environmental microbiology

Environmental Microbiology (2015) doi:10.1111/1462-2920.12914

# Comparative proteomics reveals highly and differentially expressed proteins in field-collected and laboratory-cultured blooming cells of the diatom *Skeletonema costatum*

**Hao Zhang, Da-Zhi Wang,* Zhang-Xian Xie, Shu-Fei Zhang, Ming-Hua Wang and Lin Lin**
*State Key Laboratory of Marine Environmental Science/College of the Environment and Ecology, Xiamen University, Xiamen 361005, China.*

## Summary

**Diatoms are a major phytoplankton group causing extensive blooms in the ocean. However, little is known about the intracellular biological processes occurring during the blooming period. This study compared the protein profiles of field-collected and laboratory-cultured blooming cells of *Skeletonema costatum*, and identified highly and differentially expressed proteins using the shotgun proteomic approach. A total of 1372 proteins were confidently identified with two or more peptides. Among them, 222 and 311 proteins were unique to the laboratory and field samples respectively. Proteins involved in photosynthesis, translation, nucleosome assembly, carbohydrate and energy metabolism dominated the protein profiles in both samples. However, different features of specific proteins were also found: proteins participated in light harvesting, photosynthetic pigment biosynthesis, photoprotection, cell division and redox homeostasis were highly detected in the field sample, whereas proteins involved in translation, amino acid and protein metabolic processes, and nitrogen and carbon assimilation presented high detection rates in the laboratory sample. ATP synthase cf1 subunit beta and light harvest complex protein were the most abundant protein in the laboratory and field samples respectively. These results indicated that *S. costatum* had evolved adaptive mechanisms to the changing environment, and integrating field and laboratory proteomic data should provide comprehensive understanding of bloom mechanisms.**

Received 2 November, 2013; accepted 19 May, 2015. *For correspondence. E-mail dzwang@xmu.edu.cn; Tel. (+86) 592 2186016; Fax (+86) 592 2180655.

## Introduction

Diatoms are among the most abundant and diverse group of all photosynthetic eukaryotes on Earth. They are not only the major primary producers, but also an essential part of the food chain in the ocean (Falkowski *et al*., 1998; Field *et al*., 1998; Sarthou *et al*., 2005). As a key contributor to the marine biological pump, diatoms influence ocean carbon cycling and subsequently regulate the global climate (Buesseler, 1998; Sarthou *et al*., 2005). Their central importance in the biogeochemical cycles of various nutrients, i.e. carbon, silica, nitrogen and other biogenic elements, is well studied and recognized (Morel and Price, 2003; Allen *et al*., 2006; Armbrust, 2009).

Diatoms can often form large-scale blooms and dominate the biomass of phytoplankton communities in well-mixed coastal and upwelling regions. Previous studies demonstrate that environmental factors such as water temperature, salinity, light irradiance and nutrients regulate the occurrence of diatom blooms (Yin, 2003; Ferris and Lehman, 2007; Joseph *et al*., 2008; Hu *et al*., 2011). However, the molecular mechanisms involved in diatom blooms still await discovery. Recent genomic and proteomic studies offer clues as to the success of diatoms in the marine ecosystem (Armbrust *et al*., 2004; Bowler *et al*., 2008; Nunn *et al*., 2009). These studies reveal the evolutionary origins and metabolic adaptations that may have led to their ecological success. For example, diatoms have evolved unique mechanisms to utilize carbon, nutrients and light, which enable them to outcompete other phytoplankton species in a specific environment and dominate the community (Armbrust *et al*., 2004). Multiple response biochemical strategies, such as the reallocation of cellular phosphate (P) and utilization of organic P, occur in the diatoms after P deficiency (Dyhrman *et al*., 2012). Response of diatom central carbon metabolism to nitrogen starvation also differs from other photosynthetic eukaryotes (Hockin

*et al*., 2012). Although these studies offer clues as to the success of diatoms in the marine ecosystem, the specific mechanisms that drive diatom blooms and major biological processes occurring within the blooming cells are yet to be defined.

*Skeletonema costatum* is a widely distributed diatom species in both coastal and open ocean areas, which can frequently form large-scale high biomass blooms, influencing the biogeochemical cycling of carbon and other elements, as well as ecosystem structure. Much effort has been devoted to the *in situ* investigation of environmental conditions, with the focus on physical, chemical and biological proxies during the course of *S. costatum* blooms, in order to understand the mechanisms controlling the occurrence and maintenance of these blooms at the population level (Vårum *et al*., 1986; Gao *et al*., 1993; Robertson and Alberte, 1996; Nandakumar *et al*., 2003). However, little is known regarding the biological processes occurring in the blooming cells, which regulate nutrient assimilation, carbon fixation, light utilization and other physiological activities.

Shotgun proteomics is a mass spectrometry-based technology widely used for large-scale identification of proteins in complex biological samples (Washburn *et al*., 2001; Aebersold and Mann, 2003). Typically, liquid chromatography is coupled with tandem mass spectrometry (MS/MS) resulting in high throughput peptide analysis. The MS/MS spectra are searched against a protein database to identify peptides in the sample. Shotgun proteomics is applied in a variety of organisms (Skipp *et al*., 2005; Jones *et al*., 2011) as well as environmental samples (Ram *et al*., 2005; Wilmes *et al*., 2008; Williams *et al*., 2010; Guazzaroni *et al*., 2012). Recently, shotgun proteomics has also attracted considerable attention in the field of marine science (Wang *et al*., 2014).

In the present study, we applied the MS-based shotgun proteomic approach to compare the protein expression profiles of field-collected (FC) and laboratory-cultured (LC) blooming cells of *S. costatum*, and identified highly and differentially expressed proteins as well as the major biological processes occurring in the blooming cells. The goal of this study was to gain insights into the biological processes occurring in the blooming cells of *S. costatum*, and to advance our understanding of the molecular mechanisms driving diatom blooms. We identified proteins from field and laboratory blooming cells using a combined dataset including genomes of five diatom species and various transcriptomes of diatoms. This enabled us to obtain more protein information and a more detailed picture of the biological processes occurring in the blooming cells. In addition, through comparison of field and laboratory samples, we demonstrated that *S. costatum* had evolved adaptive mechanisms to the changing environment.

## Results

### *Features of the FC and LC blooming samples*

Field surface seawater samples were collected during a phytoplankton bloom that occurred in Xiamen Bay (Fig. 1A). During the bloom period from 26 July 2011 to 2 August 2011, the phytoplankton species in the seawater samples were mainly *S. costatum*, *Akashiwo sanguinea* and *Chaetoceros* sp. (Table 1). Among them, *S. costatum* dominated the phytoplankton biomass and contributed nearly 99% of the phytoplankton biomass on July 27, before its cell density decreased rapidly on the following sampling days. The highest chlorophyll a (Chla) concentration was observed on July 28 (Fig. 1B), demonstrating that July 27 was the blooming stage of *S. costatum*. Hence, the cells on July 27 were collected as the field blooming cells for proteomic analysis.

For laboratory-culture, *S. costatum* with an initial cell density of approximately 8400 cells $ml^{-1}$ grew rapidly after a short lag phase (Fig. 1C). Cell density reached a peak in day 5 at $3.4 \times 10^5$ cells $ml^{-1}$ and was then maintained at a relatively stable level. The cells at day 4 with the maximum growth rate were collected as the cultured blooming sample for proteomic analysis.

### *Separation and identification of proteins*

Typical one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate the proteins extracted from the LC and FC *S. costatum* samples with three replicates for each (Fig. S1). Generally, a high similarity of band patterns was observed between the FC and LC samples. After trypsin digestion and MS analysis, 40 008 $\pm$ 3643 MS/MS spectra were generated from the LC samples and 39 117 $\pm$ 6164 from the FC samples. Using the combined diatom genome and transcriptome dataset, 16.44% $\pm$ 0.42% and 8.03% $\pm$ 2.25% of the MS/MS spectra obtained peptide sequences. A total of 1892 proteins were detected in the FC and LC blooming *S. costatum* samples, and 1372 of them were identified as high-confidence proteins matching two or more peptides, which were selected for further analysis in this study (Table 2). Among the high-confidence proteins, 839 proteins were shared by both samples, 222 were unique to the LC sample, whereas 311 were unique to the FC sample. Detailed information of proteins and peptides detected are listed in Tables S1 and S2.

Of all the FC distinct proteins, 26.63% were assigned to the diatom genus *Skeletonema*, 20.80% to the diatom genus *Chaetoceros* and 17.75% to the diatom genus *Thalassiosira*, followed by the genus *Pseudo-nitzschia* and other diatom species. Of all the LC distinct proteins, 36.83% were assigned to the diatom genus *Skeletonema*,

