATGGCA
AACGGAATGTAGGACCCATATATCGTAC CTAAGGTAATCGTGAGA

Tubulin alpha chain

>TR11532|c0_g1_i10
CATATTCAATGATGTGAAGGGTGGCGCATAAAATAGATACGCATATTAATCAGCACCCGCA
ACTCTTGTCACGAGAGCTTTAGCGTCTGCGAATCAAATACTCTTATCCCCTTCAAATCCTCTA
CCTTTCATCATGCGCGAATGTATCTCCATCCATATGGGCCAAGCAGGTGTCCAAACCGGTAA
CGGCTGCTGGGAACTCTACTGCCTCGAACACGGCATCCAACCCGATGGTCAAATGCCTTCCG
ACAAGACCGTCGGAGGCGGAGACGACGCCTTCAACACCTTCTTCAGCGAGACGGGCGGAGG
GAAACACGTCCCTCGCGCCATCTACGTCGATCTCGAACCCACCGTCTGCGATGAGGTTCGTT
CCGGGACGTATCGCCAGCTCTACCATCCCGAGCAGATCATCAGCGGGAAAGAGGATGCGGC
CAACAACTACGCCAGGGGACACTACACCGTCGGCAAGGAGATCGTGGATCAAGTATCGGAC
AGGATCCGCAAGCTGGCCGACAATTGCACAGGCCTGCAGGGATTCCTCATCTTCCACGCCAC
TGGCGGCGGAACCGGATCCGGCTTCGGCGCCCTCCTCCTGGAGCGTCTGTCCGTGGACTTCG
GTCGCAAGTCCAAGTTGTCGTTCGCTGTATCGCCCTCTCCCCAGGTATCCACCGCCGTGGTCG
AGCCATACAACTCCGTCCTTTCCACCCACGCTCTCCTGGAGCACACGGACGTCACCTTCTGCC
TCGACAACGAGGCCCTGTACGACGTCTGTAAGCATAGCCTCGACATAGAGCGTCCCACCTAC
ACCAATCTCAATCGCCTCATCGCCCAGGTCACCTCCTCCCTCACGGCCTCCCTTCGCTTCGAC
GGCGCGCTTAATGTCGACGTGACCGAGTTCCAGACCAACTTGGTGCCCTTCCCCCGCATCCA
CTTCATGCTCACCTCGTACGCGCCCATCATCTCGTCTGAGAAGGCCCATCACGAGAACCTCT
CTGTGGCCGAGATCACATCGTCGGTCTTCGAGCCTCAATCCGCCATGAGCAAGTGCGATGCT
CGCACTGGCAAGTACATGGCGTGCTGCCTCATGTACCGTGGGGATGTGGTGCCCACGGACGT
CAGCGCGGCTGTGACCACTGTGAAGACCAAGCGCACGATTCAGTTCGTGGATTGGTGTCCCA
CAGGATTCAAGTGCGGGATCAACTACCAGCCGCCGACGGTGGTGCCCGGGGGTGATCT

Tubulin beta chain

>TR22371|c1_g1_i1
CTTCCGATCTCAACGCCAAGCCGAAATCTTTTGACTGCTTTGCCTTTGCCAAGATCCCAACCA
TTACTCCCCCCACGCCGTATCGCCAGAGCAATCGATGCCTATAACCACACAAGAGAGAAATA
ATTAATCAATTCAACCATGAGAGAAATCGTTCATCTTCAAGCGGGACAATGCGGAAACCAA
ATCGGGTCCAAATTCTGGGAGGTTATGGCCCAGGAGCATGGGATCAACGACCAAGGAGCCT
ACCATGGTGACAACGACCTCCAACTTCAGCGCATCAACGTCTACTTCCATGAGGGCGGAAAC
GGCCGATATGTTCCCCGTGCTGTTCTCACTGATCTTGAACCGGGCACCATGGACGCCATTCG
CGCTGGACCATTCGGGGGAGTTTTCCGACCTGACAACTTTATATTTGGGCAGTCTGGGGCAG
GTAATAATTGGGCCAAGGGACATTACACTGAGGGTGCTGAGCTGGTGGAGTCCATCATGGA
TGTGACGCGTAAAGAGGCAGAGGCATGTGATATGCTCCAGGGATTTCAGATCACACATTCCA
TGGGAGGTGGAACCGGAAGTGGAATGGGAACGCTTCTTGTGAGCAAGATCAGGGAGGAATT
TCCGGATAGGATCATGACCACTTACAGTGTTGTGCCCTCCCCGAAGGTGTCTGATACCGTGG
TAGAGCCTTACAATGCGACACTATCGATCCACCAGCTTGTGGAAAACGCAGATCAGTGCTAC
GCTCTCGACAACGAGGCCCTCTACGATATCTGCTTCCGCACATTGAATCTTACCAACCCAAG
CTACGGCGACCTCAACTGCCTCATCGCGGCGGCCATCACTGGGACGACTTGCTCCCTCCGCT
TCCCTGGTCAATTAAACTGCGACTTGCGTAAGCTGTCGGTGAACATGGTCCCATTCCCCCGTC
TTCACTTCTTCCTCGTGGGATTCGCTCCCTTGACCTCGCATCGCTCCCAGGCCTTCCGTGTTCT
CACCGTCCCTGAACTGACCCAACAGGCCTTCGATGCCAAAAACATGATGTGCGCTGCTGACC
CCCGCCATGGACGATACCTCACGTGCGCGATGATGTTCCGTGGAGCAATGTCTAGCAAGGAG
GTTGATGATGAGATGCTTAAGATGACGCTGAAGAATTCTGGATACTTTGTGGAGTGGATTCC
CAATAACCTCAAGGCAAGCATTTGTGATGTACCCCCAACGGGATTTAAAATGTCGAGTGTCT
TCATTGGTAACTCTACATGCATTCAGGAAGCCTGGAAGCGTGTGGCTGAGCAGTTCACTGTT
ATGTTCCGTCGTAAGGCCTTCCTTCATTGGTACACTGGAGAGGGTATGGATGAGATGGAGTT
CACTGAGGCAGAGAGCAACCTGAATGATCTTGTGTCAGAGTATCAGCAATATCAGGATGCT
ACGGTTGGGGATGACGATATTTCGGGGGAGTTTGAGTATGAGGGTGATGACCAAAACTTTA
ATGATGACGTCGCAAACAGCCTTGAAAACAATACGGACAAAGTGGCCGATGAATTCATTGG
CATTGTGGACCAACAAACTGACCAGCTGGAATTTGATATCCCTACCGAACAACCATTGGAAT
CAGAACCTTAAGGATGAGAAACATGTAGACTTTTCGAAAGAATTTTTGTGGTTGAGGTAGAC

TTTGTTGGAATGTGAATGATATTATTTGGGGTGCTATGAATATTGCGCAAGCATTCTCTAGTG
ATCGATTTTGTGGGTTGGCGTGTGACATGTGATTGTGAGTCTTTAGTACTTGACTGTGAATGT
GTTGTTTGATGAGACAACTTGAGCGGGCCAAGATTGACACCCTTCGTGTTGGGACACCTTAT
TGCACCAATAAGTGGCTAAATATGTAGATAGCTATCGGTTCGATGGGCATGTTTTGAAAGCA
TGGATACGATTTTTCGACAATTCTCCTGCAAAGTACATGCTGAAGTGCGTTGTAAAGGATAT
CAATGAACACATAGATGTCAGGTCAAATCGAAATAGTACAGTAAACACTATATTTGTTGAAA
AACGGTGCTTCGGTGGGTTCATGGTGGTTCAAATTGTGGTATTCAATAAAAAATTAAGTCCT
ACTACAGACAAAAAAAAA

TBP (TATA box binding protein)

>TR38216|c0_g1_i2
GCTCTTCCGATCTCCTACCTCTCCCAGACGATGATGCGGCCGCATTCTTCGCACTCGCAGCCG
CCGCACTCGCTGCCTCCTCCATCGTGGGCAAACTCGCCACATTGGCCTTGCGAAATTCCAAA
AGAACCGGATAAACCTTCTTCAACCCCTCCACCAACTGCTCTTCCGTCTTGGCACCGGTAAT
AACCACCTTGCCGCTCACGAACACCAACAGCACTACCCTGGGCTGGACGAGCCTGTATATCA
ACCCCGGGAACAACTCCGGCTCGTACGACGAAAACGTGGGATGAGCGAACGCCACTCCCTC
CAACCGAATGGGATACCCCACATCCACCGTTCCCACGATGTTCTGCACTTTGAAGTTAATGT
AACTCGACGGTTGGAATCCCACCCTTTCGAGGATGTAGGCGATTTTCTTCGTTGCCAAACTG
GAATTGTGACTGCTTTTGGTTCCCGTGATGACTAATTTCCCCGAGGCGAATACGAGTGCGGT
GGCGCGGGGTTCGCGGAGGCGCATTATGACGGCGGCGAAACGACGGGGGTTGAATTCGGTG
TTTCTGCATCGGAGTGCTATTTGTTTTAAATCTAATCGCACGCCCAGGTCGGCTATTTTTGAC
ATATGTACGTGGAAGGGGAGAGATTCATGTGATGATGATGAGCGTTTTCGATAGAGTGTTAA
GGATTGGAAAATCGTTTCATTATACAAAAATC
ATCCCATCATATATCTGCACAAACATACATGCAATCCTGATCTCTCTCAC

GAPDH

>TR4255|c0_g1_i1
TCTTCCGATCTTTTTTATTCAAAACTGGCTCCTATTTTATCTATCCTTCGTAGCTTCAATCACA
TCCATTACTTATTTCTGACTCTCCATGTGCGAAATCAAATCAAGGACACGGTTCGAGTATCCG
GCCTCGTTATCATACCAGGAAACCATCTTGACAAAGTTATCATTCAATGCAATACCTGCCTTC
TTGTCAAAGATACTGGAATGGGTATCACCAATGAAGTCGGACGAGACAATGTCGTCATCATG
GAATCCCAACACTCCCTTCATCTCATTCTCGGACGCATCCTTGATCTTCTGGCAAATGGTCTC
GTAATCCGCACCCTTGTCCAACACCGCCGTCAAATCCACCACACTGACGTTGGCGGTAGGCA
CACGGAACGCCATGCCCGTCAACTTCCCATTCAAATCCGGAATGACCTTACCCACCGCCTTG
GCAGCACCGGTACCACTGGGAATAATGTTGTAACACGCGGCGCGTCCTCCCCTCCAATCCTT
CTGCGAAGGTCCATCCACCGTCTGCTGCGTGGCGGTAACGGCGTGCACCGTCGTCATGAGCG
CCTCCTTGATTCCGAACTCGTCATTGACCACCTTGGCCATGGGAGCCAGACAGTTGGTCGTA
CAGGAGGCGTTGGAGACAACATCCATGCTGTTCTCGTACTCCTTGTGGTTGACTCCCATGAC
GAACATGGGAGCGTCGCCGGAGGGAGCGGAGATGACGACCTTCTTGGCCCCGCCGGCCATG
TGCATGCTGGCCTTTTCGGTGGTGGTGAAGACTCCCGTGGATTCGATGACGTAGTCGGCTCC
AGCGTCGCCCCAGGCGATTTTGGAGGGGTCCATCTCGCCGTAGATTTGGATGGGTTTGCCAT
CGATGGAGAGGGTCTTGGCGTCCTTGTCGTGGGTGACGGTTTGACCGCATCGGCCGTGGACG
GTGTCGTACTTGTACATGTACTCCATGTAATCGACCGAGATGAAGGGATCATTGACGGCGAC
AATGTTCACATTCGGCTTGTTGGCAGCCGCGCGCATCACCAGGCGCCCGATGCGCCCAAATC
CGTTGATTCCAATATTGATAGACATGATGCTGATTATAGAGTTCTATGAAAAAGTGAATATG
AGCTAAATT

Actin

>TR18494|c0_g1_i4
CCCGACAGCCTGGAATCATGGTGCGTGACTGAGACTCGGACATTTCCATTGATGATGTTGGC
GCTCGCCCTTATCTGTCCAAGGATCTTTTGCTCACATCACACACCTCTTTTTTCACATGCGTCT
AACAGGTGGGAATGGAACAAAGAGACGCCTACGTCGGAGACGAAGCCCAGGCCAAGCGTG
GTGTCCTCACCCTGAAATACCCCATTGAGCACGGAATTGTCACCAACTGGGACGATATGGAG
AAAATCTGGCATCACACCTTCTACAACGAGCTTCGCGTCGCTCCCGAGGCCCACCCGGTCCT
CCTCACAGAAGCTCCCCAAAATCCCAAGGCGAACAGGGAGCGCATGACCCAAATCATGTTC
GAAACCTTCAACGTCCCCGCCATGTACGTGAACATCCAGGCCGTCCTTTCCCTCTACGCCTCC
GGACGTACCACGGGATGCGTCCTCGACTCCGGAGACGGTGTCACCCACACCGTCCCCATTTA

CGAGGGGTACGCGCTCCCCCACGCCGTCGTCCGTCTCGATCTCGCCGGACGCGATCTCACCG
ACTACCTCATGAAAATCCTCACGGAGCGTGGTTATTCCCTCACCACCACCGCCGAGCGCGAG
ATCGTCCGCGATATCAAGGAGTCTCTGTGCTTTGTCGCTGTGGACTTCGAGGAGGAGATGAA
GAAGGCTGCGGAGTCCTCGGCCCTTGAGAAGAGTTTCGAGCTTCCCGATGGTAACATTATTG
TTATTGGTAACGAGCGTTTCCGTTGTCCTGAGGTCCTCTTCCAGCCTAACTTGACTGGACTTG
AGATGGATGGAATTGCGGACTCCACTTTCCAGACCATCATGAAGTGCGATGTCGATATCCGT
AAGGATTTGTATGCCAACATTGTTCTGTCGGGAGGTACCACCATGTTCCCTGGAATCAGTGA
GCGTATGAGCAAGGAGATCACGGCGTTGGCCCCTGCATCTATTAAGGTGAAGATTGTTGCCC
CACCAGAGCGTAAGTACTCTGTGTGGATTGGAGGAAGTATTTTGGCAAGTCTTTCGACTTTC
CAGAGTATGTGGATTAGCAAGGAGGAGTATGATGAGAGTGGTCCCAGTATTGTGCACAGGA
AGTGCTTCTAAGTAGGCCTGAGCGCAGATGTTTCTTGGATATGTGACATCATGCGGCCCCAG
AGGCGGAATTTGTGGTATTTGTTGTGAAGCACAATAGATGAGCGTACTGGTGAATAGTGATA
GTGTTCCACAGAACAATCATCTAGGTAATTAAGTCCTACTGTTATCAAATCAAAAAAAAAAG
ATCGGAAGAGCGTCG

COPα

>TR6988|c0_g1_i2
CCATTAGACATGGTATCTAGTCATAAGACAATACGACAAACTATGTAGCTTAGGTGACCAAA
ACTAATAGATATCATCAGGTGGGCGAGCTATTGGTTTGCCTCTGAAAAGATTTTTTGAAGAT
GAAATGCTAAGGGGGGCTTAGACTTGACCCTGTGGCTCAGTGATGTGATATCTGCTTGGCCT
AAGCGAAATAGACGGATGACCAACAAGCAAAAATTAGCAAAATACTATCCAGTTACAAGCA
GTGTTTCAACGCCTTATCATAGTGCGTAAGACATCTCTCAATAAGAGTAAATAATAATTTCTT
TCTGGATGCAAGTAATAGCAGAGGAATATCATATACTCTATCATCAATTTTATATGATCAAT
TTTTCGTTAATGATAGGGCATCAGTGGGACAGCCAACAACTTATTTTCCAGTGACAAGCCCA
ATCGTTTCAACACCAATTGTGCAGAAGCCATCTGTCAGATCGACATGTCCTTTCATTTCTGAT
CCCTTGTAAGAGGATCCAGAGTAGGAGCATTTGACACTCTCTTCTCCTGTGTAAATGGGAAC
AAAATCTTTACAATCGATTACAAATGCAGCGTTCTCATTGTAGTTGAGGGTGTGTTCATTGCG
TGCCATCTGTTCACTCTTTTGAAGCACCTTTGTGGCCTTCATCCTAAGATCAGCATTTTTCGC
GCTTCTCATGTCAGGGAGCTCCAGTAGACGCCTGGCAAAACTAGCTGCAACAATGTAGTTCT
TGTGTTTGAAGGAAGTTCCCATGGCAGAGCGAAGGGCAAGAAGTTTGTGTACAGGCTGCAA
GTCGCAGTGAGTAAAAAGGGCAGCAAGTTCTGTCGATCGGACAGGATCCGCACCAGCCTCG
GCCATTGCTGCCTTGATGCGAATTGCAGTGATATATTCGCGACAGATCCCTAGCATCTCCTTC
ATTTCACTGGCTTCGTTTTTGTTCTCTGCAGCAGTGAGAGGGATCTGGGCCATGACTTCAGTA
AATGCATTCTTCGCATCGTTGAATTTTCCAGTTTGGAAGAACCTGTATCCCTGGCGGATACCT
GTGACAAGAGTCCTCATGTTAAGTGACGTGCGAGGTAAGCTGTCATTTCCTGGATGGCCTGA
CGCGTCATTCCTCATGAGAGGCACTACAATGCTTCCGCTGCCAGGGATACCTGGTAAGCTCA
CAGTGGAGCCAAGATAACAGCCAAGCATGGACCCTTTAAGGACAGAAAAGTCACTGGCGGC
AATTTGTCGATGAAGCAATTGCATAGCCGATGAAGCACCTCCAGCTGCCATGTGATCAGCGG
CATGGGAACTGTTCGCTGCCCAGCATCCAGCGGGAGGGCGCCCAGCAGCAGGCATTTGAAA
CCCTGCGTCATCTGAAATGTCATTGAGACCCGCCATTTCGTCAACTGGGGCAGCCTTGGGAG
TATCATCTCCGAAATCGTCTCCAAGATCATCCAAGTCATCACCCCAGCCCATATCATCATCAT
CACCCATATCAAAATCATCAGCTGCTCCGCCCAGGCCTCCATCAGCGTCACCATCCCCCGCG
GCAAAATCGTCCTCCCAAACAGCTGCTCCAGCAGTGTCAGCAACTGGAGCAGCTTCTTCTTC
TTCATACTCATCACCAGCATCTGCAGCTGATAAGTCTTCAAGTGTTGTCTTCTTGATCTCAAG
AGTTGGCCAGTTACCACCTCTGATGATAGGTGTTGGTGGTTGGAGGAGGCGACCACCCTTCT
TTGTGCCAACAGGGACCTTGTCCATCAGTCCTTCCACTGAACCTTCCTGGGTTTCAATGGCAA
CCTTGATTCTTTCCGCAACATCTACCAACCCATGAAGATTGGCACTGATGTAGGCCAATGGC
AAGTTTCCACTGGCCTCCAAGACAGCAACACGCTCAGCAGCATCACCAAGAAGGAGGGCAT
TGTGGTAACGGCCCATGATATCATTGCGCATGTTGGAGATCTTGAGCATTTTTCTCAACTTCT
CAGAGTCACCAGTGAGCATGTATAAGAAAGATAGTCGATCGAAATCCTTCTTGCGTTGATAG
CTCATCTCCACAACCTGATGGTTTCCTTGGCGGAGAGCTTCGCTTCCAAGTTCCCCCCAGACA
TCGCGTGCTTTGCCATCCAATCCTTGCTGCTGCTCCAAAGAAAAAGCACTCTCCATCGCAGC
CTCAATATTTCCACAAGCAAGAGCAAGACGGAACCGAGTTTGCGGATCACGGACAAAGTGG
AGTGCCACTTCAGGAAAGCCTTTAGATTGAAGGTAGGCGACAATTGCCCGGCCACAGAGAC
GGCTGTGTTTAACCATGTGCATAACTTGTCCATACTTTTTGTTTGCAAGAGCAAGCTTGAAAC
GGGCTTCAGTTGTGTCGATGCTGATGACTCTGGACCTGGCTTCCCTGTCAAGGCAGAAGAGT
TGATCCTTAACCACTCGTGTAGCATAAATGGGATTGTCCAATGTACGAATAGTACCAGTATC

GCCAGTAGAAAGGCAATATTTCACGTGGTGAAGGGTGGTGTAAATGAAAAGATTGGAGCTG
GTGCCGCCAGACGGACTGACATCCCATGCACCTGACTTGATGCGAACAGTGTCGCTGATTGA
GCATAGCTGTTCAAGCTGACGGTTGGCAATAACGACTCCATACTTGCAGAGGAGGGCAACCT
TACTGCCGTCGGGAGACCATACAACGTTCTTAATCTTGGGTGCTGTCAGTTCACCAATGACG
CGACGACTTTGGACTTCGAAGAGGACGGCGCGATCGTCCGCACGAAGAATGATTCGACCCG
AGGCACCACCATCATAAAGGCAATCAACATTGGGCAAGGGCGCCTGAACACGCTTTGTGGT
TTCGTTGTTGAGGTTCTTGACAACAATTTCACGCTTGTGCCGATCAAGGATCGCAAAGCGGT
TACGTCCAAGGAAGACTCCAGGCCCAAGACAGGATCCTCGTTTGCCATCAGTGACAGATCCC
CCAGCATTAGTGAGAGAGAAGGTGATGAGTTCATAGCATCCACCATCAACTTCAGAGCAGA
CTAGAACATTTCCTTCGGTGGGGTTGTGCATATTGTAAGAGAGGAAGCGAGGACCAGACCC
AATTCCATCCGTCTGCGCCTGCTGTCCCACCCTGCGGAGGGATGCAATGGGCACATCCACGC
CAGAACTTCCGCCCGTGGCCCGACCATAATCGTGGCAGAAGAGCTCACGTCCGCGAACATA
ATACAATTTGTCTCCTGGTCCGTTGGCGGATGCAGGACGCTCCCTCTCAAGCTTGAACACGA
TCATACCGGAGTCATGACCAGCCGCAAGCAGATTCTGAGAGCGATGAGCGGCTAAGATCCA
GAAGCGGTCCCCCTCCCGACGGAAGGTTTGCACGCCAACACGTTTGGACACATCCCAGACGC
GAATGCTGCGATCTTCTGAGTTGGAGACGACGAGTTCTTGTTTCGGGTGGAAGAGACAGCAC
GAAACATTATTGGCATGGCCACGGAGGGTGTCAACCTCCCAAGCCTTGGTCTCACTCATTCT
CCAGAGTTTGACCTGACGATCATCAGCAGCACTTGCCAATAAAGGCAGGGTAGGATGGAAG
GCTGCCCAATTCACGCCACGGTCGTGTCCCTCCAAAACATATTTGACCACAACGTCATTGGT
ACCAAAGAGCTCTGCCTGGACATTCAAAGCAGTTGCCGCCGCAGCCGATTGACCACGATTGC
CATTGGAGTCCGTCATGCCCGGATATCCTTGGTCTGCTGCCGCGGGTCCCCCTAGCTGCTTTT
TACGAAGACCAGTCGTGTCCCACACTCGAACCGTCTGATCCAAACTGGCACTAACAATAAGA
TCTTCCGATGGGTGGAAGGCTGCACACATGACATAATGGTTGTGACCCGTCAAGACACTCAA
ACAGCTCCTCTTTTGAAAATCCCACAACCTCAAAGTCTGATCGTCACTAGCACTCAAAATCC
ATGGAAATTGGGTGGCATTGGGGTGGAACTGCACCGTACGGATGTAATCCAAGTGACCCAA
CAATGTGAAAAGGCACCTCCTAAGTTTGTAATCCCAAACCTTGATTTTGTAATCATCTCCACC
AGTGACGATGAGAGGTTCGTGGAGATGAAAGTCTACGCCACGGACGGGACCATCGTGTTCT
TCAAAGCGGTCTATGACGGTGCCTACGCGGTAGTCCCATAGTTGAATGACACCATTGTGTAA
TGAGGCGCATATCCACGGCCGCACCGGGTGGAATGCGAGGCCCTTCACGCGGGCTGATTTAC
TCTCAAACTTCGTTAGC

Ubiquitin like protein

>TR31138|c0_g1_i1
TGAAGCTTGAAATACGCTGAATGCAAGCTAGTTTAATTCTGCTACTTTTCAGACGCGATTTCA
AACAGAATACCCCGGCCCATGATGATGAGGAGGTTTCTTCAATTTGTAAAAAGATCATACAT
TTTTCTACCACAGACCTACTTTCAAACCAAAACACGCCATCTCCTTTTGAATTTAGGCGAGCT
TCATTCTATTCAACGCCCGCAATAAAGTCCCTCGATTATCCCTCCTCTTGTAAATATTCGTCA
ACTTCGTCTCATCCTTCCCCAATAATCCAAATGCGCTCTTGTGCTCCACCTCCACCAACCTCG
ACCCCCTGCCGCCCGGCAAATCAAACGCCCTATCACGAGGCTGGGCCGTGCGCATGCCATAC
TCCAATTCCTCCAGGGGGGACGTGTCGCACGTACGGCGAGGTAGGGGAGCGATGCGATAAT
ACGCGCCCAGGACGGGCGCCACGCCGTCAACGACCGCCGCCTCGACTTCTTCCGCTTCGTCC
TCCTCGCTGTTGGTGCGGCCGAGGCTTTGGTCGATAAACATGTTCATCAATTCGCGTTCTTCC
TCCGTTGTTCCCTCCAAATTGGACTTGACGAAGGATTCGTCCACGCGAAAAACGCCCTTCCC
GAGCGCCTTTTCGAATGCGCTCTTGACACCTCCGCCGCCACGCAGCCTCAAAACCAAATGAA
GCGTGGCGTCATCTCCGATATTGTAATCGTCGACCGTCTTCCCGTCCTGGAGCTGCTGGCCAC
CGAATATGATCCGTTGTTGTTCAGCCGGGATGCCCTCCTTCTCGGCAATCTTTGCCTTGACGT
CCTCGATGGATTCGCCCTCCTCGATTTCGATGGAGACTGTCTTGCCGGTGAGAGTTTTGACGA
AGAGTTGCATCCCGCCGCGAAGGTCGAGGGCGGTGGTGAGTGCGGACGAGGGGCGGACGA
GGGCGCGGCTGCTGCCGATGAGGGGGAGACGATTACCACTGCATGATTGAGTTACCAGGAG
AGATGTTGCCAGGAAGAGGCAGGATTTGGAGGAAACCATAGTGCTATTGAAATGGCCTGGA
CCTTGAGATTGAG CAGAACGATGACGACTTAGTTCT

Translation elongation factor-1-alpha/EF1α

>TR44161|c0_g1_i2
CCATCATCCAGCCATCACCATGATCATCGTCACTGGCCTCCGCCGTAGAGCAATGCGGTAGT
ATGATCATCGCCTGCGGCCCTGACGGAAGGTGACGATGAAACCAAAACACACGAGATAAGA
GTGTCAATGCATATCACCTGCTGATGGGCGAATCAACGAAGAAGGGCTCAAGCTCAACAAT

GCCACATCATCTCACCCTCATCAAGCTAGCCATCTTCTTCCAAACAGCCTCGCAGGCAGGAT
GGAGTCGTTTCAGCAGCATCTACATGCTCAGCCAGGGATTCACTCCATCTCAAATCGGAAGA
ATAAAATCAACCAGTCAAATCGCCAAGGCGCTGGCTCAGCCCATTTGGGGTGTAATGAGCG
ACATAACCGATCCCCTTAGTACCATCATCGTGTCCACCGTTCTCGGTGCAGTGACACTAGCG
GGTGTTCGGGAAGCGATTGATGCGCAGAGTTTTTCCAAGCTCATGACCTGGCGAATTATACG
CTCTGCGGCCACGGCAGCATCTCCGGCGCTCGATGCGCTTGTATTACGTCTAGTCGAGAAAT
CCTCCGAGGGATATGGACCGCAGCGACTTTGGGGGAGTGTGGCATGGGGTTTATCATCGCTT
TGTGCTGGTGCAATCCTGGATCGGTTTGGCGAAGGCGCCATTTTTGGGTACACATACGTTAT
CAGCGGTGTCCTGCTGGCATTTTTCCTGGCGCTGAGAAGGGTAATTCCCAGAAGTAGTACGA
AACGGAGTGACGACGAGTCTCCCCAGAAGAAAGAATCTCAATCTACTGGGTTGATGTTGCG
GGTATGTGGATTTATTCGCACTGTAATCAGTGATCTCTTGAGAGGGACGACTTTGACAGTTT
GGGCGATTTTGATTCACAGTTTTGGATACGGCGTCATTATGGTCTTATTCAATACCATCCTTA
TGCTGCAACTCGAACGTGACTTTGGAATATCACGATCCGTACAGGGTATATTCACCTTGGTG
TCAATACTCAGCACATTGCCAGTGTACCATTATTCGAGCAAGCTCAGAGAACGCATTGGACA
TTTAGGGATGATTCAATCTTCAATAATTGTCGCAACCTTCCGCCTCTGCCTGACAAGGTTTGC
TTGCAGCGAAGCTGTCTCCGATGGCTCAAGAGCAAAATGGCTTTTGTTGCTGCAACTTCTTC
ATGGCTATCATTTTGCTGCAGATTGGGTTGCATCGGTAGAGTTACTAGACACCCTGGCAAGC
CAACGACTTCGTAGTTCTATGCAATTTCTTTTAAATATGGCATATTTCACGGCAGGGTCTGGG
GTTGGCAATTTATGGTTTGGTGTCTTATACGAGCAACAAGGGGGTCACACAACTTATACCAG
GGGTCTCTGGTTTTGCTTGATAGATTTCGCTTTCATTTCGGCATTAAAAGCGTACAATGCGCG
AGTCATGTTTACCAGGAATGATCCATTAGAAGTGGGGATGTCAGAACATGTAAGCAGCAAT
CCATGATTGGGGTGATTAATAAAAAAGAAGAAGTCTGACTGCGCAATACATGTTATCTTCTT
CCATCAAAAACTCAGCCCCTGATGAGACTGACATGCCAGTAATGATAGTTGAGGTTGAGCA
GAGCAAGCACTACAGAAGGAGTACCTAGGTGCCTCAATCCTCAATCATGAGACATAAAAAG
TCTAGCTGCTTGTTGTAAAAAAATCAACATGAGTTCTCAGGGTAGGGTCCGAACGGACAAAG
GCCAAACAGTGCCCTACTCATTCAGTCTTAGATTGATATAGGGCATCAACTCAATAATAGGC
ATGCCCCGCCGAAGATCCAACATAAGAACCTTTGTACTGTGTAATTTATTAAACTCAACACA
CTATACCTTACTTCGGCATGGTGTTGTGGATTTGTAAAGCCATAACCTATCAGCACTCACCAG
CACACAGTATTGGTCTTTGGAATGAATAATAGAGCTATGTGCTAGTTATACCAAAAATTGAA
GACTCCGCTCAAAACAACAGGTTGTTCAAACATAACCTTCTTCGCACTTCTGCGTTTCACCTT
TTCCGGCAACTCGTTAGATTCATACGCCTCGTGATCTTTCGCGCACCATATGGTAACATGCGG
GAACGTATTCAGCGGCGCAGGAAAGGAGGAAGAATGCGAGTTGTTGTCGTTGTTCTTCTTCT
TCGGAATTTCCAACTCGATTGCGGCCGATTCATCGCTGTAAAGCAAGGACACCGCTTTCACC
TCGTGCTTGGTACCAATGACATGTTGAAATGTCGTTATCATACTTTCCTGCGTTGCTTGAGAA
GCATGTGCAAACGTACAATGCAACGAAGTTATGAATCTATTCTGGCTCTTGTCATTTTCGTCG
TTGTTGTTTTTGTTATGCTCCTCCCGATCGGCGAAGTATTGCTCTACTTCCGGGAAGCTCTTCC
GAAGCTTAGCAATAATTTCGGATTGCACCGCTTTGCGATCGAGGTCCAGCGATGCAATTTTG
ATTGTTCCGGGCTTATTATTCGCTTCGGCAGCGGCCAGGATTGTTTTCGAATCTAATAATTTG
TCTGGCAGGGATGCAATCACTCTGGACAGCTCTGAGACAAATACCGTTTTAGATTCCTCCAG
CGATACTGTCAAATTCTCAATATAACTTTGATTACCGTGAATGGACGAACGAAGTCTTTTTTC
CATATCAACAACGGCATCCTTGCCAACTTTAAAGTTGTACACATCCTCTCGCGTTTGCAATAC
GATGGCGTCCTCAAGAGCCATCTTTAGATCTTTCGGTAGATCCGGAGGCATTTCTTCTCTGAA
GAATGGAGTAGTCACTGCTACAATCTCCTTCCCTCCTCGATCATGACCAATACCCAGCAGCT
TTTCCTTCAACGCCGACGTAGACATGTTACGATAAAAACAATAAAACTTCACGACCACCATA
CAAGCATTCTCGACTGCCGAATCGAGTTTGCCCTTGTGGGTATTCGGCGGACGACGTAACAC
ACGAGCTATGCAGACTGCCAAGTGTTGCAGAGAAAATGGATACACGTGCGATGATTTCGAT
GCCCCATTAGTTCCGACTACCAATGTGTCTTTCATTCCCGAAGGAAGGACGGGCAGCGCGAT
GGACTGATTCCCCTCGATGCACAACGTGTGAACAGCCGGAAAGGAGGAAGGGGGGACGTTC
TTATCCAGAATCAAAATTGAATCGAGGTTTGTCAAAATTTCCTTCTGAGCAACGTTGAAAAA
TTTCCCTTTCCCCCGTTTCTTTATTTCGTCTCCTTCCATCAGTATTAGCTTTCTATTGTCGTTCA
TAATACCGATGGATGTAATGTTTTGACATAATGATGATTTGCCCGATCCTGGTAAT
GAAGGAAAATAT

Prostaglandin E synthase 2-like_mPGES-2

>TR25058|c1_g1_i2
AGGCGGAGAAAACGGAGAGAAACCCGCTATCACGCAACCATCATTGACCGAACGACGCAA
AAGACGAATAGAAGAACAATCATTGAGAACTTCCAACTTCCAGCCGCCTCTATCATGACCAT
CTCCAAAGTCACCCCAATGGCGGGCCGCATAATCGGTAGGCACTCTTTTCTCGCGACATCCA

GGTCATCAACGGCGCTCATCGGATTGGAAACACCTTCGATATCAGCCACTCTCCCATCTGCA
GCAGCATCTCCATCATCGCGTCCAGTGCGTGCCCCATCGACTTTCAATTTTTGTTCAACTTAT
TCTACTACACGATGTTTTACAGCTGCGTCATCAACAACAATCCAACCACCCGAGATAAAACT
CTACCAATACCACATATGCCCATTCTGCAACATTACCAAAGCACTCCTATCCTATTCCAAAGT
CAACTATCAATCAGTAGAAGTCAATCCATTGACAAAGGCAGAGTTAAAACCATGGTCGGGA
GATTACAAAAAAGTTCCGATTGCTTTGATAGATGAAGGACAAATCAATGGCTCGGATAAAA
TTATCGATTCTATACTAGAGTCGCCTCATGTCCAAACTATACTGGAGGCACGATGGGCAGAA
CAGTGCAGTGGCAGTAATGACGACAACGAAAGAATGACCATGCAACAATTCCACACGTCCG
AAAATGTCCAAAAATGGCTTCAATTCGCTCGTGACGACCTAGCCGCGATACTGTATCCTAAT
ATTTGCGGCACTCTAGGCGATTCTTACGATGCATTCGGATACGTCAATAATGTGGATTCATTT
ACGGCAATGCAAAAAACTTCCATTCGATGGATGGGTTCGCTGGCAATGTATTTTGCAGCGTC
CAAGATTAAATCAAAGCGAAACATCACCGATGAACGAGCAAGCCTCCATGAAGCTCTCGAT
AAATTCGAAAGCGAAGGACTTCAGGATGGAAAGATGGCGTATTTATCAGGAGTGGATTGTC
CGGATATGGGGGATGTGGCTGTGTTTGGCGTTCTCCACTCTGTGAAGGGATTGAATGCGCAT
GATGATGCAATTGGATCACGAGGGGGTCACGTGAAGGATTGGTATGATAGGATGCATGTGC
AGGTGGTGGGAGAAGAAGCAGAATAATAAAATACCTGATCATTGATCAACATTAGAACGGA
TTTCATACTTAATCGGATGTCATGAACTCTGGTGCAACGACTCTATGGGACTGTATTCTGAAC
AATCCAATTATCGTTCTCTCAAACTGTATTGGTACGCCATCGTATCAAGAGGAAAGATGATC
AAATTGGAAGTGGATAAAAGTATCACATACCTAATGGCATCACTTTAATCGCAATGATTCCA
AACAGGGCAAGTTACGATTAAAATTAGACGCCGAGTTTTGCAGCAGAGAGTTCATCGCAAA
GAATCCTTCCATTTTATCATCACACATCCCTGCCAAACTCTCCTAATATGTTACCAATTCCTT
CACCTTTTCTCTATCCTCACCAACGCTGTCATCCCTCCAATCTGTCTCGTGCAACTCTTCTTCC
TTCACCTACAAAATAATCTACAGAGGCATCATCACTGCAACTGTTGTCATTTTACCAACCAA
GTAATTATTTGGAAATGTGCGATTGAGTAGGTGGTAGTGATGTGTAATCCATGGTGAAAGT

Prostaglandin F synthase_PGFS

>TR60026|c0_g1_i1
GCTTTTAGTATCGTACCAACGATACCATTGCAATTACCATCATTGGCGTTGTGAACTAACACC
CAACCATTACTCCTACGCAGCAACAAAATCTATTGCGCATCCCGCTATCCCTTCGAAATGAA
GCTATTTATAAAGTTAGCGATAGCGCTGACGCCAGTCGTAACAGCTCTCCAATTCCCATCGC
CCACTTCTTCATCCAATGGTGACCAGATATGTCGGCGATCGGCTGTATCAAAGTTGATCACA
TCCTCCACAGCGGCCGTGGCATCTGGCTCGGCTTCCTCTCTTTTGATGCCAACAGGTGTGGCA
AATGCAGTTGATTCTGTTACTCAAGTAACTTTGAATAGCCAAAACAAAAAGCCTTTTCCATT
GGCTTCTTTCGGTTTACAAGTTTATAGCGATGATATAGCCTATAAGCTGACTCTTACTGCGCT
GGAAGTGGGATATCGAAACTTCTTTGCAAGTGTGTTGGCAAACAATCAGAAGGGATTTGCG
AAGGCGGTGAAAGCGTCTGGAATCCCTCGCGATGAGCTGTACATATGTGGAACTGTCCTATC
CAATCGTGTTAATGGAGAAAAAGCTGCATACCAAAAAAACGAAGCAGGGCTGTCTCGAAAA
CATGGCAGCAATGGCCGCTGGCAATATTGACAAGCTCGACATGATCATGCTTGATTATCCTG
GGCCAAATGATGAGAGCGTCCGCGGTCAATGGCAGGCATTCGAGGAAATGCATTCTAAAGA
CGGAACTGTAGATGATTTGGCAGTGAGCAACTTCAGCCCAAATCAACTCGACGCAATATTGG
TAAACAAGGATGCGATTAAGCCAACCGTCAATCAGCTCCCGTTCTCCGTCGCATATCACCCC
GAGGGAATTTTGGACTACAATGCCAAACGTGATATATTGGTGCAATCGTGGTCTCCCCTATC
GAGGGTTCTCGGGCGCTATGGAAATACCCTGGATGCAATTGGGAAAAATTATGGAAAGAGT
GCAGCGCAGGTGGGACTTCGGTGGATAGTGCAGAGTGGAGCTGCGTTTTCCACGCAATCGA
AGAGTAAAAGCCATTTTGAAGAGGACTTGAACGTATTTGATTTCGCTCTTTCTGATGACGAA
ATGGCGAAGATTGCTCAATTGGCAGCTGCTTGAGTTTTGTAAGAAAATGCAGGAGCAAGTTC
AGAGCAGGGCGACTGATCATTCTTGATCCGTAGGTACACTAAATCCGTACCGAAATTTAAGA
TAATGTTTGTACTTTCAAGTTGAGTAAGAACCCCCTGATCAAGCTGACTGAGCGCTGCTATG
TCTGTGCACACGTCAACCCTAAAGCGCCCATGAAAACCAATAGATTCGCAAATAAACGCAA
AACTGAGA ATCAAAAAAATAATCCGAAGGTACTGAGATCGGA

