Characterization of Iron Acquisition and Iron Storage Mechanisms in the diatom Phaeodactylum tricornutum

Xia Gao

▶ **To cite this version:**

Xia Gao. Characterization of Iron Acquisition and Iron Storage Mechanisms in the diatom Phaeodactylum tricornutum. Subcellular Processes [q-bio.SC]. Université Paris-Saclay, 2021. English. NNT : 2021UPASB037 . tel-03741828

HAL Id: tel-03741828
https://theses.hal.science/tel-03741828

Submitted on 2 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université PARIS-SACLAY

Characterization of Iron Acquisition and Iron Storage Mechanisms in the diatom *Phaeodactylum tricornutum*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°567 : sciences du végétal : du gène à l'écosystème (SDV)
Spécialité de doctorat: Biologie
Unité de recherche : Institut de Biologie de l'École Normale Supérieure
ENS (IBENS), 75005, Paris, France
Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris, le 2 Août 2021, par

Xia GAO

Composition du Jury

Jacqui Shykoff Directrice de Recherche CNRS, Institut Diversité, Ecologie et Evolution du Vivant (IDEEV), UMR 8079	Présidente
Sébastien Thomine Directeur de Recherche CNRS, Institut de Biologie Intégrative de la Cellule de Paris Saclay (I2BC), UMR 9198	Examinateur
Angela Falciatore Directrice de Recherche CNRS, Institut de Biologie Physico-Chimique (IBPC), UMR 7238	Examinatrice
Giovanni Finazzi Directeur de Recherche CNRS, Laboratoire de Physiologie Cellulaire et Végétale de Grenoble (LPCV), UMR 5168	Rapporteur
Brigitte Gontero-Meunier Directrice de Recherche CNRS, Bioénergétique et Ingénierie des Protéines (BIP), UMR 7281	Rapportrice
Chris Bowler Directeur de Recherche CNRS, Institut de Biologie de l'École Normale Supérieure (IBENS), UMR 8197	Directeur de thèse

Thèse de doctorat

NNT : 2021UPASB037

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences du végétal : du gène à l'écosystème (SdV)

Titre : Caractérisation des mécanismes d'acquisition et de stockage du fer chez les diatomées *Phaeodactylum tricornutum*

Mots clés : diatomées, physiologie du fer, stockage du fer, absorption du fer, ferritine, ISIP

Résumé : Les diatomées réussissent à survivre dans les océans contemporains où la teneur en fer est faible. Malgré leur taille cellulaire relativement grande, les diatomées tolèrent une carence en fer et sont généralement dominantes au cours des « bloom » de phytoplancton, stimulées par le fer de manière naturelles ou artificielles. Cependant, notre compréhension de la physiologie du fer chez les diatomées reste limitée et fait l'objet de recherches en cours. L'objectif du travail présenté est d'étudier le métabolisme du fer, en examinant en premier lieu les principales stratégies de son utilisation chez les diatomées. Cette partie s'appuie sur la littérature publiée et sur la complétion des données par une analyse de 82 transcriptomes. Par la suite, la génération des lignées fluorescentes transgéniques de la ferritine (FTN) et « Iron starvation induced protein 3 » (ISIP3), ont permis d'obtenir la localisation subcellulaire de ces protéines pour la première fois. Il a été montré que la protéine FTN se localise au niveau du chloroplaste, alors que la protéine ISIP3 se localise sous forme d'agrégats à proximité du chloroplaste. ISIP3 présente ainsi une localisation similaire avec la protéine ISIP1. En outre, les phénotypes des mutants de délétion de FTN et ISIP3 cultivés sous différentes conditions en fer ont été comparés. Les mutants FTN ont montré une croissance réduite dans toutes les conditions, indiquant une fonction dans le stockage du fer. Les mutants ISIP3, quant à eux, ont montré une croissance réduite lorsque le fer est disponible, mais aucune différence lorsqu'ils sont cultivés dans un milieu sans fer, suggérant ainsi que ISIP3 est nécessaire dans l'acquisition du fer. Lorsque le cuivre a été retiré du milieu, ces phénotypes ont disparu, confirmant ainsi la dépendance au cuivre pour l'absorption du fer. La diminution de la croissance des mutants ISIP3 en présence de sidérophore (FOB) suggère que ISIP3 est également impliqué dans leur absorption. Dans l'ensemble, les résultats présentés dans cette thèse ajoutent de nouvelles perspectives dans l'étude du métabolisme du fer chez les diatomées.

Title: Characterization of Iron Acquisition and Iron Storage Mechanisms in the diatom *Phaeodactylum tricornutum*

Keywords: diatoms, iron physiology, iron storage, iron uptake, ferritin, ISIPs

Abstract: Diatoms succeed to survive the low iron conditions of the contemporary ocean. Despite their relatively large cell sizes, diatoms tolerate iron limitation and frequently dominate iron-stimulated phytoplankton blooms, both natural and artificial. However, our understanding of diatom iron physiology remains limited and is the subject of ongoing research. The objective of the present work was to investigate iron metabolism in diatoms, first, by reviewing the main iron use strategies of diatoms, relying on published literature and complementing published data with a search of 82 diatom transcriptomes. Then, through transgenic fluorescent lines, the subcellular localizations of ferritin (FTN) and iron starvation induced protein 3 (ISIP3) were Furthermore, phenotypes of FTN and ISIP3 knockout (KO) mutants grown in different conditions of iron availability and intracellular iron status were compared. The FTN-KO mutants showed decreased growth in all conditions, indicating a potential function in iron storage. The ISIP3-KO mutants showed decreased growth when iron was available and no difference when grown in iron-free medium compared with control cells, suggesting that ISIP3 was required for iron acquisition. When copper was removed from the medium, these phenotypes disappeared, confirming the copper-dependency of iron uptake. Decreased growth of iron-repleted ISIP3-KO mutants supplied with siderophore (FOB) further

visualized for the first time. It was found that FTN target to the chloroplast, whereas ISIP3 localizes to the vicinity of the chloroplast and aggregates close by exhibiting similar localization to the protein ISIP1. suggested an involvement in siderophore uptake. Overall, the results presented in this thesis have added novel insights to the study of iron metabolism in diatoms.

Acknowledgements

SPONSORS

Chinese Scholarship Council

SUPERVISORS

Chris Bowler, Elena Kazamia

COLLEAGUES

Richard Dorrell, Juan Pierella Karlusich, Nathalie Joli, Priscillia Pierre-Elies,

Maria Helena Cruz de Carvalho, Florent Charton, Shun Liu, Juliette Laude,

Khadeeja Adam, Justine Toinon, Magali Charvin

and all colleagues in IBENS that helped

COLLABORATORS

Angela Falciatore, Marianne Jaubert, Jean Pierre Bouly (IBPC)

Yannick Schwab, Karel Mocaer (EMBL, Heidelberg)

Robert Sutak (Charles University, Prague)

This thesis offered me a great opportunity to learn molecular biology and bioinformatics, which I consider as powerful tools for studying plant science that I will keep learning with enthusiasm in the future. For both scientific and personal experiences that I had in these years of research, I have built my strength, confidence, and knowledge for exploring the world of science. Now it is time to thank the persons that contributed to the accomplishment of my thesis.

First of all, I would like to express my gratitude to the Chinese Scholarship Council for the scholarship and University Paris-Saclay for offering the opportunity to perform research leading to a PhD.

I wish to pay my deepest regards to my supervisors Dr. Chris Bowler and Dr. Elena Kazamia for their thoughtful comments and recommendations on this thesis. It is wholeheartedly appreciated that your great advice for my study paved monumental for the success of this study.

I want to thank all our team members, Dr. Richard G. Dorrell and Dr. Juan Pierella Karlusich, Dr. Nathalie Joli, Dr. Maria Helena Cruz de Carvalho, Florent Charton, Shun Liu, Priscillia Pierre-Elies, and other laboratories in our institute, with all of you, I have got so much help and advice, and have always felt supported. Also, many thanks to Dr. Marianne Jaubert in Institut de Biologie Paris Seine, Dr. Angela Falciatore and Jean-Pierre Bouly in Institut de Biologie Physico-Chimique, for their help.

Many thanks to our collaborator Dr. Yannick Schwab and his colleagues in the European Molecular Biology Laboratory and Dr. Robert Sutak at Charles University.

I am also thankful to our pre-director of our doctoral school of Plant Science Dr. Jacqui Shykoff and all staff for all their considerate and helpful guidance.

At last, I would like to thank my family and friends for all the unconditional support during these very intense academic years.

Xia GAO

01/06/2021 in Paris

Table of Contents

Abbreviations

ISIP: iron-starvation induced protein

FTN: ferritin

FET: ferroxidase

FTR: iron (III) permease

FRE: ferric reductase

FBP: ferrichrome-binding protein

NRAMP: natural resistance-associated macrophage protein

ZIP: Zinc transporter

PETF: ferredoxin

FLDA: flavodoxin

PETJ: cytochrome c_6

PETE: plastocyanin

MT: wild type cell lines transformed with empty vector

KO: knock out

FP: fluorescent protein

Fv/Fm: maximum potential quantum efficiency of Photosystem II

ETRmax: electron transport rate

PSI: photosystem I

PSII: photosystem II

Chl: chlorophyll

Fe^{2+}: ferrous iron

FOB: ferrioxamine B

FOE: ferrioxamine E

DMSO: dimethyl sulfoxide

DAPI: 4',6-diamidino-2-phenylindole- a DNA stain

ESAW: enrichment solution with artificial water

PCR: polymerase chain reaction

qPCR: quantitative polymerase chain reaction

RNAseq: RNA sequencing

HGT: horizontal gene transfer

Chapter 1: Introduction

1.1 Iron Bioavailability in the ocean

1.1.1 Iron is an essential micronutrient for all forms of life

Iron (Fe) is required by all organisms. In animals, iron-containing proteins are involved in electron transfer, and participate in transport, storage and use of oxygen. In plants, iron is a vital cofactor for a range of metabolic processes, like photosynthesis, respiration, TCA cycle. Several important steps in photosynthetic pigment metabolism and chloroplast ultrastructure are dependent on iron (Briat *et al.*, 2007). The iron–sulfur clusters are also pervasive and particularly important in nitrogen metabolism. They are at the core of nitrogenase, the enzyme responsible for biological nitrogen fixation (Orme-Johnson, 1985).

These requirements are a product of the evolution of early metabolisms on this planet. When unicellular life began to evolve more than 3 billion years ago the ocean in which the early species lived was devoid of molecular oxygen and rich in ferrous iron (soluble Fe^{2+}). Iron is the fourth most abundant element on Earth (Clarke and Washington, 1924). The metabolic processes that became core to all cellular life relied on the abundance of this transition element and its ability to stably occupy multiple valence states, as cofactor to enzymes catalysing reactions involving the transfer of electrons. Most notably these include photosynthesis and respiration, but also the synthesis of essential organic molecules such as amino acids, lipids, deoxyribonucleotides and sterols. The minimum concentration of iron required by an individual cell to sustain its metabolic functions is referred to as its metabolic 'iron quota'. Within cells, iron homeostasis is carefully controlled, since overabundance of Fe^{2+} can catalyse the formation of damaging reactive oxygen species (Briat *et al.*, 2010).

1.1.2 Iron bioavailability drastically decreased in the contemporary ocean

While the preference of cells for reduced iron has remained unchanged, that is cells readily uptake ferrous iron (Fe^{2+}) and carefully regulate its concentrations intracellularly, the abiotic environment of unicellular species living in today's ocean is

vastly different. Tracing the history of early life on earth more than three billion years ago, bacteria in our planet's young oceans began to produce oxygen through photosynthesis (Figure 1.1). This oxygen combined with dissolved iron in the sea to form insoluble iron oxide, which separated out of the water and sank to the seafloor, forming undissolved iron. It is The Great Oxygenation Event, a period that is usually characterised as a transition to the blooming of primitive life. However, with the increase in the amount of oxygen in the atmosphere, the iron availability decreased, which provided a new stress in the life of algae. The contemporary surface ocean is oxygenated and well mixed, with a mildly alkaline pH (global averages are around 8.2; Lauvset *et al.*, 2015). This is an oxidizing environment that chemically shifts iron into its ferric state (Fe^{3+}), so that the most abundant form of iron is as Fe^{3+} bound by organic material or colloid particles of oxyhydroxides (Gledhill and Buck, 2012). Organisms have therefore had to evolve a variety of molecular mechanisms to make such recalcitrant forms of iron bioavailable. Furthermore, iron availability today is predicted to change rapidly with ocean acidification and global warming. In general, iron bioavailability has drastically decreased in the contemporary ocean. Fe deficiency is likely to be driving the evolution of marine microbes.

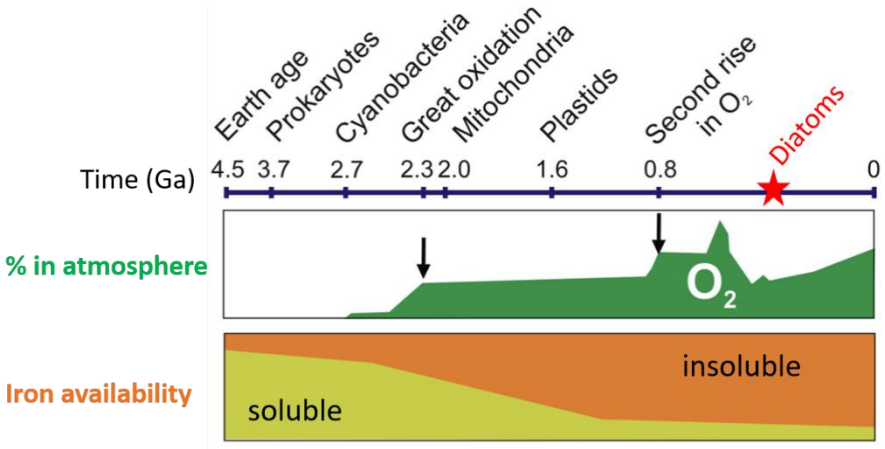


Figure 1.1 Iron availability decreased while oxygen increased in the atmosphere.

Time is measured in billions of years ago (Ga); Percentage of O2 in the atmosphere (in green) and correlated iron availability (in orange) are estimated. Diatoms arose when iron availability was low. Modified from Lodeyro *et al.*, (2012).

1.1.3 Oceanic iron biochemistry

Analysing the chemical speciation of Fe in the contemporary ocean is complicated, and consensus for which iron forms are abundant locally remains elusive. However, to the best of our understanding, the vast majority of iron present in surface ocean seawater is complexed by organic ligands (Gledhill and Buck, 2012). More than 99.9% of dissolved iron is bound to organic ligands in open ocean surface waters (Bonnain *et al.*, 2016; Hassler *et al.*, 2017). The strongest chelates are siderophore molecules, which are exceedingly strong organic iron-binding ligands, have high affinity to ferric iron, comprises more than 99% of the organic iron pool (Bundy *et al.*, 2018). They are produced by organisms, especially well-known by bacteria, although fungi, which are of lower abundance in the ocean, are also capable of the process (Renshaw *et al.*, 2002).

Siderophores can be roughly divided into three chemical classes depending on the functional group, hydroxamate, catecholate and carboxylate. Yet their relative contribution to the seawater iron pool is unknown. Hydroxamate siderophores, which have been detected both in coastal and open ocean waters (Velasquez *et al.*, 2011) are hydrophilic and are released by synthesizing species into the water column. In these microorganisms, siderophores are recaptured via specific recognition machinery. It has been proposed that it is these molecules in particular that are available to non-producing species, such as diatoms (Hopkinson and Morel, 2009). FOB is a typical hydroxamate siderophore bound iron, and is widely used in studies. Other weaker iron-binding ligand, including humic substances, exopolymeric substances, transparent exopolymers, nanofibrils, and even the biotoxin domoic acid have been increasingly recognized as an important component of a weaker iron-binding ligand pool in the oceans, particularly in the deep sea and in coastal environments (Figure 1.2). Notably, many marine viruses are also capable of chelating iron on their tails and thus may represent a large sink of marine iron (Bonnain *et al.*, 2016 and references therein). Yet determining the identity of these ligands in seawater remains a major challenge. Inorganic mineral iron is periodically available to surface communities, either through aeolian input that is arising from the action of wind, or via upwelling currents that bring re-mineralised iron from the deep ocean (Mahowald *et al.*, 2009; Boyd and Ellwood, 2010). Of these sources, the dissolved uncomplexed inorganic iron is the form most readily taken up by organisms, but present at very low concentrations, is insufficient for microbial growth

(Liu and Millero, 2002; Lis *et al.*, 2015).

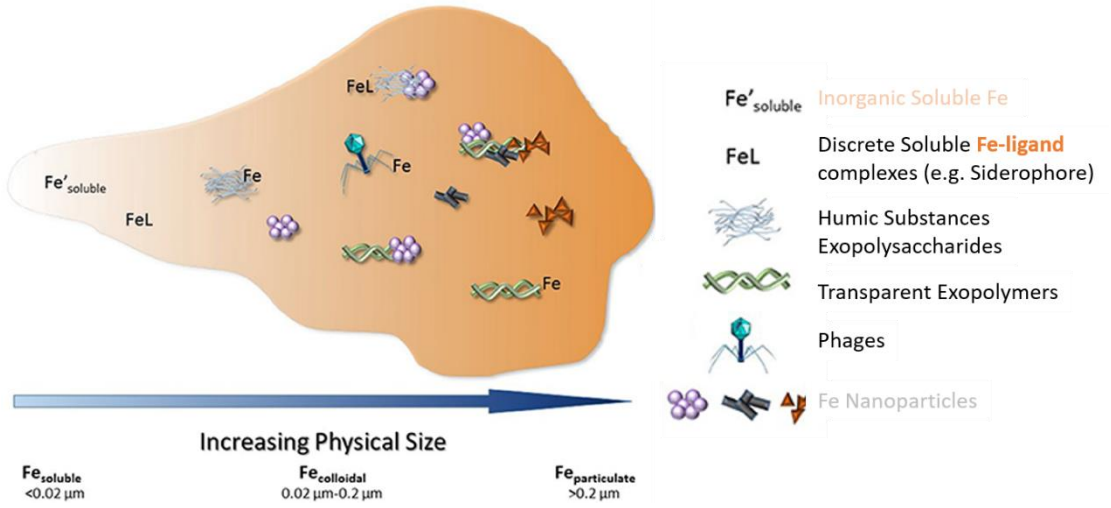


Figure 1.2 Summary of dissolved iron-binding ligands identified within seawater.

Decreasing kinetic lability of Fe within the components is represented by deeper orange background shading. From left to right, iron binding ligands are arranged according to the physical size. Figure is modified from Bonnain *et al.*, (2016).

1.2 Diatoms are a major class of phytoplankton

1.2.1 Introduction to diatoms

Diatoms comprise one of the dominant, ubiquitous groups of photosynthetic producers in the contemporary ocean, estimated to account for 20% of total global primary production (~ 20 x 10^{15} g carbon fixed per year) (Field *et al.*, 1998; Mann, 1999). Diatoms are unicellular stramenopiles (heterokont protists within the chromalveolates) and, in an evolutionary context, a relatively young form of life (Figure 1.3). Molecular phylogenies date the origin of diatoms towards the beginning of the Mesozoic Era (Sims *et al.*, 2006) while current estimates of diatom species richness vary between 40,000-200,000; counter-intuitively, the higher end is based on morphology-based estimates while the lower estimates are based on molecular markers (reviewed in (Benoiston *et al.*, 2017)).

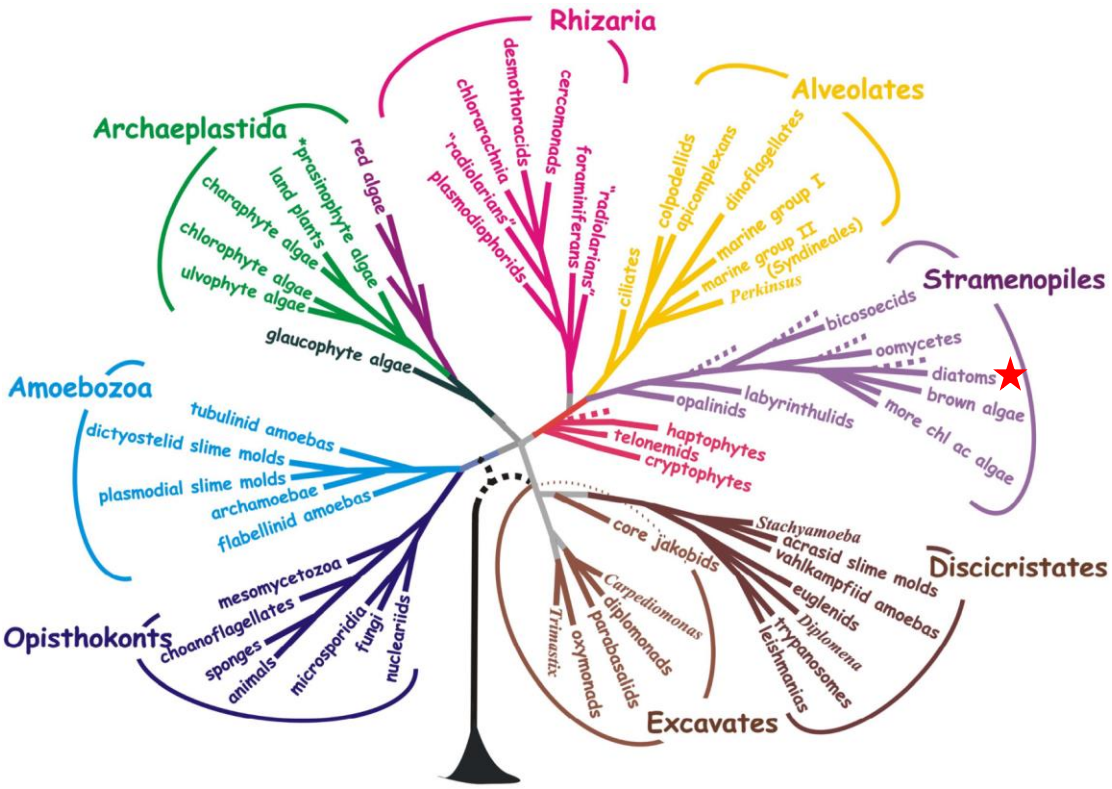


Figure 1.3 A consensus phylogeny of the major groups of eukaryotes.

The tree is based on published molecular phylogenetic and ultrastructural data. Dotted lines indicate positions of major lineages of Stramenopiles known primarily from ciPCR (Massana *et al*., 2006). The two currently proposed positions for the eukaryote root are also indicated. Figure is taken from Baldauf, (2008).

The evolutionary history of diatoms is shaped by endosymbiotic events, of which there were at least two (Figure 1.4). The first occurred an estimated 1.5 billion years ago, when a eukaryotic heterotroph assimilated or was invaded by a cyanobacterium (Yoon *et al*., 2004). Some 500 million years later, a secondary endosymbiosis occurred in which a eukaryotic heterotroph captured or was invaded by a red alga, establishing the photosynthetic stramenopile lineage to which diatoms belong (Cavalier-Smith, 1999). Recent evidence suggests that the chimeric nature of diatoms is even more complicated, as a range of nucleus encoded proteins are of green algal origin (Dorrell *et al*., 2017 and references therein).

Diatoms are subclassified into ‘centric’ species, which include two sub-orders (radial and polar), and ‘pennate’ species, which are further divided into two sub-orders (Figure 1.5). These groups are not monophyletic, as centric diatoms grade into araphid pennates, and araphid pennates grade the raphe-bearing pennate diatoms, which are a natural group (Alverson and Theriot, 2005).

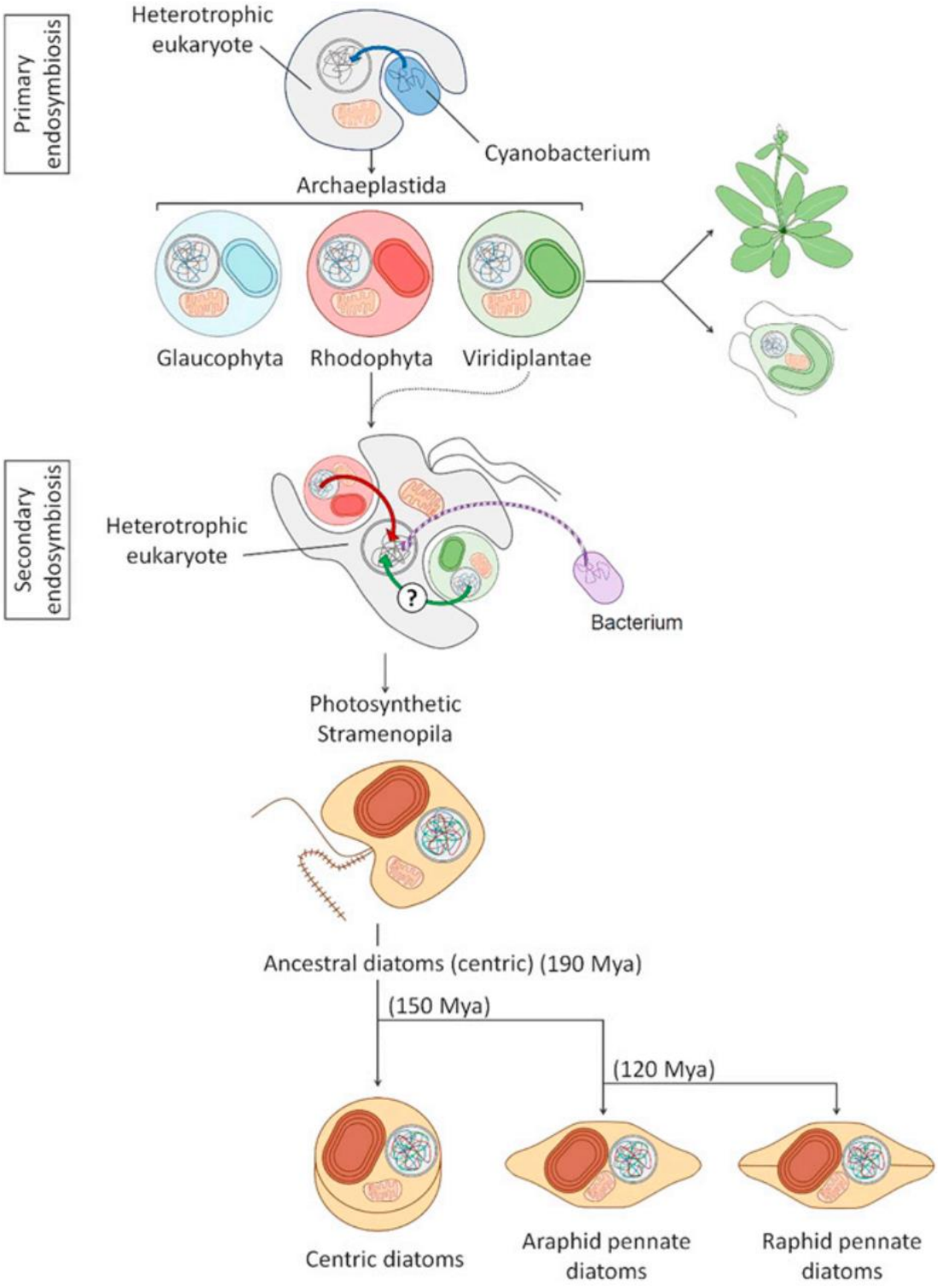


Figure 1.4 Simplified scheme of the major events leading to the evolution of diatoms through primary and secondary endosymbiosis.

The initial primary endosymbiosis occurred when a heterotrophic host engulfed a cyanobacterium (represented in blue). Over time, a large proportion of the cyanobacterial genome was transferred to the nucleus of the host, as indicated by the blue arrow. The endosymbiotic process generated the plastids of the Archaeplastida, a major group including Glaucophyta (pale blue), Rhodophyta (pink), and the Viridiplantae (the model green alga *C. reinhardtii* and plant Arabidopsis are shown). During secondary endosymbiosis, a different heterotrophic cell acquired a red alga and potentially also a green alga. The algal endosymbiont became the plastid (in brown) of the Stramenopila, a group including diatoms, but also other algae such as pelagophytes or the multicellular kelps. Algal nuclear genomes were transferred to the heterotrophic nucleus, as represented by green and red curved arrows, while the algal nucleus and mitochondria were lost. Other bacterial genes in the Stramenopila genome were derived by horizontal gene transfer events from bacterial donors (violet arrow). The figure also shows the approximate dates of diatom evolution and separation between centric and pennate diatoms based on Nakov *et al.*, (2018). Some pennate species further diversified and acquired the capacity to move by secreting mucilage through a longitudinal slit in the cell wall called raphe, hence the division between raphid (motile) and araphid (nonmotile) pennate diatoms. Mya, million years ago. Figure is taken from Falciatore *et al.*, (2020).

Diatom cells have unique structure and features (Figure 1.5). Cell sizes are large compared to other oceanic unicellular eukaryotes and are generally upwards of 20 µm in diameter (with maximum cell sizes of 2 mm recorded). Some diatoms are known to form chains. Unlike other organisms, diatom cell membranes are silicified and referred to as frustules. Frustules are porous, with the size of nanopores ranging from 250–600 nm; they are generally assumed not to interfere with diatom nutrient uptake (Bhatta *et al.*, 2009). The majority of pennate diatoms contain one central vacuole while centric species can have many. Diatom plastids are surrounded by four membranes, a vestige of the endosymbiosis events that generated the organelle. As with vacuoles, the majority of pennate species have only one plastid, but this is variable in the centrics.

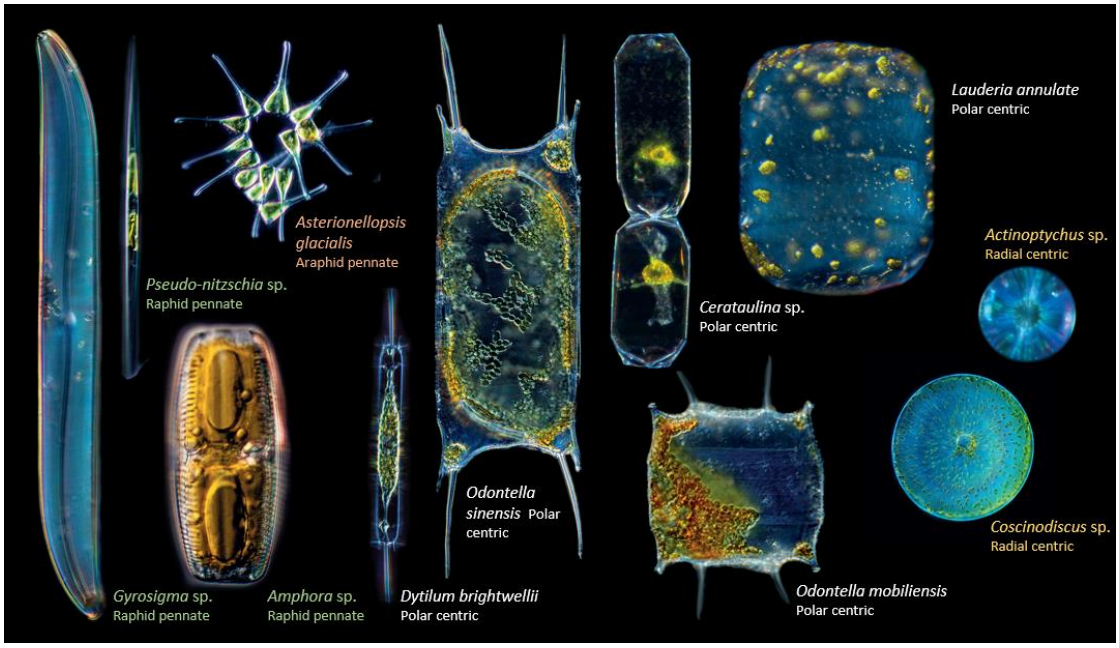


Figure 1.5 Light microscope images of representative species in raphid and araphid pennate, polar and radial centric diatom groups.

Raphid pennate (on the left) and araphid pennate (upper left), polar centric (middle) and radial centric (right). Sizes are not to scale. Diatoms collected in Roscoff Marine Station (Roscoff, France). Photography and image processing by Noé Sardet, Parafilms, Montreal.

The ecological success of diatoms is remarkable. They are present in all ocean provinces and are the most abundant photosynthetic producers in the Arctic and the Southern Ocean (Malviya *et al.*, 2016). The global distribution of diatoms has been estimated through biogeochemical/ecosystem simulation (Figure 1.6). Importantly, they are ruthless competitors for iron, often dominating iron-stimulated blooms (Boyd *et al.*, 2007). Blooms of diatoms stimulated by addition of iron filings to seawater have been detected using satellites in space (Figure 1.7). This indicates that diatoms are adapted to survive conditions of low iron and are equipped with the molecular machinery necessary to uptake and assimilate iron efficiently, without bringing damage to their cells through oxidative stress.

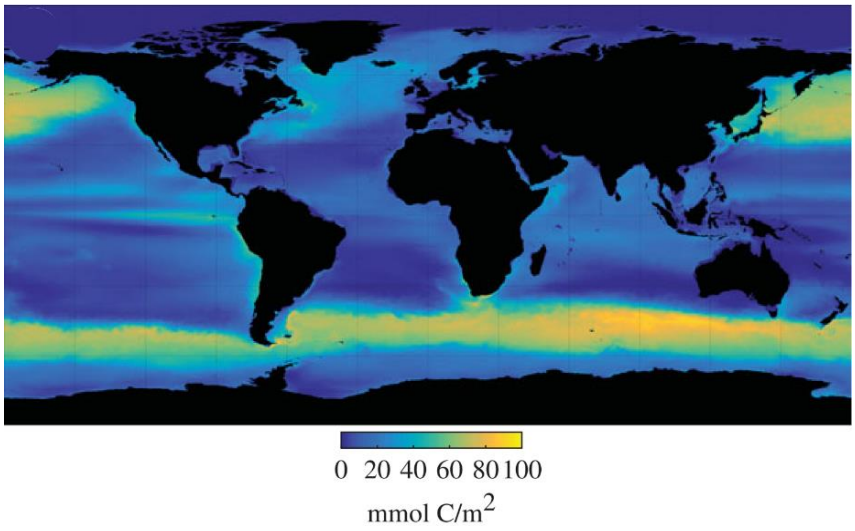


Figure 1.6 Global distribution of diatoms in the ocean. Water column inventory of diatom biomass (mmol C/m^2) from a biogeochemical/ecosystem simulation. Figure is taken from Benoiston *et al.*, (2017).

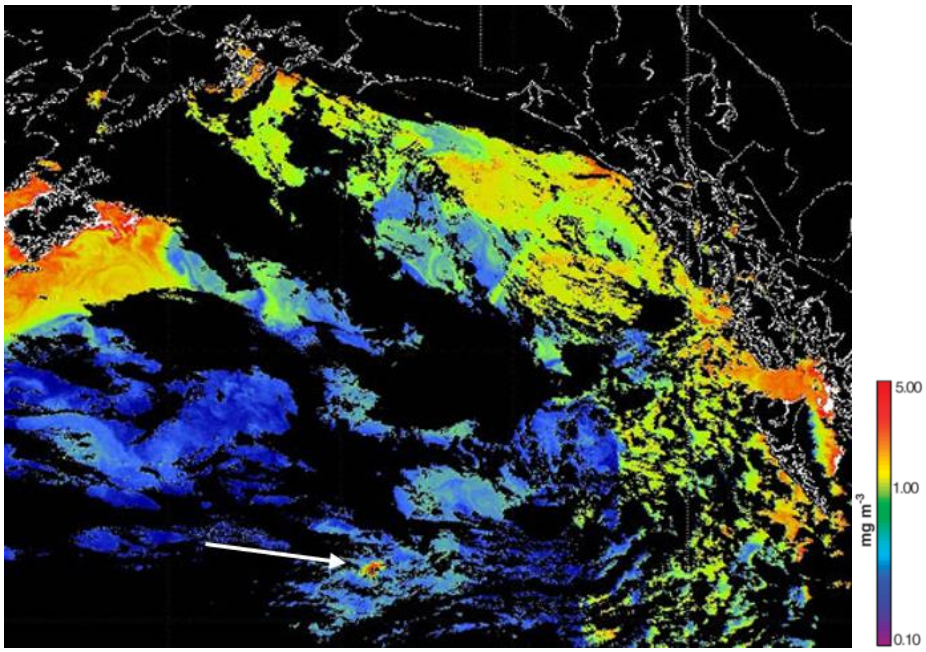


Figure 1.7 Diatom bloom induced by iron addition. Mesoscale iron fertilization experiment often results in blooms dominated by diatoms. The red spot shows a diatom bloom in the Pacific Ocean just off the coastline of Washington State in the USA, induced by Fe addition after 19 days. Figure is taken from Boyd *et al.*, (2007).

Iron starvation leads to down-regulation of nitrogen reducing enzymes such as Fe-dependent nitrate and nitrite reductases, and the up-regulation of enzymes involved in nitrogen recycling in diatoms (Nunn *et al.*, 2013). Rates of iron uptake and use are also linked to nitrogen assimilation and growth; on average diatoms with higher iron uptake increase their nitrate uptake and productivity, proportionately (Falkowski, 1997). Surprisingly, this is not reflected in their silica uptake, which remains constant under iron supplementation despite faster growth, leading to weakly silicified cells (Boyle, 1998 and references therein). Diatom iron nutrition therefore impacts not only metabolism but also the resulting morphology and physical characteristics of diatom cells.

1.2.2 'Omic resources available for the study of diatoms

Genomes

Within diatoms, nine species have published sequenced genomes: including four centric species, *Thalassiosira pseudonana* (Armbrust *et al.*, 2004), *Thalassiosira oceanica* (Lommer *et al.*, 2012), *Cyclotella cryptica*, *Skeletonema costatum*; and five pennate species, *Phaeodactylum tricornutum* (Bowler *et al.*, 2008), *Fragilariopsis cylindrus* (Mock *et al.*, 2017), *Pseudo-nitzschia multiseries* CLN47 (JGI Project ID: 16870), *Pseudo-nitzschia multistriata* strain B856 (NCBI assembly ASM90066040v1), and *Fistulifera solaris* (Tanaka *et al.*, 2015). The genome properties are summarised in Table 1.1. Currently, there are another two species with published genomes: *Synedra acus* (Galachyants *et al.*, 2015), and *Plagiostriata* sp. CCMP470 (Sato *et al.*, 2020). Overall, the pennate diatom *P. tricornutum* and the centric diatom *T. pseudonana* are the most advanced laboratory model species for molecular studies, that can be readily cultured and transformed (Apt *et al.*, 1996a; Zaslavskaia *et al.*, 2000; Poulsen *et al.*, 2006; Hopes *et al.*, 2016). Although neither is abundant in the environment, they often harbour genes that have been proposed to contribute to diatom success in the wild.

Table 1.1 Genomic features of established and emerging diatom model species.

	Centric Diatoms				Pennate Diatoms				
Features	*Thalassiosira pseudonana*[a]	*Thalassiosira oceanica*	*Cyclotella cryptica*	*Skeletonema costatum*	*Phaeodactylum tricornutum*[a]	*Fragilariopsis cylindrus*	*Pseudo-nitzschia multiseries*	*Pseudo-nitzschia multistriata*	*Fistulifera solaris*
Strain	CCMP1335	CCMP1005	CCMP332	Ariake8	CCAP 1055/1	CCMP 1102	CLN-47	B856	JPCC DA0580
Ploidy	diploid	diploid	diploid		diploid	diploid, triploidy proposed			allodiploid
Genome size (Mbp)	32.1	81.6	161.7	46.9	27.4	61.1	218.7	59	24.9
Number of:									
Chr. pairs	24				34				42
Scaffolds		51,656			55		4890		297
Haplotype divergence	0.75%				0.85%	up to 6% in 29% of genes			14 to 38%
Number of genes	11,776	34,642	21,121	16,449	12,233	21,066	19,703	12,008	11,448
Average gene size (bp)	1,740.8	1,255	1,471	1,100	1,624	1,575	1,522	2,205	
Intron									
Number/gene	1.54	1.29	1.18		0.89	1.1	1.38	0.87	
Average length	125.4 bp	150 bp	125 bp		142 bp	255.7 bp	229 bp	341 bp	
Plastid									
Genome size	129 kbp	141 kbp	129 kbp		117 kbp		111. 5 kbp		135 kbp
Number of genes	127	128	132		130		97		132
Mitochondria									
Genome size	43.8 kbp	35.3 kbp	58 kbp		77.3 kbp		46.3 kbp		>38.6 kbp
Number of genes	34	31	35		34		37		31
Annotation version	v3	v1			v3	v1	v1		
References	(Armbrust et al., 2004)	(Lommer et al., 2012)	(Traller et al., 2016)	(Ogura et al., 2018)	(Bowler, 2008; Rastogi et al., 2018)	(Mock et al., 2017)		(Basu et al., 2017)	(Tanaka et al., 2015b)

Genome portal: *T. pseudonana*, https://genome.jgi.doe.gov/Thaps3/Thaps3.home.html; *T. oceanica*, https://genome.jgi.doe.gov/Thaoce1/Thaoce1.info.html; *C. cryptica*, http://genomes.mcdb.ucla.edu/Cyclotella/download.html; *S. costatum*, https://ddbj.nig.ac.jp/DRASearch/study?acc=DRP004496; *P. tricornutum*, Phatr3:http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index; *F. cylindrus*, https://genome.jgi.doe.gov/Fracy1/Fracy1.info.html; *P. multiseries*, https://genome.jgi.doe.gov/Psemu1/P semu1.home.html; *P.multistriata*, http://apollo.tgac.ac.uk/Pseudo-nitzschia_multistriata_V1_4_browser/sequences; *F. solaris*, https://trace.ddbj.nig.ac.jp/DRASearch/submission?acc=DRA002403. Chr., chromosome.
Blank cells indicate no data available.
[a]Established diatom model species.

Other available genomes include *Synedra acus* and *Plagiostriata* sp. CCMP470. The table is taken from Falciatore *et al*., (2020).

Transcriptomes and proteomes

The availability of many transcriptomic and proteomic data sets has greatly facilitated genome annotation efforts and therefore gene discovery in model species. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP), which brought together a global effort to sequence marine phytoplankton, includes 92 diatom species with available transcriptomes (Keeling *et al.*, 2014). A further 7 transcriptomes have been generated by Genoscope and represent seven of the most globally abundant diatom genera, not covered by the MMETSP project (Dorrell *et al.*, 2021). The seven genera were determined to be abundant by analysis of data collected through the *Tara* Oceans expedition. *Tara* Oceans is an international, multidisciplinary project to assess the complexity of ocean life across comprehensive taxonomic and spatial scales. During the effort, community composition, metagenomes and metatranscriptomes of aquatic

microorganisms were collected from 210 sampling stations at depths down to 1,000 m during a global circumnavigation (Sunagawa *et al.*, 2020). These efforts illustrate how global-scale concepts and data can help to integrate biological complexity into models and serve as a baseline for assessing ecosystem changes. Furthermore, large amounts of metagenomics and metatranscriptomics data are now being generated from the sequencing of natural phytoplankton populations (Table 1.2), representing powerful resources for exploring the distribution of diatom genes and species in an environmental context. A planetary-scale understanding of the ocean ecosystem, particularly in light of climate change, is crucial. The *Tara* Oceans global circumnavigation effort can offer initial insight. However, a database centralizing all of these 'omic resources is still lacking.

Table 1.2 Large omics datasets for phytoplankton including diatoms.

Resource	Data Types	Website
The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP)	678 transcriptomes from 411 different strains, among which 92 different diatoms	https://www.imicrobe.us/#/projects/104
The Tara Ocean	MetaTranscriptomic MetaGenomic	http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/ https://doi.org/10.1594/PANGAEA.875582
Sea of Change project	MetaTranscriptomic MetaGenomic	https://genome.jgi.doe.gov/portal/AntOcenscriptome_8_FD/AntOcenscriptome_8_FD.info.html

Taken from Falciatore *et al.*, (2020).

1.2.3 Photosynthesis in diatoms

Photosystems in diatoms

Diatoms usually dominate artificially stimulated algal blooms (Boyd *et al.*, 2007) as well as many natural coastal seasonal blooms (Carstensen *et al.*, 2015). Their photosynthesis is highly efficient under dynamic light regimes. For example, in a study that compared green alga *Chlorella vulgaris* with the diatom *P. tricornutum* under a fluctuating light regime, a much higher conversion efficiency of photosynthetic energy into biomass was observed in the diatom compared with the green alga. The differences between green algae and diatoms in the efficiency of biomass production per photon absorbed are understood to be caused by the different amount of alternative electron

cycling, which is defined as the difference between fluorescence- and oxygen-base electron flow (Wagner *et al.*, 2006).

Various studies have addressed the question of what makes diatom plastids special. Their evolutionary history may be implicated. Diatom plastids have an unusual architecture. They are enclosed by four membranes (Figure 1.8) due to their origins from secondary endosymbiosis (Cavalier-Smith, 2003), unlike plastids in the green lineage, which are surrounded by only two membranes. The outermost chloroplast membrane in diatoms is connected to the endoplasmic reticulum (ER). Diatom plastids contain stacks of three thylakoids that run through the entire organelle. The photosystems are segregated, with photosystem I (PSI) found mostly within the peripheral membranes facing the stroma, and photosystem II (PSII) located within the core membranes at the junction between two thylakoid membranes (Pyszniak and Gibbs, 1992; Flori *et al.*, 2017). This structural organization is in contrast to that found in plants, in which grana stacks interspaced by stroma lamellae. Therefore, the distance between the two photosystems is short, as they face each other in the two external thylakoids, and there are physical connections between thylakoids (Flori *et al.*, 2017). This unusual segregation of the two photosystems prevents the loss of efficiency of the photosystems via possible spill over of excitons from PSII to PSI, as found in cyanobacteria and red algae (Biggins and Bruce, 1989). Further, state transitions observed in the green lineage when light harvesting complexes move from PSII to PSI is absent in diatoms. In *P. tricornutum*, the adjustments of the distribution of excitation energy between photosystems are made at the level of excitation energy transfer to the PSII reaction centre, which prevents prolonged loss of photosynthetic capacity. These differences in light-harvesting function are probably a response to the aquatic light field and may account for the success of diatoms in low and variable light environments (Owens, 1986).

In terms of pigments, diatoms use chlorophyll (Chl) *a* and Chl *c*, and the main carotenoids are b-carotene, fucoxanthin (Fx), and diadinoxanthin/diatoxanthin. The fucoxanthin Chl *a/c* binding protein (FCP), which belongs to the family of the Light Harvesting Complex (LHC) proteins, binds seven Chl *a*, two Chl *c*, seven Fx, and likely one diadinoxanthin within the protein scaffold (Wang *et al.*, 2019).

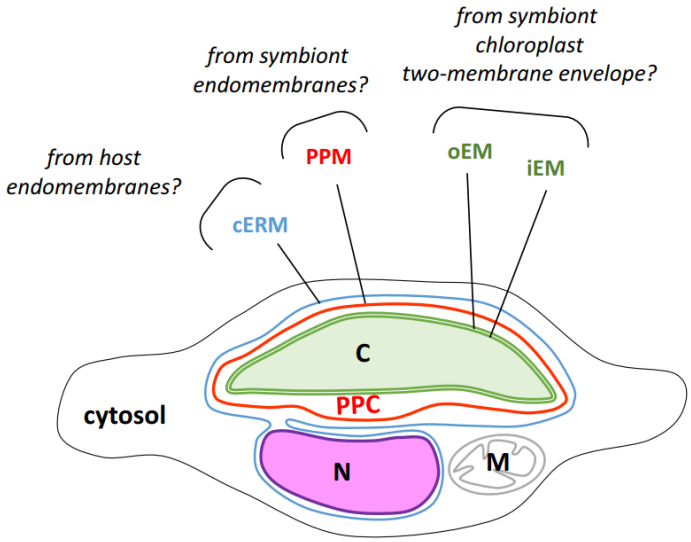


Figure 1.8 Chimeric organization of the secondary plastid in diatoms.

The schematic shows a fusiform cell of *P. tricornutum*. The plastid is surrounded by 4 membranes. This is in contrast with algae belonging to the green lineage, where chloroplasts only have two membranes (see Figure 1.4). The chloroplast endoplasmic reticulum membrane (cERM), shown in blue, is continuous with the outer nuclear envelope membrane. The periplastidial membrane (PPM) is shown red. The outer and inner envelope membrane (oEM and iEM), shown in light green, are tightly apposed. The presence of a specific periplastidial compartment (PPC) is based on the detection of blob-like structures observed by confocal microscopy. cERM and PPM, absent in green lineage, are thought to be vestiges of two events, primary and secondary endosymbiosis. C, chloroplast; N, nucleus; M, mitochondrion. Figure is taken from Flori *et al.*, (2016).

Remote sensing of diatom blooms in the ocean

All phytoplankton contain the photosynthetic pigment chlorophyll-*a* (Chl), which is often used as a proxy for estimating phytoplankton biomass. The concentration of Chl can be measured in a variety of ways, both *in situ* and remotely. *In situ* collection and filtering of water samples to measure Chl concentrations are often costly and limited

geographically to regions frequented by research cruises or established monitoring programs. In contrast, Chl fluorometers are relatively easy to obtain and deploy on a variety of *in situ* platforms, such as gliders, profiling floats, and semi-autonomous surface vehicles (Haëntjens *et al.*, 2017; Gentemann *et al.*, 2020). Chlorophyll fluorometers provide the largest *in situ* global data set for estimating phytoplankton biomass because of their ease of use, size, power consumption, and relatively low price (Roesler *et al.*, 2017). *In vivo* chlorophyll fluorescence can serve as a reasonable estimator of *in situ* phytoplankton biomass with the benefits of efficiently and affordably extending the global chlorophyll data set in time and space to remote oceanic regions where routine sampling by other vessels is uncommon.

Comparison of satellite information about cyanobacteria and diatom distributions clearly matches the concentrations based on high pressure liquid chromatography (HPLC) pigment analysis of water samples and concentrations collected from a global model analysis with the NASA Ocean Biogeochemical Model (Gregg *et al.*, 2003; Gregg and Casey 2007). Further, new approaches to discriminate dinoflagellate from diatom blooms from space and to correcting data source for known biases have been developed (Shang *et al.*, 2014; Scott *et al.*, 2020). Spatial distributions of Chl *a* fluorescence and Chl *a* concentration from HPLC pigment analyses (Taylor *et al.*, 2012) were found to be very similar across the frontal zone in the California Current ecosystem. This diatom dominated sampling area, Chl *a* fluorescence per unit Chl *a* was 50% brighter than for other samples from the euphotic layer across the frontal zone (Chekalyuk *et al.*, 2012). The quantitative assessment of the distribution of key phytoplankton groups from space (Figure 1.9) enables various biogeochemical regions to be distinguished and will be of great importance for the global modelling of marine ecosystems and biogeochemical cycles which enables the impact of climate change in the oceanic biosphere to be estimated (Bracher *et al.*, 2009).

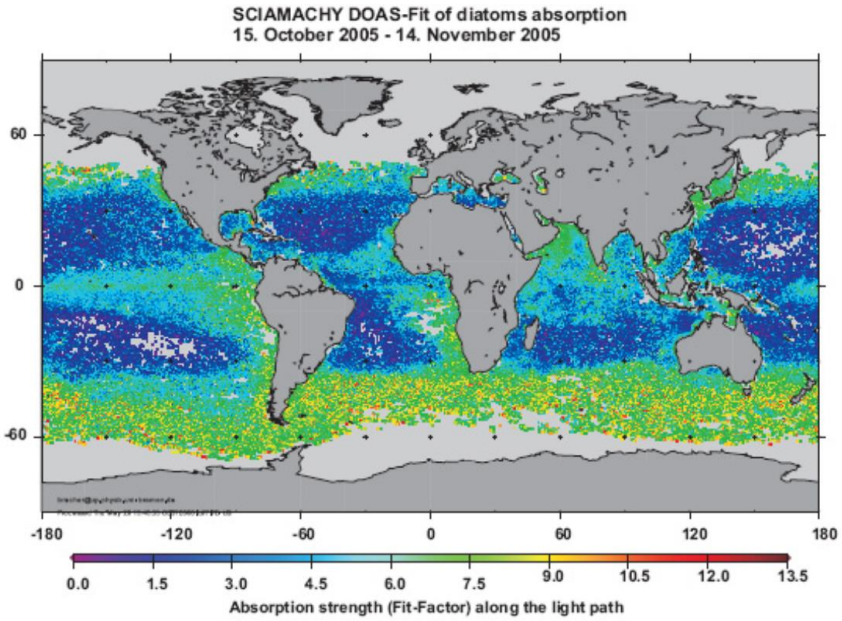


Figure 1.9 Monthly average of global distribution of diatoms Obtained as "Strength of Absorption" (=Fit-Factor) by using PhytoDOAS with SCIAMACHY (from 15 October to 14 November 2005). Figure is taken from Bracher *et al.*, (2009).

1.3 Iron metabolism strategies in diatoms

Overview

Diatoms are successful survivors and strong competitors under low iron concentrations in the ocean and are assumed to rely on multiple iron metabolism strategies for this. The strategies which are under investigation include the ability for high affinity iron uptake, high-capacity iron storage and efficient iron usage (Figure 1.10).

Iron-regulated genes in *P. tricornutum* were first identified through a combination of nontargeted transcriptomic and metabolomic approaches (Allen *et al.*, 2008). Physiological functions of these genes were studies in recent years. It has been shown that diatoms respond to available iron spontaneously, exhibit high affinity to a range of iron resources, and internalize dissolved iron and complexed iron by endocytosis without pre-processing, reduction (Kazamia *et al*., 2018; Morrissey *et al*., 2015).

Diatoms have been shown to be able to store iron exceeding their requirement many times and thus survive long-term iron deficiency (Sunda and Huntsman, 1995). FTNs are found in all domains of life, including animals, plants, and microorganisms, and are designed to accommodate large amounts of iron (Theil, 1987). Structurally, FTNs are large protein cages formed by arrays of self-assembling helices with nanocavities (5–8 nm) that catalytically couple iron and oxygen at protein sites for precursors of the cavity mineral. The mineral core of a single protein can store up to 4500 iron atoms (Liu and Theil, 2005). Iron is reversibly released from the core by reduction. However, whether ferritin functions universally as an iron storage protein is not confirmed.

In diatoms, FTN has been shown experimentally to function as an iron storage protein in the bloom-forming diatom species *Pseudo-nitzschia* (Marchetti *et al*., 2009). When iron becomes limited, diatoms with FTN have been shown to carry out more cell divisions than non-FTN-containing species (Marchetti *et al*., 2009), and FTN has been found to be critical for cells to load and unload iron efficiently (Botebol *et al.*, 2015).

Photosynthesis relies heavily on iron. While PSII requires 2 iron cofactors per unit, PSI requires 12 (Raven *et al.*, 1999; Strzepek *et al.*, 2012). In iron-limiting conditions, almost all diatoms decrease their photosynthetic activity. However, diatoms are able to

replace iron-containing proteins with iron-free equivalents to lower iron requirements. Additionally, some species are able to change photosynthesis architecture, the ratio between PSI and PSII, to maintain their photosynthetic activity (Strzepek and Harrison, 2004).

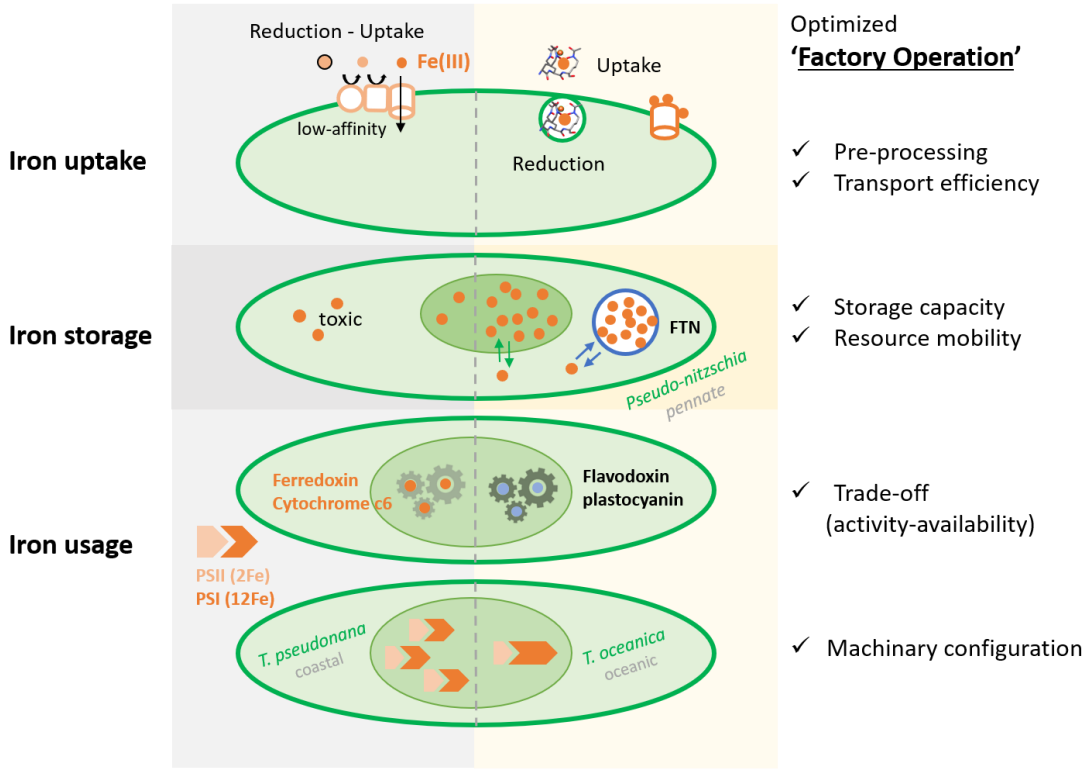


Figure 1.10 Summary of Iron Metabolism Strategies in Diatoms.

In the following sections I review these strategies in detail.

1.3.1 Iron acquisition in diatoms

A portfolio of iron uptake mechanisms shows diatoms are adapted to acquire iron from a variety of sources

Diatoms have a diverse portfolio of iron uptake mechanisms adapted to the sources of iron available in seawater (summarised in Figure 1.11). Through a combination of nontargeted transcriptomic and metabolomic approaches, biochemical strategies preferred by *P. tricornutum* at growth-limiting levels of dissolved Fe have been explored. Iron-regulated genes in *P. tricornutum* were first identified, including genes

encoding the iron starvation induced proteins (ISIPs), ferric reductase (FRE2), ferrichrome-binding protein (FBP1)(Allen *et al.*, 2008). A mechanism whose function has been experimentally verified in the model pennate diatom *P. tricornutum* is the assimilation of uncomplexed iron bound by a transferrin-like protein. The iron starvation induced protein 2A (ISIP2A) first identified in *P. tricornutum* (Allen *et al.*, 2008) was shown to function as a 'phytotransferrin', a protein directed to the cell membrane of cells with carboxylate iron binding domains, whose iron chelating properties were pH sensitive (Morrissey *et al.*, 2015; McQuaid *et al.*, 2018). *P. tricornutum* cells deficient in ISIP2A showed reduced Fe′ uptake capabilities, suggesting that this was the primary but not only mechanism of Fe′ uptake in this species (Kazamia *et al.*, 2018). ISIP2A contains the functional domain PF07692, which was first characterized in the Fe-assimilation proteins FEA1 and FEA2 of the green alga *Chlamydomonas reinhardtii* (Allen *et al.*, 2007), and additionally was functionally described in *Dunaliella salina* (Paz *et al.*, 2007). In *C. reinhardtii* FEA1 and FEA2 were highly expressed under low iron conditions and facilitated high-affinity iron uptake (Allen *et al.*, 2007), likely in its ferrous form (Narayanan *et al.*, 2011). By contrast, in diatoms, the iron binding domains of ISIP2A coordinate ferric iron (McQuaid *et al.*, 2018).

Experimental evidence suggests that colloidal iron, as particulate Fe^{2+}, is bioavailable to *P. tricornutum* (Shoenfelt *et al.*, 2017). While aeolian inputs of iron are known to stimulate algal growth, and diatoms in particular, both in short-term ecological experiments (De Baar *et al.*, 2005) and over geological timeframes (Hain *et al.*, 2014; Martínez-García *et al.*, 2014), it has been a longstanding assumption that the iron stimulation is through the increase in Fe′ (Rich and Morel, 1990). Shoenfelt and colleagues challenged this assumption by demonstrating that growth rates of *P. tricornutum* cultures improved by glacial dust stimulation beyond what is expected by the increase in Fe′ alone (the assimilation rates did not match Monod dynamics, and a fit to Michaelis-Menten kinetics resulted in unrealistic half-saturation concentrations). In the same experiments glaciogenic sediments were more bioavailable to *P. tricornutum* than non-glaciogenic sediments rich in iron, provided cells were in physical contact with the abiotic material. Together the results implicate an evolved molecular mechanism for the "mining" of mineral iron by some diatoms, which remains to be uncovered. A similar property was observed in the filamentous *Trichodesmium*

cyanobacteria, although there too the molecular underpinnings are currently unresolved (Rubin *et al.*, 2011).

There is significant evidence to show that diatoms are able to access the organic pool of ligand-bound iron. Pioneering work by Soria-Dengg and Horstmann (1995) showed that *P. tricornutum* cells were able to access iron from the tri-hydroxamate siderophores ferrioxamine E (FOE) and ferrioxamine B (FOB), albeit with different kinetics. Working on the same model species Kazamia *et al.* (2018) demonstrated that this uptake involved endocytosis of the siderophore complex into the cell, with a reduction step to disassociate the bound iron in the vicinity of the chloroplast. Through reverse genetics, the iron starvation induced protein 1 (ISIP1) was identified as being necessary for the endocytosis and siderophore assimilation process. Intriguingly, ISIP1 was shown to be largely a diatom-specific protein, implying an evolutionary innovation in this group (Kazamia *et al.*, 2018).

Additionally, Coale *et al.* (2019) verified that a previously identified putative siderophore binding protein, FBP1, which in bacteria binds the hydroxamate siderophore ferrichrome, was indeed required for siderophore assimilation. In physiological experiments on three diatom species reported in Kazamia *et al.* (2018), it was shown that siderophore recognition varied: *P. tricornutum* was unable to use the catecholate siderophore enterobactin as a source of iron, whereas *T. oceanica* assimilated this molecule, but did not respond to ferrioxamine supplementation despite having copies of *FBP1, FRE2* and *ISIP1* in its genome. *T. pseudonana*, which lacks *FBP1* and *ISIP1* genes, was unable to use either siderophore source of iron. This suggests that while the presence of *FBP1* could be a marker for siderophore uptake in diatoms, further experiments are required to understand what determines substrate specificity in these species.

Laboratory studies on model diatom species, including *P. tricornutum* and *T. oceanica* report that diatom cells exhibit ferric reductase activity (Maldonado and Price, 2001; Kazamia *et al.*, 2018; Coale *et al.*, 2019). Six genes in *P. tricornutum* have been putatively annotated as ferric reductases, but only two (*FRE1* and *FRE2*) encode two domains indicative of ferric reductase function: the ferric reductase transmembrane component PF01794 and the NAD-binding domain PF08030 (Zhang *et al.*, 2013). The *in silico* prediction for FRE2 is that it is targeted to the chloroplast, albeit with low

confidence, whereas the localisation of FRE1 is unassigned (Rastogi *et al.*, 2018).

The question arises whether these ferric reductases act on the surface of diatom cells, thus enabling the dissociation of iron from its seawater chelates prior to assimilation, or act intracellularly. On the one hand, extracellular reduction is unlikely in the highly oxidizing and diffusing environment of seawater. On the other, if it were to be coupled to efficient (that is rapid and proximal) iron uptake systems, it could allow cells to dispense with the need for discriminating between iron sources (Shaked *et al.*, 2005). The latter is a system well described in fungi, where ferric iron is displaced from weak ferric ligands by a ferrireductase, then re-oxidized by a multicopper oxidase and finally channelled across the plasma membrane through a Fe^{3+} permease (reviewed by Philpott, 2006). Transcriptional upregulation of a ferrireductase and two iron permeases was recorded under limiting iron conditions in *T. pseudonana* cells (Kustka *et al.*, 2007). However, the question of whether this reductive system acts in consort on the cell surfaces of diatom cells requires experimental verification.

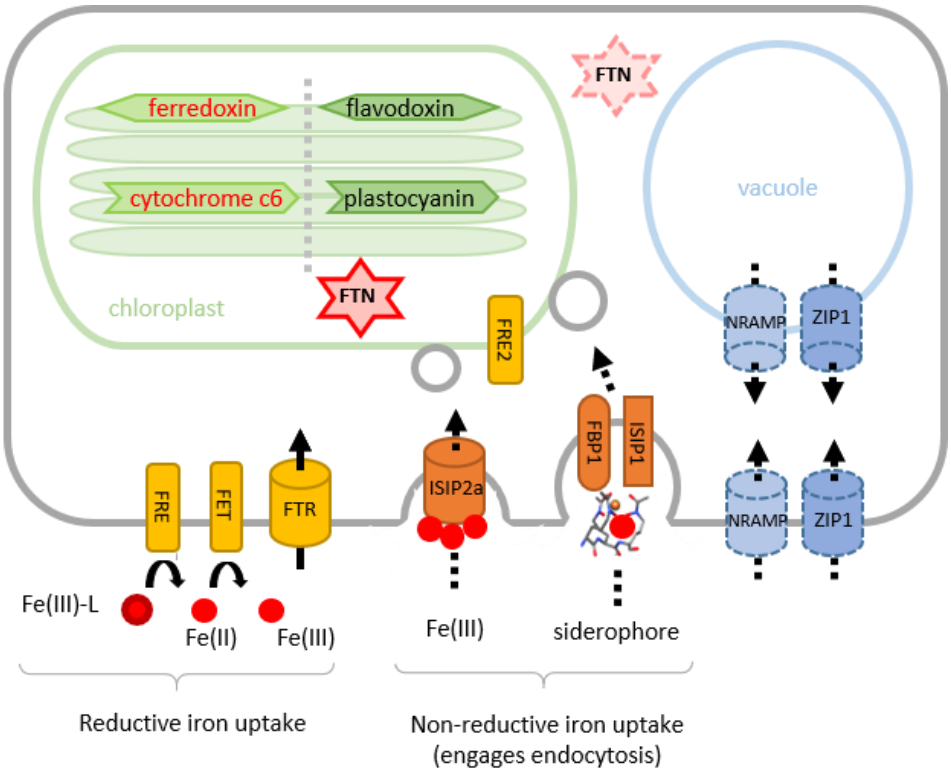


Figure 1.11 Schematic diagram of iron metabolism in diatoms.

A summary schematic of the genes discussed in this section as well as their cellular

location. In the chloroplast, I review the replacement of ferredoxin with its iron free equivalent, flavodoxin and the substitution of cytochrome c_6 with plastocyanin. Reductive and non-reductive iron uptake mechanisms were considered, the latter employing the cell's vesicular network to direct iron to the plastid. Two alternative iron storage mechanisms are reviewed: the accumulation of Fe in the mineral core of ferritin, and vacuolar storage. For the storage proteins, the cellular localization of the putative proteins has not been confirmed through microscopy analyses, indicated with dashed lines. FTN localisation has been explored in this thesis (see Chapter 3).

1.3.2 Iron homeostasis in diatoms

Low iron quota diatoms regulate the amount of Fe required for photosynthesis

The metabolic iron requirements of diatom cells differ from species to species. This is measured as Fe:C ratios (µmol Fe mol C^{-1}) and is estimated from co-measurements of iron-limited growth rates and cellular iron content (Sunda and Huntsman, 1995). For example, in *Pseudo-nitzschia* diatoms grown under low iron conditions, the Fe:C ranged from 2.8 to 3.7 µmol Fe mol C^{-1} (Marchetti *et al.*, 2006). Similar metabolic iron quotas were estimated for *T. oceanica* (at ~2 µmol Fe mol C^{-1}), whereas other members of the *Thalsasiosira* genus, such as *T. pseudonana* and *T. weissflogii* had tenfold higher requirements (Cohen *et al.*, 2018b). In the same study, the metabolic Fe:C ratio for *Corethron hystrix* was estimated at ~40 µmol Fe mol C^{-1} (Cohen *et al.*, 2018b). Diatoms with low metabolic iron quotas are remarkable for their ability to maintain uncompromised photosynthetic function and this is attributed to two main adaptations: preferred use or complete replacement of iron containing proteins with equivalents that are not dependent on iron, and re-arrangements to photosynthetic architecture.

Iron-free functional equivalents

Diatoms replace iron requiring proteins with iron-free functional equivalents

Quantitatively, iron is the most important trace metal in the photosynthetic apparatus since it is involved in PSII (which requires 2 iron atoms per subunit), the cytochrome

b$_6$*f* complex (5 iron atoms per monomer), cytochrome *c*$_6$ (1 iron atom per monomer), PSI (12 iron atoms) and ferredoxin (2 iron atoms) (Raven *et al.*, 1999; Strzepek *et al.*, 2012). Ferredoxin (PETF), is an iron-sulfur protein that is a key component of the chloroplast electron transport chain. Replacement of ferredoxin by its iso-functional carrier, the flavin-containing flavodoxin (FLDA), allows flavin rather than iron to be used for electron transport. The two proteins appear to have similar electrostatic potential profiles, although flavodoxin undergoes two successive single electron reductions, with only the second step matching ferredoxin potential (Sétif, 2001).

For species whose genomes encode both *FLDA* and *PETF*, the ratio of *PETF:FLDA* in the cell is controlled by the availability or scarcity of iron. For example, cultures of *F. cylindrus* grown at high iron concentrations produced predominantly ferredoxin, with a small amount of flavodoxin. Ferredoxin was sequentially replaced by flavodoxin in cultures grown with less iron (Pankowski and McMinn, 2009). Similarly, in *P. tricornutum*, *T. oceanica* and *Pseudo-nitzschia granii*, expression of the *FLDA* gene and FLDA protein abundance was much higher under iron limitation (La Roche *et al.*, 1995; Allen *et al.*, 2008; Lommer *et al.*, 2012; Cohen *et al.*, 2018a; Zhao *et al.*, 2018). Under iron replete conditions, PETF completely replaces FLDA, as observed in natural diatom communities (La Roche *et al.*, 1995; Erdner and Anderson, 1999; McKay *et al.*, 1999; Allen *et al.*, 2008). However, some diatoms contain multiple copies of flavodoxin, and only certain isoforms are differentially expressed in relation to iron status. Both *F. cylindrus* and *T. oceanica* have two isoforms of *FLDA*, but transcription of only one isoform is regulated by iron levels (Pankowski and McMinn, 2009; Whitney *et al.*, 2011). Whether encoding a number of flavodoxin isoforms confers an ecological advantage in environments chronically deprived of iron is an interesting question for further investigation.

The strategy of differential transcription of selected genes does not appear to be used by all diatoms. Individual diatom species may be permanently adapted to specific iron regimes in the ocean rather than maintaining transcriptional flexibility. This was supported by analysis of the *Tara* Oceans global dataset for transcriptional regulation of ferredoxin and flavodoxin across a range of algal groups. The dataset combined comprehensive bio-oceanographic genomic and transcriptomic data with iron distributions derived from two global-scale biogeochemical models (Bork *et al.*, 2015;

Carradec *et al.*, 2018). Unlike chlorophytes, haptophytes and pelagophytes, diatoms showed constitutively higher expression of flavodoxin genes than ferredoxin genes, although there was more heterogeneity in expression between species than across these four major groups (Carradec *et al.*, 2018). In *P. granii PETF* transcript abundance was appreciably lower than *FLDA* regardless of iron status (Cohen *et al.*, 2018a). The diatom *Fragilariopsis curta* appears to have lost *PETF* entirely, and transcription of the remaining *FLDA* gene is not sensitive to iron concentrations (Pankowski and McMinn, 2009). Interestingly, *PETF* is localized in the chloroplast genome in *T. pseudonana* and other diatoms but has been transferred to the nuclear genome in *T. oceanica*. Compared to its coastal relative *T. pseudonana*, the oceanic diatom *T. oceanica* is highly tolerant to iron limitation. It has been proposed that the transfer of *PETF* from the chloroplast to the nuclear genome might have contributed to the ecological success of *T. oceanica* (Lommer *et al.*, 2010). Whether or not the gene transfer described for *T. oceanica* confers a competitive advantage still needs to be assessed through experimental approaches.