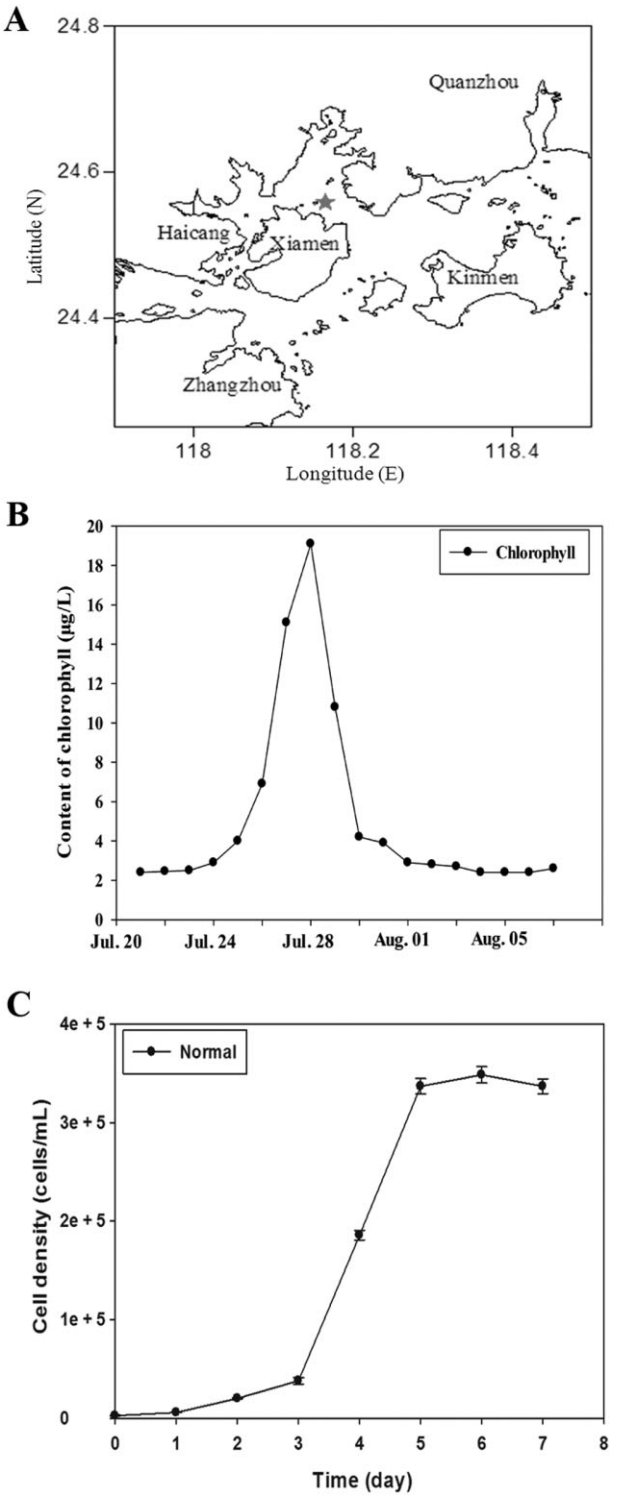


**Fig. 1.** Sampling information of the field-collected (FC) and laboratory-cultured (LC) blooming cells.
A. Field sampling location (star).
B. Variation of field chlorophyll a content during *S. costatum* blooming period (Redrawn from Yu, 2012).
C. Growth curve of *S. costatum* in laboratory culture conditions. The cells at day 4 were collected as the blooming sample for proteomic analysis.

23.51% to the diatom genus *Thalassiosira*, and 11.52% to the diatom genus *Chaetoceros*, followed by the genus *Pseudo-nitzschia* and other diatom species (Fig. S2).

### *Functional analysis of the identified proteins*

Functional categories of all the proteins identified are shown in Table S3. About 94.3% and 91.1% of the identified proteins from the FC and LC blooming samples were assigned to at least one annotation term within the GO biological process category, while the rest could not be assigned and most of them were predicted or hypothetical proteins (Table S3). Protein profiles presented high similarity between the FC and LC samples (Fig. S3). The proteins involved in photosynthesis, transport and energy metabolism, translation and nucleosome assembly were the major components not only in terms of their numbers, but also the proportions of all the peptide intensity (Fig. 2).

In the FC sample, 210 proteins involved in photosynthesis accounted for 51.08% of all the peptide intensity. Transport and energy metabolism, translation and carbohydrate metabolism-related proteins consisted of 23.25%, 8.6% and 4.91% of all peptide intensity. Interestingly, although only 13 proteins involved in nucleosome assembly were identified, they accounted for 3.35% of all peptide intensity. Other proteins, accounting for only small numbers of the total, were related to nucleobase-containing compound metabolic process (1.95%), protein metabolic process (1.16%), fatty acid biosynthesis (1.05%), oxidative phosphorylation and methylation (0.87%), stress response (0.81%), amino acid metabolism (0.79%), cell division and redox homeostasis (0.51%), phosphorus compound metabolism (0.16%), lipid metabolism (0.11%), proteolysis (0.09%) and other functions (1.19%).

In the LC sample, the proteins involved in photosynthesis (33.13%), translation (21.86%), transport and energy metabolism (19.07%), carbohydrate metabolism (4.64%) and nucleosome assembly (4.51%) were also the dominant components of all the proteins identified. Other proteins were involved in nucleobase-containing compound metabolism (4.50%), protein metabolism (2.48%), amino acid metabolism (2.38%), stress response (1.59%), oxidative phosphorylation and methylation (1.13%), proteolysis (0.36%), fatty acid biosynthesis (0.28%), phosphorus compound metabolism (0.24%), cell division and redox homeostasis (0.20%), lipid metabolism (0.13%) and other functions (3.50%).

### *Highly expressed proteins in blooming cells*

Among the high-confidence proteins, the 50 most highly expressed proteins in each sample were selected for

**Table 1.** Phytoplankton biomass and species composition of the field-collected bloom samples.

| Date | Location | *S. costatum* (cells per litre) | *A. sanguine* (cells per litre) | *Chaetoceros* sp. (cells per litre) |
|---|---|---|---|---|
| 2011/7/27 15:18 | 24°33.888 N 118°09.753 E | $6.8 \times 10^7$ | $2.0 \times 10^5$ | $1.0 \times 10^5$ |
| 2011/7/29 16:00 | 24°33.530 N 118°09.940 E | $2.7 \times 10^6$ | $4.5 \times 10^5$ | $2.0 \times 10^5$ |
| 2011/8/2 14:36 | 24°33.220 N 118°09.950 E | 0 | 0 | $1.0 \times 10^5$ |

further comparison. The top 50 abundant proteins accounted for 55.1% of the whole peptide abundance in the FC blooming cells; among them, 35 proteins were involved in photosynthesis, 10 in transport, 2 in nucleosome assembly, 2 in the nucleobase-containing compound metabolic process and 1 had an unknown function (Fig. 3A). Photosynthesis proteins were the largest proportion, whereas light utilizing proteins displayed high-frequency detection. Ten fucoxanthin chlorophyll a/c protein (FCP) homologues, three light harvest complexing (LHC) proteins, cytochrome c-550, cytochrome c-553, cytochrome b6-f complex subunit 4

**Table 2.** Proteome overview of the field-collected and laboratory-cultured blooming samples of *S. costatum*.

| Sample | Total spectra | Identified spectra | % of Identified spectra | Identified proteins (≥ 2 peptides) | Proteins (≥ 2 peptides) | Total proteins (≥ 2 peptides) | Shared proteins (≥ 2 peptides) |
|---|---|---|---|---|---|---|---|
| LC1 | 39341 | 6521 | 16.58 | 992 | | | |
| LC2 | 36746 | 5869 | 15.97 | 991 | 1061 | | |
| LC3 | 43939 | 7369 | 16.77 | 1007 | | 1372 | 839 |
| FC1 | 35225 | 3273 | 9.30 | 1091 | | | |
| FC2 | 37150 | 3473 | 9.35 | 936 | 1150 | | |
| FC3 | 44975 | 2441 | 5.43 | 822 | | | |

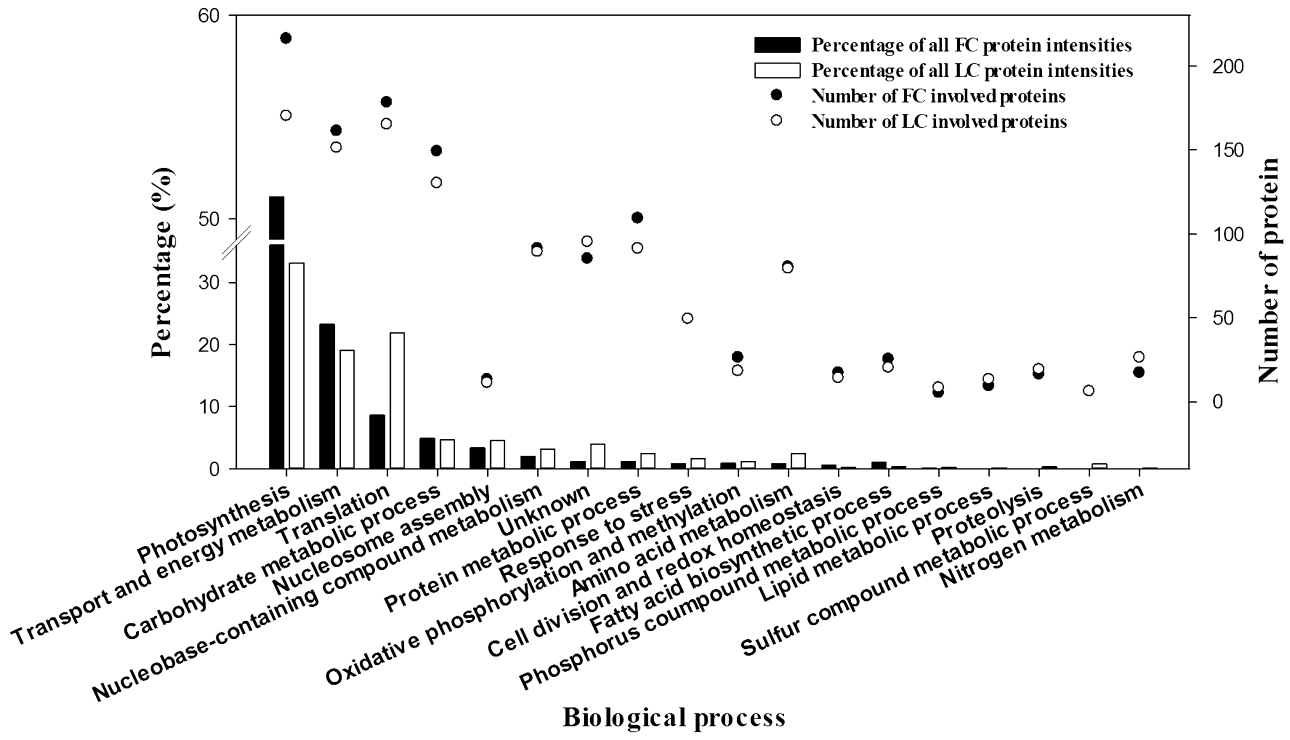


**Fig. 2.** Percentage of peptide intensity and number of proteins involved in related biological processes between the field-collected and laboratory-cultured samples.

