Unraveling the unique chloroplast transporter architecture of diatom algae Phaeodactylum tricornutum : insights from computational and experimental analyses

Shun Liu

université PARIS-SACLAY

Unraveling the unique chloroplast transporter architecture of diatom algae *Phaeodactylum tricornutum* : insights from computational and experimental analyses

Démêler l'architecture unique du transporteur de chloroplastes des algues diatomées Phaeodactylum tricornutum : aperçu des analyses informatiques et expérimentales

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°567 : sciences du végétal : du gène à l'écosystème (SEVE)
Spécialité de doctorat : Biologie
Graduate School : BioSphERA. Référent : Faculté des sciences d'Orsay

Thèse préparée à **l'Institut de Biologie de l'École Normale Supérieure (ENS Paris, CNRS, INSERM, PSL)**, sous la direction de **Chris BOWLER**, Directeur de recherche, et la co-encadrement de **Richard DORRELL**, Chargé de recherche

Thèse soutenue à Paris, le 06 Juillet 2023, par

Shun LIU

Composition du Jury

Membres du jury avec voix délibérative

Catherine de VITRY Directrice de recherche, CNRS - Sorbonne Université	Présidente
Benoît SCHOEFS Professeur, Université Le Mans	Rapporteur & Examinateur
Ellen NISBET Professeur assistant, Université de Nottingham	Rapporteure & Examinatrice
Eric MARÉCHAL Directeur de recherche, CNRS - Université Grenoble-Alpes	Examinateur

THÈSE DE DOCTORAT

NNT : 2023UPASB039

université PARIS-SACLAY

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

Titre : Démêler l'architecture unique du transporteur de chloroplastes des algues diatomées *Phaeodactylum tricornutum*: aperçu des analyses informatiques et expérimentales

Mots clés : *Phaeodactylum*, Transcriptomique, Photosynthèse, CRISPR-Cas9, auto-assemblé GFP, chloroplaste

Les diatomées sont des contributeurs clés à la photosynthèse marine. Elles sont distantes des plantes et possèdent des chloroplastes (plastes) d'origine rouge secondaire, distincts des plastes primaires des plantes, avec un métabolisme associé mal compris. Dans cette étude, nous utilisons des données transcriptomiques, phylogénétiques et environnementales pour construire un "atlas" *in silico* des transporteurs de plastes prédits, identifiés par des recherches génomiques à l'échelle du génome de l'espèce modèle de diatomée *Phaeodactylum tricornutum*. Notre analyse a conduit à la découverte d'un nouveau transporteur de la superfamille des transporteurs majeurs (MFS), désigné par la suite "Trans3". Ce gène nucléaire codant le transporteur n'a été trouvé que chez les espèces contenant des plastes rouges secondaires, et a montré une forte corrélation positive d'expression avec les gènes codés dans le génome mitochondrial de *P. tricornutum*. En utilisant la GFP auto-assemblante, nous avons localisé Trans3 à la membrane péri-plastidiale du plastid. Nous avons généré trois mutants probables hétérozygotes de Trans3 par CRISPR-Cas9 qui ont montré une réduction de la croissance cellulaire et une teneur en photosystème I P700 plus faible que les lignées témoins, ce qui implique qu'il joue des rôles importants dans la croissance cellulaire et l'efficacité photosynthétique des diatomées. Fait intéressant, nos lignées mutantes ont montré une sensibilité moindre au stress de déplétion du fer (Fe) que les contrôles avec vecteur vide, ce que nous inférons des données transcriptomiques comme étant lié à une surexpression de la synthèse des pigments et des voies de complexe de capture de la lumière, ce qui pourrait atténuer la photosensibilité du mutant causée par une déplétion du Fe. Enfin, en considérant des données structurales, nous avons suggéré que Trans3 fonctionne en tant que transporteur de molybdate similaire à la protéine de transporteur de molybdate 2 (MOT2) trouvée dans l'algue verte modèle *Chlamydomonas reinhardtii*, ce qui pourrait participer à l'importation de molybdate à travers les membranes plastidiales. Des mesures préliminaires indiquent que les mutants Trans3 avaient des taux d'absorption de molybdate plus élevés que les contrôles avec vecteur vide, bien que cela puisse être affecté par la disponibilité du nitrate (NO_3^-) et du fer dans le milieu de croissance. Dans l'ensemble, ces données fournissent de nouvelles perspectives pour comprendre le système complexe de transport dans les plastes des diatomées et leur contribution au succès remarquable de ces organismes.

université PARIS-SACLAY

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

Title : Unraveling the unique chloroplast transporter architecture of diatom algae Phaeodactylum tricornutum: insights from computational and experimental analyses

Keywords : *Phaeodactylum*; Transcriptomics; Photosynthesis; CRISPR-Cas9; Self-assembling GFP; Chloroplast

Diatoms are a key contributor to marine photosynthesis. Diatoms are distantly related to plants, and possess chloroplasts (plastid) of secondary red origin distinct from the primary plastids of plants, with a poorly understood associated metabolism. Here, we use transcriptomic, phylogenomic and environmental data to construct an *in silico* atlas of predicted plastid transporters identified by genome-wide searches of the model diatom species *Phaeodactylum tricornutum*. Our analysis led to the discovery of a novel major facilitator superfamily (MFS) transporter, henceforth termed "Trans3". This nuclear gene encoding transporter was only found in species containing secondary red plastids, and showed a strong positive correlated expression with genes encoded in the *P. tricornutum* mitochondrial genome. Using self-assembling GFP, we located Trans3 at the plastid peri-plastidial membrane (PPM). We generated three probable heterozygotic Trans3 CRISPR-Cas9 mutants which showed cell growth reduction and lower photosystem I P700 content than control lines, implying that it plays important roles for diatom cell growth and photosynthetic efficiency. Interestingly, our mutant lines showed less sensitivity to iron (Fe) depletion stress than empty vector controls, which we infer from transcriptomic data to relate to an upregulation of expression of pigment synthesis and light-harvesting complex pathways that may mitigate mutant photo-sensitivity to Fe depletion. Finally, considering structural data, we suggested Trans3 functions as a molybdate transporter similar to the molybdate transporter 2 (MOT2) protein found in the model green alga *Chlamydomonas reinhardtii*, which may participate in molybdate import through the plastid membranes. Preliminary measurements indicate that Trans3 mutants had higher molybdate uptake rates than empty vector controls, although this may be affected by the availability of nitrate (NO_3^-) and Fe in growth media. Overall, these data provide new insights into understanding the complex transport system in diatom plastids and their contribution to the striking success of diatoms.

Synthèse en français

50% de l'oxygène que nous respirons est produit dans l'océan. Il est synthétisé par les algues, qui soutiennent tous les réseaux alimentaires marins et sont des indicateurs probables des impacts du changement climatique d'origine humaine. Les diatomées, en tant qu'un groupe de phytoplancton dominant dans l'océan, jouent un rôle vital dans la photosynthèse marine. Ces organismes, distants des plantes, possèdent un plaste d'origine rouge secondaire, entouré d'une structure membranaire complexe avec un système de transport et un métabolisme associé mal compris. Cette thèse visait à éclairer les transporteurs énigmatiques associés aux plastes des diatomées.

Tout d'abord, une analyse in silico est effectuée pour profiler les fonctions probables et les voies centrales influencées de 70 transporteurs ciblant les plastes présumés, en tenant compte de la similarité phylogénétique avec les transporteurs de chloroplastes de la plante modèle *Arabidopsis thaliana*, de la co-régulation transcriptionnelle avec les génomes mitochondriaux, des histoires évolutives présumées déduites par la phylogénie d'un seul gène, et de l'expression environnementale plus large des homologues identifiés dans l'enquête *Tara* Oceans. Nous profilons la diversité biologique, métabolique et éco-physiologique des transporteurs de plastes des diatomées, et identifions de nouveaux transporteurs candidats pour des activités métaboliques spécifiques des plastes de diatomées et pour la résilience environnementale des diatomées dans l'océan moderne.

Sur la base de l'analyse, de nombreux transporteurs ciblant les plastes des diatomées avec des fonctions incertaines appartiennent à la superfamille des transporteurs majeurs (MFS). Cette superfamille code principalement pour des protéines de transport membranaire et est exprimée dans tous les règnes du vivant. Il a été rapporté que les MFS ont une grande diversité de substrats, tels que les acides organiques, les ions et les métabolites. Découvrir la fonction, la localisation et le substrat de ces transporteurs MFS sous-estimés et peu étudiés nous permettra de mieux comprendre le système de transport complexe et les voies dans le plastide des diatomées.

Grâce à notre jeu de données, une attention particulière a été portée sur un nouveau transporteur MFS encodé par le génome nucléaire, nommé Trans3, qui a été sélectionné comme cible pour de nouvelles recherches dans ma thèse. Les homologues de BLAST à l'envers (*reciprocal BLAST best-hit,* RbH) de Trans3 sont uniques aux espèces avec des plastes rouges secondaires. Les données des puces à ARN révèlent une co-régulation transcriptionnelle probablement forte de Trans3 avec les gènes codés dans le génome mitochondrial de *Phaeodactylum*.

La localisation subcellulaire de Trans3 est étudiée en utilisant la GFP auto-assemblante. Les résultats suggèrent que Trans3 se localise à la membrane péri-plastidiale (PPM). À ma connaissance, il s'agit du premier transporteur de plastide localisé à la PPM confirmé par la

GFP auto-assemblante. Cette découverte est cohérente avec le contexte évolutif des plastides secondaires et met en évidence l'utilité de la GFP auto-assemblante pour l'étude de la localisation des transporteurs.

Ensuite, la caractérisation moléculaire et fonctionnelle de Trans3 est poursuivie. La signification fonctionnelle de Trans3 a été testée en générant des lignées mutantes de *Phaeodactylum* partielles hétérozygotes par CRISPR-Cas9. La suppression partielle de l'expression de Trans3 entraîne une réduction de la croissance cellulaire et de la teneur en photosystème I P700 dans des milieux riches en fer et en nitrate. De manière intéressante, les lignées partielles des mutants de Trans3 démontrent une sensibilité réduite à la déplétion du fer, avec une diminution plus faible du taux de croissance relative et de la teneur en PSI P700. Les données transcriptomiques suggèrent une surexpression de la synthèse des pigments et des voies liées au complexe de capture de la lumière (LHCs) et aux photosystèmes, ce qui contribue potentiellement à ce phénotype observé.

Enfin, je propose Trans3 comme un potentiel transporteur de molybdate (MOT) dans les plastides de diatomées, sur la base de données structurales et biochimiques, indiquant une similarité avec la fonction confirmée du transporteur de molybdate 2 (MOT2) de l'algue verte *Chlamydomonas reinhardtii*, des motifs conservés et des la structure 3D prédite par le logiciel *AlphaFold*. L'expression du gène Trans3 est induite par des concentrations élevées de molybdate, et les lignées KO montrent une augmentation de l'absorption de molybdate, ce qui soutient son rôle potentiel dans le transport de molybdate.

En conclusion, cette thèse offre une compréhension approfondie d'un nouveau transporteur de la superfamille MFS, Trans3, dans les plastes de diatomées. Elle met en évidence la localisation de Trans3, son importance fonctionnelle dans l'efficacité photosynthétique et la croissance cellulaire, ainsi que son rôle potentiel en tant que transporteur de molybdate. Ces résultats contribuent à notre compréhension du métabolisme chloroplastique des diatomées et fournissent de nouvelles perspectives pour comprendre les transporteurs uniques associées avec leurs organites.

Acknowledgements

SPONSORS
Chinese Scholarship Council

SUPERVISORS
Chris Bowler, Richard Dorrell

COLLEAGUES
Clara Bourbousse, Martine Boccara, Maria Helena Cruz de Carvalho, Fredy Barneche, Priscillia Pierre-Elies, Aude Battistella, Fabien Richard, Anna Novák Vanclová, Nathalie Joli, Charlotte Nef, Matteo Scarsini, Xia Gao, Juliette Laude, Florent Charton, Matthias Penot, Pauline Clémente, Morgane Roquais, Tomomi Nonoyama, Amandine Baylet, and all colleagues in IBENS that helped

COLLABORATORS
Mattia Storti (CEA, Grenoble, France)
Giovanni Finazzi (CEA, Grenoble, France)

I am grateful for the opportunity to study in IBENS and delve into the fields of molecular biology. Through the scientific and personal experiences gained during my research journey, I have shaped my understanding of marine science, and have developed strength, confidence, and knowledge, all of which have equipped me to navigate the vast realm of scientific inquiry in the future. Therefore, it is imperative that I thank the individuals who have contributed to the successful completion of my thesis.

I express my appreciation to the Chinese Scholarship Council for the funding that enabled me to pursue my research. Additionally, I would like to acknowledge the University Paris-Saclay for offering me the opportunity to undertake this doctoral program cooperation.

I extend my deepest gratitude to my supervisors, Chris Bowler and Richard Dorrell, for their invaluable guidance, thoughtful comments, and recommendations throughout the development of my project and the thesis. Their expertise and unwavering support have played a pivotal role in ensuring the direction and achievement of this study. I am truly grateful for their mentorship.

I express my heartfelt gratitude for the contributions of our engineer Priscillia Pierre-Elies, Aude Battistella, and Fabien Richard, for their efforts in providing the necessary materials for molecular biology, cloning, and *Phaeodactylum* transformation. I would also like to extend my thanks to all the other laboratories within our institute. Their assistance and support have been invaluable throughout my four years journey in Paris. I also want to thank the stagiaires Pauline Clémente and Morgane Roquais for their assistance in conducting the processing experiments, as well as to Tomomi Nonoyama and Amandine Baylet for their foundational contributions to this project.

I particularly appreciate our collaborators, Giovanni Finazzi and his post-doc, Mattia Storti, from EA Grenoble, Cell & Plant Physiology Laboratory, for their assistance in measuring photosynthesis parameters. Their expertise and contributions have enhanced the quality and depth of my research.

I also extend my appreciation to my committee members, Jacqui Shykoff, Olivier Vallon, and Clara Bourbousse, for their valuable advice and guidance.

I would also like to thank Daniel Moog from the University of Marburg for generously providing the pPha_2NR plasmid used for self-assembling GFP. Moreover, I am grateful to Astou Tangara and Benjamin Mathieu (IBENS) for their training in confocal microscopy, Sandrine Adiba and Silvis de Monte (IBENS) for help with flow cytometry, as well as to Angela Falciatore and Marianne Jaubert (IBPC) for offering the use of the biolistic device. I also appreciated Magali Charvin and Fredy Barneche for help with plant and algal growth facilities.

Finally, I am profoundly grateful to my family and friends for their unwavering and unconditional support throughout these intense academic years. Their encouragement, understanding, and love have been my pillars of strength, and I am fortunate to have them by my side.

Table of Contents

Abbreviations

AO: Aldehyde Oxidase
Bip: Binding Immunoglobulin Protein
BLAST: Basic Local Alignment Search Tool
BTS: Bipartite Topogenic signal Sequence
CDS: protein Coding Sequence
cERM: chloroplast Endoplasmic Reticulum Membrane
cER: chloroplast Endoplasmic Reticulum
cPMP: cyclic pyranopterin monophosphate
Chl a: Chlorophyll a
Chl c: Chlorophyll c
Ddx/Dtx: Diadinoxanthin/Diatoxanthin
DEG: Differentially Expressed Gene
ER: Endoplasmic Reticulum
FCM: Flow Cytometer
FCP: Fucoxanthin-Chlorophyll Protein
Fe-: Iron Depletion
Fe+: Iron Replete control culture condition
FeL: Iron Depletion, Long time (14 days)
FeM: Iron Depletion, Medium time (7 days)
FeS: Iron Depletion, Short time (3 days)
Fx: Fucoxanthin
GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase protein
GFP: Green Fluorescent Protein
Glu-Gln: glutamate-glutamine
HA: empty vector control (containing Human Antigen Cas9 sequence)
HSP70: Heat Shock Protein 70
IEM: the Inner Envelope Membrane
IMS: Intermembrane Space
KO: Knockout
LHC: Light Harvesting Complex
MFS: Major Facilitator Superfamily
MGD1: Monogalactosyl Diacylglycerol Synthase 1
Mo: Molybdenum
MoCo: Molybdenum Cofactor

MO: Molybdate Oxyanions MoO_4^{-2}
MOSC: MoCo Sulfurase C-terminal domain
MOT: Molybdate Transporter
MPT: molybdopterin
mPBS: marine Phosphate-Buffered Saline
N: Nitrogen
NH_4^+: Ammonium
NO_3^-: Nitrate
NR: Nitrate Reductase
NTD: N-terminal Domain
NTT: Nucleotide Triphosphate Transporters
OEM: the Outer Plastid Membrane
ONE: Outer Nuclear Envelope
PPC: Peri-Plastidial Compartment
PPM: the Peri-Plastidial Membrane
qPCR: Quantitative Polymerase Chain Reaction
RbH: Reciprocal BLAST best Hit
SAM: S-Adenosylmethionine
SELMA: Symbiont-specific ERAD-Like Machinery
SO: Sulfite Oxidase
SP: Signal Peptide
TMD: Trans-Membrane Domain
TOM70: mitochondrial Translocase Outer Membrane 70 kDa subunit
TP: Transit Peptide-like sequence
TPM: Transcript Per Million
TPT: Triose Phosphate Transporter
WT: Wild Type
XDH: Xanthine Dehydrogenase

Chapter 1: Introduction

50% of the oxygen that we breathe is produced in aquatic habitats, principally the ocean (Chapman, 2013). This is synthesized by algae, which support all marine food webs and are probable bellwethers for the impacts of anthropogenic climate change (Smol et al., 2012). These algae are very different in evolutionary terms to plants, ranging from tiny cyanobacteria to kelps hundreds of meters long, and are spread across the tree of life (Andersen, 1992). They are however unified by the capacity to do oxygenic photosynthesis, either in thylakoid membranes (cyanobacteria), or in specialist organelles called chloroplasts (also, "plastids") (Dorrell et al., 2017; Falciatore et al., 2020). These chloroplasts have originated through several independent endosymbiosis, starting with a single primary endosymbiosis in a common ancestor of the "archaeplastid" group that includes plants, green and red algae. That said, the majority of algal groups in the oceans today, such as diatoms, haptophytes and dinoflagellates, possess chloroplasts derived from the secondary or higher endosymbiosis of a red alga (Dorrell et al., 2017; Falciatore et al., 2020). In this event, an entire eukaryotic red algal cell including its nucleus and its chloroplasts was taken up by another eukaryote, and converted into a stable organelle (Falciatore et al., 2020). In my PhD, I focus on one group of algae, diatoms, which possess secondary red chloroplasts, and belong to a group called the stramenopiles that are distantly related to plants (and to us). In this chapter, I will introduce diatoms and their plastids considering their evolution, structure and innovative metabolism.

1.1 Diatoms are a major class of phytoplankton that contribute to marine photosynthesis

1.1.1 Diatoms: dominant and diversified photosynthetic marine organisms

Diatoms, a group of eukaryotic phytoplankton, are incredibly successful in ecology, as evidenced by their abundance in the contemporary ocean (Benoiston et al., 2017; Vincent and Bowler, 2020), as well as freshwater and terrestrial environments (Liu et al., 2013; Singer et al., 2021). As the dominant and group of photosynthetic producers in the ocean, diatoms are responsible for presumed approximately provisional 20% of total planetary net primary

production and nearly 40% of marine photosynthesis (Nelson et al., 1995; Field et al., 1998; Vincent and Bowler, 2020).

Diatoms are found in all ocean provinces, with a particular prevalence in nutrient-rich coastal upwelling regions and at high latitudes (Tréguer et al., 2017). Diatoms are the most prolific photosynthetic producers in both high northern and southern latitude (Figure 1.1), e.g., the north Pacific/Atlantic and Southern Oceans (Malviya et al., 2016). These environments are characterized by abundant nutrients, due to low stratification of the water column, but are typically iron (Fe) limited (particularly in the case of the Southern Ocean), as Fe is highly insoluble and is typically only present in the oceans if it is deposited from land sources (e.g., deserts). Notably, diatoms are considered as effective competitors for iron, with efficient uptake and storage mechanisms, which can let diatoms survive under low iron conditions and dominate iron-stimulated blooms (Boyd et al., 2007). Under low iron conditions, diatoms can rely on stored cell iron and high affinity iron uptake to allow them to capitalize on other abundant nutrients (nitrogen, phosphorus). Equally, diatoms can use these same strategies to rapidly take up iron from the water, forming blooms in response to transient fertilization events (Aeolian deposition, and upwelling) and competitively exclude other species (Lampe et al., 2021). However, efficient Fe utilization and storage is not the only cellular strategy that leads to the distribution of diatom in the present-day Ocean, since the tropical Pacific is often more Fe-limited than high latitudes (Caputi et al., 2019; Ustick et al., 2021). Other factors also contribute to diatom success at high latitudes, e.g., efficient management of photosynthesis at low temperature and under continuous illumination (Dorrell et al., 2022).

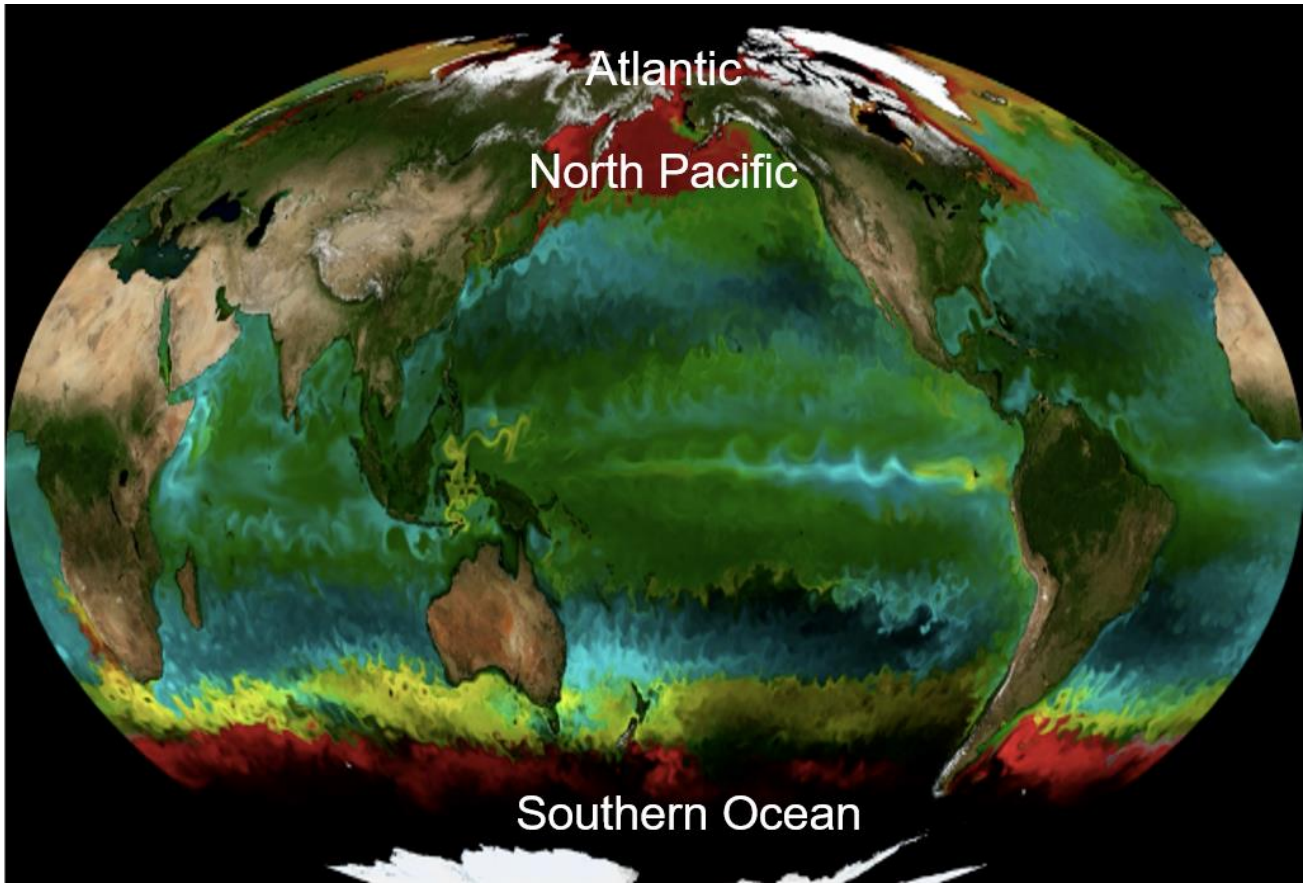

Figure 1.1 Model phytoplankton global distribution in the Ocean. Diatoms are represented in red; other eukaryotic flagellates (haptophytes, dinoflagellates, prasinophytes) are represented in yellow; and cyan and green represent different cyanobacterial groups (respectively, *Synechococcus* and *Prochlorococcus*). High latitude oceans in which diatoms show particular evolutionary prominence are labelled. This figure incorporates a screengrab of a dynamic ocean systems model obtained from the NASA website: https://svs.gsfc.nasa.gov/30669.

Diatom species are diverse, with estimated numbers ranging from 40,000 (based on molecular markers) to 200,000 (based on morphology) (Benoiston et al., 2017). According to their different symmetry characters (Alverson et al., 2006), diatoms are subclassified into 'centric' species, which include two suborders (radial and polar), and 'pennate' species, and are further divided into two suborders (raphid and araphid) (Figure 1.2). In comparison to other oceanic unicellular eukaryotes, many diatom cells are relatively large, with diameters typically exceeding 20 µm and maximum cell sizes recorded up to 2 mm (Bhatta et al., 2009), although nano-scale diatoms are also known and have important environmental roles (Leblanc et al., 2018). Diatom cells are typically silicified,

resulting in frustules with nano-scale pores ranging from 250 to 600 nm in size. These frustules are generally not believed to interfere with diatom nutrient uptake (Bhatta et al., 2009), although also imposes a silicate nutrient requirement that may limit diatom environmental distributions (Tréguer et al., 2018).