Cytochrome c_6 acts as an electron carrier between the cytochrome b_6f complex and PSI. Cytochrome c_6 (encoded by *PETJ*) may be replaced with the copper-coordinating plastocyanin (encoded by *PETE*). The replacement of cytochrome c_6 with plastocyanin is rarer than the replacement of ferredoxin by flavodoxin and is presumed to have occurred via horizontal gene transfer (Strzepek and Harrison, 2004; Peers and Price, 2006). For species that encode both *PETE* and *PETJ*, differences in regulation have been noted. *P. tricornutum* and *F. cylindrus* were found to induce plastocyanin under iron limitation, to temporarily replace cytochrome c_6, and highly express the cytochrome c_6 gene under iron-replete conditions (La Roche *et al.*, 1996; Cohen *et al.*, 2018a). In a recent study Moreno et al. (2020) investigated the response of three isoforms of *PETE* to iron in this species and found that only two were significantly overrepresented under low Fe conditions. In *T. oceanica,* which encodes *PETE* together with two functional copies of *PETJ*, relative expression of plastocyanin was sensitive to iron status, while two cytochrome c_6 genes were weakly but constitutively expressed (Lommer *et al.*, 2012). Similarly, in *P. granii* transcripts for *PETJ* were weakly abundant regardless of iron status while *PETE* was highly abundant under iron-replete conditions (Cohen *et al.*, 2018a).

Only a few studies have addressed the question of what the consequences of using plastocyanin are to the copper requirements of cells. For example, the green alga *C. reinhardtii* only switches to using cytochrome c_6 instead of plastocyanin under copper limiting conditions (Merchant *et al.*, 1991). In *T. oceanica* copper deficiency inhibited electron transport regardless of iron status, implying plastocyanin expression was not controlled by iron concentrations (Peers and Price, 2006).

Photosynthesis architecture

Low iron quota diatoms lower the ratio of photosystem I to photosystem II without compromising photosynthetic output

The PSI complex has the highest iron demand of the light-dependent reactions of photosynthesis. The optimised ratio of PSI: PSII in land plants is approximately 1:1. Comparing the coastal diatom *T. weissflogii* to the open ocean *T. oceanica*, Strzepek and Harrison (2004) found that this ratio was significantly lower in the latter. This led to the hypothesis that diatoms, and in particular those adapted to chronically iron-starved conditions, have streamlined their photosynthetic architecture, minimizing the iron quota necessary for growth. *T. weissflogii* cells contained twice as much PSII compared with PSI, and *T. oceanica* had reduced PSI demands even further, with fivefold lower PSI, that is a ratio of 1:10 of PSI : PSII (Strzepek and Harrison, 2004). In iron-replete media, the two diatoms grew at comparable rates. However, *T. oceanica* maintained high growth rates (about 80% of that in iron-replete conditions) in low-iron media that restricted the growth of *T. weissflogii*, to about 20% of iron-replete growth rates. Two studies confirmed similar streamlining in *F. kerguelensis* and *P. granii,* polar diatoms that are abundant in the Southern Ocean and Arctic Ocean, respectively (Cohen *et al.*, 2018a; Moreno et al., 2020). In *P. granii* the gene encoding PSI subunit IV (*PSAE*) was almost four times more highly expressed under iron-replete conditions, and protein levels were more abundant by 35-fold compared to conditions of iron limitation (Cohen *et al.*, 2018a).

Low iron quota diatoms such as *T. oceanica* and *P. granii* are not unique among algae and cyanobacteria to undergo substitutions and rearrangements in their photosynthetic architecture in order to conserve iron. These traits and plasticity in iron use are

widespread (Blaby-Haas and Merchant, 2013; Raven, 2013; Scheiber *et al.*, 2019). However, in an experimental comparison of 29 species of eukaryotic algae and cyanobacteria, Quigg and colleagues demonstrated that algae that contain plastids of a secondary endosymbiotic origin, the "red plastid" lineage, which includes the dinoflagellates, haptophytes and chrysophytes as well as diatoms, had lower stoichiometric quantities of iron per cell, compared with the green lineage of algae and cyanobacteria (Quigg *et al.*, 2003; Quigg *et al.*, 2011). The results from these analyses are difficult to square with measurements of metabolic iron quotas, since the former are carried out under replete Fe conditions. Nevertheless, they suggest that members of the red plastid lineage are capitalising on more than rearrangements to their photosynthetic architecture for efficient iron use. Studies that look into the adaptations of cell structure, such as the metabolic coupling of plastids and mitochondria (*e.g.,* see some pioneering work by Bailleul *et al.*, 2015) may shed further light on this question.

1.3.3 Iron storage in diatoms

Diatoms are able to store iron and regulate intracellular concentrations efficiently

It has been observed that *in situ* iron fertilisation, whether artificial or natural, results in blooms dominated by large diatoms, which are often rare in the standing microalgal community (De Baar *et al.*, 2005). These include the chain-forming members of the genera *Fragilariopsis*, *Pseudo-nitzschia* and *Chaetoceros*, in particular.

It is likely that species that respond to iron stimulation by forming blooms are not only efficient at iron uptake, but also in iron homeostasis and long-term storage. One way to assess the capacity for long-term iron storage is to measure the number of divisions that a cell is capable of when ambient iron concentrations drop back to growth-limiting conditions. This can be distinguished from short-term iron homeostasis, which is associated with mechanisms that act on a diurnal scale, where the requirement for iron depends on photosynthetic activity. An alternative is to measure the cell's metabolic iron quota and subtract this from total cellular iron carried by a cell under iron replete conditions. This has been termed "luxury iron" quota (Sunda and Huntsman, 1997; Marchetti *et al.*, 2009). For example, within the centric *Thalassiosira* genus, cellular Fe:C of coastal species reached values that were 20-30 times higher than those needed

for maximum growth; by contrast, “plateau” iron concentrations in oceanic species were only two to three times higher than maximum needed amounts (Sunda and Huntsman, 1995), indicating that coastal variants had higher capacity to accumulate iron when iron replete.

Towards a functional characterization of diatom ferritin

There are two mechanisms that have been proposed for storage of iron inside diatom cells, sequestration into the mineral core of ferritin proteins and/or vacuolar storage. Ferritins (FTN) are found in all domains of life including animals, plants and microorganisms (Figure 1.12), and are designed to accommodate large amounts of iron (Theil, 1987). Structurally, ferritins are large protein cages formed by arrays of self-assembling helices with nanocavities (5-8 nm), that catalytically couple iron and oxygen at protein sites for precursors of the cavity mineral. The mineral core of a single protein can store up to 4,500 iron atoms (Liu and Theil, 2005). Iron is reversibly released from the core by reduction.

In laboratory studies, it was shown that *P. granii,* which encodes *FTN*, was able to undergo several more cell divisions in the absence of iron than the comparably sized, oceanic centric diatom *T. oceanica*, which lacks the *FTN* gene, supporting the hypothesis that *FTN* functions in the long-term storage of iron (Marchetti *et al.*, 2009). A corollary of the hypothesis is that *FTN* transcript abundance should increase with iron concentrations. However, this has not been universally upheld in laboratory studies. While *P. granii* and *Thalassiosira* sp. strain NH16 have been observed to increase ferritin gene expression under high iron concentrations, *Amphora coffeaeformis* exhibited minimal iron storage capacities, and contained two distinct ferritin genes, one of which increased in expression under iron limitation and the second showed no variation with cellular iron status (Cohen *et al.*, 2018b). Furthermore, a recent study investigated the community-level response of open ocean plankton ecosystems to iron availability and found that, with the exception of *Pseudo-nitzschia* species, no clear pattern in ferritin gene abundance or expression and estimated iron levels could be observed, suggesting that iron storage may not be the main function of ferritin in most eukaryotic marine phytoplankton (Caputi *et al.*, 2019). Together these results suggest

that ferritin in diatoms may have evolved to serve multiple functional roles.

A proposed alternative role for ferritin in diatoms is as an iron oxidation enzyme rather than as a long-term iron storage protein. Working with *P. multiseries,* Pfaffen and colleagues showed that FTN oxidizes Fe^{2+} at its ferroxidase centres rapidly but forms iron mineral only slowly (Pfaffen *et al.*, 2013). In a later study functional mutagenesis experiments were performed showing that the protein is biochemically optimized for initial Fe^{2+} oxidation but not for mineralization. The authors argued that its primary function therefore could not be in long term iron storage, but rather in iron homeostasis (Pfaffen *et al.*, 2015). This is the situation in higher plants, where experiments on *Arabidopsis thaliana* demonstrated that *FTN* is regulated by the circadian clock cycle (Duc *et al.*, 2009) and functions to buffer the iron released by degradation from PSI, a protective role in response to photo-oxidative stress (Rossel *et al.*, 2002; Murgia *et al.*, 2007). Similarly, in *C. reinhardtii* FTN was shown to be required during high ambient iron availability or cellular degradation of iron-containing proteins and protein complexes such as ferredoxin during cellular acclimation to low iron conditions, indicative of function in iron homeostasis (La Fontaine *et al.*, 2002; Busch *et al.*, 2008; Long and Merchant, 2008). Finally, in the green picoalga *Ostreococcus tauri* mutants lacking FTN were less tolerant of low iron conditions that induced higher recycling of iron within the cell than wild-type, further underscoring the importance of this protein in homeostasis. In the same work the authors proposed that nitrate reductase functioned as an iron storage protein in this picoalga (Botebol *et al.*, 2015).

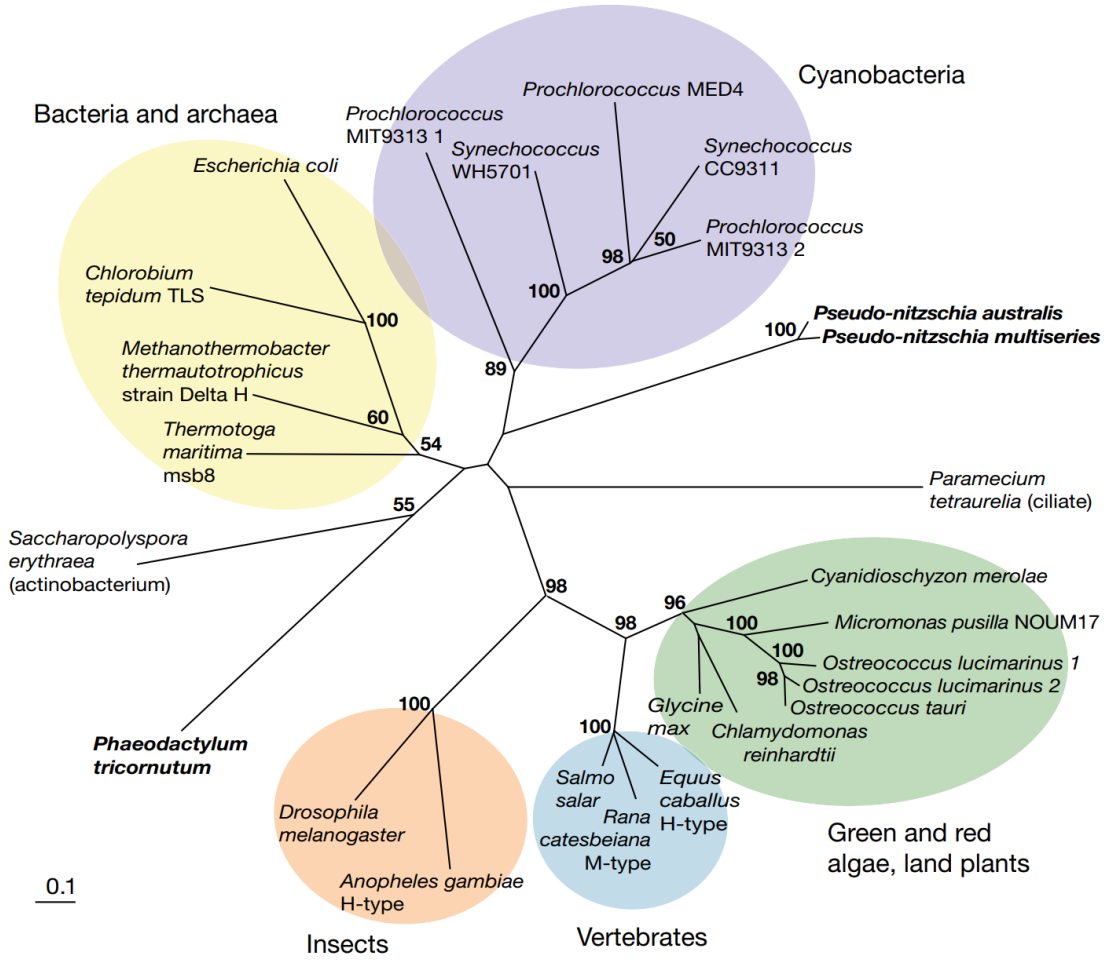


Figure 1.12 Ferritin phylogenetic tree. Bootstrap consensus tree (100 replicates) showing the evolutionary relatedness of maxi-ferritins from 26 taxa inferred using a protein distance model. Bootstrap values greater than 50 are indicated at the branch points. Diatoms are in bold. Figure is taken from Marchetti *et al.*, (2009).

Vacuolar iron storage

An alternative mechanism for long term iron storage that has been proposed is vacuolar. First identified in yeast, iron loading into and release out of the vacuole was shown to be mediated by natural resistance associated macrophage proteins (NRAMPs) in *A. thaliana* (Curie *et al.*, 2000; Lanquar *et al.*, 2005). Furthermore, some NRAMP proteins were shown to be functional equivalents of yeast FET proteins, and enabled transport of iron into the cell (Curie *et al.*, 2000). In *T. pseudonana* a vacuolar storage mechanism was proposed, when it was observed that NRAMP was dramatically regulated by

intracellular iron concentrations (Kustka *et al.*, 2007). However, it is important to note that vacuolar localisation of NRAMP was proposed but not experimentally validated. In experiments using synchrotron X-ray fluorescence element mapping, intracellular pockets of iron, indicative of storage features, were identified in both *T. pseudonana* and *T. weissflogii* (Nuester *et al.*, 2012).

A recent study investigated the responses of natural communities sampled across a gradient of iron concentrations to on-deck iron stimulation. Metatranscriptome data from incubations was used to assign relative expression of *FTN* and *NRAMP* genes onto three main diatom genera, *Chaetoceros*, *Pseudo-nitzschia* and *Thalassiosira* (Lampe *et al.,* 2018). The study found that *Pseudo-nitzschia* diatoms were unique utilizers of *FTN* (upregulating its expression when iron was supplied), whereas the response from *Thalassiosira* species suggested vacuolar storage. Transcripts for neither *FTN* nor *NRAMP* were abundant in *Chaetoceros* although iron quotas in *Chaetoceros* were often similar to those in *Pseudo-nitzschia* (Lampe *et al.*, 2018). The authors proposed that *Chaetoceros* may have an alternative divalent metal transporter that takes on the function of vacuolar iron transport in these species, a protein belonging to the ZIP family (Allen *et al.*, 2008). ZIP proteins have been shown to facilitate passive metal transport, including that of ferrous iron, in a range of species (Eide, 2005).

1.4 Outline of the thesis

Considering all phytoplankton as a group, iron bioavailability has dramatically decreased in the contemporary ocean from previous maxima. Yet diatoms arose as a young evolutionary offshoot when iron was already deficient in the vast ocean. They successfully survived low iron and outcompeted other species when iron became available, which was confirmed by the mesoscale iron fertilization experiments. In this dissertation the question posed is, what are the strategies of diatoms for obtaining iron and using it efficiently inside the cell? To answer this question, first I reviewed the studies on iron metabolism in diatoms, combining different bioinformatic analyses. Next, I conducted comparative physiological experiments on knockout strains.

In chapter 2, I present the results of a review which was published in the Journal of Experimental Botany (Gao *et al*., 2021), where together with colleagues I summarized the iron metabolism strategies in diatoms relying on the transcriptional and physiological evidence from the literature.

In this chapter I further searched for iron-related genes through bioinformatic analyses and summarized it into a heatmap showing the distribution of genes in different species. The sequences of FTN of *P. tricornutum* and other diatom species were compared by alignment and phylogenetic tree. Overall, this chapter provides novel insights into the topic of iron metabolism in diatoms.

In chapter 3, aiming to investigate the localization of ISIPs and FTN, I tagged FTN to fluorescent proteins (FP) and visualized transgenic diatom cells by confocal microscopy. Both ISIP3 and FTN localized to the vicinity of the chloroplast but with different patterns.

In chapter 4, I further investigate the function of ISIP3 and FTN in *P. tricornutum*. I generated knockout (KO) mutants and compared their growth curves under different growth conditions. Absence of the gene had an effect on the growth of cells under different conditions, indicating involvement in specific aspects of iron physiology.

In chapter 5, conclusions and perspectives were synthesized.

Chapter 2: Bioinformatics analysis of key components of iron metabolism across the diatoms

Abstract

The key components of iron metabolism strategies summarized in chapter 1 are here analysed bioinformatically for 82 diatom species. Taking advantage of the availability of genomic and transcriptomic data sets, I investigated the distribution of 14 iron related genes across these 82 diatom species, representatives of 49 diatom genera. Based on the retrieved homologs, the proportion of pennate and centric diatoms encoding PETF vs PETJ, FLDA vs PETE were compared. The ability to synthesise flavodoxin is widespread amongst diatoms. The replacement of cytochrome c_6 with plastocyanin is rarer than the replacement of ferredoxin by flavodoxin and is presumed to have occurred via horizontal gene transfer. Furthermore, a phylogenetic tree of FTN in diatoms was built and updated what was previously reported in the literature. Intriguingly, the FTN of *P. tricornutum* was found to fall within a new clade disparate from what has previously been described. Sequence features of *P. tricornutum* ferritin were compared with the proteins in *Pseudo-nitzschia* and *Thalassiosira*.

2.1 Introduction

Bioinformatics is an interdisciplinary field that combines biology, information engineering, computer science, mathematics and statistics. The availability of genomic, transcriptomic and proteomic data sets has greatly facilitated genome annotation efforts and therefore gene discovery in diatoms (see section 1.2.2).

The application of bioinformatics analysis has been widely used in all aspects of biology, on various research scales, including to understand oceanic community structure, species diversity, genetic interspecies heritability and variation, and functional gene screening. For example, diatom DNA metabarcoding applied to large scale monitoring networks for ecological assessment has become a hot topic in recent

years (Bailet *et al.*, 2020; Rivera *et al.*, 2020; Kang *et al.*, 2021). Additionally, bioinformatic inspections, led by pioneering studies on the evolutionary history of diatoms (Cavalier-Smith, 2003; Strzepek and Harrison, 2004), have revealed more and more evidence on the endosymbiotic events and horizontal gene transfer (HGT) events in diatoms, identifying the gene involved in these processes (Brembu *et al.*, 2014; Vancaester *et al.*, 2020; Dorrell *et al.*, 2021). Diatoms have recruited an enormous number of genes by horizontal gene transfer from prokaryotes, and many are likely to provide novel possibilities for metabolite management and for the perception of environmental signals (Bowler *et al.*, 2008; Dorrell *et al.*, 2021).

The Basic Local Alignment Search Tool (BLAST) is widely employed (http://www.ncbi.nlm.nih.gov/blast). The programme finds regions of local similarity between sequences, compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families (Johnson *et al.*, 2008; Madden, 2013).

Hidden Markov Model (HMM) is a powerful statistical tool for modelling, known for their wide range of applications to physics, chemistry, economics, signal processing, information theory, pattern recognition - such as speech, handwriting, gesture recognition and bioinformatics (Eddy, 1998; Liu and Cheng, 2003; Mamon and Elliott, 2007). HMMER is a developed tool using profile HMMs for searching sequence databases for sequence homologs, and for making sequence alignments. HMMER can work with query sequences, just like BLAST, to detect remote homologs as sensitively as possible. HMMER is more often used together with a profile database, such as Pfam.

To screen the distribution of iron metabolism related genes in diatom genomes and transcriptomes, a composite database combining MMETSP and Genoscope sequences were assembled and rigorously decontaminated (see further details in the Methods section 2.4).

Objectives of bioinformatic analysis of diatom genomes and transcriptomes

In this chapter I focus on the iron metabolism strategies of diatoms summarized in section 1.3, and the schematic diagram of iron-related genes in diatoms shown in Figure 1.11. I present the results that took advantage of the capabilities of bioinformatic

analyses. The objective was to investigate the distribution of iron-related genes in diatoms by retrieving homologs of genes and transcripts from available transcriptome and genome datasets. Additionally, I present an in-depth comparison of ferritin across the diatom phylogenetic tree. These analyses go beyond previously published studies by incorporating larger datasets. I identify a third clade of diatom ferritins that was not previously subclassified. The ferritin analysis stretches beyond bioinformatics, as in-depth discussion of putative functional domains and residues of the *FTN* gene is presented. The data presented in this chapter underwent peer review and is published in the Journal of Experimental Botany (Gao *et al.*, 2021).

2.2 Results

Available diatom genomes and transcriptomes for iron-related genes were searched with a profile that matched a constructed Hidden Markov Model (HMM). Results are presented as a heatmap in Figure 2.1 where the detection of homologs of iron-related genes in 82 diatom species are shown.

2.2.1 Distribution of iron acquisition related genes in diatoms

Two main strategies of iron acquisition are known in unicellular eukaryotes. One is reductive uptake, in which ferric iron is displaced from its ligand by a ferrireductase (FRE) and then re-oxidised by the multicopper oxidase (FET) and channelled through the ferric iron permease (FTR). The reductive uptake is well characterized. However, the second strategy, nonreductive uptake, including two uptake pathways, are identified in recent years. ISIP2a binds ferric iron directly on the cell surface and internalizes by endocytosis. ISIP1 and FBP1 are involved in endocytosis-mediated uptake of siderophore (Kazamia *et al.,* 2018; Coale *et al.,* 2019). Searching through the available genomes and transcriptomes, genes related to reductive uptake, in general, are more ubiquitous than genes involved in nonreductive uptake in diatoms. 65%, 71% and 99% species encode *FET*, *FTR* and *FRE* gene respectively, in parallel, only 21%, 41% and 91% species encode *FBP1*, *ISIP1* and *ISIP2a* gene (Figure 2.1).

Diatom cells have been reported to exhibit ferric reductase activity (Maldonado and

Price, 2001; Kazamia *et al.*, 2018; Coale *et al.*, 2019). Six genes in *P. tricornutum* have been putatively annotated as ferric reductases, but only two (*FRE1* and *FRE2*) encode two domains indicative of ferric reductase function. I built an HMM for putative diatom ferric reductases, FRE, based on reference genes from *P. tricornutum*, and searched the available diatom genomes and transcriptomes. The results show that all species appear to encode ferric reductases, around 2 per species on average (Figure 2.1). *Fragilariopsis kerguelensis* and *F. cylindrus* abundant diatoms in the Southern Ocean, stand out for encoding 8 and 4 distinct peptides annotated as ferric reductases, respectively. The coastal species *Ditylum brightwellii* encodes 8 *FRE* genes. Furthermore, bioinformatic analysis of transcriptome datasets confirm functional analogues of the yeast-like iron acquisition machinery ('ferrireductase (FRE), multicopper oxidase (FET), iron permease (FTR)' system) in the model green alga *Chlamydomonas reinhardtii* and across a range of diatom species, including all members of the *Pseudo-nitzschia* and *Fragilariopsis* genera (Figure 2.1 and also reviewed in Groussman *et al.*, 2015). However, the question of whether this reductive system acts in consort on the cell surfaces of diatom cells requires experimental verification. Therefore, in Figure 2.1 I have indicated the FRE genes as acting both in 'reductive' and 'non-reductive' uptake systems.

As to the non-reductive uptake pathway, FBP1 was verified previously as putative siderophore binding protein, which in bacteria binds the hydroxamate siderophore ferrichrome, was indeed required for siderophore assimilation (Coale *et al.*, 2019). ISIP1 was identified as being necessary for the endocytosis and siderophore assimilation process (Kazamia *et al.*, 2018). Diatoms exhibit species-specific recognition of different siderophores. ISIP2a was able to bind ferrous iron directly on the cell surface and was internalized by endocytosis. In our *in silico* searches, 21% of diatom genomes and transcriptomes displayed evidence of a *FBP1* gene, and 41% diatoms for ISIP1 (Figure 2.1). Although it is important to note that since ISIP1 is highly sensitive to iron status in the cell, it may have been missed from transcriptomes collected from diatoms grown in replete iron. Hits for *ISIP2A*-like genes were retrieved in most diatoms, with only 8.5% absence in 82 species, and some notable exceptions in some species of *Chaetoceros*. Further experiments are required to understand what determines substrate specificity in these species.

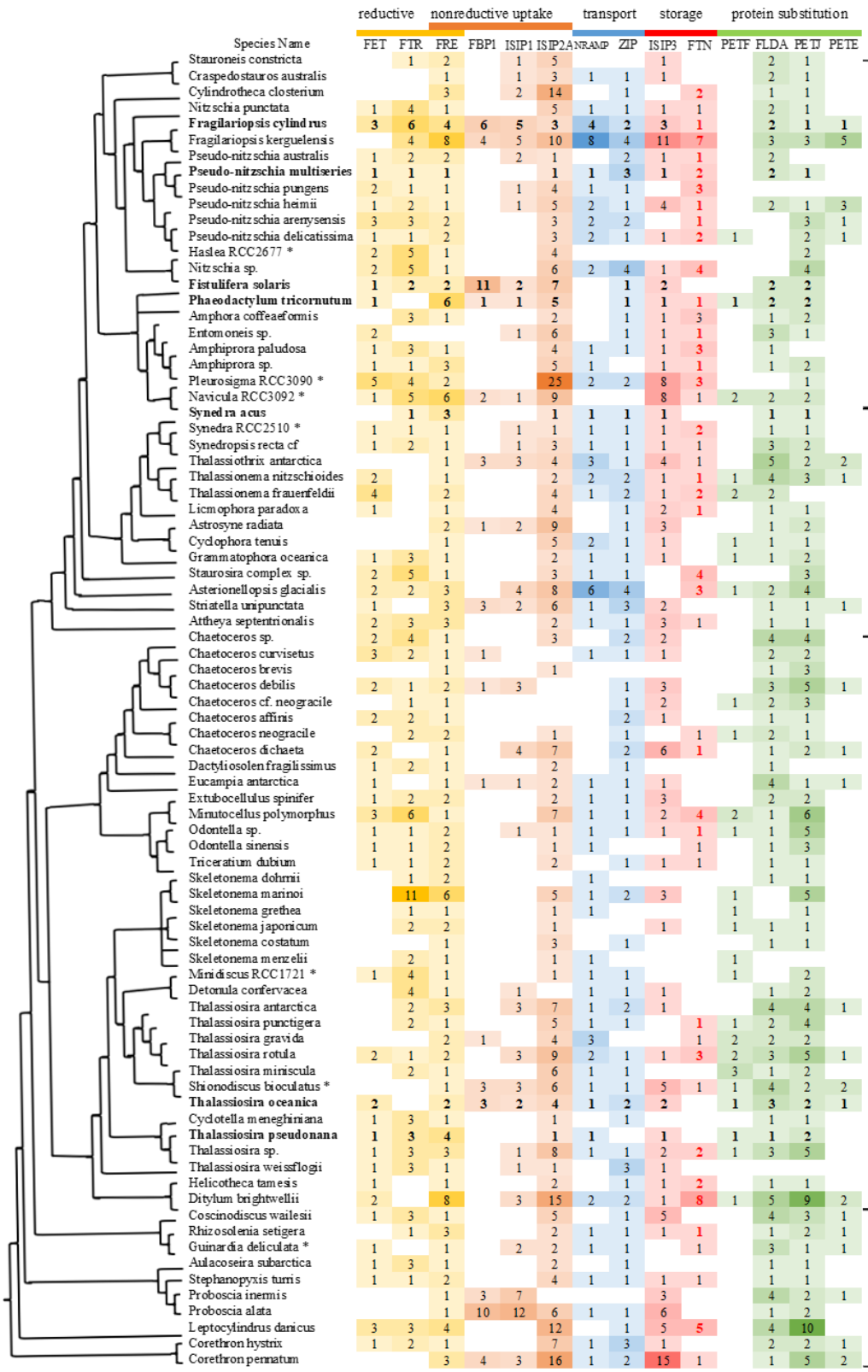


Figure 2.1 Iron-related genes in diatom transcriptomes.

The diatom species with available genomes are indicated in bold, with transcriptomes are indicated by asterisks. Where no hits were detected, the grid is blank, whereas the

shade of the colour and the number within indicates the number of distinct transcripts encoding each gene (but not the expression levels). Ferritin sequences predicted to target to the chloroplast are presented in red.

Abbreviations: ferroxidase, FET; iron (III) permease, FTR; ferric reductase, FRE; ferrichrome-binding protein, FBP; Iron-starvation induced protein, ISIP; ferritin, FTN; natural resistance-associated macrophage protein, NRAMP; Zinc transporter, ZIP; ferredoxin, PETF; flavodoxin, FLDA; cytochrome c_6, PETJ; plastocyanin, PETE.

2.2.2 Distribution of iron homeostasis related genes in diatoms

Diatoms with low metabolic iron quotas are remarkable for their ability to maintain uncompromised photosynthetic function and this is attributed to adaptation of preferred use or complete replacement of iron containing proteins with equivalents that are not dependent on iron. PETF is an iron-sulfur protein, a key component in chloroplast electron transport chain, and can be replaced by its iso-functional carrier, flavin-containing FLDA. For species that encode both *FLDA* and *PETF*, the ratio of *PETF:FLDA* in the cell is controlled by the availability or scarcity of iron (La Roche *et al.*, 1995; Allen *et al.*, 2008; Lommer *et al.*, 2012; Cohen *et al.*, 2018a; Zhao *et al.*, 2018). Under iron replete conditions, PETF completely replaces FLDA, as observed in natural diatom communities (La Roche *et al.*, 1995; Erdner and Anderson, 1999; McKay *et al.*, 1999; Allen *et al.*, 2008). Within the diatoms with sequence information, 20 out of 82 species encode both *PETF* and *FLDA* (Figure 2.1). There appears to be no phylogenetic relationship between species that encode *PETF*, implying that it was present in the diatom ancestor and that its absence from some species is likely due to gene loss (Figures. 2.1, 2.2A). The ability to synthesise flavodoxin is widespread amongst diatoms; I found *FLDA* in 70 out of 82 species (Figure 2.1). Since no photosynthetic species are known to survive without either *PETF* or *FLDA*, in transcriptomes where neither transcript was detected, such as in *Pseudo-nitzschia pungens*, I assume that this was either due to incomplete sequencing, or due to the stringency of cut-off parameters chosen for detection in our analysis. That *PETF* is often chloroplast-encoded suggests that analyses of transcriptomes such as the one presented in Figure 2.1 may be methodologically biased towards a lack of detection, as

RNA harvesting and processing select for nuclear encoded transcripts.

Cytochrome c_6 (encoded by *PETJ*) acts as an electron carrier, and may be replaced with the copper-coordinating plastocyanin (encoded by *PETE*). The replacement of cytochrome c_6 with plastocyanin is rarer than the replacement of ferredoxin by flavodoxin and is presumed to have occurred via horizontal gene transfer (Strzepek and Harrison, 2004; Peers and Price, 2006). Only 18% of centric and 9% of pennate species show evidence for the presence of a *PETE* gene (Figures. 2.1, 2.2B). For species that encode both *PETE* and *PETJ*, differences in regulation have been noted. *P. tricornutum* and *F. cylindrus* were found to induce plastocyanin under iron limitation, to temporarily replace cytochrome c_6, and highly express the cytochrome c_6 gene under iron-replete conditions (La Roche *et al.*, 1996; Cohen *et al.*, 2018a). *F. kerguelensis* is an interesting candidate for investigation since it contains multiple isoforms of plastocyanin (5 genes detected using our methods, Figure 2.1).

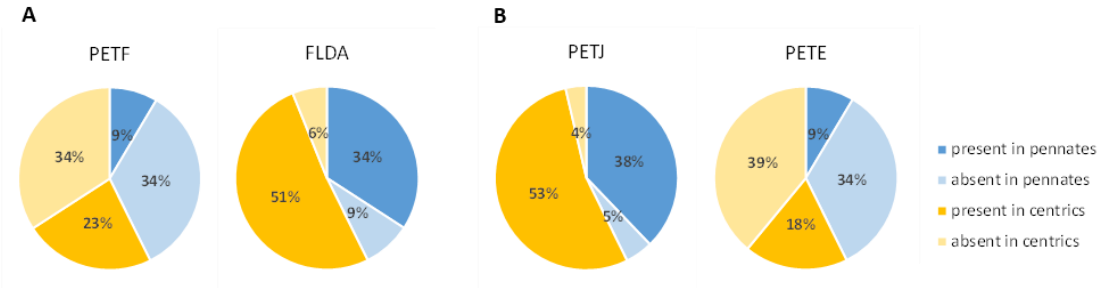


Figure 2.2 Genes encoding iron switching proteins in diatom transcriptomes.

Both ferredoxin (PETF) and cytochrome c_6 (PETJ) are iron-containing enzymes. As some diatoms are capable of replacing them with iron-free equivalents, flavodoxin (FLDA) and plastocyanin (PETE), respectively, the relative proportions of transcripts encoding these proteins may indicate the dependence of diatoms on these proteins. The pie charts indicate presence and absence of *PETF* versus *FLDA*, and *PETJ* versus *PETE* in pennate and centric diatoms in our database of 82 species. Here I show the numbers of distinct genes detected, and not transcript abundance. It is important to note that the presence/absence of genes in a transcriptome may not be a true reflection of copy numbers in genomes either because it was generated from iron replete cultures or was not sequenced to sufficient depth.

2.2.3 Distribution of iron storage related genes in diatoms

It has been observed that *in situ* iron stimulation, whether artificial or natural, results in blooms dominated by large diatoms, which are often rare in the standing microalgal community (De Baar *et al.*, 2005). These include the chain-forming members of the genera *Fragilariopsis*, *Pseudo-nitzschia* and *Chaetoceros*, in particular. Comparing their transcriptomes for the analysed genes (Figure 2.1), there is no pattern that distinguishes them readily from other diatoms, as traits appear species-specific. However, it is important to note that these transcriptomes capture the behaviour of species under laboratory conditions and may not be indicative of their behaviour during a diatom bloom in the wild. The *F. kerguelensis* transcriptome and *F. cylindrus* genome encode a more comprehensive portfolio of iron sensitive genes, with multiple isoforms of each gene compared with other species. This is intriguing since as model polar species *Fragilariopsis* representatives are considered to be particularly adapted to fluctuating environmental conditions and life in sea ice, which is not "iron-limited" (Mock *et al.*, 2017). The data suggest that *Fragilariopsis* species are interesting candidates for iron adaptation studies as well.

There are two mechanisms that have been proposed for storage of iron inside diatom cells, sequestration into the mineral core of ferritin proteins and/or vacuolar storage. FTN are found in all domains of life and identified as iron storage protein in bloom-forming diatoms species (Marchetti *et al.*, 2009). I retrieved putative transcripts for *FTN* in diatom transcriptomes and genomes by conducting a HMMER search using methods described in section 2.4. I found that around 74% of pennate diatoms contain at least one *FTN* homolog compared to 41% of centric diatoms. Within the centric diatoms the trend skewed towards the polar centrics, of which 51% miss the annotation (Figure 2.3). The apparent loss of *FTN* in the polar centrics is intriguing and requires further investigation. Looking at the transcriptomes of the three genera of frequently blooming diatoms, *Fragilariopsis* and *Pseudo-nitzchia* species appear to be rich in putative *FTN* genes (Figure 2.1). In contrast, our search for *FTN* retrieved hits for only a quarter of the sampled *Chaetoceros* strains (2 out of 8). There were notable absences amongst *Thalassiosira* species (5 species out of 9 do not encode it), including *T. oceanica* and *T. pseudonana*. Furthermore, three distinct *FTN* transcripts were detected

in *Amphora coffeaeformis* (Figure 2.1).

Another notable candidate involved in iron storage is ISIP3, the most abundant iron starvation induced protein, typified by the presence of a ferritin-like domain. Using our methods, I found transcripts for *ISIP3* across most diatoms, with some notable absences in the coastal *Skeletonema* and *Chaetocerus* genera and an overabundance of putative hits in *Fragilariopsis* (Figure 2.1).

The alternative mechanism for long term iron storage that has been proposed is vacuolar. First, NRAMPs was shown to mediate iron loading into and release out of the vacuole in yeast (Curie *et al.*, 2000; Lanquar *et al.*, 2005). Additionally, ZIP proteins have been shown to facilitate passive metal transport, including ferrous iron (Eide, 2005). I compared the portfolios and *ZIP*, *NRAMP* and *FTN* genes across diatoms using all published diatom transcriptomes and genomes, by constructing Hidden Markov Models for representative genes and scanning the dataset using HMMER, summarised in Figure 2.1. This is an amalgamated view, since I did not refine our search for subclasses of NRAMP or ZIP proteins (*e.g.,* the specificity of these genes for iron transport is not confirmed). I found that all diatoms have genes belonging to the *ZIP* family, and that the majority also encode *NRAMP* (modelled on *F. cylindrus* NRAMP protein). Notable exceptions are most members of the *Chaetoceros* genus, which miss *NRAMP* annotation. *NRAMP* genes appear to be less abundant amongst centric species than pennates, but no robust phylogenetic relationship was observed.

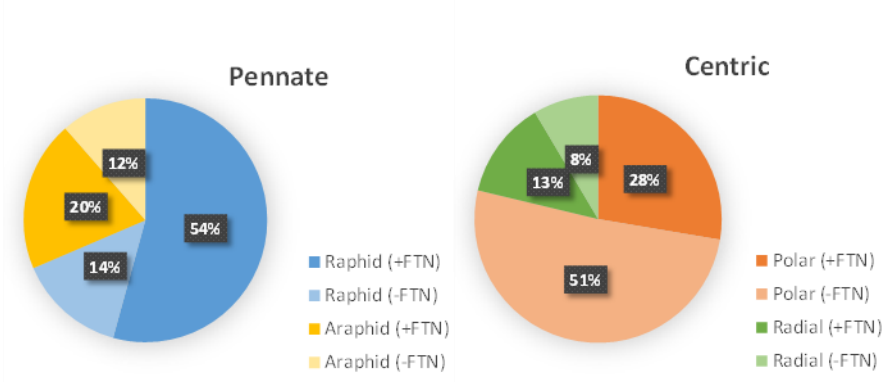


Figure 2.3 Genes encoding ferritins in diatom transcriptomes.

Homologs of *FTN* were retrieved using HMMER search within available diatom genomes and transcriptomes (see caption to Figure 2.1 for details). Pie charts

summarise the presence and absence of *FTN* in polar versus radial centrics, and raphid versus araphid pennate diatoms. It is important to note that the presence/absence of genes in a transcriptome may not be a true reflection of copy numbers in genomes either because it was generated from iron replete cultures or was not sequenced to sufficient depth.

2.2.4 FTN phylogenetic tree

A closer look at diatom FTN phylogeny, which I updated using available diatom transcriptomes (Figure 2.4), reveals that the ferritins within the diatoms resolve into several clades. The phylogeny matches that previously reported by others (*e.g.,* Marchetti *et al.*, 2009; Groussman *et al.*, 2015), but adds more species to the tree.

There are two main FTN clades, one which contains most members of the *Thalassiosira*, *Pseudo-nitzschia* and *Fragilariopsis* genera*,* although phylogenetic relationships are not robustly upheld. The second clade contains members of fourteen genera, including *Cylindrotheca* and *Leptocylindrus.* The species *Heliotheca tamensis* CCMP826*, Amphiprora paludos*a CCMP125, *Thalassionema fraundfeldii* CCMP1798, *Staurosira* complex sp. CCMP2646 and *Pleurosigma* sp. RCC3090 have annotated ferritins belonging to both clades. To my knowledge, there are no described functional differences between ferritins that belong to the two main groups. Additionally, I found a third group of putative ferritins (which I annotate as FTN clade III in Figure 2.4) with weaker bootstrap support that includes genes from *P. tricornutum*, *Staurosira complex sp*., *Guinardia deliculata*, *Chaetoceros neogracile* and *F. kerguelensis*.

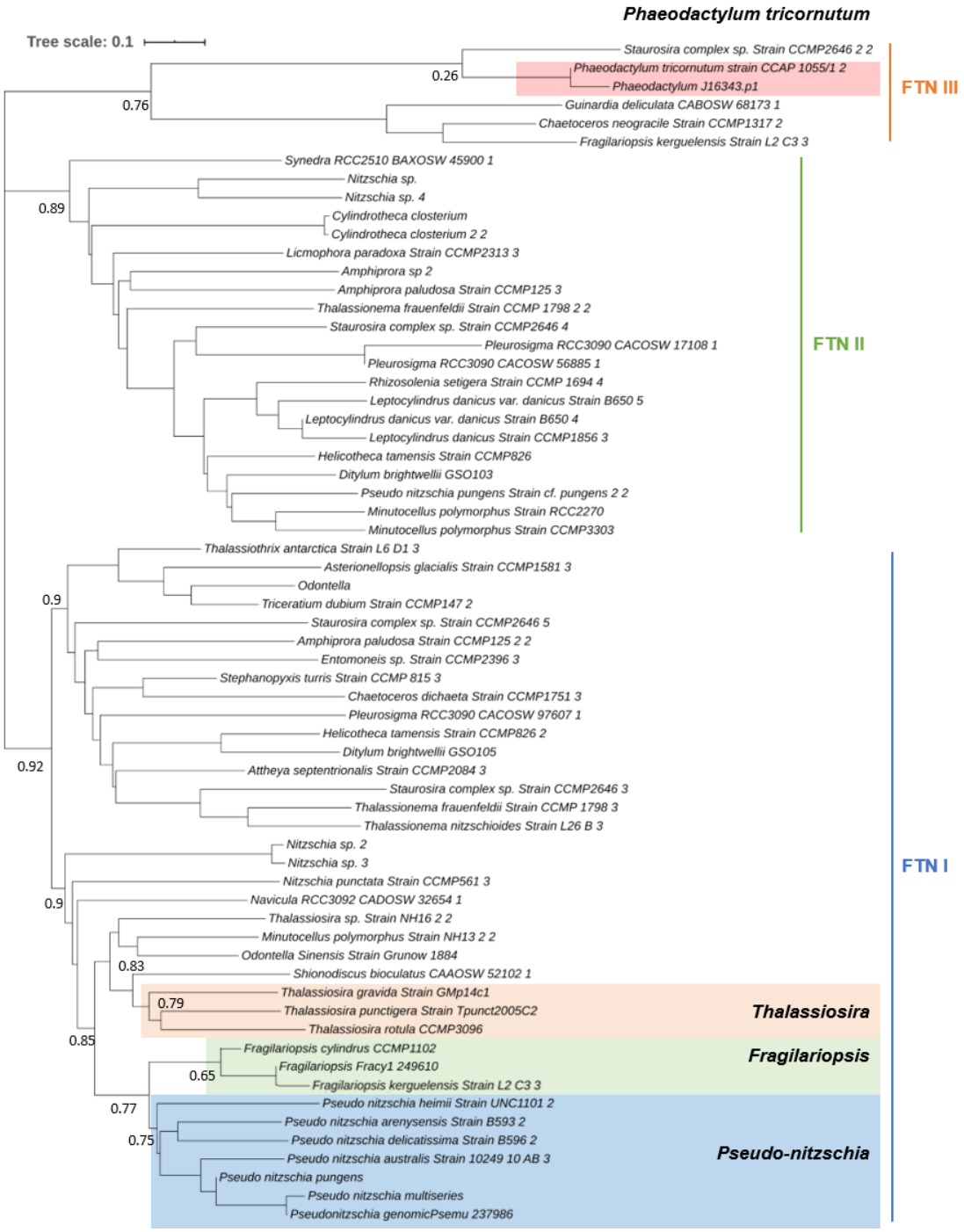


Figure 2.4 Phylogenetic tree of ferritins in diatoms. Numbers beside branches indicate RaxML bootstrap coefficients.

FTN sequence alignment

I compared the protein sequence of representative FTN from the three groups by alignment (Figure 2.5). Iron binding sites in FTN of *P. multiseries*, including several

glutamic acid residues (E15, E47/48, E94), one glutamine residue (Q127) and the histidine residue (H51) that were identified by Pfaffen *et al*. (2013), are conserved across the three different groups. However, the glutamic acid residue E44 conserved in FTNI/II is replaced with histidine in FTN III. These residues are involved in iron binding (Figure 2.5 red arrows).

Studies on the function and structure of ferritin from the frog *Rana catesbeiana* have identified conserved residue pairs essential for iron release (Jin *et al.*, 2001). These are indicated by blue arrows in Figure 2.5 and represent two amino acid pairs that are adjacent in 3D space (R62/D113 and L101/L125). R62/D113 is conserved in both FTN I and FTN III but not FTN II (which shows significant variation in these positions). Leucine is conserved in position 101 in FTN I and FTN II but not FTN III, while Leucine at position 125 is not conserved in any diatoms apart from two *Pseudo-nitzschia* species and *Synedra* sp. RCC2510. The most common replacement for leucine in this position is to a methionine, as observed in *F. cylindrus* and *Thalassiosira* sp. strain NH16.

It is possible that these modifications provide a first glimpse into functional differences in FTN between these species, however this requires experimental verification. There is evidence from on-board experiments to suggest that *Pseudo-nitzschia* ferritins serve a long-term storage role in cells, which is not the case in *Thalassiosira* and *Corethron*, where a homeostasis role is preferred as an explanation (Lampe *et al*., 2018). One structural change that may explain this, is the presence of glutamic acid at position 130 in *P. multiseries* ferritin, conserved in FTN III clades, is absent from other FTN I and II species (Figure 2.5 red arrows).

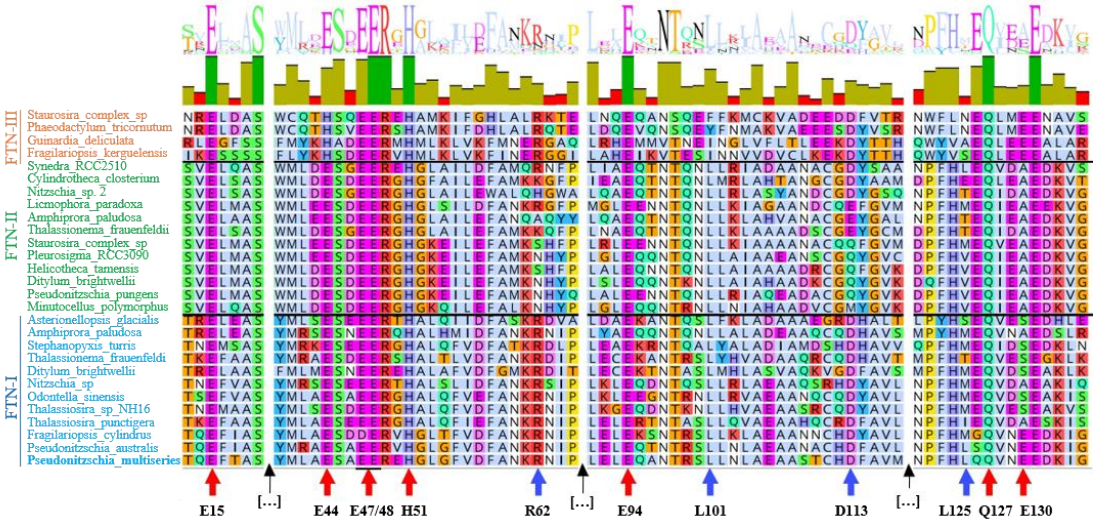


Figure 2.5 Protein sequence alignment of representative diatom ferritins.

The glutamic acid residues (E15, E47/48, E94), one glutamine residue (Q127) and the histidine residue H51 are conserved across the three different groups (red arrows). The amino acid pairs R62/D113 and L101/L125 are adjacent in 3D space (blue arrows). Black arrows indicate breaks in the sequence.

2.2.5 FTN localization prediction

The function of FTN protein in diatoms is quite controversial. It was proposed that iron storage may not be the main function of ferritin in most eukaryotic marine phytoplankton. With localization of *in vivo* FTN protein unknown in diatom cells, based on signal peptide and target peptide, localization prediction was employed. In Figure 2.1, ferritin sequences predicted to target to chloroplast are presented in red. Within the 55% species that encode a *FTN* gene, in 32 out of 45 species, FTN is predicted to be targeted to the chloroplast. This may suggest that ferritin in diatoms may have a different subcellular localization and evolved to serve multiple functional roles.

2.3 Discussion

The molecular mechanisms that have been investigated in diatoms for either iron uptake, its homeostasis or storage are summarized in Figure 1.11. Significant research attention

has focused on iron uptake in particular and it is clear that diatoms use a portfolio of strategies to access various forms of iron in the ocean. Remarkably, they adapt what appear to be ancestrally bacterial uptake proteins to eukaryotic mechanisms. The genes *FBP1* and *ISIP2A* have bacterial origins yet rely on endocytosis to function. Diatom-specific genes have also been observed (*e.g., ISIP1*), implying further innovation in this group to facilitate access to iron. Investigations into iron storage mechanisms have sparked academic debate into the role of ferritin in diatoms. This led to the discovery of species-specific differences in FTN utilization, and a phylogenetic characterisation of the *FTN* gene, highlighting the need for nuanced laboratory-based studies. The *in silico* prediction that FTN is targeted to the chloroplast in diatoms, where iron demands are highest, remains to be tested experimentally (see Chapter 3). Furthermore, it is clear that a range of molecular mechanisms require investigation – such as the functional role of ZIP and NRAMP proteins (specified by dashed lines in Figure 1.11), as well as ISIP3, to name a few.

There is evidence that the replacement of ferredoxin with flavodoxin in the photosynthetic machinery of diatom cells is one of the more important adaptations. A recent global survey of gene expression in marine phytoplankton showed that oceanic members of the diatom lineage strongly express flavodoxin over ferredoxin, while certain coastal diatoms, likely adapted to environments experiencing more frequent and larger fluctuations in iron supply, more highly expressed ferredoxin under iron-replete conditions (Caputi *et al.*, 2019). Of the seven most abundant genera in the ocean (indicated by an asterisk in Figure 2.1), four encode flavodoxin.

Similarly, the expression of ISIPs showed the highest correlation in communities, which were sampled from the lowest iron environments (Caputi *et al.*, 2019). *ISIP1* is notable for being the most sensitive to a low iron status, and likely represents a diatom-specific innovation, although its function requires further elucidation. Its role in mediating siderophore uptake is intriguing, because it suggests a community link between diatoms and siderophore producers. FBP1 is the only siderophore binding protein that has been described to date, and its distribution appears to correlate with ISIP1 in diatom transcriptomes (Figure 2.1), since most species that encode FBP1 also encode ISIP1 (15 species), with the exception of *Thalassiosira gravida* and *Chaetoceros curvisetus*. However, many more species encode ISIP1 than FBP1, so it is possible that ISIP1 has an additional role in cell iron homeostasis, or that there are

multiple siderophore binding proteins interacting with ISIP1.

An important avenue of future research is to shed a refining light on diatom niches, as the majority of studies focus either on individual model species or attempt to draw distinction between 'pennates' and 'centrics', 'open ocean' and 'coastal' diatoms while comparing only few representatives. For the iron uptake, homeostasis or storage mechanisms that I investigated here do not appear to be any remarkable patterns that distinguish pennates from centrics. To meaningfully compare open ocean and coastal diatoms, more information on the distribution of individual species in the global ocean are required. The *Tara* Oceans global circumnavigation effort that catalogued the community composition, metagenomes and metatranscriptomes of aquatic microorganisms can offer initial insight. For example, *T. oceanica*, the species most often referred to in the literature as an open ocean species in fact has a ubiquitous distribution and is frequently found in coastal areas (based on 18S data from *Tara* Oceans gene catalogue, data not shown). Of the 82 species reviewed here, only *Helicotheca tamesis* was not detected within 250 kilometers from the coast during *Tara* Oceans sampling (Juan Pierella Karlusich, Institut de Biologie de l'Ecole normale supérieure, personal communication), and may therefore be considered a true open ocean representative. By contrast, there is a considerable number of coastal species represented in the MMETSP sequence set. The species that were not detected further offshore than 250 km are *Proboscia inermis* (found exclusively in coastal zones at high latitudes), *Ditylum brightwellii*, *Skeletonema marinoi*, *Ticeratium dubium*, *Odontella aurita*, *Navicula sp., P. tricornutum*, *A. coffeaeformis* and *Cylindrotheca closterium*.

Using *Tara* Oceans data, Caputi *et al.* (2019) found that diatom species thrived across a gradient of total Fe concentrations and showed remarkable plasticity in their responses to Fe availability. The authors concluded that it was not possible to correlate species assemblages to iron levels or transcriptional responses in iron uptake systems. I believe that further progress will be made when *in situ* studies of species-specific diatom gene expression and analysis of community structure will be coupled to careful characterisation of iron sources available in seawater, since it is possible that it is the chemical nature of iron sources that complicates diatom niche separation.

2.4 Material and methods

The diatom sequence library consists of: the most recent annotation versions of all published diatom genomes (Armbrust *et al.*, 2004; Lommer *et al.*, 2012; Galachyants *et al.*, 2015; Mock *et al.*, 2017; Rastogi *et al.*, 2018); decontaminated versions of further transcriptomes from MMETSP (Keeling *et al.*, 2014; Dorrell *et al.*, 2017); and seven additional diatom transcriptomes synthesised in-lab, corresponding to the seven most abundant previously cultured taxa identified in *Tara* Oceans libraries (Malviya *et al.*, 2016). The searchable database that was used for the analyses presented comprised 82 species, which met decontamination standards described in (Johnson *et al.*, 2019). The seven species sequenced by Genoscope are among the ten most abundant diatom genera in the *Tara* Oceans dataset.

To investigate the evolutionary distribution of iron-related genes, first, representative protein sequences that are well annotated were downloaded from online databases (GenBank, Uniprot and Ensembl). Further sequences were retrieved from NCBI by BLASTp with a threshold E value of 1×10^{-5} for each gene of interest. Together these comprised the reference dataset and were aligned by MUSCLE and transformed into a Hidden Markov Model using HMMER 3.0. The profiles were used for HMMER search against the diatom transcriptomes from MMETSP alongside seven transcriptomes generated by Genoscope and diatom genomes.

It is important to note that this dataset likely misses genes that are under strong transcriptional control, which was not met by the culturing conditions during RNA harvesting (*e.g.,* low Fe induced genes in diatom cultures grown under replete Fe), as well as chloroplast encoded genes. The sequencing depth between the transcriptomes, and between genomes varied, and this may have influenced our results. The results presented here may be compared to similar published reviews, notably by Blaby-Haas and Merchant (2013); Groussman *et al.* (2015); Behnke and LaRoche (2020). Differences in numbers of hits retrieved are likely due to differences in chosen methodologies. Homologs of *FTN* were retrieved using HMMER search within available diatom genomes and transcriptomes. It is important to note that the presence/absence of genes in a transcriptome may not be a true reflection of copy numbers in genomes either because it was generated from iron replete cultures or was

not sequenced to sufficient depth.

The phylogenetic tree was constructed to clarify the evolutionary distance of *FTN* between species. 244 Sequences were first retrieved by HMMER search using the ferritin PFAM domain PF00210 as a query in the available 82 diatom transcriptomes/genomes with an E-value cut-off of 1×10^{-10}. Retrieved sequences were analysed using the CD-HIT web server and sequences that met a similarity threshold higher than 0.9 were eliminated (presumed to be duplicates). I generated an HMM for *FTN* using *P. tricornutum* and closely aligned sequences retrieved by BLASTp, and this was used as the basis for a second HMMER search of the remaining sequences, with an E-value cut-off of 1×10^{-10}, to further cut off the redundancy. This left a total of 64 representative sequences, which captured the diversity of ferritins within the diatoms. Conserved sequences were aligned using the alignment builder in Geneious v.10.2 under default criteria. The tree was drawn with ITOL (https://itol.embl.de/). Numbers beside branches indicate RaxML bootstrap coefficients. Ferritin protein sequence alignment (one-letter code) was carried out using MUSCLE in Geneious.

The subcellular localization prediction uses a combination of SignalP, ASAFind, HECTAR, MitoFates, TargetP and WolfPSort. SignalP 3.0 server predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes (Bendtsen *et al.*, 2004). ASAFind identifies nuclear-encoded plastid proteins in algae with secondary plastids of the red lineage based on the output of SignalP and the identification of conserved "ASAFAP"-motifs and transit peptides (Gruber *et al.*, 2015). HECTAR v1.3 Predict subcellular targeting for heterokont proteins (Gschloessl *et al.*, 2008). MitoFates is a prediction tool for identifying putative mitochondrial presequences and cleavage sites (Fukasawa *et al.*, 2015). TargetP-2.0 server predicts the presence of N-terminal presequences: signal peptide (SP), mitochondrial transit peptide (mTP), chloroplast transit peptide (cTP) or thylakoid luminal transit peptide (luTP). (Emanuelsson *et al.*, 2007). WoLF PSORT predicts the subcellular localization sites of proteins based on their amino acid sequences, based on both known sorting signal motifs and some correlative sequence features such as amino acid content (Horton *et al.*, 2007). I use these tools in combination with different databases in order to explore different life groups and specific subcellular predictions.

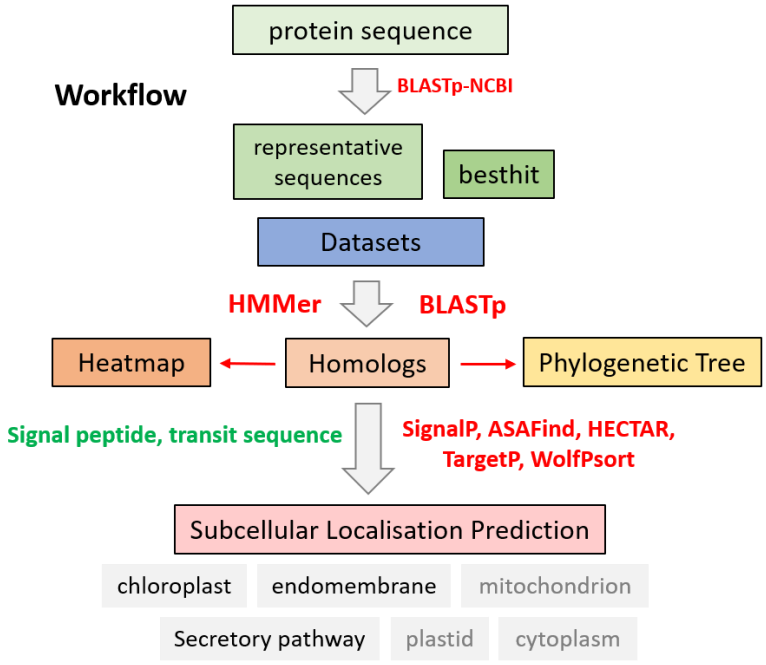


Figure 2.6 Summary of the bioinformatics workflow**.** First, a protein sequence was downloaded from the website and used as query in BLAST search. Second, HMMER and BLASTp were employed to search the profiled representative sequences or best hit through diatom datasets for homologs. Third, heatmaps and phylogenetic trees were built based on these homologs. In the end, subcellular localization prediction was carried out through a combination of several powerful tools.

Chapter 3: Localization of ISIP3 and FTN in the model diatom *P. tricornutum*

Abstract

FTN in most diatoms, including in *P. tricornutum*, is predicted to be targeted to the chloroplast, based on the sequence information encoded by its signal peptide. The function and localization of ISIP3 was unknown. In this chapter, *in vivo* observations of the localization of FTN and ISIP3 were performed for the first time. FTN was found to localize to the diatom chloroplast. ISIP3, which contains a ferritin like domain, localized into a bright spot located adjacent to the centre of the chloroplast. It is not clear whether this localization is to the periplastidial space, or the chloroplast interior. ISIP3 was also shown to co-localize with ISIP1, an unexpected result that suggests possible interaction between these two proteins. The results presented here provide valuable comparisons between ISIP3 and FTN localization and pave the way for further studies in the field of diatom iron metabolism. Further investigation on the accurate localization and mechanism in iron transport and storage are required.

3.1 Introduction

Iron stress in diatoms has been studied on a genetic level using the model pennate diatom, *P. tricornutum*, which has a sequenced genome and can be easily transformed (Apt *et al.*, 1996b; Bowler *et al.*, 2008). There are a multitude of studies which look at the effect of iron on *P. tricornutum*. In a seminal work, through a combined transcriptome and metabolome analysis it was shown that iron deprivation leads to the significant up-regulation of a group of diatom-specific genes encoding proteins designated as “iron starvation–induced proteins” (ISIPs) (Allen *et al.*, 2008). ISIP2a was shown to be a functional equivalent to phytotransferrin protein, binding dissolved ferric iron directly on the cell surface (McQuaid *et al.*, 2018, as reviewed in section 1.3.1). ISIP1 was shown to be involved in the endocytosis-mediated uptake of siderophore-bound iron (Kazamia *et al.*, 2018). The function of ISIP3 has not been

investigated physiologically, although bioinformatic annotations predict ferritin like functional domains (DUF305, PF03713) within conserved sequences. Therefore, a reasonable hypothesis is that ISIP3 may function as an iron storage protein alongside FTN in *P. tricornutum*. However, the role of FTN protein in Fe storage in *P. tricornutum* remains to be confirmed. Its function is the subject of experiments presented in this chapter. Here, I compare FTN and ISIP3 patterns of localization, which may shed light on the convergent functions of the two proteins.

Objectives and overview of the work presented

The goal of the analyses presented in this chapter of the dissertation was to clarify the localization of ISIP proteins and FTN *in vivo* in *P. tricornutum*. I constructed fluorescent lines of *P. tricornutum* by genetically engineering a fluorescent protein (FP) tag to the two genes of interest. The lines were observed using the capabilities of confocal microscopy. Specifically, I addressed the question of whether FTN localizes to the chloroplast. This was predicted bioinformatically (see section 1.3.3) but has never been confirmed *in vivo* using transformants. I generated a genetically modified *P. tricornutum* cell line, FTN tagged to GFP.

Additionally, I aimed to compare the localization of ISIP3 with ISIP1 and ISIP2a within the same transgenic line. The localization of ISIP1 and ISP2a have both been previously reported in (Morrissey *et al.*, 2015; Kazamia *et al.*, 2018; McQuaid *et al.*, 2018). Localization of ISIP3, the least studied of the ISIPs, is unknown. Understanding ISIP3 localization will provide basic clues, which can allow further study of physiological function of ISIP3 in *P. tricornutum*. A previously generated trichromatic line, P611, with ISIP1 tagged to yellow fluorescent protein (YFP), ISIP2a tagged to red fluorescent protein (RFP), and ISIP3 to blue fluorescent protein (CFP) was available in the laboratory for analysis. (The strain was a kind gift from the Allen Lab at the J. Craig Venter institute, La Jolla, USA). In this trichromatic line, each ISIP protein was tagged with a FP. The FP tag was C-terminal to the full protein sequence and the construct was driven by the native promoters of each gene.

A major strength of the study presented below is that it showcases advances in genetic engineering and fluorescence microscopy in diatoms, which are not routine to the field. Specifically, imaging of a transgenic line, which contains three proteins tagged to

different FPs in a highly auto-fluorescent cell has never been done before. The transgenic line FTN-GFP is also ground breaking in a different way – it was constructed using Universal Loop Assembly (uLoop) technology, a developed loop assembly based on a recursive approach to DNA fabrication. This is discussed in more detail in the Methods section 3.3.

The results of experiments using the trichromatic line are currently under review as part of a research paper submitted to the Journal of Experimental Botany (Kazamia *et al.*, 2021, in review). I am a contributing author to the work.

3.2Results

3.2.1 Localization of three ISIPs in a single trichromatic line

With each ISIP gene driven by its native promoter and coupled to a distinct FP, I was able to first check the response of these proteins to iron deprivation. The trichromatic line P611 was first cultivated in an iron-containing medium. Microscopic analysis showed that ISIP1-YFP and ISIP3-CFP were not visible and ISIP2a-RFP showed only very weak fluorescence (Figure 3.2A). This matches what is expected based on our understanding of the transcriptional response of *ISIP* genes to iron (Smith *et al.*, 2016). While ISIP2a is constitutively expressed, ISIP1 and ISIP3 are strongly regulated by iron availability. When iron is not available, the genes are expressed. Transcriptomic regulation of the 3 ISIPs is shown in Figure 3.1.

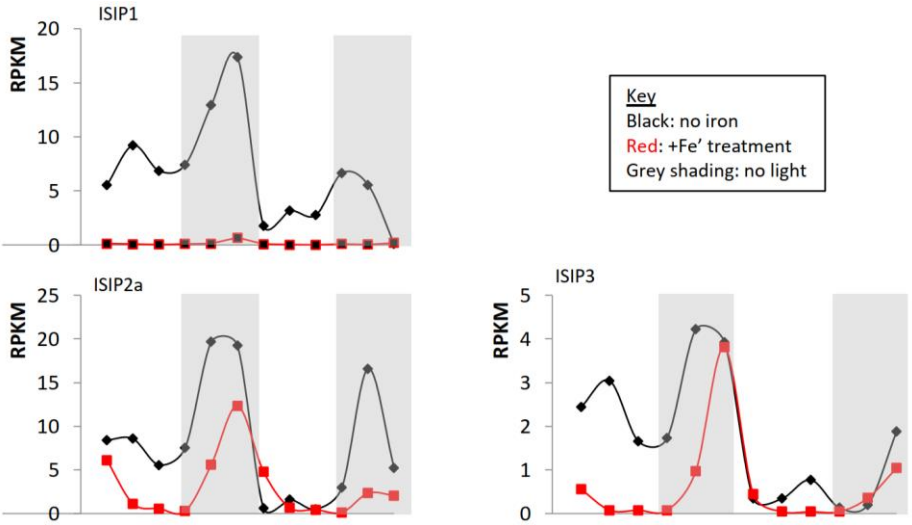


Figure 3.1 Transcriptional control of ISIPs based on results reported in Smith *et al.,* 2016. Transcript abundance of ISIPs was followed every 4 hours over a period of 48 hours for *P. tricornutum* cultures that were either iron starved (red line) or replete in iron (black line). Measurements were taken at 10 am, 2pm, 6pm, 10pm, 2 am and 6 am under a 12-hour light (white background), 12-hour darkness (grey background) regime, with lights on at 9 am and 25 lights off at 9 pm. For iron-starved cells the points represent biological duplicates and for iron-replete cells an average of 3 biological replicates were taken. Transcription abundance is measured as Reads Per Kilobase of transcript, per Million mapped reads (RPKM). Figure is shared by Kazamia *et al*., (2021, in review).

In order to induce fluorescence, I transferred the P611 cell line into iron-free medium and incubated it for two weeks. Following this treatment, ISIP1-YFP, ISIP2a-RFP and ISIP3-CFP in P611 cells all showed bright fluorescence (Figure 3.2B), indicating that ISIPs were sensitively induced by iron starvation, consistent with expectations.

A typical localisation pattern for ISIPs in iron-starved P611 cells is shown in Figure 3.2B. Here I show representative images from more than 7 biological replicates, which all showed similar localization patterns. ISIP2a was ubiquitously widespread on the cell surface, with a bright spot localized to the centre of the chloroplast. ISIP1-YFP showed bright fluorescence in the vicinity of the chloroplast and very weak fluorescence on the cell surface. Interestingly, ISIP3-CFP showed similar localization with ISIP1-CFP but with much weaker fluorescence, and had a bright central spot very similar to ISIP2a. The localization of ISIP1 in P611 cells is consistent with a previous study (Kazamia *et al.*, 2018), which reported results from a ISIP1-YFP line (*e.g.,* transformed with a single protein). My imaging analyses also corroborate findings presented in a paper submitted to the Journal of Experimental Botany, by Kazamia *et al*., (2021), a publication to which I am a contributing author. In the study, currently under review, it was found that for the P611 line, all three ISIPs localised to the vicinity of the chloroplast midpoint, in a globular region, although ISIP2A-RFP showed tighter aggregation, at the centre of the chloroplast groove. ISIP1-YFP and ISIP3-CFP co-localised. The diameter size of the bright focal point of ISIP2A-RFP was approximately 1.12 μm (SD=0.2,379 n=100), and for ISIP1/ISIP3 it was 3.06 μm (SD=0.3, n=100) (Kazamia *et al*., 2021, in review). The average plastid diameter

size was 6.21 µm (SD=0.5, n=100) measured from differential interference contrast (DIC) images rather than autofluorescence, which is variable throughout the day and does not include the chloroplast membranes (Kazamia *et al.*, 2021, in review). My findings (a representative image of which is shown in Figure 3.2), working independently on the P611 line corroborate these results. The localization and pattern of fluorescence are in agreement with the reported data.

Diatoms contain a secondary plastid that derives from a red algal symbiont, that is enveloped by four membranes (Flori *et al.*, 2016). In the experiment using the P611 line, I was unable to further resolve the localization of the 'globular region,' to which all three ISIPs localised, to distinguish whether the accumulation was in the periplastidal compartment (PPC) or inside the chloroplast. Further experiments with higher resolution, like correlative light-electron microscopy (CLEM), are planned to investigate the precise localization. Particular emphasis will be placed on identifying the 4 chloroplast membranes, and determining how many of these the ISIP-FPs had crossed.

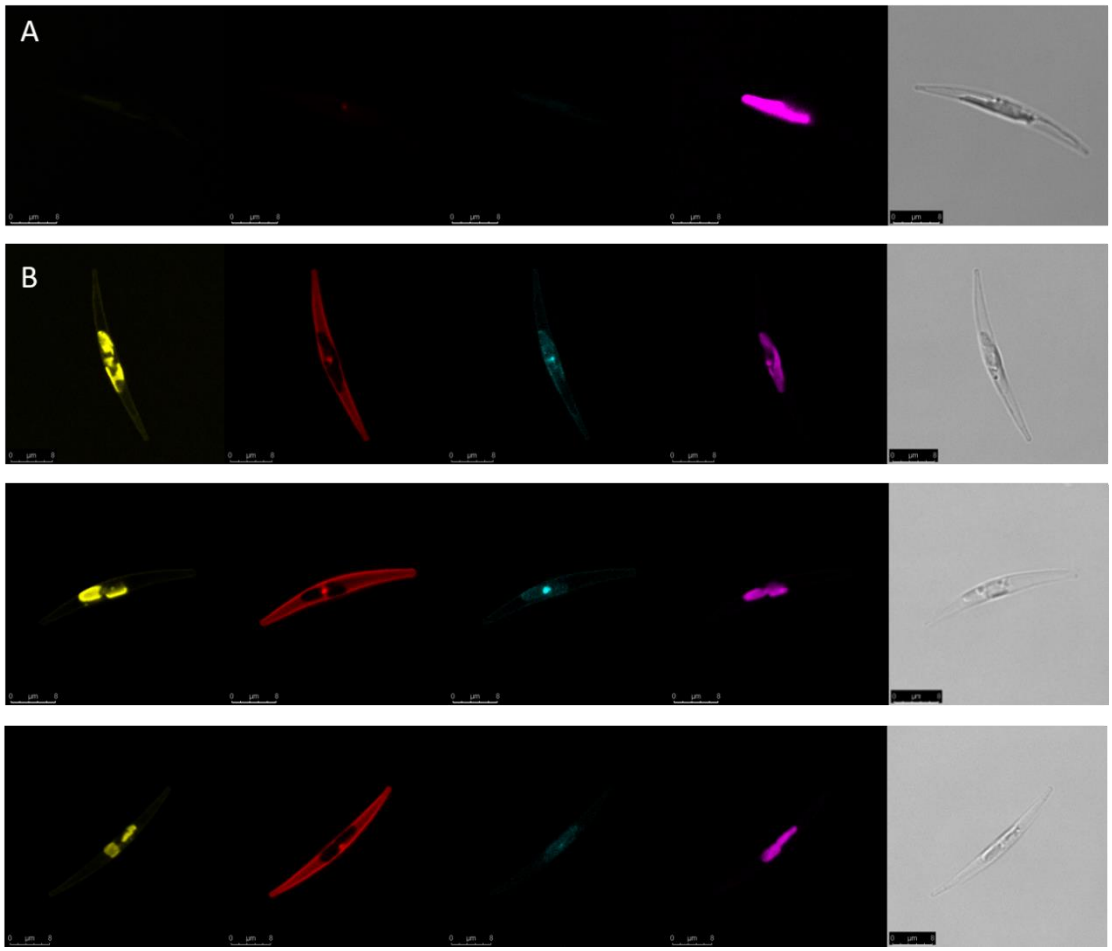


Figure 3.2 Representative images of *P. tricornutum* P611 trichromatic cell.

Images from left to right, in turns were ISIP1-YFP, ISIP2a-RFP, ISIP3-CFP, chloroplast autofluorescence and cell image under bright light. A. cells cultivated in iron-containing medium; B. cells starved for two weeks in iron-free medium. Microscopic images were recorded as described in 3.4.3 Methods. Scale bars are indicated.