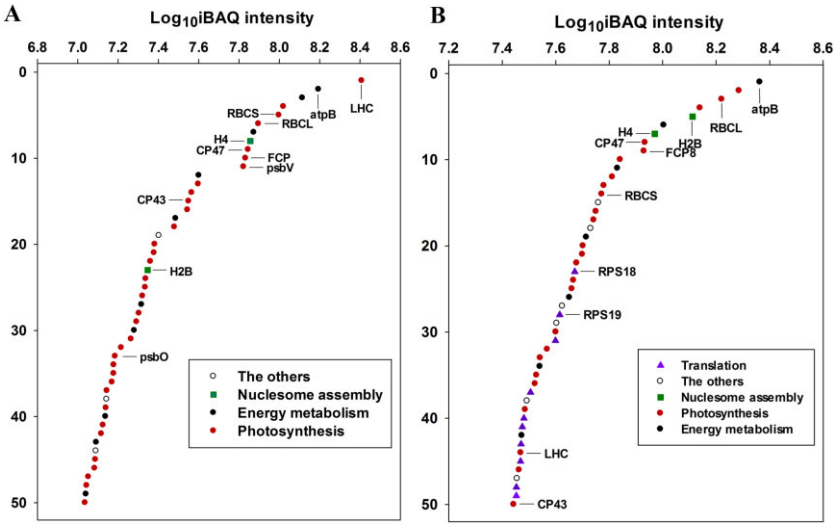


**Fig. 3.** Illustration of the 50 most abundant proteins expressed in the field-collected (A) and laboratory-cultured (B) samples. Atpb: atp synthase cf1 subunit beta; CP47: PSII cp47 chlorophyll approtein; CP43: PSII cp43 chlorophyll approtein; FCP: fucoxanthin chlorophyll a/c protein; H2A: histone 2a; H2B: histone 2b; H4: histone 4; LHC: light harvest complex protein; psbo: oxygen-evolving enhancer protein; psbv: cytochrome c550; RBCL/S: rubisco large/small subunit; RPS18: 40s ribosomal protein s18; RPS19: 40s ribosomal protein s19.

and cytochrome f were the top-ranked abundant proteins. The RuBisCO large subunit and small subunit involved in $CO_2$ assimilation were also highly detected. The other photosynthetic proteins included photosystem II (PSII) cp47 chlorophyll apoprotein, PSII cp43 chlorophyll apoprotein, PSII reaction centre protein d1, PSI reaction centre subunit II, PSII psb27 and two oxygen-evolving enhancer proteins. Ten ATP synthase subunits involved in ATP synthesis coupled proton transport were detected. Moreover, two nucleosome assembly proteins, histone 2B and histone 4 were identified.

The top 50 abundant proteins accounted for 43.7% of whole peptide abundance in the LC blooming cells and 24 of them were involved in photosynthesis (Fig. 3B). The RuBisCO large subunit and small subunit were frequently detected in photosynthetic proteins. Ten FCP homologues and three LHC proteins, cytochrome b6-f complex iron-sulfur subunit and cytochrome f were the top-ranked abundant proteins. The other six photosynthetic proteins were PSII cp47 chlorophyll apoprotein, PSI reaction centre subunit PSAE, PSI reaction centre subunit IV, extrinsic protein in PSII, PSII reaction centre protein d1, and PSII cp43 chlorophyll apoprotein. Ten ribosomal proteins (RPs) were involved in translation, eight ATP synthase subunits in ATP synthesis coupled proton transport, and two core histone proteins (histone 2B and histone 4) were detected.

Among the top 50 abundant proteins, 15 proteins were shared between the FC and LC samples, including eight photosynthesis proteins, five energy metabolism proteins, and two nucleosome assembly proteins. RuBisCO large subunit was the most abundant protein in the LC *S. costatum* cells, whereas light-harvesting complex protein was the most abundant protein in the FC *S. costatum* cells. Two histone proteins, H2B and H4, were highly expressed in both FC and LC samples, indicating their essential importance in the blooming of *S. costatum* cells.

### *Differentially expressed proteins in blooming cells*

Among the 839 shared proteins, 49 exhibited higher expressions in the FC sample, whereas 209 presented higher expressions in the LC sample (Table S4). These proteins were mainly involved in photosynthesis, translation, transport, metabolism of carbohydrate, protein and amino acid, as well as nucleobase-containing compound metabolism (Fig. 4A). LHC protein, glyceraldehyde-3-phate dehydrogenase, apocytochrome f, ATP synthase subunit, FCP protein, translation elongation factor-like protein, chaperonin and heat shock protein 90 exhibited much higher abundances in the FC sample, whereas nascent polypeptide-associated complex subunit, RPs, abc transporter substrate-binding protein, triosephosphateisomerase, ATP synthase CF1 subunit, luminal-binding protein, inositol-3-phosphate synthase, and a chain carbonic anhydrase (CA) cadmium bound domain 1 and domain 2 presented much higher abundances in the LC sample.

Besides these shared proteins, 311 proteins were identified only in the FC, and 222 only in the LC samples (Table S4). These proteins were mainly involved in photosynthesis, translation, transport, and carbohydrate, protein and amino acid metabolism (Fig. 4B). It should be pointed out that the majority of these proteins were the same proteins but matched to different diatom species. However, some specific proteins were detected in either the FC sample or the LC sample. Some light harvest proteins and pigment biosynthesis proteins, such as Mg chelatase subunit e, zeta-carotene desaturase and

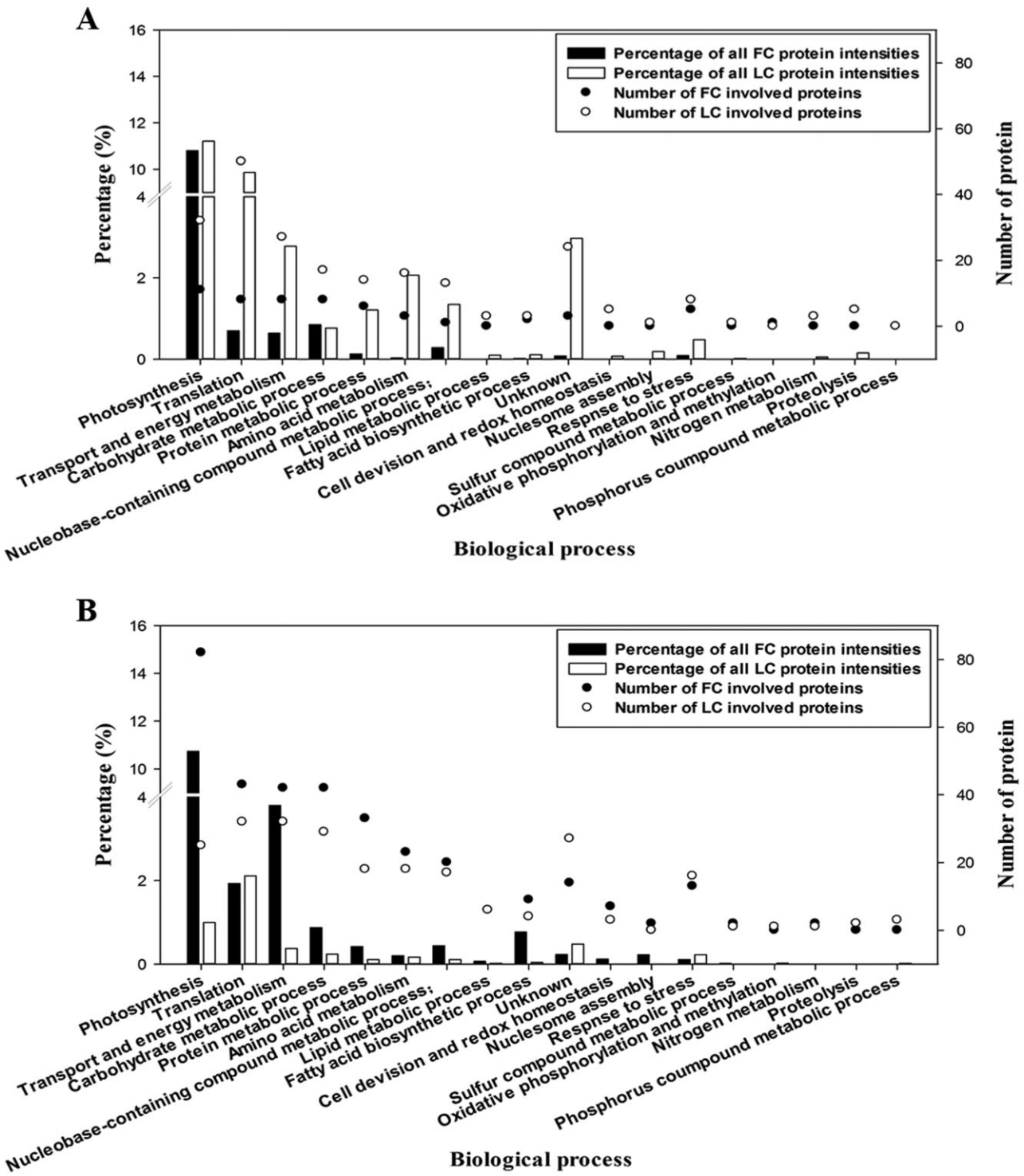


**Fig. 4.** Percentage of peptide intensity and number of proteins involved in related biological processes.
A. Highly expressed proteins in the field-collected and laboratory-cultured samples.
B. Unique expressed proteins in the field-collected and laboratory-cultured samples.

three other proteins (ras-like gtp-binding protein ypt1, DNA damage checkpoint protein rad24 and ras-related protein rab-1a) were identified only in the FC sample, whereas some RPs, amino acid and protein metabolic proteins (such as cysteine synthase, asparagine synthase, calreticulin, chaperone Dnak, and peptidylprolyl isomerase) and one ammonium transporter (AMT) protein were detected only in the LC sample.

## Discussion

The occurrence of diatom blooms is reported from the subtropical to the polar regions, and this provides us with a good chance to explore their potential ecological niche in a given aquatic environment. Some efforts have been devoted to diatom proteomics in both field and laboratory conditions in order to unveil their important physiological characteristics (Nunn *et al*., 2009; Hockin *et al*., 2012; Moore *et al*., 2012). In this study, we used the label-free shotgun proteomic approach to investigate protein expression profiles in both FC and LC blooming cells of *S. costatum* and revealed the major biological processes that occurred in the blooming cells. Proteins involved in photosynthesis, translation, nucleosome assembly, carbohydrate and energy metabolism dominated the protein profiles of both FC and LC blooming cells. However, different features of specific proteins and related biological processes were also found between the two samples, although a high similarity in the protein expression profile was observed between them.