Prostaglandin G/H synthase 2_PGHS-2

>TR754|c0_g1_i1
TCCGATCTGTGATTTCAAAGTTATCTTCAACCATTCCAACATGAAGGCTGTGCTTATTAACGA
CAAGAGACTCCCCGAAGATTCAGTGGAGTACACCTACCCCCATTCCAAGTTGACCCATTTCA
AGGAAAATATCCTCCAAAAGACCTTTTCGGCTTTGCTCAGGGCACCCATTGTCGGCAAGCCA
CTTAACCGCAAATTGACCAACAAAGCCGCAAGCAGCGGAACCAACCGCCCCTATGTACTCTC

CTGCAAGACTGACTACGCATCGTATGATTCTCTCACCGATTACTCACACTATGGAAGGCACT
TGAAGCCCGCCTCGAAGGATTGGTTGGAAAGCCTTCCCCCCATCGATGACGTTCTCAACATG
TTCGAGCGCAGAAAAAATAAGGAGGGTAAGGACGTCCAGGTCATGTGCCCCAAGAGTACCA
TGCTTTTCCCCACCTTCGCGCAGCACTTGATTGATTCCTTCATTGTTACCGCCATCAAGGAAA
ATACATCTGAAGGTGTAGAGTTCGACTGGAGGAAGACGGGATCCCCCCACGACATCGGACT
TCTTCCCCTATACGGAAAAACCATCGAACAAACTACCCAACTCCGTTTAAAGTCCGAGTTGA
AGGGCTCAAAGGGAAAATTGAAGACTCAAATCATCAAGGGAGGAGAATGGTCTCCATACCT
CTTCGACGCCAAGGGAAACAAGAAGCCTGAATTCAGCTCCCTTCCCGACCCTGAAGGATTGG
AATTTGCCTTGGGAATGGCTAAACAGCATGGCGTAGACGTCAAGAAGAACAAGGCAACTGT
CTTCGCTCTTGGTGGAAGCAGAGCCAATCTTACTCCCAACATCGTGGCATGGAATACTTTGTT
GCTTCGTGAGCACAACCGTATCGCATCGAAGATCGAGGAAGAAAACCCCAAGTGGGATGAT
GAGCGTGTGTTCCAGACCGCGCGTAATGTGAATCTTGCTGTGTACCTCAGACTTGTGATTGA
GGAGTACATCAACCACATTACTGGATTTGGTGTTGACTTTACGGTGGAGCCAGGGAAGTGGA
TGTGGGATGCCCCATGGTACAAGAGGAACTGGATCAGTGCCGAGTTCGCTGTCCTCTACCGC
TGGCATGCTGTTATCCCCAACTGCATGAAGTGGGGTGAGAAGACATTGCCCACTGCCTCATA
CCTTTACTCAAACCACCTTCTTCTTGATGAGGATGGTATGAAGGGAGATCTTCGCGAGGCAT
TCATCAATATTAGTGACCACCGTGCGACTAGTATGCAGATCCATAACACAGAAAGTTGGATG
AAAGGTCGTGATGCCAGGGCACTTCAGATGAGTCGAGCATGTGAATTGAGATCCTTCACTGA
GTACTGTGCCTATTTAGGAGCAAAGGTTCCCAAGACATTCGCCGATATCACCTCTGATGTGG
AGGTGCAGAACGAGTTGAAGGCCCTGTACGGAAAGGTGGAGAATATTGAGTTCTGGGTCGG
TTTGATTGCGAAGGACCATCCCACTGAGTCTATCATGTCTGCGGAGTTGACATCTTTTGTCGC
GAATGATGCCTTCAATCAGGCATTGACCCATCCTCTTCTTTCAGAGCATGTATGGGCAGCTG
GACCCGAGACTTTCTCGAAGGTTGGATGGGAGATGGTAATTAAGGTGCCATCAATCAAGGA
CATGCTTCAGAGGAACACTGGAGGAGCCCCAATTGAGGGTTTCGTTGGTATGACAAATCCTC
ACTACAAGATTCCTCTTTCTGAGATTACCACAAAGATTCTCCTTACCGCTTTCGTTGTCCTGG
CCGGTGTTGCTTACAAGGCTTACACTGAAATGTAAAAGCCTTGTTACACCAAATGCATCCAT
GTCTCTTTCACAGTGTACTTTTGAACTTCTCATGATGCATGATTTCAGGGAGAGAGACAACG
AATGGAACAATAGATGGGCACGTTTTTGTTTGAATGTTGAAAAACCTCTAGGTAGTGAAGTA
GGTATCGTAAGCAAGAGTATTAAGATACTATTCGATCTAGGAAAAAAAAAAGATCGGAAGA
G

Secologanin synthase_SLS

>TR26037|c1_g1_i1
ATTTCTCTCGTGTTCTGATCCGTGTCAATAAGGCTTTGTTCTCTCAATTAAGAGTAATCCAGT
AAGCTATGGCGGTCATAATTCCCAGATCTGCCGATTACAACTTCACCCATAGTTAGAAGAAA
ATAACAACGTCATGAACGTTCTGAACCGATCCGATGCACTTTGCAGTGCTCCACAAGAAGTA
GTTGCTTCAATCAATATTCCGCGACTGAGTGACACACCCATCCTGGTCAACATTCAATCACC
GTCCGCGTCAAAGTCTAATTCGAATGGCTCAATCGCACTTCAGGCTCCAACACTTCATTATTC
CATGCTCTCTATTCCGCTACTACTTCTATTTACCATAGCACTAATCGCCGCGACGCTTCGAAA
ACTCCTACGGCATTACATGCAGCAGCAACAGCCCCATGTCTTCCCAGCCTCGACAGGATTAC
CCCCCTGCGACGTGCCGTCCCATCCAATTTACGGACATCTGTGGCATGTCCTCAGCTCCCCCG
ACAGTGCCAAGTACAATTCCATATTTGGCGATCATGCGAACTCTTCTGGCTTGTCGACATTAC
GGTTCCTGAATACTCCTTGTATCGGCGTTTTGAAGGCCGAGCATGCGAGGTTCGTTCTTAGG
AATAGCATAGAACGGACTGGGACGAGCATGTTCTCTCGACATTTTAAGAGGGCTCTGGGCA
AAAATTCGATAGTGATGATTGACTGCGGAGAAACCAGCAAAGAGTTATGGAGGACTCACAG
GAACTTAATCAAGGTGGGATTCACCAAAAATGCTGTGAATAATATGGCCAACAAGGTTTGG
CAAGTAGCCAACGGGTTCACTTCTTCCTTATTGCGTGAATGCGCCAAGAATGATCAAGGTAC
AAAACAGGGGTCGTATTGTGCTGAGGCAGCAGATATCTTCAAATGGGTGACCCTCGACATCT
TTGGCAAGGTTGCCTTCAACTACAATTTTGGATGCACCGATTCGTTGTCAACGACTCCTCTTG
CACACTCACTCAACTATACACTCGAGGATAGCAATACTCGATGCAAGGCGGCCAACTTACTC
AATCCTGCGTATCAACTCTATTGGCTTCCTACCAAAAGAAATAGAGACTACAAGCACCACAG
CCAGAATGTTCGAGGCCTGATGAGAGAGCTCTGCCAGCAACGGGCGCAAGAGATAGACAAC
CAAAAGAATGTCCCGGAAGGCCATACGAATCTACCGAAGAAAGATGATCTCTTGACCATTCT
TCTCAAATCAAAAATAGAATCCCAAGGCGCCCAGAGGGATGGCACTCACGATGAATTGGTC
AAAATGCTCCTGACAGTCTTCTTTGCAGGATTTGATACATCATCCGTCCTACTCTCCATGGCC
ATGTGGTCCATAGCCACAAATCCCAGTATTCAACAAGAATGCGCCATGGAGGCGCAAGACG
CAAGCAGTACTTCCTCCGAGAACGAACCAAGTCTTCATGAGGATGCCTCACATTGGGAGTCG
AGATTAGCCTACTGCCGCGCAGTAATCTTGGAGACGTTGCGGTTGCACCCTCCGGTGCATAT

GAACAGTCGCATTCTCAACAAGGATGTAAAGTTGGGCGGATGCACCATACCAAAGAGGACG
CGCGTGTACCTCCCAATAACGCAGATCCATACAGATGAGAGGAACTTTGCAAGGCCAACAG
AGTTTTTGCCCGAACGTTGGGTTCGTAGAAATGCGACCAACAACAGATGGGTTGCCCGAGAC
TACAAGACAGAACCATCATTTGTAGAGAATGACCTGTCGTATATCCCTCCAGCCAACCCCCA
AAATATATTTACCTTCAGTGATGGTGCCCGTAACTGTGTAGGACACCGCCTGGCCCTTCAAG
AGAGCACCATGATATTTGCATGTGTTGTACGGGATTTGACTGTGAATGTACCGAAGGGCTTG
GTCATGCAAAAACGAAAAAAGTTTGCGTTGGCGCCTCCAATGGAGATGCCACTGATTTTAAA
GAGAAAAGGAAGGCAAAGCTAAGAGACTGAGAGAGGAAATCACATTTTTTTGGTTGGCCTA
ACTGCGAGTTGATCAGGCATGATCTATGATCTGATGTGCTCTCCAGTTGCTGCTGCCACGAC
CCTTCCCCAGTTTTTCCACAGGGCGTCCCAATTTTGCTTTGTGTATATGTAAATAGATATTTTC
TCCAGTTGTTGTAGCCAAGGCTCTCTCTATTCCTTGCTATCCCTCATAGGGAGGTGTAAGTTG
TCTTGATGTATAGGAGTAGTTACTTAGATAAGTGTAACAATTAACTTGGCCTGAATAATTAA
AACACACACACTACAAAAAAA

Phthiocerol synthesis polyketide synthase type I _ ppsC

>TR40329|c0_g1_i6
ACAGCAAAACCAGGATCAATCTATGCCAATGTACTTGGTACAAGTGTCATGTCTGATGGAAA
GTCAGCTTCAATCACAGCACCCAATGGCTCTGCCCAGGAAGAGTTGATCAAGAGGGCTCTGG
ATGTGTCAGGCATCAAGCCCAATGATGTAGATTACATTGAATCTCATGGGACTGGAACTCCA
CTTGGTGACCCCATTGAGGTTGAAGCTTTGGCAGAAGTTTTCGCAAATTCAAAGACTGAATC
ACATCCATTGATGATGGGCTCAGTGAAGAGCAACATTGGGCACCTAGAAGGTGCAGCTGGT
ATTGCAGGTCTGATCAAGGCTGTTTTGGTACTAGCTCATGAGCGTGTTCCATCCAATGTTGGC
CTGAAGACATTGAACCCTCTCATTGAGAAAACAATTGAATCTCATTCATTCCACATCAAGTT
TCCTACAGAGTCTCAGCCAATTATGAAAGAAGAAGATGGAGGGCTTTTGGTGGCAGGAGTG
AGCTCATTTGGCTATTCAGGCACAATTGCCCATGCAATCATTCAACAAGCTCCAGAGGATAT
GCGTCGCTGTACTGCTAATGCGAACCATTTTGGTTGTGAAGTTGATCATAATAGTTTAGTGTT
TTTGTTTACAGGTCAAGGCTCTCAATATGCAGGGATGGGGAAAGAGTTGTACCACAAAAATG
AATCATTTCGTGAAGCTATGGATAAATGTGAAAGCATATACAAGTCCCTGGTGGATGGGGA
ATCTCTTCTGGATATAATTTTTAGTTCAGGAAATGAAGGACTTGTGACATCAAATGCTCAAC
CATCTCTGATTGCCCTTGAGTGGAGTCTGGCCAGAATGTGGCAAACAAAAGGTGTTACTCCC
AGCATTGTGCTTGGTCATAGTGTGGGTGAGATAGCTGCTGCCTGTGTTGCTGGTGCAATGAC
AATTGAAACAGCATTGAAGCTGGCTGTTGCTAGACAAAGACTAGTGAATCATCTTCCTCAAA
ATAATGGCGCAATGGTTGCAGTGAAATGCAGCAAGGAAGAGGCTAACTCTACAATATCTTA
CTGTTTAAGTGAGGAAGAGCAAGATCTTGTTGGTGTAGCTAGTGTGAATGGCCCCAACAGCA
TTGTCCTCTCAGGTGTACATGATGTATTGGAAAAGCTGCTTCTTGAGCTAGAAAGGAAAGGT
GTTTATTTACAGGTATCCCATGCCTTTCATTCGCCTTTGATGAGAGGAATGGAGGATGATTTC
AGGCGTGTCATTGGTTCACTTGACCTCCAACAGCCTCTATCTATTCCACTTGCAAGCACAGTG
ACCGGTCATGTTATGACTCCTGGCAAAGTAATTGACTTAGAACATTGGGTGCAACAGCTTGC
TTCTCCAGTTCTGTTTGCAGATGCCTTCACAGAGGTCATCAATGGGAACAAGTGTGACATTTT
AGTTGAGGTTGGACCAAAGCCAATTCTATCCAAGATGGCTCAAAGCTGGTGGAAGCCAAAA
AAGAAAATTTCTCAGCCGTTGTGGGTTGCTTCTCTTGAGAAAGGGAAATTGTCGAGTCTGGA
AGAATCTTTTTTAGCTGTCGATGAGGCACTTGGCACACCGACTCATGCGGCGTCACCAAAAC
TGTGTAGTGTCTTCCCCAATCGTACAAGGTTGCCTTGGCCTGAAAGCCCACCCCACCCACTTT
TACAGCTTACTATTCCTATTGGATCACATGGCACAGAG

Polyketide Cyclase dehydrase and lipid transport protein_ PYL5

>TR3754|c0_g1_i2
CGCAGTCAATGAATCATGAACGGCACGAACACATGGCGCATGCATGGAACAACAGACGTAA
TCAGACAAAGAAAATAACAAAGAACAATGGTGACAACTGATAATATGAGGCAGCAGGGCA
GGAATGACTCTGACCAACAACATGATGACACAGCAAGTCGAAATACGGGCGCAATAAGTTT
TATCATTGCACCCCAATAGCTGAAACCCCCTTCTTCATAATAAGCTCCCAACCTCATTAACAC
CAGTACAGCGGTCCCGTACTTACTTTTACATTAGAAAAACAATGAAATTCGTGTGTCTGATC
CTAGCTCTCTCAGTACTTAAAGCCGATGCTTTCACTACGCCATCGCACCCTTTGATTCCATCC
AAAAATGCAGCCCATAATAATATCGTCACTCAGAAAACACAGCAGTCATCCACGTCCCTAA
ACGTATGGTGGTTCGGCGGGACCCCTGGGGCAGGATCAGAAACGTCCACATCTGGCGAAGA
ATGCGAACTCGTCGCCGTCCGAATTGACCGCACATCTGCTAACTCGAGACGCATCGCTGGAG
AGATCGTTGTAAACGCCCCTCCGATCGATGTATGGGCCATTCTCACTGACTACGACAATTTA

ACGACTCATGTTCCAAATTTAGTGAAAAGTAAGATAGTTAGTGATGGACCCGGGGAAGCAG
GAGATGGATCACACAAGTGCAGATTGTATCAGAGAGGAGCTCAGAAAATTATAGGATTCGA
GTTTGGAGCAGATGTAACCATGGACATGACAGAGAAGATTATCGTGGGTGGAAAGTCGGGT
AAAATTACAACATCTCAAGATACAACCTTGGGCAATGGTGTGTCCTCGACTGGACTATTTCC
CGAGGAACGACGTATTGATTTCAAGTGCGTGGATTCACAGTTCTTCTCAGAGTTTGACGGTA
CATGGCGCGTGGTACCGATGCCTGACAATCCCTTCACGGGGGAGGCTGAAACGACTGTGAG
CTACACTGTGGAGGTAAGGCCGAAAGGGCCTGTTCCAGTAGCGGCACTAGAATGGAGAATA
CGAGAGGATGTGCCAACAAATTTAAGGGCAGTGAAAAAGGCTGCGATGGAGGTAGGCGCA
GAAGGGGTCTGGTCAAAGAAAAATGGAGGCAAGATTGGAGATGCTTTGGCGAAGTTGACGG
GCGGGACAGGGCACGACGAGAGCACGCCACAGGGGAACGAGTTTGCTGATTTCATTCTTGA
TTATGAAAATGAGACATTAGGGGCGTATTTGGACTGAAGCTCACATGCGGATAAGAATTTTG
AGTTACTTAACTTTTACTGACAAACCTGAAGGTATTGGAAAAATGTTGTGATAGAATAGTGT
AAGATATGTGAAATTTTACGAGCGCATCAGCAACGATGCGAGAGAGAGTTAGTTGTGGGGG
AGTTTGTCCTATGATATCTTTCCTGACCACAAATTGAACCACAACTGAGGCGCTTGACGAAA
ATAAGATTTAAGAGTTGTATCGCTATGATGAACTACAGATACATTCCGACGAAATAGGTCAC
ACTGTTTATTTTAGATGATACTTGACGCCACACAAAATGTATTTGTGACCCCTGAGCGTGATC
GAAATCAAGTAGCTCGTTAGATCCAGCTTGTCAATTTCTCTCCGTTTGTACCTCTGCATCTGT
TTTCCATGAGTATTCTGTTTTCCGTGAGTGAATAGTACCCAAACCGGCACAGAGTTTAAACA
CTGTATCCCCCTTATCTTTTTTTCGCTCCAGCTTTTACTTTGTTCATAAGTAAGTAAGATCCTT
CAGTCCACCAGCATTCATAACTTTTGTATATCCCTTCTGTTCAAGCATTTCCTTCGCTTTTGCA
GCACGGCGTCCAGAACGGCAGAATACGATCACAGGAGCTATGGTAGGATGCAAGATGTATT
GTTATGATGAAACATATGATACGAGGTTCACAAAATAAAATACATAGATCTGATCTGACCAC
CATTTTTAATCAACGTACCATTTTTGTCGGGCATGAGTTCTTCAGCTTTTGCCATCAACTCCG
AACAATCGTCTAAATGGCAGCTCGCGTGGCGGAACTTCCTTGGAGTGAGCTGTTCCTCTTCC
AATTCCTCTTTGGTTCTCACATCGAGAATTATGGCGCCAGATACATCGACGGCGGCTTTTACT
TCGTCAGGCGTGCTGAAGGATGTCATACTTTCGGGTTTGGGCCAGATCCAAAGTTGAGCAGG
AAAAGGAGATGTGCAGGCTGGGTTGACGTCTAGGGCGTCGAACGAGATAATTCGGTAATAG
GAGTAGACGACAGACTGGCCCTCGGTTCATGAAATAATGTGACATCACCGTATCGCAATGCC
ATTCTCGATTCTCGGCGGCAATTTCGATTCTATT TCGGGT

Polyketide Enoylreductase_PKS_ER

>TR30039|c0_g1_i1
TTTTTTTTTAAAAGTAAAGTAAATTTGAAACGCTTAAATCGGAGCAGTATCTACTTGTGGGGT
TCATTATTGTACACTTTTAGGATACCGAGACTCCTATCATTATCATCCATCTCCATTCTATTCG
TGTTGCCACTTTCATTTGTTGCAGAAGGTTGGTACTATCACAGTGTTTTGTCACATCTCCTAC
AGATTACTCTTGACGTTACTTTCTGTGAGGCTATGGTCCAATACAACCTTTCCAGAAGATTCT
CCCTTCAAAAACTTCTTCAGATAGTCGTCCAGTTGATGTCGTCCGATACTGACAGCAGAAGC
CTCTGTCTTGGGACAAGACCACTCGCCCGCCAGCTTCTCCCAAACAGCTACTTTCTGTAGCA
AGGGAATCTCGACCGAGTCCACTCCCAGCAAATTGACGCCTCGCAGAATAAAGGGCAAAAC
AGTAGTATCCAACCCCATGCCGGCCACCAAACCGCAGCAGGCCACCGAACCACCTGGCCGA
ATTTGTTTTAACAATTCTGCTAGGGGACCTCCACCCACGGTATCTACGGCATGCGCAAACAG
CGGCTTGACCAGCGGTTTCTTGTTCGCTTCCAAAGCTTCTCTGCCCACCACTTCCGTTGCGCC
CAATTCCAGAAGTCTGCCTTTTTCAGATACTTTCCCCGAAATGGCAACAACTTTGAATCCTAG
TTTGGACAGCAACTCCACGGCCACGCTTCCCACGGCGCCGCTGGCCCCCGTCACGGCCACCC
GGCCTCCATGTTGGGACAGATCG

Polyketide hydroxylase-like_ PK_Hxl

>TR21208|c0_g1_i1
TCGAGGAACGTCGCGTTGGTGGGGGATGCGGCGCATTCGTTCCCTCCCGCGGGAGGATTCGG
GATGAATACCGGTTTGCAGGACGCGCACAATATCGCGTGGAGGCTCGCTTTGTTGCTGCTCG
ACGGAGGGAGAGGGAATTCGTCCGAGGCGTCCTTCTTCTCGGAGGACGTGCTGTCCAAGTAC
GACGCGGAACGAAGACCCATCGCCACCCAAAACGCGGCGCTCAG

PART 2

Supplementary Information

Chapter 2

Variation in prostaglandin metabolism during *T. rotula* growth

Adapted from the published article:

Di Dato, V., Barbarinaldi, R., Amato, A., Di Costanzo, F., Fontanarosa, C., Perna, A., Amoresano, A., Esposito, F., Cutignano, A., Ianora, A., Romano, G. Variation in prostaglandin metabolism during growth of the diatom *Thalassiosira rotula*. *Scientific Reports* (2020) **10**, 5374.

Supplementary Information

Supplementary Figures.

Supplementary Figure 2.1: *Phyre2* best hits structures from TrotCOX protein blast.

Supplementary Figure 2.2: *Phyre2* alignment between *Thalassiosira rotula* (TrotCOX) and *Bos taurus* lactoperoxidase (BtLPO).

Supplementary Figure 2.3: *Phyre2* alignment between *Thalassiosira rotula* (TrotCOX) and *Ovis aries* cyclooxygenase (OaCOX).

Supplementary Figure 2.4: Alignment of the diatoms sequences used in the phylogenetic analysis.

External standard calibration method for PG quantization.

Supplementary Figure 2.5: Calibration curve of deuterated prostaglandin E_2 (PGE_2-d4) used as internal standard.

Supplementary Figure 2.6: Calibration curve of prostaglandin E_2 (PGE_2) used as standard.

Supplementary Figure 2.7: Calibration curve of prostaglandin B_2 (PGB_2) used as standard.

Supplementary Figure 2.8: Calibration curve of 15-Deoxy-Delta-12,14-prostaglandin J_2 ($15dPGJ_2$) used as standard.

Supplementary Figure 2.9: TIC chromatogram (A) of a representative sample and XIC chromatogram (B) of PGEM (327 m/z precursor ion; 291 m/z and 309 m/z daughter ion).

Supplementary Tables.

Supplementary Table 2.1: List of species and sequences ID used in the phylogenetic analysis.

Supplementary Table 2.2: List of PGs identified in the culture medium with their LC-MS/MS parameters.

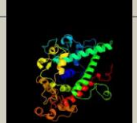
Aligned sequences, with gap, used in the phylogenetic analysis

Sequences used in the phylogenetic analysis, without gap

References

Supplementary Figures.

Supplementary Figure 2.1. *Phyre2*[1] best hits structures from TrotCOX protein blast.

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c2qjmA_	Alignment		100.0	19	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**lactoperoxidase; **PDBTitle:** crystal structure of buffalo lactoperoxidase at 2.75a resolution
2	c2oyuP_	Alignment		100.0	28	**PDB header:**oxidoreductase **Chain:** P: **PDB Molecule:**prostaglandin g/h synthase 1; **PDBTitle:** indomethacin-(s)-alpha-ethyl-ethanolamide bound to cyclooxygenase-1
3	c1pggB_	Alignment		100.0	28	**PDB header:**oxidoreductase **Chain:** B: **PDB Molecule:**prostaglandin h2 synthase-1; **PDBTitle:** prostaglandin h2 synthase-1 complexed with 1-(4-iodobenzoyl)-5-2 methoxy-2-methylindole-3-acetic acid (iodoindomethacin), trans model
4	c1ht8B_	Alignment		100.0	28	**PDB header:**oxidoreductase **Chain:** B: **PDB Molecule:**prostaglandin h2 synthase-1; **PDBTitle:** the 2.7 angstrom resolution model of ovine cox-1 complexed with2 alclofenac
5	d1q4ga1	Alignment		100.0	28	**Fold:**Heme-dependent peroxidases **Superfamily:**Heme-dependent peroxidases **Family:**Myeloperoxidase-like
6	c6ercA_	Alignment		100.0	26	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**peroxinectin a; **PDBTitle:** peroxidase a from dictyostelium discoideum (ddpoxa)
7	c5mfaA_	Alignment		100.0	20	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**myeloperoxidase; **PDBTitle:** crystal structure of human promyeloperoxidase (prompo)
8	c1ddxA_	Alignment		100.0	27	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**protein (prostaglandin h2 synthase-2); **PDBTitle:** crystal structure of a mixture of arachidonic acid and prostaglandin2 bound to the cyclooxygenase active site of cox-2: prostaglandin3 structure
9	c3pghD_	Alignment		100.0	27	**PDB header:**oxidoreductase **Chain:** D: **PDB Molecule:**cyclooxygenase-2; **PDBTitle:** cyclooxygenase-2 (prostaglandin synthase-2) complexed with a non-2 selective inhibitor, flurbiprofen
10	d1cvua1	Alignment		100.0	27	**Fold:**Heme-dependent peroxidases **Superfamily:**Heme-dependent peroxidases **Family:**Myeloperoxidase-like
11	c1d2vD_	Alignment		100.0	20	**PDB header:**oxidoreductase **Chain:** D: **PDB Molecule:**myeloperoxidase; **PDBTitle:** crystal structure of bromide-bound human myeloperoxidase isoform c at2 ph 5.5

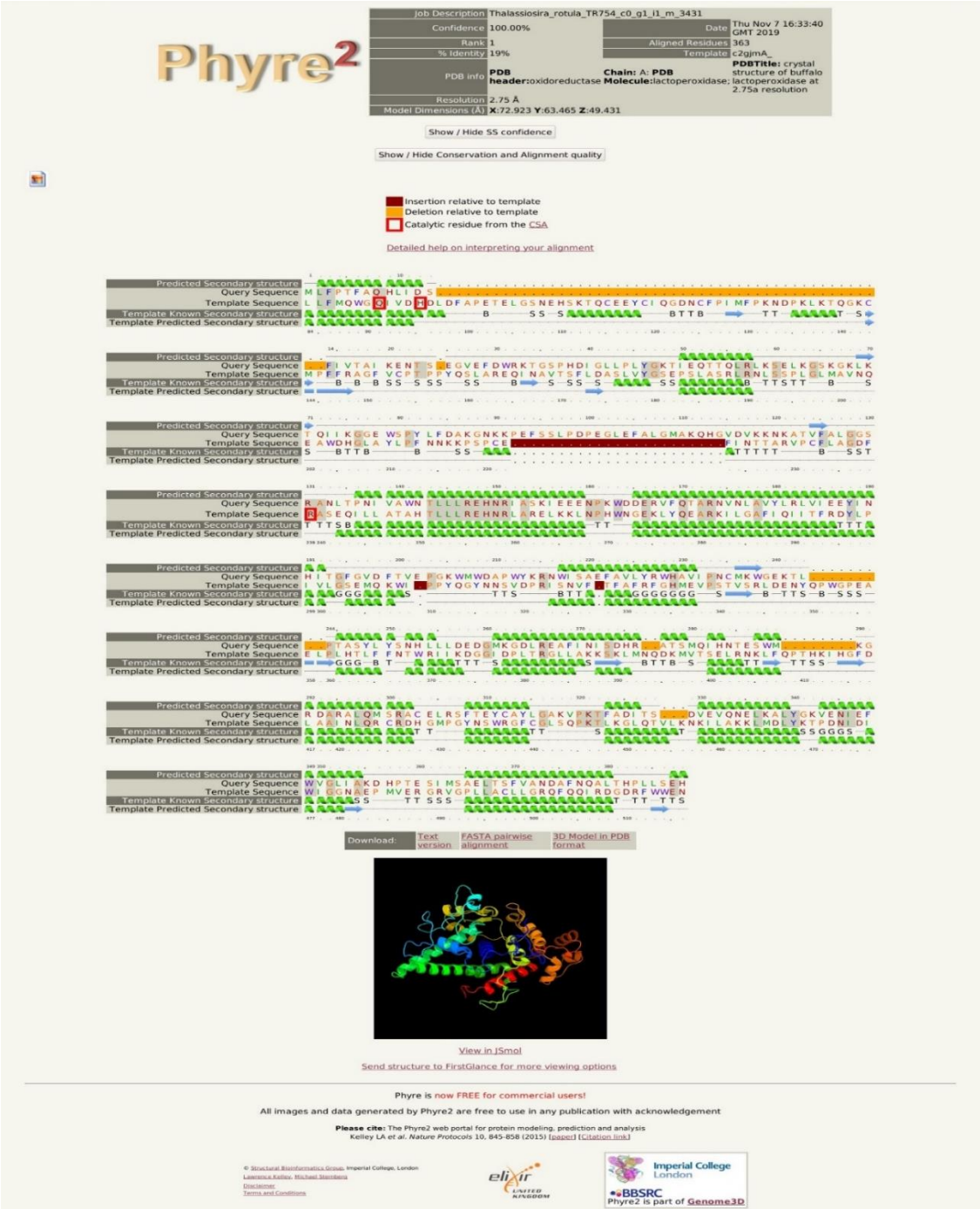
12	c4hhsA_	Alignment		100.0	19	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**alpha-dioxygenase; **PDBTitle:** crystal structure of fatty acid alpha-dioxygenase (arabidopsis2 thaliana)
13	c4kvjA_	Alignment		100.0	19	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**fatty acid alpha-oxidase; **PDBTitle:** crystal structure of oryza sativa fatty acid alpha-dioxygenase with2 hydrogen peroxide
14	c5gvrA_	Alignment		36.3	17	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**probable atp-dependent rna helicase ddx41; **PDBTitle:** crystal structure of the ddx41 dead domain in an apo closed form
15	c3brcA_	Alignment		35.0	18	**PDB header:**structural genomics, unknown function **Chain:** A: **PDB Molecule:**conserved protein of unknown function; **PDBTitle:** crystal structure of a conserved protein of unknown function from2 methanobacterium thermoautotrophicum
16	c1vkc5A_	Alignment		33.5	20	**PDB header:**structural genomics, unknown function **Chain:** A: **PDB Molecule:**expressed protein; **PDBTitle:** x-ray structure of gene product from arabidopsis thaliana at3g22680
17	d1vkc5a_	Alignment		33.5	20	**Fold:**Hypothetical protein At3g22680 **Superfamily:**Hypothetical protein At3g22680 **Family:**Hypothetical protein At3g22680
18	d1u7ka_	Alignment		28.8	33	**Fold:**Retrovirus capsid protein, N-terminal core domain **Superfamily:**Retrovirus capsid protein, N-terminal core domain **Family:**Retrovirus capsid protein, N-terminal core domain
19	c3fe2B_	Alignment		22.4	13	**PDB header:**hydrolase **Chain:** B: **PDB Molecule:**probable atp-dependent rna helicase ddx5; **PDBTitle:** human dead-box rna helicase ddx5 (p68), conserved domain i in complex2 with adp
20	c4px9C_	Alignment		21.3	12	**PDB header:**translation, rna binding protein **Chain:** C: **PDB Molecule:**atp-dependent rna helicase ddx3x; **PDBTitle:** dead-box rna helicase ddx3x domain 1 with n-terminal atp-binding loop
21	c3evyB_	Alignment	not modelled	17.7	9	**PDB header:**hydrolase **Chain:** B: **PDB Molecule:**putative type i restriction enzyme r protein; **PDBTitle:** crystal structure of a fragment of a putative type i restriction2 enzyme r protein from bacteroides fragilis
22	d1ou9a_	Alignment	not modelled	17.3	21	**Fold:**SspB-like **Superfamily:**SspB-like **Family:**Stringent starvation protein B, SspB
23	c2nqxP_	Alignment	not modelled	17.3	14	**PDB header:**oxidoreductase **Chain:** P: **PDB Molecule:**tryptophan 2,3-dioxygenase; **PDBTitle:** crystal structure of tryptophan 2,3-dioxygenase from ralstonia2 metallidurans
24	c6hwwA_	Alignment	not modelled	16.6	33	**PDB header:**viral protein **Chain:** A: **PDB Molecule:**putative gag polyprotein; **PDBTitle:** immature miv capsid hexamer structure in intact virus particles
25	d1zszc1	Alignment	not modelled	16.5	16	**Fold:**SspB-like **Superfamily:**SspB-like **Family:**Stringent starvation protein B, SspB
26	c6hwyB_	Alignment	not modelled	15.9	33	**PDB header:**viral protein **Chain:** B: **PDB Molecule:**putative gag polyprotein; **PDBTitle:** mature miv capsid pentamer structure in intact virus particles
27	c3fhcB_	Alignment	not modelled	15.6	9	**PDB header:**transport protein/hydrolase **Chain:** B: **PDB Molecule:**atp-dependent rna helicase ddx19b; **PDBTitle:** crystal structure of human dbp5 in complex with nup214
28	d1qr7b2	Alignment	not modelled	15.0	32	**Fold:**Sigma2 domain of RNA polymerase sigma factors **Superfamily:**Sigma2 domain of RNA polymerase sigma factors **Family:**Sigma2 domain of RNA polymerase sigma factors

29	c4cmxB_	Alignment	not modelled	14.8	14	**PDB header:**nuclear protein **Chain:** B: **PDB Molecule:**rv3378c; **PDBTitle:** crystal structure of rv3378c
30	d1g1xc_	Alignment	not modelled	14.3	25	**Fold:**DNA/RNA-binding 3-helical bundle **Superfamily:**Ribosomal protein S18 **Family:**Ribosomal protein S18
31	c4qkcB_	Alignment	not modelled	14.2	18	**PDB header:**signaling protein/dna binding protein **Chain:** B: **PDB Molecule:**anti-sigma factor nepr; **PDBTitle:** co-crystal structure of anti-anti-sigma factor phyr complexed with2 anti-sigma factor nepr from bartonella quintana
32	d1mn3a_	Alignment	not modelled	13.8	14	**Fold:**RuvA C-terminal domain-like **Superfamily:**UBA-like **Family:**CUE domain
33	d1ou8a_	Alignment	not modelled	12.8	21	**Fold:**SspB-like **Superfamily:**SspB-like **Family:**Stringent starvation protein B, SspB
34	c2c2xB_	Alignment	not modelled	12.6	14	**PDB header:**oxidoreductase **Chain:** B: **PDB Molecule:**methylenetetrahydrofolate dehydrogenase- **PDBTitle:** three dimensional structure of bifunctional methylenetetrahydrofolate2 dehydrogenase-cyclohydrolase from mycobacterium tuberculosis
35	c1fcuA_	Alignment	not modelled	12.6	36	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**hyaluronoglucosaminidase; **PDBTitle:** crystal structure (trigonal) of bee venom hyaluronidase
36	d1fcqa_	Alignment	not modelled	12.5	36	**Fold:**TIM beta/alpha-barrel **Superfamily:**(Trans)glycosidases **Family:**Bee venom hyaluronidase
37	c3fmpD_	Alignment	not modelled	12.2	9	**PDB header:**oncoprotein/hydrolase **Chain:** D: **PDB Molecule:**atp-dependent rna helicase ddx19b; **PDBTitle:** crystal structure of the nucleoporin nup214 in complex with the dead-2 box helicase ddx19
38	d1yfna1	Alignment	not modelled	11.9	21	**Fold:**SspB-like **Superfamily:**SspB-like **Family:**Stringent starvation protein B, SspB
39	c2b19A_	Alignment	not modelled	11.5	42	**PDB header:**neuropeptide **Chain:** A: **PDB Molecule:**neuropeptide k; **PDBTitle:** solution structure of mammalian tachykinin peptide,2 neuropeptide k
40	c3rfaB_	Alignment	not modelled	11.0	13	**PDB header:**oxidoreductase **Chain:** B: **PDB Molecule:**ribosomal rna large subunit methyltransferase n; **PDBTitle:** x-ray structure of rlmn from escherichia coli in complex with s-2 adenosylmethionine
41	c2kneB_	Alignment	not modelled	10.8	32	**PDB header:**metal transport **Chain:** B: **PDB Molecule:**atpase, ca++ transporting, plasma membrane 4; **PDBTitle:** calmodulin wraps around its binding domain in the plasma2 membrane ca2+ pump anchored by a novel 18-1 motif
42	d1upka_	Alignment	not modelled	10.6	12	**Fold:**alpha-alpha superhelix **Superfamily:**ARM repeat **Family:**Mo25 protein
43	c3rfaA_	Alignment	not modelled	10.2	13	**PDB header:**oxidoreductase **Chain:** A: **PDB Molecule:**ribosomal rna large subunit methyltransferase n; **PDBTitle:** x-ray structure of rlmn from escherichia coli in complex with s-2 adenosylmethionine
44	d2bmfa1	Alignment	not modelled	10.1	21	**Fold:**P-loop containing nucleoside triphosphate hydrolases **Superfamily:**P-loop containing nucleoside triphosphate hydrolases **Family:**RNA helicase
45	d2qajr1	Alignment	not modelled	10.1	29	**Fold:**DNA/RNA-binding 3-helical bundle **Superfamily:**Ribosomal protein S18 **Family:**Ribosomal protein S18
46	c3df1R_	Alignment	not modelled	10.1	29	**PDB header:**ribosome **Chain:** R: **PDB Molecule:**30s ribosomal protein s18; **PDBTitle:** crystal structure of the bacterial ribosome from escherichia coli in2 complex with hygromycin b. this file contains the 30s subunit of the3 first 70s ribosome, with hygromycin b bound. the entire crystal4 structure contains two 70s ribosomes.
47	d1ho8a_	Alignment	not modelled	9.9	21	**Fold:**alpha-alpha superhelix **Superfamily:**ARM repeat **Family:**Regulatory subunit H of the V-type ATPase
48	c3cuzA_	Alignment	not modelled	9.9	9	**PDB header:**transferase **Chain:** A: **PDB Molecule:**malate synthase a; **PDBTitle:** atomic resolution structures of escherichia coli and2 bacillis anthracis malate synthase a: comparison with3 isoform g and implications for structure based drug design
49	c2g2bA_	Alignment	not modelled	9.9	7	**PDB header:**immune system **Chain:** A: **PDB Molecule:**allograft inflammatory factor 1; **PDBTitle:** nmr structure of the human allograft inflammatory factor 1
50	c5gjuA_	Alignment	not modelled	9.5	26	**PDB header:**hydrolase **Chain:** A: **PDB Molecule:**atp-dependent rna helicase dead; **PDBTitle:** dead-box rna helicase
51	c5cofA_	Alignment	not modelled	9.4	11	**PDB header:**unknown function **Chain:** A: **PDB Molecule:**uncharacterized protein; **PDBTitle:** crystal structure of uncharacterised protein q1r1x2 from escherichia2 coli uti89
52	c5o5jR_	Alignment	not modelled	9.3	31	**PDB header:**ribosome **Chain:** R: **PDB Molecule:**30s ribosomal protein s18 2; **PDBTitle:** structure of the 30s small ribosomal subunit from mycobacterium2 smegmatis
53	c3a0bT_	Alignment	not modelled	8.9	20	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure of br-substituted photosystem ii complex