3.2.2 FTN is targeted to the chloroplast

The fluorescent FTN-GFP cell lines were constructed by universal loop assembly (uLoop) and conjugation. At the beginning, dozens of colonies were harvested. Then, four cell lines were selected for further experiments relying on their performance observed under confocal microscope (as described in section 3.4).

With the FTN protein operating under an ammonium sensitive cassette of promoter and terminator (*Pnr / Tnr*), the fluorescence of FTN-GFP is expected to be 'off' when cells are grown in medium with ammonium as the N resource and be switched 'on' when nitrate was available, in which the expression of FTN-GFP is induced. The *Pnr / Tnr* cassette enables inducible gene expression and control of both the level and timing of mRNA and protein expression in transgenic diatoms (Poulsen and Kröger, 2005).

The generated FTN-GFP line behaved as predicted; cells showed weak FTN-GFP fluorescence when grown in ammonium-containing medium (Figure 3.3A) and bright fluorescence when transferred and grown in nitrate-containing medium (Figure 3.3B). The FTN-GFP exhibited consistent localization, in each case localizing to the chloroplast in four independent transformed cell lines when nitrate was available in the medium. This result was consistent in all biological replicates, of which there were at least 3 for each cell line. FTN-GFP aggregated as a bright spot, of differing size. The chloroplast localization is consistent with the *in silico* localization prediction of FTN. This is the first time that localization of FTN has been confirmed *in vivo* in *P. tricornutum*.

The distribution of FTN-GFP within the chloroplast is not even. The pattern appears speckled, with granules of different sizes. Further investigations by microscopy with higher resolution are planned to investigate the precise distribution and structure.

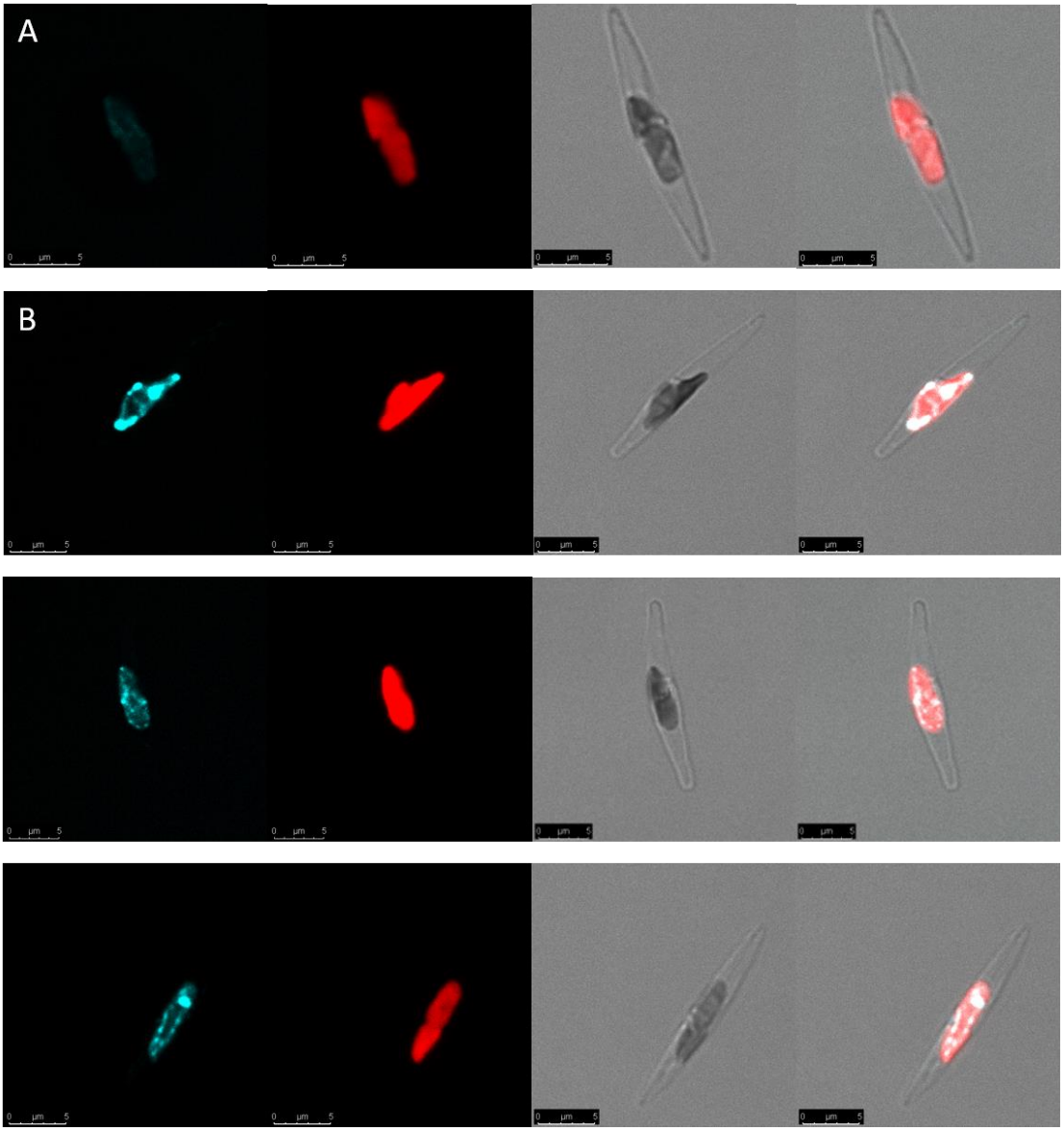


Figure 3.3 Representative images from *P. tricornutum* P611 trichromatic cells expressing FTN-GFP.

Images from left to right, in turns were FTN-GFP, chloroplast autofluorescence, cell image under bright light and overlapped image. A. cells cultivated in ammonium containing medium; B. cells grow in medium with nitrate as N resource.

3.3 Discussion

The functional role of FTN in diatoms is disputed. In laboratory studies, it was shown that *P. granii*, a chain forming diatom whose genome encodes FTN (Figure 2.4), was able to undergo several more cell divisions in the absence of iron than the comparably sized, oceanic centric diatom *T. oceanica*, which lacks the FTN gene, supporting the hypothesis that FTN functions in the long-term storage of iron (Marchetti *et al*., 2009). However, others have argued that FTN functions instead as an iron oxidation enzyme, in iron buffering, and iron homeostasis (Murgia *et al.*, 2007; Duc *et al.*, 2009; Pfaffen *et al.*, 2013). Additionally, the function of ISIP3 is completely unknown, but annotation of ferritin-like domains suggests a possible overlap in function with FTN.

In this chapter, by the advantage of the available trichromatic line P611 and the FTN-GFP line that I generated, it was possible to compare the localization of ISIP3 and FTN *in vivo* in the model diatom species, *P. tricornutum*, for the first time.

FTN was shown to function in long-term iron storage in *P. granii* (Marchetti *et al.*, 2009). The subcellular localization of FTN in *P. tricornutum* was predicted to be chloroplastic *in silico* (see section 2.2.2). The observations of the FTN-GFP line presented here show that FTN is indeed localized to the *P. tricornutum* chloroplast, as predicted bioinformatically. In the chloroplast, iron is of highest requirement, as iron ions function in the transfer of electrons. The results are consistent with the phylogenetic relationship between FTN clades, in which PtFTN is a variant of the more typical PmFTN, and forms a distinct clade. It provides the basic evidence for our hypothesis that FTN functions as an iron storage protein in *P. tricornutum*. Further work with *Pseudo-nitzschia* ferritins will be of great interest. Namely, it will be interesting to see whether *P. granii* ferritin also localizes to the chloroplast in a similar pattern established here for *P. tricornutum*. Unfortunately, this diatom has not yet been genetically transformed.

ISIP1 was shown to be involved in endocytosis-mediated siderophore uptake, localized to cell surface and chloroplast membrane in *P. tricornutum* (Kazamia *et al*., 2018). ISIP2a was shown to bind ferrous iron on the cell surface, internalized by endocytosis and localized to cell surface and intracellular vesicles (Morrissey *et al.*, 2015; McQuaid *et al.*, 2018). The results were basically consistent with previous studies, with small

differences in the localisation pattern of ISIP2a-RFP with respect to previous results, which aggregated as a bright distinct spot close to the chloroplast. ISIP3, the iron-starvation induced protein whose function is completely unknown, despite it being the most abundant of the three ISIPs in global datasets, encodes a conserved domain belonging to the ferritin superfamily (Behnke and LaRoche, 2020). Therefore, we hypothesize that ISIP3 may be involved in iron storage. Unexpectedly, ISIP3 turned out to localize in the vicinity of the chloroplast and to aggregate as a bright central spot in the chloroplast. Rather than showing similarity with FTN localization, it showed near identical localization with ISIP1. While the hypothesis that ISIP3 may function in iron storage remains to be fully tested, the findings provide an interesting possibility – that ISIP3 may play a role in iron metabolism through direct interaction with ISIP1 and ISIP2a. This is a valuable avenue for future investigation.

The confocal microscopy experiments presented here are not able to identify intracellular membrane structures. The results presented here are not conclusive of function, which naturally requires further physiological investigation. The localization analysis can significantly benefit from comparisons at higher resolution. Experiments employing advanced microscopy techniques with higher resolution, which could identify the ultrastructure of the cell interior are enormously beneficial. A technique that can provide further insight is correlative light-electron microscopy (CLEM), which combines the capabilities of two typically separate microscopy platforms: light (or fluorescent) microscopy and electron microscopy, affording the dual advantages of ultrafast fluorescent imaging with the high-resolution of electron microscopy. Further experiments performed in various growth conditions with different iron bioavailabilities are being planned in collaboration with Yannick Schwab at the European Molecular Biology Laboratory (EMBL).

One further avenue for future work is to compare localization profiles of ISIPs and FTN in *P. tricornutum* cells exposed to different iron conditions, and over longer time periods. All the experiments presented here were surveying cells that were harvested at mid-exponential growth phase. The localization of these proteins throughout a culture's growth – at the beginning and in stationary phase - are also of interest. In order to parse the relationship between ISIP3 and FTN it is interesting to compare the localization profiles under supplementation of a various range of iron sources, including

siderophores, as well as in response to short-term Fe^{2+} spikes.

For further investigation into the function and mechanism of ISIPs and FTN in *P. tricornutum*, protein-protein interaction analysis combining biochemical and immunological characterization is also required. It is important to reproduce localization patterns with the native proteins in wild type cells, using antibodies, immunofluorescence and immunogold labelling. Antibody-based techniques are important because they will permit the study of the native protein, deemed essential to unequivocally define subcellular protein localisation.

Diatom plastids are of high complexity, surrounded by four membranes (Cavalier-Smith, 1999). As most proteins localizing to the different plastid compartments are nuclear-encoded, they have to be imported across several membranes via sophisticated translocon systems (Bolte *et al*., 2009). In recent years, important subunits of these translocators have been identified and used to clarify the exact location of proteins (Gould *et al*., 2015; Lau *et al.*, 2016; Chen, 2019). To test protein–protein interactions, the split GFP assay, or bimolecular fluorescence complementation (BiFC) assay are often applied, which allows detection of protein–protein interactions *in vivo* as the two GFP fragments can only re-assemble if they are in close proximity to each other (Kerppola, 2006) and not if they are located in the same compartment only without being fused to interacting proteins (Hempel *et al.*, 2009).

Another promising method of exploring the protein-protein interactions (or protein-DNA interactions) *in situ* is yeast two hybrid (Y2H) assay. It tests physical interactions (such as binding) between two proteins or a single protein and a DNA molecule, respectively. The premise behind the test is the activation of downstream reporter gene(s) by the binding of a transcription factor onto an upstream activating sequence (Caufield *et al*., 2012). Furthermore, heterologous expression of ISIP3 and FTN in yeast strain can tell if these proteins bind iron. While techniques such as Y2H could be valuable, they may however be challenging because the ISIP proteins are likely to be membrane localized.

3.4 Material and methods

To generate the FTN-GFP line, the FTN gene was first tagged to fluorescent protein GFP (mNeonGreen) and then cloned into the plasmid with fragment PtCv2 through the universal loop assembly technique (Pollak *et al.*, 2019). PtCv2 included many well-defined cassettes, the antibiotic resistance cassettes for *E. coli* selection, a diatom cassette for diatom selection on phleomycin, and an origin of transfer enabling conjugative transfer of episomes from bacteria into diatoms (Diner *et al.*, 2016). Then the plasmid was transferred to *P. tricornutum* by conjugation and cell lines were cultivated in enrichment solution with artificial water (ESAW), selected by phleomycin and observed by confocal microscopy.

3.4.1 Universal loop assembly

Universal loop assembly is a developed Loop assembly based on a recursive approach to DNA fabrication. The system makes use of two Type IIS restriction endonucleases and corresponding vector sets for efficient and parallel assembly of large DNA circuits. Standardized level 0 (L0) parts can be assembled into circuits containing 1, 4, 16 or more genes by looping between the two vector sets. The vectors also contain modular sites for hybrid assembly using sequence overlap methods. It provides a simple generalized solution for DNA construction with standardized parts (Pollak *et al.*, 2019). The procedures for constructing the FTN-GFP plasmid are shown in Figure 3.4.

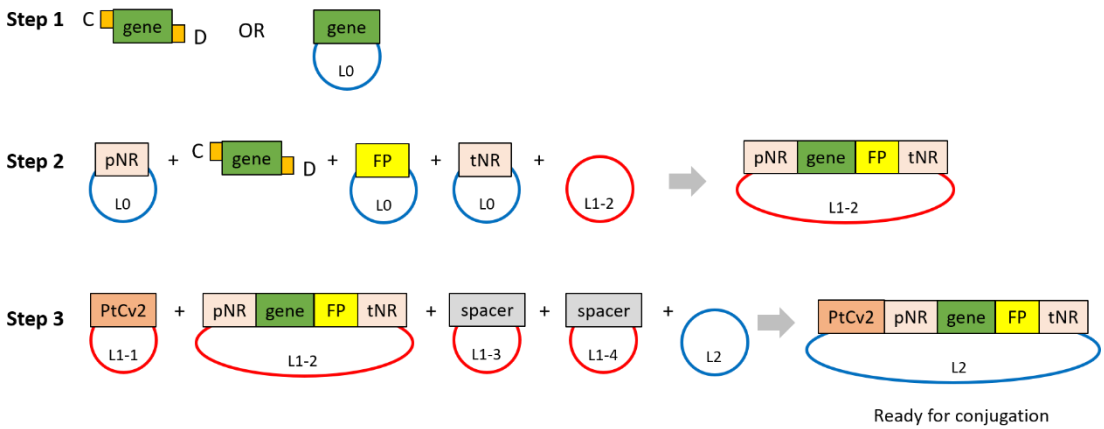


Figure 3.4 Procedures for construction of FTN-GFP plasmid in uLoop.

First, the *FTN* gene was cloned from a cDNA library using Phusion high fidelity

polymerase (New England Biolabs) and then constructed with the C and D overhangs on the 5' and 3' termini, respectively. Then, the L0 level units, promoter pNR, gene with CD overhangs, FP (GFP, mNeonGreen), terminator tNR and L1 level receiver were assembled into one plasmid L1-2_pNR_FTN_GFP_tNR. Last, the L1 level units, L1-1_PtCv2, L1-2_pNR_FTN_GFP_tNR, L1-3_spacer, L1-4_spacer and L2 level receiver were assembled into L2_FTN-GFP, which are ready for conjugation.

Design of Oligonucleotides for CD overhangs

The primers for constructing genes with CD overhangs consisted of UNS fragment, *BsaI* site and overlap sequence. The nucleotides were as below.

UNS1-full Fw: catta ctcgc atcca ttctc aggct gtctc gtctc gtctc (40 bp)

UNS1_half_20 bp *BsaI*_site overhang 18 bp_sequence

C.Fw aggct gtctc gtctc gtctc aGGTCTCa **AATG** NNNNNNNNNNNNNNNNN

UNSX-full Rv: ggtgg aaggg ctcgg agttg tggta atcta tgtat cctgg (40 bp)

UNSX_half_20 bp *BsaI*_site overhang 18 bp_sequence+-start/stop

D.Rv tggta atcta tgtat cctgg tGGTCTCt **ACCTGA** NNNNNNNNNNNNNNNN

Since the C and D overhangs include the start codon ATG and stop codon TGA, respectively. The start and stop codons of the gene of interest were removed.

Loop assembly

The Type IIS assembly protocol used in Loop assembly was adapted from Pollak *et al.*, (2019). Changes were performed to enable simpler calculation of required concentrations, and streamlining of the reaction set-up by the use of master mixes and pre-diluted plasmids at working concentrations. Plasmids in Loop assembly were diluted according to their lengths. For 'donor' plasmids, the target concentration was length/100, for a 'receiver' plasmid length/200. This resulted in a 15 fmol/mL concentration for donor parts and 7.5 fmol/mL for the receiver plasmid. Then, 1 mL of each part was mixed resulting in a final volume of 5 mL. Assembly of L0 parts into a L1 construct might require aliquoting to 5 mL by adding water or by removing a volume

of the mixed DNA due to the variable number of parts that can be composed. Then, 5 mL of 2X reaction master mix (2X LBL for odd level assemblies "*e.g.,* L0 to L1", and 2X LSL for even level assemblies) were mixed with the DNA at room temperature and cycled between 37°C and 16°C in a thermocycler.

LBL mix: 2X Master Mix (*BsaI*) per reaction:

- 3 µL of HPLC grade H2O
- 1 µL of 10x T4 DNA ligase buffer (NEB)
- 0.5 µL of purified BSA 1 mg/mL
- 0.25 µL of T4 DNA Ligase at 400 U/mL (NEB)
- 0.25 µL of *BsaI* at 10 U/mL (NEB).

LSL mix: 2X Master Mix (*SapI*) per reaction:

- 3.5 µL of HPLC grade H2O
- 0.5 µL of 10x T4 DNA ligase buffer (NEB)
- 0.5 µL of CutSmart buffer 10X (NEB)
- 0.25 µL of T4 DNA Ligase at 400 U/mL (NEB)
- 0.25 µL of *SapI* at 10 U/mL (NEB)

PCR cycling parameter:

- Step 1: 3 minutes at 37°C
- Step 2: 4 minutes at 16°C
  Repeat step 1 and 2 for 25 cycles
- Step 3: 5 minutes at 50°C
- Step 4: 10 minutes at 80°C

3.4.2 Conjugation

The resulting episome was introduced to *P. tricornutum* using the conjugative plasmid pTA-MOB, as described in Diner *et al.*, (2016). Positive transformants were selected on zeocin antibiotic plates (100 µg/ml). The selected cell lines were grown in liquid cultures using ESAW medium supplemented with zeocin (100 µg/ml) during microscopy experiments.

3.4.3 Confocal microscopy

Images were acquired using a confocal imaging microscope, an inverted Leica SP8 (Leica Microsystems, Germany) equipped with 63x/1.2 W objectives (HCLP APOC52), a pulsed white light Acousto-Optical Beam Splitter (AOBS) laser and internal hybrid single-photon counting detectors, HyD SMD 1-4. Notch filters were used for each laser, the speed of acquisition was 200 frames per second, with optimised pixel range set to 3608x3608, 8-line accumulations and pixel development time of 100ns per frame.

Scanning was sequential between frames.

To observe the trichromatic line, an optimised three-step fluorescence image acquisition process was employed:

Sequence 1: cells were excited by 20 μW of 470 nm wavelength and fluorescence was detected in the window 475-509 nm gated at 1.2 ns. This sequence allowed the measurement of CFP.

Sequence 2: cells were excited by a laser pulse at 20 μW of 514 nm light, with fluorescence detected in the 519-572 nm range. This sequence measured fluorescence of YFP.

Sequence 3: cells were excited at 590 nm (20 μW) with fluorescence detected at 595-660 nm as well as at 689-730 nm, the latter using a photomultiplier tube (PMT) detector. The shorter wavelengths detected are indicative of RFP fluorescence, and the PMT detector measured chlorophyll autofluorescence. The signal between RFP and autofluorescence was separated by applying different time gates (0.9-1.7 ns for RFP and 3.6-12.5 ns for autofluorescence). The time gates were established through fluorescence lifetime imaging analysis (Figure 3.5). Chlorophyll autofluorescence was successfully 'gated-out' from interfering with FP emission measurements and did not interfere with FP detection during experimental procedures. Therefore, chlorophyll autofluorescence did not interfere with FP detection during experimental procedures.

Expected emission and excitation spectra of the FPs is shown in Figure 3.6. The fluorescence acquisition was done in this order, so that the dimmest fluorophore, CFP, would be detected first, followed by YFP and then the brightest, RFP. Desiccation and bleaching from prolonged laser exposure occurred after approximately 10 minutes of continuous exposure, such as during a z-stack collection. Images interpreted in this manuscript are from alive cells only. No FP-associated fluorescence was detected in wildtype cells under these parameters.

To observe the FTN-GFP line, Cells were excited by 20 μW of 470 nm wavelength and fluorescence of CFP was detected in the window 475-509 nm gated at 1.2 ns. Cells were excited at 590 nm (20 μW) with chlorophyll autofluorescence detected at 595-660 nm (HyD SM4) as well as at 689-730 nm. The time gates were established through

fluorescence lifetime imaging analysis. No FP-associated fluorescence was detected in wildtype cells under these parameters.

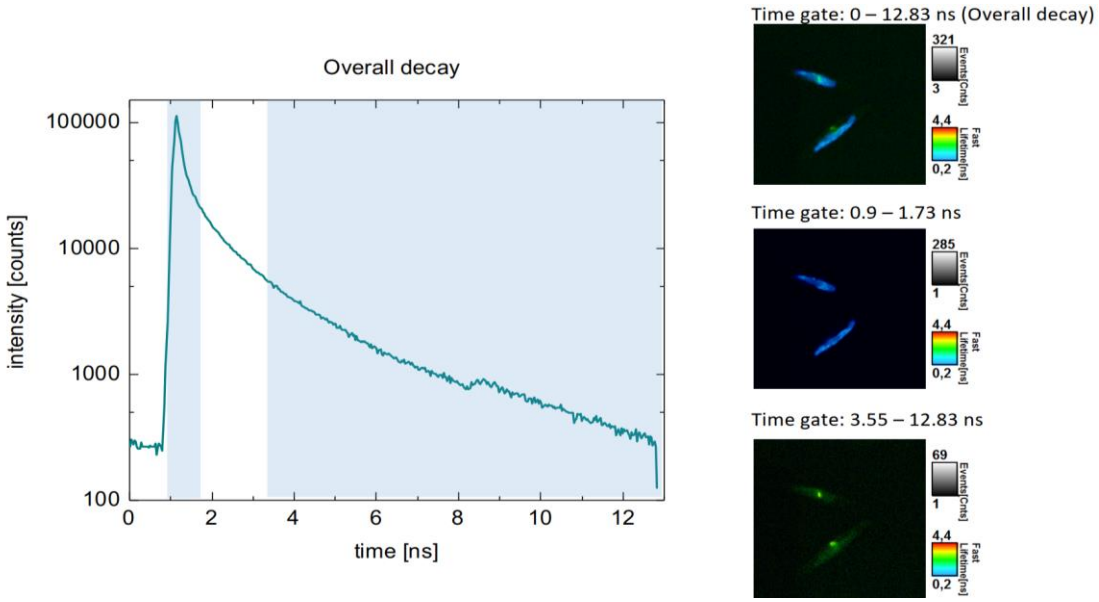


Figure 3.5 FLIM of P611 cells excited at 470 nm and a repetition frequency of 80 MHz. The left-hand panel shows the overall emission intensity (in total counts) of P611 cells over a period of 12 seconds. The righthand panels show images of the cells fluorescence within a defined time gate, either 0-12.83 seconds (total), or an early (0.9-1.73 ns) and late (3.44-12.83 ns) gate. The initial steep drop-off in intensity is largely attributed to chlorophyll autofluorescence, which decays significantly faster than the fluorescence associated with FPs. Figure is shared by Kazamia *et al*., (2021, in review).

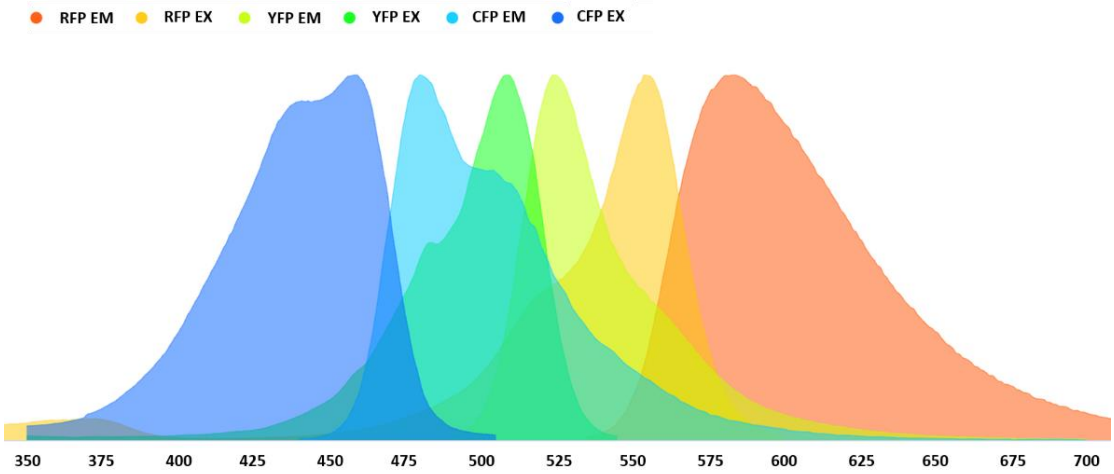


Figure 3.6 Emission and excitation spectra for the fluorescent proteins used in this study. These are the expected excitation and emission spectra of YFP, RFP and CFP. Figure is generated on FPbase (https://www.fpbase.org/).

Chapter 4: A functional investigation of ISIP3 and FTN in *P. tricornutum* using knockout mutants

Abstract

To explore further the role of ISIP3 and FTN in *P. tricornutum*, knockout mutants were generated by CRISPR-Cas9, selected and confirmed by genotyping and by protein detection through western blotting. The phenotypes of KO mutants and the control cell lines (wild type transformed with empty vector, MT) were compared. The absence of FTN reduced the growth of the cell cultures. The most obvious effect was observed when iron-deplete cells were grown in iron-free medium. This indicates that FTN is involved in iron storage and iron homeostasis. The absence of ISIP3 exhibited a completely different effect on the cells, with no phenotype observed in KO mutants when these were grown in iron-free medium, irrespective of the initial iron-replete or iron-deplete status of cells. Furthermore, there was a significant decrease in growth of KO mutants compared with MT cells when cultures were grown in iron-containing medium. The results suggest that ISIP3 is not involved in iron storage, but rather iron acquisition. The loss of phenotype of KO mutants grown in copper-free medium further supported the conclusion. Further experiments on unicellular iron quota and iron uptake rates are required.

4.1 Introduction

In the previous chapter, the localization of ISIP3 and FTN proteins were investigated for the first time. FTN was shown to localize to the chloroplast. Rather than showing similarity with FTN, ISIP3 showed a much more complex pattern, localizing in a bright central spot in the center of the chloroplast. This could either be the periplastidial space or the chloroplast interior. ISIP3 showed near identical localization with ISIP1. This indicates a likely different mode of action for ISIP3 compared with FTN. Instead, possible involvement of ISIP3 in iron acquisition and interaction with ISIP1 and ISIP2a is proposed. To further explore and compare the possible functions of FTN and ISIP3

in iron uptake and iron storage in *P. tricornutum*, *in vivo* physiological experiments are required.

In the current chapter, further investigation into the role of ISIP3 and FTN in of iron metabolism have been performed. To investigate the function of ISIP3 and FTN in *P. tricornutum*, I first generated knockout (**KO**) mutants by CRISPR-Cas9 (see section 4.4.2) and compared their phenotypes with control cell lines containing an empty vector (i.e., without inserted gene cassette), which henceforth I refer to as **MT** (empty vector control lines). These empty cassette lines are resistant to and were maintained by selection in zeocin, allowing for a fair comparison to the mutants. After I obtained the colonies, the mutants were genotyped by sequencing target genes. Protein presence or absence was verified through western-blotting. With the confirmed KO strains, I carried out growth curve experiments to compare the physiology of KO mutants with MT.

4.1.1 The four physiological states of cells in relation to iron

Since iron is the vital micronutrient being investigated, the experimental design I employed varied iron availability, through presence or absence of ferric iron in the medium ESAW (plus and minus ferric iron refer to ESAW+Fe, ESAW-Fe). As diatoms are known to store iron intracellularly, the second factor taken into account during comparisons was the intracellular iron status of cells. Growth curves were started either with iron-replete or iron-deplete cells. The latter had been iron-starved for two weeks prior to the start of experiments. In the following sections, iron-replete cells are referred to as 'Cells+Fe' and iron-deplete cells as 'Cells-Fe').

Combining the two factors - different iron availability and intracellular iron status – gave a total of four experimental conditions, summarized in Figure 4.1, abbreviated as condition i (Cells+Fe, ESAW+Fe), condition ii (Cells+Fe, ESAW-Fe), condition iii (Cells-Fe, ESAW+Fe) and condition iv (Cells-Fe, ESAW-Fe).

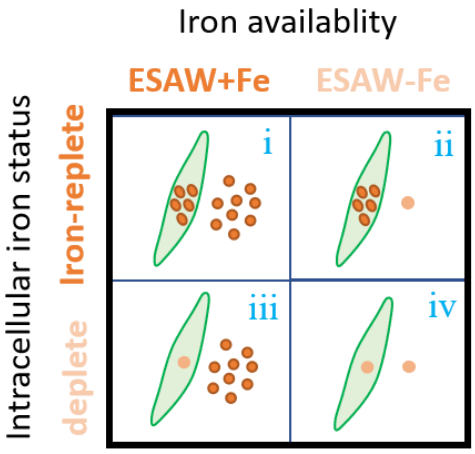


Figure 4.1 Schematic of iron availability and intracellular iron status as a two-dimensional setting.

The green fusiform represents *P. tricornutum* cells. The orange dots represent the relative abundance of iron content, intracellularly and extracellularly. There were four growth conditions. Condition i. iron-replete cells grown in iron-containing medium (Cells+Fe, ESAW+Fe); Condition ii. iron-replete cells grown in iron-free medium (Cells+Fe, ESAW-Fe); Condition iii. iron-deplete cells grown in iron-containing medium (Cells-Fe, ESAW+Fe); Condition iv. iron-deplete cells grown in iron-containing medium (Cells-Fe, ESAW-Fe).

4.1.2 The role of copper in iron acquisition

The reductive iron uptake pathway is well characterized in diatoms. It is composed of putative ferric iron reductases, multi–copper(Cu)-containing ferrous iron oxidases, and ferric iron permeases (Maldonado and Price, 2001; Shaked *et al.*, 2005; Maldonado *et al.*, 2006). In addition, oceanic diatoms may have additional copper requirements associated with the substitution of the iron-containing enzyme cytochrome c_6 by copper-containing plastocyanin, as was shown in *T. oceanica* (Peers and Price, 2006). There is a hypothesis, therefore, that copper is required for efficient reductive iron uptake. While copper can be limiting to growth of diatoms on its own, its role in Fe uptake is expected to have a confounding effect on iron limitation. When the effect of Fe and Cu availability was tested on the physiology of a range of diatoms, both had a significant effect on the growth rates of both coastal and oceanic diatoms, however, the

interaction between Fe and Cu availability and growth rates was only observed for oceanic diatoms (Annett *et al*., 2008). Copper limitation was shown to be more severe in diatoms grown in low- than in high-Fe seawater. *T. weissflogii* and *T. oceanica* had lower Fe quotas and slower rates of Fe uptake when Cu was reduced in the medium. Brief exposure of Cu-limited cells to 10 nmol/L Cu increased the instantaneous Fe uptake rate by 1.5 times in *T. oceanica* (Peers *et al*., 2005). The ferrous iron oxidase activity was dependent on copper (Cu) availability and was diminished by exposure to a multi-Cu oxidase (MCO) inhibitor in *T. oceanica*. The rates of Fe uptake from FOB by Fe-limited *T. oceanica* were also dependent on Cu availability (Maldonado *et al*., 2006).

Further, it has been proposed that Fe-limited *Pseudo-nitzschia* spp. may produce the potent neurotoxin domoic acid (DA) to access Cu, needed at the core of a high-affinity Fe transport system (Lelong *et al*., 2013). Additionally, the crystal structure of CopM (a cyanobacterial metallochaperone) in apo and copper-bound forms were reported (Zhao *et al*., 2016), further implicating the element in Fe uptake. Strikingly, ISIP3 of *P. tricornutum* turned out to be the only gene retrieved by BLASTp search of CopM. ISIP3 shows 37% sequence identity with CopM.

Taking these observations together, I therefore hypothesized that Cu was required for Fe transport in *P. tricornutum.* I therefore tested the effect of Cu on the iron physiology of the mutants and MT cells by removing the micronutrient from the growth medium. Where this treatment was applied, cells were either exposed to iron or iron was eliminated from the medium (abbreviated as ESAW+Fe-Cu and ESAW-Fe-Cu).

4.1.3 Investigating siderophores as organic iron source for Fe.

An early study showed that *P. tricornutum* cells were able to access iron from the tri-hydroxamate siderophores ferrioxamine E (FOE) and ferrioxamine B (FOB), albeit with different kinetics (Soria-Dengg and Horstmann, 1995). Experiments with FOB confirmed this recently, and ISIP1 was shown to be involved in endocytosis of the siderophore complex into the cell, with a reduction step to disassociate the bound iron in the vicinity of the chloroplast (Kazamia *et al.*, 2018). Since ISIP3 showed a similar localization to ISIP1 (see section 3.2.1) in my microscopy work, an interesting

hypothesis is that ISIP3 is involved in siderophore uptake and/or interacts directly with ISIP1. Thus, in the experiments that follow I compared the growth of ISIP3-KO mutants and MT in medium that supplied the siderophores FOB and FOE as an iron source instead of ferric iron chelated with EDTA (abbreviation as ESAW+FOB, ESAW+FOE).

For treatments where Fe-EDTA was the iron source, the concentrations of $FeCl_3$ and Na_2EDTA added to ESAW were 8.69×10^{-6} and 8.30×10^{-6} mol/L respectively. EDTA is a common synthetic chelator, in addition to ferric iron; it complexes other metal ions (manganese, iron, zinc, copper). In this scenario, only a fraction of the iron is available to diatoms as dissolved Fe' – approximately $1/50^{th}$ of the added $FeCl_3$ (as described in Sunda *et al*., 2005). For treatments where siderophores were the source of iron, the concentration of FOB and FOE added were 2.64×10^{-8} and 1.15×10^{-7} mol/L, respectively. In the scenario where the siderophores were added, no freely dissolved Fe' is available in solution, so if any iron is taken up, it is from the siderophores directly.

$$Fe^{3+} + EDTA \leftrightharpoons FeEDTA \quad (1)$$

4.1.4 Growth curve measurements

To characterize the phenotypes of KO mutants, I recorded a range of parameters (summarised in Figure 4.2). First, I counted cell numbers every two days. Based on cell density I calculated the specific growth rate and doubling time when cultures were at the very beginning of exponential phase, and compared the cell numbers at the end of the experiments when cultures reached the stationary phase. Further, I compared the average cell sizes at the end of growth curves, when cultures had reached stationary phase. As photosynthesis is vital for diatoms and its efficiency is directly linked to Fe availability, I measured the chlorophyll content and chlorophyll auto-fluorescence in the late exponential phase. For each strain, three cultures were set as biological replicates for the measurements. The values were represented as mean ± SD. The statistical significance was calculated, * indicated p-value < 0.05, ** refers to p-value < 0.001. The number of independently generated cell lines of MT, FTN-KO mutants, ISIP3-KO mutants were 2, 3 and 2, respectively. For each cell line, cultures were run in triplicate, *e.g.,* with three biological replicates.

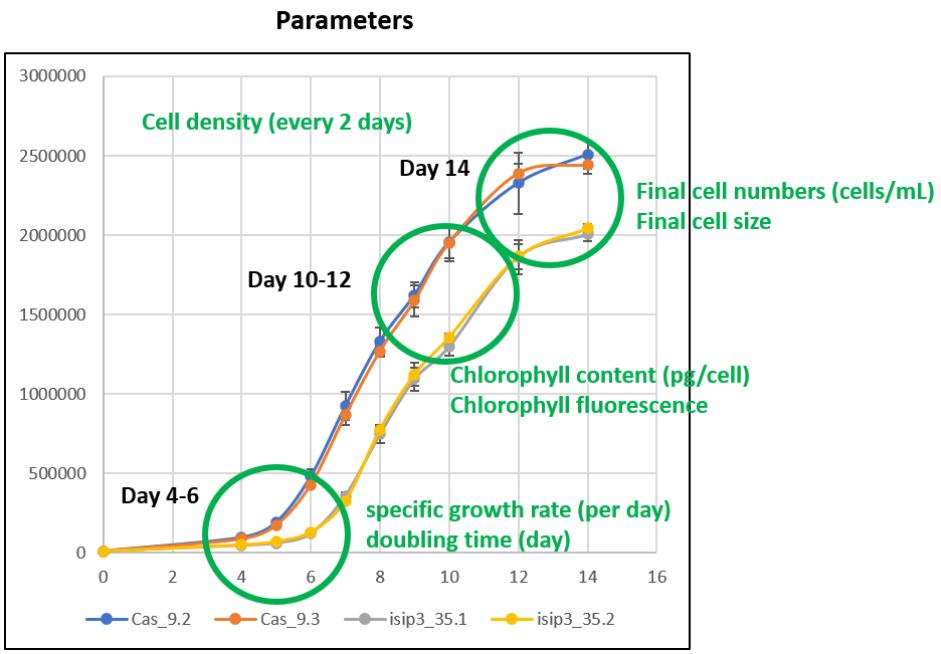


Figure 4.2 Physiological parameters to be recorded during growth curves.

The monitored experimental data included first, biomass: cell density (cells/mL), specific growth rate (day-1) and doubling time (day); second, chlorophyll: chlorophyll content (pg/cell) and chlorophyll fluorescence; third, relative cell size.

4.2 Results

With combined molecular biology and cell physiology, I investigated the role of ISIP3 and FTN in the model diatom species *P. tricornutum*. First, knockout mutants were generated by CRISPR-Cas9 and bombardment, selected and validated by genotyping and western-blot. Then growth curves of MT and KO-mutants were carried out to screen the phenotypes under different conditions, with different iron availability and initial intracellular iron status.

4.2.1 Knockout mutants confirmed by genotyping and western-blot

ISIP3-KO mutants

To generate ISIP3-KO mutants by CRISPR-Cas9 techniques, five initial sites were designed to target to the first exon of gene *ISIP3* in *P. tricornutum* (Figure 4.3, more

details see Methods 4.4). Among five targets, only isip3_35 worked. Two knockout mutants were selected and verified by genotyping and western-blot experiments. Mutant isip3_35.1 contained a deletion of one nucleotide whereas isip3_35.2 had a deletion of seven nucleotides. As expected, interruption of the *ISIP3* gene resulted in the complete deletion of ISIP3 protein in both mutants (Figure 4.4).

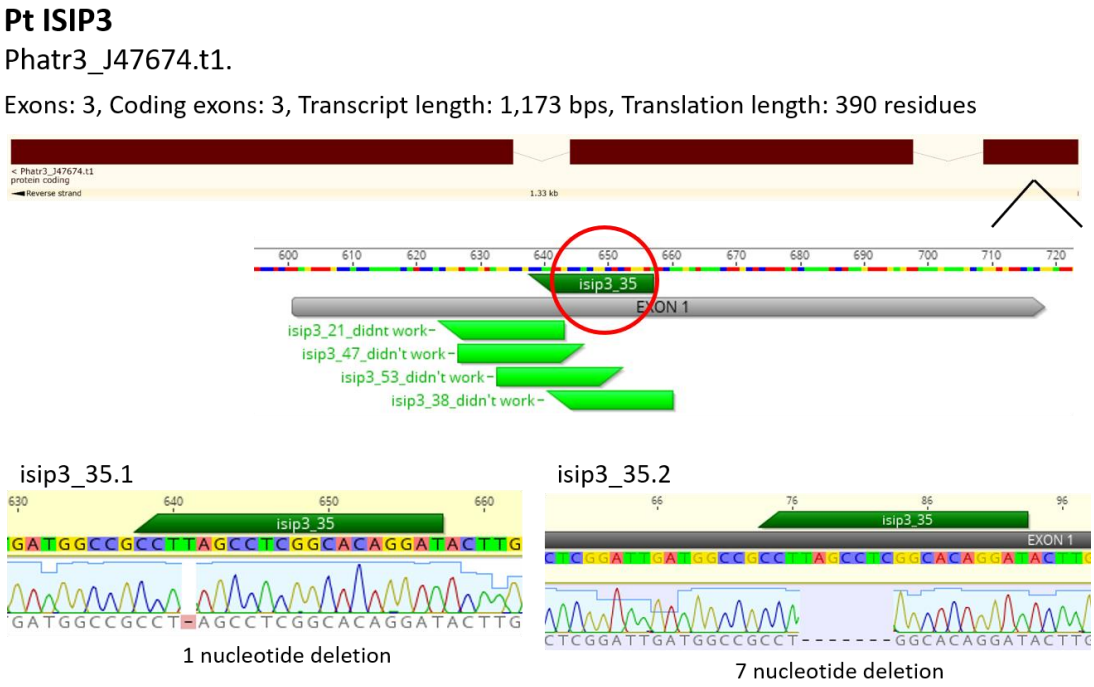


Figure 4.3 Genotypes of ISIP3-KO mutants, determined by sequencing.

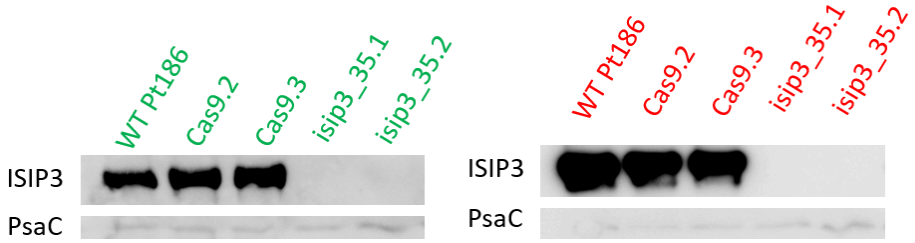


Figure 4.4 Deletion of ISIP3 protein in KO mutants confirmed by western blot.

Above the figure are the names of cells. In each lane, the same amount of protein extract was loaded. On the left panel are protein samples extracted from cells growing in iron-containing medium (green); on the right, under iron starvation (red). Protein extracted from wild type cells and two MT lines (Cas9.2 and Cas9.3) are used as positive control. ISIP3 and PsaC antibodies are used separately to detect the proteins. PsaC is a chloroplast protein that is ubiquitously expressed in diatoms and therefore broadly used as reference gene (Takahashi, 1991).

FTN-KO mutants

The target site of CRISPR-Cas9 in FTN was localized to the second exon because the first exon was too short (Figure 4.5). Two targeting sites worked among three. More than eight cell lines were selected and verified as knockout mutants, but I chose the three most promising lines with clear mutations.

To check the mutants at protein level, because a diatom or plant anti-Ferritin antibody was not available, a human Anti-Ferritin Heavy Chain antibody (ab89787) on ABCAM was tried. However, this antibody did not bind anything in my western blot experiments, not even in protein samples extracted from human cells (Hela, A549 and HEK_293 Figure 4.6).

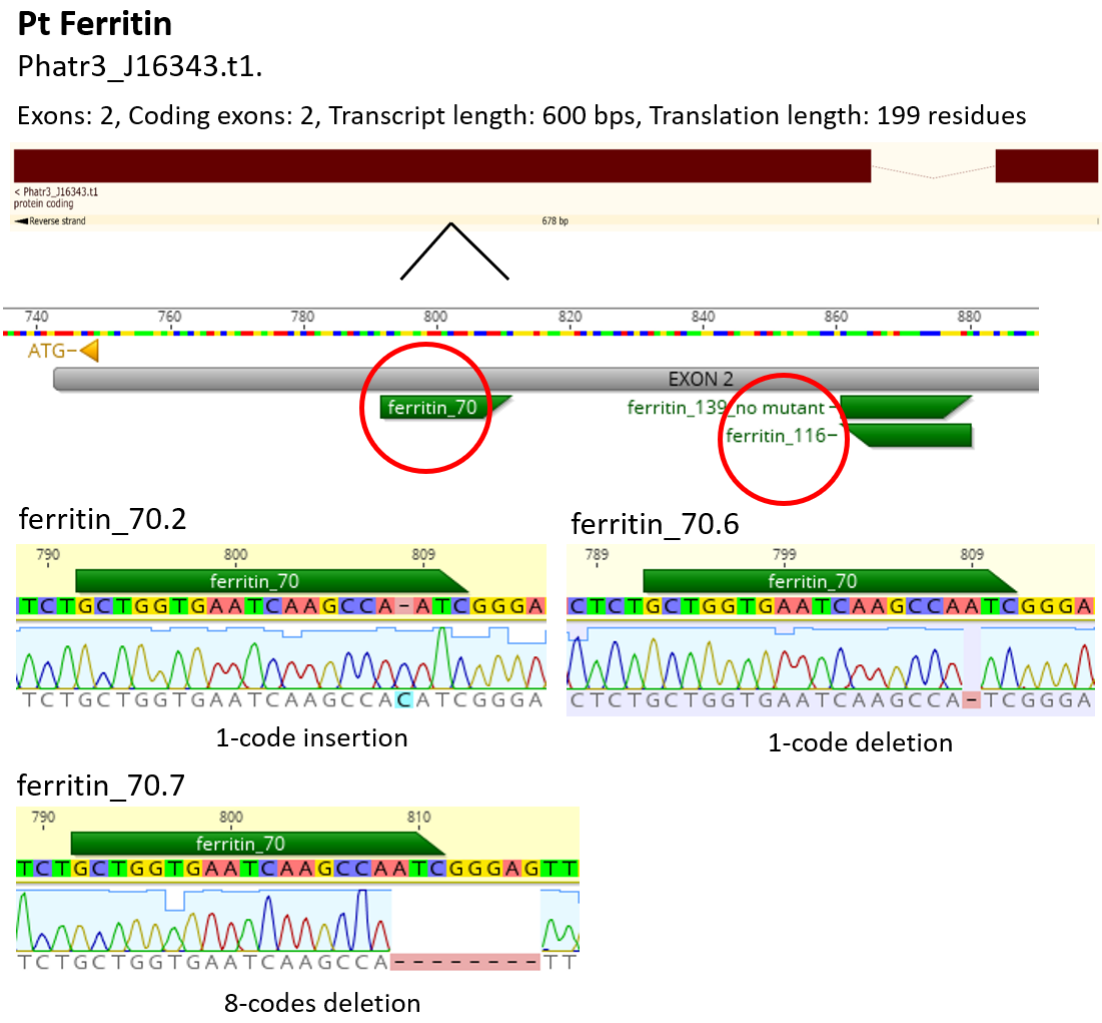


Figure 4.5 Genotypes of FTN-KO mutants, checked by sequencing.

Genotypes correspond to 1 nucleotide insertion, 1 nucleotide deletion and 8 nucleotides deletion in mutation lines ferritin_70.2, ferritin 70.6 and ferritin 70.7 respectively. Therefore, in each cassette full length sequence of FTN in *P. tricornutum* is expected to be interrupted by the mutation.

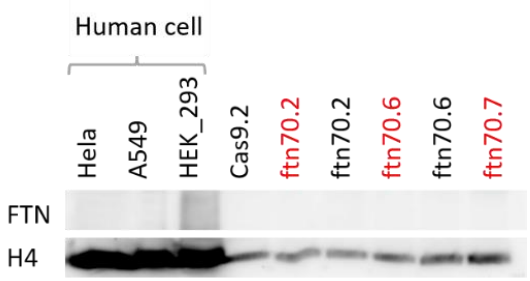


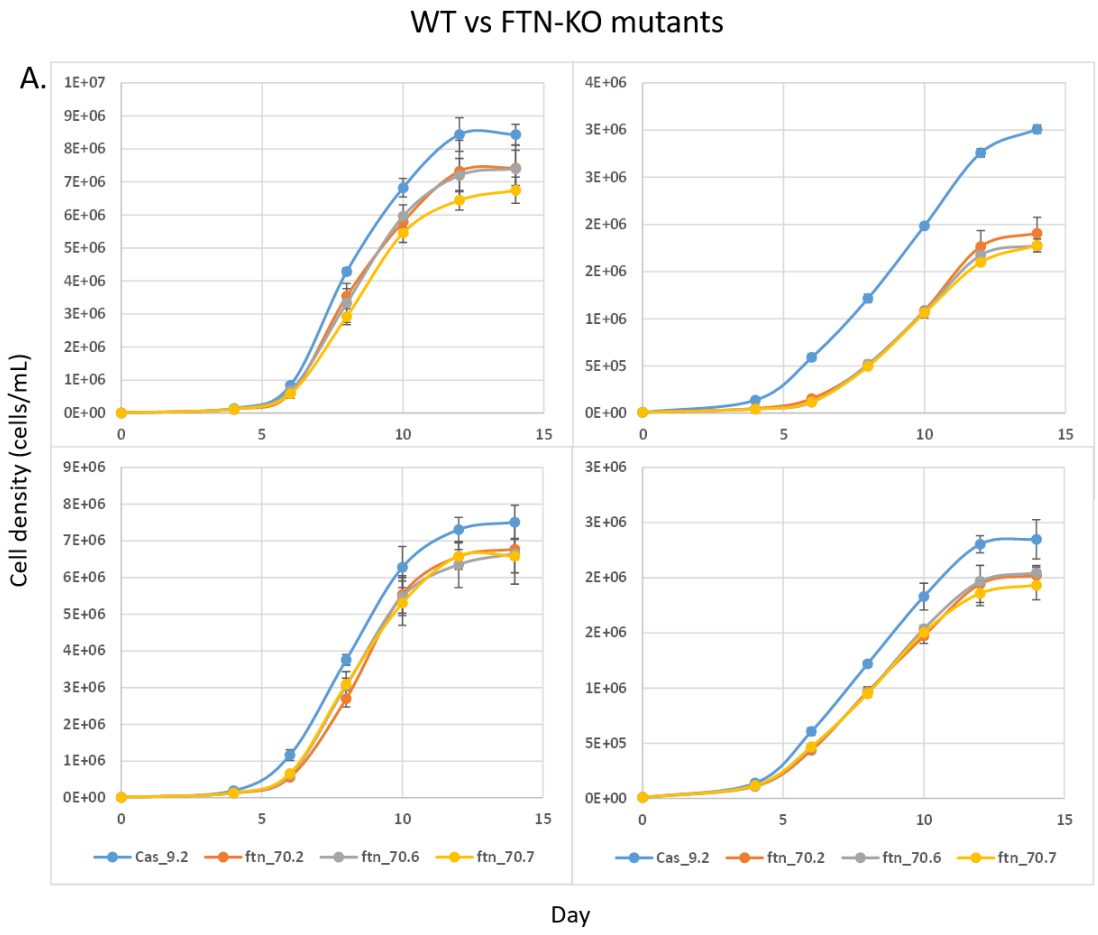
Figure 4.6 Western-blot result of FTN-KO mutants.

In each lane, the same amount of protein samples is added. Protein extracted from human cells that are used as positive control, including cancer cell line Hela, lung cell A549 and kidney cell HEK_293. Protein samples extracted from the knockout mutants grown either with iron (black) or under iron starvation (red) are tested. Anti-Histone H4 antibody is used as reference gene.

4.2.2 Phenotypes of ferritin knockout (FTN-KO) mutants

In all growth conditions, regardless of iron availability and intracellular iron status, FTN-KO mutants showed lower final cell density compared with MT (wild type cell lines transformed with empty vector). This carrying capacity of the cultures of KO mutants was lower by 16.7%, 40.2%, 8.2% and 15.2% in condition i, ii, iii and iv, respectively, compared to MT (Figure 4.7A). The biggest difference was observed in condition ii (Cells+Fe ESAW-Fe), in which the final cell density of FTN-KO mutants was $1.6\times10^6 \pm 7.4\times10^4$ cells/mL, and was 40.2% lower than MT, $2.8\times10^6 \pm 4.3\times10^4$ cells/mL (Figure 4.7Aii). In other words, the absence of FTN had an overall effect on growth regardless of iron availability and intracellular iron status. Furthermore, when iron was not available for iron-replete cells, the cell density of FTN-KO mutants was significantly reduced compared with the MT cultures.

The specific growth rate and doubling time were calculated from the cell density, from day 4 to day 6, when cells were at the early exponential phase. Both of them were consistently related to the final cell density measured during the growth curve. The most obvious difference between FTN-KO mutants and MT was observed in condition ii (Cells+Fe ESAW-Fe). The specific growth rates of MT and FTN-KO mutants were 0.69 ± 0.02 and 0.45 ± 0.03 day^{-1}, respectively, in condition ii, showing 34.8% decrease in FTN-KO mutants compared with MT ($p<0.01$, Figure 4.7Bii). The doubling time of MT and FTN-KO mutants were 1.02 ± 0.04 and 1.64 ± 0.10 day, respectively, in condition ii, the doubling time in FTN-KO mutants was 60.3% longer than MT ($p<0.01$, Figure 4.7Cii). In contrast, the difference was neg in conditions i and iv (Figure 4.7B i, iv and 10C i, iv). It meant that the absence of FTN affected the growth but didn't affect the survival of cells, at least not in terms of cell division.



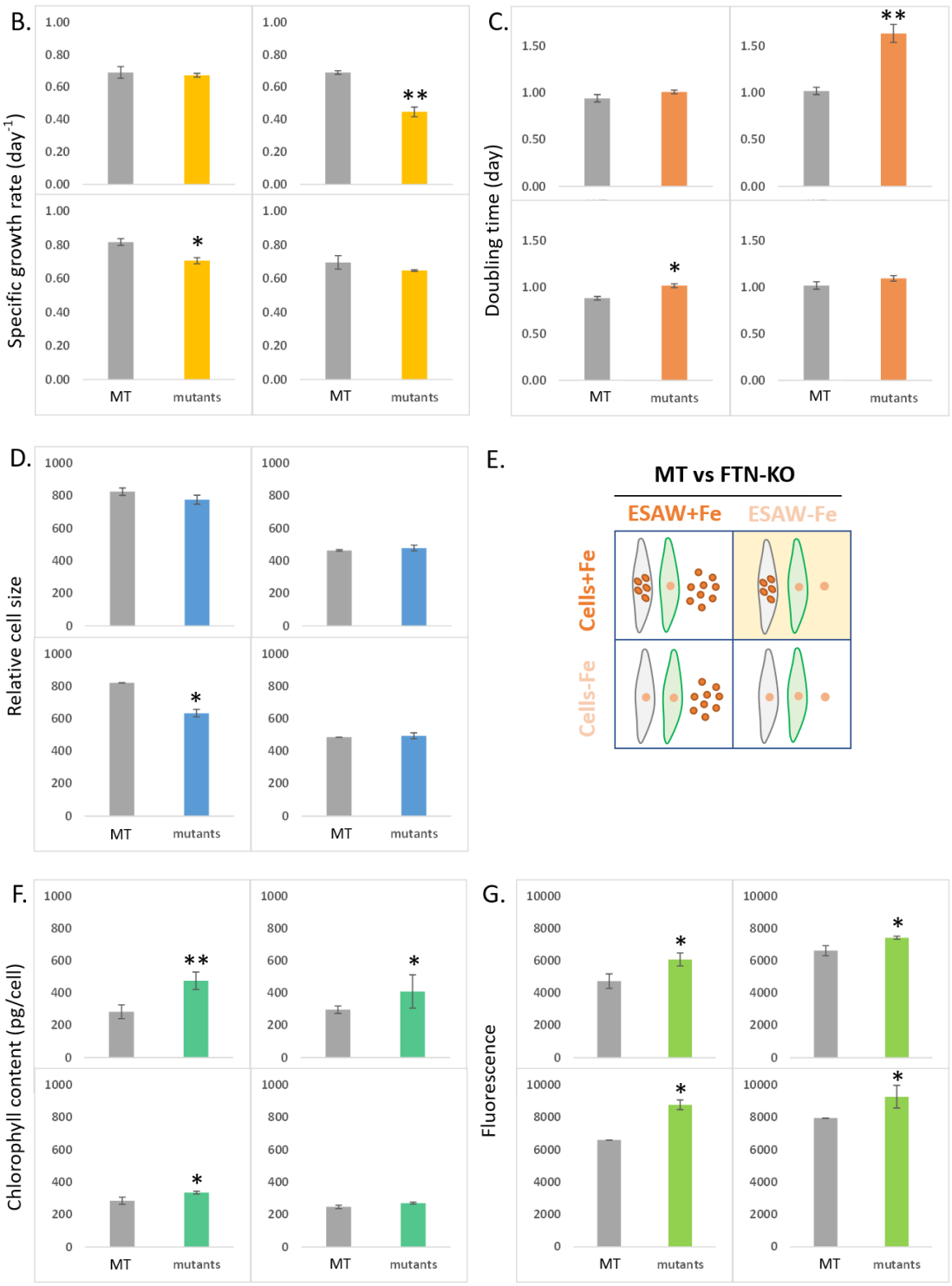


Figure 4.7 FTN-KO mutants versus MT cells grown in ESAW+Fe and ESAW-Fe.

A. growth curve; B. specific growth rate; C. doubling time; D. relative cell size; E. a diagram showing my hypothesis with MT cells in grey and KO mutants in green; F. chlorophyll content; G. chlorophyll auto-fluorescence.

The biggest difference between FTN-KO mutants and MT in final cell density, specific growth rate and doubling time were recorded in condition ii (Cells+Fe ESAW-Fe). However, the difference in relative cell size was observed in condition iii (Cells-Fe, ESAW+Fe), the relative cell size of FTN-KO mutants was 22.8% smaller than MT (Figure 4.7Diii). Different parameters (cell density and cell size) showed remarkable differences between FTN-KO mutants and MT across all conditions, suggesting that the overall effect on the cell was multifaceted and extended beyond growth physiology, and also affected cell morphology.

Iron is an essential component for the production of chlorophyll. In plants, iron deficiency often causes chlorosis, a condition in which leaves produce insufficient chlorophyll and turn yellow (Abadía *et al.*, 2011). Therefore, I hypothesized that chlorophyll synthesis in the FTN-KO mutants might be affected and may be always lower than MT, especially if ferritin is used as an iron storage protein. Strikingly, the chlorophyll content of FTN-KO mutants was always higher than MT in all conditions. Chlorophyll content was greater in FTN KOs than MT by 67.4%, 37.9%, 17.8% and 9.3% in condition i, ii, iii and iv, respectively, regardless of the iron availability and intracellular iron status (Figure 4.7F). In condition i (Cells+Fe, ESAW+Fe), the chlorophyll content of MT was 284.6 ± 43 pg/cell compared with 476.6 ± 54.0 pg/cell in FTN-KO mutants. In condition ii (Cells+Fe, ESAW-Fe), it was 298.3 ± 23 and 410.6 ± 103.7 pg/cell respectively in MT and FTN-KO mutants. The chlorophyll fluorescence of FTN-KO mutants was consistent with chlorophyll content. It was consistently and significantly higher in FTN than MT. It was greater by 28.4%, 12.0%, 32.9% and 6.6% in condition i, ii, iii and iv, respectively, in mutants compared with MT (Figure 4.7G). In condition i, the chlorophyll fluorescence of MT and FTN-KO mutants was 4747 ± 451 and 6093 ± 402 arbitrary unit/cell, respectively. In condition iii (Cells-Fe, ESAW+Fe), it was 6604 ± 118 and 8779 ± 302 arbitrary unit/cell, respectively.

Overall, both chlorophyll content and fluorescence were higher in FTN KO mutants than in MT. While this disagreed with my original hypothesis, a possible explanation is that mutant cells may be compensating for the mutation's effect on growth by accumulating more chlorophyll. Intriguingly, the difference between FTN-KO mutants and MT in chlorophyll content was most remarkable when intracellular iron was replete, and in chlorophyll fluorescence when iron was available in the medium. Further

experiments are required to explain these phenotypes.

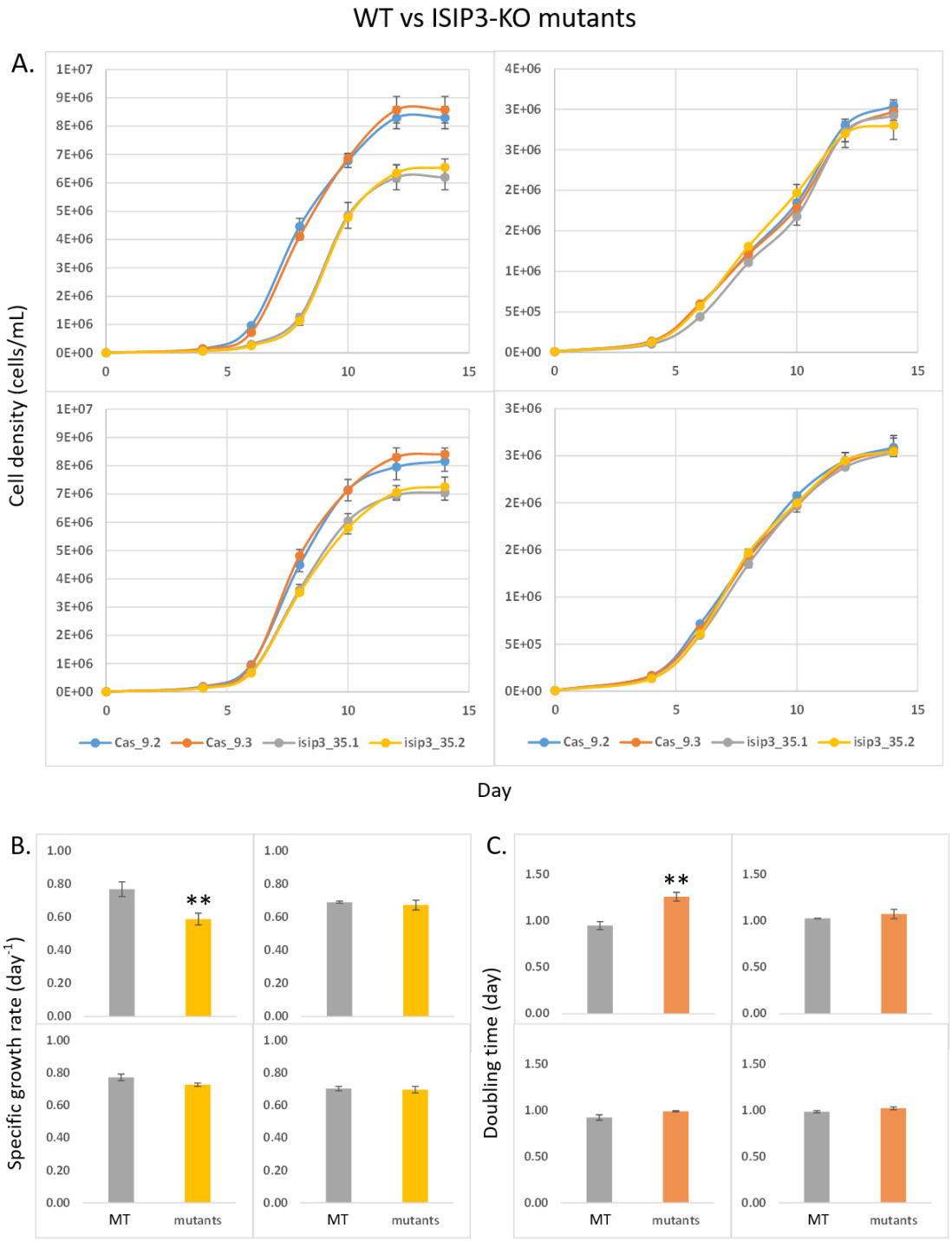
In conclusion, the marked decrease of cell density in FTN-KO mutants when iron was not available for iron-replete cells provides initial evidence that FTN may function as an iron storage protein in *P. tricornutum*. The effect of absence of FTN on diatom growth was observed regardless of iron status and availability, and the different effects of iron status and availability on chlorophyll might be the side-effect of interruption of iron storage (Figure 4.7E). The alternative hypothesis, that FTN is used in iron homeostasis, would suggest that a less consistent pattern would be observed, where availability of Fe would play a role in whether a decreased growth phenotype was observed, which was not the case in my experiments. Further experiments measuring iron cell quota directly in cells (both MT and KO) are needed to identify the role of FTN more precisely.

4.2.3 Phenotypes of ISIP3-KO mutants

ISIP3-KO mutants performed completely differently compared with FTN-KO mutants. FTN-KO lines showed significant decreases in growth compared with MT in condition iii, and slight decreases in the other three conditions, however, for ISIP3-KO mutants, I observed decreases in growth in conditions i and iii, but no difference in conditions ii and iv. The final cell density of MT and ISIP3-KO mutants in condition i (Cells+Fe, ESAW+Fe) were $8.4\times10^6 \pm 3.4\times10^5$ and $6.3\times10^6 \pm 4.5\times10^4$ cells/mL, respectively. KO mutants decrease 25.7% in cell density compared with MT (Figure 4.8Ai). In condition iii (Cells+Fe, ESAW-Fe), it was $8.3\times10^6 \pm 2.9\times10^5$ and $7.2\times10^6 \pm 3.0\times10^4$ cells/mL, respectively, decreased by 13.5% in KO mutants (Figure 4.8Aiii). In contrast, in conditions ii (Cells-Fe, ESAW+Fe) and iv (Cells-Fe, ESAW-Fe), there wer no obvious differences between MT and ISIP3-KO mutants.

Consistently with carrying capacity differences, I observed that specific growth rate, doubling time and cell size showed the biggest difference between ISIP3-KO mutants and MT in condition i (Cells+Fe, ESAW+Fe). In this condition, the specific growth rates of MT and ISIP3-KO mutants were 0.77 ± 0.04 and 0.59 ± 0.04 day-1, respectively, decreased by 23.4% in KO mutants (Figure 4.8Bi) compared with MT. The doubling time of MT and ISIP3-KO mutants were 0.95 ± 0.04 and 1.26 ± 0.05 day, which was

32.6% longer in KO than MT (Figure 4.8Ci); the relative cell size of ISIP3-KO mutants was 20.7% smaller than MT (Figure 4.8Di).



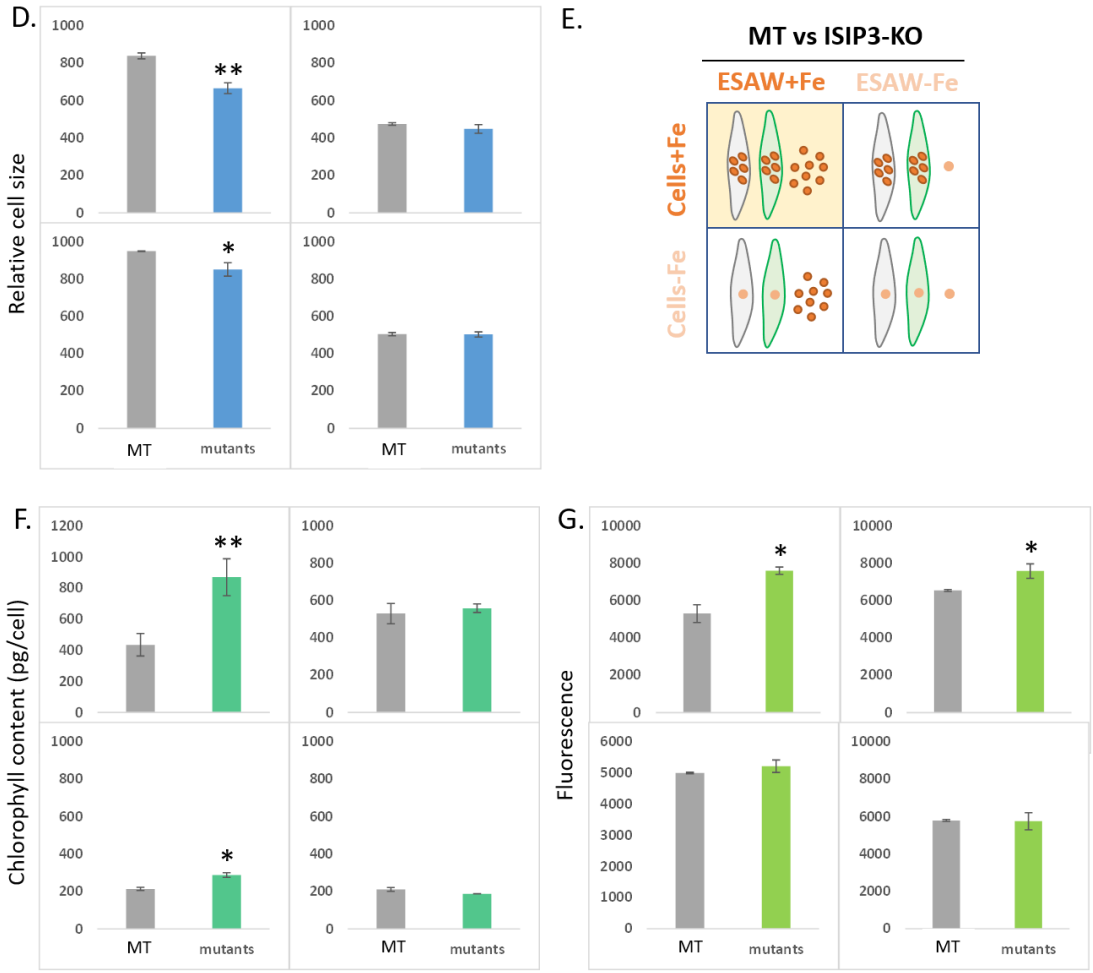


Figure 4.8 ISIP3-KO mutants versus MT cells grown in ESAW+Fe and ESAW-Fe.

A. growth curve; B. specific growth rate; C. doubling time; D. relative cell size; E. a diagram showing my hypothesis with MT cells in grey and KO mutants in green; F. chlorophyll content; G. chlorophyll auto-fluorescence.

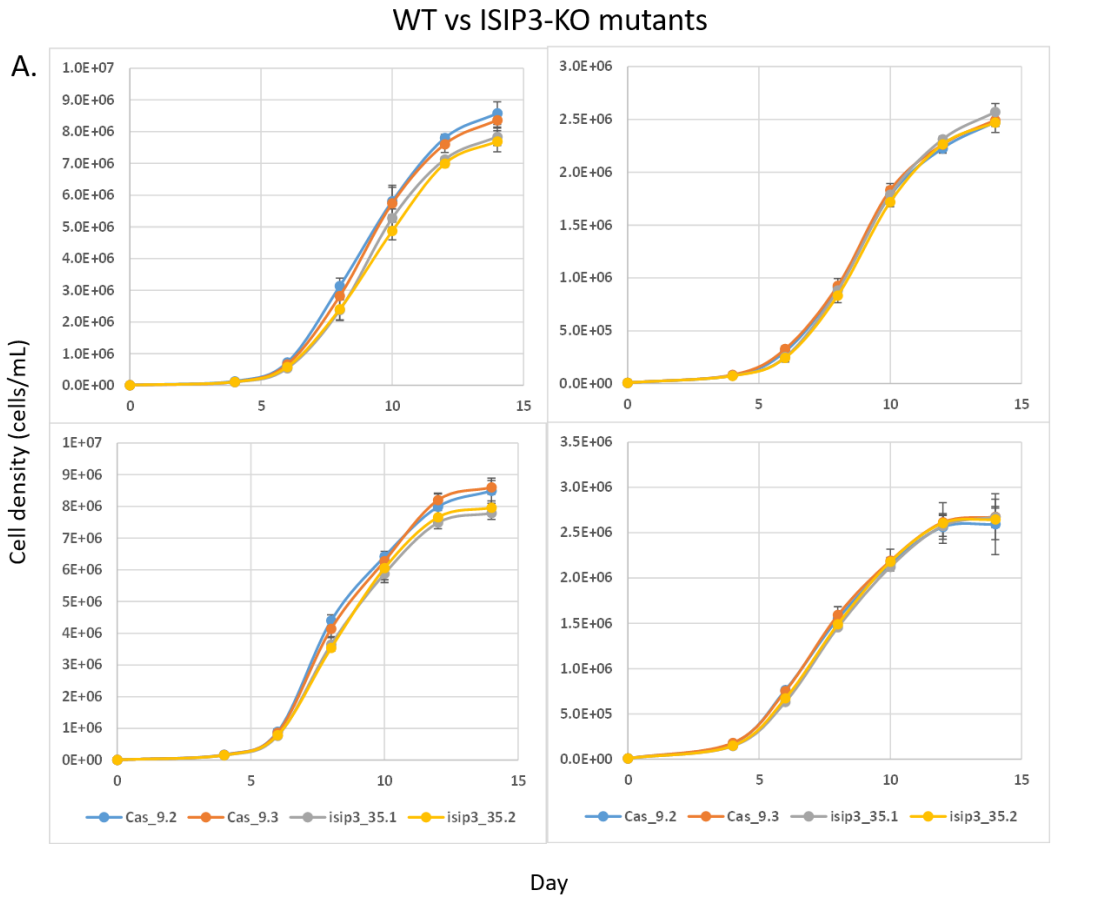
The chlorophyll content of ISIP3-KO mutants was higher than MT in conditions i and iii, increased by 97.7% and 34.6%, respectively. For conditions ii and iv, where Fe was not supplied into the cultures, the chlorophyll content was similar with MT, with variations smaller than 5.5%. The most remarkable differences were in condition i (Cells+Fe, ESAW+Fe), in which the chlorophyll contents were 436 ± 72 and 871 ± 118 pg/cell in MT and KO mutants, respectively, increased by 222.6% (Figure 4.8Fi). This obvious phenotype in chlorophyll content was consistent with the growth curve observations, with significant decrease in KO mutants in condition i.