### *Light-harvesting and photosynthetic proteins*

Diatoms rely on light to photosynthetically fix carbon dioxide into organic carbon, and so their light-harvesting ability partly determines their niche in the complex marine environment. In this study, FCPs, LHC proteins, PSII cp47 chlorophyll apoprotein and PSII cp43 chlorophyll apoprotein presented high expressions in both FC and LC blooming cells, and one LHC protein was the most abundant protein in the FC sample. In plants and microalgae, the PSII light-harvesting system consists of two distinct types of pigment-protein complexes, chlorophyll-related proteins (CP43 and CP47) and LHC proteins. The former can bind Chla and carotene for light utilization, whereas the latter can bind Chla, chlorophyll b and xanthophylls for light harvesting (Bianchi *et al*., 2010). As the most important light-harvesting antennae in light energy utilization, FCP can bind fucoxanthin, Chla and chlorophyll c to harvest light energy and transfer it equally and efficiently to the reaction centres of PSI and PSII. The gene expression of FCP increases with the enhancement of light irradiance below the saturation light flux, and its expression is regulated by a circadian clock in some species (Grossman *et al*., 1995; Passaquet and Lichtle, 1995). FCP is also the most abundant protein in *Thalassiosira pseudonana* in optimal laboratory conditions (Nunn *et al*., 2009). More LHC genes with high expressions are also reported in the blooming cells of *Aureococcus anophagefferens* in order to overwhelm other microalgae (Gobler *et al*., 2011). High expressions of FCPs, LHC proteins, PSII CP43 and CP47 in both FC and LC blooming samples enabled cells to harvest more light and subsequently provided sufficient energy for the cells to achieve maximal cell growth rate.

Light is essential for photosynthesis in phytoplankton, and the wide spectral composition and high intensity of daylight in the field must be responsible for the variations of photosynthetic protein expression (Huppertz *et al*., 1990; Hanelt, 1992). In our study, high expressions of three proteins belonging to PSII (cytochrome c-550, cytochrome b6-f complex and oxygen-evolving enhancer protein) were found in both FC and LC blooming cells. These proteins play a substantial role in maintaining the stability and functioning in algal PSII, and their lack of results in the complete loss of photosynthetic oxygen evolution (Mayfield, 1991; Shen *et al*., 1998; Suh *et al*., 2000). The cytochrome b6f complex also regulates the transfer of electrons from PSII to PSI during photosynthesis, and electron transport via cytochrome b6f creates the proton gradient that drives the synthesis of ATP in the chloroplast and it is essential for photosynthesis in maintaining the proper ratio of ATP/NADPH (nicotinamide adenine dinucleotide phosphate hydrogen) production for carbon fixation (Bendall and Manasse, 1995; Kurisu *et al*., 2003; Munekage *et al*., 2004; Shikanai, 2007). The high expression of photosynthetic proteins suggested that active photosynthesis was occurring in the blooming cells, and this provided sufficient energy for various intracellular metabolic activities.

RuBisCO is a bifunctional enzyme that catalyses both the initial carboxylation reaction in the photosynthetic carbon reduction cycle and the initial oxygenation reaction in the photorespiratory carbon oxidation cycle. RuBisCO is reported as the most abundant protein in nutrient-replete plants and microalgae (Dhingra *et al*., 2004), and our results supported this. In our study, both large and small subunits of RuBisCO were identified to be highly expressed in the FC and LC blooming cells, indicating that active photosynthetic $CO_2$ assimilation and photorespiratory carbon oxidation were occurring in the blooming cells of *S. costatum*, and so provided sufficient macromolecular compounds for sustaining rapid cell division and growth during the blooming period. The expression level of RuBisCO is regulated by the ambient $CO_2$ concentration, and some phytoplankton species can synthesize more RuBisCO at low $CO_2$ in order to support photosynthesis (Tortell, 2000). Thus, the high expression of RuBisCO in both FC and LC samples indicated that the blooming cells might be experiencing carbon stress resulting in the enhanced biosynthesis of RuBisCO.

### *Nucleosome assembly*

As the basic unit of chromatin, each nucleosome consists of about 148 bp of DNA wrapped around an octamer

containing two copies of histone 2A (H2A), H2B, H3 and H4 (Luger *et al*., 1997). Histones play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. In our study, all four core histones except H3 were identified in both FC and LC blooming cells of *S. costatum* and two of them ranked among the top 50 abundant protein profiles, indicating that a typical eukaryotic nucleosomal organization existed in the diatoms. Nucleosomes regulate cellular protein biosynthesis and the expression levels of core histones vary with the cell cycle and other conditions (Hirakawa *et al*., 2011). H2A.1 and H4 originating from diatoms dominate the protein profile not only in the water column sample, but also in an overwintered shelf sediment sample from the Bering Sea after a diatom bloom, and this may partly be due to its high abundance and specific structure (Moore *et al*., 2012). Lower expression levels of histones are reported in nitrogen-starved *Ostreococcus tauri* and they undergo posttranslational modification that affects chromatin structure, and so respond to stress efficiently (Bihan *et al*., 2011). H2A phosphorylation contributes significantly to DNA damage of a budding yeast (Chambers and Downs, 2007), and phosphorylate histone H2AX is reported to regulate the cell cycle of the reproductive cysts before their release in dinoflagellates (Litaker *et al*., 2003). In our study, high abundance of histones in the blooming cells of *S. costatum* indicated that they might play key roles in repairing the transcribed mistakes and maintaining active transcription as well as chromosomal stability in the active blooming cells grown in complex environmental conditions.

*Energy metabolism*

Several ATP synthase cf1 subunit alpha and beta proteins involved in ATP synthesis coupled proton transport ranked among the top abundant protein profiles. ATP synthase cf1 subunit, integrated into the thylakoid membrane, is an important enzyme to catalyse the synthesis of ATP and so provide energy for cells (Yoshida *et al*., 2001; Itoh *et al*., 2004). Not only taking the central role in energy transduction in the chloroplasts and mitochondria, ATP synthase also functions in the alleviation of stress. Overexpression of the ATP synthase enhances tolerance to drought in *Arabidopsis* and cotton (Zhang *et al*., 2008; Deeba *et al*., 2012). In our study, the high abundance of the ATP synthase cf1 subunit in the blooming cells of *S. costatum* revealed that high energy demands and active energy metabolism occurred within the blooming cells to support the rapid cell division. ATP synthase cf1 subunit beta is also identified in an overwintered shelf sediment sample from the Bering Sea after a diatom bloom (Moore *et al*., 2012), demonstrating the abundance and role of this protein in diatoms.

*Specific expressed proteins in the FC blooming cells*

Our study revealed that proteins involved in light-harvesting, photosynthetic pigment biosynthesis, photoprotection, cell division and redox homeostasis occupied a greater proportion of the FC blooming cells not only in terms of the protein number, but also their abundance.

In our study, more specific light harvest proteins, such as LHC proteins, and photosynthetic pigment biosynthesis proteins, i.e. zeta-carotene desaturase, were identified only in the FC cells. Not only acting as accessory molecules for light harvesting, LHC proteins also function to prevent cells from photo damage and environmental stresses (Lichtenthaler, 2007). In phytoplankton including *S. costatum*, LHC proteins play essential roles in photoprotection through multiple pathways, such as Chl singlet energy dissipation, Chl triplet quenching and scavenging of reactive oxygen species (Post and Larkum, 1993; Horton *et al*., 1996; Wu and Gao, 2009). Zeta-carotene desaturase is an essential enzyme involved in catalysing the biosynthesis of carotenoids, and its specific and high expression in the FC blooming cells indicated active carotenoid biosynthesis in the FC cells. Similar to LHC proteins, carotenoids also function as light-harvesting and photoprotection accessory molecules (Lichtenthaler, 2007). Phytoplankton can increase their carotenoid contents and change their cellular pigment composition to adapt to varying UV exposure (Post and Larkum, 1993; Wu and Gao, 2009). The high ratio of carotenoid/chlorophyll in diatoms increases their ability to absorb blue–green light, crucial for growth in the aquatic environment (Stauber and Jeffrey, 1988; Beer *et al*., 2006). Previous genetic analysis reveals that diatoms likely perceive blue and red, but not green light (Armbrust *et al*., 2004). However, the identification of zeta-carotene desaturase in the FC blooming cells indicated that diatoms might be able to utilize green light as the energy source through synthesizing more carotenoids. Overall, the high and specific expressions of LHCs and photosynthetic pigment biosynthesis proteins in the FC blooming cells suggested that FC cells suffered more light stress compared with LC cells: FC cells launched more light utilization and photoprotection mechanisms in order to fulfil the light requirement of the blooming cells for highly active intracellular biosynthesis.

Interestingly, DNA damage checkpoint protein rad24 and ras-related protein rab-1a, involved in cell cycle and redox homeostasis, were detected only in the FC cells. The rad24 activates the DNA damage checkpoint when DNA damage is detected, which then leads to cell cycle arrest (Majka and Burgers, 2003). Besides checkpoint activation, rad24 is also involved in DNA repair, the transcriptional program and cell apoptosis (Hirao *et al*., 2000;

Wu *et al*., 2000; Zhou and Elledge, 2000). Ras proteins are members of a superfamily of small GTPases that are involved in many aspects of cell growth control and are possibly connected with cell apoptosis (McCormick, 1995). The specific identification of rad24 and rab-1a in the FC blooming cells indicated that more DNA damage and cell apoptosis might be caused by harsh environmental conditions such as high light intensity, UV exposure and other abiotic stresses in the field when compared with the LC cells.

### *Specific expressed proteins in the LC blooming cells*

As with the FC blooming cells, some specific proteins involved in translation, the amino acid and protein metabolic process, and nitrogen and carbon assimilation were found in the LC blooming cells.

