GEOMAR
Helmholtz Centre for Ocean Research Kiel

Acclimation and adaptation to low-iron conditions in the marine diatoms *Phaeodactylum tricornutum* and *Thalassiosira oceanica*

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Markus Lommer

Kiel, im Mai 2012

Referentin: Prof. Dr. Julie LaRoche

Korreferent: Prof. Dr. Philip Rosenstiel

Tag der mündlichen Prüfung: 28. Juni 2012

Zum Druck genehmigt: 28. Juni 2012

gez. Der Dekan

Table of Contents

1 Summary

1.1 Abstract

In the open ocean phytoplankton growth is widely limited by the availability of iron, an essential element of the photosynthetic electron transport system. Survival under these conditions requires sophisticated strategies to maintain growth, e.g. lowering iron requirements and enhancing cellular iron affinity. In this work we used genomic and transcriptomic data to unravel acclimation (transcriptomics) and adaptation (genomics) to low-iron conditions in the two diatoms *Thalassiosira oceanica* and *Phaeodactylum tricornutum*, who both are highly tolerant to iron limitation.

The ***acclimation*** response to low ambient iron concentrations is very similar in the two diatoms. Both undergo an extensive cellular retrenchment, best visible from chloroplast reduction and from the concomitant pigment loss that imposes a chlorotic phenotype on the cells. Growth rates are very low with a high degree of photosynthetic energy dissipation. The differential regulation of the genes for ferredoxin (PETF) and flavodoxin (FLDA) and for class II and class I fructose-bisphosphate aldolases (FBA) indicates that cellular iron requirements are lowered by replacing abundant iron-rich proteins with iron-free substitutes.

Genetically fixed features of *T. oceanica* and *P. tricornutum* that may represent beneficial ***adaptations*** to low-iron are the small cell sizes found for these species and the possession of class I FBA genes not present in the coastal diatom species *Thalassiosira pseudonana*. As adaptations specific to *T. oceanica* we observe the constitutive expression of the plastocyanin gene (PCY) that encodes for a substitute of iron-containing cytochrome c_6. The ferredoxin gene (PETF) has been transferred from the chloroplast to the nuclear genome in this species, facilitating its co-regulation with the flavodoxin gene (FLDA).

The acclimation of diatoms to low-iron resembles the processes running off in green plants and algae upon low-iron stress. Thus, the response to low-iron represents an ancient cellular mechanism common to most, if not all photosynthetic groups. Genetic adaptation to a persistent shortage of iron such as in the open ocean likely occurs by exploring strategies to optionally or permanently replace abundant iron-rich proteins by iron-free substitutes, thereby approximating the cellular iron requirements to a lowest possible level. The adaptive significance of substitution strategies is strengthened by their absence in coastal diatoms like *T. pseudonana* which is adapted to iron-rich coastal waters and is highly sensitive to low ambient iron concentrations.

1.2 Zusammenfassung

Im offenen Ozean ist das Phytoplankton-Wachstum häufig limitiert durch die Verfügbarkeit von Eisen, einem wesentlichen Bestandteil der photosynthetischen Elektronen-Transportkette. Damit Wachstum und Vermehrung aufrechterhalten werden können, muss der zelluläre Eisenbedarf gesenkt bzw. das Binde- und Aufnahmevermögens für Eisen erhöht werden. In der vorliegenden Arbeit werden Genom- und Transkriptomdaten genutzt, um Akklimatisierung (Transkriptomics) sowie Adaptation (Genomics) der Kieselalgen *Thalassiosira oceanica* und *Phaeodactylum tricornutum* an eisenarme Bedingungen zu untersuchen. Beide Spezies sind sehr tolerant gegenüber Eisenmangel.

Die ***Akklimatisierung*** an eisenarme Bedingungen verläuft in beiden Spezies sehr ähnlich. Die gesamte Zell-Biomasse wird stark reduziert, erkennbar an der Reduktion der Chloroplasten und dem damit einhergehenden Pigmentverlust, welcher der Zelle einen ausgeblichenen, chlorotischen Phänotyp verleiht. Wachstumsraten sind extrem niedrig bei gleichzeitig hohem Anteil photosynthetisch nichtverwertbarer Lichtenergie, die ungenutzt abgeleitet wird. Die differentielle Regulation der Gene für Ferredoxin (PETF) und Flavodoxin (FLDA) sowie von Genen für Klasse II und Klasse I Fructosebisphosphataldolasen (FBA) weist darauf hin, daß der zelluläre Eisenbedarf gesenkt wird, indem eisenreiche durch eisenfreie Proteine ersetzt werden.

Genetisch manifestierte Merkmale von *T. oceanica* und *P. tricornutum*, welche vorteilhafte ***Adaptationen*** an Eisenmangelbedingungen darstellen könnten, sind sowohl die geringen Zellgrößen beider Arten, als auch der Besitz von Klasse I FBA-Genen, welche nicht in der küstennah beheimateten Art *Thalassiosira pseudonana* vorkommen. Als für *T. oceanica* spezifische Adaptationen beobachten wir die konstitutive Expression des Plastocyanin-Gens (PCY), das ein Substitut für das eisenhaltige Cytochrom c_6 kodiert. Das Ferredoxin-Gen (PETF) wurde in dieser Spezies vom Chloroplasten- zum Kerngenom verlagert, was seine Koregulation mit dem Flavodoxin-Gen (FLDA) erleichert.

Die Akklimatisierung von Kieselalgen an Eisenmangelbedingungen ähnelt den Reaktionen von grünen Pflanzen und Algen auf Eisenstress und repräsentiert einen evolutionär alten Mechanismus, der den meisten, wenn nicht allen, photosynthetischen Organismen gemein ist. Genetische Adaptation an dauerhaften Eisenmangel, wie er im offenen Ozean vorherrscht, geschieht wahrscheinlich vorwiegend durch fortgesetztes Ausloten neuer Möglichkeiten, wie eisenreiche durch eisenarme Proteine ersetzt werden können, um somit den zellulären Eisenbedarf auf ein Minimum zu reduzieren. Die mögliche Bedeutung solcher Ersetzungsstrategien für evolutionäre Anpassung an Eisenmangel wird unterstrichen durch deren Fehlen in Küsten-Spezies wie *T. pseudonana*, die an eisenreiche Standorte angepasst ist.

2 Introduction

The pivotal role of diatoms in marine primary production

Biogeochemical elemental cycles are largely driven by primary production of biomass via phototrophic phytoplankton growth. A major fraction of marine primary production is accomplished by diatoms (Nelson et al. 1995). During temporary absence of nutrient limitation a diatom population can bloom and accumulate a massive biomass that may exceed the capacity of the microbial loop responsible for recycling organic matter in surface waters. In these cases excess biomass is exported to the deep sea floor as "marine snow" aggregates forming the basis of the biological carbon pump. This blooming potential of diatoms points towards an adaptation to the feast & famine open ocean regimes, where nutrients are only provided sporadically. Survival of diatoms under these conditions requires a constant presence of a silent stock of cells and their ability to weather periods of nutrient shortage in a metabolic stand-by mode of lowered activity and growth.

From an evolutionary perspective diatoms can be characterized as heterotrophs that acquired the infrastructure for phototrophic carbon assimilation as an additional nutritional mode from the secondary endosymbiosis with a red algal cell (Grzebyk 2003). Acquisition of phototrophy provided the prerequisites for colonizing even oligotrophic open ocean sites like the Sargasso Sea, where non-phototrophic eukaryotes – for reason of their small surface to volume ratio – would not be able to compete with the diverse and specialized group of heterotrophic bacteria. At the same time, it seems unlikely that diatoms would have lost their genuine heterotrophic capacity, as it may be especially beneficial under conditions where photosynthesis is impaired as under iron limited conditions or during temporary periods of darkness. While there has been considerable research on diatom photosynthesis, which has lead to a better understanding of the phototrophic aspect, less information is available for the "dark" side of such mixotrophic nutrition, i.e. the heterotrophic capacity. However, diatom mixotrophy is known since long (Lewin & Lewin 1960) and has been explored for biotechnological application (Garcia et al. 2006). An implicit capacity of diatoms for heterotrophic nutrition is also seen from the required supplementation of axenic lab cultures with organic vitamins – in contrast to green plants that are able to synthesize all essential organic components from the scratch. However, under conditions where iron or other nutrients limit the build-up of biomass like in the oligotrophic Sargasso Sea (N and P limited, low iron;

isolation site of *Thalassiosira oceanica*) or the southern ocean (iron limited; isolation site of *Fragilariopsis cylindrus*) the major contribution to cellular growth can be expected to stem from photosynthetic carbon assimilation.

The iron-limited open ocean

Iron is an indispensable nutrient for all phototrophic organisms. It functions as a powerful electron carrier in iron-sulfur and heme-containing proteins and as such is an essential component of the photosynthetic electron transport system. The solubility of iron in water decreases exponentially with increasing salinity. While dissolved iron in riverine or coastal surface waters can be present at concentrations in the micromolar range, concentrations in oceanic sea water are usually in the subnanomolar range (Figuères et al. 1978). Consequentially, in wide areas of the worlds' oceans phytoplankton growth is limited through iron availability. This is best illustrated by HNLC (high-nitrate low-chlorophyll) "deserts", oceanic areas that lack any form of regular iron supply and suffer from a persistent shortage of this nutrient. Here, although other commonly limiting nutrients like nitrate or phosphate are present at high concentrations, primary productivity – and biomass as a whole – is low (Martin et al. 1994).

Numerous large-scale iron fertilization experiments have confirmed the growth limiting effect of iron in HNLC regions (Boyd et al. 2007). Induced phytoplankton blooms were found to be dominated by diatoms and carbon export to the deep sea floor could be observed in some cases (e.g. at EIFEX, an eddy fertilization experiment in the southern ocean; reviewed by Sachs 2008). The strong responsiveness of diatoms to the input of iron in iron-limited environments has been a motivation for exploring large-scale iron fertilization as a possible bioengineering strategy to sequester CO_2 into the ocean. The fertilizing effect of iron on marine habitats is also found in nature. Prominent cases of such *natural* iron fertilization can be observed in the iron-limited Southern Ocean (Blain et al. 2007) as well as in the N limited waters west of Sahara near the Cape Verde Islands, where massive input of iron-rich atmospheric dust from dust storms serves the extensive iron needs of diazotrophs (Maranon et al. 2010).

Reductive iron uptake

Phytoplankton has evolved sophisticated strategies for the uptake of iron. The applicable mechanisms of iron uptake crucially depend on the nature of the available iron sources. In oxygenated sea water iron is largely present in its oxidized Fe^{3+} form. The extremely low solubility of Fe^{3+} ions implies that most iron may not be present as freely dissolved ions but rather as part of organic complexes that bind Fe^{3+} in a specific or unspecific manner. For the purpose of iron uptake it must

be extracted from its organic carriers. It has been found that for several algae incl. diatoms – as well as for fungi like yeast) – the reduction of Fe^{3+} to Fe^{2+} represents an essential step in the uptake of organically complexed iron (Shaked et al. 2005). The iron transmembrane transport in diatoms is hypothesized to occur via a specific permease upon previous reoxidation of iron by a multicopper oxidase (Maldonado et al. 2006) as described for yeast (reviewed by Van Ho et al. 2002). Moreover, Paz et al. (2007) could show that in *Dunaliella salina* iron is temporarily stored on the cellular surface by specialized iron-binding molecules and finally taken up via endocytotic vesicular transport in acidic vacuoles. Though the organisms used for characterizing the single steps in iron-uptake are somewhat diverse, the presence of the respective orthologous genes – like e.g. ferric reductases and multicopper oxidases – in diatoms suggests that these also use a reductive mechanism for iron uptake, potentially also combined with endocytotic cycling. Accordingly, major cellular players involved in diatom iron uptake can be expected to comprise receptors for organic iron complexes, redox enzymes needed for extracting the iron from its complexes, ferric reductases and (multicopper) ferrous oxidases responsible for interconversion of the iron redox species +III and +II, and finally iron permeases for iron import.

Acclimation and adaptation to low-iron conditions

The crucial impact of iron availability on phytoplankton growth had lead to the early evolution of strategies to counteract iron limitation. The cellular strategies to cope with decreased iron availability can principally be differentiated into (I) decreasing the cellular iron requirements and (II) increasing the iron uptake kinetics (Morel et al. 1991).

Cellular iron requirements can be lowered by

- Reducing the iron-rich photosynthetic machinery. This often is accompanied by a decrease in chloroplast size and number and a bleached phenotype (“chlorosis”).

- Reducing the cell size or the inner cellular biomass. Inner biomass reduction as response to low-iron can be observed in diatoms from the decrease in cellular protein content and from the expansion of the vacuolar compartments. We found that in *Thalassiosira oceanica* the reduction of the chloroplast compartments is roughly proportional to the reduction of the cellular remainder.

- Replacing proteins that use iron as cofactor with iron-free substitutes that are functional equivalent. Examples for this strategy are the mutual substitution of ferredoxin/flavodoxin

(LaRoche et al. 1996) or cytochrome c_6/plastocyanin (Peers & Price 2006).

The iron uptake kinetics can be enhanced by

- Increasing the abundance of iron binding sites on the cell surface. This can be achieved by an enhanced expression of the involved genes. On the genomic level it may be supported by duplication of the respective iron-binding genes or domains.
- Increasing the ratio between cell surface area and cell volume. This can result from either a general decrease in cellular size or from modification of the cellular shape (ratio between height and width).
- Inducing a high-affinity iron-uptake system whose components have a higher affinity to iron (iron complexes) compared to the default uptake machinery.
- Accessing additional iron sources. A possible scenario for this would be the production and secretion of iron chelating siderophores for the purpose of recruiting remote iron to the cell.

How these strategies are combined and regulated by an organism under conditions of decreased iron availability determines its capacity for ***acclimation*** to low-iron in the light of its respective genomic background. The genome (the complement of an organism's genes) on the other hand provides the framework for such acclimation and determines which of these strategies are available for acclimation to low-iron. The outcome of the continuous refinement of this genomic framework in response to a persistent selective pressure of growth-limiting iron concentrations can be regarded as an organism's ***adaptation*** to these conditions.

Tolerance to iron limitation

Macro- and micronutrient availability in the open ocean is fundamentally different from the situation in coastal areas. As a result of specific adaptation, oceanic and neritic phytoplankton species can be distinguished from another by their growth characteristics and their tolerance to nutrient limitation (Brand et al. 1983).

The pennate diatom species *Phaeodactylum tricornutum* and the centric diatom *Thalassiosira oceanica* have both been shown to be tolerant to iron-limitation and were used in the projects presented in this work as models for a comprehensive analysis of their low-iron response (acclimation) in the light of their respective genomic

backgrounds (adaptation). Species-specific peculiarities in terms of acclimation or adaptation to low-iron conditions can only be fully appreciated from their absence in diatoms that are sensitive to iron-limitation. An especially well-suited control for this purpose is *Thalassiosira pseudonana*, a coastal centric diatom species that requires the relatively high iron concentrations found in coastal waters, originating from continuous replenishment through coastal and riverine iron input.

The age of "omics"

The ever-decreasing costs of large-scale DNA sequencing due to innovative technologies (former "next generation sequencing" technologies) make it feasible to address questions of the cellular stress response or of adaptive genomic features on a global cellular level ("omics" approaches). Accordingly, the presented projects integrate whole-genome (gen***omics***), whole-transcriptome (transcript***omics***) and whole-proteome (prote***omics***) data for unraveling the cellular strategies to reduce iron requirements during phototrophic growth and to overcome the physiological restrictions coming along with a reduced photosynthetic capacity.

The analysis of an organism's **genome** reveals its functional **potential** and enables the detection of species-specific adaptive features (genes). The analysis of **transcriptome** and **proteome** data on the other hand allows the investigation of what part of this functional potential is actually realized under the tested conditions in terms of a functional **response**, thereby constituting a reversible acclimation to these conditions on the level of gene expression.

The aim of this work is to use the complex sequence data obtained from "omics" approaches for a refined characterization of the acclimation of diatoms to low-iron and for the identification of genomic features that may be regarded as mediating a specific adaptation to these conditions. For this purpose we use two diatoms that are known to be tolerant to iron-limited conditions and that can be regarded as adapted to the low iron concentrations found in the open ocean, *P. tricornutum* and *T. oceanica*. In each case cultures from iron-limited versus iron-replete conditions were grown and compared to assess the acclimation response, while genome analysis revealed putative adaptive features.

Acclimation and adaptation to low iron in P. tricornutum

For characterization of its genome and low-iron response we used the strain *Phaeodactylum tricornutum* Bohlin CCAP1055/1 (Bohlin 1897). The work on *P. tricornutum* has lead to the following publications and manuscripts:

- *P. tricornutum* Genome Paper:

  Bowler C, et al.: **The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**:239-244.

- *P. tricornutum* Iron Paper:

  Allen AE, et al.: **Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation.** *PNAS* 2008, **105**:10438-10443.

- *P. tricornutum* FCP Paper (Predraft):

  Lommer M, et al.: **The *Phaeodactylum tricornutum* superfamily of FCP-like proteins: Evolution, differential stress response and interaction model.**

Iron-limited *P. tricornutum* has a bleached, chlorotic phenotype with a strongly reduced two-lobed chloroplast. Attached to each chloroplast lobe there is a chlorophyll-rich vesicle assumed to be the Thylakoid-Organizing Body (TOB) which serves as a reservoir for chlorophyll and/or chlorophyll-binding proteins and which is not observed in iron-replet cells. In electron micrographs we observe an expansion of the vacuolar system indicating a lower biomass compared to nutrient-replete cells. The fusiform cells are also smaller in size under low-iron conditions (Allen et al. 2008: "*P. tricornutum* Iron Paper"). Iron-limited *P. tricornutum* proliferates with a minimal growth rate of 0.2.

Prerequisite for efficient molecular work on *P. tricornutum* was the publication of its 28Mb genome (Bowler et al. 2008: "*P. tricornutum* Genome Paper"). From the analysis of Expressed Sequence Tag (EST) libraries of iron-limited versus iron-replete cultures and from qPCR analyses we found a differential regulation for ferredoxin and ***flavodoxin***, both functioning as photosynthetic electron carriers between photosystem I and the NADPH oxidoreductase and known for their mutual substitution upon change in cellular iron-supply. Further we could identify the induction of two ***class I fructose-bisphosphate***

aldolases (FBAs) in this data set. Class I FBAs are involved in the conversion of carbohydrates and use a Schiff-base catalysis instead of the metal catalysis typical for class II FBA enzymes.

With respect to putative components of the iron-uptake system the induction of a gene for a ***receptor for ferrichrome*** (a hydroxamate-type ferrisiderophore complex) was observed. Siderophores are small organic molecules that function in chelating soluble iron. The resulting ferrisiderophore complex is better accessible to the cell than the free ion. The high affinity of siderophores for iron further prevents the unspecific association of iron with other organic matter. The gene for this ferrichrome receptor (FCR) is lacking in other diatoms sequenced so far. Further, a total of four ***ferric reductases*** that may be involved in reductive iron-uptake appeared up-regulated. Three genes that encode for putative cell surface proteins appeared strongly up-regulated upon low-iron. These have been named after there respective gene products ***Iron Starvation Induced Proteins*** (ISIPs). Though their function is unknown it appears possible that these ISIPs (ISIP1, ISIP2 and ISIP3) – just like the ferrichrome receptor – may be involved in iron-uptake. Accordingly, their induction would increase the abundance of iron binding sites on the cell surface, thereby enhancing the kinetics of iron binding and uptake.

The observation of differential regulation of ***Fucoxanthin-Chlorophyll-binding Proteins*** (FCPs, mainly photosynthetic light antenna proteins) motivated a study leading to a comprehensive characterization of the whole protein family (Lommer et al. – predraft manuscript "*P. tricornutum* FCP Paper"). FCPs constitute the interface between solar energy input and photosynthetic electron transport and their ability to either forward or quench such energy makes them important regulatory units in photosynthesis. Their interaction with each other and with the photosystems helps to adjust the photosynthetic energy supply to the actual needs. At the same time, the ability of certain FCPs to quench light energy helps to avoid excessive production of Reactive Oxygen Species (ROS), thereby protecting the photosynthetic machinery that is strongly reduced and impaired under low-iron conditions.

The differential response of FCP gene expression to the (interfering) stresses imposed by light intensity, iron supply and growth phase was tested with Reverse Transcription quantitative Polymerase Chain Reaction (RT-qPCR). Therein we could confirm the prominent role of certain FCP-like proteins (***CPQ*** – former LI818; ***FCPAQ***) in the general and/or specific stress response of the photosynthetic system (Peers et al. 2009). Moreover, we could show that the light antenna system of diatoms (in this case *P. tricornutum*) indeed has evolved members of

the FCP family that are specifically responsive to iron limitation (e.g. LHC9 – going to be renamed to ***FCPF9***).

In the qPCR experiment addressing the differential stress response of FCPs we included several of the low-iron responsive genes identified earlier (Allen et al. 2008) for a refined characterisation of their regulation. The most relevant findings resulted from the addition of the ***nuclear RNA-Polymerase II*** (RPB1) gene as second standard besides the 18S ribosomal RNA gene. As the expression of RPB1 is more constant per cell, it addresses not only the regulation relative to the active biomass (that is highly variable in response to changing iron-availability and indicated by 18S) but also the regulation per cell. This is especially important for genes whose products are targeted to compartments that do not change as much as the inner cellular biomass, namely the cell surface or the nuclear compartment. Specifically we could confirm the strong and specific differential regulation of ***ISIP1*** and ***ISIP3***. ISIP2 on the other hand does not appear to be specifically responsive to low-iron, but may rather represent a constitutive cell surface protein. Moreover, we could show that of the four genes that contain a ferric reductase domain only a single one is specifically induced by low-iron (***FRE2***). This suggests that the other members of this group may not be involved in reductive iron-uptake and act in a different functional context. With regard to the group of FCP-like proteins we conclude that most FCPs that appear up-regulated at low-iron when normalized to 18S (***FCPAQ1***, ***FCP1***, LHC/***FCPF9***), are rather retained than induced when normalized to RPB1, while all other FCPs are down-regulated (consistent with the chlorotic phenotype). FCPs that are maintained at iron-limitation may be found in the TOBs and could be responsible for the deep green staining of these bodies.

Unlike the diverse facets that make up the acclimation response of *P. tricornutum* to low-iron, the evidence for adaptive features is scarce. The genome of *P. tricornutum* is relatively small suggesting a limited capacity for hosting DNA from external sources. However, Bowler et al. have found a significant fraction of *P. tricornutum* genes to be derived from horizontal gene transfer, i.e. from the uptake and integration of external DNA (Bowler et al. 2008).

Genes that may have an adaptive significance with regard to the adaptation to low-iron are the genes for ***flavodoxin*** (FLDA), for the ***pyrenoid class I FBA***, for the ***ferrichrome receptor*** (FCR) and for the ***putative receptor ISIP1***. None of these are found in the genome of *T. pseudonana*.

Acclimation and adaptation to low iron in T. oceanica

For characterization of its genome and low-iron response we used the strain *Thalassiosira oceanica* (Hustedt) Hasle et Heimdal CCMP1005 (Hasle 1983). The work on *T. oceanica* has lead to the following publications and manuscripts:

- *T. oceanica* PETF Paper:

  Lommer M, et al.: **Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation.** *BMC Genomics* 2010, **11**:718.

- *T. oceanica* Genome Paper (submitted manuscript):

  G**enome and low-iron response of an oceanic diatom adapted to chronic iron limitation.**

Iron-limited *T. oceanica* has a reduced chloroplast system with two instead of four organelles and a strongly reduced volume of the chloroplasts, whereas the bleaching known from iron-limited *P. tricornutum* is less pronounced. The cellular protein content is decreased by 50 % and from microscopy images we find an increased vesiculation in the cellular interior. Both indicate a lower biomass compared to nutrient-replete cells as has been found for *P. tricornutum*. Further, *T. oceanica* cells show an elongated phenotype with decreased width and increased height under low-iron conditions, thereby increasing the cellular surface to volume ratio. Iron-limited *T. oceanica* proliferates with a minimal growth rate of 0.2.

As a major part of the project the genome of *T. oceanica* has been sequenced and is estimated to be approx. 80 Mb in size (haploid). The genome assembly is presented in a browser environment with all additional experimental data from transcriptomics and proteomics approaches mapped to the genomic scaffolds at http://bose.geomar.de/cgi-bin/gbrowse/Toceanica/ (Lommer et al. – submitted manuscript "*T. oceanica* Genome Paper"). From the analysis of Expressed Sequence Tag (EST) libraries of iron-limited versus iron-replete cultures and from qPCR analyses we found a differential regulation for ferredoxin and ***flavodoxin***, a constitutive expression of ***plastocyanin*** and the differential regulation of ***class I and class II FBAs*** at low-iron. Plastocyanin functions as photosynthetic electron carrier between the cytochrome b_6/f complex and photosystem I and is known to replace cytochrome c_6 upon decreased iron availability in some algae. With respect to putative components of the iron-uptake system the genes for ***ISIP1***, ***ISIP3*** and for a single ferric reductase (***FRE1***) were found to be induced on the level of transcript abundance

by low-iron. Moreover, an increased abundance of the ISIP1, ISIP3 and class I FBA *protein* products could be shown from *proteomics* data, suggesting that expression of these genes may be regulated on the transcript level.

The genome of *T. oceanica* contains a variety of genes for which a lateral origin is assumed. Despite the larger genome size the estimated fraction of such laterally transferred genes with 5 % is very similar to that found for *P. tricornutum.*

A refined analysis of the FBA protein family revealed that *T. oceanica* contains for each of the compartments **cytosol**, **chloroplast stroma** and **chloroplast pyrenoid** both a ***class II*** and a ***class I FBA*** gene. This strongly suggests that *T. oceanica* may be able to replace each class II FBA with a class I isoform when required.

Genes that may have an adaptive significance with regard to the adaptation to low-iron are the genes for ***flavodoxin*** (FLDA), for ***plastocyanin*** (PETE), for the ***pyrenoid*** and the ***stromal class I FBAs***, and for the ***putative receptor ISIP1***. None of these are found in the genome of *T. pseudonana*. A further peculiarity of *T. oceanica* is the ***transfer*** of the chloroplast ***petF gene*** to the nuclear compartment (Lommer et al. 2010 "*T. oceanica* PETF Paper"). The adaptive significance of this transfer with respect to low-iron is unclear, though a co-location of the ferredoxin and the flavodoxin gene in the nuclear genome may facilitate their differential regulation.

Concluding Remarks

The ability to reduce iron requirements is key to the ecological success in iron poor environments. The **acclimation** response to low-iron is very similar between *T. oceanica* and *P. tricornutum*. Both minimize their active biomass – best visible from the concurrent reduction of the chloroplast system – and thrive at low growth rates. Such retrenchment to a ***low-activity-low-biomass state*** readily represents a significant decrease in iron requirements.

On the molecular level we observe the strong induction of ***iron-free proteins capable of functionally replacing abundant enzymes that contain iron as a co-factor***, namely the photosynthetic electron carrier flavodoxin, plastocyanin and class I fructose-bisphosphate aldolases. ***LI818-like components of the antenna system*** that are capable of quenching excess energy are induced in terms of a general stress response. With respect to the iron-uptake system, both diatoms specifically induce a ***single ferric reductase*** and few additional cell-surface proteins of unknown function (***ISIP1 and ISIP3***). For ISIP1 we provide strong evidence for a role as a cell-surface receptor. If

ISIP1 is indeed a receptor for iron, its strong induction would increase the abundance of iron binding sites on the cell surface, thereby enhancing iron affinity and iron uptake kinetics.

Several marine-derived fungi are known to produce siderophores under iron-limited conditions (Holinsworth & Martin 2009). The question related to whether or not diatoms are capable of producing siderophores has attracted a high interest in the field of diatom research. Diatoms contain a conserved ***non-ribosomal peptide synthase*** (NRPS) that is homologous to those of ascomycetes capable of producing hydroxamate-type siderophores like ferrichrome. *T. oceanica* contains a second NRPS gene homologous to bacterial-type NRPS that are responsible for production of diverse antimicrobial peptides. Minor modifications of the peptide products determine the functional identity of these molecules as antimicrobial peptides or siderophores, and currently there is no way of predicting the precise product of NRPS solely from sequence information of the NRPS gene. Further, we could not find a specific induction of NRPS genes in response to low-iron, when normalized to a cell constant standard like RPB1. However, the presence of a ***ferrichrome receptor*** in *P. tricornutum* indicates that siderophores are an iron source for at least some diatoms. If these iron chelators are also produced by diatoms or just consumed in terms of siderophore piracy, remains an open question.

A ***comparative genomics*** analysis of iron-regulated genes between the four diatoms whose genomes are known to date sheds light on the genomic **adaptation** to persistent iron-limitation. While *T. oceanica* (from the oligotrophic Sargasso Sea) and *F. cylindrus* (from the iron-limited Southern Ocean) are known to be tolerant to severe iron-limitation, *T. pseudonana* (a coastal species) is adapted to the high iron concentrations found in coastal waters and sensitive to low-iron. *P. tricornutum* has an intermediate phenotype with regard to its low-iron tolerance. The differential complement of genes for these species (presented in the "*T. oceanica* Genome Paper", Table 3) shows strong similarities between *T. oceanica* and *F. cylindrus*: a common property is the presence of all five genes whose products are iron-free substitutes for abundant iron-enzymes. Further, the duplication of ISIP1 in both diatom genomes is not found in *T. pseudonana* or *P. tricornutum,* suggesting an adaptive significance for this gene. An additional gene found in multiple copies in *T. oceanica* and *F. cylindrus* is that for a multicopper oxidase. Multicopper oxidases can function as ascorbase or in the oxidation of Fe^{2+} to Fe^{3+}. In yeast a multicopper oxidase is specifically associated with an iron permease and accomplishes iron reoxidation prior to import (reviewed by Van Ho et al. 2002). Maldonado et al. (2006) present evidence that like yeast,

diatoms are capable of extracellular iron oxidation in a copper-dependent manner and they hypothesize that a multicopper ferrous oxidase may be a regular component of the diatom iron-uptake system. The multiple copies found in *T. oceanica* and *F. cylindrus,* but not in *T. pseudonana* or *P. tricornutum* suggest a possible relevance for iron uptake in these species. Moreover, the presence of multiple ferrous oxidases could indicate a more complex and flexible iron-uptake system.

While for cyanobacteria it has been shown that adaptation to marine niches can happen via the acquisition of key genes (Martiny et al. 2006), this has not been shown for eukaryotic organisms, probably due to the complexity of eukaryotic genome data. In this work we show that the degree of adaptation to low-iron might indeed be inferred from a ***limited list of indicator genes*** that are presumably found in species with high tolerance to iron-limitation and absent in others like coastal diatoms. Genome data for more ecologically diverse species is needed to test this hypothesis in an extended comparative approach.

Moreover, the data suggests that the ***substitution of abundant iron-proteins*** represents a key strategy in adaptation to low-iron. While the cytosolic FBA may also be found as an adaptive feature in non-phototrophic organisms, the genes for flavodoxin, plastocyanin and the two chloroplast FBAs can be expected to be limited to phototrophic organisms. It becomes evident that much of the (known) strategies to reduce cellular iron requirements have been evolved in phototrophs as a result from the high iron demand of this metabolic pathway.

Such adaptation is not limited to diatoms or phototrophs. An ecosystem is characterized by a certain ***gene pool*** composed of a basic set needed for "ordinary" cellular maintenance common to all organisms and a more specific set that might be regarded as a signature for the specific selective pressure present in a certain habitat. And indeed, diatoms appear to be in extensive exchange of genetic material with other species as seen from the large fractions of laterally acquired genes. To identify these ***environment-specific signatures*** and to unveil how the respective genes are distributed among different prokaryotic and eukaryotic species, will be crucial to understand ecosystem function and is required for diagnostic purposes with respect to changing ecosystems. The application of metagenomics approaches will be needed to address these questions.

References

- Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C: **Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation.** *PNAS* 2008, **105**:10438-10443.
- Blain S, Quéguiner B, Armand L, Belviso S, Bombled B, et al.: **Effect of natural iron fertilization on carbon sequestration in the Southern Ocean.** *Nature* 2007, **446**:1070–1074.
- Bohlin K: **Zur Morphologie und Biologie einzelliger Algen.** *Öfvers af K Vet Acad Förhandl Stockholm* 1897, **54**:519-522.
- Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJ, Follows M, Harvey M, Lancelot C, Levasseur M, Owens NP, Pollard R, Rivkin RB, Sarmiento J, Schoemann V, Smetacek V, Takeda S, Tsuda A, Turner S, Watson AJ: **Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions.** *Science* 2007, **315**:612-617.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, et al.: **The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**:239-244.
- Brand LE, Sunda WG, Guillard RRL: **Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron.** *Limnol Oceanogr* 1983, **28**:1182-1198.
- Doolittle WE: **You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes.** *Trends Genet* 1998, **14**:307-311.
- Figuères G, Martin JM, Meybeck M: **Iron behaviour in the Zaire estuary.** *Neth J Sea Res* 1978, **12**:329-337.
- Garcia MCC, Camacho FG, Miron AS, Sevilla JMF, Chisti Y, Grima EM: **Mixotrophic Production of Marine Microalga *Phaeodactylum tricornutum* on Various Carbon Sources.** *J Microbiol Biotechnol* 2006, **16**:689-694.
- Grzebyk D, Schofield O, Vetriani C, Falkowski PG: **The mesozoic radiation of eukaryotic algae: The portable plastid hypothesis.** *J Phycol* 2003, **39**:259-267.
- Hasle GR: **The Marine, Planktonic Diatoms *Thalassiosira-Oceanica* Sp-Nov and *Thalassiosira-Partheneia*.** *J Phycol* 1983, **19**:220-229.

- Holinsworth B, Martin JD: **Siderophore production by marine-derived fungi.** *Biometals* 2009, **22**:625-632.
- LaRoche J, Boyd PW, McKay RML, Geider RJ: **Flavodoxin as an in situ marker for iron stress in phytoplankton.** *Nature* 1996, **382**:802-805.
- Lewin JC, Lewin RA: **Auxotrophy and Heterotrophy in Marine Littoral Diatoms.** *Can J Microbiol* 1960, **6**:127-134.
- Maldonado MT, Allen AE, Chong JS, Lin K, Leus D, Karpenko N, Harris SL: **Copper-dependent iron transport in coastal and oceanic diatoms.** *Limnol Oceanogr* 2006, **51**:1729-1743.
- Maranon E, Fernandez A, Mourino-Carballido B, Martinez-Garcia S, Teira E, Cermeno P, Choucino P, Huete-Ortega M, Fernandez E, Calvo-Diaz A, Moran XAG, Bode A, Moreno-Ostos E, Varela MM, Patey MD, Achterberg EP: **Degree of oligotrophy controls the response of microbial plankton to Saharan dust.** *Limnol Oceanogr* 2010, **55**:2339-2352.
- Martin JH, Coale KH, Johnson KS, Fitzwater SE, Gordon RM, Tanner SJ, Hunter CN, Elrod VA, Nowicki JL, Coley TL, Barber RT, Lindley S, Watson AJ, Van Scoy K, Law CS, Liddicoat MI, Ling R, Stanton T, Stockel J, Collins C, Anderson A, Bidigare R, Ondrusek M, Latasa M, Millero FJ, Lee K, Yao W, Zhang JZ, Friederich G, Sakamoto C, Chavez F, Buck K, et al.: **Testing the Iron Hypothesis in Ecosystems of the Equatorial Pacific-Ocean.** *Nature* 1994, **371**:123-129.
- Martiny AC, Coleman ML, Chisholm SW: **Phosphate acquisition genes in Prochlorococcus ecotypes: evidence for genome-wide adaptation.** *PNAS* 2006, **103**:12552-12557.
- Morel FMM, Rueter JG, Price NM: **Iron nutrition of phytoplankton and its possible importance in the ecology of ocean regions with high nutrient and low biomass.** *Oceanography* 1991, **4**:56-61.
- Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B: **Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation.** *Global Biogeochem Cycles* 1995, **9**:359-372.
- Paz Y, Shimoni E, Weiss M, Pick U: **Effects of iron deficiency on iron binding and internalization into acidic vacuoles in *Dunaliella salina*.** *Plant Phys* 2007, **144**:1407-1415.
- Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK: **An ancient light-harvesting protein is critical for the regulation of algal photosynthesis.** *Nature* 2009, **462**:518-522.
- Peers G, Price NM: **Copper-containing plastocyanin used for**

electron transport by an oceanic diatom. *Nature* 2006, **441**:341-344.

- Sachs O: **Benthic organic carbon fluxes in the southern ocean: Regional differences and links to surface primary production and carbon export.** *Reports on polar and marine research* 2008, **578**.
- Shaked Y, Kustka AB, Morel FMM : **A general kinetic model for iron acquisition by eukaryotic phytoplankton.** *Limnol Oceanogr* 2005, **50**:872-882.
- Van Ho A, Ward DM, Kaplan J: **Transition metal transport in yeast.** *Annu Rev Microbiol* 2002, **56**:237-261.

3 Publications and Manuscripts

3.1 P. tricornutum Genome Paper

Title:

The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes.

Published in:

Nature. 2008 Nov 13; 456(7219):239-44. Epub 2008 Oct 15.

Author list:

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kröger N, Kroth PG, LaRoche J, Lindquist E, Lommer M, Martin-Jézéquel V, Lopez PJ, Lucas S, Mangogna M, Mc Ginnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV.

Abstract:

Diatoms are photosynthetic secondary endosymbionts found throughout marine and freshwater environments, and are believed to be responsible for around one-fifth of the primary productivity on Earth. The genome sequence of the marine centric diatom *Thalassiosira pseudonana* was recently reported, revealing a wealth of information about diatom biology. Here we report the complete genome sequence of the pennate diatom *Phaeodactylum tricornutum* and compare it with that of *T. pseudonana* to clarify evolutionary origins, functional significance and ubiquity of these features throughout diatoms. In spite of the fact that the pennate and centric lineages have only been diverging for 90 million years, their genome structures are dramatically different and a substantial fraction of genes (approximately 40%) are not shared by these representatives of the two lineages. Analysis of molecular divergence compared with yeasts and metazoans reveals rapid rates of gene diversification in diatoms. Contributing factors include selective gene family expansions,

differential losses and gains of genes and introns, and differential mobilization of transposable elements. Most significantly, we document the presence of hundreds of genes from bacteria. More than 300 of these gene transfers are found in both diatoms, attesting to their ancient origins, and many are likely to provide novel possibilities for metabolite management and for perception of environmental signals. These findings go a long way towards explaining the incredible diversity and success of the diatoms in contemporary oceans.

Statement on my contribution to this publication:

As a member of the *Phaeodactylum* genome annotation jamboree I annotated much of the specifically low-iron responsive genes of this diatom species.

3.2 P. tricornutum Iron Paper

Title:

Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation.

Published in:

Proc Natl Acad Sci U S A. 2008 Jul 29; 105(30):10438-43. Epub 2008 Jul 24.

Author list:

Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C.

Abstract:

Marine primary productivity is iron (Fe)-limited in vast regions of the contemporary oceans, most notably the high nutrient low chlorophyll (HNLC) regions. Diatoms often form large blooms upon the relief of Fe limitation in HNLC regions despite their prebloom low cell density. Although Fe plays an important role in controlling diatom distribution, the mechanisms of Fe uptake and adaptation to low iron availability are largely unknown. Through a combination of non-targeted transcriptomic and metabolomic approaches, we have explored the biochemical strategies preferred by *Phaeodactylum tricornutum* at growth-limiting levels of dissolved Fe. Processes carried out by components rich in Fe, such as photosynthesis, mitochondrial electron transport, and nitrate assimilation, were down-regulated. Our results show that this retrenchment is compensated by nitrogen (N) and carbon (C) reallocation from protein and carbohydrate degradation, adaptations to chlorophyll biosynthesis and pigment metabolism, removal of excess electrons by mitochondrial alternative oxidase (AOX) and non-photochemical quenching (NPQ), and augmented Fe-independent oxidative stress responses. Iron limitation leads to the elevated expression of at least three gene clusters absent from the *Thalassiosira pseudonana* genome that encode for components of iron capture and uptake mechanisms.

Statement on my contribution to this publication:

I collected much of the physiological data including ^{14}C carbon fixation assays and NPQ measurements. As a member of the *Phaeodactylum* genome annotation jamboree I annotated much of the specifically low-iron responsive genes and prepared the overview of low-iron responsive gene clusters (Figure 2). Further I contributed gene expression data from RT-qPCR experiments.

3.3 T. oceanica PETF Paper

Title:

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

Published in:

BMC Genomics. 2010 Dec 20; 11:718.

Author list:

Lommer M, Roy AS, Schilhabel M, Schreiber S, Rosenstiel P, LaRoche J.