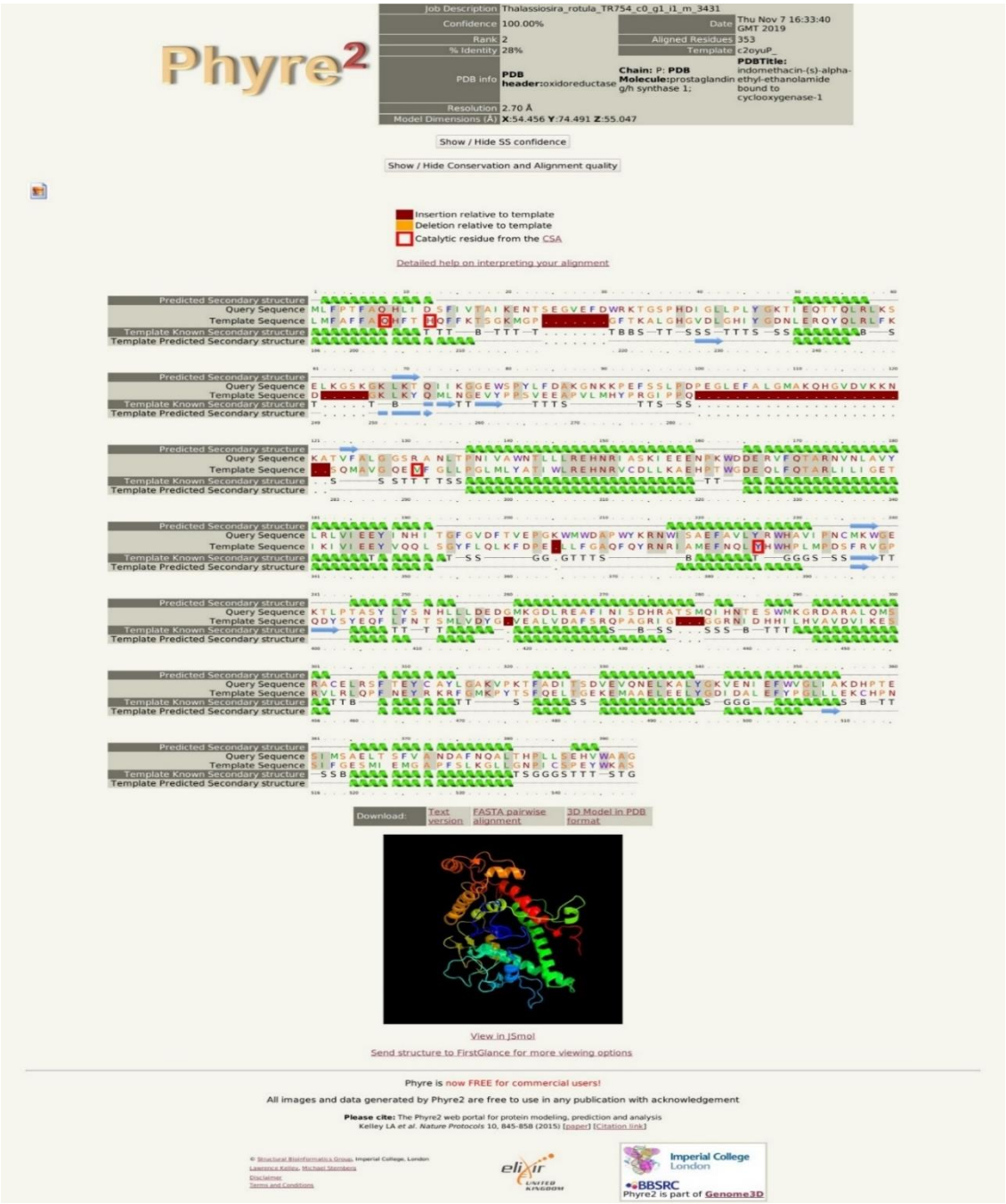
54	c5mx2T_	Alignment	not modelled	8.2	18	**PDB header:**oxidoreductase **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** photosystem ii depleted of the mn4cao5 cluster at 2.55 a resolution
55	c2hv1J_	Alignment	not modelled	8.2	17	**PDB header:**lyase **Chain:** J: **PDB Molecule:**cyanate hydratase; **PDBTitle:** site directed mutagenesis of key residues involved in the catalytic2 mechanism of cyanase
56	d1rfya_	Alignment	not modelled	8.2	24	**Fold:**Long alpha-hairpin **Superfamily:**Transcriptional repressor TraM **Family:**Transcriptional repressor TraM
57	c5aj3R_	Alignment	not modelled	8.0	6	**PDB header:**ribosome **Chain:** R: **PDB Molecule:**mitoribosomal protein bs18m, mrps18c; **PDBTitle:** structure of the small subunit of the mammalian mitoribosome
58	c2c0zA_	Alignment	not modelled	7.9	22	**PDB header:**isomerase **Chain:** A: **PDB Molecule:**novw; **PDBTitle:** the 1.6 a resolution crystal structure of novw: a 4-keto-6-2 deoxy sugar epimerase from the novobiocin biosynthetic3 gene cluster of streptomyces spheroides
59	c3j0xU_	Alignment	not modelled	7.8	31	**PDB header:**ribosome **Chain:** U: **PDB Molecule:**30s ribosomal protein s18; **PDBTitle:** structural characterization of mrna-trna translocation intermediates2 (30s ribosome of class 4b of the six classes)
60	d1wrka1	Alignment	not modelled	7.8	7	**Fold:**EF Hand-like **Superfamily:**EF-hand **Family:**Calmodulin-like
61	c1upiA_	Alignment	not modelled	7.8	16	**PDB header:**epimerase **Chain:** A: **PDB Molecule:**dttdp-4-dehydrorhamnose 3,5-epimerase; **PDBTitle:** mycobacterium tuberculosis rmlc epimerase (rv3465)
62	c5b66T_	Alignment	not modelled	7.7	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure analysis of photosystem ii complex
63	c3bbnR_	Alignment	not modelled	7.7	27	**PDB header:**ribosome **Chain:** R: **PDB Molecule:**ribosomal protein s18; **PDBTitle:** homology model for the spinach chloroplast 30s subunit fitted to 9.4a2 cryo-em map of the 70s chlororibosome.
64	c5gtht_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (dark dataset)
65	c5ws5t_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (preflash dark dataset)
66	c5ws6t_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (preflash two-flash dataset
67	c5ws5T_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (preflash dark dataset)
68	c5gthT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (dark dataset)
69	c3wu2T_	Alignment	not modelled	7.6	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure analysis of photosystem ii complex
70	c5h2fT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure of the psbm-deletion mutant of photosystem ii
71	c5kait_	Alignment	not modelled	7.6	18	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** nh3-bound rt xfel structure of photosystem ii 500 ms after the 2nd2 illumination (2f) at 2.8 a resolution
72	c5h2ft_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure of the psbm-deletion mutant of photosystem ii
73	c4ubBt_	Alignment	not modelled	7.6	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native structure of photosystem ii (dataset-2) by a femtosecond x-ray2 laser
74	c3wu2t_	Alignment	not modelled	7.6	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure analysis of photosystem ii complex
75	c5gtlT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (two flash dataset)
76	c5v2cT_	Alignment	not modelled	7.6	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** re-refinement of crystal structure of photosystem ii complex
77	c5gtlt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native xfel structure of photosystem ii (two flash dataset)
78	c5v2ct_	Alignment	not modelled	7.6	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** re-refinement of crystal structure of photosystem ii complex
79	c6jlpt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (3f state, dataset2)
						PDB header:photosynthesis

80	c6jlpT_	Alignment	not modelled	7.6	18	**Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (3f state, dataset2)
81	c6jlmT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (dark state, dataset2)
82	c6jlkT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (1f state, dataset1)
83	c6jloT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (2f state, dataset2)
84	c6jlT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (dark state, dataset1)
85	c6jlT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (2f state, dataset1)
86	c6jlot_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (2f state, dataset2)
87	c6jlnt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (1f state, dataset2)
88	c6jlnT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (1f state, dataset2)
89	c6jlt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (2f state, dataset1)
90	c6jlkt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (1f state, dataset1)
91	c6jlmt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (dark state, dataset2)
92	c6jlt_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** xfel structure of cyanobacterial photosystem ii (dark state, dataset1)
93	c5tisT_	Alignment	not modelled	7.6	18	**PDB header:**photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** room temperature xfel structure of the native, doubly-illuminated2 photosystem ii complex
94	c5zznt_	Alignment	not modelled	7.6	18	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure of photosystem ii from an sqdg-deficient mutant of2 thermosynechococcus elongatus
95	c5kaiT_	Alignment	not modelled	7.6	18	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** nh3-bound rt xfel structure of photosystem ii 500 ms after the 2nd2 illumination (2f) at 2.8 a resolution
96	c5kafT_	Alignment	not modelled	7.6	18	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** rt xfel structure of photosystem ii in the dark state at 3.0 a2 resolution
97	c4ub6T_	Alignment	not modelled	7.6	18	**PDB header:**electron transport, photosynthesis **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** native structure of photosystem ii (dataset-1) by a femtosecond x-ray2 laser
98	c5zznT_	Alignment	not modelled	7.6	18	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** crystal structure of photosystem ii from an sqdg-deficient mutant of2 thermosynechococcus elongatus
99	c4il6t_	Alignment	not modelled	7.6	18	**PDB header:**electron transport **Chain:** T: **PDB Molecule:**photosystem ii reaction center protein t; **PDBTitle:** structure of sr-substituted photosystem ii

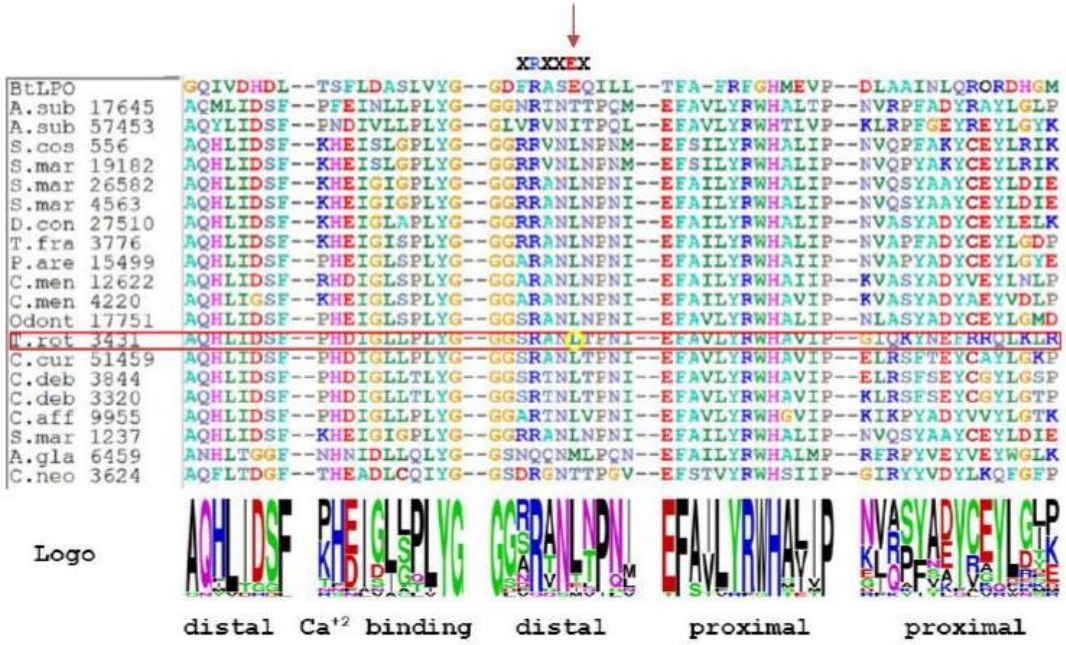
Supplementary Figure 2.2. *Phyre2* alignment between *Thalassiosira rotula* (TrotCOX) and *Bos taurus* lactoperoxidase (BtLPO).



Supplementary Figure 2.3. *Phyre2* alignment between *Thalassiosira rotula* (TrotCOX) and *Ovis aries* cyclooxygenase (OaCOX).



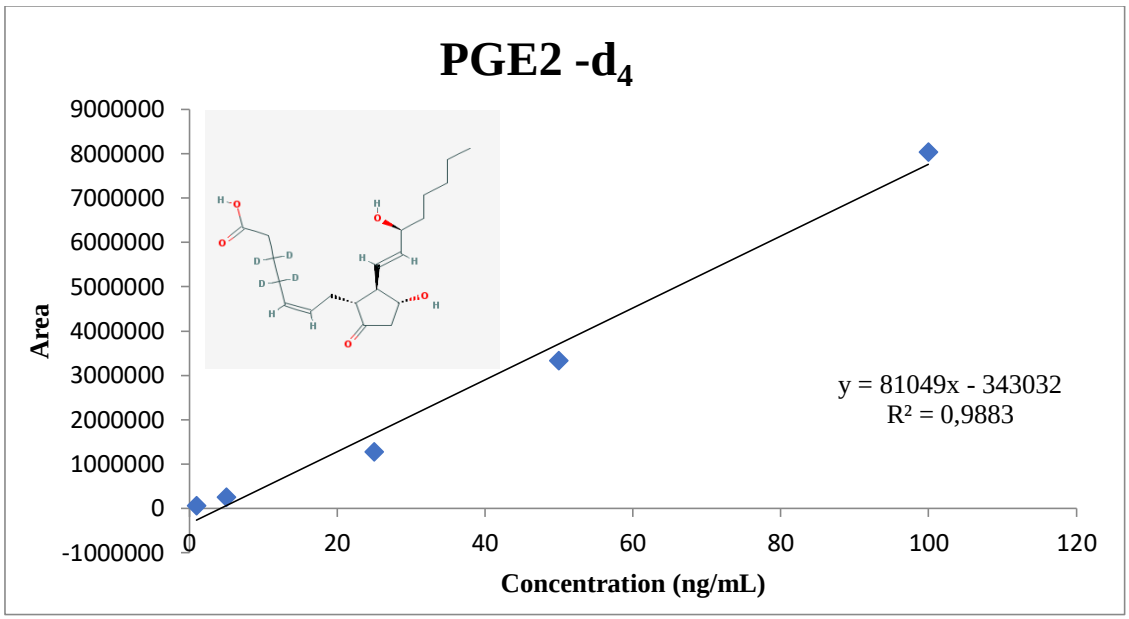
Supplementary Figure 2.4. Alignment of the diatoms sequences used in the phylogenetic analysis



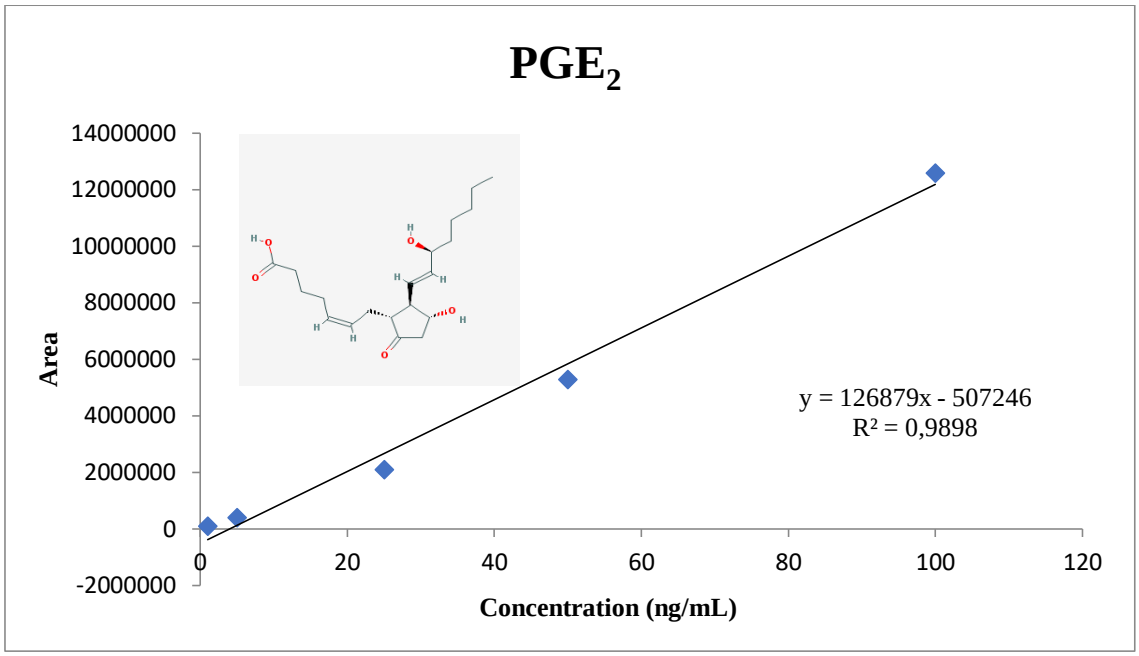
External standard calibration method for PG quantization.

For an external standard quantitation, known data from a calibration standard and unknown data from the sample are combined to generate a quantitative report. Working standard solutions containing all the analytes at 1 µg/mL were prepared, from stock solutions of 100 µg/mL for each standard: PGE_2, 15-d-PGJ_2, PGB_2, PGD_2, PGJ_2 (Sigma-Aldrich, St. Louis, Missouri, USA), 20-oh-PGE_2, 6-keto-$PGF_1\alpha$, DHK (CAYMAN CHEMICAL, Michigan, USA). From these standard solutions, five solutions at different concentration levels were prepared: 1, 5, 25, 50, 100 ng/mL for the construction of the calibration curves. The coefficients of determination (R2) were used to evaluate the linearity of the calibration ranges. All R2 were greater than 0.99, demonstrating that the method was linear in the range of calibration for all analytes. The amount of prostaglandins was normalized by the sampled volume extracted for each daily replicate and then by the number of cells per mL for each replicate. The final value reported in the graph were pg/cell.

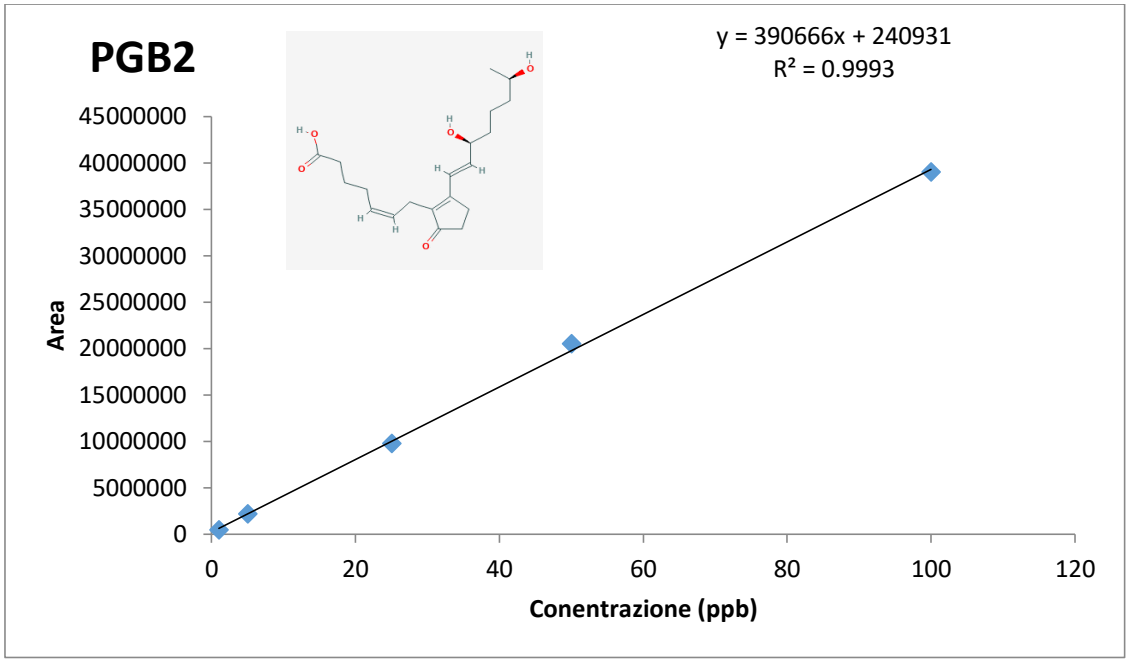
Supplementary Figure 2.5: calibration curve of deuterated prostaglandin E_2 (PGE_2-d4) used as internal standard.



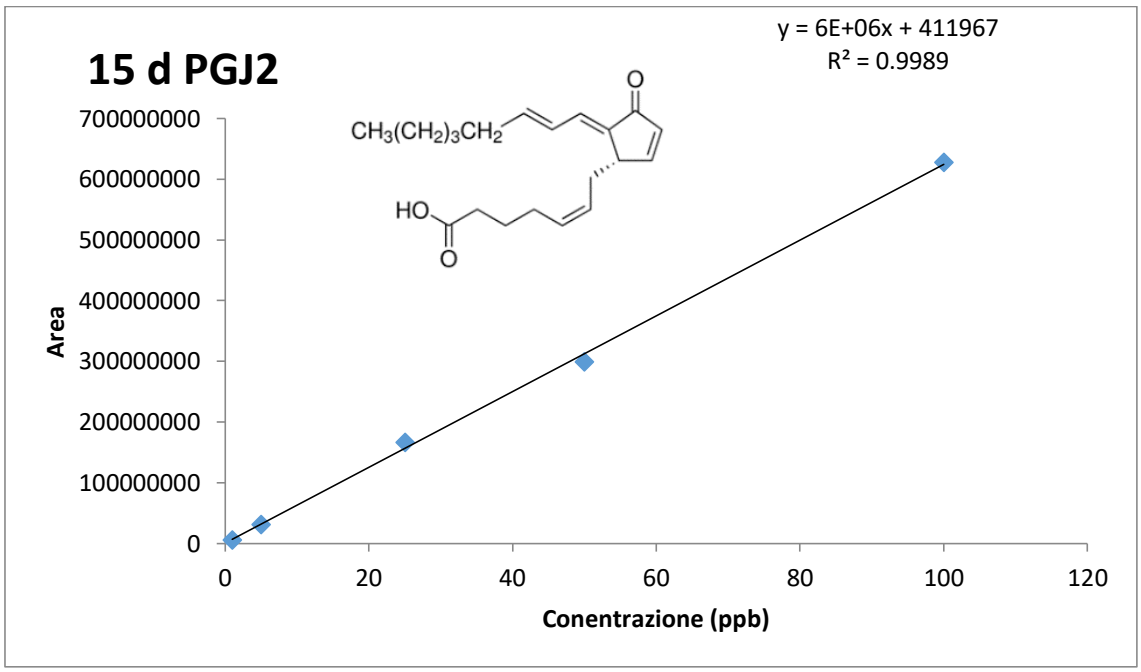
Supplementary Figure 2.6: calibration curve of prostaglandin E_2 (PGE_2) used as standard.



Supplementary Figure 2.7: calibration curve of prostaglandin B_2 (PGB_2) used as standard.



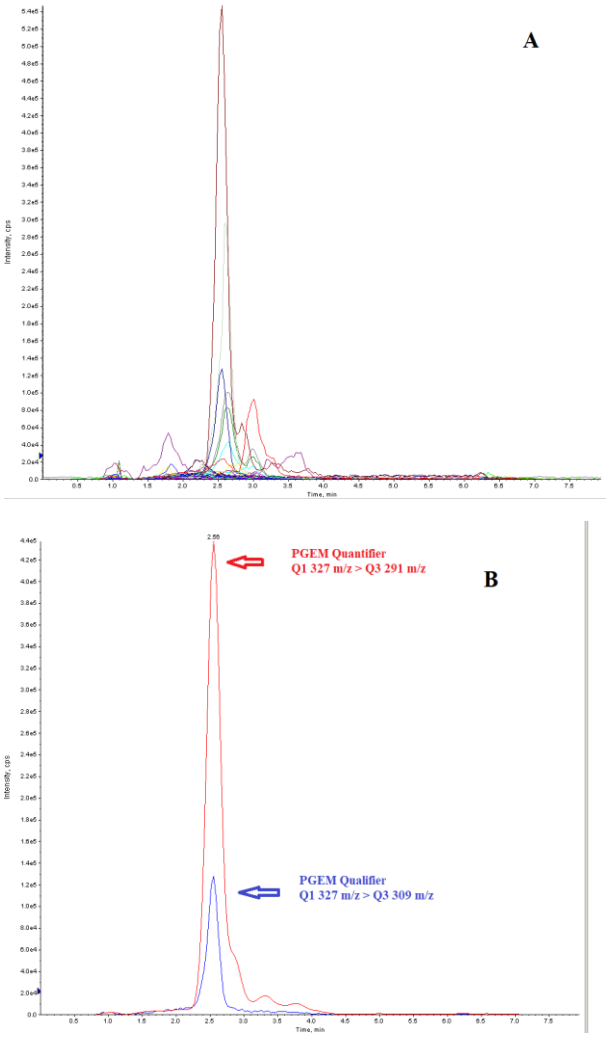
Supplementary Figure 2.8: calibration curve of 15-Deoxy-Delta-12,14-prostaglandin J_2 ($15dPGJ_2$) used as standard.



Supplementary Figure 2.9: examples of TIC chromatograms.

A. Representative sample from the LC/MSMS analysis of the culture medium and XIC chromatogram;

B. PGEM (327 m/z precursor ion; 291 m/z and 309 m/z daughter ion) peaks.



Supplementary Tables.

Supplementary Table 2.1: List of species and sequences IDs used in the phylogenetic analysis.

Species	Annotation	Sequence ID
Chaetoceros neogracile		MMETSP0751-20121128_3624
Asterionellopsis glacialis		MMETSP0707-20130614_6459
Aulacoseira subarctica		MMETSP1064-20121228_17645
Aulacoseira subarctica		MMETSP1064-20121228_57453
Chaetoceros affine		MMETSP0090-20130426_9955
Chaetoceros curvisetus		MMETSP0717-20131115_51459
Chaetoceros debilis		MMETSP0149-20130528_3320
Chaetoceros debilis		MMETSP0150-20130528_3844
Odontella		MMETSP0015_2-20120614_17751
Cyclotella meneghiniana		MMETSP1057-20121228_4220
Cyclotella meneghiniana		MMETSP1057-20121228_12622
Pseudo-nitzschia arenysensis		MMETSP0329-20121206_15499
Thalassionema frauenfeldii		MMETSP0786-20121207_3776
Skeletonema marinoi		MMETSP0320-20121206_19182
Skeletonema costatum		MMETSP0013_2-20120614_556
Detonula confervacea		MMETSP1058-20130122_27510
Skeletonema marinoi		MMETSP1039-20121108_4563
Skeletonema marinoi		MMETSP0319-20121206_26582
Skeletonema marinoi		MMETSP0562-20121206_1237
Skeletonema marinoi		MMETSP1428-20130617_783
Synedropsis recta		MMETSP1176-20130426_10672
Nostoc sp.	COX-1	WP015113127
Chaetoceros cf. *neogracile*		MMETSP1336-20130426_1470
Grammatophora oceanica		MMETSP0009_2-20130614_5553

Gersemia fruticosa	COX-A	AAF93168
Gersemia fruticosa	COX-B	AAS48061
Plexaura homomalla	5S-specific-Cyclooxygenase	AAU87497
Plexaura homomalla	COX-1	AAF93169
Gammarus sp.	COX-1	ADB65786
Caprella sp.	COX-1	ADB65785
Nitzschia sp		MMETSP0014_2-20120614_6887
Skeletonema marinoi		MMETSP0319-20121206_1511
Pseudo-nitzschia pungens		MMETSP1061-20121228_74475
Thalassiosira rotula		TR754|c0_g1_i1|m.3431
Skeletonema marinoi		MMETSP0920-20130426_33289
Fragilariopsis kerguelensis		MMETSP0907-20130614_13332
Skeletonema grethea		MMETSP0578-20130828_1168
Staurosira sp.		MMETSP1361-20130828_23824
Skeletonema marinoi		MMETSP1428-20130617_31509
Homo sapiens	COX-1	NP000953
Ovis aries	COX-1	NP001009476
Homo sapiens	COX-2	NP000954
Coccotylus truncatus	COX-1	AFN20596
Ciona intestinalis	COX-a	XP002127674
Ciona intestinalis	COX-b	XP002123273
Oncorhynchus mykiss	COX-1	CAC10360.1
Fistulifera solaris	Alpha-dioxygenase	GAX23950.1
Cyanothece sp.	Heme peroxidase	ACB53655.1
Fistulifera solaris		GAX18717.1
Alcanivorax nanhaiticus		WP_035233415.1

Hordeum vulgare subsp. *vulgare*		BAJ90503.1
Triticum aestivum		CDM84254.1
Rhizophagus clarus		GBC07128.1
Emiliania huxleyi		XP_005780718.1
Rhodococcus gordoniae		WP_064063209.1
Herbidospora mongoliensis		WP_066371138.1
Eutrema halophilum		BAJ34623.1
Pisum sativum	Alpha-dioxygenase	CAH05011.1

Supplementary Table 2.2: List of prostaglandin identified in the culture medium with their LC/MSMS parameter.

Molecule ID	rt	Q1	Q3	DP	CE
PGE_1	1,13	353	235	-60	-29
15-d-PGD_2	2,14	333	271	-60	-22
15-d-PGJ_2*	2,21	315	203	-60	-20
PGE_3	2,28	349	269	-60	-24
PGE-M	2,56	327	291	-60	-23
PGA_2	2,64	333	271	-60	-20
PGB_2*	2,69	333	271	-60	-25
PGF-M	2,7	329	293	-60	-25
2,3-dinor-11b-PGF_2	3,05	325	227	-60	-22
PGD_1	3,14	353	271	-60	-29
PGD_3	3,14	349	269	-60	-20
PGE_2*	3,23	351	272	-60	-23

ID: official abbreviated prostaglandin names; Q1: precursor ion (*m/z*) selected in quadrupole Q1; Q3: fragment ion (*m/z*) selected in quadrupole Q3; Q1>Q3: transition pairs utilized for quantification; DP: declustering potential; CE: collision energy for each molecule; rt: Liquid Chromatography retention time in minutes. Asterisks indicate prostaglandins identified by comparison with their own standard reference. The non-asterisk one indicates the prostaglandins identified by comparison with literature data.

Aligned sequences, with gap, utilized in the phylogenetic analysis

***Chaetoceros neogracile* MMETSP0751-20121128_3624**

MVFPNFAQFLTDGFAK-----SIPG----------TNFLRNNATHEADLCQIYGRNKDQTDCLRLMS-
SSIGEKGRLKSCMING-----EEWPMPYFLESGEVDP----------
QFERLDPPKFSFDHVIEQLSRIDPNGRLGLVQ----------
KIKKNIFAVGSDRGNTTPGVAAFASLFLREHNRLAAEVENR---------NPA------
WDDGQIFQTARNINIVIYIRIIMEDYLNQLSHAP----------
FDFKLDPGPWTWDAKWNKK------FRVAVEFSTVYR-WHSIIPNAIQIGE----------
GDPLPIMQAMFNNELLIDR------TLLKSFEDSSSQRATSFTPFNTAD----------
IMLPREYNTILQARQAGIRYYVDYLKQFGFPVDLPKNYSDITS-DEKVQ----------
KMLEEMYGKG--NVDKVEFYIGAICATHERN-A----
PFSLFMNLQVAHDAIKAFYMNPLLQRSSWK----

***Asterionellopsis glacialis* MMETSP0707-20130614_6459**

MILPTFANHLTGGFIK-----STRG--PDGS------
YQYGRTHSNHNIDLLQLYGRTQEQTLALRLRDNTTFGKKGRLKSQILNG----
GEEWPPFLYKDGRIDPQ----------FEVLDPIAHGMSHLVDIMEGD----------
----------RSKFFAMGSNQQNMLPQNIAMTTLFLREHNRVAGELERR---------YSD------
WDDDRIFETARNIVIVIYLKIVIEEYINHISPLP----------
DIVKFLVDPGPWMWNAEWNKP------NWISAEFAILYR-WHALMPNKVMHA----------
GIVFDLSSTLFQHDLLLGED----RSLKKTFVDMSAQRSASCECLNTAG----------
PLVEREVESLRYSRALRFRPYVEYVEYWGLK--KPRRFEDITK-NAEVA----------
SMLRDLYGT----VDKVEFFVGLIASDHSKN-G----IFGSSMNIGVGLDAFSQALTNPLLSEHVWK----

***Aulacoseira subarctica* MMETSP1064-20121228_17645**

MLFPVFAQMLIDSFIS-----TVTYTNTTTK---KLTFDWKRTNSPFEINLLPLYGRFEYQTDALRLKS-
EKS--RGRLKSQIING-----EEYAEFLYNDLGEIKD----------EFQVLGPPQA-
FNQILSKVDKTEEEKR----------
KIKSKIFAFGGNRTNTTPQMAALNTLFLREHNRLAGQLEKH---------NED------
WDDDRVFQTARNINLVIYLKIIIEEYINHITSS----------
GARFKVQPKEWVWNADWNKP------NWIAVEFAVLYR-WHALTPNSYLWE----------
GKRVKISDDLFNNALLLET----SGGLRQAIAEISKNPATIMAPFNTAL----------
ELLSQEQAAFAQTRQANVRPFADYRAYLGLP--LVKTFKDITQ-DKEVQ----------
QKLEELYKT----PDRVEFWAGLVAEDKDPK-A----IFGPTLSTLVALDAFSQALTHPLLSEQVFN----

***Aulacoseira subarctica* MMETSP1064-20121228_57453**

MLFPVFAQYLIDSFIL-----TKRIKTEGGT---VSTIDWKRSDSPNDIVLLPLYG-FKNATDALRLKS-
EEAQKRGKLKSQLIDG-----EEYAPFLYDESGMVKE----------EFKALG-
EPESLGEILKAMMADSNTVK----------

KYKSKIFAFGLVRVNITPQLAAMNTLFLREHNRLAGVLEEN---------NPE------
WDDERVFQTARNINLVIYLKIVIEEYINHISS----------
GVQFKAKPEKWIWHASWNKP------NWISVEFAVLYR-WHTLVPNSDEWN----------
GIRYDFQNEIFNNEILLD-----AGGLRHSFASISANRAPSICPFNTGG----------
FLLPREASALQQSRVNKLRPFGEYREYLGYK--KTKEFSDISS-DPQVQ----------ETLRKLYKT-
---PDRVEFYAGLIAEDHVAK-G----ILAPIMTDLVAKDAFTQALTHPLLSENVFN----

***Chaetoceros affine* MMETSP0090-20130426_9955**

MLFPTFAQHLIDSFIV-----TAVKSDGGSG----TEFEWKKTDSPHDIGLLPLYGRTFDQTKQLRVQN-
PPRGKYGQLKSQIIHG-----EEYAPYLYDADGKVKK----------EFDLLE-
TPQGLERSLSMLSPE--DAKA----------K-
KSNIFAFGGARTNLVPNITAWNTLLLREHNRIAQTIEKE---------EPT------
WDDERVFQTARNVLLVIYLKLVVEEYINHITGY----------
GIDFTVDPGKWMWNAPWYKR------NWISAEFAVLYR-WHGVIPSCMKWG----------
DKTLSTHESLFNNAVLTED---MKGSLRDTFINISNHRATQMNLFNTES----------
MMVLRDMAALKQCRACKIKPYADYVVYLGTKE-RPTKFSDISK-DKEVQ----------
EALEKVYKK----VENVEFWTGLLASDNPPE-G----IMSPEMTTFVANDAFNQALCHPLLSENVWS--
--

***Chaetoceros curvisetus* MMETSP0717-20131115_51459**

MLFPTFAQHLIDSFIV-----TAIKENSSEG----VVFNWRKTGSPHDIGLLPLYGKTIEQTTQLRLKS-
EVKGSKGRMKTQIIKG-----GEWSPYLFDSKGNKKA----------EFSSLP-
DPEGLEFALMEAKKFGVDAEK----------N-
KASVFALGGSRANLTPNIVAWNTLLLREHNSIAAKIEAE---------NPE------
WDDERVFQTARNVNLAVYLRLVIEEYINHITAF----------
GVDFTVEPGKWMWDAPWYKR------NWISAEFAVLYR-WHAVIPNCMKWG----------
EKTLPTAAYLYSNHLLLDDT-GLKGDLREAFINISNHRATSMEIHNSEK----------
WMTGRDSRALQMSRECELRSFTEYCAYLGKP--VPKTFADITS-DVEVQ----------
NELKALYGK----VENVEFWVGLIAKDHPTE-A----IMSAELTTFVANDAFNQALTHPLLSEHVWP---
-

***Chaetoceros debilis* MMETSP0149-20130528_332**0

MLFPTFAQHLIDSFID-----TVYHYDDDGN----VVFDWKRTETPHDIGLLTLYGKTIPETKQLRKQS-
ETSGEKGKLKSQLVNG-----EEWAPFLYDSNGKVKE----------EFNELP-
VPQGIDEKMYAARPQ-VQAKL----------K-
ESIFAFFGGSRTNLTPNISAWNTLLLREHNRIAGLIEEE---------NPT------
WDDERVFQTARNCTLVIYLRLVIEEYINHITAY----------
GVDFEIEPEKWMWDAPWYKR------NWISAEFAVLYR-WHAVIPSLMKWG----------
KNTHTTMDYLFSNNLLLSDD-GMKGNLRDCFHNICDHRATNMQLHNSEGG----------
FMVGRDKSALEMSRSCKLRSFSEYCGYLGTP--APESFADITQ-DKDLQ----------
KELKDVYGE----VKNVEFWTGLIAKDHSCE-A----IMSAELTKFVANDAFNQALTHPLLSEHVFN---
-

***Chaetoceros debilis* MMETSP0150-20130528_3844**

MLFPTFAQHLIDSFID-----TVYHYENDGN----IVFDRKLTGTPHDIGLLTLYGKTIPETKQLRKQS-
DTSGEKEKICPNLLKE-----KSXAPFLYDSNGKVKE----------EFNKLP-
VPQGIDGKMYAARPQ-VRAKL----------K-
ESIFAFFGGSRTNLTPNISAWNTLLLREHNRIAGLIEED---------NPT------
WDDERVFQTARNCTLVIYLRLVIEEYINHITTY----------
GVDFKIEPEKWMWDAPWYKR------NWISAEFAVLYR-WHAVIPSLMKWG----------
KNTHTTMEYLFSNNLLLSDD-GMKGNLRDCFHNICDHRATNMQLHNSEGG----------
FMVGRDKSALEMSRSCELRSFSEYCGYLGSP--APESFADITQ-DKDLQ----------
KELKDVYGE----VKNVEFWTGLIAKDHSCE-A----IMSAELTKFVANDAFNXGFD----SSSPFR----

***Odontella* MMETSP0015_2-20120614_17751**

LLFPTVAQHLIDSFIN-----TKYTKTING-----TFFDWARTDSPHEIGLSPLYGCNPKQTSQLRELN-
EKIGCKGRLKTQILGD---GGEEWAPFLYGDSGKIKA----------EFNELD-
EPVALDHILGFNSENAEMMR----------
KSIFAFGGSRANLNPNIVAWNTLLLREHNRVASKIEAS---------EPS------
WDDERVFQTARNVVLVIYLKIVIEDYVAHIS----------
GVNFKVQPGEWMWNAKWYKR------NWMSVEFAVLYR-WHALIPNTTFWG----------
SKTASTKDALYNNTLLLNKEQGCAGNLRNIMVQISEQRITSFQLNNTED----------
WLVDRDLNALRQSRECNLASYADYCEYLGMD--RPKTFTDISL-YPEVQ----------
EKLRSLYKT----ADKVEFYVGLIAADHSPGGK----LFSEAMTRFVANDAFNQALANPLLSENVWK-
---

***Cyclotella meneghiniana* MMETSP1057-20121228_4220**

LLFPTFAQHLIDSFIN-----
TRIDHEATEKNGGRPVFDWARTDSKHEIGLSPLYGDTEEQLNQLRLLS-NVNGFKGLMKTQIINE-
----EEWAPFLYEVDGSKKK----------EFSAIW-DPSGASYVLGMRRDVGERQK----------
----------RTLFAFGGARANLNPNIVAWNTLLLREHNRIAGEIEKS---------EPS------
WDDERVFQTARNVLVVIYLKLVIEEYIAHIA----------
GVNFKVDPGEWMWNAPWAKT------NWMSTEFAILYR-WHAVIPNTSSWGK----------
AKNLKVLDTLFNNDLLLDTKEGLSGNLRDAFVSISEERVTAHQLFNTEE----------
WMVDRELAAIKQGRANKVASYADYAEYVDLP--RPKTFADISL-YPEVQ----------
KALEEVYGT----VDRVEFYVGLIAVDMGAGGK----IFSP-
MTKFVANDAFNQALTNPLLSMNVWN----

***Cyclotella meneghiniana* MMETSP1057-20121228_12622**

LLFPTFAQHLIGSFIN-----
TKIDHEATEKNGGQVIFDWAKTYSRHDIGLSPLYGDTTEQTDQLRLMS-
NVTGFKGRMKTQLIND-----EEWAPFLYRVDGTKKP----------EFNAIW-
DPAGASHVLKMRGEEGFKQK----------
RTLFAFGGARANLNPNIVAWNTLLLREHNRLAGEIEKS---------EPS------
WDDERVFQTARNVNIVIYLKLVIEEYIAHIS----------

GADFKVDPGEWMWNADWNKA------NWMSVEFAILYR-WHAIIPNTINWG----------
TSNMKVSDILFNNDLLVKETEGLDANLRDVFVQMSEQRATAHQLFNTEE----------
WLLDRELAAVTQGRANKVASYADYVEYLNLP--RPKTFADISL-YPEVQ----------
NALKEVYGT----VARVELYVGLIAADMGAGGK----IFSL-
MTKFVANDAFNQALTNPLLSQNVWK----