However, the chlorophyll content was higher in ISIP3-KO mutants than MT in conditions i and ii, being increased by 43.2% and 15.8%, respectively. For conditions iii and iv, it was similar between mutants and MT, with variation smaller than 4.4%. In condition i, the chlorophyll fluorescence of MT and ISIP3-KO mutants was 4995 ± 123 and 5213 ± 205 arbitrary unit/cell, respectively, increased by 43.2% in KO mutants (Figure 4.8Gi).

Overall, in all growth conditions, different parameters showed very similar trends, with differences between MT and KO cell lines only in iron-replete media (EASW+Fe), and no differences between MT and KO lines in iron-free media (ESAW-Fe). In iron-free media, iron was not available externally for cells to take up, so cells are only expected to be able to use stored intracellular iron. In the conditions when iron was unavailable, the growth of ISIP3-KO mutants was not affected by the absence of ISIP3, indicating that the iron storage capacities of ISIP3-KO mutants was not affected. Therefore, I conclude that ISIP3 is not involved in iron storage. Instead, the marked decrease of growth of ISIP3-KO mutants in iron-containing medium might indicate involvement of ISIP3 in ferric iron acquisition (Figure 4.8E).

4.2.4 ISIP3-KO mutants grown in copper-free conditions

In the previous section I investigated the phenotypes of ISIP3-KO mutants. ISIP3-KO mutants showed significant decrease of cell density compared with MT in conditions with iron-containing media (conditions i and iii). Compared with that, in this experiment, I removed copper from all the media, both iron-replete and iron-free. Strikingly, I observed only a slight decrease of growth in ISIP3-KO mutants compared with MT, decreased by 8.3% and 7.9%, respectively, in conditions i and iii (Figure 4.9A). The specific growth rate and doubling times were consistent with growth curves, showing no obvious differences between MT and KO in conditions i and iii (variation < 5.4%, Figure 4.9 B and C). However, the relative cell size of ISIP3-KO mutants was 13.1% smaller than MT in condition iii (917 ± 14 and 797 ±37, respectively).


WT vs ISIP3-KO mutants
A.
Cell density (cells/mL)
Cas_9.2
Cas_9.3
isip3_35.1
isip3_35.2
Day
B.
Specific growth rate (day-1)
MT
mutants
C.
Doubling time (day)
MT
mutants

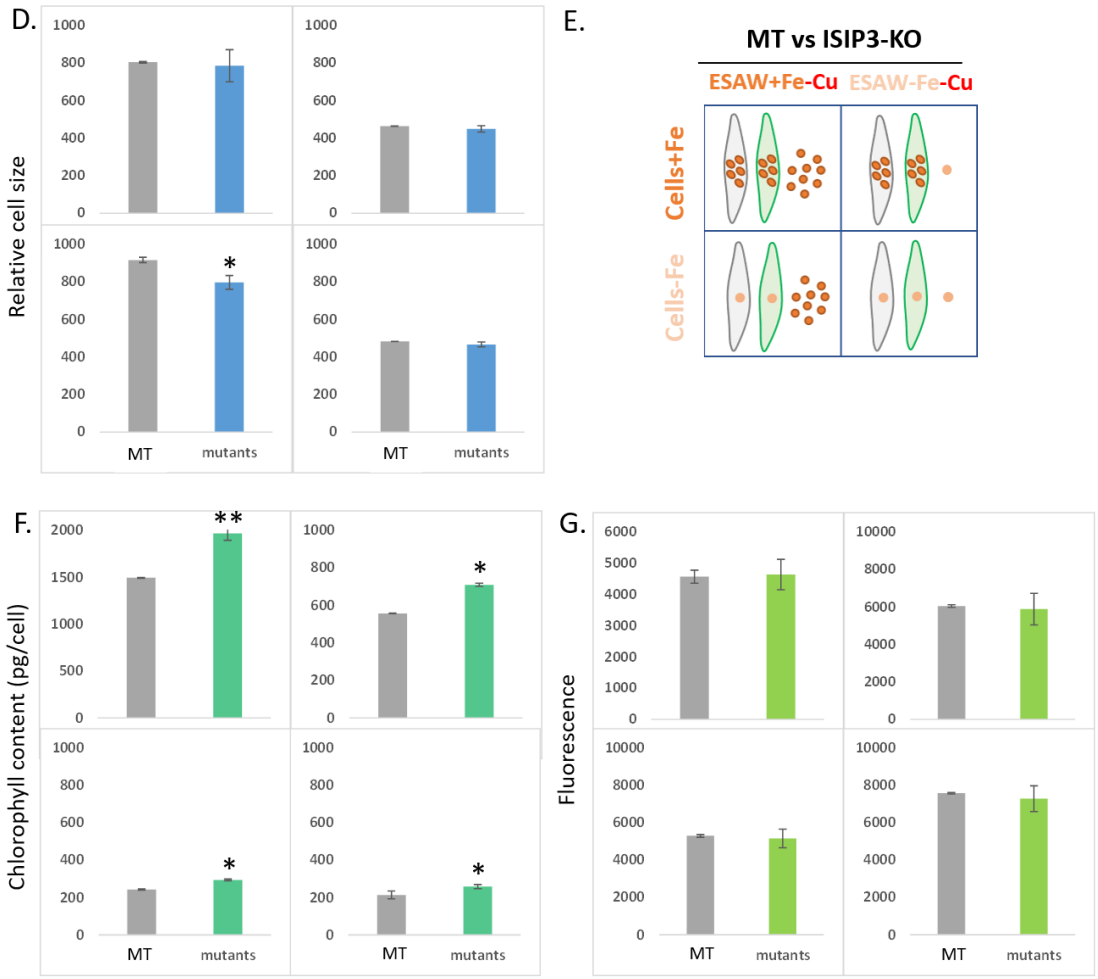


Figure 4.9 ISIP3-KO mutants versus MT cells grown in Cu-free medium.

A. growth curve; B. specific growth rate; C. doubling time; D. relative cell size; E. a diagram showing my hypothesis with MT cells in grey and KO mutants in green; F. chlorophyll content; G. chlorophyll auto-fluorescence.

I detected increased chlorophyll content in all growth conditions. The chlorophyll content in ISIP3-KO mutants increased by 31.7%, 27.4%, 21.4% and 20.8% compared with MT, respectively, in conditions i, ii, iii and iv (Figure 4.9F). Surprisingly, the chlorophyll fluorescence didn't show obvious differences in any conditions (variation < 4%, Figure 4.9G).

Copper has been suggested to be required for high-affinity iron uptake in many centric oceanic diatom species (Peers *et al.*, 2005; Annett *et al.*, 2008; Lelong *et al.*, 2013). In

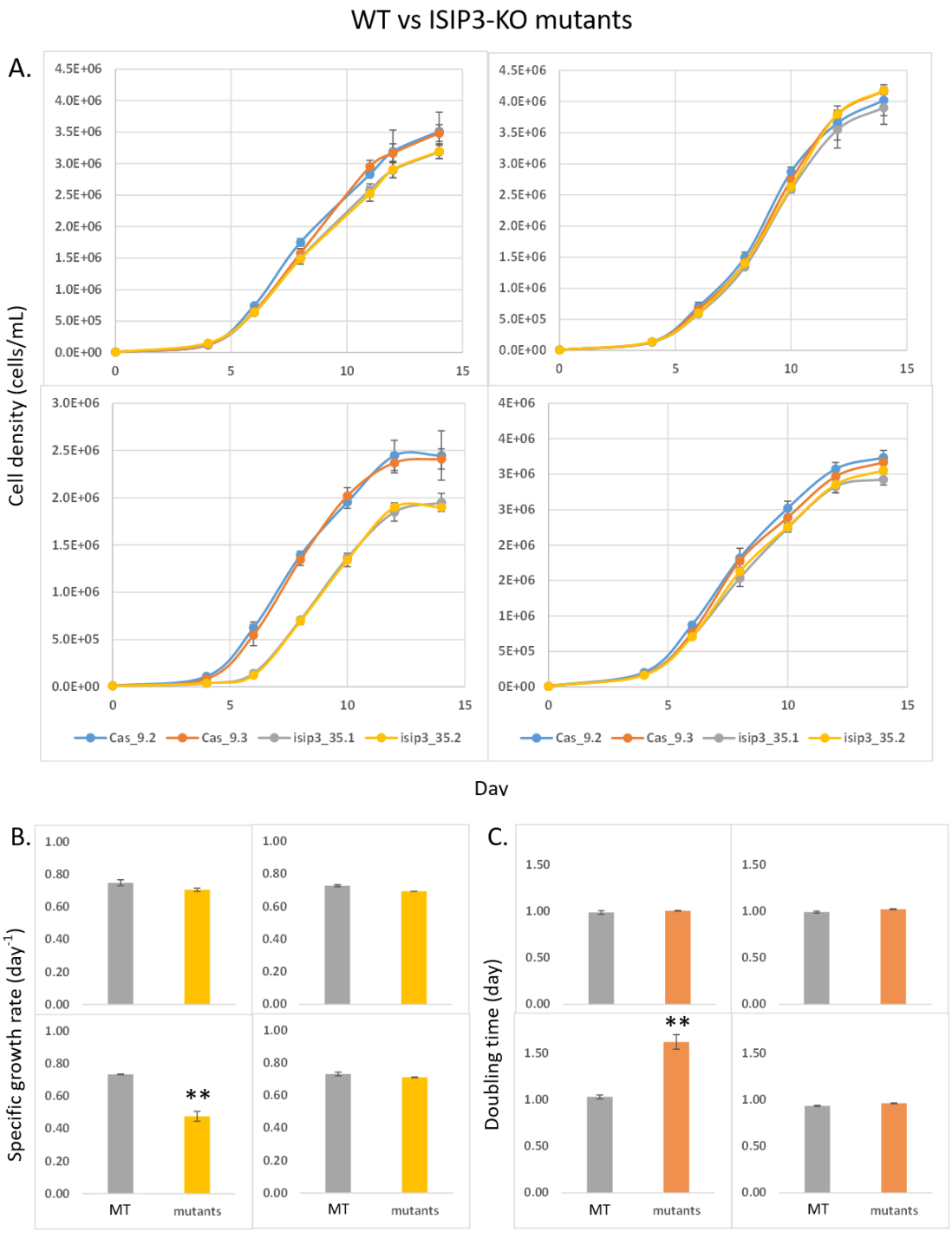
section 3.2.2, I determined that ISIP3-KO mutants showed negligible growth differences compared with MT in conditions where iron was unavailable, and significant decreases in growth when iron was available. I therefore concluded that ISIP3 was not functional as an iron storage protein but might be involved in iron acquisition. In this experiment, the phenotypes of ISIP3-KO mutants previously detected by comparison with MT, disappeared when copper was removed from the medium, suggesting the phenotypes were copper-dependent. This confirmed the hypothesis concerning ISIP3's involvement in iron acquisition in *P. tricornutum*. Further experiments on the effect of copper in *P. tricornutum* wild type strains is necessary to verify the correlation between copper and iron acquisition in *P. tricornutum*.

4.2.5 Siderophore as an iron resource for ISIP3-KO mutants

As ISIP1 was experimentally shown to be involved in endocytosis-mediated siderophore uptake (Kazamia *et al.*, 2018), and ISIP3 showed co-localization with ISIP1, I further asked whether ISIP3 is required for siderophore uptake specifically (and not just in reductive Fe uptake). Aiming to clarify this, I substituted ferric iron with siderophore FOB and FOE as iron resources (see section 4.1). Referring to Figure 4.10E, the left-hand panels were conditions supplied with FOB-containing medium, and on the right FOE (condition i: Cells+Fe, ESAW+FOB; condition iii: Cells-Fe, ESAW+FOB; condition ii: Cells+Fe, ESAW+FOE; condition iv: Cells-Fe, ESAW+FOE).

In section 4.2.2, ISIP3-KO mutants showed a decrease in growth compared with MT by 25.7% and 13.5%, respectively, with different intracellular iron status (condition i: Cells+Fe, ESAW+Fe; condition iii: Cells-Fe, ESAW+Fe). In this experiment, when FOB was supplied as iron source, the final cell density of ISIP3-KO mutants slightly decreased by 8.8% in condition i (Cells+Fe, ESAW+FOB) and significantly decreased by 20.8% compared with MT ($1.9\times10^6 \pm 5.4\times10^4$ and $2.4\times10^6 \pm 1.3\times10^5$ cells/mL, respectively) in condition iii (Cells-Fe, ESAW-FOB). This is a significant change compared with phenotypes observed in section 4.2.2. However, when supplied with FOE, the final cell density of ISIP3-KO mutants decreased by 1.4%and 6.6%, respectively, in conditions ii and iv (Figure 4.10A).

Both specific growth rate and doubling time displayed huge differences between MT and ISIP3-KO mutants in condition iii. The specific growth rate of ISIP3-KO mutants was 0.48 ± 0.03 day^{-1}, decreased by 35.2% compared with MT, which was 0.73 ± 0.01 day^{-1} (Figure 4.10B). The doubling time of ISIP3-KO mutants and MT was 1.62 ± 0.08 and 1.03 ±0.02 day, respectively, increased by 57.3% in KO mutants (Figure 4.10C). There was no obvious variation in relative cell size (<5% in all conditions, Figure 4.10D) at the end of the experiments in MT and KO cultures.



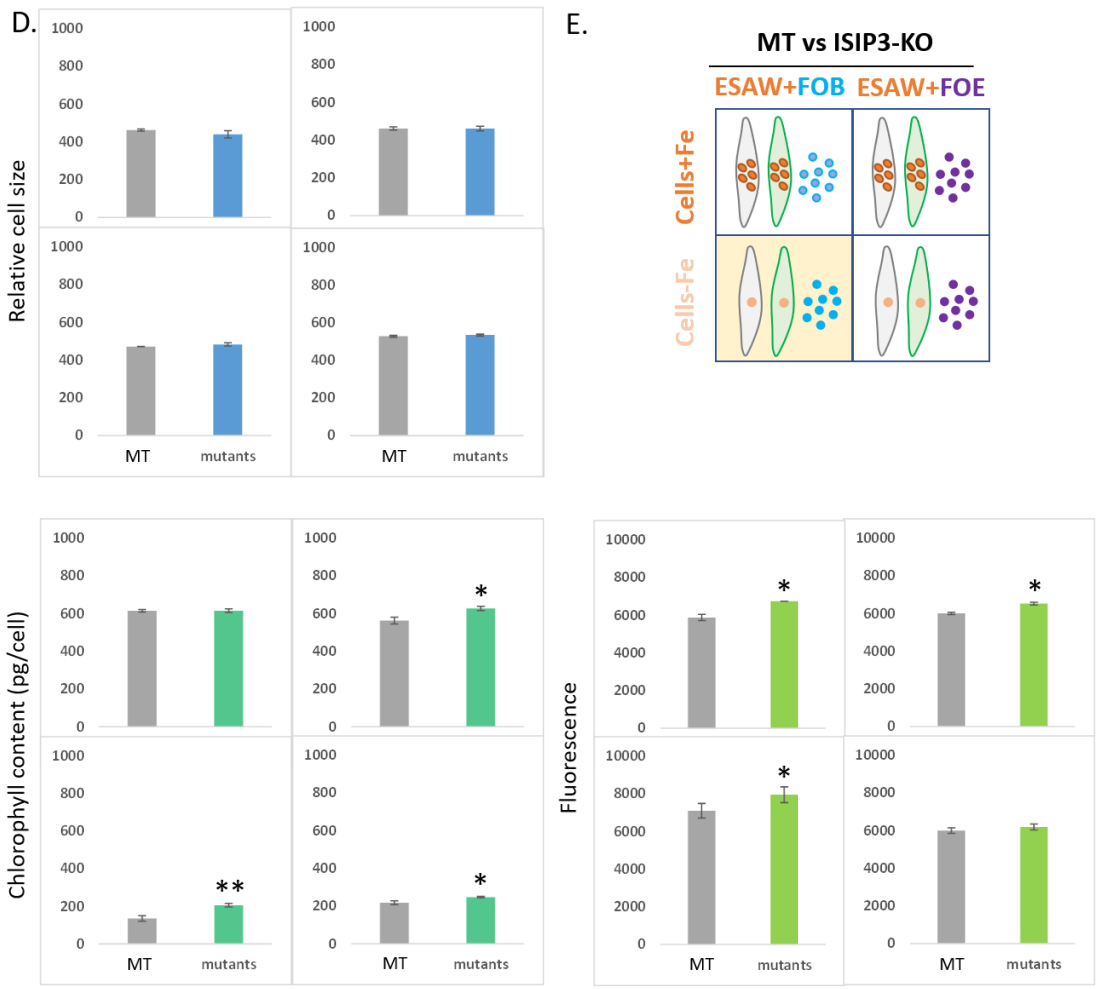


Figure 4.10 ISIP3-KO mutants versus MT cells grown in ESAW+ FOB and ESAW+FOE.

A. growth curve; B. specific growth rate; C. doubling time; D. relative cell size; E. a diagram showing my hypothesis with MT cells in grey and KO mutants in green; F. chlorophyll content; G. chlorophyll auto-fluorescence.

The chlorophyll content and fluorescence of ISIP3-KO mutants were higher than MT in all growth conditions when siderophore was supplied as the iron resource, with the most obvious increase in chlorophyll content (52.0%) in condition iii (Cells-Fe, ESAW-FOB) and in fluorescence in conditions ii and iii (14.10% and 12.0%, respectively, Figure 4.10F and G).

Since I detected a striking phenotype when siderophore was available for iron-deplete ISIP3-KO mutant cells, I therefore propose that ISIP3 is involved in siderophore uptake alongside ISIP1. Iron-replete cells mainly depended on intracellular iron for growth and

cell division, however for iron-deplete cells, the main resource of iron in experiments shown in Figure 4.10 was siderophore in the medium. By comparison, when ferric iron is provided as the iron source (Figure 4.8), the final cell density of MT reached to ~8.0×10^6 cells/mL, and with siderophores, it was ~3.5×10^6 cells/mL. I therefore confirm that *P. tricornutum* is capable of utilizing siderophore-bound iron, and exhibits species-specific recognition of siderophores (Kazamia *et al.*, 2018). The phenotypes of iron-replete ISIP3-KO mutants grown in medium with FOB but not in media with FOE indicate that *P. tricornutum* discriminates between these two siderophores. FOE is a circularised version of FOB. This adds a further nuance to our understanding of the involvement of siderophores in the bioavailable iron pool for diatoms.

Further fundamental experiments comparing the differences in growth of *P. tricornutum* wild type strains growing in medium supplied with FOB and FOE are necessary to investigate the iron acquisition mechanism of different siderophore types.

4.3 Discussion

4.3.1 Heatmap overview of phenotypes

For an easier and clearer overview of the phenotypes studied, I constructed a heatmap (Figure 4.11), in which only percentages of the increased or decreased growth of the KO mutants compared to the control, the MT, were recorded. Heatmaps have the advantage of quickly and easily identifying differences in large-scale data analysis from parallel experiments.

The phenotypes of FTN-KO and ISIP3-KO mutants were significantly different. FTN-KO mutants showed a significant decrease in growth in conditions compared to MT when iron was not available for iron-replete cells, and only a slight decrease in growth in all other conditions. However, for ISIP3-KO mutants, phenotypes appeared only when iron was available, for either iron-replete or iron-deplete cells. There was no discernible phenotype in the mutants compared to MT empty vector controls when iron was unavailable.

Phenotypes in final cell density, specific growth rate and doubling time were based on

growth curves and were consistently correlated with each other. As to the relative cell size, chlorophyll content and fluorescence, in ISIP3-KO mutants, they were largely consistent with growth, however, in FTN-KO mutants, they were not predictable, with most significant decreases in cell density in condition ii (Cells+Fe, ESAW-Fe), whereas cell size and chlorophyll fluorescence differences were apparent in condition iii (Cells-Fe, ESAW+Fe). Metabolisms are linked through complex intracellular pathways, therefore further investigations are required to map these coregulations between iron acquisition, iron storage and photo-physiology.

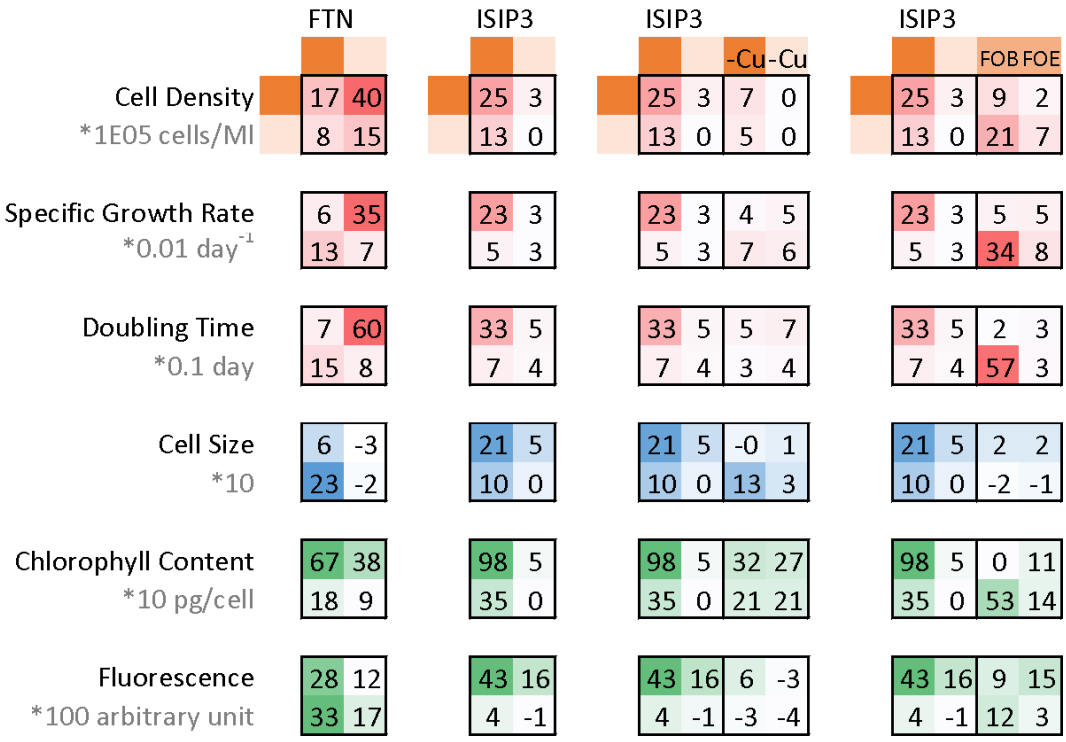


Figure 4.11 Heatmap of iron-dependent phenotypes of FTN and ISIP3 KO mutants.

The shade of colour and numbers within indicate the percentage of increase or decrease of each parameter in KO mutants comparing to MT. The first column on the left shows the phenotypes of FTN-KO mutants in four conditions (iron availability and intracellular iron status described in 2.1); then the second column, ISIP3-KO mutants; in the third column, the left grids (normal medium) was compared with the right grids (copper-free); in the last column, the left grids (ferric iron as iron resource) was compared with the right grids (siderophore FOB and FOE).

4.3.2 Conclusions and perspectives

In conclusion, the work described in Chapter 4 of this dissertation provides insights into the study of function of FTN and ISIP3 in *P. tricornutum*. The absence of FTN has an overall dampening effect on the growth of the culture, with lower growth rates at mid-exponential phase and lower overall carrying capacity reached in stationary phase. The most obvious effect was observed when iron deplete cells were grown in iron-free medium. It indicates that FTN is involved in iron storage. ISIP3 KO mutants showed a completely different effect, with no phenotype in KO mutants when grown in iron-free medium, no matter the initial iron-replete or iron-deplete starting conditions for the cells, and a significant decrease of growth observed for KO mutants grown in iron-containing medium. This suggests that ISIP3 is not involved in iron storage, but more likely in iron acquisition. The loss of phenotype of KO mutants grown in copper-free medium further supported the conclusion.

However, further experiments are required. For FTN-KO mutants, total cell iron quota will be strong evidence to confirm the function as iron storage protein. A further important experiment will be to complement FTN-KO mutants of *P. tricornutum* with FTN from *Pseudo-nitzschia,* in order to test whether PmFTN can fully replace PtFTN. FTN-KO mutants. For ISIP3-KO mutants, experiments on iron uptake rates using radiolabelled Fe, are a powerful tool for garnering evidence for the protein's hypothesized involvement in iron acquisition. The determination of both cellular iron quota and cellular iron uptake rates requires a trace metal clean experimental environment, which is challenging and outside the scope of capabilities at IBENS.

4.4 Materials and methods

4.4.1 Diatom cultivation

Diatom strain

Experiments were carried out with strain *Phaeodactylum tricornutum* CCAP1055/1.

Culture conditions

Cultivation of wild-type and mutants of *P. tricornutum* were performed under a 12-hour light/12-hour dark cycle, with an average intensity of 100 µmol photons/m^2/s during the light phase on incubator with 100 rpm shaking speed. Cultures were maintained at a temperature of 18°C at a volume of 10 mL in single use transparent plastic flasks.

Culture medium

The growth medium was enriched artificial seawater, ESAW (Falciatore *et al.*, 2000). Its molecular composition is given in (Table 4.1).

Table 4.1 Molecular composition of medium ESAW

ESAW	Composition	Concentration (moles/L)	g/mol	g/L	STOCK	mL	g			
	Ca.Cl2	9.14E-03	147.00	1.344						
	K.Cl	8.03E-03	74.55	0.599						
	Mg.Cl2·6H2O	4.72E-02	203.30	9.592						
	Na2.SO4	2.50E-02	142.04	3.550						
	Na.Cl	3.63E-01	58.44	21.194						
N source	Na.NO3	5.49E-04	84.99		1:1000	50	2.334			
P source	NaH2.PO4	2.58E-05	119.98		1:1000	50	0.155			
C source	NaHCO3	2.07E-03	84.01		1:1000	50	8.700			
chelators	Na2.EDTA	8.30E-06	372.24		1:1000	50	0.155			
halogens	K.Br	7.25E-04	119.00		1:1000	50	4.315			
	Sr.Cl2	8.18E-05	266.62		1:1000	50	1.090			
	NaF	6.67E-05	41.99		1:1000	50	0.140			
	H3.B.O4	3.72E-04	61.83		1:1000	50	1.150			
silica	Na2.SiO3-9H2O	1.75E-04	284.20		1:1000	50	2.487			
Fe	ferric EDTA	8.69E-06	345.07		1:1000	50	0.150			
vitamins					1:1000	50				
	cobalamin (B12)	7.38E-07	1355.38				0.050			
	biotin (B7)	4.09E-06	244.31				0.050			
	thiamine (B1)	5.93E-04	337.26				10.00			
trace metal					1:1000	50		STOCK	mL	g
	Mn.Cl2	2.73E-06	197.91					1:100	50	2.700
	Zn.Cl2	2.54E-07	136.28					1:1000	50	1.730
	Na2.MoO4	6.12E-09	241.97					1:10000	50	0.740
	Co.Cl2	6.72E-08	237.93					1:1000	50	0.800
	Ni.SO4	6.27E-09	280.88					1:10000	50	0.880
	Na2.SeO3	1.00E-09	172.95					1:100000	50	0.865
	Cu.SO4	3.92E-08	249.72					1:1000	50	0.490

To prepare the medium, all reagents were handled only using plasticware in a trace metal clean environment to avoid iron contamination, because glass bottles can absorb heavy metals and medium can be contaminated by metal tools, like iron spatulas. Dense solutions were stored in Falcon tubes for 6 months maximum at room temperature

avoiding light, except the vitamins, which were stored at 4°C. I used filtration units (Nalgene™ Rapid-Flow™ Disponsible Sterile Filtration Units with PES Membrane) to sterilize the medium instead of autoclaving, since the autoclave process is also known to contaminate media solutions with iron.

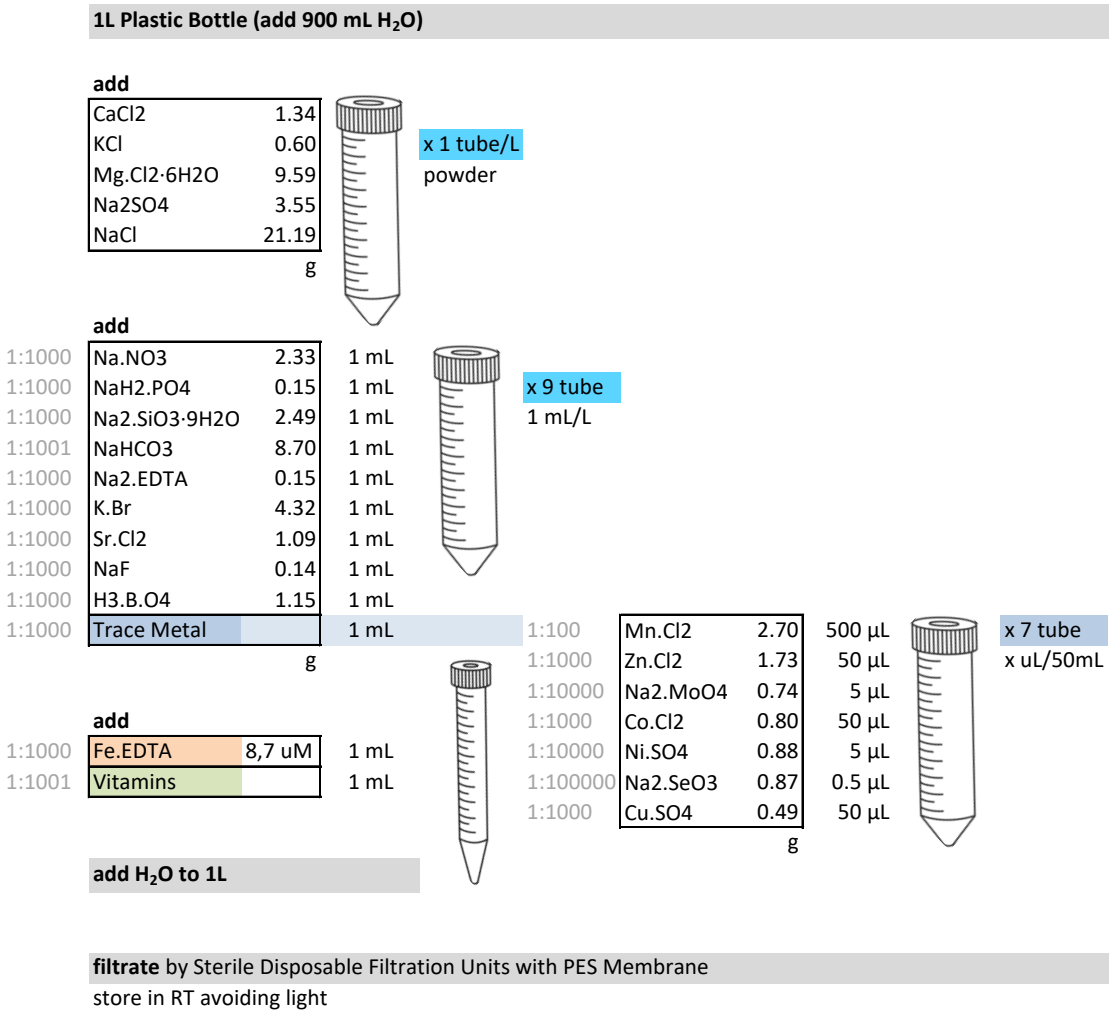


Figure 4.12 The procedures for preparing the medium

Salts required for media preparation were measured out, brought into solution and kept in Falcon tubes. Concentrated solutions of microelements and trace elements were kept in Falcon tubes at room temperature in the dark. Fresh media was always mixed from the stocks and sterilized just before the experiments.

DMSO cryopreservation

Once the knockout mutants had been selected, cryopreserved stocks were made from cultures. These were important as back-ups in case of laboratory accidents (for example

bacterial contamination of cultures) and make it possible to recover cultures when phenotypes change after many generations of cultivation due to acclimation.

First, 30% DMSO solution was made from 10 mL DMSO and 23.33 mL ESAW, mixed in 50 mL falcon tube. Fresh and healthy cultures in exponential phase for cryopreservation were pre-cultivated. Then 4 mL 30% DMSO solution and 6 mL healthy exponential cultures were mixed in 10 mL falcon tube and aliquoted to 1.8 mL cryotubes. So, the final concentration of DMSO was 12%. Following the standard slow cooling cryopreservation protocol, the cryotubes were first put in freezer 4℃ for 1 hour, then put in -20℃ for 1 hour and stored at -80℃. The last but optional step was to test the recoverability by recovering cells from these cryotubes.

Culture maintenance and pre-cultivation

All cell lines were maintained in medium ESAW supplemented with zeocin. Cultures were propagated continuously by refreshing the culture with a dilution of 1:100 every week. Contamination tests were carried out every three months. For the growth curve comparisons, cells were pre-cultivated with different treatments for two weeks before being used as original cells in growth curve experiments.

Cells were washed with iron- and copper-free medium by centrifugation at 2500 rpm at 15°C for 6 min (which was much gentle for cells than 4000 rpm 10 min) for three times. For the last wash, one drop of 2 mol/L Na_2EDTA was added to collect any free iron in solution. Subsequently, cells were counted using a Malassez counting chamber under the microscope. The original cell density of growth curve was 10^4 cells/mL.

Contamination test and decontamination

Maintaining good growth of algae cells is a prerequisite for any experiment. So, it is necessary and important to check the contamination regularly to make sure the cells are growing well. Both bacteria and the nuclei of diatom cells can be stained by DAPI. For contamination test, 3 µL cell culture was added to the slide with 1 µL 20 ug/mL DAPI, mixed and left for 10 min until cells were stained before being observed under a fluorescent microscope. The wavelength of excitation and emission were 360 nm and 475 nm, respectively. Contaminated cultures were reselected by replating diluted cultures on plates with antibiotics.

4.4.2 Knockout mutants

CRISPR-Cas9 is a powerful tool for genome manipulation and has been adapted for gene editing in marine algae (Nymark *et al.*, 2016). Knockout mutants of *ISIP3* and *FTN* were generated by CRISPR-Cas9 and bombardment techniques, verified by genotyping and western-blot.

CRISPR-Cas9

Coding sequences of genes were downloaded from the website Ensembl Protists (http://protists.ensembl.org/Phaeodactylum_tricornutum/Gene/Sequence). CRISPR-Cas9 specific targets for guide RNA were designed on CRIPOR (*http://crispor.tefor.net/crispor.py*). Primer pairs were designed to replace the fragment of the target site within the pU6-sgRNA vector, consisting of target and overlapping sequences. The sequence of pU6-target-sgRNA was checked before bombardment.

Bombardment

A *P. tricornutum* cell culture (500 µL) adjusted to 1×10^8 cells/ml was plated on ½ ESAW agar plates and grown for 1 day. The pU6-target-sgRNA vector and pDEST-hCAS9 vector were introduced to the cells by biolistic bombardment using the Biolistic Particle Delivery System (Bio-Rad, Hercules, CA, USA) as previously described (Falciatore *et al.*, 1999). The cells were co-transfected with the Paf6/pPhat1 vector to enable selection of transformants on ½ ESAW agar plates containing 100 µg/ml zeocin. Tungsten M17 microcarriers (Bio-Rad) were coated with 2.5 µg of each vector following the manufacturer's instructions (Bio-Rad). Cells were transferred to selection plates 1 day after bombardment. Resistant colonies appearing 3–4 weeks after bombardment were transferred to liquid ESAW containing 50µg/ml zeocin.

Screening for gene targeted mutations

For screening, zeocin resistant clones showing sequence "ambiguities" in the target region (based on sequencing of PCR products) were transferred to liquid cultures and spread on agar plates to obtain single colonies.

Western blot

The Anti-PtISIP3 antibody was a gift from Andrew E. Allen's group (J. Craig Venter

Institute, Scripps Institution of Oceanography). The Anti-Ferritin Heavy Chain antibody (ab89787) was ordered from ABCAM. The positive control was a protein extract from human cells (Hela, A549 and HEK_293). Anti-Histone H4 antibody was used as reference.

4.4.3 Flow cytometry

The Flow Cytometer (FCM) that I used is The Partec CyFlow® Cube, which is fully equipped with operation CyView® software, and runs with an internal PC. Data acquisition, instrument control, and data analysis are controlled and performed by the software (Figure 4.13).

Configuration

Channels: FSC (front scatter) and SSC (side scatter) recorded as relative particle size were activated by 532 nm wavelength laser light and FL3 as green fluorescence by 488 nm wavelength laser light.

Plots: FL3-FSC dot plot is set to present the relative size (FSC) on a logarithmic scale in X and fluorescence on a linear scale in Y.

Results: The particle numbers, the mean, CV and median of cell size and fluorescence of certain regions, and the volume that was tested are shown on the bottom-right.

Regions: The dark blue circle up in FL3-FSC dot plot was set as the region of diatom cells, by which the fluorescence was separated from other particles in the medium.

Script - Measure mode: Plates 40-100, autostart autosave

Process: 2 uL/sec, 40 uL, Maximum Acquisition Speed 25,000 events/sec. Dilute the culture when the events exceed the maximum acquisition speed.

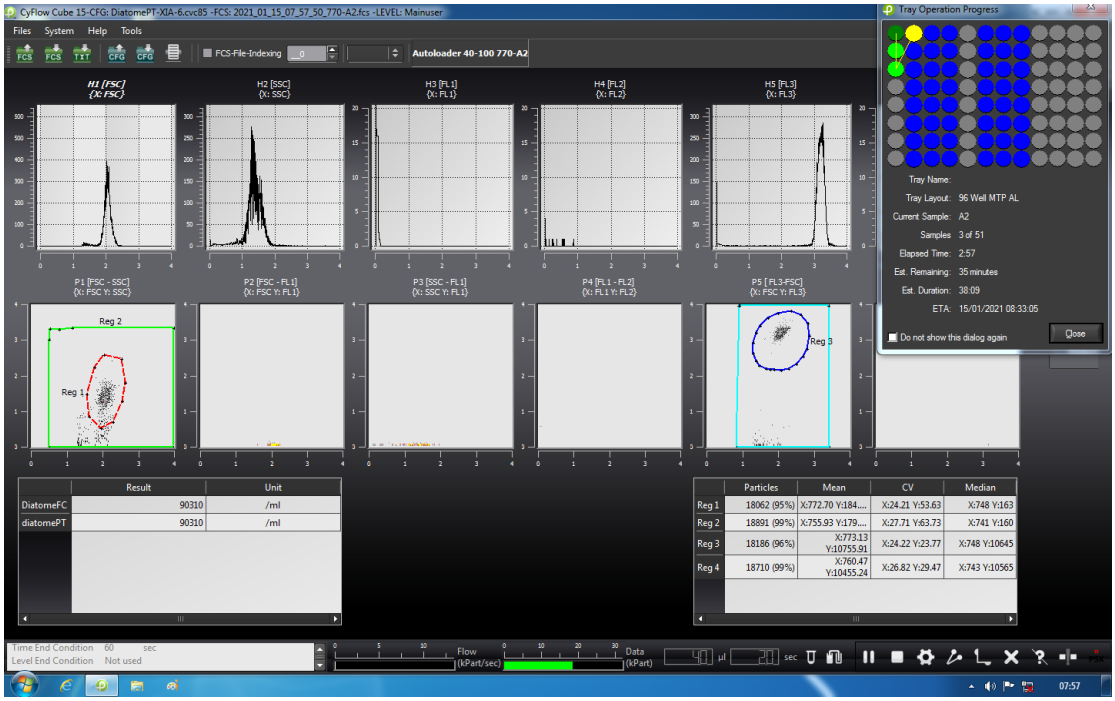


Figure 4.13 Flow Cytometer Configuration

4.4.4 Chlorophyll content

The chlorophyll a content was determined using the method described in (Xie *et al.*, 2014). Instead of using ethanol or methanol, DMSO was used to extract chlorophyll, which was toxic but of high efficiency. A cell pellet of 2 mL cell culture was collected by centrifuge. Chlorophyll was extracted by adding 1 mL DMSO to the cell pellet and incubating at 60 °C for 10 min. The dissolved chlorophyll in solution was centrifuged again and supernatant was ready for measurement. The absorptions at wavelengths (664nm, 750nm, 647nm, 630nm) were measured and the chlorophyll contents were calculated according to modified equation:

$$\text{Chl } a = 1.98\,[11.85(OD_{664} - OD_{750}) - 1.54(OD_{647} - OD_{750}) - 0.08(OD_{630} - OD_{750})]$$

4.4.5 Standard curves

To test the reliability of the methods, diatom cultures were diluted into a gradient of concentration for flow cytometer and chlorophyll content measurements (Figure 4.14A).

Figure 4.14B shows the standard curve of the particle size, the relative cell size and fluorescence recorded by flow cytometer, from which I can see great dependence of cell counting to concentration, and the reading of cell size and fluorescence are very stable. Figure 4.15 shows the correspondence of chlorophyll content to cell concentration. Therefore, I was confident that the methods of flow cytometry and chlorophyll content are reliable for tracking the growth curve.

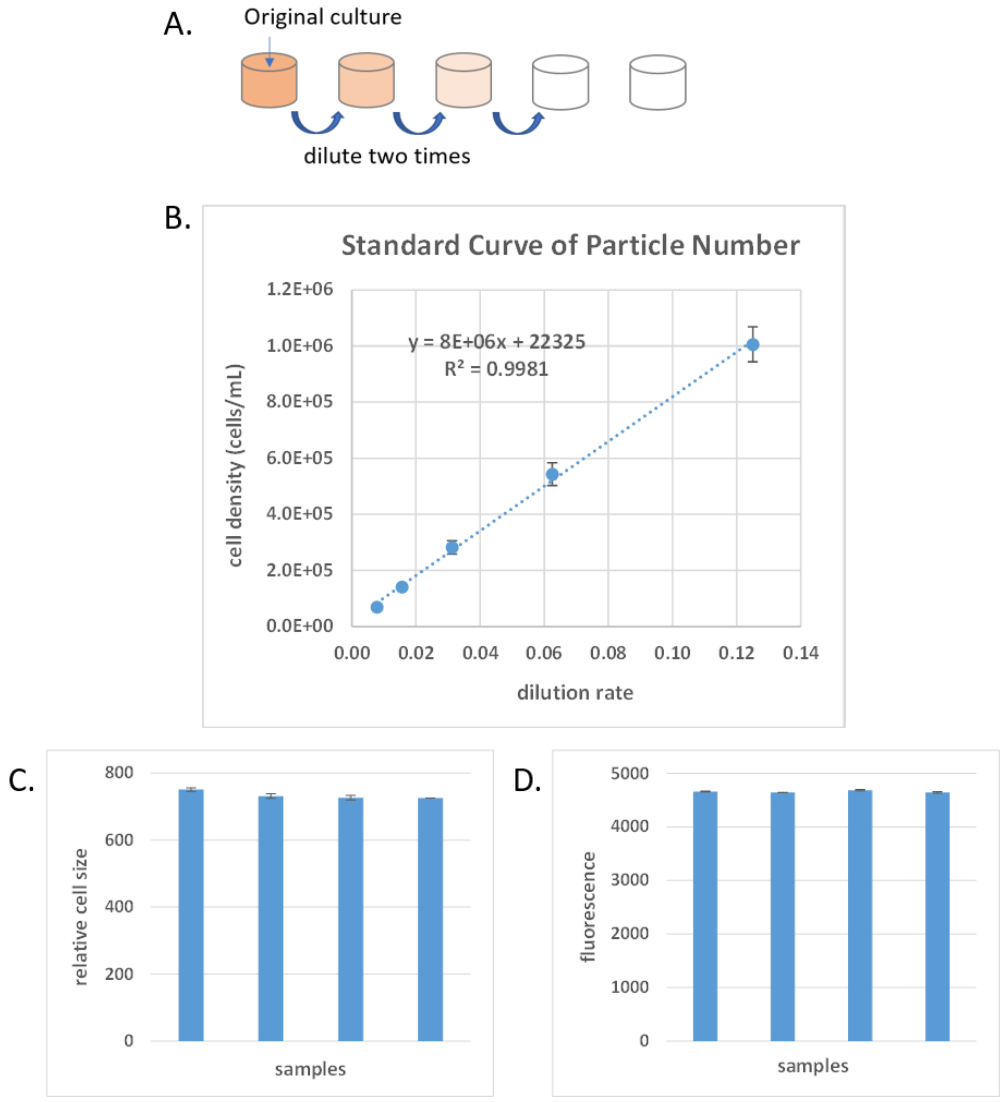


Figure 4.14 Standard curves from low cytometry

A. dilution of diatom cultures; B. standard curve of particle numbers; C. relative cell size values; D. fluorescence values.

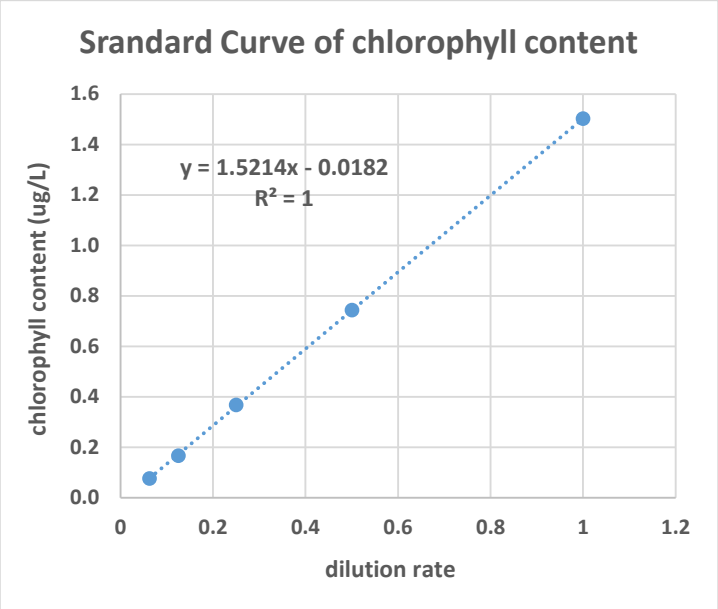


Figure 4.15 Standard curve of chlorophyll content.

Chapter 5: Conclusion and perspectives

5.1 Conclusions

5.1.1 FTN functions in iron storage in *P. tricornutum*

The bioinformatic analysis presented in Chapter 2 revealed that FTN of *P. tricornutum* fell into a third distinct clade in a FTN phylogenetic tree comprising all sequenced diatoms, whereas FTN from *Pseudo-nitzschia* and *Thalassiosira* species were mainly found in clades I and II. This distinction updates the existing consensus on FTN phylogeny in the diatoms, and begs the question whether all diatom ferritins share the same function. Are there functional differences between the FTNs belonging to different clades?

In initial experiments I identified that the cell density of FTN-KO mutants decreased significantly when iron-replete cells were grown in iron-free medium. The effect of absence of FTN on diatom growth was observed regardless of iron status and availability, and on chlorophyll regarding different iron status and availability. It indicates that FTN may function as an iron storage protein in *P. tricornutum*, as in *P. granii* (Marchetti *et al.*, 2009), and may be used in iron homeostasis, as observed in another diatom species, *P. multiseries* (Pfaffen *et al.*, 2013; Pfaffen *et al.*, 2015). These studies are not conclusive and further analysis into total cell iron quota are required. But KO mutants are a great resource for the community of diatom specialists and the findings here pave the way forward.

Subcellular localization of most FTNs in diatoms was predicted by targeting software to be in the chloroplast. These analyses however have all been bioinformatic inferences and to the best of my knowledge, no physiological validation has been provided in any diatom. The generation of the FTN-GFP line allowed me to test this hypothesis directly. For the first time, I was able to confirm that FTN of *P. tricornutum* localized to the chloroplast, where iron is mainly required for photosynthesis. The pattern of localization is intriguing. The FTN looks to be aggregated in ‘granule’-like structures, and the ultrafine structure of these putative storage structures is highly interesting. The findings here provide insight into the function and localization of FTN in diatoms.

Further experiments with higher resolution, like correlative light-electron microscopy (CLEM), are planned to investigate the precise structure.

The physiological growth curve experiments presented in chapter 4 compared the growth of MT and FTN-KO mutants in media either with iron or without iron. They shed light on the possible function of FTN inside the cell. There is a competing hypothesis for the role of FTN in diatoms. Some argue that FTN functions as a storage protein, like it does in humans and other animals, whereas others suggest it may have a role in iron homeostasis. The biggest difference is in the time period over which the protein is involved. For storage, the effect would be expected to be seen over a longer term, and presence of FTN could in theory allow diatoms to survive prolonged periods of iron scarcity. A homeostasis hypothesis however, suggests a continuous need for FTN use and function. In this case the role of FTN would be more directly linked to the short-term availability of iron to the cell, *e.g.,* its immediate microenvironment. My observations of studying the physiological behaviour of FTN-KO mutants compared with MT strongly suggest that FTN is implicated in long-term iron storage, rather than short-term homeostasis. The strongest growth impairment of mutants was observed in my experiments in the condition when iron was unavailable to already starved, iron-depleted cells. In other words, when iron storage would be most important to cell physiology. My results are preliminary evidence that FTN is functional as an iron storage protein.

5.1.2 ISIP3 is required for iron acquisition in *P. tricornutum*

Since ISIP3 contains a ferritin-like domain, it was a reasonable first assumption that this protein may also function as an iron storage protein. In chapter 3 I demonstrated experimentally that ISIP3 localizes to the centre of the chloroplast, but in a manner different to FTN. While FTN is found within the chloroplast in ‘granules’, ISIP3 aggregates at the centre of the chloroplast, possibly in the periplastidial space. This is also the place where I detect both ISIP1 and ISIP2A (Figure 3.2, refer to section 3.2.1). ISIP3 has almost perfect co-localization with ISIP1. Furthermore, when comparing the phenotypes of ISIP3-KO mutants with FTN-KO mutants, it was obvious that the two proteins have vastly different functions in the cell. In Chapter 4 I presented data that

show the growth of ISIP3-KO mutants significantly decreased in iron-containing medium, especially in iron-replete cells. In contrast, I found no phenotypes when cells were cultivated in iron-free medium, regardless of the intracellular iron status of the cells. This was the condition where FTN-KO mutants were most sensitive, compared to MT.

The conclusion therefore is that despite the presence of ferritin-like domains in ISIP3, there is little similarity in action between ISIP3 and the canonical ferritin protein in *P. tricornutum*. My results indicate that ISIP3 is not required for iron storage, but instead is likely to be involved in iron acquisition. The role of ISIP3 in iron acquisition is further supported by growth curve experiments in copper-free media. Copper is known to be required for iron acquisition, and in experiments using ISIP3 KO mutants, previously observed phenotypes in response to iron availability disappeared when copper was removed from the growth media. This means that the iron sensitive phenotypes detected, were copper dependent, supporting the hypothesis that ISIP3 is involved in iron uptake. As further supporting evidence, there is intriguing protein alignment between ISIP3 and the copper-binding metallochaperone, CopM (section 4.1.2). Further analysis into iron uptake rates is required. But the findings provide the first evidence on the physiological role of ISIP3 in diatoms and pave the way forward.

In experiments where the source of iron provided to diatoms was siderophore instead of dissolved ferric iron, the significant decrease in cell density of ISIP3-KO mutants compared with MT, was observed when siderophore was provided to iron-deplete cells. I think this was because cells tend to use intracellular iron when ferric iron was unavailable, so uptake of siderophore in ISIP3-KO mutants was affected as well. The localization of ISIP3, targeting to chloroplast and aggregating at the centre, further indicates that ISIP3 may be involved in endocytosis-mediated iron acquisition, like ISIP1 with which it perfectly co-localizes. But these studies are not conclusive. Further visualization of these proteins by microscopy at higher resolution, and comparison of the localization profiles under supplementation of a various range of iron sources, including siderophore, as well as short-term Fe^{2+} spikes, are required.

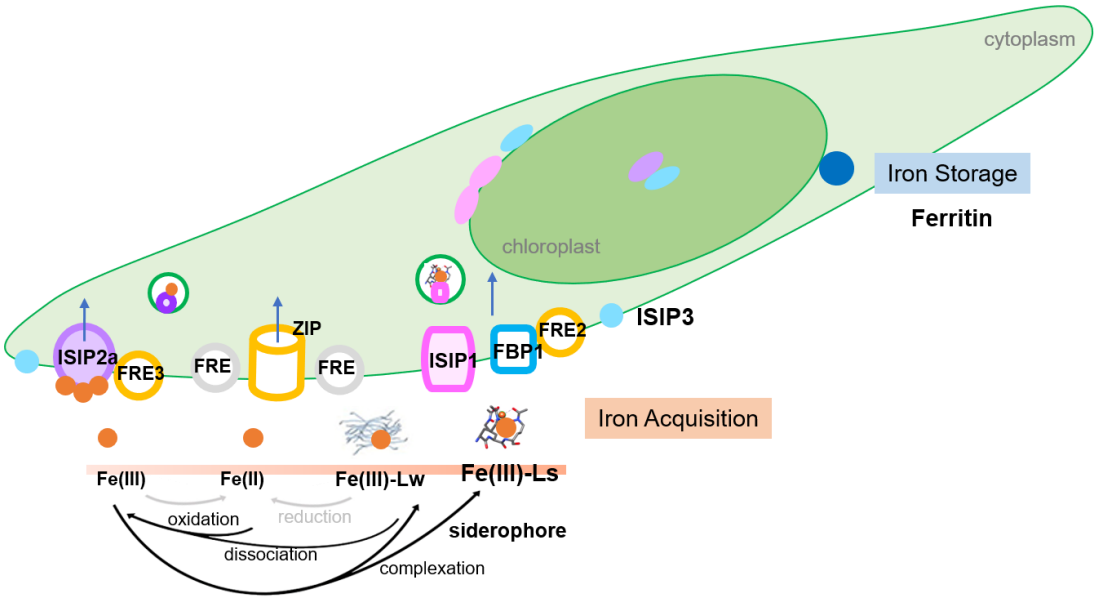


Figure 5.1 Schematic diagram of ISIPs and FTN localisation in *P. tricornutum, based on current data*.

5.1.3 Portfolios of iron proteins are not consistent with diatom phylogeny

In addition to *ISIP3* and *FTN*, I investigated 12 additional iron metabolism related genes across 87 diatom species through bioinformatic analysis on transcriptomes and genomes (presented in Chapter 2). I found that the replacement of cytochrome c_6 with plastocyanin is rarer than the replacement of ferredoxin by flavodoxin, and the ability to synthesise flavodoxin is widespread amongst diatoms. However, there appears to be no phylogenetic relationship between species in encoding for these genes, suggesting a complex evolutionary history of gene gain and gene loss.

In the literature as a whole, the majority of studies focus either on individual model species or attempt to draw distinction between ‘pennates’ and ‘centrics’, ‘open ocean’ and ‘coastal’ diatoms. For the iron uptake, homeostasis or storage mechanisms that I investigated, there appear to be no remarkable patterns that distinguish pennates from centrics. Additionally, there is no consistent pattern of phylogenetic relationships between diatom groups. This indicates that there is no “one rule fits all” scenario for diatom iron strategies, even at genus level. Species belonging to the same genus can

exhibit very different portfolios of iron proteins and iron strategies in the wild. For example, in the *Thalassiosira* genus, the ratio of PSI : PSII was found to be significantly lower in open ocean diatom species *T. oceanica* compared with the coastal diatom *T. weissflogii. T. weissflogii* cells contained twice as much PSII compared with PSI, and *T. oceanica* had reduced PSI demands even further, with fivefold lower PSI, that is a ratio of 1:10 of PSI : PSII (Strzepek and Harrison, 2004). A further example from the *Pseudo-nitzschia* genus, of which there were 6 representatives in our study is noteworthy (Figure 2.1). Half (3 species) encoded ISIP1 whereas 3 missed the annotation. Species with widely different iron portfolios, within the same genus, might indicate the rapid evolution and diversification driven by geographic separation in the ocean. Rapid evolution is also likely especially in the context of iron metabolism, since adaptation to a fluctuating iron supply is considered to drive diatom success in the wild. To meaningfully compare open ocean and coastal diatoms, more information on the distribution of individual species in the global ocean is required. The *Tara* Oceans global circumnavigation effort that catalogued the community composition, metagenomes and metatranscriptomes of aquatic microorganisms can offer initial insight.

5.2 Perspectives

5.2.1 qPCR and RNAseq

Molecular methods are being increasingly applied to detect, quantify and study gene function in diatoms. Among these methods, PCR-based techniques have been the subject of considerable focus. More particularly, quantitative PCR (qPCR) is considered as a method of choice for the detection and quantification of the expression level of specific genes. One of its major advantages is to be faster than other methods. It is also highly sensitive, specific and enables simultaneous detection of cells grown in different conditions.

With the full suite of FTN-KO mutants, ISIP3-KO mutants and MT (transformed with empty vector) strains, it will be interesting to screen the regulation network of iron-related genes in *P. tricornutum*. I have undertaken initial steps to compare the

expression level of some iron-related genes of the aforementioned strains using qPCR. The strains were grown in iron-containing and iron-free media (data not shown). Such work follows directly as a consequence of the experiments presented in this thesis. Once I have qPCR data the path forward is to consider RNAseq whole transcriptome experiments as an advanced holistic way to investigate and compare diatom iron metabolism between MT and mutant strains.

5.2.2 CLEM for fluorescence visualization

The visualization of fluorescent proteins in living cells is a powerful approach to study intracellular dynamics. A limitation of fluorescence imaging, however, is that it lacks fine structural information; a fluorescent spot could represent an entire organelle, an organellar subdomain or even aggregates of proteins or membranes. These limitations can be overcome by correlative light-electron microscopy (CLEM), which bridges the gap between live-cell imaging and electron microscopy (EM). Fluorescently-tagged proteins are first imaged by light microscopy and then visualized via EM.

CLEM work using the fluorescent lines generated as part of my doctoral work is underway. Here, I am taking advantage of a collaboration with the European Molecular Biology Laboratory (EMBL), which I helped establish. The objective for CLEM is to monitor the kinetics and localization of FTN, ISIP1, ISIP2a and ISIP3 at nanometer resolution. For the ISIP proteins a question of particular interest is whether the consistent aggregation of these proteins to the vicinity of the chloroplast occurs in a dedicated organelle, the periplastidial space, or a vesicle. I plan to uncover a precise structure from electron microscopy with the localisation of the ISIP proteins, which will show whether the proteins are in a structure surrounded by membranes – either self confined, or associated with the chloroplast membranes? Using a complementary method, I plan to quantify the levels of iron within this microenvironment.

Additionally, I am interested in revealing the structure of the ferritin globules/granules. In an ongoing effort I am attempting to resolve a technical issue, cell auto-fluorescence, which is currently too bright and interrupts the CLEM imaging process severely.

5.2.3 Iron uptake and iron quota

The phenotypes of FTN and ISIP3 KO mutants are very intriguing and require further physiological experimental verification. The current assumption as a result of experiments presented in this dissertation is that FTN functions as an iron storage protein in *P. tricornutum*. Intracellular iron quota of cells growing in medium with available iron compared with MT will be the most direct and strong supporting evidence to this hypothesis. As to ISIP3, I assume it to be involved in iron acquisition. Comparison of iron uptake rate of cells of ISIP3-KO mutants and MT under different iron treatments combining isotope tracer technology can address this hypothesis directly. However, the determination of both cellular iron quota and cellular iron uptake rate provides a very strict demand for experimental environment, which are outside the scope of ENS capabilities. I plan to carry out these experiments through our collaboration with Dr. Robert Sutak at the Biotechnology and Biomedicine Center in Vestec (BIOCEV) institute of Charles University in Prague.

5.2.4 Photosynthetic activity

Photosynthesis is the fundamental mechanism for cell growth. Although the photosynthesis process is well understood in plants, the mechanisms evolved by phytoplankton to achieve extremely efficient photosynthesis are just starting to come to light.

Diatoms with low metabolic iron quotas are remarkable for their ability to maintain uncompromised photosynthetic function and this is attributed to two main adaptations: preferred use or complete replacement of iron containing proteins with equivalents that are not dependent on iron (Pankowski and McMinn, 2009; La Roche *et al.*, 1995; Erdner and Anderson, 1999; McKay *et al.*, 1999; Allen *et al.*, 2008), and re-arrangements to photosynthetic architecture (Strzepek and Harrison, 2004). Diatoms additionally utilize a different strategy in the dissipation of excessively absorbed energy compared with green algae (Wagner *et al.*, 2006). Thylakoid architecture might evolve independently in the land and the ocean. The photosynthetic complexes in typical diatoms are segregated in the loosely stacked thylakoid membranes. Separation of photosystems within subdomains minimizes their physical contacts, as required for

improved light utilization (Flori *et al*., 2017). It has been found that light-harvesting chlorophyll (LHC) proteins and the large and small subunits of ribulose bisphosphate carboxylase were not affected by Fe deficiency (Greene *et al*., 1991).

Chlorophylls trap light energy that is used in photosynthesis. Chlorophyll fluorescence is light re-emitted by chlorophyll molecules during return from excited to non-excited states, used as an indicator of photosynthetic energy conversion. An intriguing finding of my work has been the detection of higher chlorophyll content in ISIP3-KO mutants. It is not clear how this increase in chlorophyll affects the output of the photosynthetic machinery overall in the cells. How the photosynthesis apparatus as a whole is impacted by the KO of ISIP3 is an interesting avenue for further research. With the availability of the mutants, I hope to collaborate with the Institut de biologie physico-chimique (IBPC) to continue this avenue of work. Photosynthetic activity (Fv/Fm and ETR), maximum fluorescence yield, PSI : PSII ratio, and Chl *a* : C ratio will be promising parameters to profile the photosynthesis architecture and photophysiological performance of mutants versus MT control cell lines.

5.2.5 Concluding remarks

About one-fifth of the photosynthesis on Earth is carried out by diatoms (Field *et al*., 1998; Mann, 1999). Each year, diatom photosynthesis in the sea generates about as much organic carbon as all the terrestrial rainforests combined (Nelson, 1995). These photosynthetic workhorses are found in waters worldwide, survive extreme conditions (low light intensities, low temperature, high salinities and limited supply of nutrients, *e.g.*, iron). In particular, diatoms are characterized by their ability to survive prolonged periods of iron scarcity (Sunda and Huntsman, 1995). Therefore, it is critical to understand the special machinery of how these essential organisms adapted to the exceedingly conditions. Furthermore, diatoms today generate most of the organic matter that serves as food for life in the sea. They greatly influence global climate, atmospheric carbon dioxide concentration and marine ecosystem function (Denman, 2008). How diatoms will respond to the rapidly changing conditions in today's oceans is critical for the health and the future of our environment. This thesis provides novel insight into the mechanisms underlying the iron acquisition and storage in diatoms.

References

Abadía J, Vázquez S, Rellán-Álvarez R, El-Jendoubi H, Abadía A, Álvarez-Fernández A, López-Millán AF. 2011. Towards a knowledge-based correction of iron chlorosis. Plant Physiology and Biochemistry **49**, 471-482.

Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C. 2008. Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation. Proceedings of the National Academy of Sciences **105**, 10438-10443.

Allen MD, del Campo JA, Kropat J, Merchant SS. 2007. FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with FOX1 and FTR1 in iron-deficient Chlamydomonas reinhardtii. Eukaryotic cell **6**, 1841-1852.

Alverson AJ, Theriot EC. 2005. Comments on recent progress toward reconstructing the diatom phylogeny. Journal of Nanoscience and Nanotechnology **5**, 57-62.

Annett AL, Lapi S, Ruth TJ, Maldonado MT. 2008. The effects of Cu and Fe availability on the growth and Cu: ℃ ratios of marine diatoms. Limnology and Oceanography **53**, 2451-2461.

Apt KE, Grossman A, Kroth-Pancic P. 1996a. Stable nuclear transformation of the diatomPhaeodactylum tricornutum. Molecular and General Genetics MGG **252**, 572-579.

Apt KE, Grossman A, Kroth-Pancic P. 1996b. Stable nuclear transformation of the diatom Phaeodactylum tricornutum. Molecular and General Genetics MGG **252**, 572-579.

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M. 2004. The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science **306**, 79-86.

Bailet B, Apothéloz-Perret-Gentil L, Baričević A, Chonova T, Franc A, Frigerio J-M, Kelly M, Mora D, Pfannkuchen M, Proft S. 2020. Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. Science of the Total Environment **745**, 140948.

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D. 2015. Energetic coupling between plastids and mitochondria drives CO 2 assimilation in diatoms. Nature **524**, 366-369.

Baldauf SL. 2008. An overview of the phylogeny and diversity of eukaryotes. Journal of Systematics and Evolution **46**, 263-273.

Behnke J, LaRoche J. 2020. Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). European Journal of Phycology **55**, 339-360.

Bendtsen JD, Nielsen H, Von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of molecular biology **340**, 783-795.

Benoiston A-S, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, Bowler C. 2017. The evolution of diatoms and their biogeochemical functions. Philosophical Transactions of the Royal Society B: Biological Sciences **372**, 20160397.

Bhatta H, Kong TK, Rosengarten G. 2009. Diffusion through diatom nanopores. *Journal of Nano Research*, Vol. 7: Trans Tech Publ, 69-74.

Biggins J, Bruce D. 1989. Regulation of excitation energy transfer in organisms

containing phycobilins. Photosynthesis research **20**, 1-34.
Blaby-Haas CE, Merchant SS. 2013. Iron sparing and recycling in a compartmentalized cell. Current opinion in microbiology **16**, 677-685.
Bonnain C, Breitbart M, Buck KN. 2016. The ferrojan horse hypothesis: iron-virus interactions in the ocean. Frontiers in Marine Science **3**, 82.
Bork P, Bowler C, de Vargas C, Gorsky G, Karsenti E, Wincker P. 2015. Tara Oceans. Tara Oceans studies plankton at planetary scale. Introduction. Science (New York, NY) **348**, 873-873.
Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature **456**, 239.
Boyd PW, Jickells T, Law C, Blain S, Boyle E, Buesseler K, Coale K, Cullen J, De Baar HJ, Follows M. 2007. Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. Science **315**, 612-617.
Boyd PW, Ellwood MJ. 2010. The biogeochemical cycle of iron in the ocean. Nature Geoscience **3**, 675-682.
Boyle E. 1998. Pumping iron makes thinner diatoms. Nature **393**, 733-734.
Bolte K, Bullmann L, Hempel F, Bozarth A, Zauner S, and Maier UG. 2009. Protein targeting into secondary plastids. J Eukaryot Microbiol **56**: 9–15.
Bracher A, Vountas M, Dinter T, Burrows J, Röttgers R, Peeken I. 2009. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. Biogeosciences **6**, 751-764.
Brembu T, Winge P, Tooming-Klunderud A, Nederbragt AJ, Jakobsen KS, Bones AM. 2014. The chloroplast genome of the diatom Seminavis robusta: new features introduced through multiple mechanisms of horizontal gene transfer. Marine genomics **16**, 17-27.
Briat J-F, Curie C, Gaymard F. 2007. Iron utilization and metabolism in plants. Current opinion in plant biology **10**, 276-282.
Briat J-F, Ravet K, Arnaud N, Duc C, Boucherez J, Touraine B, Cellier F, Gaymard F. 2010. New insights into ferritin synthesis and function highlight a link between iron homeostasis and oxidative stress in plants. Annals of botany **105**, 811-822.
Bundy RM, Boiteau RM, McLean C, Turk-Kubo KA, McIlvin MR, Saito MA, Van Mooy BA, Repeta DJ. 2018. Distinct siderophores contribute to iron cycling in the mesopelagic at station ALOHA. Frontiers in Marine Science **5**, 61.
Busch A, Rimbauld B, Naumann B, Rensch S, Hippler M. 2008. Ferritin is required for rapid remodeling of the photosynthetic apparatus and minimizes photo-oxidative stress in response to iron availability in Chlamydomonas reinhardtii. The Plant Journal **55**, 201-211.
Caputi L, Carradec Q, Eveillard D, Kirilovsky A, Pelletier E, Pierella Karlusich JJ, Rocha Jimenez Vieira F, Villar E, Chaffron S, Malviya S. 2019. Community-level responses to iron availability in open ocean plankton ecosystems. Global Biogeochemical Cycles **33**, 391-419.
Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, Lima-Mendez G, Rocha F, Tirichine L, Labadie K. 2018. A global ocean atlas of eukaryotic genes. Nature communications **9**, 1-13.
Carstensen J, Klais R, Cloern JE. 2015. Phytoplankton blooms in estuarine and coastal waters: Seasonal patterns and key species. Estuarine, Coastal and Shelf Science **162**, 98-

109.
Caufield JH, Sakhawalkar N, Uetz P. 2012. A comparison and optimization of yeast two-hybrid systems. Methods. **58**, 317-24.
Cavalier-Smith T. 2003. Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **358**, 109-134.
Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree 1, 2. Journal of Eukaryotic Microbiology **46**, 347-366.
Chekalyuk AM, Landry MR, Goericke R, Taylor AG, Hafez MA. 2012. Laser fluorescence analysis of phytoplankton across a frontal zone in the California Current ecosystem. Journal of Plankton Research **34**, 761-777.
Chen Z, Wang X, Li S, Yao J, Shao Z, Duan D. 2019. Verification of the Saccharina japonica translocon Tic20 and its localization in the chloroplast membrane in diatoms. International journal of molecular sciences. **16**, 4000.
Clarke FW, Washington HS. 1924. *The composition of the earth's crust*: US Government Printing Office.
Coale TH, Moosburner M, Horák A, Oborník M, Barbeau KA, Allen AE. 2019. Reduction-dependent siderophore assimilation in a model pennate diatom. Proceedings of the National Academy of Sciences **116**, 23609-23617.
Cohen NR, Gong W, Moran DM, McIlvin MR, Saito MA, Marchetti A. 2018a. Transcriptomic and proteomic responses of the oceanic diatom Pseudo-nitzschia granii to iron limitation. Environmental microbiology **20**, 3109-3126.
Cohen NR, Mann E, Stemple B, Moreno CM, Rauschenberg S, Jacquot JE, Sunda WG, Twining BS, Marchetti A. 2018b. Iron storage capacities and associated ferritin gene expression among marine diatoms. Limnology and Oceanography **63**, 1677-1691.
Curie C, Alonso JM, Jean ML, Ecker JR, Briat J-F. 2000. Involvement of NRAMP1 from Arabidopsis thaliana in iron transport. Biochemical Journal **347**, 749-755.
De Baar HJ, Boyd PW, Coale KH, Landry MR, Tsuda A, Assmy P, Bakker DC, Bozec Y, Barber RT, Brzezinski MA. 2005. Synthesis of iron fertilization experiments: from the iron age in the age of enlightenment. Journal of Geophysical Research: Oceans **110**.
Denman KL. 2008. Climate change, ocean processes and ocean iron fertilization. Marine Ecology Progress Series. **25**, 364-219.
Diner RE, Bielinski VA, Dupont CL, Allen AE, Weyman PD. 2016. Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. Frontiers in bioengineering and biotechnology **4**, 65.
Dorrell RG, Gile G, Mccallum G, Méheust R, Bapteste EP, Klinger CM, Brillet-Guéguen L, Freeman KD, Richter DJ, Bowler C. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. elife **6**, e23717.
Dorrell RG, Villain A, Perez-Lamarque B, de Kerdrel GA, McCallum G, Watson AK, Ait-Mohamed O, Alberti A, Corre E, Frischkorn KR. 2021. Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. Proceedings of the National Academy of Sciences **118**.
Duc C, Cellier F, Lobréaux S, Briat J-F, Gaymard F. 2009. Regulation of iron homeostasis in Arabidopsis thaliana by the clock regulator time for coffee. Journal of Biological

Chemistry **284**, 36271-36281.
Eddy SR. 1998. Profile hidden Markov models. Bioinformatics (Oxford, England) **14**, 755-763.
Eide DJ. 2005. The ZIP family of zinc transporters. *Zinc Finger Proteins*: Springer, 261-264.
Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature protocols **2**, 953.
Erdner DL, Anderson DM. 1999. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. Limnology and Oceanography **44**, 1609-1615.
Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C. 1999. Transformation of nonselectable reporter genes in marine diatoms. Marine Biotechnology **1**, 239-251.
Falciatore A, d'Alcalà MR, Croot P, Bowler C. 2000. Perception of environmental signals by a marine diatom. Science **288**, 2363-2366.
Falciatore A, Jaubert M, Bouly J-P, Bailleul B, Mock T. 2020. Diatom molecular research comes of age: model species for studying phytoplankton biology and diversity. The Plant Cell **32**, 547-572.
Falkowski PG. 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO 2 in the ocean. Nature **387**, 272-275.
Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281**, 237-240.
Flori S, Jouneau P-H, Finazzi G, Maréchal E, Falconet D. 2016. Ultrastructure of the periplastidial compartment of the diatom Phaeodactylum tricornutum. Protist **167**, 254-267.
Flori S, Jouneau P-H, Bailleul B, Gallet B, Estrozi LF, Moriscot C, Bastien O, Eicke S, Schober A, Bártulos CR. 2017. Plastid thylakoid architecture optimizes photosynthesis in diatoms. Nature communications **8**, 1-9.
Fukasawa Y, Tsuji J, Fu S-C, Tomii K, Horton P, Imai K. 2015. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Molecular & Cellular Proteomics **14**, 1113-1126.
Galachyants YP, Zakharova YR, Petrova D, Morozov A, Sidorov I, Marchenkov A, Logacheva M, Markelov M, Khabudaev K, Likhoshway YV. 2015. Sequencing of the complete genome of an araphid pennate diatom Synedra acus subsp. radians from Lake Baikal. *Doklady Biochemistry and Biophysics*, Vol. 461: Springer, 84-88.
Gao X, Bowler C, Kazamia E. 2021. Iron metabolism strategies in diatoms. Journal of experimental botany **72**, 2165-2180.
Gentemann C, Scott JP, Mazzini PL, Pianca C, Akella S, Minnett PJ, Cornillon P, Fox-Kemper B, Cetinić I, Chin TM. 2020. Saildrone: Adaptively sampling the marine environment. Bulletin of the American Meteorological Society **101**, E744-E762.
Gledhill M, Buck KN. 2012. The organic complexation of iron in the marine environment: a review. Frontiers in Microbiology **3**, 69.
Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. The Plant Journal **81**, 519-528.
Gschloessl B, Guermeur Y, Cock JM. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. BMC bioinformatics **9**, 393.