The ribosome, the cellular organelle responsible for protein synthesis in cells, is universal and essential for all organisms. In our study, 29 unique RPs as well as 46 highly expressed RPs were found, indicating active protein synthesis in the LC blooming cells. Apart from protein synthesis, many RPs also present various extra ribosomal functions, including DNA replication, transcription and repair, RNA splicing and modification, cell growth and proliferation, regulation of apoptosis and development, and cellular transformation (Lai and Xu, 2007; Bhavsar *et al*., 2010). We know little about RPs in the diatoms at present, but the expression of more specific RPs in the LC blooming cells suggested that RPs might play much more complex roles than protein synthesis. Moreover, 39 amino acid metabolism proteins and 47 protein metabolic process-related proteins presented higher expressions in the LC blooming cells, indicating that a more active protein metabolic process occurred in the LC cells, which maintained active cell division, apoptosis and ageing during the blooming stage.

Nitrogen is generally believed to be the main element limiting phytoplankton growth and a major selective pressure imprinted on ocean microorganism genomes (Grzymski and Dussaq, 2011). Much effort has been devoted to nitrogen metabolism within diatom cells, and the biological pathways have been proposed (Fig. 5). Diatom cells are able to actively uptake $NO_3^-$ and $NH_4^+$ from the ambient environment and assimilate inorganic N into organic molecules through the coordinating activities of assimilatory nitrate reductase, nitrite reductase, glutamine synthetase and glutamate synthase (Takabayashi *et al*., 2005). In our study, 18 proteins involved in nitrogen assimilation and metabolism were detected in the blooming cells; however, AMT, assimilating external $NH_4^+$ into cells, was identified only in the LC blooming cells, which might be caused by different ambient nitrogen conditions. The field data from this area also proved this point, in that the major nitrogen nutrient species were nitrate and ammonium in the field seawater and the concentration of dissolved inorganic nitrogen is about 58.6 μM (Yu, 2012). However, in laboratory culture conditions, nitrate was the sole nitrogen source that might induce the expression of AMT. Ammonium is the preferential nitrogen species for most phytoplankton and microalgal species because of its lower energetic costs for cells to metabolize even when nitrate is sufficient (Bloom *et al*., 1992; Kronzucker *et al*., 1999). The highest mRNA levels of AMT are detected in nitrogen-starved cells, followed by nitrate-grown cells and then ammonium-grown cells in the diatom *Cylindrotheca fusiformis* (Hildebrand, 2005). These results supported our finding that AMT was induced when ambient ammonium was lacking. It should be pointed out that urease is also reported to be a reliable physiological marker to characterize nitrogen status, and a negative relationship between urea transporter abundance and inorganic nitrogen concentration is proved in Pacific oceans and nitrogen-starved *T. pseudonana* (Lomas, 2004; Hockin *et al*., 2012; Saito *et al*., 2014). However, urease was not detected in our study, although one important enzyme, argininosuccinate synthase, involved in the urea cycle was identified, indicating that the expression of urease was inhibited by the enrichment of ambient inorganic nitrogen nutrients.

Diatoms possess highly efficient carbon concentrating mechanisms compared with two other eukaryotic marine phytoplankton groups: the coccolithophores and dinoflagellates (Reinfelder, 2011). Cadmium-specific CA is reported in *Thalassiosira weissflogii* and other diatom species (Park *et al*., 2007), and its high expression highlights the exhaustion of $CO_2$ in the ambient environment. In our study, the expressions of two CA cadmium-bound domain proteins were much higher in the LC blooming cells than in the FC blooming cells, which might be caused by the availability of ambient $CO_2$. In the laboratory, *S. costatum* cells were incubated in sealed flasks without aeration, and the blooming cells faced high $CO_2$ stress compared with the field *S. costatum* cells, which induced a high expression of cadmium-specific CA to concentrate limited $CO_2$ from the culture media for supporting cell growth.

### *Protein identification by MS/MS database search*

Shotgun proteomics is a MS-based approach to identify and characterize the entire protein complement within a complex biological mixture (Wu and MacCoss, 2002). The ultimate success of protein identification using this approach relies on protein sample complexity, target organism and accuracy of peptide mass assignment (Liu *et al*., 2007). In this study, approximately 40 000 MS/MS

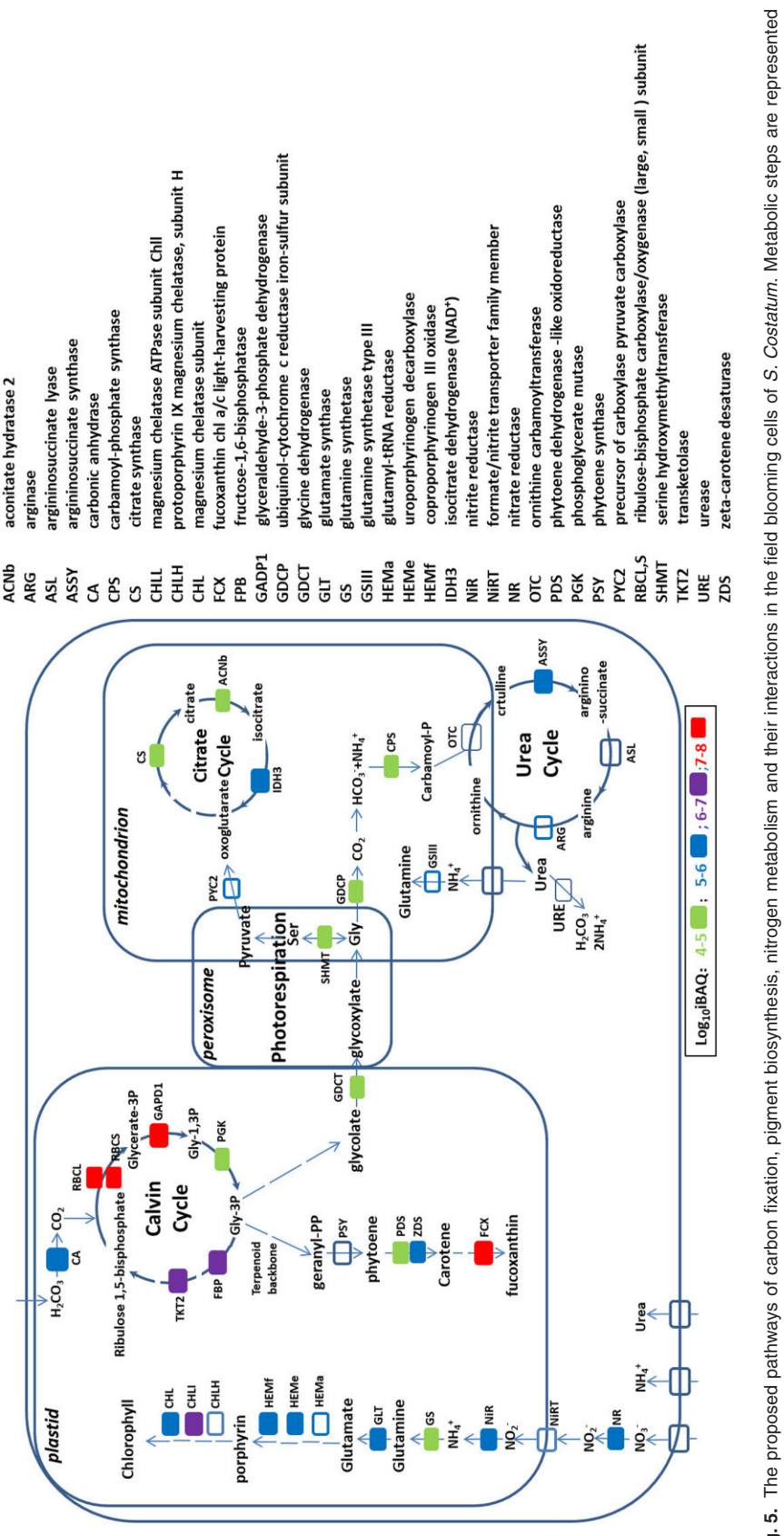


**Fig. 5.** The proposed pathways of carbon fixation, pigment biosynthesis, nitrogen metabolism and their interactions in the field blooming cells of *S. Costatum*. Metabolic steps are represented by arrows: solid arrows indicate direct steps, dashed arrows indicate multiple steps. The solid squares indicate proteins involved in the pathway, whereas blank squares indicate proteins unidentified in this study. Colours represent protein abundances based on the $Log_{10}$iBAQ value: the bigger the value, the more abundant the protein.

spectra were generated from each sample. Using the combined diatom genome and transcriptome dataset, 8.03% and 16.44% of MS/MS spectra generated from the FC and LC samples were assigned to peptide sequences. The result of the FC sample was consistent with previous studies on field marine samples (Powell *et al.*, 2005; Dong *et al.*, 2013). The lower spectra utilization in the FC sample might be caused by the complexity of the field samples that generated complicated peptides. However, lower spectra utilization in the FC sample did not influence the protein identification, and 1150 high-confidence distinct proteins were detected, which was higher than that in the LC samples (1061 proteins). Moreover, 61.2% of them were also detected in the LC samples with a high repetitive rate of 0.922 based on the triple biological repeat intensities of the FC proteome. Thus, MS/MS spectra utilization of this study was bearable despite the interference of field sample complexity.

Genetic diversity within a species will also affect protein identification because a single amino acid mutation could significantly alter the peptide mass and resulting fragmentation pattern. Genetic diversity among populations has been identified in the planktonic diatoms (Rynearson and Armbrust, 2000; Evans and Hayes, 2004; Rynearson *et al.*, 2009). However, over short-time intervals, such as within a season, an individual species may be defined by the relative proportion of a few genotypes that are best suited to the conditions present at any given time (Rynearson and Armbrust, 2000), and form blooms under distinct environmental conditions (Rynearson and Armbrust, 2006). Laboratory culture experiments also prove that the isolate with the fastest growth rate numerically dominates the population within a few days in a stable condition (Rynearson and Armbrust, 2000). Moreover, little genetic differentiation is found on a small geographic scale, suggesting that the region is occupied by a single population (Evans *et al.*, 2005). For our study, it could be postulated that one genotype of *S. costatum* suitable for the *in situ* environment might dominate the *S. costatum* populations during the blooming period; meanwhile, the subclone isolated from the simultaneous blooming sample should possess the identical genetic feature as the field genotype. Thus, differentially expressed proteins between the field and laboratory samples were most likely caused by different ambient conditions rather than by the different genotypes. Of course, we could not rule out the possibility that different genotypes might exist in the field *S. costatum* populations, and future proteomic studies should consider genetically distinct populations in the field.