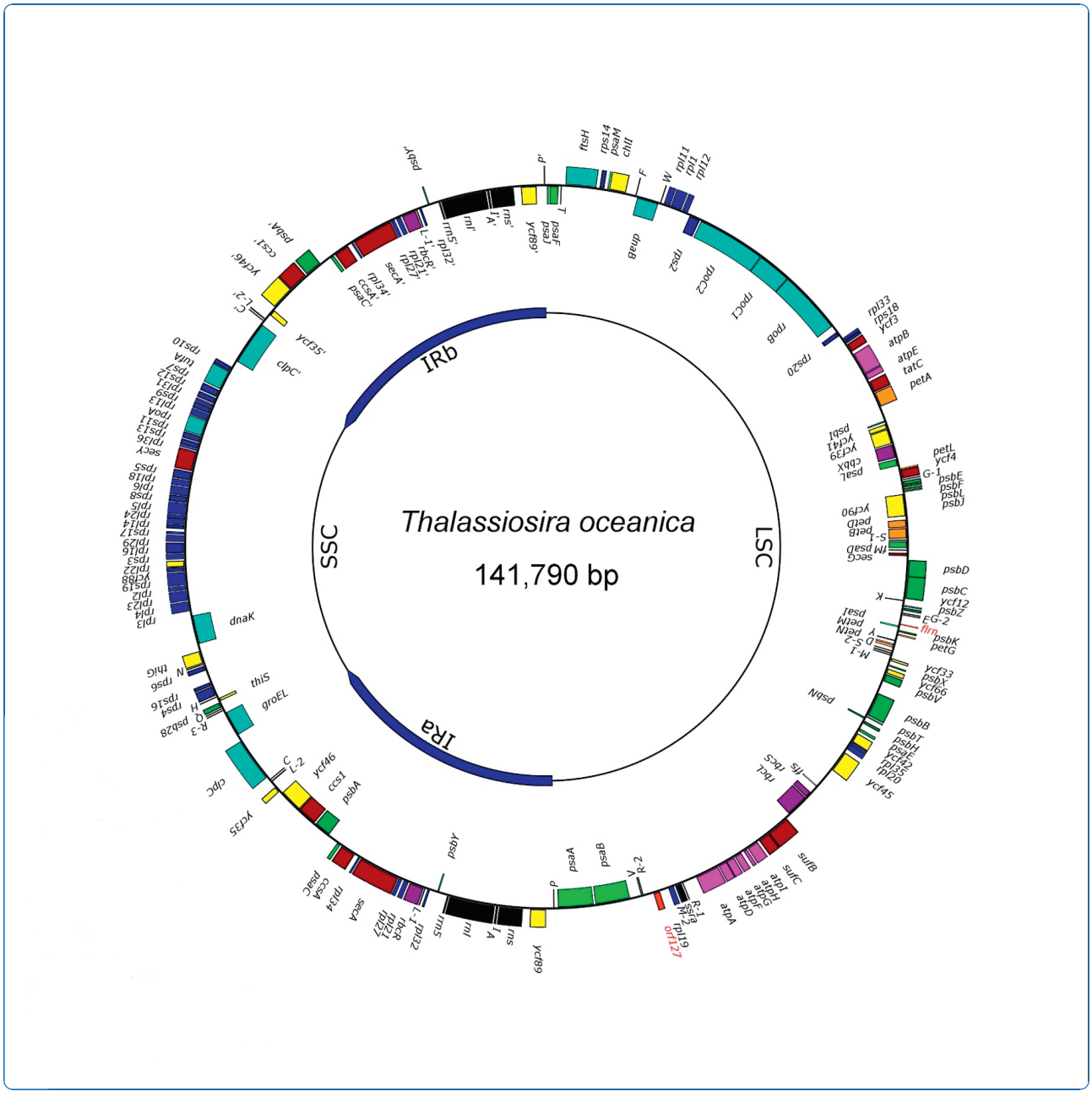
Abstract:

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

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

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

BMC Genomics



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

Lommer *et al.*

BMC Genomics

RESEARCH ARTICLE Open Access

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

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

Abstract

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

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

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

Background

In contemporary oceans, diatoms account for approximately 40% of the oceanic primary production and play a critical role in the sequestration of atmospheric CO_2 into the deep ocean [1]. The high diversity of diatoms and their cosmopolitan distribution in the marine environment reflect the ecological success endured by this group since their first appearance more than 150 Ma ago [2].

Although diatoms thrive in coastal areas where dissolved nutrients are high, many species of diatoms are prevalent in high-nutrient, low-chlorophyll (HNLC) oceanic regions where primary production is chronically iron-limited [3]. Iron fertilization experiments in HNLC regions have repeatedly demonstrated the ability of opportunistic diatom species to bloom once iron is no longer growth limiting [4]. In contrast, some diatom species such as *Thalassiosira oceanica* thrive equally well in the presence or absence of iron [5]. A key determinant for the survival and growth of phytoplankton under iron limitation must be the ability to carry out photosynthesis efficiently, despite the high iron requirements of the photosynthetic infrastructure.

The photosynthetic apparatus, largely contained in the chloroplasts, is jointly coordinated by the plastid and

* Correspondence: jlaroche@ifm-geomar.de
[1]Leibniz Institute of Marine Sciences at Kiel University IFM-GEOMAR, Kiel, Germany
Full list of author information is available at the end of the article

nuclear genomes, involving more than 700 genes [6]. The chloroplast genome of most species generally retained less than 200 of the genes contributing to chloroplast function, as the majority of the endosymbiont's chloroplast genes have been lost or incorporated into the host nuclear genome. The plastids of diatoms and other chromalveolates, originated from a secondary endosymbiosis with a red alga, have retained a higher proportion of the symbiont's genes in their genomes relative to their green counterparts, which derived from a primary endosymbiosis with a cyanobacterium. The red origin of the chloroplasts [7] and the lower cellular iron requirements of the red lineage [8] may have contributed to the ecological success of diatoms in the marine environment in terms of a putative evolutionary-based pre-adjustment to iron-deplete conditions.

The retention of a core set of chloroplast genes and the factors preventing their transfer to the nucleus are the subject of ongoing debates [9]. The known chloroplast genomes of diatoms are circular with an extended inverted repeat region (IR) and are subject to internal rearrangements such as inversions [10]. Occasional organelle lysis and free release of organellar DNA is considered an important first step in the transfer of organelle-encoded genes to the nuclear genome. Indeed high quantities of chloroplast (cp) and mitochondrial (mt) DNA are frequently transferred and inserted into the nuclear genome, thereafter referred to as nuclear plastid or nuclear mitochondrial DNA (NUPTs, NUMTs) [11-14]. However, stable replacement of a plastid gene by its nuclear copy requires a retargeting of the nuclear gene product back to the chloroplast compartment as well as a functional expression and regulation. Until this multi-step development has been accomplished the chloroplast version cannot be discarded and the gene exists in two copies, which might even overlap in function in a differentially regulated manner [15]. Genes that provide a dual targeting sequence enabling import to both mitochondria and chloroplasts at the same time are also known [16].

Here, we present the complete *T. oceanica* CCMP1005 chloroplast genome sequence [GenBank: GU323224], which we assembled from a massively parallel pyrosequencing data set. Assembled genomic shotgun reads show that in the *T. oceanica* genome, the ferredoxin *petF* gene has been transferred to the nucleus. Ferredoxin is a photosynthetic redox protein that contains iron, and in diatoms it can be replaced by the iron-free flavodoxin, when iron-limited growth conditions prevail. The *petF* transfer to the nuclear genome may enable a refined regulation of this gene in response to iron availability in *T. oceanica*. Through comparative genomics between the coastal *Thalassiosira pseudonana* and the oceanic *T. oceanica*, we can trace the *T. oceanica PETF* gene back to its chloroplast origin and identify elements that may have played a role in the transfer of this important photosynthetic gene.

Results

Characteristics of the *T. oceanica* chloroplast genome

The cp genome of *T. oceanica* has a physical size of 141790 bp and maps to a circular topology (Figure 1), slightly larger in size than the published cp genomes of the closely related *T. pseudonana* and the more distantly related *P. tricornutum* (Table 1). The cp genome of *P. tricornutum* shows a significantly higher G+C content and contains three protein coding genes not present in the cp genomes of *T. oceanica* or *T. pseudonana*, while the 6.9 Kbp inverted repeat region is significantly smaller than in the two *Thalassiosira* species. The gene composition of the large and small single copy regions (LSC, SSC) and inverted repeat subdomains (IRa, IRb) is nearly identical between *T. oceanica* and *T. pseudonana*. The differences in the overall size of their cp genomes can be solely attributed to the expansion of the inverted repeat region in *T. oceanica* leading to a *de facto* duplication of the three genes *clpC*, *trnC* and *trnL*, and to the loss of the *petF* gene from the large single copy region. Novel features of the *T. oceanica* cp genome are the appearance of an expressed *orf127* at the site affected by the *petF* gene transfer event, and the partial duplication of the RNA gene *ffs* referred to as *flrn* (ffs-like RNA). Synteny between *T. oceanica* and *T. pseudonana* is weakly conserved (data not shown) and indicates a high degree of dynamic restructuring of chloroplast genomes mainly in form of small scale inversions, though restricted to the respective subdomains (LSC, SSC, IR) without crossing the borders of these elements (compare [10]).

Transfer of *petF* gene to the nuclear genome in *T. oceanica*

The ferredoxin *petF* gene is not present in the cp genome of *T. oceanica* indicating a loss or transfer of that gene to the nucleus. To address this question, BLAST analysis of 454 whole genome shotgun data was performed and revealed the presence of a nuclear genomic contig containing a complete *PETF* gene, confirming the gene transfer event. The nuclear genomic context of the *PETF* region could be assembled manually from the 454 sequence read data and was compared to the syntenic region of *T. pseudonana* (Figure 2). In the *T. pseudonana* cp genome, the *petF* gene is located between the genes *rpl19* and *psaB*; the comparable site in *T. oceanica* shows the *T. pseudonana petF* gene being replaced by a sequence block consisting of the novel *orf127* and the tRNA genes *trnV* and *trnR*. Interestingly, the latter unit is surrounded by an unusually large dispersed 74 bp

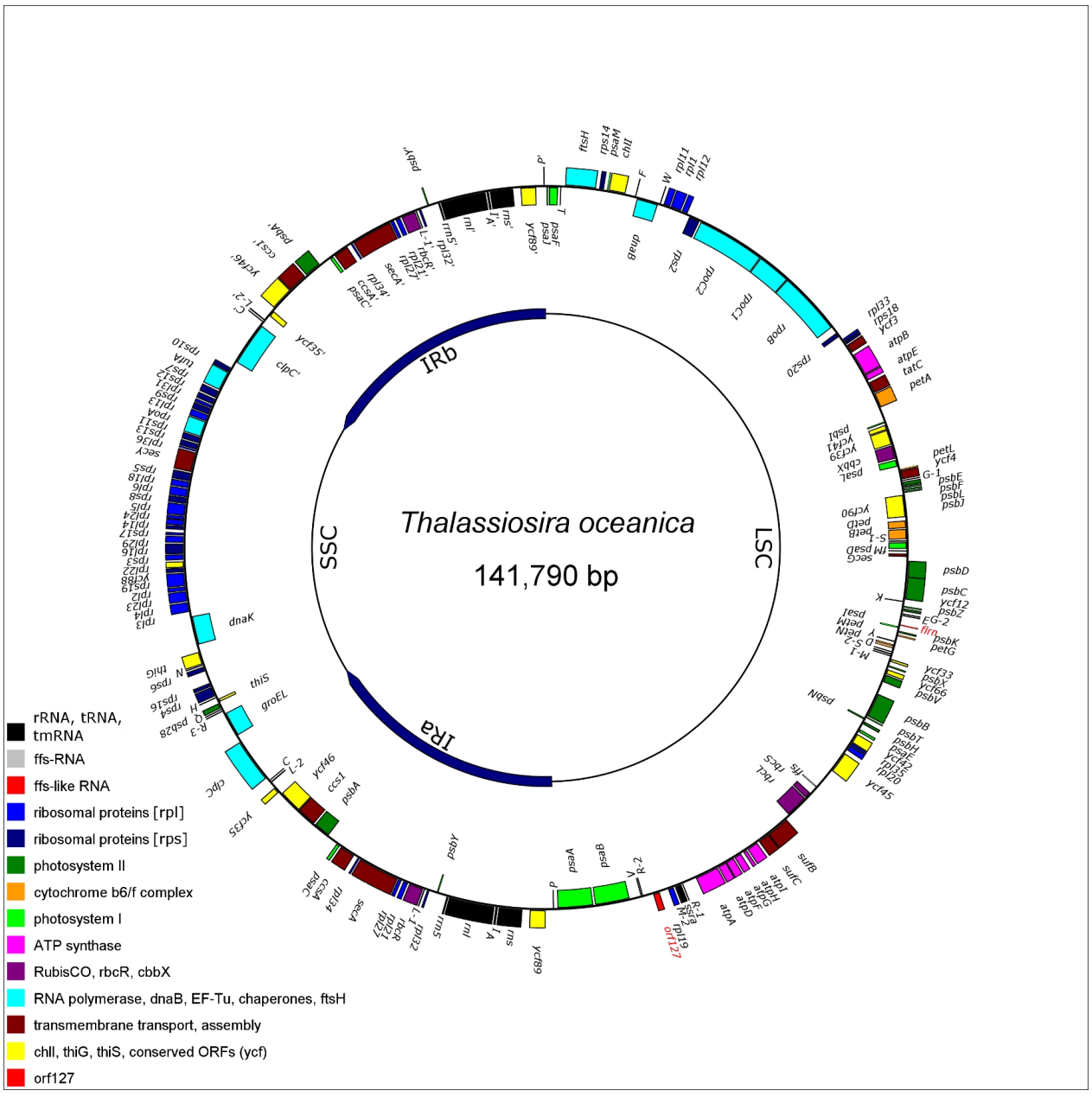


Figure 1 Circular Map of the *Thalassiosira oceanica* CCMP1005 Chloroplast Genome. Genes on the outer *forward* strand are transcribed clockwise, genes on the inner *reverse* strand counterclockwise. Gene symbols are colour-coded according to functional groups as indicated, tRNA genes are represented by single-letter code. Features not present in *T. pseudonana* CCMP 1335 are marked red (***orf127***, ***flrn***). **LSC** large single-copy region; **SSC** small single-copy region; **IRa** and **IRb** inverted repeat regions.

inverted repeat, indicating major genomic rearrangements in that region mediated by these elements and involving a recruitment of the two tRNA genes from a more distant site. Mapping of the *PETF* gene on the nuclear genomic fragment revealed a *PETF* gene model composed of two exons separated by an intron. The first exon encodes a chloroplast targeting peptide, while the second exon carries the information for the conserved functional part of the ferredoxin gene product. The predicted gene model reveals a nuclear promoter as well as a polyA-signal. The *PETF* gene and the adjacent *PDD* gene encoding for a pyridoxal-dependent decarboxylase [17] are surrounded by larger non-related repeat elements (>1000 bp) likely representing remnants of recombinational genomic rearrangements involved in the *petF* gene transfer event. A highly conserved ortholog of the *PDD*

Table 1 Chloroplast Genome Features of *T. oceanica* in Comparison with *T. pseudonana* and *P. tricornutum*

	Thalassiosira oceanica CCMP 1005	*Thalassiosira pseudonana* CCMP 1335	*Phaeodactylum tricornutum* CCAP 1055/1
genome size [bp]	**141 790**	**128 814**	**117 369**
LSC	70 298	65 250	63 674
SSC	24 106	26 889	39 871
IR	**23 693**	IRa: **18 338**	**6 912**
		IRb: 18 337	
G+C content [%]	**30,39**	**30,66**	**32,56**
A+T content [%]	69,61	69,34	67,44
gene content [#]	**158+2**	**159**	**162**
protein coding	126 +*orf127*	127	130
rRNA	3	3	3
tRNA	27	27	27
other RNAs	2 +*flrn*	2	2
in ***T. oceanica***, but not in *T. pseudonana*	***orf127***[a], *flrn*		
in ***T. pseudonana***, but not in *T. oceanica*		***petF***[a]	
in **IR of *T. oceanica***, but not in IR of *T. pseudonana*	***clpC, trnC, trnL***		
in ***P. tricornutum*** only			***tsf, syfB, acpP***

[a] presence or absence of *petF* and *orf127* in *T. oceanica* is the result of a gene transfer event.

Genes duplicated in the IR are counted once. The increased genome size of *T. oceanica* can conveniently be explained by an expansion of the IR region. *Orf127* and *flrn* are potential pseudogenes and are listed as additional gene content (+2).

gene exists in *T. pseudonana* and its genomic context is shown as a virtual destination for the gene transfer in *T. oceanica*. The chloroplast's contribution to the newly formed nuclear *PETF* gene must at least be the conserved part of the *petF* ORF encoding for ferredoxin. Indeed the 5' and 3' borders of the conserved main coding region of the *petF* gene in the *T. pseudonana* chloroplast match comparable sequences in the nuclear *PETF* gene of *T. oceanica* (Figure 2 center). The situation at the 3' end illustrates that only a few changes are needed to transform parts of the AT-rich stem-loop terminating many chloroplast transcription units into a functional polyA-signal (C_2) favouring termination of nuclear transcription. However, other conserved elements (C_1, C_3) are maintained as well and support the chloroplast origin of a *de novo* nuclear polyA-signal. For functional transcription, the 5' end of the *T. oceanica PETF* exon 2 had to acquire a splice acceptor site linking it to the exon 1, encoding for the cp transit peptide. We observe that the *T. pseudonana petF* sequence contains motifs in its coding region that may act as a functional splice acceptor site (coloured bars) in a nuclear genomic context. In summary, the *petF* sequence as present in the *T. pseudonana* cp genome already contains several sequence motifs that would permit an incorporation of the *petF* gene as a nearly functional exon into a nuclear genomic context, thereby facilitating the modular acquisition of a second exon and promoter through further rearrangements.

An important step towards establishing a nuclear *PETF* gene is a functional retargeting of the ferredoxin gene product for import into the chloroplast and, hence, the acquisition of a chloroplast transit peptide. In *T. oceanica* this transit peptide is encoded as a functional unit by a separate exon, suggesting the presence of a donor gene encoding for another protein of plastid destination as a potential source for such a module by duplication and exon shuffling events. An alignment of the ferredoxin gene products, showing the exon structure of the *T. oceanica PETF* and *LI818* (a gene encoding for a chlorophyll-binding light-harvesting protein) as such a possible donor gene for the transit peptide, is provided in Figure 3. The transit peptide in diatoms [18] has a length of approx. 28 aa and, while the *T. oceanica PETF* exon 2 encodes the highly conserved part of the protein, the part encoded by the 3' half of its exon 1 is less conserved and could have evolved from a sequence similar to the *LI818* exon 1.

Functional expression and differential regulation of *PETF*

Ferredoxin contains iron and is often replaced by the iron-free flavodoxin in iron-limited growth conditions. Relative quantification studies using RT-qPCR confirm the functional expression of the transferred *T. oceanica PETF* gene. The down-regulation of *PETF* under iron-limited growth conditions was similar to that of *PCY*, the gene encoding for plastocyanin, another photosynthetic protein involved in electron transfer (Figure 4). *PETF* and *PCY* show a

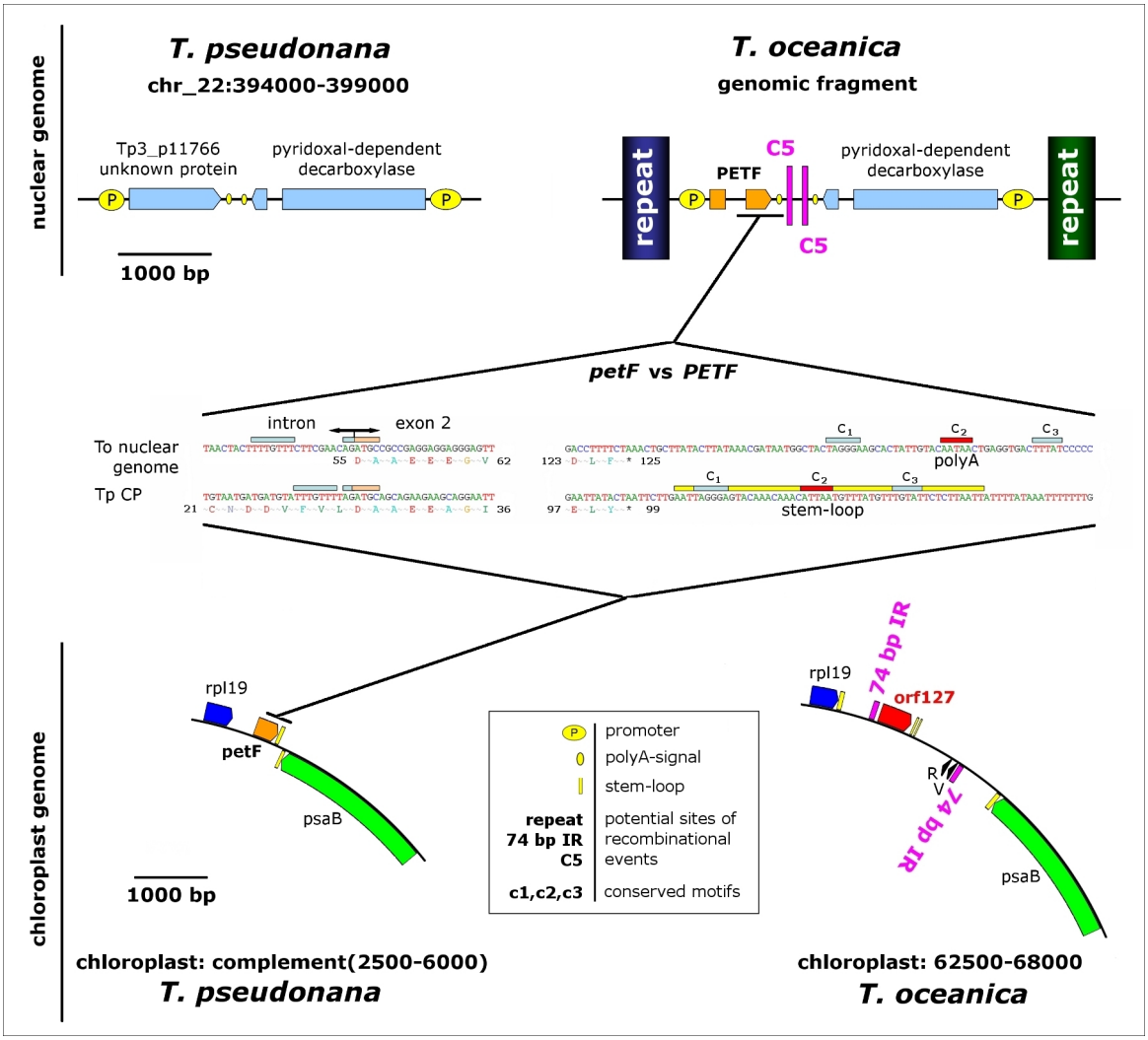


Figure 2 Genomic Context of the *Thalassiosira petF* Gene Transfer. The genomic context of the ferredoxin genes in *T. oceanica* and *T. pseudonana* indicates a recent gene transfer event. The *T. pseudonana* ***petF*** gene is part of the cp genome and is located between ***rpl19*** and ***psaB*** (bottom left), serving as a model for the genomic situation in a common ancestor of the two species. The *T. oceanica* ***PETF*** gene is found in the nuclear genome next to a *PDD* gene encoding for a pyridoxal-dependent decarboxylase (upper right) with a conserved ortholog present in *T. pseudonana* (upper left). The sites of excision (chloroplast) and insertion (nuclear fragment) in *T. oceanica* show major genome rearrangements in conjunction with the presence of a novel protein coding gene ***orf127*** in the chloroplast. Both nuclear and chloroplast sites in *T. oceanica* contain various forms of repeat elements („repeat", „74 bp IR") and motifs („C5") that point to the recombinational events underlying the complex multi-step gene transfer (upper and bottom right); these are not present at the comparable sites in *T. pseudonana*. An alignment between the *T. pseudonana* and *T. oceanica* ferredoxin genes reveals sequence similarities at what might be the 5' and 3' ends of the transferred DNA segment (center). The *T. pseudonana* ***petF*** coding sequence already contains a potential functional splice acceptor site near the N-terminus of the reading frame. A comparable situation in the ancestral *T. oceanica* cp genome would have facilitated the acquisition of an intron in a new genomic context leaving the major conserved part of the protein unaffected by the process (center left); the stem-loop (yellow bar), that serves as a transcriptional terminator for ***petF*** in the *T. pseudonana* cp genome, contains motifs that are retained in the nuclear genomic context of *T. oceanica* ***PETF***, one of them being slightly modified to become an eukaryotic polyA-signal (center right).

concerted down-regulation upon iron limitation as expected for constitutive photosynthesis genes, and in agreement with the observed down-regulation of photosynthesis in response to iron limitation. As expected, *FLDA* appears strongly up-regulated under iron-limited growth conditions, thereby supporting the idea of a mutual substitution of the ferredoxin and flavodoxin proteins.

Novel features of the *T. oceanica* cp genome

The genomic rearrangements in the *T. oceanica* cp genome accompanying the *petF* transfer led to a novel *orf127* appearing at the inferred *petF* excision site. This gene encodes a protein of 140 aa containing two predicted transmembrane helices (Figure 5a) with no similarity to any known proteins in the NCBI nr database.

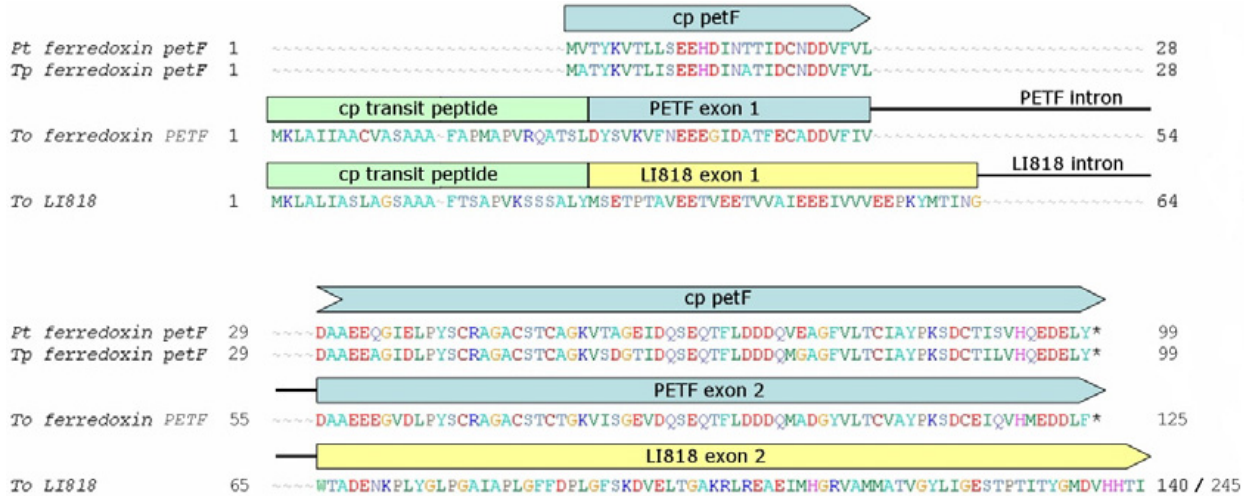


Figure 3 Chloroplast Retargeting of Ferredoxin. The *T. oceanica* nuclear ferredoxin protein sequence is aligned to its chloroplast counterparts from *T. pseudonana* and *P. tricornutum*. The ***PETF*** gene is composed of two exons, the second of which encodes the complete conserved part of the protein and is clearly of chloroplast origin, while the first exon contributes the chloroplast retargeting part and is of nuclear origin. Such a modular gene structure is not uncommon in genes encoding chloroplast targeted products as can be seen at the *T. oceanica* ***LI818*** gene belonging to the large group of nuclear ***FCP*** genes. The ***LI818*** gene encodes a cp transit peptide in a structurally separated exon and is presented as a potential donor gene candidate to provide the freshly transferred *PETF* DNA segment with the required targeting module.

The size and overall topology appear somewhat similar to the small transmembrane proteins from the FCP group of chlorophyll binding proteins, though sequence similarity or characteristic motifs are lacking. The gene is placed in a well-defined genomic context with an obvious ribosomal binding sequence GGGAGGG at -15 and two small inverted repeats serving as a putative rho-independent transcriptional terminator.

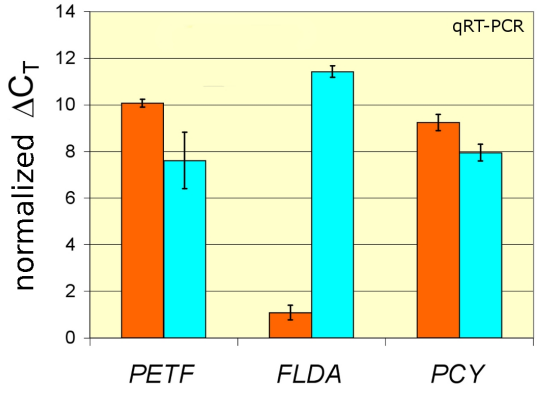


Figure 4 Functional Expression and Differential Regulation of *PETF* as a Function of Iron Level. Relative transcript abundances of the genes *FLDA*, *PETF* and *PCY* in triplicate iron-limited (light blue bars) and iron-replete cultures (orange bars) of *T. oceanica* were measured using a RT-qPCR approach with gene-specific primers pairs (Table 2). The normalized ΔC_T values is calculated as 22-ΔC_T, where 22 is an arbitrary number chosen to provide a comparable positive logarithmic scale for the transcript abundance of the three genes tested, where a high normalized ΔC_T represents a high transcript level.

As a second novel feature in the *T. oceanica* chloroplast genome we identified a duplication of the *ffs* RNA gene referred to as *flrn*. The ffs RNA adopts a stem-loop structure and is part of a signal recognition particle that might play a role at insertion of proteins into the inner chloroplast membrane [19]. An alignment of the *flrn* gene with the *ffs* genes of *T. oceanica*, *T. pseudonana* and *P. tricornutum* (Figure 5b) reveals that the *flrn* gene is truncated and contains a nucleotide polymorphism in the conserved loop region. Although the similarity between *flrn* and *ffs* is high, the truncated nature of *flrn* suggests that this sequence may represent a pseudogene.

Discussion

With new sequencing technologies emerging and an enormous increase of available sequence data, chloroplast genomes have drawn considerable interest not only for the purpose of phylogenetic reconstructions [20], but also for elucidating basic principles of genome organisation and dynamics of structural recombination events in comparative approaches [10]. Chloroplast genomes show a remarkable diversity among phototrophic eukaryotes [21]. However, their coding potential, ranging between 50 and 200 genes (Figure 6), is not even closely approaching the size and complexity of the plastid proteome needed to carry out the metabolic and physiological functions of the chloroplast. The plastid proteome is estimated to contain more than 700 proteins, many of uncertain structure and function [6]. Regulation of photosynthesis might even

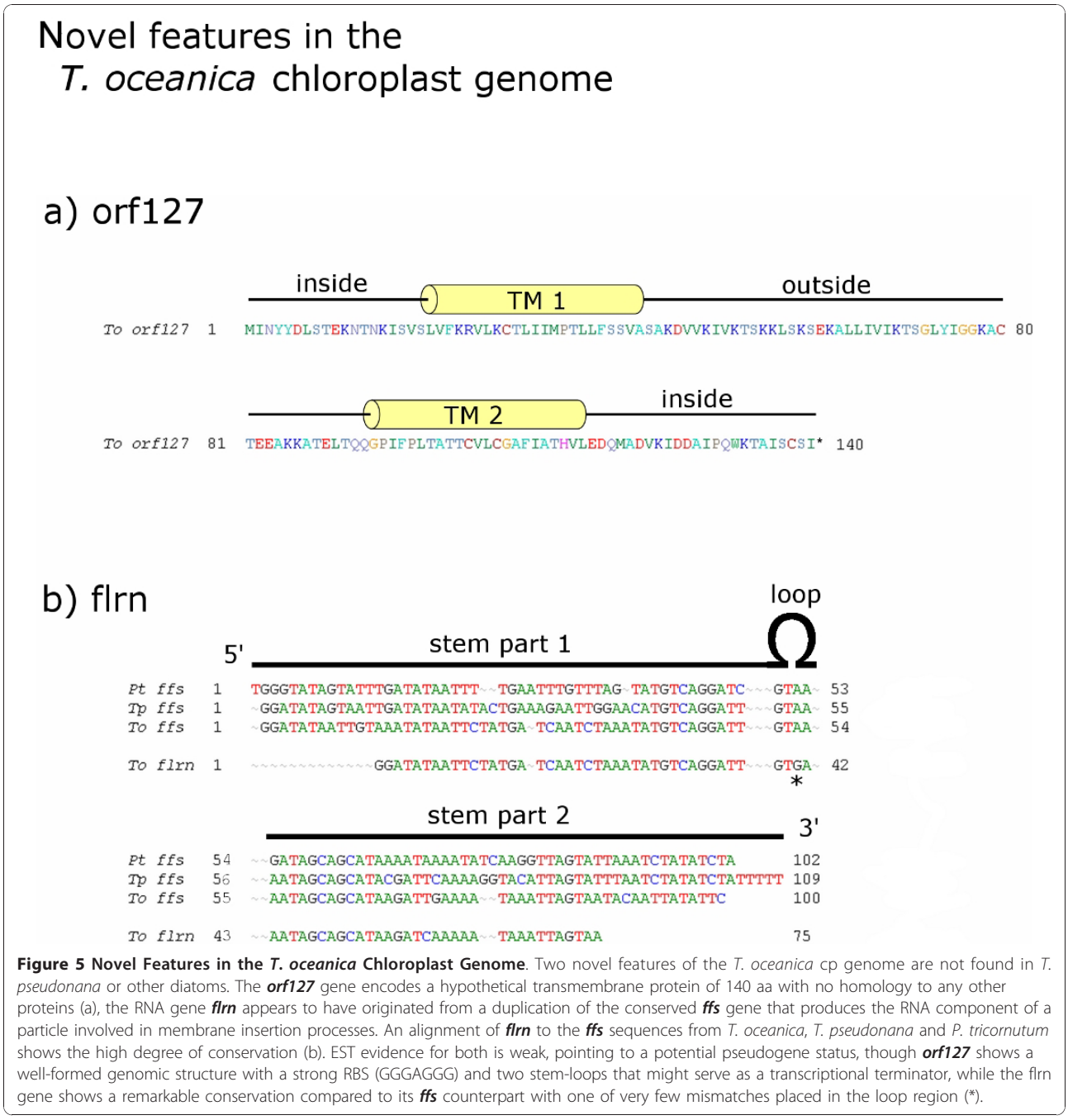


Figure 5 Novel Features in the *T. oceanica* Chloroplast Genome. Two novel features of the *T. oceanica* cp genome are not found in *T. pseudonana* or other diatoms. The ***orf127*** gene encodes a hypothetical transmembrane protein of 140 aa with no homology to any other proteins (a), the RNA gene ***flrn*** appears to have originated from a duplication of the conserved ***ffs*** gene that produces the RNA component of a particle involved in membrane insertion processes. An alignment of ***flrn*** to the ***ffs*** sequences from *T. oceanica*, *T. pseudonana* and *P. tricornutum* shows the high degree of conservation (b). EST evidence for both is weak, pointing to a potential pseudogene status, though ***orf127*** shows a well-formed genomic structure with a strong RBS (GGGAGGG) and two stem-loops that might serve as a transcriptional terminator, while the flrn gene shows a remarkable conservation compared to its ***ffs*** counterpart with one of very few mismatches placed in the loop region (*).

affect more than 4000 genes in *Arabidopsis* and novel proteins involved in elemental steps of photosynthesis are still being discovered [22].

The chloroplast genome is a dynamic structure in which the collinear binding of the inverted repeat regions likely leads to a handle-like structure isolating the LSC and SSC regions in separate domains and serving as a basis for occasional recombinational events that result in an inversion of the SSC region [23]. The weak conservation of synteny between *T. oceanica* and *T. pseudonana* cp genomes implies high structural dynamics of the circular genomic molecule, best explained by frequent inversion events. The remarkable lack of genomic rearrangements across the borders of the single copy regions, as observed by comparison between *T. oceanica* and *T. pseudonana*, could be explained by the structural separation of the LSC and SSC in a handle-like structure and supports this structural model.

The establishment of the primary endosymbiont as a chloroplast led to the loss of many of the endosymbiont's

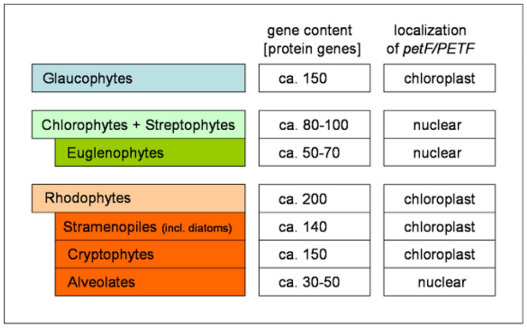


Figure 6 Subcellular Localization of the *petF/PETF* Gene in the Phylogenetic Context. A comparison of the chloroplast gene content between the main phylogenetic groups shows a reduced coding potential in the green line (Chlorophytes + Streptophytes, Euglenophytes) and in the Alveolates indicating extensive gene transfer to the nucleus or gene loss. The localization of the ferredoxin ***petF/PETF*** gene correlates with the extent of genome reduction, hence, ***petF*** is generally retained in the larger cp genomes of Glaucophytes, Rhodophytes, Stramenopiles (incl. diatoms) and Cryptophytes. Protein gene numbers are taken from the Chloroplast Genome Database [54] and represent the majority of species in the respective groups, though exceptions can deviate from the given range. Phylogenetic groups originating from secondary endosymbiosis are indented.

genes or their transfer to the host's nucleus. Red algal and Glaucophyte chloroplasts retain about 200 protein coding genes, while most members of the green lineage exhibit a further genome reduction to less than 100 genes, indicating functional transfer of several essential genes to the nuclear genome. The substantial retention of coding potential in the primary plastids of the red algal lineage has been hypothesized to have an impact on the chloroplast portability during secondary endosymbiosis, preferentially facilitating further endosymbiotic events of autotrophs from the red lineage over their more reduced green counterparts [7]. Gene content of extant cp genomes derived from the red-algal lineage indicate that further gene loss or transfer proceeded after secondary endosymbiosis events, leading to a gene content of approx. 140 genes in the Stramenopiles and Cryptophytes. Stronger reductions can be found within the Apicomplexa and Dinophytes, referred to as Alveolates, perhaps because the known members of the Apicomplexa and several of the Dinophytes are non-photosynthetic organisms, and sometimes parasitic [24]. All of these subgroups are regarded as descendants of an ancient red algal endosymbiont and its host [7].

Occasional free release of cp genomic DNA upon chloroplast lysis is considered an important source of chloroplast DNA for integration into the nuclear genome. This mechanism can operate only in organisms containing multiple plastids. Indeed the higher degree of gene retention in *P. tricornutum* compared to *Thalassiosira* species might be partially explained by the presence of a single, unique chloroplast in *P. tricornutum* which would not be available for lysis in the context of the living cell. *T. oceanica* and *T. pseudonana* cells on the other hand are equipped with approx. 4 plastids each [25,26].

As for other evolutionary processes, the transfer of chloroplast-encoded genes to the nuclear genome is difficult, if not impossible, to observe. Experimental approaches with transgenic tobacco have demonstrated that plastid genes can be successfully transferred to the nuclear genome [27], but most of the evidence for the inferred transfers is derived from phylogenetic and phylogenomic arguments [20]. Why certain genes are preferentially retained or transferred is puzzling and several theories have emerged ranging from a simple economic hypothesis in favour of gene transfer [28] to the CORR (CO-location for Redox Regulation) hypothesis in favour of gene retention [29]. The ferredoxin *petF* gene is not essentially retained in the CP genome, as a nuclear location is observed in several phylogenetic groups. So far, with the exception of *T. oceanica*, the *petF* gene has been retained in large cp genomes (Figure 6) such as the ones from red algae and descendants (including diatoms), and has generally been transferred to the nucleus in groups with reduced cp genomes such as Chlorophytes and Streptophytes. Several intrinsic features of the *petF* gene, conserved in the cp genomes of the closely related species *T. pseudonana* and *T. weissflogii* [30], may have facilitated the functional transfer of *petF* from the chloroplast to the nuclear genome in *T. oceanica*. Thus, in the coastal species *T. pseudonana* and *T. weissflogii*, the *petF* gene transfer may be waiting to happen as well, given appropriate environmental selection pressures. However, at this point, the establishment of the *PETF* gene in the nuclear genome of *T. oceanica* represents an exception within the diatoms for which genome information is available so far.

It has been demonstrated that the photosynthetic architecture in *T. oceanica* is better adapted to iron-limited areas than its coastal counterpart *T. weissflogii* [5]. Tolerance to iron limitation most likely arises from a combination of several genetic adaptations that contribute to a better streamlining of the photosynthetic apparatus towards low iron requirements. The low abundance of PSI relative to PSII in *T. oceanica* is clearly important in reducing the cellular iron requirements [5]. The substitution of the iron-requiring cytochrome c_6 by the copper-containing plastocyanin in *T. oceanica* is an additional strategy to further reduce photosynthetic iron requirements, and, hence, the iron quota of this diatom [31]. Likewise, the flavodoxin

expression is up-regulated upon iron limitation in many diatom species [32] while a simultaneous down-regulation of the expression of the ferredoxin *petF* gene is observed (e.g. [33]). In our study, we could confirm the expression and iron-dependent regulation of the ferredoxin *PETF*, the plastocyanin *PCY* and the flavodoxin *FLDA* gene in *T. oceanica* (Figure 4). The concerted down-regulation of *PETF* and *PCY* upon iron limitation is expected as part of the general down-regulation of the photosynthetic apparatus under iron limitation. In contrast, *FLDA* is strongly up-regulated in order to substitute ferredoxin with flavodoxin, consequently keeping intact the electron-transport interface between membrane-bound light reactions and dark reactions in the cp stroma.