***Pseudo-nitzschia arenysensis* MMETSP0329-20121206_15499**

LLFPTVAQHLIDSFIN-----THVDRKATEAKG-
SPVFNWAKTDSPHEIGLSPLYGDSKEQTDQLRERS-EEVGRKGRLKSQMIEG-----
EEWSPFLYDSKGNKKT----------EFSKIP-DPSGMSHILKHVNAATAKKA----------
----------SIFAFGGARANLNPNIVAWNVLLLREHNRLASMIEES---------EPT------
WDDERVFQTARNVSIVIYLKIVIEEYIKHIS----------
GANFRVAPGKWLWNAQWNKT------NWMSVEFAILYR-WHAIIPNSSSWGP----------
SKDVEVRESLFNNTLLLDKTKGMGAKLADIFVQISNERTTSFELNNTEK----------
WLVDREMAAIKQGRTNNVAPYADYCEYLGYE--RPKTFADINR-DPKVQ----------
EQLKELYGT----PDKVEFYVGLIAGEHPSGGK----IFSK-MTSFVANDAFNQALGNPLLSQNVWE---
-

***Thalassionema frauenfeldii* MMETSP0786-20121207_3776**

LLFPTFAQHLIDSFIN-----TRIKADTDKNK--PPEFEWTRTDSKHEIGISPLYGDEPAQTKQLRELS-
ETSGRKGRLKTQMIEG-----EEWAPFLYNGN-NKKP----------EFSDIP-
EPDGVRMIEKHG---FGDRS----------
TIFAFGGRRANLNPNIVAWNTLLLREHNRLAAEIEKS---------EPE------
WDDERVFQTARNVLTVIYSKIVVEEYVGHIS----------
GVPFEVQPGPWMWNAEWNKT------NWMSVEFAILYR-WHALIPDTIRWGP----------
KVDIGIMKQLFNNTLLLSKENGMGANLKDCFTEISRNRVTSFELFNTEGS----------
YMATREMQAIRQCRAANVAPFADYCEYLGDP--RPKTFEDISR-KPEVQ----------
KVLKELYGT----PDRVEFYVGLIAQDHSAGPK----IFGDVMTKFVANDAFNQALANPLLSQNVWE-
---

***Skeletonema marinoi* MMETSP0320-20121206_19182**

LLFPTFAQHLIDSFIN-----TKLN-TETG------QFEWDQTESKHEISLGPLYGDDVGQTNQLREKS-
EVTGRRGRLKTQILDG----GEEWAHFLYDETGTKKE----------EFNLIR-
DPDGMKHILKALYNSDPAAKS----------
SIMQTIFAFGGRRVNLNPNMVAWNTLLLREHNRLAGEIERS---------EPS------
WDDERVFQTARNVLIVIYLKLVIQEYIGHIS----------
GVKFKLDPGEWMWNAPWYKT------NWMSTEFSILYR-WHALIPNQNGLGP----------
SKDLGVLKSLFNNPLLLDEEKGLGGNLRDVFVDICKARITSFQLFNTEK----------
WMVGREAATISQGRANNVQPYAKYCEYLRIK--PPKTFKDISL-VPEVQ----------
QALEELYGT----PDRVEFYVGLIAADHAPG-K----IFSL-MTKFVANDAFNQALTNPLLSQNVWE----

***Skeletonema costatum* MMETSP0013_2-20120614_556**

LLFPTFAQHLIDSFIN-----TKLN-TETG------KFEWDQTESKHEISLGPLYGDDVGQTNQLREKS-
EVTGRRGRLKTQILDG----GEEWAHFLYDETGTKKE----------EFNLIR-
DPDGMKHILKALYSSDPAAKS----------
SIMQTIFAFGGRRVNLNPNMVAWNTLLLREHNRLAGEIERS---------EPS------
WDDERVFQTARNVLIVIYLKLVIQEYIGHIS----------
GVKFKLDPGEWMWNAPWYKT------NWMSTEFSILYR-WHALIPNQNGLGP----------
SKDLGVLKSLFNNPLLLDKEKGLGGNLRDVFVDICKARITSFQLFNTEK----------
WMVGREAATINQGRANNVQPFAKYCEYLRIK--PPKTFKDISL-VPEVQ----------
QALEELYGT----PDRVEFYVGLIAADHAPG-K----IFSL-MTKFVANDAFNQALTNPLLSQNVWE----

***Detonula confervacea* MMETSP1058-20130122_27510**

LLFPTFAQHLIDSFIN-----TKTKPLESGRL----GFEWDQTFSKHEIGLAPLYGDYEEETLQLREKS-
EASGRKGRLKTQVLDG----GEEWAPFLYDSEGTKKP----------EFSLLP-DPDGLQHIL-
---AMQPETME----------
KKKKSIFAFGGRRANLNPNIVAWNTLLLREHNRLASEIEKS---------EPS------
WDDERIFQTSRNVVIVMYCKIIIEEYIKHIS----------
GVNLRVEPGEWMWNASWYKP------NWISTEFAILYR-WHALIPNTSSWGA----------
SKNIEVVDALFNNNLLISKDTGMGGNLRDAFVEISKTRITSFELFNTEK----------
AMVQRETAAIEQGRFNNVASYADYCEYLELK--RPETFEDISL-KPEVQ----------
QALKELYGS----PDRVEFYIGLIAADHPAGGK----VFSAAMTKFVANDAFNQALTNPLLSQNAWN-
---

***Skeletonema marinoi* MMETSP1039-20121108_4563**

LLFPTFAQHLIDSFIN-----TKIN-VETG------EFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKS-
EKPGRRGRLKTQVLEG----GEEWAPFLYNEDGTKKK----------EFSAIH-
DPDGMKTILGLVYSSDPTTKS----------
SIEQSIFAFGGRRANLNPNIVAWNTLLLREHNRLAGEIEKS---------EPG------
WDDERVFQTARNVLIVMYCKIVIEEYIKHIS----------
GVNFKVEPGPWMWNAPWYKT------NWMSTEFAILYR-WHALIPNEAGLGP----------
SKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEK----------
WMVERESAAINQGRANNVQSYAAYCEYLDIE--PPKTFEXISM-VPERQ----------
QALKELYGT----PDRVEFYVGLIAADHPAGGK----IFSE-MTKFVANDAFNQALTNPLLSQNVWE---
-

***Skeletonema marinoi* MMETSP0319-20121206_26582**

LLFPTFAQHLIDSFIN-----TKIN-VETG------EFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKS-
EKPGRRGRLKTQVLEG----GEEWAPFLYNEDGTKKKE----------FSAIEDGTKKEEFSAIH-
DPDGMKTILGLVYSSDPTTKS----------
SIEQSIFAFGGRRANLNPNIVAWNTLLLREHNRLAGEIEKS---------EPG------
WDDERVFQTARNVLIVMYCKIVIEEYIKHIS----------
GVNFKVEPGPWMWNAPWYKT------NWMSTEFAILYR-WHALIPNEAGLGP----------

SKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEK----------
WMVERESAAINQGRANNVQSYAAYCEYLDIE--PPKTFEDISM-VPERQ----------
QALKELYGT----PDRVEFYVGLIAADHPAGGK----IFSE-MTKFVANDAFNQALTNPLLSQNVWE---
-

***Skeletonema marinoi* MMETSP0562-20121206_1237**

LLFPTFAQHLIDSFIN-----TKIN-AETG------EFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKS-
EKPGRRGRLKTQVREG----GEEWAPFLYNKDGTKKE----------EFSAIH-
DPDGMKTILGLVYSSDPTTKS----------
SIEQSIFAFGGRRANLNPNIVAWNTLLLREHNRLAGEIEKS---------EPG------
WDDERVFQTARNVLIVMYCKIVIEEYIKHIS----------
GVNFKVEPGPWMWNAPWYKT------NWMSTEFAILYR-WHALIPNEAGLGP----------
SKDAGVMEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSFQLFNTEK----------
WMVYRESAAINQGRANNVQSYAAYCEYLDIE--PPKTFEDITM-VPERQ----------
QALKELYGT----PDRVEFYVGLIAADHPAGGK----IFSEAMTKFVANDAFNQALTNPLLSQNVWE-
---

***Skeletonema marinoi* MMETSP1428-20130617_783**

LLFPTFAQHLIDSFIN-----TKIN-AKTG------EFEWDKNDSKHEIGIGPLYGDEVEQTNQLREKS-
EAPGRRGRLKTQVL-E---GGEEWAPFLYNEDGTKKK----------EFSAIH-
DPDGMKTILGLVYSSDPTTKS----------
SIEQSIFAFGGRRANLNPNIVAWNTLLLREHNRLAGEIEKS---------EPG------
WDDERVFQTARNVLIVMYCKIVIEEYIKHIS----------
GVNFKVDPGPWMWNAPWYKT------NWMSTEFAILYR-WHALIPNEAGLGP----------
SKDAGVMEALFNNPMLLDDKTGLGGNLRDIFVDISQTRVTSFQLFNTEK----------
WMVYRESAAINQGRANNVQSYAAYCEYLDIE--PPKTFEDITM-VPERQ----------
QALKELYGT----PDRVEFYVGLIAADHPAGGK----IFSEAMTKFVANDAFNQALTNPLLSQNVWE-
---

***Synedropsis recta* MMETSP1176-20130426_10672**

PTKPWLGQGCRGMALR---
FPKRPLPDNLPDAKTLVEDFCIRPDDDASFTPCANGVNSLTPYFALCVIHDFFRSDTGRSKVGKLD
R---PWVNLHSSYLDLQTVYGYN----------KKSCASVRTQKHG----------
----------KLTEKKIVDHRLERMAVCHALILVMVKHHNFICDQLMER---------YPD----
KFKTDEMIFQTARLINCGVYINLIIELYSCLFHVWN----------
EDGSNPVELRGLDYPRDIQGY-------HLSYEFNIMYR-FHAFIPKEWTPFYG----------
LKEAIAYLKKGDFTHSVPALSDDAIKTLLVDAMSHRAGANHVPSNVPR----------
VMGPAEVKGIEDARLLGISTYNDFKEAVGEP--KYDTFLDMSGNRPELA----------
AKLEKHYPT----VDDVEFAVGMRVETRTPLRG----AGFSTVGRAILADAFSSIRFDRFYSQSEYH----

Nostoc sp **WP015113127**

MLFPYWVQWFTDSFLR-----LDHTN----------KLKNTSNHEIDLCNVYGLTRKQTHLLRSFQ------
GGKFKTQKLKRQDGVEEEYPLFYYADPAQG----------IVDPQFAGLYE-PVNDE---
KRQPAD----------KKQYLFAMGVERANVQIGYVMLNTLCLREHNRLCDVLASN--
-------YPD------WDDERLFQTSRNILMAIILKIIMEEYINHITPY----------
----------HFKLFADPEAFTK-ESWYRT------NHMAIEFDFVYR-WHSAIPETFNCNG----------
KPTHVADTLWNNKILIDQG------LGALMEETCSQAGTRIGLFNTPNI----------
LVDLAELPSIKLGRQLQLASYNDYRELCGFP--RVTSFDQITG-DEFTQ----------QKLKEFYGH-
---VDNIEFFVGLYSEDVRQN------STIPPVARLIGIDAFSQALTNPLLSPKIFN----

Chaetoceros cf neogracile **MMETSP1336-20130426_1470**

LMLPIFAQWFTDSFLR-----TKFRVDGPQD--------FKENESNHEIDLCQIYGMTETQTAMLRSME----
--NGFLKSQIIDK-----EEFPAHLFEETASG---LV----------QIKPEFAGLYT-
EANVQRVFRRASKE----------
HKLNCFAVGIEHGNSTMGNTLMNIIWMREHNRIAREIAKA---------HPT------
WDDERLFQTARNVNIVLLLKVVVEDYIFHIS----------
GFPFKMDN-KIAEGERWYRS------NWMAAEFALLYR-WHDLIPDTVEFGG----------
ERKDSSALNKNNRWLMKVG------VDRACLDASQEPSGKIMLGNTPNF----------
LVEVGKMS-LKQGRVSGLASFNDYCEHYGLK--RKKNFMDLTGNNKEVS----------
KKLEDVYGS----IDNLEWFVGLWAEAYDDKTK----CMGP-
LTYMVGNDAFTQALTNPLLAAEVFN----

Grammatophora oceanica **MMETSP0009_2-20130614_5553**

LLLPMFAQWFTDGFLR-----TKWKPPTEQD--------YVENESTHQIDLNQIYGATEVQTNILRAKK----
--GGKLRSQMIKG-----EEYPEYLFDTET------L----------QIKPEFEGLYS-
DVNFTRVFQNVSDE----------
QKKVTFAVGLEHGNSTVANTIMNTLFLREHNRLAGIIEAA---------HKD------
WDDERIFQTTRNCLIVILMKIVINDYVAHIA----------
DAPVFLDAGGFAEDQPWYRE------NWMSVEFNLLYR-WHDMIPDQVSFTSADQVSFS---
SEDEYWDMESSKTLVNNNTLMLELG------VEQILLDASKQRAGQLGLFNTPDF----------
LLHVHKET-MKQCRMINLAPYVDYCKEYGLT--VPKTFRELTGGNLEYA----------
KALESVYDS----VEHVEWFVGLWAGG-KALGQ----FSSDLLILMVGHDAVTQLFTNPLLSRRVYN-
---

Gersemia fruticosa **AAF93168**

WLFMFFAQHFTHEFFK-----TIYHS-PAFT----------WGN-HGVDVSHIYGQDMERQNKLRSFE------
DGKLKSQTING-----EEWPPYLKDVDNVTMQY----------PPNT----------
----------PEDQKFALGHPFYSMLPGLFMYASIWLREHNRVCTILRKE---------HPH------
WVDERLYQTGKLIITGELIKIVIEDYVNHLAN----------
YNLKLTYNP-ELVFDHGYDYD------NRIHVEFNHMYH-WHPFSPDEYNISG----------
STYSIQDFMYHPEIVVKHG------MSSFVDSMSKGLCGQMS-HHNHGA----------

YTLDVAVEVIKHQRELRMQSFNNYRKHFALE--PYKSFEELTG-DPKMS----------
AELQEVYGD----VNAVDLYVGFFLEKGLTT-S----PFGITMIAFGAPYSLRGLLSNPVSSPTYWK----

***Gersemia fruticosa* AAS48061**

WLFMFFAQHFTHEFFK-----TIYHS-PAFT----------WGN-HGVDVSHIYGQDIERQNKLRSFQ------
DGKLKSQMING-----EEWPPYLKDVDNVTMQY----------PPNT----------
----------SEDQKFALGHPFYGMLPGLFMYASIWLREHNRVCTILRKE---------HPH------
WEDERLYHTGKLIITGELIKIVIEDYVNHLAN----------
YNLKITYDP-ELVFDHGYDYD------NRIHVEFNHMYH-WHPFSPDEYNISG----------
STYSIHEFMYHPEIVVKHG------MSSFVDSMSKGLCGQMS-HHNHGA----------
YTLDVAVEVIKHQRELRMQSFNNYREHFALE--PYKSFEEMTG-DPKMA----------
AELQETYGD----VNAVDLYVGFFLEKGLTT-S----PFGITMIAFGAPYSLRGLLSNPVSSPTYWK----

***Plexaura homomalla* AAU87497**

WLFMFFAQHFTHQFFK-----TVHHS-PAFS----------WGN-HGVDVSHIYGQGVERENKLRAFK------
DGKLKSQMING-----EEYPPYLKDVDDLKMQY----------LENT----------
----------AEEQKFALGHPFFSMLPGLFMYATIWLREHNRVCMILRKE---------HPH------
WEDERIYQTGKLIITGELIKIVIEDYVNHLAN----------
YNMKLRYDP-QLVFSRNYDYD------NRIHLEFNHLYH-WHPFSPDQFNISG----------
TTYAIKDFMYHPEIVVKHG------MSSFVNAMSSGLCGKMS-HHNHGQ----------
YTLDVAVEVIKYQRELRMQSFNXYRRHFGLH--AYKSFEEMTG-NPKMA----------
AELKEVYGD----VNAVDFYVGFFLEKSLTT-S----PFGITMIASGAPYSLRGLLSNPVSSPTYWK----

***Plexaura homomalla* AAF93169**

WLFMFFAQHFTHQFFK-----TVHHS-PAFS----------WGN-HGVDVSHIYGQGVERENKLRAFK------
DGKLKSQMING-----EEYPPYLKDVDGLKMQY----------LENT----------
----------AEEQKFALGHPFFSMLPGLFMFATLWLREHNRVCMILRKE---------HPH------
WEDERIYQTAKLIITGETIKIVIEDYINHLAN----------
YNMKLRYDP-QLVFSRNYDYD------NRIHLEFNHLYH-WHPFSPDQFNISG----------
TTYTINDFMYHPEIVVKHG------MSSFVNAMSSGLCGKMS-HHNHGQ----------
YTLDVAVEVIKYQRKLRMQSFNNYRRHFGLP--AYKSFEEMTG-DPKLA----------
AELKEVYGD----VNAVDFYVGFFLEKSLPT-S----PFGITMIASGAPYSLRGLLSNPVSSPTYWK----

***Gammarus sp* ADB65786**

LLFQYYAQHFTHQFFR-----TNYTKGPQFT----------KGN-GGVDVSNIYGLTERQRRALRSNV------
DGKLKFQIING-----EHFPPYLKDVPGISMEY----------PPHLP----------
----------ITEDNKFALGHPFFALLPGLFVYSTIWMREHNRVCEVLKEQ---------HPH------
WDDERLYHTAKLIITGEVIKITIEDYVQHLSQ----------
YKVDLKFKP-QVVHGTRFQFH------NRINVEFDHLYH-WHPLIPEGIKVED----------
SYYSLMDMAFSTKSVFTHG------LDAFVKALVTNRAGKLT-SRNHSP----------
VTVPVLKKMLENSRILRFQGVNQYRKKFNMR--PFRDFLDLTG-DEELA----------

RDMEEMYGD----INAVEYYVGLIAEKDSP--S----LTPLTMVNVGGPWSVKGLIANPICSPHWWK---
-

***Caprella sp* ADB65785**

LLFQYYAQHFTHQFFR-----TNYTMCPQFT----------KGN-GGVDVSNIYGLTEQHRRAIRMNS------
DGKLKYQVIND-----EHYPPYLRDVQGIEMDY----------PPHIP----------
----------ITEDNKFALGHPFFALLPGLFVFSTIWMREHNRVCDVLKNQ---------HPD------
WDDERLYQTAKLIITGEVIKITIEDYVQHLSQ----------
YKVDLKFKP-QVVHGTRFQFD------NRINAEFNHLYH-WHPLIPDGIQVED----------
KYYSLMDMAFSTKSVFTHG------LDKFIESMATSRAGKLS-HSNHPL----------
VTLPVLKKMMENGRKLRYQGINEYRKRFALK--PFKDFMDLTG-DEALA----------
KDLQELYGH----VDAVEFYVGLLTEKDSP--S----LTPLTMVNVGGPWSVKGLIANPICSPHWWK---
-

***Nitzschia sp* MMETSP0014_2-20120614_6887**

PQLNFWMLSFVNWFHD-----DNFRTLPDTDG-------
AFTWSDRGSLHMTHLYGHTEYRQAALRTMA-----GDGKMKTSSRLG----
WDYYPPLLMDVQADFPD----------FDMWTSQRGSKHKSTSAGQTSEQADEN----------
----------MPYYFAIGDPRFNLHLGHILWTSVGLYLHNTACVILQRE---------DPG------
LTDEDIFQRARVIVFHIIQKIRLQDFVMDSISSTR----------
DHIRIPYDPKLLREEFAHHFAYSGGN-QPNFLEFNHLYQAWHALIPNGLVLNEDVD---------
DGEKDILPIRKTLWAPKLMTTNFT-----IGEMATSFASTPLTLYSPHNFPVF----------
LRGVTEAALKDERAQRMAPYNSYRELIGLD--PITSFEQFAVDDP----------QKMAELYNN---
DVDSVDFIAGILADSNPHLPG----NFFGVQLVLVALFALQDLANNPLILDPVMS----

***Skeletonema marinoi* MMETSP0319-20121206_1511**

EADVALVANIMKRDTN--SFAPFNQLASAWIQFMT--HDWFQHDASSS--QGLK----------
----------MKNVVTHWWDASQMYGS----------SQEEVDAVRAE-
GGKLHLDVNDEIDYN----------
ASLPITGFRENWWAGLHILHTIFAREHNHIVDILAQS---------YPS------
MTEDELFGTARNIIAVLLAKIHTVEWTPTLLDNAVSDMALNINWHGLQTVTSMYFKGKNIPDEV
KDIIDEMK----------VPSVFGSNYTTEQTLFNTP------FYMTEEFVSVYR-
MHTLLPDAMILEGGKT----------VSLQELAFTDARNLVSDPSKTTAT-------
LLQAFAHTPAQALSLKNYPKSLFNLQIGNGK----------SINLAEIDISRDRARGIPRYNDARRQLLLT-
-PYKSMDDLTSDKE----------ELKLLKSVY----TDIEQVDFLVGCLVDKDRPDGF----
AFGIVPYYIFVVMASRRLLSDRFFQEGLTEENYS

***Pseudo-nitzschia pungens* MMETSP1061-20121228_74475**

DNVQVIAQRLLAREGFKPAGNQLNIIAAAWIQAQV--
HSWIQHLDGAPTSIEATAEAVGPVCPVKK--MNFFETE-ERPD----GE----
YNSFRTQWWDASFVYGQ----------NREQVHLGRLYKDGKLKVNESNPDTLSFV------
----------EEGKSKIDVVGDQSNSWVGVTVLQVLFIKEHNYCAEMIKKE---------NPQ------

LTDEEIYGHCRNIISALVAKIHTIDWTVELLKTEQLRVAMEINWKG----------
ATKAVFGDKAPF----------HP--LRLINKPKADNKGVP------FCLTEEFAAVYR-
LHPLLPPGLVVEHG----------EGKEEFIGIEKLLTTKGRDKMREPG---
MAKKIMFSAFHYSCGHLKSSNYPFIMRKFTPTDHKGVDLQPPEDRVVDMAAIDLHRDRERGIQK
YNEFRRQLKLR--PFKTWEALTGEENKSDN----------
ELTDAKKLELIYGPAPEGIEKIDLLVGDLYERKIADGF----
ALSETSFMIFLLMASRRIDSDPYLNEYYNEEYYT

***Thalassiosira rotula* TR754|c0_g1_i1|m.3431**

MLFPTFAQHLIDSFIV-----TAIKENTSEG----VEFDWRKTGSPHDIGLLPLYGKTIEQTTQLRLKS-
ELKGSKGKLKTQIIKG-----GEWSPYLFDAKGNKKP----------EFSSLP-
DPEGLEFALGMAKQHGVDVKK----------N-
KATVFALGGSRANLTPNIVAWNTLLLREHNRIASKIEEE---------NPK------
WDDERVFQTARNVNLAVYLRLVIEEYINHITGF----------
GVDFTVEPGKWMWDAPWYKR------NWISAEFAVLYR-WHAVIPNCMKWG----------
EKTLPTASYLYSNHLLLDED-GMKGDLREAFINISDHRATSMQIHNTES----------
WMKGRDARALQMSRACELRSFTEYCAYLGAK--VPKTFADITS-DVEVQ----------
NELKALYGK----VENIEFWVGLIAKDHPTE-S----
IMSAELTSFVANDAFNQALTHPLLSEHVWAAGPE

***Skeletonema marinoi* MMETSP0920-20130426_33289**

LLFPTFAQHLIDSFIN-----TKIN-VETG------EFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKS-
EKPGRRGRLKTQVLEG----DEEWAPFLYNEDGTKKE----------EFSAIH-
DPDGMKTILGLVYSSDPTTKS----------
SIEQSIFAFGGRRANLNPNIVAWNTLLLREHNRLAGEIEKS---------EPG------
WDDERVFQTARNVLIVMYCKIVIEEYIKHIS----------
GVNFKVDPGPWMWNAPWYKT------NWMSTEFAILYR-WHALIPNEAGLGP----------
SKDAGVMESLFNNPVLLDDETGLGGNLRDIFVDISQTRITSFQLFNTEK----------
WMVERESAAINQGRANNVQSYAAYCEYLDIE--PPKTFEDISM-VPERQ----------
QALKELYGT----PDRVEFYVGLIAADHPAGGK----IFSEAMTKFVANDAFNQALTNPLLSQNVWE-
---

***Fragilariopsis kerguelensis* MMETSP0907-20130614_13332**

PNVQVVAQRLLARESFTPAGDQLNIVAAGWIQAMV--
HDWMKHEDGKKTSMEVTPAVVGSQCPLHR--FNLFETK-ERPDGH--------
YNSERTNWWDASFVYGQ----------NAEQVQNSRAFVGGKLKVNEKNPDTLPSR------
----------DDG---TDLTGDQSNSWVGVSVLQTLFLYEHNYCAEQIAKE---------NPN------
LTDHQIYGHCRNIIAALVAKIHTIDWTVELLKTPQLKIGMRTNWMG----------IIQAITGLKIPF---
----------LDRLLRLIKKKENNNEGVP------FCLTEEFAAVYR-LHPLLPPGLIIEGEGA-----
GDKDEDEDTFIDLRDTLTTKGRDLMRKSG---
MAKKVMKAVFTYPCGNMAPSNYPDVMRDFHPTDLLGNNL----
DDRIDLAAIDLFRDRERGIQYFNNFRRKLSMK--

PFQTWEELTGDDKMTEEALAAFVAGTGNLTNAKKLELVYGAAPKGIEHCDLLVGDLYEKKIP-
GF----AISETSFMIFLLMASRRLDADPYLNEYFDEEHYT

***Skeletonema grethea* MMETSP0578-20130828_1168**

ILVAYSIXSFQYVLDFKPAASQLNVLAASWIQAMV--HDWIGHFDGDET--
ETLDRGGESLCPFAKSPFSFKNTKTEKIDGVPF------SPSERTNWWDASFVYGN----------
---NSEQIDRARTMQGGKMVTSD-IPHALAED----------KDG---
VYFAGDNKNSWVGVALLQDLFIREHNYICDQIAAE---------AKEEGK--
EMTDEELFGKARVVVAALVAKIHTVDWTVELLKTKLLAIGMKTNWDG----------LLKAVG-
IPIPG----------ILSQMGEKKGRVSDNEGTP------FCLTEEFAAVYR-LHSLSPPGLILGDG------
-----DAKDKFIGLEDLLGDEGRKQMRETKT--
RPKEMMKSCLHWPCGALMSSNYPNAFRDVAPTDDYGKDLK---
SQNIDLAALDLFRDRERGILKFNEFRRQLNLK--PYRTWLELTENEE----------
DARKLELIYGPGQEGIERCDLLVGDMYERKVQPSF----
ALSETSFIIFLLMASRRLSADPFLNELYNEETYS

***Staurosira sp* MMETSP1361-20130828_23824**

LLFPFFAQWFVDSFLR-----TKWKPLAEQD--------FKENESNHDIDLNQIYGTSEIQTDMLRSMK-----
-GGRLKSQIIDG-----EEYPVFLFDQKT------A----------TLKPEFRGLYT-
EENFKRVFGNASKE----------
HKLHSFAVGLEHGNSTIGNTVMNTLFLREHNRVAGVISAA---------HPE------
WDDERVFQTTRNVMIVLLIKIVLADYIYHIS----------
GAAVFADPGGFAEDELWYRE------NWMSVEFSLLYR-WHDLIPSSVTFDG----------
ETRDAVDLQNNNRWLLKAG------MDSVIQDASNQKAGVMGLGNTPDF----------
LLSVTKMS-LHMARTCKLSSYVEYCKEYGQD--PPEDFMDLTG-DQDSA----------
SKLERVYGS----IDKVEWFVGLFAQK-RDV------FSGLLMTLMVGNDAVTQAFTNPLLAKRVYN---
-

***Skeletonema marinoi* MMETSP1428-20130617_31509**

EADVALVANMMKRDTN--SFAPFNQLASAWIQFMT--HDWFQHDASSS--QGLK----------
----------MQNVVTHWWDASQMYGS----------SQEEVDAVRAE-
GGKLHLDVNDEIDYN----------
ASLPITGFRENWWAGLHILHTVFAREHNHIVDILAQS---------YPS------
MTEDELFGTARNIIAALLAKIHTVEWTP----------KNIPDEVKDIIDEMK----------
----VPSVFGSNYTTEQTLFNTP------FYMTEEFVSVYR-MHHLLPDEMILEGGK----------
LQELAFTDARNLVSDPSKTTAT-------LLQAFAHTPAQALSLKNYPKSLFNLQIGNGK----------
LINLAEIDISRDRARGIPRYNDARRQLLLT--PYKSMDDLTSDKE----------ELKLLKSVY----
ADIEQVDFLVGCLVDKDRPDGF----AFGI--YYIFVVMASRRLLSDRFFQ---------

***Homo sapiens* NP000953**

LMFAFFAQHFTHQFFK-----TSGKMGPGFT----------KALGHGVDLGHIYGDNLERQYQLRLFK-----
-DGKLKYQVLDG-----EMYPPSVEEAP-VLMHY----------PRGI----------

----------PPQSQMAVGQEVFGLLPGLMLYATLWLREHNRVCDLLKAE---------HPT------
WGDEQLFQTTRLILIGETIKIVIEEYVQQLSG----------
YFLQLKFDP-ELLFGVQFQYR------NRIAMEFNHLYH-WHPLMPDSFKVGS----------
QEYSYEQFLFNTSMLVDYG------VEALVDAFSRQIAGRIGGGRNMDH----------
HILHVAVDVIRESREMRLQPFNEYRKRFGMK--PYTSFQELVG-EKEMA----------
AELEELYGD----IDALEFYPGLLLEKCHPN-S----IFGESMIEIGAPFSLKGLLGNPICSPEYWK----

***Ovis aries* NP001009476**

LMFAFFAQHFTHQFFK-----TSGKMGPGFT----------KALGHGVDLGHIYGDNLERQYQLRLFK-----
-DGKLKYQMLNG-----EVYPPSVEEAP-VLMHY----------PRGI----------
----------PPQSQMAVGQEVFGLLPGLMLYATIWLREHNRVCDLLKAE---------HPT------
WGDEQLFQTARLILIGETIKIVIEEYVQQLSG----------
YFLQLKFDP-ELLFGAQFQYR------NRIAMEFNQLYH-WHPLMPDSFRVGP----------
QDYSYEQFLFNTSMLVDYG------VEALVDAFSRQPAGRIGGGRNIDH----------
HILHVAVDVIKESRVLRLQPFNEYRKRFGMK--PYTSFQELTG-EKEMA----------
AELEELYGD----IDALEFYPGLLLEKCHPN-S----IFGESMIEMGAPFSLKGLLGNPICSPEYWK----

***Homo sapiens* NP000954**

MMFAFFAQHFTHQFFK-----TDHKRGPAFT----------NGLGHGVDLNHIYGETLARQRKLRLFK-----
-DGKMKYQIIDG-----EMYPPTVKDTQ-AEMIY----------PPQV----------
----------PEHLRFAVGQEVFGLVPGLMMYATIWLREHNRVCDVLKQE---------HPE------
WGDEQLFQTSRLILIGETIKIVIEDYVQHLSG----------
YHFKLKFDP-ELLFNKQFQYQ------NRIAAEFNTLYH-WHPLLPDTFQIHD----------
QKYNYQQFIYNNSILLEHG------ITQFVESFTRQIAGRVAGGRNVPP----------
AVQKVSQASIDQSRQMKYQSFNEYRKRFMLK--PYESFEELTG-EKEMS----------
AELEALYGD----IDAVELYPALLVEKPRPD-A----IFGETMVEVGAPFSLKGLMGNVICSPAYWK----

***Coccotylus truncatus* AFN20596**

VLIAYYAQWVTHQFFN-----TDESDPTGHS----------VKQPVGVNMSMLYGSKQEVEKSVRAYK---
---GGLLKSTIKNG-----QEFPEIMPCQEGSRIP----------
----------GKEMFNMPILIANMIPGFAAIHVLFFRRHQYICRELAKW---------AEAQGK--
NIDDEELFQKAKLIVTVNMLRITMHDYVSRALQSSHAKMR----------
---FDQKVKQSRIWKMFGPDYFPPS-----NAIQFEFNIFYR-WHQFYPDTTKIMKR--------
IDDLKFPKSKQQLDEKWNAVRWIADE---PDGMERVLFSASSQRAGKLSLLNTNQW----------
--IVEHVVKPGLARCREHQLASYNDYREKVGFP--RLTTFEQVTS-NPALL----------
EKLKRVYRN----VDQIEYYPGVFAEDKHFG-------NVHGPLTFGSSMTFTGIFSSRLFETALDE----

***Ciona intestinalis* XP002127674**

VLFPFFAQHFTHMFFK-----TDPMKGMPYQ----------WGD-QLVDLSQIYGHGEKRQHELRSHV-----
-NGKLKVSLVDG-----HEFPP-LSNQTTANMSN----------INLL----------
----------PQEYQFVFGHQGFSLMPTFLIWSTIWLREHNRICDLIKEE---------NPA------
WDDERIFQTARLVLTGETIKVVIEDYVQHLSG----------

FHYKLLYDP-ELVQGGSHSFH------NQIHVEFQLLYH-WHALMPDQIEFNG----------
KSYTMKRLLFNPEPVVKGG------LKRTIEDLSNQWAGQVAGGKTQGA----------
ATLHVAGLAIKNGRDLRMQSFNAYKEKFEMK--KYTTFQELTG-EEEMA----------
AELQKLYGD----IDAVEYYIGIMLEKRRSP-Q----LFGETLTEMGSPYSLKGLYSNPINHKDWWK----

***Ciona intestinalis* XP002123273**

VFFAFFAQHFTHQFFK-----TNTIKGMPFQ----------WGE-HSVDLSHVYGHTIQRQHELRSHI------
DGKLKVFETNG-----EVFPP-LTESANVTMSG----------EKLM----------
----------RGR-KFAIGHPGFGAFPSFFVIATLWLREHNRVCDILKDL---------HPD------
WDDERLFQTARLILTGETLKIIVEDYVQHVSG----------
FHFQLSYDP-EILHKSTFSYN------NQIHAEFHILYH-WHMLMPDFIELGE----------
HVYPLKELLFNVDPVVEIG------METVLKQLSNQFAGKVVGGRNQGP----------
ELVAVVELALKQTRQMRMCSFNKYRERFGMK--PYTSFEELTG-ETEVA----------
ALLRNLYYD----IDALELFVGYFVEHRRNR-Q----VLGATMLEMGAPYSLKGVFGNPIGSPAWWK-
---

***Oncorhynchus mykiss* CAC10360.1**

LMFAFFAQHFTHQFFK-----TRNSMGLGFT----------RALGHGVDAGNVYGDNLVRQLNLRLLK----
--DGKMKYQVVKG-----EVYPPTVAEAA-VNMRY----------PQET----------
----------PVGQRMAIGQEVFGLLPGLTMYATLWLREHNRVCDILKAE---------HPT------
WGDEQLFQTARLIVIGETIRIVIEEYVQHLSG----------
YLLDLKFDP-VLLFKSTFQYR------NRIAVEFKQLYH-WHPLMPDSFHIDG----------
DEVPYSQFIFNTSIVTHYG------VEKLVDAFSRQCAGQIGGGRNIHP----------
VVTNVAEGVIEESRTLRLQPFNEYRKRFNLK--PYTSFSDFTG-EEEMA----------RELEELYGD-
---IDALEFYPAIMLEKTRPN-A----IFGESMVEMGAPFSLKGLLGNPICSPEYWK----

***Fistulifera solaris* GAX23950.1**

LPPVKDVAAILRRPMNPTSVAPFNQIAVAWIQMMT--HDWFQHDPAHPDQK----------
----------MNRVTHWWDASQLYGS----------
TLAQQTAVRVPNTGKVRLDQHQELNYTT----------
TGIPITGFADNWWAGLHMMHTLFVREHNWLVDQFERQ---------YPG-----
VYTANDKFQLARLCLSALLAKIHTVEWTPTLLDNPVAALGLHTNWRG---
VDAILEYGTRFELQLAYRIVGGD----------QSVPHAGNGTTRETLYNTT------
FAMTEEFVAVYR-MHPLLPDEMEIEGKTFS----------LNDLSFVDARTLTKS-VKTTQT-------
LLQAFGMTPANTLSLQNYPRQLYGLEKPGMS---------
QPVNLAEIDLQRDRERNLPRYNDMRRQLLLK--PYKRLEDLTDDET----------ELNLLKSVY---
-QDMDQVDLMVGCLVDKDRPYGF----AFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT

***Cyanothece sp* ACB53655.1**

PNPRVVSRTLMTR-EDFKPATILNLLAAAWIQFEN--
HDWFSHGDNKPEDKLEIPLEANDPWPEEYRPLEVGKTLPDTSRPEGAKPP--
TFINTVTHWWDGSQIYGS----------NPETVDQLRSHEDGKLIIGENGLLPVDP----------

----------ETGVDITGFNDNWWIGLGMLHTLFTREHNLICDHLKQE---------YSQ------
WSDDDLFDHARLINAALMAKIHTVDWTPAILPLPATDIALNVNWNGFLG----------
EDIKQVLGTVG----------EGEIADLLTGIIGS-DKNHHTAP------YYLTEEFVSVYR-
MHPLIPDELEFRSLEGD---------KFLQEVNFFEMSGKRTRALLESISL----
PDLFYSFGITHPGAITLHNYPRFLQQLVRDNGE----------
VFDLAAVDILRDRERGVPRYNRFREIMGRG--RVKCFEEISSNKQ----------WVEEMRRVYN---
DNIDQVDLMVGLFAEDT-PEGF----GFSDTAFRVFILMASRRLKSDRFFTTDYRAEIYT

***Fistulifera solaris* GAX18717.1**

LPPVKDVAAILRRPENPSTVAPFNQIAVAWIQMMT--HDWFQHDPTNPDSK----------
----------MNRVTHWWDASQLYGS----------
TLTQQRAVRVPNTGKLRLDKHQELNYTS----------
TGIPITGFADNWWAGLHMMHTLFVREHNWLADQFELQ---------YPG-----
VYTANDKFQLTRLCLSALLAKIHTIEWTPTLLDNPVAALGLHTNWRG---
VDAILEYGTQFELQLAYRIVGGD----------QSVPHAGNGTTRETMYNTT------
FAMTEEFVAVYR-MHPLLPDDMEIEGITLT----------LNDLSFVDARKLTKS-VKTTQT-------
LLQAFGTTPANTLSLQNYPRQLYGLEKPGMS---------
QPINLAEIDLQRDRERNLPRYNDMRRQLLLK--PYKRLEDLTDDET----------ELNLLKSVY----
QDIDQVDLMVGCLVDKDRPYGF----AFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT

***Alcanivorax nanhaiticus* WP_035233415.1**

PNPREVSNLIMSRGGDFKPATTLNFIATSWIQFMV--
HDWFDHGPRTDANPIEFPLPAGDVLGGGTMSVQRTRPDPDVSGDEGIITY--ENIN--
THWWDGSQLYGS----------SKEKNDEVRSFVDGKLKVDGDGRLPTEF----------
----------FSGKPVTGFNENWWVGLSMLHHIFTQEHNAIADMLVAN---------YPG------
QSDQWYFDKARLINSALMAKIHTVEWTPAILANPVLERAMYANWWGLGGDRDKRDKYQDDL
DNLNNNLGQIGGLLDLVGIDNGLGDSPTGSLEHALAGLVGSRTPNNYNVP------
YTLTEEFVSVYR-MHPLLRDEIKVYDIGSN---------VVDEEIALEDTRNGDAEDLLGDIGQ----
DRLWYSFGITHPGALTLNNYPDFLRNLSMPLIG----------
DIDMAAIDILRDRERGVPRYNEFRRQIGLK--PLTSFEQLTSDPQ----------LLADLKALYN---
NDIELVDTLVGQLGEETRPEGF----GFGETSFQIFILNASRRLMTDRFFTTDYTDEVYT

***Hordeum vulgare* subsp. Vulgare BAJ90503.1**
PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMV--
HDWMDHMEDTKQIEITAPKEVANECPLKS--FKFYATKEQPTNSDGIKT---
GYHNIRTAWWDGSAVYGN----------NEKQEKKIRTYADGKLVIGDD-GLLLHE--------
----------ENG---VPLSGDVRNGWVGISILQALFVKEHNAVCDAIKEE---------HPN------
LSDEELYRYAKLVTSAVIAKIHTIDWTVELLKTKTLRAGMRANWYG----------
LLGKKIKDTFGHIG----------GTALGGLVGLKKPINHGVP------YSLTEEFTSVYR-
MHPLIPSTLKLRDPTGQPA--ADNSPPYLEDIDIGELVGLKGEDQLSKIG----
FEKQTLSMGYQACGALELWNYPSFFRDLIPQNLDGTNRS----