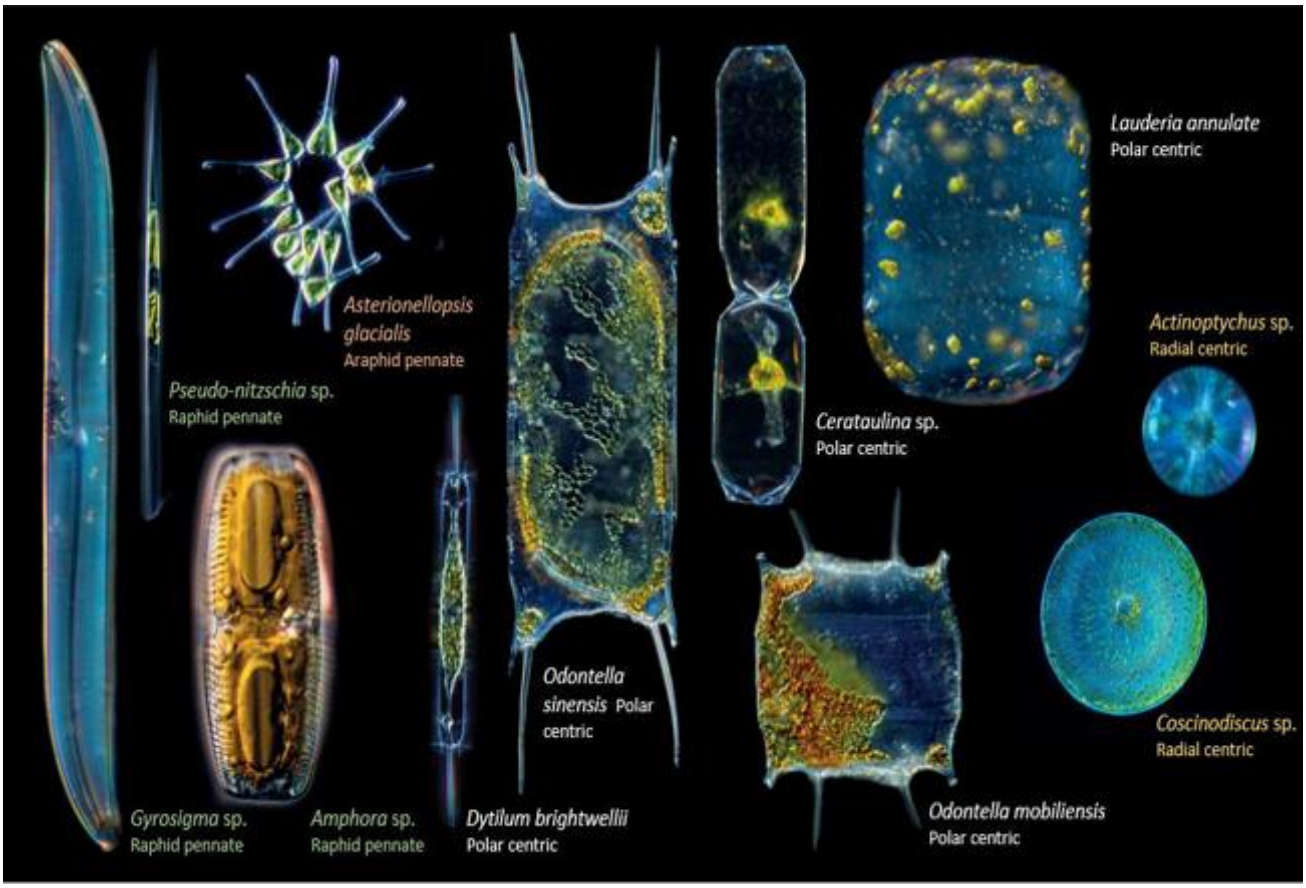

Figure 1.2 Cells shape diversity in diatom species. The figure is cited from Gao et al. (2021). Light microscope images showed different diatom species belonging to suborders as: raphid pennate (on the left) and araphid pennate (upper left), polar centric (middle) and radial centric (right).

1.1.2 Diatoms have a unique evolutionary history that defines their cell biology

The evolutionary context of diatoms within the eukaryotic tree of life is shown in Figure 1.3. As represented in the picture, diatoms belong to a group termed the stramenopiles (also "heterokonts", which contain multiple other photosynthetic groups (e.g., chrysophytes and kelps) and non-photosynthetic members (e.g., the plant pathogens oomycetes). The stramenopiles are distantly

related to land plants, animals and fungi, and indeed to many other eukaryotic algal groups (e.g., primary red and green algae), belonging instead to the "SAR" clade of Stramenopiles, Alveolates and Rhizaria (Figure 1.3) (Dorrell and Smith, 2011). The SAR Clade is ancient, extending at least into the pre-Cambrian, but diatoms themselves have a relatively recent evolutionary origin, with fossil dates indicating that they originated about 250 million years ago (Sims et al., 2006; Strassert et al., 2021).

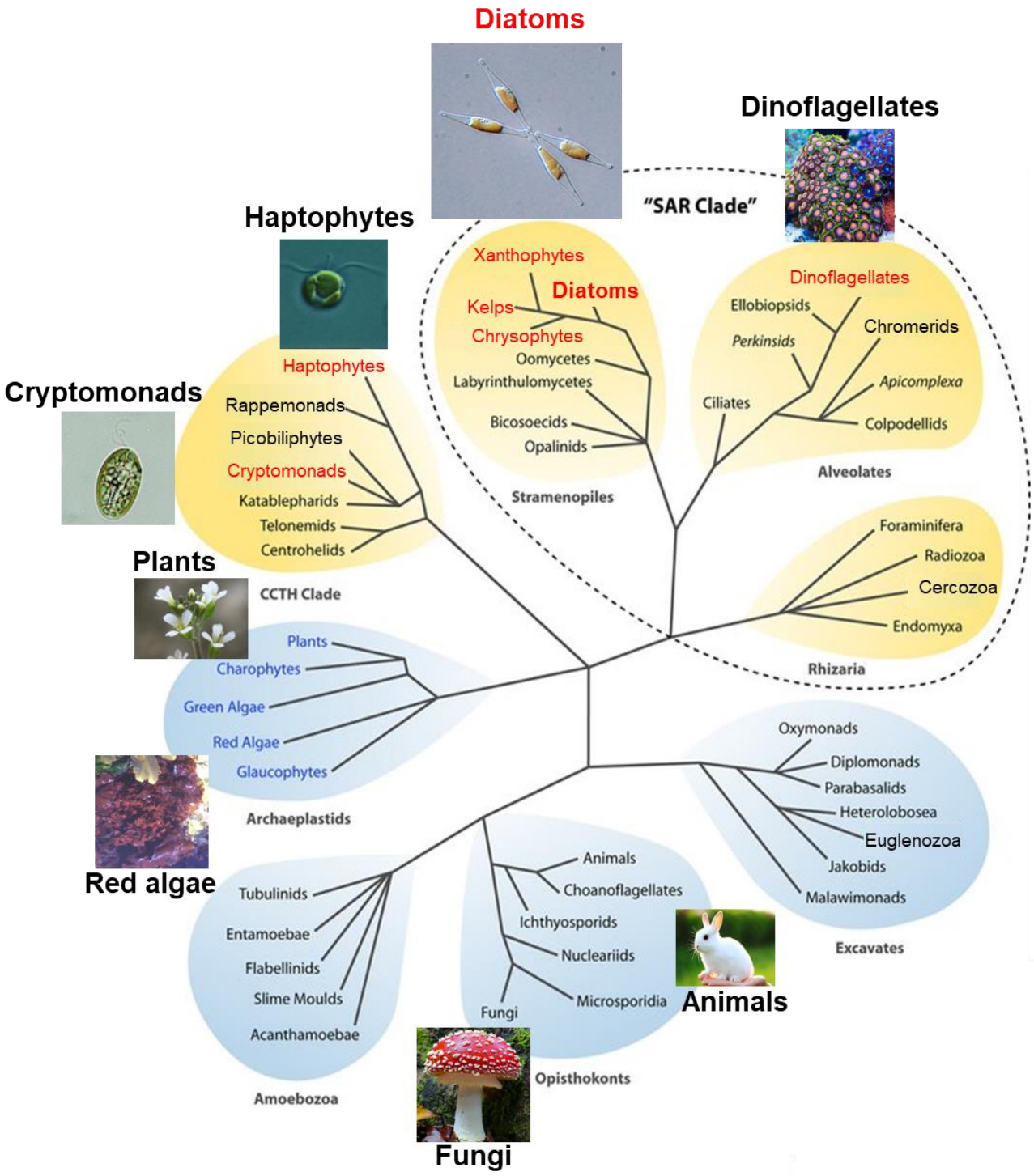

Figure 1.3 Diatom evolutionary position and relationships in the eukaryotic tree of life. The original tree figure is modified from Dorrell and Smith, 2011. Representative members of some ecologically important and well-studied groups are shown with pictures. Red color represents species that possess secondary red chloroplasts obtained via secondary or higher endosymbiosis. SAR: the "SAR" clade of Stramenopiles (including diatoms), Alveolates and Rhizaria (Dorrell and Smith, 2011).

The evolutionary history of diatoms involves at least two (or more) endosymbiotic events. A simplified scheme of the major events in diatom endosymbiotic evolution are shown in Figure 1.4. First, a cyanobacterium was acquired in a primary endosymbiosis by a common ancestor of the Archaeplastids, regrouping red algae, green algae and plants. This first endosymbiosis is believed from molecular data to have occurred approximately 1.5 billion years ago (Yoon et al., 2004). A large portion of the cyanobacterial genome was integrated into the host nucleus (with fewer than 250 genes retained in the chloroplast genome), and Archaeplastid chloroplasts are thus supported predominantly by nucleus-encoded and plastid-targeted proteins (Bhattacharya et al., 2004; Yoon et al., 2004; Falciatore et al., 2020; Dorrell et al., 2017).

Approximately half a billion years later, a second endosymbiotic event occurred, in which a eukaryotic red alga (and potentially also a green alga) containing a chloroplast and nucleus was captured by or invaded into a eukaryotic heterotroph (Moustafa et al., 2009; Dorrell et al., 2017). This "secondary red chloroplast" is present in stramenopiles, dinoflagellates, haptophytes and cryptomonads, which are nonetheless distantly related to one another in the tree of life (Figure 1.3, Figure 1.4). It is debated whether these red secondary chloroplasts were acquired through one or multiple secondary endosymbiosis events, or have a more complex history (involving tertiary or quaternary endosymbioses) but they are clearly related to one another and share similar genetic and biochemical features (Dorrell et al., 2017; Stiller et al., 2014). The origin of secondary red chloroplasts was accompanied by some degree of endosymbiotic gene relocation to the host nucleus (Oudot-le-Secq 2007; Green, 2011). Genes from the endosymbiont nuclear genome were either relocated to the host nucleus, with a projected maximum 450 genes in diatom nuclear genomes of red algal origin, while the endosymbiont mitochondrial genome was lost entirely (Rastogi et al., 2018; Dorrell et al., 2021).

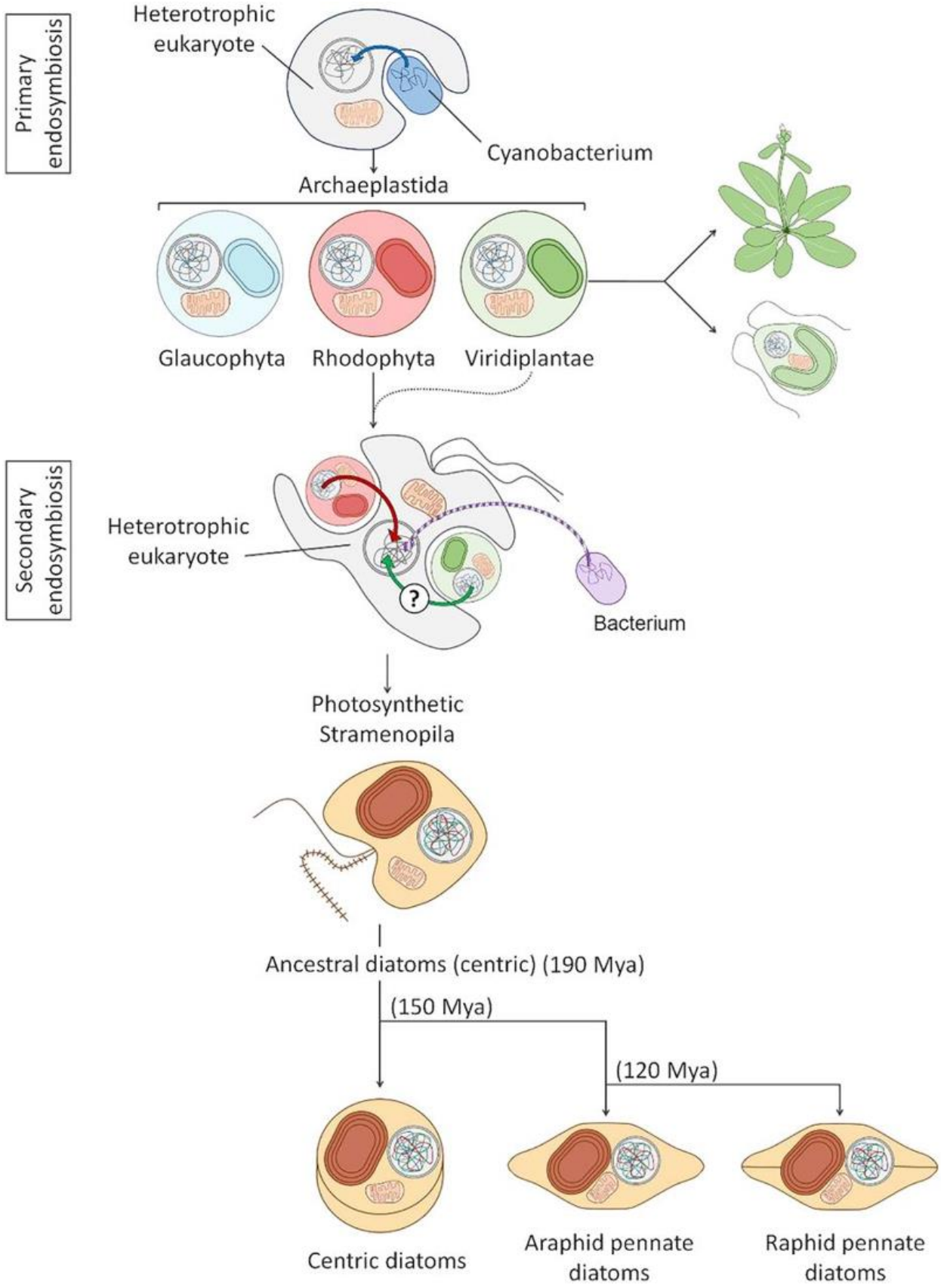

Figure 1.4 Simplified sketch map of the Major Events Leading to the Evolution of Diatoms through Primary and Secondary Endosymbiosis. The figure is cited from

Falciatore et al. (2020), which illustrates the process of primary and secondary endosymbiosis events. In primary endosymbiosis, a heterotrophic host engulfed a cyanobacterium (blue), leading to the transfer of a significant portion of the cyanobacterial genome into the host nucleus (blue arrow). This event gave rise to chloroplasts in the Archaeplastida group, including Glaucophyta (pale blue), Rhodophyta (pink), and the Viridiplantae (green). During secondary endosymbiosis, a separate heterotrophic cell acquired a red alga and potentially a green alga. The red alga became the chloroplast (brown) of the Stramenopila group (including diatoms). The transfer of algal nuclear genes to the heterotrophic nucleus is indicated by the green and red curved arrows, while the algal nucleus and mitochondria were lost. The Stramenopila genome also acquired bacterial genes through horizontal gene transfer events from bacterial donors (violet arrow). The approximate dates of separation between centric and pennate diatoms are based on Nakov et al. (2018).

1.2 *Phaeodactylum tricornutum*: the model species for diatom study

Much modern research on diatom biology has focused on one species, *Phaeodactylum tricornutum* (Bowler et al., 2019). *P. tricornutum* is the only member of the genus *Phaeodactylum,* although multiple genetically distinct ecotypes have been isolated (Rastogi, 2016; Rastogi et al., 2020). *P. tricornutum* is further known for its capability of altering its cell shape into different morphotypes, such as fusiform, triradiate, and oval, in response to environmental stimulation (De Martino et al., 2011). *P. tricornutum* is a probable coastal or intertidal species, although is rare elsewhere (e.g., is not detected in *Tara* Oceans data, Malviya et al., 2016).

Nonetheless, *P. tricornutum* harbors genes that have been proposed to contribute to diatom success in the wild, and serves as a *de facto* laboratory system for studying diatom eco-evolutionary physiology (Apt et al., 1996; Zaslavskaia et al., 2000; Siaut et al., 2007). *P. tricornutum* can easily be maintained under continuous culture in the lab with a fast cell division speed (approximately 1.5 times/day in mid-exponential phase under replete culture conditions). Moreover, unlike most diatoms *P. tricornutum* can be grown on solid growth media in the absence of silicate, and does not need to produce a silicate frustule to survive, rendering it amenable to experimental techniques such as transformation on agar plates.

As a model species in diatom research, various biotechnologies have been developed in *P. tricornutum*. The complete nuclear genome of *P. tricornutum* was first published in 2008 (Bowler et al., 2008), and the most recent (version 3) annotation was released in 2018 (Rastogi et al., 2018). The development of experimental transformation strategies, beginning with biolistic (bombardment) (Apt et al., 1996) and subsequently conjugation (Karas et al., 2015) and electroporation techniques has opened up the possibility for reverse genetic characterization of individual diatom genomes. Protein localization in *Phaeodactylum* has been rendered routine by GFP localization, with known reporter proteins for most cellular compartments (Siaut et al., 2007; Liu et al., 2016). Advancements in genetic modification techniques and knockout mutant generation have also been achieved in *P. tricornutum*, beginning with antisense, knockdown and overexpression strategies (De Riso et al., 2009), and subsequently utilizing CRISPR-Cas9 (Nymark et al., 2017) and TALENs-mediated systems (Weyman et al., 2015; Serif et al., 2017). Furthermore, *P. tricornutum* benefits from an extensive source of gene expression resources, beginning with microarray studies as synthesized by DiatomPortal (https://networks.systemsbiology.net/diatom-portal/), (Ashworth et al., 2016). More recently, RNAseq based approaches as synthesized by PhaeoNet (Ait-Mohamed et al., 2020) have provided greater clarity into absolute as well as relative expression trends of different *P. tricornutum* genes in response to environmental changes. Both DiatomPortal and PhaeoNet, as meta-studies of *P. tricornutum* microarray and RNAseq data respectively, have further shed light on transcriptional coordination among different genes, allowing a holistic view of the entire *P. tricornutum* genome as a series of co-expressed gene "modules".

1.3 The Chloroplast (Plastid): a metabolic factory integral to the diatom cell

1.3.1 Diatom plastids have a complex, four membrane-enclosed structure

Diatom chloroplasts, which I will henceforth refer to as "plastids", because of their difference in structure and evolutionary history to those of land plants, are fundamental to diatom cell biology. The majority of pennate diatoms contain one plastid, while the number of plastids in centric species is more variable. In

contrast to the double membrane-surrounded chloroplasts of primary endosymbiotic origin found in archaeplastid groups, the diatom plastid is surrounded by four membranes as shown in Figure 1.5 (Bhattacharya et al., 2004; Pfeil et al., 2014; Dorrell and Bowler, 2017; Murik et al., 2019). These membranes are, from the outside to inside (Figure 1.5): the plastid endoplasmic reticulum (cERM), the peri-plastidial membrane (PPM), the outer plastid membrane (OEM) and the inner envelope membrane (IEM). The outermost membrane cERM is connected to the diatom endoplasmic reticulum (ER) and nuclear membrane (Solymosi, 2012; Marchand et al., 2018).

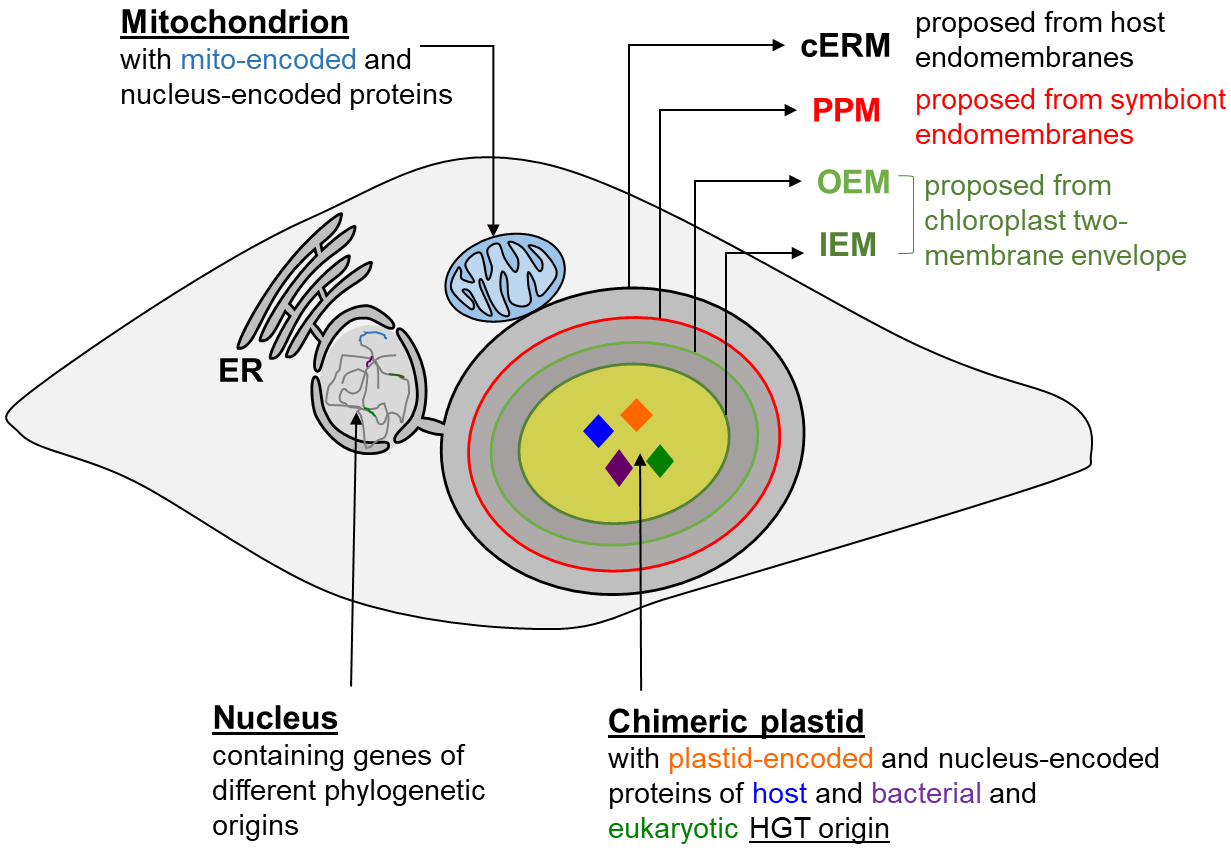

Figure 1.5 Structural diagram of the secondary red plastid in diatom cells. This figure was modified according to an original figure from Nonoyama et al. (2019). The plastid is surrounded by four membranes. The plastid endoplasmic reticulum membrane (cERM), shown in black, is continuous with the outer nuclear envelope membrane. The peri-plastidial membrane (PPM) is shown in red. The outer and inner envelope membrane (OEM and IEM) are shown in green. The chimeric origin of the diatom plastid proteome with plastid-encoded (orange) and nucleus-encoded proteins of host (blue) and bacterial (purple) and eukaryotic HGT origin (green) is shown by different color rhombuses. ER: endoplasmic reticulum.

1.3.2 Mosaic origin and complex import of diatom nucleus-encoded plastid proteins

A further level of complexity contributed by the multiple endosymbiosis events is the chimeric origin of the diatom chloroplast proteome (Dorrell et al., 2017; Nonoyama et al., 2019). Although diatoms contain some plastid-targeted proteins that clearly originate from red algae, they also possess a significant portion that originate from other sources such as the host, green algae and bacteria (Figure 1.5), alongside proteins that lack obvious homologues outside the diatoms, and may have specifically evolved within this lineage (Dorrell et al., 2017; Nonoyama et al., 2019; Falciatore et al., 2020). Although diatom plastids retain their own genomes, these are far smaller (< 150 genes) than their proteomes. Most of the plastid-associated proteins of diatoms (approximately 2000), are encoded in the nuclear genome, and transported into plastids following translation (Green, 2011; Gruber et al., 2015; Dorrell et al., 2017).

Diatom plastid-targeted proteins commence with an N-terminal signal peptide, recognized by a conventional Sec protein derived ER import machinery in the cERM (Kilian and Kroth, 2005). This is followed by a consensus "ASAFAP" motif with the the aromatic amino acid phenylalanine (F) (this can also be a Tryptophan; Tyrosine or Leucine) in the cleavage site +1 position (Kilian and Kroth, 2005; Gruber et al., 2007), allowing recognition by a specialized ERAD-derived protein import complex, SELMA, that resides in the PPM (Hempel et al., 2010; Lau et al., 2015; Lau et al., 2016). Finally, stromal preproteins are presumably transported across the IEM and OEM by a TOC/TIC-mediated translocon similar to that of higher plants (Bolte et al., 2009; Gruber and Kroth, 2017). Some stromal proteins may further be imported into thylakoids by the conserved ribo-nucleoprotein complex Signal Recognition Particle (SRP), Twin-Arginine Translocation delta pH-dependent (TAT) and thylakoid Sec pathways, as occurs in plants (Jarvis et al., 2004; Rosenblad et al., 2013). Other diatom proteins may also potentially be trafficked into the plastid following glycosylation in the ER/Golgi body or may be dual-targeted to the plastid and other compartments, such as the mitochondria (Peschke et al., 2013; Gile et al., 2015; Dorrell et al., 2017; Nonoyama et al., 2019). More details of the diatom plastid protein import system will be introduced in Chapter 3.