The contrast in photosynthetic physiology between *T. oceanica* and its coastal relatives *T. pseudonana* and *T. weissflogii* make an attractive case to infer an adaptive significance for the transfer of *petF*, an iron-regulated cp gene, to the nuclear genome. Whether the transfer of *petF* to the nuclear genome is simply a byproduct of evolutionary trends towards chloroplast genome reduction or truly confers an ecological advantage with respect to the response to iron remains uncertain, though the observed high tolerance of *T. oceanica* to severe iron limitation relative to its close relatives *T. pseudonana* and *T. weissflogii* suggests that the latter hypothesis is worthy of further investigation.

Single gene transfers are the elemental steps of cp genome reduction and may confer benefits to single species in the context of niche adaptation. It is tempting to speculate that larger phylogenetic groups whose members share a reduced cp genome (as in the green lineage) likewise emerged from a founder species that profited from the benefits of cp genome reduction and improved nuclear control over organelle function. Centralized and synchronized regulation of cp metabolism is assumed to be a potential driving force for intracellular gene transfers. Organisms that already experienced large-scale cp genome reduction should benefit as well from an improved regulation of the transferred genes. The uniform genomic situation in larger phylogenetic groups raises the question whether such competitive advantages might even apply to these groups as a whole. A comparative evaluation of this effect between the red and green lineage is complicated by the interference with different extents of gene losses in both groups as well as the presence of distinct types of photosynthetic physiology in general. However, the terrestrial environment has been conquered exclusively by members of the green lineage, and prerequisite for this achievement might (at least partially) have been the improvement of regulatory capacities linked to cp genome reduction by large scale gene transfer. The settlement of the land represented the occupation of a new ecological niche rich in abiotic stresses of a novel type. It remains to be elucidated to which extent such specific environmental stresses exert a selective pressure favouring gene transfer events [34], ultimately leading to competitive advantage and enhanced fitness.

Conclusions

Although the chloroplast genomes of some closely related marine phytoplankton species have been sequenced, the differences between species within a genus have been small and restricted to gene reshuffling. Our findings, reporting a traceable single gene transfer from the chloroplast to the nuclear genome, are unique so far, in part because of the availability of both plastid and nuclear genome sequences for *T. oceanica* and *T. pseudonana*. The example of *petF* shows that chloroplast and nuclear genomes are of remarkable plasticity. Whether or not the gene transfer described for *T. oceanica* confers a competitive advantage still needs to be assessed through experimental approaches. Future analyses of cp genomes from a wider range of ecologically diverse species will likely reveal other surprising patterns of cp gene content, loss and regulation, and further enhance our understanding of their impacts on the evolutionary fitness of species.

Methods

Strains and Cultures

T. oceanica Hasle [25] strain CCMP1005 was grown from an axenic clonal isolate, obtained from serial dilutions of a stock culture to extinction. *T. oceanica* cells were grown in 8 l batch cultures using iron-free f/2 nutrients [35] in ASW (artificial seawater medium [36]) supplied with 10 μM $FeCl_3$ at 100 μE, 25°C and a 14/10 h light/dark cycle. Cells were harvested by filtration on 47 mm 5 μ-PC [polycarbonate]-filters, resuspended into a small volume of media, followed by centrifugation at 4°C for 10 min at 11000 rpm. Cell pellets were frozen in liquid N_2 and stored at -80°C. Genome comparison was conducted with the genome data available at JGI and NCBI for *T. pseudonana* Hasle & Heimdal CCMP1335 [26,37] and *Phaeodactylum tricornutum* Bohlin CCAP1055/1 [38,39].

Nucleic Acid Extraction and Sequencing

Total genomic DNA for sequencing of the *T. oceanica* genome was extracted from nutrient-replete cells using the QIAGEN DNeasy kits. The quality of nucleic acids was assessed by NanoDrop UV absorption profiles and agarose gel electrophoresis. Next generation 454 sequencing technology [40] was applied to the gDNA as follows. After mechanical shearing, specific sequencing adaptors were ligated and the genomic DNA fragments were shotgun sequenced using massively parallel

pyrosequencing on a 454 gs-flx instrument (Roche, Penzberg, Germany) according to the manufacturer's protocol. The resulting libraries were sequenced on a gs-flx sequencer using the standard manufacturer's protocol.

Cp Genome Sequence Assembly and Gap Closure

1.2 Mio flx pyrosequencing reads were assembled into contigs with the TGICL assembler using the CAP3 algorithm [41]. The quality of the resulting contigs was manually confirmed by inspection using the CLVIEW cluster viewer program. Using the local BLAST package from NCBI [42], we identified 9 contigs with high sequence coverage as elements of the cp genome. Additional information extracted from the contig ends of the *ace*-file of the original assembly, enabled the manual assembly of the cp genome to near completeness. The two remaining small gaps were targeted by PCR amplification of the contig ends and demonstrated physical continuity of the gap regions. Bridging fragments were cloned in a TOPO cloning vector and their sequences determined by Sanger sequencing.

Sequence Analysis and Annotation

The assembled chloroplast genome sequence was analyzed by BLAST against the related *T. pseudonana* cp genome sequence, the NCBI nr (non-redundant) protein database and the NCBI CDD Conserved Domain Database. The BLAST analysis revealed few obvious artificial frameshifts within the original contigs, allowing correction of the chloroplast scaffold for single nucleotide errors placed in low complexity regions of single nucleotide repeats that generally appear to be critical in 454 data. The final chloroplast scaffold was annotated using Artemis [43] and the submission software Sequin [44]. Inverted repeats at the 3' end of genes representing putative rho-independent transcriptional terminators [45,46] were identified with the EINVERTED tool from the EMBOSS software package [47]. Ribosomal Binding Sites (RBS) were determined manually from similarity to the AGGAGGT consensus sequence [48] and close proximity to the respective translation start. The contig containing the nuclear ferredoxin gene *PETF* was assembled manually from raw genomic 454 reads using local BLAST, database sequence retrieval and BioEdit [49]. Gene modelling was done with the GENSCAN webserver [50] and confirmed by NCBI blastx against the nr protein database. The derived transcript encoded a protein with high homology to *T. pseudonana* and *P. tricornutum* petF protein orthologs.

Circular Map Construction

The circular genomic map was constructed from the primary embl-annotation file using CGVIEW [51] as follows: The *embl*-file was converted into an xml-file with the perl-script cgview_xml_builder.pl, which is enclosed in the CGVIEW package; the xml-file containing the formatting details for the circular map was then customized manually by adding functional categorization and appropriate gene symbol shapes and dimensions. The circular map was constructed from the *xml*-file as *png*-graphic, using the CGVIEW main function. Gene names were added using the open source graphics program GIMP [52].

RT-qPCR data

Based on the transcript sequences of the 18S-rDNA, *PETF*, *FLDA* and *PCY* genes derived from the assembly data, sets of primers were designed and optimized to detect gene specific amplicons of approx. 100 bp with uniformly high amplification efficiency (>95%, Table 2). A local BLASTN analysis of the primers against all sequences available for *T. oceanica* confirmed the specificity of the primers for their respective genes. cDNA template was prepared from 1 μg RNA by reverse transcription using the QuantiTect Rev. Transcription Kit (QIAGEN), followed by digestion of residual DNA using the included gDNA wipeout reagents. The cDNA was diluted to 0.5 ng μl^{-1} and 2.5 ng were used per qPCR reaction run on an ABI Prism 7000 (Applied Biosystems). Cycling conditions were 2 min at 50°C (once), 2 min at 95°C (once), and 40 cycles of 95°C for 0:15 min, followed by 0:30 min at 60°C. The qPCR mixtures contained 12.5 μl SYBR qPCR SUPERMIX W/ROX (Invitrogen), 0.5 μl of 10 μM forward and reverse primer each, 6.5 μl H_2O and 5 μl of cDNA template. Gene expression was assessed as the mean from the C_T values of 2-4 replicate reactions at a threshold level of 0.2. Relative expression of genes with respect to the 18S-rDNA gene was calculated using the ΔC_T method (ΔC_T [geneX] = C_T[gene*X*] - C_T[18S]). Final data are presented as the mean ± s.e.m. of the ΔC_T values from three biological replicates.

Nomenclature of Gene Names

The nomenclature of gene names follows the recommendations for *Chlamydomonas reinhardtii* [53]. Gene names are typed in italic and uppercase for nuclear genes (*PETF*) or italic and lower case for organelle genes (*petF*).

Table 2 Oligonucleotide Primers used for RT-qPCR Analyses

Gene	Forward Primer (5'→3')	Reverse Primer (5'→3')
PETF	AGGCCACCTCCCTCGACTAC	GCGGCATCGACGATGAAG
FLDA	CCGGCCTTTTCTACTCGACC	TTGACGTCTCCGATGTCCTTC
PCY	CTCCGCCCCTGCTTACG	TCCCTTGCAGACAGTGACCTT

Acknowledgements
We thank Prof. Stefan Rose-John (Department of Biochemistry, Christian-Albrechts-Universität, Kiel, Germany) for advice in the isolation of nuclear genomic DNA from *T. oceanica* and the access to his laboratory and equipment. This work was supported in part by a DFG grant to JLR (RO2138/6-1) and by the DFG Cluster of Excellence Future Ocean (EXC 80). Prof. T. Bosch and Dr. Georg Hemmrich provided help with the initial cp contig assembly. We thank Tania Klüver for help with the laboratory experiment and culturing of the algae.

Author details
[1]Leibniz Institute of Marine Sciences at Kiel University IFM-GEOMAR, Kiel, Germany. [2]Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany.

Authors' contributions
ML performed the gap closure cloning, the cp genome assembly and annotation, the analysis of the *petF* gene transfer, prepared the figures and wrote a major portion of the manuscript.
ASR cultured the algae, carried out the RT-qPCR work and commented on the manuscript.
MS prepared the gDNA libraries and performed the 454 sequencing.
SS provided the sequencing technology and supply.
PR coordinated the sequencing and contributed to manuscript writing.
JLR coordinated the study, isolated the *T. oceanica* gDNA, and made major contributions to discussion and manuscript writing.
All authors read and approved the final manuscript.

Received: 1 June 2010 Accepted: 20 December 2010
Published: 20 December 2010

References

1. Dugdale RC, Wilkerson FP: **Silicate regulation of new production in the equatorial Pacific upwelling.** *Nature* 1998, **391**:270-273.
2. Finkel ZV, Katz ME, Wright JD, Schofield OME, Falkowski PG: **Climatically driven macroevolutionary patterns in the size of marine diatoms over the cenozoic.** *Proc Natl Acad Sci USA* 2005, **102**:8927-8932.
3. Martin JH, Coale KH, Johnson KS, Fitzwater SE, Gordon RM, Tanner SJ, Hunter CN, Elrod VA, Nowicki JL, Coley TL, Barber RT, Lindley S, Watson AJ, Van Scoy K, Law CS, Liddicoat MI, Ling R, Stanton T, Stockel J, Collins C, Anderson A, Bidigare R, Ondrusek M, Latasa M, Millero FJ, Lee K, Yao W, Zhang JZ, Friederich G, Sakamoto C, Chavez F, Buck K, Kolber Z, Greene R, Falkowski P, Chisholm SW, Hoge F, Swift R, Yungel J, Turner S, Nightingale P, Hatton A, Liss P, Tindale NW: **Testing the Iron Hypothesis in Ecosystems of the Equatorial Pacific-Ocean.** *Nature* 1994, **371**:123-129.
4. Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJ, Follows M, Harvey M, Lancelot C, Levasseur M, Owens NP, Pollard R, Rivkin RB, Sarmiento J, Schoemann V, Smetacek V, Takeda S, Tsuda A, Turner S, Watson AJ: **Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions.** *Science* 2007, **315**:612-617.
5. Strzepek RF, Harrison PJ: **Photosynthetic architecture differs in coastal and oceanic diatoms.** *Nature* 2004, **431**:689-692.
6. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S: **The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions.** *Curr Biol* 2004, **14**:354-362.
7. Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR: **The evolution of modern eukaryotic phytoplankton.** *Science* 2004, **305**:354-360.
8. Quigg A, Finkel ZV, Irwin AJ, Rosenthal Y, Ho TY, Reinfelder JR, Schofield O, Morel FMM, Falkowski PG: **The evolutionary inheritance of elemental stoichiometry in marine phytoplankton.** *Nature* 2003, **425**:291-294.
9. Timmis JN, Ayliffe MA, Huang CY, Martin W: **Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes.** *Nat Rev Genet* 2004, **5(2)**:123-135.
10. Oudot-Le Secq M-P, Grimwood J, Shapiro H, Armbrust EV, Bowler C, Green BR: **Chloroplast genomes of the diatoms Phaeodactylum tricornutum and Thalassiosira pseudonana: comparison with other plastid genomes of the red lineage.** *Mol Genet Genomics* 2007, **277**:427-439.
11. Martin W: **Gene transfer from organelles to the nucleus: Frequent and in big chunks.** *Proc Natl Acad Sci USA* 2003, **100**:8612-8614.
12. Richly E, Leister D: **NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs.** *Mol Biol Evol* 2004, **21**:1972-1980.
13. Bock R, Timmis JN: **Reconstructing evolution: gene transfer from plastids to the nucleus.** *BioEssays* 2008, **30**:556-566.
14. Sheppard AE, Timmis JN: **Instability of Plastid DNA in the Nuclear Genome.** *PLoS Genet* 2009, **5**:e1000323.
15. Fujita K, Ehira S, Tanaka K, Asai K, Ohta N: **Molecular phylogeny and evolution of the plastid and nuclear encoded cbbX genes in the unicellular red alga Cyanidioschyzon merolae.** *Genes Genet Syst* 2008, **83**:127-133.
16. Ueda M, Nishikawa T, Fujimoto M, Takanashi H, Arimura S, Tsutsumi N, Kadowaki K: **Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal.** *Mol Biol Evol* 2008, **25**:1566-1575.
17. Jackson FR: **Prokaryotic and Eukaryotic Pyridoxal-Dependent Decarboxylases Are Homologous.** *J Mol Evol* 1990, **31**:325-329.
18. Kilian O, Kroth PG: **Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids.** *Plant J* 2005, **41**:175-183.
19. Ulbrandt ND, Newitt JA, Bernstein HD: **The E-coli signal recognition particle is required for the insertion of a subset of inner membrane proteins.** *Cell* 1997, **88**:187-196.
20. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D: **Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms.** *Science* 2009, **324**:1724-1726.
21. Simpson CL, Stern DB: **The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications.** *Plant Physiol* 2002, **129**:957-966.
22. Eberhard S, Finazzi G, Wollman F-A: **The Dynamics of Photosynthesis.** *Annu Rev Genet* 2008, **42**:463-515.
23. Cattolico RA, Jacobs MA, Zhou Y, Chang J, Duplessis M, Lybrand T, McKay J, Ong HC, Sims E, Rocap G: **Chloroplast genome sequencing analysis of Heterosigma akashiwo CCMP452 (West Atlantic) and NIES293 (West Pacific) strains.** *BMC Genomics* 2008, **9**:211.
24. Hackett JD, Anderson DM, Erdner DL, Bhattacharya D: **Dinoflagellates: A remarkable evolutionary experiment.** *Am J Bot* 2004, **91**:1523-1534.
25. Hasle GR: **The Marine, Planktonic Diatoms Thalassiosira-Oceanica Sp-Nov and Thalassiosira-Partheneia.** *J Phycol* 1983, **19**:220-229.
26. Hasle GR, Heimdal BR: **Some species of the centric diatom genus Thalassiosira studied in the light and electron microscopes.** *Beih zur Nova Hedwigia* 1970, **31**:543-581.
27. Huang CY, Ayliffe MA, Timmis JN: **Direct measurement of the transfer rate of chloroplast DNA into the nucleus.** *Nature* 2003, **422**:72-76.
28. Douglas AE, Raven JA: **Genomes at the interface between bacteria and organelles.** *Philos Trans R Soc B* 2003, **358**:5-17.
29. Allen JF, Puthiyaveetil S, Ström J, Allen CA: **Energy transduction anchors genes in organelles.** *BioEssays* 2005, **27**:426-435.
30. Gueneau P, Morel F, LaRoche J, Erdner D: **The petF region of the chloroplast genome from the diatom *Thalassiosira weissflogii*: sequence, organization and phylogeny.** *Eur J Phycol* 1998, **33**:203-211.
31. Peers G, Price NM: **Copper-containing plastocyanin used for electron transport by an oceanic diatom.** *Nature* 2006, **441**:341-344.
32. LaRoche J, Murray H, Orellana M, Newton J: **Flavodoxin Expression as an Indicator of Iron Limitation in Marine Diatoms.** *J Phycol* 1995, **31**:520-530.
33. Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C: **Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation.** *Proc Natl Acad Sci USA* 2008, **105**:10438-10443.
34. Cullis CA, Vorster BJ, Van Der Vyver C, Kunert KJ: **Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants?** *Ann Bot* 2009, **103**:625-633.
35. Guillard RRL, Ryther JH: **Studies of marine planktonic diatoms. I. Cyclotella nana Hustedt and Detonula confervacea Cleve.** *Can J Microbiol* 1962, **8**:229-239.
36. Goldman JC, McCarthy JJ: **Steady-State Growth and Ammonium Uptake of a Fast-Growing Marine Diatom.** *Limnol Oceanogr* 1978, **23**:695-703.
37. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N,

Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS: **The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism.** *Science* 2004, **306**:79-86.

38. Bohlin K: **Zur Morphologie und Biologie einzelliger Algen.** *Öfvers af K Vet Acad Förhandl Stockholm* 1897, **54**:519-522.
39. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kröger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jézéquel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV: **The Phaeodactylum genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**:239-244.
40. Wolff J, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D: **Nucleotide divergence versus gene expression differentiation: 454 transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow.** *Mol Ecol* 2010, **19(Suppl 1)**:162-75.
41. Pertea G, Huang XQ, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J: **TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets.** *Bioinformatics* 2003, **19**:651-652.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool.** *J Mol Biol* 1990, **215**:403-410.
43. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16**:944-945.
44. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2006, **34**:D16-D20.
45. Farnham PJ, Platt T: **Rho-Independent Termination - Dyad Symmetry in DNA Causes Rna-Polymerase to Pause During Transcription Invitro.** *Nucleic Acids Res* 1981, **9**:563-577.
46. Wilson KS, Von Hippel PH: **Transcription Termination at Intrinsic Terminators - the Role of the RNA Hairpin.** *Proc Natl Acad Sci USA* 1995, **92**:8793-8797.
47. Rice P, Longden I, Bleasby A: **EMBOSS: The European molecular biology open software suite.** *Trends Genet* 2000, **16**:276-277.
48. Shine J, Dalgarno L: **Determinant of Cistron Specificity in Bacterial Ribosomes.** *Nature* 1975, **254**:34-38.
49. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95-98.
50. **The GENSCAN Web Server at MIT.** [http://genes.mit.edu/GENSCAN.html].
51. Stothard P, Wishart DS: **Circular genome visualization and exploration using CGView.** *Bioinformatics* 2005, **21**:537-539.
52. **GNU Image Manipulation Program.** [http://www.gimp.org/].
53. **Chlamy Center - An Online Informatics Resource for Chlamydomonas.** [http://www.chlamy.org/nomenclature.html].
54. Cui L, Veeraraghavan N, Richter A, Wall K, Jansen RK, Leebens-Mack J, Makalowska I, dePamphilis CW: **ChloroplastDB: the chloroplast genome database.** *Nucleic Acids Res* 2006, **34**:D692-D696.

doi:10.1186/1471-2164-11-718
Cite this article as: Lommer *et al.*: Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics* 2010 **11**:718.

3.4 T. oceanica Genome Paper

Title:

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

Manuscript submitted to: Genome Biology

Author list:

Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, LaRoche J.

Abstract:

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

RESULTS: The combined efforts in genomics, transcriptomics and proteomics reveal an unexpected metabolic flexibility in response to iron availability for *Thalassiosira oceanica* CCMP1005. The complex response comprises cellular retrenchment as well as remodeling of bioenergetic pathways, where iron-rich photosynthetic proteins are degraded, whereas iron-rich mitochondrial proteins are not. As a consequence of iron-deprivation the photosynthetic machinery undergoes a remodelling to adjust the light energy utilization with the overall decrease in photosynthetic electron transfer complexes.

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

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

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

[1]Helmholtz Centre for Ocean Research Kiel (GEOMAR), Kiel, Germany
[2]Institute of Plant Biology and Biotechnology, University of Münster, Münster, Germany
[3]Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany
[4]Department of Biology, University of Bergen, Bergen, Norway
[5]Estonian Biocentre, Tartu, Estonia
[6]Institute of Physiology, Christian-Albrechts-University Kiel, Kiel, Germany
[7]Faculty of Computer Science, Dalhousie University, Nova Scotia B3H 1W5, Canada

[§]Corresponding author
[*]Current address: Department of Biology Dalhousie University, Nova Scotia B3H 1W5, Canada

Email addresses:
ML: mlommer@geomar.de
MS: micha.specht@gmail.com
ASR: sroy@geomar.de
LK: l.kraemer@ikmb.uni-kiel.de
RA: reidar.andreson@ut.ee
MAG: m.gutowska@physiologie.uni-kiel.de
JW: j.wolf@imb.uq.edu.au
SVB: s.bergner@uni-muenster.de
MBS: m.schilhabel@ikmb.uni-kiel.de
UCK: u.klostermeier@ikmb.uni-kiel.de
RGB: beiko@cs.dal.ca
PR: p.rosenstiel@mucosa.de
MH: mhippler@uni-muenster.de
JLR: jlaroche@geomar.de

Abstract

Background

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

Results

The combined efforts in genomics, transcriptomics and proteomics reveal an unexpected metabolic flexibility in response to iron availability for *Thalassiosira oceanica* CCMP1005. The complex response comprises cellular retrenchment as well as remodeling of bioenergetic pathways, where iron-rich photosynthetic proteins are degraded, whereas iron-rich mitochondrial proteins are not. As a consequence of iron-deprivation, the photosynthetic machinery undergoes a remodelling to adjust the light energy utilization with the overall decrease in photosynthetic electron transfer complexes.

Conclusions

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

Keywords

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

Background

Diatoms are important primary producers in the ocean [1], contributing approximately 40% to global marine productivity. Although diatoms often dominate phytoplankton communities in nutrient-rich ecosystems, members of this diverse group are also adapted to survive and persist in nutrient-limited conditions. The development of large diatom blooms upon nutrient resupply demonstrates the metabolic plasticity inherent to their ability to recover rapidly from nutrient limitation.

Iron is an essential nutrient for all organisms and in particular for photoautotrophic organisms. It functions as a powerful electron carrier in iron-sulfur- and heme-containing proteins and as such is a required component of the photosynthetic apparatus. Solubility of iron in seawater is low and phytoplankton growth in marine habitats is often limited by iron availability. This is best illustrated in high-nitrate low-chlorophyll (HNLC) regions, remote oceanic areas that lack any form of regular iron supply and suffer from a persistent shortage of this micronutrient. Here, although other commonly limiting nutrients like nitrate or phosphate are present at high concentrations, primary productivity – and biomass as a whole – is low [2].

Numerous large-scale iron fertilization experiments have confirmed that iron is the limiting nutrient in HNLC regions [3]. Phytoplankton blooms induced by iron fertilization were dominated by diatoms and carbon export to the deep-sea floor could be observed in some cases. The strong response of diatoms to the input of iron in HNLC regions has been a motivation for exploring large-scale iron fertilization as a possible bioengineering strategy to sequester CO_2 into the ocean in HNLC regions, which are otherwise rich in macronutrients.

Genome projects on the model organisms *Thalassiosira pseudonana* [4] and *Phaeodactylum tricornutum* [5], have already generated a wealth of insights into the metabolic complexity of diatoms [6], a consequence of the secondary endosymbiosis event that gave rise to this group [7]. This secondary endosymbiosis brought together the benefits of a heterotrophic host and the "red"-type photosynthesis of red alga cells, which already have an elemental composition low in iron [8].

The impact of iron availability on phytoplankton growth has lead to the evolution of strategies to counteract iron limitation. Well established parts of the low-iron response found in diverse phytoplankton species are the reduction of the chloroplast system, the corresponding development of a chlorotic phenotype, compensation mechanisms (replacement of iron-rich elements with iron-poor substitutes) and the activation of high-affinity iron-uptake systems [9]. The substitution of ferredoxin by flavodoxin [10], the use of

plastocyanin instead of cytochrome c_6 [11] and a variant stoichiometry of photosynthetic complexes [12] are notable adaptive strategies to facilitate diatom growth at low iron.

Oceanic and neritic phytoplankton species can be distinguished from each other by their growth characteristics and their tolerance to nutrient-limitation [13]. Unlike many other *Thalassiosira* species that are predominantly found in coastal waters, *T. oceanica* is adapted to oligotrophic conditions and is highly tolerant to iron limitation in particular. Therefore, we chose *T. oceanica* CCMP1005 as a model for a comprehensive analysis of its low-iron response in the context of genomic information.

Here, we explore the complex cellular response of *T. oceanica* to low-iron conditions with genomics, transcriptomics and proteomics approaches complemented by RT-qPCR analyses. We present a metabolic reconstruction of the iron limitation response based on the transcriptomics data from cells grown under iron-deplete vs iron-replete conditions. A metabolic isotope labeling approach using $^{14}N/^{15}N$ was established for *T. oceanica* and showed the response to iron limitation at the protein expression level in a marine diatom for the first time. General characteristics of the "diatom" low-iron response and its ecological implications are discussed, as well as the constraints for species-specific adaptations to low-iron environments.

Results

Characteristics of the T. oceanica *genome*

The genome of the centric diatom *T. oceanica* CCMP1005 (Fig. 1) was *de novo* assembled from 725 Mb of Roche 454 sequence read information, using nuclear genomic (g)DNA of an axenic clonal culture as substrate [14]. The current assembly version comprises 51656 contigs of total size 92.15 Mb at N50 = 3623 (i.e. 50% of the genomic sequence information is present as contigs equal or larger than 3623 bases). From a median 8.7-fold coverage of long contigs (>= 10 Kb) we estimated a true haploid nuclear genome size of 81.6 Mb, suggesting some redundancy in the current assembly. This estimate is in good agreement with the 159 Mb measured by van Dassow et al. [15] for the diploid G1 DNA content The gene finder tool AUGUSTUS [16] predicts 37921 protein gene models that cluster into a non-redundant set of 29306 models including pseudogenes and short ORFs. 10109 models have BLAST hits to NCBI nr protein database at a conservative E-value cut-off of 1.0E-30 and thus are more indicative of the expected true protein-coding gene number (i.e. expressed genes excl. pseudogenes and short ORFs). Best BLAST hits are listed in Additional file 1 (Table S1 "AUGUSTUS_models_vs_nr+CDD").

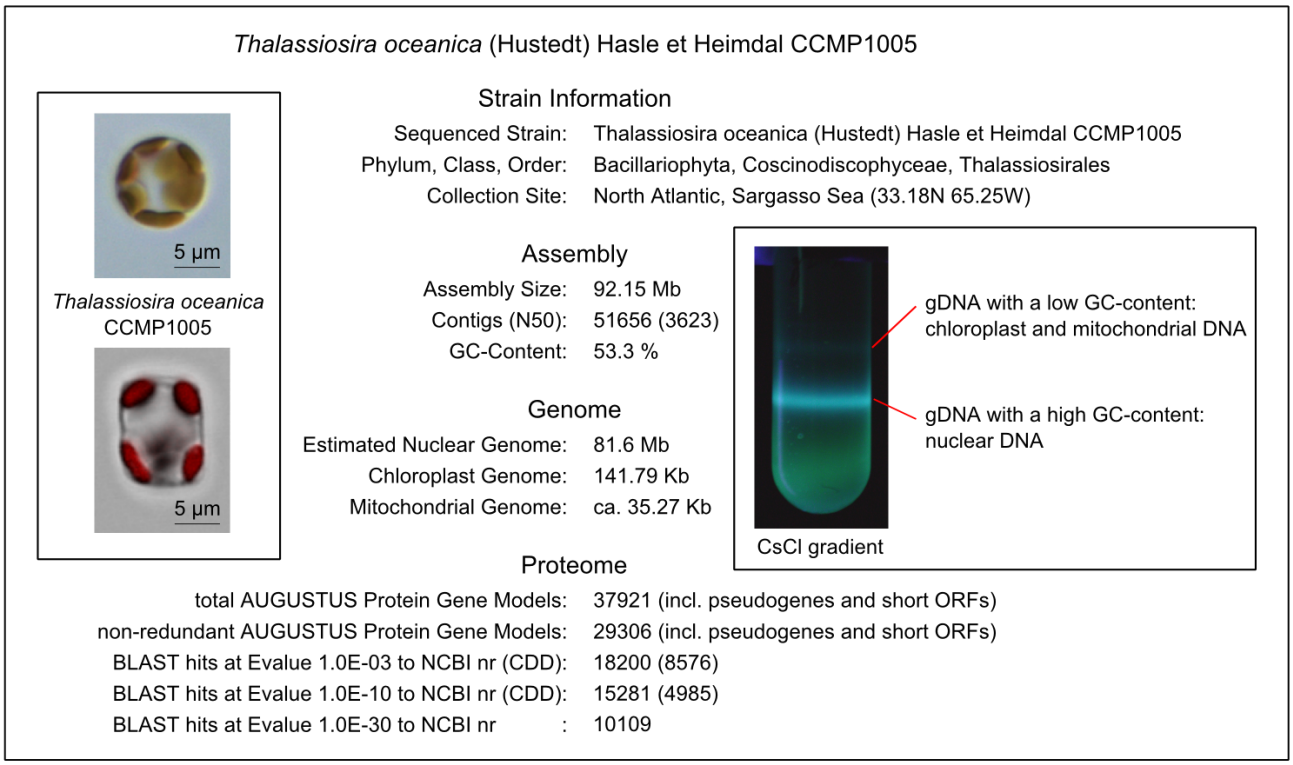


Figure 1 - *T. oceanica* CCMP1005 genome statistics

The sequenced strain *T. oceanica* CCMP1005 belongs to the Centrales group of radially symmetric diatoms and was first isolated from the oligotrophic Sargasso Sea by R. Guillard. With 92.15 Mb our genome assembly is slightly larger than the expected haploid genome size of 81.6 Mb suggesting some redundancy of the current assembly. The genuine AUGUSTUS gene model predictions include a large fraction of pseudogenes and short ORFs that show no homology to any proteins from NCBI nr database at a reasonable Evalue cut-off.

Left inset contains light microscopy images of the sequenced organism in valve view (upper image, chloroplasts brown) and girdle view (lower image, chloroplasts red from overlay of chlorophyll autofluorescence).

Right inset shows the separation of nuclear and organellar DNA in a CsCl density gradient. Stained DNA emits blue fluorescence upon excitation with UV light.

In Table 1 we present an overview of the most abundant Clusters of Orthologous Groups (COG) domains in *T. oceanica*. The abundances of diverse groups of ATPases were overall very similar to those for other diatoms. A group of 19 chitinases is shared between the two centric *Thalassiosira* species.

Table 1 - Most abundant protein domains (based on Clusters of Orthologous Groups COG with an Evalue threshold of 1.0E-10) in the four diatom species *T. oceanica* (To), *T. pseudonana* (Tp), *P. tricornutum* (Pt) and *F. cylindrus* (Fc).

COG	ATPases	To	Tp	Pt	Fc
COG0515	SPS1, Serine/threonine protein kinase	115	132	119	137
COG0464	SpoVK, ATPases of the AAA+ class	90	43	38	44
COG1132	MdlB, ABC-type multidrug transport system, ATPase and permease components	54	44	47	51
COG1222	RPT1, ATP-dependent 26S proteasome regulatory subunit	50	41	37	42
COG0465	HflB, ATP-dependent Zn proteases	49	37	35	39
COG1223	Predicted ATPase (AAA+ superfamily)	44	39	34	41
COG3899	Predicted ATPase	42	48	11	2
COG2274	SunT, ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain	42	52	50	61
COG5265	ATM1, ABC-type transport system involved in Fe-S cluster assembly, permease and ATPase components	40	33	30	33
COG4618	ArpD, ABC-type protease/lipase transport system, ATPase and permease components	39	41	42	43
COG4987	CydC, ABC-type transport system involved in cytochrome bd biosynthesis, fused ATPase and permease components	31	50	46	56
COG4988	CydD, ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components	29	52	49	60
COG0488	Uup, ATPase components of ABC transporters with duplicated ATPase domains	29	50	46	53
COG1131	CcmA, ABC-type multidrug transport system, ATPase component	22	56	52	65
COG	**basic cellular functions**	**To**	**Tp**	**Pt**	**Fc**
COG0513	SrmB, Superfamily II DNA and RNA helicases	46	48	44	54
COG0553	HepA, Superfamily II DNA/RNA helicases, SNF2 family	35	27	24	36
COG5059	KIP1, Kinesin-like protein	24	25	15	14
COG1643	HrpA, HrpA-like helicases	21	14	9	20
COG0443	DnaK, Molecular chaperone	18	14	9	10
COG5021	HUL4, Ubiquitin-protein ligase	15	7	8	8
COG5022	Myosin heavy chain	14	11	9	9
COG	**chitinases**	**To**	**Tp**	**Pt**	**Fc**
COG3325	ChiA, Chitinase	19	19	1	0

The chloroplast genome has been published previously [17]. The mitochondrial genome encodes for 31 protein genes and is represented by two contigs at a total of 35.3 Kb (excl. the characteristic mitochondrial repeats). The current genome assembly, AUGUSTUS protein gene models, ESTs and proteomics peptides as well as updated

versions thereof are publicly accessible in a browsable form at http://bose.geomar.de/cgi-bin/gbrowse/Toceanica/.

With an estimated haploid size of ~80 Mb, the genome of *T. oceanica* is significantly larger than those of *T. pseudonana* (~34 Mb) or *P. tricornutum* (~28 Mb), and rather comparable to that of *Fragilariopsis cylindrus* (~80 Mb) [18]. The genome expansion has occurred by DNA recruitment from both internal and external DNA sources.

Of the 9579 vertically inherited genes (i.e. any genes that have not been acquired by a horizontal transfer event) most of the genes (88%) mapped to *T. pseudonana*, the closest relative species for which a sequenced genome is available (Fig. 2). However, a significant fraction (10%) mapped to *P. tricornutum* instead. This could result from frequent gene loss/replacement events in the genome of *T. pseudonana*, thereby reflecting the overall high capacity for horizontal gene transfer in diatoms [5]. Alternatively, the small genome size of *T. pseudonana* may as well have arisen from reductional trends in this species.

The broad diversity of source taxa for laterally acquired genes suggests that this was the result of frequent events. The taxonomic distribution of the best BLAST hits at a conservative E-value cut-off of 1.0E-30 is presented in Fig. 2. 5% of 10109 AUGUSTUS gene models were assigned to non-stramenopile groups and were likely acquired through lateral gene transfer. This group of genes can be assigned equally to prokaryotic (45%) and eukaryotic (55%) taxa expected to be present in the ecological niche of *T. oceanica*, e.g. the green algal genuses *Micromonas* and *Ostreococcus*. The presented taxonomic assignments for genes of lateral origin are supported through phylogenetic trees, which additionally show a heterogenous mix of taxa in sister lineages, suggesting frequent transfer between major groups (Additional file 2 - "PhyloTrees_for_all_To_LGT_candidates.tre"; Additional file 3 – Fig. S1 "Taxonomy of genes acquired by lateral gene transfer (2)").

Genomic expansion originating from internal DNA sources may happen from genomic duplication events or transposon activity. In *T. oceanica* we observe several paralogous gene pairs that could be the result of either mechanism (Additional file 3 – Fig. S2 "Duplications"). Notably, several iron-regulated genes have either been duplicated (e.g. the *ISIP1* genes *ISIP1A* & *ISIP1B* and the flavodoxin genes *FLDA1* & *FLDA2*) or contain domain duplications (e.g. *CREGx2*) as discussed below.

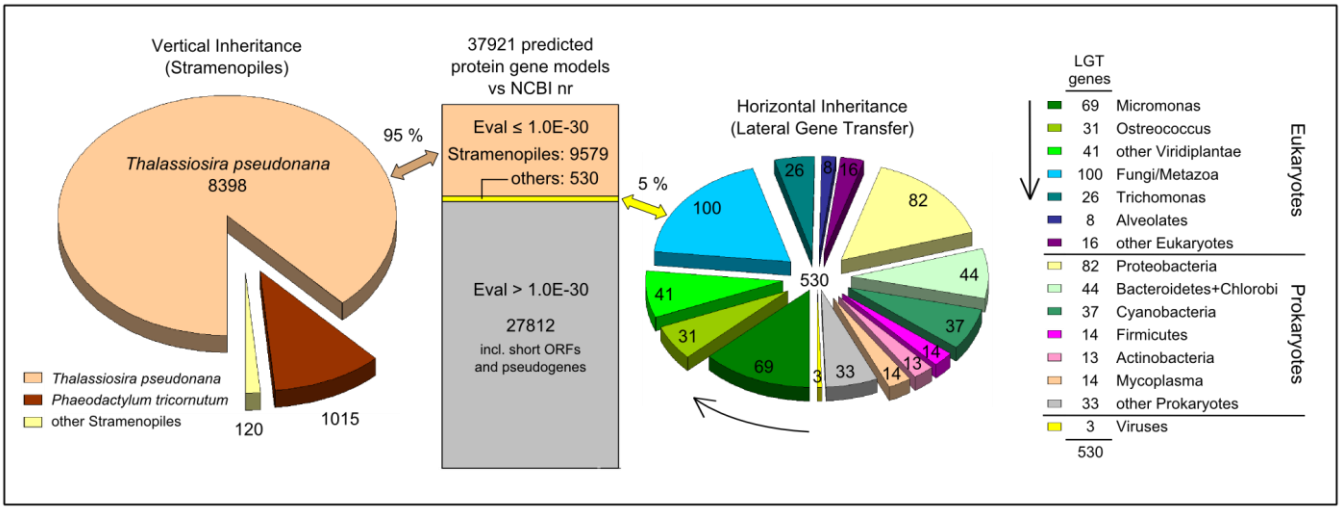


Figure 2 - Vertical vs horizontal inheritance of genes

For evaluation of the extent of laterally acquired genes we focussed on the 10109 AUGUSTUS gene models that have homologs in the NCBI nr protein database at a conservative E-value cut-off of 1.0E-30 (middle bar). A significant fraction of the vertically inherited genes (left) is not shared with the closest relative *T. pseudonana*, but rather with *P. tricornutum*. Genes acquired through mechanisms of lateral gene transfer (right) appear to be derived from diverse prokaryotic and eukaryotic taxons with highest contribution by the green algal genus *Micromonas*.

Physiology of the low-iron response: Fe(-) vs Fe(+)

Variable fluorescence F_v/F_m, an indicator of Fe-limitation in the laboratory [19], was used as a rapid measure of the physiological status in Fe-replete and Fe-limited cultures of *T. oceanica* harvested in late exponential growth phase. The growth rate of iron-limited cells in exponential phase was accordingly much smaller than for iron-replete cells (Table 2), and cellular protein content was 50 % lower.

Table 2 - Physiology of the low-iron response. Cellular dimensions and physiological parameters are compared between nutrient replete Fe(+) and iron-limited Fe(-) cells of exponentially growing *T. oceanica* cultures.

Fe (+)		Fe (–)
0.5 - 0.6	**F_v / F_m**	0.2 - 0.3
0.73 ± 0.01	**Growth Rate μ [d^{-1}]**	0.28 ± 0.02
4	**Chloroplasts / Cell**	2
409 ± 48	**Chlorophyll a / Cell [fg]**	58 ± 27
122 ± 3	**Cell Surface Area [μm^2]**	140 ± 5
100 ± 4	**Cell Volume [μm^3]**	80 ± 5
15.3 ± 0.9	**Single CP Surface Area [μm^2]**	12.0 ± 0.8
4.7 ± 0.3	**Single CP Volume [μm^3]**	3.5 ± 0.2

The cell volume of iron-limited *T. oceanica* was smaller than that of iron-replete cells, whereas the cells had a larger surface area at low-iron due to a smaller diameter (4.7±0.1 vs 5.9±0.1 μm) and a larger height (7.0±0.4 vs 5.5±0.2 μm). This imposed an elongated phenotype to the cells, but at the same time increased the surface/volume ratio by 43 % (1.75 vs 1.22) and favoured cellular nutrient (i.e. iron) supply [20]. The intracellular space of iron-limited cells exhibited increased vesiculation (Fig. 3).