DRIDLAALEVYRDRERSVPRYNEFRRRLFLI--PIKCWEDLTSDND----------AIEAIRAIYG---
DDVEKLDLLVGLLAEKKIK-GF----AISETAFNIFILMASRRLEADRFLTSNFNEKTYT

***Triticum aestivum* CDM84254.1**

PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMV--
HDWMDHMEDTKQIEITAPKEVANECPLKS--FKFYATKEQPTNSDGIKT---
GYNNVRTAWWDGSAVYGN----------NEKQAEKTRTYVDGKLVIGDD-GLLLHE------
----------ENG---VPLSGDVRNGWVGVSILQALFVKEHNAVCDAIKED---------HPN------
LSDEELYRYAKLVTSAVIAKIHTIDWTVELLKTKTMRAAMRANWYG----------
LLGKKFKDTFGHIG----------GTALGGLVGLKKPINHGVP------YSLTEEFTSVYR-
MHSLIPTTLKLRDPTGQPA--ANNSPPYLEDIDIGELVGLKGEDQLSKIG----
FEKQTLSMGYQACGALELWNYPSFFRDLIPQNLDGTNRS----
DRIDLAALEVYRDRERSVPRYNEFRRRLFLI--PIKSWEDLTSDKD----------AIESIRAIYG---
DDVEKLDLLVGLMAEKKIK-GF----AISETAFNIFILMASRRLEADRFITSNFNEKTYT

***Rhizophagus clarus* GBC07128.1**

LELEHYITALTQLPLN----
TPEEKKLLSILERLLVTQLWSDISKPPAMIAGDIYRSSDGSGYNRLIQSLGKANSRYSLTSRIQYPIK
LSVLPKSEDIFDKIMVQGGDFVEHPSGISSMLFYLAIIITHDLFHTSFADPNINLTSSYLDLTPLYGS
NDQEQKSIRTLKG----------
GLLKPDTFADSRILLQPPGVSALVILFSRNHNFIAQTLLEKNELGRFSVTNPNDPEQLKKQDEDLF
QTARLINCGFYINVILHNYLRTILGLDR----------
TNSKWFVDPTVPYNKRGQLEPLPSGIGNIVSLEFNYIYR-WHAATSKDDSKFVEDEFKTIFG------
DDWENITIDEFKEKMGVWGRSIPKDPSKWKFNHIERGSDNRFKDTDIAKEIINGTKKVSGAFGAN
RIPKVFRPIELLGIESARVLGLSSLNDFRRSLNLK--PYESFMEMNP-DPLIA----------
KKLEELYGS----
IENVELYPGLMTEKTKPDMLGSAIALPFTISRAILSDAVNLVRNDRYYTNDFSPRNLT

***Emiliania huxleyi* XP_005780718.1**

PFQALHNLPSADEVVR------LLYKRDAFK----------KAPYGVNSLATWFANVAIHDFFRTAT----
GTDGGTHPERGSDK--EWVNLHSSYLDLQPLYGYS----------KTTADATREWS--
----------GGKLKAFAEDRMRRIPESRVIVELLRREHNYVAEQLAQR---------
YPAQ----FATDEELYQQARLIMGGVYINIILRAYGCQMFG----------
---------EIAPDGSGFCELRQGYGGAGVG---NMCTFNFNLIYR-FHTSIPVEWSATD----------
PPPIDTDEQMR----------TLLNGILNWESGGFGPNNVPDSILG----------
ERARVSQRAIEAARLMGAPTLNTFRRRFTSG---YSSFEDMTGGDQATA----------
DTLRQLYPG---GIEDVELVVGCQVEKCMSGGW----
ALPSTIGQAIVADAFASIRQDRFYTQDWGASSYT

***Rhodococcus gordoniae* WP_064063209.1**

VWFAFFAQWFTDSFLR-----TNSKD----------PRKNDSNHEIDLCQIYGVNQAKTTMLRAGY------
GGRLDSQVIDK-----KEYPPFLFAARTPGEEL----------

RFVPKFEGLHDREYLLDTVLRLCPDE----------

RKKSVFAVGLEHGNSTIGNTVLNVLFLREHNRIAGILEGA---------YPEWD------

DDRLFETTRNIMIVILLKIVIEEYIKHIG----------

PFDFPIEFVP-FMADNAPWNRT------NWCAIEFNLLYR-WHSLIPDTVVFDS----------

QRVSTRILVDNNPLVLDRG------IESIIDQCSRQKASRIGLGNTPAFLIDRHPMCPD---------

RESVEERT-VGLMRQARLRSFNDYREAFGLG--RLTSFTELTG-DVEVQ----------

QKLARLYGD----VDAVEWYVGIFAED-YPRHR----

MMGELLTTMVAHDAFTQAFTNPLLARHVYHEDTF

Herbidospora mongoliensis **WP_066371138.1**

VMFMFFAQWFTDSFLR-----TSRDD----------FRRNTSTQEIDLCQIYGLTEEKTHLLRAHR------

DGRLKSQLIDG-----EEYPEFLFRPRSSGEPP----------

VFKPEFEGLHDSEFIVSRLLEDAPEK----------

QKDTFFAVGLEHGNSTIGNTIMNVVFLREHNRIAGLLKQA---------

HPEWAERPEGADARLFETTRNIMIVLLLKLVVEEYIRHIS----------

-------PYDFPLETVP-FIADGKRWNRS------NWISIEFNLLYR-WHSLVPSTIGSGP----------

DRLDSTDFRNNNPLVLARG------IESLVSQCSGERAGRIGLMNTPGFLVDRDPAHPE---------

RPSVQERT-IALSRKARLASYNDYRENFGLG--RLKDFGRLTQ-DAELR----------ERLEKLYDD-

---IDKLEWYVGIFAED-YSRDE----MMGRLMTTMVAYDAFTQALTNPLLARDVYNEATF

Eutrema halophilum **BAJ34623.1**

---MVVATKLLTRRKMIDTGKQFNMIAASWIQFMI--HDWVDHLEDTDQIELSAPKEAAKGCPLSS-

-FRFFKTKEVPTGFFEIKT---GSLNTRTPWWDSSVIYGS----------

NSKTLERVRTYKDGKLKISEETGLLLHD----------DDG---

LAISGDIRNSWVGVSALQALFIKEHNAVCDLLKKE---------YED------

LEDEDLYRHARLVTSAVIAKIHTIDWTVELLKTDTLLAGMRANWYG----------

LLGKKFKDTFGHVG----------SSIFGGVVGMKKPQNHGVP------YSLTEEFTSVYR-

MHSLLPDQLHMRDIDVTPG--PNKSLPLTQEVSMEKLIGREGEETMSQIG----

FTKLMVSMGHQACGALELMNYPAWFRDLVPQDPNGHDRP----

DHIDLAALEIYRDRERNVARYNDFRRAMFMI--PIKTWEDLTDDKE----------AIELLDDVYG--

-GDVDELDLLVGLMAEKKIK-GF----AISETAFNIFLLMATRRLEADRFFTSDFNEMTYT

Pisum sativum **CAH05011.1**

PDPMVVVTKLLERKTYKDTGTQFNVIAASWIQFMI--

HDWIDHMEDTKQVELSAPSEVASQCPLKS--FKFFKTKEIPTGFYDIKT---

GHANVRTPWWDGSVVYGS----------NEQVLNKVRTFKDGKLKISKE-GHLLHN-------

----------EDG---TAISGDIRNSWAGVTTLQTLFVQEHNAVCDALKKE---------NSD------

LEDEDLYRHARLVTSAVIAKIHTIDWTVELLKTDTLLAGMRANWYG----------

LLGKQFKDRFGHVG----------NSILSGFVGMKRSENHGVP------YSLTEEFATVYR-

MHPLLPDSLHLRDISASPG--PNKSPPLIKEIPMNDLIGLQGEKTLLEIG----

NAKKLVSMGHQACGALELWNYPSWLRNLVPHNIDGTERS----

DHVDLAALEVYRDRERNVARYNQFRRGLLLI--PISKWEDLTDDEE----------AIKVLEEVYG---DDVEELDVLVGLMAEKKIK-GF----AISETAFVIFLLMASRRLEADRFFTSNFNEETYT

Original sequences, ungapped, utilized in the phylogenetic analysis

***Chaetoceros neogracile* MMETSP0751-20121128_3624**

MVFPNFAQFLTDGFAKSIPGTNFLRNNATHEADLCQIYGRNKDQTDCLRLMSSSIGEKGRLKSCMINGEEWPMPYFLESGEVDPQFERLDPPKFSFDHVIEQLSRIDPNGRLGLVQKIKKNIFAVGSDRGNTTPGVAAFASLFLREHNRLAAEVENRNPAWDDGQIFQTARNINIVIYIRIIMEDYLNQLSHAPFDFKLDPGPWTWDAKWNKKFRVAVEFSTVYRWHSIIPNAIQIGEGDPLPIMQAMFNNELLIDRTLLKSFEDSSSQRATSFTPFNTADIMLPREYNTILQARQAGIRYYVDYLKQFGFPVDLPKNYSDITSDEKVQKMLEEMYGKGNVDKVEFYIGAICATHERNAPFSLFMNLQVAHDAIKAFYMNPLLQRSSWK

***Asterionellopsis glacialis* MMETSP0707-20130614_6459**

MILPTFANHLTGGFIKSTRGPDGSYQYGRTHSNHNIDLLQLYGRTQEQTLALRLRDNTTFGKKGRLKSQILNGGEEWPPFLYKDGRIDPQFEVLDPIAHGMSHLVDIMEGDRSKFFAMGSNQQNMLPQNIAMTTLFLREHNRVAGELERRYSDWDDDRIFETARNIVIVIYLKIVIEEYINHISPLPDIVKFLVDPGPWMWNAEWNKPNWISAEFAILYRWHALMPNKVMHAGIVFDLSSTLFQHDLLLGEDRSLKKTFVDMSAQRSASCECLNTAGPLVEREVESLRYSRALRFRPYVEYVEYWGLKKPRRFEDITKNAEVASMLRDLYGTVDKVEFFVGLIASDHSKNGIFGSSMNIGVGLDAFSQALTNPLLSEHVWK

***Aulacoseira subarctica* MMETSP1064-20121228_17645**

MLFPVFAQMLIDSFISTVTYTNTTTKKLTFDWKRTNSPFEINLLPLYGRFEYQTDALRLKSEKSRGRLKSQIINGEEYAEFLYNDLGEIKDEFQVLGPPQAFNQILSKVDKTEEEKRKIKSKIFAFGGNRTNTTPQMAALNTLFLREHNRLAGQLEKHNEDWDDDRVFQTARNINLVIYLKIIIEEYINHITSSGARFKVQPKEWVVWNADWNKPNWIAVEFAVLYRWHALTPNSYLWEGKRVKISDDLFNNALLLETSGGLRQAIAEISKNPATIMAPFNTALELLSQEQAAFAQTRQANVRPFADYRAYLGLPLVKTFKDITQDKEVQQKLEELYKTPDRVEFWAGLVAEDKDPKAIFGPTLSTLVALDAFSQALTHPLLSEQVFN

***Aulacoseira subarctica* MMETSP1064-20121228_57453**

MLFPVFAQYLIDSFILTKRIKTEGGTVSTIDWKRSDSPNDIVLLPLYGFKNATDALRLKSEEAQKRGKLKSQLIDGEEYAPFLYDESGMVKEEFKALGEPESLGEILKAMMADSNTVKKYKSKIFAFGLVRVNITPQLAAMNTLFLREHNRLAGVLEENNPEWDDERVFQTARNINLVIYLKIVIEEYINHISSGVQFKAKPEKWIWHASWNKPNWISVEFAVLYRWHTLVPNSDEWNGIRYDFQNEIFNNEILLDAGGLRHSFASISANRAPSICPFNTGGFLLPREASALQQSRVNKLRPFGEYREYLGYKKTKEFSDISSDPQVQETLRKLYKTPDRVEFYAGLIAEDHVAKGILAPIMTDLVAKDAFTQALTHPLLSENVFN

***Chaetoceros affine* MMETSP0090-20130426_9955**

MLFPTFAQHLIDSFIVTAVKSDGGSGTEFEWKKTDSPHDIGLLPLYGRTFDQTKQLRVQN

PPRGKYGQLKSQIIHGEEYAPYLYDADGKVKKEFDLLETPQGLERSLSMLSPEDAKAKKSNIFAF
GGARTNLVPNITAWNTLLLREHNRIAQTIEKEEPTWDDERVFQTARNVLLVIYLKLVVEEYINHI
TGYGIDFTVDPGKWMWNAPWYKRNWISAEFAVLYRWHGVIPSCMKWGDKTLSTHESLFNNAV
LTEDMKGSLRDTFINISNHRATQMNLFNTESMMVLRDMAALKQCRACKIKPYADYVVYLGTKE
RPTKFSDISKDKEVQEALEKVYKKVENVEFWTGLLASDNPPEGIMSPEMTTFVANDAFNQALCH
PLLSENVWS

Chaetoceros curvisetus **MMETSP0717-20131115_51459**

MLFPTFAQHLIDSFIVTAIKENSSEGVVFNWRKTGSPHDIGLLPLYGKTIEQTTQLRLKSE
VKGSKGRMKTQIIKGGEWSPYLFDSKGNKKAEFSSLPDPEGLEFALMEAKKFGVDAEKNKASVF
ALGGSRANLTPNIVAWNTLLLREHNSIAAKIEAENPEWDDERVFQTARNVNLAVYLRLVIEEYIN
HITAFGVDFTVEPGKWMWDAPWYKRNWISAEFAVLYRWHAVIPNCMKWGEKTLPTAAYLYSN
HLLLDDTGLKGDLREAFINISNHRATSMEIHNSEKWMTGRDSRALQMSRECELRSFTEYCAYLG
KPVPKTFADITSDVEVQNELKALYGKVENVEFWVGLIAKDHPTEAIMSAELTTFVANDAFNQAL
THPLLSEHVWP

Chaetoceros debilis **MMETSP0149-20130528_3320**

MLFPTFAQHLIDSFIDTVYHYDDDGNVVFDWKRTETPHDIGLLTLYGKTIPETKQLRKQS
ETSGEKGKLKSQLVNGEEWAPFLYDSNGKVKEEFNELPVPQGIDEKMYAARPQVQAKLKESIFA
FFGGSRTNLTPNISAWNTLLLREHNRIAGLIEEENPTWDDERVFQTARNCTLVIYLRLVIEEYINHI
TAYGVDFEIEPEKWMWDAPWYKRNWISAEFAVLYRWHAVIPSLMKWGKNTHTTMDYLFSNNL
LLSDDGMKGNLRDCFHNICDHRATNMQLHNSEGGFMVGRDKSALEMSRSCKLRSFSEYCGYLG
TPAPESFADITQDKDLQKELKDVYGEVKNVEFWTGLIAKDHSCEAIMSAELTKFVANDAFNQAL
THPLLSEHVFN

Chaetoceros debilis **MMETSP0150-20130528_3844**

MLFPTFAQHLIDSFIDTVYHYENDGNIVFDRKLTGTPHDIGLLTLYGKTIPETKQLRKQSD
TSGEKEKICPNLLKEKSXAPFLYDSNGKVKEEFNKLPVPQGIDGKMYAARPQVRAKLKESIFAFF
GGSRTNLTPNISAWNTLLLREHNRIAGLIEEDNPTWDDERVFQTARNCTLVIYLRLVIEEYINHITT
YGVDFKIEPEKWMWDAPWYKRNWISAEFAVLYRWHAVIPSLMKWGKNTHTTMEYLFSNNLLL
SDDGMKGNLRDCFHNICDHRATNMQLHNSEGGFMVGRDKSALEMSRSCELRSFSEYCGYLGSP
APESFADITQDKDLQKELKDVYGEVKNVEFWTGLIAKDHSCEAIMSAELTKFVANDAFNXGFDS
SSPFR

Odontella **MMETSP0015_2-20120614_17751**

LLFPTVAQHLIDSFINTKYTKTINGTFFDWARTDSPHEIGLSPLYGCNPKQTSQLRELNEK
IGCKGRLKTQILGDGGEEWAPFLYGDSGKIKAEFNELDEPVALDHILGFNSENAEMMRKSIFAFG
GSRANLNPNIVAWNTLLLREHNRVASKIEASEPSWDDERVFQTARNVVLVIYLKIVIEDYVAHIS
GVNFKVQPGEWMWNAKWYKRNWMSVEFAVLYRWHALIPNTTFWGSKTASTKDALYNNTLLL
NKEQGCAGNLRNIMVQISEQRITSFQLNNTEDWLVDRDLNALRQSRECNLASYADYCEYLGMD
RPKTFTDISLYPEVQEKLRSLYKTADKVEFYVGLIAADHSPGGKLFSEAMTRFVANDAFNQALA
NPLLSENVWK

***Cyclotella meneghiniana* MMETSP1057-20121228_4220**

LLFPTFAQHLIDSFINTRIDHEATEKNGGRPVFDWARTDSKHEIGLSPLYGDTEEQLNQL
RLLSNVNGFKGLMKTQIINEEEWAPFLYEVDGSKKKEFSAIWDPSGASYVLGMRRDVGERQKRT
LFAFGGARANLNPNIVAWNTLLLREHNRIAGEIEKSEPSWDDERVFQTARNVLVVIYLKLVIEEYI
AHIAGVNFKVDPGEWMWNAPWAKTNWMSTEFAILYRWHAVIPNTSSWGKAKNLKVLDTLFN
NDLLLDTKEGLSGNLRDAFVSISEERVTAHQLFNTEEWMVDRELAAIKQGRANKVASYADYAE
YVDLPRPKTFADISLYPEVQKALEEVYGTVDRVEFYVGLIAVDMGAGGKIFSPMTKFVANDAFN
QALTNPLLSMNVWN

***Cyclotella meneghiniana* MMETSP1057-20121228_12622**

LLFPTFAQHLIGSFINTKIDHEATEKNGGQVIFDWAKTYSRHDIGLSPLYGDTTEQTDQL
RLMSNVTGFKGRMKTQLINDEEWAPFLYRVDGTKKPEFNAIWDPAGASHVLKMRGEEGFKQK
RTLFAFGGARANLNPNIVAWNTLLLREHNRLAGEIEKSEPSWDDERVFQTARNVNIVIYLKLVIE
EYIAHISGADFKVDPGEWMWNADWNKANWMSVEFAILYRWHAIIPNTINWGTSNMKVSDILFN
NDLLVKETEGLDANLRDVFVQMSEQRATAHQLFNTEEWLLDRELAAVTQGRANKVASYADYV
EYLNLPRPKTFADISLYPEVQNALKEVYGTVARVELYVGLIAADMGAGGKIFSLMTKFVANDAF
NQALTNPLLSQNVWK

***Pseudo-nitzschia arenisensis* MMETSP0329-20121206_15499**

LLFPTVAQHLIDSFINTHVDRKATEAKGSPVFNWAKTDSPHEIGLSPLYGDSKEQTDQLR
ERSEEVGRKGRLKSQMIEGEEWSPFLYDSKGNKKTEFSKIPDPSGMSHILKHVNAATAKKASIFA
FGGARANLNPNIVAWNVLLLREHNRLASMIEESEPTWDDERVFQTARNVSIVIYLKIVIEEYIKHI
SGANFRVAPGKWLWNAQWNKTNWMSVEFAILYRWHAIIPNSSSWGPSKDVEVRESLFNNTLLL
DKTKGMGAKLADIFVQISNERTTSFELNNTEKWLVDREMAAIKQGRTNNVAPYADYCEYLGYE
RPKTFADINRDPKVQEQLKELYGTPDKVEFYVGLIAGEHPSGGKIFSKMTSFVANDAFNQALGNP
LLSQNVWE

***Thalassionema frauenfeldii* MMETSP0786-20121207_3776**

LLFPTFAQHLIDSFINTRIKADTDKNKPPEFEWTRTDSKHEIGISPLYGDEPAQTKQLRELS
ETSGRKGRLKTQMIEGEEWAPFLYNGNNKKPEFSDIPEPDGVRMIEKHGFGDRSTIFAFGGRRAN
LNPNIVAWNTLLLREHNRLAAEIEKSEPEWDDERVFQTARNVLTVIYSKIVVEEYVGHISGVPFE
VQPGPWMWNAEWNKTNWMSVEFAILYRWHALIPDTIRWGPKVDIGIMKQLFNNTLLLSKENG
MGANLKDCFTEISRNRVTSFELFNTEGSYMATREMQAIRQCRAANVAPFADYCEYLGDPRPKTF
EDISRKPEVQKVLKELYGTPDRVEFYVGLIAQDHSAGPKIFGDVMTKFVANDAFNQALANPLLS
QNVWE

***Skeletonema marinoi* MMETSP0320-20121206_19182**

LLFPTFAQHLIDSFINTKLNTETGQFEWDQTESKHEISLGPLYGDDVGQTNQLREKSEVT
GRRGRLKTQILDGGEEWAHFLYDETGTKKEEFNLIRDPDGMKHILKALYNSDPAAKSSIMQTIFA
FGGRRVNLNPNMVAWNTLLLREHNRLAGEIERSEPSWDDERVFQTARNVLIVIYLKLVIQEYIGH
ISGVKFKLDPGEWMWNAPWYKTNWMSTEFSILYRWHALIPNQNGLGPSKDLGVLKSLFNNPLL
LDEEKGLGGNLRDVFVDICKARITSFQLFNTEKWMVGREAATISQGRANNVQPYAKYCEYLRIK
PPKTFKDISLVPEVQQALEELYGTPDRVEFYVGLIAADHAPGKIFSLMTKFVANDAFNQALTNPL
LSQNVWE

***Skeletonema costatum* MMETSP0013_2-20120614_556**

LLFPTFAQHLIDSFINTKLNTETGKFEWDQTESKHEISLGPLYGDDVGQTNQLREKSEVT
GRRGRLKTQILDGGEEWAHFLYDETGTKKEEFNLIRDPDGMKHILKALYSSDPAAKSSIMQTIFA
FGGRRVNLNPNMVAWNTLLLREHNRLAGEIERSEPSWDDERVFQTARNVLIVIYLKLVIQEYIGH
ISGVKFKLDPGEWMWNAPWYKTNWMSTEFSILYRWHALIPNQNGLGPSKDLGVLKSLFNNPLL
LDKEKGLGGNLRDVFVDICKARITSFQLFNTEKWMVGREAATINQGRANNVQPFAKYCEYLRIK
PPKTFKDISLVPEVQQALEELYGTPDRVEFYVGLIAADHAPGKIFSLMTKFVANDAFNQALTNPL
LSQNVWE

***Detonula confervacea* MMETSP1058-20130122_27510**

LLFPTFAQHLIDSFINTKTKPLESGRLGFEWDQTFSKHEIGLAPLYGDYEEETLQLREKSE
ASGRKGRLKTQVLDGGEEWAPFLYDSEGTKKPEFSLLPDPDGLQHILAMQPETMEKKKKSIFAF
GGRRANLNPNIVAWNTLLLREHNRLASEIEKSEPSWDDERIFQTSRNVVIVMYCKIIIEEYIKHISG
VNLRVEPGEWMWNASWYKPNWISTEFAILYRWHALIPNTSSWGASKNIEVVDALFNNNLLISKD
TGMGGNLRDAFVEISKTRITSFELFNTEKAMVQRETAAIEQGRFNNVASYADYCEYLELKRPETF
EDISLKPEVQQALKELYGSPDRVEFYIGLIAADHPAGGKVFSAAMTKFVANDAFNQALTNPLLSQ
NAWN

***Skeletonema marinoi* MMETSP1039-20121108_4563**

LLFPTFAQHLIDSFINTKINVETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEKPG
RRGRLKTQVLEGGEEWAPFLYNEDGTKKKEFSAIHDPDGMKTILGLVYSSDPTTKSSIEQSIFAFG
GRRANLNPNIVAWNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCKIVIEEYIKHIS
GVNFKVEPGPWMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDAGVMEALFNNPML
LDDETGLGGNLRDIFVDISQTRVTSLQLFNTEKWMVERESAAINQGRANNVQSYAAYCEYLDIE
PPKTFEXISMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKIFSEMTKFVANDAFNQALTNP
LLSQNVWE

***Skeletonema marinoi* MMETSP0319-20121206_26582**

LLFPTFAQHLIDSFINTKINVETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEKPG
RRGRLKTQVLEGGEEWAPFLYNEDGTKKKEFSAIEDGTKKEEFSAIHDPDGMKTILGLVYSSDPT
TKSSIEQSIFAFGGRRANLNPNIVAWNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMY
CKIVIEEYIKHISGVNFKVEPGPWMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDAGV
MEALFNNPMLLDDETGLGGNLRDIFVDISQTRVTSLQLFNTEKWMVERESAAINQGRANNVQSY
AAYCEYLDIEPPKTFEDISMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKIFSEMTKFVAN
DAFNQALTNPLLSQNVWE

***Skeletonema marinoi* MMETSP0562-20121206_1237**

LLFPTFAQHLIDSFINTKINAETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEKPG
RRGRLKTQVREGGEEWAPFLYNKDGTKKEEFSAIHDPDGMKTILGLVYSSDPTTKSSIEQSIFAFG
GRRANLNPNIVAWNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCKIVIEEYIKHIS
GVNFKVEPGPWMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDAGVMEALFNNPML
LDDETGLGGNLRDIFVDISQTRVTSFQLFNTEKWMVYRESAAINQGRANNVQSYAAYCEYLDIE
PPKTFEDITMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKIFSEAMTKFVANDAFNQALT
NPLLSQNVWE

***Skeletonema marinoi* MMETSP1428-20130617_783**

LLFPTFAQHLIDSFINTKINAKTGEFEWDKNDSKHEIGIGPLYGDEVEQTNQLREKSEAPG
RRGRLKTQVLEGGEEWAPFLYNEDGTKKKEFSAIHDPDGMKTILGLVYSSDPTTKSSIEQSIFAFG
GRRANLNPNIVAWNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCKIVIEEYIKHIS
GVNFKVDPGPWMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDAGVMEALFNNPML
LDDKTGLGGNLRDIFVDISQTRVTSFQLFNTEKWMVYRESAAINQGRANNVQSYAAYCEYLDIE
PPKTFEDITMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKIFSEAMTKFVANDAFNQALT
NPLLSQNVWE

***Synedropsis recta* MMETSP1176-20130426_10672**

PTKPWLGQGCRGMALRFPKRPLPDNLPDAKTLVEDFCIRPDDDASFTPCANGVNSLTPY
FALCVIHDFFRSDTGRSKVGKLDRPWVNLHSSYLDLQTVYGYNKKSCASVRTQKHGKLTEKKIV
DHRLERMAVCHALILVMVKHHNFICDQLMERYPDKFKTDEMIFQTARLINCGVYINLIIELYSCL
FHVWNEDGSNPVELRGLDYPRDIQGYHLSYEFNIMYRFHAFIPKEWTPFYGLKEAIAYLKKGDFT
HSVPALSDDAIKTLLVDAMSHRAGANHVPSNVPRVMGPAEVKGIEDARLLGISTYNDFKEAVGE
PKYDTFLDMSGNRPELAAKLEKHYPTVDDVEFAVGMRVETRTPLRGAGFSTVGRAILADAFSSI
RFDRFYSQSEYH

***Nostoc* WP015113127**

MLFPYWVQWFTDSFLRLDHTNKLKNTSNHEIDLCNVYGLTRKQTHLLRSFQGGKFKTQ
KLKRQDGVEEEYPLFYYADPAQGIVDPQFAGLYEPVNDEKRQPADKKQYLFAMGVERANVQIG
YVMLNTLCLREHNRLCDVLASNYPDWDDERLFQTSRNILMAIILKIIMEEYINHITPYHFKLFADP
EAFTKESWYRTNHMAIEFDFVYRWHSAIPETFNCNGKPTHVADTLWNNKILIDQGLGALMEETC
SQAGTRIGLFNTPNILVDLAELPSIKLGRQLQLASYNDYRELCGFPRVTSFDQITGDEFTQQKLKE
FYGHVDNIEFFVGLYSEDVRQNSTIPPVARLIGIDAFSQALTNPLLSPKIFN

***Chaetoceros cf neogracile* MMETSP1336-20130426_1470**

LMLPIFAQWFTDSFLRTKFRVDGPQDFKENESNHEIDLCQIYGMTETQTAMLRSMENGF
LKSQIIDKEEFPAHLFEETASGLVQIKPEFAGLYTEANVQRVFRRASKEHKLNCFAVGIEHGNSTM
GNTLMNIIWMREHNRIAREIAKAHPTWDDERLFQTARNVNIVLLLKVVVEDYIFHISGFPFKMDN
KIAEGERWYRSNWMAAEFALLYRWHDLIPDTVEFGGERKDSSALNKNNRWLMKVGVDRACLD
ASQEPSGKIMLGNTPNFLVEVGKMSLKQGRVSGLASFNDYCEHYGLKRKKNFMDLTGNNKEVS
KKLEDVYGSIDNLEWFVGLWAEAYDDKTKCMGPLTYMVGNDAFTQALTNPLLAAEVFN

***Grammatophora oceanica* MMETSP0009_2-20130614_5553**

LLLPMFAQWFTDGFLRTKWKPPTEQDYVENESTHQIDLNQIYGATEVQTNILRAKKGG
KLRSQMIKGEEYPEYLFDTETLQIKPEFEGLYSDVNFTRVFQNVSDEQKKVTFAVGLEHGNSTVA
NTIMNTLFLREHNRLAGIIEAAHKDWDDERIFQTTRNCLIVILMKIVINDYVAHIADAPVFLDAGG
FAEDQPWYRENWMSVEFNLLYRWHDMIPDQVSFTSADQVSFSSEDEYWDMESSKTLVNNNTL
MLELGVEQILLDASKQRAGQLGLFNTPDFLLHVHKETMKQCRMINLAPYVDYCKEYGLTVPKT
FRELTGGNLEYAKALESVYDSVEHVEWFVGLWAGGKALGQFSSDLLILMVGHDAVTQLFTNPL
LSRRVYN

***Gersemia fruticosa* AAF93168**

WLFMFFAQHFTHEFFKTIYHSPAFTWGNHGVDVSHIYGQDMERQNKLRSFEDGKLKSQ
TINGEEWPPYLKDVDNVTMQYPPNTPEDQKFALGHPFYSMLPGLFMYASIWLREHNRVCTILRK
EHPHWVDERLYQTGKLIITGELIKIVIEDYVNHLANYNLKLTYNPELVFDHGYDYDNRIHVEFNH
MYHWHPFSPDEYNISGSTYSIQDFMYHPEIVVKHGMSSFVDSMSKGLCGQMSHHNHGAYTLDV
AVEVIKHQRELRMQSFNNYRKHFALEPYKSFEELTGDPKMSAELQEVYGDVNAVDLYVGFFLE
KGLTTSPFGITMIAFGAPYSLRGLLSNPVSSPTYWK

***Gersemia fruticosa* AAS48061**

WLFMFFAQHFTHEFFKTIYHSPAFTWGNHGVDVSHIYGQDIERQNKLRSFQDGKLKSQ
MINGEEWPPYLKDVDNVTMQYPPNTSEDQKFALGHPFYGMLPGLFMYASIWLREHNRVCTILR
KEHPHWEDERLYHTGKLIITGELIKIVIEDYVNHLANYNLKITYDPELVFDHGYDYDNRIHVEFN
HMYHWHPFSPDEYNISGSTYSIHEFMYHPEIVVKHGMSSFVDSMSKGLCGQMSHHNHGAYTLD
VAVEVIKHQRELRMQSFNNYREHFALEPYKSFEEMTGDPKMAAELQETYGDVNAVDLYVGFFL
EKGLTTSPFGITMIAFGAPYSLRGLLSNPVSSPTYWK

***Plexaura homomalla* AAU87497**

WLFMFFAQHFTHQFFKTVHHSPAFSWGNHGVDVSHIYGQGVERENKLRAFKDGKLKS
QMINGEEYPPYLKDVDDLKMQYLENTAEEQKFALGHPFFSMLPGLFMYATIWLREHNRVCMIL
RKEHPHWEDERIYQTGKLIITGELIKIVIEDYVNHLANYNMKLRYDPQLVFSRNYDYDNRIHLEF
NHLYHWHPFSPDQFNISGTTYAIKDFMYHPEIVVKHGMSSFVNAMSSGLCGKMSHHNHGQYTL
DVAVEVIKYQRELRMQSFNXYRRHFGLHAYKSFEEMTGNPKMAAELKEVYGDVNAVDFYVGF
FLEKSLTTSPFGITMIASGAPYSLRGLLSNPVSSPTYWK

***Plexaura homomalla* AAF93169**

WLFMFFAQHFTHQFFKTVHHSPAFSWGNHGVDVSHIYGQGVERENKLRAFKDGKLKS
QMINGEEYPPYLKDVDGLKMQYLENTAEEQKFALGHPFFSMLPGLFMFATLWLREHNRVCMIL
RKEHPHWEDERIYQTAKLIITGETIKIVIEDYINHLANYNMKLRYDPQLVFSRNYDYDNRIHLEFN
HLYHWHPFSPDQFNISGTTYTINDFMYHPEIVVKHGMSSFVNAMSSGLCGKMSHHNHGQYTLD
VAVEVIKYQRKLRMQSFNNYRRHFGLPAYKSFEEMTGDPKLAAELKEVYGDVNAVDFYVGFFL
EKSLPTSPFGITMIASGAPYSLRGLLSNPVSSPTYWK

***Gammarus sp* ADB65786**

LLFQYYAQHFTHQFFRTNYTKGPQFTKGNGGVDVSNIYGLTERQRRALRSNVDGKLKF
QIINGEHFPPYLKDVPGISMEYPPHLPITEDNKFALGHPFFALLPGLFVYSTIWMREHNRVCEVLK
EQHPHWDDERLYHTAKLIITGEVIKITIEDYVQHLSQYKVDLKFKPQVVHGTRFQFHNRINVEFD
HLYHWHPLIPEGIKVEDSYYSLMDMAFSTKSVFTHGLDAFVKALVTNRAGKLTSRNHSPVTVPV
LKKMLENSRILRFQGVNQYRKKFNMRPFRDFLDLTGDEELARDMEEMYGDINAVEYYVGLIAE
KDSPSLTPLTMVNVGGPWSVKGLIANPICSPHWWK

***Caprella sp.* ADB65785**

LLFQYYAQHFTHQFFRTNYTMCPQFTKGNGGVDVSNIYGLTEQHRRAIRMNSDGKLKY
QVINDEHYPPYLRDVQGIEMDYPPHIPITEDNKFALGHPFFALLPGLFVFSTIWMREHNRVCDVL
KNQHPDWDDERLYQTAKLIITGEVIKITIEDYVQHLSQYKVDLKFKPQVVHGTRFQFDNRINAEF

NHLYHWHPLIPDGIQVEDKYYSLMDMAFSTKSVFTHGLDKFIESMATSRAGKLSHSNHPLVTLP
VLKKMMENGRKLRYQGINEYRKRFALKPFKDFMDLTGDEALAKDLQELYGHVDAVEFYVGLL
TEKDSPSLTPLTMVNVGGPWSVKGLIANPICSPHWWK

Nitzschia sp. **MMETSP0014_2-20120614_6887**

PQLNFWMLSFVNWFHDDNFRTLPDTDGAFTWSDRGSLHMTHLYGHTEYRQAALRTM
AGDGKMKTSSRLGWDYYPPLLMDVQADFPDFDMWTSQRGSKHKSTSAGQTSEQADENMPYYF
AIGDPRFNLHLGHILWTSVGLYLHNTACVILQREDPGLTDEDIFQRARVIVFHIIQKIRLQDFVMD
SISSTRDHIRIPYDPKLLREEFAHHFAYSGGNQPNFLEFNHLYQAWHALIPNGLVLNEDVDDGEK
DILPIRKTLWAPKLMTTNFTIGEMATSFASTPLTLYSPHNFPVFLRGVTEAALKDERAQRMAPYN
SYRELIGLDPITSFEQFAVDDPQKMAELYNNDVDSVDFIAGILADSNPHLPGNFFGVQLVLVALF
ALQDLANNPLILDPVMS

Skeletonema marinoi **MMETSP0319-20121206_1511**

EADVALVANIMKRDTNSFAPFNQLASAWIQFMTHDWFQHDASSSQGLKMKNVVTHW
WDASQMYGSSQEEVDAVRAEGGKLHLDVNDEIDYNASLPITGFRENWWAGLHILHTIFAREHN
HIVDILAQSYPSMTEDELFGTARNIIAVLLAKIHTVEWTPTLLDNAVSDMALNINWHGLQTVTSM
YFKGKNIPDEVKDIIDEMKVPSVFGSNYTTEQTLFNTPFYMTEEFVSVYRMHTLLPDAMILEGGK
TVSLQELAFTDARNLVSDPSKTTATLLQAFAHTPAQALSLKNYPKSLFNLQIGNGKSINLAEIDIS
RDRARGIPRYNDARRQLLLTPYKSMDDLTSDKEELKLLKSVYTDIEQVDFLVGCLVDKDRPDGF
AFGIVPYYIFVVMASRRLLSDRFFQEGLTEENYS

Pseudo-nitzschia pungens **MMETSP1061-20121228_74475**

DNVQVIAQRLLAREGFKPAGNQLNIIAAAWIQAQVHSWIQHLDGAPTSIEATAEAVGPV
CPVKKMNFFETEERPDGEYNSFRTQWWDASFVYGQNREQVHLGRLYKDGKLKVNESNPDTLSF
VEEGKSKIDVVGDQSNSWVGVTVLQVLFIKEHNYCAEMIKKENPQLTDEEIYGHCRNIISALVAK
IHTIDWTVELLKTEQLRVAMEINWKGATKAVFGDKAPFHPLRLINKPKADNKGVPFCLTEEFAA
VYRLHPLLPPGLVVEHGEGKEEFIGIEKLLTTKGRDKMREPGMAKKIMFSAFHYSCGHLKSSNYP
FIMRKFTPTDHKGVDLQPPEDRVVDMAAIDLHRDRERGIQKYNEFRRQLKLRPFKTWEALTGEE
NKSDNELTDAKKLELIYGPAPEGIEKIDLLVGDLYERKIADGFALSETSFMIFLLMASRRIDSDPYL
NEYYNEEYYT

Thalassiosira rotula **TR754|c0_g1_i1|m.3431**

MLFPTFAQHLIDSFIVTAIKENTSEGVEFDWRKTGSPHDIGLLPLYGKTIEQTTQLRLKSE
LKGSKGKLKTQIIKGGEWSPYLFDAKGNKKPEFSSLPDPEGLEFALGMAKQHGVDVKKNKATVF
ALGGSRANLTPNIVAWNTLLLREHNRIASKIEEENPKWDDERVFQTARNVNLAVYLRLVIEEYIN
HITGFGVDFTVEPGKWMWDAPWYKRNWISAEFAVLYRWHAVIPNCMKWGEKTLPTASYLYSN
HLLLDEDGMKGDLREAFINISDHRATSMQIHNTESWMKGRDARALQMSRACELRSFTEYCAYL
GAKVPKTFADITSDVEVQNELKALYGKVENIEFWVGLIAKDHPTESIMSAELTSFVANDAFNQAL
THPLLSEHVWAAGPE

Skeletonema marinoi **MMETSP0920-20130426_33289**

LLFPTFAQHLIDSFINTKINVETGEFEWDRTDSKHEIGIGPLYGDEVEQTSQLREKSEKPG
RRGRLKTQVLEGDEEWAPFLYNEDGTKKEEFSAIHDPDGMKTILGLVYSSDPTTKSSIEQSIFAFG

GRRANLNPNIVAWNTLLLREHNRLAGEIEKSEPGWDDERVFQTARNVLIVMYCKIVIEEYIKHIS
GVNFKVDPGPWMWNAPWYKTNWMSTEFAILYRWHALIPNEAGLGPSKDAGVMESLFNNPVLL
DDETGLGGNLRDIFVDISQTRITSFQLFNTEKWMVERESAAINQGRANNVQSYAAYCEYLDIEPP
KTFEDISMVPERQQALKELYGTPDRVEFYVGLIAADHPAGGKIFSEAMTKFVANDAFNQALTNP
LLSQNVWE