1.3.3 Diatom plastids possess unique photosystem and pigment architectures

The process of absorption, transfer and conversion of light energy in photosynthetic organisms primarily occurs across the thylakoid membrane, and involves four transmembrane macro-complexes, including Photosystem II (PSII), cytochrome b_6f ($cytb_6f$), Photosystem I (PSI), and ATP synthase (Nelson et al., 2015). Through light energy absorption, water photolysis, electron transfer, light energy is finally converted into chemical energy in the forms of ATP and NADPH (Nelson et al., 2015). Diatom plastids possess thylakoid membranes inside the stroma (Pfeil et al., 2014), but these are arranged in a concentric "girdle lamella" around the stromal periphery (Figure 1.6), as opposed to the stacked and unstacked thylakoids found in plant plastids (Andersen et al., 2013; Flori et al., 2017). The photosystems seem to be 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 thylakoids (Pyszniak and Gibbs, 1992; Flori et al., 2017). This structural organization is in contrast to that found in plants, where the two photosystems are separated over short distances and face each other in two external thylakoids with physical connections (Flori et al., 2017). The unusual segregation of the two photosystems may prevent the loss of reducing potential spillover from PSII to PSI, as found in cyanobacteria and red algae (Biggins and Bruce, 1985).

Diatom photosynthesis is highly efficient under dynamic light regimes. Diatoms contain pigments chlorophyll a (Chl a), chlorophyll c (Chl c) and fucoxanthin (Fx), as light-harvesting pigments (Kuczynska et al., 2015), similar to most other secondary red plastid-containing species with a very few exceptions (e.g., *Nannochloropsis*, which only binds Chl a and violaxanthin). A further subset of photoprotective carotenoids is uniquely found in a subset of secondary red-containing plastids, including diatoms close relatives (dictyochophytes, pelagophytes), and potential endosymbiotic derivatives (haptophytes), such as the carotenoids fucoxanthin (Fx) and diadinoxanthin/diatoxanthin (Ddx/Dtx) in diatom plastid (Kuczynska et al., 2015). Other xanthophylls like violaxanthin (Vx), astaxanthin (Ax), and zeaxanthin (Zx) may also occur in diatoms under specific stress conditions (Kuczynska et al., 2015).

Photosynthetic efficiency is an important trait for determining the ecological success of primary producers. Photosystem reaction centers surround a series of peripheral antenna-shaped light harvesting complexes (LHCs) (Figure 1.6), which play important roles in the absorption and dissipation of excess light energy (Nelson et al., 2015; Buck et al., 2019). Similar to the situation in other algae, diatom LHCs are attached to photosystem complexes PSI and PSII. Unlike the arrangement of these LHCs found in red algae and some of their endosymbiotic derivatives (e.g., cryptomonads), phycobilisomes are not known in diatoms.

The *P. tricornutum* light-harvesting antenna fucoxanthin Chl a/c binding protein (FCP) has a dimeric structure, centered around the encoded products of the Lhcf3 or Lhcf4 genes (Wang et al., 2019), and is structurally quite different from the trimeric LHCII of green-lineage organisms. Each FCP monomer binds seven Chl a, two Chl c, seven Fx, and one Ddx within the protein scaffold, with two calcium cations, one phosphatidyl-glycerol, and one digalactosyl-diacylglycerol. (Wang et al., 2019). The number of constitutive pigments in FCP differ across diatom species (Pi et al., 2019), although this basic structure is quite well-conserved. Diatom FCPs are further divided into three groups (Gundermann and Büchel, 2014). LHCF is the main light-harvesting antenna; LHCR is the specific complex of PSI (Gundermann and Büchel, 2014); and LHCX has been reported to involved in photoprotection under high and fluctuating light levels (Buck et al., 2019).

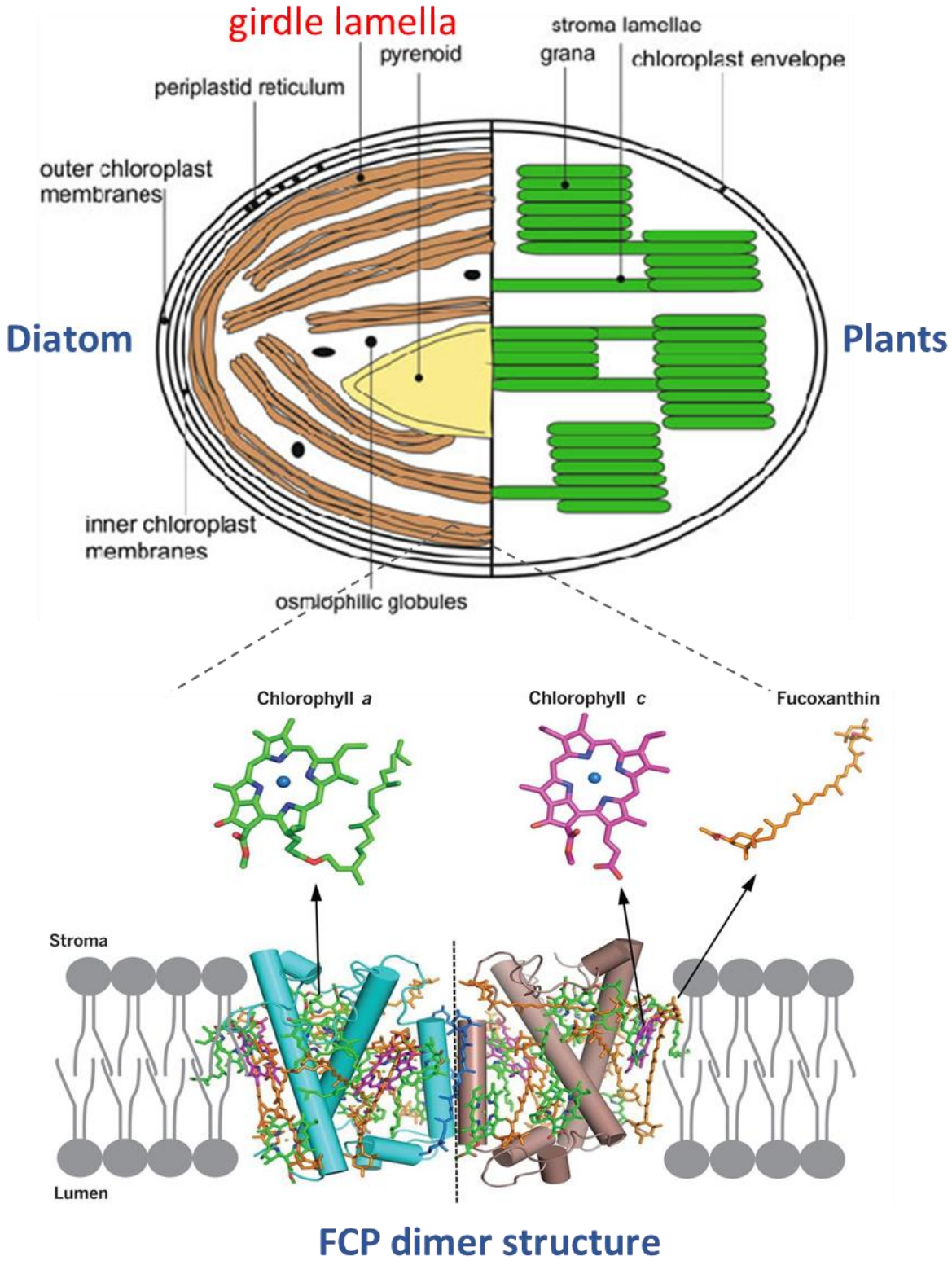

Figure 1.6 Comparisons of diatom and plant plastid structures (top), and anatomy of a diatom LHC (bottom). The diatom girdle lamella, FCP dimer structure, and pigment (Chla, Chlc, Fx) compositions are based on those present in *P. tricornutum*. This figure is modified from Grouneva et al. (2013) and Wang et al. (2019).

1.3.4 Metabolic innovation and crosstalk in the diatom plastid

The diatom plastid performs effectively all of the essential primary metabolism pathways identified in plant chloroplasts, including photosynthesis, central carbon and lipid metabolism, and the synthesis of plastidial amino acids, chlorophyll and carotenoids (Nonoyama et al., 2019, Ait-Mohamed et al., 2020). The diatom plastid also has a complete plastid-encoded and plastid-targeted machinery for its own upkeep, e.g., biosynthesis of the plastid and replication of its encoded genome, and transcription and translation of plastid genes (Oudot-le-Secq et al., 2007; Schober et al., 2019).

Nonetheless, diatoms possess many metabolic innovations different to those of higher plants, and these metabolic characteristics may contribute to their dominance in the contemporary ocean. These include a complex suite of plastid-targeted proteins involved in iron acquisition, mobilizing, storage, and stress tolerance. Diatoms may replace iron-dependent proteins in chloroplast photosystems with iron-free alternatives such as flavodoxin to tolerate chronic iron depletion stress (Pankowski and McMinn, 2009). Potential iron storage proteins (e.g., ferritin) are predicted to function in diatom plastids for temporal differences in iron availability (Gao et al., 2021, Nonoyama et al., 2019). This protein has been proposed to be used to maintain an Fe reserve for times of Fe starvation, or equally be used to "buffer" or remove excess Fe from the water column and competitively exclude other species (Gao et al., 2021, Nonoyama et al., 2019).

Diatoms can further utilize CO_2 concentrating mechanisms (CCMs) to enrich carbon dioxide availability to RuBisCO (Reinfelder et al., 2000; Roberts et al., 2007; Kikutani et al., 2016). Unusual plastid located theta- and iota-class carbonic anhydrases have been reported in *P. tricornutum* that augment the efficiency of CO_2 supply into RuBisCO (Jensen et al., 2019). The primary carbon metabolism of diatom plastids includes a complete glycolysis-gluconeogenesis pathway from glucose-6-phosphate to pyruvate, which may be used to remove excess triose phosphate in response to long days and potentially supply it at low temperature (Dorrell et al., 2022). Finally, a potential C4 photosynthesis pathway encoded in diatom genomes is of unresolved function, although has recently been proposed to function principally in the mitochondria to allow efficient

recycling of respired CO_2 (Yu et al., 2022).

Another interesting phenomenon of diatom plastids is their intricate crosstalk with mitochondria. These two organelles typically show close proximity to one another in transmission electron micrograph (TEM) images (Figure 1.7). Diatom plastids and mitochondria may possess specific connection points that are not associated (at least found to the same extent) with plants or other algae (Prihoda et al., 2012; Bailleul et al., 2015; Uwizeye et al., 2020).

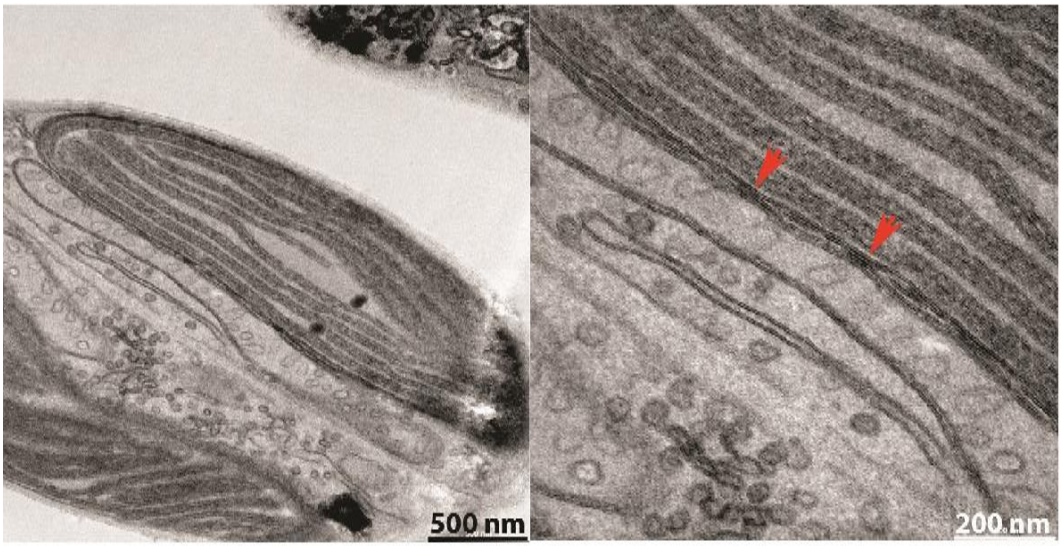

Figure 1.7 Electron microscopy pictures showing contact points between diatom plastids and mitochondria. This picture is cited from Bailleul et al. (2015). The red arrows show the contact point between the plastid and the mitochondrion.

At a metabolic level, diatom plastids and mitochondria directly interact with one another via a complete plastid-targeted ornithine cycle, and mitochondria-targeted urea cycle, which facilitate the recycling of excess amino acids and nitrogen assimilation by equilibrating cellular glutamine and ornithine reserves (Allen et al., 2011; McCarthy et al., 2017; Nonoyama et al., 2019). Diatom plastids and mitochondria may further interact via a shared glyoxylate pathway, succinic acid cycle and branched-chain amino acid synthesis and catabolism pathway (Prihoda et al., 2012; Bailleul et al., 2015; Broddrick et al., 2019; Murik et al., 2019). Certain metabolic pathways exist in duplicated or redundant copies between diatom plastids and mitochondria, such as the lower half of glycolysis and the glutamate-glutamine (Glu-Gln) cycle performed by Glutamine synthase (GS) and glutamate synthase/glutamine oxoglutarate aminotransferase (GOGAT) (Dorrell et al., 2022; Broddrick et al., 2019; Smith et al., 2019; Ait-Mohamed et al., 2020).

Most unusually, a direct energetic coupling pathway has been observed between diatom plastids and mitochondria, based on an electro-chromic shift (ECS) that relates to the proton motive forces identified across organelle membranes (Bailleul et al., 2015). This interaction has been interpreted to be due to a re-routing of excess plastid reducing power, either ATP/NADPH towards the mitochondria, and the corresponding import of mitochondrial ATP into the plastid to support photosynthetic metabolism (Bailleul et al., 2015). This pathway has been documented elsewhere in green algae and plants, but only in mutant lines from which other alternative electron flow pathways endogenous to each organelle (e.g., the water-water cycle) have been suppressed, whereas it appears to be the predominant form of alternative electron transport in diatoms (Kinoshita et al., 2011; Cardol et al., 2011; Kaye et al., 2019). The exact molecular actors of these exchanges remain debated, although mitochondrial alternative oxidase (AOX) is believed to play a significant role (Broddrick et al., 2019; Murik et al., 2019).

Recently, in light of the distinct organelle metabolism observed in diatoms compared to plants, Ait-Mohamed et al. (2020) analyzed a global WGCNA dataset of co-regulated genes across the *P. tricornutum* genome for 372 curated nuclear-encoded proteins targeted to either the plastids and/or to the mitochondria (Figure 1.8). The co-regulated gene modules that were most frequently associated with organelle metabolism in this study, e.g., blue, lightcyan1, lightsteelblue1, and paleturquoise, show different functions to one another related to, photosynthesis, carbon fixation, and core biosynthetic pathways. For example, amino-acyl tRNA synthetase genes, which are to be dual-targeted to both chloroplasts and mitochondria in diatoms (Gile et al., 2015; Dorrell et al., 2017, Dorrell et al., 2019), were frequently retrieved within the plastid-related blue module (Figure 1.8) that otherwise encodes large components of photosynthesis and the Calvin cycle, providing a possible transcriptional mechanism for the coordination of plastid metabolism to expression of the mitochondrial genome. Other modules, such as the magenta, steelblue and cyan module, contain proteins inferred to form metabolic connections between diatom plastids and mitochondria, such as the Glu-Gln cycle and glycine cleavage system, as illustrated in Figure 1.8.

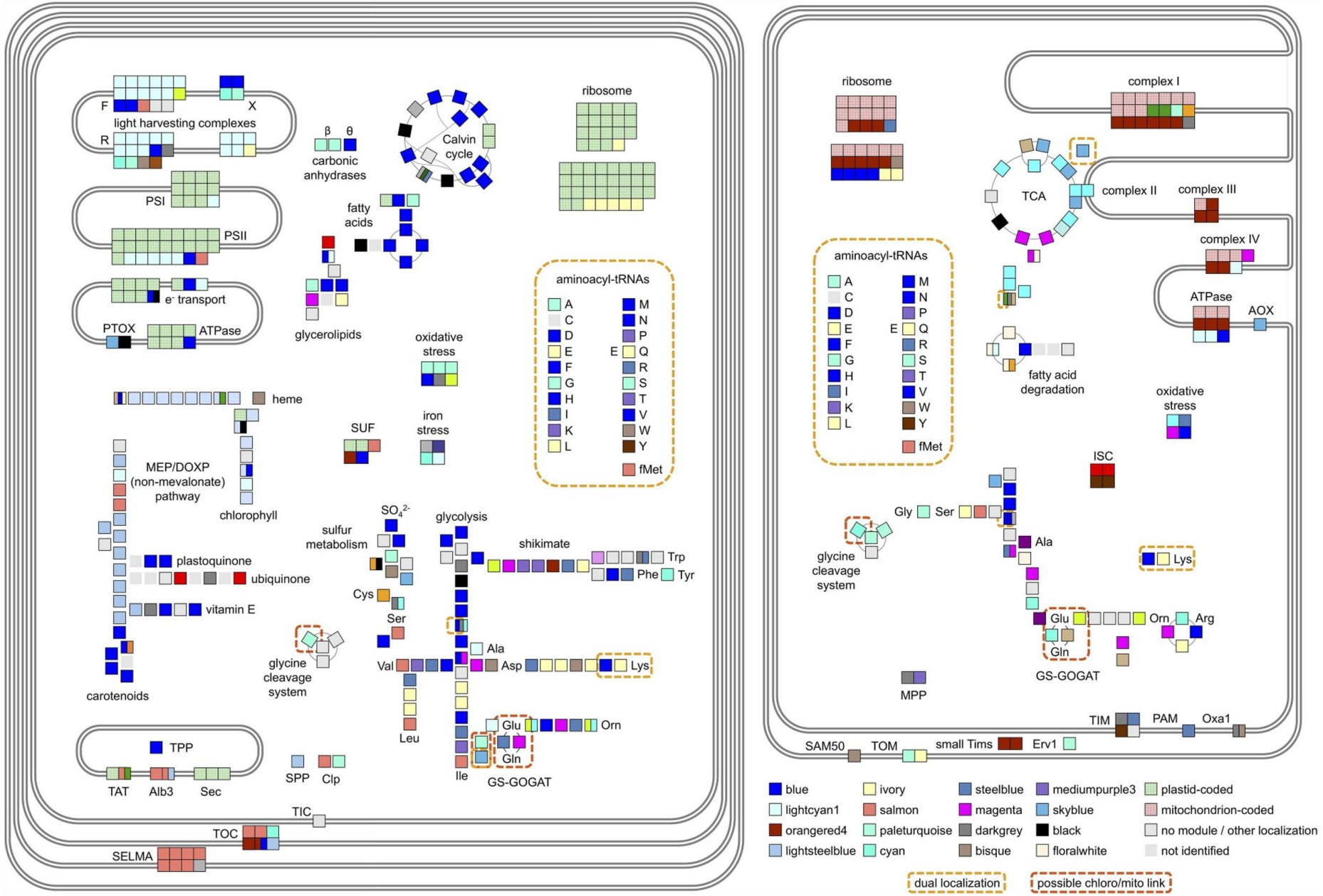
F
X
light harvesting complexes
R
PSI
PSII
e- transport
PTOX
ATPase
β θ
carbonic anhydrases
Calvin cycle
fatty acids
glycerolipids
ribosome
oxidative stress
aminoacyl-tRNAs
A C D E F G H I K L
M N P E Q R S T V W Y
fMet
heme
SUF
iron stress
MEP/DOXP (non-mevalonate) pathway
chlorophyll
plastoquinone
ubiquinone
vitamin E
carotenoids
SO4 2-
sulfur metabolism
glycolysis
shikimate
Trp
Phe Tyr
Cys
Ser
glycine cleavage system
Ala
Val
Asp
Lys
Leu
Glu
Orn
Gln
Ile
GS-GOGAT
TPP
TAT
Alb3
Sec
SPP
Clp
TIC
TOC
SELMA
ribosome
complex I
TCA
complex II
complex III
complex IV
ATPase
AOX
aminoacyl-tRNAs
A C D E F G H I K L
M N P E Q R S T V W Y
fMet
fatty acid degradation
oxidative stress
ISC
Gly
Ser
glycine cleavage system
Ala
Lys
Glu
Orn
Arg
Gln
GS-GOGAT
MPP
TIM
PAM
Oxa1
SAM50
TOM
small Tims
Erv1
blue
ivory
steelblue
mediumpurple3
plastid-coded
lightcyan1
salmon
magenta
skyblue
mitochondrion-coded
orangered4
paleturquoise
darkgrey
black
no module / other localization
lightsteelblue
cyan
bisque
floralwhite
not identified
dual localization
possible chloro/mito link

Figure 1.8 Main metabolic pathways and functional complexes of *P. tricornutum* plastid (left) and mitochondria (right) and their coregulated modules in PhaeoNet. This figure is cited from Ait-Mohamed et al. (2020). Each square represents a protein-coding gene, colored by transcriptional profiles in different WGCNA modules. Definitions of each module are provided in the figure key. Clusters of adjacent squares indicate genes encoding different components of a specific multi-unit enzyme or complex, while split squares represent genes encoding functional homologs of one specific protein. Merged modules are color-coded to indicate the associated WGCNA merged module. Proteins from organellar genomes are depicted as dotted green or red squares, unknown targeting or unassigned merged modules are shown as light gray squares. Enzymatic steps not identified in the genome are depicted by light gray squares without borders. Dual-localized proteins are marked by checkered yellow boxes, and orange boxes highlight potential interconnection points between organelles. Note: Blue model links central components of plastid and mitochondial metabolism carbon and amino acid, lightsteelblue and lightcyan1 are principally confined to photosynthesis, other modules (salmon, for example) are principally associated with mitochondrial metabolism. Abbreviations please check in Ait-Mohamed et al. (2020).

1.4 Diatom plastid transporter research status in diatoms

The compartmental organization of the diatom plastid necessitates a diverse range of transporters to transport ions, substrates and products across its membranes (Marchand et al., 2018; Marchand et al., 2020). The plastid-related metabolic activities unique to diatoms may depend on transporter innovations across the four membranes, particularly in the context of metabolite and ion exchanges between plastid and cytoplasm, or between organelles (Shai et al., 2016; Dorrell et al., 2017; Mix et al., 2018).

Although transporters play essential roles in the diatom plastid, little is known specifically concerning the evolutionary diversity, molecular functions and environmental relevance and localizations. Previously, Marchand et al. (2018) presented overviews of the localization and activity of ion and metabolite plastid transporters from algae to land plants; Brownlee et al. (2022) offered a comprehensive summary of key transport mechanisms involved in key nutrient (nitrogen, carbon, phosphorus, silicon, and iron) acquisition and transport in marine diatoms. In our recent work (Liu et al., 2022), we utilized a combination of transcriptomic, phylogenomic, and environmental data to construct an *in*

silico survey of predicted plastid transporters identified through genome-wide searches in the model diatom *Phaeodactylum tricornutum*. This paper is provided in Chapter 2 of this thesis.

Potential diatom plastid ion and metabolic transporters are listed in Table 1.1. As represented in the table, diatom plastids possess multiple probable ion transporters, including phosphate, nitrite and potassium (Table 1.1). Based on my knowledge, to date, no diatom plastid transporters specifically for iron and manganese have been identified (Table 1.1). The phosphate transporter Pho4, which contains a signal peptide, is probably involved in phosphate transport, although it has been localized by eGFP to the cER (Dell'Aquila et al., 2020). Regarding nitrogen transport, a plastid-localized nitrite transporter has been identified, playing a crucial role in the exchange of NO_2^- between the plastid and cytoplasm, thereby contributing to nitrogen assimilation (McCarthy et al., 2017). A homologue of this transporter has also been detected in an experimentally realized plastid proteome of the diatom *T. pseudonana* (Schober et al., 2019). A presumed plasma membrane nitrate transporter, NRT2, has also been reported (McCarthy et al., 2017; Kang et al., 2019). Seydoux et al. (2022) have recently generated *P. tricornutum* mutants of the H^+/K^+ antiporter KEA3, suggesting that it serves as a key regulator of PET/PMF coupling. Finally, SLC4 family proteins may participate in the active transport of bicarbonate into diatom plastids. The *P. tricornutum* genome encodes one SLC4 transporter, PtSLC4-2, with a predicted plastid-targeting sequence, although it has also been experimentally localized to the plasma membrane (Nakajima et al., 2013).

Specific studies of diatom plastid metabolite transporters have mainly focused on triose phosphate transporters (TPT) and nucleotide triphosphate transporters (NTT) (Table 1.1). Ast et al. (2009) conducted a comprehensive search for potential NTTs in diatoms *T. pseudonana* and *P. tricornutum*, characterizing two isoforms of nucleotide transporter, NTT1 and NTT2. Their heterologous expression experiments confirmed that NTT1 acts as a proton-dependent adenine nucleotide importer, while NTT2 facilitates the counter exchange of (deoxy-) nucleoside triphosphates. Moog et al. (2015) investigated the presence, localization, and phylogeny of putative triose phosphate translocators (TPTs) in the diatom plastids. They identified four potential TPTs, named TPT1, TPT2, TPT4a, and TPT4b, located in the plastid complex transmembrane system.