Under low-iron conditions *T. oceanica* cells show a severe decrease in chlorophyll content (Table 2). This chlorosis response of iron-limited *T. oceanica* is further accompanied by a decrease in cellular chloroplast volume and in total cellular chloroplast surface area (Fig. 3). Iron-limited cells have reduced the number of chloroplasts to 2 instead of 4 in the iron-replete counterpart, and these are also smaller in size. Total chloroplast dimensions for individual cells were distributed over a range spanning the 2-fold increase in volume and surface, thereby reflecting chloroplast duplication during cellular growth.

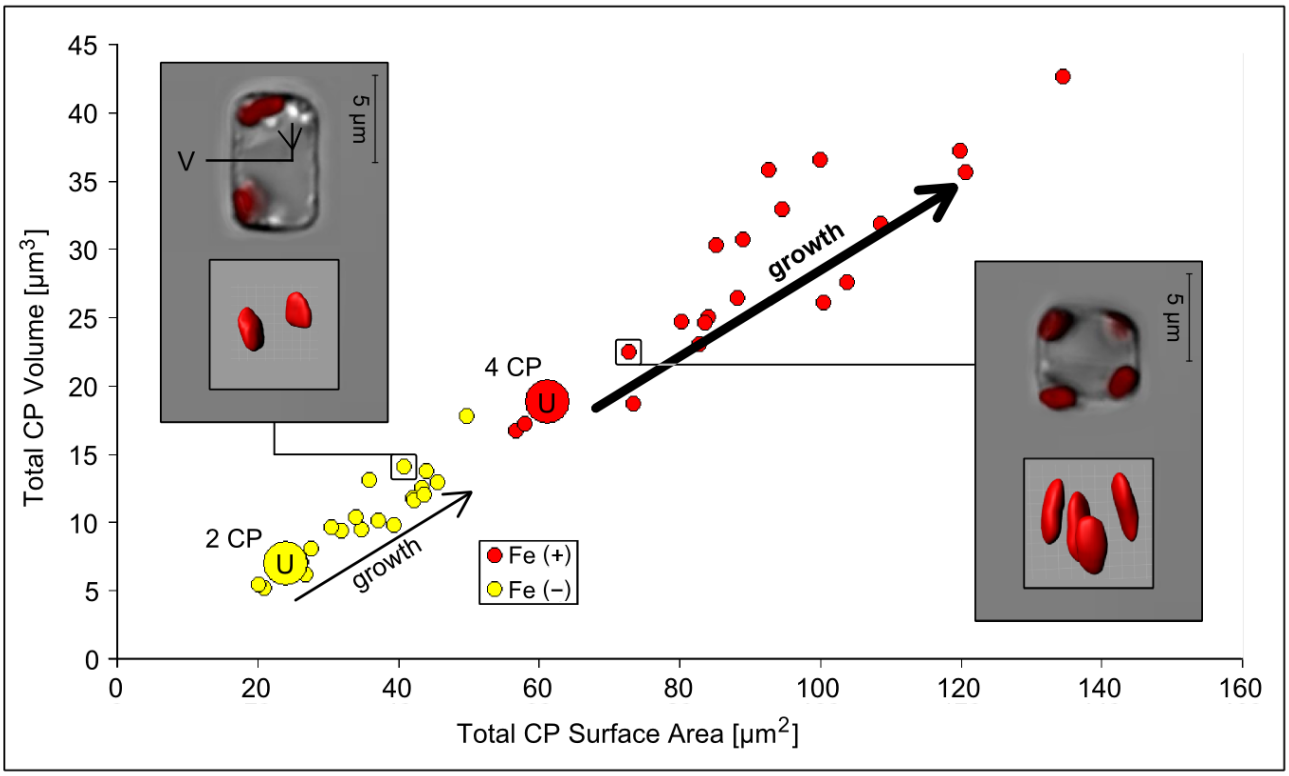


Figure 3 - Reduction of the chloroplast system

The approximate dimensions of the photosynthetic machinery were assessed using confocal laser scanning microscopy and subsequent 3D reconstruction of the chlorophyll autofluorescence signal. A plot of total cellular chloroplast volume vs total cellular chloroplast surface area shows a reduction of the chloroplast system in iron-limited *T. oceanica* cells. Iron-limited cells have a reduced number of 2 chloroplasts instead of 4. Total chloroplast dimensions for individual cells (small circles) are distributed over a range spanning the 2-fold increase in volume and surface that accompanies chloroplast duplication during cellular growth.

Inserts show an overlay transmission and chlorophyll autofluorescence image (top) and the respective 3D chloroplast reconstruction (bottom). The left insert illustrates an iron-limited cell close to dividing with two nearly duplicated chloroplasts. Note the characteristic increase in vesiculation of the cellular interior at low-iron. An iron-replete cell at the beginning of its cell cycle (shortly after division) contains 4 chloroplasts (right insert).

CP chloroplasts; U cell at the beginning of its cell cycle ("unit cell"); V vesicles

---

Transcriptomics

For an in-depth analysis of the *T. oceanica* low-iron response, we focused on approx. 300 genes that were identified from a log-likelihood ratio test statistic [21] as significantly differentially regulated and that could be assigned a specific function (Fig. 4A). Some additional genes for paralogous proteins were added. These were

selected on the basis of their involvement in substitution between related proteins under iron-limited and replete conditions, or as members of a protein family exhibiting a differential response to low-iron. In such cases, the response of a specific gene is better understood in the context of its respective group or family. The complement of organellar genes (encoded by the chloroplast and mitochondrial genome) was added as representative for the two well-defined and important pathways of photosynthetic and respiratory electron transport, or as proxy for organellar activity respectively. All sequences of the selected proteins are provided in Additional file 4 – Supplemental Sequences S1 "protein_user_models", the corresponding annotation is provided in Additional file 5 - Table S2 (Transcriptome_Manual_Annotation) GB.xls.

To determine the major metabolic differences found in iron-limited compared to iron-replete growth conditions, all annotated (see Additional files section) gene products together with their respective expression data were mapped on a cellular scheme. The major cellular trends that could be deduced are summarized in Fig. 4C. Identifier and detailed information on the discussed proteins (Additional file 6 – Supplemental Sequences S2 "selection") are given in Additional file 7 – Table S3 "Low-Iron_Responsive_Genes". In the following proteins are referred to as exemplary (HSF1, p271) with HSF1 reflecting the gene name (or shortcut) and p271 being the identifier of its respective manually improved protein model (Additional file 4 – Supplemental Sequences S1 "protein_user_models").

Under stress conditions, maintaining cellular integrity is crucial to survival. During iron limitation, the electron flow through the impaired photosynthetic machinery leads to enhanced production of reactive oxygen species (ROS) that damage biomolecules located near the thylakoid membranes [22]. The need for protein repair and refolding induces an "oxidative stress response" that is presumably coordinated by up-regulated heat shock factors (HSF1, p271; HSF2, p256). While all other chloroplast-encoded transcripts were down-regulated in the course of the general chlorosis response, the chloroplast chaperones dnaK and clpC were up-regulated. Additionally, an LHCSR ortholog (LI818, p170), belonging to the FCP (fucoxanthin-chlorophyll a/c-binding protein) family of light-harvesting proteins and implicated in efficient non-photochemical quenching (NPQ) [23, 24], showed an increased transcript level.

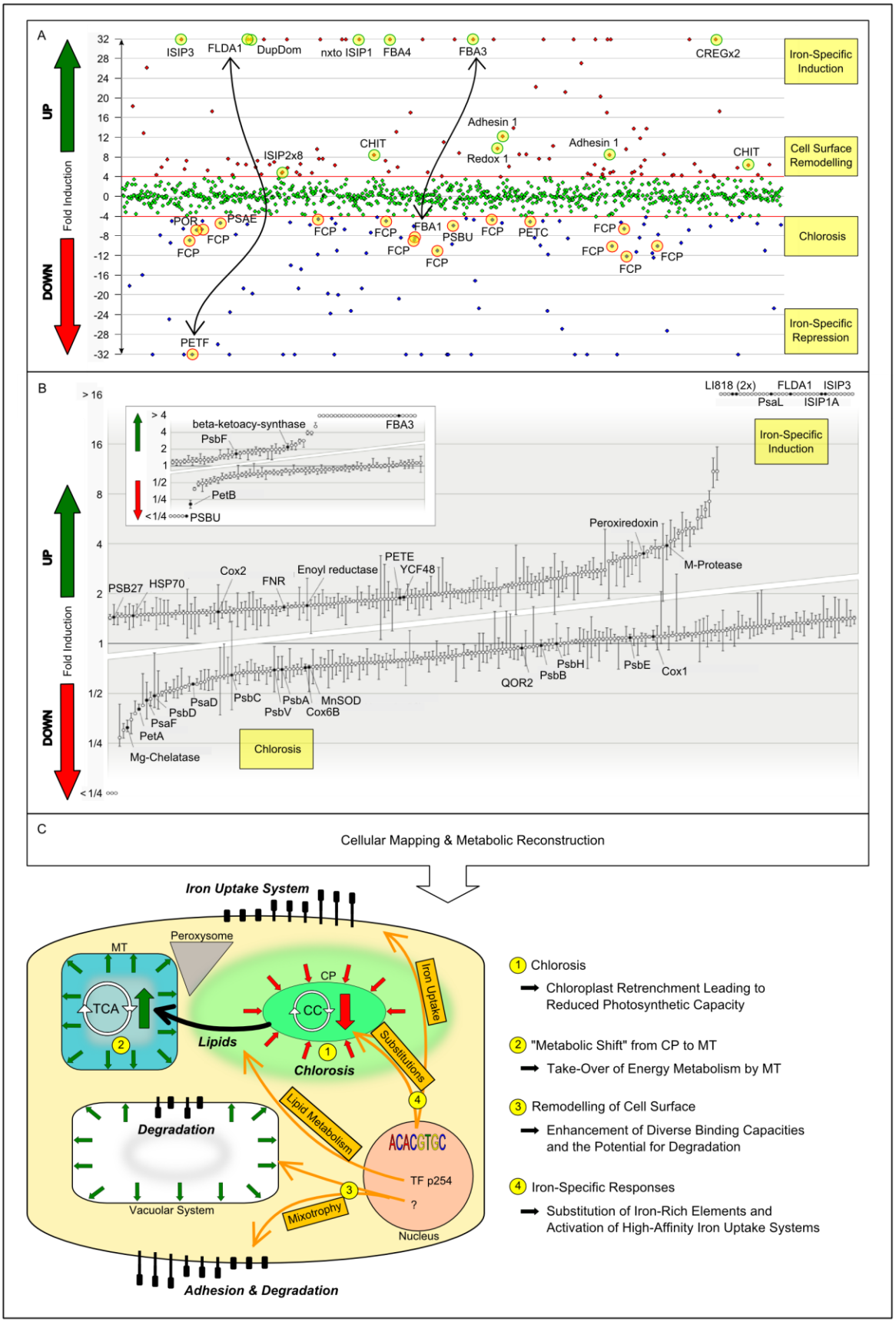


Figure 4 - Basic cellular changes at low-iron

Figure 4 *(continued)*

Differential gene expression of exponentially growing iron-limited versus iron-replete *T. oceanica* cells was assessed from global transcriptomics and proteomics approaches. (A) Transcriptomics data was screened with T-ACE, a transcriptome database browser that plots the assembled transcript fragments according to their differential regulation as inferred from differential read contribution of Fe(-) and Fe(+) libraries to each transcript contig. (B) For the proteomics data the differential regulation of each gene product is represented by the median of all PBC (peptide/SDS-PAGE band/charge) ratios assigned to it, with error bars constructed from the first and third quartiles. The main plot shows proteins with at least two PBC values, inset contains proteins with a single PBC value. (C) Only a subset of low-iron responsive genes could be assigned a robust annotation and were suitable for mapping to a cellular scheme. Accordingly, the cellular response of *T. oceanica* to low-iron was inferred from the mapping of a representative selection of genes (see text) and their respective differential regulation on transcript and protein level.

The most pronounced elements of the complex response are chloroplast retrenchment (chlorosis) and the consequential take-over of the energy metabolism by the mitochondrial system (metabolic shift). Diverse surface-related binding capacities and the potential for degrading organic matter are enhanced suggesting a putative mixotrophic response (mixotrophy). The strongest transcriptional response is seen from genes involved in iron-uptake or compensational substitutions (4). This iron-specific part of the cellular response may be mediated by a conserved promoter motif identified in this work.

The development of a chlorotic phenotype and the corresponding retrenchment of the chloroplast system is the most pronounced cellular response to low iron. Accordingly we find substantial changes in organellar transcript levels which suggests that major functions related to the cellular energy metabolism are adopted by the mitochondrial system instead ("metabolic shift"). Chloroplast transcript levels decreased (2026 [Fe(-)] vs 14931 [Fe(+)] total chloroplast reads), while mitochondrial transcripts showed a 2-fold increase (31261 [Fe(-)] vs 18136 [Fe(+)] total mitochondrial reads). Much of this effect can be attributed to the organellar rRNA operons, whose transcription is indicative of organellar translational activity (Additional file 3 – Fig. S3 "Metabolic Shift"). In parallel, diverse nuclear-encoded, but chloroplast-targeted gene products were down-regulated. These included genes coding for enzymes involved in chlorophyll biosynthesis and the Calvin cycle, as well as components of the light reaction such as photosystem (PS) subunits and several FCPs. Conversely, components of the mitochondrial respiratory chain, like cytochrome c oxidase, cytochrome b and several subunits of the NADH dehydrogenase were up-regulated. This was also seen for a

mitochondrial ATP/ADP-translocase (p242) involved in the transport of energy equivalents.

Cellular retrenchment, i.e. the reduction of cellular biomass and activity, and decreased growth rates are general responses of nutrient-limited cells [13]. While chloroplast reduction was readily observable in iron limitation due to the visual predominance of these organelles in the cells, we also saw indications of a general cellular retrenchment in the transcriptional response. The expression level of the 18S rRNA gene (represented by 1154 [Fe(-)] vs 2691 [Fe(+)] reads) suggests a lower translation rate under iron limitation. Though such inferences must be taken with care, this would be in agreement with the decreased growth rate and lower biomass, as cellular rRNA correlates with cellular biomass. The strong up-regulation of mitochondrial isocitrate lyase (ICL, p419) and glutamine synthetase (GS, p302) suggests biomass recycling strategies to avoid losing fixed carbon and nitrogen during the metabolite conversions associated with enhanced respiration. The isocitrate lyase bridges the two decarboxylation steps of the mitochondrial citric acid cycle TCA (carried out by isocitrate dehydrogenase and α-ketoglutarate dehydrogenase), thereby preserving carbon as glyoxylate. The glutamine synthetase reincorporates free ammonium, preserving nitrogen as glutamine. Under low-iron conditions, utilization of ammonium is energetically advantageous due to the high iron-requirements of the nitrate assimilation pathway [25]. The concerted action of cellular retrenchment and biomass recycling allows for prolonged growth despite reduced carbon assimilation, thereby increasing the probability of cell survival.

Diverse genes, whose products are targeted to the secretory pathway, are up-regulated under iron limitation, suggesting an extensive cell-surface remodelling as also observed for iron-limited *P. tricornutum* [19]. Many of these genes are assigned adhesive or degradative functions. An enhanced capacity for adhesion favours recruitment of organic matter to the cell. As organic matter can be a rich and complex source for various nutrients including iron, its recruitment to the cellular surface represents a required first step in iron uptake. Besides providing a source of iron, the bound organic matter could also serve as a source for other nutrients like N or P in the context of a facultative mixotrophy. Example genes assignable to such a hypothetical scenario and highly responsive to low-iron are given in Fig. 5 with *Adhesin 1* (p329), *CB* (Carbohydrate-binding 1, p230), *CHIT* (chitinase, p88), *M-Phosphoesterase* (p323), *M-Protease* (p279), *Redox 1* (p232). However, under the photoautotrophic experimental conditions, the cultures lacked any external organic carbon source except the essential vitamins.

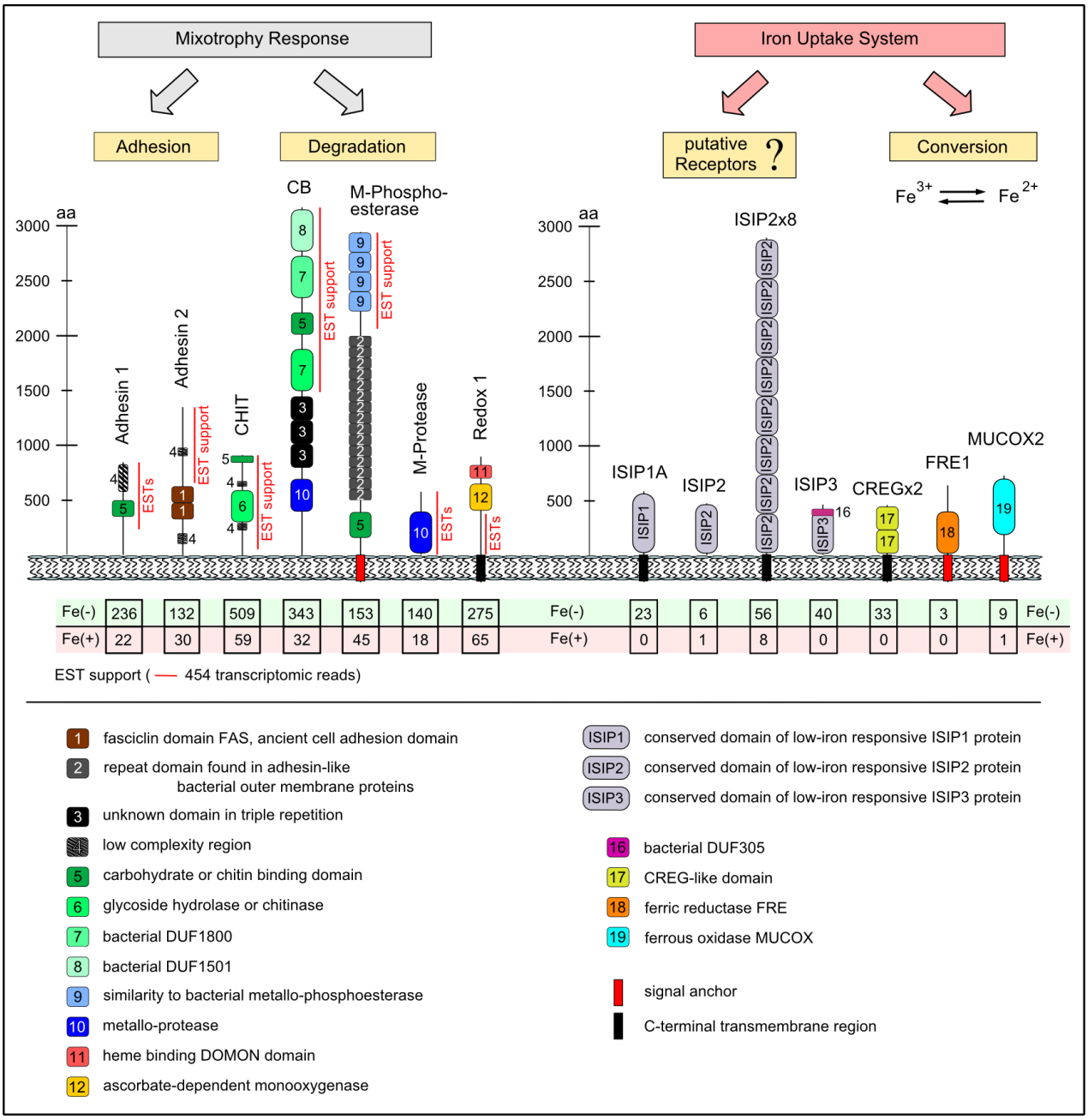


Figure 5 - Hypothetical categorization of low-iron inducible cell surface proteins

In low-iron conditions we find an up-regulation of diverse genes, whose products are targeted to the secretory pathway, suggesting an extensive cell-surface remodelling. Many of these are predicted to be involved in adhesion or degradation processes and might contribute to enhancing the overall cellular capacity to bind and process external organic matter. We provide a hypothetical categorization for highly responsive genes that can be assigned to this function. **[continued on next page]**

Figure 5 *(continued)*

While some of the gene products can be placed in the context of iron-uptake (right), others are less well defined, but contain a variety of conserved domains involved in adhesion or degradation of organic matter (left). Especially for larger genes EST support is patchy suggesting possible inaccuracies in AUGUSTUS gene modelling. Differential read contribution from the Fe(-) and Fe(+) libraries to each transcript contig (ESTs) is taken as a measure for the differential transcription of the respective gene.

A straight-forward strategy to survive at low-iron is to lower the cellular iron requirements by replacing components that are rich in iron with iron-free substitutes that are functionally equivalent, like the substitution of the chloroplast electron carrier ferredoxin with flavodoxin [10]. The genome of *T. oceanica* encodes for two cytochrome c_6 genes and one plastocyanin gene. While the cytochrome c_6 genes *CYTC6A* and *CYTC6B* are found to be weakly expressed, the plastocyanin gene *PETE* shows high expression under high-iron with a characteristic decrease in low-iron conditions as seen from many constitutively expressed chloroplast genes in the course of the chlorosis response. This suggests a constitutive use of plastocyanin (PETE, p175) instead of cytochrome c_6 for photosynthetic electron transport and is consistent with prior findings [11, 12]. A constitutive expression of plastocyanin could certainly be regarded as a specific adaptation to low-iron regimes, although the retention of the cytochrome c_6 genes suggests that these may play a role under specific environmental conditions. Fructose-bisphosphate aldolase (FBA) genes are redundant in some diatoms and have recently been described in more detail for *P. tricornutum* [26]. The *T. oceanica* genome, too, was found to encode for several FBA enzymes, with the cytosol, the chloroplast stroma and the chloroplast pyrenoid harbouring two FBA each (CP pyrenoid: FBA1, p380, and FBA3, p153; CP stroma: FBA2, p381, and FBA5, AUG_g19407; cytosolic: FBA6, AUG_g24977, and FBA4, p154). As is the case in *Phaeodactylum*, one of the *T. oceanica* FBAs from each compartment (FBA1, FBA2, FBA6) appears to act through metal catalysis (class II) while the second (FBA3, FBA5, FBA4) is predicted to use a Schiff-base catalysis (class I) instead. While the metal cofactor of different class II FBAs was found to be Mn^{2+} [27], Zn^{2+} [28] or Cd^{2+} [29] in *E. coli*, the orthologous FBAs of *T. oceanica* apparently are differentially regulated through the availability of iron, suggesting the involved metal in these enzymes might be Fe^{2+}, and implying a pairwise substitution by class I enzymes.

An essential part of iron-uptake systems are ferric reductases (FRE) and ferrous oxidases (MUCOX) that act on the interconversion of the two ionic species Fe^{3+} and Fe^{2+}. In the iron-limited transcriptome we

find an up-regulated putative ferric reductase (FRE1, p157) and an up-regulated multicopper oxidase (MUCOX2, p67) that shows characteristics of a ferrous oxidase. Their differential regulation with respect to iron availability makes them candidates for iron-specific reductase and oxidase involved in iron-uptake (Fig. 5). Iron uptake requires an initial binding of iron and/or iron-complexes. The involved receptors are presently unknown, though a number of genes, exclusively expressed under iron limitation, are targeted to the cell-surface, making them candidates for iron-binding receptors. The low-iron responsive gene *ISIP1* ("Iron-Starvation Induced Protein") was first identified in *P. tricornutum*, but has conserved orthologs in *T. oceanica* (*ISIP1A*, p159) and *F. cylindrus*. We provide further evidence for a role of the ISIP1 protein as putative receptor below. Additional members in this group are ISIP2 (p160) and ISIP3 (p161), both represented by orthologs in *P. tricornutum* as well. Further, we list some proteins that contain duplicated domains known from *P. tricornutum* low-iron responsive genes, like an 8-fold duplicated ISIP2-like subdomain (ISIP2x8, p84) or a duplicated CREG-like domain (CREGx2, p90). Duplication of iron-binding domains would directly enhance the capacity for iron binding and enable increased uptake kinetics [25].

Non-ribosomal peptide synthases (NRPS) [30] are responsible for the production of peptide antibiotics or – in some cases – siderophores that are capable of binding iron [31]. In addition to a conserved fungal NRPS (NRPS1, p174) with orthologs in *T. pseudonana* and *P. tricornutum* we find a putatively cytosolic NRPS of bacterial origin (NRPS2, p173) up-regulated at low-iron. Co-regulated with this bacterial NRPS is a multidrug resistance-associated protein (MRP, p57) that might be involved in the export of the respective peptide products. The up-regulation of NRPS likely indicates a defense mechanism in response to enhanced competition (either for iron or – under the premise of facultative mixotrophy – for organic matter).

We observe the induction of a reverse transcriptase (RT, p222) and a CRE-like recombinase (CRE, p321), potentially indicating an activation of mobile elements under iron limitation. These enzymes might also be involved in gene and/or domain duplication events through reverse transcription and genomic integration of cellular mRNA copies. Thereby, this molecular system may provide a link between environmental stresses and the structural dynamics of the diatom genome.

Proteomics

The transcriptomic data of *T. oceanica* unveils extensive changes in cellular transcript levels in response to iron limitation. Although informative, transcript abundance do not necessarily reflect cellular

protein levels [32]. We therefore supplemented the transcriptomic data with proteomic data to determine the protein complement in action under the defined iron-replete and iron-limited growth conditions. Fig. 4B illustrates the dynamic range of differential abundances for all proteins detected by LC-MS/MS relative to equal amounts of total cellular protein for both conditions. The induction of flavodoxin is a hallmark of iron-deficiency responses in many diatoms and cyanobacteria (see above). In accordance with the transcriptome response, flavodoxin as well as ISIPs or class I FBAs could only be found under iron limitation. The extent of correlation between proteomics and transcriptomics data was assessed through plotting the relative abundance data from peptides (proteomics P) against the relative abundance data from their corresponding transcripts (transcriptomics T) (Additional file 3 – Fig. S4 "Correlation Plot P vs T"). A stretched cluster along the y-axis indicates a high dynamic range of the transcriptomics data, while the proteomics data is more uniform for this group.

Both transcriptomics and proteomics data are biased towards highly abundant transcripts/proteins. Especially the proteomics data despite its relatively high number of signals could resolve only a subset of the protein complement. Accordingly, we interpret the complement of differentially regulated genes and proteins recovered from both approaches as complementary in the information that they provide, and we do not expect them to show a complete overlap. However, the overlap in the response for the specifically induced proteins ISIP1 and class I FBAs shows that the data from both approaches are in general in good agreement with each other.

In the proteomics data it is of specific interest to have a closer look at proteins of the photosynthetic machinery. An appropriate internal reference for the regulation of chloroplast proteins is given with the chloroplast ribosomal proteins that indicate a down-regulation of the ribosomes at a ratio of 0.8 relative to the iron-replete proteome. Protein subunits of photosystem I (PS I) were reduced about two-fold under low iron (0.45), except PsaL, which was only found under iron limitation. In cyanobacteria, PsaL, generally important for trimer formation, facilitates the formation of IsiA rings around PS I monomers under iron-deprivation [33]. We speculate that PsaL might be involved in the organization of PS I light-harvesting structures specifically formed under low-iron conditions and/or oligomerization of PS I in iron-limited *T. oceanica*. Subunits of the iron-containing cytochrome b_6/f (cyt b_6/f) complex, were down-regulated, with ratios of 0.2 and 0.32. In contrast, photosystem II (PS II) subunits PsbB, PsbC, PsbE, PsbH and PsbV remained almost constant with ratios at about 1.1. While the PS II core complex seems to be retained to some extent, the

labile D1 protein is down-regulated at 0.7 thereby probably reflecting a proportional decrease in functional PSII. The differential regulation of the two photosystems (0.45 for PS I vs 0.7 for PS II D1 protein) supports an adaptive significance for the remodelling of the photosynthetic architecture under iron limitation, conversely to earlier findings [12].

While PS I and cyt b_6/f complexes were down-regulated two- to threefold, it was still possible to detect the iron-rich mitochondrial complexes under iron limitation. Relative protein quantitation was possible for subunits of complex III, complex IV and the ATPase with low-to-high-iron ratios ranging from 0.95 for QOR2 to 1.7 for the beta subunit of the mitochondrial cytochrome c oxidase (Fig. 4B). This is in agreement with the transcriptomic data and supports the idea that mitochondrial electron transfer protein complexes are preserved under iron limitation relative to photosynthetic electron transfer protein complexes.

While the magnesium chelatase, involved in chlorophyll synthesis, is down-regulated at 0.35, the numerous FCP light-harvesting proteins showed very diverse responses to iron limitation (Additional file 3 - Fig. S5 "FCP proteomics"). Some FCPs showed down-regulation under low-iron whereas others were up-regulated. In particular LHCSR-like FCPs, involved in photoprotection, were highly abundant under iron limitation corroborating the transcriptome analysis. Notably, the xanthophyll cycle enzyme violaxanthin de-epoxidase showed significant up-regulation at 3.1 suggesting a possible linkage to the group of FCPs, which accumulate under iron limitation.

Comparative genomics reveals extensive genomic plasticity in T. oceanica

We used the genome information of *T. oceanica*, *T. pseudonana*, *P. tricornutum* and *F. cylindrus* to investigate central issues of the diatom low-iron response in a comparative genomics approach:

Taxonomic distribution of iron-regulated genes

We screened the four diatom genomes known to date (*T. oceanica*, *T. pseudonana*, *P. tricornutum* and *F. cylindrus*) for the highly conserved iron-regulated genes *ISIP1*, *ISIP3, PETF, FLDA, CYTC6, PETE* and *FBAs* of class II and class I (Table 3). Phylogenetic trees for the important groups of flavodoxin and FBA proteins are provided in Additional file 3 - Fig. S6 and S7.

The short flavodoxin isoform, plastocyanin and the class I FBAs or known or assumed to replace iron-containing counterparts under low-iron conditions. The two oceanic diatoms *T. oceanica* and *F. cylindrus* that have some of the highest tolerance to low-iron both contain all 5 of the respective genes while *P. tricornutum* lacks 2 of these. The

typical coastal species *T. pseudonana* lacks all except the gene for the cytosolic class I FBA, while at the same time having the highest requirement for iron in the group of diatoms for which genome information is currently available. Further we find multiple copies of the ISIP1 gene in *T. oceanica* and *F. cylindrus*, while these gene is absent in *T. pseudonana*. The presence or copy number of these genes in the tested diatom genomes suggests an adaptive significance with respect to the low-iron condition found in oceanic waters.

Table 3 - Presence and copy number of iron-regulated genes in the four diatom species *T. oceanica* (To), *T. pseudonana* (Tp), *P. tricornutum* (Pt) and *F. cylindrus* (Fc). Listed are also the respective counterparts whose products are subject of substitution under iron-limited conditions. The conserved paralogous genes of FLDA [34] and CYTC6 are predicted to contain a signal peptide and are assumed to act in a different functional context. SP secretory pathway; ER endoplasmic reticulum; CP chloroplast; protein identifier are provided in Additional file 8 (Table S4 "Iron-Regulated Genes").

gene	product	destination	mutual substitution at low-iron	To	Tp	Pt	Fc
PETF	ferredoxin	chloroplast	ferredoxin -> flavodoxin (short)	1	1	1	1
FLDA(s)	flavodoxin (short)	chloroplast	ferredoxin -> flavodoxin (short)	2	0	1	1
FLDA(l)	flavodoxin (long)	SP (ER ?)	none (distinct functional context)	1	1	1	1
CYTC6 (type A)	cytochrome c6	chloroplast	cytochrome c6 (type A) -> plastocyanin	2	1	1	1
CYTC6 (type B)	cytochrome c (?)	SP (ER ?)	none (distinct functional context)	1	1	1	0
PETE/PCY	plastocyanin	chloroplast	cytochrome c6 (type A) -> plastocyanin	1	0	0	1
class II FBA (type A)	class II fructose-bisphosphate aldolase	CP pyrenoid (Pt FBAC1)	class II FBA (type A) -> class I FBA (type A)	1	1	1	1
class II FBA (type B)	class II fructose-bisphosphate aldolase	CP stroma (Pt FBAC2)	class II FBA (type B) -> class I FBA (type B)	1	1	1	1
class II FBA (type C)	class II fructose-bisphosphate aldolase	cytosolic (Pt FBA3)	class II FBA (type C) -> class I FBA (type C)	1	1	1	1
class I FBA (type A)	class I fructose-bisphosphate aldolase	CP pyrenoid (Pt FBAC5)	class II FBA (type A) -> class I FBA (type A)	1	0	1	1
class I FBA (type B)	class I fructose-bisphosphate aldolase	CP stroma	class II FBA (type B) -> class I FBA (type B)	1	0	0	1
class I FBA (type C)	class I fructose-bisphosphate aldolase	cytosolic (Pt FBA4)	class II FBA (type C) -> class I FBA (type C)	1	1	1	1
gene	**product**	**destination**	**putative role in iron-uptake**	**To**	**Tp**	**Pt**	**Fc**
ISIP1	iron starvation induced protein 1	cell surface	receptor (?)	2	0	1	3
ISIP3	iron starvation induced protein 3	cell surface	co-receptor (?)	1	1	1	2

Domain duplications of iron-regulated cell-surface proteins

While differentially regulated genes for cell-surface proteins, identified from the low-iron response of *P. tricornutum* [19], like *ISIP1*, *ISIP2*, *FLDA* or *CREG*, represent single-copy genes encoding for well-defined single-domain proteins, the situation in *T. oceanica* is different (Additional file 3 – Fig. S2 "Duplications"). Here, we find additional paralogous versions of several iron-regulated genes (*ISIP1*, *FLDA*), as well as diverse examples of domain duplications (*CREGx2*, *ISIP2x8*). In case of iron-binding proteins the duplication of domains might provide benefits under iron limitation through a higher density of exposed domains, thereby increasing the affinity for iron at the cell-surface [25].

With respect to the selective pressure encountered in the low-iron open ocean the duplication of complete genes allows one of the two copies to vary, improve and optimize its iron-binding themes/motifs. This may potentially result in more efficient iron-uptake. RT-qPCR allowed us to distinguish iron-regulated genes from their closely related paralogs (Additional file 3 - Fig. S8 "qPCR $\Delta\Delta C_T$").

Iron uptake and the cell-surface protein ISIP1

Conservation between the predicted protein orthologs of ISIP1 in *T. oceanica*, *P. tricornutum* and *F. cylindrus* was high, and the orthologs exhibited an identical prediction of secondary structure (Fig. 6). We found an N-terminal signal peptide targeting the protein to the secretory pathway, while a C-terminal transmembrane domain anchors the protein to a membrane. The major part of the protein is represented by a domain rich in β-strands that likely folds into a β-propeller-like structure. A clue to the structure and function of ISIP1 could be the low-density lipoprotein receptor LDLR, an important cell-surface receptor in humans [35]. Although its extracellular domains differ from the single β-propeller domain of ISIP1, the remainder of the protein is strikingly similar with regard to amino acid composition and secondary structure prediction. Hence, we may transfer the respective LDLR annotation to the ISIP1 protein model.

Accordingly, the ISIP1 protein would represent a cell-surface receptor, anchored to the plasma membrane by a C-terminal transmembrane helix. A small C-terminal tail without well-defined secondary structure contains a conserved endocytosis motif responsible for endocytotic cycling. An α-helical region N-terminal from the transmembrane helix is predicted to be O-glycosylated and would thereby serve to expose the large β-propeller as putative receptor domain to the extracellular space.

An alignment of the ISIP1 proteins from *T. oceanica*, *P. tricornutum* and *F. cylindrus* illustrates that the extracellular β-propeller domain

contains a cysteine-rich center (Fig. 6), whose pattern is reminiscent of cysteines found in Fe-S cluster proteins and might be involved in complexing Fe. The cysteine-rich center is not found in the orthologous p130B of *D. salina* that is thought to have undergone an evolutionary change in function and interacts with transferrin-like proteins [36].

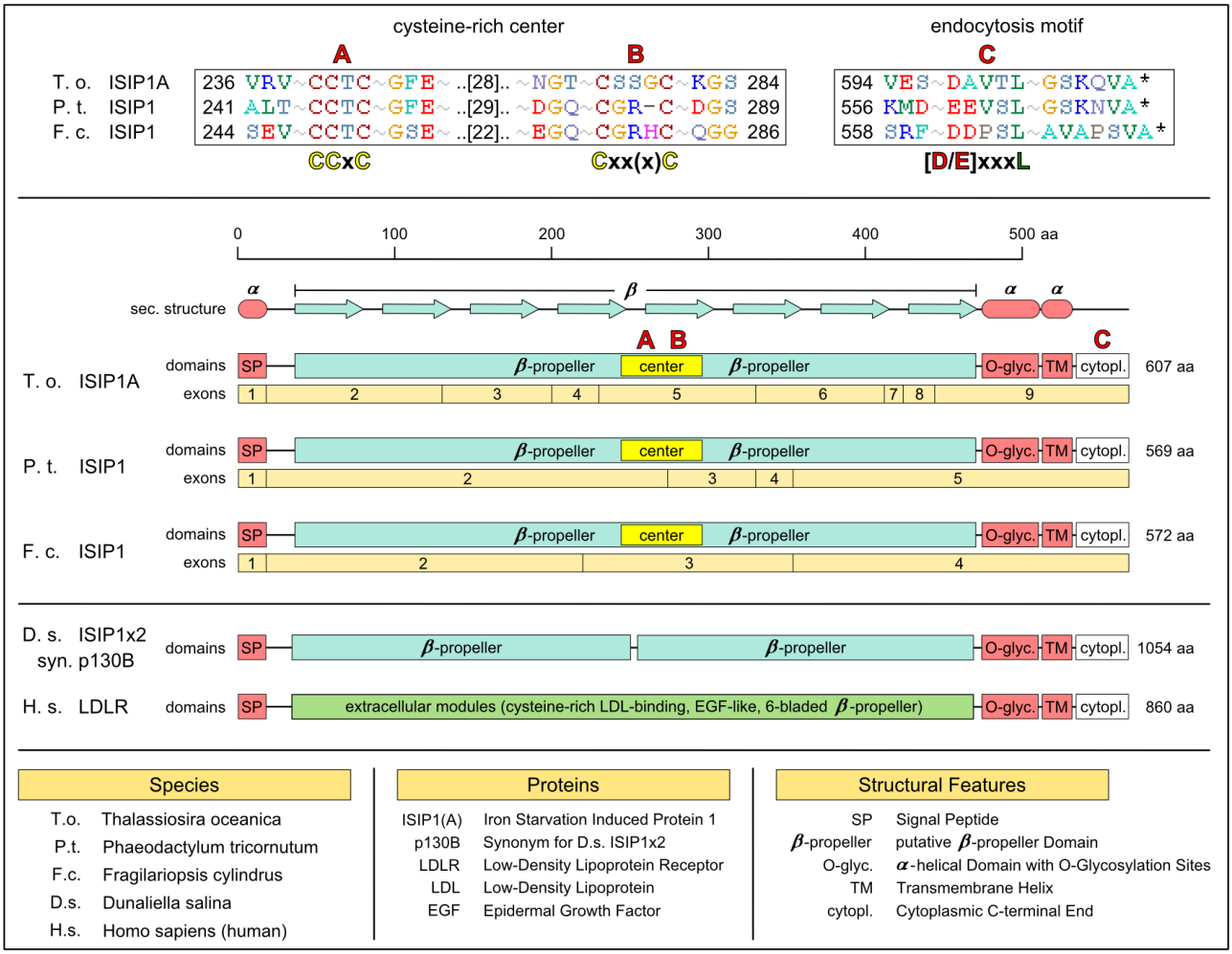


Figure 6 - The low-iron inducible receptor ISIP1

ISIP1 protein models and secondary structure from *T. oceanica*, *P. tricornutum* and *F. cylindrus* are compared.

Conservation between the protein orthologs is high with an identical prediction of secondary structure (center). We find an N-terminal signal peptide targeting the protein to the secretory pathway, while a C-terminal transmembrane domain anchors the protein to a membrane. The major part of the protein is represented by a domain rich in β-strands that likely folds into a β-propeller-like structure. **[continued on next page]**

Figure 6 *(continued)*

While in *D. salina* p130B (bottom) this β-propeller domain is duplicated and only distantly related to the respective diatom domains, the remainder of the protein shows a clear homology to the group of diatom ISIP1 proteins. A clue to the structure and function of ISIP1 could be the human low-density lipoprotein receptor LDLR due to its detailed characterization as a human cell-surface receptor: While its extracellular domains are very different from the single β-propeller domain of ISIP1, the remainder of the protein is again strikingly similar, what allows us to transfer the respective annotation from LDLR to the ISIP1 protein model. Accordingly, the ISIP1 protein would represent a cell-surface receptor that is anchored to the plasmamembrane by a C-terminal transmembrane helix. A small C-terminal tail without well-defined secondary structure contains a conserved endocytosis motif C (top, right) responsible for endocytotic cycling of ISIP1. An α-helical region N-terminal from the transmembrane helix is predicted to be O-glycosylated and thereby would serve to expose the large β-propeller as putative receptor domain to the extracellular space.