***Fragilariopsis kerguelensis* MMETSP0907-20130614_13332**

PNVQVVAQRLLARESFTPAGDQLNIVAAGWIQAMVHDWMKHEDGKKTSMEVTPAVV
GSQCPLHRFNLFETKERPDGHYNSERTNWWDASFVYGQNAEQVQNSRAFVGGKLKVNEKNPD
TLPSRDDGTDLTGDQSNSWVGVSVLQTLFLYEHNYCAEQIAKENPNLTDHQIYGHCRNIIAALV
AKIHTIDWTVELLKTPQLKIGMRTNWMGIIQAITGLKIPFLDRLLRLIKKKENNNEGVPFCLTEEF
AAVYRLHPLLPPGLIIEGEGAGDKDEDEDTFIDLRDTLTTKGRDLMRKSGMAKKVMKAVFTYPC
GNMAPSNYPDVMRDFHPTDLLGNNLDDRIDLAAIDLFRDRERGIQYFNNFRRKLSMKPFQTWEE
LTGDDKMTEEALAAFVAGTGNLTNAKKLELVYGAAPKGIEHCDLLVGDLYEKKIPGFAISETSF
MIFLLMASRRLDADPYLNEYFDEEHYT

***Skeletonema grethea* MMETSP0578-20130828_1168**

ILVAYSIXSFQYVLDFKPAASQLNVLAASWIQAMVHDWIGHFDGDETETLDRGGESLCP
FAKSPFSFKNTKTEKIDGVPFSPSERTNWWDASFVYGNNSEQIDRARTMQGGKMVTSDIPHALA
EDKDGVYFAGDNKNSWVGVALLQDLFIREHNYICDQIAAEAKEEGKEMTDEELFGKARVVVAA
LVAKIHTVDWTVELLKTKLLAIGMKTNWDGLLKAVGIPIPGILSQMGEKKGRVSDNEGTPFCLT
EEFAAVYRLHSLSPPGLILGDGDAKDKFIGLEDLLGDEGRKQMRETKTRPKEMMKSCLHWPCG
ALMSSNYPNAFRDVAPTDDYGKDLKSQNIDLAALDLFRDRERGILKFNEFRRQLNLKPYRTWLE
LTENEEDARKLELIYGPGQEGIERCDLLVGDMYERKVQPSFALSETSFIIFLLMASRRLSADPFLNE
LYNEETYS

***Staurosira sp* MMETSP1361-20130828_23824**

LLFPFFAQWFVDSFLRTKWKPLAEQDFKENESNHDIDLNQIYGTSEIQTDMLRSMKGGR
LKSQIIDGEEYPVFLFDQKTATLKPEFRGLYTEENFKRVFGNASKEHKLHSFAVGLEHGNSTIGNT
VMNTLFLREHNRVAGVISAAHPEWDDERVFQTTRNVMIVLLIKIVLADYIYHISGAAVFADPGGF
AEDELWYRENWMSVEFSLLYRWHDLIPSSVTFDGETRDAVDLQNNNRWLLKAGMDSVIQDAS
NQKAGVMGLGNTPDFLLSVTKMSLHMARTCKLSSYVEYCKEYGQDPPEDFMDLTGDQDSASK
LERVYGSIDKVEWFVGLFAQKRDVFSGLLMTLMVGNDAVTQAFTNPLLAKRVYN

***Skeletonema marinoi* MMETSP1428-20130617_31509**

EADVALVANMMKRDTNSFAPFNQLASAWIQFMTHDWFQHDASSSQGLKMQNVVTHW
WDASQMYGSSQEEVDAVRAEGGKLHLDVNDEIDYNASLPITGFRENWWAGLHILHTVFAREHN
HIVDILAQSYPSMTEDELFGTARNIIAALLAKIHTVEWTPKNIPDEVKDIIDEMKVPSVFGSNYTTE
QTLFNTPFYMTEEFVSVYRMHHLLPDEMILEGGKLQELAFTDARNLVSDPSKTTATLLQAFAHTP
AQALSLKNYPKSLFNLQIGNGKLINLAEIDISRDRARGIPRYNDARRQLLLTPYKSMDDLTSDKEE
LKLLKSVYADIEQVDFLVGCLVDKDRPDGFAFGIYYIFVVMASRRLLSDRFFQ

***Homo sapiens* NP000953**

LMFAFFAQHFTHQFFKTSGKMGPGFTKALGHGVDLGHIYGDNLERQYQLRLFKDGKLK

YQVLDGEMYPPSVEEAPVLMHYPRGIPPQSQMAVGQEVFGLLPGLMLYATLWLREHNRVCDLL
KAEHPTWGDEQLFQTTRLILIGETIKIVIEEYVQQLSGYFLQLKFDPELLFGVQFQYRNRIAMEFN
HLYHWHPLMPDSFKVGSQEYSYEQFLFNTSMLVDYGVEALVDAFSRQIAGRIGGGRNMDHHIL
HVAVDVIRESREMRLQPFNEYRKRFGMKPYTSFQELVGEKEMAAELEELYGDIDALEFYPGLLL
EKCHPNSIFGESMIEIGAPFSLKGLLGNPICSPEYWK

***Ovis aries* NP001009476**

LMFAFFAQHFTHQFFKTSGKMGPGFTKALGHGVDLGHIYGDNLERQYQLRLFKDGKLK
YQMLNGEVYPPSVEEAPVLMHYPRGIPPQSQMAVGQEVFGLLPGLMLYATIWLREHNRVCDLL
KAEHPTWGDEQLFQTARLILIGETIKIVIEEYVQQLSGYFLQLKFDPELLFGAQFQYRNRIAMEFN
QLYHWHPLMPDSFRVGPQDYSYEQFLFNTSMLVDYGVEALVDAFSRQPAGRIGGGRNIDHHILH
VAVDVIKESRVLRLQPFNEYRKRFGMKPYTSFQELTGEKEMAAELEELYGDIDALEFYPGLLLEK
CHPNSIFGESMIEMGAPFSLKGLLGNPICSPEYWK

***Homo sapiens* NP000954**

MMFAFFAQHFTHQFFKTDHKRGPAFTNGLGHGVDLNHIYGETLARQRKLRLFKDGKM
KYQIIDGEMYPPTVKDTQAEMIYPPQVPEHLRFAVGQEVFGLVPGLMMYATIWLREHNRVCDV
LKQEHPEWGDEQLFQTSRLILIGETIKIVIEDYVQHLSGYHFKLKFDPELLFNKQFQYQNRIAAEF
NTLYHWHPLLPDTFQIHDQKYNYQQFIYNNSILLEHGITQFVESFTRQIAGRVAGGRNVPPAVQK
VSQASIDQSRQMKYQSFNEYRKRFMLKPYESFEELTGEKEMSAELEALYGDIDAVELYPALLVE
KPRPDAIFGETMVEVGAPFSLKGLMGNVICSPAYWK

***Coccotylus truncatus* AFN20596**

VLIAYYAQWVTHQFFNTDESDPTGHSVKQPVGVNMSMLYGSKQEVEKSVRAYKGGLL
KSTIKNGQEFPEIMPCQEGSRIPGKEMFNMPILIANMIPGFAAIHVLFFRRHQYICRELAKWAEAQ
GKNIDDEELFQKAKLIVTVNMLRITMHDYVSRALQSSHAKMRFDQKVKQSRIWKMFGPDYFPPS
NAIQFEFNIFYRWHQFYPDTTKIMKRIDDLKFPKSKQQLDEKWNAVRWIADEPDGMERVLFSAS
SQRAGKLSLLNTNQWIVEHVVKPGLARCREHQLASYNDYREKVGFPRLTTFEQVTSNPALLEKL
KRVYRNVDQIEYYPGVFAEDKHFGNVHGPLTFGSSMTFTGIFSSRLFETALDE

***Ciona intestinalis* XP002127674**

VLFPFFAQHFTHMFFKTDPMKGMPYQWGDQLVDLSQIYGHGEKRQHELRSHVNGKLK
VSLVDGHEFPPLSNQTTANMSNINLLPQEYQFVFGHQGFSLMPTFLIWSTIWLREHNRICDLIKEE
NPAWDDERIFQTARLVLTGETIKVVIEDYVQHLSGFHYKLLYDPELVQGGSHSFHNQIHVEFQLL
YHWHALMPDQIEFNGKSYTMKRLLFNPEPVVKGGLKRTIEDLSNQWAGQVAGGKTQGAATLH
VAGLAIKNGRDLRMQSFNAYKEKFEMKKYTTFQELTGEEEMAAELQKLYGDIDAVEYYIGIML
EKRRSPQLFGETLTEMGSPYSLKGLYSNPINHKDWWK

***Ciona intestinalis* XP002123273**

VFFAFFAQHFTHQFFKTNTIKGMPFQWGEHSVDLSHVYGHTIQRQHELRSHIDGKLKVF
ETNGEVFPPLTESANVTMSGEKLMRGRKFAIGHPGFGAFPSFFVIATLWLREHNRVCDILKDLHP
DWDDERLFQTARLILTGETLKIIVEDYVQHVSGFHFQLSYDPEILHKSTFSYNNQIHAEFHILYHW
HMLMPDFIELGEHVYPLKELLFNVDPVVEIGMETVLKQLSNQFAGKVVGGRNQGPELVAVVEL

ALKQTRQMRMCSFNKYRERFGMKPYTSFEELTGETEVAALLRNLYYDIDALELFVGYFVEHRR
NRQVLGATMLEMGAPYSLKGVFGNPIGSPAWWK

***Oncorhynchus mykiss* CAC10360.1**

LMFAFFAQHFTHQFFKTRNSMGLGFTRALGHGVDAGNVYGDNLVRQLNLRLLKDGKM
KYQVVKGEVYPPTVAEAAVNMRYPQETPVGQRMAIGQEVFGLLPGLTMYATLWLREHNRVCD
ILKAEHPTWGDEQLFQTARLIVIGETIRIVIEEYVQHLSGYLLDLKFDPVLLFKSTFQYRNRIAVEF
KQLYHWHPLMPDSFHIDGDEVPYSQFIFNTSIVTHYGVEKLVDAFSRQCAGQIGGGRNIHPVVTN
VAEGVIEESRTLRLQPFNEYRKRFNLKPYTSFSDFTGEEEMARELEELYGDIDALEFYPAIMLEKT
RPNAIFGESMVEMGAPFSLKGLLGNPICSPEYWK

***Fistulifera solaris* GAX23950.1**

LPPVKDVAAILRRPMNPTSVAPFNQIAVAWIQMMTHDWFQHDPAHPDQKMNRVTHW
WDASQLYGSTLAQQTAVRVPNTGKVRLDQHQELNYTTTGIPITGFADNWWAGLHMMHTLFVR
EHNWLVDQFERQYPGVYTANDKFQLARLCLSALLAKIHTVEWTPTLLDNPVAALGLHTNWRG
VDAILEYGTRFELQLAYRIVGGDQSVPHAGNGTTRETLYNTTFAMTEEFVAVYRMHPLLPDEME
IEGKTFSLNDLSFVDARTLTKSVKTTQTLLQAFGMTPANTLSLQNYPRQLYGLEKPGMSQPVNL
AEIDLQRDRERNLPRYNDMRRQLLLKPYKRLEDLTDDETELNLLKSVYQDMDQVDLMVGCLV
DKDRPYGFAFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT

***Cyanothece sp* Heme peroxidase ACB53655.1**

PNPRVVSRTLMTREDFKPATILNLLAAAWIQFENHDWFSHGDNKPEDKLEIPLEANDPW
PEEYRPLEVGKTLPDTSRPEGAKPPTFINTVTHWWDGSQIYGSNPETVDQLRSHEDGKLIIGENGL
LPVDPETGVDITGFNDNWWIGLGMLHTLFTREHNLICDHLKQEYSQWSDDDLFDHARLINAAL
MAKIHTVDWTPAILPLPATDIALNVNWNGFLGEDIKQVLGTVGEGEIADLLTGIIGSDKNHHTAP
YYLTEEFVSVYRMHPLIPDELEFRSLEGDKFLQEVNFFEMSGKRTRALLESISLPDLFYSFGITHPG
AITLHNYPRFLQQLVRDNGEVFDLAAVDILRDRERGVPRYNRFREIMGRGRVKCFEEISSNKQW
VEEMRRVYNDNIDQVDLMVGLFAEDTPEGFGFSDTAFRVFILMASRRLKSDRFFTTDYRAEIYT

***Fistulifera solaris* GAX18717.1**

LPPVKDVAAILRRPENPSTVAPFNQIAVAWIQMMTHDWFQHDPTNPDSKMNRVTHWW
DASQLYGSTLTQQRAVRVPNTGKLRLDKHQELNYTSTGIPITGFADNWWAGLHMMHTLFVREH
NWLADQFELQYPGVYTANDKFQLTRLCLSALLAKIHTIEWTPTLLDNPVAALGLHTNWRGVDAI
LEYGTQFELQLAYRIVGGDQSVPHAGNGTTRETMYNTTFAMTEEFVAVYRMHPLLPDDMEIEGI
TLTLNDLSFVDARKLTKSVKTTQTLLQAFGTTPANTLSLQNYPRQLYGLEKPGMSQPINLAEIDL
QRDRERNLPRYNDMRRQLLLKPYKRLEDLTDDETELNLLKSVYQDIDQVDLMVGCLVDKDRPY
GFAFGIVPFHVFLVMASRRILNDRFFMEDFNAKVYT

***Alcanivorax nanhaiticus* WP_035233415.1**

PNPREVSNLIMSRGGDFKPATTLNFIATSWIQFMVHDWFDHGPRTDANPIEFPLPAGDVL
GGGTMSVQRTRPDPDVSGDEGIITYENINTHWWDGSQLYGSSKEKNDEVRSFVDGKLKVDGDG
RLPTEFFSGKPVTGFNENWWVGLSMLHHIFTQEHNAIADMLVANYPGQSDQWYFDKARLINSA
LMAKIHTVEWTPAILANPVLERAMYANWWGLGGDRDKRDKYQDDLDNLNNNLGQIGGLLDL
VGIDNGLGDSPTGSLEHALAGLVGSRTPNNYNVPYTLTEEFVSVYRMHPLLRDEIKVYDIGSNVV

DEEIALEDTRNGDAEDLLGDIGQDRLWYSFGITHPGALTLNNYPDFLRNLSMPLIGDIDMAAIDIL
RDRERGVPRYNEFRRQIGLKPLTSFEQLTSDPQLLADLKALYNNDIELVDTLVGQLGEETRPEGF
GFGETSFQIFILNASRRLMTDRFFTTDYTDEVYT

***Hordeum vulgare* subsp. vulgare BAJ90503.1**

PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMVHDWMDHMEDTKQIEITAPKEVA
NECPLKSFKFYATKEQPTNSDGIKTGYHNIRTAWWDGSAVYGNNEKQEKKIRTYADGKLVIGD
DGLLLHEENGVPLSGDVRNGWVGISILQALFVKEHNAVCDAIKEEHPNLSDEELYRYAKLVTSA
VIAKIHTIDWTVELLKTKTLRAGMRANWYGLLGKKIKDTFGHIGGTALGGLVGLKKPINHGVPY
SLTEEFTSVYRMHPLIPSTLKLRDPTGQPAADNSPPYLEDIDIGELVGLKGEDQLSKIGFEKQTLSM
GYQACGALELWNYPSFFRDLIPQNLDGTNRSDRIDLAALEVYRDRERSVPRYNEFRRRLFLIPIKC
WEDLTSDNDAIEAIRAIYGDDVEKLDLLVGLLAEKKIKGFAISETAFNIFILMASRRLEADRFLTSN
FNEKTYT

***Triticum aestivum* CDM84254.1**

PDPFVVATKLLARREYKDTGKQFNILAAAWIQFMVHDWMDHMEDTKQIEITAPKEVA
NECPLKSFKFYATKEQPTNSDGIKTGYNNVRTAWWDGSAVYGNNEKQAEKTRTYVDGKLVIGD
DGLLLHEENGVPLSGDVRNGWVGVSILQALFVKEHNAVCDAIKEDHPNLSDEELYRYAKLVTS
AVIAKIHTIDWTVELLKTKTMRAAMRANWYGLLGKKFKDTFGHIGGTALGGLVGLKKPINHGV
PYSLTEEFTSVYRMHSLIPTTLKLRDPTGQPAANNSPPYLEDIDIGELVGLKGEDQLSKIGFEKQTL
SMGYQACGALELWNYPSFFRDLIPQNLDGTNRSDRIDLAALEVYRDRERSVPRYNEFRRRLFLIPI
KSWEDLTSDKDAIESIRAIYGDDVEKLDLLVGLMAEKKIKGFAISETAFNIFILMASRRLEADRFIT
SNFNEKTYT

***Rhizophagus clarus* GBC07128.1**

LELEHYITALTQLPLNTPEEKKLLSILERLLVTQLWSDISKPPAMIAGDIYRSSDGSGYNR
LIQSLGKANSRYSLTSRIQYPIKLSVLPKSEDIFDKIMVQGGDFVEHPSGISSMLFYLAIIITHDLFHT
SFADPNINLTSSYLDLTPLYGSNDQEQKSIRTLKGGLLKPDTFADSRILLQPPGVSALVILFSRNHN
FIAQTLLEKNELGRFSVTNPNDPEQLKKQDEDLFQTARLINCGFYINVILHNYLRTILGLDRTNSK
WFVDPTVPYNKRGQLEPLPSGIGNIVSLEFNYIYRWHAATSKDDSKFVEDEFKTIFGDDWENITID
EFKEKMGVWGRSIPKDPSKWKFNHIERGSDNRFKDTDIAKEIINGTKKVSGAFGANRIPKVFRPIE
LLGIESARVLGLSSLNDFRRSLNLKPYESFMEMNPDPLIAKKLEELYGSIENVELYPGLMTEKTKP
DMLGSAIALPFTISRAILSDAVNLVRNDRYYTNDFSPRNLT

***Emiliania huxleyi* XP_005780718.1**

PFQALHNLPSADEVVRLLYKRDAFKKAPYGVNSLATWFANVAIHDFFRTATGTDGGTH
PERGSDKEWVNLHSSYLDLQPLYGYSKTTADATREWSGGKLKAFAEDRMRRIPESRVIVELLRR
EHNYVAEQLAQRYPAQFATDEELYQQARLIMGGVYINIILRAYGCQMFGEIAPDGSGFCELRQG
YGGAGVGNMCTFNFNLIYRFHTSIPVEWSATDPPPIDTDEQMRTLLNGILNWESGGFGPNNVPDS
ILGERARVSQRAIEAARLMGAPTLNTFRRRFTSGYSSFEDMTGGDQATADTLRQLYPGGIEDVEL
VVGCQVEKCMSGGWALPSTIGQAIVADAFASIRQDRFYTQDWGASSYT

***Rhodococcus gordoniae* WP_064063209.1**

VWFAFFAQWFTDSFLRTNSKDPRKNDSNHEIDLCQIYGVNQAKTTMLRAGYGGRLDSQ

VIDKKEYPPFLFAARTPGEELRFVPKFEGLHDREYLLDTVLRLCPDERKKSVFAVGLEHGNSTIG
NTVLNVLFLREHNRIAGILEGAYPEWDDDRLFETTRNIMIVILLKIVIEEYIKHIGPFDFPIEFVPFM
ADNAPWNRTNWCAIEFNLLYRWHSLIPDTVVFDSQRVSTRILVDNNPLVLDRGIESIIDQCSRQK
ASRIGLGNTPAFLIDRHPMCPDRESVEERTVGLMRQARLRSFNDYREAFGLGRLTSFTELTGDVE
VQQKLARLYGDVDAVEWYVGIFAEDYPRHRMMGELLTTMVAHDAFTQAFTNPLLARHVYHED
TF

Herbidospora mongoliensis **WP_066371138.1**

VMFMFFAQWFTDSFLRTSRDDFRRNTSTQEIDLCQIYGLTEEKTHLLRAHRDGRLKSQLI
DGEEYPEFLFRPRSSGEPPVFKPEFEGLHDSEFIVSRLLEDAPEKQKDTFFAVGLEHGNSTIGNTIM
NVVFLREHNRIAGLLKQAHPEWAERPEGADARLFETTRNIMIVLLLKLVVEEYIRHISPYDFPLET
VPFIADGKRWNRSNWISIEFNLLYRWHSLVPSTIGSGPDRLDSTDFRNNNPLVLARGIESLVSQCS
GERAGRIGLMNTPGFLVDRDPAHPERPSVQERTIALSRKARLASYNDYRENFGLGRLKDFGRLT
QDAELRERLEKLYDDIDKLEWYVGIFAEDYSRDEMMGRLMTTMVAYDAFTQALTNPLLARDV
YNEATF

Eutrema halophilum **BAJ34623.1**

MVVATKLLTRRKMIDTGKQFNMIAASWIQFMIHDWVDHLEDTDQIELSAPKEAAKGCP
LSSFRFFKTKEVPTGFFEIKTGSLNTRTPWWDSSVIYGSNSKTLERVRTYKDGKLKISEETGLLLH
DDDGLAISGDIRNSWVGVSALQALFIKEHNAVCDLLKKEYEDLEDEDLYRHARLVTSAVIAKIHT
IDWTVELLKTDTLLAGMRANWYGLLGKKFKDTFGHVGSSIFGGVVGMKKPQNHGVPYSLTEEF
TSVYRMHSLLPDQLHMRDIDVTPGPNKSLPLTQEVSMEKLIGREGEETMSQIGFTKLMVSMGHQ
ACGALELMNYPAWFRDLVPQDPNGHDRPDHIDLAALEIYRDRERNVARYNDFRRAMFMIPIKT
WEDLTDDKEAIELLDDVYGGDVDELDLLVGLMAEKKIKGFAISETAFNIFLLMATRRLEADRFFT
SDFNEMTYT

Pisum sativum **CAH05011.1**

PDPMVVVTKLLERKTYKDTGTQFNVIAASWIQFMIHDWIDHMEDTKQVELSAPSEVAS
QCPLKSFKFFKTKEIPTGFYDIKTGHANVRTPWWDGSVVYGSNEQVLNKVRTFKDGKLKISKEG
HLLHNEDGTAISGDIRNSWAGVTTLQTLFVQEHNAVCDALKKENSDLEDEDLYRHARLVTSAVI
AKIHTIDWTVELLKTDTLLAGMRANWYGLLGKQFKDRFGHVGNSILSGFVGMKRSENHGVPYS
LTEEFATVYRMHPLLPDSLHLRDISASPGPNKSPPLIKEIPMNDLIGLQGEKTLLEIGNAKKLVSM
GHQACGALELWNYPSWLRNLVPHNIDGTERSDHVDLAALEVYRDRERNVARYNQFRRGLLLIPI
SKWEDLTDDEEAIKVLEEVYGDDVEELDVLVGLMAEKKIKGFAISETAFVIFLLMASRRLEADRF
FTSNFNEETYT

Reference:

1. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).

PART 3

Supplementary Information

Chapter 3

Reconstruction of the phytosterols biosynthetic pathway and evaluation of sterol content in *Thalassiosira rotula*

Manuscript in preparation for publication

Di Costanzo, F., Di Dato, V., Cutignano, A., Romano, G.

Supplementary Information

Supplementary Figures.

Supplementary Figure 3.1: *Thalassiosira rotula* growth curves in the different experimental conditions tested.

Supplementary Figure 3.2: TLC of the fractionated lipids extracts.

Supplementary Figure 3.3: GC-MS calibration curves, obtained using cholesterol as internal standard, for the quantization of the 24-methylenecholesterol (24-MC) in the different experimental conditions tested.

Supplementary Figure 3.4: Complete images of the GC-MS chromatograms of the *T. rotula* sterol fractions from cultures grown in different nutritional conditions.

Supplementary Figure 3.5: Confirmation of the absence of gDNA contamination in the cDNA extracted from the *T. rotula* cultures grown in the different experimental conditions tested.

Supplementary Figure 3.6: Analysis of the best reference gene using the RefFinder tool: **a)** comprehensive gene stability; **b)** gene stability by Delta Ct method; **c)** gene stability by NormFinder; **d)** gene stability by BestKeeper; **e)** gene stability by Genorm.

Supplementary Figure 3.7: qPCR standard curve of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR, SiDepl and NDepl samples mix).

Supplementary Figure 3.8: qPCR standard curve of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR and PDepl samples mix).

Supplementary Tables.

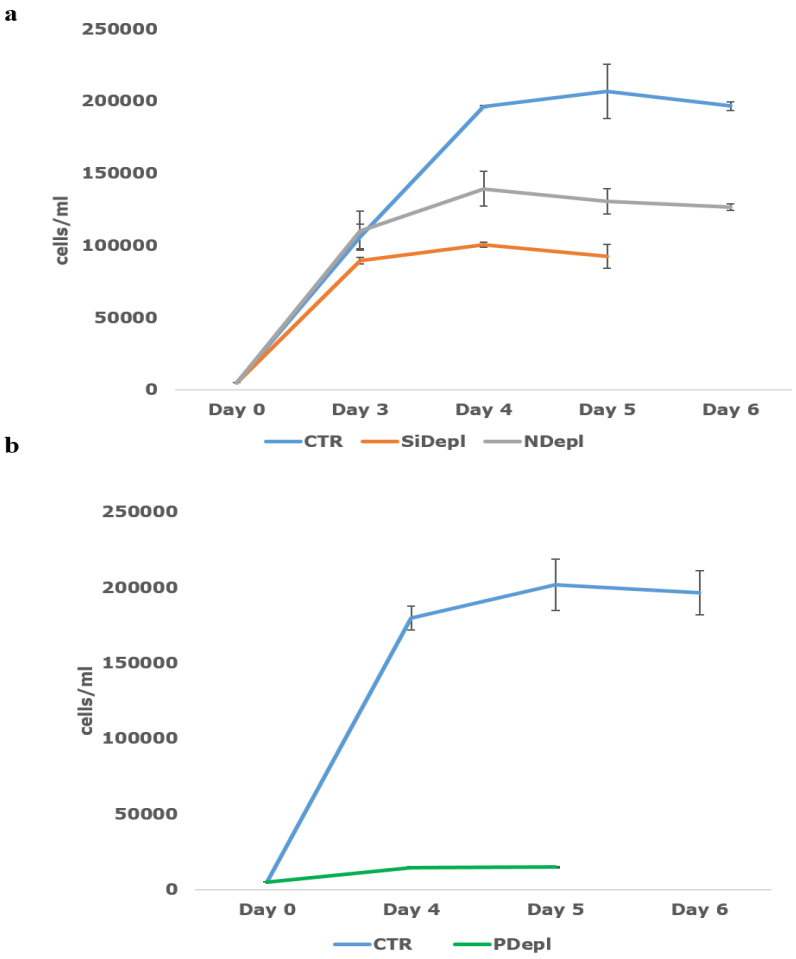
Supplementary Table 3.1: RPKM expression values of: **a)** MEV pathway enzymes; **b)** MEP pathway enzymes; **c)** Iriroid pathway enzymes.

Supplementary Table 3.2: qPCR slope, efficiency (%) and R_2 of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR, SiDepl and NDepl samples mix).

Supplementary Table 3.3: qPCR slope, efficiency (%) and R_2 of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR and PDepl samples mix).

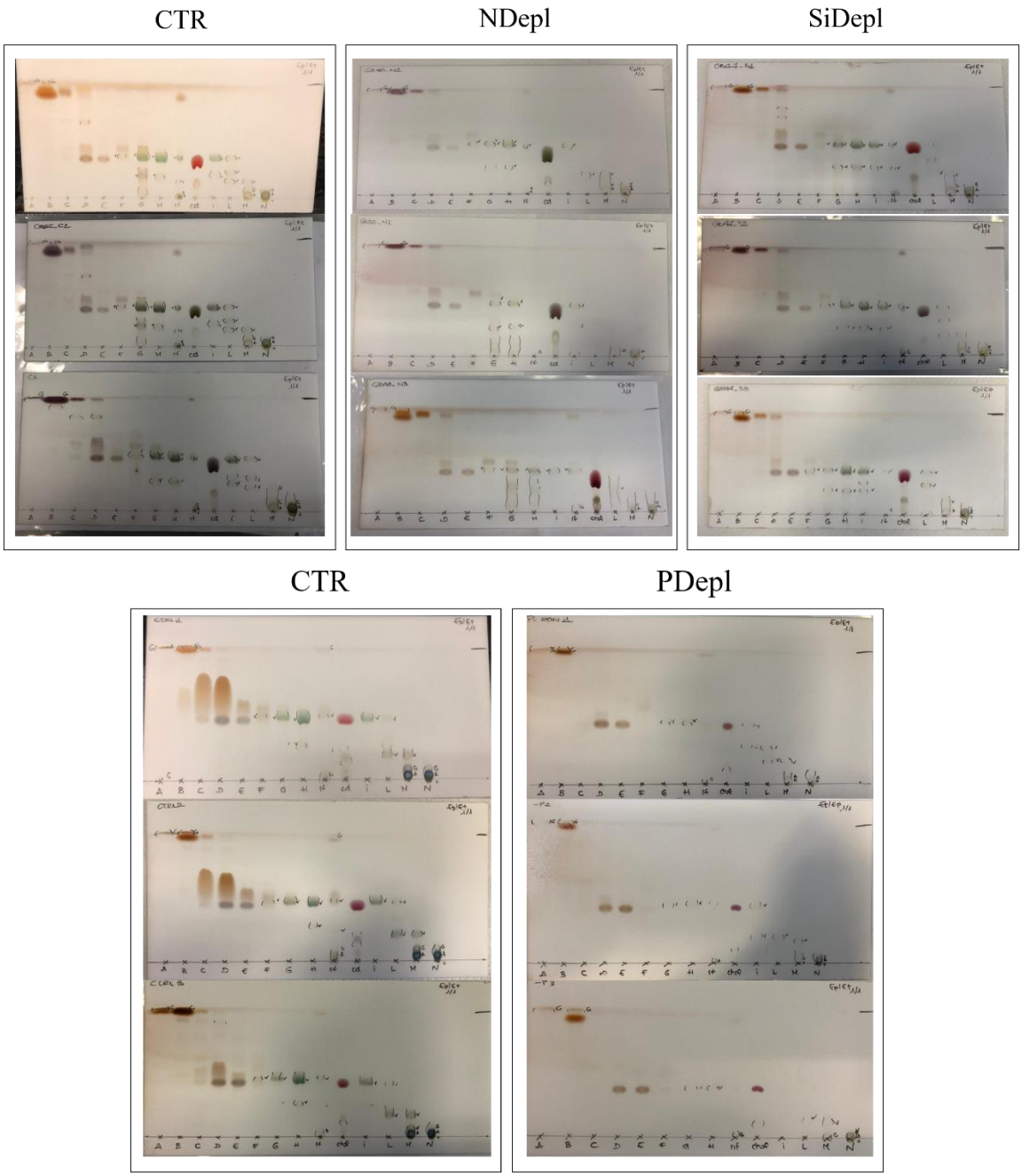
Supplementary Figures.

Supplementary Figure 3.1. *Thalassiosira rotula* growth curves in the different experimental conditions tested. The graphs report mean and standard deviation of the culture triplicates cells/mL in each growth condition. **a)** f/2 medium (CTR), silica depletion medium (SiDepl) and nitrogen depletion medium (NDepl); **b)** f/2 medium (CTR) and phosphate depletion medium (PDepl).



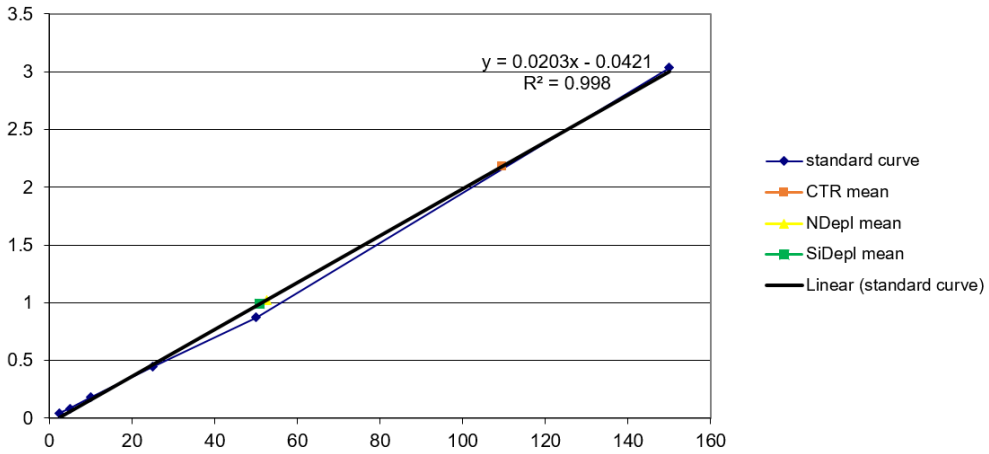
Supplementary Figure 3.2. TLC of the fractionated lipids extracts.

CTR= controls, NDepl= nitrogen starved cultures, SiDepl= silica starved cultures, PDepl= phosphate starved cultures. Letters A to N= different fractions (see materials and methods). Cd/Chol= cholesterol, used as standard. Rf= unfractionated extract. The solvent used for TLC was prepared with petroleum ether (Ep) and diethyl ether (Et) in a 1:1 ratio.

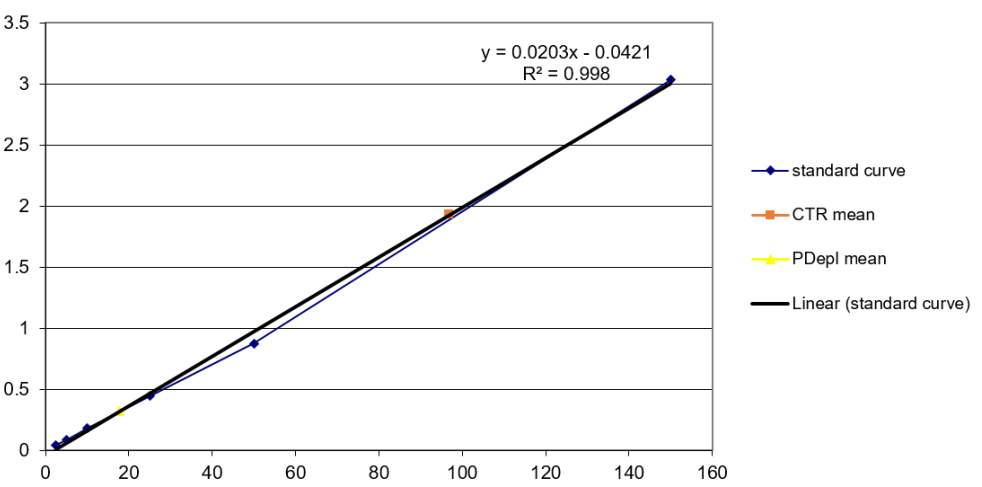


Supplementary Figure 3.3. GC-MS calibration curves, obtained using cholesterol as internal standard, for the quantization of the 24-methylenecholesterol (24-MC) in the different experimental conditions tested. **a)** f/2 medium (CTR), nitrogen starved cultures (NDepl) and silica starved cultures (SiDepl). **b)** f/2 medium (CTR) and phosphate starved cultures (PDepl).

a

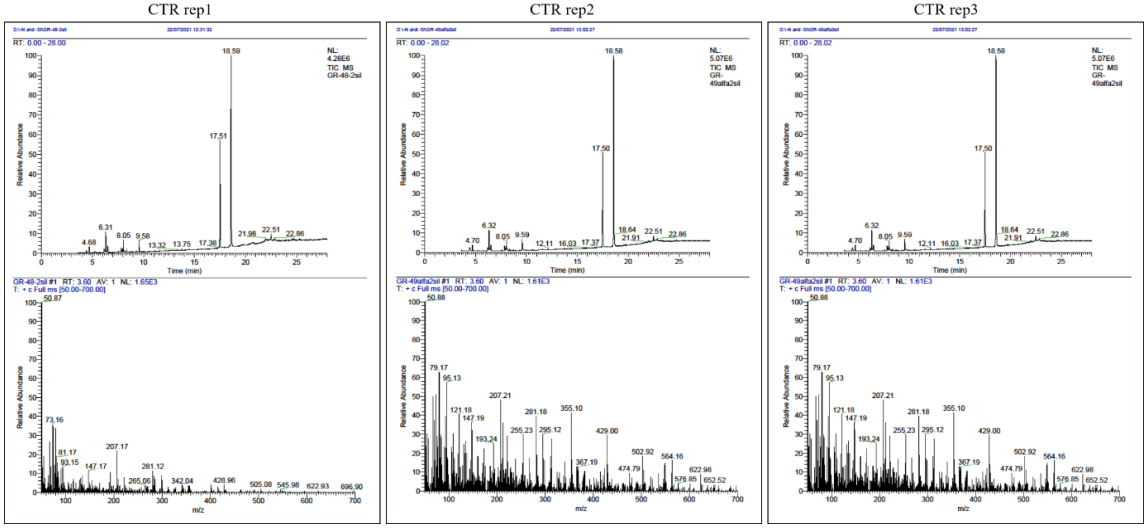


b

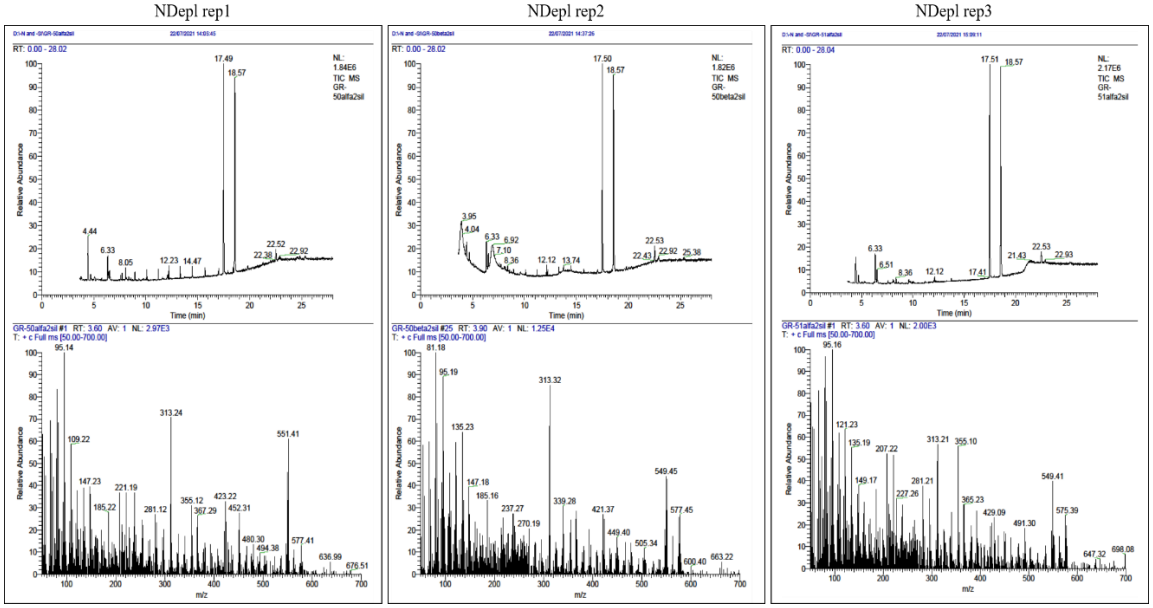


Supplementary Figure 3.4. Complete images of the GC-MS chromatograms of the *T. rotula* sterol fractions from cultures grown in different nutritional conditions. For each growth condition (CTR, SiDepl, NDepl and PDepl described above) is reported the GC-MS chromatogram of each member of the triplicate (rep 1-3). GC-MS peaks fragmentation patterns are reported at the bottom of each figure panel.