It should be pointed out that although some unique proteins were identified only in the field sample (311 proteins) or only in the laboratory sample (222 proteins), a substantial amount of them were the same proteins but matched to different diatom species. This bias might be caused by the database used in this study. Because of the lack of an *S. costatum* genome, we integrated all available diatom genome and transcriptome databases into a large dataset. This integration provided more gene information for protein identification and minimized the bias towards highly conserved proteins, but the complex database also increased the potential for false-positive matches to variable residue sequences of homologous protein from different diatom species (Nesvizhskii and Aebersold, 2005). Thus, the combination of different databases might not be an ideal choice for protein identification. To overcome this bias, genomic and/or transcriptomic data derived from the same sample can be used as the reference (Verberkmoes *et al.*, 2009; Williams and Cavicchioli, 2014). In addition, multiple protein isoforms existing in the cells might also cause this bias. Two systems of molecular mechanism are responsible for protein isoform diversity: one gene among the members of a multigene family is expressed in a particular cell, developmental stage or physiological condition; and different isoforms are generated from a single gene that is caused by DNA rearrangement and alternative RNA splicing (Breitbart *et al.*, 1987). Different roles of protein isoforms are revealed within cellular metabolic activities in eukaryotes (Bark *et al.*, 1995; Barnier *et al.*, 1995; Wetering *et al.*, 2014). In our study, the same proteins from different diatom species might have been different protein isoforms, indicating that *S. costatum* might have evolved comprehensive adaptive mechanisms to environmental changes. Thus, future work should be devoted to condition-specific experiments to reveal the roles of these specific proteins.

### *Conclusions*

This study, for the first time, applied the MS-based proteomic approach to investigate the global protein expression profiles in both FC and LC blooming cells of *S. costatum*. Hundreds of proteins were identified using the combined diatom genome and transcriptome database, gaining a new insight into the cellular biological processes occurring in the blooming cells. The proposed major biological processes related to carbon fixation and metabolism, nitrogen assimilation and their interactions in the blooming cells are shown in Fig. 5. Proteins involved in photosynthesis, transport and energy metabolism, translation and nucleosome assembly presented a high detection ratio, indicating their importance in the blooming cells. However, different features of specific proteins were also found in both FC and LC blooming cells, indicating that *S. costatum* had evolved adaptive mechanisms to the changing environment, which might explain their dominant status in taking the niche in the harsh and variable

marine environment. Besides this, a typical eukaryotic nucleosomal organization and protein biosynthesis apparatus existed in *S. costatum* cells through the core histones and all RPs, and the high expressions of these proteins revealed their essential importance in regulating cell division and proliferation during the blooming period.

A comparison of protein expression profiles between the LC and FC blooming cells revealed that the protein expressions and biological processes between the two samples presented some differences, although the majority were similar. Therefore, integrating proteomic data of both field and laboratory samples should help us to unveil the molecular mechanisms involved in the regulation of diatom blooms. The shotgun proteomic approach provides a powerful tool to investigate the *in situ* physiological and metabolic status of marine diatoms in complex aquatic environments. It also lays the background for future comprehensive proteomic studies of blooming cells in order to identify the specific expressed proteins and biological processes involved in the phytoplankton blooming mechanism in both the coastal seas and open oceans.

## Experimental procedures

### *Field-blooming sample collection*

In 2011, a phytoplankton bloom occurred in Xiamen Bay, China, (24°33.530′N, 118°09.940′E) from July 26 to August 2 (Fig. 1A). During the bloom period, surface seawater was collected every day for monitoring the phytoplankton species and cell density as well as for proteomic analysis. To analyse the phytoplankton species composition and cell density, three 1 ml of surface seawater samples were collected and fixed with Lugol's solution for subsequent microscopic examination. For proteomic analysis, three 2 l of surface seawater samples were collected and filtered on-site immediately with a 10 μm of nylon net, washed twice with autoclaved seawater, and then centrifuged (3000*g* for 10 min) at room temperature and stored at −80°C until proteomic analysis. Physico-chemical parameters of the seawater including temperature, salinity, inorganic nitrogen and phosphorus, as well as the *in situ* solar irradiance on July 27 (blooming stage) were also monitored (Table S5).

### *Laboratory culture of* S. costatum

A single chain containing more than 50 cells of *S. costatum* was isolated from the blooming stage on July 27 and was routinely maintained in f/2 medium at 20°C under a 14:10 h light : dark photoperiod at a light intensity of approximately 100 μmol photons $m^{-2}$ $s^{-1}$ provided by fluorescent lamps. For laboratory culture, vegetative cells in the mid-exponential growth phase were inoculated into freshly prepared f/2 medium at an initial concentration of 8400 cells $ml^{-1}$. When the cells entered the exponential growth phase, approximately $10^{10}$ vegetative cells were collected as the LC blooming sample using centrifugation at 3000*g* for 10 min at room temperature. The pellets were subsequently transferred to 1.5 ml of microcentrifuge tubes, rinsed twice with sterile seawater and then stored at −80°C for subsequent proteomic analysis.

### *Protein extraction*

Protein extraction was performed following the procedure described by Wang and colleagues (2011). Briefly, the cell pellet was sonicated on ice with 1 ml of Trizol reagent. Subsequently, 200 μl of chloroform was added to the cell lysate, and after being vortexed for 15 s, the mixture was held at room temperature for 5 min, and then centrifuged at 12 000*g* for 15 min at 4°C. After removing the top pale-yellow or colourless layer, 300 μl of ethanol was added to re-suspend the reddish bottom layer. The mixture was vortexed and then centrifuged at 2000*g* for 5 min at 4°C, and then the supernatant was transferred to a new tube and 1 ml of isopropanol added. The mixture was stored at −20°C for at least 2 h for protein precipitation, then centrifuged at 14 000*g* for 30 min at 4°C. After washing with 1 ml of 95% ethanol, the pellet obtained was dissolved in 30–50 μl of rehydration buffer (7M urea, 2M thiourea, 4% w/v CHAPS). Protein quantification was performed using a 2D Quant kit (GE Healthcare, USA).

### *Protein separation and digestion*

Proteins (150 μg) were applied to each lane of a 5–12% Bis–Tris gel (13 cm × 13 cm). Electrophoresis was performed in 12% separation gel at a constant voltage of 20 V, then at 40 V in a 5% stacking gel in the electrode buffer solution (25 mM Tris, 192 mM glycine and 0.1% SDS) on a Hoefer™ SE 600 apparatus (Amersham). Pre-stained protein molecular weight standards were used as the reference. After electrophoresis, the proteins on the gel were visualized with colloidal Coomassie Brilliant Blue G-250 (Bio-Rad). The gel was cut into seven pieces to isolate molecular weight fractions, followed by reduction, alkylation and in-gel digestion with trypsin as described previously (Wilm *et al*., 1996). After trypsin digestion, each slice was separately analysed using an LTQ Orbitrap mass spectrometer, and the peak spectral list files were combined in order to search for the proteins in the data set.

### *Liquid chromatography-MS/MS analysis*

Nano-flow high-performance liquid chromatography coupled with an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific) was used to separate and identify the peptides. Each sample was repeated twice with triplicates of each condition. The peptides were extracted from the seven gel pieces separately with 60% acetonitrile and 0.1% trifluoroacetic acid. The extracts were first dried, re-dissolved in 20 μl of 0.1% formic acid and injected onto a peptide trap (CapTrap, Bruker), and then desalted with 0.1% formic acid at a flow rate of 20 μl $min^{-1}$ for 4 min. Peptides were eluted from the trap and separated on a reverse phase C18 column (0.075 mm × 150 mm, Column Technology) with a 60 min linear gradient from 5% to 45% buffer B (90% acetonitrile, 0.1% formic acid) in buffer A (0.1% formic acid) at 500 μl $min^{-1}$. After the gradient, the column was washed with 90% buffer B and re-equilibrated with buffer A.

Mass spectra were acquired in a data-dependent mode, with an automatic switch between MS and MS/MS scans using a top 10 method. The LTQ mass spectrometer operated in the data-dependent mode with the following parameters: spray voltage 1.8 kV, spray temperature 180°C, full scan m/z range 400–2000 and a target value of $10^6$ ions. Peptide fragmentation was performed using the higher-energy C-trap dissociation method with the target value of 40 000 ions. The ion selection threshold was set to 5000 counts. The whole liquid chromatography-MS system was fully automated and under the direct control of an Xcalibur software system (Thermo Finnigan).

### *Data analysis*

Raw MS files were analysed using MaxQuant version 1.2.0.28, and the MS/MS spectra were searched using the Andromeda search engine against the Bacillariophyta genome and transcriptome database. The database for this study was combined by downloading data from the National Center for Biotechnology Information, Joint Genome Institute and Marine Microbial Eukaryote Transcriptome Sequencing Project websites on 10 March 2015. Details of the combined data set are shown in Table S6. An initial search with a precursor mass tolerance of 20 ppm. was completed for mass recalibration. The mass tolerances for precursor mass and fragment mass were 6 and 20 ppm. in the main Andromeda search. Variable modification of cysteine residues and methionine oxidation was selected in the search. Minimal peptide length was set to six amino acids and a maximum of two miscleavages was allowed. The false discovery rate was set to 0.01 for peptide and protein identification. Proteins matching two or more peptides were selected for further discussion in our study. For classification, proteins of all the identified peptides shared between two proteins were combined and reported as one protein group. After comparison, the redundant proteins in a group were removed as previously described, and so only one protein remained in a group (He *et al*., 2005). For further analysis, only the proteins that remained in a group were considered and calculated.