Additionally, a study of non-photosynthetic diatom *Nitzschia sp.* NIES-3581 (Moog et al., 2022) proposed that TPT homologues of complex plastids derived from red algae potentially have the ability to transport triose phosphates (TPs) and phosphoenolpyruvate (PEP), considering the absence of detected glucose 6-phosphate transporter (GPT), xylulose 5-phosphate transporter (XPT), and phosphoenolpyruvate transporter (PPT) homologues (Moog et al., 2020; Liu et al., 2022).

Table 1.1 Important ion and metabolic plastid transporters in *P. tricornutum*

Categories	Type of transport	Potential plastid transporter	Reference	Localization confirmed?
Ion transport	Sodium/hydrogen	J10409: Na+/H+ antiporter	(Liu et al., 2022)	–
		J14403: Na+/H+ antiporter	(Liu et al., 2022)	–
		J1764: Na+/H+ antiporter	(Liu et al., 2022)	–
	Chloride transport	J43785: CLC superfamily/chloride channel	(Pfeil et al., 2014; Liu et al., 2022)	–
	Phosphate transport	J23830: Phosphate transport Pho4	(Dell'Aquila et al., 2020; Liu et al., 2022)	GFP: ER?
	Sulfur transport	EG00808: Sulfate transporter	(Liu et al., 2022)	–
	Nitrite transport	J13076: Nitrite transporter NAR1	(McCarthy et al., 2017; Schober et al., 2019)	–
	Potassium transport	J39274: Potassium transport KEA3	(Pfeil et al., 2014;Seydoux et al., 2022)	GFP: plastid
		J38821: Ion channel protein (K+)	(Liu et al., 2022)	–
		J43171: Ion channel protein (K+)	(Liu et al., 2022)	–
	Magnesium transport	J38786: Magnesium transporter	(Liu et al., 2022)	–
	Calcium transport	J39402: Calcium transport	(Liu et al., 2022)	–
	Manganese transport	–	–	–
	Iron transport	–	–	–
	Copper transport	J47805: Copper transport	(Liu et al., 2022)	–
Metabolite transport	Nucleotide transport	J49533: Nucleotide transporters NTT1	(Ast et al., 2009)	GFP: plastid
		J11615: Nucleotide transporters NTT2	(Ast et al., 2009)	GFP: plastid?
	Carbohydrate transport	J50742: Triose phosphate translocator TPT1	(Moog et al., 2015)	self-assembling GFP: cER?
		EG00747: Triose phosphate translocator TPT2	(Moog et al., 2015)	self-assembling GFP: PPM?
		J54017: Triose phosphate translocator TPT4a	(Moog et al., 2015)	self-assembling GFP: IEM
		J24610: Triose phosphate translocator TPT4b	(Moog et al., 2015)	self-assembling GFP: IEM
	Bicarbonate transport	J45656: HCO3- transporter SLC4	(Liu et al., 2022)	–
	Amino acid transport	J43194: Amino acid transporter	(Liu et al., 2022)	–
		J45969: Amino acid transporter	(Liu et al., 2022)	–
		J49985: Amino acid transporter	(Liu et al., 2022)	–
		J50071: Amino acid transporter	(Liu et al., 2022)	–
		J45852: Amino acid transporter	(Liu et al., 2022)	–
	Glycolipid transport	J36726: Glycolipid transporter	(Liu et al., 2022)	–
Other Superfamily	MFS contained	J18945: MFS contain	My project target	self-assembling GFP: PPM
		J16421: MFS contain	(Liu et al., 2022)	–
		J37674: MFS contain	(Liu et al., 2022)	–
		J42868: MFS contain	(Liu et al., 2022)	–
		J44665: MFS contain	(Liu et al., 2022)	–
		J49936: MFS contain	(Liu et al., 2022)	–
	ABC transporter family	J41857: ABC	(Liu et al., 2022)	–
		J22803: ABC	(Liu et al., 2022)	–
		J16142: ABC	(Liu et al., 2022)	–
		J22325: ABC	(Liu et al., 2022)	–
		J22332: ABC	(Liu et al., 2022)	–

In Table 1.1, transporters are listed based on predicted functional annotations and inferred plastid-targeting sequences (ASAFind and HECTAR prediction). Transporters that have previously been investigated experimentally are shown in color, with red showing transporters with coherent localizations to plastid membranes, and blue transporters inferred to localize to other membranes.

1.5 Diatom 'omics datasets and their availability

1.5.1 Genome sequencing resources of diatom species

Beyond *P. tricornutum*, multiple diatom species have published sequenced genomes, which will help the study for their metabolism and evolution. These include: *Thalassiosira pseudonana* (Armbrust et al., 2004), *Thalassiosira oceanica* (Lommer *et al.*, 2012), *Cyclotella cryptica* (Traller et al., 2016), *Skeletonema costatum* (Ogura et al., 2018), *Fragilariopsis cylindrus* (Mock et al., 2017), *Pseudo-nitzschia multiseries* CLN47 (JGI Project ID: 16870), *Pseudo-nitzschia multistriata* strain B856 (Basu et al., 2017), *Fistulifera solaris* (Tanaka et al., 2015b), *Synedra acus* (Galachyants et al., 2015), *Plagiostriata* sp. CCMP470 (Sato et al., 2020) and a non-photosynthetic species *Nitzschia putrida* (Kamikawa et al., 2022). Moreover, the 100 diatom genomes sequencing project of the Joint Genomes Institute (JGI) (https://jgi.doe.gov/csp-2021-100-diatom-genomes/), will allow us to fully understand diatom genomic diversity.

1.5.2 Large 'omic datasets: MMETSP and *Tara* Oceans resources

The availability of many transcriptomic and proteomic data sets has greatly facilitated genome annotation efforts and therefore gene discovery in model species. Beyond single diatom species transcriptome and proteome sequencing efforts (Rastogi et al., 2018; Wang et al., 2018), large-scale 'omic datasets such as The Marine Microbial Eukaryote Transcriptome Sequencing (MMETSP) (Keeling et al., 2014) and *Tara* Oceans projects (Sunagawa et al., 2020) have illuminated diatom evolutionary and environmental diversity.

MMETSP brought together a global effort to sequence marine phytoplankton, including 678 transcriptomes from 411 different strains among 92 diatom

species (Keeling et al., 2014). Further transcriptomes have been generated by Genoscope and represent seven of the most globally abundant diatom genera, but not covered by MMETSP (Dorrell et al., 2021).

Tara Oceans is an international, multidisciplinary project to assess the complexity of ocean life across comprehensive taxonomic and spatial scales from 210 sampling stations (Figure 1.9) at depths down to 1,000 m during a global circumnavigation (Sunagawa et al., 2020). In this project, multiple environmental factors associated with sampling stations have been measured, and plankton separated according to different size fractions have been analyzed using meta-barcoding, meta-genomic, meta-transcriptomic and high-throughput imaging approaches (Figure 1.9). *Tara* Oceans provides a powerful resource to explore the distribution of diatom species and genes in a global-scale environmental context.

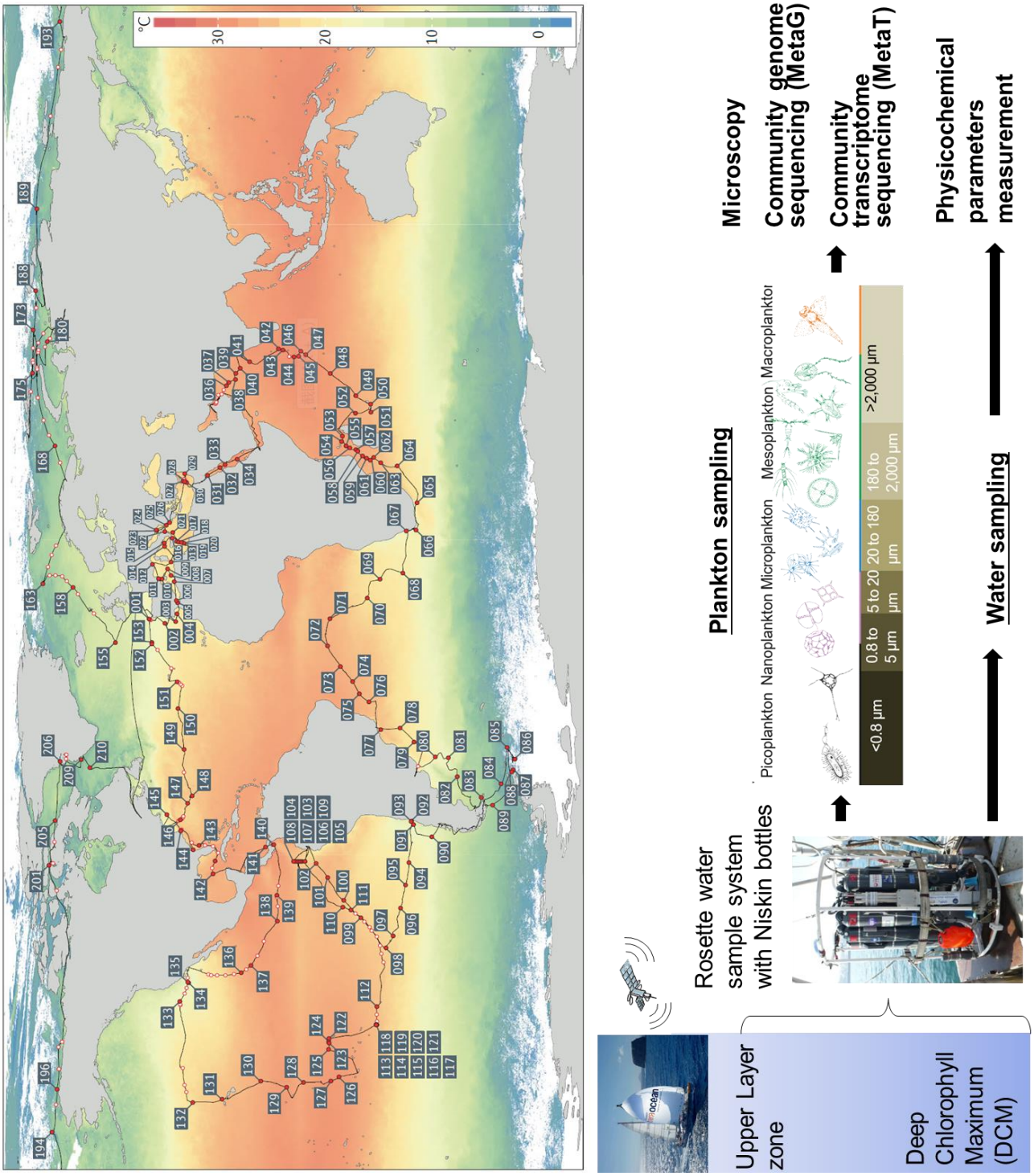

Figure 1.9. Overview of the *Tara* Oceans expeditions. The map shows the cruise track and 210 sample stations of *Tara* Oceans and *Tara* Polar Circle expeditions. Sea surface temperature is shown on a color gradient. At each open-ocean station, *Tara* sampled plankton and from the upper layer zone Surface (SRF) and Deep Chlorophyll Maximum layer (DCM), separated over multiple size fractions (0.8- 3/5 µm; 3/5-20 µm; 20-180 µm; 180-2000 µm). Sample community meta-genomes (MetaG) and meta-transcriptomes (MetaT) were then sequenced, and water from each station was tested for key physico-chemical and environmental parameters. This figure is adapted from Sunagawa et al. (2020).

1.6 Outline of the thesis

As outlined above, diatoms are a major component of marine phytoplankton primary production, and their global success in particular in iron-limited regions of the world ocean may relate in part to their unusual four-membrane bound plastids of secondary red origin that show close proximity to the mitochondria. Nonetheless, the specific metabolic effectors and in particular the transporters essential for diatom plastid function remain poorly understood. My PhD has sought to characterize novel plastid transporters that underpin diatom-specific metabolic innovations, centered around the model species *Phaeodactylum tricornutum*.

Chapter 2 presents a previously published study I have led of 70 predicted plastid transporters encoded by the *P. tricornutum* genome (Liu et al., 2022), which leverages transcriptional, phylogenetic and *Tara* Oceans data to assign novel functions to these transporters. I participated in the design of this project, led the functional analysis, and co-wrote the paper in collaboration with my co-authors.

The remainder of my thesis focuses on a novel Major Facilitator Superfamily (MFS) transporter, named Trans3. A brief overview of this transporter is presented in Chapter 2.2. The gene encoding Trans3 was strongly transcriptionally coregulated with the mitochondrial genome, which may suggest that this transporter participates in plastid-mitochondria crosstalk; and based on Reciprocal Blast best-Hit (RbH) analysis, Trans3 homologues were unique to species with secondary red plastids.

Chapter 3 presents experimental evidence of Trans3 subcellular localization and orientation in the *P. tricornutum* plastid. I present *in vivo* observations of Trans3 by self-assembling GFP, suggesting that it is located at the peri-plastidial membrane (PPM).

Chapter 4 presents the generation of partial Trans3 knockout lines by CRISPR-Cas9, alongside an assessment of its physiological functions considering growth rate, photo-physiology and transcriptome analysis. These data suggest Trans3 is essential for *P. tricornutum* cell growth and efficient Photosystem I activity. By

performing comparative analyses in media with different iron depletion and nitrogen source conditions, I propose that Trans3 interacts with cellular Fe and NO_3^-.

Chapter 5 presents Trans3 structural similarity with the Molybdate Transporter 2 (MOT2) protein from the model green alga *Chlamydomonas reinhardtii*. Through qRT-PCR analysis and experimental tests of mutant lines molybdate uptake, I propose that Trans3 acts as a novel, plastid-targeted molybdate transporter.

In Chapter 6, I present the final conclusions and perspectives of my PhD, alongside suggestions for future work to study the functions of Trans3.

Chapter 2: Target transporter (Trans3) discovery though *in silico* analysis of 70 predicted plastid transporters in the *P. tricornutum* genome

Abstract

We performed *in silico* analyses of 70 predicted plastid transporters identified by a genome-wide search of *Phaeodactylum tricornutum*. We considered similarity with *Arabidopsis thaliana* chloroplast transporters, transcriptional co-regulation with genes encoding core plastid metabolic pathways and with genes encoded in the mitochondrial genomes, inferred evolutionary histories using single-gene phylogeny, and environmental expression trends using *Tara* Oceans meta-transcriptomics and meta-genomes data. Our data reveal some of the ion, nucleotide and sugar plastid transporters associated with plants are conserved in diatoms, such as non-specific triose phosphate transporters implicated in the transport of phosphorylated sugars, nucleoside triphosphates (NTP)/nucleoside diphosphates (NDP) and cation exchange transporters. However, our data also highlight the presence of diatom-specific transporter functions, such as carbon and amino acid transporters implicated in intricate plastid-mitochondria crosstalk events. These confirm previous observations that substrate non-specific triose phosphate transporters (TPT) may exist as principal transporters of phosphorylated sugars into and out of the diatom plastid, alongside suggesting probable agents of NTP exchange. We additionally provide evidence from environmental meta-transcriptomic/metagenomic data that plastid transporters may underpin diatom sensitivity to ocean warming.

We identified an interesting major facilitator superfamily (MFS) transporter termed as Trans3 with uncertain function. Trans3 was inferred to be distributed across and unique to species containing secondary red plastids (Dorrell et al., 2021), and showed strong positive coregulation with genes from the mitochondrial genome. This novel transporter was selected as a target for further localization and functional investigation.

2.1 Annex I: "A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species *Phaeodactylum*"

Liu, S., Storti, M., Finazzi, G., Bowler, C., & Dorrell, R. G. (2022). A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species *Phaeodactylum*. Frontiers in Plant Science, 3640. https://doi.org/10.3389/fpls.2022.950467

My contributions:

Project design: 50% (with RGD)
Data analysis: 70% (Figure 1 50%, Figure 2, Figure 4 and Figure 5, with MS and RGD)
Paper writing: 70% (with all co-authors)

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 September 2022
DOI 10.3389/fpls.2022.950467

OPEN ACCESS

EDITED BY
Ben Lucker,
Prosel Biosciences, United States

REVIEWED BY
Ilka Haferkamp,
University of Kaiserslautern, Germany
Luiz Felipe Benites,
NASA Postdoctoral Program - Rutgers University, United States

*CORRESPONDENCE
Richard G. Dorrell
dorrell@bio.ens.psl.eu

SPECIALTY SECTION
This article was submitted to Marine and Freshwater Plants, a section of the journal Frontiers in Plant Science

RECEIVED 22 May 2022
ACCEPTED 02 September 2022
PUBLISHED 22 September 2022

CITATION
Liu S, Storti M, Finazzi G, Bowler C and Dorrell RG (2022) A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species *Phaeodactylum*. *Front. Plant Sci.* 13:950467. doi: 10.3389/fpls.2022.950467

A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species *Phaeodactylum*

Shun Liu[1,2], Mattia Storti[3], Giovanni Finazzi[3], Chris Bowler[1,2] and Richard G. Dorrell[1,2]*

[1]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Ecole Normale Supérieure, Centre National De La Recherche Scientifique (CNRS), Institut National De La Santé Et De La Recherche Médicale (INSERM), Université Paris Sciences et Lettres (PSL), Paris, France, [2]CNRS Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, 3 rue Michel-Ange, Paris, France, [3]Univ. Grenoble Alpes (UGA), Centre National Recherche Scientifique (CNRS), Commissariat Energie Atomique Energies Alternatives (CEA), Institut National Recherche Agriculture Alimentation Environnement (INRAE), Interdisciplinary Research Institute of Grenoble (IRIG), Laboratoire de Physiologie Cellulaire et Végétale (LPCV), Grenoble, France

Diatoms are an important group of algae, contributing nearly 40% of total marine photosynthetic activity. However, the specific molecular agents and transporters underpinning the metabolic efficiency of the diatom plastid remain to be revealed. We performed *in silico* analyses of 70 predicted plastid transporters identified by genome-wide searches of *Phaeodactylum tricornutum*. We considered similarity with *Arabidopsis thaliana* plastid transporters, transcriptional co-regulation with genes encoding core plastid metabolic pathways and with genes encoded in the mitochondrial genomes, inferred evolutionary histories using single-gene phylogeny, and environmental expression trends using *Tara* Oceans meta-transcriptomics and meta-genomes data. Our data reveal diatoms conserve some of the ion, nucleotide and sugar plastid transporters associated with plants, such as non-specific triose phosphate transporters implicated in the transport of phosphorylated sugars, NTP/NDP and cation exchange transporters. However, our data also highlight the presence of diatom-specific transporter functions, such as carbon and amino acid transporters implicated in intricate plastid-mitochondria crosstalk events. These confirm previous observations that substrate non-specific triose phosphate transporters (TPT) may exist as principal transporters of phosphorylated sugars into and out of the diatom plastid, alongside suggesting probable agents of NTP exchange. Carbon and amino acid transport may be related to intricate metabolic plastid-mitochondria crosstalk. We additionally provide evidence from environmental meta-transcriptomic/meta-genomic data that plastid transporters may underpin diatom sensitivity to ocean warming, and identify a diatom plastid transporter (J43171) whose expression may be positively correlated with temperature.

KEYWORDS

Bacillariophyta, metabolite import-export, plastid-targeted proteins, meta-genomics, *in silico* prediction, RNAseq, microarray, Protist

Introduction

Diatoms are one of the most abundant eukaryotic phytoplankton groups in the contemporary ocean (Benoiston et al., 2017; Vincent and Bowler, 2020), especially in nutrient-rich coastal upwelling regions and at high latitudes (Tréguer et al., 2017), and are responsible for 20% of total planetary net primary production (Nelson et al., 1995; Field et al., 1998; Vincent and Bowler, 2020). Understanding why diatoms have risen to a position of such unique ecological prominence is fundamental to understanding the function and future dynamics of the ocean ecosystem. Diatom plastids (*i.e.*, "chloroplasts"), as host organelles of photosynthesis (Marchand et al., 2018) and other central biosynthetic pathways (Terashima et al., 2011; Solymosi, 2012; Heydarizadeh et al., 2013; Dorrell et al., 2017; Dorrell et al., 2019; Nonoyama et al., 2019), play a critical role in marine primary production, and may be the key to explaining the dominance of diatoms in the ocean.

In contrast to the double membrane-surrounded plastids of primary endosymbiotic origin found in red algae, green algae and land plants, diatoms possess plastids which were acquired by a secondary (or higher) endosymbiosis event, in which a non-photosynthetic host eukaryote acquired a plastid by combining with a red algal endosymbiont (or one of its direct endosymbiotic descendants) (Bhattacharya et al., 2004; Dorrell and Bowler, 2017; Murik et al., 2019). As a result of their eukaryotic origins, diatom plastids are enclosed by four membrane layers (Bhattacharya et al., 2004; Pfeil et al., 2014). These membranes are, from the outside to inside: the plastid endoplasmic reticulum (cERM), the peri-plastidial membrane (PPM), the outer plastid membrane (OEM) and the inner envelope membrane (IEM) (Solymosi, 2012; Marchand et al., 2018). Diatom plastids possess thylakoid membranes inside the stroma (Pfeil et al., 2014), but these are arranged in a concentric "girdle lamella" around the stromal periphery, as opposed to the stacked and unstacked thylakoids found in plant plastids (Flori et al., 2017). The compartmental organization of the diatom plastid necessitates a diverse range of transporters to transport ions, substrates and products across its membranes (Marchand et al., 2018; Marchand et al., 2020). Although diatom plastids retain their own genomes, these are far smaller (< 150 genes) than their proteomes. Most of the approximately 2000 plastid-associated proteins) of diatoms are encoded in the nuclear genome, and transported into plastids following translation (Green, 2011; Gruber et al., 2015; Dorrell et al., 2017; Gruber and Kroth, 2017).

Previous phylogenomic studies of diatom genomes have revealed a chimeric origin of the diatom nucleus-encoded and plastid-targeted proteome (Oborník and Green, 2005; Nonoyama et al., 2019). These plastid-targeted proteins include not only proteins of red algal (i.e., endosymbiont) origin, but also proteins derived from green algae, bacteria, the host cell, and even a large number of proteins specific to this lineage without obvious homologues elsewhere in the tree of life (Dorrell et al., 2017; Nonoyama et al., 2019). This diverse combination of genes from different origins has also contributed to metabolic innovations in the diatom plastid distinctive from those of plants (Marchand et al., 2018; Nonoyama et al., 2019). These include a complete plastid-targeted ornithine cycle, which interacts with a complete mitochondria-targeted urea cycle in diatom amino acid metabolism and recycling (Allen et al., 2011; Nonoyama et al., 2019); a complex suite of plastid-targeted proteins involved in iron storage, acquisition, and stress tolerance (Gao et al., 2021); and elaborate CO_2 concentration and carbon metabolism systems (e.g., a complete glycolysis-gluconeogenesis pathway) not known in plant plastids (Marchand et al., 2018; Nonoyama et al., 2019). Moreover, many of these diatom-associated plastid metabolic innovations, including but not limited to central nitrogen and carbon metabolism, depend on intricate crosstalk between diatom plastids and mitochondria, which typically show close proximity to one another in diatom cells (Prihoda et al., 2012; Bailleul et al., 2015; Uwizeye et al., 2020). These plastid-related novel metabolic activities may depend on transporter innovations across the four membranes, or metabolite exchanges between organelles (Shai et al., 2016; Dorrell et al., 2017; Mix et al., 2018).

Although previous studies have focused on diatom plastid metabolic innovations (Bailleul et al., 2015; Marchand et al., 2018; Nonoyama et al., 2019; Gao et al., 2021), little is known specifically concerning their plastid transporter diversity, molecular and environmental functions. Marchand et al. presented overviews of the localization and activity of ion and metabolite plastid transporters from algae to land plants (Marchand et al., 2018; Marchand et al., 2020); and Brownlee et al. recently summarized key mechanisms of diatom key nutrient transport and acquisition (Brownlee et al., 2022). Specific studies have further elaborated on the diversity of diatom plastid sugar transporters (which typically transport triose phosphates) (Moog et al., 2015; Moog et al., 2020); low CO_2 induced HCO_3^- transporters (SLC) (Nakajima et al., 2013; Matsuda et al., 2017; Tsuji et al., 2017); nitrate/peptide transporters (Santin et al., 2021); and nucleotide triphosphate transporters (Ast et al., 2009).

Finally, diatoms and their plastid metabolism may act as important bellwethers of environmental and climate change (Spetea et al., 2014). Previous studies have demonstrated that both geochemical factors (e.g., iron and copper concentrations) (Hippmann et al., 2020; Kong and Price, 2021; Turnšek et al., 2021) alongside physical factors (pH, CO_2 availability, and temperature) (Tong et al., 2021; Zhong et al., 2021) directly influence diatom photosynthetic activity and abundance. Understanding the roles of specific plastid transporters in these responses, such as the production and accumulation of plastid metabolites and compatible ions implicated in stress tolerance (Marchand et al., 2018), and intracellular

communication between various subcellular compartments may allow more nuanced prediction of diatom responses to dynamically changing environments (Murik et al., 2019; Hippmann et al., 2020).

In this study, we use bioinformatic techniques to profile the predicted functions of 70 plastid-targeted transporters inferred from the well-characterized genome of the model diatom *Phaeodactylum tricornutum*, with complete models realised for > 99% of its probable encoded genes *via* transcriptomic and proteogenomic reannotations (Bowler et al., 2008; Rastogi et al., 2018; Yang et al., 2018); alongside extensive gene expression and epigenomic resources to understand its encoded functions (Veluchamy et al., 2013; Veluchamy et al., 2015; Ashworth et al., 2016; Ait-Mohamed et al., 2020; Zhao et al., 2020). Considering *Phaeodactylum*-specific gene expression trends, phylogenetic similarity to transporters from other organisms, and the broader environmental expression dynamics of homologues identified within the *Tara* Oceans survey (Villar et al., 2018), we profile the probable biological, metabolic and eco-physiological diversity of the diatom plastid transporter repertoire, and identify new candidate transporters for diatom-specific plastid metabolic activities and environmental resilience. Our study is the first systematic exploration of plastid transporter diversity across diatom algae.