Gould SB, Maier UG, and Martin WF. 2015. Protein Import and the Origin of Red Complex Plastids. Curr Biol **25**, 515–521.
Haëntjens N, Boss E, Talley LD. 2017. Revisiting Ocean Color algorithms for chlorophyll a and particulate organic carbon in the Southern Ocean using biogeochemical floats. Journal of Geophysical Research: Oceans **122**, 6583-6593.
Hain M, Sigman D, Haug G. 2014. 8.18–The biological Pump in the Past. Reference Module in Earth Systems and Environmental Sciences, Treatise on Geochemistry (Second Edition), The Oceans and Marine Geochemistry **8**, 485-517.
Hassler CS, van den Berg CM, Boyd PW. 2017. Toward a regional classification to provide a more inclusive examination of the ocean biogeochemistry of iron-binding ligands. Frontiers in Marine Science **4**, 19.
Hopes A, Nekrasov V, Kamoun S, Mock T. 2016. Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. Plant Methods **12**, 49.
Hopkinson BM, Morel FM. 2009. The role of siderophores in iron acquisition by photosynthetic marine microorganisms. Biometals **22**, 659-669.
Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier C, Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic acids research **35**, W585-W587.
Jin W, Takagi H, Pancorbo B, Theil EC. 2001. "Opening" the ferritin pore for iron release by mutation of conserved amino acids at interhelix and loop sites. Biochemistry **40**, 7525-7532.
Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. Nucleic acids research **36**, W5-W9.
Kang W, Anslan S, Börner N, Schwarz A, Schmidt R, Künzel S, Rioual P, Echeverría-Galindo P, Vences M, Wang J. 2021. Diatom metabarcoding and microscopic analyses from sediment samples at Lake Nam Co, Tibet: The effect of sample-size and bioinformatics on the identified communities. Ecological indicators **121**, 107070.
Kazamia E, Sutak R, Paz-Yepes J, Dorrell RG, Vieira FRJ, Mach J, Morrissey J, Leon S, Lam F, Pelletier E. 2018. Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. Science advances **4**, eaar4536.
Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS biology **12**.
Kerppola TK. 2006. Visualization of molecular interactions by fluorescence complementation. Nat Rev Mol Cell Biol **7**, 449–456.
Kustka AB, Allen AE, Morel FM. 2007. Sequence analysis and transcriptional regulation of iron acquisition genes in two marine diatoms 1. Journal of Phycology **43**, 715-729.
La Fontaine S, Quinn JM, Nakamoto SS, Page MD, Göhre V, Moseley JL, Kropat J, Merchant S. 2002. Copper-dependent iron assimilation pathway in the model photosynthetic eukaryote Chlamydomonas reinhardtii. Eukaryotic cell **1**, 736-757.
La Roche J, Boyd PW, McKay RML, Geider RJ. 1996. Flavodoxin as an in situ marker for iron stress in phytoplankton. Nature **382**, 802-805.
La Roche JL, Murray H, Orellana M, Newton J. 1995. Flavodoxin Expression as an Indicator of Iron Limitation in Marine Diatoms1. Journal of Phycology **31**, 520-530.
Lanquar V, Lelièvre F, Bolte S, Hamès C, Alcon C, Neumann D, Vansuyt G, Curie C, Schröder A, Krämer U. 2005. Mobilization of vacuolar iron by AtNRAMP3 and

AtNRAMP4 is essential for seed germination on low iron. The EMBO journal **24**, 4041-4051.

Lau JB, Stork S, Moog D, Schulz J, Maier UG. 2016. Protein–protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids. Molecular microbiology. **100**, 76-89.

Lauvset SK, Gruber N, Landschützer P, Olsen A, Tjiputra JF. 2015. Trends and drivers in global surface ocean pH over the past 3 decades. Biogeosciences **12**, 1285-1298.

Lelong A, Bucciarelli E, Hégaret H, Soudant P. 2013. Iron and copper limitations differently affect growth rates and photosynthetic and physiological parameters of the marine diatom Pseudo-nitzschia delicatissima. Limnology and Oceanography **58**, 613-623.

Lis H, Shaked Y, Kranzler C, Keren N, Morel FM. 2015. Iron bioavailability to phytoplankton: an empirical approach. The ISME journal **9**, 1003-1013.

Liu X, Millero FJ. 2002. The solubility of iron in seawater. Marine Chemistry **77**, 43-54.

Liu X, Cheng T. 2003. Video-based face recognition using adaptive hidden markov models. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, Vol. 1: IEEE, I-I.

Liu X, Theil EC. 2005. Ferritins: dynamic management of biological iron and oxygen chemistry. Accounts of chemical research **38**, 167-175.

Lodeyro AF, Ceccoli RD, Karlusich JJP, Carrillo N. 2012. The importance of flavodoxin for environmental stress tolerance in photosynthetic microorganisms and transgenic plants. Mechanism, evolution and biotechnological potential. FEBS letters **586**, 2917-2924.

Lommer M, Roy A-S, Schilhabel M, Schreiber S, Rosenstiel P, LaRoche J. 2010. Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. BMC genomics **11**, 718.

Lommer M, Specht M, Roy A-S, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC. 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome biology **13**, R66.

Long JC, Merchant SS. 2008. Photo-oxidative stress impacts the expression of genes encoding iron metabolism components in Chlamydomonas. Photochemistry and photobiology **84**, 1395-1403.

Madden T. 2013. The BLAST sequence analysis tool. *The NCBI Handbook [Internet]. 2nd edition*: National Center for Biotechnology Information (US).

Mahowald NM, Engelstaedter S, Luo C, Sealy A, Artaxo P, Benitez-Nelson C, Bonnet S, Chen Y, Chuang PY, Cohen DD. 2009. Atmospheric iron deposition: global distribution, variability, and human perturbations. Annual review of marine science **1**, 245-278.

Maldonado MT, Price NM. 2001. Reduction and transport of organically bound iron by Thalassiosira oceanica (Bacillariophyceae). Journal of Phycology **37**, 298-310.

Maldonado MT, Allen AE, Chong JS, Lin K, Leus D, Karpenko N, Harris SL. 2006. Copper-dependent iron transport in coastal and oceanic diatoms. Limnology and Oceanography **51**, 1729-1743.

Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, Wincker P, Iudicone D, de Vargas C, Bittner L. 2016. Insights into global diatom distribution and diversity in the world's ocean. Proceedings of the National Academy of Sciences **113**, E1516-E1525.

Mamon RS, Elliott RJ. 2007. *Hidden Markov models in finance*: Springer.
Mann DG. 1999. The species concept in diatoms. Phycologia **38**, 437-495.
Marchetti A, Maldonado MT, Lane ES, Harrison PJ. 2006. Iron requirements of the pennate diatom Pseudo-nitzschia: Comparison of oceanic (high-nitrate, low-chlorophyll waters) and coastal species. Limnology and Oceanography **51**, 2092-2101.
Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, Ribalet F, Murphy ME, Maldonado MT, Armbrust EV. 2009. Ferritin is used for iron storage in bloom-forming marine pennate diatoms. Nature **457**, 467-470.
Martínez-García A, Sigman DM, Ren H, Anderson RF, Straub M, Hodell DA, Jaccard SL, Eglinton TI, Haug GH. 2014. Iron fertilization of the Subantarctic Ocean during the last ice age. Science **343**, 1347-1350.
McKay RML, La Roche J, Yakunin AF, Durnford DG, Geider RJ. 1999. Accumulation of ferredoxin and flavodoxin in a marine diatom in response to Fe. Journal of Phycology **35**, 510-519.
McQuaid JB, Kustka AB, Oborník M, Horák A, McCrow JP, Karas BJ, Zheng H, Kindeberg T, Andersson AJ, Barbeau KA. 2018. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. Nature **555**, 534-537.
Merchant S, Hill K, Howe G. 1991. Dynamic interplay between two copper-titrating components in the transcriptional regulation of cyt c6. The EMBO journal **10**, 1383-1389.
Mock T, Otillar RP, Strauss J, McMullan M, Paajanen P, Schmutz J, Salamov A, Sanges R, Toseland A, Ward BJ. 2017. Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. Nature **541**, 536-540.
Moreno CM, Gong W, Cohen NR, DeLong K, Marchetti A. 2020. Interactive effects of iron and light limitation on the molecular physiology of the Southern Ocean diatom Fragilariopsis kerguelensis. Limnology and Oceanography.
Morrissey J, Sutak R, Paz-Yepes J, Tanaka A, Moustafa A, Veluchamy A, Thomas Y, Botebol H, Bouget F-Y, McQuaid JB. 2015. A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. Current Biology **25**, 364-371.
Murgia I, Vazzola V, Tarantino D, Cellier F, Ravet K, Briat J-F, Soave C. 2007. Knock-out of ferritin AtFer1 causes earlier onset of age-dependent leaf senescence in Arabidopsis. Plant Physiology and Biochemistry **45**, 898-907.
Narayanan NN, Ihemere U, Chiu WT, Siritunga D, Rajamani S, Singh S, Oda S, Sayre RT. 2011. The iron assimilatory protein, FEA1, from Chlamydomonas reinhardtii facilitates iron-specific metal uptake in yeast and plants. Frontiers in plant science **2**, 67.
Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B. 1995. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. Global Biogeochemical Cycles. **9**(3), 359-72.
Nuester J, Vogt S, Twining BS. 2012. LOCALIZATION OF IRON WITHIN CENTRIC DIATOMS OF THE GENUS THALASSIOSIRA 1. Journal of Phycology **48**, 626-634.
Nunn BL, Faux JF, Hippmann AA, Maldonado MT, Harvey HR, Goodlett DR, Boyd PW, Strzepek RF. 2013. Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. PLoS One **8**, e75653.
Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P. 2016. A CRISPR/Cas9 system adapted for gene editing in marine algae. Scientific reports **6**, 1-6.

Orme-Johnson W. 1985. Molecular basis of biological nitrogen fixation. Annual review of biophysics and biophysical chemistry **14**, 419-459.
Owens TG. 1986. Light-harvesting function in the diatom Phaeodactylum tricornutum: II. Distribution of excitation energy between the photosystems. Plant physiology **80**, 739-746.
Pankowski A, McMinn A. 2009. Iron availability regulates growth, photosynthesis, and production of ferredoxin and flavodoxin in Antarctic sea ice diatoms. Aquatic Biology **4**, 273-288.
Paz Y, Katz A, Pick U. 2007. A multicopper ferroxidase involved in iron binding to transferrins in Dunaliella salina plasma membranes. Journal of Biological Chemistry **282**, 8658-8666.
Peers G, Quesnel SA, Price NM. 2005. Copper requirements for iron acquisition and growth of coastal and oceanic diatoms. Limnology and Oceanography **50**, 1149-1158.
Peers G, Price NM. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. Nature **441**, 341-344.
Pfaffen S, Abdulqadir R, Le Brun NE, Murphy ME. 2013. Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom. Journal of Biological Chemistry **288**, 14917-14925.
Pfaffen S, Bradley JM, Abdulqadir R, Firme MR, Moore GR, Le Brun NE, Murphy ME. 2015. A diatom ferritin optimized for iron oxidation but not iron storage. Journal of Biological Chemistry **290**, 28416-28427.
Philpott CC. 2006. Iron uptake in fungi: a system for every source. Biochimica et Biophysica Acta (bba)-molecular cell research **1763**, 636-645.
Pollak B, Cerda A, Delmans M, Álamos S, Moyano T, West A, Gutiérrez RA, Patron NJ, Federici F, Haseloff J. 2019. Loop assembly: a simple and open system for recursive fabrication of DNA circuits. New Phytologist **222**, 628-640.
Poulsen N, Kröger N. 2005. A new molecular tool for transgenic diatoms: control of mRNA and protein biosynthesis by an inducible promoter–terminator cassette. The FEBS journal **272**, 3413-3423.
Poulsen N, Chesley PM, Kröger N. 2006. Molecular genetic manipulation of the diatom Thalassiosira pseudonana (Bacillariophyceae) 1. Journal of Phycology **42**, 1059-1065.
Pyszniak A, Gibbs SP. 1992. Immunocytochemical localization of photosystem I and the fucoxanthin-chlorophylla/c light-harvesting complex in the diatomPhaeodactylum tricornutum. Protoplasma **166**, 208-217.
Quigg A, Finkel ZV, Irwin AJ, Rosenthal Y, Ho T-Y, Reinfelder JR, Schofield O, Morel FM, Falkowski PG. 2003. The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. Nature **425**, 291-294.
Quigg A, Irwin AJ, Finkel ZV. 2011. Evolutionary inheritance of elemental stoichiometry in phytoplankton. Proceedings of the Royal Society B: Biological Sciences **278**, 526-534.
Rastogi A, Maheswari U, Dorrell RG, Vieira FRJ, Maumus F, Kustka A, McCarthy J, Allen AE, Kersey P, Bowler C. 2018. Integrative analysis of large-scale transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of diatoms. Scientific reports **8**, 1-14.
Raven JA, Evans MC, Korb RE. 1999. The role of trace metals in photosynthetic electron transport in O 2-evolving organisms. Photosynthesis research **60**, 111-150.
Raven JA. 2013. Iron acquisition and allocation in stramenopile algae. Journal of

experimental botany **64**, 2119-2127.
Renshaw JC, Robson GD, Trinci AP, Wiebe MG, Livens FR, Collison D, Taylor RJ. 2002. Fungal siderophores: structures, functions and applications. Mycological Research **106**, 1123-1142.
Rich HW, Morel FM. 1990. Availability of well-defined iron colloids to the marine diatom Thalassiosira weissflogii. Limnology and Oceanography **35**, 652-662.
Rivera SF, Vasselon V, Bouchez A, Rimet F. 2020. Diatom metabarcoding applied to large scale monitoring networks: Optimization of bioinformatics strategies using Mothur software. Ecological indicators **109**, 105775.
Roesler C, Uitz J, Claustre H, Boss E, Xing X, Organelli E, Briggs N, Bricaud A, Schmechtig C, Poteau A. 2017. Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors. Limnology and Oceanography: Methods **15**, 572-585.
Rossel JB, Wilson IW, Pogson BJ. 2002. Global changes in gene expression in response to high light in Arabidopsis. Plant physiology **130**, 1109-1120.
Rubin M, Berman-Frank I, Shaked Y. 2011. Dust-and mineral-iron utilization by the marine dinitrogen-fixer Trichodesmium. Nature Geoscience **4**, 529-534.
Sathyendranath S, Watts L, Devred E, Platt T, Caverhill C, Maass H. 2004. Discrimination of diatoms from other phytoplankton using ocean-colour data. Marine Ecology Progress Series **272**, 59-68.
Sato S, Nanjappa D, Dorrell RG, FR JV, Kazamia E, L T, Veluchamy A HR, Aury J-M, Jaillon O, Wincker P, Fussy Z, Obornik M, Muñoz-Gómez SA, Mann DG, Bowler ℃ & Zingone A. 2020. Genome enabled phylogenetic and functional reconstruction of an araphid pennate diatom Plagiostriata sp. CCMP470, previously assigned as a radial centric diatom, and its bacterial commensal. Scientific reports **10**.
Scheiber IF, Pilátová J, Malych R, Kotabova E, Krijt M, Vyoral D, Mach J, Léger T, Camadro J-M, Prášil O. 2019. Copper and iron metabolism in Ostreococcus tauri–the role of phytotransferrin, plastocyanin and a chloroplast copper-transporting ATPase. Metallomics **11**, 1657-1666.
Scott JP, Crooke S, Cetinić I, Del Castillo CE, Gentemann CL. 2020. Correcting non-photochemical quenching of Saildrone chlorophyll-a fluorescence for evaluation of satellite ocean color retrievals. Optics express **28**, 4274-4285.
Sétif P. 2001. Ferredoxin and flavodoxin reduction by photosystem I. Biochimica et Biophysica Acta (BBA)-Bioenergetics **1507**, 161-179.
Shaked Y, Kustka AB, Morel FM. 2005. A general kinetic model for iron acquisition by eukaryotic phytoplankton. Limnology and Oceanography **50**, 872-882.
Shang S, Wu J, Huang B, Lin G, Lee Z, Liu J, Shang S. 2014. A new approach to discriminate dinoflagellate from diatom blooms from space in the East China Sea. Journal of Geophysical Research: Oceans **119**, 4653-4668.
Shoenfelt EM, Sun J, Winckler G, Kaplan MR, Borunda AL, Farrell KR, Moreno PI, Gaiero DM, Recasens C, Sambrotto RN. 2017. High particulate iron (II) content in glacially sourced dusts enhances productivity of a model diatom. Science advances **3**, e1700314.
Sims PA, Mann DG, Medlin LK. 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia **45**, 361-402.
Smith SR, Gillard JT, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont

CL, Obata T, Fernie AR. 2016. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS genetics **12**.
Soria-Dengg S, Horstmann U. 1995. Ferrioxamines B and E as iron sources for the marine diatom Phaeodactylum tricornutum. Marine Ecology Progress Series **127**, 269-277.
Strzepek RF, Harrison PJ. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. Nature **431**, 689-692.
Strzepek RF, Hunter KA, Frew RD, Harrison PJ, Boyd PW. 2012. Iron-light interactions differ in Southern Ocean phytoplankton. Limnology and Oceanography **57**, 1182-1200.
Sunagawa S, Acinas SG, Bork P, Bowler C, Eveillard D, Gorsky G, Guidi L, Iudicone D, Karsenti E, Lombard F. 2020. Tara Oceans: towards global ocean ecosystems biology. Nature Reviews Microbiology **18**, 428-445.
Sunda WG, Huntsman SA. 1995. Iron uptake and growth limitation in oceanic and coastal phytoplankton. Marine Chemistry **50**, 189-206.
Sunda WG, Huntsman SA. 1997. Interrelated influence of iron, light and cell size on marine phytoplankton growth. Nature **390**, 389-392.
Sunda WG, Price NM, Morel FM. 2005. Trace metal ion buffers and their use in culture studies. Algal culturing techniques, **4**, 35-63.
Tanaka T, Maeda Y, Veluchamy A, Tanaka M, Abida H, Maréchal E, Bowler C, Muto M, Sunaga Y, Tanaka M. 2015. Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. The Plant Cell **27**, 162-176.
Theil EC. 1987. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. Annual review of biochemistry **56**, 289-315.
Vancaester E, Depuydt T, Osuna-Cruz CM, Vandepoele K. 2020. Comprehensive and functional analysis of horizontal gene transfer events in diatoms. Molecular biology and evolution **37**, 3243-3257.
Velasquez I, Nunn BL, Ibisanmi E, Goodlett DR, Hunter KA, Sander SG. 2011. Detection of hydroxamate siderophores in coastal and Sub-Antarctic waters off the South Eastern Coast of New Zealand. Marine Chemistry **126**, 97-107.
Wagner H, Jakob T, Wilhelm C. 2006. Balancing the energy flow from captured light to biomass under fluctuating light conditions. New Phytologist **169**, 95-108.
Wang W, Yu L-J, Xu C, Tomizaki T, Zhao S, Umena Y, Chen X, Qin X, Xin Y, Suga M. 2019. Structural basis for blue-green light harvesting and energy dissipation in diatoms. Science **363**.
Whitney LP, Lins JJ, Hughes MP, Wells ML, Chappell PD, Jenkins BD. 2011. Characterization of putative iron responsive genes as species-specific indicators of iron stress in thalassiosiroid diatoms. Frontiers in Microbiology **2**, 234.
Xie J, Bai X, Li Y, Sun C, Qian H, Fu Z. 2014. The effect of glufosinate on nitrogen assimilation at the physiological, biochemical and molecular levels in Phaeodactylum tricornutum. Ecotoxicology **23**, 1430-1438.
Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Molecular biology and evolution **21**, 809-818.
Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. 2000. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable

marker and reporter genes. Journal of Phycology **36**, 379-386.
Zhang X, Krause K-H, Xenarios I, Soldati T, Boeckmann B. 2013. Evolution of the ferric reductase domain (FRD) superfamily: modularity, functional diversification, and signature motifs. PLoS One **8**.
Zhao P, Gu W, Huang A, Wu S, Liu C, Huan L, Gao S, Xie X, Wang G. 2018. Effect of iron on the growth of Phaeodactylum tricornutum via photosynthesis. Journal of Phycology **54**, 34-43.

Annexes

Annex I: Iron Metabolism Strategies in Diatoms

Journal of Experimental Botany, Vol. 72, No. 6 pp. 2165–2180, 2021
doi:10.1093/jxb/eraa575

REVIEW PAPER

Iron metabolism strategies in diatoms

Xia Gao, Chris Bowler, and Elena Kazamia

Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France

Correspondence: kazamia@bio.ens.psl.eu

Received 20 July 2020; Editorial decision 2 December 2020; Accepted 3 March 2021

Editor: Janneke Balk, John Innes Centre, UK

Abstract

Diatoms are one of the most successful group of photosynthetic eukaryotes in the contemporary ocean. They are ubiquitously distributed and are the most abundant primary producers in polar waters. Equally remarkable is their ability to tolerate iron deprivation and respond to periodic iron fertilization. Despite their relatively large cell sizes, diatoms tolerate iron limitation and frequently dominate iron-stimulated phytoplankton blooms, both natural and artificial. Here, we review the main iron use strategies of diatoms, including their ability to assimilate and store a range of iron sources, and the adaptations of their photosynthetic machinery and architecture to iron deprivation. Our synthesis relies on published literature and is complemented by a search of 82 diatom transcriptomes, including information collected from seven representatives of the most abundant diatom genera in the world's oceans.

Keywords: Diatoms, ferritin, iron physiology, iron quotas, iron starvation-induced proteins (ISIPs), iron storage, iron uptake.

Introduction

Unicellular life began to evolve more than 3 billion years ago in an ocean that was devoid of molecular oxygen and rich in soluble ferrous iron (Fe^{2+}). The metabolic processes that became core to all cellular life relied on the abundance of this transition element and its ability to stably occupy multiple valence states, as cofactor to enzymes catalysing reactions involving the transfer of electrons. Most notably, these processes include photosynthesis and respiration, as well as the synthesis of essential organic molecules such as amino acids, lipids, deoxyribonucleotides, and sterols. The minimum concentration of iron required by an individual cell to sustain its metabolic functions is referred to as its metabolic 'iron quota'. Within cells, iron homeostasis is carefully controlled, since overabundance of Fe^{2+} can catalyse the formation of damaging reactive oxygen species. While the preference of cells for reduced iron has remained unchanged, that is, cells readily take up Fe^{2+} and carefully regulate its concentration intracellularly, the abiotic environment of unicellular species living in today's ocean is vastly different from that existing at the start of these organisms' evolution. The contemporary surface ocean is oxygenated and well mixed, with a mildly alkaline pH (global averages are around 8.2; Lauvset *et al.*, 2015). This is an oxidizing environment that chemically shifts iron into its ferric state (Fe^{3+}), so that the most abundant form of iron is as Fe^{3+} bound by organic material or colloid particles of oxyhydroxides (Gledhill and Buck, 2012). Organisms have therefore had to evolve a variety of molecular mechanisms to make such recalcitrant forms of iron bioavailable.

One of the dominant, ubiquitous groups of photosynthetic producers in the contemporary ocean is diatoms, which are

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Experimental Biology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

estimated to account for 20% of total global primary production (~20×10^{15} g carbon fixed per year) (Field *et al.*, 1998; Mann, 1999). Diatoms are unicellular stramenopiles (heterokont protists within the chromalveolates) and are, in an evolutionary context, a relatively young form of life. Molecular phylogenies date the origin of diatoms towards the beginning of the Mesozoic Era (Sims *et al.*, 2006), while current estimates of the richness of diatom species vary between 40 000 and 200 000; counterintuitively, the higher end of this range is based on morphology-based estimates while the lower estimates are based on molecular markers (reviewed in Benoiston *et al.*, 2017). The evolutionary history of diatoms is shaped by endosymbiotic events, of which it has been proposed that there were at least two. The first occurred an estimated 1.5 billion years ago, when a eukaryotic heterotroph assimilated or was invaded by a cyanobacterium (Yoon *et al.*, 2004). Some 500 million years later, a secondary endosymbiosis occurred, in which a eukaryotic heterotroph captured or was invaded by a red alga, establishing the photosynthetic stramenopile lineage to which diatoms belong (Cavalier-Smith, 1999). Recent evidence suggests that the chimeric nature of diatoms is even more complicated, as a range of nucleus-encoded proteins are of green algal origin (Dorrell *et al.*, 2017 and references therein). Diatoms are subclassified into 'centric' species, which include two suborders (radial and polar), and 'pennate' species, which are further divided into two suborders (raphid and araphid) (Fig. 1). These groups are not monophyletic, as centric diatoms grade into araphid pennates, and araphid pennates grade into the raphe-bearing pennate diatoms, which are a natural group (Alverson and Theriot, 2005).

Diatom cells have a unique structure and features (Fig. 1). Cell sizes are large compared with other oceanic unicellular eukaryotes and are generally in excess of 20 μm in diameter (with maximum cell sizes of 2 mm recorded). Some diatoms are known to form chains. Unlike other organisms, diatom cell membranes are silicified and are referred to as frustules. Frustules are porous, with their nanopores ranging in size from 250 nm to 600 nm; they are generally assumed not to interfere with diatom nutrient uptake (Bhatta *et al.*, 2009). The majority of pennate diatoms contain one central vacuole, while centric species can have many vacuoles. Diatom plastids are surrounded by four membranes, a vestige of the endosymbiosis events that generated the organelle. As with vacuoles, the majority of pennate species have only one plastid, but the number of plastids is variable in the centric species. Nine species of diatoms have had their genomes sequenced: *Thalassiosira pseudonana* (Armbrust *et al.*, 2004),

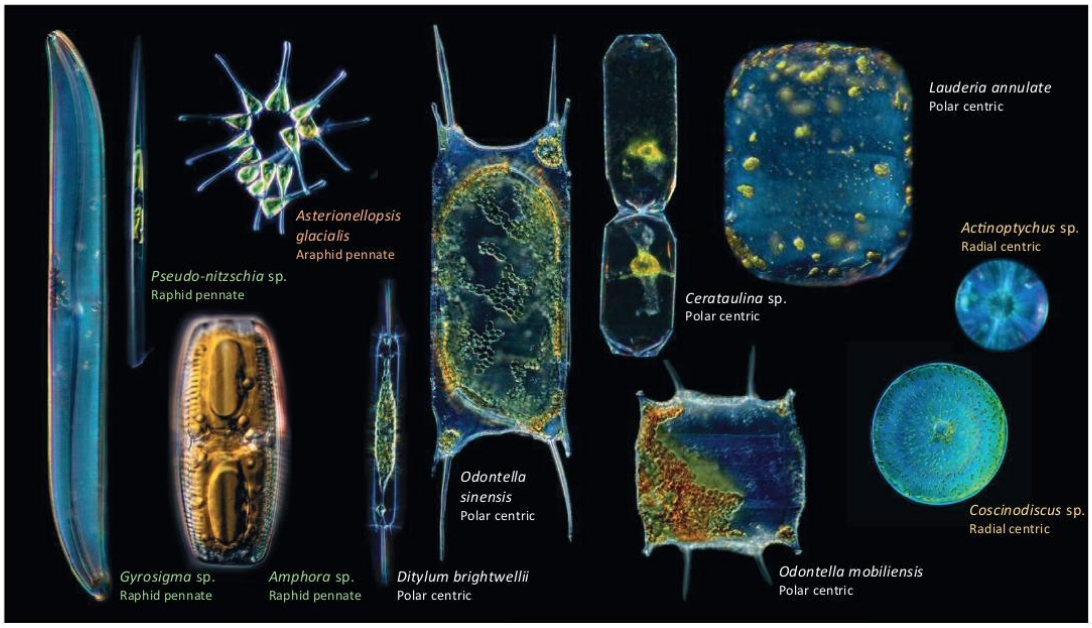


Fig. 1. Light microscopy images of representative species in the raphid and araphid pennate, and polar and radial centric groups of diatoms. Sizes are not to scale. Diatoms were collected and photographed on board *Tara*, in Roscoff, Villefranche-sur-Mer Marine Stations (France), and Bigelow Laboratory, Boothbay (USA). Photography by Christian and Noé Sardet, rearranged from C. Sardet, 2015. Plankton: Wonders of the Drifting World. Republished with permission of University of Chicago Press; permission conveyed through Copyright Clearance Center, Inc.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Phaeodactylum tricornutum (Bowler *et al.*, 2008), *Thalassiosira oceanica* (Lommer *et al.*, 2012), *Pseudo-nitzschia multistriata* strain B856 (NCBI assembly ASM90066040v1), *Pseudo-nitzschia multiseries* CLN47 (JGI Project ID: 16870), *Synedra acus* (Galachyants *et al.*, 2015), *Fistulifera solaris* (Tanaka *et al.*, 2015), *Fragilariopsis cylindrus* (Mock *et al.*, 2017), and *Plagiostriata* sp. CCMP470 (Sato *et al.*, 2020), while 92 species have available transcriptomes collected through the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) (Keeling *et al.*, 2014). A further seven transcriptomes have been generated by Genoscope and represent seven of the most globally abundant diatom genera not covered by the MMETSP project (Dorrell *et al.*, 2021). The pennate diatom *P. tricornutum* and the centric diatom *T. pseudonana* are laboratory models that can be readily cultured and transformed (Apt *et al.*, 1996; Zaslavskaia *et al.*, 2000; Poulsen *et al.*, 2006; Hopes *et al.*, 2016). Although neither is abundant in the environment, they often harbour genes that have been proposed to contribute to the success of diatoms in the wild.

The ecological success of diatoms is remarkable. They are present in all ocean provinces and are the most abundant photosynthetic producers in the Arctic and the Southern oceans (Malviya *et al.*, 2016). Importantly, they are ruthless competitors for iron, often dominating iron-stimulated blooms (Boyd *et al.*, 2007). This indicates that diatoms are adapted to survive conditions of low iron and are equipped with the molecular machinery necessary to take up and assimilate iron efficiently, without causing damage to their cells through oxidative stress. Rates of iron uptake and use are also linked to nitrogen assimilation and growth: on average, diatoms with higher iron uptake show proportionately greater nitrate uptake and productivity (e.g. Falkowski, 1997). Surprisingly, this is not reflected in their silica uptake, which remains constant under iron supplementation despite faster growth, leading to weakly silicified cells (Boyle, 1998 and references therein). Diatom iron nutrition therefore affects not only metabolism but also the resulting morphology and physical characteristics of diatom cells.

Comparing the cell metal concentrations of diatoms with those of autotrophic flagellates, heterotrophic flagellates, and autotrophic picoplankton in the equatorial Pacific Ocean, Twining *et al.* (2011) reported that phosphorus-normalized Mn, Fe, Ni, and Zn ratios were significantly higher in diatoms than in other cell types. Furthermore, total diatom Fe concentrations per cell exceeded minimum subsistence levels.

Here, we review what is known about iron physiology in diatoms. As far as is possible, we draw comparisons between different diatom species (Sunda and Huntsman, 1995; Strzepek and Harrison, 2004; Kustka *et al.*, 2007). We take a broad perspective that considers how diatoms take up, store, and use iron, in order to shed light on the question of what makes diatoms competitive in handling this micronutrient.

A portfolio of iron uptake mechanisms shows diatoms are adapted to acquire iron from a variety of sources

The vast majority of iron present in surface ocean seawater is complexed by organic ligands (Gledhill and Buck, 2012). These are classified based on their affinity for iron. The strongest chelates are siderophore molecules, thought to be synthesized largely by bacteria, although fungi, which are of lower abundance in the ocean, are also capable of siderophore synthesis (Renshaw *et al.*, 2002). There are three chemical classes of siderophores—hydroxamate, catecholate, and carboxylate—and their relative contribution to the iron pool of seawater is unknown. Hydroxamate siderophores, which have been detected both in coastal and open ocean waters (Mawji *et al.*, 2008; Boiteau *et al.*, 2016), are hydrophilic and are released by synthesizing species into the water column. In these microorganisms, siderophores are recaptured by specific recognition machinery. It has been proposed that it is these molecules in particular that are available to non-producing species, such as diatoms (Hopkinson and Morel, 2009). Amphiphilic siderophores, which are attached by lipid tails to the cell membranes of producing bacterial species, have received particular attention in the literature on account of their chimeric chemical structures (Martinez *et al.*, 2000; Vraspir *et al.*, 2011). However, they are an unlikely source of iron for non-producers. Other chelates include extracellular polysaccharides and humic acids, which bind iron weakly. Notably, many marine viruses are also capable of chelating iron on their tails and thus may represent a large sink of marine iron (Bonnain *et al.*, 2016). Inorganic mineral iron is periodically available to surface communities, either through aeolian input (i.e. arising from the action of wind) or via upwelling currents that bring remineralized iron from the deep ocean (Mahowald *et al.*, 2008; Boyd and Ellwood, 2010). The most bioavailable and scarcest source of iron is dissolved, uncomplexed iron (Fe') (Liu and Millero, 2002; Lis *et al.*, 2015).

Diatoms have a diverse portfolio of iron uptake mechanisms adapted to the sources of iron available in seawater (summarized in Fig. 2). A mechanism whose function has been experimentally verified in the model pennate diatom *P. tricornutum* is the assimilation of uncomplexed iron bound by a transferrin-like protein. The iron starvation-induced protein 2A (ISIP2A), first identified in *P. tricornutum* (Allen *et al.*, 2008), was shown to function as a 'phytotransferrin', a protein directed to the cell membrane of cells with carboxylate iron-binding domains, whose iron-chelating properties were pH sensitive (Morrissey *et al.*, 2015; McQuaid *et al.*, 2018). *P. tricornutum* cells deficient in ISIP2A showed reduced Fe' uptake capabilities, suggesting that this was the primary but not the only mechanism of Fe' uptake in this species (Kazamia *et al.*, 2018). ISIP2A contains the functional domain PF07692, which was first characterized in the iron-assimilation proteins FEA1 and FEA2 of the green alga *Chlamydomonas reinhardtii* (Allen *et al.*, 2007), and

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

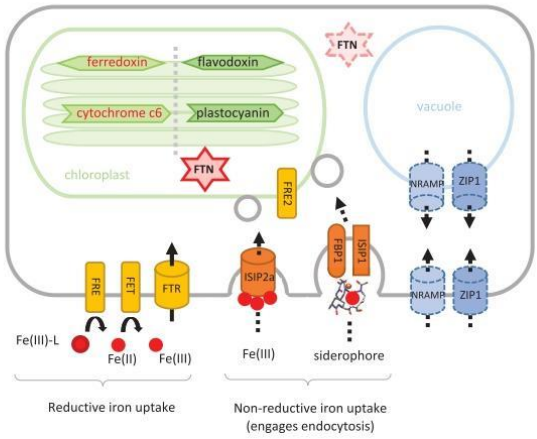


Fig. 2. Schematic diagram of iron uptake mechanisms and metabolism in diatoms. A summary schematic of the proteins discussed in this review and of their cellular location. In the chloroplast, we review the replacement of ferredoxin with its iron-free equivalent, flavodoxin, and the substitution of cytochrome c_6 with plastocyanin. Reductive and non-reductive iron uptake mechanisms are considered, the latter employing the cell's vesicular network to direct iron to the plastid. Two alternative iron storage mechanisms are reviewed: the accumulation of iron in the mineral core of ferritin (FTN), and vacuolar storage. For the storage proteins, the cellular localization of the putative proteins has not been confirmed through microscopy analyses (indicated with dashed lines).

additionally was functionally described in *Dunaliella salina* (Paz *et al.*, 2007). In *C. reinhardtii*, FEA1 and FEA2 were highly expressed under low-iron conditions and facilitated high-affinity iron uptake (Allen *et al.*, 2007), likely in its ferrous form (Narayanan *et al.*, 2011). By contrast, in diatoms, the iron-binding domains of ISIP2A coordinate ferric iron (McQuaid *et al.*, 2018). We searched the available diatom genomes and transcriptomes for genes with a profile that matched a constructed Hidden Markov Model (HMM) for ISIP2A. Our results are presented as a heatmap in Fig. 3, where we show the detection of *ISIP2A*-like genes in 82 diatom species. We retrieved matches for *ISIP2A*-like genes in most diatoms, with some notable exceptions in some species of *Chaetoceros*.

Experimental evidence suggests that colloidal iron, as particulate Fe^{2+}, is bioavailable to *P. tricornutum* (Shoenfelt *et al.*, 2017). While aeolian inputs of iron are known to stimulate algal growth, and growth of diatoms in particular, both in short-term ecological experiments (De Baar *et al.*, 2005) and over geological timeframes (Hain *et al.*, 2014; Martínez-García *et al.*, 2014), it has been a long-standing assumption that the iron stimulation is through the increase in Fe′ (Rich and Morel, 1990). Shoenfelt *et al.* (2017) challenged this assumption by demonstrating that the growth rates of *P. tricornutum* cultures improved by glacial dust stimulation beyond what would be expected by the increase in Fe′ alone (the assimilation rates did not match Monod dynamics, and a fit to Michaelis–Menten kinetics resulted in unrealistic half-saturation concentrations). In the same experiments, glaciogenic sediments were more bioavailable to *P. tricornutum* than non-glaciogenic sediments rich in iron, provided cells were in physical contact with the abiotic material. Together, these results implicate an evolved molecular mechanism for the 'mining' of mineral iron by some diatoms, which remains to be uncovered. A similar property was observed in the filamentous cyanobacterium *Trichodesmium*, although there too the molecular underpinnings are currently unresolved (Rubin *et al.*, 2011).

There is significant evidence to show that diatoms are able to access the organic pool of ligand-bound iron. Pioneering work by Soria-Dengg and Horstmann (1995) showed that *P. tricornutum* cells were able to access iron from the trihydroxamate siderophores ferrioxamine E (FOE) and ferrioxamine B (FOB), albeit with different kinetics. Working on the same model species, Kazamia *et al.* (2018) demonstrated that this uptake involved endocytosis of the siderophore complex into the cell, with a reduction step to disassociate the bound iron in the vicinity of the chloroplast. Through reverse genetics, the iron starvation-induced protein 1 (ISIP1) was identified as being necessary for the process of endocytosis and siderophore assimilation. Intriguingly, ISIP1 was shown to be largely a diatom-specific protein, implying an evolutionary innovation in this group (Kazamia *et al.*, 2018). The presence of *ISIP1* in diatom genomes and transcriptomes is shown in Fig. 3, although it is important to note that since this gene is highly sensitive to iron status in the cell, it may have been missed from transcriptomes collected from diatoms grown in iron-replete conditions.

Additionally, Coale *et al.* (2019) verified that a previously identified putative siderophore-binding protein, FBP1, which in bacteria binds the hydroxamate siderophore ferrichrome, was indeed required for siderophore assimilation. In our *in silico* searches, 21% of diatom genomes and transcriptomes displayed evidence of a *FBP1* gene (Fig. 3). In physiological experiments on three diatom species reported in Kazamia *et al.* (2018), it was shown that siderophore recognition varied: *P. tricornutum* was unable to use the catecholate siderophore enterobactin as a source of iron, whereas *T. oceanica* assimilated this molecule, but did not respond to ferrioxamine supplementation despite having copies of *FBP1*, *FRE2*, and *ISIP1* in its genome; *T. pseudonana*, which lacks *FBP1* and *ISIP1* genes, was unable to use either siderophore as a source of iron. These observations suggest that while the presence of *FBP1* could be a marker for siderophore uptake in diatoms, further experiments are required to understand what determines substrate specificity in these species.

Laboratory studies on model diatom species, including *P. tricornutum* and *T. oceanica*, report that diatom cells exhibit ferric reductase activity (Maldonado and Price, 2001; Kazamia *et al.*, 2018; Coale *et al.*, 2019). Six genes in *P. tricornutum* have

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

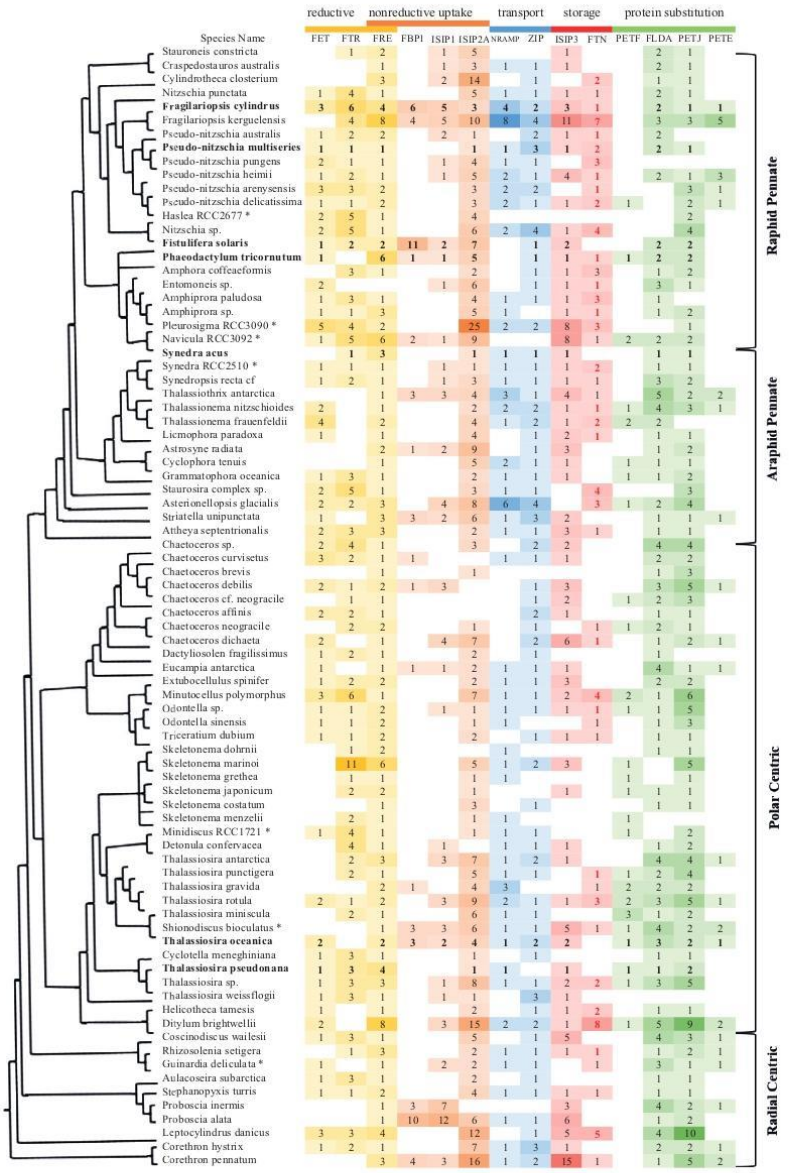


Fig. 3. Iron-related genes in diatom transcriptomes. To investigate the evolutionary distribution of iron-related genes, first, representative protein sequences that are well annotated were downloaded from online databases (GenBank, Uniprot, and Ensembl). Further sequences were retrieved from NCBI by BLASTp with a threshold E value of 1×10^{-5} for each gene of interest. Together, these comprised the reference dataset; they were aligned by MUSCLE and transformed into a Hidden Markov Model using HMMER 3.0. The profiles were used for HMMER searches against the diatom transcriptomes from MMETSP alongside seven transcriptomes generated by Genoscope (indicated with asterisks) and diatom genomes (indicated in bold text). In total, our searchable database comprised 82 species, which met decontamination standards described in Johnson *et al.* (2019). The seven species sequenced by Genoscope are among the 10 most abundant diatom genera in the *Tara* Oceans dataset. Where no hits were detected, the grid is blank, otherwise the shade of the colour and the number within indicates the number of distinct transcripts encoding each gene (but not the

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

been putatively annotated as ferric reductases, but only two (*FRE1* and *FRE2*) encode two domains indicative of ferric reductase function: the ferric reductase transmembrane component PF01794 and the NAD-binding domain PF08030 (Zhang *et al.*, 2013). The *in silico* prediction for FRE2 is that it is targeted to the chloroplast, albeit with low confidence, whereas the localization of FRE1 is unassigned (Rastogi *et al.*, 2018). We built a HMM for putative diatom ferric reductases based on reference genes from *P. tricornutum*, and searched the available diatom genomes and transcriptomes. Our results show that all species appear to harbour genes encoding ferric reductases, with around two per species on average (Fig. 3). *Fragilariopsis kerguelensis* and *F. cylindrus*, abundant diatoms in the Southern Ocean, stand out because their genomes encode eight and four distinct peptides annotated as ferric reductases, respectively. The coastal species *Ditylum brightwellii* has eight *FRE* genes.

The question arises whether these ferric reductases act on the surface of diatom cells, thus enabling the dissociation of iron from its seawater chelates before assimilation, or act intracellularly. On the one hand, extracellular reduction is unlikely in the highly oxidizing and diffusing environment of seawater. On the other hand, if it were to be coupled to efficient (i.e. rapid and proximal) iron-uptake systems, extracellular reduction could allow cells to dispense with the need to discriminate between iron sources (Shaked *et al.*, 2005). The latter is a system well described in fungi, where ferric iron is displaced from weak ferric ligands by a ferrireductase, then re-oxidized by a multicopper oxidase and finally channelled across the plasma membrane through a Fe^{3+} permease (reviewed by Philpott, 2006). Transcriptional up-regulation of a ferrireductase and two iron permeases was recorded under limiting iron conditions in *T. pseudonana* cells (Kustka *et al.*, 2007). Furthermore, bioinformatic analysis of transcriptome datasets confirm the existence of functional analogues of the yeast-like iron acquisition machinery [the 'ferrireductase (FRE), multicopper oxidase (FET), iron permease (FTR)' system] in the model green alga *C. reinhardtii* and across a range of diatom species, including all members of the *Pseudo-nitzschia* and *Fragilariopsis* genera (Fig. 3; also reviewed in Groussman *et al.*, 2015). However, the question of whether this reductive system acts in consort on the surfaces of diatom cells requires experimental verification. Therefore, in Fig. 3 we have indicated the *FRE* genes as acting in both 'reductive' and 'non-reductive' uptake systems.

Low iron quota diatoms regulate the amount of iron required for photosynthesis

The metabolic iron requirement (or metabolic iron quota) of diatom cells differs from species to species. This is measured as the Fe:C ratio (μmol Fe mol C^{-1}) and is estimated from co-measurements of iron-limited growth rates and cellular iron content (Sunda and Huntsman, 1995). For example, in *Pseudo-nitzschia* diatoms grown under low-iron conditions, the Fe:C ratio ranged from 2.8 to 3.7 μmol Fe mol C^{-1} (Marchetti *et al.*, 2006). Similar metabolic iron quotas were estimated for *T. oceanica* (at ~2 μmol Fe mol C^{-1}), whereas other members of the *Thalassiosira* genus, such as *T. pseudonana* and *T. weissflogii*, had 10-fold higher requirements (Cohen *et al.*, 2018b). In the same study, the metabolic Fe:C ratio for *Corethron hystrix* was estimated at ~40 μmol Fe mol C^{-1} (Cohen *et al.*, 2018b). Diatoms with low metabolic iron quotas are remarkable for their ability to maintain uncompromised photosynthetic function, which is attributed to two main adaptations: (i) preferred use or complete replacement of iron-containing proteins with equivalents that are not dependent on iron, and (ii) re-arrangements of the photosynthetic architecture.

Diatoms replace iron-requiring proteins with iron-free functional equivalents

Quantitatively, iron is the most important trace metal in the photosynthetic apparatus because it is involved in the Photosystem II complex (PSII) (which requires two iron atoms per subunit), the cytochrome b_6f complex (five iron atoms per monomer), cytochrome c_6 (one iron atom per monomer), the Photosystem I complex (PSI) (12 iron atoms), and ferredoxin (two iron atoms) (Raven *et al.*, 1999; Strzepek *et al.*, 2012). Ferredoxin (PETF) is an iron–sulfur protein that is a key component of the chloroplast electron transport chain. Replacement of ferredoxin by its iso-functional carrier, the flavin-containing protein flavodoxin (FLDA), allows flavin rather than iron to be used for electron transport. The two proteins appear to have similar electrostatic potential profiles, although flavodoxin undergoes two successive single electron reductions, with only the second step matching ferredoxin potential (Sétif, 2001). Among the diatoms with sequence information available, 20 out of 82 species harbour both *PETF* and *FLDA* genes (Fig. 3). There appears to be no phylogenetic

expression levels). The subcellular localization of FTN was predicted using ASAFind, HECTAR, MitoFates, TargetP, and WolfPSort (Bendtsen *et al.*, 2004; Emanuelsson *et al.*, 2007; Horton *et al.*, 2007; Gschloessl *et al.*, 2008; Fukasawa *et al.*, 2015; Gruber *et al.*, 2015). Sequences predicted to target to the chloroplast are shown in red. It is important to note that this dataset likely misses genes that are under strong transcriptional control, which was not met by the culture conditions during RNA harvesting (e.g. low-Fe-induced genes in diatom cultures grown under Fe-replete conditions), as well as chloroplast-encoded genes. The sequencing depth between the transcriptomes and between genomes varied, and this may have influenced our results. The results presented here may be compared to similar published reviews, notably those by Blaby-Haas and Merchant (2013), Groussman *et al.* (2015), and Behnke and LaRoche (2020). Differences in the numbers of hits retrieved between analyses are likely due to differences in chosen methodologies. FET, ferroxidase; FTR, iron (III) permease; FRE, ferric reductase; FBP, ferrichrome-binding protein; ISIP, iron starvation-induced protein; FTN, ferritin; NRAMP, natural resistance-associated macrophage protein; ZIP, zinc transporter; PETF, ferredoxin; FLDA, flavodoxin; PETJ, cytochrome c_6; PETE, plastocyanin.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

relationship between species that encode *PETF*, implying that it was present in the diatom ancestor and that its absence from some species is likely due to gene loss (Figs 3, 4A). The ability to synthesize flavodoxin is widespread among diatoms; we found *FLDA* in 70 out of 82 species (Fig. 3). Since no photosynthetic species are known to survive without either *PETF* or *FLDA*, in transcriptomes where neither transcript was detected, such as that of *Pseudo-nitzschia pungens*, we assume that this was due either to incomplete sequencing or to the stringency of the cut-off parameters chosen for detection in our analysis.

For species that encode both *FLDA* and *PETF*, the ratio of *PETF*:*FLDA* in the cell is controlled by the availability or scarcity of iron. For example, cultures of *F. cylindrus* grown at high iron concentrations produced predominantly ferredoxin, with a small amount of flavodoxin. Ferredoxin was sequentially replaced by flavodoxin in cultures grown with less iron (Pankowski and McMinn, 2009). Similarly, in *P. tricornutum*, *T. oceanica*, and *Pseudo-nitzschia granii*, expression of the *FLDA* gene and the abundance of FLDA protein was much higher under iron limitation (La Roche *et al.*, 1995; Allen *et al.*, 2008; Lommer *et al.*, 2012; Cohen *et al.*, 2018a; Zhao *et al.*, 2018). Under iron-replete conditions, PETF completely replaces FLDA, as observed in natural diatom communities (La Roche *et al.*, 1995; Erdner and Anderson, 1999; McKay *et al.*, 1999; Allen *et al.*, 2008). However, some diatoms contain multiple copies of genes encoding flavodoxin, and only certain isoforms are differentially expressed in relation to iron status. Both *F. cylindrus* and *T. oceanica* have two isoforms of *FLDA*, but transcription of only one isoform is regulated by iron levels (Pankowski and McMinn, 2009; Whitney *et al.*, 2011). Whether possessing a number of flavodoxin isoforms confers an ecological advantage in environments chronically deprived of iron is an interesting question for further investigation.

The strategy of differential transcription of selected genes does not appear to be used by all diatoms. Individual diatom species may be permanently adapted to specific iron regimes in the ocean rather than maintaining transcriptional flexibility. This was supported by analysis of the *Tara* Oceans global dataset for transcriptional regulation of ferredoxin and flavodoxin across a range of algal groups. The dataset combined comprehensive bio-oceanographic genomic and transcriptomic data with iron distributions derived from two global-scale biogeochemical models (Bork *et al.*, 2015; Carradec *et al.*, 2018). Unlike chlorophytes, haptophytes, and pelagophytes, diatoms showed constitutively higher expression of flavodoxin genes than ferredoxin genes, although there was more heterogeneity in expression between species than across these four major groups (Carradec *et al.*, 2018). In *P. granii*, the transcript abundance of *PETF* was appreciably lower than that of *FLDA* regardless of iron status (Cohen *et al.*, 2018a). The diatom *Fragilariopsis curta* appears to have lost *PETF* entirely, and transcription of this species' remaining *FLDA* gene is not sensitive to iron concentrations (Pankowski and McMinn, 2009). Interestingly, *PETF* is localized in the chloroplast genome in *T. pseudonana* and other diatoms, but has been transferred to the nuclear genome in *T. oceanica*. Compared with its coastal relative *T. pseudonana*, the oceanic diatom *T. oceanica* is highly tolerant to iron limitation. It has been proposed that the transfer of *PETF* from the chloroplast to the nuclear genome might have contributed to the ecological success of *T. oceanica* (Lommer *et al.*, 2010). Whether or not the gene transfer described for *T. oceanica* confers a competitive advantage still needs to be assessed through experimental approaches. That *PETF* is often chloroplast-encoded suggests that analyses of transcriptomes such as the one presented in Fig. 3 may be methodologically biased towards a lack of detection, as RNA harvesting and processing select for nuclear-encoded transcripts.

Cytochrome c_6 acts as an electron carrier between the cytochrome b_6f complex and PSI. Cytochrome c_6 (encoded by *PETJ*) may be replaced with the copper-coordinating protein plastocyanin (encoded by *PETE*). The replacement of cytochrome c_6 with plastocyanin is rarer than the replacement of ferredoxin by flavodoxin, and is presumed to have occurred via horizontal gene transfer (Strzepek and Harrison, 2004;

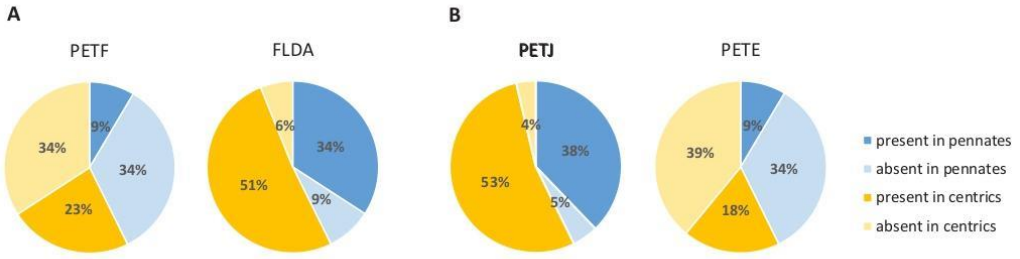


Fig. 4. Genes encoding iron-switching proteins in diatom transcriptomes. Both ferredoxin (PETF) and cytochrome c_6 (PETJ) are iron-containing enzymes. As some diatoms are capable of replacing them with the iron-free equivalents flavodoxin (FLDA) and plastocyanin (PETE), respectively, the relative proportions of transcripts encoding these proteins may indicate the dependence of diatoms on these proteins. The pie charts indicate the presence and absence of *PETF* versus *FLDA* (A), and *PETJ* versus *PETE* (B) in pennate and centric diatoms in our database of 82 species. It is important to note that our comparisons here rely on the numbers of genes detected in the transcriptomes of species, and therefore may not be a true reflection of species' genomes. It is possible that genomes contain genes that were not detected in the transcriptomes, either because the transcriptomes were generated from iron-replete cultures or because species were not sequenced to sufficient depth.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Peers and Price, 2006). Only fifteen centric species and eight pennate species show evidence for the presence of a *PETE* gene (Figs 3, 4B). For species that encode both *PETE* and *PETJ*, differences in regulation have been noted. *P. tricornutum* and *F. cylindrus* were found to induce plastocyanin under iron limitation, to temporarily replace cytochrome c_6, and to highly express the gene encoding cytochrome c_6 under iron-replete conditions (La Roche *et al.*, 1996; Cohen *et al.*, 2018a). *F. kerguelensis* is an interesting candidate for investigation because it contains multiple isoforms of plastocyanin (five genes detected using our methods; Fig. 3). In a recent study, Moreno *et al.* (2020) investigated the response of three isoforms of *PETE* to iron in this species, and found that only two were significantly over-represented under low-iron conditions. In *T. oceanica,* which encodes *PETE* together with two functional copies of *PETJ*, the relative expression of plastocyanin was sensitive to iron status, while two genes encoding cytochrome c_6 were weakly but constitutively expressed (Lommer *et al.*, 2012). Similarly, in *P. granii*, transcripts for *PETJ* were weakly abundant regardless of iron status, while *PETE* was highly abundant under iron-replete conditions (Cohen *et al.*, 2018a).

Only a few studies have addressed the question of what the consequences of using plastocyanin are for the copper requirements of cells. For example, the green alga *C. reinhardtii* switches to using cytochrome c_6 instead of plastocyanin only under copper-limiting conditions (Merchant *et al.*, 1991). In *T. oceanica*, copper deficiency inhibited electron transport regardless of iron status, implying that plastocyanin expression was not controlled by iron concentrations (Peers and Price, 2006).

Low iron quota diatoms lower the ratio of PSI to PSII without compromising photosynthetic output

The PSI complex has the highest iron demand of the light-dependent reactions of photosynthesis. The optimized ratio of PSI: PSII in land plants is approximately 1:1. Comparing the coastal diatom *T. weissflogii* with the open ocean diatom *T. oceanica*, Strzepek and Harrison (2004) found that this ratio was significantly lower in the latter species. This led to the hypothesis that diatoms, in particular those adapted to chronically iron-starved conditions, have streamlined their photosynthetic architecture, minimizing the iron quota necessary for growth. *T. weissflogii* cells contained twice as much PSII than PSI, and *T. oceanica* had reduced PSI demands even further, with a ratio of PSI:PSII of 1:10 (Strzepek and Harrison, 2004). In iron-replete media, the two diatoms grew at comparable rates. However, *T. oceanica* maintained high growth rates (~80% of the rate in iron-replete conditions) in low-iron media that restricted the growth of *T. weissflogii* to ~20% of its iron-replete growth rate. Two studies confirmed similar streamlining in *F. kerguelensis* and *P. granii,* polar diatoms that are abundant in the Southern Ocean and the Arctic Ocean, respectively (Cohen *et al.*, 2018a; Moreno *et al.*, 2020). In *P. granii*, the gene encoding PSI subunit IV (*PSAE*) was almost 4-fold more highly expressed under iron-replete conditions, and protein levels were more abundant by 35-fold compared with their levels under conditions of iron limitation (Cohen *et al.*, 2018a).

Low iron quota diatoms such as *T. oceanica* and *P. granii* are not unique among algae and cyanobacteria in having undergone substitutions and rearrangements in their photosynthetic architecture in order to conserve iron. These traits and plasticity in iron use are widespread (Blaby-Haas and Merchant, 2013; Raven, 2013; Scheiber *et al.*, 2019). However, in an experimental comparison of 29 species of eukaryotic algae and cyanobacteria, Quigg *et al.* (2003, 2011) demonstrated that algae that contain plastids of a secondary endosymbiotic origin, the 'red plastid' lineage, which includes the dinoflagellates, haptophytes, and chrysophytes as well as diatoms, had lower stoichiometric quantities of iron per cell compared with the 'green lineage' of algae and cyanobacteria. The results from these analyses are difficult to square with measurements of metabolic iron quotas, since the former are carried out under iron-replete conditions. Nevertheless, they suggest that members of the red plastid lineage are capitalizing on more than rearrangements of their photosynthetic architecture for efficient iron use. Studies that look into the adaptations of cell structure, such as the metabolic coupling of plastids and mitochondria (e.g. see some pioneering work by Bailleul *et al.*, 2015), may shed further light on this question.

Diatoms are able to store iron and regulate its intracellular concentration efficiently

It has been observed that *in situ* iron fertilization, whether artificial or natural, results in blooms dominated by large diatoms, which are often rare in the standing microalgal community (De Baar *et al.*, 2005). These include the chain-forming members of the genera *Fragilariopsis*, *Pseudo-nitzschia*, and *Chaetoceros*, in particular. Comparing their transcriptomes for the analysed genes (Fig. 3), there is no pattern that distinguishes them readily from other diatoms, as traits appear to be species-specific. However, it is important to note that these transcriptomes capture the behaviour of species under laboratory conditions and may not be indicative of their behaviour during a diatom bloom in the wild. The *F. kerguelensis* transcriptome and the *F. cylindrus* genome encode a more comprehensive portfolio of iron-sensitive genes, with multiple isoforms of each gene compared with other species. This is intriguing because, as model polar species, *Fragilariopsis* representatives are considered to be particularly adapted to fluctuating environmental conditions and life in sea ice, which is not 'iron-limited' (Mock *et al.*, 2017). Our data suggest that *Fragilariopsis* species are interesting candidates for iron adaptation studies as well.

It is likely that species that respond to iron stimulation by forming blooms are efficient not only at iron uptake, but also in iron homeostasis and long-term storage. One way to

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

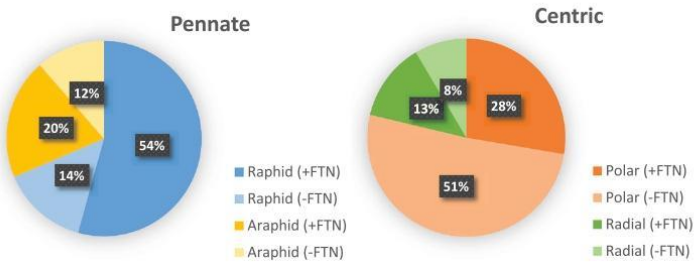


Fig. 5. Genes encoding ferritins (FTNs) in diatom transcriptomes. Homologs of *FTN* were retrieved using a HMMER search within available diatom genomes and transcriptomes (see the caption of Fig. 3 for details). Pie charts summarize the presence and absence of *FTN* in raphid versus araphid pennate diatoms, and in polar versus radial centric species. It is important to note that the presence or absence of genes in a transcriptome may not be a true reflection of copy numbers in a genome, either because it was generated from iron-replete cultures or because it was not sequenced to sufficient depth.

assess the capacity for long-term iron storage is to measure the number of divisions that a cell is capable of when ambient iron concentrations drop back to growth-limiting conditions. This can be distinguished from short-term iron homeostasis, which is associated with mechanisms that act on a diurnal scale, where the requirement for iron depends on photosynthetic activity. An alternative is to measure the cell's metabolic iron quota and subtract this from the total cellular iron carried by a cell under iron-replete conditions. This has been termed the 'luxury iron' quota (Sunda and Huntsman, 1997; Marchetti *et al.*, 2009). For example, within the centric *Thalassiosira* genus, the cellular Fe:C ratio of coastal species reached values that were 20–30 times higher than those needed for maximum growth; by contrast, 'plateau' iron concentrations in oceanic species were only two to three times higher than maximum needed amounts (Sunda and Huntsman, 1995), indicating that coastal variants had greater capacity to accumulate iron when iron replete.

Two mechanisms have been proposed for the storage of iron inside diatom cells: sequestration into the mineral core of ferritin (FTN) proteins and/or vacuolar storage. FTNs are found in all domains of life, including animals, plants, and microorganisms, and are designed to accommodate large amounts of iron (Theil, 1987). Structurally, FTNs are large protein cages formed by arrays of self-assembling helices with nanocavities (5–8 nm) that catalytically couple iron and oxygen at protein sites for precursors of the cavity mineral. The mineral core of a single protein can store up to 4500 iron atoms (Liu and Theil, 2005). Iron is reversibly released from the core by reduction.

We retrieved putative transcripts for *FTN* in diatom transcriptomes and genomes by conducting a HMMER search using the methods described in Fig. 3. We found that ~74% of pennate diatoms contain at least one *FTN* homolog compared with 41% of centric diatoms. Among the centric diatoms the trend for lacking *FTN* skewed towards the polar centric species, of which approximately two thirds were missing the annotation (Fig. 5). The apparent loss of *FTN* in the polar centric diatoms is intriguing and requires further investigation. Looking at the transcriptomes of the three genera of frequently blooming diatoms, *Fragilariopsis* and *Pseudo-nitzschia* species appear to be rich in putative *FTN* genes (Fig. 3). In contrast, our search for *FTN* retrieved hits for only a quarter of the sampled *Chaetoceros* strains (two out of eight). There were notable absences among *Thalassiosira* species (among which five species out of nine do not have a *FTN* gene), including *T. oceanica* and *T. pseudonana*.

In laboratory studies, it was shown that *P. granii,* which harbours a gene encoding *FTN*, was able to undergo several more cell divisions in the absence of iron than the comparably sized oceanic centric diatom *T. oceanica*, which lacks the *FTN* gene, supporting the hypothesis that *FTN* functions in the long-term storage of iron (Marchetti *et al.*, 2009). A corollary of the hypothesis is that *FTN* transcript abundance should increase with iron concentration. However, this has not been universally upheld in laboratory studies. While *P. granii* and *Thalassiosira* sp. NH16 have been observed to increase *FTN* gene expression under high iron concentrations, *Amphora coffeaeformis* exhibited minimal iron-storage capacities and contained two distinct *FTN* genes, one of which increased in expression under iron limitation whereas the second showed no variation with cellular iron status (Cohen *et al.*, 2018b). In fact, we detected three distinct *FTN* transcripts in this species using our methods (Fig. 3). Furthermore, a recent study investigated the community-level response of open ocean plankton ecosystems to iron availability and found that, with the exception of *Pseudo-nitzschia* species, no clear pattern in *FTN* gene abundance or expression and estimated iron levels could be observed, suggesting that iron storage may not be the main function of FTN in most eukaryotic marine phytoplankton (Caputi *et al.*, 2019). Together, these results suggest that FTN in diatoms may have evolved to serve multiple functional roles.

A proposed alternative role for FTN in diatoms is as an iron oxidation enzyme rather than as a long-term iron-storage

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

protein. Working with *P. multiseries*, Pfaffen *et al.* (2013) showed that FTN oxidizes Fe^{2+} at its ferroxidase centres rapidly but forms iron mineral only slowly. In a subsequent study, functional mutagenesis experiments showed that the protein is biochemically optimized for initial Fe^{2+} oxidation but not for mineralization. The authors argued that its primary function therefore might not be in long-term iron storage, but rather in iron homeostasis (Pfaffen *et al.*, 2015). This is the situation in higher plants, where experiments on *Arabidopsis thaliana* demonstrated that *FTN* is regulated by the circadian clock cycle (Duc *et al.*, 2009) and functions to buffer the iron released by degradation from PSI, a protective role in response to photo-oxidative stress (Rossel *et al.*, 2002; Murgia *et al.*, 2007). Similarly, in *C. reinhardtii*, FTN was shown to be required during high ambient iron availability or cellular degradation of iron-containing proteins and protein complexes such as ferredoxin during cellular acclimation to low-iron conditions, indicative of a function in iron homeostasis (La Fontaine *et al.*, 2002; Busch *et al.*, 2008; Long and Merchant, 2008). Finally, in the green picoalga *Ostreococcus tauri*, mutants lacking FTN were less tolerant of low-iron conditions, which induced greater recycling of iron within the cell than in the wild type, further underscoring the importance of this protein in homeostasis. In the same work, the authors proposed that nitrate reductase functioned as an iron-storage protein in this picoalga (Botebol *et al.*, 2015).

A closer look at diatom FTN phylogeny, which we updated using available diatom transcriptomes (Fig. 6), reveals that the FTNs within the diatoms resolve into several clades. Our phylogeny matches that previously reported by others (e.g. Marchetti *et al.*, 2009; Groussman *et al.*, 2015), but adds more species to the tree. There are two main FTN clades, one of which contains most members of the *Thalassiosira*, *Pseudo-nitzschia*, and *Fragilariopsis* genera, although phylogenetic relationships are not robustly upheld. The second clade contains members of 14 genera, including *Cylindrotheca* and *Leptocylindrus*. The species *Helicotheca tamesis* CCMP826, *Amphiprora paludosa* CCMP125, *Thalassionema fraundfeldii* CCMP1798, *Staurosira* complex sp. CCMP2646, and *Pleurosigma* sp. RCC3090 have annotated FTNs belonging to both clades. To our knowledge, there are no described functional differences between FTNs that belong to the two main groups. Additionally, we found a third group of putative FTNs (which we have annotated as FTN clade III in Fig. 6) with weaker bootstrap support, which includes genes from *P. tricornutum*, *Staurosira* complex sp., *Guinardia deliculata*, *Chaetoceros neogracile*, and *F. kerguelensis*.

We compared the protein sequence of representative FTNs from the three groups by alignment (Fig. 7). We focused on a comparison of iron-binding sites identified by Pfaffen *et al.* (2013) working on *P. multiseries* and key residues identified by Jin *et al.* (2001) working on FTN isolated from the frog *Rana catesbeiana*. It would appear that essential residues involved in iron binding are more strictly conserved in FTN I and II (Fig. 7, red arrows), whereas residues involved in iron release are preserved more across FTN I and III. However, experimental verification is required to validate any implied functional differences.

There is evidence from on-board experiments (which use environmental samples) to suggest that *Pseudo-nitzschia* ferritins serve a long-term storage role in cells, which is not the case in *Thalassiosira* and *Corethron*, in which a role in iron homeostasis is preferred as an explanation (Lampe *et al.*, 2018). Broadly, this is also supported by *Tara* Oceans transcriptomic analyses, which detected up-regulated *FTN* expression in *Pseudo-nitzschia* cells in locations that are rich in iron (Caputi *et al.*, 2019). One structural change that may explain this is the presence of glutamic acid at position 130 in *P. multiseries* FTN, which is absent from other FTN clade I and II species (Pfaffen *et al.*, 2015). It should be noted that E130 is conserved in FTN III diatoms (Fig. 7).