A sequence alignment of the ISIP1 proteins from *T. oceanica*, *P. tricornutum* and *F. cylindrus* illustrates that the extracellular β-propeller domain contains a cysteine-rich center, A and B (top, left). The pattern of cystein residues is reminiscent of patterns found in Fe-S cluster proteins and might also be involved in complexing iron.

A conserved promoter motif associated with diverse iron-regulated genes

At the core of an organism's low-iron response are transcription factors (repressors or enhancers) and their respective binding sites (specific promoter motifs) that mediate the cellular response at the gene expression level. From pairwise promoter comparisons between the exclusively iron-regulated *ISIP1*, *FLDA* and *FBA3* genes of *T. oceanica*, *P. tricornutum* and *F. cylindrus* using dotlet [37] and MEME [38] we identified a conserved palindromic motif "ACACGTGC" located around pos. -200 from the translation start. Upon genome-wide screening, a total of 45 gene models contained the complete motif (perfect match) at a position of 150-250 bases before translation start. Functional assignments for genes with positive matches were rarely possible (mostly hypothetical genes of unknown function and without significant regulation). However, the accumulation of low-iron responsive genes (ISIP1 and 3 FBAs) in this group is remarkable. In Figure S8 (Additional file 3 – Fig. S9 "Promoter Motif") we present only those genes whose orthologs in other species carry the motif in their promoters, too.

The complexity of the identified motif ($A_2T_1C_3G_2$) is high; its palindromic structure suggests binding of a homo- or heterodimeric protein factor. The remarkable conservation of this motif and its

position (-200) relative to the translation start across three diatom species reinforces the suggestion that this motif plays a prominent role in iron-dependent gene regulation.

Discussion

The cellular response to iron limitation

As the most prominent part of the complex low-iron response of *T. oceanica* we observe clear indications for an extensive cellular retrenchment, best seen in the reduced number and size of the chloroplasts. The proteomics and qPCR results indicate that not only the iron-rich photosynthetic machinery is affected, but that other cellular components also encounter a large-scale reduction, resulting in a
decreased growth rate. In addition, low-iron cells have a significantly lower protein biomass with roughly half of the iron-replete value. We speculate that during the transformation from a high-iron/high-biomass cell to a low-iron/low-biomass cell the cellular biomass itself may serve as a supplementary energy source to compensate for the decrease in photosynthetic performance, i.e. carbon fixation and generation of ATP. This is in line with the observation of increased cellular vesiculation at low-iron (light microscopy) and increased lipid metabolism (transcriptomics and proteomics). It is also consistent with the relative increase of the mitochondrial respiratory machinery as deduced from the transcriptomics and proteomics data. As phototrophy and thereby the chloroplast energy metabolism is largely impaired at low-iron, mitochondrial respiration may provide a more constant and robust energy source that is retained, i.e. excluded from the observed cellular retrenchment. Hence, we can describe the principal cellular changes observed at iron limitation as a metabolic shift with a gradual take-over of the energy metabolism by the mitochondrial system. The cellular maintenance under iron-limited conditions is further supported from biomass recycling through the action of isocitrate lyase and glutamine synthetase.

Moreover, changes in the photosynthetic machinery are likely a consequence of a coordinated remodelling process indicating an intricate regulatory network that adjusts cellular energy demand in response to the availability of iron. The active remodelling as a response to low iron is an unexpected result since it was proposed earlier that photosynthesis of *T. oceanica* is constitutively adapted to a low-iron environment [12]. The observation of distinct cellular phenotypes in iron-limited versus iron-replete cultures may be due to the differing iron levels used in iron-replete control cultures, with ~60 nM applied by Strzepeck and Harrison compared to saturating 10 µM $FeCl_2$ used in this work, though not expected to ever occur under

natural conditions. Nevertheless our data demonstrate that *T. oceanica* possesses the potential to remodel its bioenergetics pathways in response to iron-availability.

While the cultures in the present experimental setup were held under axenic photoautotrophic conditions, the situation in the natural context of the open ocean is very different. A major difference found between the artificially induced iron limitation of photoautotrophically growing cultures and the iron limitation encountered by diatoms in their natural habitats is the ubiquitous presence of particulate and dissolved organic matter (POM/DOM) in the latter, albeit dilute as in the case of the oligotrophic Sargasso Sea [39].

Some of the strongest up-regulated transcripts at low-iron were found to be targeted to the secretory pathway, i.e. with the cell-surface or the vacuolar system as their destination. The functional annotation of the respective protein models reveals a complex suite of molecules capable of adhesion or degradative functions, suggesting a possible role in a mixotrophic context (Fig. 5). Observations of diatom mixotrophy have been reported for five decades [40]. This characteristic has recently been explored for biotechnological application [41]. A principle metabolic competency for heterotrophic nutrition was demonstrated in transgenic *P. tricornutum* that were able to grow on sugar upon expression of a transgenic sugar transporter [42].

A hypothetical capability for mixotrophic nutrition and the conjoined ability to feed on dissolved and/or particulate organic matter would place diatoms in close proximity to the microbial loop responsible for recycling organic matter in the marine food web. Moreover, this would contribute to resolving the paradox of diatom survival in a low-iron world, as DOM/POM can be expected to be a relevant iron source. On the other hand, utilization of organic carbon/iron sources in low-iron conditions will immediately put the cell into competition with the bacterial community of diverse and specialized heterotrophs. In line with this scenario we find an up-regulation of both identified NRPS enzymes (Additional file 3 – Fig. S8 "qPCR $\Delta\Delta C_T$") that may mediate a bacterial defence response under iron-limited conditions.

Evolutionary roots and the impact of genome plasticity on adaptation to low-iron

Diatoms are the outcome of an endosymbiotic event that brought together two nutritional modes, each somewhat pre-adapted to low-iron conditions: Phototrophy was contributed by the red-type plastids of the red algal endosymbiont, whose elemental composition shows a reduced iron requirement relative to the green-type plastids [8]; the fungal-type host (related to ancient Oomycetes) contributed an efficient heterotrophic machinery that might be retained to some

extent in today's diatoms (compare [43]), thereby enabling exploitation of POM/DOM as a supplementary carbon and iron source. Survival under iron-limited conditions would have benefited from both the ancestral host's and symbiont's characteristics.

Additional adaptive strategies to the low-iron environment have probably also originated from the high degree of horizontal gene transfer, with approximately 5% of the conserved genes encoded in the *T. oceanica* genome being assigned to diverse taxonomic groups. This is consistent with the findings from the *Phaeodactylum* genome project [5] and appears to be a recurrent topic in diatom genomics. A prerequisite for integration of foreign DNA into a genome is its uptake, and an explanation for how lateral gene transfer might occur in heterotrophic cells has been proposed [44]. Accordingly, heterotrophic organisms would continuously take up and integrate new genes along with the organic matter they feed on, thereby creating some genetic redundancy that eventually leads to the permanent replacement of the genuine counterparts as long as no disadvantage is encountered. We take the proposed mechanism of lateral gene transfer, together with the extent of laterally acquired genes observed in *T. oceanica,* as additional evidence for mixotrophic potential in *T. oceanica*.

In low-iron conditions we observe the joint up-regulation of a reverse transcriptase and a CRE-like recombinase, thereby providing an appropriate mechanistic basis for genome rearrangements via transposon mobilisation. Moreover, these are activated at a time where enhanced DNA input through increased uptake of organic matter might be expected under natural conditions. Stress-induced transposon activation has been reported for higher organisms as well, e.g. [45].

The observed plasticity of diatom genomes clearly has ecological implications, as bacterial inventions such as genes that are beneficial in a competitive context might quickly find their way into diatom species and strains. We therefore state that the hypothesized close integration of diatoms within the microbial loop (due to mixotrophy) and their remarkable genomic plasticity (as seen from lateral gene transfer) are keys to diatoms' ecological success: While mixotrophy opens up complex sources for carbon, energy and nutrients, the high capacity for lateral acquisition of genetic material facilitates adaptation in the context of the resulting competition with bacteria for organic matter, nutrients and iron.

Iron uptake, cellular iron requirements and adaptation of species to low-iron

For a better characterization of the complex cellular low-iron response it is necessary to distinguish the iron-specific aspects directed at counteracting the shortage of the limiting nutrient

(substitution of iron-proteins, induction of high-affinity iron uptake) from the rather general stress response directed at the situation of impaired growth and cellular starvation (cellular retrenchment, chlorosis, metabolic shift). With regard to the cellular iron economy we can define some constraints for beneficial adaptations to low-iron [25].

Each evolutionary innovation that lowers the cellular iron requirements will favour survival under iron-limited conditions. Besides general biomass retrenchment, i.e. the cellular adoption of a state with decreased biomass, the cell can actively lower its iron requirements through substitution of iron-containing proteins like ferredoxin or other metal enzymes that use iron as a cofactor. In addition to the well-known substitution pair ferredoxin/flavodoxin [10], we present evidence for the replacement of three metal-containing FBA enzymes by substitutes that use an aminoacid-based Schiff-base catalysis instead. Differential regulation of diatom FBA was recently described for *P. tricornutum* and appears to be an evolutionary conserved feature of several diatoms [26]. In *T. oceanica* we even find the apparently permanent functional replacement of an iron-rich counterpart as in the case of plastocyanin, though at the same time the genome harbours two genes encoding for cytochrome c_6 whose function and regulation remain unknown. Cytochrome c_6 might be retained in the genome for replacing plastocyanin under specific circumstances like copper-limitation, though this could not yet be observed [11]. Benefits may also arise from an improved regulation of mutual substitution pairs, e.g. as in the case of the transfer of the usually organellar ferredoxin gene *petF* to the nuclear compartment as found in *T. oceanica* [17].

Further, any improvement of the cellular iron-affinity and/or iron-uptake system will improve the competitive fitness under low-iron conditions. The activation of a specific high-affinity uptake system as observed in yeast is expected to occur in diatoms as well. It has been found that reduction of Fe^{3+} to Fe^{2+} represents an essential step in uptake of organically complexed iron [46], suggesting that iron is extracted from its complexes prior to uptake. Accordingly, major players involved in iron uptake can be expected to be iron-complex-binding receptors, redox enzymes needed for extracting the iron from its complexes, possibly also specific iron-binding cell-surface molecules for short-term iron storage (like *D. salina* transferrins), ferric reductases and ferrous oxidases responsible for interconversion of the iron redox species +III and +II, and finally iron permeases for iron import. In this work on *T. oceanica* we were able to identify several candidate elements for the above groups. However, critical for iron-uptake kinetics is the overall iron-binding capacity of the cell-surface that directly depends on the sheer amount of iron-binding sites exposed to the cellular exterior [25]. A straight-forward strategy to enhance the cellular capacity for iron-binding is seen in the remarkable

extent of domain duplications in iron-regulated cell surface proteins. From the strong and exclusive expression in iron-limited conditions we speculate that ISIP1/ISIP3 are part of a specialized high-affinity iron-uptake system,with ISIP1 as the putative receptor (Fig. 6). C-terminal endocytosis motifs as seen in ISIP1 (Fig. 6) can also be found in other iron-regulated proteins (*T. oceanica* CREGx2, *P. tricornutum* ISIP2) and biochemical work is needed to confirm the location and proposed receptor function for these components. In how far such features are species-specific adaptations or rather common to diatoms can only be clarified by a comparative genomic approach. Notably, a screening for highly conserved iron-regulated genes in the genomes of *T. oceanica*, *T. pseudonana*, *P. tricornutum* and *F. cylindrus* revealed a correlation between the extent of a diatom's tolerance to low-iron and the presence of *ISIP1* and *PETE*, which directly impact cellular iron economy and uptake.

Conclusions

From their evolutionary roots diatoms already appear to be pre-adapted to low-iron conditions through the endosymbiotic acquisition of a "red"-type photosynthetic machinery. While they have retained an implicit capacity for mixotrophy, the main contribution to cellular growth under conditions where iron or other nutrients limit the build-up of biomass, like in the iron-limited southern ocean or the oligotrophic Sargasso Sea, can be expected to stem from photosynthetic carbon assimilation.

The combined efforts in genomics, transcriptomics and proteomics reveal an unexpected metabolic flexibility in response to iron availability for *Thalassiosira oceanica* CCMP1005. These responses include an extensive cellular retrenchment, the pronounced remodeling of bioenergetic pathways and an intrinsic shift to a mixotrophic life style. As a consequence of iron-deprivation, the photosynthetic machinery undergoes a remodelling into a photo-protected mode to cope with the overall decrease in photosynthetic electron transfer complexes. From the genomic and transcriptomic data we identify candidate components of a diatom high-affinity iron-uptake system, and we present a novel cellular strategy to enhance iron economy of phototrophic growth with the iron-regulated mutual substitution of three metal-containing FBAs.

The enormous genomic plasticity of *T. oceanica*, as seen from the large fraction of genes acquired through horizontal gene transfer, provides the platform for complex adaptations to the iron-limited ocean. The inferred dynamic exchange of genes between marine microbes suggests the genome of *T. oceanica* may not be an exceptional evolutionary invention, but rather that it may be seen as one possible outcome from a larger metagenomic gene pool. The

future comprehensive characterisation of this gene pool constitutes the ultimate challenge in appreciating the solutions that marine life found for defying the persistent shortage of iron in the open ocean.

Methods

Strains, cultures and physiology

The sequenced strain *Thalassiosira oceanica* (Hustedt) Hasle et Heimdal CCMP1005 [47] was obtained from CCMP [48].

T. oceanica cells were grown in 8 l batch cultures using iron-free f/2 nutrients in ASW (i.e. artificial seawater medium [49]) at 100 µE, 25°C and a 14/10 h light/dark cycle. Iron-replete cultures ("control") were supplied in excess of other essential nutrients (10 µM $FeCl_3$); no iron was added to the iron-limited cultures ("stress"), except for residual iron from the ASW salts, promoting iron-limited growth. Cells were harvested at late exponential growth phase by filtration on 5µm polycarbonate filters of 47 mm diameter, resuspended into a small volume of media, followed by centrifugation at 4°C for 10 min at 11000 rpm. Cell pellets were frozen in liquid N_2 and stored at -80°C.

For iron-limited cultures iron-free techniques were applied as follows. Culture bottles were composed of plastic material, washed and incubated for some days with 1 N HCl and rinsed with ultrapure MilliQ water. All additional supply for iron-free work was washed with 1 N HCl and stored in closure bags until use. Iron-limited cultures were best achieved from a minimal inoculation volume of 10-20 µl or less than 10 000 cells.

Throughout this work we compare cells from late exponential growth phase, though we recommend to use iron-limited cells from late stationary phase when working on specifically low-iron responsive genes (e.g. from compensation pairs or involved in iron uptake). For these, the expression level at late stationary phase is found to be even higher than in late exponential phase.

Total cellular protein was determined for iron-replete and iron-deplete cells as follows. 85.5 million cells each were concentrated by centrifugation. The resulting pellets were frozen in liquid N_2 and stored at -80°C. For protein determination pellets were redissolved and lysed in 200-300 µl SDS/CO_3 buffer with additional application of ultrasonication. Cell debris was precipitated by 4 min centrifugation at RT and max rpm. 5-10 µl of the supernatant served as input for BCA protein assay (Pierce). SDS/CO_3 buffer was composed of 4 % SDS, 68 mM Na_2CO_3 and 0.4 mM PMSF. PMSF was dissolved in EtOH to obtain a 100 mM stock solution.

The variable chlorophyll fluorescence F_v/F_m of *T. oceanica* cells was measured from fresh cultures with a PhytoPAM (PHYTO-PAM

Phytoplankton Analyzer, Heinz Walz GMBH, Effeltrich, Germany) [50] upon 5 minutes of dark incubation.

Comparative genomics was done with genome data available at JGI for *Thalassiosira pseudonana* (Hustedt) Hasle & Heimdal CCMP1335 [4], *Phaeodactylum tricornutum* Bohlin CCAP1055/1 [5] and *Fragilariopsis cylindrus* (Grunow) Krieger CCMP1102 [18].

Microscopy and confocal microscopy

T. oceanica was imaged *in vivo* using confocal laser scanning microscopy LSM 510 (Zeiss). Chlorophyll autofluorescence was excited at 488 nm (1% laser intensity), and emission recorded with a LP 650 nm filter. Images were made using a Plan-Neofluar 40x1.3 oil objective (Zeiss). Z-section image series were captured with LSM 510 v3.2 software (Zeiss). 3D reconstructions of the chlorophyll fluorescence signal were made using the cell surface area-/cell volume-analyzing "Surpass" program module in Imaris 7.1.1 (Bitplane). Images were segmented using consistent threshold values. Surface area grain size was set at 0.1µm. In 20 cells from both iron-replete and iron-deplete cultures, cell dimensions were calculated from transmission image measurements based on a cylinder model. Chloroplast dimensions were calculated from 3D chlorophyll autofluorescence signal reconstructions.

Despite imaging in a narrow time window (12:00±2 hr), growth in the *T. oceanica* cultures was not synchronized. Thus, the observed values for the parameters with a growth-dependent variability are accordingly distributed over a characteristic growth range. The representative cellular and chloroplast dimensions provided in Table 2 were therefore determined from the statistical mean of 20 cells by calculating back to a cell at the beginning of its cell cycle. For this purpose we used a regular cylinder that expands through gradual 2-fold increase of its height as a model for the diatom cell. Around the statistical mean from all cells (of a variable parameter like cell volume) we created a range whose higher end differs from the lower end by the factor of 2, thereby representing a 2-fold increase of that parameter during growth of the diatom cell. The lower end of that range is given in Table 2 as a representative value for a "cellular unit", i.e. a freshly divided cell at the beginning of its cell cycle.

Nucleic acid extraction and sequencing

T. oceanica CCMP1005 was grown as axenic clonal culture from a single cell isolate obtained from serial dilutions of a stock culture to extinction. Nuclear gDNA for sequencing of the *T. oceanica* genome was extracted from nutrient-replete cells and separated from organellar gDNA in a CsCl gradient (Fig. 1) using the alternative CTAB protocol for algae [51].The quality of nucleic acids was assessed from

NanoDrop UV absorption profiles and agarose gel electrophoresis. Second generation sequencing technology was applied to the gDNA as follows. After mechanical shearing by nebulization, followed by an end repair specific sequencing adaptors were ligated. The genomic DNA fragments were shotgun sequenced using massive parallel pyrosequencing [52] on a Roche 454 GS-FLX instrument (Roche, Penzberg, Germany) according to the manufacturer's protocol.

Total RNA for transcriptome sequencing was extracted from frozen pellets of iron-replete and iron-limited *T. oceanica* cells from late exponential growth phase using the QIAGEN RNeasy kit. RNA quality was assessed from NanoDrop UV absorption profiles and agarose gel electrophoresis. For reverse transcription of total RNA the SMART cDNA synthesis kit from Clontech was used with 1µg input material and 15 rounds of amplification. The size distribution of the obtained cDNA libraries were controlled with agarose gel electrophoresis and then subjected to Roche 454 sequencing as described above for gDNA.

Transcriptomics

Global gene expression was assessed through Roche 454 massive parallel pyrosequencing of cDNA libraries prepared from total RNA extracted from iron-deplete and iron-replete cultures. The 2x 95000 sequence reads from both libraries were pooled, cleaned from adapter ends and processed in a combined assembly revealing 11264 contigs (i.e. transcript fragments) that map to approx. 6500 distinct AUGUSTUS gene models (Additional file 9 – Table S5 "Transcriptome_Mapping"). The differential read contribution from the Fe(-) and Fe(+) libraries to each contig is taken as a measure for the differential transcription of the respective gene. For the purpose of statistically evaluating the gene expression level across our two cDNA libraries we applied a log-likelihood ratio test statistic as described in Stekel et a. 2000 [21].

Differentially regulated genes were first screened with T-ACE (Transcriptome analyses and comparison explorer; E Philipp, L Kraemer and P Rosenstiel, unpubl.), a transcriptome database browser that plots the assembled transcript fragments according to their differential regulation (Fig. 4A) and provides information from BLAST analyses against NCBI nr protein (NR) and conserved domain database (CDD).

Models from selected genes were manually curated and annotated for protein function and location as described above (Additional file 5 – Table S2 "Transcriptome_Manual_Annotation").

RT-qPCR

RT-qPCR was done as in [17] and is described in the Additional file 3 (Supplemental Methods). Primers are listed in Additional file 10

(Supplemental Table S6 "RT-qPCR_Primers"). The differential regulation between high and low iron conditions with respect to 18S and RPB1 (threshold level) is shown in a $\Delta\Delta C_T$ plot with $\Delta\Delta C_T$ = [ΔC_T(Gene x) @ Fe(+)] - [ΔC_T(Gene x) @ Fe(-)] (Additional file 3 – Fig. S8 "qPCR $\Delta\Delta C_T$").

Proteomics

The proteomes of iron-replete and iron-deplete cells were analysed in a mass spectrometry approach. Differentially labelled iron-sufficient and iron-deficient cells were mixed at equal protein concentration and fractionated by SDS-PAGE. Protein bands were excised, digested in-gel with trypsin and subjected to LC-MS/MS analyses. Identification and quantitation of peptides and proteins was performed using Proteomatic data evaluation pipelines [53] as follows. To provide candidate peptides in the database search step using OMSSA [54], several sequence sources were used: (1) the AUGUSTUS gene models, (2) GPF peptides [55], (3) high quality chloroplast protein models, (4) a set of manually curated protein models. Resulting peptide/spectral matches (PSM) were filtered with a hit distinctiveness filter, using a threshold of 2. PSM were further filtered with a dynamically determined E-value threshold to achieve an estimated false discovery rate (FDR) of 1% [56]. Finally, all PSM with a precursor mass deviation greater than 5 ppm were discarded.

A total of 1695 peptides could be identified from two independent biological experiments and assigned to 767 protein groups. All identified peptides were subsequently quantified using qTrace [57], resulting in the quantitation of 633 protein groups and an additional set of 88 quantified peptides which were identified exclusively via GPF. For the determination of Fe(-)/Fe(+) protein ratios, all resulting combinations of peptide, SDS-PAGE band and charge state were grouped and all group ratios were combined into a total protein group ratio by calculating the median and interquartile range (IQR).

Data for differential protein expression as revealed from the mass spectrometry approach refers to equal amounts of total protein. For relating the observed regulation to a "cellular unit", the ratio of cellular protein biomass [Fe(-)/Fe(+)] as determined from a BCA protein assay, needs to be taken into account.

Bioinformatics

Bioinformatic analyses are described in the Additional file 3 (Supplemental Methods).

Data access

Thalassiosira oceanica CCMP1005 Whole Genome Shotgun Assembly is registered as bioproject 36595, *Thalassiosira oceanica* CCMP1005

Transcriptome Shotgun Assembly is registered as bioproject 73029. The genomic and transcriptomic Roche 454 GS FLX sequence reads from this study have been submitted to the NCBI Sequence Read Archive [58] under accession numbers SRA045826 and SRA045825 respectively. The Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AGNL00000000. The version described in this paper is the first version, AGNL01000000. AUGUSTUS gene models deduced from the genome assembly have been assigned the gene loci accession numbers THAOC_00001-THAOC_37921. The transcriptome assembly has been submitted to NCBI Transcriptome Shotgun Assembly database [59] under accession numbers JP288099-JP297710. The data from this study can be accessed in an integrated form at http://bose.geomar.de/cgi-bin/gbrowse/Toceanica/.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ML prepared the *T. oceanica* RNA used for transcriptomics, was involved in genomics, transcriptomics and proteomics data analysis, performed manual gene modelling and annotation, contributed the comparative genomics section, prepared the figures, and drafted the manuscript. MS developed the bioinformatic tools for analysis of the proteomics data and prepared Figure 4B. ML and MS conducted the gene prediction on the *T. oceanica* genome assembly using AUGUSTUS. ASR cultured the algae, carried out the RT-qPCR work and commented on the manuscript. LK performed the assembly of the Roche 454 reads and a general blast and domain annotation. RA set up the web-server and GMOD database, annotated gene models, and helped with genome and proteome analysis. MAG carried out the confocal microscopy and the 3D rendering of chlorophyll autofluorescence signals. JW and SVB carried out the proteomics experiment and participated in the proteomics analysis. MBS prepared the gDNA libraries, performed the Roche 454 sequencing and generated the initial assemblies. UCK prepared the cDNA libraries for transcriptomics. RGB and RA contributed the phylogenetic analysis of LGT genes. PR coordinated the sequencing and contributed to manuscript writing. MH coordinated the proteomics and contributed to manuscript writing. JLR coordinated the study, isolated the *T. oceanica* gDNA and contributed to manuscript writing. All authors read and approved the final manuscript.

Acknowledgements

We thank Prof. Stefan Rose-John (Institute of Biochemistry, Christian-Albrechts-University Kiel, Kiel, Germany) for advice in the isolation of nuclear genomic DNA from *T. oceanica* and the access to his laboratory and equipment, and Prof. Stefan Schreiber (Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany) for extensive help in building up sequencing resources in Kiel. Prof. Thomas Bosch (Institute of Zoology, Christian-Albrechts-University Kiel, Kiel, Germany) and Dr. Georg Hemmrich-Stanisak (Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany) provided help with the initial contig assembly.

We thank Tania Klüver (Leibniz Institute of Marine Sciences at Kiel University IFM-GEOMAR, Kiel, Germany) for help with the laboratory experiments and culturing of the algae. Dr. Dhwani Desai (Leibniz Institute of Marine Sciences at Kiel University IFM-GEOMAR, Kiel, Germany) helped with bioinformatics analyses and setup of the genome browser.

The upper left light micrograph in Fig. 1 showing *T. oceanica* in valve view is friendly courtesy of CCMP. Sequence data from *Fragilariopsis cylindrus* were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community.

This work was supported in part by a DFG grant to JLR (RO2138/6-1) and by funding from the DFG Cluster of Excellence "Future Ocean" (EXC 80) to JLR and PR. MH and JLR acknowledge funding from the BMBF "BIOACID" (03F0608N) program. RA received funding from EC FP7/2007-2011 under grant agreement n° PITN-GA-2008-215157 and EC FP7 Grant #205419 (ECOGENE).

References

1. Armbrust EV: **The life of diatoms in the world's oceans.** *Nature* 2009, **459**: 185-192.
2. Martin JH, Coale KH, Johnson KS, Fitzwater SE, Gordon RM, Tanner SJ, Hunter CN, Elrod VA, Nowicki JL, Coley TL, Barber RT, Lindley S, Watson AJ, Van Scoy K, Law CS, Liddicoat MI, Ling R, Stanton T, Stockel J, Collins C, Anderson A, Bidigare R, Ondrusek M, Latasa M, Millero FJ, Lee K, Yao W, Zhang JZ, Friederich G, Sakamoto C, et al.: **Testing the Iron Hypothesis in Ecosystems of the Equatorial Pacific-Ocean.** *Nature* 1994, **371**:123-129.
3. Boyd PW, Jickells T, Law CS, Blain S, Boyle EA, Buesseler KO, Coale KH, Cullen JJ, de Baar HJ, Follows M, Harvey M, Lancelot C, Levasseur M, Owens NP, Pollard R, Rivkin RB, Sarmiento J, Schoemann V, Smetacek V, Takeda S, Tsuda A, Turner S, Watson

AJ: **Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions.** *Science* 2007, **315**:612-617.

4. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, et al.: **The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism.** *Science* 2004, **306**:79-86.
5. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, et al.: **The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**:239-244.
6. Bowler C, Vardi A, Allen AE : **Oceanographic and Biogeochemical Insights from Diatom Genomes.** *Annual Review of Marine Science* 2010, **2**:333-365.
7. Grzebyk D, Schofield O, Vetriani C, Falkowski PG: **The mesozoic radiation of eukaryotic algae: The portable plastid hypothesis.** *J Phycol* 2003, **39**:259-267.
8. Quigg A, Finkel ZV, Irwin AJ, Rosenthal Y, Ho TY, Reinfelder JR, Schofield O, Morel FMM, Falkowski PG: **The evolutionary inheritance of elemental stoichiometry in marine phytoplankton.** *Nature* 2003, **425**:291-294.
9. Geider RJ, Julie LaRoche J: **The role of iron in phytoplankton photosynthesis, and the potential for iron limitation of primary productivity in the sea.** *Photosynth Res* 1994, **39**:275-301.
10. LaRoche J, Boyd PW, McKay RML, Geider RJ: **Flavodoxin as an in situ marker for iron stress in phytoplankton.** *Nature* 1996, **382**:802-805.
11. Peers G, Price NM: **Copper-containing plastocyanin used for electron transport by an oceanic diatom.** *Nature* 2006, **441**:341-344.
12. Strzepek RF, Harrison PJ: **Photosynthetic architecture differs in coastal and oceanic diatoms.** *Nature* 2004, **431**:689-692.
13. Brand LE, Sunda WG, Guillard RRL: **Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron.** *Limnol Oceanogr* 1983, **28**:1182-1198.
14. Rothberg JM, Leamon JH: **The development and impact of 454 sequencing.** *Nat Biotechnol* 2008, **26**:1117-1124.

15. von Dassow P, Petersen TW, Chepurnov VA, Armbrust EV: **Inter- and intraspecific relationships between nuclear DNA content and cell size in selected members of the centric diatom genus *Thalassiosira* (Bacillariophyceae).** *J Phycol* 2008, **44**:335-349.
16. Stanke M, Tzvetkova A, Morgenstern B: **AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome.** *Genome Biology* 2006, **7(Suppl 1)**:S11.
17. Lommer M, Roy AS, Schilhabel M, Schreiber S, Rosenstiel P, LaRoche J: **Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation.** *BMC Genomics* 2010, **11**:718.
18. ***Fragilariopsis cylindrus* - JGI Genome Browser** [http://genome.jgi-psf.org/Fracy1/Fracy1.download.html]
19. Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C: **Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation.** *PNAS* 2008, **105**:10438-10443.
20. Morel FMM, Rueter JG, Price NM: **Iron nutrition of phytoplankton and its possible importance in the ecology of ocean regions with high nutrient and low biomass.** *Oceanography* 1991, **4**:56-61.
21. Stekel DJ, Git Y, Falciani F: **The comparison of gene expression from multiple cDNA libraries.** *Genome Res* 2000, **10**:2055-2061.
22. Apel K, Hirt H: **Reactive oxygen species: metabolism, oxidative stress, and signal transduction.** *Annu Rev Plant Biol* 2004, **55**:373-399.
23. Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, Falciatore A, Finazzi G: **An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light.** *PNAS* 2010, **107**:18214-18219.
24. Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK : **An ancient light-harvesting protein is critical for the regulation of algal photosynthesis.** *Nature* 2009, **462**:518-521.
25. Morel FMM, Rueter JG, Price NM: **Iron Nutrition of Phytoplankton and its Possible Importance in the Ecology of Ocean Regions With High Nutrient and Low Biomass.** *Oceanography* 1991, **4**:56-61.
26. Allen AE, Moustafa A, Montsant A, Eckert A, Kroth PG, Bowler C: **Evolution and Functional Diversification of Fructose**

Bisphosphate Aldolase Genes in Photosynthetic Marine Diatoms. *Mol Biol Evol* 2011, doi:10.1093/molbev/msr223.
27. Donahue JL, Bownas JL, Niehaus WG, Larson TJ: **Purification and characterization of glpX-encoded fructose 1,6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*.** *J Bacteriol* 2000, **182**:5624-5627.
28. Blom NS, Tetreault S, Coulombe R, Sygusch J: **Novel active site in *Escherichia coli* fructose 1,6-bisphosphate aldolase.** *Nat Struct Biol* 1996, **3**:856-862.
29. Hall DR, Kemp LE, Leonard GA, Marshall K, Berry A, Hunter WN: **The organization of divalent cations in the active site of cadmium *Escherichia coli* fructose-1,6-bisphosphate aldolase.** *Acta Crystallogr D* 2003, **59**:611-614.
30. Hutchinson CR: **Polyketide and non-ribosomal peptide synthases: Falling together by coming apart.** *PNAS* 2003, **100**:3010-3012.
31. Hudson RJM, Morel FMM: **Trace metal transport by marine microorganisms: Implications of metal coordination kinetics.** *Deep Sea Res Part I Oceanogr Res Pap* 1992, **40**:129-150.
32. de Godoy LMF, Olsen JV, Cox J, Nielsen ML, Hubner NC, Froehlich F, Walther TC, Mann M: **Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast.** *Nature* 2008, **455**:1251-1254.
33. Chitnis VP, Chitnis PR: **PsaL subunit is required for the formation of photosystem I trimers in the cyanobacterium *Synechocystis* sp. PCC 6803.** *FEBS Lett* 1993, **336**:330-334.
34. Whitney LP, Lins JJ, Hughes MP, Wells ML, Chappell PD and Jenkins BD: **Characterization of Putative Iron Responsive Genes as Species-Specific Indicators of Iron Stress in Thalassiosiroid Diatoms.** *Front Microbio* 2011, **2**:234. doi: 10.3389/fmicb.2011.00234
35. Jeon H, Blacklow SC: **Structure and physiologic function of the low-density lipoprotein receptor.** *Annu Rev Biochem* 2005, **74**:535-562.
36. Paz Y, Katz A, Pick U : **Multicopper ferroxidase involved in iron binding to transferrins in Dunaliella salina plasma membranes.** *J Biol Chem* 2007, **282**:8658-8666.
37. Junier T, Pagni M: **Dotlet: diagonal plots in a web browser.** *Bioinformatics* 2000, **16**:178-179.
38. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, AAAI Press, Menlo Park, California* 1994, pp. 28-36.

39. Lomas MW, Burke AL, Lomas DA, Bell DW, Shen C, Dyhrman ST, Ammerman JW: **Sargasso Sea phosphorus biogeochemistry: an important role for dissolved organic phosphorus (DOP).** *Biogeosciences* 2010, **7**:695-710.
40. Lewin JC, Lewin RA: **Auxotrophy and Heterotrophy in Marine Littoral Diatoms.** *Can J Microbiol* 1960, **6**:127-134.
41. Garcia MCC, Camacho FG, Miron AS, Sevilla JMF, Chisti Y, Grima EM: **Mixotrophic Production of Marine Microalga *Phaeodactylum tricornutum* on Various Carbon Sources.** *J Microbiol Biotechnol* 2006, **16**:689-694.
42. Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR, Apt KE: **Trophic obligate conversion of an photoautotrophic organism through metabolic engineering.** *Science* 2001, **292**:2073-2075.
43. Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu HH, Fernie AR, Bowler C: **Evolution and metabolic significance of the urea cycle in photosynthetic diatoms.** *Nature* 2011, **473**:203-209.
44. Doolittle WE: **You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes.** *Trends Genet* 1998, **14**:307-311.
45. Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa AP, Le QH, Melayah D, Petit M, Poncet C, Tam SM, Van Sluys MA, Mhiri C: **Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae.** *Cytogenet Genome Res* 2005, **110**:229-241.
46. Shaked Y, Kustka AB, Morel FMM : **A general kinetic model for iron acquisition by eukaryotic phytoplankton.** *Limnol Oceanogr* 2005, **50**:872-882.
47. Hasle GR: **The Marine, Planktonic Diatoms *Thalassiosira-Oceanica* Sp-Nov and *Thalassiosira-Partheneia*.** *J Phycol* 1983, **19**:220-229.
48. Guillard RRL: **The Center for Culture of Marine Phytoplankton: History, Structure, Function and Future.** *J Protozool* 1988, **35**:255-256.
49. Goldman JC, McCarthy JJ: **Steady-State Growth and Ammonium Uptake of a Fast-Growing Marine Diatom.** *Limnol Oceanogr* 1978, **23**:695-703.
50. Schreiber U: **Chlorophyll fluorescence: new instruments for special applications.** *In: Garab G. (ed) "Photosynthesis: Mechanisms and Effects. Vol. V", Kluwer Academic Publishers, Dordrecht* 1998, pp. 4253-4258.
51. Lang BF, Burger G: **Purification of mitochondrial and plastid DNA.** *Nat Protoc* 2007, **2**:652-660.

52. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, et al.: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376-380.
53. Specht M, Kuhlgert S, Fufezan C, Hippler M: **Proteomics to go: Proteomatic enables the user-friendly creation of versatile MS/MS data evaluation workflows.** *Bioinformatics* 2011, **27**:1183-1184.
54. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH: **Open mass spectrometry search algorithm.** *J Proteome Res* 2004, **3**:958-964.
55. Specht M, Stanke M, Terashima M, Naumann-Busch B, Janssen I, Hohner R, Hom EF, Liang C, Hippler M: **Concerted action of the new Genomic Peptide Finder and AUGUSTUS allows for automated proteogenomic annotation of the *Chlamydomonas reinhardtii* genome.** *Proteomics* 2011, **11**:1814-1823.
56. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.** *Nat Methods* 2007, **4**:207-214.
57. Terashima M, Specht M, Naumann B, Hippler M: **Characterizing the Anaerobic Response of *Chlamydomonas reinhardtii* by Quantitative Proteomics.** *Mol Cell Proteomics* 2010, **9**:1514-1532.
58. **NCBI Sequence Read Archive** [http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi]
59. **NCBI Transcriptome Shotgun Assembly database** [http://www.ncbi.nlm.nih.gov/genbank/TSA.html]

Additional files

Bioinformatics methods (genome assembly, gene modelling, genome analysis and setup of the *T. oceanica* genome browser) as well as RT-qPCR and the respective references are provided as Additional file 3 (Supplemental Methods and Figures).

Though we achieved an estimated 8.7-fold coverage and though the genomic information is expected to be complete, the *T. oceanica* CCMP1005 genome remains fragmented and is by definition an "incomplete genome". The protein gene predictions from AUGUSTUS provide a solid starting point for more specific analyses, though these models often – especially for larger genes – suffer from inaccuracies in intron or 5'-end predictions (corresponding to the N-terminal end of the respective protein). While AUGUSTUS provides alternative

predictions for genes of uncertain structure, we submitted just one prediction model per gene (the first alternative prediction "t1") to NCBI, thereby avoiding redundance in the public database. These "t1"-models represent the entire *T. oceanica* proteome and have been assigned gene loci THAOC_0001-THAOC_37921. Best BLAST hits from a BLASTP analysis against NCBI nr protein and CDD databases are listet in Additional file 1 (Table S1 "AUGUSTUS_models_vs_nr+CDD").

While many of these AUGUSTUS *ab initio* predictions appear to reflect the correct gene structure, several others (especially at longer genes) apparently contain minor inaccuracies. For the purpose of functional analysis of transcriptomic and proteomic data a larger set of protein genes was selected for manual inspection and protein model improvement as described in the text. This set of 436 manually curated protein sequences from nuclear genes has been assigned custom identifiers (x1-x12, p1-p455) and is provided together with proteins encoded by the organellar genomes as Additional file 4 (Supplemental Sequences S1 "protein_user_models"). Their detailed annotation can be found in Additional file 5 (Supplemental Table S2 "Transcriptome_Manual_Annotation"). The sequences of proteins that we explicitly refer to in the discussion of the low-iron response are provided with Additional file 6 (Supplemental Sequences S2 "selection"), their annotation and regulation is presented in Additional file 7 (Supplemental Table S3 "Low-Iron_Responsive_Genes"). Deviation of a custom curated model from its respective official "t1"-AUGUSTUS prediction is indicated as "mod".

Additional file 9 (Supplemental Table S5 "Transcriptome_Mapping") provides a comprehensive overview for all 11264 transcript fragments with read statistics, best BLASTX hits to NCBI nr and CDD, and additional best TBLASTX hits to the genomes of *T. pseudonana*, *P. tricornutum* and *F. cylindrus*, notably including www link-outs to the respective orthologous genes of these species.

Phylogenetic trees for genes assumed to be acquired from lateral gene transfer are provided as Additional file 2 ("PhyloTrees_for_all_To_LGT_candidates.tre"). The file comprises 254 unrooted, Newick-formatted trees containing candidate LGT genes as described in the main manuscript. Trees and corresponding support values were generated using FastTree. Novel *T. oceanica* genes are identified in the trees by a reference number followed by '_TO'. Homologs from the 'nr' database identified using BLAST are given in the form 'Genus_GI', except for matches to unnamed genera which are shown as refseq GI number followed by 'X'. Visual inspection of trees was performed by importing this file into FigTree (Andrew Rambaut; http://tree.bio.ed.ac.uk/software/figtree/) and assigning a root as described in the Supplementary Methods.

Identifiers for iron-regulated diatom genes listed in Table 3 are provided in Additional file 8 (Supplemental Table S4 "Iron-Regulated Genes").

Primers used in the RT-qPCR experiments are listed in Additional file 10 (Supplemental Table S6 "RT-qPCR_Primers").

Genes and abbreviations used in this paper are listed in Additional file 11 ("Genes and Abbreviations").

3.5 P. tricornutum FCP Paper

Title:

The *Phaeodactylum tricornutum* superfamily of FCP-like proteins: Evolution, differential stress response and interaction model.

Predraft manuscript prepared for submission to:

Genome Biology

Author list:

Lommer M, Klüver T, LaRoche J.

Abstract:

BACKGROUND: Diatoms represent a major group of marine primary producers. Their photosynthetic performance crucially depends on the balance between light harvesting and excess energy dissipation. Both functions are mediated by a complex light antenna system composed of pigment-binding proteins that all belong to the same group of Fucoxanthin-Chlorophyll a/c-binding Proteins (FCPs). Members of this group are numerous and diverse, while their structural and biochemical similarity as well as their dynamic interactions impose major problems on a concise functional characterization.