Materials and methods

Comparison of plastid transporters from *Arabidopsis thaliana* and *Phaeodactylum tricornutum*

A list of 77 A. *thaliana* plastid transporters (Supplementary Table S1) was obtained from ChloroKB (Gloaguen et al., 2017; Gloaguen et al., 2021) and integrated with other characterized inorganic ion transporters from the literature (Finazzi et al., 2015; Marchand et al., 2018). Transporter gene functions were determined from the literature (Finazzi et al., 2015; Marchand et al., 2018), as integrated into TAIR (accessed 06/2021) (Huala et al., 2001).

A list of 70 plastid transporter genes (Supplementary Table S2) was found in the version 3 annotation of the *P. tricornutum* genome (Bowler et al., 2008; Yang et al., 2018). These transporters were identified through the presence of plastid-targeting sequences, inferred *in silico* using ASAFind version 2.0 (Gruber et al., 2015) with SignalP 3.0 (Bendtsen et al., 2004) and HECTAR (Gschloessl et al., 2008), run under default conditions; and the presence of transporter functions inferred annotated by KEGG (using BLASTkoala) (Kanehisa et al., 2016), Gene Ontology Annotation (Harris et al., 2004) or PFAM (using InterProScan) (Mulder and Apweiler, 2007).

BLASTp v 2.12.0 (Altschul et al., 1990; Hernández-Salmerón and Moreno-Hagelsieb, 2020) was used to find potential homologues of candidate *P. tricornutum* plastid transporters in the predicted protein models encoded by the complete genome of *A. thaliana* with threshold percentage identity ≥ 30% and E-value ≤ 1e-05, and vice-versa. The localization of *P. tricornutum* homologues of *A. thaliana* transporters was inferred based on *in silico* prediction, using ASAFind and HECTAR as before; MitoFates, using threshold value 0.35 (Fukasawa et al., 2015); and WolfPSort (Horton et al., 2006), using the consensus prediction of animal, fungi and plant models. Homologue localizations were divided into mitochondrial (M), plastid (P), plasma membrane or endoplasmic reticulum (PM/ER), and homologues with unclear localizations (e.g., N-incomplete gene annotations) were listed as undefined (U). The localizations of *A. thaliana* homologues of *P. tricornutum* plastid transporters, as inferred by BLASTp, were based on experimental data (https://suba.plantenergy.uwa.edu.au/; https://www.rostlab.org/services/locDB/) (Rastogi and Rost, 2011; Hooper et al., 2017), and when not available inferred from WolfPSORT (https://wolfpsort.hgc.jp/) using a plant reference dataset (Horton et al., 2006). Localizations were sorted as mitochondria (M), plastid (C), Vacuole (V), Golgi (G) and plasma membrane (PM).

A. thaliana plastid transporters were searched by SUBA (https://suba.live/), and manually checked in the literature if not cataloged. For *P. tricornutum*, plastid transporters that have been experimentally verified by GFP were identified from the literature (Gruber et al., 2015; Moog et al., 2015; Dell'Aquila et al., 2020), alongside *P. tricornutum* homologues (identified by genome-to-genome RbH) to plastid transporters experimentally identified from the purified plastid proteome of the diatom *Thalassiosira pseudonana* (Schober et al., 2019). All 67 A. *thaliana* plastid transporters and 7 P. *tricornutum* plastid transporters that have been experimentally verified (GFP localization, plastid proteome, mass spectroscopy) are noted in Supplementary Tables S1, S2.

Identification of metabolic pathways correlated with *P. tricornutum* plastid transporters

The complete version 3 annotation of the *P. tricornutum* genome (Bowler et al., 2008; Yang et al., 2018) was filtered for genes showing probable linked functions to each of the 70 plastid transporters, using a composite approach based on seven filtered conditions to generate a composite "score" for probable pathway linkage to each transporter. Four of these conditions related to gene co-regulation patterns: (i) the repartition of each gene into WGCNA (weighted genome correlation network analyses) modules in a previous meta-study of *P. tricornutum* RNAseq data (Ait-Mohamed et al., 2020), with genes that were retrieved in the same module as the query transporter scored as +1; and the crude Pearson Correlation coefficient of each genes to each

the query transporter calculated in (ii) *P. tricornutum* RNA-seq data (Ait-Mohamed et al., 2020), (iii) *P. tricornutum* microarray data integrated into DiatomPortal (http://networks.systemsbiology.net/diatom-portal/) (Ashworth et al., 2016), and (iv) the *Thalassiosira pseudonana* homologues (as defined by RbH search) of *P. tricornutum* genes in *T. pseudonana* microarray data (Ashworth et al., 2016), with genes retrieving any correlation coefficient > 0.5 scored as +1. In addition, three conditions relating to localization and evolution were considered: (v) the consensus *in silico* targeting predictions inferred with ASAFind and HECTAR (Gschloessl et al., 2008; Gruber et al., 2015), with proteins showing chloroplast targeting predictions scored as +1; and (vi) the inferred origin point and (vii) pattern of inferred loss of each transporter across the ochrophytes in a previously published RbH study (Nonoyama et al., 2019), with only genes that showed the same pattern of origin as the query gene (i.e., earliest ochrophyte common ancestor for which the presence of gene homologues could be inferred) or loss (i.e., monophyletic ochrophyte clades descended from the inferred common ancestor of the given gene in which the gene was not detected) scored as +1. Genes in the *P. tricornutum* genome for which a composite score of at least three was calculated were deemed to be associated to each query transporter (Supplementary Table S3).

Plastid metabolic pathways were annotated from previously constructed datasets (Nonoyama et al., 2019; Ait-Mohamed et al., 2020), PFAM, KEGG and KOG. Pathways for which more than three genes recovered a score ≥ +3 to a given transporter were considered to show a functional association with the transporter gene.

Correlation calculations of *P. tricornutum* plastid transporters with mitochondria-encoded genes

Pearson correlation coefficients were calculated using DiatomPortal microarray values for 34 genes in the *P. tricornutum* mitochondrial genome (Oudot-Le Secq et al., 2007; Ashworth et al., 2016), and 66 P. *tricornutum* plastid transporter genes for which corresponding microarray data were available. Microarray values were drawn from studies of light: dark acclimation dynamics in 17 different conditional processing groups (Nymark et al., 2009; Nymark et al., 2013; Valle et al., 2014); and the average of the correlation coefficients observed for each gene in the mitochondrial genomes was calculated for each transporter gene. Equivalent calculations were not performed for RNAseq data (Ait-Mohamed et al., 2020) due to the exclusion of mitochondrial transcripts from polyA-selected RNA libraries; or from *Thalassiosira* equivalents of *Phaeodactylum* genes due to the absence of generated mitochondrial microarray probes for this species (Ashworth et al., 2016).

Evolutionary conservation of *P. tricornutum* plastid transporters

70 plastid transporters were searched by RbH against 287 genomes and transcriptomes belonging to 11 taxonomic groups across the stramenopiles (diatoms, ochrophytes, and aplastidic groups), cryptomonads and haptophytes (Supplementary Table S4), with threshold E-value< 1e-05 and percentage identity ≥ 30%. *In silico* targeting predictions were performed for each transporter using ASAFind, HECTAR and MitoFates as defined above. The deeper evolutionary history of each transporter was summarized from previously published data (Dorrell et al., 2021), considering horizontal, endosymbiosis-associated and endosymbiotic gene transfers into and out of the diatom and ochrophyte common ancestors, from both prokaryotic and eukaryotic sources.

Calculation of environmental expression trends associated with *P. tricornutum* plastid transporters

The relationships between *P. tricornutum* plastid transporters and environmental factors were analyzed by correlating the relative abundances of transporter homologous sequences to environmental factors from *Tara* Oceans Version 1 (Carradec et al., 2018) on the Ocean Gene Atlas (OGA) website (http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/) (Villar et al., 2018). The FASTA-formatted protein sequence of each plastid transporter was searched on the OGA portal by BLASTp with threshold E-value $1x10^{-05}$. Gene abundance estimates were calculated from the meta-genomes (MetaG) database, and transcript abundance estimates were calculated from the meta-transcriptomes (MetaT) database. Complete details of homologue sequences, taxonomic annotations, and environmental distributions of each diatom transporter meta-gene, as inferred by Ocean Gene Atlas website, are provided in the linked osf.io repository https://osf.io/89vm3/ in the folder "Transporters"

The resulting raw homologues downloaded from OGA website were filtered to only retrieve those homologue sequences to diatom plastid-targeted transporters. First, a reference dataset was constructed for each gene from all plastid-targeted transporters identified in stramenopile, haptophyte and cryptomonad libraries as defined above, and the best-matching sequence identified from 151 non-redundant taxonomic categories from a previously constructed whole-tree of life dataset (Dorrell et al., 2021), inferred by BLASTp using the query with threshold e-value 10^{-05}. The *Tara* Oceans sequences were searched against the reference dataset by BLASTp with the –max_target_seqs 1 threshold applied, and only sequences with diatom plastid-targeted best hits were retained. A second BLASTp was performed against the complete protein sequence

annotation models from the *P. tricornutum* genome using the same criteria, and only sequences that obtained best hits with the specific queried gene ID were retained. The intersection of each set of retained genes were selected and combined with the reference dataset, and aligned using MAFFT (Katoh et al., 2002) with the –auto setting. Poorly aligned sequences were manually removed, defined visually by <50% overlap with the conserved domain region of the query transporter, and a guide phylogenetic tree was generated for each curated alignment using the NJ tree-building function in Geneious v 10.0.9 (Kearse et al., 2012) and 100 replicates. Finally, Tara meta-genes corresponding to diatom plastid transporters in the NJ tree topology, to the exclusion of all other cultured homologues, were extracted for 65 plastid transporter families, and used for quantitative analysis (Supplementary Table S5). Complete fasta files of *Tara* Oceans homologues for each transporter, along with raw and curated alignments and tree topologies are provided in the linked osf supporting database https://osf.io/89vm3/ in the folder "Transporters".

MetaT and MetaG abundances of verified transporter homologues were calculated for all stations, all size fragments (0.8-5, 5-20, 20-180 and 180-2000µm) and different depths (DCM: deep chlorophyll maximum layer; SRF: upper layer zone). Absolute correlation coefficients were calculated by Pearson correlation analysis between the sum MetaT total relative abundance, the sum MetaG total relative abundance (with the exception of transporter J12788, for which no MetaG data were available), and measured environmental factors for each station, following data from PANGAEA (Pesant et al., 2015). A two-tailed *t*-test was used to determine if the MetaT-MetaG total was significantly positively or negative correlated to each environmental variable (Supplementary Table S5). P-values were shown as positive if $R^2_{metaT} > R^2_{metaG}$, and negative value if $R^2_{metaT} > R^2_{metaG}$; with significance threshold P= 0.05.

Finally, environmental parameters which were shown to be strongly positively or negatively correlated in MetaT compared to MetaG data (MetaT-MetaG) in at least three of the ten combinations of different size fractions and depth conditions (DCM<0.8; DCM 0.8-5; DCM 5-20; DCM 20-180; DCM 180-2000; SRF<0.8; SRF 0.8-5; SRF 5-20; SRF 20-180; SRF 180-2000), were inferred to be concluded to significantly related to the expression of the transporter. To allow a more intuitive view of the results, five pigment concentration parameters related to photosynthetic potential (Liu et al., 2009; Puissant et al., 2021); and net primary production, particulate inorganic, organic and total carbon, previously used as metrics for modeling carbon export from the ocean surface layer, were merged together as primary production-related categories (Guidi et al., 2016; Leblanc et al., 2018). A threshold of significant (P< 0.05) correlations in ≥ 3 depth and size fractions combinations for ≥ 8 of the ten selected parameters across this merged dataset was deemed to correspond to a significant relationship with primary production.

Results

Limited conservation of plastid transporters between *Arabidopsis thaliana* and *P. tricornutum*

To gain an overall appreciation of the similarity of the plastid transporters in diatoms to those of the more well-studied but less complex plastids of higher plants, the total plastid transporter proteomes of *A. thaliana* and *P. tricornutum* were compared using reciprocal BLASTp-BLASTp searches. A total of 77 A. *thaliana* plastid transporters were defined from the literature (Finazzi et al., 2015; Marchand et al., 2018), ChloroKB (Gloaguen et al., 2017), and NCBI (Wheeler et al., 2008) (Supplementary Table S1), whereas 70 predicted *P. tricornutum* plastid transporters were found based on (i) the presence of a transport-associated KEGG, GO or PFAM domain (Rastogi et al., 2018; Ait-Mohamed et al., 2020), and (ii) the presence of a plastid targeting sequence inferred by *in silico* prediction with ASAFind (Gruber et al., 2015) and HECTAR (Gschloessl et al., 2008) from the version 3 genome annotation (Supplementary Table S2). The 70 inferred *in silico Phaeodactylum* transporters include the small number of plastid transporters that have been experimentally verified by GFP localization or mass spectroscopy in diatoms (Supplementary Tables S1, S2), whereas only ten of the *A. thaliana* transporter localizations are still putative and have yet to be formally experimentally confirmed (Supplementary Table S1).

48 *A. thaliana* transporters, from a diverse range of families, were found to have possible homologues (BLASTp: E-value ≤ 1e-05, identity ≥ 30%) in the *P. tricornutum* proteome, but only 20 of these proteins have at least one homologue that was predicted to contain plastid targeting sequences in *P. tricornutum* using ASAFind or HECTAR as above (Figure 1A; Supplementary Table S1). In turn, while 36 of the *P. tricornutum* plastid transporters had probable BLASTp homologues in the *A. thaliana* genome, only 11 of these have homologues previously annotated as transporters in the *A. thaliana* plastid proteome (Figure 1A, Supplementary Table S1). Our understanding of which transporters are plastid-targeted in *P. tricornutum* depends largely on the quality of genome annotation in the absence of an experimentally resolved purified plastid proteome, although the version 3 annotation used is inferred to be largely N-terminally correct considering both transcriptomic and proteomic data (Rastogi et al., 2018; Schober et al., 2018; Yang et al., 2018).

The 20 A. *thaliana* plastid transporters that were found to correspond directly to 11 P. *tricornutum* plastid transporters, matched with both one-to-one and one-to-many correspondences (Figures 1A, B, Supplementary Tables S1, S2). These one-to-many correspondences probably reflect independent duplications of genes encoding plastid transporters in both the *A. thaliana* and *P. tricornutum*

10.3389/fpls.2022.950467

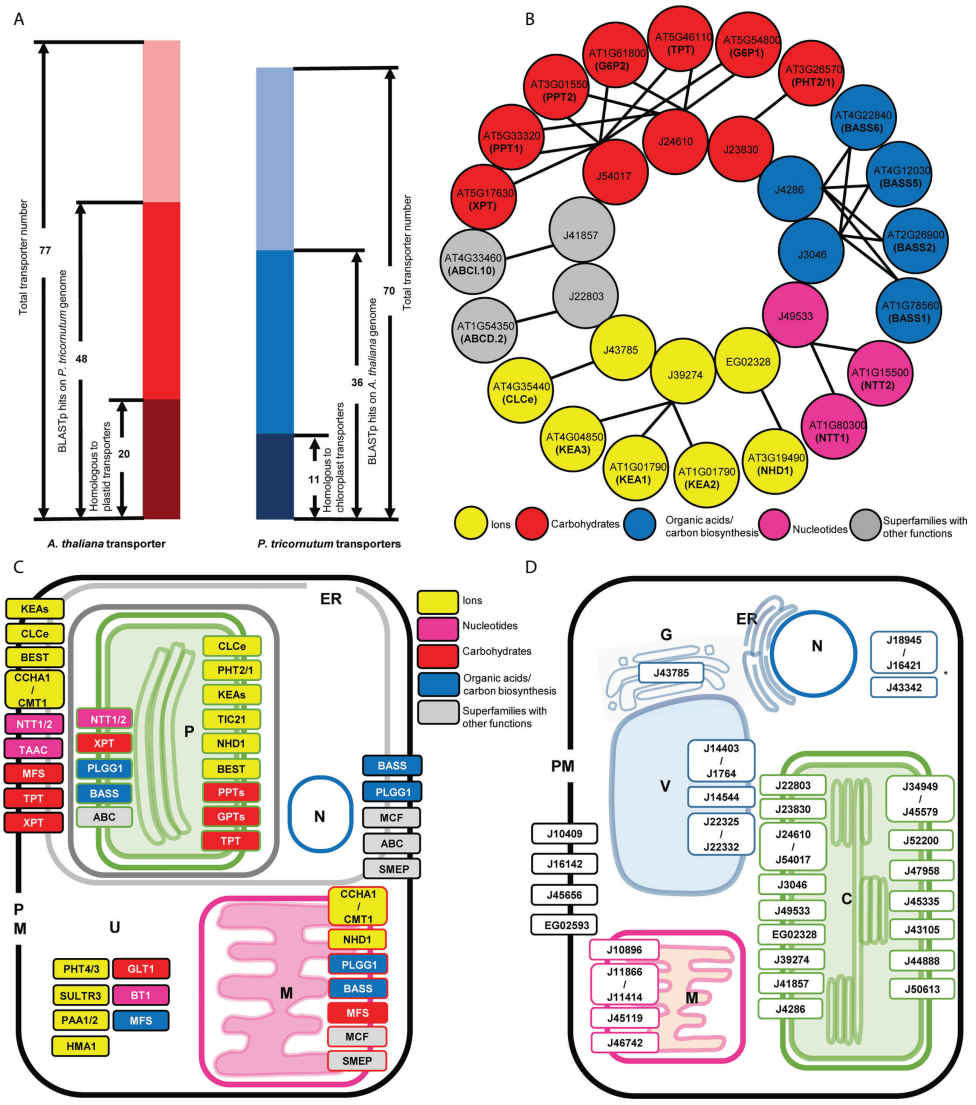

FIGURE 1
Comparison of *thaliana* and *P. tricornutum* plastid transporters. **(A)** The number of BLASTp homologues identified between 77 A. *thaliana* plastid transporters and 70 P. *tricornutum* plastid transporters. Dark red and blue represent transporters that had BLASTp homologue annotated as transporters within the searched genome; red and blue represent the number of plastid transporters that had BLASTp homologues (e-value < 1E-5; identity >30%) not annotated as transporters within the searched genome; and light red and light blue represent the number of transporters that do not have identifiable BLASTp homologues in the searched genome. **(B)** Transporter BLASTp equivalent matches in detail. One-to-one and one-to-many correspondences are indicated by arrow lines; circle colors represent functional categories, with grey indicating transporter superfamilies that have different functions and thus have uncertain function. **(C)** Predicted localization of *thaliana* plastid transporter homologues in *P. tricornutum*. Transporters are named according to the common abbreviation/gene family name in *thaliana*, and their localization in *P. tricornutum* are shown according to homologue predictions from ASAFind, HECTAR and MitoFates, ER, endoplasmic reticulum, P, plastid, N, nucleus M, mitochondria, PM, plasma membrane, U, unknown/undefined. Colored and grey boxes represent as above. Multiple localisations indicate transporters for which *P. tricornutum* homologues were inferred to localise to more than one organelle. **(D)** Predicted localization of *P. tricornutum* plastid transporter homologues in *thaliana*. Transporters are named according to the gene ID in the version 3 *P. tricornutum* genome annotation, and their localization in *thaliana* are shown according to experimental data (https://suba.plantenergy.uwa.edu.au/ or https://www.rostlab.org/services/locDB/), or when not available inferred from WolfPSORT, C, plastid, ER, endoplasmic reticulum, G, Golgi, M, mitochondria, PM, plasma membrane, V, Vacuole, N, nucleus, *, undefined targeting.

genomes (Jaillon et al., 2007; Parks et al., 2018). The homologue equivalent transporters found between *A. thaliana* and *P. tricornutum* perform key plastid functions, including the nucleotide transporter family (NTT1/2, matching *Phaeodactylum* transporter J49533) essential for the import of NTPs from the cytoplasm into plastid (Ast et al., 2009) (Figure 1B). Three plastid transporters belonging to the AtKEA family (KEA1, KEA2, KEA3) that function as K^+/H^+

antiporters and play essential roles in plastid development, ion homeostasis, and photosynthesis (DeTar et al., 2021), correspond to one *P. tricornutum* plastid transporter J39274. Four members of the BASS family (Bile acid/sodium symporter-like transporter; BASS1, BASS2, BASS5, BASS6) which exhibit a wide range of substrate specificities, including non-bile acid organic compounds such as pyruvate, steroids, methionine-derived glucosinolates (GSL) and exogenous substances (South et al., 2017), showed equivalence to one sodium/bile acid transporter (J4286) and one sodium/pyruvate transporter (J3046) in the *P. tricornutum* genome (Figure 1B).

Two annotated *P. tricornutum* triose phosphate transporters (J24610, J54017) retrieved BLASTp results with the *A. thaliana* TPT, glucose 6-phosphate transporter (GPT), xylulose 5-phosphate transporter (XPT), and phosphoenolpyruvate transporters (PPT; Figure 1B). This latter result supported a recent study of plastid phosphate transporters in the non-photosynthetic diatom *Nitzschia* sp. NIES-3581, which suggests that TPT transporters are capable of transporting multiple metabolites including triose phosphates (TPs) and phosphoenolpyruvate (PEP) (Moog et al., 2020), whereas no recognizable hexose phosphate transport system is associated with diatom plastids (Moog et al., 2015). An ABC supergroup transporter ABCI.10 (AT4G33460) putatively involved in metal homeostasis retrieved the *P. tricornutum* plastid transporter J41857 as the equivalent (Figure 1B).

The RbH results provide insights into the putative functions of specific *P. tricornutum* transporters that have not previously been annotated. These include J39274 which was annotated within the version 3 P. *tricornutum* genome as an Na^+/H^+ transporter by CDD search and PFAM; but in our search shows more similarity to KEA family transporters. We note that J39274 has recently and independently been functionally characterised as the diatom KEA3 (Seydoux et al., 2022). We further identify a phosphate permease J23830, which was found to show homology to the *A. thaliana* phosphate transporter PHT2/1, previously believed to be restricted to green and red algae and not found elsewhere in the tree of life (Pfeil et al., 2014; Marchand et al., 2018). Other key *A. thaliana* plastid transporters were either not detected, or only possess non-plastid targeted homologues in *P. tricornutum* such as the $SO4^{2-}/H^+$ antiporter superfamily SULTR (Cao et al., 2013) and metal transporter P-type ATPases (PAA1, PAA2) (Hanikenne and Baurain, 2013).

Differential localizations of diatom homologues of plant plastid transporters

Over their evolutionary history, proteins can undergo re-localization: proteins currently targeted to the plastid may have originally been recruited from other compartments inside the host cell (Liu et al., 2014), while plastid-targeted proteins may also be relocated to support other host organelles (Martin et al., 2002; Dorrell et al., 2019). This phenomenon may have occurred with *A. thaliana* and *P. tricornutum* plastid transporters, where some homologues (BLASTp: E-value ≤ 1e-05, identity ≥ 30%) apparently displayed a different localization in each species (Figure 1A; Supplementary Table S1, S2). Based on *in silico* localization predictions and experimental data, the predicted localizations of the *P. tricornutum* homologous sequences of *A. thaliana* plastid transporters were mapped in the *P. tricornutum* cell. These were found to be distributed across the plastid (P), mitochondria (M), plasma membrane (PM) and endoplasmic reticulum (ER) (Figure 1C; Supplementary Table S1).

Previous studies of plastid localizations have been inferred experimentally for certain TPTs (J24610 and J54017) (2005; Kilian and Kroth, 2004; Moog et al., 2015) and NTT1 (J49533) (Ast et al., 2009) in *P. tricornutum*, coherent with *in silico* predictions. Moreover, by reciprocal BLAST best hit (RbH), six *P. tricornutum* plastid transporters (a probable Na^+/H^+ antiporter family protein EG02328, Tic110 J50540, TPT J24610, a formate/nitrite transporter J13076, NTT1 J49533, and a sodium/pyruvate transporter J3046) were found to be homologous to plastid transporters (Supplementary Table S2) identified in an experimental plastid proteome study of the diatom *Thalassiosira pseudonana* (Schober et al., 2019), The closest *P. tricornutum* homologues of the *A. thaliana* glucose-6-phosphate and phosphate transporters (PHT2/1, PPTs, GPTs: AT3G26570, AT5G33320 and AT3G01550, AT5G54800 and AT1G61800) (Supplementary Table S2) were further inferred to have exclusively plastid localizations by *in silico* prediction (Figure 1C), although as discussed above these homologues may be triose phosphate transporters (Moog et al., 2015; Moog et al., 2020).

Other *P. tricornutum* homologues of *A. thaliana* transporters did not show exclusive plastid localization, but were also inferred to localize to the PM/ER (Figure 1C), such as probable homologues of the KEA K^+/H^+ antiporters family (AT1G01790, AT4G04850 and AT1G01790). The *P. tricornutum* homologues of a thylakoid ADP/ATP carrier protein (AT5G01500), implicated in both ATP/ADP and 3'-phosphoadenosine 5'-phosphosulfate (PAPS) transport across the *A. thaliana* plastid envelope (Thuswaldner et al., 2007; Gigolashvili et al., 2012), were only predicted to be localized to the *P. tricornutum* PM/ER (Figure 1C). At least some of these transporters may localize to the *P. tricornutum* cERM and therefore still be implicated in metabolite exchange with the *P. tricornutum* plastid. In the case of the *P. tricornutum* PAPS transporter, a complete plastid cysteine synthesis pathway from PAPS reductase to cysteine synthase is known for diatoms, but no plastid-targeted adenosine sulphate kinase is known, indicating PAPS is likely to be imported from the cytoplasm

across the cERM (Dorrell et al., 2017). Finally, we note that one putative phosphate permease (J23830) in our dataset has previously been GFP localized to the *P. tricornutum* cERM, and not the plastid (Dell'Aquila et al., 2020). It remains to be determined if this discrepancy relates to a possible dual-localisation or mis-targeting considering either overexpression or *in silico* prediction.