It is likely that a range of proteins that have not yet been explored are involved in iron storage and homeostasis in diatom cells. One notable candidate is the iron starvation-induced protein ISIP3, the most abundant of the three ISIPs in global datasets (Carradec *et al.*, 2018). ISIP3 is typified by a conserved domain of unknown function, DUF305, which belongs to the FTN superfamily (Behnke and LaRoche, 2020). Using our methods, we found transcripts for *ISIP3* across most diatoms, with some notable absences in the coastal *Skeletonema* and *Chaetocerus* genera and an overabundance of putative hits in *Fragilariopsis* (Fig. 3).

An alternative mechanism for long-term iron storage that has been proposed is vacuolar. Iron loading into and release out of the vacuole was first identified in yeast, and was shown to be mediated by natural resistance-associated macrophage proteins (NRAMPs) in *A. thaliana* (Curie *et al.*, 2000; Lanquar *et al.*, 2005). Furthermore, some NRAMP proteins were shown to be functional equivalents of yeast FET proteins, and enabled the transport of iron into the cell (Curie *et al.*, 2000). In *T. pseudonana*, a vacuolar storage mechanism was proposed when it was observed that NRAMP was dramatically regulated by intracellular iron concentrations (Kustka *et al.*, 2007). However, it is important to note that vacuolar localization of NRAMP was proposed but not experimentally validated. In experiments using synchrotron X-ray fluorescence element mapping, intracellular pockets of iron, indicative of storage features, were identified in both *T. pseudonana* and *T. weissflogii* (Nuester *et al.*, 2012).

A recent study investigated the responses of natural communities, sampled across a gradient of iron concentrations, to on-deck iron stimulation. Metatranscriptome data from incubations was used to assign relative expression of *FTN* and *NRAMP* genes on to three main diatom genera, *Chaetoceros*, *Pseudo-nitzschia*, and *Thalassiosira* (Lampe *et al.*, 2018). The study found that *Pseudo-nitzschia* diatoms were unique utilizers of *FTN* (up-regulating its expression when iron was supplied), whereas the response from *Thalassiosira* species suggested

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

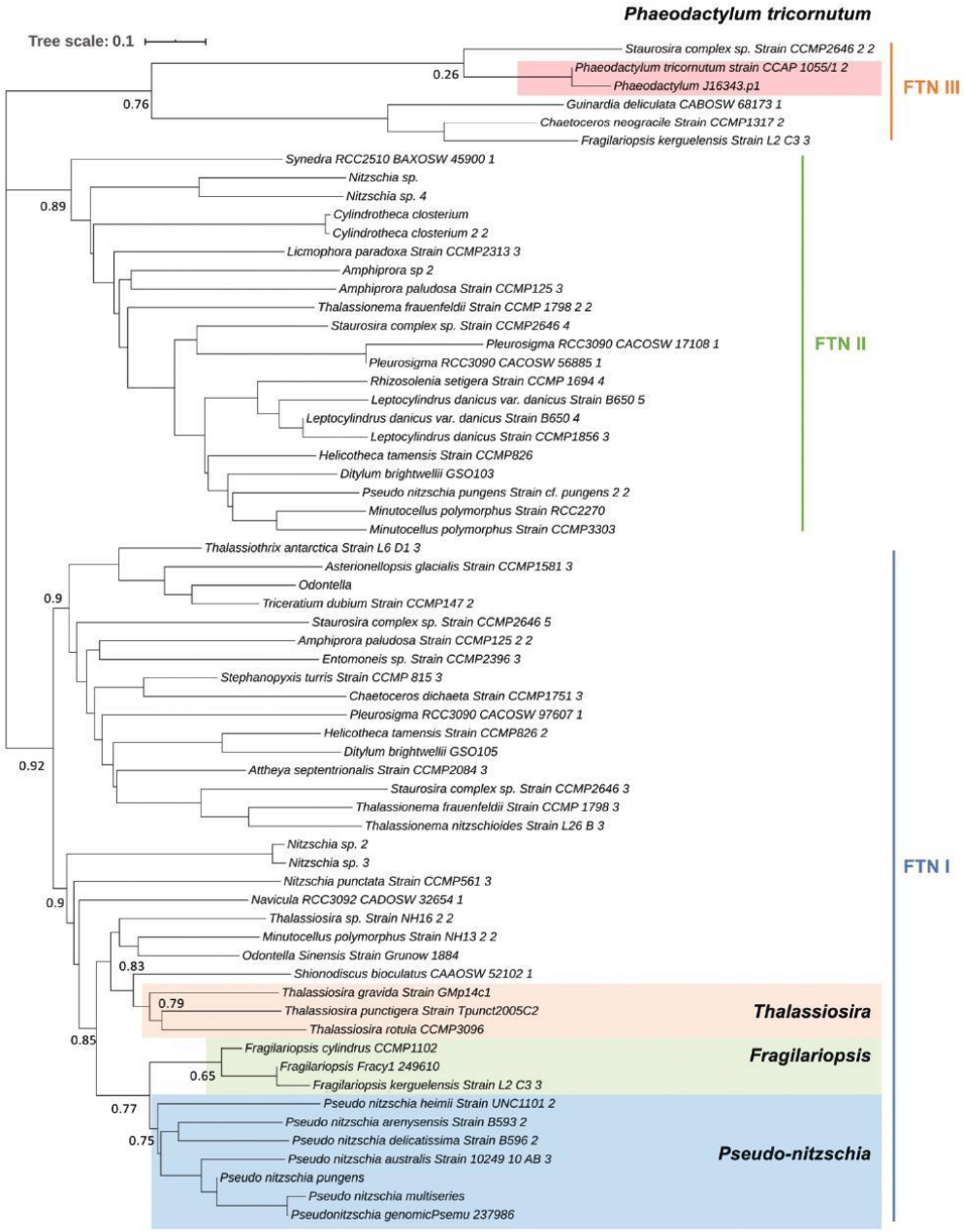


Fig. 6. Phylogenetic tree of ferritins (FTNs) in diatoms. The phylogenetic tree was constructed to clarify the evolutionary distance of *FTN* between species. A total of 244 sequences were first retrieved by a HMMER search using the FTN PFAM domain PF00210 as a query in the available 82 diatom transcriptomes/genomes with an E-value cut-off of 1×10^{-10}. Retrieved sequences were analysed using the CD-HIT web server and sequences that met a

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

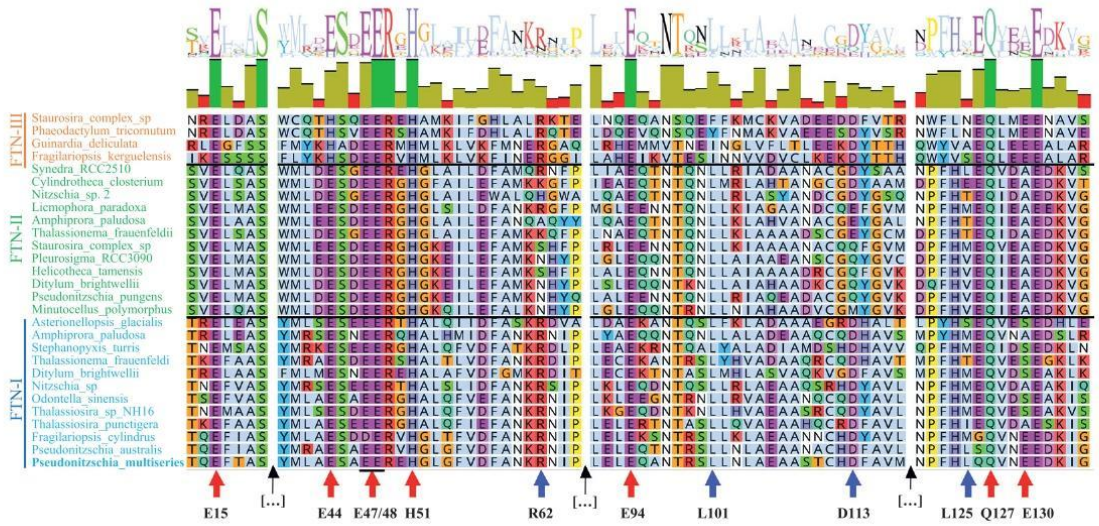


Fig. 7. Protein sequence alignment of representative diatom ferritins (FTNs). FTN protein sequence alignment (one-letter code) was carried out using MUSCLE in Geneious v.10.2. At the top is the sequence logo and identity. At the bottom, red arrows highlight the ferroxidase residue sites first identified in *Pseudo-nitzschia multiseries* (Pfaffen *et al.*, 2013). The glutamic acid residues (E15, E47/48, and E94), one glutamine residue (Q127), and the histidine residue H51 are conserved across the three different clades (red arrows). However, the glutamic acid residue E44, which is conserved in FTN I/II, is replaced with histidine in FTN III. These residues are involved in iron binding. Blue arrows show conserved residue pairs essential for iron release in the frog species *Rana catesbeiana* (Jin *et al.*, 2001). The amino acid pairs R62/D113 and L101/L125 are adjacent in three-dimensional space. R62/D113 is conserved in both FTN I and FTN III but not FTN II (which shows significant variation in these positions). Leucine is conserved in position 101 in FTN I and II but not FTN III, while leucine at position 125 is not conserved in any diatoms apart from two *Pseudo-nitzschia* species and *Synedra* sp. RCC2510. The most common replacement for leucine in this position is methionine, as observed in *F. cylindrus* and *Thalassiosira* sp. NH16. Black arrows indicate breaks in the sequence.

vacuolar storage. In *Chaetoceros*, transcripts of neither *FTN* nor *NRAMP* were abundant, although iron quotas in *Chaetoceros* were often similar to those in *Pseudo-nitzschia* (Lampe *et al.*, 2018). The authors proposed that *Chaetoceros* may have an alternative divalent metal transporter, a protein belonging to the ZIP family, that takes on the function of vacuolar iron transport in these species (Allen *et al.*, 2008). ZIP proteins have been shown to facilitate passive metal transport, including that of ferrous iron, in a range of species (Eide, 2005).

We compared the portfolios of *ZIP*, *NRAMP*, and *FTN* genes across diatoms using all published diatom transcriptomes and genomes, by constructing HMMs for representative genes and scanning the dataset using HMMER (summarized in Fig. 3). This is an amalgamated view, since we did not refine our search for subclasses of NRAMP or ZIP proteins (i.e. the specificity of these genes for iron transport is not confirmed). We found that all diatoms have genes belonging to the *ZIP* family, and that the majority also have *NRAMP* genes (modelled on *F. cylindrus* NRAMP protein). Notable exceptions are most members of the *Chaetoceros* genus, which lack a *NRAMP* annotation. *NRAMP* genes appear to be less abundant among centric species than pennate species, but no robust phylogenetic relationship was observed.

Synthesis and future directions

There are multiple strands of evidence suggesting that diatoms as a group are successful competitors for iron. Significant research attention has focused on iron uptake and storage, and it is clear that diatoms use a portfolio of strategies to access various forms of iron in the ocean. Remarkably, they have adapted what appear to be ancestrally bacterial uptake proteins to eukaryotic mechanisms. The genes *FBP1* and *ISIP2A* have bacterial origins yet rely on endocytosis to function.

similarity threshold >0.9 were presumed to be duplicates and eliminated. We generated a Hidden Markov Model for *FTN* using *P. tricornutum* and closely aligned sequences retrieved by BLASTp, and this was used as the basis for a second HMMER search of the remaining sequences, with an E-value cut-off of 1×10^{-10}, to further reduce the redundancy. This left a total of 64 representative sequences, which captured the diversity of FTNs within the diatoms. Conserved sequences were aligned using the alignment builder in Geneious v.10.2 under default criteria. The tree was drawn with ITOL (https://itol.embl.de/). Numbers beside branches indicate RaxML bootstrap coefficients.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Diatom-specific genes have also been observed (e.g. *ISIP1*), implying further innovation in this group to facilitate access to iron. However, experimental verification of the molecular basis of proposed traits remains a major challenge. For example, validation of the *in silico* prediction that FTN is targeted to the chloroplast in diatoms, where iron demands are highest, should be an experimental priority. Additionally, multiple studies have indicated that ZIP and NRAMP proteins, as well as ISIP3, are important for iron physiology, making them clear targets for forward and reverse genetics.

One important avenue for future work is gauging which of these mechanisms is most strongly related to the success of diatoms in the wild. There is evidence that the replacement of ferredoxin with flavodoxin in the photosynthetic machinery of diatom cells is one of the more important adaptations. A recent global survey of gene expression in marine phytoplankton showed that oceanic members of the diatom lineage strongly express flavodoxin over ferredoxin, whereas certain coastal diatoms, which are likely adapted to environments experiencing more frequent and larger fluctuations in iron supply, expressed ferredoxin more highly under iron-replete conditions (Caputi *et al.*, 2019). Of the seven most abundant genera in the ocean (indicated by asterisks in Fig. 3), four contain genes encoding flavodoxin.

Similarly, the expression of ISIPs showed the highest correlation in communities that were sampled from the lowest-iron environments (Kazamia *et al.*, 2018; Caputi *et al.*, 2019). *ISIP1* is notable for being the most sensitive to low-iron status, and likely represents a diatom-specific innovation, although its function requires further elucidation. Its role in mediating siderophore uptake is intriguing, because it suggests a community link between diatoms and siderophore producers. FBP1 is the only siderophore-binding protein that has been described to date, and its distribution appears to correlate with that of ISIP1 in diatom transcriptomes, since most species that encode FBP1 also encode ISIP1, with the exception of *Thalassiosira gravida* and *Chaetoceros curvisetus* (Fig. 3). However, many more species encode ISIP1 than FBP1, so it is possible that ISIP1 has an additional role in cell iron homeostasis or that there are multiple siderophore-binding proteins interacting with ISIP1.

As we garner information on how diatoms put their metabolic portfolios to use in the wild, we will shed a refining light on diatom niches. This will be an important breakthrough in moving away from studies that either focus on individual model species or attempt to draw distinctions between 'pennates' and 'centrics', or 'open ocean' and 'coastal' diatoms, while comparing only a few representatives. For the iron uptake, homeostasis, or storage mechanisms that we have reviewed here, there do not appear to be any remarkable patterns that distinguish pennates from centrics. To meaningfully compare open ocean and coastal diatoms, we require more information on the distribution of individual species in the global ocean. The *Tara* Oceans global circumnavigation effort, which catalogued the community composition, metagenomes, and metatranscriptomes of aquatic microorganisms, can offer initial insight. For example, *T. oceanica*, the species most often referred to in the literature as an open ocean species, in fact has a ubiquitous distribution and is frequently found in coastal areas (based on 18S data from the *Tara* Oceans gene catalogue; data not shown). Of the 82 species reviewed here, only *H. tamesis* was not detected within 250 km from the coast during *Tara* Oceans sampling (J. Pierella Karlusich, Institut de Biologie de l'École Normale Supérieure, personal communication), and this species may therefore be considered a true open ocean representative. By contrast, there is a considerable number of coastal species represented in the MMETSP sequence set. The species that were not detected further offshore than 250 km are *Proboscia inermis* (found exclusively in coastal zones at high latitudes), *D. brightwellii*, *Skeletonema marinoi*, *Ticeratium dubium*, *Odontella aurita*, *Navicula* sp., *P. triornutum*, *A. coffeaeformis*, and *Cylindrotheca closterium*.

Using *Tara* Oceans data, Caputi *et al.* (2019) found that diatom species thrived across a gradient of total iron concentrations and showed remarkable plasticity in their responses to iron availability. The authors concluded that it was not possible to correlate species assemblages to iron levels or transcriptional responses in iron-uptake systems. We believe that further progress will be made when *in situ* studies of species-specific diatom gene expression and analysis of community structure will be coupled to careful characterization of iron sources available in seawater, since it is possible that it is the chemical nature of iron sources that complicates diatom niche separation.

Acknowledgements

CB acknowledges funding from the European Research Council Advanced Awards Diatomite and Diatomic under the European Union's Horizon 2020 research and innovation programme (grant agreement nos. 294823 and 835067), the ANR project PhytoMet (grant ANR-16-CE01-0008-02), the HFSP project Green Life in the Dark (grant RGP0003/2016), and the French Government 'Investissements d'Avenir' Programmes MEMO LIFE (grant ANR-10-LABX-54), Université de Recherche Paris Sciences et Lettres (Université PSL) (grant ANR-1253 11-IDEX-0001-02), and OCEANOMICS (grant ANR-11-BTBR-0008). XG acknowledges funding from the Chinese Scholarship Council (CSC-201706910081). We thank Dr Juan Pierella Karlusich and Dr Richard G. Dorrell for their advice on the bioinformatic analysis.

Author contributions

XG and EK wrote the original draft of the manuscript. EK and CB were responsible for reviewing and editing the writing. XG curated and analysed the data and was responsible for the visualization of the work. EK and CB conceptualized and supervised the project. CB and XG secured the funding support.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Conflict of interest

The authors declare no conflict of interest.

References

Allen MD, del Campo JA, Kropat J, Merchant SS. 2007. *FEA1*, *FEA2*, and *FRE1*, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with *FOX1* and *FTR1* in iron-deficient *Chlamydomonas reinhardtii*. Eukaryotic Cell **6**, 1841–1852.

Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G, Fernie AR, Bowler C. 2008. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. Proceedings of the National Academy of Sciences, USA **105**, 10438–10443.

Alverson AJ, Theriot EC. 2005. Comments on recent progress toward reconstructing the diatom phylogeny. Journal of Nanoscience and Nanotechnology **5**, 57–62.

Apt KE, Grossman A, Kroth-Pancic P. 1996. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. Molecular Genetics and Genomics **252**, 572–579.

Armbrust EV, Berges JA, Bowler C, *et al.* 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. Science **306**, 79–86.

Bailleul B, Berne N, Murik O, *et al.* 2015. Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. Nature **524**, 366–369.

Behnke J, LaRoche J. 2020. Iron uptake proteins in algae and the role of Iron Starvation-Induced Proteins (ISIPs). European Journal of Phycology **55**, 339–360.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology **340**, 783–795.

Benoiston A-S, Ibarbalz FM, Bittner L, Guidi L, Jahn O, Dutkiewicz S, Bowler C. 2017. The evolution of diatoms and their biogeochemical functions. Philosophical Transactions of the Royal Society B: Biological Sciences **372**, 20160397.

Bhatta H, Kong TK, Rosengarten G. 2009. Diffusion through diatom nanopores. Journal of Nano Research **7**, 69–74.

Blaby-Haas CE, Merchant SS. 2013. Iron sparing and recycling in a compartmentalized cell. Current Opinion in Microbiology **16**, 677–685.

Boiteau RM, Mende DR, Hawco NJ, McIlvin MR, Fitzsimmons JN, Saito MA, Sedwick PN, DeLong EF, Repeta DJ. 2016. Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean. Proceedings of the National Academy of Sciences, USA **113**, 14237–14242.

Bonnain C, Breitbart M, Buck KN. 2016. The Ferrojan Horse hypothesis: iron-virus interactions in the ocean. Frontiers in Marine Science **3**, 82.

Bork P, Bowler C, de Vargas C, Gorsky G, Karsenti E, and Wincker P. 2015. *Tara* Oceans studies plankton at planetary scale. Science **348**, 873.

Botebol H, Lesuisse E, Šuták R, *et al.* 2015. Central role for ferritin in the day/night regulation of iron homeostasis in marine phytoplankton. Proceedings of the National Academy of Sciences, USA **112**, 14652–14657.

Bowler C, Allen AE, Badger JH, *et al.* 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature **456**, 239.

Boyd PW, Ellwood MJ. 2010. The biogeochemical cycle of iron in the ocean. Nature Geoscience **3**, 675–682.

Boyd PW, Jickells T, Law C, *et al.* 2007. Mesoscale iron enrichment experiments 1993–2005: synthesis and future directions. Science **315**, 612–617.

Boyle E. 1998. Pumping iron makes thinner diatoms. Nature **393**, 733–734.

Busch A, Rimbauld B, Naumann B, Rensch S, Hippler M. 2008. Ferritin is required for rapid remodeling of the photosynthetic apparatus and minimizes photo-oxidative stress in response to iron availability in *Chlamydomonas reinhardtii*. The Plant Journal **55**, 201–211.

Caputi L, Carradec Q, Eveillard D, *et al.* 2019. Community-level responses to iron availability in open ocean plankton ecosystems. Global Biogeochemical Cycles **33**, 391–419.

Carradec Q, Pelletier E, Da Silva C, *et al.* 2018. A global ocean atlas of eukaryotic genes. Nature Communications **9**, 1–13.

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. Journal of Eukaryotic Microbiology **46**, 347–366.

Coale TH, Moosburner M, Horák A, Oborník M, Barbeau KA, Allen AE. 2019. Reduction-dependent siderophore assimilation in a model pennate diatom. Proceedings of the National Academy of Sciences, USA **116**, 23609–23617.

Cohen NR, Gong W, Moran DM, McIlvin MR, Saito MA, Marchetti A. 2018a. Transcriptomic and proteomic responses of the oceanic diatom *Pseudo-nitzschia granii* to iron limitation. Environmental Microbiology **20**, 3109–3126.

Cohen NR, Mann E, Stemple B, Moreno CM, Rauschenberg S, Jacquot JE, Sunda WG, Twining BS, Marchetti A. 2018b. Iron storage capacities and associated ferritin gene expression among marine diatoms. Limnology and Oceanography **63**, 1677–1691.

Curie C, Alonso JM, Jean ML, Ecker JR, Briat J-F. 2000. Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport. Biochemical Journal **347**, 749–755.

De Baar HJ, Boyd PW, Coale KH, *et al.* 2005. Synthesis of iron fertilization experiments: from the Iron Age in the Age of Enlightenment. Journal of Geophysical Research: Oceans **110**, C09S16.

Dorrell RG, Gile G, Mccallum G, *et al.* 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. eLife **6**, e23717.

Dorrell RG, Villain A, Perez-Lamarque B, *et al*. 2021. Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. Proceedings of the National Academy of Sciences **118**.

Duc C, Cellier F, Lobréaux S, Briat JF, Gaymard F. 2009. Regulation of iron homeostasis in *Arabidopsis thaliana* by the clock regulator time for coffee. Journal of Biological Chemistry **284**, 36271–36281.

Eide DJ. 2005. The ZIP family of zinc transporters. In Iuchi S, Kuldell N, eds. Zinc finger proteins: from atomic contact to cellular function. Boston: Springer, 261–264.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols **2**, 953–971.

Erdner DL, Anderson DM. 1999. Ferredoxin and flavodoxin as biochemical indicators of iron limitation during open-ocean iron enrichment. Limnology and Oceanography **44**, 1609–1615.

Falkowski PG. 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature **387**, 272–275.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. Science **281**, 237–240.

Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. 2015. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Molecular & Cellular Proteomics **14**, 1113–1126.

Galachyants YP, Zakharova YR, Petrova D, *et al.* 2015. Sequencing of the complete genome of an araphid pennate diatom *Synedra acus* subsp. *radians* from Lake Baikal. Doklady Biochemistry and Biophysics **461**, 84–88.

Gledhill M, Buck KN. 2012. The organic complexation of iron in the marine environment: a review. Frontiers in Microbiology **3**, 69.

Groussman RD, Parker MS, Armbrust EV. 2015. Diversity and evolutionary history of iron metabolism genes in diatoms. PLoS One **10**, e0129081.

Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. The Plant Journal **81**, 519–528.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Gschloessl B, Guermeur Y, Cock JM. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics **9**, 393.

Hain M, Sigman D, Haug G. 2014. 8.18 The biological pump in the past. In Holland HD, Turekian KK, eds. Treatise on geochemistry (2nd edition), Vol. **8**: The oceans and marine geochemistry. Amsterdam: Elsevier, 485–517.

Hopes A, Nekrasov V, Kamoun S, Mock T. 2016. Editing of the urease gene by CRISPR-Cas in the diatom *Thalassiosira pseudonana*. Plant Methods **12**, 49.

Hopkinson BM, Morel FM. 2009. The role of siderophores in iron acquisition by photosynthetic marine microorganisms. Biometals **22**, 659–669.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Research **35**, W585–W587.

Jin W, Takagi H, Pancorbo B, Theil EC. 2001. "Opening" the ferritin pore for iron release by mutation of conserved amino acids at interhelix and loop sites. Biochemistry **40**, 7525–7532.

Johnson LK, Alexander H, Brown CT. 2019. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes. GigaScience **8**, giy158.

Kazamia E, Sutak R, Paz-Yepes J, *et al.* 2018. Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. Science Advances **4**, eaar4536.

Keeling PJ, Burki F, Wilcox HM, *et al.* 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biology **12**, e1001889.

Kustka AB, Allen AE, Morel FM. 2007. Sequence analysis and transcriptional regulation of iron acquisition genes in two marine diatoms. Journal of Phycology **43**, 715–729.

La Fontaine S, Quinn JM, Nakamoto SS, Page MD, Göhre V, Moseley JL, Kropat J, Merchant S. 2002. Copper-dependent iron assimilation pathway in the model photosynthetic eukaryote *Chlamydomonas reinhardtii*. Eukaryotic Cell **1**, 736–757.

Lampe RH, Mann EL, Cohen NR, Till CP, Thamatrakoln K, Brzezinski MA, Bruland KW, Twining BS, Marchetti A. 2018. Different iron storage strategies among bloom-forming diatoms. Proceedings of the National Academy of Sciences, USA **115**, E12275–E12284.

Lanquar V, Lelièvre F, Bolte S, *et al.* 2005. Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. The EMBO Journal **24**, 4041–4051.

La Roche J, Boyd PW, McKay RML, Geider RJ. 1996. Flavodoxin as an *in situ* marker for iron stress in phytoplankton. Nature **382**, 802–805.

La Roche JL, Murray H, Orellana M, Newton J. 1995. Flavodoxin expression as an indicator of iron limitation in marine diatoms. Journal of Phycology **31**, 520–530.

Lauvset SK, Gruber N, Landschützer P, Olsen A, Tjiputra JF. 2015. Trends and drivers in global surface ocean pH over the past 3 decades. Biogeosciences **12**, 1285–1298.

Lis H, Shaked Y, Kranzler C, Keren N, Morel FM. 2015. Iron bioavailability to phytoplankton: an empirical approach. The ISME Journal **9**, 1003–1013.

Liu X, Millero FJ. 2002. The solubility of iron in seawater. Marine Chemistry **77**, 43–54.

Liu X, Theil EC. 2005. Ferritins: dynamic management of biological iron and oxygen chemistry. Accounts of Chemical Research **38**, 167–175.

Lommer M, Roy AS, Schilhabel M, Schreiber S, Rosenstiel P, LaRoche J. 2010. Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. BMC Genomics **11**, 718.

Lommer M, Specht M, Roy A-S, *et al.* 2012. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. Genome Biology **13**, R66.

Long JC, Merchant SS. 2008. Photo-oxidative stress impacts the expression of genes encoding iron metabolism components in Chlamydomonas. Photochemistry and Photobiology **84**, 1395–1403.

Mahowald NM, Engelstaedter S, Luo C, *et al.* 2008. Atmospheric iron deposition: global distribution, variability, and human perturbations. Annual Review of Marine Science **1**, 245–278.

Maldonado MT, Price NM. 2001. Reduction and transport of organically bound iron by *Thalassiosira oceanica* (Bacillariophyceae). Journal of Phycology **37**, 298–310.

Malviya S, Scalco E, Audic S, *et al.* 2016. Insights into global diatom distribution and diversity in the world's ocean. Proceedings of the National Academy of Sciences, USA **113**, E1516–E1525.

Mann DG. 1999. The species concept in diatoms. Phycologia **38**, 437–495.

Marchetti A, Maldonado MT, Lane ES, Harrison PJ. 2006. Iron requirements of the pennate diatom *Pseudo-nitzschia*: comparison of oceanic (high-nitrate, low-chlorophyll waters) and coastal species. Limnology and Oceanography **51**, 2092–2101.

Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, Ribalet F, Murphy ME, Maldonado MT, Armbrust EV. 2009. Ferritin is used for iron storage in bloom-forming marine pennate diatoms. Nature **457**, 467–470.

Martinez J, Zhang G, Holt P, Jung H-T, Carrano C, Haygood M, Butler A. 2000. Self-assembling amphiphilic siderophores from marine bacteria. Science **287**, 1245–1247.

Martínez-García A, Sigman DM, Ren H, Anderson RF, Straub M, Hodell DA, Jaccard SL, Eglinton TI, Haug GH. 2014. Iron fertilization of the Subantarctic Ocean during the last ice age. Science **343**, 1347–1350.

Mawji E, Gledhill M, Milton JA, Tarran GA, Ussher S, Thompson A, Wolff GA, Worsfold PJ, Achterberg EP. 2008. Hydroxamate siderophores: occurrence and importance in the Atlantic Ocean. Environmental Science & Technology **42**, 8675–8680.

McKay RML, La Roche J, Yakunin AF, Durnford DG, Geider RJ. 1999. Accumulation of ferredoxin and flavodoxin in a marine diatom in response to Fe. Journal of Phycology **35**, 510–519.

McQuaid JB, Kustka AB, Oborník M, *et al.* 2018. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. Nature **555**, 534–537.

Merchant S, Hill K, Howe G. 1991. Dynamic interplay between two copper-titrating components in the transcriptional regulation of cyt c6. The EMBO journal **10**, 1383–1389.

Mock T, Otillar RP, Strauss J, *et al.* 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. Nature **541**, 536–540.

Moreno CM, Gong W, Cohen NR, DeLong K, Marchetti A. 2020. Interactive effects of iron and light limitation on the molecular physiology of the Southern Ocean diatom *Fragilariopsis kerguelensis*. Limnology and Oceanography **65**, 1511–1531.

Morrissey J, Sutak R, Paz-Yepes J, *et al.* 2015. A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates uptake. Current Biology **25**, 364–371.

Murgia I, Vazzola V, Tarantino D, Cellier F, Ravet K, Briat JF, Soave C. 2007. Knock-out of ferritin *AtFer1* causes earlier onset of age-dependent leaf senescence in *Arabidopsis*. Plant Physiology and Biochemistry **45**, 898–907.

Narayanan NN, Ihemere U, Chiu WT, Siritunga D, Rajamani S, Singh S, Oda S, Sayre RT. 2011. The iron assimilatory protein, FEA1, from *Chlamydomonas reinhardtii* facilitates iron-specific metal uptake in yeast and plants. Frontiers in Plant Science **2**, 67.

Nuester J, Vogt S, Twining BS. 2012. Localization of iron within centric diatoms of the genus *Thalassiosira*. Journal of Phycology **48**, 626–634.

Pankowski A, McMinn A. 2009. Iron availability regulates growth, photosynthesis, and production of ferredoxin and flavodoxin in Antarctic sea ice diatoms. Aquatic Biology **4**, 273–288.

Paz Y, Katz A, Pick U. 2007. A multicopper ferroxidase involved in iron binding to transferrins in *Dunaliella salina* plasma membranes. Journal of Biological Chemistry **282**, 8658–8666.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Peers G, Price NM. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. Nature **441**, 341–344.

Pfaffen S, Abdulqadir R, Le Brun NE, Murphy ME. 2013. Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom. Journal of Biological Chemistry **288**, 14917–14925.

Pfaffen S, Bradley JM, Abdulqadir R, Firme MR, Moore GR, Le Brun NE, Murphy ME. 2015. A diatom ferritin optimized for iron oxidation but not iron storage. Journal of Biological Chemistry **290**, 28416–28427.

Philpott CC. 2006. Iron uptake in fungi: a system for every source. Biochimica et Biophysica Acta - Molecular Cell Research **1763**, 636–645.

Poulsen N, Chesley PM, Kröger N. 2006. Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae). Journal of Phycology **42**, 1059–1065.

Quigg A, Finkel ZV, Irwin AJ, Rosenthal Y, Ho TY, Reinfelder JR, Schofield O, Morel FM, Falkowski PG. 2003. The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. Nature **425**, 291–294.

Quigg A, Irwin AJ, Finkel ZV. 2011. Evolutionary inheritance of elemental stoichiometry in phytoplankton. Proceedings of the Royal Society B: Biological Sciences **278**, 526–534.

Rastogi A, Maheswari U, Dorrell RG, *et al.* 2018. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. Scientific Reports **8**, 4834.

Raven JA. 2013. Iron acquisition and allocation in stramenopile algae. Journal of Experimental Botany **64**, 2119–2127.

Raven JA, Evans MC, Korb RE. 1999. The role of trace metals in photosynthetic electron transport in O_2-evolving organisms. Photosynthesis Research **60**, 111–150.

Renshaw JC, Robson GD, Trinci AP, Wiebe MG, Livens FR, Collison D, Taylor RJ. 2002. Fungal siderophores: structures, functions and applications. Mycological Research **106**, 1123–1142.

Rich HW, Morel FM. 1990. Availability of well-defined iron colloids to the marine diatom *Thalassiosira weissflogii*. Limnology and Oceanography **35**, 652–662.

Rossel JB, Wilson IW, Pogson BJ. 2002. Global changes in gene expression in response to high light in Arabidopsis. Plant Physiology **130**, 1109–1120.

Rubin M, Berman-Frank I, Shaked Y. 2011. Dust-and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. Nature Geoscience **4**, 529–534.

Sardet, C. 2015. Plankton : wonders of the drifting world. University of Chicago Press – Books.

Sato S, Nanjappa D, Dorrell RG, *et al.* 2020. Genome enabled phylogenetic and functional reconstruction of an araphid pennate diatom *Plagiostriata* sp. CCMP470, previously assigned as a radial centric diatom, and its bacterial commensal. Scientific Reports **10**, 9449.

Scheiber IF, Pilátová J, Malych R, *et al.* 2019. Copper and iron metabolism in *Ostreococcus tauri* – the role of phytotransferrin, plastocyanin and a chloroplast copper-transporting ATPase. Metallomics **11**, 1657–1666.

Sétif P. 2001. Ferredoxin and flavodoxin reduction by photosystem I. Biochimica et Biophysica Acta - Bioenergetics **1507**, 161–179.

Shaked Y, Kustka AB, Morel FM. 2005. A general kinetic model for iron acquisition by eukaryotic phytoplankton. Limnology and Oceanography **50**, 872–882.

Shoenfelt EM, Sun J, Winckler G, *et al.* 2017. High particulate iron (II) content in glacially sourced dusts enhances productivity of a model diatom. Science Advances **3**, e1700314.

Sims PA, Mann DG, Medlin LK. 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia **45**, 361–402.

Soria-Dengg S, Horstmann U. 1995. Ferrioxamines B and E as iron sources for the marine diatom *Phaeodactylum tricornutum*. Marine Ecology Progress Series **127**, 269–177.

Strzepek RF, Harrison PJ. 2004. Photosynthetic architecture differs in coastal and oceanic diatoms. Nature **431**, 689–692.

Strzepek RF, Hunter KA, Frew RD, Harrison PJ, Boyd PW. 2012. Iron–light interactions differ in Southern Ocean phytoplankton. Limnology and Oceanography **57**, 1182–1200.

Sunda WG, Huntsman SA. 1995. Iron uptake and growth limitation in oceanic and coastal phytoplankton. Marine Chemistry **50**, 189–206.

Sunda WG, Huntsman SA. 1997. Interrelated influence of iron, light and cell size on marine phytoplankton growth. Nature **390**, 389–392.

Tanaka T, Maeda Y, Veluchamy A, *et al.* 2015. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. The Plant Cell **27**, 162–176.

Twining BS, Baines SB, Bozard JB, Vogt S, Walker EA, Nelson DM. 2011. Metal quotas of plankton in the equatorial Pacific Ocean. Deep Sea Research Part II: Topical Studies in Oceanography **58**, 325–341.

Theil EC. 1987. Ferritin: structure, gene regulation, and cellular function in animals, plants, and microorganisms. Annual Review of Biochemistry **56**, 289–315.

Vraspir JM, Holt PD, Butler A. 2011. Identification of new members within suites of amphiphilic marine siderophores. Biometals **24**, 85–92.

Whitney LP, Lins JJ, Hughes MP, Wells ML, Chappell PD, Jenkins BD. 2011. Characterization of putative iron responsive genes as species-specific indicators of iron stress in thalassiosiroid diatoms. Frontiers in Microbiology **2**, 234.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Molecular Biology and Evolution **21**, 809–818.

Zaslavskaia LA, Lippmeier JC, Kroth PG, Grossman AR, Apt KE. 2000. Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. Journal of Phycology **36**, 379–386.

Zhang X, Krause K-H, Xenarios I, Soldati T, Boeckmann B. 2013. Evolution of the ferric reductase domain (FRD) superfamily: modularity, functional diversification, and signature motifs. PLoS One **8**, e58126.

Zhao P, Gu W, Huang A, Wu S, Liu C, Huan L, Gao S, Xie X, Wang G. 2018. Effect of iron on the growth of *Phaeodactylum tricornutum* via photosynthesis. Journal of Phycology **54**, 34–43.

Downloaded from https://academic.oup.com/jxb/article/72/6/2165/6161194 by guest on 23 April 2021

Annex II: Metabolic Innovations Underpinning the Origin and Diversification of the Diatom Chloroplast

biomolecules

MDPI

Review

Metabolic Innovations Underpinning the Origin and Diversification of the Diatom Chloroplast

Tomomi Nonoyama [1,2], Elena Kazamia [1], Hermanus Nawaly [3], Xia Gao [1], Yoshinori Tsuji [3,4], Yusuke Matsuda [3], Chris Bowler [1], Tsuyoshi Tanaka [2,*] and Richard G. Dorrell [1,*]

1 Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
2 Division of Biotechnology and Life Science, Institute of Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan
3 Department of Bioscience, Kwansei Gakuin University, Sanda, Hyogo 669-1337, Japan
4 Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan
* Correspondence: tsuyo@cc.tuat.ac.jp (T.T.); dorrell@biologie.ens.fr (R.G.D.)

Received: 29 June 2019; Accepted: 25 July 2019; Published: 30 July 2019

check for updates

Abstract: Of all the eukaryotic algal groups, diatoms make the most substantial contributions to photosynthesis in the contemporary ocean. Understanding the biological innovations that have occurred in the diatom chloroplast may provide us with explanations to the ecological success of this lineage and clues as to how best to exploit the biology of these organisms for biotechnology. In this paper, we use multi-species transcriptome datasets to compare chloroplast metabolism pathways in diatoms to other algal lineages. We identify possible diatom-specific innovations in chloroplast metabolism, including the completion of tocopherol synthesis via a chloroplast-targeted tocopherol cyclase, a complete chloroplast ornithine cycle, and chloroplast-targeted proteins involved in iron acquisition and CO_2 concentration not shared between diatoms and their closest relatives in the stramenopiles. We additionally present a detailed investigation of the chloroplast metabolism of the oil-producing diatom *Fistulifera solaris*, which is of industrial interest for biofuel production. These include modified amino acid and pyruvate hub metabolism that might enhance acetyl-coA production for chloroplast lipid biosynthesis and the presence of a chloroplast-localised squalene synthesis pathway unknown in other diatoms. Our data provides valuable insights into the biological adaptations underpinning an ecologically critical lineage, and how chloroplast metabolism can change even at a species level in extant algae.

Keywords: MMETSP; stramenopiles; biotechnology; secondary endosymbiosis; isoprenoids; plastid proteome

1. Diatoms: Powerhouses and Bellweathers in the Contemporary Ocean

Against the backdrop of rising atmospheric carbon dioxide concentrations and anthropogenic climate change, it is increasingly important to understand how photosynthesis functions in the ocean [1]. Among the diverse photosynthetic marine organisms in the contemporary ocean, diatoms are the most prominent and are estimated to be directly responsible for 40% total marine photosynthesis ([2] and references therein). In particular, diatoms dominate in polar waters and provinces that are chronically low in iron [3] and are the main photosynthetic producers in waters classified as 'high in nutrients, low in chlorophyll' (HNLC), notably large swathes of the Southern Ocean, equatorial Pacific Ocean, and north Pacific Ocean. However, they are not obligately oligotrophic species and bloom seasonally when nutrient or physical conditions change. This occurs mostly in coastal regions, following seasonal thermocline breaks, nutrient influx from the land, or aeolean fertilization.

The environmental abundance of diatoms is in contrast to other related algal groups [4,5]. This may reflect underlying physiological innovations in diatoms, which are better able to tolerate physiological stresses, such as abnormal light regimes [6,7] and carbon dioxide limitation [6], than other algal lineages. In addition, diatoms display superior abilities to capture and utilise nutrients, including nitrogen [8] and iron [9,10]. Conversely, diatom abundance is being adversely affected by global environmental changes that reduce nutrient availability in the ocean. For example, in high latitude environments in which diatoms currently flourish, increased freshwater input from melting ice caps will depress nutrient concentration, likely favouring species with smaller cell sizes than diatoms [11]. Ocean acidification may also adversely affect nutrient acquisition and photoprotection in diatoms, although this remains debated [9].

Alongside their environmental abundance, the diatom group includes various phenotypes with possible agricultural and industrial applications, many of which may be lineage- or even species-specific. Diatoms can be applied as aquafeeds for fish because they contain suitable nutrients, and particularly lipid profiles such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) [12,13]. Prior field trials with *Phaeodactylum*, *Nanofrustulum*, and *Navicula* have demonstrated the success of up to 5–10% replacement of conventional fishmeal with diatom stocks, and these may form the basis of next-generation 'circular' aquaculture techniques with increased capacity and reduced environmental impact [12,14].

Potential biotechnological applications of diatoms include the use of pigments (e.g., from *Phaeodactylum*) in cosmetics [15,16], diatom-derived polyunsaturated fatty acids as food supplements [17, 18], diatom frustules (e.g., from benthic pennate diatoms) as UV-resistant coatings for photovoltaic cells [19], and even use in next-generation biomedical technologies, e.g., the use of diatom derived ice-binding proteins (from the polar species *Navicula glaciei*) in blood cryopreservation [20]. Much of current research is focused on oleaginous diatom species, which might be of particular interest for the production of algal biodiesel; for example, the raphid pennate species *Fistulifera solaris* is a can contain up to 60% of its dry mass as lipids [21], including high amounts of the valuable polyunsaturated fatty eicosapentenoic acid [22]. Techniques for mass cultivation [21,23] and transformation are established in some diatom species, including *F. solaris* and *Phaeodactylum tricornutum* [4,21,24], opening up significant windows for synthetic engineering and practical use of these species.

Understanding diatom biology, and particularly which factors of their metabolism make them unique, may help us understand better how they will respond to environmental perturbation and offer new routes for the modification and exploitation of algal systems for industrial aquaculture. In this paper, we focus on which pathways delineate and optimise metabolic functions in the diatom chloroplast. To elucidate which of these processes have specifically contributed to the success of the diatom lineage, we reconstruct the probable metabolic functions contained in the diatom common ancestor, following its divergence from other algal lineages. We also explore specific biochemical adaptations associated with the chloroplasts of one diatom species, the biofuel producer *F. solaris*.

2. Taxonomic and Ecological Diversity of Diatoms

Diatoms arose within the stramenopile supergroup, which otherwise includes both non-photosynthetic and photosynthetic members [4,25]. Diatoms possess a silicon-rich cell wall called the frustule, which is an important trait for taxonomic classification [26,27]. Diatoms are classified into two major morphological categories (Figure 1A). The centric diatoms have radial symmetry, generally undergo anisogamous sexual reproduction, and contain multiple chloroplasts per cell [4,28] (Figure 1A). In contrast, the pennate diatoms have linear symmetry, typically have bigger frustules than centric diatoms [29], produce isomorphic gametes, and only have one chloroplast per cell (Figure 1A). Both centric and pennate diatoms are further divided into two sub-groups: the centrics into radial and polar types, based on descriptors of frustule symmetry, and the pennates into raphid and araphid types, based on the presence or absence, respectively, of a slit known as raphe at the centre of the frustule that facilitates motility on surfaces (Figure 1A).

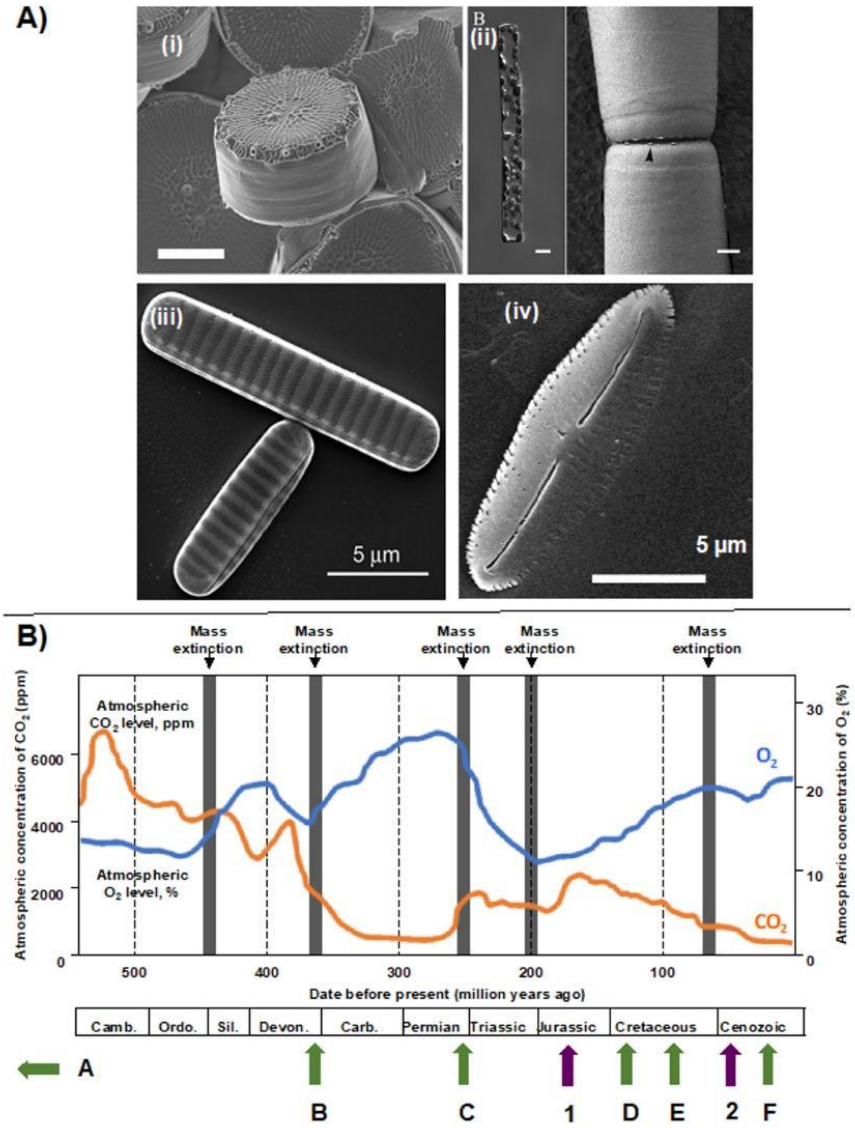


Figure 1. Diatom diversity and origins. Panel (**A**) shows scanning electron micrograph images of (**i**) *Thalassiosira pseudonana* (polar centric diatom, scale bar: 2 µm; [30]); (**ii**) *Leptocylindrus hargravesii* (radial centric diatom, scale bar: 5 µm); (**iii**) *Fragilariopsis cylindrus* (raphid pennate diatom, [31]); (**iv**) *Fistulifera solaris* JPCC DA0580 (raphid pennate diatom, authors' own image). Panel (**B**) shows a global timeline of atmospheric CO_2 (orange) and O_2 concentrations (blue), adapted from [32], with key events in diatom

evolution, as described in [33–35]. Abbreviated geological epochs are as follows: Camb., Cambrian; Ordo., Ordovician; Sil., Silurian; Dev., Devonian; Carb., Carboniferous. Green arrows show key points in diatom evolution: (**A**), median inferred radiation from molecular clock data of photosynthetic stramenopiles (>600 MYA); (**B**), median inferred date from molecular clock data for divergence of diatoms and bolidophytes (350 MYA); (**C**), earliest identifiable diatom fossils (250 MYA); (**D**), earliest identifiable pennate diatom fossils (125 MYA); (**E**), earliest identifiable polar centric diatom fossils (95 MYA); (**F**), earliest identifiable raphid pennate diatom fossils (20 MYA). Purple arrows show corresponding events in geological history that may have impacted diatom evolution and diversification: **1**, subduction of Tethyan trench (150 MYA); **2**, uplift of Himalayan plateau and opening of Drake passage and Southern circumpolar current (between 55 and 41 MYA). Figure 1Ai: Reproduced with permission from Eike Brunner, Patrick Richthammer, Hermann Ehrlich, Silvia Paasch, Paul Simon, Susanne Ueberlein, Karl-Heinz van Pée, Angewandte Chemie International Edition; published by John Wiley and Sons, 2009. Figure 1Aii: Reproduced with permission from Deepak Nanjappa, Wiebe H. C. F. Kooistra, Adriana Zingone, Journal of Phycology; published by John Wileyand Sons, 2013. Figure 1Aiii: Reproduced under the Creative Commons license from Thomas Mock, Robert P. Otillar, Jan Strauss, Mark McMullan, Pirita Paajanen, Jeremy Schmutz, Asaf Salamov, Remo Sanges, Andrew Toseland, Ben J. Ward, Andrew E. Allen, Christopher L. Dupont, Stephan Frickenhaus, Florian Maumus, Alaguraj Veluchamy, Taoyang Wu, Kerrie W. Barry, Angela Falciatore, Maria I. Ferrante, Antonio E. Fortunato, Gernot Glöckner, Ansgar Gruber, Rachel Hipkin, Michael G. Janech, Peter G. Kroth, Florian Leese, Erika A. Lindquist, Barbara R. Lyon, Joel Martin, Christoph Mayer, Micaela Parker, Hadi Quesneville, James A. Raymond, Christiane Uhlig, Ruben E. Valas, Klaus U. Valentin, Alexandra Z. Worden, E. Virginia Armbrust, Matthew D. Clark, Chris Bowler, Beverley R. Green, Vincent Moulton, Cock van Oosterhout, Igor V. Grigoriev, Nature; published by Springer Nature Publishing AG, 2017. Figure 1B: Reproduced with permission from Richard G. Dorrell, Alison G. Smith, Eukaryotic Cell; published by American Society for Microbiology, 2011.

Phylogenetic evidence places the radial centric diatoms as paraphyletic to all other lineages, the polar centric diatoms as paraphyletic to the pennate diatoms, the araphid pennates as paraphyletic to the raphid species, and the raphid pennates as monophyletic [36]. Fossil-calibrated molecular clock estimates place diatom origins in the Permian period (~320 Mya) with a shared common ancestor to another stramenopile group, the bolidophytes, at ~375 Mya [34]. Radial centric diatoms arose at least 250 million years before the present (Figure 1B), potentially placing them on a similar degree of evolutionary antiquity as flowering plants [34,37]. Polar centric and araphid pennate diatoms appear to have arisen in the Jurassic and Cretaceous (Figure 1B) [3,26], alongside falling atmospheric carbon dioxide concentrations [38]. Finally, the raphid pennate diatoms diversified in the early Oligocene (~40 Mya) (Figure 1B), alongside rising oceanic silicate concentrations, following the opening of the Drake Passage between South America and Antarctica [38].

Genome sequences have been completed for both centric (*Thalassiosira pseudonana, T. oceanica*) [39,40] and pennate diatom species (e.g., *Phaeodactylum*, *Fistulifera*) [21,24], revealing that these two distinct diatom morphogroups are more genetically distant to one another than humans are to pufferfish [24]. However, both diatom morphogroups are ecologically successful, with the most abundant genera (each comprising at least 4% of total diatom OTUs) in the *Tara* Oceans library including radial centric (*Corethron, Leptocylindrus*), polar centric (*Thalassiosira*), and pennate taxa (*Fragilariopsis*, *Pseudo-nitzschia*) [3]. Typically, centric diatoms are found predominantly in the open oceans, and pennate species are found in coastal and benthic waters [3,27]. By comparison, their immediate sister group within the stramenopiles, the bolidophytes, which do not ever form more 4% of the total photosynthetic abundance in any one *Tara* Oceans station [5].

3. Diatom Chloroplast Structure and Genomes

Diatoms and other photosynthetic members of the stramenopile algae possess a chloroplast derived from the secondary endosymbiosis of a red alga (Figure 2a). This chloroplast is closely related to the chloroplasts found in other lineages with secondary red chloroplasts (i.e., cryptomonads, haptophytes,

and alveolates). Alternative origins have also been proposed, in particular a tertiary endosymbiotic origin, in which a red alga was acquired through secondary endosymbiosis through another algal group (e.g., cryptomonads [4,25]), which were in turn engulfed by the photosynthetic stramenopile ancestor (Figure 2b). The diatom chloroplast has a distinctive structure, consisting of an annular set of unstacked thylakoids around the periphery of the stroma (« girdle lamella »), enclosing a diffuse pyrenoid, and a ring-like genophore, and contains chlorophyll a, c1, and c2, as well as fucoxanthin and hexanoylfucoxanthin as typical pigments (Figure 2c) [4,41]. All of these features are shared with the bolidophytes and hence are not diatom-specific.

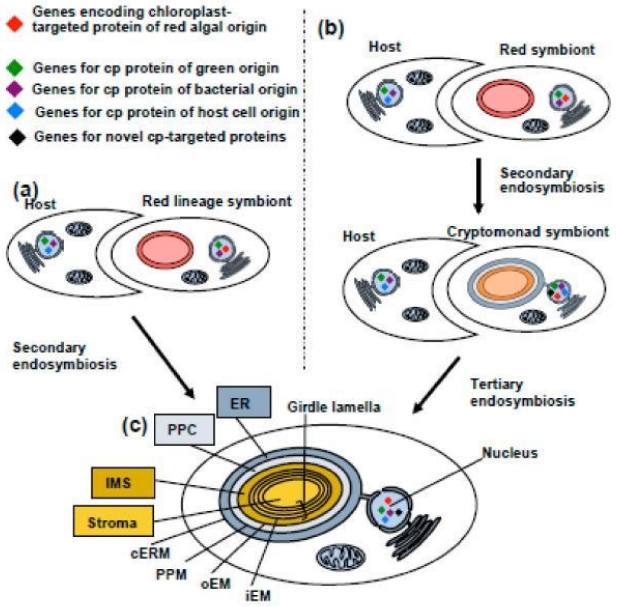


Figure 2. Origins and structure of the diatom chloroplast. This schematic figure shows two alternative hypotheses for the origins of the diatom chloroplast [4,25,42]: (**a**) secondary endosymbiosis of a red alga by a common ancestor of photosynthetic stramenopiles or (**b**) tertiary endosymbiosis of a cryptomonad-like organism, itself harbouring a chloroplast of secondary, red algal endosymbiotic origin. Either the host or the endosymbionts may have possessed genes retained from a cryptic endosymbiont of green algal origin [25,43–45], although this remains debated [44,45]. Other chloroplast-targeted proteins (indicated by the presence of coloured diamonds) may have been recruited from bacterial sources, either in the host or symbiont, proteins or paralogous copies of proteins previously targeted to other host cell organelles or may even have evolved de novo at the point of endosymbiosis [25]. (**c**) shows a schematic diagram of the four membranes surrounding the diatom chloroplast, adapted from [4]. Abbreviations are as follows: cERM; chloroplast endoplasmic reticular membrane; ER, endoplasmic reticulum; iEM, inner envelope membrane; IMS, intermembrane space; oEM, outer envelope membrane; PPC, periplastid compartment; PPM, periplastid membrane.

The diatom chloroplast retains a genome containing between 164 (*Astrosyne radiata*) and 204 genes (*Eunotia naegelii*) [4,46], with functions in photosynthesis, genome expression and protein import, cofactor (chlorophyll, thiamine, Fe-S cluster) biosynthesis, and a select number of conserved *ycf* genes of unknown function [46]. Some diatom species retain additional chloroplast-encoded functions, e.g., acetolactate synthase typically found in centric diatom chloroplast genomes [47], and a chloroplast-encoded light-independent protochorophyllide reductase so far known to be chloroplast-encoded solely in the polar centric diatom *Toxarium* [48].

Some diatoms, in particular within the genus *Nitzschia*, have lost the capacity to photosynthesize [49,50]. These species lack most chloroplast genes involved in photosynthesis and associated metabolisms, although intriguingly they retain components of a chloroplast ATP synthase. This complex has been proposed to function in the catabolism, rather than synthesis, of ATP, generating a trans-thylakoid proton gradient facilitating protein import through the *tat* translocase complex [50].

4. Import and Mosaic Origin of Diatom Nucleus-Encoded Chloroplast Proteins

As with other photosynthetic eukaryotes, the overwhelming majority of proteins that function in the diatom chloroplast are not encoded in its genome but are instead encoded in the nucleus and imported from the cytoplasm [25]. The diatom chloroplast, like those of other stramenopiles, haptophytes, and cryptomonads, is surrounded by four membranes which are, from inside out, the inner envelope membrane (iEM), the outer envelope membrane (oEM), the periplasmic membrane (PPM), and the chloroplast endoplasmic reticular membrane (cERM), which is contiguous with the endoplasmic reticulum (Figure 2c) [4].

Proteins encoded in the diatom nucleus are transported across each chloroplast membrane through individual transporter complexes, each of which recognises a different component of the chloroplast presequence. Diatom chloroplast-targeted proteins commence with an *N*-terminal signal peptide, recognised by a conventional ER import machinery in the cERM; followed by an aromatic amino acid or leucine, allowing recognition by a specialised protein import complex, « SELMA », that resides in the PPM; and finally a hydrophilic chloroplast transit peptide that allows import through the iEM and oEM [51,52].

Different proteins localise to different sub-compartments within this chloroplast, including the space between the PPM and oEM (the « perichloroplast compartment ») [53,54]; the pyrenoid and thylakoids [25,54]; and the intermembrane space between the iEM and oEM [10,25]. Other proteins may be trafficked into the chloroplast following glycosylation in the ER/Golgi body [55] or may be dual-targeted to the chloroplast and other compartments, such as the mitochondria [25,56].

A further level of complexity is contributed by the chimeric origins of the diatom chloroplast proteome (Figure 2B). Previously, we have characterised 770 nucleus-encoded and chloroplast-targeted proteins that are shared across, hence ancestral to, photosynthetic stramenopiles [4,25]. We noted that although 60% of the proteins with clearly attributable origins resolved with red algae, consistent with the endosymbiotic origin of the chloroplast, the remaining 40% originated from other sources [25]. These include a sizeable number of chloroplast-targeted proteins of green algal origin that are not found in red algae [25,57], alongside chloroplast-targeted proteins of bacterial origin, and proteins repurposed from other cellular compartments within the stramenopile host [25]. Most intriguingly, over 90 of the proteins identified in this study did not have obvious homologues (as inferred using BLAST top hit searches, with threshold evalue 1×10^{-5}) outside of algae with secondary chloroplasts [25] and thus might have specifically evolved within this lineage [10,25].

5. Metabolic and Evolutionary Complexity of the Diatom Chloroplast

In the time that diatoms have separated from related stramenopile groups, they have accumulated innovations that have allowed this group to rise to ecological prominence. This begs the question, how are diatom chloroplasts different to their nearest relatives with which they share ancestry? Here, we study stramenopile sequence libraries to infer innovations in the diatom chloroplast that may enable enhanced primary productivity, resistance to photochemical damage, and ability to effectively utilise key organic and inorganic nutrients in ocean environments [4]. We focus on four themes: iron metabolism, biosynthesis of organic metabolites, photoprotection, and CO_2 concentrating mechanisms.

5.1. The Chloroplast Proteome of the Diatom Common Ancestor

We profiled the metabolic innovations that are likely to underpin the origins of the diatom chloroplast using methodology adapted from previous studies [25,58]. This involved subsets of

chloroplast-targeted proteins from different stramenopile transcriptomes and genomes, identified via in silico prediction, for homologues of 9531 evolutionarily non-redundant chloroplast-targeted proteins identified in species with chloroplasts of secondary red algal origin via a composite BLAST pathway, based on reciprocal BLAST best-hit searches, using a floating *e*-value threshold.

We mapped the distributions identified onto previously calculated diatom and stramenopile phylogenetic trees [4,25] to identify the most probable origin points of each protein, inferred to be the last common ancestor of all species in which the protein was detected. We also identified possible loss events, defined if a chloroplast-targeted protein was not detected in any members of the clade, the clade contained multiple species and at least one genome sequence, and the protein was detected in at least two successive sister-groups to the clade. A KEGG map of two salient time points—the common ancestor of pennate and polar centric diatoms and the common ancestor of all diatoms, bolidophytes, and hypogyristea (pelagophytes and dictyochophytes)—is provided in Figure 3A. A comparison of the proteins identified through this approach, via reciprocal BLAST best-hit analysis, to analogous chloroplast protein datasets from plants, red algae, and their closest relatives [59–62] is provided in Figure 3B. Complete outputs are provided in Table S1.

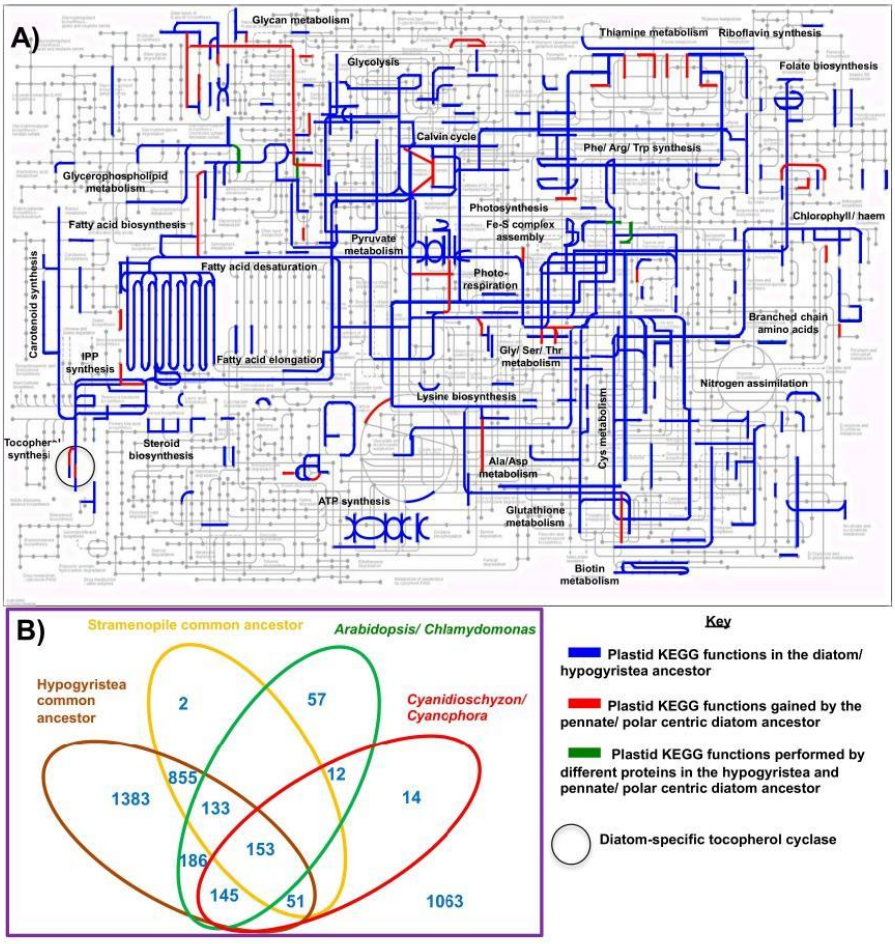


Figure 3. Innovations in metabolic complexity in the diatom chloroplast. Panel (**A**) shows KEGG

(Kyoto Encyclopedia of Genes and Genomes) pathways that could be assigned to the chloroplasts of diatoms and their close relatives, based on the evolutionary distributions of 9531 chloroplast-targeted proteins identified by Dorrell et al. [25,58]. KEGG functions were identified using KEGG mapper [63], and key chloroplast metabolism pathways are annotated following [60]. Blue lines show metabolic pathways present in the common ancestor of diatoms, bolidophytes, pelagophytes, and dictyochophytes; red lines show pathways subsequently gained by a common ancestor of the chloroplasts of pennate and polar centric diatoms; and green lines show pathways conserved between but performed by different enzymes in each lineage. Panel (**B**) compares the protein sets identified in the ochrophyte common ancestor; the common ancestor of diatoms, bolidophytes, pelagophytes, and dictyochophytes; and the common ancestor of pennate and polar centric diatoms to published experimental and phylogenomic plastid protein datasets for *Arabidopsis* [60], the green alga *Chlamydomonas* [61], the red alga *Cyanidioschyzon* [62], and the glaucophyte *Cyanophora* [59]. Protein co-occurrence is detected by reciprocal BLASTp/BLASTp search best-hit between each dataset with bidirectional threshold evalue 1×10^{-5}. Complete plastid proteome outputs for each species, and reciprocal BLAST best hit analysis between proteome datasets, are provided in Table S1.

The diatom chloroplast performs effectively all of the essential pathways identified in plant chloroplasts, including photosynthesis, central carbon and lipid metabolism, synthesis of plastidial amino acids (e.g., glutamate/ glutamine, cysteine, lysine, branched chain, and aromatic amino acids), chlorophyll and carotenoid synthesis, and essential plastid biogenesis pathways associated with expression of the chloroplast genome and protein import [25,58,61] (Figure 3A). However, it differs substantially in protein content from that of its ancestors, including (for example) substantial numbers of proteins exclusively shared with green algae and plants and not detectable in red algae and large numbers not present in primary chloroplast lineage (Figure 3B). Moreover, the diatom chloroplast is not identical to that of other stramenopiles, possessing a wider range of predicted metabolism pathways (Figure 3A), and as much as 1063 proteins not found in either the hypogyristea or stramenopile common ancestor (Figure 3B).

Below, we outline key innovations in the chloroplasts of diatoms compared to close relatives within the stramenopiles, focusing on iron and nitrogen metabolism, photoprotection, and carbon dioxide acquisition.

5.2. Iron Metabolism

Iron is essential for cellular life. Its flexible redox chemistry is at the heart of the fundamental metabolic processes of photosynthesis and nitrogen fixation, including three of the four major photosystem complexes (photosystem II, cytochrome b6/f, and in particular in photosystem I). The majority of iron in the contemporary ocean is in the insoluble form of Fe^{3+} complexed to oxyhydroxides, which is generally not as readily bioavailable as the more soluble Fe^{2+} or organically bound iron. The ecological success of diatoms is frequently linked to their unique and highly adapted iron physiologies, which match the chemical speciation of iron in the water column [4,36].

Diatoms may be able to better tolerate chronic iron deprivation by replacing iron-dependent proteins in chloroplast photosystems with iron-free alternatives (Figure 4A). These can include plastocyanin (which contains copper) instead of cytochrome c6 [64,65] and flavodoxin as an iron-free alternative to the photosystem I acceptor ferredoxin [66]. While plastocyanin genes are widespread across diatom species and are upregulated under Fe-limiting conditions [67,68], very few of the identified diatom plastocyanin sequences contain recognisable chloroplast-targeting peptides, and it remains to be confirmed where in the cell this protein functions [64]. Diatom flavodoxin genes, by contrast, are universally distributed, and many have chloroplast-targeting sequences. However, chloroplast-targeted flavodoxin protein at least arose in the common ancestor of diatoms, pelagophytes, and dictyochophytes (Figure 4A), rather than being an adaptation specifically confined to diatoms.

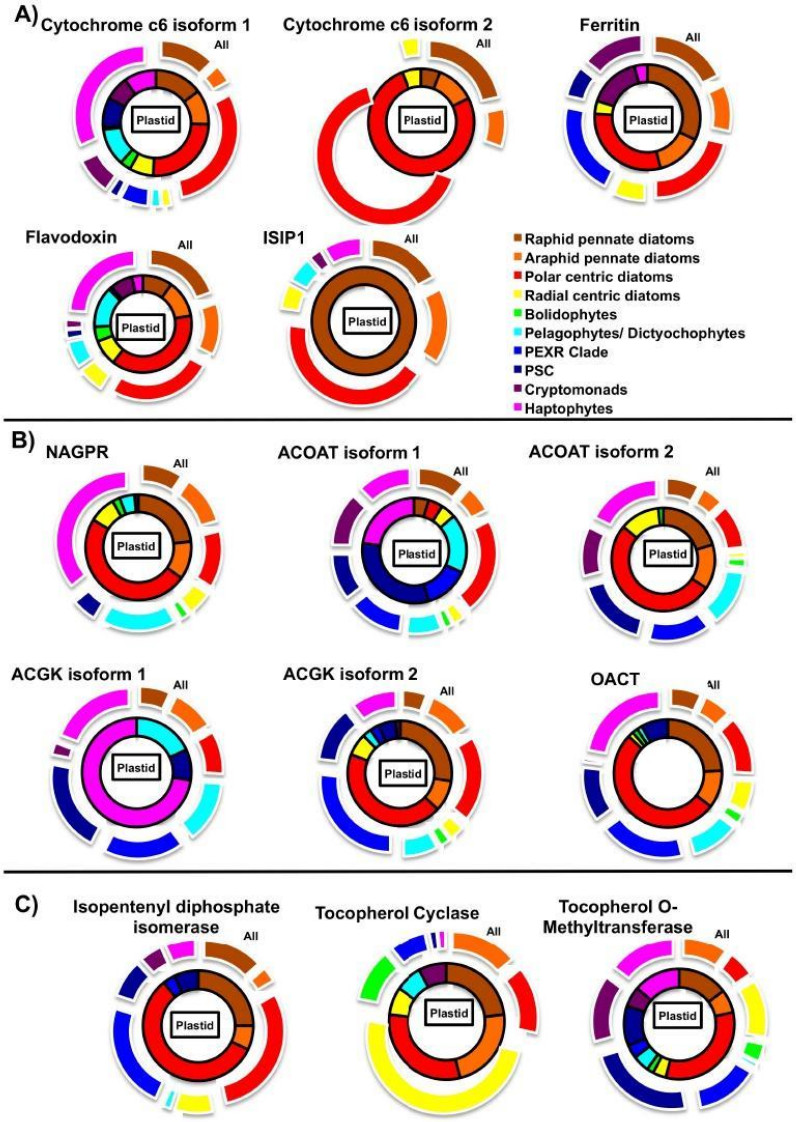


Figure 4. Evolutionary distributions of chloroplast proteins restricted to diatoms and their close relatives. This figure shows the evolutionary distributions of chloroplast-targeted proteins involved in (**A**) iron metabolism, (**B**) ornithine metabolism, and (**C**) organic compound synthesis, inferred from the pipeline used in Figure 3A, across photosynthetic stramenopiles, haptophytes, and cryptomonads. Lineages are arranged into monophyletic taxonomic groups, following published phylogenies [36,69]. The inner wheel of each diagram (bounded with black lines) shows the total numbers of species within a given lineage for which chloroplast-targeted proteins were identified, and the outer wheel of each diagram shows the number of species within a given lineage for any protein homologue (including those with non-chloroplast or ambiguous targeting signals) could be identified. Exemplar GFP localisations for enzymes associated with the probable diatom chloroplast ornithine cycle are shown in Figure S1. Presence and absence distributions of each protein, listed by species, are provided in Table S1.

Diatoms may additionally be able to tolerate iron limitation by efficiently mobilising iron, when it becomes available in the water column. This is evidenced during periods of transient enrichment of Fe, e.g., experimentally during artificial Fe fertilization in situ, which results in diatom-dominated blooms [4]. This effective iron acquisition is partly due to a novel pathway, allowing diatom chloroplasts to utilise iron complexed to siderophore molecules. Siderophores are iron-chelators, are typically biosynthesized by bacteria and fungi, comprise 99% of the organic iron pool, and were previously not considered accessible to photosynthetic eukaryotes [70]. Siderophore-bound iron is taken up from the diatom cell surface via an endocytosis pathway mediated by ISIP1, an 'iron-starvation induced protein' [10,71]. The ISIP1-siderophore complex is then trafficked toward the chloroplast, and possibly into the PPC, where reduction occurs, releasing iron from its organic chelator [10,72]. Notably, ISIP1 does not have conventional chloroplast targeting peptide, and the exact mechanism by which it is transported from the plasma membrane to this compartment remains unknown (Figure 4A) [10]. However, ISIP1 is only found in diatoms, with a distantly divergent homologue found in some pelagophyte, haptophyte, and dinoflagellate libraries, suggesting that it may be a diatom-specific strategy to facilitate iron delivery to the chloroplast (Figure 4) [10].

Finally, diatom chloroplasts may possess large internal reserves of iron, allowing them to maintain a large reserve of iron resources and correct for temporal differences in iron availability. For example, the iron storage protein ferritin [10] is predicted to be chloroplast-targeted across a wide range of diatoms, and is absent from other stramenopiles, although chloroplast-targeted equivalents are detectable in haptophytes and cryptomonads (Figure 4A). What is unclear is the cellular function of ferritin in cells, as a storage protein (as evidenced for *Pseudo-nitzschia* [73]) or as a homeostatic buffer. We should stress it is unlikely to be the only storage protein or the only mechanism for homeostasis, as other whole-cell mechanisms for iron homeostasis have been proposed [72].