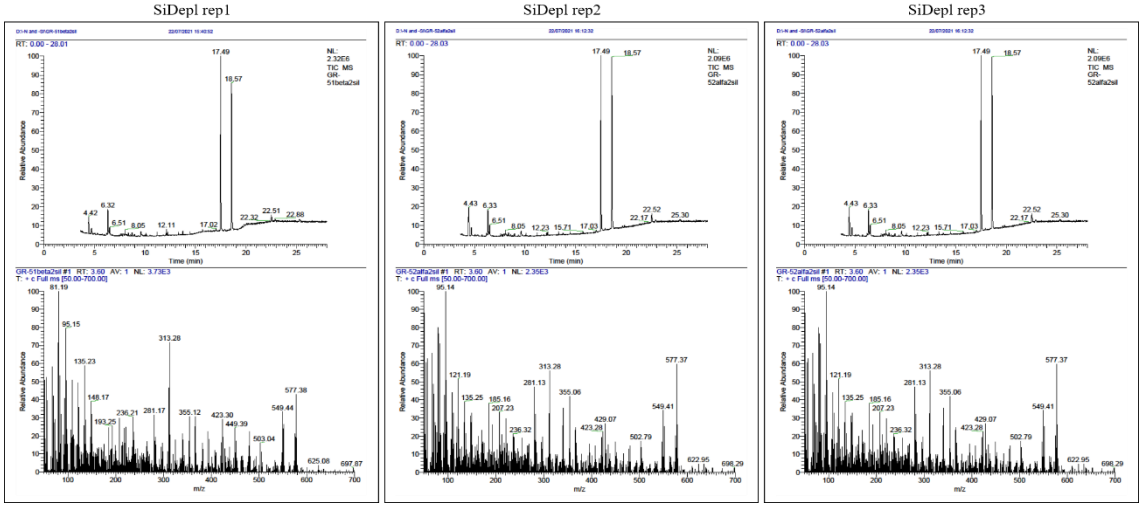
CTR rep1 CTR rep2 CTR rep3

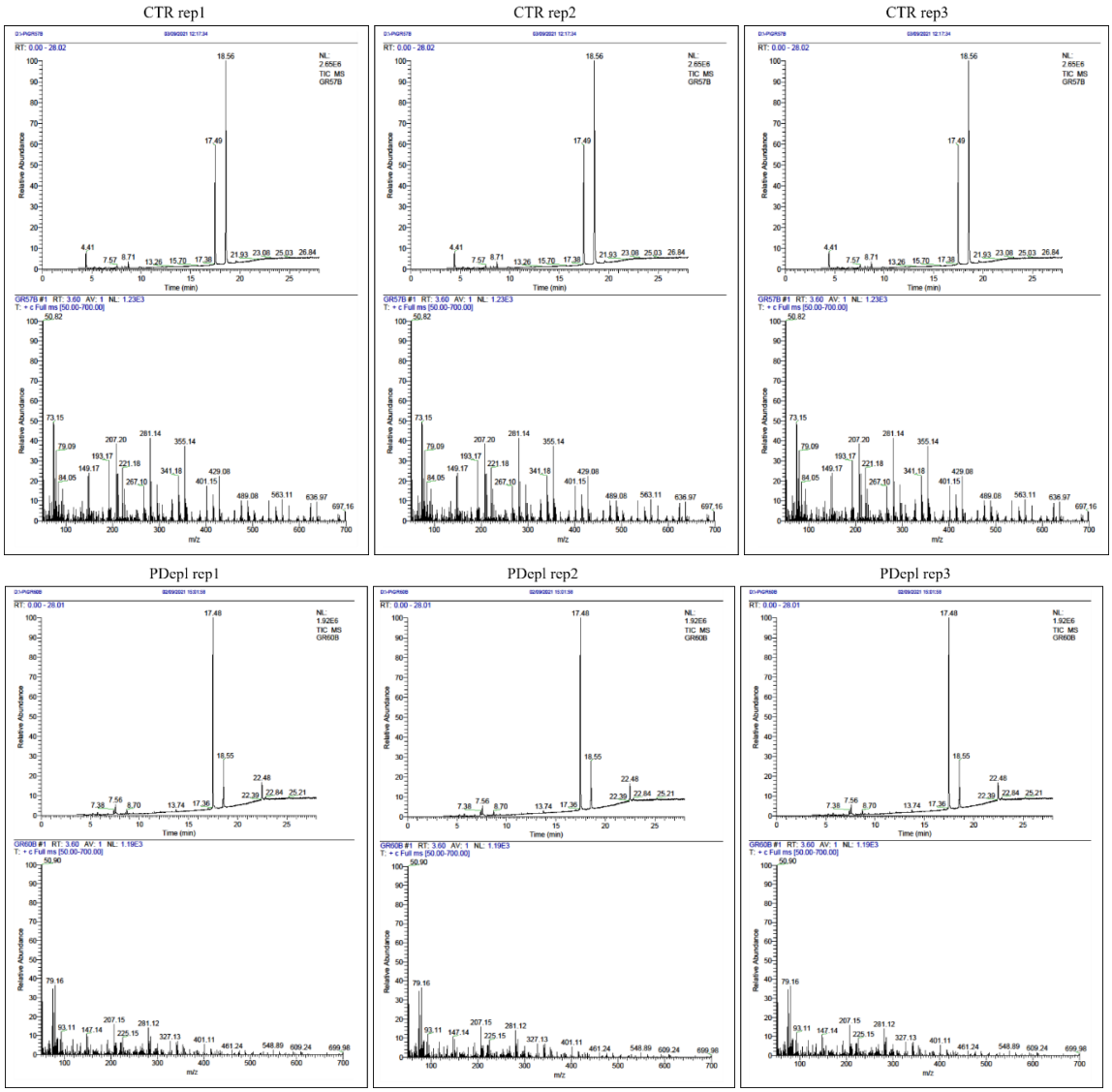


NDepl rep1 NDepl rep2 NDepl rep3

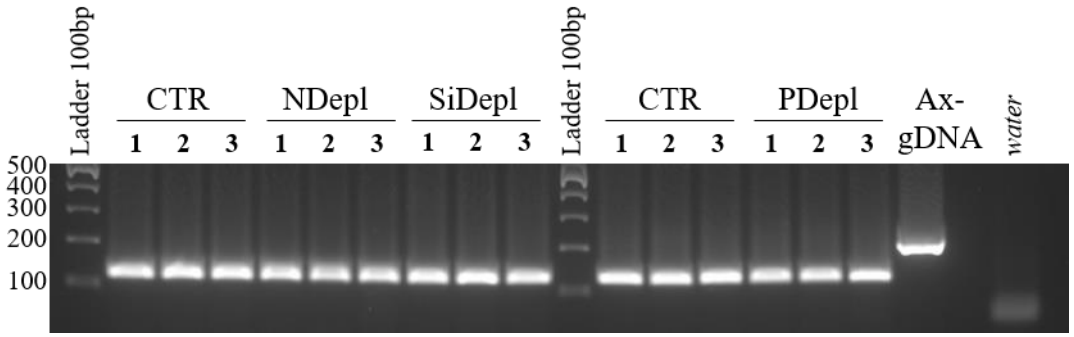


SiDepl rep1 SiDepl rep2 SiDepl rep3



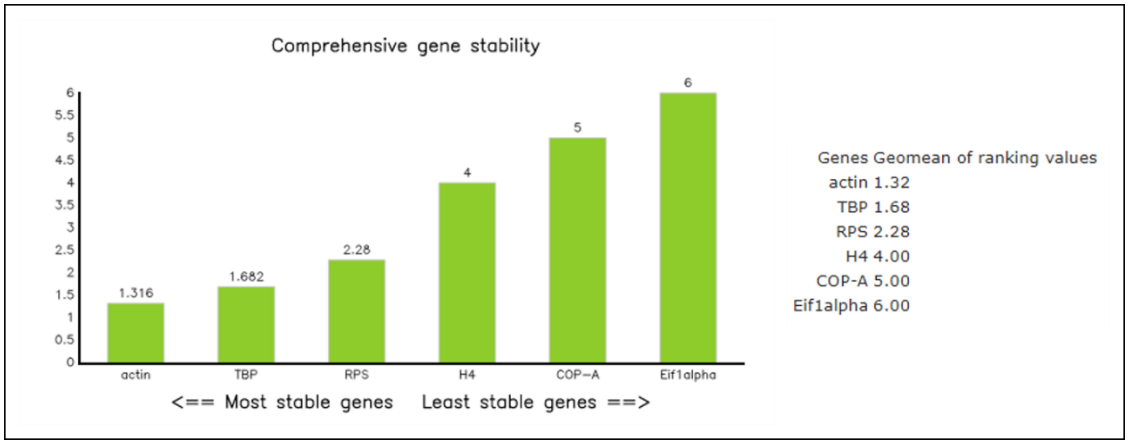


Supplementary Figure 3.5. Confirmation of the absence of gDNA contamination in the cDNA extracted from the *T. rotula* cultures grown in the different experimental conditions tested. 1.5 % agarose gel of a PCR performed with 1 µg of each cDNA using an intron-spanning primer couple for the COP-A gene. As control, the same PCR was performed on the axenic gDNA of *T. rotula* (Ax-gDNA). The presence of a single amplicon in the cDNA samples of different size with respect to gDNA one demonstrates the absence of gDNA contamination in the cDNA samples. Ladder 100 bp: from bottom to top: 100, 200, 300, 400, 500 bps.

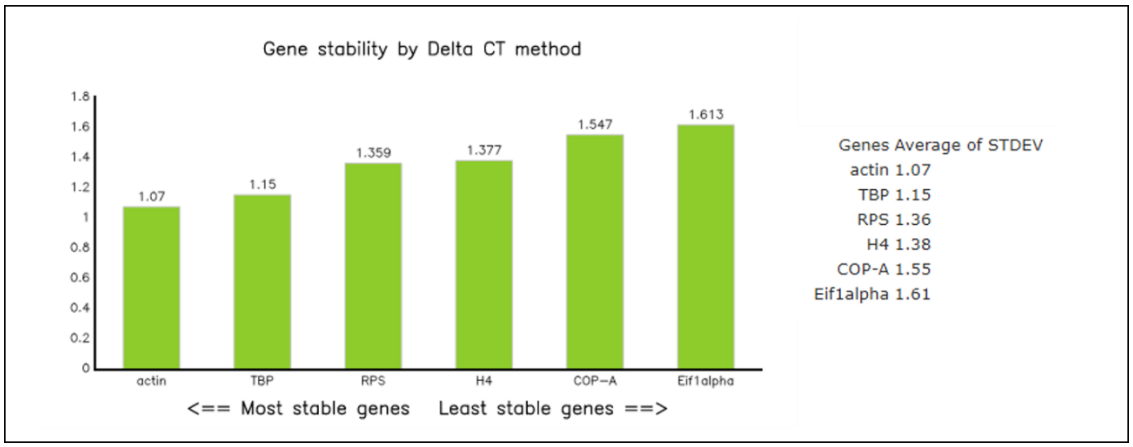


Supplementary Figure 3.6: Analysis of the best reference gene using the RefFinder tool:

a) comprehensive gene stability.

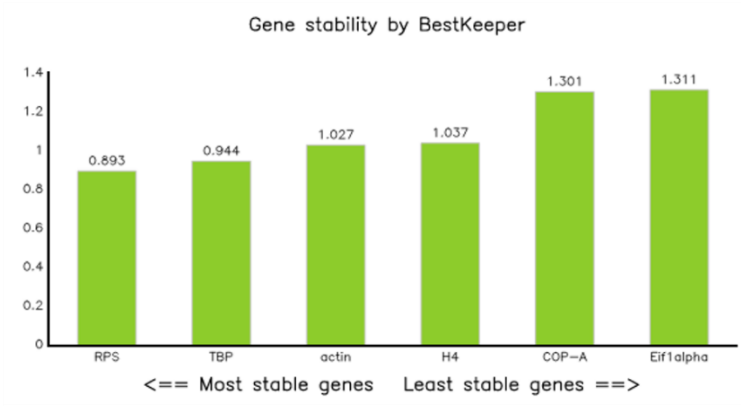


b) gene stability by Delta Ct method.



c) gene stability by NormFinder.

d) gene stability by BestKeeper.



CP data of housekeeping Genes by BEST KEEPER

	actin	RPS	TBP	COP-A	Eif1alpha	H4
n	36	36	36	36	36	36
geo Mean [CP]	23.77	26.84	28.91	29.95	33.07	27.78
AR Mean [CP]	23.80	26.87	28.93	30.00	33.11	27.80
min [CP]	22.44	24.93	27.09	25.76	30.40	25.85
max [CP]	26.64	28.70	31.59	34.05	37.04	29.73
std dev [+/- CP]	1.03	0.89	0.94	1.30	1.31	1.04
CV [% CP]	4.31	3.32	3.26	4.34	3.96	3.73
min [x-fold]	-2.52	-3.78	-3.53	-18.28	-6.36	-3.81
max [x-fold]	7.29	3.62	6.42	17.11	15.64	3.88
std dev [+/- x-fold]	2.04	1.86	1.92	2.46	2.48	2.05

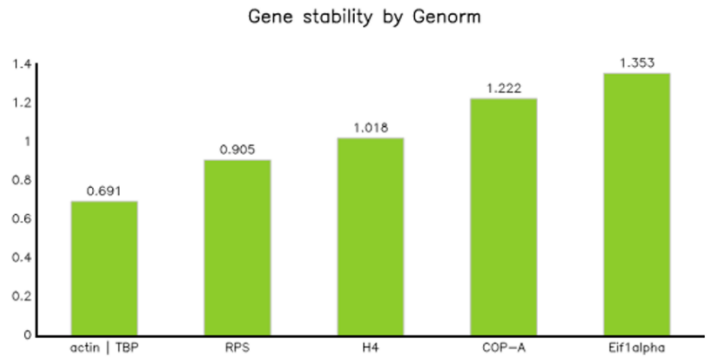
Pearson correlation coefficient (r) by BEST KEEPER

	actin	RPS	TBP	COP-A	Eif1alpha	H4
RPS	0.779	-	-	-	-	-
p-value	0.001	-	-	-	-	-
TBP	0.849	0.476	-	-	-	-
p-value	0.001	0.003	-	-	-	-
COP-A	0.620	0.345	0.639	-	-	-
p-value	0.001	0.039	0.001	-	-	-
Eif1alpha	0.492	0.136	0.576	0.531	-	-
p-value	0.002	0.431	0.001	0.001	-	-
H4	0.684	0.450	0.510	0.422	0.127	-
p-value	0.001	0.006	0.001	0.010	0.461	-

Pearson correlation coefficient (r)

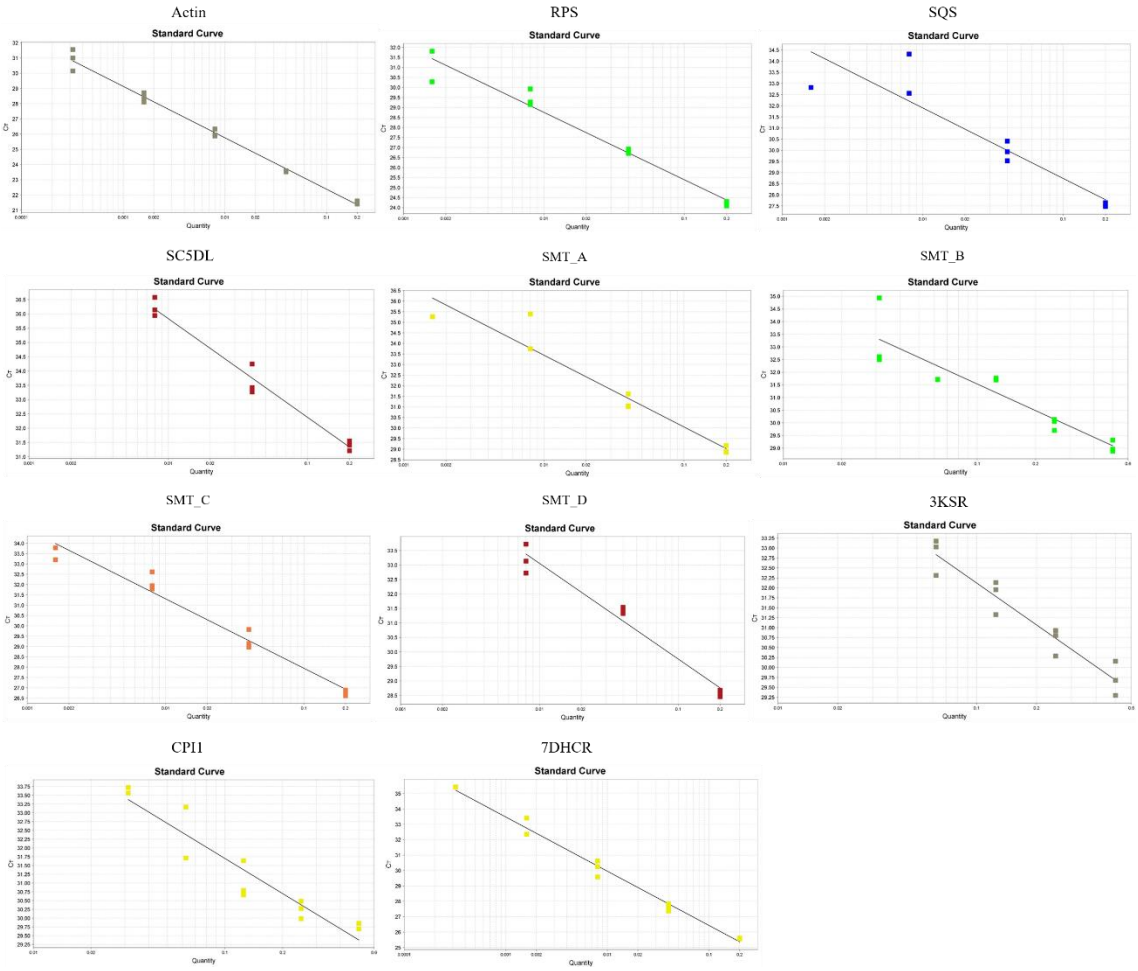
BestKeeper vs.	actin	RPS	TBP	COP-A	Eif1alpha	H4
coeff. of corr. [r]	0.951	0.677	0.873	0.799	0.630	0.676
p-value		0.001	0.001	0.001	0.001	0.001

e) gene stability by Genorm.

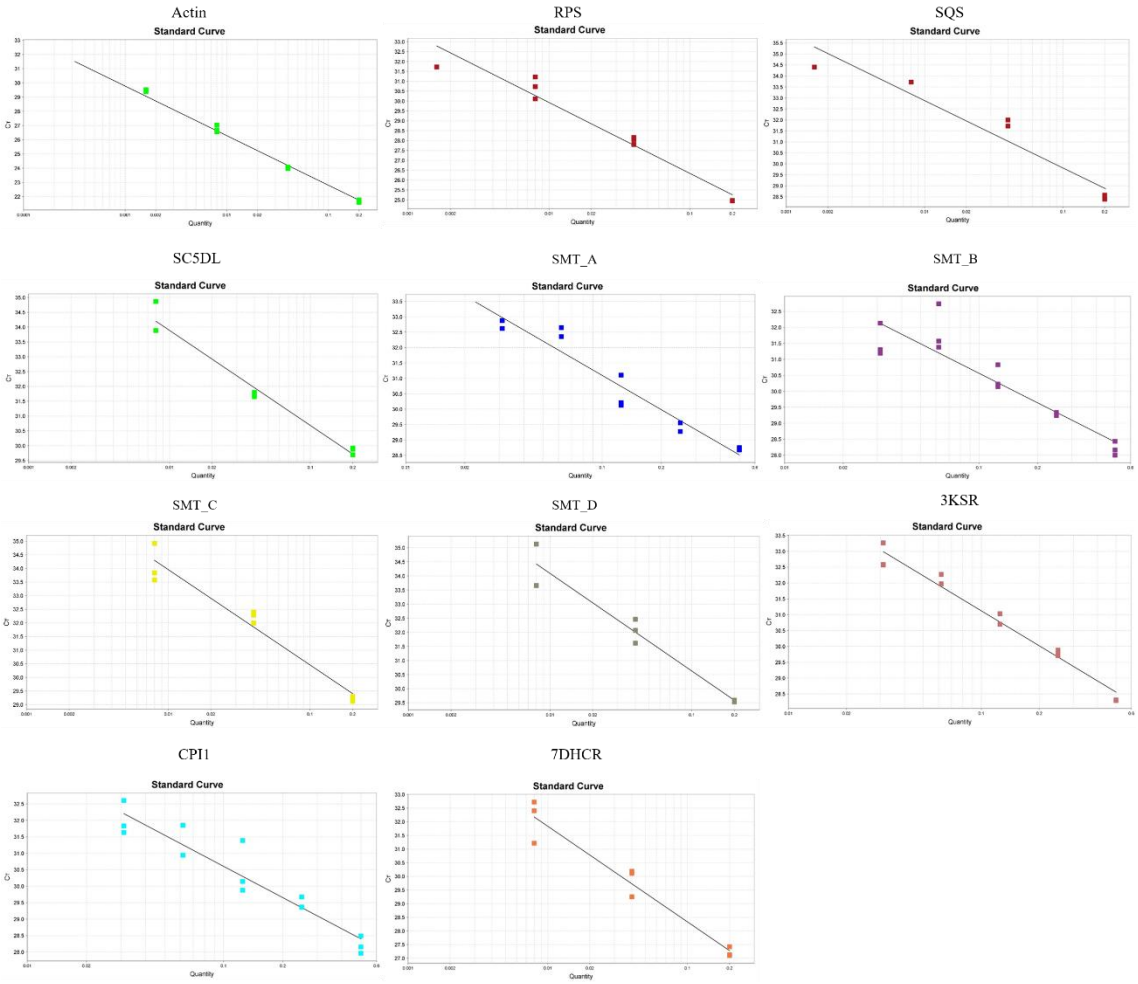


Gene name	Stability value
actin | TBP	0.691
RPS	0.905
H4	1.018
COP-A	1.222
Eif1alpha	1.353

Supplementary Figure 3.7: qPCR standard curve of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR, SiDepl and NDepl samples mix).



Supplementary Figure 3.8: qPCR standard curve of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR and PDepl samples mix).



Supplementary Tables.

Supplementary Table 3.1: RPKM expression values of:

a) MEV pathway enzymes;

Transcript ID	Enzyme	Complete Medium	Silica Depleted Medium
		RPKM	
TR29102|c1_g1_i1	AACT	2,57	2,39
TR16704|c0_g1_i2	HMGS	2,50	1,23
TR1112|c0_g1_i3	HMGR	62,01	34,51
TR8551|c0_g1_i1	MK	31,63	36,88
TR55782|c0_g1_i1	PMK	6,08	5,76
TR6289|c0_g1_i1	MPDC	26,35	21,86

b) MEP pathway enzymes;

Transcript ID	Enzyme	Complete Medium	Silica Depleted Medium
		RPKM	
TR37205|c0_g1_i1	DXS	4,3	2,7
TR2820|c0_g1_i1	DXR	4,7	5,4
TR46467|c0_g1_i1	MCT	1,1	1,8
TR2889|c0_g1_i1	CMK	0,82	0,77
TR30|c0_g2_i1	MECS	2,1	4,9
TR12957|c0_g1_i2	HDS	6,43	7,3
TR12704|c0_g1_i1	HDR	51,14	44,94

c) Iriroid pathway enzymes.

Transcript ID	Enzyme	Complete Medium	Silica Depleted Medium
		RPKM	
TR34822|c0_g3_i1	IDI	5,3	4,2
TR59462|c0_g2_i1	GDPS	13,5	23
	GES		

TR45707|c0_g2_i1	G10H	6,9	3,73
	IS		
	7-DL		
TR47837|c0_g1_i1	7-DLGT	7,9	11,72
TR45636|c1_g1_i1	LAMT	2,2	2,97
TR26037|c1_g1_i2	SLS	1,9	22,2
TR7289|c0_g1_i2	TDC	16,4	18,6
TR52240|c0_g1_i1	STR	92,52	156,3

Supplementary Table 3.2. qPCR slope, efficiency (%) and R_2 of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR, SiDepl and NDepl samples mix).

Name	**Slope**	**Efficiency (%)**	**R_2**
Actin	-3.38	97.62	0.991
RPS	-3.368	98.129	0.967
TBP	-3.633	88.475	0.824
SQS	-3.163	107.069	0.849
SMT_A	-3.395	97.05	0.929
SMT_B	-3.493	93.23	0.853
SMT_C	-3.366	98.201	0.968
SMT_D	-3.03	100.8	0.965
SC5DL	-3.454	94.765	0.976
CYC	-3.322	100.01	0.888
7DHCR	-3.522	92.282	0.986

Supplementary Table 3.3. qPCR slope, efficiency (%) and R_2 of the primer pairs designed for the different transcripts in the sterol biosynthetic pathway (CTR and PDepl samples mix).

Name	Slope	Efficiency (%)	R_2
Actin	-3.489	93.478	0.97
RPS	-3.592	89.861	0.956
TBP	-3.633	88.475	0.824
SQS	-3.067	111.866	0.922
SMT_A	-3.694	86.529	0.926
SMT_B	-3.072	111.618	0.842
SMT_C	-3.501	93.019	0.953
SMT_D	-3.464	94.413	0.952
SC5DL	-3.21	104.916	0.969
CYC	-3.152	107.596	0.902
7DHCR	-3.502	93.008	0.949

PART 4

Supplementary Information

Chapter 4

Characterization of three novel bacteria associated with *T. rotula* and *Skeletonema marinoi*

Adapted from the published article:

Di Costanzo, F., Di Dato, V., van Zyl, L. J., Cutignano, A., Esposito, F., Trindade, M., Romano, G. Three novel bacteria associated with two centric diatom species from the Mediterranean Sea, *Thalassiosira rotula* and *Skeletonema marinoi. International Journal of Molecular Sciences* **22**, 13199 (2021)

Supplementary Information

Supplementary Figures.

Supplementary Figure 4.1: Analysis of MAG secondary metabolite pathways predicted using antiSMASH.

Supplementary Figure 4.2: Confirmation of Cls presence in diatoms xenic cultures.

Supplementary Figure 4.3: Uncropped image showing the detection of MAG following the introduction of *S. marinoi* FE7-FL bacteria into axenic *T. rotula* FE80 cultures.

Supplementary Figure 4.4: qPCR Standard Curve and its Melt Curve Plot with specific primers for the evaluation of Cl-1 abundance in SA fraction of cultures grown in different salinities.

Supplementary Figure 4.5: qPCR Standard Curve and its Melt Curve Plot with specific primers for the evaluation of MAG abundance in SA-FL culture fractions.

Supplementary Figure 4.6: Uncropped image showing the detection of MAG growth in different types of media.

Supplementary Figure 4.7: OD measurement and visual inspection of cultures grown in different types of media.

Supplementary Figure 4.8: Uncropped image showing MAG detection in a medium F/2¼ with and without DNase treatment over nine days.

Supplementary Figure 4.9: qPCR Standard Curve and associated melt curve plot with specific primers for the evaluation of MAG abundance in different media.

Supplementary Tables.

Supplementary Table 4.1: General genome features and MIxS of the Cluster 1

Supplementary Table 4.2: General genome features and MIxS of the Cluster 2

Supplementary Table 4.3: General genome features and MIxS of the Cluster 8

Supplementary Table 4.4: PRISM *T. rotula* Metagenome-Assembled Genomes (MAG) Cluster annotation.

Supplementary Table 4.5: qPCR quantization of Cl-1 abundance in SA fraction of cultures grown in different salinities.

Supplementary Table 4.6: qPCR Slope, R^2 and % of Efficiency with specific primers for the evaluation of Cl-1 abundance in cultures with different salinities.

Supplementary Table 4.7: qPCR quantization of MAG abundance in SA and FL fractions.

Supplementary Table 4.8: qPCR Slope, R^2 and % of Efficiency with specific primers for the evaluation of MAG abundance in SA-FL culture fractions samples.

Supplementary Table 4.9: qPCR quantization of MAG abundance in FE7-FL grown in different media.

Supplementary Table 4.10: qPCR quantization of MAG abundance in FE7-FL grown in F/2¼ medium supplemented with DNase-treated sonicate.

Supplementary Table 4.11: qPCR Slope, R^2 and % of Efficiency with specific primers for the evaluation of MAG abundance in different media.

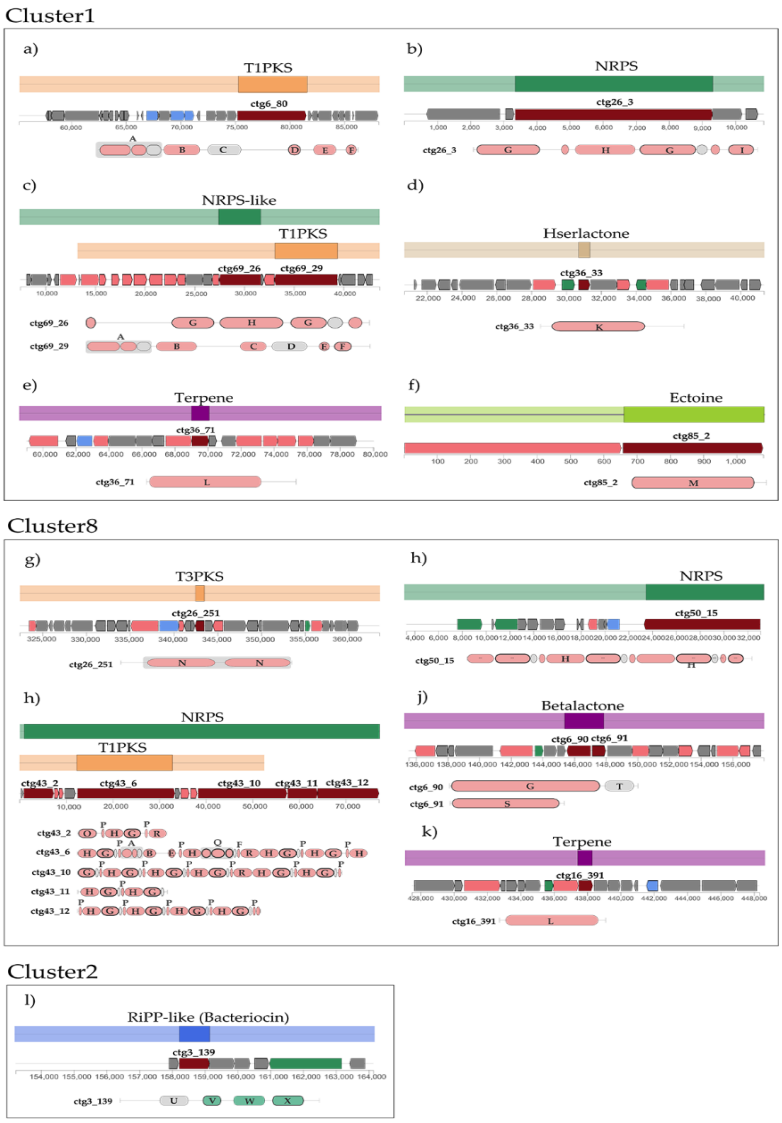
Supplementary Table 4.12: antiSMASH detection of biosynthetic clusters in the whole metagenome.

Supplementary Table 4.13: List of primers used for amplification of the selected gene clusters.

Supplementary Data: Sequences of the clusters-specific genes.

Supplementary Figures.

Supplementary Figure 4.1. Analysis of MAG secondary metabolite pathways predicted using antiSMASH. Representative clusters organization of pathways related to Cluster1 *(a-f)*, Cluster8 *(g-k)* and Cluster2 (l). The genes constituting the clusters in the pathways are represented in different colors: red for core biosynthetic genes, pink for additional biosynthetic genes, blue for transport-related genes, green for regulatory genes and grey for other genes. The pfam domains constituting each core biosynthetic gene are also represented under the total cluster and are indicated by the letters ctg associated to numbers (i.e. ctg3_139). Domain abbreviations: A= ketoacyl synthase (KS), B= acyltransferase (AT), C= polyketide synthase dehydratase (PS-DH), D= Enoylreductase (ER), E= ketoreductase (KR), F= phosphopantetheine attachment site (PP), G= AMP-binding, H= condensation, I= thioesterase (TE), J= Bac_luciferase, K= Autoinducer synthase, L= squalene/phytoene synthase (SQS_PSY), M= Ectoine synthase, N= Chalchone and stylbene synthase, O= Co-enzyme A ligase (CAL), P= Peptidyl carrier protein (PCP), Q= Adenylation domain with integrated oxidase domain (A-OX), R= Epimerization, S=HMGL-like, T= C-terminal, U= Pkinase domain-containing protein 9 (PAS_9), V= His Kinase A (phosphoacceptor) domain (HisKA), W= Histidine kinase-like ATPases (HATPase c), X= Response regulator.

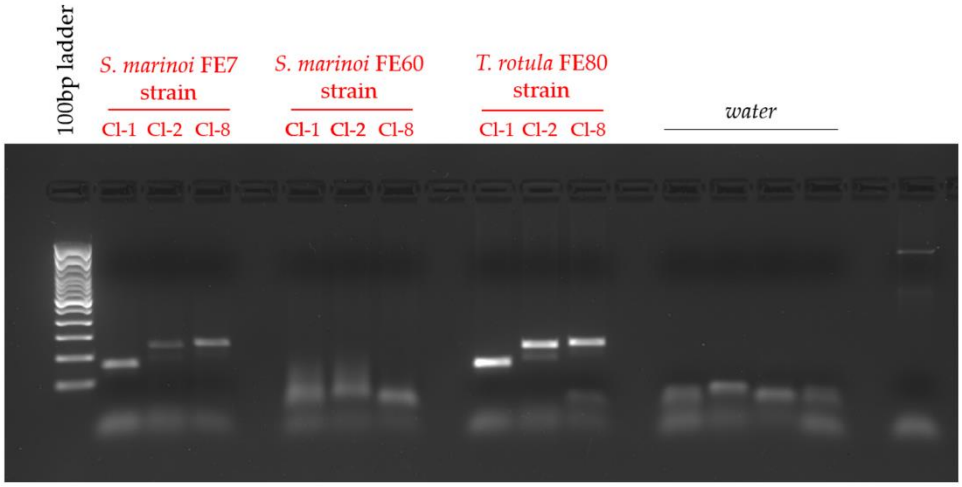


Supplementary Figure 4.2. Confirmation of Cls presence in diatoms xenic cultures.

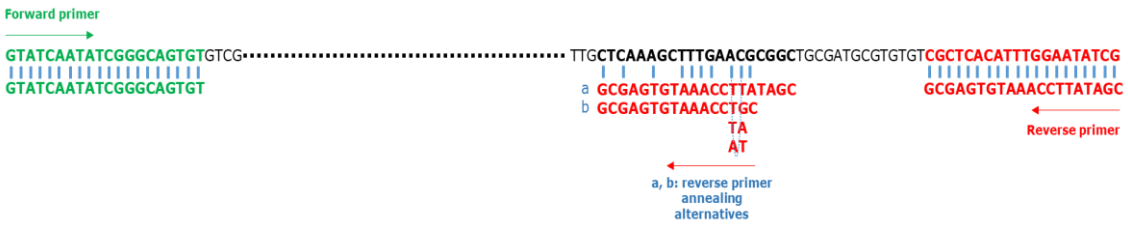
a. Uncropped image showing the MAG presence in different diatom species and strains. PCR amplification from 50ng of free-living bacterial fraction (FL) genomic DNA (gDNAs) with MAG-specific primers. gDNAs were extracted from the FL fraction of *Thalassiosira rotula* FE80 strain, *Skeletonema marinoi* FE7 and FE60 strains harvested at day 7 of growth. Abbreviations: Cl-1= Cluster1, Cl-2= Cluster2, Cl-8= Cluster8. 100bp ladder: from bottom to top: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 3000 bps.

b. Cl-2 reverse primer alternative annealing sites responsible for the lower band visible in panel **a-** lanes *S. marinoi* FE7 strain and *T. rotula* FE80 strain, Cl-2. The reverse primer has some aspecific annealing possibilities forward in the sequence, in respect to the original primer site chosen for the amplification. Sequencing of the lower band confirmed its specificity as Cl-2 contig_82 and its derivation from a second annealing site of the reverse primer.

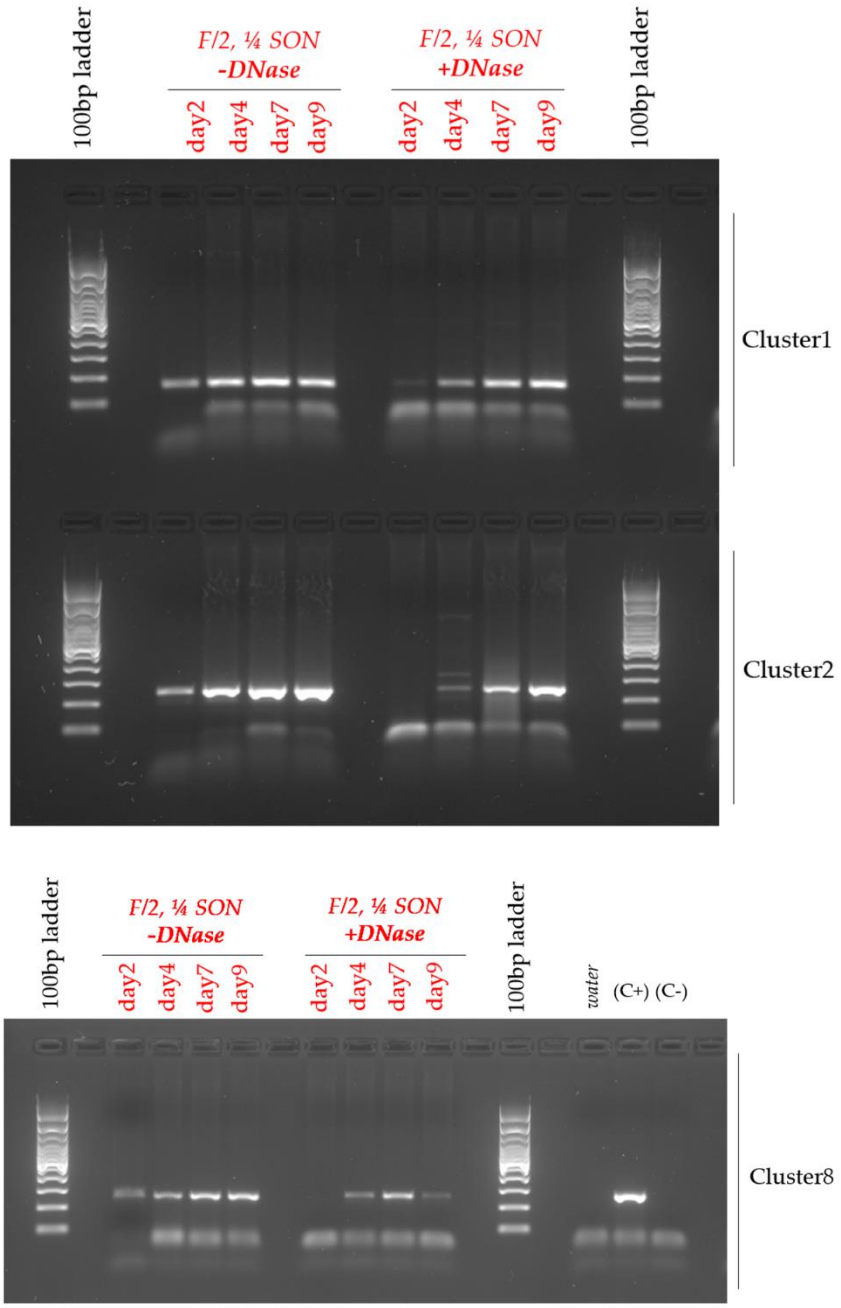
a



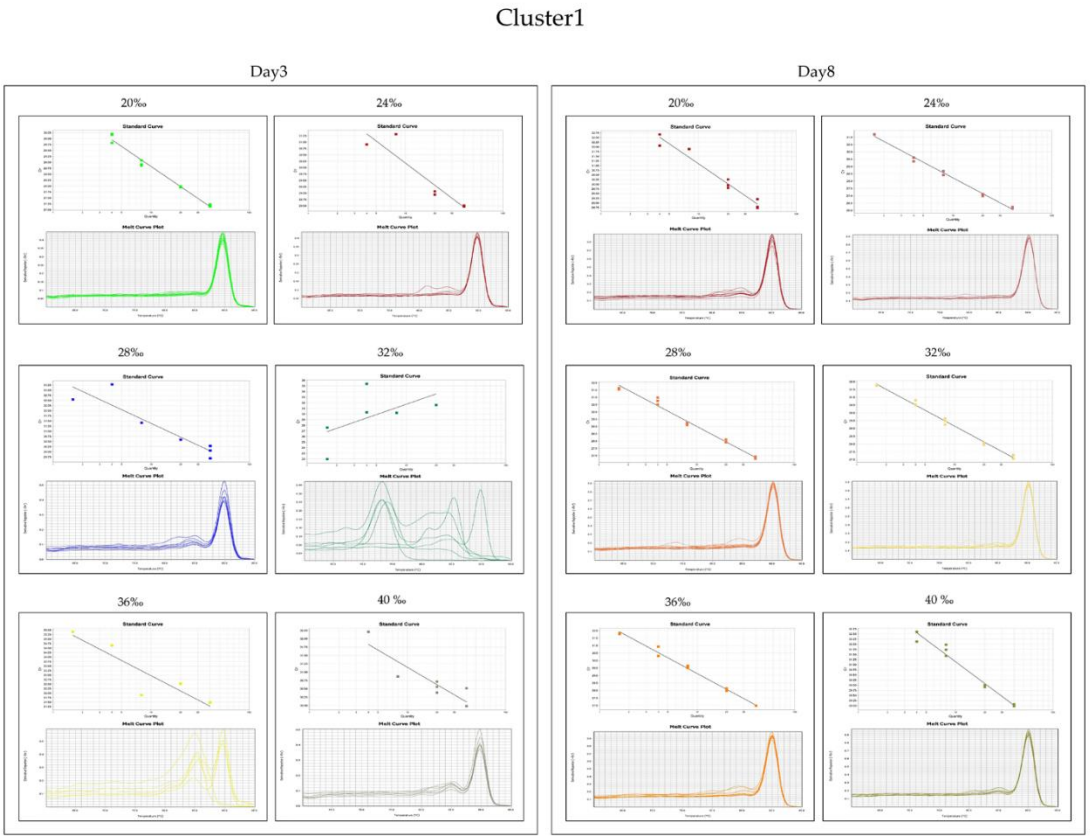
b



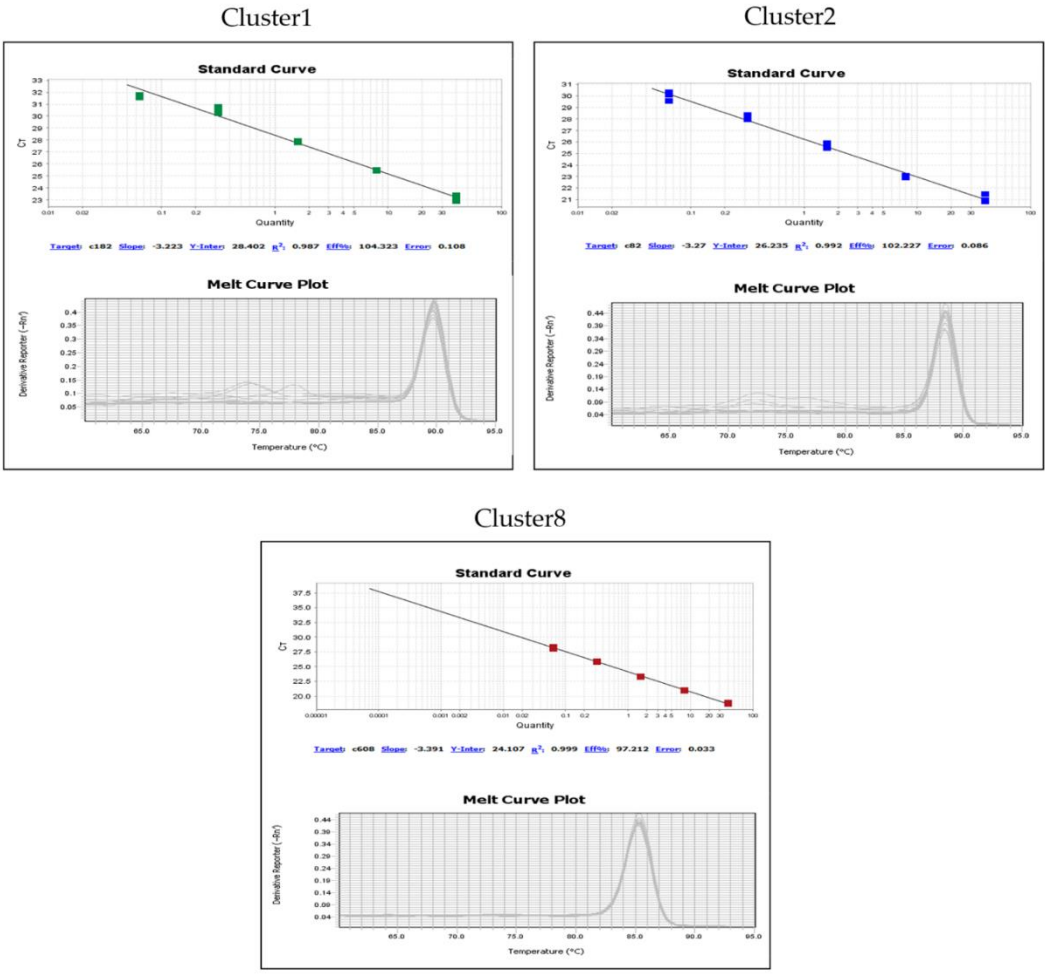
Supplementary Figure 4.3. Uncropped image showing the detection of MAG following the introduction of *S. marinoi* FE7-FL bacteria into axenic *T. rotula* FE80 cultures. MAG PCR amplification from 50ng of strictly-associated (SA) and FL fractions gDNAs collected from *T. rotula* FE80 cultures in which FE7-FL bacteria were reintroduced. MAG presence was inspected after 1 (FE80_1), 3 (FE80_3), 6 (FE80_6) and 8 (FE80_8) months from the bacterial reintroduction. All the cultures were harvested at day 7 of growth. Abbreviations: (C-) = mixed gDNA from FE80 culture in which bacteria were not added, (C+) = mixed gDNA from FE7 culture. 100bp ladder: from bottom to top: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 3000 bps.