Moreover, some *P. tricornutum* homologues of *A. thaliana* plastid transporters showed mitochondrial localization predictions (Figure 1C), e.g., homologues of glycolate/glycerate translocator 1 (PLGG1; AT1G32080) (Pick et al., 2013), the GDT1-family Ca^{2+}/H^{+} antiporter CCHA1 and CMT1 (AT1G64150 and AT4G13590) (Wang et al., 2016), and members of the BASS superfamily (AT2G26900, AT4G22840, AT4G12030, AT1G78560) (Figure 1C). These may represent distant mitochondrial homologues of *A. thaliana* plastid transporters, with either plastid, mitochondrial, or alternative ancestral localizations.

The subcellular locations of the *A. thaliana* homologues of 36 P. *tricornutum* plastid transporters included the plastid (C), but also the plasma membrane (PM), mitochondria (M), Golgi body (G) and vacuole (V) (Figure 1D, Supplementary Table S2).

Five *P. tricornutum* transporters were inferred to possess dual plastid/mitochondria-targeting sequences by *in silico* prediction (Gile et al., 2015; Dorrell et al., 2017), i.e., a plastid-targeting prediction with either ASAFind or HECTAR, and a mitochondria-targeting prediction with either HECTAR or MitoFates. These proteins were: a probable Na^{+}/H^{+} antiporter family protein (EG02328); a transporter of unknown function (EG02514); a multidrug and toxic compound extrusion family/MATE efflux family protein (J35587); a glycolipid transporter (J36726); and a sodium/bile acid transporter (J4286) (Supplementary Table S2), suggesting more intricate metabolic interactions between both organelles in the diatom cell.

Transcriptional co-regulation reveals metabolic pathway linkages of *P. tricornutum* plastid transporters

To classify putative roles for the diverse range of transporters associated with the *P. tricornutum* plastid with no clear homologues in *A. thaliana*, we inferred probable biochemical functions for each transporter. Large-scale microarray and RNA sequencing data (Ashworth et al., 2016; Ait-Mohamed et al., 2020); alongside other annotations such as inferred evolutionary origin and loss (Nonoyama et al., 2019; see Materials and Methods) were availed to identify metabolic pathways associated with each plastid transporter in the *P. tricornutum* genome (Supplementary Table S3).

Phaeodactylum plastid-related processes, established from previous studies (Ait-Mohamed et al., 2020) were divided into five functional groups (Supplementary Table S3): photosynthesis-related (relating either to light acquisition, photosynthetic electron transport, chlorophyll or carotenoid biosynthesis), amino acid-related (lysine, chorismate, branched-chain amino acid and cysteine synthesis; alongside Fe-S cluster synthesis, as a necessary precursor to cysteine synthesis, and aminoacyl-tRNA activation as necessary for protein synthesis), biogenesis-related (ribosomal proteins, protein import, and DNA-associated proteins), carbon-related (carbohydrate transport, Calvin cycle and glycolysis/gluconeogenesis pathways) and lipid-related (fatty acid synthesis and lipid head group exchange) (Figure 2). 22 transporters were found to be associated with all five major processes considered, suggesting that they are likely to have central roles in diatom plastid metabolism (Figure 2). Other transporters showed associations only with specific pathways. For example, J46742 and J34949 were found to only be associated with amino acid-related pathways; J32535, J36726, J36253, J44665 and J4286 were associated only with photosynthesis-related pathways; EG00042, J32491, J46051, J43785 were associated only with biogenesis-related functions; and EG02514 was associated only with fatty acid/lipid related functions. No transporters were found to be associated exclusively with carbon-related functions (Figure 2), reflecting the importance of central carbon metabolism for other plastid metabolic processes (e.g., amino acid and lipid biosynthesis) (Nonoyama et al., 2019; Ait-Mohamed et al., 2020).

Identification of *P. tricornutum* plastid transporters potentially associated with mitochondrial crosstalk

The intricate metabolic connections between diatom plastids and mitochondria (Prihoda et al., 2012; Bailleul et al., 2015; Uwizeye et al., 2020) likely depend on transporters that transfer metabolites either between the two organelles, or with adjacent compartments such as the peroxisome (Shai et al., 2016; Dorrell et al., 2017; Mix et al., 2018). In order to find potential plastid transporters that may be related to this cross-talk, the average Pearson correlation coefficient values were calculated between 66 plastid transporters and 34 genes from the mitochondria genome, for which equivalent relative fold-expression change data were present in published microarray data (Oudot-Le Secq et al.; Ashworth et al., 2016).

A total of 37 plastid transporters were found to have an average mitochondrial Pearson correlation > 0, and eight plastid transporters showed average values > 0.5, considering both correlation to mitochondrial respiratory complex and biogenesis genes (Figure 3). Only three plastid transporters (NTT1 ATP/ADP transporter J49533, triose phosphate

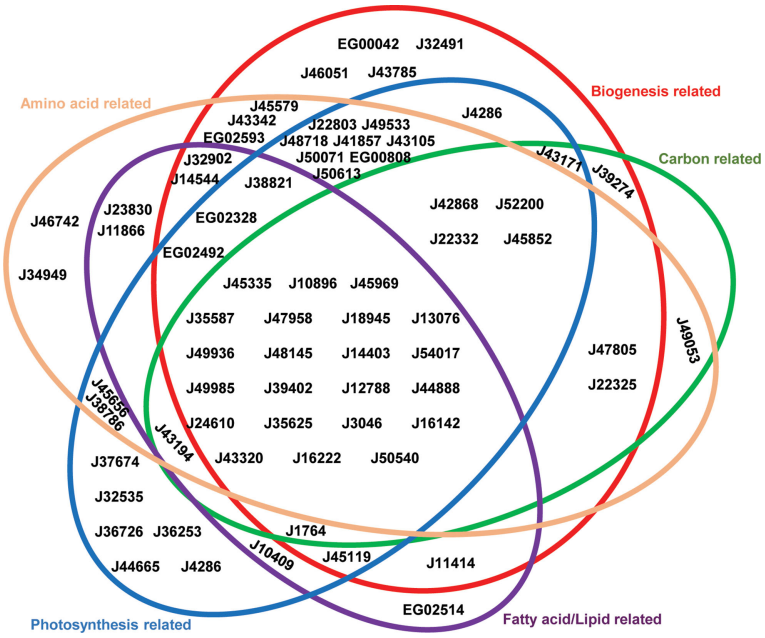

FIGURE 2
Venn Diagram showing associated pathways of *P. tricornutum* plastid transporters. Plastid-transporter associated pathways were identified from a set of associated genes in the *P. tricornutum* genome, considering seven co-regulation, evolutionary and localization conditions (details in Materials and Methods). Each metabolic pathway annotated from PFAM, KEGG, KOGG, was manually merged into five main categories, represented as different colored circles. These are: photosynthesis-related = photosynthetic and light-harvesting proteins, enzymes involved in chlorophyll and carotenoid biosynthesis, and mitochondrial respiration-related pathways; biogenesis-related = protein subunits of the plastid ribosome, protein import, division, DNA transcription and replication machineries; carbon-related = enzymes with KOGG annotations related to carbohydrate transport and metabolism; fatty acid/lipid-related = enzymes with KOGG annotations related to fatty acid or lipid metabolism; and amino acid-related = enzymes with KOGG annotations related to amino acid transport and metabolism, tRNA activation and Fe-S cluster synthesis. Transporters are shown as gene IDs in the version 3 annotation of the *P. tricornutum* genome.

transporter J24610, MFS transporter J18945) that showed Pearson correlation values > 0.5 retrieved BLASTp homologues with E-value ≤ 1e-05 against the *A. thaliana* genome (Figure 3; Supplementary Table S2). The remaining five had no obvious homology to *A. thaliana* proteins, suggesting that many of the transporters implicated in plastid-mitochondria crosstalk may be unique to diatoms, or at least not conserved with plants.

The strongly mitochondrially co-regulated transporters included four containing the MFS domain (J42868, J18945, J49533, J49936), and a triose phosphate transporter (J24610) (Figure 3; Supplementary Table S2). These results underline the probable roles of carbon metabolism in mediating diatom plastid-mitochondria interactions (Broddrick et al., 2019). The majority of these strongly mitochondrially co-regulated plastid transporters seem to be associated with more than one fundamental category in *P. tricornutum* metabolism, with five of them (J50540, J18945, J45969, J49936, J24610) linked to all five pathway categories (Figures 2, 3). None of the remaining strongly mitochondrially co-regulated transporters showed functional associations with lipid-related metabolism (Figures 2, 3), suggesting that lipids play more limited roles in diatom mitochondria-plastid crosstalk.

Identification of diatom-specific plastid transporters from multispecies sequence datasets

Given the limited conservation observed with *A. thaliana*, we wished to understand the deeper evolutionary conservation and gene transfer events underpinning the origins of *P. tricornutum* plastid transporters. First, to assess the immediate evolutionary conservation of each transporter, we identified homologous sequences of all 70 P. *tricornutum* plastid transporters by RbH amongst a previously assembled composite library of more than two hundred species (Nonoyama et al., 2019; Dorrell et al., 2021) sampled from across stramenopiles, the broader group of organisms including diatoms. The composite library was further divided into three groups: diatoms (106 species), other members of the ochrophytes that contain plastids (94 species) and are

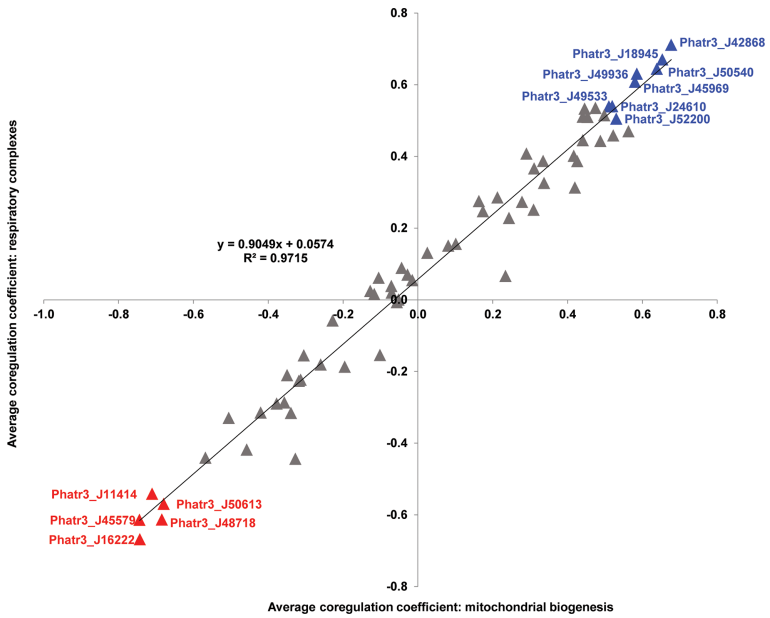

FIGURE 3
Scatterplot of Pearson correlation coefficients between plastid transporters and mitochondria-encoded genes. Horizontal axes correspond to average correlation coefficients calculated between *Phaeodactylum* transporter genes and mitochondrial genes involved in biogenesis (encoding ribosomal and translocase proteins), and vertical axes correspond to average correlation coefficients calculated between *Phaeodactylum* transporter genes and mitochondrial respiratory complex genes across 17 conditional groups in published microarray data studies (Ashworth et al, 2016; *ibid*). The strong positive correlation coefficient between both axes (*t*-test $P < 10^{-05}$) indicates strong overall reproducibility of correlation, independent of mitochondrial gene function. Eight transporter genes that recover average correlation coefficients > 0.5 and five that recover average correlation coefficients < - 0.5 for both axes are labelled in blue and red text, respectively.

predominantly photosynthetic, and aplastidic groups (23 species) that lack functional plastids. The composite library was further enriched with members of the cryptomonads (27 species) and haptophytes (37 species) which, like the photosynthetic members of the stramenopiles, have plastids derived from the secondary endosymbiosis of red algae and may share a common origin with the ochrophyte plastid, and are surrounded by four membranes (Dorrell and Smith, 2011). As each of these lineages possess similar plastid structures, plastid-targeted proteins can be predicted for them *in silico* using common tools HECTAR (Gschloessl et al., 2008) and ASAFind (Gruber et al., 2015), and RbH searches were also performed explicitly for proteins with inferred plastid-targeted sequences in each species.

Unsurprisingly, given their evolutionary proximity, raphid pennate diatoms (other than *P. tricornutum*) were found to possess the greatest number (63) of homologues of *P. tricornutum* plastid transporters (Figure 4A); followed by other diatoms within the Thalassiosirales, Chaetocerotales, araphid pennate and radial centric lineages, for which 48-58 transporter homologues were detected (Figure 4A). The next highest numbers (40-44) of transporters with homologues were found in pelagophyte/dictyochophyte, pinguiophyte/chrysophyte, and haptophyte groups (Figure 4A). Fewer transporters were inferred to have homologues in ochrophytes within the PX/raphidophytes (33) and cryptomonads (31), and only five were found to have homologues in plastid-lacking stramenopiles (Figure 4A). As this last group is projected to have never acquired a red algal plastid (Dorrell et al., 2017; Stiller et al., 2009), this likely indicates that the majority of *P. tricornutum* plastid transporters have explicitly plastid-related evolutionary origins.

Using this approach, five transporters (J45852, J50540, J42868, EG02492, and J35625) were inferred to be conserved across, but unique to diatoms (Figure 4A). None of these five diatom unique transporters retrieved homologues in the *A. thaliana* genome (Figure 1A; Supplementary Table S2), while two (J42868: MFS domain transporter; J50540: plastid transport protein Tic110) showed the strongest and third strongest positive transcriptional correlations with the mitochondrial genome (Figure 3), suggesting possible diatom-specific effectors or regulators of plastid-mitochondria crosstalk. All of these diatom-specific transporters were transcriptionally co-regulated with at least four of the five functional categories of diatom plastid metabolism studied (Figure 2), suggesting that they play central roles in diatom cell metabolism.

Finally, we reanalyzed single-gene trees presented in a previously published phylogenetic study, which considered all transfer events, including genes received from and donated into

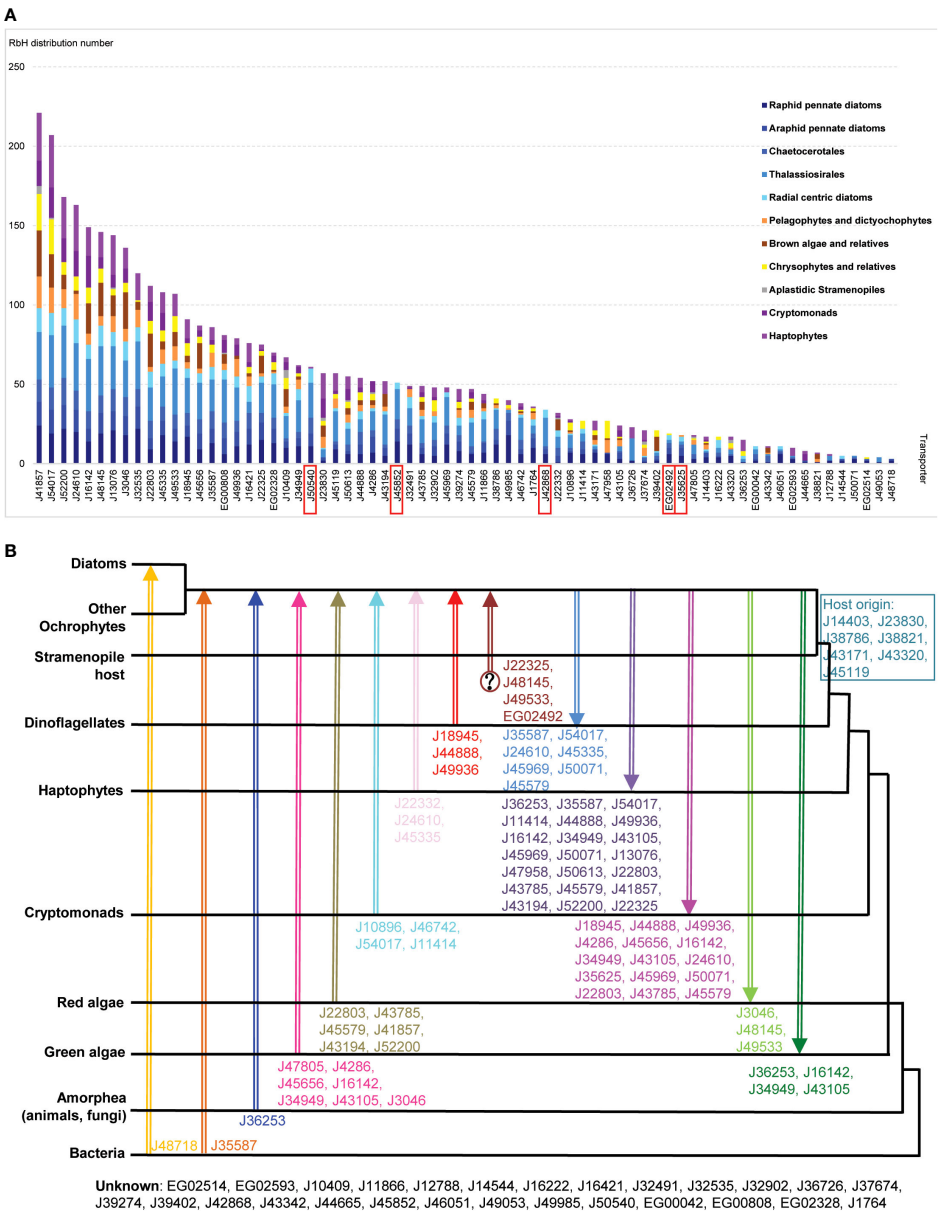

FIGURE 4
Evolutionary relationships between *P. tricornutum* plastid transporters. **(A)** Bar chart of the numbers of *P. tricornutum* plastid transporters that retrieved RbH over at least two species from 11 taxonomic groups (Materials and Methods; Supplementary Table S4). Taxonomic categories are grouped as diatoms (blue); other ochrophytes (yellow/orange); aplastidic stramenopiles (grey) or other lineages with secondary red plastids surrounded by four membranes (purple). Transporter genes inferred to be conserved across diatoms (present in >1 species in all diatom taxonomic categories considered) but not found elsewhere (present in < 4 non-diatom species, and no more than two in any one non-diatom taxonomic category) are labelled with red rectangles. **(B)** Phylogram of *P. tricornutum* transporters based on single-gene trees for each transporter with a taxonomically exhaustive reference dataset, per a prior phylogenomic study (Dorrell et al., 2021). Different gene transfer events between different taxonomic groups are shown by arrow colour and orientation, with transporter predicted to have undergone this gene transfer event are shown next to each arrow. Transporter genes inferred to have been vertically inherited from the aplastidic stramenopile host are displayed in the blue box, and transporters with unknown origins are shown at the bottom.

other lineages, visible in the *P. tricornutum* genome (Dorrell et al., 2021) and extending to the last common ochrophyte ancestor. The results are presented in Figure 4B and Supplementary Table S4. Six transporter genes (J22803, J43785, J45579, J41857, J43194, J52200) were predicted to have arisen *via* transfers from red algae into ochrophytes (Figure 4B). A further ten transporters were predicted to be transferred from algae with secondary red plastids (cryptomonads, haptophytes

or dinoflagellates) into the ochrophyte common ancestor, reflecting the red algal origin of the *P. tricornutum* plastid (Figure 4B). Seven transporter genes (J47805, J4286, J45656, J16142, J34949, J43105, J3046) likely originated from green algae into ochrophytes, while two transporters (J48718: a permease, and J35587: a multidrug and toxic compound extrusion family protein) were inferred to have arisen from horizontal transfers from bacteria into, respectively, diatoms specifically, or all ochrophyte lineages. These latter transporter groups reflect the chimeric composition of the diatom plastid proteome, which is supported by nucleus-encoded and plastid-targeted proteins of red, green, host and bacterial origin (Oborník and Green, 2005; Nonoyama et al., 2019).

Environmental correlations with abundance of diatom plastid transporters in *Tara* Oceans data

We wished to identify the expression trends of diatom plastid transporters under different environmental factors, and find transporters that show unique trends with environmental variation. We extracted meta-transcriptome homologues of *P. tricornutum* plastid transporters from version 1 of the *Tara* Oceans Expedition Ocean Gene Atlas (Villar et al., 2018), and identified sequences that corresponded specifically to diatom plastid-targeted proteins by phylogenetic approaches (Supplementary Table S5). A total of 65 of the 70 P. *tricornutum* transporters were found to possess multiple environmental homologues, and were selected for subsequent quantitative analysis of meta-transcriptome (MetaT) and meta-genomes (MetaG) abundance against sampled environmental variables at each depth and size fraction. Krona Plots concerning the nearest homologue from cultured species of all phylogenetically identified diatom transporter meta-gene are provided in Supplementary Figure S1. Confirming the accuracy of the phylogenetic identification approach employed, the majority (90%) of these homologues correspond to stramenopiles, and 97% of these stramenopiles belong to diatoms. None were precisely annotated as *P. tricornutum*, reflecting its rarity in environmental samples (Malviya et al., 2016).

Many of the diatom plastid transporters showed similar correlations to different environmental factors (Figure 5A), e.g., positive transcriptional correlations with primary production-related parameters (pigments, net primary production, particulate inorganic, organic and total carbon), and negative correlations with iron, pH, and temperature (Figure 5A). These trends likely reflect diatom environmental preferences for high-latitude, nutrient-rich but iron-limited, and highly productive environments (Malviya et al., 2016; Benoiston et al., 2017; Nonoyama et al., 2019; Young and Schmidt, 2020).

Other environmental factors showed more variable relationships with diatom plastid transporters. For example, colored dissolved organic matter (CDOM), showed a strong positive correlation ($0< P< 0.05$ observed between a parameter and transporter expression abundance across ≥ three combinations of depth and size fraction) with two transporters, and strong negative correlations ($-0.05< P< 0$ observed between a parameter and transporter expression abundance across ≥ three combinations of depth and size fraction) with five transporters (Figure 5A; Supplementary Table S5). The transporters that showed strong positive environmental correlations to CDOM (EG02593, J32535) appeared to have more specific functional relationships in our Venn diagram of *P. tricornutum* cell metabolism, while the transporters that showed strong negative relationships to CDOM (EG00808, J32902, J16421, J43105, J48145) were found to be associated with multiple *P. tricornutum* plastid metabolism pathways (Figures 2, 5). These relationships may identify novel biomarkers of diatom plastid functional health, and of carbon export from the ocean surface layer (Guidi et al., 2016; Sunagawa et al., 2020).

Identification of specific diatom plastid transporters linked to individual marine processes

Despite these global biases, we noted distinctive environmental expression trends for different diatom plastid transporters. We summarize the significant relationships for each transporter in Figure 5B, grouping the environmental parameters into four functional categories: temperature, iron, pH and ten merged primary production related parameters (see Materials and Methods). No transporter was significantly correlated with all of the environmental parameters studied (Figure 5B), indicating probable physiological partitioning of transporter functions in relation to different environmental factors. The two transporter genes showing the strongest positive and negative relationships to primary production were, respectively, J50540 (plastid transport protein Tic110) and J42868 (MFS domain transporter) (Figure 5A; Supplementary Table S5). These two genes were also among the two most strongly correlated plastid transporters to the *P. tricornutum* mitochondrial genome (Figure 3), underpinning the importance for plastid-mitochondrial crosstalk in supporting diatom photosynthetic activity, and potentially even in antagonistic interactions with primary production. Other plastid transporter genes showed strong negative relationships with pH (mitochondrial carrier protein J46742 and ABC transporter J41857) and iron (J48145, an EamA-like/drug-metabolite transporter) providing potential new biomarkers for understanding diatom responses to ocean acidification and biological competition (Figure 5A; Supplementary Table S5).