RESULTS: The differential stress response of all 43 FCP-like genes present in *Phaeodactylum tricornutum* was assessed from RT-qPCR on cultures acclimated to combinations of three different stresses that induce enhanced energy dissipation - light, low iron and stationary phase. From protein sequence alignments and phylogenetic clustering of FCPs from *Phaeodactylum* and *Thalassiosira* we develop an evolutionary trajectory for the complete FCP-like superfamily. The combination of phylogenetic patterns with the differential stress response allows for a hypothetical functional categorization and the construction of a mechanistic model of FCP interactions.

CONCLUSIONS: The differential gene expression reveals distinct functional groups of FCP-like proteins that appear to be consistent with their phylogenetic clustering patterns. Here we distinguish seven major groups of *Phaeodactylum* FCP-like proteins: 3 ELIPs, 6 ancient FCPs, 4 PS II-type quenchers (CPQ), 5 PS II-type antennae (FCPB), 9 LHC (LHC), 9 PS I antennae (FCPA) and 7 PS I-type quenchers (FCPAQ). We present a consistent evolutionary trajectory for these groups. Our mechanistic model of FCP interactions proposes roles for most FCPs and awaits biochemical proof.

The *Phaeodactylum tricornutum* superfamily of FCP-like proteins: Evolution, differential stress response and interaction model

Markus Lommer[1], Tania Klüver[1], Julie LaRoche[1§]

[1] Helmholtz Centre for Ocean Research Kiel (GEOMAR), Kiel, Germany

[§] Corresponding author

Email addresses:
ML: mlommer@geomar.de
TK: tkluever@geomar.de
JLR: jlaroche@geomar.de

Abstract

Background

Diatoms represent a major group of marine primary producers. Their photosynthetic performance crucially depends on the balance between light harvesting and excess energy dissipation. Both functions are mediated by a complex light antenna system composed of pigment-binding proteins that all belong to the same group of Fucoxanthin-Chlorophyll a/c-binding Proteins (FCPs). Members of this group are numerous and diverse, while their structural and biochemical similarity as well as their dynamic interactions impose major problems on a concise functional characterization.

Results

The differential stress response of all 43 FCP-like genes present in *Phaeodactylum tricornutum* was assessed from RT-qPCR on cultures acclimated to combinations of three different stresses that induce enhanced energy dissipation - light, low iron and stationary phase. From protein sequence alignments and phylogenetic clustering of FCPs from *Phaeodactylum* and *Thalassiosira* we develop an evolutionary trajectory for the complete FCP-like superfamily. The combination of phylogenetic patterns with the differential stress response allows for a hypothetical functional categorization and the construction of a mechanistic model of FCP interactions.

Conclusions

The differential gene expression reveals distinct functional groups of FCP-like proteins that appear to be consistent with their phylogenetic clustering patterns. Here we distinguish seven major groups of *Phaeodactylum* FCP-like proteins: 3 ELIPs, 6 ancient FCPs, 4 PS II-type quenchers (CPQ), 5 PS II-type antennae (FCPB), 9 LHC (LHC), 9 PS I antennae (FCPA) and 7 PS I-type quenchers (FCPAQ). We present a consistent evolutionary trajectory for these groups. Our mechanistic model of FCP interactions proposes roles for most FCPs and awaits biochemical proof.

Background

Diatoms contribute approx. 40 % to global marine productivity [1]. Their strong impact on global carbon cycling has attracted increasing interest in the efficient phototrophic metabolism of this group. Diatoms originated from an ancient secondary endosymbiosis event between a fungal-like host and a red algal-like endosymbiont [2].

The raphide pennate species *Phaeodactylum tricornutum* [3] is a rather unusual diatom with respect to its reduced silicification and the polymorphic appearance with fusiform, triradiate and oval phenotypes. It can be cultivated axenically and has become a favorite diatom model due to its high tolerance to light stress and iron limitation [4]. *Phaeodactylum* was the second diatom and first pennate species to have its genome sequenced [5], favouring sophisticated molecular and transgenic approaches.

Diatom photosynthesis differs in several aspects significantly from the well characterized photosynthesis of green plants, not only because of the endosymbiotic origin of diatoms, but also because of the evolutionary early divergence of *green*-type (e.g. green algae and plants) and *red*-type (red algae and their descendants like stramenopiles, e.g. diatoms) photosynthetic groups. Pioneering work was done with experiments on the micronutrient-dependency of diatom growth [4] and the analysis of photosynthetic energy transfer by application of fluorescence techniques to *Phaeodactylum* [6]. However, a closer look to the underlying molecular machinery of *Phaeodactylum* photosynthesis has only become feasible with the publication of its 28 Mb genome [5].

The initial step in the photochemical reaction chain is the absorption of light through pigments that are coordinated on protein scaffolds. These proteins constitute – as a complex light antenna system – the interface between solar energy input and photochemical charge separation in the reaction centres of photosystem I (PS I) and photosystem II (PS II). These pigment-protein assemblages are capable of capturing solar energy as excited pigment states in terms of harvesting light for serving the photochemical reaction. At the same time they are involved in mechanisms to get rid of this energy – if in excess – by means of thermal energy dissipation. Besides highly conserved subunits that are part of the photosystem cores and have an intrinsic capability of light harvesting and energy dissipation, *green*- as well as *red*-type organisms have evolved specialized groups of accessory pigment-binding proteins that experienced a major radiation since the evolutionary divergence of these lineages. As a result, nowadays' *green* algae and plants are found to encode approx. 20 Chlorophyll A/B-binding proteins (CABs) most of which could be assigned to PS I antennae, trimeric light harvesting complexes (LHC) and PS II adapters for these LHC [7]. However, the fine regulation and especially the dynamic interaction between the different flavours of LHC proteins and the two photosystems remain obscure. The situation is worse in *red*-type diatoms like *Phaeodactylum*, where approx. 40 genes for Fucoxanthin-Chlorophyll a/c-binding Proteins (FCPs) are observed. They are assumed to confer the same functions of light harvesting and excess energy dissipation as seen

from the green CABs, though in an even more complex and dynamic context.

CABs and FCPs bind chlorophyll a as the main pigment, share similarities on the sequence as well as on the structural level and are believed to have a common origin [8]. Each protein is embedded in the thylakoid membrane with three transmembrane helices, of which the apparently homologous helices 1 and 3 contain highly conserved motifs that are involved in the coordination of chlorophyll and carotenoid pigments. While chlorophyll b and lutein are characteristic accessory pigments of plant and green algal CABs, the diatom FCPs coordinate chlorophyll c and fucoxanthin instead [8].

The kinetics of energy transfer from chlorophyll-binding proteins to photosystem II can be assessed through fluorescence-based approaches, enabling us to infer basic mechanisms of cellular energy processing in terms of light harvesting, exciton transfer and energy dissipation on the molecular level. While the harvesting of light energy can be regarded to be a rather passive event, its transfer or dissipation allows for some regulatory intervention. Traditionally, the possible fates for excited states at the site of photosystem II are distinguished as PS II *fluorescence* and mechanisms capable of lowering ("quenching") a maximal fluorescence level to a variable extent, i.e. photochemistry ("photochemical *fluorescence quenching*") and non-fluorescent energy dissipation NPQ ("Non-Photochemical *fluorescence* Q*uenching*") [9]. The dissipation of excess energy through NPQ is subject to cellular regulatory action and represents an essential property of the photosynthetic machinery, being indispensable for phototrophic survival in variable light regimes. A major component of energy dissipation via NPQ is energy (ΔpH) dependent and has been related to xanthophyll cycles (XC) [10] as well as to the presence of specific chlorophyll-binding proteins that appear to be required for NPQ. These proteins are up-regulated upon excess light and might control the connectivity between PS II, LHC and/or XC components [11]. In green plants an unusual member of the ELIP group (PSBS) has been found to be required for NPQ [12], whereas algae seem to use a presumably ancient chlorophyll-binding protein (LI818/LHCSR) for the same purpose [13].

The case of NPQ and PSBS/LI818 suggests that regulation of light harvesting and dissipation occurs largely on the protein level and crucially depends on the dynamic interactions between chlorophyll-binding proteins. However, much about these interactions and how competition between energy transfer and dissipation is influenced by regulatory action remains unclear to date.

The research on PSBS and LI818 further shows how difficult it is to unambiguously assign specific functions to chlorophyll-binding

proteins, and – more importantly – how different solutions have evolved in divergent evolutionary lineages to accomplish the very same task of NPQ. Therefore, research on diatom photosynthesis has started to look at the entirety of cellular chlorophyll-binding proteins to assess the systems' response rather than the response of single elements. More recently, a microarray-based transcriptional profile of *Phaeodactylum* FCPs during the course of high light acclimation has been presented [14] and Lepetit et al. could prove the expression of nearly all *Phaeodactylum* FCPs on the protein level by subjecting isolated FCP fractions to mass spectrometry [15]. Updated classifications of the light harvesting protein superfamily have been presented for eukaryotic algae [16, 17] or more specifically for (*red*-type) chlorophyll c-containing algae [18]. Macroevolutionary trajectories for these proteins have been proposed [16, 17, 19].

Here we focus on a single diatom species, *Phaeodactylum tricornutum*, to develop a mechanistic model of FCP interactions from the differential stress response of FCPs as observed upon induction of enhanced energy dissipation by light, low iron or stationary phase conditions. A comprehensive evolutionary trajectory for *Phaeodactylum* FCP-like proteins (ELIPs and FCPs) is inferred from phylogenetic clustering patterns and from the core motifs embedded in the conserved helices 1 and 3 of FCPs. The relationship between different groups of chlorophyll-binding proteins that are responsive to enhanced energy stress ("excitation pressure") and therefore are assumed to be involved in energy dissipation processes is discussed.

Results

Physiological acclimation of P. tricornutum to different light levels, low iron and stationary phase

Axenic cultures of *P. tricornutum* Bohlin CCAP1055/1 [5] were grown under different combinations of ambient light intensity and iron supply and were sampled in exponential and late stationary growth phase (Table 1). Growth rates for iron replete cultures indicate light limitation of growth at low light conditions and the capacity of *Phaeodactylum* for acclimation to high ambient light, while for iron limited cultures the highest growth rate is observed at low light conditions with a gradual decrease upon increase of ambient light. Besides its strong impact on the capacity for high light acclimation, low iron induces a characteristic chlorotic phenotype with much of the chloroplast membrane system retrenched into thylakoid organizing bodies (Additional file 1 - Figure S1).

Table 1 - *P. tricornutum* Samples, treatments and growth rates

sample #	**1**	**2**	**3**	**4**	**5**	**6**
light	LL	NL	HL	LL	NL	HL
iron	high	high	high	low	low	low
growth phase	exp.	exp.	exp.	exp.	exp.	exp.
growth rate μ [d^{-1}]	0.67 ± 0.06	0.83 ± 0.03	0.80 ± 0.02	0.23 ± 0.04	0.16 ± 0.08	0.09 ± 0.07
stress level	**MIN**	**control**	**HL**		**low iron**	**combined**
RT-qPCR	x	x	x	–	x	x

sample #	**7**	**8**	**9**	**10**	**11**	**12**
light	LL	NL	HL	LL	NL	HL
iron	high	high	high	low	low	low
growth phase	stat.	stat.	stat.	stat.	stat.	stat.
stress level						**MAX**
RT-qPCR	–	–	–	–	–	x

LL (low light) 50 μE $m^{-2}*s^{-1}$; NL (normal light) 150 μE $m^{-2}*s^{-1}$; HL (high light) 450 μE $m^{-2}*s^{-1}$.
high iron 10 μM $FeCl_3$; low iron 5 nM $FeCl_3$.
exp. exponential growth; stat. stationary phase.
The growth rate μ [d^{-1}] for exponentially growing cultures is provided as mean ± SD of two samples.
The overall stress level is assumed to increase from sample 1 to 6 and from sample 7 to 12 respectively.
Selection of samples for RT-qPCR analysis as indicated.

Pigment analysis with HPLC reveals a gradual light-dependent decrease in the main pigments (Chl a, Chl c and Fucoxanthin) under all conditions, with the absolute amount of xanthophyll cycle pigments (Ddx, Dtx) staying unaffected (Table 2). The deepoxidation state of the XC pigments is indicative of the cellular energy dissipation status and increases slightly with light. In agreement with the chlorotic phenotype, low iron cells have much lower pigment content compared to high iron controls. XC pigments are less affected and, hence, show a relative increase under these iron limited conditions. At stationary phase cellular pigment content is comparable to exponential phase for main pigments, but the XC pigment pool is enlarged. For the pigment-rich iron replete cells an extraordinary high deepoxidation state is observed.

Table 2 - *P. tricornutum* Cellular pigment content and xanthophyll de-epoxidation state

exponential growth	LL high iron	NL high iron	HL high iron	LL low iron	NL low iron	HL low iron	(low iron)$_{light}$ / (high iron)$_{light}$
Chl a [amol/cell]	445	305	176	41	22	14	0.08 ± 0.01
Chl c_1 + Chl c_2 [amol/cell]	130	77	44	22	10	6	0.15 ± 0.02
Fucoxanthin [amol/cell]	452	287	178	82	45	27	0.16 ± 0.02
Ddx + Dtx [amol/cell]	93	103	116	22	31	27	0.26 ± 0.04
Diadinoxanthin [amol/cell]	93	99	105	22	30	25	
Diatoxanthin [amol/cell]	0	4	11	0	1	2	
DES [Dtx / (Ddx + Dtx)]	0.00	0.04	0.09	0.00	0.03	0.07	
(Ddx + Dtx) / Fucoxanthin	0.21	0.36	0.65	0.27	0.69	1.00	

stationary phase	LL high iron	NL high iron	HL high iron	LL low iron	NL low iron	HL low iron	(low iron)$_{light}$ / (high iron)$_{light}$
Chl a [amol/cell]	444	117	108	37	17	18	0.13 ± 0.04
Chl c_1 + Chl c_2 [amol/cell]	116	68	48	23	8	4	0.13 ± 0.06
Fucoxanthin [amol/cell]	469	263	188	90	39	23	0.15 ± 0.04
Ddx + Dtx [amol/cell]	151	134	140	34	35	34	0.24 ± 0.02
Diadinoxanthin [amol/cell]	116	114	94	33	34	31	
Diatoxanthin [amol/cell]	35	20	46	1	1	3	
DES [Dtx / (Ddx + Dtx)]	0.23	0.15	0.33	0.03	0.03	0.09	
(Ddx + Dtx) / Fucoxanthin	0.32	0.51	0.74	0.38	0.90	1.48	

The photosynthetic performance of *Phaeodactylum* was assessed from the relative electron transport rates (rETR) measured during light curves with incident light levels increasing from 44 to 900 $\mu E*m^{-2}*s^{-1}$ (Figure 1). Maximal operating efficiency of PS II (operational quantum yield) is achieved in exponentially growing cells under iron replete conditions at the lowest light level, maximal rETRs are achieved under same conditions at highest non-inhibitory light levels with a strong impact of the experienced degree of light acclimation. Low iron and stationary phase both induce a strong

decrease in the PS II operating efficiency and, accordingly, rETRs are low. Minimal values are observed under the impact of both stresses combined (i.e. for chlorotic and severely iron starved cells).

The cellular light utilization is best characterized by the fractionation of absorbed light transferred to PS II into photochemistry ($\Phi_{PS\ II}$), constitutive ($\Phi_{f,\ D}$) and regulated energy dissipation (Φ_{NPQ}) at the site of PS II [20] (Additional file 1 - Figure S2). The fraction assigned to photochemistry was integrated across 55 to 800 $\mu E*m^{-2}*s^{-1}$ to obtain a representative value for the gross photosynthetic performance of *Phaeodactylum*. The distribution of these values across the parameters growth phase, iron supply and light conditions was visualized through box whisker plots (Additional file 1 - Figure S3) and reveals that the impact of iron supply and growth phase on photochemistry dominates over light acclimation effects. The dependence of photochemistry on iron and growth phase is summarized in an interaction plot (Additional file 1 - Figure S4).

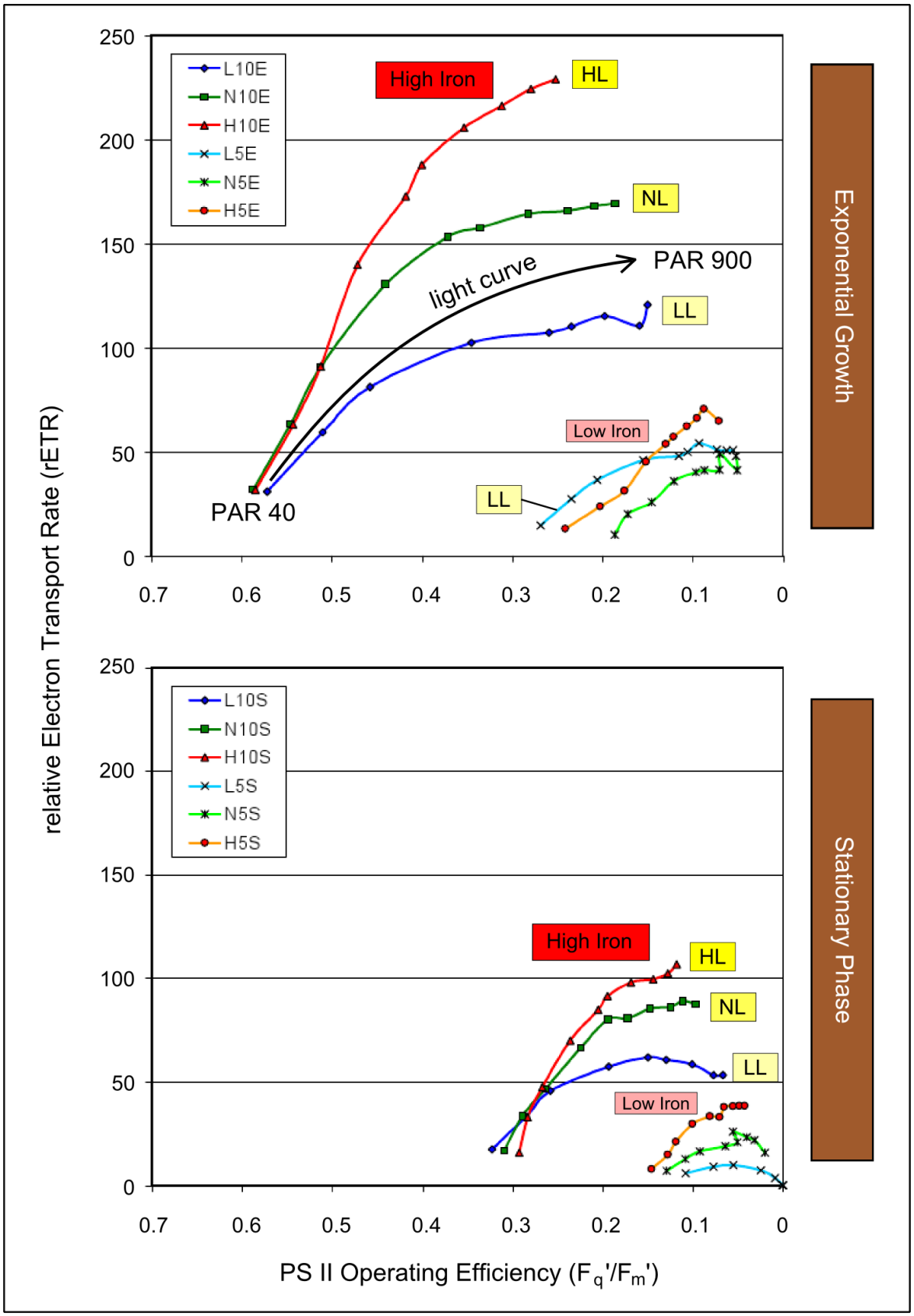


Figure 1 - Photosynthetic performance upon acclimation to different growth conditions

Figure 1 *(continued)*

The photosynthetic performance of *Phaeodactylum* varies upon acclimation to different combinations of ambient light intensity, iron supply and growth phase. The development of the relative Electron Transport Rate (rETR) and PS II quantum yield during increasing levels of ambient light (light curve) was determined from variable chlorophyll fluorescence and is shown for all cultures. The impact of iron supply and growth phase on photochemistry dominates over light acclimation effects.

L (LL) - low light 50 µE; N (NL) - normal light 150 µE; H (HL) - high light 450 µE; 10 - high iron 10 µM $FeCl_3$; 5 - low iron 5 nM $FeCl_3$; E – exponential growth; S – stationary phase; PAR – photosynthetically active radiation [µE].

Differential stress response of FCP-like, photosynthesis-related and low iron-responsive genes

The differential stress response of *Phaeodactylum* with respect to the parameters ambient light level, iron supply and growth phase was assessed at the transcriptional level by RT-qPCR on RNA from selected cultures that show an increasing need for energy dissipation (Table 1). Tested genes comprise the complete group of FCP-like proteins (listed and defined in Table 3), an additional group of genes involved in photosynthesis and mitochondrial metabolism, as well as a range of genes that are known to be highly responsive to low iron from earlier experiments on iron-limited *Phaeodactylum* [21] (Additional file 2 - Table S1). The 18S normalized expression level is provided for the control sample (NL) as relative transcript number on an arbitrary $\log_2$ scale, while differential gene expression is represented as the difference between the 18S normalized ΔC_T values of a sample and its control (Figure 2).

Among the FCP-like group the highest expression levels under control conditions (normal light, high iron, exponential growth) are achieved by members FCPB2, LHC5, LHC 8 and CPQ2 (expression levels >15.5). The majority of FCPs (subgroups FCP, FCPA, FCPB, LHC) exhibit expression levels between 12 and 15. Low expression between 7 and 10 is shown by ELIP1-2 and subgroups FCPAQ and CPQ. Among photosynthetic genes highest expression levels are achieved by the labile D1 protein (20), RubisCO large subunit (17) and carbonic anhydrase (16), while main components of the photosynthetic electron transport chain show a range of relatively high expression levels (11 to 15). Expression of tested mitochondrial enzymes is rather low (<8) under control conditions, as well as expression of many low iron-responsive genes like ISIP1, FCR and FRE2, whose products are all targeted to the cell surface (9 to 11). Of specific interest is the comparably high expression level (>15) of CPQ2 from the LI818-like FCP clade.

The impact of acclimation to different levels of ambient light on differential gene expression was assessed by comparing expression levels of iron replete and exponentially growing low light (LL) and high light (HL) samples to a normal light (NL) control (Figure 2). Among the FCP-like group we observe regulation patterns that correlate with the different FCP subgroups, just as for the expression level. Members of the subgroups FCP, FCPA, FCPB and LHC are up-regulated upon low light acclimation and down-regulated upon high light acclimation, whereas ELIP2 and subgroups FCPAQ and CPQ are further down at LL and up or unaffected at HL. The expression levels of the tested photosystem center subunits PsbA (PS II) and PsaA (PS I) are rather not light responsive with respect to the chosen standard 18S. The tested mitochondrial enzymes, as well as most genes from the low iron-responsive group show a clear low light response (down-regulation), but are not significantly responsive to high light.

The low iron response of exponentially growing cells was assessed by comparing the low iron NL and HL samples to the high iron NL control, with low iron HL expected to show the combined impact of high light and low iron stresses on cellular gene expression. Among the FCP-like group low iron leads to strong specific up-regulation of members FCP1, FCPB1&2, LHC3&9, FCPAQ1&3&4 and CPQ4. Conversely, genes for subunits involved in photosynthetic electron transport are rather down-regulated, while genes for mitochondrial enzymes show strong up-regulation. Genes that are known for their strong induction upon low iron show accordingly the highest effect. Additional high light stress increases 18S-normalized relative expression levels observed at low iron further for almost all genes tested.

Characteristics of the stationary growth phase were assessed by comparing a maximally stressed sample STAT (high light, low iron, stationary phase) to its exponentially growing control (high light, low iron, exponential growth). Here, a large-scale down-regulation of transcript levels is observed for almost all genes tested. Notable exceptions are the specific induction of CPQ4 and members of the FCPAQ subgroup.

Besides these general trends the stress response of some specific genes is of special interest for further discussion. Among the FCP-like group we observe co-regulation of ELIP1 & ELIP2, co-regulation of FCP1 & FCPAQ1, the differential regulation of CPQ3 & CPQ4 in response to low iron, the light-specific regulation of LHC8, and a low light response of CPQ4 & FCPAQ3 that is also found for genes under iron-specific control. Among the group of photosynthesis-related genes we observe the differential regulation of ferredoxin (PetF) & flavodoxin (FldA) in response to low iron, and the induction of an alternative ferredoxin-NADPH oxidoreductase (PetH2) at low iron or high light conditions. In the group of low iron-responsive genes we

observe the uniformly high expression level of ISIP1 in all low iron samples, the differential regulation of the bicarbonate transporters BCT1 & BCT2 in response to low iron or low light, the strong induction of the heat-shock transcription factor HSF2 upon low iron, and a significant low light response of genes that are suggested to be part of the high-affinity iron-uptake system (ISIP1, FCR and FRE2).

The gene for the RNA polymerase II large subunit RPB1, commonly used as a molecular standard, is differentially regulated at low iron with respect to 18S. This indicates that normalization to different commonly used standard genes will produce different results. RPB1 transcript levels are generally assumed to be more constant per cell, while expression of ribosomal components like 18S rRNA is rather correlating with the gross cellular activity or growth rate. While 18S is a proper reference standard for genes, whose activity correlates with growth rate, the better standard for genes, whose products are part of rather constitutive compartments like the cellular surface, is RPB1. Accordingly, for cell surface genes like FRE, FCR, ISIPs a true up-regulation (per cell) upon low iron is only observed, when their regulation overtops the relative regulation of RPB1 (2.1 for low iron, normal light, exponential growth). Notably, the only *Phaeodactylum* FRE gene, whose regulation matches these criteria of low iron-specific induction is FRE2. Moreover, the ISIP2 gene that has been regarded to be strongly induced upon low iron does not match these criteria. If expression on transcript level correlates with protein expression, ISIP2 rather represents a constitutive cell surface protein showing a virtual up-regulation in response to low iron as a result of its normalization to the growth rate dependent reference gene 18S.

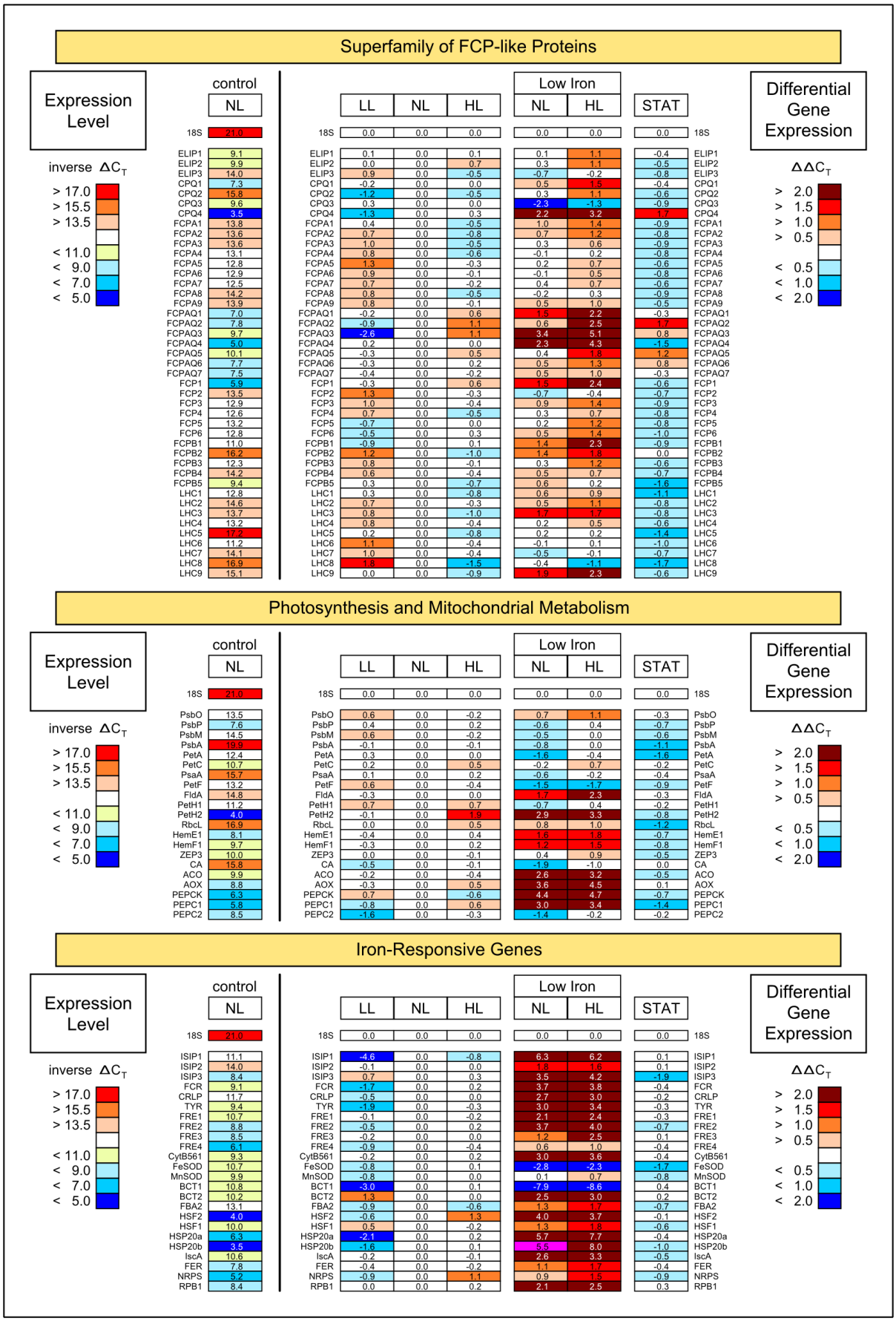


Figure 2 - Differential gene expression in response to light, iron supply and stationary phase

Figure 2 *(continued)*

Differential gene expression was tested with RT-qPCR on cultures that show an increasing need for energy dissipation. Relative transcript levels are shown for the "NL" control sample by inverting 18S-normalized ΔC_T values to obtain an arbitrary $\log_2$ scale (left). The differential expression with respect to the control sample is provided as $\Delta\Delta C_T$ with the exception of "STAT" that is compared to the "Low Iron & HL" sample instead (right) for the purpose of isolating the stationary phase effect. Expression data for the RPB1 gene (bottom), whose cellular transcript level is assumed to be relatively constant, suggests much of the up-regulation observed in the Low Iron samples for mitochondrial or cell surface protein genes are rather 18S-based normalization artefacts (see text). The FCP group shows a complex differential regulation suggesting a major remodelling of the photosynthetic apparatus. Most stress responsive are FCP genes from the CPQ and FCPAQ groups.

LL – low light, high iron, exponential growth; NL – normal light, high iron, exponential growth; HL – high light, high iron, exponential growth; Low Iron & NL – normal light, low iron, exponential growth; Low Iron & HL – high light, low iron, exponential growth; STAT – high light, low iron, stationary phase; gene abbreviations defined in Additional File 2.

Rational constraints for FCP categorization and for the assignment of a putative functional context to FCPs

Growing experimental evidence suggests many functional features of photosynthetic energy management be mediated by proteins of the FCP superfamily (FCPs and ELIPs). Among these functions is the harvesting of light energy as well as the different flavours of energy dissipation, whose regulated und thus more variable component NPQ can be detected by fluorescence techniques at the site of PS II, but not at the site of PS I or in the fraction of free FCPs. The main purpose of this work is to use the information derived from phylogenetic clustering and gene expression analyses as well as from biochemical findings [15] for proposing candidate FCPs for each of these mandatory functional groups as a working hypothesis for future biochemical approaches.

While the hypothetical assignment of a possible functional context to most FCPs inherently will remain highly speculative within this theoretical framework, we nevertheless can follow some rational objectives for providing a consistent categorization of FCP, for developing an evolutionary trajectory and for speculating on FCP interactions.

The underlying rational constraints are:

- FCPs preferentially form stable trimeric complexes.
- Co-occurence of FCPs in the same trimer or functional interaction may preferentially be expected from
  - FCPs co-located in the same phylogenetic cluster (similar sequence and structure),
  - FCPs that show a similar level of expression (RT-qPCR),
  - FCPs that show co-regulation or oppositional regulation with respect to their differential stress response (RT-qPCR).
- Each photosystem is expected to have a specific set of inner antenna FCPs, while additional FCPs may be recruited to the photosystems from a larger pool of accessory FCPs (in analogy to green plants).
- High conservation on sequence level (short phylogenetic branches) implies high conservation of function and is expected for rather ancient FCPs (e.g. FCPs that are closely associated with the two photosystems), while longer phylogenetic branches characterize derived and more modern FCPs.
- We propose that light harvesting as well as the two different flavours of energy dissipation, the rather constitutive and the regulated component, are mediated specifically by different FCPs.
- This should be the case at each of the different sites PS II, PS I and the accessory FCP fraction, though we experimentally can only detect the situation at PS II due to its specific fluoescence properties.
- Relative transcript levels are assumed to correlate with expression on the protein level for the majority of genes incl. FCPs.

Phylogeny and evolutionary trajectory for Phaeodactylum FCP-like proteins

The majority of *Phaeodactylum* FCP-like proteins share a conserved topology. Their precursors contain a group-specific chloroplast targeting sequence [22] and three transmembrane domains, two of which contain highly conserved motifs involved in pigment coordination [8]. A phylogenetic neighbour joining tree of these 40 different sequences (Figure 3) together with available experimental evidence for their regulation (RT-qPCR) or association to the two photosystems [15] provided the basis for our hypothetical FCP classification indicated in Figure 3A and presented in more detail in Table 3. Herein we attempt to make a basic distinction between FCPs that may preferentially act as light harvesting units (LHC, FCPB, FCP, FCPA) and FCPs that may rather be involved in energy dissipation processes (ELIP1-2, CPQ, FCPAQ, FCP1).

New categories for FCPs were assigned with regard to diverse lines of evidence that are presented in more detail in the following sections: A dense and short-branching cluster of FCPs at the bottom of the tree (Figure 3A) contains sequences that show the strongest expression (LHC5, LHC8) among all FCPs and that are found in the "free FCP fraction" upon biochemical fractionation [15]. We suggest these constitute the main light harvesting FCPs (Light-Harvesting Complex FCPs, *LHC*) in *Phaeodactylum*. A small group of FCPs that are similar to and constitute a monophyletic cluster with LHCs includes FCPB2, an FCP that – despite its strong expression – is neither found in the free FCP fraction nor associated to PS I [15]. We suggest this group rather to be specifically associated with PS II (photosystem II associated FCPs, *FCPB*) and to possibly interact with LHCs. FCPs that are co-purified with PS I are found in two clusters and are assumed to be specifically associated with PS I (photosystem I associated FCPs, *FCPA*). A group of FCPs that constitute a monophyletic cluster with FCPA, but are not co-purified with PS I, show low expression and a sophisticated differential stress response. We suggest these act as putative energy quenchers at the site of PS I (Quencher-type photosystem I associated FCPs, *FCPAQ*). Further we have a small cluster of FCPs that share low expression and differential stress response with FCPAQ, but are related to a *green*-lineage chlorophyll-binding protein found to be associated with energy dissipation at the site of PS II [13]. We suggest that these act as putative energy quenchers at the site of PS II (Quencher-type Chlorophyll-binding Protein CPQ). A small set of remaining irregularly branching FCPs is unspecifically denoted as FCP1-6, but assumed to rather be associated with PS II, as co-purification with PS I could not be achieved [15].

A joint tree of *P. tricornutum* and *Thalassiosira pseudonana* [23] (a distantly related, central diatom species) FCPs (Figure 3B) reveals that most bifurcate ortholog pairs between both species are found in the FCPA group, while other FCP categories show higher evolutionary dynamics. The combined tree shows a better separation of categories FCPA and FCPAQ as two parts of a monophyletic cluster. LHC proteins are placed in species-specific clusters indicating a major radiation of these proteins upon divergence of central (*T. p.*) and pennate (*P. t.*) diatoms. A comparison between this tree of *red*-type chlorophyll-binding proteins with a joint tree of *green*-lineage chlorophyll-binding proteins from *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* (Additional file 1 - Figure S5) reveals similar phylogenetic macropatterns with proteins that are assigned to the two photosystems being found in separate clusters. Like their *red*-type relatives the LHC proteins are located in species-specific subclusters indicating a major radiation of these proteins upon divergence of

Chlamydomonas and *Arabidopsis* lineages. Most importantly, the NPQ-related LHCSR proteins of *Chlamydomonas* (relatives of diatom CPQ) are isolated on a long phylogenetic branch with short ends, thereby suggesting a rather recent lateral acquisition for LHCSR genes.

Of specific interest is a higher resolution of the *Phaeodactylum* CPQ cluster (Figure 3C), because its members are assumed to be involved in NPQ. The observed branching pattern suggests CPQ3 be the most conserved CPQ protein and thus the most ancient. It is not responsive to light, but appears to be differentially regulated with CPQ4, the protein with next highest conservation and strongest stress response with respect to low iron. We suggest that CPQ3 may confer constitutive energy quenching to photosystem II ($\Phi_{f,\,D}$), while CPQ4 may replace CPQ3 to some extent under low iron stress and confer NPQ to photosystem II (Φ_{NPQ}). CPQ2 is less conserved thus evolutionary younger and its constitutively high expression level is comparable to that of main LHC subunits. We suggest that CPQ2 co-evolved with LHC and may confer constitutive energy quenching to the main LHC complexes (contributing to $\Phi_{f,\,D}$, if connected to PS II). The least conserved CPQ member is CPQ1. Its low expression and moderate stress response indicates a possible role as accessory energy quencher.

Further we provide here a higher resolution of the LHC cluster that includes some of the main light harvesting subunits found in *Phaeodactylum* chloroplast membranes (Figure 3D). *Phaeodactylum* LHC5 and LHC8 show highest expression and highest conservation. We suggest these constitute the main trimeric LHC complexes and are complemented by LHC9 under low iron stress conditions. Remaining LHC proteins are less conserved and may have a role as accessory LHC. Of special interest is the high conservation of FCPB1, FCPB3 and FCPB4 proteins adding evidence to the idea that these be assigned to photosystem II, possibly as adapters for more ancient light-harvesting FCP. Conversely, FCPB2 has experienced evolutionary change and its expression is comparable to that of main LHC proteins. We suggest this be a result of co-evolution with the LHC group and FCPB2 be an adapter for these modern LHC proteins.

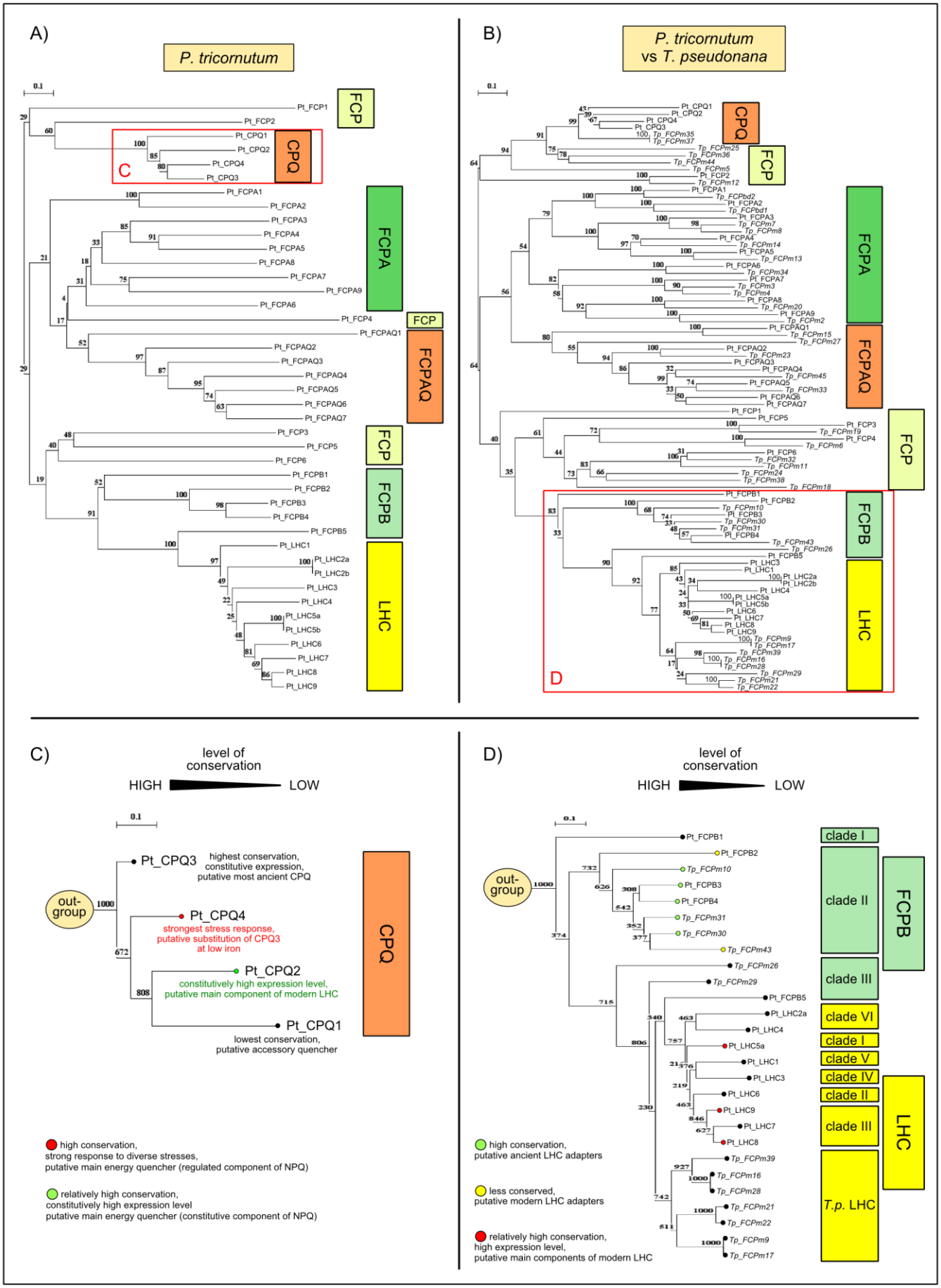


Figure 3 - Phylogeny of FCPs

Figure 3 *(continued)*
Phylogenetic relationships within the FCP family are presented as FCP neighbour joining trees for *P. tricornutum* (A) and *P. tricornutum* & *T. pseudonana* (B). New FCP categories were assigned to phylogenetic clusters from their expected association with PS II (FCPB, LHC, CPQ) or PS I (FCPA, FCPAQ) as developed in the text. Subtrees of higher resolution are provided for the *P. tricornutum* CPQ group (C) and the joint FCPB/LHC group of *P. tricornutum* and *T. pseudonana* (D). Increasing branch lengths indicate evolutionary progression and diversification (CPQ), while uniformly short branches rather suggest an evolutionary freezing of a functional invariant interaction (Pt_FCPB1/3/4).