5.3. Nitrogen Metabolism

Recently, it was proposed that diatoms possess a chloroplast-targeted glutamine-ornithine cycle, allowing them to recycle chloroplastidial glutamine to ornithine, which could then be exchanged with the cytoplasm [4]. This is particularly intriguing as it complements the finding that diatoms possess a mitochondria-targeted urea cycle, which would conversely allow them to synthesize glutamine in the mitochondria from ornithine [74]. Working in parallel, the organelles can sensitively equilibrate the availabilities of free amino acids, achieving a level of cellular integration not observed in other species [8,75]. Crucially, enzymes in this pathway are upregulated under nitrogen limitation and are coregulated with chloroplast-targeted nitrate reductase, suggesting roles in mediating nitrate assimilation [8,74,76].

We have localised all four of the involved enzymes in this cycle to the chloroplast of *P. tricornutum*, either through the localisation of full-length GFP constructs (acetylornithine transaminase, acetylglutamate kinase, glutamate N-acetyltransferase) or N-terminal GFP constructs (N-acetyl-δ-glutamyl-phosphate reductase; Figure S1A), following previous methodology [25]. Chloroplast-targeted copies of all four enzymes can be detected in other stramenopiles, cryptomonads, and haptophytes, although the complete cycle of four enzymes appears to only be chloroplast-targeted in diatoms (Figure 4B). Understanding the exact function of this pathway will depend on the characterisation of mutant lines for each protein.

5.4. Photoprotection

Diatoms possess a number of photo-physiological strategies for tolerating excess light that are shared with green algae, but not with the red antecedents of their chloroplasts [76]. These include an inferred xanthophyll cycle, which allows diatoms to effect non-photochemical quenching through the conversion of the accessory pigment diadinoxanthin into diatoxanthin and an expanded set of LHCx/Li818 family light harvesting proteins, which are implicated in the tolerance of light stress, not least by inducing a functional xanthophyll cycle [77]. The exact evolutionary origin of these

pathways, including whether they can be truly considered as absent from red algae, has been considered elsewhere [78]. However, the effector enzymes of the xanthophyll cycle (violaxanthin de-epoxidase and zeaxanthin epoxidase), and LHCx/Li818 family proteins, are broadly conserved across algae with secondary red algal chloroplasts and are therefore not a diatom-specific innovation [57,79].

Recently, an additional photoprotective strategy has been evidenced in diatoms, in which excess chloroplast reducing potential is translocated into the mitochondria and dissipated through respiratory electron transport [25,79]. Excess mitochondrial ATP might conversely be imported into the chloroplast and used to activate ATP-dependent metabolic strategies (such as the Calvin Cycle) in the absence of chloroplast ATP, e.g., following the dark-light transition of a seasonal or circadian light cycle [7,80]. This process has been documented in other lineages (e.g., the green model alga *Chlamydomonas*), but only in cases where alternative photochemical quenching strategies have diminished function [7,75]. Thus, diatoms depend on this metabolite exchange much more substantially than other lineages of algae. Ultimately, identifying which effector proteins are involved in this process may provide clues into whether it is a diatom-specific or more widespread chloroplast evolutionary adaptation.

Finally, we have uncovered evidence for a diatom-specific chloroplast antioxidant strategy. We note that, in our KEGG map of chloroplast functions, the last common ancestor of the pennate and polar centric diatom lineages (Figure 3A, circled) is predicted to possess a complete chloroplast-targeted tocopherol (vitamin E) biosynthesis pathway. In contrast, other stramenopile algae lack a chloroplast tocopherol cyclase, with one possible exception in dictyochophytes (Figure 4C; Table S1). Tocopherol is a known antioxidant, having overlapping functions to xanthophylls in scavenging of singlet oxygen and preventing the propagation of lipid peroxidation in plant chloroplasts [7,81]. Understanding the physiological significance of this pathway will depend first on identifying whether tocopherol is indeed present in diatom chloroplast preparations [82], and phenotyping mutants of the tocopherol cyclase, and/or its downstream methyltransferase (Figure 4C).

5.5. CO_2 Concentrating Mechanisms (CCM)

Photosynthetic efficiency is a likely candidate trait for determining ecological success in marine environments. Minor adjustments that increase the proportion of fixed carbon could quickly translate to enhanced growth and standing biomass. In particular, carbon dioxide has limited solubility in water; this may therefore limit the carbon catalytic activity, and increase the relative proportion of wasteful oxygenase activity of RuBisCO in aquatic photosynthesis [32,83]. As a result, many photosynthetic algae utilise CO2 concentrating mechanisms that enrich carbon dioxide availability to RuBisCO.

To date, it has been demonstrated that diatoms, including *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, have a biophysical CCM (using carbonic anhydrases and bicarbonate transporters) [54,83,84]. Some diatoms may additionally utilise biochemical strategies such as C4 photosynthesis [85], although an evident C4 operation has so far been detected solely in one centric species, *T. weissflogii* [83,85]. Previously, we and others have shown that diatoms possess a diverse set of carbonic anhydrases, at least some of which localise to different chloroplast sub-compartments [86], and which may confer an advantage in dynamic open ocean environments. These enzymes include an unusual, theta-class carbonic anhydrase, which in *P. tricornutum* localises within the lumen of pyrenoid-penetrating thlaykoid membranes and is directly implicated in the release of CO2 into chloroplast RuBisCO [54].

To gain a sense of how these proteins evolved, we profiled the distribution of all experimentally characterised carbonic anhydrases from *P. tricornutum* and *T. pseudonana* across stramenopile, haptophyte and cryptomonad genomes and transcriptomes, as above (Table S2; Figure 5A) [54,86]. Through this approach, we identified 434 chloroplast-targeted theta-carbonic anhydrases in 48 of the 86 species studied, including the early-diverging radial centric genus *Leptocylindrus* (Figure S2). This implicates chloroplast theta-carbonic anhydrases as being present in the diatom common ancestor. We observed much more restricted evolutionary distributions for chloroplast-targeted isoforms of other diatom chloroplast carbonic anhydrases, including the *Phaeodactylum* pyrenoidal

beta-carbonic anhydrases PtCA1 and PtCA2 (Phatr3_45443, 51305), and the *Thalassiosira* stromal alpha carbonic anhydrase TpCA1 [54,86–88] (Figure S2). Although we could identify chloroplast-targeted theta-carbonic anhydrases in other stramenopile, cryptomonad, and haptophyte lineages, we noted that diatom theta-carbonic anhydrases contained a much larger proportion of chloroplast-targeted proteins than orthologues identified in other lineages, suggesting this predominance is a specific innovation within the diatom chloroplast (Figure 5A).

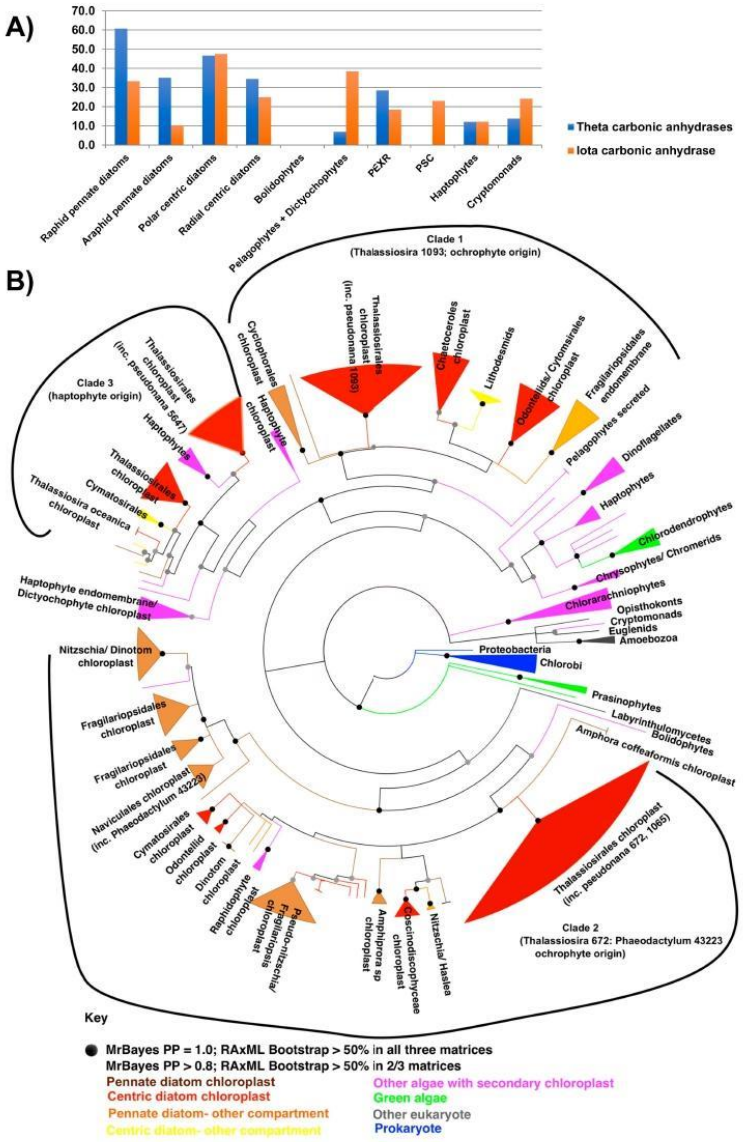


Figure 5. Evolution of chloroplast-targeted diatom carbonic anhydrases. Panel (**A**) shows the number

(labels) and proportion of chloroplast-targeted sequences (bars) identified for theta and iota carbonic anhydrases in diatoms and other stramenopile, cryptomonad and haptophyte lineages, by reciprocal BLAST best hit analysis. Detailed outputs are provided for the distribution of theta, alpha, and beta carbonic anhydrases and iota carbonic anhydrases, across diatom species in Figures S2 and S3, respectively. Panel (**B**) shows the consensus Bayesian topology of a 204 taxa × 210 aa alignment of theta-carbonic anhydrase sequences across the tree of life. Three clades of diatom chloroplast-targeted proteins are shown with circular brackets. Sequence names are shaded by taxonomic origin and are labelled with predicted localisations (chloroplast, endomembrane system, and/or mitochondria). Alignment and tree topologies under each programme and matrix considered are shown in Table S2.

A consensus Bayesian and RAxML tree of theta carbonic anhydrases revealed three distinct groups of diatom chloroplast proteins (Figure 5B). Two of the clades (clades 1 and 2) possess immediate sister-groups within the stramenopiles (respectively, pelagophytes and bolidophytes), suggesting probable vertical inheritances, although the sister-group sequences appear not to be chloroplast-targeted (Figure 5). The final clade of diatom chloroplast theta carbonic anhydrases (clade 3) resolves within a paraphyletic clade of haptophyte algae, suggesting a probable horizontal transfer event from haptophytes into the diatom common ancestor. At least some of the haptophyte sequences have chloroplast-targeting sequences, so it is possible that this anhydrase already possessed chloroplast functions prior to its acquisition by the diatom ancestor (Figure 5A,B). Overall, our data is consistent with theta carbonic anhydrases being recruited through horizontal gene transfer and the recycling of enzymes from other cellular compartments in the diatom ancestor.

Finally, using similar methodology, we considered the distribution and inferred subcellular localisation of the new, iota-class of carbonic anhydrases, recently reported to function in the *T. pseudonana* chloroplast (Figure 5A and Figure S3) [89]. We noted that chloroplast-targeted copies of this enzyme are widespread across diatoms and are present in extremely high numbers in certain species, including 10 or more in *F. solaris*, *Odontella aurita*, and *Ditylum brightwellii* (Figure S3). We could not detect chloroplast-targeted copies of this enzyme in the genus *Proboscia*, and a subclade of *Chaetoceros* (sp., UNC1202, *curvisetus*, *brevis*) from which chloroplast-targeted theta-carbonic anhydrases were also globally absent (Figures S2 and S3). In contrast to the situation for theta-carbonic anhydrases, there was no apparent bias in favour of chloroplast-targeted proteins in diatoms compared to other stramenopile, haptophyte or cryptomonad groups, with a greater proportion of pelagophyte and dictyochophyte sequences possessing chloroplast-targeting peptides (Figure 5A; Table S2), suggesting it is an evolutionary innovation that precedes the origins of the diatom chloroplast.

6. Industrial Futures: Remodelling Chloroplast Metabolism in the Oil-Producing Diatom *Fistulifera solaris*

Diatoms show diverse phenotypes, and some species produce valuable products for industrial application. For instance, the raphid pennate diatom *F. solaris* accumulates large quantities of lipids, which is of interest for conversion to biodiesel (Figure 6A) [21]. Since photosynthetic eukaryotes including diatoms typically synthesize glycerolipids in the chloroplast and ER [21], *F. solaris* may have unique alterations to its chloroplast metabolism to facilitate lipid production.

6.1. Remodelling of Lipid and Amino Acid Metabolism in the F. solaris Chloroplast

Using the methodology described above, we compared the complement of nucleus-encoded and chloroplast-targeted proteins inferred from the *F. solaris* genome to those of other diatom species. We noted distinct differences between *F. solaris* and other diatoms. These include the probable absence of a chloroplast-targeted pyruvate carboxylase (PC), pyruvate phosphate dikinase (PPDK), and phosphoenolpyruvate synthase (PEPS), which catalyze the conversion of pyruvate to phosphoenolpyruvate (Figure 6(Bi,Ci)) (Table S3) and are otherwise widespread across diatoms.

Conversely, only *F. solaris* among diatoms has a chloroplast-targeted pyruvate/2-oxoglutarate dehydrogenase complex (PDC), which generates acetyl-CoA, a substrate for fatty acid synthesis

(Figure 6(Bi,Ci)) (Table S3). We confirmed that the *F. solaris* PDC protein transit peptide cleavage site is upstream of the conserved PDC domain (Figure S4A) and is therefore unlikely to be an annotation artefact or internal region of a non-chloroplast-targeted protein. The chloroplast-targeted PDC might participate in supplying acetyl CoA for chloroplast fatty acid synthesis, as has previously been proposed to occur in higher plants and *C. reinhardtii* [90], contributing to its oleaginous phenotype.

We found a possible remodeling of branched-chain amino acid (BCAA) synthesis pathways, which convert 2-oxybutyrate and pyruvate into isoleucine, valine, and leucine, in *F. solaris*. Unlike *P. tricornutum* and *T. pseudonana*, no chloroplast-targeted homologs of acetohydroxyacid isomeroreductase (AHAIR) and branched-chain amino acid transferase (BCAT) were identified in *F. solaris* (Figure 6(Bii,6Cii)) (Table S3). We could identify equivalents of these enzymes functioning in the mitochondria (BACT) and the cytoplasm (AHAIR), suggesting that these essential amino acids are synthesized elsewhere in the *F. solaris* cell.

The absence of chloroplast-targeted copies of each enzyme might relate to the oleaginous phenotype of *F. solaris*, as both are related to NADPH metabolism. AHAIR directly utilises NADPH as a reducing agent, and BCAT requires NADPH to regenerate its substrate glutamate from 2-oxoglutarate by glutamate synthase [91]. The absence of these pathways from the *F. solaris* chloroplast might indirectly increase NADPH availability for alternative pathways, such as chloroplast fatty acid synthesis. It has previously been shown that both the downregulation of plastidial amino acid biosynthesis and overexpression of NADPH-producing enzymes in diatoms accelerates neutral lipid accumulation in diatoms [92,93]. Ultimately, the absence of these enzymes, and the phenotypic consequences for the *F. solaris* chloroplast, awaits experimental verification.

6.2. Modified Chloroplast Isoprenoid Synthesis in Specific Diatoms

Unlike other diatoms, *F. solaris* encodes a chloroplast-targeted squalene synthase in its nuclear genome (Figure 6(Biii,Ciii)) (Table S3). Squalene is synthesized from farnesyl and geranylgeranyl diphosphate, which are converted by squalene synthase (SQS) and farnesyl diphosphate synthase (FPPS), respectively. As per the situation in most other diatoms [92], SQS in *F. solaris* is fused with isopentenyl diphosphate isomerase (IDI) (Figure 6Ciii). We confirmed that the *F. solaris* FPPS target sequence was positioned upstream of the conserved domain, indicating that it is genuinely chloroplast-targeted (Figure S4C). We localized SQS in *F. solaris* using a full-length intron-containing C-terminal fusion GFP overexpression construct, and it localized in the vicinity of the chloroplast (Figure 6D). This localization pattern is similar to PPM proteins in *P. tricornutum* [94].

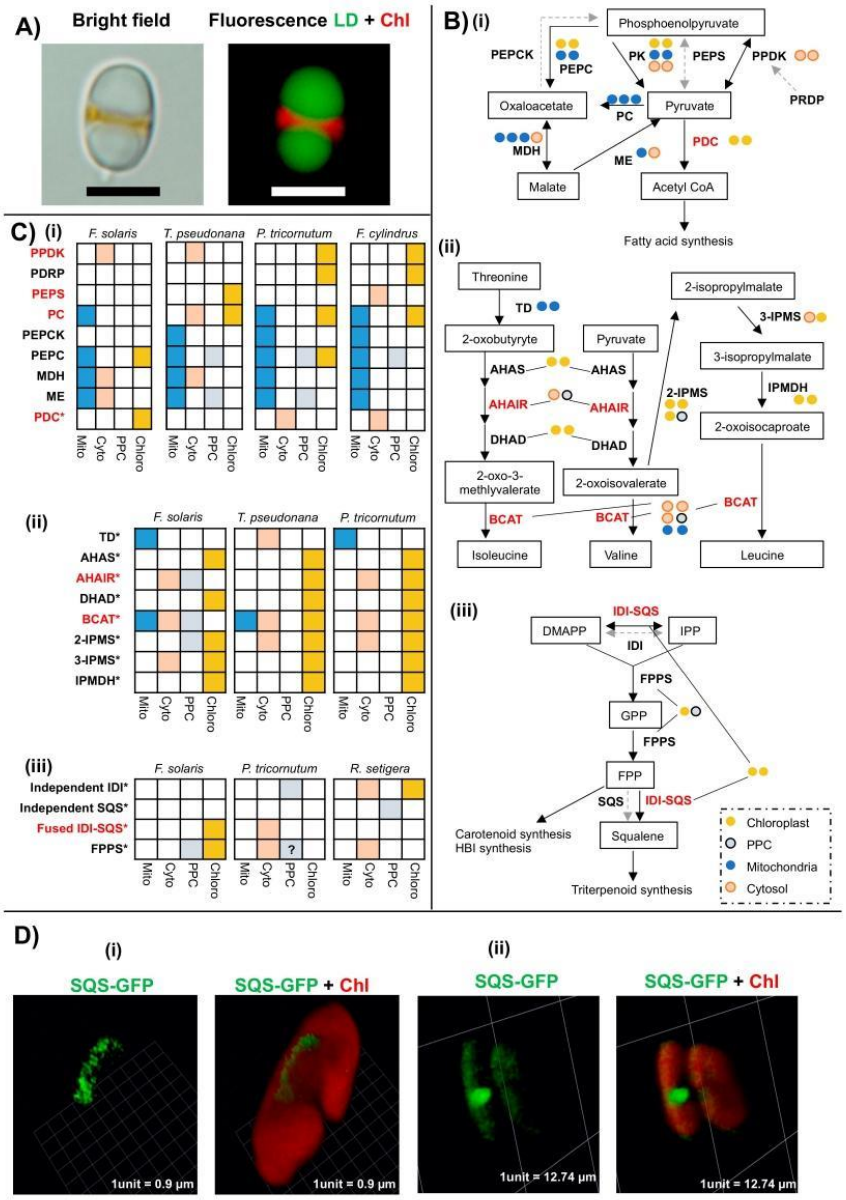


Figure 6. Innovations in the *Fistulifera solaris* chloroplast. Panel (**A**) shows a *Fistulifera solaris* cell visualized under bright field and fluorescence microscopy. The conspicuous lipid droplets (LD) are visualized using BODIPY (boron-dipyrromethene) 505/515 staining (in green) and chlorophyll fluorescence (Chl) is shown in red; scale bar = 5 µm. Panel (**B**) shows schematic pathway diagrams for

(**i**) pyruvate hub metabolism; (**ii**) branched chain amino acid synthesis; and (**iii**) isoprenoid metabolism as inferred from the *F. solaris* genome [21]. Each circle equates to one distinct copy of each gene. Genes labelled in red are those for which *F. solaris* has different metabolic arrangements to other diatom species considered. Panel (**C**) shows detailed localisation predictions for homologues of key enzymes implicated in modified chloroplast metabolism in *F. solaris* [21], alongside equivalent data for (**i**) *T. pseudonana, P. tricornutum, and F. cylindrus* [94], (**ii**) *T. pseudonana* and *P. tricornutum* [91], and (**iii**) *P. tricornutum* and *R. setigera* [24,95]. The symbol '?' depicts a sequence without a cleavable signal peptide but predicted as a transmembrane protein. Sequence IDs shown in this figure are provided in Table S3. Panel (**D**) shows a three-dimensional fluorescence microscope image of a line of *F. solaris* expressing a C-terminal GFP full length fusion construct of native squalene synthase protein, following previous methodology [21,25,96]. Cells visualised show two localisation patterns, both previously indicated to the periplastid compartment [97]—either (**i**) elongated localisation along the chloroplast periphery or (**ii**) across the mid-band of dividing chloroplasts. A diagram of the targeting sequence of this protein is provided in Figure S4.

Elsewhere within our dataset, we identified chloroplast-targeted isoprenoid biosynthesis enzymes in *Pseudo-nitzschia* sp., *Synedra acus*, and *Rhizosolenia setigera* (Figure 4, Figure 6Ciii); chloroplast FPP synthesis has also been predicted in *Haslea ostrearia*. By contrast, plant FPPS are found to localise in the cytosol, mitochondria, peroxisome, and possibly the chloroplast [98], while SQS localizes to the ER membrane in *Arabidopsis thaliana* [99]. Squalene production has been documented in other microalgae (e.g., the green alga *Botryococcus braunii* [100]), although the localisations of the enzymes involved remain unknown. Notably, the enzymes contributing to these metabolic fluxes vary between each species, with only *F. solaris* possessing a chloroplast-targeted IDI-SQS and FPPS, compared to, e.g., a chloroplast-targeted IDI in *R. setigera*. Thus, a complex chloroplast isoprenoid metabolism may have evolved in parallel in different diatoms.

The exact function of these enzymes in diatom chloroplasts remains to be determined. The enzymatic activity of isopentenyl diphosphate isomerase can be compensated by another enzyme, hydroxymethylbillance-CoA reductase, which is chloroplast-targeted across the stramenopiles, so it is therefore not essential for carotenoid synthesis [25,100]. However, the presence of chloroplast-targeted IDI-SQS and FPPS might facilitate the production of secondary metabolites that would allow diatoms to survive diverse biotic and abiotic pressures, similar to the functions of polyprenoid biosynthesis in plants [77] Isopentenyl diphosphate synthase activity in *R. setigera* has been proposed to be involved in the synthesis of highly branched isoprenoids, which may function in cryoprotection 92, and may include the diatom-specific isoprenoid IP25, which can be used as a biomarker for primary production in sea ice [92,101]. The adoption of different polyprenoid biosynthesis pathways in different diatoms might therefore confer different functions to these species.

7. Concluding Remarks

In this review, we discussed the mosaic origins, complex structure, and metabolic diversity of the diatom chloroplast. Focusing on innovations that might help explain the success of diatoms in the contemporary ocean, we have noted that diatoms possess unique chloroplast features, specifically within iron metabolism, biosynthesis of organic metabolites, photoprotection, and CO2 concentrating mechanisms. However, there are many other chloroplast proteins that may be unique to diatoms, with limited structural homology to other annotated proteins [25]. Characterisation of mutant lines of these proteins remains imperative to understanding their function. Environmental sampling projects (e.g., the *Tara* Oceans Survey [3]) may be instrumental in allowing us to deduce transcriptional patterns and the environmental relevance of understudied diatom proteins in current and future climates.

Comparing the chloroplast-targeted proteins of species with industrially desirable phenotypes, such as *F. solaris*, may reveal candidate pathways for engineering in other organisms. For example, it will be interesting to determine if suppressing chloroplast pyruvate or BCAA metabolism in *P. tricornutum* can enhance lipid production, which would support a link between these proteins and the

oleaginous phenotype of *F. solaris*. Similarly, heterologous expression of chloroplast-targeted proteins aimed at modifying isoprenoid metabolism in *F. solaris* and other diatoms may reveal the function of these enzymes and might also allow us to generate new and industrially relevant metabolites in easily cultivable species.

Overall, the unique chloroplast metabolic pathways identified within multi-species transcriptome datasets help us understand more clearly the features underpinning diatom biology.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/9/8/322/s1, Figure S1: Experimental localisations of diatom ornithine cycle enzymes, Figure S2: Evolutionary distribution and localisation of diatom theta-carbonic anhydrases, Figure S3: Evolutionary distributions of diatom iota-carbonic anhydrases, Figure S4: Chloroplast targeting sequence of Fistulifera solaris chloroplast proteins, Table S1: Overview of plastid protein content across diatom and related algal lineages, Table S2: Evolutionary and subcellular distributions of diatom carbonic anhydrases, Table S3: Tabulated sequences of Fistulifera and other diatom chloroplast pyruvate hub, branched chain amino-acid, and isoprenoid-associated synthesis enzymes.

Funding: This research was funded by a CNRS Momentum Fellowship to RGD, and a JSPS PhD studentship to TN. The APC was funded by Tokyo University of Agriculture and Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hinder, S.L.; Hays, G.C.; Edwards, M.; Roberts, E.C.; Waine, A.W.; Gravenor, M.B. Changes in marine dinoflagellate and diatom abundance under climate change. *Nat. Clim. Chang.* **2012**, *2*, 271–275. [CrossRef]
2. Field, C.B.; Behrenfeld, M.J.; Randerson, J.T.; Falkowski, P.A. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **1998**, *281*, 4. [CrossRef] [PubMed]
3. Malviya, S.; Scalco, E.; Audic, S.; Vincent, F.; Veluchamy, A.; Poulain, J.; Wincker, P.; Iudicone, D.; de Vargas, C.; Bittner, L.; et al. Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E1516–E1525. [CrossRef] [PubMed]
4. Dorrell, R.G.; Bowler, C. Secondary Plastids of Stramenopiles. In *Adv Bot Res: Secondary Endosymbiosis*; Hirakawa, Y., Ed.; Elsevier: Amsterdam, The Netherlands, 2017; Volume 84, pp. 59–104.
5. Ichinomiya, M.; Dos Santos, A.L.; Gourvil, P.; Yoshikawa, S.; Kamiya, M.; Ohki, K.; Audic, S.; de Vargas, C.; Noël, M.H.; Vaulot, D.; et al. Diversity and oceanic distribution of the Parmales (Bolidophyceae), a picoplanktonic group closely related to diatoms. *ISME J.* **2016**, *10*, 2419–2434. [CrossRef] [PubMed]
6. Lepetit, B.; Sturm, S.; Rogato, A.; Gruber, A.; Sachse, M.; Falciatore, A.; Kroth, P.G.; Lavaud, J. High light acclimation in the secondary plastids containing diatom *Phaeodactylum tricornutum* is triggered by the redox state of the plastoquinone pool. *Plant Physiol.* **2013**, *161*, 853–865. [CrossRef] [PubMed]
7. Bailleul, B.; Berne, N.; Murik, O.; Petroutsos, D.; Prihoda, J.; Tanaka, A.; Villanova, V.; Bligny, R.; Flori, S.; Falconet, D.; et al. Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* **2015**, *524*, 366. [CrossRef] [PubMed]
8. Levering, J.; Broddrick, J.; Dupont, C.L.; Peers, G.; Beeri, K.; Mayers, J.; Gallina, A.A.; Allen, A.E.; Palsson, B.O.; Zengler, K. Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS ONE* **2016**, *11*, e0155038. [CrossRef]
9. McQuaid, J.B.; Kustka, A.B.; Oborník, M.; Horák, A.; McCrow, J.P.; Karas, B.J.; Zheng, H.; Kindeberg, T.; Andersson, A.J.; Barbeau, K.A.; et al. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature* **2018**, *555*, 534–537. [CrossRef]
10. Kazamia, E.; Sutak, R.; Paz-Yepes, J.; Dorrell, R.G.; Vieiria, F.R.J.; Mach, J.; Morrissey, J.; Leon, S.; Lam, F.; Pelletier, E.; et al. Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Sci. Adv.* **2018**, *4*, 4536.
11. Li, W.K.; McLaughlin, F.A.; Lovejoy, C.; Carmack, E.C. Smallest algae thrive as the Arctic Ocean freshens. *Science* **2009**, *326*, 539. [CrossRef]
12. Yarnold, J.; Karan, H.; Oey, M.; Hankamer, B. Microalgal aquafeeds as part of a circular bioeconomy. *Trends Plant Sci.* **2019**, in press. [CrossRef] [PubMed]
13. Sprague, M.; Betancor, M.B.; Tocher, D.R. Microbial and genetically engineered oils as replacements for fish oil in aquaculture feeds. *Biotechnol. Lett.* **2017**, *39*, 1599–1609. [CrossRef]

14. Sørensen, M.; Gong, Y.; Bjarnason, F.; Vasanth, G.K.; Dahle, D.; Huntley, M.; Kiron, V. *Nannochloropsis oceania*-derived defatted meal as an alternative to fishmeal in Atlantic salmon feeds. *PLoS ONE* **2017**, *12*, e0179907. [CrossRef] [PubMed]
15. Abida, H.; Ruchaud, S.; Rios, L.; Humeau, A.; Probert, I.; De Vargas, C.; Bach, S.; Bowler, C. Bioprospecting marine plankton. *Mar. Drugs* **2013**, *11*, 4594–4611. [CrossRef]
16. Hamilton, M.L.; Warwick, J.; Terry, A.; Allen, M.J.; Napier, J.A.; Sayanova, O. Towards the industrial production of omega-3 long chain polyunsaturated fatty acids from a genetically modified diatom *Phaeodactylum tricornutum*. *PLoS ONE* **2015**, *10*, e0144054. [CrossRef]
17. Kim, S.M.; Jung, Y.J.; Kwon, O.N.; Cha, K.H.; Um, B.H.; Chung, D.; Pan, C.H. A potential commercial source of fucoxanthin extracted from the microalga *Phaeodactylum tricornutum*. *Appl. Biochem. Biotechnol.* **2012**, *166*, 1843–1855. [CrossRef]
18. Kuczynska, P.; Jemiola-Rzeminska, M.; Strzalka, K. Photosynthetic pigments in diatoms. *Mar. Drugs* **2015**, *13*, 5847–5881. [CrossRef] [PubMed]
19. Allert, S.; Wulff, A. Frustules Extracted from Benthic Pennate Diatoms Harvested from an Industrial Biofilm Process. U.S. Patent US20190106672A1, 18 April 2019.
20. Kang, J.S.; Raymond, J.A. Reduction of freeze-thaw-induced hemolysis of red blood cells by an algal ice-binding protein. *Cryoletters* **2004**, *25*, 307–310.
21. Tanaka, T.; Maeda, Y.; Veluchamy, A.; Tanaka, M.; Abida, H.; Marechal, E.; Bowler, C.; Muto, M.; Sunaga, Y.; Yoshino, T.; et al. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *Plant Cell* **2015**, *27*, 162–176. [CrossRef]
22. Tanaka, T.; Yabuuchi, T.; Maeda, Y.; Nojima, D.; Matsumoto, M.; Yoshino, T. Production of eicosapentaenoic acid by high cell density cultivation of the marine oleaginous diatom *Fistulifera solaris*. *Bioresour. Technol.* **2017**, *245 (Pt A)*, 567–572. [CrossRef]
23. Matsumoto, M.; Nojima, D.; Nonoyama, T.; Ikeda, K.; Maeda, Y.; Yoshino, T.; Tanaka, T. Outdoor cultivation of marine diatoms for year-round production of biofuels. *Mar. Drugs* **2017**, *15*, 94. [CrossRef] [PubMed]
24. Rastogi, A.; Maheswari, U.; Dorrell, R.G.; Vieira, F.R.J.; Maumus, F.; Kustka, A.; McCarthy, J.; Allen, A.E.; Kersey, P.; Bowler, C.; et al. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* **2018**, *8*, 4834. [CrossRef]
25. Dorrell, R.G.; Gile, G.; McCallum, G.; Méheust, R.; Bapteste, E.P.; Klinger, C.M.; Brillet-Guéguen, L.; Freeman, K.D.; Richter, D.J.; Bowler, C. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* **2017**, *6*, 23717. [CrossRef] [PubMed]
26. McNair, H.M.; Brzezinski, M.A.; Krause, J.W. Diatom populations in an upwelling environment decrease silica content to avoid growth limitation. *Environ. Microbiol.* **2018**, *20*, 4184–4193. [CrossRef] [PubMed]
27. De Tommasi, E.; Gielis, J.; Rogato, A. Diatom Frustule Morphogenesis and Function: A Multidisciplinary Survey. *Mar. Genom.* **2017**, *35*, 1–18. [CrossRef] [PubMed]
28. Armbrust, E.V.; Chisholm, S.W.; Olson, R.J. Role of light and the cell cycle on the induction of spermatogenesis in a centric diatom. *J. Phycol.* **1990**, *26*, 470–478. [CrossRef]
29. Finkel, Z.V.; Katz, M.E.; Wright, J.D.; Schofield, O.M.; Falkowski, P.G. Climatically driven macroevolutionary patterns in the size of marine diatoms over the Cenozoic. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8927–8932. [CrossRef]
30. Brunner, E.; Richthammer, P.; Ehrlich, H.; Paasch, S.; Simon, P.; Ueberlein, S.; van Pée, K.H. Chitin-based organic networks: An integral part of cell wall biosilica in the diatom *Thalassiosira pseudonana*. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9724–9727. [CrossRef]
31. Mock, T.; Otillar, R.P.; Strauss, J.; McMullan, M.; Paajanen, P.; Schmutz, J.; Salamov, A.; Sanges, R.; Toseland, A.; Ward, B.J.; et al. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* **2017**, *541*, 536–540. [CrossRef]
32. Dorrell, R.G.; Smith, A.G. Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryot. Cell* **2011**, *10*, 856–868. [CrossRef]
33. Lewitus, E.; Bittner, L.; Malviya, S.; Bowler, C.; Morlon, H. Clade-specific diversification dynamics of marine diatoms since the Jurassic. *Nat. Ecol. Evol.* **2018**, *2*, 1715–1723. [CrossRef] [PubMed]

34. Brown, J.W.; Sorhannus, U. A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): Substantive underestimation of putative fossil ages. *PLoS ONE* **2010**, *5*, 12759. [CrossRef] [PubMed]
35. Kooistra, W.; Forlani, G.; De Stefano, M. Adaptations of araphid pennate diatoms to a planktonic existence. *Mar. Ecol.* **2009**, *30*, 1–15. [CrossRef]
36. Parks, M.B.; Wickett, N.J.; Alverson, A.J. Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (Diatoms, Bacillariophyta). *Mol. Biol. Evol.* **2018**, *18*, 80–93. [CrossRef] [PubMed]
37. Kooistra, W.; Medlin, L.K. Evolution of the diatoms (bacillariophyta): IV. Reconstruction of their age from small subunit rRNA coding regions and the fossil record. *Mol. Phylogenet. Evol.* **1996**, *6*, 391–407. [CrossRef] [PubMed]
38. Guidi, L.; Chaffron, S.; Bittner, L.; Eveillard, D.; Larhlimi, A.; Roux, S.; Darzi, Y.; Audic, S.; Berline, L.; Brum, J.R.; et al. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **2016**, *532*, 465–470. [CrossRef] [PubMed]
39. Lommer, M.; Specht, M.; Roy, A.S.; Kraemer, L.; Andreson, R.; Gutowska, M.A.; Wolf, J.; Bergner, S.V.; Schilhabel, M.B.; Klostermeier, U.C.; et al. Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genom. Biol.* **2012**, *13*, R66. [CrossRef] [PubMed]
40. Armbrust, E.V.; Berges, J.A.; Bowler, C.; Green, B.R.; Martinez, D.; Putnam, N.H.; Zhou, S.G.; Allen, A.E.; Apt, K.E.; Bechner, M.; et al. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* **2004**, *306*, 79–86. [CrossRef]
41. Andersen, R.A. Biology and systematics of heterokont and haptophyte algae. *Am. J. Bot.* **2004**, *91*, 1508–1522. [CrossRef]
42. Stiller, J.W.; Schreiber, J.; Yue, J.; Guo, H.; Ding, Q.; Huang, J. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* **2014**, *5*, 5764. [CrossRef]
43. Moustafa, A.; Beszteri, B.; Maier, U.G.; Bowler, C.; Valentin, K.; Bhattacharya, D. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **2009**, *324*, 1724–1726. [CrossRef] [PubMed]
44. Deschamps, P.; Moreira, D. Re-evaluating the green contribution to diatom genomes. *Genome Biol. Evol.* **2012**, *4*, 683–688. [CrossRef] [PubMed]
45. Burki, F.; Flegontov, P.; Oborník, M.; Cihlár, J.; Pain, A.; Lukes, J.; Keeling, P.J. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol. Evol.* **2012**, *4*, 626–635. [CrossRef] [PubMed]
46. Yu, M.; Ashworth, M.; Hajrah, N.; Khiyami, M.; Sabir, M.; Alhebshi, A.; Al-Malki, A.; Sabir, J.; Theriot, E.C.; Jansen, R. Evolution of the Plastid Genomes in Diatoms. In *Plastid Genome Evolution*; Chaw, S., Jansen, R., Eds.; Elsevier: Amsterdam, The Netherlands, 2018; Volume 85, pp. 129–155.
47. Sabir, J.S.M.; Yu, M.J.; Ashworth, M.P.; Baeshen, N.A.; Baeshen, M.N.; Bahieldin, A.; Theriot, E.C.; Jansen, R.K. Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *PLoS ONE* **2014**, *9*, 107854. [CrossRef] [PubMed]
48. Ruck, E.C.; Linard, S.R.; Nakov, T.; Theriot, E.C.; Alverson, A.J. Hoarding and horizontal transfer led to an expanded gene and intron repertoire in the plastid genome of the diatom, *Toxarium undulatum* (Bacillariophyta). *Curr. Genet.* **2017**, *63*, 499–507. [CrossRef] [PubMed]
49. Kamikawa, R.; Yubuki, N.; Yoshida, M.; Taira, M.; Nakamura, N.; Ishida, K.; Leander, B.S.; Miyashita, H.; Hashimoto, T.; Mayama, S.; et al. Multiple losses of photosynthesis in Nitzschia (Bacillariophyceae). *Phycol. Res.* **2015**, *63*, 19–28. [CrossRef]
50. Kamikawa, R.; Tanifuji, G.; Ishikawa, S.A.; Ishii, K.I.; Matsuno, Y.; Onodera, N.T.; Ishida, K.I.; Hashimoto, T.; Miyashita, H.; Mayama, S.; et al. Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Mol. Biol. Evol.* **2015**, *32*, 2598–2604. [CrossRef] [PubMed]
51. Gruber, A.; Rocap, G.; Kroth, P.G.; Armbrust, E.V.; Mock, T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **2015**, *81*, 519–528. [CrossRef] [PubMed]
52. Felsner, G.; Sommer, M.S.; Gruenheit, N.; Hempel, F.; Moog, D.; Zauner, S.; Martin, W.; Maier, U.G. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. *Genom. Biol. Evol.* **2011**, *3*, 140–150. [CrossRef]

53. Gould, S.B.; Sommer, M.S.; Kroth, P.G.; Gile, G.H.; Keeling, P.J.; Maier, U.G. Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Mol. Biol. Evol.* **2006**, *23*, 2413–2422. [CrossRef]
54. Kikutani, S.; Nakajima, K.; Nagasato, C.; Tsuji, Y.; Miyatake, A.; Matsuda, Y. Thylakoid luminal theta-carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9828–9833. [CrossRef] [PubMed]
55. Peschke, M.; Moog, D.; Klingl, A.; Maier, U.G.; Hempel, F. Evidence for glycoprotein transport into complex plastids. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 10860–10865. [CrossRef] [PubMed]
56. Gile, G.H.; Moog, D.; Slamovits, C.H.; Maier, U.G.; Archibald, J.M. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genom. Biol. Evol.* **2015**, *7*, 1728–1742. [CrossRef] [PubMed]
57. Dautermann, O.; Lohr, M. A functional zeaxanthin epoxidase from red algae shedding light on the evolution of light-harvesting carotenoids and the xanthophyll cycle in photosynthetic eukaryotes. *Plant. J.* **2017**, *92*, 879–891. [CrossRef] [PubMed]
58. Dorrell, R.G.; Azuma, T.; Nomura, M.; Audren de Kerdrel, G.; Paoli, L.; Yang, S.; Bowler, C.; Ishii, K.I.; Miyashita, H.; Gile, G.H.; et al. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6914–6923. [CrossRef] [PubMed]
59. Qiu, H.; Price, D.C.; Weber, A.P.; Facchinelli, F.; Yoon, H.S.; Bhattacharya, D. Assessing the bacterial contribution to the plastid proteome. *Trends Plant Sci.* **2013**, *18*, 680–687. [CrossRef]
60. 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. [CrossRef]
61. Terashima, M.; Specht, M.; Hippler, M. The chloroplast proteome: A survey from the *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. *Curr. Genet.* **2011**, *57*, 151–168. [CrossRef]
62. Suzuki, K.; Miyagishima, S. Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses. *Mol. Biol. Evol.* **2010**, *27*, 581–590. [CrossRef]
63. Karlsen, E.; Schulz, C.; Almaas, E. Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformat.* **2018**, *19*, 467. [CrossRef]
64. Marchetti, A.; Schruth, D.M.; Durkin, C.A.; Parker, M.S.; Kodner, R.B.; Berthiaume, C.T.; Morales, R.; Allen, A.E.; Armbrust, E.V. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E317–E325. [CrossRef] [PubMed]
65. Morrissey, J.; Bowler, C. Iron utilization in marine cyanobacteria and eukaryotic algae. *Front. Microbiol.* **2012**, *3*, 43. [CrossRef] [PubMed]
66. Peers, G.; Price, N.M. Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **2006**, *441*, 341–344. [CrossRef] [PubMed]
67. Groussman, R.D.; Parker, M.S.; Armbrust, E.V. Diversity and Evolutionary History of Iron Metabolism Genes in Diatoms. *PLoS ONE* **2015**, *10*, 25. [CrossRef] [PubMed]
68. Lommer, M.; Roy, A.S.; 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 Genom.* **2010**, *11*, 718. [CrossRef] [PubMed]
69. Burki, F.; Kaplan, M.; Tikhonenkov, D.V.; Zlatogursky, V.; Minh, B.Q.; Radaykina, L.V.; Smirnov, A.; Mylnikov, A.P.; Keeling, P.J. Untangling the early diversification of eukaryotes: A phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. Biol. Sci.* **2016**, *283*, 20152802. [CrossRef] [PubMed]
70. Boyd, P.W.; Jickells, T.; Law, C.S.; Blain, S.; Boyle, E.A.; Buesseler, K.O.; Coale, K.H.; Cullen, J.J.; de Baar, H.J.; Follows, M.; et al. Mesoscale iron enrichment experiments 1993–2005: Synthesis and future directions. *Science* **2007**, *315*, 612–617. [CrossRef] [PubMed]
71. Gledhill, M.; Buck, K.N. The organic complexation of iron in the marine environment: A review. *Front. Microbiol.* **2012**, *3*, 69. [CrossRef]
72. Allen, A.E.; Laroche, J.; Maheswari, U.; Lommer, M.; Schauer, N.; Lopez, P.J.; Finazzi, G.; Fernie, A.R.; Bowler, C. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 10438–10443. [CrossRef]

73. Lampe, R.H.; Mann, E.L.; Cohen, N.R.; Till, C.P.; Thamatrakoln, K.; Brzezinski, M.A.; Bruland, K.W.; Twining, B.S.; Marchetti, A. Different iron storage strategies among bloom-forming diatoms. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E12275–E12284. [CrossRef]
74. Allen, A.E.; Dupont, C.L.; Obornik, M.; Horak, A.; Nunes-Nesi, A.; McCrow, J.P.; Zheng, H.; Johnson, D.A.; Hu, H.; Fernie, A.R.; et al. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* **2011**, *473*, 203–207. [CrossRef] [PubMed]
75. Broddrick, J.T.; Du, N.; Smith, S.R.; Tsuji, Y.; Jallet, D.; Ware, M.A.; Peers, G.; Matsuda, Y.; Dupont, C.L.; Mitchell, B.G.; et al. Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricornutum*. *New Phytol.* **2019**, *222*, 1364–1379. [CrossRef] [PubMed]
76. McCarthy, J.K.; Smith, S.R.; McCrow, J.P.; Tan, M.; Zheng, H.; Beeri, K.; Roth, R.; Lichtle, C.; Goodenough, U.; Bowler, C.P.; et al. Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell* **2017**, *29*, 2047–2070. [CrossRef] [PubMed]
77. Coesel, S.; Obornik, M.; Varela, J.; Falciatore, A.; Bowler, C. Evolutionary origins and functions of the carotenoid biosynthetic pathway in marine diatoms. *PLoS ONE* **2008**, *3*, e2896. [CrossRef] [PubMed]
78. Lepetit, B.; Gelin, G.; Lepetit, M.; Sturm, S.; Vugrinec, S.; Rogato, A.; Kroth, P.G.; Falciatore, A.; Lavaud, J. The diatom *Phaeodactylum tricornutum* adjusts nonphotochemical fluorescence quenching capacity in response to dynamic light via fine-tuned Lhcx and xanthophyll cycle pigment synthesis. *New Phytol.* **2017**, *214*, 205–218. [CrossRef] [PubMed]
79. Büchel, C. Evolution and function of light harvesting proteins. *J. Plant Physiol.* **2015**, *172*, 62–75. [CrossRef]
80. Murik, O.; Tirichine, L.; Prihoda, J.; Thomas, Y.; Araújo, W.L.; Allen, A.E.; Fernie, A.R.; Bowler, C. Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. *New Phytol.* **2019**, *221*, 1303–1316. [CrossRef]
81. Dang, K.V.; Plet, J.; Tolleter, D.; Jokel, M.; Cuiné, S.; Carrier, P.; Auroy, P.; Richaud, P.; Johnson, X.; Alric, J.; et al. Combined increases in mitochondrial cooperation and oxygen photoreduction compensate for deficiency in cyclic electron flow in *Chlamydomonas reinhardtii*. *Plant Cell* **2014**, *26*, 3036–3050. [CrossRef]
82. Dłużewska, J.; Szymańska, R.; Gabruk, M.; Kós, P.; Nowicka, B.; Kruk, J. Tocopherol Cyclases—substrate specificity and phylogenetic Relations. *PLoS ONE* **2016**, *11*, 0159629. [CrossRef]
83. Roberts, K.; Granum, E.; Leegood, R.C.; Raven, J.A. Carbon acquisition by diatoms. *Photosynth. Res.* **2007**, *93*, 79–88. [CrossRef]
84. Holdsworth, E.S.; Colbeck, J. The pattern of carbon fixation in the marine unicellular alga *Phaeodactylum tricornutum*. *Mar. Biol.* **1976**, *3*, 11. [CrossRef]
85. Reinfelder, J.R.; Kraepiel, A.M.; Morel, F.M. Unicellular C4 photosynthesis in a marine diatom. *Nature* **2000**, *407*, 996–999. [CrossRef] [PubMed]
86. Tanaka, Y.; Nakatsuma, D.; Harada, H.; Ishida, M.; Matsuda, Y. Localization of soluble beta-carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*. Sorting to the chloroplast and cluster formation on the girdle lamellae. *Plant Physiol.* **2005**, *138*, 207–217. [CrossRef] [PubMed]
87. Samukawa, M.; Shen, C.; Hopkinson, B.M.; Matsuda, Y. Localization of putative carbonic anhydrases in the marine diatom, *Thalassiosira pseudonana*. *Photosynth Res.* **2014**, *121*, 235–249. [CrossRef] [PubMed]
88. Tachibana, M.; Allen, A.E.; Kikutani, S.; Endo, Y.; Bowler, C.; Matsuda, Y. Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Photosynth Res.* **2011**, *109*, 205–221. [CrossRef] [PubMed]
89. Jensen, E.L.; Clement, R.; Kosta, A.; Maberly, S.C.; Gontero, B.A. A new widespread subclass of carbonic anhydrase in marine phytoplankton. *ISME J.* **2019**, *13*, 2094–2106. [CrossRef]
90. Shtaida, N.; Khozin-Goldberg, I.; Boussiba, S. The role of pyruvate hub enzymes in supplying carbon precursors for fatty acid synthesis in photosynthetic microalgae. *Photosynth. Res.* **2015**, *125*, 407–422. [CrossRef]
91. Bromke, M.A. Amino Acid biosynthesis pathways in diatoms. *Metabolites* **2013**, *3*, 294–311. [CrossRef]
92. Osada, K.; Maeda, Y.; Yoshino, T.; Nojima, D.; Bowler, C.; Tanaka, T. Enhanced NADPH production in the pentose phosphate pathway accelerates lipid accumulation in the oleaginous diatom *Fistulifera solaris*. *Algal Res.* **2017**, *23*, 126–134. [CrossRef]

93. Ge, F.; Huang, W.; Chen, Z.; Zhang, C.; Xiong, Q.; Bowler, C.; Yang, J.; Xu, J.; Hu, H. Methylcrotonyl-coA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum*. *Plant Cell* **2014**, *26*, 1681–1697. [CrossRef]
94. Smith, S.R.; Abbriano, R.M.; Hildebrand, M. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Res.* **2012**, *1*, 2–16. [CrossRef]
95. Ferriols, V.M.E.N.; Yaginuma-Suzuki, R.; Fukunaga, K.; Kadono, T.; Adachi, M.; Matsunaga, S.; Okada, S. An exception among diatoms: Unique organization of genes involved in isoprenoid biosynthesis in *Rhizosolenia setigera* CCMP 1694. *Plant J.* **2017**, *92*, 822–833. [CrossRef] [PubMed]
96. Muto, M.; Fukuda, Y.; Nemoto, M.; Yoshino, T.; Matsunaga, T.; Tanaka, T. Establishment of a genetic transformation system for the marine pennate diatom *Fistulifera* sp. strain JPCC DA0580–a high triglyceride producer. *Mar. Biotechnol.* **2013**, *15*, 48–55. [CrossRef] [PubMed]
97. Liu, X.J.; Hempel, F.; Stork, S.; Bolte, K.; Moog, D.; Heimerl, T.; Maier, U.G.; Zauner, S. Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Res.* **2016**, *20*, 249–257. [CrossRef]
98. Szkopińska, A.; Plochocka, D. Farnesyl diphosphate synthase; regulation of product specificity. *Acta Biochim. Pol.* **2005**, *52*, 45–55. [PubMed]
99. Busquets, A.; Keim, V.; Closa, M.; Del Arco, A.; Boronat, A.; Arró, M.; Ferrer, A. *Arabidopsis thaliana* contains a single gene encoding squalene synthase. *Plant Mol. Biol.* **2008**, *67*, 25–36. [CrossRef] [PubMed]
100. Metzger, P.; Largeau, C. *Botryococcus braunii*: A rich source for hydrocarbons and related ether lipids. *Appl Microbiol. Biotechnol.* **2005**, *66*, 486–496. [CrossRef] [PubMed]
101. Brown, T.A.; Belt, S.T.; Tatarek, A.; Mundy, C.J. Source identification of the Arctic sea ice proxy IP25. *Nat. Commun.* **2014**, *5*, 4197. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Annex III: *In vivo* localisation of Iron Starvation Induced Proteins under variable iron (in reveiw)

Title: *In vivo* localisation of Iron Starvation Induced Proteins under variable iron supplementation regimes in *Phaeodactylum tricornutum*.

Authors:

Elena Kazamia[1*], Jan Mach[2], Jeffrey B. McQuaid[3,^], Xia Gao[1], Tyler H. Coale[4], Ronald Malych[2], Jean-Michel Camadro[5], Emmanuel Lesuisse[5], Andrew E. Allen[3,4], Chris Bowler[1], Robert Sutak[2].

Affiliations

1. Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France.
2. Department of Parasitology, Faculty of Science, Charles University, BIOCEV, 25250 Vestec, Czech Republic.
3. Microbial & Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037
4. Scripps Institution of Oceanography, Integrative Oceanography Division, University of California, San Diego, La Jolla, CA 92093.
5. Université de Paris, CNRS, Institut Jacques Monod, UMR7592, 15 rue Hélène Brion, 75013 Paris.

Current address

^ The Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, 27570 Bremerhaven, Germany.

Email addresses

EK : kazamia@bio.ens.psl.eu
JM : mach@natur.cuni.cz
JMcQ : jmcquaid@ucsd.edu
TC : tcoale@ucsd.edu
XG: xgao@bio.ens.psl.eu
RM : malychr@natur.cuni.cz
J-MC : jean-michel.camadro@ijm.fr
EL : emmanuel.lesuisse@ijm.fr
AEA : aallen@jcvi.org
CB: cbowler@bio.ens.psl.eu
RS : robert.sutak@natur.cuni.cz

Corresponding author

Elena Kazamia
Tel: +1(619)-394-4843

Date of initial submission 15 June 2020

Date of resubmission 12 April 2021

Word Count: 6780

Total number of figures: 4 main text figures (colour) and 5 supplementary figures (colour), 3 supplementary tables and 10 supplementary videos.

Title *In vivo* localisation of Iron Starvation Induced Proteins under variable iron supplementation regimes in *Phaeodactylum tricornutum*.

Running title Dynamics of Iron Starvation Induced Proteins in *Phaeodactylum tricornutum.*

Highlight The marine diatom *P. tricornutum* engages a vesicular network to traffic siderophores and phytotransferrin from the cell membrane directly to a putative iron processing site in the vicinity of the chloroplast.

Abstract

The model pennate diatom *Phaeodactylum tricornutum* is able to assimilate a range of iron sources. It therefore provides a platform to study different mechanisms of iron processing concomitantly in the same cell. In this study we follow the localisation of three iron starvation induced proteins (ISIPs) *in vivo*, driven by their native promoters and tagged by fluorophores in an engineered line of *P. tricornutum*. We find that the localisation patterns of ISIPs are dynamic and variable depending on the overall iron status of the cell and the source of iron it is exposed to. Notwithstanding, a shared destination of the three ISIPs both under ferric iron and siderophore-bound iron supplementation, is a globular compartment in the vicinity of the chloroplast. In a proteomic analysis, we identify that the cell engages endocytosis machinery involved in the vesicular trafficking as a response to siderophore molecules, even when these are not bound to iron. Our results suggest that there may be a direct vesicle traffic connection between the diatom cell membrane and the PPC that co-opts clathrin mediated endocytosis and the "cytoplasm to vacuole" (Cvt) pathway, for proteins involved in iron assimilation. Proteomics data are available via ProteomeXchange with identifier PXD021172.

Keywords

Diatoms, Iron Starvation Induced Proteins, fluorescent proteins, siderophores, iron, proteome, *Phaeodactylum tricornutum*

1. Introduction

Iron is essential for cellular life. Its flexible redox chemistry confers reactivity to the fundamental metabolic processes of respiration, photosynthesis and nitrogen fixation. However, the redox flexibility that makes iron desirable for metabolism complicates its chemical speciation in the environment. Both the ionic state and ligand binding of iron depend on the pH and mineral chemistry of its ambient environment. By contrast, when unicellular life first evolved in the ocean, surface waters were less oxygenated, and iron was abundantly present as uncomplexed Fe^{2+}. In today's oxygenated ocean there is significantly less dissolved iron (<0.2nM; Boyd and Ellwood, 2010) and this predominantly in its oxidized, ferric form (Fe^{3+}). Of this dissolved pool, the vast majority (an estimated 99%) is organically complexed, and the remaining 1% is labile inorganic iron (Fe′) coupled to oxyhydroxides (Gledhill and Buck, 2012). Whether the different chemical forms of iron are bioavailable to marine microorganisms depends on the molecular machinery encoded by individual species. The processing of iron inside cells is subject to careful homeostasis. An excess of iron may be toxic, due to an imbalance of the redox chemistry of the cell and accumulation of reactive oxygen species.

Diatoms are the most successful marine microbial eukaryotes in the contemporary ocean, ubiquitous and abundant, estimated to contribute up to 40% of marine carbon fixation (Granum *et al.*, 2005). Their evolutionary origin involves at least two endosymbiotic events, the first of which occurred 1.5 billion years ago, when a eukaryotic heterotroph assimilated or was invaded by a cyanobacterium (a point in evolutionary history diatoms share with green plants or Viridiplantae). Divergence came an estimated 500 million years later, when a secondary endosymbiosis occurred in which a eukaryotic heterotroph captured a red alga, which subsequently established the Stramenopile lineage to which diatoms belong (Bhattacharya *et al.*, 2004).

It is widely considered that the ecological success of diatoms is in large part due to their iron physiologies, that are highly adapted to the contemporary ocean's iron chemistry (e.g. Morrissey and Bowler, 2012; Marchetti *et al.*, 2012). Open ocean diatoms are able to survive in waters that are chronically iron poor, but often dominate algal blooms induced by iron fertilization, for example through aeolian input (reviewed in Maher *et al.*, 2010). In the field of diatom research, three types of iron physiology adaptations have been proposed: decreased iron requirements, efficient Fe homeostasis and storage, and superior Fe uptake strategies. With respect to decreased iron quotas, functional replacements of ferroproteins

with non-iron requiring alternatives have been widely investigated, such as the replacement of ferredoxin with flavodoxin (La Roche *et al.*, 2005), which uses flavin rather than iron as a redox cofactor in the electron carrier, or the substitution of the iron-requiring cytochrome c_6 with the copper-coordinating plastocyanin (Peers and Price, 2006). Nevertheless, such replacements are not ubiquitous amongst diatom species, and the reconfiguring of cell photosynthesis and carbon storage in response to iron stress is enacted across a wide range of metabolic processes that are often species-specific (Groussman *et al.*, 2015; Smith *et al.*, 2018; Caputi *et al.*, 2019).

A portfolio of Iron Starvation Induced Proteins (ISIPs) was identified as likely protein candidates for conferring the ability of diatoms to survive periods of prolonged iron deprivation. In early transcriptomic analyses of *Phaeodactylum tricornutum,* a model pennate diatom with a sequenced genome (Bowler *et al.,* 2008), grown under iron replete and iron deplete conditions, transcripts of genes encoding the ISIP1, ISIP2A and ISIP3 proteins exhibited the highest fold-change in response to Fe′ limitation (Allen *et al.,* 2008). Their environmental importance has been verified by comprehensive analysis of the global ocean microbial metacommunity where *ISIP* transcript abundance was strongly anti-correlated with modelled Fe′ concentrations (Caputi *et al.,* 2019). The roles of the three ISIP proteins remain to be fully elucidated. Thus far, ISIP2A has received the most research attention, and was experimentally and phylogenetically shown to be a phytotransferrin protein, binding dissolved Fe^{3+} using pH-sensitive carboxylate functional groups and thus mediating the uptake of Fe′ into the cell (Morrissey et al., 2015; McQuaid *et al.,* 2018). The binding of iron by transferrins is a basal trait, widespread amongst all domains of life, although the amino acid sequence of diatom ISIP2A has little in common with the canonical transferrin structure.

ISIP1 was shown to be involved in the uptake of siderophore-bound iron via an endocytosis pathway (Kazamia *et al.,* 2018). Mutants deficient in *ISIP1* were unable to use siderophore-bound iron, likely due to aborted endocytosis. In the same work, experiments confirmed the cellular destination of the siderophore products to be the chloroplast interior. The phylogenetic origin of ISIP1 appears to be largely diatom-specific, although it was also detected in a small subset of pelagophyte and prasinophyte algae (Kazamia *et al.,* 2018). A related protein, p130B identified in *Dunaliella salina*, enabled iron uptake by associating with transferrin on the plasma membrane of cells (Paz et al., 2007). Additionally, it has been suggested that ISIP1 is an iron receptor, based on partial sequence similarity with the low-density lipoprotein receptor LDLR in humans (Lommer *et al.,* 2010). The function of ISIP3 has not been tested physiologically, although bioinformatic annotations predict ferritin-like

150 functional domains (DUF305, PF03713) within conserved sequences (Behnke and LaRoche,
151 2020).
152 In the present study we investigated the localisation of three ISIP proteins *in vivo* in a
153 genetically modified *P. tricornutum* cell line. With each gene driven by its native promoter
154 and coupled to a distinct fluorescent protein (FP) we were able to establish the localisation
155 pattern of the three ISIPs in the same cell under conditions of iron stress. We compared this
156 localisation pattern with the predicted localisation based on state of the art *in silico*
157 bioinformatic analyses. Next, we identified that the localisation of ISIPs is not static and is
158 instead highly dependent on the source of iron that was provided to cells experimentally. Our
159 observations of *in vivo* ISIP localisations together with published functional studies on these
160 and other proteins suggest a highly nuanced sensing of various iron sources by diatom cells
161 and at least four iron uptake pathways in *P. tricornutum*. Focusing on one of those uptake
162 processes, the assimilation of siderophores, using a proteomic analysis we shed light on
163 putative protein partners to ISIP1. We identify candidate transporter proteins as well as
164 components of endocytosis machinery likely to be specially adapted to function in diatom
165 cells, which have complicated intracellular compartmentalisation, a product of their
166 endosymbiotic origins.
167

2. Materials and Methods

Culture conditions of *P. tricornutum*

The wild-type (CCMP 2561) and engineered (P611) FP-tagged strain of *P. tricornutum* were grown under a 12-hour light/12-hour dark cycle, with an average intensity of 100 µmol photons m^{-2} s^{-1} during the light phase in a Panasonic MLR-352-PE incubator outfitted with with Osram Fluora T8 36W fluorescent lamps (with peak illumination at ~440 nm and ~630 nm, optimised for plant growth). Cultures were maintained at a temperature of 18 °C at a volume of 20 ml in sterile vented single use polystyrene flasks (Corning). Culture axenicity was confirmed by examining the cultures under the microscope and by standard plating analyses. The growth medium was modified artificial seawater, Mf (Sutak *et al.*, 2010). Its molecular composition is given in Supplementary Table S1, which also provides a comparison to Aquil media (Morel *et al.,* 1979). The pH was kept at 7.8 throughout growth and during experimental procedures. Reagents, media, and cultures were handled only using plasticware to avoid metal contamination. Where growth is reported, cell counts were taken daily in triplicate using a Malassez chamber.

To starve cultures of iron, cells in mid-exponential phase were collected, separated from growth medium by centrifugation (5000 rpm for 10 mins) and triple washed with Mf growth medium spiked with 1 mM EDTA to chelate iron. Cells were resuspended in fresh culture medium with no iron and allowed to grow for a week before the procedure was repeated. Cells that had undergone two rounds of iron starvation and showed a significant decrease in their growth rate compared with wild type were considered 'iron deplete' (denoted as '-Fe' treatments in figures).

Chemical stocks

Stock solutions of siderophores were prepared as follows: a 10 % excess of desferri-siderophore solution was added to $FeCl_3$ in 0.1 M HCl to give a 5 mM Fe solution with a Fe:ligand ratio of 1:1.1; after 10 minutes at room temperature, the pH was adjusted to 7 with 1 M HEPES, and the final 1 mM stock solution was kept at 20°C until use. The same procedure was followed to obtain siderophore analogues bound to gallium, instead of iron. It is important to note that two siderophores were used in our studies, desferrioxamine B (DFB) and desferrichrome (DFCH), which when coupled to iron are noted as ferrioxamine B (FOB) and ferrichrome (FCH), and as Ga-DFCH and Ga-FDB when bound to gallium. The siderophores can be used alternatively, as they both belong to the hydroxamate class of siderophores and are equally bioavailable to *P. tricornutum*, as previously shown by Kazamia

et al. (2018). For ferric EDTA (the Fe′ source in our experiments), the ligand to iron ratio was 20, and we checked (by ultracentrifugation) that the complex remained fully soluble during the time of experiments.

All other chemicals including growth medium reagents and bathophenanthroline disulfonate were purchased from Sigma-Aldrich.

Generation of trichromatic *Phaeodactylum tricornutum* line P611

To generate the trichromatic *P. tricornutum* line, three independent pUC-19 constructs were built for each ISIP gene by amlpifying 600 base-pair promoter + full length gene, fluorescent protein, and 300 bp terminator regions from iron-starved cDNA libraries. Amplifications were performed with Phusion high fidelity polymerase (New England Biolabs) and assembled into pUC-19 using Gibson Assembly cloning. YFP, RFP and CFP protein tags were used with *ISIP1*, *ISIP2A* and *ISIP3* genes, respectively; all primer sequences are available in Supplementary Table S2. Full length constructs were then re-amplified using extra-long primers with 40 base-pair sticky ends – these were complementary to each other as well as the destination cargo plasmid p0521s. p0521s contains a selection marker for bleomycin as well as the URA3 gene flanked by I-SceI sites - this latter marker provides an efficient counter selection to insert DNA sequences using yeast assembly methods (Karas *et al.,* 2015). For assembly, yeast cells were spheroplasted for 20 minutes prior to adding the three ISIP constructs with the linearized p0521s plasmid (Karas *et al.,* 2015). The final product was assembled in yeast, purified and reintroduced into *Escherichia coli* containing the conjugation plasmid pRL443 and screened for size with restriction enzymes. Sequence-verified clones were then introduced into *P. tricornutum* by plate mating (conjugation) with *E. coli* for 90 minutes in the dark at 30°C as described in Karas *et al.* 2015. Positive transformants were selected on zeocin antibiotic plates (100 µg ml^{-1}) and verified by PCR and sequencing. The resulting P611 line was grown in liquid cultures using Mf culture medium supplemented with zeocin (100 µg ml^{-1}) during microscopy experiments.

Confocal microscopy imaging

Detection of three fluorophores in the line P611

Images were acquired using a confocal imaging microscope, an inverted Leica SP8 (Leica Microsystems, Germany) equipped with 20x/0.75 and 63x/1.2 W objectives (HCLP APOC52), a pulsed white light Acousto-Optical Beam Splitter (AOBS) laser and internal hybrid single-photon counting detectors, HyD SMD 1-4. A water objective was used to

minimise light distortion during fluorescence capture, since cells were suspended in the Mf aqueous growth medium. Notch filters were used for each laser, the speed of acquisition was 200 frames per second, with optimised pixel range set to 3608 x 3608, 8 line accumulations and pixel development time of 100ns per frame. Scanning was sequential between frames. Fig. S1 shows expected emission and excitation spectra of the FPs.

We optimised image acquisition settings for measuring fluorescence of P611 cells, which were used in all time course experiments. This comprised the following three-step fluorescence acquisition process:

Sequence 1: cells were excited by 20 µW of 470 nm wavelength and fluorescence was detected in the window 475 nm-509 nm gated at 1.2 ns. This sequence allowed the measurement of CFP.

Sequence 2: cells were excited by a laser pulse at 20 µW of 514 nm light, with fluorescence detected in the 519-572 nm range. This sequence measured fluorescence of YFP.

Sequence 3: cells were excited at 590 nm (20 µW) with fluorescence detected at 595 nm- 660 nm as well as at 689 nm-730 nm, the latter using a photomultiplier tube (PMT) detector. The shorter wavelengths detected are indicative of RFP fluorescence, and the PMT detector measured chlorophyll autofluorescence. The signal between RFP and autofluorescence was separated by applying different time gates (0.9-1.7 ns for RFP and 3.6-12.5 ns for autofluorescence). The time gates were established through fluorescence lifetime imaging analysis (Fig. S2).

The fluorescence acquisition was done in this order, so that the dimmest fluorophore, CFP would be detected first, followed by YFP and then the brightest, RFP. Desiccation and bleaching from prolonged laser exposure occurred after approximately 10 minutes of continuous exposure, such as during a z-stack collection. Images interpreted in this manuscript are from alive cells only.

No FP-associated fluorescence was detected in wildtype cells under these parameters.

Fluorescence Lifetime Imaging (FLIM)

FLIM images were acquired during confocal imaging on inverted Leica SP8 TCS SMD FLIM (Leica Microsystems, Germany) equipped with 63x/1.2 W objective, a pulsed white light laser (selected excitation wavelength 470 nm, repetition frequency 80 MHz), AOBS and HyD detectors (measured emission range: 499-601 nm). Single-photon counting signal was detected using the HydraHarp400 module (PicoQuant, Berlin, Germany). FLIM

images were processed in SymPhoTime64 software (PicoQuant, Germany). The signals of ISIP2-RFP and chloroplast autofluorescence were separated by Pattern Matching analysis in SymPhoTime software. Pattern matching analysis can separate two populations based on different average lifetime and number of exponential components attributing to decay (Fig. S3).

Time courses

For a chosen population of cells, we took 10 images, either every minute between 0-10 minutes, every ten minutes over a 0-100 minute interval, or in hourly measurements 0-10 hours. Ten frames capture was the maximum number of images we could acquire of the cell population before significant laser damage occurred under our optimised imaging settings. The cells were suspended in their growth medium and placed inside a magnetic cover slip chamber (1 well Chamlide CMS, Quorum Technologies), which decreased the rate of evaporation. The maximum period over which we were able to track a population of cells was ten hours. After that time cells desiccated due to culture medium evaporation.

Spectral unmixing

Data for spectral unmixing analysis were acquired on an inverted confocal scanning microscope Carl Zeiss LSM 880 (Carl Zeiss, Germany) equipped with 63x/1.20 W objective, continuous wave lasers, and a spectral 32-channel GaAsP PMT detector. Spectral scans were acquired using 32 channels in the range 410-695 nm, each spectral channel has a width of 9 nm. For the final acquisition, the sample was excited simultaneously with 405 nm, 488 nm, and 561 nm lasers. For reference spectra CFP was excited by 405 nm laser only, YFP was excited by 488 nm laser only and RFP was excited by 561 nm only. The spectrum of autofluorescence was selected from the chloroplast region after excitation with a 405 nm laser. However, emission spectra of chloroplast autofluorescence at all the applied excitation wavelengths were identical. Reference emission spectra were obtained by spectral scans of structures labelled with only one fluorescent protein and were used to compile a spectra database (Fig. S4A). Spectral scans of samples containing CFP-, YFP- and RFP-labelled ISIP proteins were spectrally decomposed into four components (CFP, YFP, RFP, and autofluorescence) based on saved reference spectra by linear unmixing analysis in ZEN BLACK software (Carl Zeiss, Germany) (Fig. S4B).

The spatial resolution of the spectral unmixing method is diffraction-limited. Lateral spatial resolution depends on the numerical aperture of the objective (1.2) and the applied

303 excitation and emission wavelengths. Based on the Abbe formula, the achievable lateral
304 resolution in our analysis is ~170 nm for 405 nm excitation, ~200 nm for 488 nm excitation,
305 and ~230 nm for 561 nm excitation (Abbe, 1873).
306
307 **Proteomics**
308 *Sample preparation for proteomics analyses*
309 For global proteomic analyses, the cells were grown for one week in Mf medium
310 containing 0.1 µM iron (as ferric EDTA, Fe'), and then transferred either into iron-deficient
311 Mf medium or into iron-deficient Mf medium containing 1 µM Ga(III)-DFCH Ga(III)-DFCH,
312 which is the desferrisiderophore bound to gallium ions. After 3 days, cells from 3
313 independent cultures were harvested for each treatment, and cell extracts were prepared by
314 sonication in the presence of 0.1% digitonin, and the protein content of the samples was
315 measured, in order to quantify the amount of proteins to be precipitated with acetone.
316 *LC-MS/MS acquisition*
317 Protein extracts (40 µg) were precipitated with acetone at -20 °C (4 vol/vol sample).
318 The pellet of proteins was collected by centrifugation. The acetone supernatant was carefully
319 removed and the protein pellet was heated for 20 minutes at 95 °C and then cooled on ice for
320 20 minutes. It was resuspended in 20 µl of 25 mM NH_4HCO_3 containing sequencing-grade
321 trypsin (12.5 µg/ml, Promega) and incubated overnight at 37 °C. The resulting peptides were
322 desalted using ZipTip µ-C18 Pipette Tips (Millipore) and analysed on a Q-Exactive Plus
323 coupled to a Nano-LC Proxeon 1000 equipped with an Easy Spray ion source (all from
324 Thermo Scientific). Peptides were separated by chromatography with the following
325 parameters: Acclaim PepMap100 C18 pre-column (2 cm, 75 µm i.d., 3 µm, 100 Å), LC
326 EASY-Spray™ C18 column (50 cm, 75 µm i.d., 2 µm bead size, 100 Å pore size) operated at
327 55 °C, 300 nl/min flow rate, gradient from 95% solvent A (water, 0.1% formic acid) to 35%
328 solvent B (100% acetonitrile, 0.1% formic acid) over a period of 97 min, followed by a
329 column regeneration for 23 min, giving a total run time of 2 hrs. Peptides were analysed in the
330 Orbitrap cell, in full ion scan mode, at a resolution of 70,000 (at m/z 200), with a mass range
331 of m/z 375-1500 and an Automatic Gain Control (AGC) target of $3x10^6$. Fragments were
332 obtained by higher-energy C-trap dissociation (HCD) activation with a collisional energy of
333 30%, and a quadrupole isolation window of 1.4 Da. MS/MS data were acquired in the
334 Orbitrap cell in Top20 mode, at a resolution of 17,500 with an AGC target of $2x10^5$, with a
335 dynamic exclusion of 30 sec. MS/MS of most intense precursor were acquired first. Peptides

with unassigned charge state or singly charged were excluded from the MS/MS acquisition. The maximum ion accumulation times were set to 50 ms for MS acquisition and 45 ms for MS/MS acquisition.