One transporter, J43171, encoding an ion channel protein with an EF-hand calcium-binding motif, uniquely showed a weak positive relationship with temperature in the *Tara* Oceans data (Figures 5A, B). This transporter was found to be associated to

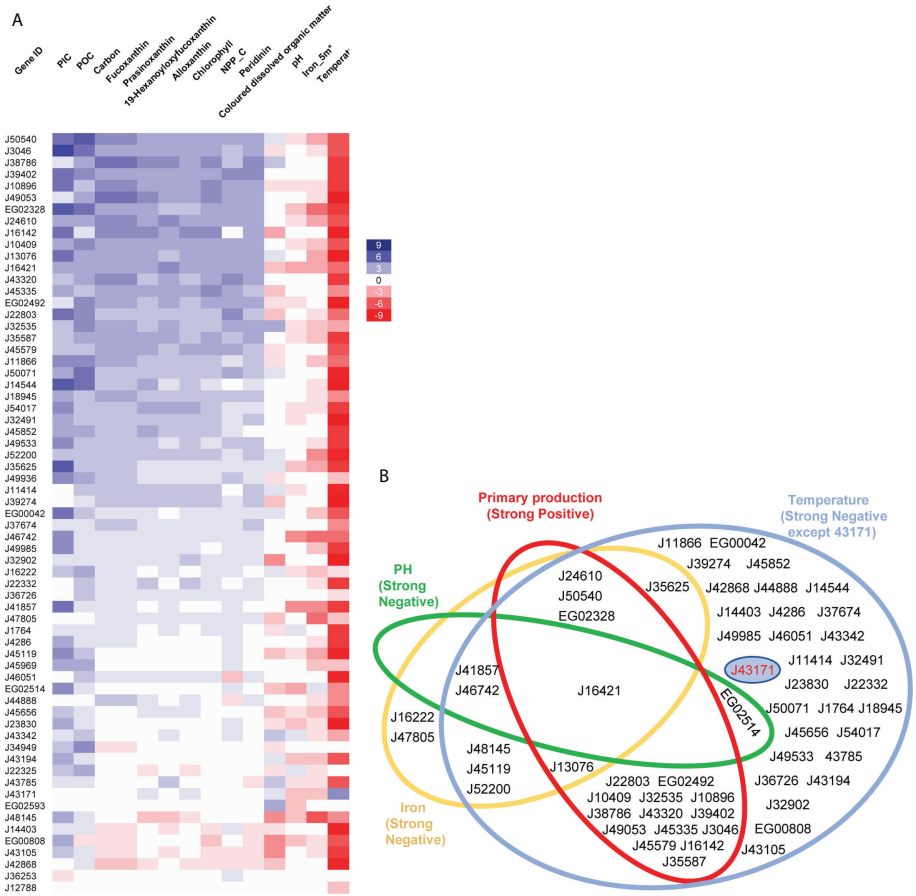

FIGURE 5
Relationships between abundance of diatom plastid transporters and environmental parameters. **(A)** Heatmap showing relationships of environmental factors with 65 P. *tricornutum* transporter homologues in *Tara* Oceans meta-genome (MetaG) and meta-transcriptome (MetaT) datasets. The calculations for each transporter were performed using data from all *Tara* Oceans sampling stations in version 1; with independent calculations for each combination of size fractions (0.8 to 5, 5 to 20, 20 to 180, 180 to to 2000; and 0.8 to 2000 µm) and depths (DCM, deep chlorophyll maximum layer; SRF, upper layer zone) performed. Heatmap rows correspond to transporter gene IDs in *P. tricornutum*, columns represent different environmental parameters. Blue denotes positive relationships between transporters and environmental factors, red denotes negative relationships, with shade proportional to the number of depth and size fraction combinations (maximum 10) showing a significant ($P< 0.05$) Pearson correlation. Aggregate values are shown, that is to say that a significant positive correlation (score +1) in one depth and size fraction combination, and a significant negative correlation (score -1) in another would equate to an aggregate score of 0. POC: Particulate Organic Carbon (µg/kg), PIC: Particulate Inorganic Carbon (mol/m^3), Carbon: Carbon Total (µmol/kg), NCC_C: Net primary production of carbon (mg/m^2/day). **(B)** Plastid transporters showing strong relationships ($P< 0.05$ in at least three different combinations of depths and size fractions; for merged primary production, significant correlations in ≥ 3 depth and size fractions combinations for ≥38 of the ten selected parameters) with four environmental parameters, these being: temperature (blue, negative related except J43171), iron (yellow, negative related), pH (green, negative related) and primary production related parameters (red, positive related, merged by Chlorophyll c3 (mg/m^3), Peridinin (mg/m^3), Fucoxanthin (mg/m^3), Prasinoxanthin (mg/m^3), 19-Hexanoyloxyfucoxanthin (mg/m^3), Alloxanthin (mg/m^3), NCC (mg/m^2/day), PIC (mol/m^3), POC (µg/kg), Carbon Total (µmol/kg). J43171 uniquely displays a positive relationship with temperature and is separately displayed by a blue circle in the temperature circle. Transporters are shown as gene IDs from *P. tricornutum*.

plastidial carbon, photosynthesis and biogenesis-related pathways in *P. tricornutum* transcriptome data (Figure 2) albeit with limited co-regulation with the mitochondria (Supplementary Table S3); showed a sporadic distribution across photosynthetic stramenopile and haptophyte transcriptomes (Figure 4); and lacks any observable *A. thaliana* homologue (Figure 1). *Tara* homologue abundance distribution maps of transporter J43171 and J50540 (as a representative transporter strongly positively correlated to marine primary production) are shown in Supplementary Figure S2 and Supplementary Figure S3. The more specific functions of these transporters, particularly their functions in relation to oceanic temperature and photosynthetic activity remain to be confirmed e.g., through the phenotyping of mutant lines in transformable model species such as *Phaeodactylum*.

Discussion

Although plastid transporters play important roles in energy supply and many other physiological and biochemical reactions in eukaryotic photosynthetic organisms, existing research has largely focused on those found in higher plants, while comprehensive analyses of plastid transporters in eukaryotic algae are still relatively rare (Marchand et al., 2018; Marchand et al., 2020). Here, we profile diverse bioinformatic datasets constructed around the model diatom *P. tricornutum*, alongside environmental sequence data from *Tara* Oceans, to close the knowledge gap concerning the plastid transporters associated with diatoms, and other related algal groups with four-membranes surrounded plastids derived from the secondary endosymbiosis of red algae (e.g., stramenopiles, cryptomonads, and haptophytes). Complementing more detailed studies from other groups on specific proteins (Marchand et al., 2018; Marchand et al., 2020), our data provide holistic insights into the metabolic, evolutionary and environmental functions of diatom plastid transporters.

We note that the transporters identified in our approach (i.e., *via* the presence of a bipartite cleavable targeting sequence, Gruber et al., 2015) may localize to one of multiple membranes (i.e., the plastidial inner or outer envelopes, and the periplastidial membrane; Figure 1C, Figure 1D). The precise resolution of membrane localization and orientation of specific diatom plastid transporters awaits detailed functional characterization, e.g. *via* self-assembling GFP or cross-linking approaches (Moog et al., 2015; Moog et al., 2020). We also note that there are likely to be further transporters necessary for diatom plastidial function not detected by this approach (e.g., transporters targeted to the chloroplast endoplasmic reticulum, and transporters targeted to the plastid *via* non-conventional pathways that do not utilize conventional plastid targeting sequences). These additional transporters may be best revealed by targeted proteomics either of isolated diatom chloroplasts (Schober et al., 2019), or quantitative proteomics of whole-cell separations (e.g., LOPIT, Mulvey et al., 2017 and Barylyuk et al., 2020).

Our analyses reveal similarities and differences in the transporter profiles of diatom and plant plastids (Figure 1A, Figure 1B), reflecting their deep evolutionary divergence. We found clear homologue equivalents of key *A. thaliana* ion, nucleotide and sugar plastid transporters in the *P. tricornutum* genome (Figure 1B). Our data corroborates possible differences in substrate specificity between homologous *A. thaliana* and *P. tricornutum* plastid metabolite transporters (Figure 1B), e.g., the triose phosphate transporters J24610 and J54017, which appear to be nearest equivalents of *A. thaliana* TPT, GPT, XPT and PPT transporters (Figure 1B). It has been reported that TPT homologues in other red alga-derived composite plastids, e.g. those of the non-photosynthetic diatom *Nitzschia putrida*, may have bifunctional TP and PEP transport capabilities (Moog et al., 2020), although the case of in *P. tricornutum* this awaits specific functional characterization, e.g., by electrophysiology or *via* micelle incorporation and *in vitro* assays (Shukla et al., 2017; Wright et al., 2018; Nielsen et al., 2019). Alongside this, we identify clear differences in plant and diatom plastid transporter architecture, including at least six unique to diatoms and not found in other stramenopile, cryptomonad or haptophyte groups (Figure 4). The exact evolutionary origins of these diatom-specific plastid transporters will be best revealed by detailed phylogenies of each protein, incorporating densely sampled phylogenetic reference datasets and *in silico* localization predictions, to identify their probable histories of gene transfer, duplication and re-localization within the diatom cell.

The transporter architecture of the diatom plastid may underpin its unique metabolic potential (Prihoda et al., 2012; Marchand et al., 2018; Nonoyama et al., 2019). Previously, for example, Bailleul et al. have proposed that import of mitochondrial ATP and export of plastid NADPH facilitates diatom photo-acclimation to high light and post-illumination conditions (Bailleul et al., 2015). Our data highlight plastid transporters that show strong transcriptional co-regulation with the mitochondrial genome (Figure 3), and may be involved in plastid-to-mitochondria metabolite exchange. We also identify by *in silico* prediction at least five transporters that are potentially dual-targeted to the plastids and mitochondria (Supplementary Table S2). Finally, we note that at least two of the plastid transporters that show the strongest transcriptional coordination to the *P. tricornutum* mitochondria (J42868: MFS domain transporter; J50540: plastid transport protein Tic110) also had strong correlations to primary production in *Tara* Oceans data (Figures 3, 5), underlining the importance of plastid/mitochondria metabolite exchange for diatom fitness in the environment. The localizations and functions of transporters putatively involved in *P. tricornutum* plastid/mitochondria crosstalk may be best explored by experimental localization, and functional characterisation of mutant lines.

Our environmental data provide insights into which plastid transporters may direct diatom sensitivity to anthropogenic environmental impacts. Rising seawater temperatures (i.e., caused by global warming) and pH decreases (ocean acidification), will most likely individually or synergistically influence diatom photosynthesis, respiration (Goldman et al., 2017; Shi et al., 2019) and primary metabolism (Bermúdez et al., 2015; Novak et al., 2019); although with rising temperatures predicted to have greater overall impact on diatom biomass and photosynthesis than acidification (Sommer et al., 2015; Zhong et al., 2021). Concordant with this, our *Tara* Oceans-based analysis of environmental factors indicates that most diatom plastid transporters showed negative relationships with temperature but no clear response to pH (Figure 5).

We see contrasting relationships between the expression parameters of different diatom plastid transporters with CDOM. The physiological links between transporter expression and this environmental parameter, which is influenced by both primary production, photodegradation and

decomposition (Zhong and Wang, 2009; Zou et al., 2012), and in turn is likely to influence algal photosynthesis and nutrient uptake (White et al., 2003; Helms et al., 2013) alongside algal-bacterial interactions (Klug, 2005), are more complex, and may best be explored by mesocosm-related environmental expression measurements. Finally, the uniquely temperature-associated transporter J43171 (Figure 5B) may provide new insights into diatom thermal adaptation and climate resilience, which may be explored either by cellular (mutant phenotyping, experimental evolution) or environmental approaches (e.g., sequencing of diatoms isolated from subtropical and warm-water habitats).

In conclusion, our holistic study is significant to broader studies of diatom plastid metabolic, evolutionary and environmental functions, allowing the pinpointing of targets linked to key physiological and environmental processes for functional and experimental characterization.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: osf.io repository https://osf.io/89vm3/ in the folder "Transporters".

Author contributions

SL and RGD conceived the project and designed research; SL, MS, and RGD performed research; SL wrote the paper; CB, RGD, and GF contributed to funding acquisition; all authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Chinese Scholarship Council (awarded 2019-2023, scholarship file No. 201904910555) to SL, a CNRS Momentum Fellowship (awarded 2019-2021), and an ANR JCJC (ANR-21-CE02-0014-01) to RGD, an ANR Collaborative grant (ANR-19-CE20-0020) of France to CB, and a European Research Council Advanced Grant (Chloro-mito, Grant No. 833184) to GF.

Acknowledgments

SL acknowledges a Chinese Scholarship Council PhD student (awarded 2019-2023, scholarship file No. 201904910555). RD acknowledges a CNRS Momentum Fellowship (awarded 2019-2021) and an ANR t-ERC (ChloroMosaic). MS and GF acknowledge the support by the European Research Council (ERC) Chloro-mito (Grant No. 833184). CB acknowledges an ANR collaborative grant (Browncut, Grant No. ANR-19-CE20-0020). The authors thank Morgane Roquais and Juan Pierella Karlusich (IBENS) for assistance in the retrieval of Tara Oceans homologues for Phatr3_J43171. This is article # ABCDEF in the *Tara* Oceans series.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.950467/full#supplementary-material

SUPPLEMENTARY TABLE 1
List of *A. thaliana* plastid transporters inferred from ChloroKB, BLASTp outputs and localization predictions.

SUPPLEMENTARY TABLE 2
List of *P. tricornutum* plastid transporters inferred *via in silico* prediction, BLASTp outputs and localization predictions.

SUPPLEMENTARY TABLE 3
List of genes in *P. tricornutum* genome inferred to show transcriptional co-regulation, localization and evolutionary associations with plastid transporters, and their inferred metabolic and biological functions.

SUPPLEMENTARY TABLE 4
List probable evolutionary conservations of transporters inferred from phylogeny and RbH analyses.

SUPPLEMENTARY TABLE 5
Raw data for transporter environmental relationships.

SUPPLEMENTARY FIGURE 1
Pie chart of the taxonomic distribution of all identified diatom transporter meta-gene homologs.

SUPPLEMENTARY FIGURE 2
Tara homologue MetaT and MetaG abundance distribution maps of environmental homologues of the transporter J43171.

SUPPLEMENTARY FIGURE 3
Tara homologue MetaT and MetaG abundance distribution maps of environmental homologues of the transporter J50540.

SUPPLEMENTARY DATA SHEET 1
Fasta format sequences of all 70 inferred *Phaeodactylum* transporter proteins used in this study.

References

Ait-Mohamed, O., Novák Vanclová, A. M. G., Joli, N., Liang, Y., Zhao, X., Genovesio, A., et al. (2020). PhaeoNet: A holistic RNAseq-based portrait of transcriptional coordination in the model diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 11, 590949. doi: 10.3389/fpls.2020.590949

Allen, A. E., Dupont, C. L., Oborník, M., Horák, A., Nunes-Nesi, A., McCrow, J. P., et al. (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473 (7346), 203–207. doi: 10.1038/nature10074

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410. doi: 10.1016/S0022-2836(05)80360-2

Ashworth, J., Turkarslan, S., Harris, M., Orellana, M. V., and Baliga, N. S. (2016). Pan-transcriptomic analysis identifies coordinated and orthologous functional modules in the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. *Mar. Genomics* 26, 21–28. doi: 10.1016/j.margen.2015.10.011

Ast, M., Gruber, A., Schmitz-Esser, S., Neuhaus, H. E., Kroth, P. G., Horn, M., et al. (2009). Diatom plastids depend on nucleotide import from the cytosol. *Proc. Natl. Acad. Sci. U.S.A.* 106 (9), 3621–3626. doi: 10.1073/pnas.0808862106

Bailleul, B., Berne, N., Murik, O., Petroutsos, D., Prihoda, J., Tanaka, A., et al. (2015). Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* 524 (7565), 366–369. doi: 10.1038/nature14599

Barylyuk, K., Koreny, L., Ke, H., Butterworth, S., Crook, O. M., Lassadi, I., et al. (2020). A comprehensive subcellular atlas of the toxoplasma proteome *via* hyperLOPIT provides spatial context for protein functions. *Cell Host Microbe* 28 (5), 752–766.e9. doi: 10.1016/j.chom.2020.09.011

Bendtsen, J. D., Nielsen, H., Von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340 (4), 783–795. doi: 10.1016/j.jmb.2004.05.028

Benoiston, A. S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., et al. (2017). The evolution of diatoms and their biogeochemical functions. *Philos. Trans. R Soc. Lond B Biol. Sci.* 372 (1728), 20160397. doi: 10.1098/rstb.2016.0397

Bermúdez, R., Feng, Y., Roleda, M. Y., Tatters, A. O., Hutchins, D. A., Larsen, T., et al. (2015). Long-term conditioning to elevated pCO_2 and warming influences the fatty and aamino acid composition of the diatom *Cylindrotheca fusiformis*. *PloS One* 10 (5), e0123945. doi: 10.1371/journal.pone.0123945

Bhattacharya, D., Yoon, H. S., and Hackett, J. D. (2004). Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *Bioessays* 26 (1), 50–60. doi: 10.1002/bies.10376

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456 (7219), 239–244. doi: 10.1038/nature07410

Broddrick, J. T., Du, N., Smith, S. R., Tsuji, Y., Jallet, D., Ware, M. A., et al. (2019). Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricornutum*. *New Phytol.* 222 (3), 1364–1379. doi: 10.1111/nph.15685

Brownlee, C., Helliwell, K. E., Meeda, Y., McLachlan, D., Murphy, E. A., and Wheeler, G. L. (2022). Regulation and integration of membrane transport in marine diatoms. *Semin. Cell Dev. Biol.* S1084-9521(22)00072-6 ISSN 1084-9521. doi: 10.1016/j.semcdb.2022.03.006

Cao, M. J., Wang, Z., Wirtz, M., Hell, R., Oliver, D. J., and Xiang, C. B. (2013). SULTR3;1 is a chloroplast-localized sulfate transporter in *Arabidopsis thaliana*. *Plant J.* 73 (4), 607–616. doi: 10.1111/tpj.12059

Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nat. Commun.* 9 (1), 373. doi: 10.1038/s41467-017-02342-1

Dell'Aquila, G., Zauner, S., Heimerl, T., Kahnt, J., Samel-Gondesen, V., Runge, S., et al. (2020). Mobilization and cellular distribution of phosphate in the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00579

DeTar, R. A., Barahimipour, R., Manavski, N., Schwenkert, S., Höhner, R., Bölter, B., et al. (2021). Loss of inner-envelope K^+/H^+ exchangers impairs plastid rRNA maturation and gene expression. *Plant Cell* 33 (7), 2479–2505. doi: 10.1093/plcell/koab123

Dorrell, R. G., Azuma, T., Nomura, M., Audren de Kerdrel, G., Paoli, L., Yang, S., et al. (2019). Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc. Natl. Acad. Sci. U.S.A.* 116 (14), 6914–6923. doi: 10.1073/pnas.1819976116

Dorrell, R. G., and Bowler, C. (2017). *Advances in botanical research: secondary endosymbiosis* Vol. 84. Ed. Y. (Hirakawa Cambridge, Massachussetts, USA: Academic Press), 59–103.

Dorrell, R. G., Gile, G., McCallum, G., Méheust, R., Bapteste, E. P., Klinger, C. M., et al. (2017). Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *Elife* 6, 23717. doi: 10.7554/eLife.23717

Dorrell, R. G., and Smith, A. G. (2011). Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell* 10 (7), 856–868. doi: 10.1128/ec.00326-10

Dorrell, R. G., Villain, A., Perez-Lamarque, B., Audren de Kerdrel, G., McCallum, G., Watson, A. K., et al. (2021). Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc. Natl. Acad. Sci. U.S.A.* 118 (4), 2009974118. doi: 10.1073/pnas.2009974118

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281 (5374), 237–240. doi: 10.1126/science.281.5374.237

Finazzi, G., Petroutsos, D., Tomizioli, M., Flori, S., Sautron, E., Villanova, V., et al. (2015). Ions channels/transporters and chloroplast regulation. *Cell Calcium* 58 (1), 86–97. doi: 10.1016/j.ceca.2014.10.002

Flori, S., Jouneau, P. H., Bailleul, B., Gallet, B., Estrozi, L. F., Moriscot, C., et al. (2017). Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nat. Commun.* 8 (1), 1–9. doi: 10.1038/ncomms15885

Fukasawa, Y., Tsuji, J., Fu, S. C., Tomii, K., Horton, P., and Imai, K. (2015). MitoFates: Improved prediction of mitochondrial targeting sequences and ttheir ccleavage sites. *Mol. Cell Proteomics* 14 (4), 1113–1126. doi: 10.1074/mcp.M114.043083

Gao, X., Bowler, C., and Kazamia, E. (2021). Iron metabolism strategies in diatoms. *J. Exp. Bot.* 72 (6), 2165–2180. doi: 10.1093/jxb/eraa575

Gigolashvili, T., Geier, M., Ashykhmina, N., Frerigmann, H., Wulfert, S., Krueger, S., et al. (2012). The *Arabidopsis thylakoid* ADP/ATP carrier TAAC has an additional role in supplying plastidic phosphoadenosine 5'-phosphosulfate to the cytosol. *Plant Cell* 24 (10), 4187–4204. doi: 10.1105/tpc.112.101964

Gile, G. H., Moog, D., Slamovits, C. H., Maier, U. G., and Archibald, J. M. (2015). Dual organellar ttargeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome Biol. Evol.* 7 (6), 1728–1742. doi: 10.1093/gbe/evv095

Gloaguen, P., Bournais, S., Alban, C., Ravanel, S., Seigneurin-Berny, D., Matringe, M., et al. (2017). ChloroKB: A web aapplication for the integration of knowledge related to chloroplast metabolic network. *Plant Physiol.* 174 (2), 922–934. doi: 10.1104/pp.17.00242

Gloaguen, P., Vandenbrouck, Y., Joyard, J., and Curien, G. (2021). ChloroKB, a cell metabolism reconstruction of the model plant *Arabidopsis thaliana*. *C R Biol.* 344 (2), 157–163. doi: 10.5802/crbiol.49

Goldman, J. A., Bender, M. L., and Morel, F. M. (2017). The effects of pH and pCO_2 on photosynthesis and respiration in the diatom *Thalassiosira weissflogii*. *Photosynth Res.* 132 (1), 83–93. doi: 10.1007/s11120-016-0330-2

Green, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* 66 (1), 34–44. doi: 10.1111/j.1365-313X.2011.04541.x

Gruber, A., and Kroth, P. G. (2017). Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research. *Philos. Trans. R Soc. Lond B Biol. Sci.* 372 (1728), 20160402. doi: 10.1098/rstb.2016.0402

Gruber, A., Rocap, G., Kroth, P. G., Armbrust, E. V., and Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* 81 (3), 519–528. doi: 10.1111/tpj.12734

Gschloessl, B., Guermeur, Y., and Cock, J. M. (2008). HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinf.* 9, 393. doi: 10.1186/1471-2105-9-393

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532 (7600), 465–470. doi: 10.1038/nature16942

Hanikenne, M., and Baurain, D. (2013). Origin and evolution of metal p-type ATPases in plantae (Archaeplastida). *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00544

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32 (Database issue), D258–D261. doi: 10.1093/nar/gkh036

Helms, J. R., Stubbins, A., Perdue, E. M., Green, N. W., Chen, H., and Mopper, K. (2013). Photochemical bleaching of oceanic dissolved organic matter and its effect on absorption spectral slope and fluorescence. *Mar. Chem.* 155 (sep.20), 81–91. doi: 10.1016/j.marchem.2013.05.015

Hernández-Salmerón, J. E., and Moreno-Hagelsieb, G. (2020). Progress in quickly finding orthologs as reciprocal best hits: Comparing blast, last, diamond and MMseqs2. *BMC Genomics* 21 (1), 741. doi: 10.1186/s12864-020-07132-6

Heydarizadeh, P., Poirier, I., Loizeau, D., Ulmann, L., Mimouni, V., Schoefs, B., et al. (2013). Plastids of marine phytoplankton produce bioactive pigments and lipids. *Mar. Drugs* 11, 3425–3471. doi: 10.3390/md11093425

Hippmann, A. A., Schuback, N., Moon, K. M., Mccrow, J. P., and Maldonado, M. T. (2020). Proteomic analysis of metabolic pathways shows chloroplast-mitochondria cross-talk in a Cu-limited diatom. *Plant Direct* 6 (1), e376. doi: 10.1101/2020.09.15.298752

Hooper, C. M., Castleden, I. R., Tanz, S. K., Aryamanesh, N., and Millar, A. H. (2017). SUBA4: the interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res.* 45 (D1), D1064–D1074. doi: 10.1093/nar/gkw1041

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., et al (2006). WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* 35, W585–7. doi: 10.1093/nar/gkm259.