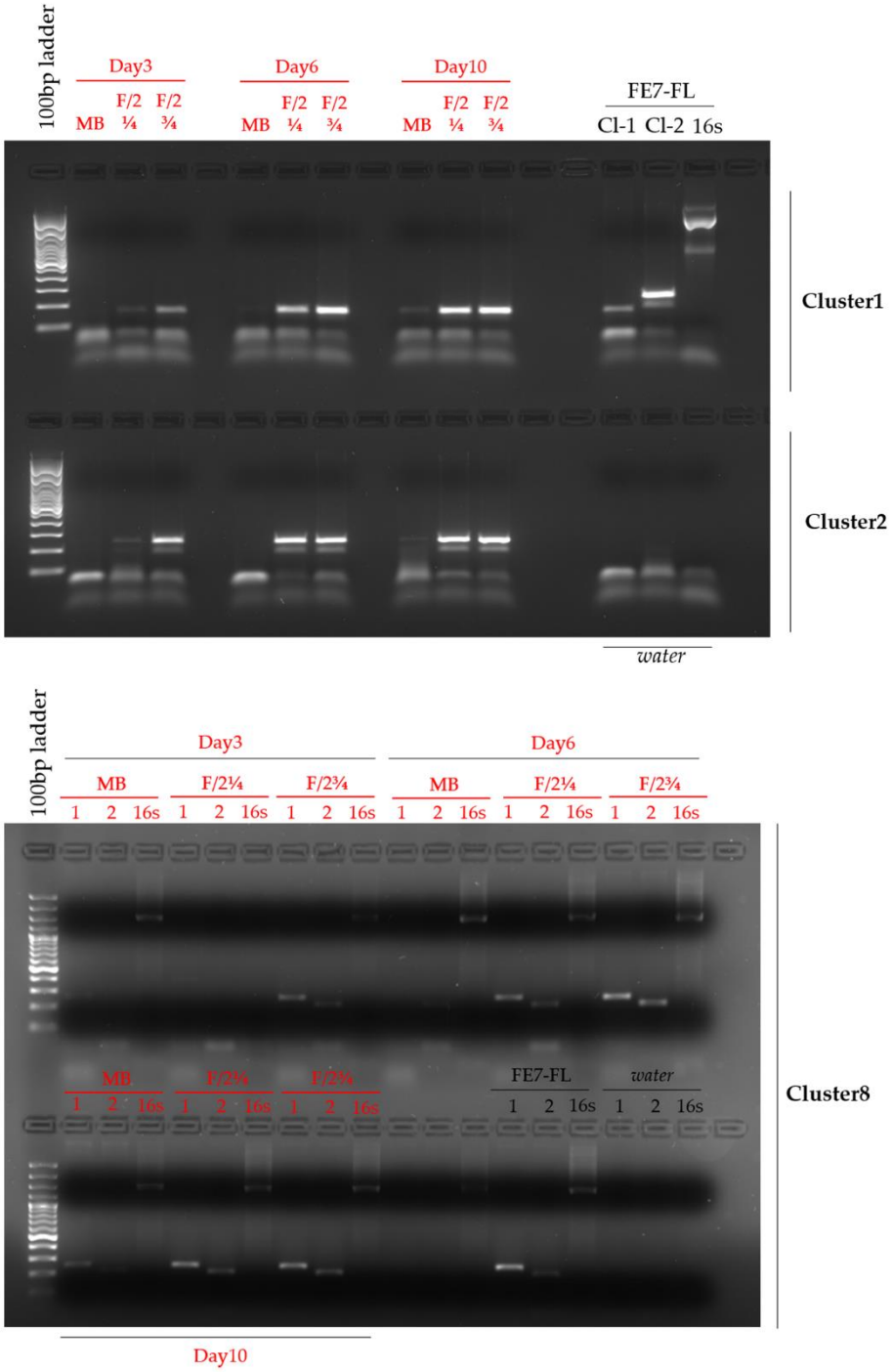
Supplementary Figure 4.4. qPCR Standard Curves and their Melt Curve Plots with specific primers for the evaluation of Cl-1 abundance in SA fraction of cultures grown in different salinities. A standard curve was made for each salinity both at day 3 and day 8 of growth.



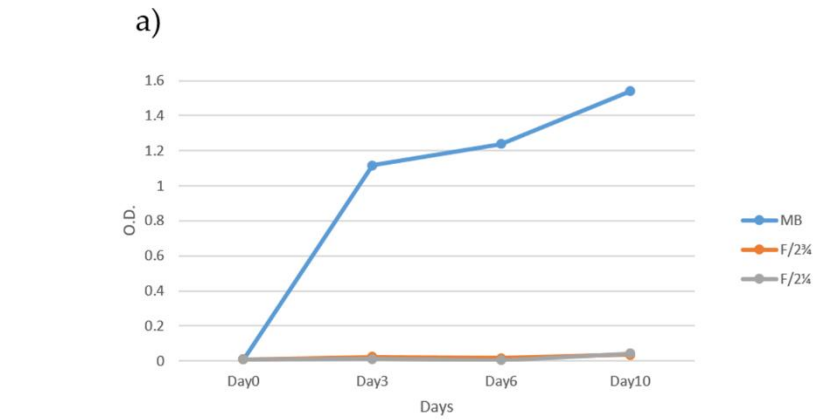
Supplementary Figure 4.5. qPCR Standard Curve and its Melt Curve Plot with specific primers for the evaluation of MAG abundance in SA-FL culture fractions.

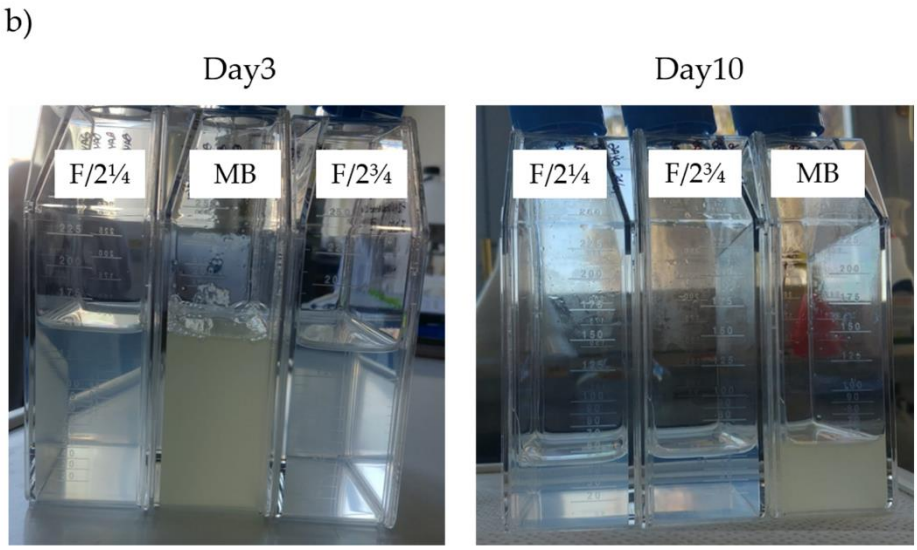


Supplementary Figure 4.6. Uncropped image showing the detection of MAG growth in different types of media. PCR amplifications with MAG-specific primers on gDNA (50ng) of FE7-FL grown in the following types of medium: Marine Broth supplemented with FE7 spent culture medium (MB), F/2 supplemented with two different sonicated diatoms concentrations (¼ and ¾). Bacterial growth was followed along ten days. Abbreviations: Cl-1= Cluster1, Cl-2= Cluster2, 16s= E9/ U1510 universal primer. For Cluster8 were used two primer pairs (1= C8_c988; 2= C8_c450). 100bp ladder: from bottom to top: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 3000 bps.

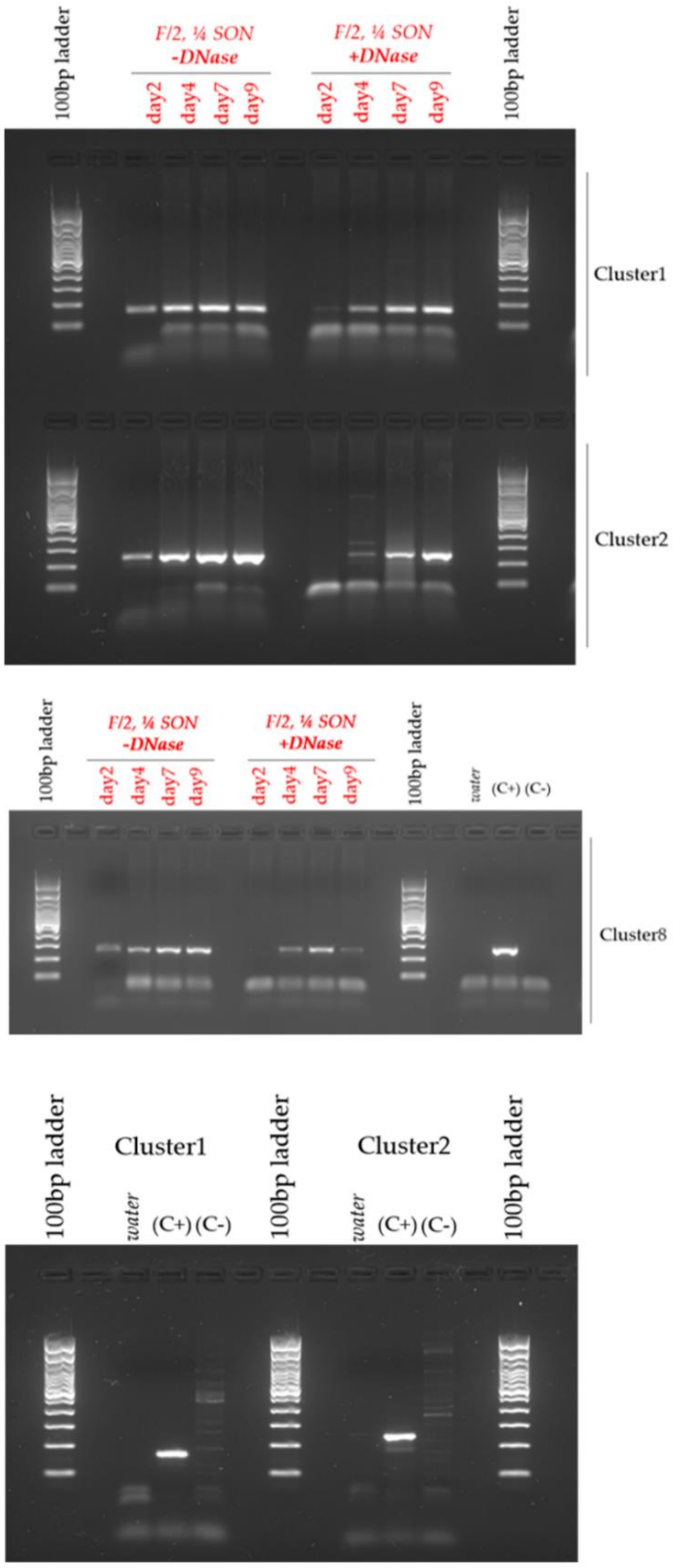


Supplementary Figure 4.7. Optical density (OD) measurement and visual inspection of cultures grown in different types of media. **a)** OD reading of the FE7-FL total bacteria grown in MB or F/2 with the two sonicate concentrations. The OD were normalized on the blank (only MB or only F/2 without bacteria) and the means of three reads for each time point are reported in the graph; **b)** Visual inspection of the total FE7-FL bacteria growing in the three different media at day3 (starting sampling point) and day10 (final sampling point).

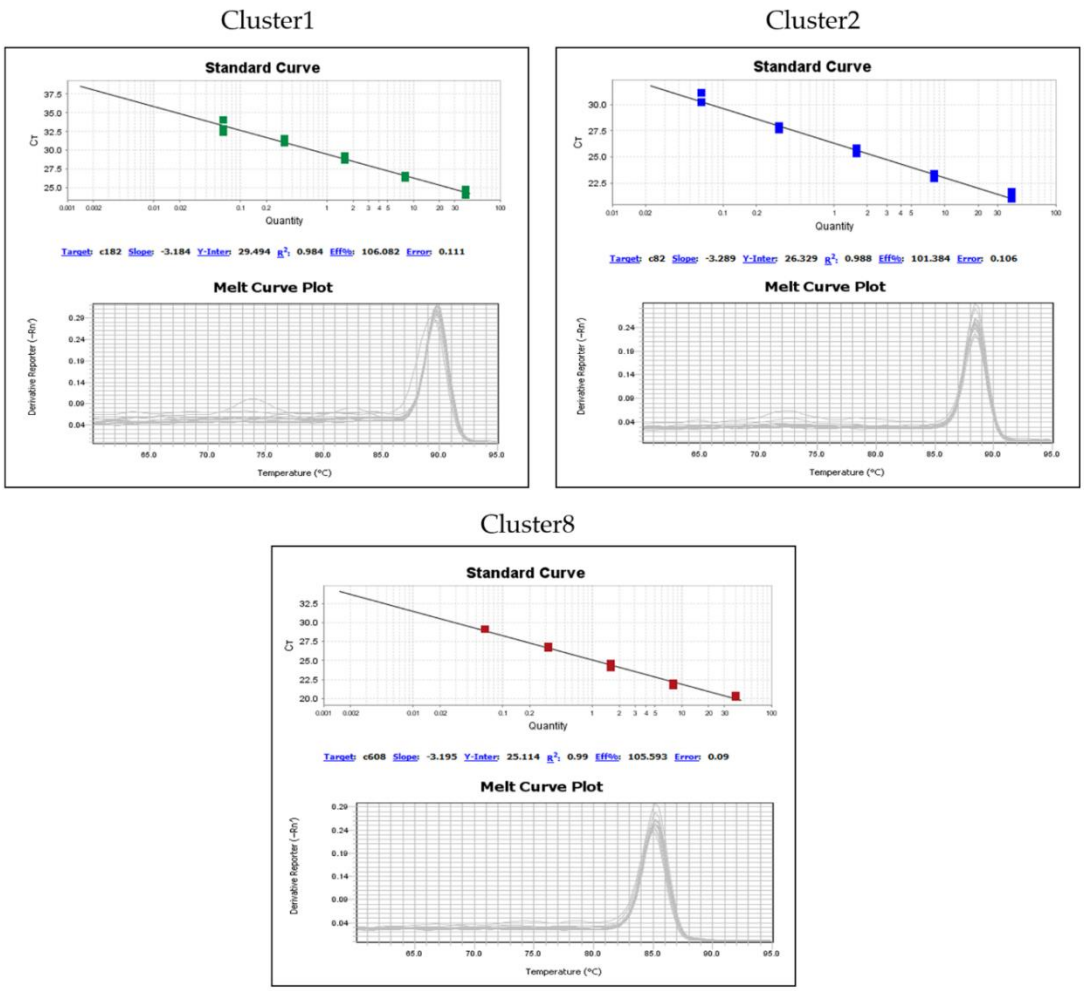




Supplementary Figure 4.8. Uncropped image showing MAG detection in a medium F/2¼ with and without DNase treatment over nine days. PCR-based MAG detection in medium F/2¼ with and without DNase treatment over nine days. PCR amplifications with MAG-specific primers on gDNA (50ng) from FE7-FL bacteria grown in F/2¼ medium with (+DNase) and without (-DNase) the addiction of DNase. MAG presence was evaluated along nine days of culture growth. Abbreviations: (C-)= mixed gDNA from FE80 culture before the reintroduction of the bacteria, (C+)= mixed gDNA from FE7 culture. 100bp ladder: from bottom to top: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 3000 bps.



Supplementary Figure 4.9. qPCR Standard Curve and its Melt Curve Plot with specific primers for the evaluation of MAG abundance in different media cultures.



Supplementary Tables.

Supplementary Table 4.1. General genome features and MIxS of the Cluster 1.

Items	Description
General features Classification Cluster 1	Domain *Bacteria* Phylum *Proteobacteria* Class *Alphaproteobacteria* Order *Rhodobacterales* Family *Rhodobacteraceae* Genus *Aestuariivita* Species *sp.*
Gram stain	Negative
Cell shape	NA
Motility	Motile
Sporulation	NA
Temperature Range; Optimum	18°C
pH range; optimum	NA
Salinity range; optimum	36ppm
Carbon source	NA
MIGS data	
Investigation_type	Bacteria
Project_name	Genome sequence of three bacteria associated with the diatom *Thalassiosira rotula*
Lat_lon	40°48.5′N, 14°15′E
Depth	NA
Alt_elev	NA
Geo_loc_name	Mediterranean sea: Central Tyrrhenian sea: East Sector
Collection_date	2011
Env_biome	marine biome [ENVO:00000447]
Env_feature	alga [ENVO:02500019]
Env_material	water [ENVO:00002006]
Env_package	host-associated environment [ENVO:01001000]
Ref_biomaterial	NA
Pathogenicity	NA
Biotic_relationship	Both free living and strictly associated
Trophic_level	Photoheterotroph
Rel_to_oxygen	Aerobic
Isol_growth_condt	NA
Samp_store_temp	18°C
Sediment_type	NA
Genome assembly data	
Seq_meth	Illumina MiSeq
Assembly	CLC Genomics
Finishing_strategy	N/A
Annot_source	PGAP (NCBI)
MAG Genome attribute	

Total number of reads before trimming	20,505,104
Total number of reads after trimming	19,519,539
Average read length	239,5
Number of reads in contigs	17,286,414
Number of contigs >5kb	416
N50 (bp)	146,992
Assembly size (Mb)	22,15
Maximum contig length (bp)	952,220

Supplementary Table 4.2. General genome features and MIxS of the Cluster 2.

Items	Description
General features Classification Cluster 2	Domain *Bacteria* Phylum *Proteobacteria* Class *Alphaproteobacteria* Order *Parvibaculales* Family *Parvibaculaceae* Genus *Mf105b01* Species *sp.*
Gram stain	Negative
Cell shape	NA
Motility	Motile
Sporulation	NA
Temperature Range; Optimum	18°C
pH range; optimum	NA
Salinity range; optimum	36ppm
Carbon source	NA
MIGS data	
Investigation_type	Bacteria
Project_name	Genome sequence of three bacteria associated with the diatom *Thalassiosira rotula*
Lat_lon	40°48.5′N, 14°15′E
Depth	NA
Alt_elev	NA
Geo_loc_name	Mediterranean sea: Central Tyrrhenian sea: East Sector
Collection_date	2011
Env_biome	marine biome [ENVO:00000447]
Env_feature	alga [ENVO:02500019]
Env_material	water [ENVO:00002006]
Env_package	host-associated environment [ENVO:01001000]
Ref_biomaterial	NA
Pathogenicity	NA
Biotic_relationship	Both free living and strictly associated
Trophic_level	
Rel_to_oxygen	Aerobic

Isol_growth_condt	NA
Samp_store_temp	18°C
Sediment_type	NA
Genome assembly data	
Seq_meth	Illumina MiSeq
Assembly	CLC Genomics
Finishing_strategy	N/A
Annot_source	PGAP (NCBI)
MAG Genome attribute	
Total number of reads before trimming	20,505,104
Total number of reads after trimming	19,519,539
Average read length	239,5
Number of reads in contigs	17,286,414
Number of contigs >5kb	416
N50 (bp)	146,992
Assembly size (Mb)	22,15
Maximum contig length (bp)	952,220

Supplementary Table 4.3. General genome features and MIxS of the Cluster 8.

Items	Description
General features **Classification Cluster 8**	Domain *Bacteria* Phylum *Bacteroidota* Class *Bacteroidia* Order *Cytophagales* Family *Cyclobacteriaceae* Genus *Roseivirga_A* Species *sp.*
Gram stain	Negative
Cell shape	NA
Motility	Nonmotile
Sporulation	NA
Temperature Range; Optimum	18°C
pH range; optimum	NA
Salinity range; optimum	36ppm
Carbon source	NA
MIGS data	
Investigation_type	Bacteria
Project_name	Genome sequence of three bacteria associated with the diatom *Thalassiosira rotula*
Lat_lon	40°48.5′N, 14°15′E
Depth	NA
Alt_elev	NA
Geo_loc_name	Mediterranean sea: Central Tyrrhenian sea: East Sector
Collection_date	2011
Env_biome	marine biome [ENVO:00000447]
Env_feature	alga [ENVO:02500019]

Env_material **Env_package**	water [ENVO:00002006] host-associated environment [ENVO:01001000]
Num_replicons	
Ref_biomaterial	NA
Pathogenicity	NA
Biotic_relationship	Both free living and strictly associated
Trophic_level	Chemoorganotroph
Rel_to_oxygen	Aerobic
Isol_growth_condt	NA
Samp_store_temp	18°C
Sediment_type	NA
Genome assembly data	
Seq_meth	Illumina MiSeq
Assembly	CLC Genomics
Finishing_strategy	N/A
Annot_source **MAG Genome attribute**	PGAP (NCBI)
Total number of reads before trimming	20,505,104
Total number of reads after trimming	19,519,539
Average read length	239,5
Number of reads in contigs	17,286,414
Number of contigs >5kb	416
N50 (bp)	146,992
Assembly size (Mb)	22,15
Maximum contig length (bp)	952,220

Supplementary Table 4.4. PRISM *T. rotula* Metagenome-Assembled Genomes (MAG) Cluster annotation.

***T. rotula* MAG-Cluster 1**							
Biosynthetic Cluster	PK		NRP	Acyl homoserine lactone	Acyl homoserine lactone	PK, NRP	Ectoine
Biosynthetic assemblies (ORF #)	71, 75	56, 60, 62	37626, 3	33, 50261	25	23, 24, 26	2
ORFs Domains	**71:** Phosphopantetheinyltransferase **75:** Ketosynthase, Acyltransferase, Dehydratase, Enoylreductase, Ketoreductase, Thiolation.	**56:** Ketosynthase **60:** Thiolation **62:**Acyltransferase	**37626:** Lasso peptide precursor. **3:** Acyl adenylating enzyme, Thiolation, Condensation, Adenylation, Thiolation, Thioesterase.	**33:**Acyl homoserine lactone synthase; **50261:**Dehydratase.	**25:** Acyl homoserine lactone synthase.	**23:**Phosphopantetheinyltransferase; **24:**Formyltransferase, Adenylation, Adenylation, Thiolation; **26:**Ketosynthase, Acyltransferase, Ketoreductase, Dehydratase, Thiolation, Thioesterase.	**2:**Ectoine synthase.
***T. rotula* MAG-Cluster 2**							
Biosynthetic Cluster	PK						
Biosynthetic assemblies (ORFs #)	352, 355, 356						
ORFs Domains	**352:** Acyltransferase; **355:** Thiolation; **356:** Ketosynthase.						

	***T. rotula* MAG-Cluster 8**		
Biosynthetic Cluster	PK, NRP	NRP	lasso peptide
Biosynthetic assemblies (ORFs #)	2, 3,_6,_8,_10,_11, 12	13, 17	178, 80751, 188, 80664, 191, 192
ORFs Domains	**2:** Acyl adenylating enzyme, Thiolation, Condensation, Adenylation, Thiolation, Epimerization. **3:** 2,3-diaminopropionate biosynthesis protein SbnA. **6:** Condensation, Adenylation, Thiolation, Ketosynthase, Acyltransferase, Ketoreductase, Ketoreductase, Thiolation, Condensation, Adenylation, Adenylation, Thiolation, Epimerization, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation. **8:** Amidotransferase. **10:** Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Epimerization, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation. **11:** Condensation, Adenylation, Thiolation, Condensation, Adenylation. **12:** Thiolation, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation.	**13:** Thioesterase. **17:** Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thiolation, Condensation, Adenylation, Thioesterase.	**178:** Lasso peptide transglutaminase homolog; **80751:** Putative lasso peptide precursor; **188:** Sulfotransferase; **80664:** Putative lasso peptide precursor; **191:** Sulfotransferase; **192:** Lasso peptide asparagine synthase homolog

Supplementary Table 4.5. qPCR quantization of Cl-1 abundance in SA fraction of cultures grown in different salinities.

Cluster 1						
Day3						
	20 ‰		**24 ‰**		**28 ‰**	
Sample name	Quantity mean	Quantity SD	Quantity mean	Quantity SD	Quantity mean	Quantity SD
Repl_1	14.5112409591674	0.798255980014801	8.16201210021972	0.606226086616516	N/D	N/D
Repl_2	11.9483919143676	2.14905190467834	11.9473886489868	2.53087162971496	N/D	N/D
Repl_3	12.2064056396484	3.06914186477661	12.0288705825805	3.67823576927185	N/D	N/D
	32 ‰		**36 ‰**		**40 ‰**	
Sample name	Quantity mean	Quantity SD	Quantity mean	Quantity SD	Quantity mean	Quantity SD
Repl_1	N/D	N/D	N/D	N/D	N/D	N/D
Repl_2	N/D	N/D	N/D	N/D	N/D	N/D
Repl_3	N/D	N/D	N/D	N/D	N/D	N/D
Day8						
	20 ‰		**24 ‰**		**28 ‰**	
Sample name	Quantity mean	Quantity SD	Quantity mean	Quantity SD	Quantity mean	Quantity SD
Repl_1	31.6183595657348	5.80254477802653	6.68693971633911	0.831853449344635	11.7782106399536	3.32517790794372
Repl_2	10.9883317947387	1.28338074684143	20.596580505371	1.56980264186859	39.6703758239746	1.72131192684173
Repl_3	10.3919868469238	1.23993575572967	8.58546638488769	1.17350924015045	12.7557649612426	1.61978256702423
	32 ‰		**36 ‰**		**40 ‰**	
Sample name	Quantity mean	Quantity SD	Quantity mean	Quantity SD	Quantity mean	Quantity SD
Repl_1	22.0469570159912	0.88727593421936	13.7929029464721	1.13168704509735	20.7034702301025	2.02587294578552
Repl_2	14.0322103500366	3.17058825492858	30.1682376861572	3.14465661499737	11.9415206909179	0.597728669643402
Repl_3	15.5249853134155	1.89186143875122	26.3027057647705	2.68924212455749	20.1522417068481	1.68568107645698

N/D – Not detected

Supplementary Table 4.6. qPCR Slope, R^2 and % of Efficiency with specific primers for the evaluation of Cl-1 abundance in cultures with different salinities.

Cluster 1

	Day3			Day8		
Standard curves	**Slope**	**R^2**	**Efficiency (%)**	**Slope**	**R^2**	**Efficiency (%)**
20 ‰	-2.816	0.983	126.497	-3.574	0.962	90.458
24 ‰	-2.877	0.872	122.655	-3.564	0.994	90.792
28 ‰	-2.228	0.852	181.147	-3.536	0.982	91.763
32 ‰	6.121	0.332	-31.351	-3.366	0.991	98.197
36 ‰	-2.812	0.803	126.801	-3.503	0.992	92.965
40 ‰	-1.722	0.799	280.820	-3.538	0.971	91.717

Supplementary Table 4.7. qPCR quantization of MAG abundance in SA and FL fractions. Quantity Mean and Quantity Standard Deviations (SD) of the three MAG calculated via qPCR data analysis. Abbreviations: SA= strictly associated bacterial fraction; FL= free-living bacterial fraction; Day4= fourth day of culture growth; Day7= seventh day of culture growth.

Cluster 1

	SA		FL	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day4	2.850962639	0.088266894	3.204114914	0.439069539
Day7	9.169445038	1.758309007	1.82853353	0.187807962

Cluster 2

	SA		FL	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day4	4.156836987	0.520205617	0.114571519	0.026290052
Day7	12.14747334	0.741565466	0.185813785	0.033655368

Cluster 8

	SA		FL	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day4	5.457804203	0.176898703	0.44285664	0.11048381
Day7	9.304740906	1.28759563	3.954962969	0.157003194

Supplementary Table 4.8. qPCR Slope, R^2 and % of Efficiency with specific primers for the evaluation of MAG abundance in SA-FL culture fractions samples.

Primer name	**Slope**	**R^2**	**Efficiency (%)**
Cluster8_c608	-3.291	0.999	97.212
Cluster2_c82	-3.270	0.992	102.227
Cluster1_c182	-3.223	0.987	104.323

Supplementary Table 4.9. qPCR quantization of MAG abundance in FE7-FL grown in different media.

Cluster 1

	MB		F/2¼		F/2¾	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day3	1.280346274	0.130211171	0.027051291	0.036026332	1.242784977	0.166221648
Day6	0.029285605	0.006642076	13.66777325	0.425333649	16.20821953	1.267161489
Day10	0.009822866	0.011881083	29.6822567	3.296260118	42.63631821	2.170769453

Cluster 2

	MB		F/2¼		F/2¾	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day3	0.032683714	0.016084398	0.024166935	0.002645122	4.377933025	0.241250709
Day6	0.061404854	0.039326753	18.35878563	0.782420158	30.27388763	2.298482418
Day10	0.090489812	0.004574142	24.60587311	0.367527306	39.92933273	3.180368423

Cluster 8

	MB		F/2¼		F/2¾	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day3	0.1362793	0.038577229	0.008	0.00915	2.689586878	0.452325106
Day6	0.129597649	0.034835324	18.79345703	3.886913193	14.20588303	1.957057476
Day10	0.095607392	0.018735146	27.99685478	2.574785471	39.20956039	2.05914402

Supplementary Table 4.10. qPCR quantization of MAG abundance in FE7-FL grown in F/2¼ medium supplemented with DNase-treated sonicate.

f/2¼ (sonicate treated with DNase)

	Cluster 1		Cluster 2		Cluster 8	
Days	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD	Quantity Mean	Quantity SD
Day 2	0.096502073	0.042068169	N/D	N/D	0.000123611	6.98944E-05
Day 4	0.410697669	0.012779716	N/D	N/D	0.195641135	0.02719987
Day 7	1.694224715	0.283904761	4.922961235	0.426027685	0.537181735	0.017694673
Day 9	2.03311801	0.131902337	8.805378914	0.84562999	0.418001503	0.036180586

N/D – Not detected

Supplementary Table 4.11. qPCR Slope, R^2 and % of Efficiency with specific primers for the evaluation of MAG abundance in different media.

Primer name	Slope	R^2	Efficiency (%)
Cluster8_c608	-3.195	0.990	105.593
Cluster2_c82	-3.289	0.988	101.384
Cluster1_c182	-3.184	0.984	106.082

Supplementary Table 4.12. antiSMASH detection of biosynthetic clusters in the whole metagenome.

Whole metagenome biosynthetic clusters	
n° of Biosynthetic clusters	**Type of Biosynthetic clusters**
2	Ectoine
5	Terpenes
3	T1PKS
1	T3PKS
9	NRPS
3	Bacteriocin
1	RRE-containing
2	hserlactone
1	betalactone
1	Ranthipeptide
1	arylpolyene
1	Redox-cofactor

Supplementary Table 4.13. List of primers used for amplification of the selected gene clusters.

Primer Name	Forward Sequence	Reverse Sequence	Tm (°C)	Amplicon length (bp)
C8_c988	AGCTTCAAACGTATCGATCA	CTCAAACAGGCAATTGGATG	59.9/63.1	206
C8_c608	GCTCCAGTGTTTTAACCGG	CCATCTATTCTGCCGACC	62.1/60.7	251
C8_c450	TCGCCAATACTGATTATGCT	GTCGTAGTTCCTAAGGTCAC	59.7/55.3	169
C1_c182	CTGATCTGTTATATGATGCGGA	GACATGACAGTGATGCATTG	61.3/60.2	161
C2_c82	GTATCAATATCGGGCAGTGT	CGATATTCCAAATGTGAGCG	58.9/63.0	243
E9/ U1510 (16s)	GAGTTTGATCCTGGCTCAG	GGCTTACCTTGTTACGACTT	60/53.1	1500

Supplementary Data: Sequences of the cluster-specific genes.

C8_c988

>Diatom_S1_L001_R1_001_(paired)_unmapped_reads_[Diatom_S1_L001_R1_001] (paired)_trimmed_(paired)_not_merged_contig_988 Average coverage: 35.51

AGCTTCAAACGTATCGATCAAAAGCCACTTGGTCATACCCTCATATTTACGATAGTGGTGGT
AAGAGTAGGTTTTTCCATCTTCTGAAAAGACATGCTCCCTGCGTACCTCAAACTTTCCGGTCT
CGTCCTTCACAGAGTCTTCATACAGGAGTTTTTTACCTTTGAATTTCAAAGAGATTTGATCCA
TCCAATTGCCTGTTTGAG

C8_c608

>Diatom_S1_L001_R1_001_(paired)_unmapped_reads_[Diatom_S1_L001_R1_001] (paired)_trimmed_(paired)_not_merged_contig_608 Average coverage: 28.01

GCTCCAGTGTTTTAACCGGTTCTGGTTCAGGAGCTGGGGTGATTGCCTCTTTTGTTTCCTGTC
CTGCTTCCGTTTTAGCCTCTTCTTCTTTTTGCTCTTTTAGTGCCGCTGGTGCCGGCACTATAGG
CGCTACTGCCACTGGCGTTTCAACCTGAGCAACGGCAACGACCGTCTCTTCTACAACGGTTG
CAAAATCATGCTCATCCATAATAATGGCTTTGAGCAAGGCACGGTCGGCAGAATAGATGG

C8_c450

>Diatom_S1_L001_R1_001_(paired)_unmapped_reads_[Diatom_S1_L001_R1_001] (paired)_trimmed_(paired)_not_merged_contig_450 Average coverage: 39.94

TCGCCAATACTGATTATGCTTTAAGTAATGCTACCATTACCATTCTGGACGGAGCCACAACC
GGTATGATCACTTTTACTATTCAGGATGATGCGGATATTGAAGGAGACGAGACCGCTACTTT
GACGCTCAGCAATCCATCAGCAGGTGTGACCTTAGGAACTACGAC

C1_c182

>Diatom_S1_L001_R1_001_(paired)_unmapped_reads_[Diatom_S1_L001_R1_001] (paired)_trimmed_(paired)_not_merged_contig_182 Average coverage: 738.28

CTGATCTGTTATGATGCGGAATTCCCCTTGCTGGCGCGCGCGCTGGCGGAGGCCGACATCAT
TCTGGTGCCCTCGTGCACCGACGCGTTGTCGGGCTATTGGCGGGTGCGAATCGGGGCTATGG
CCCGCGCGTTGGAAGGGCAATGCATCACTGTCATGTC

C2_c82

>Diatom_S1_L001_R1_001_(paired)_unmapped_reads_[Diatom_S1_L001_R1_001] (paired)_trimmed_(paired)_not_merged_ contig_82 Average coverage: 66.46

GTATCAATATCGGGCAGTGTGTCGAGAAGGGCGACGCGGGCGTCTTTCAGATTGGCGCGGG
TATCGGCGAGCGCATGTTCCGTTGACCAGCGCACATTGTCGAAGATCGAGCGCACATGAGG
AAAACGCTTCAGGCCCACCAGCCAATAGCCGCCGTCTTCCGCCGGCCCAAAGACAGCATCAT
GATTGCTCAAAGCTTTGAACGCGGCTGCGATGCGTGTGTCGCTCACATTTGGAATATCG

Appendix B

List of Publications

- Di Costanzo, F., Di Dato, V., Ianora, A. & Romano, G. Prostaglandins in Marine Organisms: A Review. *Mar Drugs* **17**, (2019).
- Di Dato, V., Di Costanzo, F., Barbarinaldi, R., Perna, A., Ianora, A., Romano G. Unveiling the presence of biosynthetic pathways for bioactive compounds in the *Thalassiosira rotula* transcriptome. *Scientific Reports* **9**, *9893* (2019).
- Di Dato, V., Barbarinaldi, R., Amato, A., Di Costanzo, F., Fontanarosa, C., Perna, A., Amoresano, A., Esposito, F., Cutignano, A., Ianora, A., Romano, G. Variation in prostaglandin metabolism during growth of the diatom *Thalassiosira rotula*. *Scientific Reports* **10**, 5374 (2020).
- Di Costanzo, F., Di Dato, V., van Zyl, L. J., Cutignano, A., Esposito, F., Trindade, M., Romano, G. Three novel bacteria associated with two centric diatom species from the Mediterranean Sea, *Thalassiosira rotula* and *Skeletonema marinoi*. *International Journal of Molecular Sciences* **22**, 13199 (2021)