A small and presumably ancient group of chlorophyll-binding proteins differs significantly from the FCP family and is treated separately as *ELIP*s (in analogy to green plants' *Early Light Induced Proteins* that differ from the CAB family). A schematic alignment of the three ELIPs of *Phaeodactylum* (Figure 4A) suggests the FCP-like three transmembrane domain (3-TM) protein ELIP3 be the product of a gene fusion between the 2-TM protein ELIP1 and the 1-TM protein ELIP1. ELIP1 and ELIP2 represent a co-regulated pair, while the expression of ELIP3 is rather in the range of major FCP proteins. All three ELIPs are highly conserved in the two distantly related diatoms *P. tricornutum* and *T. pseudonana* as well as in the cryptophyte *Guillardia theta* (Additional file 1 - Figure S6). Besides ELIP3 another important chlorophyll-binding protein appears to have evolved from an ELIP-like gene (Figure 4B). The green plant PSBS protein shows strong sequence similarities to the ELIP1 protein and could be the result of an ELIP1 gene duplication.

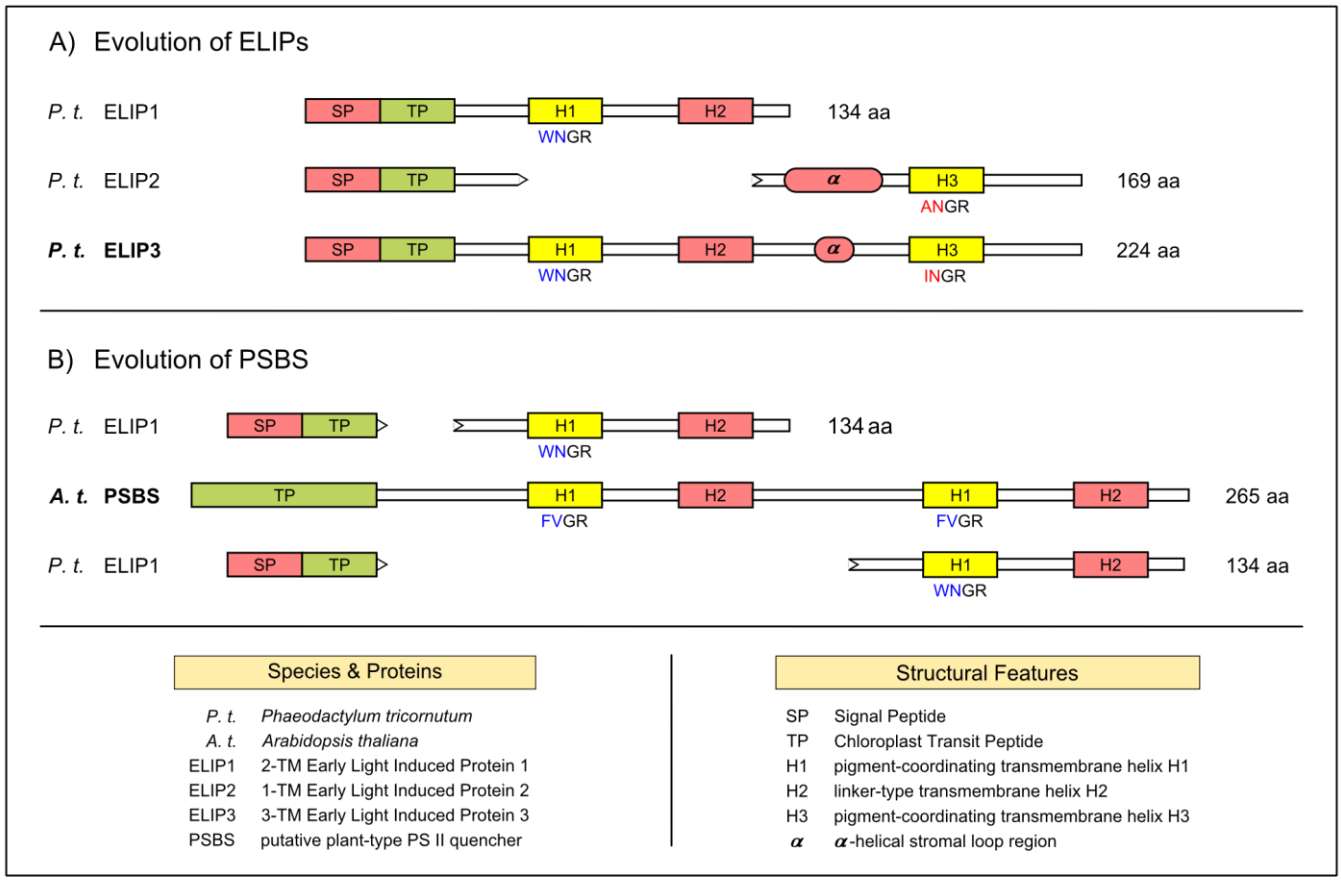


Figure 4 - Evolution of ELIPs and PSBS

The three homologous proteins presented as ELIP1-3 share characteristic features of chlorophyll-binding proteins with FCPs and CABs incl. the conserved transmembrane helices, but are sufficiently deviant on the sequence level to justify a separation as an own group. Extensive similarities on sequence and secondary structure level provide strong evidence for *P. tricornutum* ELIP3 to have evolved from a fusion of ELIP1 and ELIP2 (A). Moreover, the major part of the famous PSBS protein involved in NPQ of *green* plants is homologous to *Phaeodactylum* ELIP1 protein and could have evolved by duplication and fusion thereof (B). Characteristic amino acid motifs conserved in the transmembrane helices 1 and 3 of most FCPs and CABs are indicated for the respective orthologous helices of ELIPs with variable residues in blue for helix 1-type domains and in red for helix 3-type domains. Accordingly, ELIP3 shows a regular domain organization of helices as found in FCPs and CABs, while PSBS is lacking a helix 3-type domain as is the case for ELIP1.

The hypothetical categorization and the phylogenetic context of the *red*-type chlorophyll-binding proteins found in *Phaeodactylum* is summarized in an evolutionary trajectory for ELIPs, FCPs and CABs of the *red* and *green* lineages (Figure 5). On top we place the three ELIPs. We suggest ELIP1 be the most ancient component of all chlorophyll-binding proteins. The 1-TM protein ELIP2 may have evolved from the 2-TM protein ELIP1 by gene duplication, diversification and loss of the second transmembrane region, while ELIP3 is likely the product of a gene fusion of ELIP1 and ELIP2. ELIP1 and ELIP2 have orthologs in the cyanobacterial group and thus their evolutionary origin predates the primary endosymbiosis. ELIP3 already shows the membrane topology and domain patterns found in FCPs and CABs and it appears possible that evolution of these two groups of *red*-type and *green*-type chlorophyll-binding proteins is linked by a common ancestor (ProtoCP) that could have evolved from the ELIP3 gene. CABs and FCPs are very uniform protein families and likely originate from ancestral founder genes upon divergence of the *red* and *green* lineages of phototrophic algae (ProtoCAB and ProtoFCP, respectively). Subsequently the ProtoFCP diversified into the different clades of FCPs present to date (FCPA, FCPAQ, FCPB, CPQ and few other FCP clades). While we cannot resolve the secondary endosymbiosis leading to the stramenopile group (e.g. diatoms), there is evidence that this must have occurred after emergence of the FCPA clade, because orthologs are found in today's red algal genomes. LHCs are directly derived from the FCPB clade and experienced a major radiation upon divergence of the major diatom clades Centrales (incl. *Thalassiosira*) and Pennales (incl. *Phaeodactylum*). A special chlorophyll-binding protein implied in energy dissipation processes at PS II in green algae (LI818/LHCSR) is suggested to be acquired by means of lateral gene transfer from diatoms (CPQ). Additional support for the proposed evolutionary trajectory comes from the highly conserved aminoacid motifs embedded in the conserved transmembrane helices 1 and 3 of FCPs and CABs or the homologous helices found in ELIPs (characteristic four-residue consensus motifs are indicated on top of the FCP categories as "motif_helix1 x motif_helix3" with variable residues marked in blue and red, respectively).

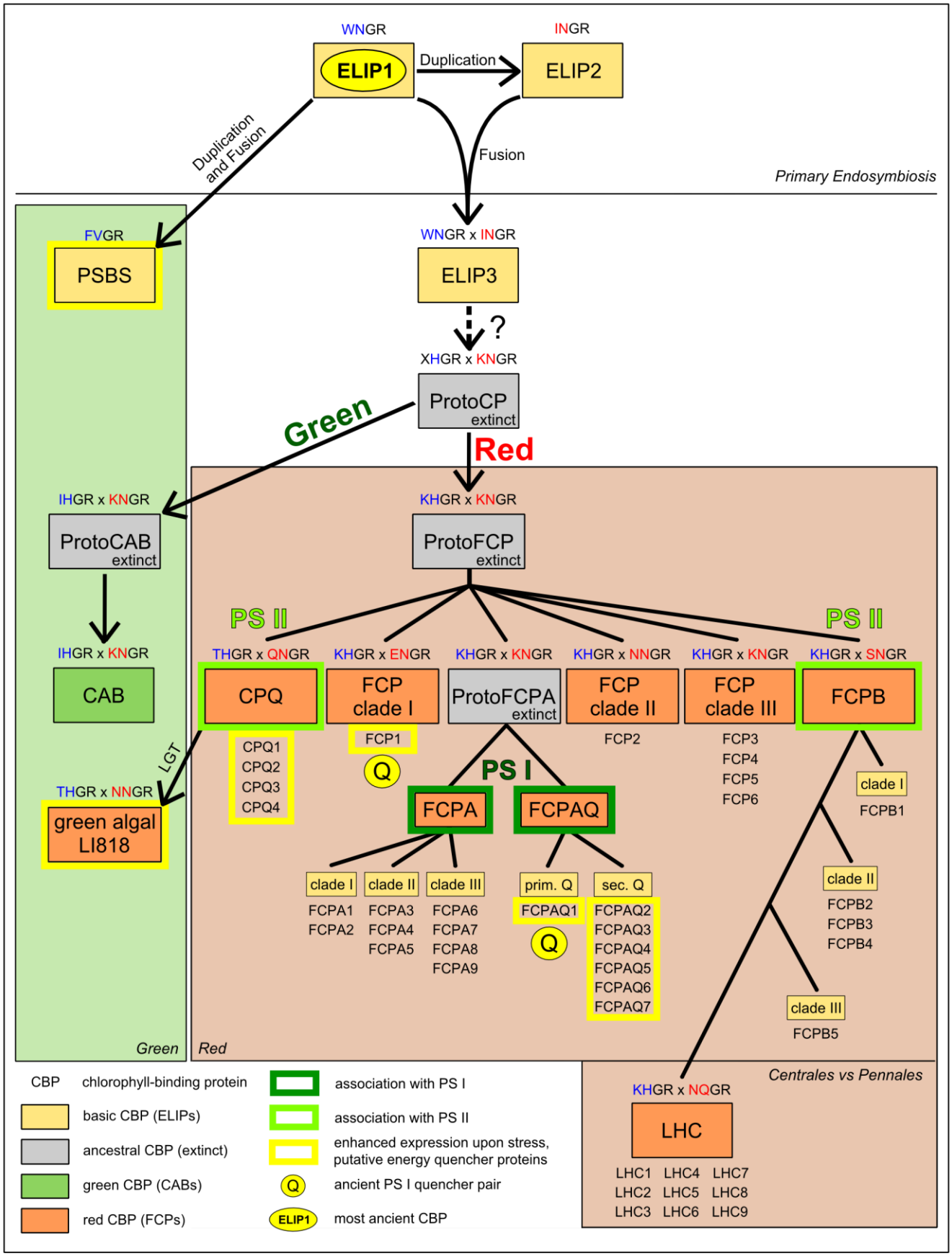


Figure 5 - Evolutionary trajectory for *P. tricornutum* FCP-like proteins

Figure 5 *(continued)*

Phylogenetic relationships among FCPs as well as similarities on sequence and secondary structure level between FCPs and ELIPs were translated into a consistent evolutionary trajectory for all FCP-like proteins present in *Phaeodactylum*. Here, the ELIP1 protein is considered to be the most ancient member, with ELIP2 originating from a duplicate ELIP1 gene by diversification and loss of the second helix.. Both are highly conserved across phylogenetic classes and their origin may predate the primary endosymbiosis event. ELIP3 is likely the product of a gene fusion between ELIP1 and ELIP2 that brought together the two distinct types of conserved chlorophyll-binding helices 1 and 3 present in all modern FCPs and CABs. Thus, it may have been an ELIP3-like ancestral protein that – after divergence of green- and red-type photosynthetic lineages – gave rise to FCPs and CABs, who subsequently experienced a major radiation into different clades.

While in the *red* branch the oldest FCPs may have served PS I as is the case in red algae (FCPA), the continuous expansion of FCPs may have led to groups specific for PS II (FCPB) and eventually to the extensive accessory antenna system that presumably is responsible for much of the light harvesting capacity of today's diatoms (LHC). LHC have undergone further radiation after divergence of central and pennate diatom lineages. Hypothetical association of FCP clades with specific photosystems is indicated by a dark (PS I) or light (PS II) green rim. Stress-responsive clades or FCPs (yellow rim, "Q") are suggested to be involved in energy dissipation processes.

Besides the CABs that make up the antenna system in the *green* lineage we find the PSBS protein that may have evolved from an ancestral ELIP1-like protein and that is present in all plants and several green algae (*Chlamydomonas*). In *Chlamydomonas*, but not in plants, we also observe a CPQ-like protein that appears to have functionally replaced a PSBS protein and that may have been acquired by horizontal gene transfer. A heterogeneous group of shorter chlorophyll-binding proteins (also called "ELIPs") that may have functional relevance for the antenna and/or XC system of green plants and green algae is not considered here.

Characteristic amino acid motifs conserved in the transmembrane helices 1 and 3 of most FCPs and CABs are indicated above each clade and support the presented scheme (variable residues in blue for helix 1-type domains and in red for helix 3-type domains). Specific members are listed below each FCP clade.

A mechanistic model of FCP interactions

FCPs are not autonomous components, but form a sophisticated system of closely interacting proteins ensuring proper perception, distribution and dissipation of light energy. The information derived from the phylogenetic context and differential regulation was used to speculate on specific interaction patterns between FCPs and the two photosystems in their native microenvironment, i.e. the chloroplast membrane (Figure 6).

From the lack of photosystem I PsaG and PsaH subunits as well as from the presence of a cyanobacterial-type PsaL subunit in diatom genomes we would expect the native state of PS I to be trimeric [24, 25]. However, experimental evidence from the diatom *Chaetoceros* rather suggests a monomeric PS I surrounded by approx. 8 FCP trimers [26]. In our model PS I is surrounded by an inner antenna system composed of FCPA1-FCPA9 that are co-purified with PS I, while the dimeric PS II may be closely associated with some FCPs from the FCPB group that could act as inner antennae and/or interface (adapter) for main and accessory LHC trimers. We distinguish the co-regulated pair ELIP1/ELIP2 from the remaining group of 3-TM FCPs who likely combine to form trimeric complexes functioning in light harvesting (LHC) or energy dissipation (quencher) at the site of PS II or PS I. By means of their high expression level we suggest LHC5, LHC8 constitute the main LHC complex responsible for the gross PS II energy supply. In this scenario the optional presence of CPQ2 may transform trimers into a "quenched" state with CPQ2 biochemically detectable as a 19 KDa subunit in contrast to the 18 KDa LHC5 and LHC8 subunits (Additional file 1 - Figure S7). FCPs with intermediate expression levels may form accessory LHC complexes composed of ancient (FCP subgroup) or rather modern (LHC subgroup) FCPs. Further, we assume co-regulated (RT-qPCR) or co-clustering (phylogeny) FCPs may be co-located in a shared trimer as indicated. Mature FCP trimers may be recruited to the photosystems according to specific needs for energy acquisition or dissipation. Induction of specific FCPs upon unfavourable conditions like high light, low iron or stationary phase may lead to specific substitution of FCPs (CPQ4 ←→ CPQ3) or increased integration of single FCPs (e.g. LHC9) into trimers or induction of certain trimer species, respectively. The complementary charged loops in FCPA9 and FCPAQ1 suggest a strong and specific interaction (putative PS I adapter site for accessory FCP trimers). Notably, the expression levels for ELIP1/ELIP2 and for FCPs assumed to be involved in energy dissipation processes (CPQ, FCPAQ, FCP1) are in the range of the zeaxanthin epoxidase ZEP3 expression, a major XC enzyme.

A major obstacle in the unambiguous biochemical identification of FCPs is a lack of precise expectations for the molecular weight of specific mature FCP proteins. Processing of an ER/chloroplast targeting peptide at a conserved motif "ASAFAP" has been described [22], but resulting molecular masses are still well above the 18 and 19 KDa frequently observed upon biochemical isolation of FCPs. Here we propose a second cleavage site at the next N-terminal bulky aromatic residue (phenylalanine F or tyrosin Y) for final removal of the remaining chloroplast transit peptide upon completion of chloroplast import (Additional file 1 - Figure S7). Especially with respect to the FCPs that show highest expression

levels and thus can be expected to represent a major fraction of FCPs in isolation procedures (LHC5, LHC8, FCPB2, CPQ2), resulting molecular weights are precisely around the values of 18 and 19 KDa reported for major *Phaeodactylum* FCP fractions [27].

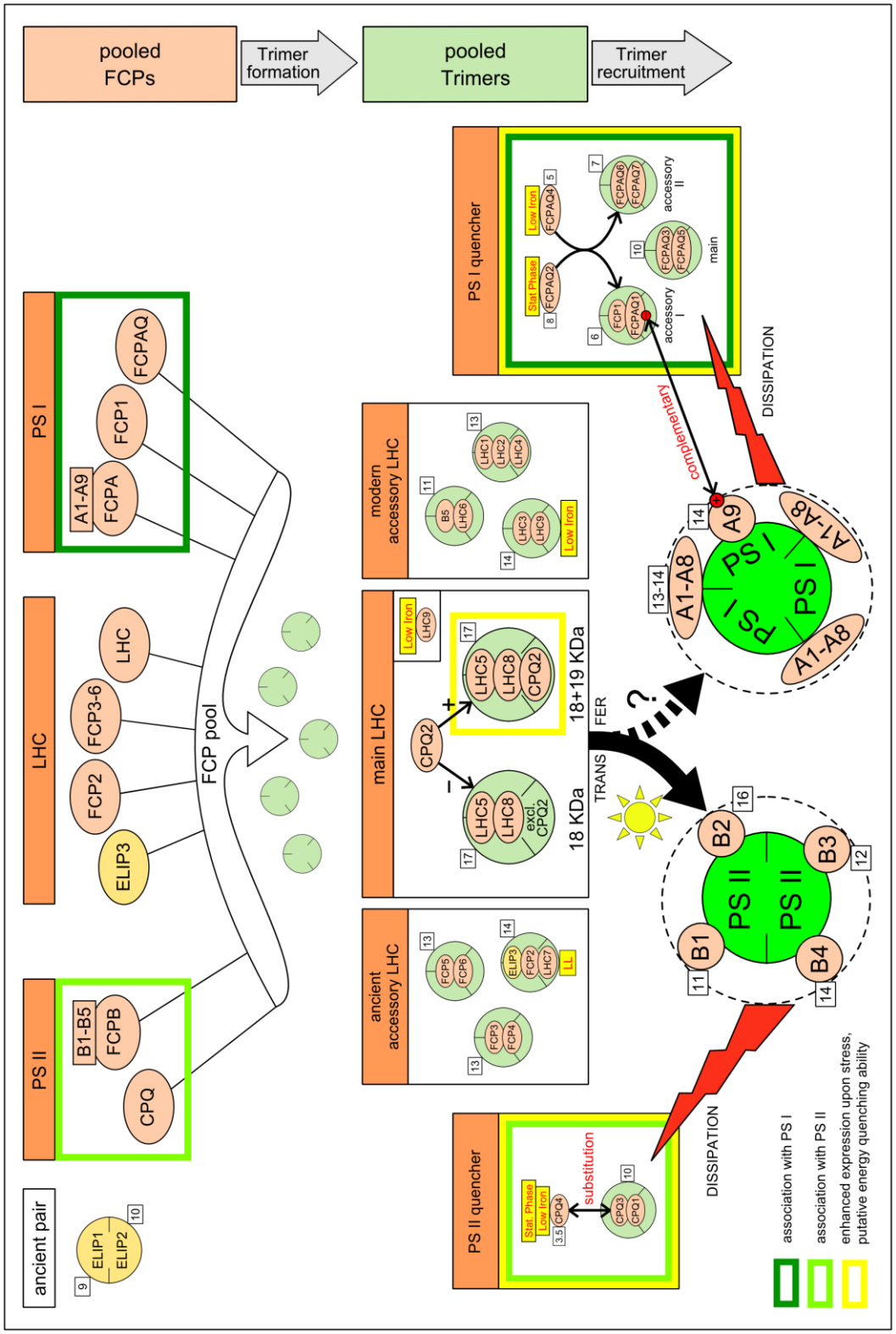


Figure 6 - Mechanistic model of FCP interactions

Figure 6 *(continued)*

Correlations between FCPs and ELIPs with respect to their expression level, differential regulation, phylogenetic clustering and their successful co-purification with specific photosystems were translated into a scheme of hypothetical FCP interactions that co-locates co-regulated FCPs to the same (virtual) trimer. FCPs and ELIPs are recruited from a common pool of chlorophyll-binding proteins to form trimers of a defined or rather random composition (top). Mature FCP trimers may be recruited to the photosystems according to the actual needs (center). Specific FCPs may form inner antennae for PS I (FCPA) and PS II (FCPB), respectively (bottom).

Phylogenetic FCP clades are assumed to either be specifically associated with photosystem I (FCPA, FCPAQ, FCP1) or photosystem II (FCPB, CPQ) or to form an external accessory antenna system (LHC, FCP2-6, ELIP3) – a tripartite organization as found in green plants. The observation of two FCP populations that show fundamentally different expression levels suggests FCPs with high expression levels may function in light-harvesting (FCPA, FCPB, LHC, FCP2-6, ELIP3), while FCPs with low expression levels may rather be involved in other processes like energy dissipation (FCPAQ, FCP1, CPQ, ELIP1-2). FCPB2, LHC5, LHC8 and CPQ2 show highest expression levels and may form the main accessory LHC antenna, with FCPB2 as a putative PS II adapter and CPQ2 as an optional quencher component (see text).

Several FCPs are specifically responsive to certain stress conditions like Low Light, Low Iron or Stationary Phase as indicated. The complementary charged loops of FCPA9 and FCPAQ1 suggest a specific interaction of these proteins. Relative expression of FCPs is indicated next to FCP symbols for normal growth conditions and is provided as relative transcript levels on an arbitrary $\log_2$ scale. Details are discussed in the text.

Table 3 - *P. tricornutum* superfamily of FCP-like proteins

NAME	CATEGORY	ID (NCBI)	SYNONYM [Nymark et al. 2009]	CLADES [Dittami et al. 2010]	CLADES [Neilson et al. 2010]	CLADES [Hoffman et al. 2011]
ELIP1	**ELIP**	XP_002181847			2TM-LIL (SEP-like)	
ELIP2	**ELIP**	XP_002184741			1TM-LIL (OHP-like)	
ELIP3	**ELIP**	XP_002177121			3TM-LIL (ELIP-like)	
CPQ1	**CPQ**	XP_002182760	LHCX4	LI818 / LHCX	LHCSR / LI818	V b2
CPQ2	**CPQ**	XP_002176987 mod	LHCX2	LI818 / LHCX	LHCSR / LI818	V b2
CPQ3	**CPQ**	XP_002179760	LHCX1	LI818 / LHCX	LHCSR / LI818	V b2
CPQ4	**CPQ**	XP_002178699	LHCX3	LI818 / LHCX	LHCSR / LI818	V b2
FCPA1	**FCPA clade I**	XP_002182329	LHCR14	LhcaR / LHCR	LhcaR	III
FCPA2	**FCPA clade I**	XP_002182162	LHCR13	LhcaR / LHCR	LhcaR	III
FCPA3	**FCPA clade II**	XP_002184127	LHCR11	LhcaR / LHCR	LhcaR	III a1
FCPA4	**FCPA clade II**	XP_002176857	LHCR12	LhcaR / LHCR	LhcaR	III a1
FCPA5	**FCPA clade II**	XP_002177385	LHCR4	LhcaR / LHCR	LhcaR	III a1
FCPA6	**FCPA clade III**	XP_002178624	LHCR1	LhcaR / LHCR	LhcaR	III
FCPA7	**FCPA clade III**	XP_002183608	LHCR2	LhcaR / LHCR	LhcaR	III
FCPA8	**FCPA clade III**	XP_002178019	LHCR3	LhcaR / LHCR	LhcaR	III
FCPA9	**FCPA clade III**	XP_002182909 mod	LHC6062	LhcaR / LHCR	LhcaR	III
FCPAQ1	**FCPAQ**	XP_002183911 mod	LHC15820	LhcaR / LHCR	LhcaR	III
FCPAQ2	**FCPAQ**	XP_002176917 mod	LHCR8	LHCZ	LHCZ	I
FCPAQ3	**FCPAQ**	XP_002184869	LHCR10	LHCZ	LHCZ	I
FCPAQ4	**FCPAQ**	XP_002177668	LHCR7	LHCZ	LHCZ	I
FCPAQ5	**FCPAQ**	XP_002181976	LHCR6	LHCZ	LHCZ	I
FCPAQ6	**FCPAQ**	XP_002186024	LHCR9	LHCZ	LHCZ	I
FCPAQ7	**FCPAQ**	XP_002182761 mod	LHCR5	LHCZ	LHCZ	I
FCP1	**FCP clade I**	XP_002177056		FCPs / LHCF	Chlorophyll a/c	
FCP2	**FCP clade II**	XP_002176735	LHC17531	FCPs / LHCF	Chlorophyll a/c	VI
FCP3	**FCP clade III**	XP_002183454		FCPs / LHCF	Chlorophyll a/c	
FCP4	**FCP clade III**	XP_002181795 mod	LHC13877	FCPs / LHCF	Chlorophyll a/c	
FCP5	**FCP clade IV**	XP_002185437	LHC24119	FCPs / LHCF	Chlorophyll a/c	
FCP6	**FCP clade IV**	XP_002178860	LHCF16	FCPs / LHCF	Chlorophyll a/c	IV b

[mod] Protein sequence used in this work was derived from a modified version of the respective gene model …

Table 3 - *P. tricornutum* superfamily of FCP-like proteins *(cont.)*

NAME	HYPOTHETICAL ANNOTATION	COMMENT	conserved core motif [helix 1]	conserved core motif [helix 3]
ELIP1	basic FCP assembly function (?)	coregulation of ELIP1 and ELIP2, ortholog in plants (PsbS)	WNGR	x
ELIP2	basic FCP assembly function (?)	coregulation of ELIP1 and ELIP2, ortholog in plants	x	ANGR
ELIP3	low light LHC	coregulation of ELIP3, FCP2 and LHC7, specific low light response	WNGR	INGR
CPQ1	acc CPQ	x	THGR	QNGR
CPQ2	LHC quenching 19 KDa subunit	coregulation of CPQ2 and FCPB2	THGR	QNGR
CPQ3	substitute of PS II main quencher	differential regulation of CPQ3 and CPQ4	THGR	QNGR
CPQ4	PS II main quencher	differential regulation of CPQ3 and CPQ4, strong response to high light, low iron and stationary phase	THGR	QNGR
FCPA1	PS I antenna	copurified with PS I	KHGR	KNGR
FCPA2	PS I antenna	copurified with PS I	RHGR	LNGR
FCPA3	PS I antenna	copurified with PS I	KHCR	NNGR
FCPA4	PS I antenna	copurified with PS I	KHGR	MNGR
FCPA5	PS I antenna	copurified with PS I	KHGR	TNGR
FCPA6	PS I antenna	copurified with PS I	KHGR	KNGR
FCPA7	PS I antenna	copurified with PS I	KHCR	KNGR
FCPA8	PS I antenna	copurified with PS I	KHGR	VHCR
FCPA9	PS I adapter for FCPAQ1	structural complementation between FCPA9 and FCPAQ1	MHGR	KHAR
FCPAQ1	ancient PS I quencher	structural complementation between FCPA9 and FCPAQ1, coregulation of FCP1 and FCPAQ1	RHGR	FNGR
FCPAQ2	PS I stat phase quencher	strong response to stationary phase	KHCR	KNGR
FCPAQ3	PS I main quencher	strong response to high light, low iron and stationary phase	KHGR	KNGR
FCPAQ4	PS I low iron quencher	strong response to low iron	KHAR	KHGR
FCPAQ5	PS I acc quencher	x	KHAR	KNGR
FCPAQ6	PS I quencher pair AQ6/AQ7	coregulation of FCPAQ6 and FCPAQ7	KHAR	KNGR
FCPAQ7	PS I quencher pair AQ6/AQ7	coregulation of FCPAQ6 and FCPAQ7	KHGR	KHGR
FCP1	ancient PS I quencher	coregulation of FCP1 and FCPAQ1	KHGR	ENGR
FCP2	ancient low light LHC	coregulation of ELIP3, FCP2 and LHC7, specific low light response	KHGR	NNGR
FCP3	ancient LHC	phylogenetic clustering of FCP3 and FCP4	SNGR	KNGR
FCP4	ancient LHC	phylogenetic clustering of FCP3 and FCP4	KHGR	KNGR
FCP5	ancient acc LHC	coregulation of FCP5 and FCP6, responsive to cumulative stress	KHGR	QNGR
FCP6	ancient acc LHC	coregulation of FCP5 and FCP6, responsive to cumulative stress	KHSR	NNGR

… and differs from the NCBI reference.

Table 3 - *P. tricornutum* superfamily of FCP-like proteins *(cont.)*

NAME	CATEGORY	ID (NCBI)	SYNONYM [Nymark et al. 2009]	CLADES [Dittami et al. 2010]	CLADES [Neilson et al. 2010]	CLADES [Hoffman et al. 2011]
FCPB1	**FCPB clade I**	XP_002184763	LHCF17	FCPs / LHCF	Chlorophyll a/c	
FCPB2	**FCPB clade II**	XP_002183381	LHCF15	FCPs / LHCF	Chlorophyll a/c	VII b
FCPB3	**FCPB clade II**	XP_002186206	LHCF14	FCPs / LHCF	Chlorophyll a/c	VII b
FCPB4	**FCPB clade II**	XP_002183291	LHCF13	FCPs / LHCF	Chlorophyll a/c	VII b
FCPB5	**FCPB clade III**	XP_002184765 mod	LHCF12	FCPs / LHCF	Chlorophyll a/c	VII e1
LHC1	**LHC clade V**	XP_002182937	LHCF8	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC2	**LHC clade VI**	XP_002182305 XP_002184540	LHCF6 LHCF7	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC3	**LHC clade IV**	XP_002182219	LHCF10	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC4	**LHC clade VI**	XP_002183709	LHCF9	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC5	**LHC clade I**	XP_002177868 XP_002177869	LHCF3 LHCF4	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC6	**LHC clade II**	XP_002177870	LHCF2	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC7	**LHC clade III**	XP_002177871	LHCF1	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC8	**LHC clade III**	XP_002184620	LHCF5	FCPs / LHCF	Chlorophyll a/c	VII e2.2
LHC9	**LHC clade III**	XP_002184619	LHCF11	FCPs / LHCF	Chlorophyll a/c	VII e2.2

[mod] Protein sequence used in this work was derived from a modified version of the respective gene model …

Table 3 - *P. tricornutum* superfamily of FCP-like proteins *(cont.)*

NAME	HYPOTHETICAL ANNOTATION	COMMENT	conserved core motif [helix 1]	conserved core motif [helix 3]
FCPB1	ancient FCPB-like low iron LHC	strong response to low iron	KHGR	SNGR
FCPB2	PS II adapter for main LHC (?)	coregulation of CPQ2 and FCPB2, not found in free FCP fraction [Lepetit et al. 2010]	KNGR	SNGR
FCPB3	PS II adapter for acc LHC (?)	phylogenetic clustering with FCPB2	KHGR	QNGR
FCPB4	PS II adapter for acc LHC (?)	phylogenetic clustering with FCPB2	KHGR	QHGR
FCPB5	Proto-LHC	LHC-like core motif, same expression level as CPQ3	KHGR	NNGR
LHC1	acc LHC	x	KHGR	NQGR
LHC2	acc LHC	x	KHGR	NQGR
LHC3	low iron LHC	strong response to low iron (like LHC9)	KHGR	NQGR
LHC4	acc LHC	x	KHGR	NQGR
LHC5	LHC basic 18 KDa component	main antenna subunit, LHC scaffold (?)	KHGR	NQGR
LHC6	acc low light LHC	specific low light response (like ELIP3, FCP2 and LHC7)	KHGR	NQGR
LHC7	low light LHC	coregulation of ELIP3, FCP2 and LHC7, specific low light response	KHGR	NQGR
LHC8	LHC light harvesting 18 KDa subunit	main antenna subunit, specific light response	KHGR	NQGR
LHC9	low iron LHC	strong response to low iron	KHGR	NQGR

… and differs from the NCBI reference.

Discussion

Impact of different stresses on the physiology of P. tricornutum

The relative fitness of species or populations is best reflected by their respective growth rate. This growth rate will be maximal, when biotic and abiotic factors influencing growth and reproduction are least unfavorable, i.e. optimal. In our experimental setup and for the chosen strain *P. tricornutum* CCAP1055/1 this is the case for the combination of an intermediate light regime and a sufficiently high, non-limiting nutrient and iron supply. While low light conditions have a limiting effect on photoautotrophic growth through the relative shortage of light energy input, all other treatments (high light, low iron, stationary phase) lead to an oversupply with light energy and cause an increased need for excess energy dissipation. Of these fundamentally different stress conditions the variation of ambient light levels appears to have the lowest impact on cellular growth. Increasing ambient light, though leading to adjustment of pigmentation in terms to optimize light utilization, do not impose major problems on the system, as seen from the uniformly high growth rates at non-limiting light conditions (at high iron, exponential phase). Conversely, shortage of iron heavily impacts *Phaeodactylum* growth, visible from the very low growth rates and the chlorotic phenotype. Accordingly, low iron may exert its limiting effect not only via direct physical limitation of the iron-rich photosynthetic machinery but also via the chlorotic pigment loss. This may secondarily induce light limitation in exponentially growing cells, ultimately resulting in low growth rates (low iron, exponential phase). In the context of the photosynthetic energy metabolism stationary phase can also be regarded a stress-like condition, because cellular pigmentation is retained and passively causes a constitutively high energy input to the photosynthetic system. This energy is not converted into growth, but needs to be dissipated. The cellular "energy dissipation mode" observed at stationary phase is reflected by maximal values for the de-epoxidation state of XC pigments.

In each case of enhanced "excitation pressure" on the system (high light, low iron, stationary phase) the dissipation of excess energy is assumed to occur at the level of the proteins that capture this energy, i.e. the complex FCP light antenna system serving the photosystems I and II and closely interacting with xanthophyll cycle components. Here we postulate that different FCPs may be involved in the fundamentally different processes of light harvesting and the two different flavours of energy dissipation, the rather constitutive and the regulated component.

Differential gene expression in response to light, low iron and stationary phase

Acclimation to different light levels can be achieved by balancing the expression of FCPs that are rather involved in light harvesting and FCPs that are rather involved in XC-related energy dissipation processes. We suggest FCPs that share a high expression level and are up-regulated upon low light, but down-regulated at high light, be part of the former group (ELIP3, FCPA, FCP2-6, FCPB, LHC), and FCPs, that share a low expression level and are down-regulated upon low light, but up-regulated at high light, be part of the latter group (ELIP1-2, FCPAQ, FCP1, CPQ). A strong and specific response to varying light combined with a high expression level is observed for LHC8 that we therefore regard as the main light harvesting subunit. Down-regulation upon low light acclimation is found e.g. for genes implied in iron-uptake mechanisms (ISIP1, FCR, FRE2) suggesting a decreased iron uptake-machinery be sufficient for iron uptake at a lower growth rate. Accordingly, genes involved in metabolic processes, whose activity is growth rate dependent, may show such a low light-dependent response.

Iron limitation has a more severe impact on the whole system, inducing a chlorotic cellular phenotype with lower biomass and lower overall transcript levels. This is best visible from the differential regulation (2.1) of the constitutively expressed RPB1 with respect to 18S. As a result we observe a virtual up-regulation for genes, whose cellular transcript levels relative to RPB1 haven't changed (e.g. FRE1). True up-regulation (per cell) is only observed, when gene regulation overtops the relative regulation of RPB1, as is the case for iron-regulated genes like ISIP1, ISIP3, FCR and FRE2. One of the two tested heatshock transcription factors HSF2 appears to be regulated by light and iron, while two small heatshock proteins HSP20, who are implied in protein repair/refolding, are specifically induced by iron. This heatshock response indicates an increase in non-functional/damaged cellular proteins upon low iron and the onset of repair mechanisms. Relative to RPB1 the whole photosynthetic machinery including the FCP antenna system is retrenched as expected from chlorotic cells. The differential regulation of ferredoxin PetF and flavodoxin FldA indicates the known substitution of the abundant iron-rich electron carrier ferredoxin by flavodoxin [28] with the effect of lowering the cellular iron quota. The FCP light antenna system shows more complex regulation patterns. Most prominent is the selective up-regulation of FCPs belonging to the CPQ and FCPAQ subgroups suggested to be implied in XC-related energy dissipation processes. The LI818/LHCSR ortholog CPQ4 is implied to be involved in NPQ at the site of photosystem II. The differential regulation of CPQ3 and CPQ4 suggests a possible mutual substitution. A specific induction by low

iron is also found for LHC3 and LHC9, suggesting major rearrangements occur in the main light harvesting complexes, too.

The gene expression at stationary phase is characterized by a large-scale decrease of transcript levels, a possible starvation effect observed upon long-term incubation of a chlorotic culture under high light and low iron conditions. Of interest is the specific induction of CPQ4 and four members of the FCPAQ subgroup, adding evidence to a proposed role in energy dissipation processes. Most importantly, we note that in the CPQ subgroup we do not observe a significant light-dependent regulation, while the response to low iron and stationary phase is mediated by the same gene (CPQ4). Conversely, the FCPAQ subgroup appears to originate from a radiation with subsequent specialization of its members to the different stresses high light, low iron and stationary phase. What the qualitative differences between these "stresses" are with respect to the underlying molecular processes of perception, transfer and dissipation of light energy, what the functional connection to the XC may be, and what the nature of the evolutionary selection pressure for the observed differentiation between FCPAQ genes may be, remain highly cryptic questions.