Quantitative analysis in label-free experiments

Label-free quantification in-between subject analysis was performed on raw data with Progenesis-Qi software 4.1 (Nonlinear Dynamics Ltd, Newcastle, U.K.) using the following procedure: (i) chromatograms alignment, (ii) peptide abundances normalisation, (iii) statistical analyses of features, and (iv) peptides identification using the Mascot server through Proteome Discoverer 2.1 (Thermo Scientific). A decoy search was performed and the significance threshold was fixed to 0.01. The resulting files were imported into Progenesis-LC software. Normalised abundances of proteins from trypsin digests with similar normalised abundance variations of the corresponding peptides (ANOVA p value < 0.05) were classified together by the AutoClass Bayesian clustering system (Achcar *et al.,* 2009) and visualised with Javatreeview (http://jtreeview.sourceforge.net/).

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol et al., 2019) with the dataset identifier PXD021172.

3. Results

Visualising the three ISIPs in the same cell

We generated a trichromatic transgenic *P. tricornutum* line we called P611, with ISIP1 tagged to YFP, ISIP2A tagged to RFP, and ISIP3 to CFP. For each protein the FP tag was C-terminal to the full protein sequence and the construct was driven by native promoters for each gene (see Materials and Methods for further details). The goal was to observe the patterns of localisation for the three proteins in the same cell and compare it to *in silico* targeting predictions.

The *in silico* targeting prediction for the three proteins is the same; in each case a signal peptide is identified bioinformatically (SignalP 3.0, Bendsten *et al.*, 2004), which does not have the motif necessary for chloroplast targeting (ASAFind, Gruber *et al.,* 2015), with likely protein destination therefore assigned to the cell endomembrane (Rastogi *et al.*, 2018).

To determine the best time of day to take *in vivo* measurements of fluorescence, we synchronised the cells to a 12 hr light/12 hr darkness regime and starved them of iron, since this is the condition which promotes *ISIP* expression (Allen *et al.*, 2008). We used the

supplementary data to a time-resolved transcriptomic study of *P. tricornutum* under iron limitation (Smith *et al.,* 2016) to anticipate maximum expression and fluorescence levels (Fig. S5). *ISIP* expression peaked during darkness and, allowing time for translation, we confirmed maximum fluorescence detection within the first hours after illumination, at the beginning of the light phase.

A typical localisation pattern for ISIPs in iron-starved P611 cells imaged as early as possible into the light phase is shown in Fig. 1. All three ISIPs localised to the vicinity of the chloroplast midpoint, in a globular region (Fig. 1A, D-E), although ISIP2A-RFP showed tighter aggregation, at the centre of the chloroplast groove (Fig. 1E). ISIP1-YFP and ISIP3-CFP co-localise (whether this could be due to a detection error was further investigated). The diameter size of the bright focal point of ISIP2A-RFP was approximately 1.12 µm (SD=0.2, n=100), and for ISIP1/ISIP3 it was 3.06 µm (SD=0.3, n=100). The average plastid diameter size was 6.21 µm (SD=0.5, n=100) measured from differential interference contrast (DIC) images (Fig. 1C) rather than autofluorescence, which is variable throughout the day and does not include the chloroplast membranes. The average cell diameter of trichromatic cells was 23.2 µm (SD=4.0, n=100), and did not deviate significantly from wild type (*wt* diameter= 23.0 µm, n=100). In their majority cells were fusiform, with a small subpopulation of oval cells (<1% of the total population). Chlorophyll autofluorescence is shown in Fig. 1B; and it was successfully 'gated-out' from interfering with FP emission measurements (see Materials and Methods and Fig. S2). Chlorophyll autofluorescence did not interfere with FP detection during experimental procedures.

ISIP1-YFP and ISIP3-CFP showed near perfect co-localisation in P611 cells. We performed an additional set of experiments using a Zeiss confocal microscope equipped with 34 detection channels, and a lower, 405 nm excitation wavelength, followed by a procedure of spectral unmixing, to parse emission spectra of YFP from CFP. While we were able to isolate detection of CFP in the cells, its localisation under conditions of iron deprivation fully overlapped either one or both of the other fluorophores (Fig. S4). It is possible that this is because ISIP3 interacts with these proteins. Further functional analyses are required to test this hypothesis.

Localisation of ISIPs subject to Fe supply

In previous work we showed that ISIP1 has a dynamic cell localisation and can be detected in endocytic vesicles, the chloroplast and the cell membrane (Kazamia *et al.*, 2018).

401 A similar behaviour has also been noted for ISIP2A (McQuaid *et al.,* 2018). To test how ISIP
402 proteins respond dynamically to iron supplementation *in vivo*, first we conducted time-course
403 experiments with iron pulses. Iron-starved cells were dosed either with Fe′ or with the
404 hydroxamate siderophore, ferrioxamine B (FOB) at analogous concentrations. We did not
405 observe significant differences in ISIP localisation over a period 0-10 minutes, 0-1hr or 0-10
406 hrs (representative 0-10 mins time courses shown in Supplementary Videos SV1-SV10).
407 Previously we had quantified that release of siderophore-bound iron delivered via endocytosis
408 to the chloroplast occurs within 4 minutes of iron addition (Kazamia *et al.* 2018,
409 Supplementary Video). Therefore, it is possible that even our fastest time course analysis (0-
410 10 minutes) missed this dynamic protein movement, where the first frame is taken after 30 s.
411 It was not possible to take measurements with a faster detection frequency using the available,
412 state-of-the-art microscopy platforms. We conclude that all images we obtained therefore
413 showed cells with a steady state protein localisation, which did not change in the first 10
414 hours of iron supplementation.
415 In addition, we took the approach of pulsing *P. tricornutum* cells with a fluorescently
416 labelled siderophore, a chemical conjugate of FOB with nitrobenzoxadiazole (FOB-NBD).
417 We found that FOB-NBD could recapitulate the observed pattern of localisation reported by
418 Kazamia *et al.* (2018) for wild type cells, although NBD fluorescence was masked by the
419 stronger YFP signal emerging from the tagged ISIP1 in the P611 line.
420 While we were unable to track the same population of cells over a time-course period
421 longer than 10 hours because of cell desiccation, we tracked differences in fluorescence in
422 population averages over a period of 5 days following iron supplementation. Based on
423 transcriptomic analyses of the three *ISIPs* (Smith *et al.*, 2016) we expected Fe′
424 supplementation to decrease ISIP1 fluorescence and to have no significant effect on ISIP2A-
425 RFP or ISIP3-CFP (Fig. S5). Intriguingly, we observed that either the addition of siderophore
426 chelated iron (FOB) or labile Fe^{3+} (Fe′) both led to a significant decrease in fluorescence of
427 ISIP1-YFP and ISIP2A-RFP in the cell population, 48 hours following supplementation (Fig.
428 2). In that time there was no overall change in average levels of chlorophyll fluorescence.
429 FOB had the strongest influence, resulting in near complete loss of fluorescence associated
430 with ISIP1 and ISIP2A (Fig. 2). Significantly, supplementation of desferrioxamine chelated to
431 gallium instead of iron (+Ga-FDB) had no equivalent effect (Fig. 2, fourth row). Instead, the
432 membrane localisation of ISIP1-YFP under that condition appears stronger than in iron
433 starved cells, suggesting that the presence of the siderophore induces the movement of ISIP1-
434 YFP to the membrane. With Fe′ supplementation, FP fluorescence disappeared altogether

435 after a further 12 hours, that is ~60 hrs from the pulse (data not shown), while average
436 chlorophyll autofluorescence remained similar. ISIP3-CFP was hardest to image, and it was
437 weakly fluorescent even in iron-starved cells (when fluorescence is expected to be high). We
438 observed a slight increase in fluorescence associated with ISIP3-CFP under the Ga-FDB
439 treatment, but this requires further verification.
440 To revisit the role of reduction in iron uptake at the cell surface of *P. tricornutum*
441 (discussed most recently in Kazamia *et al.,* 2018, McQuaid *et al.,* 2018 and Coale *et al.,*
442 2019), we additionally supplemented iron-starved *P. tricornutum* cells with Fe′ in the
443 presence of the Fe^{2+} chelator bathophenanthroline disulfonate (BPS) and recorded
444 fluorescence (Fig. 2, bottom row). We observed that the localisation of ISIP2A-RFP in the
445 presence of BPS remained similar to conditions of iron starvation ('no Fe' treatment). Our
446 proposed explanation for the difference between this localisation and Fe′ supplementation
447 without BPS is that ISIP2A continues to be required and is functional when cells are hindered
448 in reductive iron uptake. In the +Fe′ treatment we are thus observing the global cell response
449 due to reductive Fe′ uptake as well as endocytosis mediated by phytotransferrin (ISIP2A)
450 assimilation, which quickly replenishes the iron stocks inside the cell and lowers the
451 requirement of ISIP2A.
452 Intriguingly, our imaging results suggest that subject to conditions all ISIP proteins at
453 some point in their lifetime converge on a globular compartment in the vicinity of the
454 chloroplast, identifying this region as a possible processing site for iron inside diatom cells.
455 This is visible for ISIP1-YFP and ISIP2A-RFP in 'no Fe' treatment and in the '+Ga-FDB'
456 treatment for ISIP3-CFP (Fig. 2).
457
458 **Summary of iron uptake pathways and likely roles of ISIPs**
459 Overall, our results support the view that there are at least four independent processes
460 of iron assimilation in *P. tricornutum*: direct uptake of ferrous iron at the cell surface
461 (Schallenberg *et al.*, 2015), non-reductive uptake of ferric iron via phytotransferrin (Morrissey
462 *et al.,* 2015, McQuaid *et al.,* 2018), reductive uptake of ferric iron at the cell surface (Lis *et*
463 *al.*, 2015), and non-reductive uptake of siderophore-bound iron (Kazamia *et al.,* 2018, Coale
464 *et al.,* 2019). Uptake of Fe(II) preceded by reduction of Fe(III) on the cell surface of diatoms
465 has been additionally demonstrated in *Thalassiosira* diatoms (Anderson and Morel, 1980;
466 Shaked *et al*., 2004). A similar process has also been described in the green alga
467 *Chalmydomonas reinhardtii* (Eckhardt and Buckhout, 1998).

468 We summarise the role of known proteins in these pathways in Fig. 3A. ISIP2A is
469 linked directly to only one of these, whereas ISIP1 is likely to function in two pathways,
470 involved in the assimilation of bound ferric iron either by siderophores or transferrin. The role
471 of ISIP3 remains elusive. Although of the three ISIP proteins it is the most globally abundant
472 (Caputi et al., 2019), it is not highly expressed in *P. tricornutum* (Figure S5) and was the
473 hardest to image, exhibiting fluorescence levels just above detection limits. It is likely that
474 either ISIP3 has a role in iron acquisition not captured by our experimental conditions, or that
475 *P. tricornutum* does not rely heavily on this protein.
476 It is important to note that the rate and affinity for iron of each type of uptake system
477 is predicted to be dependent on the abundance of the iron source and the ambient pH (Fig.
478 3B). Under low pH, reductive uptake of Fe′ is predicted to be significantly lower compared to
479 non-reductive iron uptake. This was evidenced in Kazamia *et al.* (2018), where growth culture
480 medium was kept at a pH of 7.8 and reductive iron uptake processes were slow, and more
481 thoroughly in a mechanistic investigation into the role of carbonate ions (whose
482 concentrations change with pH) in directing ISIP2A-mediated uptake of ferric irons by
483 McQuaid *et al.* (2018). That others find the phytotransferrin-mediated uptake to be faster
484 under higher pH (Morrissey et al., 2015, Coale *et al.,* 2019) is not surprising, and is in
485 agreement with this synthesis (Fig. 5B).
486
487 **Proteomics analysis of the response of *P. tricornutum* to iron-free siderophore**
488 **stimulation**
489 We decided to shed further light on the siderophore uptake pathway by conducting an
490 exploratory proteomics study. Our aim was to compare the protein portfolios of diatom cells
491 that had been pulsed with a siderophore chelated to gallium (Ga-DFCH) with control iron-
492 limited cells (that had not been pulsed with DFCH). In both conditions the cells were iron
493 starved, which eliminates the confounding factor of the intracellular iron status, which is
494 difficult to measure and control. In the Ga-DFCH treatment iron starved *P. tricornutum* cells
495 were spiked with the small organic molecule, desferrichrome bound to gallium instead of
496 iron. This comparison allowed us to identify the siderophore response of the cell, in the
497 absence of an iron stimulus or change to the overall iron state of the cell. The study was
498 designed to parse out proteins that are involved solely in siderophore uptake, and not in the
499 sensing of iron. We expected the highest concentration of ISIPs under this treatment, based on
500 our microscopy results, which showed the highest fluorescence of all three ISIPs in cells
501 pulsed with a gallium desferrisiderophore (Fig. 2).

Supplementary Table S3 describes the top 100 proteins that showed either an upregulated or repressed response under the two conditions ranked by the p values of an analysis of variance (see Methods for more details). Of these, one fifth were of unknown function and approximately one third (35/100) were involved in housekeeping cell functions such as cell division. Seven annotated transporter proteins showed a differential response under Ga-DFCH supplementation compared to control treatments. Of these, one transporter protein (NCBI protein ID 219118307), annotated as a 'fungal purine transporter', showed the highest fold difference with respect to the proteome from control iron starved cells, albeit with low statistical confidence. It is possible that this transporter is specific to siderophore transport into the cell and is a potential candidate for further study of the siderophore uptake system in diatoms. Both ISIP1 and FBP1 showed significant increases after the addition of the siderophore analogue (Fig. 4A), as anticipated. Two anion exchange proteins (NCBI IDs 219114320, 441476376) were significantly repressed ($p<0.05$) and are both annotated as belonging to the $HCO3^-$ transporter family. This suggests that siderophore processing is linked with intricate intracellular pH homeostasis. A paralog of ISIP2A, more similar to Fea1 (NCBI ID 219122261) also showed statistically significant increases compared to controls.

Major changes were detected in putative endocytosis and membrane trafficking pathways of the cell, whose proteins were consistently more abundant in the Ga-DFCH treated cells (Fig. 2B). Both clathrin and its major adaptor protein, epsin, showed striking quantitative increases, albeit of low statistical significance, compared to the proteome of control cells, further corroborating the observations that siderophore uptake occurs via endocytosis (Kazamia *et al.,* 2018). Two putative components of exocytosis were repressed with their fold changes shown in Fig. 4B.

Finally, 16 proteins linked to micronutrient metabolism showed significant changes in their abundance in iron starved cells stimulated with Ga-DFCH. Of these, 13 were induced and are likely involved in siderophore processing. For example, the non-heme iron binding, demethylating protein (NCBI protein ID 219121067) showed a two-fold increase in abundance, as did a putative siderophore synthase (NCBI protein ID 219122041), which may be involved in the breakdown and remodelling of the siderophore molecule (Supplementary Table S3).

4. Discussion

Understanding how diatoms process iron is an important question in cell physiology that has consequences on global biogeochemistry. As the contemporary ocean acidifies and warms, the inorganic iron pool is rapidly changing. Geological reconstructions of iron speciation in the ocean indicate that such shifts lead to dramatic changes to photosynthetic life (e.g. Burke *et al.,* 1993) and an understanding of the phytoplankton response is paramount to our predictions of the impacts of climate change. *P. tricornutum* is an excellent model for the study of diatom iron physiology as it appears to harbour the full suite of iron uptake and processing mechanisms that have been described for any diatom (Smith *et al.*, 2016). In this model species we have the possibility to compare changing iron supplementation regimes on different uptake mechanisms concomitantly in the same cell.

The morphology of *P. tricornutum* cells is thoroughly documented, including through electron microscopy imaging (starting with Borowizska and Volcani, 1978). While there are three morphotypes of *P. tricornutum*, the majority of cells of the model ecotype, which we use here, are fusiform, weakly silicified, and harbour a single plastid, which as in all diatoms is derived from a red algal symbiont (Cavalier-Smith, 2003). The plastid is separated from the rest of the cell by four membranes. The two outermost membranes are the chloroplast endoplasmic reticulum membrane, which is continuous with the host outer nuclear envelope, and the periplastidial membrane. The two innermost membranes correspond to the outer and inner envelope membranes (iEM and oEM respectively) of the symbiont's chloroplast. Between the periplastidial membrane and the oEM lies a minimized symbiont cytoplasm, the periplastidial compartment (PPC). This space has been described as home to "blob-like structures", with certain proteins residing in this space permanently (Moog *et al.*, 2011), likely directed here by a modified chloroplast signal peptide (Kilian and Kroth, 2005; Gruber *et al.*, 2007). The cell functions carried out within the PPC remain elusive, since proteins from this compartment have not been separated from the rest of the cell in the lab. In recent work, a vesicular network was revealed via electron microscopy to exist within the PPC, although canonical components involved in vesicular lipid trafficking were not predicted to be directed here *in silico* (Flori *et al*., 2016).

There are three iron starvation induced proteins annotated in diatoms, ISIP1, ISIP2A and ISIP3 (Allen *et al.,* 2008) and we were able to visualise their localisation in a trichromatic line of *P. tricornutum,* P611, under a range of scenarios of iron supplementation. It is clear that the localisation of the proteins is dynamic; all ISIPs showed an endomembrane localisation and a likely PPC localisation, which varied depending on the iron

supplementation regime. We propose that the aggregation of ISIP proteins we visualise in the PPC of P611 (Fig. 3) represents a dedicated iron processing space within this reduced cytoplasm of the secondary endosymbiont. ISIP2A showed a narrower distribution, whereas ISIP1/ISIP3 were slightly more diffuse, but directly adjacent. Others have imaged FRE2 and FBP1 tagged with FPs aggregating in an analogous location (Coale *et al.*, 2019). We propose that the confluence of proteins related to iron uptake in this space is to serve as the site of iron reduction, prior to the distribution of ferrous iron as co-factor to the iron-requiring photosynthetic and respiratory machineries. Our hypothesis is that this site is under strict pH regulation, which in turn controls the release of ferrous iron for cell use. It is possible that the bicarbonate transporters identified in the proteomics study we conducted (Table S3 and Fig. 4) are involved in the pH regulation of this space. Further physiological experiments are required to test this hypothesis – our work serves merely as a preliminary investigation for further detailed study.

The dynamic localisation patter for ISIP1 and ISIP2A further corroborates the observation that these proteins make use of endocytosis pathways (Kazamia et al., 2018; McQuaid et al., 2018). How the vesicle network operates continuously between the diatom endomembrane and the PPC in diatoms requires further investigation, but since ISIP proteins do not reside in this space permanently, it is not surprising that they lack the PPC-specific targeting motif identified by Gruber *et al.* (2007). Some of the proteins involved in the endocytosis-mediated uptake of siderophores were identified by our proteomics study. They include canonical clathrin-mediated endocytosis proteins epsin, SNARE, copine and the microtubule affinity regulating kinase, (MARK) (Fig. 4). Intriguingly, our study also picked up a two-fold increase in autophagy-related protein 13, ATG13, during siderophore supplementation. ATG13 is an adaptor protein for the "cytoplasm to vacuole" Cvt pathway originally described in yeast, but present in all eukaryotes (Reggiori *et al.,* 2004). A major difference between the Cvt pathway and the canonical endocytosis network is that it employs double membrane vesicles and by-passes the Golgi apparatus. Further research is required into the physical connection between the vesicular network identified inside the PPC by Flori *et al.* (2016) and the cytoplasm/cell membrane of diatoms. Our studies on ISIP proteins suggest, albeit tentatively, that there may be a direct vesicle traffic connection between the diatom cell membrane and the PPC that co-opts clathrin mediated endocytosis and the Cvt pathway. An unique proposition that would be apt for diatoms, which have complicated internal structures due to their complex endosymbiotic origins.

We propose that ISIP1 is as an adaptor protein for the vesicular network, engaged to deliver chelated ferric iron bound to proteins to the dedicated iron homeostasis site for processing. In our hands, of the three ISIPs, ISIP1 showed the most dynamic localisation, sensitive both to the overall iron status of the cell and the source of iron that the cell is exposed to. In iron replete trichromatic P611 cells we detected little fluorescence associated with ISIP1-YFP on the cell membrane, which was significantly increased when cells were supplied with siderophores. This is particularly apparent when gallium analogues of the siderophore are used (Fig. 2). Supplementation with iron laden siderophores increases membrane fluorescence in the short term, but within 48 hours the iron status of the cell changes so total fluorescence is lower, so the effect is easy to miss.

In previous work we showed experimentally that siderophore uptake and assimilation are inhibited in cells deficient in *ISIP1* (Kazamia *et al.*, 2018). While knocking down this protein does not impact cell membrane uptake of Fe′ via ISIP2A, it has not been investigated whether cells deficient in ISIP1 process intracellular ISIP2A-bound ferric iron in the same way as wild type cells, so an intracellular interaction between ISIP1 and ISIP2A is a likely possibility. A recent study investigating the interactions of ISIP2A with other proteins through engineered ascorbate peroxidase APEX2-based subcellular proteomics identifies ISIP1 as a highly likely interacting partner (Turnsek *et al.*, 2021).

In conclusion, our study shed light on the localisation of three ISIP proteins within the same diatom cell and revealed a dynamic pattern of localisation that depends on the source of iron and the overall iron status of the cell. It is likely that siderophore bound and transferrin bound iron is taken for processing to a globular compartment in the vicinity of the chloroplast via a non-canonical endocytosis route, which combines elements of clathrin-mediated and Cvt pathways. Higher resolution studies are required to confirm this and to clarify whether the destination is the PPC or perhaps even a dedicated iron processing organelle inside the cells, which can be maintained in careful pH balance, subject to the needs of the mitochondria and the chloroplast for ferrous iron. Furthermore, while it was our intention to push *in vivo* fluorescence imaging in an auto-fluorescent organism to its limits by imaging three FPs in the same cell, further studies can refine our initial findings here through analyses of single fluorescent lines.

5. Acknowledgements

632 We thank Marie Olsinova and Radek Machan from BIOCEV for their technical help with
633 collecting and analysing microscopy images, supported by Czech-BioImaging (LM2018129
634 and LM2015062) ERDF (CZ.02.1.01/0.0/0.0/16_013/0001775). EK and JMQ were supported
635 by Marie Curie Individual Fellowships H2020 IRONCOMM and IRONBIND. RS
636 acknowledges the Czech Science Foundation project 18-07822S and
637 CZ.02.1.01/0.0/0.0/16_019/0000759 CePaViP provided by ERDF and MEYS. Funding is
638 acknowledged from the Agence Nationale de la Recherche "Phytomet" (ANR-16-CE01-0008
639 to CB and EL). CB additionally acknowledges funding from the European Research Council
640 (ERC) under the European Union's Horizon 2020 research and innovation programme
641 (Diatomic; grant agreement No. 835067), as well as the French Government "Investissements
642 d'Avenir" programs MEMO LIFE (ANR-10-LABX-54), PSL* Research University (ANR-
643 11-IDEX-0001-02), and Oceanomics (ANR-11-BTBR-0008). This study was supported by
644 the National Science Foundation Grants NSF-OCE-1756884 and NSF-MCB-1818390 (to
645 A.E.A.); United States Department of Energy Genomics Science program Grant DE-
646 SC0018344 (to A.E.A.); Gordon and Betty Moore Foundation Grants GBMF3828 and
647 GBMF5006 (to A.E.A.).
648

649 **6. Conflict of Interest**

650 The authors declare no conflict of interest.

7. Figure legends

Fig. 1

Representative images of Fe-starved *P. tricornutum* P611 trichromatic cell. In this line ISIP1 is tagged to YFP, ISIP2A to RFP, ISIP3 to CFP. (A) A field view of P611 cells under the optimised settings on a Leica SP8 confocal microscope (see methods). Chlorophyll fluorescence does not interfere with the FPs. (B) Magnified image of the cell region indicated by the white square in Panel A (chl). Differential interference contrast (DIC) also shown for comparison. YFP and CFP show near perfect co-localization. (C) A representative image of P611 cells captured on the Zeiss confocal microscope using linear unmixing (see Methods), for comparison.

Fig. 2

Trichromatic iron-starved line, spiked with a range of iron sources, imaged 48 hours after spike.

Treatments (top row to bottom): control (no Fe spike); labile iron supplemented as iron citrate (1 μM); siderophore spike (1 μM FOB); gallium analogue of siderophore (1 μM Ga-FDB); Fe′/BPS (1 μM/ 100 μM) spike to remove Fe^{2+} from the cell surface.

Fig. 3

Summary of iron uptake systems in *P. tricornutum*.

(A) Synthesis schematic of the known iron uptake and processing mechanisms in *P. tricornutum*. The iron pool is partitioned into three components: ligand bound Fe (Fe_L), Fe^{2+} and Fe^{3+}. Uptake of Fe^{2+} has not been measured directly but is presumed to be fast and prevalent in all diatoms. This is the smallest iron pool available, as the majority of iron is either chelated or oxidised. Soluble Fe^{3+} (as in an Fe-EDTA system or Fe^{3+} supplementation through iron citrate addition) is taken-up either through a phytotransferrin mediated uptake system or a reductive process on the cell surface (indicated here by FRE). The former is preferred under higher pH when the concentration of dissolved carbonate is lower and phytotransferrin (ISIP2A) is more efficient. Siderophore uptake is via extracellular binding to specialised proteins, including FBP1 and mediated via endocytosis, coordinated by ISIP1. The proposed intracellular dynamics of ISIP1 are shown by arrows and include a possible association with ISIP2A. The ultimate localisation of all protein-ligated Fe^{3+} iron (siderophore- and phytotransferrin- bound) is a cite (perhaps a putative organelle, indicated in

685 dashed lines) in the vicinity of the chloroplast (shown in green). This is the site of reduction,
686 by FRE2 in the case of siderophores (Coale *et al.,* 2019) but possibly a wider suite of
687 reductases. **(B)** Summary of the uptake pathways preferred under low and high pH scenarios,
688 experimentally verified in *P. tricornutum*.
689
690 **Fig. 4**
691 **Proteome comparison of iron limited *P. tricornutum* cells with and without Ga-DFCH**
692 **supplementation.**
693 Iron limited cells (control) compared to cells spiked with Ga-DFCH (treatment). Protein
694 extraction was 3 days after siderophore spike/ iron limitation. (A**)** Putative siderophore or iron
695 processing proteins. (B) Changes to proteins associated with the endocytosis
696 machinery/membrane trafficking of the cell. Black bars: proteins that increased in abundance;
697 red bars: those that showed repression. Further information/annotation provided in
698 Supplementary Table S2.
699
700
701
702
703

704 **Supplementary Figure Legends**

705

706 **Fig. S1 Emission and excitation spectra for the fluorescent proteins used in this study.**
707 These are the expected excitation (dashed lines) and emission spectra (solid lines) of YFP,
708 RFP and CFP. However, cell types and growth medium affect the observed spectra of these
709 fluorophores.

710

711 **Fig. S2 FLIM of P611 cells excited at 470 nm and a repetition frequency of 80 MHz.**
712 The lefthand panel shows the overall emission intensity (in total counts) of P611 cells over a
713 period of 12 seconds. The righthand panels show images of the cells fluorescence within a
714 defined time gate, either 0-12.83 seconds (total), or an early (0.9-1.73 ns) and late (3.44-12.83
715 ns) gate. The initial steep drop-off in intensity is largely attributed to chlorophyll
716 autofluorescence, which decays significantly faster than the fluorescence associated with FPs.

717

718 **Fig. S3 Pattern matching analysis in SymPhoTime software.**
719 This shows the analysis output for a representative image of P611 cells excited at 470nm
720 using SymPhoTime software. RFP associated fluorescence (red) and chloroplast
721 autofluorescence (green) is separated by pattern matching analysis based on separation of
722 average lifetimes and number of exponential components attributing to decay (decay diversity
723 parameter).

724

725 **Fig. S4 Spectral unmixing of fluorescence measured in P611 cells.**
726 (A) Measured emission spectra for CFP, YFP, RFP and autofluorescence in the trichromatic
727 P611 cells using 34 detection channels (9nm width each) at an excitation of 405nm (20μW).
728 (B) Nine frames of a z-stack imaging an iron starved (-Fe) P611 cell. The fluorescence
729 measured is assigned into 4 categories, CFP, YFP, RFP or autofluorescence based on the
730 emission spectra shown in (A).

731

732 **Fig. S5**
733 **Transcriptional control of ISIPs based on results reported in Smith *et al.*, 2016).**
734 Transcript abundance of ISIPs was followed every 4 hours over a period of 48 hours for *P.*
735 *tricornutum* cultures that were either iron starved (red line) or replete in iron (black line).
736 Measurements were taken at 10 am, 2pm, 6pm, 10pm, 2 am and 6 am under a 12 hour light
737 (white background), 12 hour darkness (grey background) regime, with lights on at 9 am and

738 lights off at 9 pm. For iron starved cells point represents biological duplicates and for iron
739 replete an average of 3 biological replicates. Transcription abundance is measured as Reads
740 Per Kilobase of transcript, per Million mapped reads (RPKM).
741
742
743
744

745 **Supplementary Table Legends**

746

747

748 **Supplementary Table S1.**

749 A culture medium composition comparison between Mf and Aquil.

750

751 **Supplementary Table S2.**

752 Primer sequences used for Gibson assembly cloning for the construction of P611 line.

753

754 **Supplementary Table S3.**

755 Information on the top 100 proteins that showed quantitative differences between control iron
756 starved *P. tricornutum* wild type cells and cells that were stimulated with gallium analogues
757 of ferrichrome.

758
759

760 **Supplementary Videos SV1- SV10**

761 Iron starved *P. tricornutum* P611 cells were spiked either with Fe′ (videos 1-5) or with FOB
762 (videos 6-10). In each case the cells were imaged on a Leica confocal microscope for 10
763 minutes continuously following the spike, according to protocol for image acquisition
764 outlined in the methods section of this manuscript. Fluorescence from YFP (video 5 and 10),
765 RFP (video 1 and 9), CFP (video 2 and 6), autofluorescence (video 3 and 7) was recorded, as
766 well as DIC (video 5 and 8).

767

768

769 **References**

770 **Abbe, E.** (1873) Beiträge zur Theorie des Mikroskops und der micrkoskopischen
771 Wahrnehmung. Arch. Mikrosk. Anat. **9,** 413–468.

772 **Achcar F, Camadro JM, Mestivier D**. 2009. AutoClass@ IJM: a powerful tool for Bayesian
773 classification of heterogeneous data in biology. Nucleic acids research **1**;37(suppl_2):W63-7.

774 **Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, Lopez PJ, Finazzi G,**
775 **Fernie AR, Bowler C.** 2008. Whole-cell response of the pennate diatom *Phaeodactylum*
776 *tricornutum* to iron starvation. Proceedings of the National Academy of Sciences
777 **105**(30):10438-10443.

778 **Anderson MA, Morel FM.** 1980. Uptake of Fe (II) by a diatom in oxic culture medium.
779 Marine Biology Letters **1**(5), 263-268.

780 **Behnke J, LaRoche J.** 2020. Starvation-Induced Proteins (ISIPs). European Journal of
781 Phycology **55(**3), 339-360.

782 **Bendtsen JD, Nielsen H, von Heijne G, Brunak S.** 2004. Improved prediction of signal
783 peptides: SignalP 3.0. Journal of molecular biology **340**(4):783-95.

784 **Bhattacharya D, Yoon HS, Hackett JD**. 2004. Photosynthetic eukaryotes unite:
785 endosymbiosis connects the dots. Bioessays **26**(1):50-60.

786 **Borowitzka MA, Volcani BE.** 1978. The polymorphic diatom *Phaeodactylum tricornutum*:
787 ultrastructure of ITS morphotypes. Journal of Phycology **14**(1):10-21.

788 **Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U,**
789 **Martens C, Maumus F, Otillar RP, Rayko E**. 2008. The *Phaeodactylum* genome reveals
790 the evolutionary history of diatom genomes. Nature **456** (7219):239.

791 **Boyd PW, Ellwood MJ.** 2010. The biogeochemical cycle of iron in the ocean. Nature
792 Geoscience **3**, 675–682.

793 **Burke DH, Hearst JE, Sidow A.** 1993. Early evolution of photosynthesis: clues from
794 nitrogenase and chlorophyll iron proteins. Proceedings of the National Academy of Sciences
795 **90**(15):7134-8.

796 **Caputi L, Carradec Q, Eveillard D, Kirilovsky A, Pelletier E, Pierella Karlusich JJ,**
797 **Rocha Jimenez Vieira F, Villar E, Chaffron S, Malviya S, Scalco E. el al.** 2020.
798 Community-level responses to iron availability in open ocean plankton ecosystems. Global
799 Biogeochemical Cycles **33**(3):391-419.

800 **Cavalier-Smith T**. 2003. Genomic reduction and evolution ofnovel genetic membranes and
801 protein-targeting machinery ineukaryote-eukaryote chimaeras (meta-algae). Philosophical
802 Transactions of the Royal Society of London B Biological Sciences **358**:109–133.

803 **Coale TH, Moosburner M, Horák A, Oborník M, Barbeau KA, Allen AE. 2019**.
804 Reduction-dependent siderophore assimilation in a model pennate diatom. Proceedings of the
805 National Academy of Sciences **116** (47), 23609-23617.
806 **Eckhardt U, Buckhout TJ.** 1998. Iron assimilation in *Chlamydomonas reinhardtii* involves
807 ferric reduction and is similar to Strategy I higher plants. Journal of Experimental Botany
808 **49**(324), 1219-1226.
809 **Gledhill M, Buck KN.** 2012. The organic complexation of iron in the marine environment: a
810 review. Frontiers in microbiology **3**, 69.
811 **Granum E, Raven JA, Leegood RC**. 2005. How do marine diatoms fix 10 billion tonnes of
812 inorganic carbon per year?Canadian journal of botany **83**(7), 898-908.
813 **Groussman RD, Parker MS, Armbrust EV.** 2016. Diversity and evolutionary history of
814 iron metabolism genes in diatoms. PLoS One **10**(6).
815 **Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG.** 2007. Protein
816 targeting into complex diatom plastids: functional characterisation of a specific targeting
817 motif. Plant Molecular Biology **64**:519–530.
818 **Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T**. 2015. Plastid proteome prediction
819 for diatoms and other algae with secondary plastids of the red lineage. The Plant Journal
820 **81**(3):519-28.
821 **Flori S, Jouneau PH, Finazzi G, Maréchal E, Falconet D.** 2016. Ultrastructure of the
822 periplastidial compartment of the diatom Phaeodactylum tricornutum. Protist **167**(3):254-67.
823 **Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips APR, Noddings CM, Brunson**
824 **JK Valas RE, Deerinck TJ, Jablanovic J, Gillard JT.** 2015. Designer diatom episomes
825 delivered by bacterial conjugation. Nature communications **6** (1): 1-10.
826 **Kazamia E, Sutak R, Paz-Yepes J, Dorrell RG, Vieira FRJ, Mach J, Morrissey J, Leon**
827 **S, Lam F, Pelletier E, Camadro JM, Bowler C, Lessuisse E.** 2018. Endocytosis-mediated
828 siderophore uptake as a strategy for Fe acquisition in diatoms. Science advances **4**(5),
829 p.eaar4536.
830 **Kilian O, Kroth PG**. 2005. Identification and characterization of a new conserved motif
831 within the presequence of proteins targeted into complex diatom plastids. Plant Journal
832 **41**:175–183.
833 **La Roche J, Boyd PW, McKay RML, Geider RJ.** 1996. Flavodoxin as an *in situ* marker for
834 iron stress in phytoplankton. Nature **382**:802–805.
835 **Lis H, Shaked Y, Kranzler C, Keren N, Morel FM**. 2015. Iron bioavailability to
836 phytoplankton: an empirical approach. The ISME journal **9**(4):1003-13.

837 **Lommer M, Roy AS, Schilhabel M, Schreiber S, Rosenstiel P, LaRoche J.** 2010. Recent
838 transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom
839 adapted to chronic iron limitation. BMC genomics **11**(1):718.
840 **Maher BA, Prospero JM, Mackie D, Gaiero D, Hesse PP, Balkanski Y.** 2010. Global
841 connections between aeolian dust, climate and ocean biogeochemistry at the present day and
842 at the last glacial maximum. Earth-Science Reviews **99**(1-2): 61-97.
843 **Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, Berthiaume CT,**
844 **Morales R, Allen AE, Armbrust VE.** 2012. Comparative metatranscriptomics identifies
845 molecular bases for the physiological responses of phytoplankton to varying iron availability.
846 Proceedings of the National Academy of Sciences **109**(6), E317-E325.
847 **McQuaid JB, Kustka AB, Oborník M, Horák A, McCrow JP, Karas BJ, Zheng H,**
848 **Kindeberg T, Andersson AJ, Barbeau KA, Allen AE.** 2018. Carbonate-sensitive
849 phytotransferrin controls high-affinity iron uptake in diatoms. Nature **555**(7697):534-5377.
850 **Moog D, Stork S, Zauner S, Maier UG**. 2011. *In silico* and *in vivo* investigations of proteins
851 of a minimized eukaryotic cyto-plasm. Genome Biology and Evolution **3**:375–382.
852 **Morel FM, Rueter JG, Anderson DM, Guillard RR.** 1979. Aquil: a chemically defined
853 phytoplankton culture medium for trace metal studies. Journal of Phycology, **15**(2):135-41.
854 **Morrissey J, Bowler C.** 2012. Iron utilization in marine cyanobacteria and eukaryotic algae.
855 Frontiers in microbiology **3**, 43.
856 **Morrissey J, Sutak R, Paz-Yepes J, Tanaka A, Moustafa A, Veluchamy A, Thomas Y,**
857 **Botebol H, Bouget FY, McQuaid JB, Tirichine L, Bowler C.** 2015. A novel protein,
858 ubiquitous in marine phytoplankton, concentrates iron at the cell surface and facilitates
859 uptake. Current Biology **25**(3), 364-371.
860 **Paz Y, Katz A, Pick U.** 2007. A multicopper ferroxidase involved in iron binding to
861 transferrins in *Dunaliella salina* plasma membranes. Journal of Biological Chemistry
862 **282**(12):8658-66.
863 **Peers G, Price NM. 2006.** Copper-containing plastocyanin used for electron transport by an
864 oceanic diatom. Nature **441**(7091), 341.
865 **Rastogi A, Maheswari U, Dorrell RG, Vieira FR, Maumus F, Kustka A, McCarthy J,**
866 **Allen AE, Kersey P, Bowler C, Tirichine L.** 2018. Integrative analysis of large-scale
867 transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome
868 and evolutionary origin of diatoms. Scientific reports **8**(1):1-4.

869 **Reggiori F, Tucker KA, Stromhaug PE, Klionsky DJ.** 2004. The Atg1-Atg13 complex
870 regulates Atg9 and Atg23 retrieval transport from the pre-autophagosomal structure.
871 Development of the Cell **6**, 79 –90.
872 **Reggiori F, Wang CW, Nair U, Shintani T, Abeliovich H, Klionsky DJ.** 2004. Early stages
873 of the secretory pathway, but not endosomes, are required for Cvt vesicle and autophagosome
874 assembly in *Saccharomyces cerevisiae*. Molecular Biology of the Cell **15**, 2189-204.
875 **Schallenberg C, Davidson AB, Simpson KG, Miller LA, Cullen JT.** 2015. Iron (II)
876 variability in the northeast subarctic Pacific Ocean. Marine Chemistry **177**, 33-44.
877 **Shaked Y, Kustka AB, Morel FM, Erel Y.** 2004. Simultaneous determination of iron
878 reduction and uptake by phytoplankton. Limnology and Oceanography: methods **2**(5), 137-
879 145.
880 **Smith SR, Gillard JT, Kustka AB, McCrow JP, Badger JH, Zheng H, New AM, Dupont**
881 **CL, Obata T, Fernie AR, Allen AE.** 2016. Transcriptional orchestration of the global
882 cellular response of a model pennate diatom to diel light cycling under iron limitation. PLoS
883 genetics **12**(12).
884 **Sutak R, Botebol H, Blaiseau PL, Léger T, Bouget FY, Camadro JM, Lesuisse E.** 2012.
885 A comparative study of iron uptake mechanisms in marine microalgae: Iron binding at the cell
886 surface is a critical step. Plant Physiology **160**, 2271–2284.
887 **Sutak R, Šlapeta J, Roman MS, Camadro JM, Lesuisse E.** 2010. Nonreductive iron uptake
888 mechanism in the marine alveolate Chromera velia. Plant Physiology **154**, (2010): 991-1000.
889 **Turnšek J, Brunson JK, Martinez Viedma M de P, Deerinck TJ, Horák A, Oborník M,**
890 **Bielinski VA, Allen AE** 2021. Proximity proteomics in a marine diatom reveals a putative
891 cell surface-to-chloroplast iron trafficking pathway. eLife **10**, e52770
892 **Véron B, Billard C, Dauguet J.-C, Hartmann M.-A.** 1996. Sterol composition of
893 *Phaeodactylum tricornutum* as influenced by growth temperature and light spectral quality.
894 Lipids, **31**(9), 989–994.
895 **Wang LH, Rothberg KG, Anderson RG**. 1993. Mis-assembly of clathrin lattices on
896 endosomes reveals a regulatory switch for coated pit formation. The Journal of cell biology
897 **123**(5):1107-17.
898 **Zuhorn IS, Kalicharan R, Hoekstra D.** 2002. Lipoplex-mediated transfection of
899 mammalian cells occurs through the cholesterol-dependent clathrin-mediated pathway of
900 endocytosis. Journal of Biological Chemistry **277**(20):18021-8.

Fig. 1

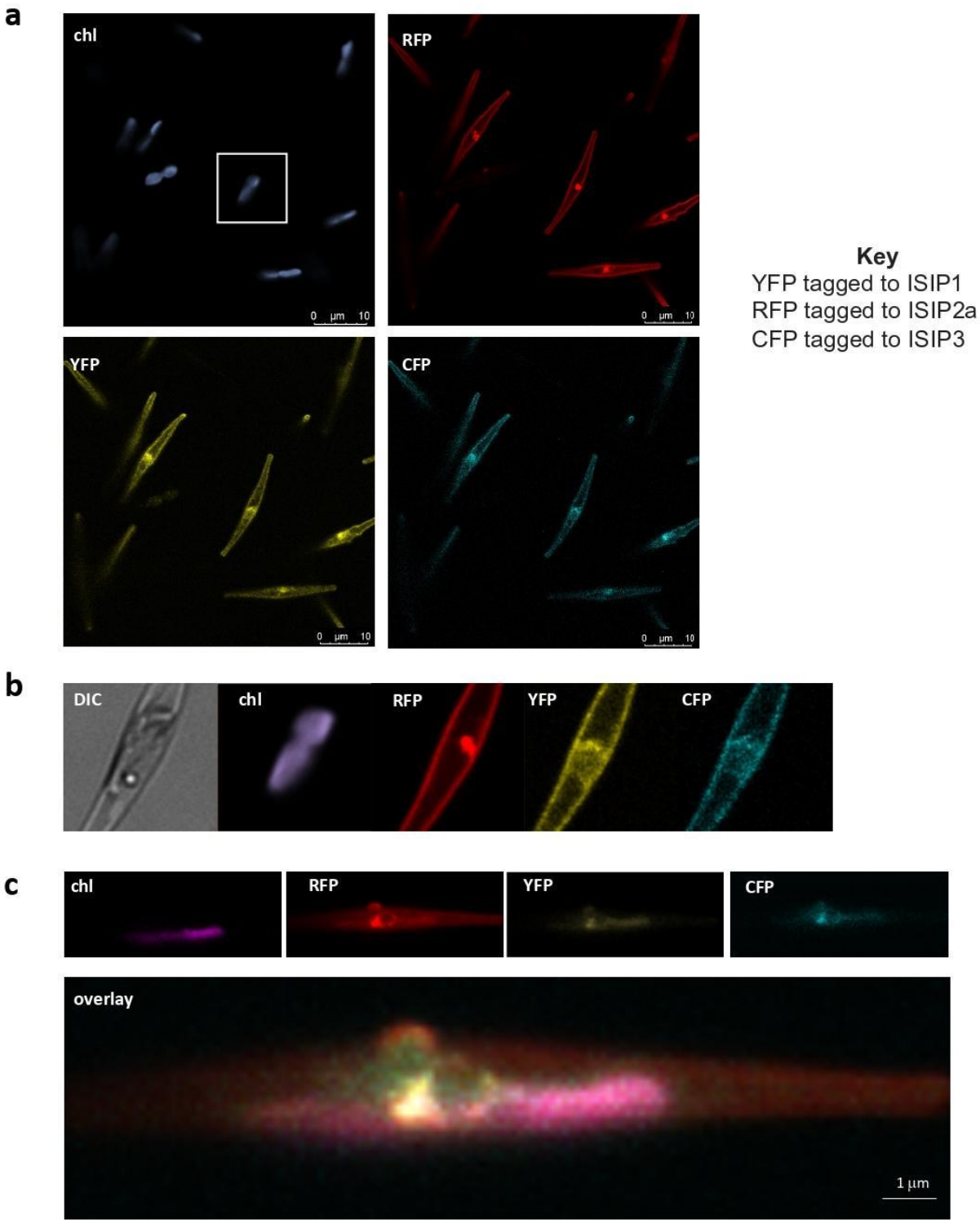

a
chl
RFP
YFP
CFP
0 µm 10
Key
YFP tagged to ISIP1
RFP tagged to ISIP2a
CFP tagged to ISIP3
b
DIC
chl
RFP
YFP
CFP
c
chl
RFP
YFP
CFP
overlay
1 µm

Fig. 2

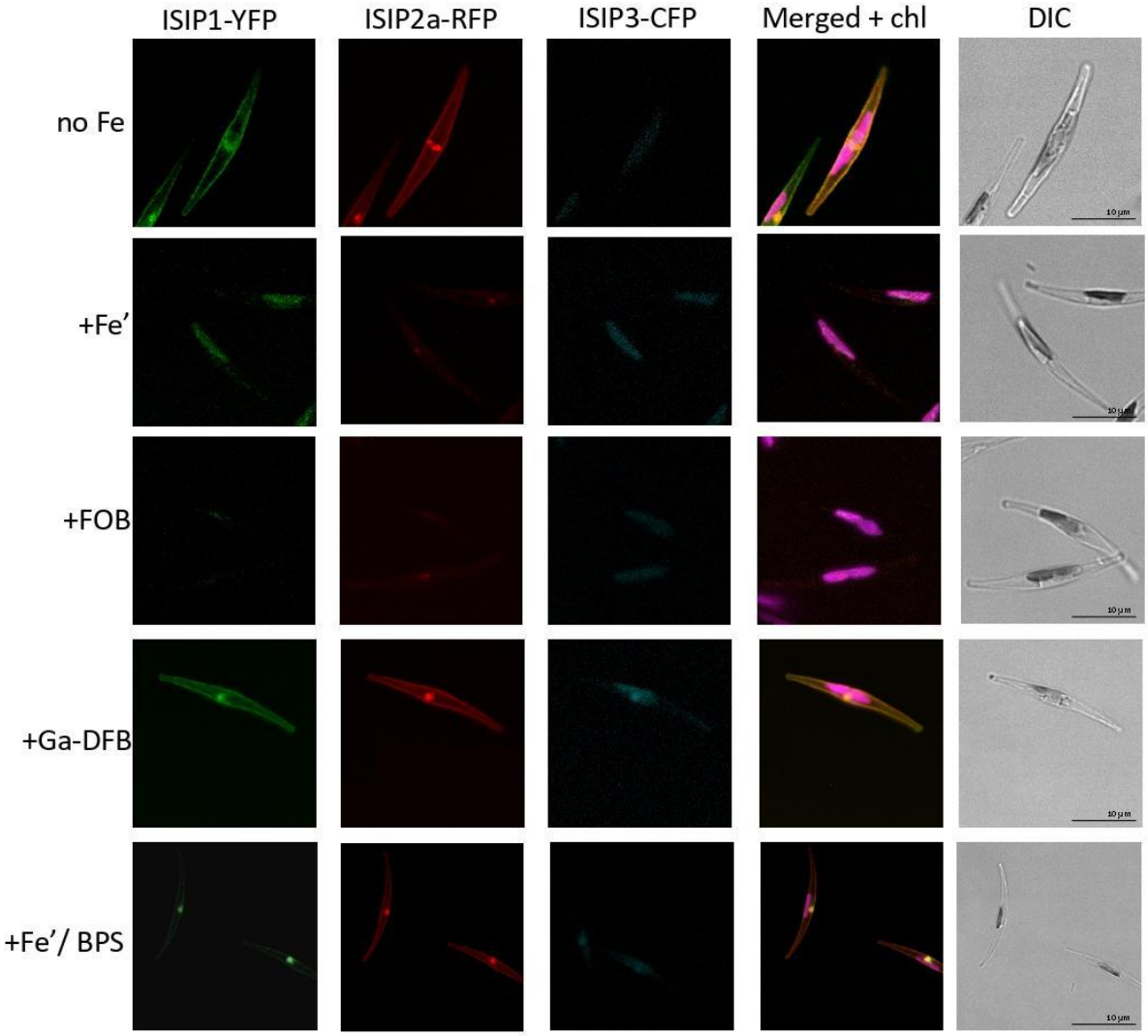
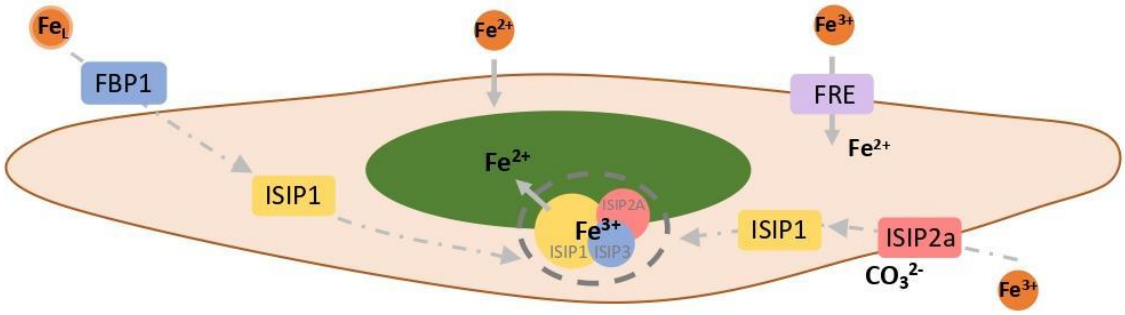

ISIP1-YFP
ISIP2a-RFP
ISIP3-CFP
Merged + chl
DIC
no Fe
+Fe'
+FOB
+Ga-DFB
+Fe'/ BPS

Fig. 3

a



b

pH 8.2

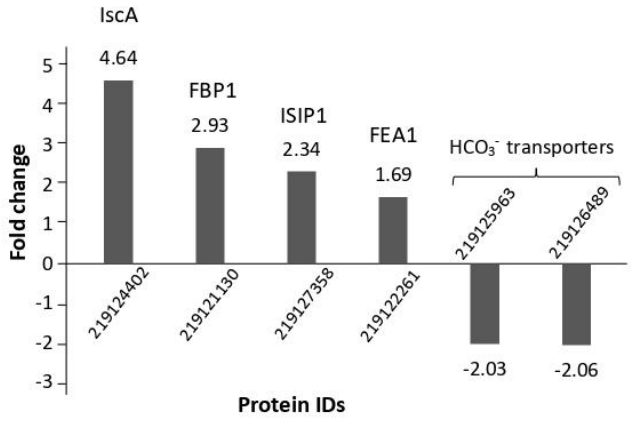
- high $[CO_3^{2-}]$
- high phytotransferrin-mediated Fe^{3+} uptake
- low siderophore-mediated Fe_L uptake
- low reductive uptake at cell surface?
- low direct Fe^{2+} uptake?
- intracellular reduction maintained by homeostasis

pH 7.8

- low $[CO_3^{2-}]$
- low phytotransferrin-mediated Fe^{3+} uptake
- high siderophore-mediated FeL uptake
- high reductive uptake at cell surface?
- high direct Fe^{2+} uptake?
- intracellular reduction maintained by homeostasis

Fig. 4

a **Siderophore processing proteins**



b **Endocytosis and membrane trafficking proteins**

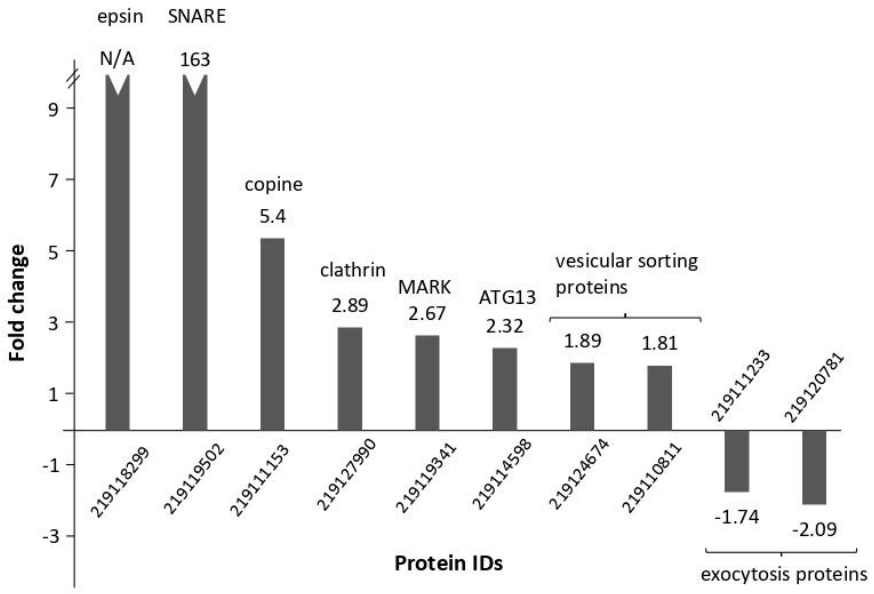


Fig. S1

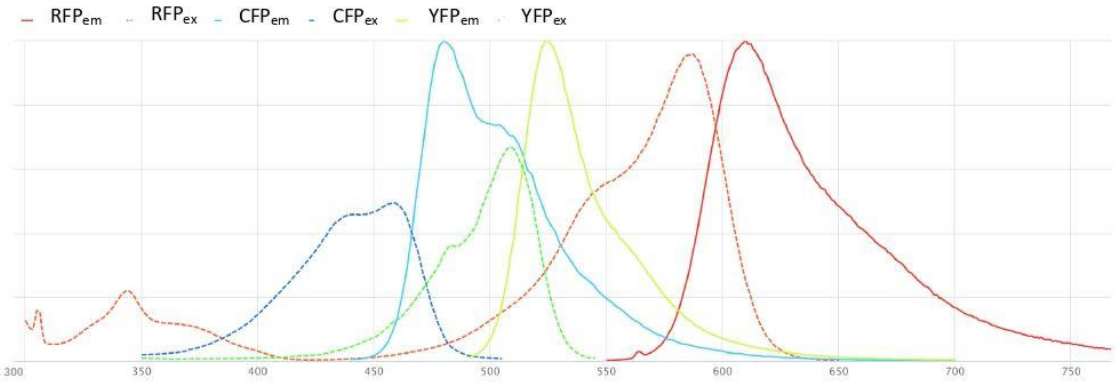

RFPem
RFPex
CFPem
CFPex
YFPem
YFPex
300
350
400
450
500
550
600
650
700
750

Fig. S2

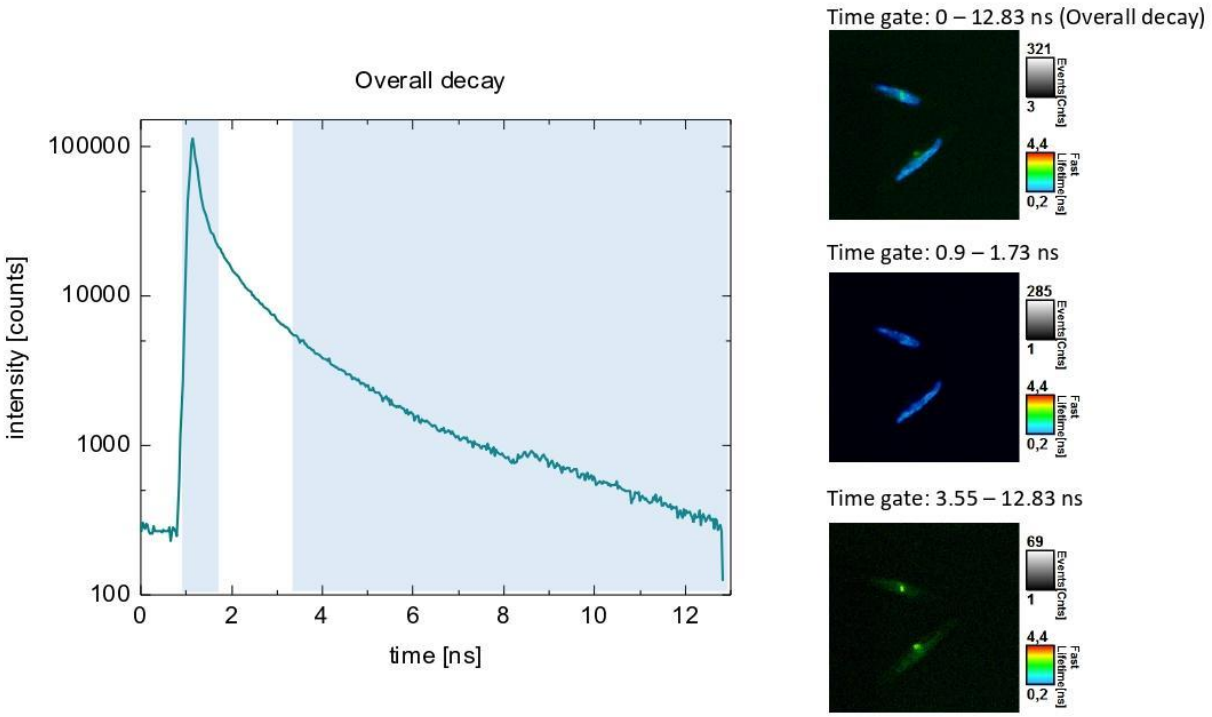

Overall decay
intensity [counts]
100000
10000
1000
100
0
2
4
6
8
10
12
time [ns]
Time gate: 0 – 12.83 ns (Overall decay)
321
Events[Cnts]
3
4,4
Fast Lifetime[ns]
0,2
Time gate: 0.9 – 1.73 ns
285
Events[Cnts]
1
4,4
Fast Lifetime[ns]
0,2
Time gate: 3.55 – 12.83 ns
69
Events[Cnts]
1
4,4
Fast Lifetime[ns]
0,2

Fig. S3

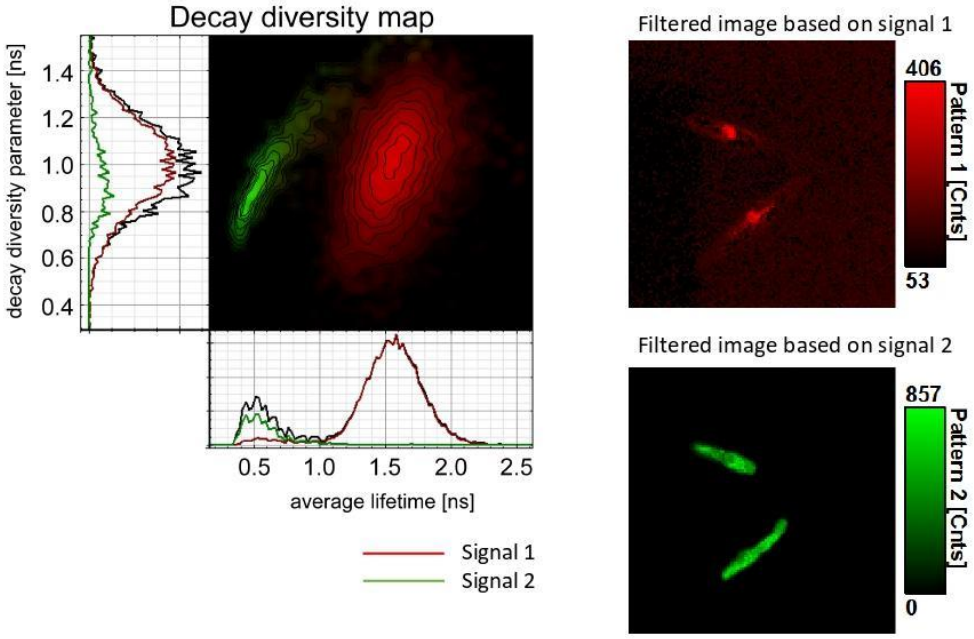
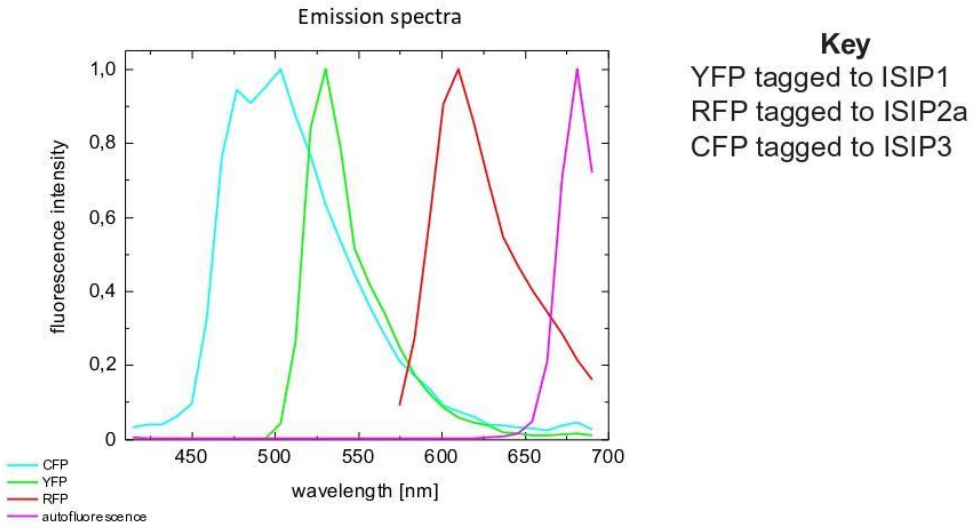

Decay diversity map
decay diversity parameter [ns]
1.4
1.2
1.0
0.8
0.6
0.4
0.5
1.0
1.5
2.0
2.5
average lifetime [ns]
Signal 1
Signal 2
Filtered image based on signal 1
406
Pattern 1 [Cnts]
53
Filtered image based on signal 2
857
Pattern 2 [Cnts]
0

Fig. S4

a


Emission spectra
fluorescence intensity
1,0
0,8
0,6
0,4
0,2
0
450
500
550
600
650
700
wavelength [nm]
CFP
YFP
RFP
autofluorescence
Key
YFP tagged to ISIP1
RFP tagged to ISIP2a
CFP tagged to ISIP3

b

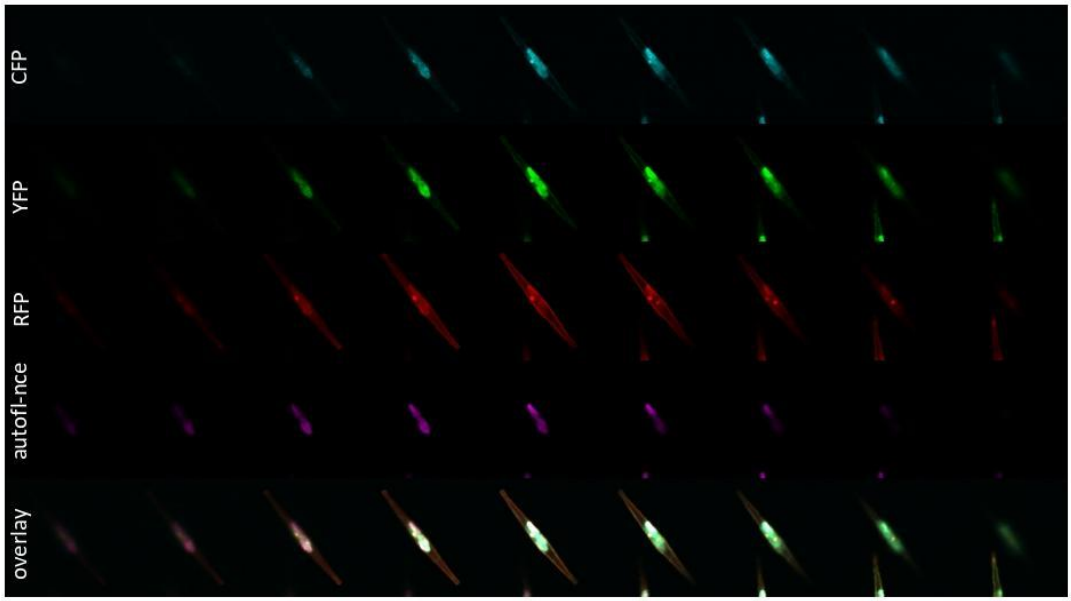
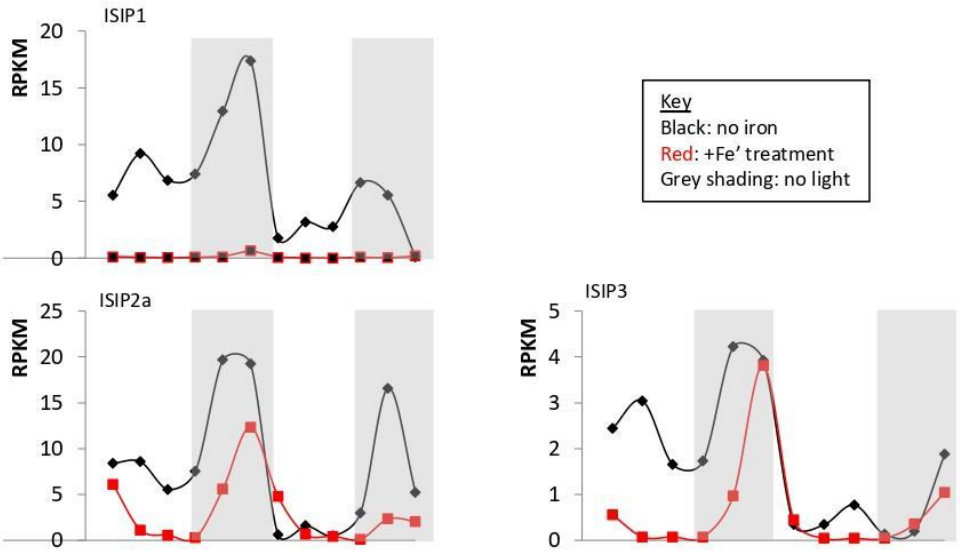

CFP
YFP
RFP
autofl-nce
overlay

Fig. S5


ISIP1
RPKM
20
15
10
5
0
ISIP2a
RPKM
25
20
15
10
5
0
ISIP3
RPKM
5
4
3
2
1
0
Key
Black: no iron
Red: +Fe' treatment
Grey shading: no light

Synthèse en français

Titre: Caractérisation des mécanismes d'acquisition et de stockage du fer chez les diatomées Phaeodactylum tricornutum

Mots clés: diatomées, physiologie du fer, stockage du fer, absorption du fer, ferritine, ISIP

Résumé: Les diatomées réussissent à survivre dans les océans contemporains où la teneur en fer est faible. Malgré leur taille cellulaire relativement grande, les diatomées tolèrent une carence en fer et sont généralement dominantes au cours des « bloom » de phytoplancton, stimulées par le fer de manière naturelles ou artificielles. Cependant, notre compréhension de la physiologie du fer chez les diatomées reste limitée et fait l'objet de recherches en cours. L'objectif du travail présenté est d'étudier le métabolisme du fer, en examinant en premier lieu les principales stratégies de son utilisation chez les diatomées. Cette partie s'appuie sur la littérature publiée et sur la complétion des données par une analyse de 82 transcriptomes. Par la suite, la génération des lignées fluorescentes transgéniques de la ferritine (FTN) et « Iron starvation induced protein 3 » (ISIP3), ont permis d'obtenir la localisation subcellulaire de ces protéines pour la première fois. Il a été montré que la protéine FTN se localise au niveau du chloroplaste, alors que la protéine ISIP3 se localise sous forme d'agrégats à proximité du chloroplaste. ISIP3 présente ainsi une localisation similaire avec la protéine ISIP1. En outre, les phénotypes des mutants de délétion de FTN et ISIP3 cultivés sous différentes conditions en fer ont été comparés. Les mutants FTN ont montré une croissance réduite dans toutes les conditions, indiquant une fonction dans le stockage du fer. Les mutants ISIP3, quant à eux, ont montré une croissance réduite lorsque le fer est disponible, mais aucune différence lorsqu'ils sont cultivés dans un milieu sans fer, suggérant ainsi que ISIP3 est nécessaire dans l'acquisition du fer. Lorsque le cuivre a été retiré du milieu, ces phénotypes ont disparu, confirmant ainsi la dépendance au cuivre pour l'absorption du fer. La diminution de la croissance des mutants ISIP3 en présence de sidérophore (FOB) suggère que ISIP3 est également impliqué dans leur absorption. Dans l'ensemble, les résultats présentés dans cette thèse ajoutent de nouvelles perspectives dans l'étude du métabolisme du fer chez les diatomées.

Aperçu de la Thèse

Considérant tout le phytoplancton en tant que groupe, la biodisponibilité du fer a considérablement diminué dans l'océan contemporain par rapport aux maxima précédents. Pourtant, les diatomées sont apparues comme une jeune ramification évolutive lorsque le fer était déjà déficient dans le vaste

océan. Ils ont survécu avec succès à une faible teneur en fer et ont supplanté d'autres espèces lorsque le fer est devenu disponible, ce qui a été confirmé par les expériences de fertilisation en fer à mésoéchelle. Dans cette thèse, la question posée est : quelles sont les stratégies des diatomées pour obtenir le fer et l'utiliser efficacement à l'intérieur de la cellule ? Pour répondre à cette question, j'ai d'abord passé en revue les études sur le métabolisme du fer chez les diatomées, en combinant différentes analyses bioinformatiques. Ensuite, j'ai mené des expériences physiologiques comparatives sur des souches knock-out.

Dans le chapitre 2, je présente les résultats d'une revue publiée dans le Journal of Experimental Botany (Gao et al., 2021), où, avec des collègues, j'ai résumé les stratégies du métabolisme du fer chez les diatomées en m'appuyant sur les preuves transcriptionnelles et physiologiques de la Littérature.

Dans ce chapitre, j'ai recherché davantage les gènes liés au fer par le biais d'analyses bioinformatiques et je les ai résumés dans une carte thermique montrant la distribution des gènes dans différentes espèces. Les séquences de FTN de P. tricornutum et d'autres espèces de diatomées ont été comparées par alignement et arbre phylogénétique. Dans l'ensemble, ce chapitre fournit de nouvelles perspectives sur le sujet du métabolisme du fer chez les diatomées.

Dans le chapitre 3, visant à étudier la localisation des ISIP et des FTN, j'ai marqué FTN sur des protéines fluorescentes (FP) et visualisé des cellules de diatomées transgéniques par microscopie confocale. ISIP3 et FTN localisés à proximité du chloroplaste mais avec des motifs différents.

Dans le chapitre 4, j'étudie plus en détail la fonction d'ISIP3 et de FTN chez P. tricornutum. J'ai généré des mutants knock-out (KO) et comparé leurs courbes de croissance dans différentes conditions de croissance. L'absence du gène a eu un effet sur la croissance des cellules dans différentes conditions, indiquant une implication dans des aspects spécifiques de la physiologie du fer.

Dans le chapitre 5, les conclusions et perspectives ont été synthétisées.