### *Bioinformatic analysis*

The matched peptide sequences were annotated using Blast 2GO. Categorical annotation was supplied in the form of the GO biological process, molecular function and cellular component, as well as its participation in a KEGG pathway. The iBAQ algorithm was used to rank the absolute abundance of different proteins within a single sample (Schwanhäusser *et al*., 2011; Geiger *et al*., 2012). The average values of triplicates of each sample were calculated and the estimated absolute abundances of all identified proteins were ranked. Label-free quantification was used to determine the relative amount of proteins and compare the differentially expressed proteins between FC and LC samples. Abundance of each protein was normalized by calculating its iBAQ proportion in all protein intensities within one sample before comparison between LC and FC samples. Proteins with at least two identifications were considered for further analysis. A *t*-test was applied to compare the variation of each protein after normalization, and only proteins with fold change $\geq 4$ and $P$ value $\leq 0.01$ were selected as significantly varied between the two samples.

## Acknowledgements

This work was partially supported by research grants from the National Natural Science Foundation of China (Project No. 41230961 and 41425021), the Ministry of Science and Technology of the People's Republic of China (Project No. 2010CB428703). D.-Z. Wang was also supported by the 'Ten Thousand Talents Program' for leading talents in science and technological innovation. We thank Mr. Yue Gao for his assistance during field sample collection, and Professor John Hodgkiss from The University of Hong Kong for his help with English.

## References

Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422:** 198–207.

Allen, A.E., Vardi, A., and Bowler, C. (2006) An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Curr Opin Plant Biol* **9:** 264–273.

Armbrust, E., Berges, J.A., Bowler, C., Green, B.R., Martinez, D., Putnam, N.H., *et al*. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306:** 79–86.

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

Bark, I.C., Hahn, K.M., Ryabinin, A.E., and Wilson, M.C. (1995) Differential expression of SNAP-25 protein isoforms during divergent vesicle fusion events of neural development. *Proc Natl Acad Sci USA* **92:** 1510–1514.

Barnier, J.V., Papin, C., Eychene, A., Lecoq, O., and Calothy, G. (1995) The mouse B-raf gene encodes multiple protein isoforms with tissue-specific expression. *J Biol Chem* **270:** 23381–23389.

Beer, A., Gundermann, K., Beckmann, J., and Büchel, C. (2006) Subunit composition and pigmentation of fucoxanthin-chlorophyll proteins in diatoms: evidence for a subunit involved in diadinoxanthin and diatoxanthin binding. *Biochemistry* **45:** 13046–13053.

Bendall, D.S., and Manasse, R.S. (1995) Cyclic photophosphorylation and electron transport. *Biochim Biophys Acta* **1229:** 23–38.

Bhavsar, R.B., Makley, L.N., and Tsonis, P.A. (2010) The other lives of ribosomal proteins. *Hum Genomics* **4:** 327–344.

Bianchi, S.D., Ballottari, M., Dall'Osto, L., and Bassi, R. (2010) Regulation of plant light harvesting by thermal dissipation of excess energy. *Biochem Soc Trans* **38:** 651–660.

Bihan, T.L., Martin, S.F., Chirnside, E.S., Ooijen, G.V., Barrios-Llerena, M.E., O'Neill, J.S., *et al*. (2011) Shotgun proteomic analysis of the unicellular alga *Ostreococcus tauri*. *J Proteomics* **74:** 2060–2070.

Bloom, A.J., Sukrapanna, S.S., and Warner, R.L. (1992) Root respiration associated with ammonium and nitrate absorption and assimilation by barley. *Plant Physiol* **99:** 1294–1301.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., *et al*. (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456:** 239–244.

Breitbart, R.E., Andreadis, A., and Nadal-Ginard, B. (1987) Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu Rev Biochem* **56:** 467–495.

Buesseler, K.O. (1998) The decoupling of production and particulate export in the surface ocean. *Global Biogeochem Cycle* **12:** 297–310.

Chambers, A., and Downs, J. (2007) The contribution of the budding yeast histone H2A C-terminal tail to DNA-damage responses. *Biochem Soc Trans* **35:** 1519–1524.

Deeba, F., Pandey, A.K., Ranjan, S., Mishra, A., Singh, R., Sharma, Y.K., *et al*. (2012) Physiological and proteomic responses of cotton (*Gossypoum herbaceum* L.) to drought stress. *Plant Physiol Biochem* **53:** 6–18.

Dhingra, A., Portis, A.R., Jr, and Daniell, H. (2004) Enhanced translation of a chloroplast-expressed RbcS gene restores small subunit levels and photosynthesis in nuclear RbcS antisense plants. *Proc Natl Acad Sci USA* **101:** 6315–6320.

Dong, H.P., Wang, D.Z., Xie, Z., X., Dai, M.H., and Hong, H.S. (2013) Metaproteomic characterization of high molecular weight dissolved organic matter in surface seawaters in the South China Sea. *Geochim Cosmochim Acta* **109:** 51–61.

Dyhrman, S.T., Jenkins, B.D., Rynearson, T.A., Saito, M.A., Mercier, M.L., Alexander, H., *et al*. (2012) The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS ONE* **7:** e33768.

Evans, K.M., and Hayes, P.K. (2004) Microsatellite markers for the cosmopolitan marine diatom *Pseudo-nitzschia pungens*. *Mol Ecol Notes* **4:** 125–126.

Evans, K.M., Kühn, S.F., and Hayes, P.K. (2005) High levels of genetic diversity and low levels of genetic differentiation in North Sea *Pseudo-nitzschia punges* (bacillariophyceae) populations. *J Phycol* **41:** 506–514.

Falkowski, P.G., Barber, R.T., and Smetacek, V. (1998) Biogeochemical controls and feedbacks on ocean primary production. *Science* **281:** 200–206.

Ferris, J.A., and Lehman, J.T. (2007) Interannual variation in diatom bloom dynamics: roles of hydrology, nutrient limitation, sinking, and whole lake manipulation. *Water Res* **41:** 2551–2562.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281:** 237–240.

Gao, Y., Smith, G.J., and Alberte, R.S. (1993) Nitrate reductase from the marine diatom *Skeletonema costatum* (biochemical and immunological characterization). *Plant Physiol* **103:** 1437–1445.

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol Cell Proteomics* **11:** M111.014050.

Gobler, C.J., Berry, D.L., Dyhrman, S.T., Wilhelm, S.W., Salmov, A., Lobanov, A.V., *et al*. (2011) Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Natl Acad Sci USA* **108:** 4352–4357.

Grossman, A.R., Bhaya, D., Apt, K.E., and Kehoe, D.M. (1995) Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annu Rev Genet* **29:** 231–288.

Grzymski, J.J., and Dussaq, A.M. (2011) The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J* **6:** 71–80.

Guazzaroni, M.E., Herbst, F.A., Lores, I., Tamames, J., Peláez, A.I., López-Cortés, N., *et al*. (2012) Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation. *ISME J* **7:** 122–136.

Hanelt, D. (1992) Photoinhibition of photosynthesis in marine macrophytes of the South China Sea. *Mar Ecol Prog Ser* **82:** 199–206.

He, P., He, H.Z., Dai, J., Wang, Y., Sheng, Q.H., Zhou, L.P., *et al*. (2005) The human plasma proteome: analysis of Chinese serum using shotgun strategy. *Proteomics* **5:** 3442–3453.

Hildebrand, M. (2005) Cloning and functional characterization of ammonium transporters from the marine diatom *Cylindrotheca fusiformis* (Bacillariophyceae). *J Phycol* **41:** 105–113.

Hirakawa, Y., Burki, F., and Keeling, P.J. (2011) Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. *Mol Microbiol* **80:** 1439–1449.

Hirao, A., Kong, Y.Y., Matsuoka, S., Wakeham, A., Ruland, J., Yoshida, H., *et al*. (2000) DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* **287:** 1824–1827.

Hockin, N.L., Mock, T., Mulholland, F., Kopriva, S., and Malin, G. (2012) The response of diatom central carbon metabolism to nitrogen starvation is different from that of green algae and higher plants. *Plant Physiol* **158:** 299–312.

Horton, P., Ruban, A.V., and Walters, R.G. (1996) Regulation of light harvesting in green plants. *Annu Rev Plant Physiol Plant Mol Biol* **47:** 665–684.

Hu, H., Zhang, J., and Chen, W. (2011) Competition of bloom-forming marine phytoplankton at low nutrient concentrations. *J Environ Sci (China)* **23:** 656–663.

Huppertz, K., Hanelt, D., and Nultsch, W. (1990) Photoinhibition of photosynthesis in the marine brown alga *Fucusserratus* as studied in field experiments. *Mar Ecol Prog Ser* **66:** 175–182.

Itoh, H., Takahashi, A., Adachi, K., Noji, H., and Yasuda, R. (2004) Mechanically driven ATP synthesis by F1-ATPase. *Nature* **427:** 465–468.

Jones, B.M., Edwards, R.J., Skipp, P.J., O'Connor, C.D., and Iglesias-Rodriguez, M.D. (2011) Shotgun proteomic analysis of *Emiliania huxleyi*, a marine phytoplankton species of major biogeochemical importance. *Mar Biotechnol (NY)* **13:** 496–504.

Joseph, T., Shaiju, P., Laluraj, C., Balachandran, K., Nair, M., George, R., *et al*. (2008) Nutrient environment of red tide-infested waters off south-west coast of India. *Environ Monit Assess* **143:** 355–361.

Kronzucker, H.J., Siddiqi, M.Y., Glass, A.D.M., and Kirk, G.J.D. (1999) Inhibition of nitrate uptake by ammonium in barley: analysis of component fluxes. *Plant Physiol* **120:** 283–292.

Kurisu, G., Zhang, H., Smith, J.L., and Cramer, W.A. (2003) Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science* **302:** 1009–1014.

Lai, M.D., and Xu, J. (2007) Ribosomal proteins and colorectal cancer. *Curr Genomics* **8:** 43–49.

Lichtenthaler, H.K. (2007) Biosynthesis, accumulation and emission of carotenoids, α-tocopherol, plastoquinone, and isoprene in leaves under high photosynthetic irradiance. *Photosynth Res* **92:** 163–179.

Litaker, R.W., Vandersea, M.W., Kibler, S.R., Madden, V.J., Noga, E.J., and Tester, P.A. (2003) Life cycle of the heterotrophic dinoflagellate *Pfiesteria piscicida* (Dinophyceae). *J Phycol* **38:** 442–463.