Evolution of FCPs

In expanding protein families novel functions emerge in a gradual process as a result of gene duplication and diversification. This implies that in phylogenetic clustering patterns we should find genes of a related function contained in the same cluster and *vice versa*. We use this simple rational objective for the presented categorization of FCPs and ELIPs. The high conservation of the ELIP1 and ELIP2 protein across phylogenetic classes and their tight co-regulation argue for a very basic function in photosynthesis predating the primary endosymbiosis event. The related ELIP3 protein, that is regarded to have originated from a gene fusion between ELIP1 and ELIP2, appears to be co-regulated with other major FCPs and shares the membrane topology and a relatively high expression level with them. We therefore consider the origin of the FCP family (and the CAB family as well) may be found at an ELIP3-like ancestor. FCPA-like FCPs are found in red algae and, therefore, their evolutionary emergence may predate the secondary endosymbiosis event leading to organisms with secondary plastids incl. diatoms. Additional support for this scenario comes from the co-purification of FCPA proteins with *Phaeodactylum* PS I. This indicates that these proteins are associated with PS I in diatoms as is the case for the homologs in red algae. Soon after the secondary endosymbiosis event an ancestral FCPAQ protein may have evolved from FCPA and subsequently have undergone radiation and diversification for the hypothetical purpose of regulating energy dissipation at the site of PS I. While the prospective light antennae

FCPA show strong expression, the rather enzymatic task of XC-related energy dissipation would require lower transcript/protein levels as found for FCPAQ. FCPB and LHC form a common cluster, and among the LHC proteins are the most abundant FCPs found in biochemical preparations. Therefore we suggest these constitute the main light harvesting trimers serving PS II. The phylogeny shows that the LHC proteins are derived from an FCPB-like ancestor and thus we postulate a role for FCPB proteins in mediating the interaction between LHC trimers and PS II, possibly as adapter-like components. For an unknown reason the LHC group underwent a major radiation upon evolutionary divergence of the radiate and pennate diatoms with an analogous event observed for the green lineage LHC proteins (CABs). Specific FCPB and LHC proteins have retained a relatively high degree of conservation and at the same time show high expression levels. We suggest that these now constitute the main light harvesting machinery at the site of PS II, possibly as inner (FCPB1, FCPB3, FCPB4) and trimeric accessory antennae (FCPB2, LHC5, LHC6, LHC8, LHC9). Notably phylogenetic branches inbetween these two groups share the same length suggesting an evolutionary freezing of FCP sequences upon establishment of a relevant antenna interaction. Green lineage orthologs of the CPQ proteins (green algal LI818 and LHCSR) have been found to be associated with the ability for NPQ at the site of PS II [13]. Consequentially, we suggest CPQ be involved in energy dissipation processes at PS II as well, while the extreme conservation of sequence implies an ancient and conserved functionality and, accordingly, a close association with the photosystem. However, the exceptional high expression of CPQ2 indicates that it may rather have co-evolved with LHC proteins for regulating energy dissipation processes in the abundant accessory antennae. The remaining FCPs, FCP1-6, are of uncertain evolutionary origin, but their expression and regulation suggests FCP1 be involved in energy dissipation processes (co-regulated with FCPAQ1) and FCP2-6 be ancient light antennae at PS II (phylogenetic position more close to FCPB). Though widely hypothetical, this evolutionary trajectory draws a consistent picture of the complete system with all mandatory functions of a light antenna system (light harvesting, energy dissipation) needed at any specific location (PS I, PS II, free FCP trimers) assigned to specific groups of FCPs. Notably, this scheme is supported by protein phylogeny, conserved motifs of helices 1 and 3, transcript expression levels, differential regulation, and by most available biochemical and molecular data.

The non-uniform phylogenetic situation observed for FCPs in the diatoms *Phaeodactylum* and *Thalassiosira* shows that species, though members of the same phylogenetic group, continuously proceed to change and adjust their antenna systems to their actual

needs, what implicitly must lead to vast differences in light harvesting and energy dissipation strategies between species. Here, it is of special interest that FCPs assigned to PS I (FCPA) are evolutionary relatively constant and invariable with many conserved ortholog pairs between *P. tricornutum* and *T. pseudonana* being observed. Conversely, evolution of the prospective PS II antenna system (light harvesting proteins FCP, FCPB, LHC; energy quenching proteins CPQ) appears to have occurred in an apparently convergent, but more dynamic fashion. Further this may indicate that differentiation of the PS I antennae was completed before divergence of central and pennate diatoms, while PS II may just have acquired a primitive progenitor system of PS II antennae that subsequently underwent expansion and optimization independently in both diatom lineages. Mutual horizontal gene transfers could have ensured evolutionary convergence of the PS II antenna evolution.

Quench it ! – Growing evidence for an evolutionary diversification of photosynthetic energy quenchers

The ability for excess energy dissipation is a fundamental property of all organisms capable of oxygenic photosynthesis. At the site of PS II, but not at PS I or in the accessory antenna fraction, we are able to detect the regulation of energy dissipation due to the intrinsic variable chlorophyll fluorescence of PS II. The regulated part of this PS II energy dissipation (NPQ, qE, Φ_{NPQ}) was found to be an energy (ΔpH) dependent process in plants and algae and has been correlated with the mutual conversion of XC pigments [29] and with the presence of stress responsive proteins of CPQ-type (LI818, LHCSR) or ELIP-type (PSBS). This correlation suggests NPQ might be mediated in charged thylakoid membranes by certain chlorophyll-binding proteins through their interaction with XC pigments. While the induction of NPQ implicitly has been assigned to the strong differential regulation of CPQ-like or ELIP-like "quencher" proteins, the constitutive component of PS II energy dissipation ($\Phi_{f,\,D}$) has not been linked to protein action. We here hypothesize that part of this constitutive component may also be mediated by FCP-like proteins (CPQ, ELIP), but remains invisible due to a lack of differential regulation of the respective proteins. Further, we cannot detect energy dissipation processes associated with PS I or the free FCP fraction, as these do not yield a variable fluorescence signal (with the potential exception of some PS I contribution to Fo fluorescence). Nevertheless it appears straight-forward to postulate energy dissipation action at these sites and an involvement of chlorophyll-binding proteins in these processes. From their expression level, differential stress response and phylogenetic background it is tempting to speculate that FCPAQ proteins may be involved at PS I-related energy dissipation, while specific CPQ proteins (CPQ2) may have evolved to act on the free

fraction of accessory PS II antennae that appear to have evolved more recently from FCPB-like proteins (see above). In the analogous *green* light-harvesting system the ability for self-quenching has been suggested for free LHCII of *Chlamydomonas* during state transitions [30].

The combined phylogenetic tree featuring *P. tricornutum* and *T. pseudonana* FCPs translates into a consistent evolutionary trajectory for diatom FCPs and ELIPs. Stress-responsive FCPs (of *Phaeodactylum*) are grouped into discrete clusters as a result of evolutionary diversification of FCPAQ (PS I) and CPQ (PS II). Conversely, proteins that are found to act in energy dissipation processes in *Chlamydomonas* (LHCSR) and *Arabidopsis* (PSBS) are not part of monophyletic protein family, but rather are descendents of a CPQ-like (LHCSR) or ELIP1-like (PSBS) precursor. Experimental evidence suggests these proteins confer the very same functionality of NPQ to PS II. With respect to the evolutionary process it appears that PSBS is an ancient feature that more recently was functionally replaced by a CPQ-like protein (LHCSR) in *Chlamydomonas*. While the multicellular plant *Arabidopsis* has an intrinsically reduced ability for lateral acquisition of new genes, the unicellular *Chlamydomonas* may have benefited from assimilation of external genetic material (LHCSR) in terms to further optimize the highly specialized and reduced photosynthetic machinery characteristic for the green photosynthetic lineage.

The observation that very different groups of chlorophyll-binding proteins are capable of conferring NPQ to PS II (CPQ-like LHCSR/LI818, ELIP1-like PSBS) raises the question for functional differences and possible interference at the site of PS II, as both types of proteins occur in diatoms (CPQ, ELIP1) and *Chlamydomonas* (LHCSR, PSBS) simultaneously. In both cases (diatoms and *Chlamydomonas*) NPQ has experimentally been assigned to the CPQ-like component (CPQ, LHCSR) rather than to the ELIP-like component (ELIP1, PSBS). Due to the lack of any CPQ-like genes green plants use PSBS instead (discussed in [31]).

This protein-based view of NPQ does not conflict with models that focus on the interconversion of XC pigments as the crucial event in dissipative processes, like the quencher-antiquencher concept of XC carotenoids [32]. Accordingly, interconversion would trigger loss of absorbance properties and dissociation of XC pigments resulting in a functional switch from energy perception to dissipation and/or associated conformational changes in the affected FCPs. In such a scenario, the primary FCPs interacting with XC carotenoids would be the stress-responsive minor FCPs (CPQ, FCPAQ, FCP1) as energetic sinks “on demand”. Conformational changes in these FCPs may subsequently be passed over to the larger antenna complexes consisting of major FCPs, thereby leading to higher order structural

rearrangements. In the analogous *green* antenna system a pH-inducible monomerization was found for dimeric Maize PSBS [33].

FCP dynamics - a mechanistic model of FCP interactions

The hypothetical functional assignments and assumptions for FCP interactions heavily depend on the validity of the underlying rational objectives. These may not apply in all cases, as we e.g. do not know, if expression on the protein level indeed does correspond to the transcript levels determined by RT-qPCR, though a strong transcriptional component of regulation was described for CABs [34]. Further, the suggested trimer formation might not apply for all FCPs. How FCP apoproteins obtain their pigments [35] and how single FCPs are assembled into trimers is unknown to date. In Figure 6 we present a scenario, where co-regulated FCPs may specifically be combined in one trimer, though the alternative of a random assembly of FCPs might apply just as well for main LHC trimers and other accessory antennae. In the former case the functional units would consist of well defined trimers of constant composition. In the latter case, FCPs would exert their functional effect through modification of their relative abundance in trimers. The same applies for the prospective "quencher" groups CPQ and FCPAQ. For this reason the overview in Figure 6 shall rather be regarded as a visualization of correlations between FCPs with respect to their expression levels, differential regulation, phylogenetic clustering and successful co-purification with specific photosystems [15].

From the group of FCPs with high expression levels we propose candidates for the inner antennae of PS I (FCPA) and PS II (FCPB). As possible accessory PS II antennae we distinguish main LHC trimers serving PS II (subunits LHC5, LHC8) from accessory PS II antennae of rather ancient (ELIP3, FCP2-5) or rather modern (FCPB5, LHC) evolutionary origin. While from co-purification experiments [15] we expect a defined architecture for the inner antennae of PS I (FCPA) and PS II (FCPB), this may not apply for accessory antennae. These might be recruited to the photosystems according to the actual needs of the system, though experimental evidence for a targeted energy supply directed to a specific photosystem (II or I) seams to be lacking [6].

Some diatom FCPs assignable to the accessory antenna fraction tend to assemble into higher oligomeric states [36]. Accordingly, the "heavy" multimeric and fucoxanthin-rich FCP conjugates that account for 75 % of all FCPs in biochemical preparations of *Phaeodactylum* [37] may be formed by the main LHC antennae as an external accessory complex in analogy to green LHC, whose mobility enables state transitions in plants and green algae. The candidate subunits for main diatom LHC (LHC5, LHC8) as proposed

from qPCR data have already been identified as major LHC subunits (syn. FCPC/D/E) by subjecting 18 and 19 KDa FCP fractions to mass spectrometry [27]. LHC5 (syn. FCPE) may serve as a basic trimer scaffold for other LHC subunits as well, e.g. for LHC7 (FCPA) [38].

FCPs and ELIPs with low expression levels are often stress-responsive and are assigned to the broad functional context of XC-related energy dissipation processes. NPQ was hypothesized to occur at distinct sites. For *Arabidopsis* the presence of two distinct PSBS- and XC-dependent quenching sites (inner PS II antennae, detached LHC trimers) was described [39] and multiple quenching sites have also been proposed for *Phaeodactylum* [6, 40]. As the two distinct quenching sites in *Phaeodactylum* that have experimental support we propose PS II and external FCP aggregates (accessory antenna fraction incl. higher LHC oligomers) that both may rely on specific CPQ-like proteins for NPQ. FCPAQ-like proteins and FCP1 may confer similar functionality to PS I, though experimental support for this is lacking so far.

Future outlook

Prerequisite for a concise functional characterization of specific FCPs and their involvement in ecologically relevant molecular processes like the perception and dissipation of light energy by diatoms is a comprehensive conceptional framework for the complete FCP antenna system as a working model to build on. The evolutionary trajectory for FCPs and the hypothetical model of FCP interactions presented in this work are meant to provide no less and no more than such a working model that is lacking to date. Assessing the dynamic interaction between ELIPs, FCPs, photosystems and XC pigments on a biophysical and biochemical basis remains the major challenge for future work. Given the wealth of information that is already available on the XC and FCP system of specific diatoms, the analysis of specific FCP knock-out strains represents a promising approach for extending research on diatom photosynthesis [41]. Although the specific FCP players involved in light harvesting and energy dissipation might vary between diatom species, and although differences in NPQ kinetics have been reported between diatoms [42], there is hope and evidence that underlying principles of photosynthetic light perception are more universal, allowing the transfer of most experimental findings to related evolutionary groups.

Conclusions

The *Phaeodactylum tricornutum* superfamily of FCP-like proteins ensures highly efficient photosynthesis through balancing light harvesting vs excess energy dissipation and through proper distribution of energy between the two photosystems.

The differential gene expression in response to light, low-iron and stationary phase conditions reveals distinct functional groups of FCPs that appear to be consistent with the observed phylogenetic clustering patterns. Here we distinguish seven major groups of *Phaeodactylum* FCP-like proteins: 3 ELIPs, 6 ancient FCPs, 4 PS II-type quenchers (CPQ), 5 PS II-type antennae (FCPB), 9 accessory antennae (LHC), 9 PS I-type antennae (FCPA) and 7 PS I-type quenchers (FCPAQ). A consistent evolutionary trajectory is presented for these groups. Our mechanistic model of FCP interactions proposes defined roles for most FCPs in terms of a basic distinction between FCPs acting as light harvesting proteins (high expression levels) and FCPs that may rather be involved in XC-related energy dissipation processes (low expression levels).

For the PSBS protein, shown to be required for NPQ in plants, we provide strong evidence to be directly derived from a duplication of an ELIP1 protein, the presumably most ancient of all chlorophyll-binding proteins (ELIPs, FCPs and CABs). This implies an independent evolution of PS II energy quenchers in the green and red evolutionary lineages on distinct, divergent routes. The *green* lineage evolved the PSBS protein directly from an ELIP1-like protein. Conversely, algae from the *red* lineage (e.g. diatoms) primarily rely on CPQ-like proteins that more recently may have been transferred horizontally to green algae to functionally replace a PSBS-like element. Diatoms evolved an additional group FCPAQ of unknown affiliation, though phylogenetic patterns suggest they be related to PS I antenna proteins (FCPA) and thus might serve for dissipating excess energy at the site of PS I.

Methods

Culturing, microscopy and PhytoPAM

Phaeodactylum tricornutum Bohlin CCAP1055/1 [3, 5] was grown as axenic clonal culture in 1 l batch cultures using iron-free f/2 nutrients [43] in ASW (artificial seawater medium [44]) supplied with 10 µM $FeCl_3$ (high iron) or 5 nM $FeCl_3$ (low iron) at 50 µE (low light LL), 150 µE (normal light NL) or 450 µE (high light HL), at 20°C and a 16/8 h light/dark cycle. For RNA extraction cells were harvested at maximal growth rate (late exponential growth phase) or after 1 week at stationary phase (late stationary growth phase) by centrifugation in 35 ml tubes at 4°C for 5 min at 11000 rpm, resuspension into a small volume of media, and a second centrifugation step in 1.5 ml cryo-tubes at 4°C for 1 min at 13000 rpm. Cell pellets were frozen in liquid N_2 and stored at -80°C.

For iron-limited cultures iron-free techniques were applied as follows. Culture bottles were composed of plastic material, washed and incubated for some days with 1 N HCl and rinsed with ultrapure

MilliQ water. All additional supply for iron-free work was washed with 1 N HCl and stored in closure bags until use.

P. tricornutum was imaged *in vivo* in an Utermöhl chamber using an Axiovert 100 inverse microscope (Zeiss) with a Neofluar 100/1.30 oil immersion objective (Zeiss). Images were captured with a Canon PowerShot S50 digital camera.

The variable chlorophyll fluorescence F_v/F_m of *P. tricornutum* cells from fresh cultures was measured with a PhytoPAM (PHYTO-PAM Phytoplankton Analyzer, Heinz Walz GMBH, Effeltrich, Germany) [45] upon 5 minutes of dark incubation. Light curves were conducted in steps of increasing actinic background illumination with actinic light increasing from 1 to 907 µE as follows: At each step chlorophyll fluorescence was measured along intervals of 5 min actinic background light followed by a 4000 µE red (655 nm) saturation pulse (→ F_q', F_m') and 2 min dark (→ F_o'). Actinic light was white (470, 520, 645 and 665 nm) for intensities between 1 and 44 µE and red (655 nm) for intensities between 54 and 907 µE. To avoid negative values for resulting quenching parameters, values for F_o and F_m were taken from the absolute minimum and maximum fluorescence observed during the entire light curve, instead of those observed upon dark acclimation. F_o' was calculated as outlined in [46] and matched the values observed during the dark intervals. As the change of light quality at the transition from 44 to 54 µE actinic background light introduced a discontinuity in chlorophyll fluorescence values, we only use data obtained with red actinic light for assessing the photosynthetic efficiency of *Phaeodactylum*. The relative Electron Transport Rate shown in Figure 1 was calculated as rETR = (F_q'/F_m') * (actinic light intensity [µE]). The cellular light utililization (fractionation of absorbed light into $\Phi_{PS\ II}$, $\Phi_{f,\ D}$, Φ_{NPQ}) presented in Additional file 1 (Figure S2) was calculated as outlined in [20]. Quantum efficiencies of PS II as measured along the light curve were integrated for each sample between 54 and 802 µE after the trapezoidal rule to obtain a representative value for the gross photosynthetic efficiency of *P. tricornutum*. The dependence of PS II quantum efficiencies on the different types of stress was visualized through Box-Whisker-Plots (Additional file 1 - Figure S3). Box-Whisker-Plots and the interaction plot in Additional file 1 (Figure S4) were constructed using R [47].

HPLC

Dependent on the individual chlorophyll content 10-50 ml of culture were filtered on a 25 mm GF/F glass fibre filter (Whatman). Filters were stored at -80 °C. Pigments were extracted in 5 ml 90% acetone during a 5 min disruption step with glass beads in a cell mill. Debris and beads were separated from the homogenate by 10 min centrifugation at 0 °C and 5500 rpm. 2 ml of supernatant were passed through a syringe with 0.2 µm PTFE prefilter and 100 µl of a

canthaxanthin standard were added to each sample. 110 µl of sample were transferred into an HPLC vial and subjected to reverse phase HPLC modified after Barlow et al. [48]. Measurements were conducted by a Waters 600 Controller in combination with a Waters 996 PDA Detector and a Waters 717 Plus Autosampler.

RNA extraction and RT-qPCR

Sets of primers were designed with Primer Express Software to detect gene specific amplicons of approx. 100 bp for all tested genes (Additional file 2 - Table S1). Specifically, primers for FCP genes were designed in the most variable transcript region corresponding to the central protein region between the first and third transmembrane domain. Moreover, primers were not designed to cross exon-intron borders, so that primer-specific amplification efficiency could be assessed in a comparative qPCR approach on genomic DNA by testing three different primer pairs on the same amount of sheered gDNA template. The primer pair showing the lowest C_T value, indicating the highest amplification efficiency, was selected as optimized gene-specific probe for RT-qPCR. List of primers and transcripts, results from a specificity BLASTN search of primers against *P. tricornutum* genome and C_T values from the comparative qPCR test are provided as additional files (Additional files 3-6 - Files S01, S02, S03 and Table S2).

Total RNA was extracted from frozen pellets of *P. tricornutum* cells using the QIAGEN RNeasy kit. RNA quality was assessed from NanoDrop UV absorption profiles. cDNA template was prepared from 1µg RNA by reverse transcription using the QuantiTect Rev. Transcription Kit (QIAGEN) after digestion of residual DNA using the included gDNA wipeout reagents. The purified RNA was tested for possible residual DNA by qPCR using primers specific for the 18S-rRNA gene. For qPCR the cDNA template was diluted down to 0.5 ng μl^{-1} with 1x TE buffer and 5 µl equaling 2.5 ng cDNA were used as template per 25 µl qPCR reaction. qPCR was done on an ABI Prism 7000 (Applied Biosystems) qPCR cycler. Cycling conditions were 2 min at 50°C (once), 2 min at 95°C (once), and 40 cycles of 95°C for 0:15 min, followed by 0:30 min at 60°C. The qPCR mixtures contained 12.5 µl SYBR qPCR SUPERMIX W/ROX (Invitrogen), 0.5 µl of 10 µM forward and reverse primer each, 6.5 µl H_2O and 5 µl of dilute cDNA template. C_T values were calculated from SDS raw data using auto-baseline setting and a manual fluorescence threshold of 0.2.

Due to strong machine-specific systematic biases towards higher C_T values, observed at the edges of qPCR plates and indicative of lower amplification efficiencies in these wells, upper and lower plate edges (rows A and H) were excluded from qPCR and each reaction was run as four replicates in a blockwise 2x2 well arrangement. Remaining biased outliers were efficiently removed by using only

the lowest C_T values from the four technical replicates for calculating a representative mean C_T for the respective reaction. Data resulting from this analysis (MAN) is compared in Additional file 7 (Table S3) to alternative analysis strategies of taking the minimal (MIN) or mean (MEAN) C_T value of all technical replicates and reveals minor differences between MAN and MIN, but strong biases in the MEAN data set due to the positional biases mentioned above, e.g. for the rightmost sample (HL, low iron, stat. phase).

Relative expression of genes with respect to the 18S-rRNA gene was determined using the ΔC_T method (ΔC_T[geneX] = C_T[geneX] - C_T[18S]). Two biological replicates were combined as mean ΔC_T ± s.d. (or s.e.m. respectively). Resolution of qPCR data on the level of biological replication was substantially improved by applying a normalization step that follows the rational objective that overall transcriptional regulation of the entirety of an organism's genes should show minimal variation in two biological replicates, when grown under identical conditions. Therefore, we adjusted the C_T value for 18S in one of the two biological replicates, so that the sum of ΔC_T errors (s.d.) over all tested genes was minimal between the two biological replicates. By this we account for the variability of the 18S transcription and avoid that this variability carries over to all target genes. Resulting errors between biological replicates are very low for most genes confirming a strict transcriptional regulation on system level. Final ΔC_T data is presented as ΔC_T[inverted] = 21 - ΔC_T to reveal relative expression levels on an arbitrary positive $\log_2$ scale, where high values indicate high cellular transcript levels (Figure 2). Differential regulation of transcription between two samples is given as $\Delta\Delta C_T$ = ΔC_T[sample] - ΔC_T[control].

A comparison between FCP expression levels as determined by qPCR and relative EST abundances in existing *Phaeodactylum* EST libraries (integrated in JGI *P. tricornutum* genome browser at http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) reveals that the sensitivity of qPCR resolves even the low expression levels of minor FCPs (e.g. FCPAQ) or enzymes like PEPCK, while even the largest EST library ("standard") is merely able to resolve the high expression levels of major FCPs (Additional file 8 - Table S4). Implicitly all other EST libraries sequenced for the purpose of characterizing differential transcriptional responses to various stresses and treatments rather represent statistical noise, as they cannot even resolve the high expression levels of major FCPs.

Sequence retrieval, BLAST searches and phylogenetic analysis

Protein and transcript sequences were retrieved from JGI genome browser interfaces for *P. tricornutum* [49] and *T. pseudonana* [50] or from NCBI nr protein database. In case of gene model inaccuracies improved versions were derived manually from

alternative gene predictions as indicated ("modified"). Official NCBI identifiers for all protein sequences are provided in Additional file 9 (File S04) except for those subjected to manual correction ("modified"), where the identifier for the closest relative protein model in NCBI public database is provided instead.

For manual improvement of gene models and manual annotation we made use of the GENSCAN webserver [51], InterproScan [52] and diverse bioinformatics tools like TMHMM (transmembrane regions), SIGNALP (signal peptides or anchors) and TARGETP (targeting to cellular compartments) [53]. Secondary structure □-helices of ELIP2 and ELIP3 (Figure 4) were identified by PSIPRED [54]. All BLAST [55] searches were done using the local BLAST package from NCBI. BioEdit [56] was used for manual sequence manipulation and visual inspection of sequences and sequence alignments.

For construction of phylogenetic trees protein sequences were aligned with CLUSTAL W v1.83 [57]. Resulting alignments were trimmed for hypervariable regions (processed signal peptide, variable loops) and converted to Phylip 4 format. Neighbour joining trees were constructed with TREECON v1.3b [58] using midpoint rooting, so that major clusters of sequences assigned to photosystem I ("A") or photosystem II ("B") were separated best possible and most branches showed similar length, thereby reflecting an even evolutionary progression for most proteins (Additional files 10-13 - Files S05-S08). For high-resolution subtrees presented in Figures 3 C and D alignments of CPQ or FCPB/LHC were further trimmed to retain solely residues of homologous origin (Additional files 14-15 - Files S09-S10). Outgroups for CPQ and FCPB/LHC trees were LHC and CPQ sequences, respectively.

Authors' contributions

ML designed the study, performed PhytoPAM measurements, optimized and carried out the RT-qPCR, performed phylogenetic analyses, developed the evolutionary trajectory and interaction model for *Phaeodactylum* FCPs, prepared the figures and drafted the manuscript. TK cultured and harvested the algae, performed physiological analyses and processed the HPLC data. JLR inspired and coordinated the study, contributed to experimental design and manuscript writing. All authors read and approved the final manuscript.

Acknowledgements

We thank Diana Gill (Helmholtz Centre for Ocean Research Kiel (GEOMAR), Kiel, Germany) for performing extensive RT-qPCR work during the preparatory experimental phase. The contribution of

Phaeodactylum microscopy images by Michaela Günther and Jaqueline Heinrich is kindly acknowledged. Alexandra-Sophie Roy gave helpful feedback to experimental issues and commented on the manuscript. This work was supported by a DFG grant to JLR (RO2138/6-1).

References

1. Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B: **Production and dissolution of biogenic silica in the ocean: Revised global estimates, comparison with regional data and relationship to biogenic sedimentation.** *Global Biogeochem Cycles* 1995, **9**:359-372.
2. Grzebyk D, Schofield O, Vetriani C, Falkowski PG: **The mesozoic radiation of eukaryotic algae: The portable plastid hypothesis.** *J Phycol* 2003, **39**:259-267.
3. Bohlin K: **Zur Morphologie und Biologie einzelliger Algen.** *Öfvers af K Vet Acad Förhandl Stockholm* 1897, **54**:519-522.
4. Brand LE, Sunda WG, Guillard RRL: **Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron.** *Limnol Oceanogr* 1983, **28**:1182-1198.
5. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, et al.: **The Phaeodactylum genome reveals the evolutionary history of diatom genomes.** *Nature* 2008, **456**:239-244.
6. Owens TG: **Light-Harvesting Function in the Diatom Phaeodactylum tricornutum (II. Distribution of Excitation Energy between the Photosystems).** *Plant Physiol* 1986, **80**:739-746.
7. Elrad D, Grossman AR: **A genome's-eye view of the light-harvesting polypeptides of *Chlamydomonas reinhardtii*.** *Curr Genet* 2004, **45**:61-75.
8. Green BR, Durnford DG: **The chlorophyll-carotenoid proteins of oxygenic photosynthesis.** *Annu Rev Plant Physiol Plant Mol Biol* 1996, **47**:685-714.
9. Müller P, Li XP, Niyogi KK: **Non-photochemical quenching. A response to excess light energy.** *Plant Physiol* 2001, **125**:1558-1566.
10. Goss R, Jakob T: Regulation and function of xanthophyll cycle-dependent photoprotection in algae. *Photosynth Res* 2010, **106**:103-122.
11. Moseley JL, Allinger T, Herzog S, Hoerth P, Wehinger E, Merchant S, Hippler M: **Adaptation to Fe-deficiency requires remodeling of the photosynthetic apparatus.** *EMBO J* 2002, **21**:6709-6720.
12. Li XP, Phippard A, Pasari J, Niyogi KK: **Structure-function analysis of photosystem II subunit S (PsbS) in vivo.** *Funct Plant Biol* 2002, **29**:1131-1139.

13. Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, Hippler M, Niyogi KK: **An ancient light-harvesting protein is critical for the regulation of algal photosynthesis.** *Nature* 2009, **462**:518-522.
14. Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM: **An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom Phaeodactylum tricornutum.** *PLoS One* 2009, **4**:e7743.
15. Lepetit B, Volke D, Gilbert M, Wilhelm C, Goss R: **Evidence for the Existence of One Antenna-Associated, Lipid-Dissolved and Two Protein-Bound Pools of Diadinoxanthin Cycle Pigments in Diatoms.** *Plant Physiol* 2010, **154**:1905-1920.
16. Dittami SM, Michel G, Collen J, Boyen C, Tonon T: **Chlorophyll-binding proteins revisited - a multigenic family of light-harvesting and stress proteins from a brown algal perspective.** *BMC Evol Biol* 2010, **10**:365.
17. Neilson JAD, Durnford DG: **Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes.** *Photosynth Res* 2010, **106**:57-71.
18. Hoffman GE, Puerta MVS, Delwiche CF: **Evolution of light-harvesting complex proteins from Chl c-containing algae.** *BMC Evol Biol* 2011, **11**:101.
19. Engelken J, Brinkmann H, Adamska I: **Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily.** *BMC Evol Biol* 2010, **10**:233.
20. Hendrickson L, Furbank RT, Chow WS: **A simple alternative approach to assessing the fate of absorbed light energy using chlorophyll fluorescence.** *Photosynth Res* 2004, **82**:73-81.
21. Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C: **Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation.** *PNAS* 2008, **105**:10438-10443.
22. Kilian O, Kroth PG: **Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids.** *Plant J* 2005, **41**:175-183.
23. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane

TW, Larimer FW, Lippmeier JC, Lucas S, et al.: **The genome of the diatom *Thalassiosira* pseudonana: Ecology, evolution, and metabolism.** *Science* 2004, **306**:79-86.
24. Amunts A, Nelson N: **Functional organization of a plant photosystem I: Evolution of a highly efficient photochemical machine.** *Plant Physiol Biochem* 2008, **46**:228-237.
25. Chitnis VP, Chitnis PR: **PsaL subunit is required for the formation of photosystem I trimers in the cyanobacterium Synechocystis sp. PCC 6803.** *FEBS Lett* 1993, **336**:330-334.
26. Ikeda Y, Komura M, Watanabe M, Minami C, Koike H, Itoh S, Kashino Y, Satoh K: **Photosystem I complexes associated with fucoxanthin-chlorophyll-binding proteins from a marine centric diatom, *Chaetoceros gracilis*.** *Biochim Biophys Acta Bioenerg* 2008, **1777**:351-361.
27. Lepetit B, Volke D, Szabo M, Hoffmann R, Garab GZ, Wilhelm C, Goss R: **Spectroscopic and molecular characterization of the oligomeric antenna of the diatom *Phaeodactylum tricornutum*.** *Biochemistry* 2007, **46**:9813-9822.
28. LaRoche J, Boyd PW, McKay RML, Geider RJ: **Flavodoxin as an in situ marker for iron stress in phytoplankton.** *Nature* 1996, **382**:802-805.
29. Bilger W, Björkman O: **Role of the xanthophyll cycle in photoprotection elucidated by measurements of light-induced absorbance changes, fluorescence and photosynthesis in leaves of *Hedera canariensis*.** *Photosynth Res* 1990, **25**:173-185.
30. Iwai M, Yokono M, Inada N, Minagawa J: **Live-cell imaging of photosystem II antenna dissociation during state transitions.** *PNAS* 2010, **107**:2337-2342.
31. Barros T, Kühlbrandt W: **Crystallisation, structure and function of plant light-harvesting Complex II.** *Biochim Biophys Acta Bioenerg* 2009, **1787**:753-772.
32. Frank HA, Cua A, Chynwat V, Young A, Gosztola D, Wasielewski MR: **The lifetimes and energies of the first excited singlet states of diadinoxanthin and diatoxanthin: The role of these molecules in excess energy dissipation in algae.** *Biochim Biophys Acta Bioenerg* 1996, **1277**:243-252.
33. Bergantino E, Segalla A, Brunetta A, Teardo E, Rigoni F, Giacometti GM, Szabo I: **Light- and pH-dependent structural changes in the PsbS subunit of photosystem II.** *PNAS* 2003, **100**:15265-15270.
34. Escoubas JM, Lomas M, LaRoche J, Falkowski PG: **Light-Intensity Regulation of CAB Gene-Transcription is**

Signaled by the Redox State of the Plastoquinone Pool. *PNAS* 1995, **92**:10237-10241.
35. Hoober JK, Eggink LL, Chen M: **Chlorophylls, ligands and assembly of light-harvesting complexes in chloroplasts.** *Photosynth Res* 2007, **94**:387-400.
36. Büchel C: **Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19 kDa subunits assemble into different oligomeric states.** *Biochemistry* 2003, **42**:13027-13034.
37. Guglielmi G, Lavaud J, Rousseau B, Etienne AL, Houmard J, Ruban AV: **The light-harvesting antenna of the diatom *Phaeodactylum tricornutum* - Evidence for a diadinoxanthin-binding subcomplex.** *FEBS J* 2005, **272**:4339-4348.
38. Joshi-Deo J, Schmidt M, Gruber A, Weisheit W, Mittag M, Kroth PG, Büchel C: **Characterization of a trimeric light-harvesting complex in the diatom *Phaeodactylum tricornutum* built of FcpA and FcpE proteins.** *J Exp Bot* 2010, **61**:3079-3087.
39. Holzwarth AR, Miloslavina Y, Nilkens M, Jahns P: **Identification of two quenching sites active in the regulation of photosynthetic light-harvesting studied by time-resolved fluorescence.** *Chem Phys Lett* 2009, **483**:262-267.
40. Miloslavina Y, Grouneva I, Lambrev PH, Lepetit B, Goss R, Wilhelm C, Holzwarth AR: **Ultrafast fluorescence study on the location and mechanism of non-photochemical quenching in diatoms.** *Biochim Biophys Acta Bioenerg* 2009, **1787**:1189-1197.
41. Bailleul B, Rogato A, de Martino A, Coesel S, Cardol P, Bowler C, Falciatore A, Finazzi G: **An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light.** *PNAS* 2010, **107**:18214-18219.
42. Grouneva I, Jakob T, Wilhelm C, Goss R: **A new multicomponent NPQ mechanism in the diatom *Cyclotella meneghiniana*.** *Plant Cell Physiol* 2008, **49**:1217-1225.
43. Guillard RRL, Ryther JH: **Studies of marine planktonic diatoms. I. Cyclotella nana Hustedt and Detonula confervacea Cleve.** *Can J Microbiol* 1962, **8**:229-239.
44. Goldman JC, McCarthy JJ: **Steady-State Growth and Ammonium Uptake of a Fast-Growing Marine Diatom.** *Limnol Oceanogr* 1978, **23**:695-703.
45. Schreiber U: **Chlorophyll fluorescence: new instruments for special applications.** *In: Garab G. (ed) "Photosynthesis: Mechanisms and Effects. Vol. V", Kluwer Academic Publishers, Dordrecht* 1998, pp. 4253-4258.

46. Baker NR: **Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo.** *Annu Rev Plant Biol* 2008, **59**:89-113.
47. R Foundation for Statistical Computing: *R: A language and environment for statistical computing.* Vienna; 2008.
48. Barlow RG, Cummings DG, Gibb SW: **Improved resolution of mono- and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC.** *Mar Ecol: Prog Ser* 1997, **161**:303-307.
49. **JGI Genome Portal for *Phaeodactylum tricornutum* v2.0** [http://genome.jgi-psf.org/Phatr2/Phatr2.home.html]
50. **JGI Genome Portal for *Thalassiosira pseudonana*** [http://genome.jgi-psf.org/Thaps3/Thaps3.home.html]
51. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA.** *J Mol Biol* 1997, **268**:78-94.
52. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, et al.: **InterPro: the integrative protein signature database.** *Nucleic Acids Res* 2009, **37**:D211-D215.
53. **CBS (Center for Biological Sequence analysis) Prediction Servers** [http://www.cbs.dtu.dk/services/]
54. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT: **Protein annotation and modelling servers at University College London.** *Nucl Acids Res* 2010, **38**(Suppl):W563-W568.
55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic Local Alignment Search Tool.** *J Mol Biol* 1990, **215**:403-410.
56. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser* 1999, **41**:95-98.
57. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucl Acids Res* 1994, **22**:4673-4680.
58. Van de Peer Y, De Wachter R: **TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment.** *Comput Appl Biosci* 1994, 10:569-570.

4 Supplemental Data (DVD Content)

4.4 T. oceanica Genome Paper

Additional file 1: Table S1

AUGUSTUS models vs nr+CDD.xls

Additional file 2: Phylogenetic Trees for LGT Candidate Genes

Phylogenetic_trees.tre

Additional file 3: Supplemental Methods and Figures

Supplemental methods and figures.pdf

Additional file 4: Supplemental Sequences S1

Protein_user_models.fas

Additional file 5: Table S2

Transcriptome manual annotation.xls

Additional file 6: Supplemental Sequences S2

Selection.fas

Additional file 7: Table S3

Low-iron responsive genes.xls

Additional file 8: Table S4

Iron regulated genes.xls

Additional file 9: Table S5

Transcriptome mapping.xls

Additional file 10: Table S6

RT-qPCR primers.xls

Additional file 11: Overview of Genes and Abbreviations

Genes and abbreviations.xls

4.5 P. tricornutum FCP Paper

Additional file 1: Supplemental Figures

Supplemental figures S1-S7.pdf

Additional file 2: Table S1

RT-qPCR genes and primers.xls

Additional file 3: File S1

Pt_qPCR_primers_F+R.fas

Additional file 4: File S2

Pt qPCR primers F+R vs Pt genome blastn10.html

Additional file 5: File S3

Pt_transcripts.fas

Additional file 6: Table S2

Relative primer efficiencies.xls

Additional file 7: Table S3

Alternative RT-qPCR Evaluations (MAN vs MIN vs MEAN).xls

Additional file 8: Table S4

FCP expression qPCR vs ESTs.xls

Additional file 9: File S4

FCPs+CABs+ELIPs_incl_IDs.fas

Additional file 10: File S5

Pt_FCPs.phy

Additional file 11: File S6

Pt+Tp_FCPs.phy

4 Supplemental Data (DVD Content)

Additional file 12: File S7

At+Cr_CABs.phy

Additional file 13: File S8

Pt+Tp+Gt_ELIPs.phy

Additional file 14: File S9

Pt_CPQs_(high_resolution).phy

Additional file 15: File S10

Pt+Tp_FCPBs_(high_resolution).phy

Declaration

Hiermit erkläre ich, daß diese Abhandlung unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden ist. Sie ist nach Inhalt und Form meine eigene Arbeit. Alle wörtlich oder inhaltlich übernommenen Stellen sind als solche gekennzeichnet.

Ich versichere außerdem, daß ich diese Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe, und daß diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Kiel, 31. Mai 2012	Markus Lommer
Ort, Datum	Unterschrift

Acknowledgements

Finishing a project like this PhD is not possible without the help and support of many people. Throughout my scientific career I came across dedicated teachers and professors, whose prime aim it was to carry forward their fascination for the mysteries of life.

Accordingly, I am grateful to Mr Kutschrad and Mrs Neuber for motivating me to study life sciences. At Kiel university the botany courses of Prof. Jörg Sauter, Prof. Uhlarz and Prof. Usinger were most enjoyable. During my diploma thesis on custom-made microarrays I had a fantastic and very productive time in the working group of Prof. Schulz. A great many thanks go to Jens Appel for his scientific enthusiasm and for preparing me best possible for the challenges that were to come thereafter. I'm looking forward to our next canoe ride !

Starting my PhD at Julie's group certainly was the beginning of a new scientific era for me. I appreciate a lot to have been supplied with some of the biggest scientific challenges that I can think of – e. g. interpretation of large-scale sequence data and experimental optimization of RT-qPCR. This finally led me to a point of biological understanding that I have always been aiming for. Given that scientific efficiency may be defined as scientific progress per time, I still think my PhD was efficient, because the scientific achievements therein are more than sufficient to balance the time needed ☺ ! I am grateful to Julie for the extensive funding that enabled me to follow my projects towards a very good end. Working with her has always been fun, while at the same time I benefited a lot from her scientific intuition and experience. Julie also organized the contacts to my co-authors that I had an enjoyable and productive time with. In this context I would like to send sincere greetings to Reidar, Micha and Magda. Throughout my PhD I felt myself embedded in a very professional scientific environment, and I am thankful to have Philip Rosenstiel as second supervisor and Michael Hippler as co-author. The little diatom *P. tricornutum* is thanked for having a small-sized genome and for being such a grateful organism to work with.

Special thanks go to Tania for the fantastic scientific support with all the diatom cultures. The teamwork with Sophie was just very enjoyable, and I benefited a lot from her critical feedback to some problematic qPCR issues. Finally, I believe that some of my best ideas emerged from the unconscious incubation of data during leisure time. Accordingly, my thanks also go to the private side of my life: all my Badminton folks, my fellow lodgers, "postman" Sven Rogalski, the amazing Tori Amos, my coffee sponsor Frank Höpfner and my wonderful parents. Thank you all !