Liu, T., Belov, M.E., Jaitly, N., Qian, W.J., and Smoth, R.D. (2007) Accurate mass measurements in proteomics. *Chem Rev* **107:** 3621–3653.

Lomas, M. (2004) Nitrate reductase and urease enzyme activity in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae): interactions among nitrogen substrates. *Mar Biol* **144:** 37–44.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251–260.

McCormick, F. (1995) Ras-related proteins in signal transduction and growth control. *Mol Reprod Dev* **42:** 500–506.

Majka, J., and Burgers, P.M.J. (2003) Yeast rad 17/mec3/dac1: a sliding clamp for the DNA damage checkpoint. *Proc Natl Acad Sci USA* **100:** 2249–2254.

Mayfield, S.P. (1991) Over-expression of the oxygen-evolving enhancer 1 protein and its consequences on photosystem II accumulation. *Planta* **185:** 105–110.

Moore, E.K., Nunn, B.L., Goodlett, D.R., and Harvey, H.R. (2012) Identifying and tracking proteins through the marine water column: insights into the inputs and preservation mechanisms of protein in sediments. *Geochim Cosmochim Acta* **83:** 324–359.

Morel, F., and Price, N. (2003) The biogeochemical cycles of trace metals in the oceans. *Science* **300:** 944–947.

Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K.-I., Endo, T., Tasaka, M., and Shikanai, T. (2004) Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature* **429:** 579–582.

Nandakumar, K., Obika, H., Shinozaki, T., Ooie, T., Utsumi, A., and Yano, T. (2003) Pulsed laser irradiation impact on two marine diatoms *Skeletonema costatum* and *Chaetoceros gracilis*. *Water Res* **37:** 2311–2316.

Nesvizhskii, A.I., and Aebersold, A. (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **4:** 1419–1440.

Nunn, B.L., Aker, J.R., Shaffer, S.A., Tsai, S., Strzepek, R.F., Boyd, P.W., *et al*. (2009) Deciphering diatom biochemical pathways via whole-cell proteomics. *Aquat Microb Ecol* **55:** 241–253.

Park, H., Song, B., and Morel, F.M. (2007) Diversity of the cadmium-containing carbonic anhydrase in marine diatoms and natural waters. *Environ Microbiol* **9:** 403–413.

Passaquet, C., and Lichtle, C. (1995) Molecular study of a light-harvesting apoprotein of *Giraudyopsis stellifer* (Chrysophyceae). *Plant Mol Biol* **29:** 135–148.

Post, A., and Larkum, A. (1993) UV-absorbing pigments, photosynthesis and UV exposure in Antarctica: comparison of terrestrial and marine algae. *Aquat Bot* **45:** 231–243.

Powell, M.J., Sutton, J.N., Del Castillo, C.E., and Timperman, A.I. (2005) Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. *Mar Chem* **95:** 183–198.

Ram, R.J., VerBerkmoes, N.C., Thelen, M.P., Tyson, G.W., Baker, B.J., Blak, R.C., *et al*. (2005) Community proteomics of a natural microbial biofilm. *Science* **308:** 1915–1920.

Reinfelder, J.R. (2011) Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Ann Rev Mar Sci* **3:** 291–315.

Robertson, D.L., and Alberte, R.S. (1996) Isolation and characterization of glutamine synthetase from the marine diatom *Skeletonema costatum*. *Plant Physiol* **111:** 1169–1175.

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

Rynearson, T.A., and Armbrust, E.V. (2006) Succession, bloom development and genetic variation in *Ditylum brightwellii* during the spring growth season. *Limnol Oceanogr* **51:** 1249–1261.

Rynearson, T.A., Lin, E.O., and Armbrust, E.V. (2009) Metapopulation structure in the planktonic diatom *Ditylum brightwellii* (Bacillariophyceae). *Protist* **160:** 111–121.

Saito, M.A., McIlvin, M.R., Moran, D.M., Goepfert, T.J., DiTullio, G.R., Post, A.F., and Lamborg, C.H. (2014) Multiple nutrient stresses at intersecting pacific ocean biomes detected by protein biomarkers. *Science* **345:** 1173–1177.

Sarthou, G., Timmermans, K.R., Blain, S., and Tréguer, P. (2005) Growth physiology and fate of diatoms in the ocean: a review. *J Sea Res* **53:** 25–42.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., *et al*. (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337–342.

Shen, J.R., Qian, M., Inoue, Y., and Burnap, R.L. (1998) Functional characterization of *Synechocystis* sp. PCC 6803 Δ psb U and Δ psb V mutants reveals important roles of cytochrome c-550 in cyanobacterial oxygen evolution. *Biochemistry* **37:** 1551–1558.

Shikanai, T. (2007) Cyclic electron transport around photosystem I: genetic approaches. *Annu Rev Plant Biol* **58:** 199–217.

Skipp, P., Robinson, J., O'Connor, C.D., and Clarke, I.N. (2005) Shotgun proteomic analysis of *Chlamydia trachomatis*. *Proteomics* **5:** 1558–1573.

Stauber, J.L., and Jeffrey, S. (1988) Photosynthetic pigments in fifty-one species of marine diatoms. *J Phycol* **24:** 158–172.

Suh, H.J., Kim, C.S., and Jung, J. (2000) Cytochrome b6/f complex as an indigenous photodynamic generator of singlet oxygen in thylakoid membranes. *Photochem Photobiol* **71:** 103–109.

Takabayashi, M., Wilkerson, F.P., and Robertson, D. (2005) Response of Glutamine synthetase gene transcription and

enzyme activity to external nitrogen sources in the diatom *Skeletonema costatum* (Bacillariophyceae). *J Phycol* **41:** 84–94.
Tortell, P.D. (2000) Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnol Oceanogr* **45:** 744–750.
Vårum, K.M., Østgaard, K., and Grimsrud, K. (1986) Diurnal rhythms in carbohydrate metabolism of the marine diatom *Skeletonema costatum* (Grev.) Cleve. *J Exp Mar Bio Ecol* **102:** 249–256.
Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarsonm, J., *et al*. (2009) Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3:** 179–189.
Wang, D.Z., Li, C., Xie, Z.X., Dong, H.P., Lin, L., and Hong, H.S. (2011) Homology-driven proteomics of dinoflagellates with unsequenced genomes using MALDI-TOF/TOF and automated de novo sequencing. *Evid Based Complement Alternat Med* doi:10.1155/2011/471020.
Wang, D.Z., Xie, Z.X., and Zhang, S.F. (2014) Marine metaproteomics: current status and future directions. *J Proteomics* **97:** 27–35.
Washburn, M.P., Wolters, D., and Yates, J.R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19:** 242–247.
Wetering, M.V.D., Castrop, J., Korinek, V., and Clevers, H. (2014) Extensive alternative splicing and dual promoter usage generate Tcf-1 protein isoforms with differential transcription control properties. *Mol Cell Biol* **16:** 745–752.
Williams, M.A., Taylor, E.B., and Mula, H.P. (2010) Metaproteomic characterization of a soil microbial community following carbon amendment. *Soil Biol Biochem* **42:** 1148–1156.
Williams, T.J., and Cavicchioli, R. (2014) Marine metaproteomics: deciphering the microbial metabolic food web. *Trends Microbiol* **22:** 248–260.
Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., *et al*. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379:** 466–469.
Wilmes, P., Wexler, M., and Bond, P.L. (2008) Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS ONE* **3:** e1778.
Wu, C.C., and MacCoss, M.J. (2002) Shotgun proteomics: tools for the analysis of complex biological systems. *Curr Opin Mol Ther* **4:** 242–250.
Wu, H., and Gao, K. (2009) Responses of a marine red tide alga *Skeletonema costatum* (Bacillariophyceae) to long-term UV radiation exposures. *J Photochem Photobiol B* **94:** 82–86.
Wu, X.H., Ranganathan, V., Weisman, D.S., Heine, W.F., Ciccone, D.N., O'Neill, T.B., *et al*. (2000) ATM phosphorylation of Nijmegen breakage syndrome protein is required in a DNA damage response. *Nature* **405:** 477–482.
Yin, K. (2003) Influence of monsoons and oceanographic processes on red tides in Hong Kong waters. *Mar Ecol Prog Ser* **262:** 27–41.
Yoshida, M., Muneyuki, E., and Hisabori, T. (2001) ATP synthase – a marvelous rotary engine of the cell. *Nat Rev Mol Cell Biol* **2:** 669–677.
Yu, X.X. (2012) A triphase red tide in Tongan Bay of Xiamen seas and its causes analysis. *J Fujian Fish* **34:** 203–207.
Zhang, X., Liu, S., and Takano, T. (2008) Overexpression of a mitochondrial ATP synthase small subunit gene (AtMtATP6) confers tolerance to several abiotic stresses in *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. *Biotechnol Lett* **30:** 1289–1294.
Zhou, B.B.S., and Elledge, S.J. (2000) The DNA damage response: putting checkpoints in perspective. *Nature* **408:** 433–439.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Fig. S1.** One-dimensional SDS-PAGE electrophoretogram of the field-collected and laboratory-cultured samples. Lane M is a molecular weight standard. Lanes X1, X2 and X3 represent three biological repeats of the field blooming sample. Lanes A1, A2 and A3 represent the laboratory blooming sample. The gel was stained with CCB and cut into seven pieces.
**Fig. S2.** Taxonomic distribution of all proteins identified in the field-collected (A) and laboratory-cultured (B) blooming samples using the Bacillariophyta genome and transcriptome database. Number and proportion of proteins belonging to each genus are shown.
**Fig. S3.** Relative label-free quantification correlation based on normalized protein intensities between the field-collected and laboratory-cultured blooming samples.
**Table S1.** Proteins identified during triplicate analysis of *S. costatum* whole cells from the laboratory-cultured and field-collected blooming samples.
**Table S2.** Peptides identified in the triplicate analysis of *S. costatum* whole cells from the laboratory-cultured and field-collected samples.
**Table S3.** Functional analysis of all proteins identified from the field-collected and laboratory-cultured samples.
**Table S4.** Significantly varied proteins between the field-collected and laboratory-cultured samples.
**Table S5.** Physico-chemical parameters of the field seawater on July 27 under field and laboratory culture conditions.
**Table S6.** Sequences of the combined database used in this study.