Huala, E., Dickerman, A. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., et al. (2001). The *Arabidopsis* information resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* 29 (1), 102–105. doi: 10.1093/nar/29.1.102

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449 (7161), 463–467. doi: 10.1038/nature06148

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428 (4), 726–731. doi: 10.1016/j.jmb.2015.11.006

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30 (14), 3059–3066. doi: 10.1093/nar/gkf436

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28 (12), 1647–1649. doi: 10.1093/bioinformatics/bts199

Kilian, O., and Kroth, P. G. (2004). Presequence acquisition during secondary endocytobiosis and the possible role of introns. *J. Mol. Evol.* 58 (6), 712–721. doi: 10.1007/s00239-004-2593-z

Kilian, O., and Kroth, P. G. (2005). Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J.* 41 (2), 175–183. doi: 10.1111/j.1365-313X.2004.02294.x

Klug, J. L. (2005). Bacterial response to dissolved organic matter affects resource availability for algae. *Can. J. Fish Aquat. Sci.* 62 (2), 472–481. doi: 10.1139/f04-229

Kong, L., and Price, N. M. (2021). Transcriptomes of an oceanic diatom reveal the initial and final stages of acclimation to copper deficiency. *Environ. Microbiol.* doi: 10.1111/1462-2920.15609

Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu de Madron, X., et al. (2018). Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. *Nat. Commun.* 9 (1), 953. doi: 10.1038/s41467-018-03376-9

Liu, S. L., Pan, A. Q., and Adams, K. L. (2014). Protein subcellular relocalization of duplicated genes in *Arabidopsis*. *Genome Biol. Evol.* 6 (9), 2501–2515. doi: 10.1093/gbe/evu191

Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., et al. (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc. Natl. Acad. Sci. U.S.A.* 106 (31), 12803–12808. doi: 10.1073/pnas.0905841106

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., et al. (2016). Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. U.S.A.* 113 (11), E1516–E1525. doi: 10.1073/pnas.1509523113

Marchand, J., Heydarizadeh, P., Schoefs, B., and Spetea, C. (2018). Ion and metabolite transport in the chloroplast of algae: Lessons from land plants. *Cell Mol. Life Sci.* 75 (12), 2153–2176. doi: 10.1007/s00018-018-2793-0

Marchand, J., Heydarizadeh, P., Schoefs, B., and Spetea, C. (2020). "Chloroplast ion and metabolite transport in algae," in *Photosynthesis in algae: Biochemical and physiological mechanisms*, vol. 2020. (New York, New York USA: Springer), 107–139. doi: 10.1007/978-3-030-33397-3_6

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., et al. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 99 (19), 12246–12251. doi: 10.1073/pnas.182432999

Matsuda, Y., Hopkinson, B. M., Nakajima, K., Dupont, C. L., and Tsuji, Y. (2017). Mechanisms of carbon dioxide acquisition and CO_2 sensing in marine diatoms: a gateway to carbon metabolism. *Philos. Trans. R Soc. Lond B Biol. Sci.* 372 (1728), 20160403. doi: 10.1098/rstb.2016.0403

Mix, A. K., Cenci, U., Heimerl, T., Marter, P., Wirkner, M. L., and Moog, D. (2018). Identification and localization of peroxisomal biogenesis proteins indicates the presence of peroxisomes in the cryptophyte *Guillardia theta* and other "Chromalveolates". *Genome Biol. Evol.* 10 (10), 2834–2852. doi: 10.1093/gbe/evy214

Moog, D., Nozawa, A., Tozawa, Y., and Kamikawa, R. (2020). Substrate specificity of plastid phosphate transporters in a non-photosynthetic diatom and its implication in evolution of red alga-derived complex plastids. *Sci. Rep.* 10 (1), 1167. doi: 10.1038/s41598-020-58082-8

Moog, D., Rensing, S. A., Archibald, J. M., Maier, U. G., and Ullrich, K. K. (2015). Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. *Genome Biol. Evol.* 7 (11), 2955–2969. doi: 10.1093/gbe/evv190

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol. Biol.* 396, 59–70. doi: 10.1007/978-1-59745-515-2_5

Mulvey, C. M., Breckels, L. M., Geladaki, A., Britovšek, N. K., Nightingale, D. J. H., Christoforou, A., et al. (2017). Using hyperLOPIT to perform high-resolution mapping of the spatial proteome. *Nat. Protoc.* 12 (6), 1110–1135. doi: 10.1038/nprot.2017.026

Murik, O., Tirichine, L., Prihoda, J., Thomas, Y., Araújo, W. L., Allen, A. E., et al. (2019). Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. *New Phytol.* 221 (3), 1303–1316. doi: 10.1111/nph.15479

Nakajima, K., Tanaka, A., and Matsuda, Y. (2013). SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. *Proc. Natl. Acad. Sci. U.S.A.* 110 (5), 1767–1772. doi: 10.1073/pnas.1216234110

Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A., and 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 biogeochem cycles* 9 (3), 359–372. doi: 10.1029/95GB01070

Nielsen, A. K., Möller, I. R., Wang, Y., Rasmussen, S. G. F., Lindorff-Larsen, K., Rand, K. D., et al. (2019). Substrate-induced conformational dynamics of the dopamine transporter. *Nat. Commun.* 10 (1), 2714. doi: 10.1038/s41467-019-10449-w

Nonoyama, T., Kazamia, E., Nawaly, H., Gao, X., Tsuji, Y., Matsuda, Y., et al. (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules* 9 (8), 322. doi: 10.3390/biom9080322

Novak, T., Godrijan, J., Pfannkuchen, D. M., Djakovac, T., Medić, N., Ivančić, I., et al. (2019). Global warming and oligotrophication lead to increased lipid production in marine phytoplankton. *Sci. Total Environ.* 668, 171–183. doi: 10.1016/j.scitotenv.2019.02.372

Nymark, M., Valle, K. C., Brembu, T., Hancke, K., Winge, P., Andresen, K., et al. (2009). An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PloS One* 4 (11), e7743. doi: 10.1371/journal.pone.0007743

Nymark, M., Valle, K. C., Hancke, K., Winge, P., Andresen, K., Johnsen, G., et al. (2013). Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*. *PloS One* 8 (3), e58722. doi: 10.1371/journal.pone.0058722

Oborník, M., and Green, B. R. (2005). Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol. Biol. Evol.* 22 (12), 2343–2353. doi: 10.1093/molbev/msi230

Oudot-Le Secq, M. P., Grimwood, J., Shapiro, H., Armbrust, E. V., Bowler, C., and Green, B. R. (2007). Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: Comparison with other plastid genomes of the red lineage. *Mol. Genet. Genomics* 277 (4), 427–439. doi: 10.1007/s00438-006-0199-4

Parks, M. B., Nakov, T., Ruck, E. C., Wickett, N. J., and Alverson, A. J. (2018). Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *Am. J. Bot.* 105 (3), 330–347. doi: 10.1002/ajb2.1056

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Bescot, N. L., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of Tara oceans data. *Sci. Data* 2, 150023. doi: 10.1038/sdata.2015.23

Pfeil, B. E., Schoefs, B., and Spetea, C. (2014). Function and evolution of channels and transporters in photosynthetic membranes. *Cell Mol. Life Sci.* 71 (6), 979–998. doi: 10.1007/s00018-013-1412-3

Pick, T. R., Bräutigam, A., Schulz, M. A., Obata, T., Fernie, A. R., and Weber, A. P. (2013). PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. *Proc. Natl. Acad. Sci. U.S.A.* 110 (8), 3185–3190. doi: 10.1073/pnas.1215142110

Prihoda, J., Tanaka, A., de Paula, W. B., Allen, J. F., Tirichine, L., and Bowler, C. (2012). Chloroplast-mitochondria cross-talk in diatoms. *J. Exp. Bot.* 63 (4), 1543–1557. doi: 10.1093/jxb/err441

Puissant, A., El Hourany, R., Charantonis, A. A., Bowler, C., and Thiria, S. (2021). Inversion of phytoplankton pigment vertical profiles from satellite data using machine learning. *Remote Sens.* 13 (8), 1445. doi: 10.3390/rs13081445

Rastogi, A., Maheswari, U., Dorrell, R. G., Vieira, F. R. J., Maumus, F., Kustka, A., et al. (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. *Sci. Rep.* 8, 4834. doi: 10.1038/s41598-018-23106-x

Rastogi, S., and Rost, B. (2011). LocDB: Experimental annotations of localization for *Homo sapiens* and *Arabidopsis thaliana*. *Nucleic Acids Res.* 39 (suppl_1), D230–D234. doi: 10.1093/nar/gkq927

Santin, A., Caputi, L., Longo, A., Chiurazzi, M., Ribera d'Alcalà, M., Russo, M. T., et al. (2021). Integrative omics identification, evolutionary and structural analysis of low affinity nitrate transporters in diatoms, diNPFs. *Open Biol.* 11 (4), 200395. doi: 10.1098/rsob.200395

Schober, A. F., Flori, S., Finazzi, G., Kroth, P. G., and Bártulos, C. R. (2018). Isolation of plastid fractions from the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. *Methods Mol. Biol.* 1829, 189–203. doi: 10.1007/978-1-4939-8654-5_13

Schober, A. F., Río Bártulos, C., Bischoff, A., Lepetit, B., Gruber, A., and Kroth, P. G. (2019). Organelle studies and proteome analyses of mitochondria and plastids fractions from the diatom *Thalassiosira pseudonana*. *Plant Cell Physiol.* 60 (8), 1811–1828. doi: 10.1093/pcp/pcz097

Seydoux, C., Storti, M., Giovagnetti, V., Matuszyńska, A., Guglielmino, E., Zhao, X., et al. (2022). Impaired photoprotection in *Phaeodactylum tricornutum* KEA3 mutants reveals the proton regulatory circuit of diatoms light acclimation. *New Phytol.* 234 (2), 578–591. doi: 10.1111/nph.18003

Shai, N., Schuldiner, M., and Zalckvar, E. (2016). No peroxisome is an island - peroxisome contact sites. *Biochim. Biophys. Acta* 1863 (5), 1061–1069. doi: 10.1016/j.bbamcr.2015.09.016

Shi, D., Hong, H., Su, X., Liao, L., Chang, S., and Lin, W. (2019). The physiological response of marine diatoms to ocean acidification: Differential roles of seawater pCO_2 and pH. *J. Phycol* 55 (3), 521–533. doi: 10.1111/jpy.12855

Shukla, S., Abel, B., Chufan, E. E., and Ambudkar, S. V. (2017). Effects of a detergent micelle environment on p-glycoprotein (ABCB1)-ligand interactions. *J. Biol. Chem.* 292 (17), 7066–7076. doi: 10.1074/jbc.M116.771634

Solymosi, K. (2012). Plastid structure, diversification and interconversions i. algae. *Curr. Chem. Biol.* 6 (3), 167–186. doi: 10.2174/2212796811206030002

Sommer, U., Paul, C., and Moustaka-Gouni, M. (2015). Warming and Ocean Acidification Effects on Phytoplankton—From Species Shifts to Size Shifts within Species in a Mesocosm Experiment. *PLoS One* 10(5):e0125239. doi: 10.1371/journal.pone.0125239

South, P. F., Walker, B. J., Cavanagh, A. P., Rolland, V., Badger, M., and Ort, D. R. (2017). Bile acid sodium symporter BASS6 can transport glycolate and is involved in photorespiratory metabolism in *Arabidopsis thaliana*. *Plant Cell* 29 (4), 808–823. doi: 10.1105/tpc.16.00775

Spetea, C., Rintamäki, E., and Schoefs, B. (2014). Changing the light environment: Chloroplast signalling and response mechanisms. *Philos. Trans. R Soc. Lond B Biol. Sci.* 369 (1640), 20130220. doi: 10.1098/rstb.2013.0220

Stiller, J. W., Huang, J., Ding, Q., Tian, J., and Goodwillie, C. (2009). Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses?. *BMC Biol* 10, 484. doi: 10.1186/1471-2164-10-484

Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Eveillard, D., Gorsky, G., et al. (2020). Tara Oceans: Towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* 18 (8), 428–445. doi: 10.1038/s41579-020-0364-5

Terashima, M., Specht, M., and Hippler, M. (2011). The chloroplast proteome: a survey from the *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. *Curr. Genet.* 57 (3), 151–168. doi: 10.1007/s00294-011-0339-1

Thuswaldner, S., Lagerstedt, J. O., Rojas-Stütz, M., Bouhidel, K., Der, C., Leborgne-Castel, N., et al. (2007). Identification, expression, and functional analyses of a thylakoid ATP/ADP carrier from *Arabidopsis*. *J. Biol. Chem.* 282 (12), 8848–8859. doi: 10.1074/jbc.M609130200

Tong, S., Xu, D., Wang, Y., Zhang, X., Li, Y., Wu, H., et al. (2021). Influence of ocean acidification on thermal reaction norms of carbon metabolism in the marine diatom *Phaeodactylum tricornutum*. *Mar. Environ. Res.* 164, 105233. doi: 10.1016/j.marenvres.2020.105233

Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., et al. (2017). Influence of diatom diversity on the ocean biological carbon pump. *Nat. Geosci* 11, 27–37. doi: 10.1038/s41561-017-0028-x

Tsuji, Y., Nakajima, K., and Matsuda, Y. (2017). Molecular aspects of the biophysical CO_2-concentrating mechanism and its regulation in marine diatoms. *J. Exp. Bot.* 68 (14), 3763–3772. doi: 10.1093/jxb/erx173

Turnšek, J., Brunson, J. K., Viedma, M., Deerinck, T. J., Horák, A., Oborník, M., et al. (2021). Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway. *Elife* 10, 52770. doi: 10.7554/eLife.52770

Uwizeye, C., Decelle, J., Jouneau, P. H., Flori, S., Gallet, B., Keck, J. B., et al (2021). Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. *Nat Commun* 12, 1049. doi: 10.1038/s41467-021-21314-0

Valle, K. C., Nymark, M., Aamot, I., Hancke, K., Winge, P., Andresen, K., et al. (2014). System responses to equal doses of photosynthetically usable radiation of blue, green, and red light in the marine diatom *Phaeodactylum tricornutum*. *PloS One* 9 (12), e114211. doi: 10.1371/journal.pone.0114211

Veluchamy, A., Lin, X., Maumus, F., Rivarola, M., Bhavsar, J., Creasy, T., et al. (2013). Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nat. Commun.* 4, 2091. doi: 10.1038/ncomms3091

Veluchamy, A., Rastogi, A., Lin, X., Lombard, B., Murik, O., Thomas, Y., et al. (2015). An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. *Genome Biol.* 16, 102. doi: 10.1186/s13059-015-0671-8

Villar, E., Vannier, T., Vernette, C., Lescot, M., Cuenca, M., Alexandre, A., et al. (2018). The ocean gene atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res.* 46 (W1), W289–w295. doi: 10.1093/nar/gky376

Vincent, F., and Bowler, C. (2020). Diatoms are selective segregators in global ocean planktonic communities. *mSystems* 5 (1), e00444–19. doi: 10.1128/mSystems.00444-19

Wang, C., Xu, W., Jin, H., Zhang, T., Lai, J., Zhou, X., et al. (2016). A putative chloroplast-localized Ca^{2+}/H^{+} antiporter CCHA1 is involved in calcium and pH homeostasis and required for PSII function in *Arabidopsis*. *Mol. Plant* 9 (8), 1183–1196. doi: 10.1016/j.molp.2016.05.015

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., et al. (2008). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 36, D13–D21. doi: 10.1093/nar/gkm1000

White, E. M., Vaughan, P. P., and Zepp, R. G. (2003). Role of photo-fenton reaction in the production of hydroxyl radicals and photobleaching of coloured dissolved organic matter in a coastal river of the southern united states. *Aquat Sci.* 65 (4), 402–414. doi: 10.1007/s00027-003-0675-4

Wright, J., Muench, S. P., Goldman, A., and Baker, A. (2018). Substrate polyspecificity and conformational relevance in ABC transporters: new insights from structural studies. *Biochem. Soc. Trans.* 46 (6), 1475–1484. doi: 10.1042/BST20180146

Yang, M., Lin, X., Liu, X., Zhang, J., and Ge, F. (2018). Genome annotation of a model diatom *Phaeodactylum tricornutum* using an integrated proteogenomic pipeline. *Mol. Plant* 11 (10), 1292–1307. doi: 10.1016/j.molp.2018.08.005

Young, J. N., and Schmidt, K. (2020). It's what's inside that matters: Physiological adaptations of high-latitude marine microalgae to environmental change. *New Phytol.* 227 (5), 1307–1318. doi: 10.1111/nph.16648

Zhao, X., Deton Cabanillas, A. F., Veluchamy, A., Bowler, C., Vieira, F. R. J., and Tirichine, L. (2020). Probing the diversity of polycomb and trithorax proteins in cultured and environmentally sampled microalgae. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00189

Zhong, J., Guo, Y., Liang, Z., Huang, Q., Lu, H., Pan, J., et al. (2021). Adaptation of a marine diatom to ocean acidification and warming reveals constraints and trade-offs. *Sci. Total Environ.* 771, 145167. doi: 10.1016/j.scitotenv.2021.145167

Zhong, H., and Wang, W. X. (2009). Controls of dissolved organic matter and chloride on mercury uptake by a marine diatom. *Environ. Sci. Technol.* 43 (23), 8998–9003. doi: 10.1021/es901646k

Zou, L. M., Li, J. H., Jiang, D. G., and Wang, Y. L. (2012). Spectral characteristics in the degradation process of colored dissolved organic matter from large filamentous algae sources in saltwater lake. *Adv. Mat Res.* 356-360, 2536–2541. doi: 10.4028/www.scientific.net/AMR.356-360.2536

2.2 The discovery of a novel MFS transporter: Trans3

Previously, we have performed an *in silico* analysis to profile the probable functions and influenced core pathways of 70 assumed plastid-targeted transporters (Liu et al., 2022), considering similarity to *Arabidopsis thaliana* chloroplast transporters, transcriptional co-regulation with mitochondrial genomes, presumptive evolutionary histories inferred by single-gene phylogeny, and the broader environmental expression dynamics of homologues identified within the *Tara* Oceans survey. We profile the probable biological, metabolic and eco-physiological diversity of diatom plastid transporters, and identify new candidate transporters for diatom-specific plastid metabolic activities and environmental resilience (Annex I).

Based on the analysis, six plastid-targeted transporters with uncertain functions belonged to the major facilitator superfamily (MFS). This superfamily mainly encodes membrane transport proteins and is found in all kingdoms of life (Pao et al., 1998; Niño-González et al., 2019). It has been reported that MFS transporters have a great diversity of substrates, e.g. organic acids, ions and metabolites (Pao et al., 1998; Niño-González et al., 2019). Discovering the function, localization and substrate of these underrated and not well-studied MFS transporters will let us better understand the complex transport activities associated with diatom plastids.

Our analysis then led to the focus of a novel MFS superfamily transporter, henceforth termed "Trans3". This gene showed a central place to be associated with all five major processes (Annex I Figure 2), suggesting that Trans3 is likely to have central roles in diatom plastid metabolism.

Considering reciprocal BLAST best hit (RbH) analysis amongst a previously assembled composite library of the tree of life (Dorrell et al., 2021; Nonoyama et al., 2019), Trans3 was found to possess potential homologues in a range of organisms harvesting secondary red plastids. These include diatoms, other stramenopiles (bolidophytes, pelagophytes, dictyochophytes, and the "PSC" clade of pinguiophytes, synchromophytes and chrysophytes) (Annex I Figure 4, Figure 2.1); and cryptomonads, haptophytes and dinoflagellates, whose nuclear genomes are distantly related to stramenopiles but also possess secondary red

plastids (Chapter 1 Figure 1.3). No other homologues were retrieved from other eukaryotes (e.g., in primary red chloroplast contained algae) or prokaryotes (bacteria and archaea) suggesting Trans3 is unique to secondary red plastid-containing species.

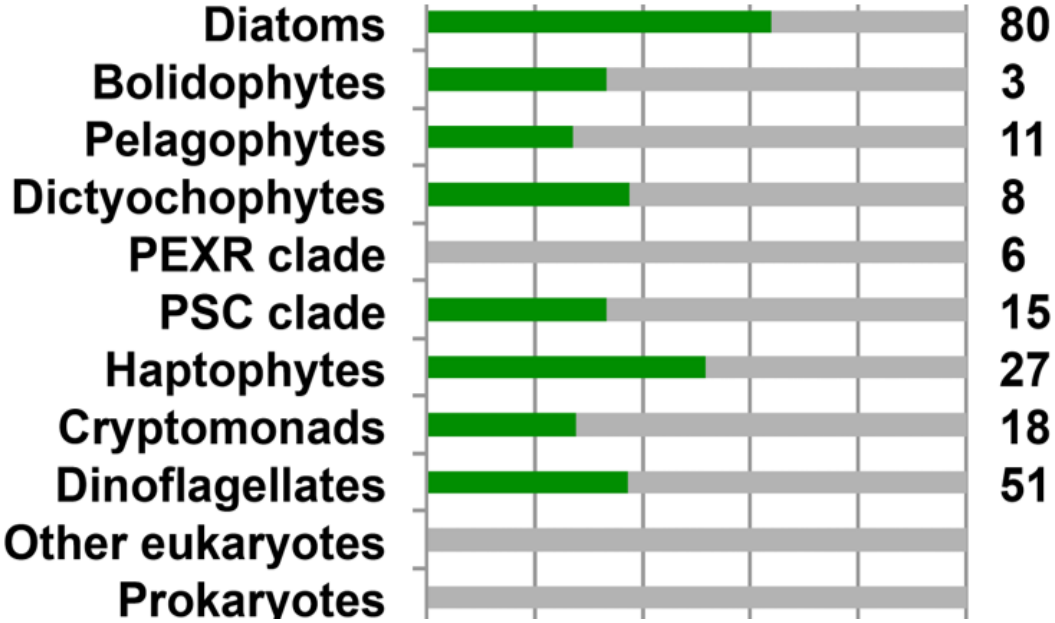

Figure 2.1 RbH homologue distribution of Trans3. The previous RbH distribution analysis was done by Richard Dorrell (Dorrell et al., 2021) prior to the start of my PhD. Green color represents the percentage of species searched in a given group containing Trans3 RbH homologues, and grey the percentage that do not. The absolute number of species considered in each category is noted on the right-hand side.

Interestingly, when we tried to find plastid transporters whose genes showed strong coregulation with the mitochondrial genome, four plastid-targeted transporters identified with mean coregulation coefficients > 0.5 were MFS transporters (Annex I Figure 3, Figure 2.2). Trans3 was the second most positively plastidial transporter to genes encoded in *P. tricornutum* mitochondrial genome (Annex I Figure 3, Figure 2.2). Trans3 was first investigated therefore as a potential mediator of plastid-mitochondria crosstalk.

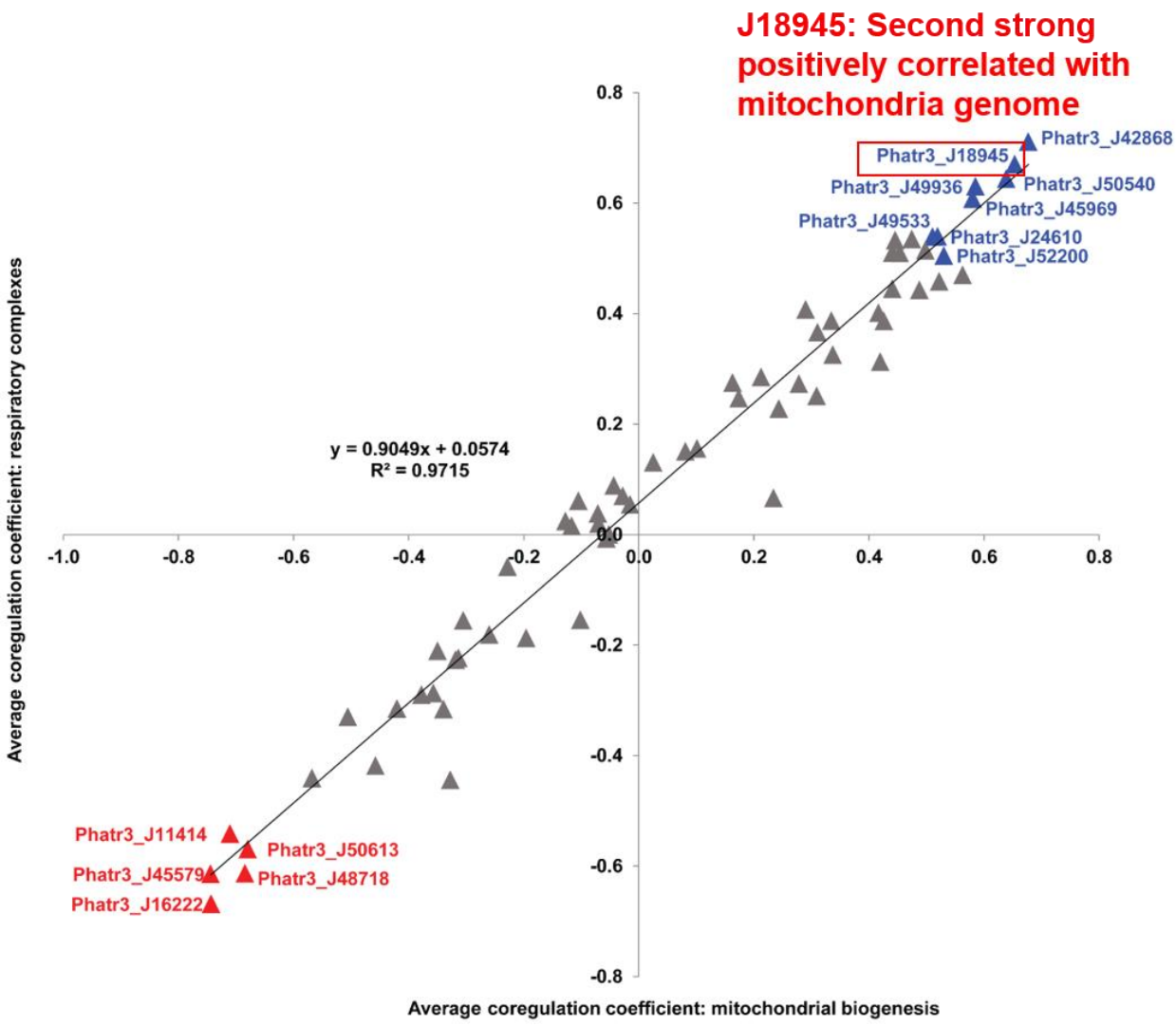

Figure 2.2 Scatterplot of Pearson correlation coefficients between plastid transporters and mitochondria-encoded genes. Horizontal axes correspond to average correlation coefficients calculated between *Phaeodactylum* transporter genes and mitochondrial genes involved in biogenesis (encoding ribosomal and translocase proteins), and vertical axes correspond to average correlation coefficients calculated between *Phaeodactylum* transporter genes and mitochondrial respiratory complex genes across 17 conditional groups in published microarray data studies (Oudot-Le Secq et al., 2007; Ashworth et al, 2016; Nymark et al., 2009; Nymark et al., 2013; Valle et al., 2014). The strong positive correlation coefficient between the expression trends of plastid transporters and mitochondrially-encoded genes (t-test $P < 10^{-05}$) indicates strong overall reproducibility of correlation (and strong internal positive co-regulation of the mitochondrial genome), independent of mitochondrial gene function. Eight transporter genes that recover average correlation coefficients > 0.5 and five that recover average correlation coefficients < - 0.5 for both axes are labelled in blue and red text, respectively. The Figure is cited from Liu et al. (2022).

We further explored Trans3 relative abundance in meta-transcriptome (metaT) and meta-genome (metaG) data in *Tara* Oceans (Sunagawa et al., 2020). This analysis used meta-genes that resolved with diatom Trans3 isoforms, which were mapped onto a previously published Trans3 phylogeny using RAxML phylogenetic analysis (Dorrell et al., 2021). As shown in Figure 2.3, the environmental homologues of this gene showed greatest relative abundance in the Arctic and Southern Oceans, where diatoms predominate.

Armed with these comprehensive data, I made a deliberate decision to delve deeper into investigating the precise functions of Trans3 in *Phaeodactylum tricornutum*. This involved a meticulous examination encompassing experimental localization, analysis of mutant phenotypes, and the exploration of potential biochemical activities.

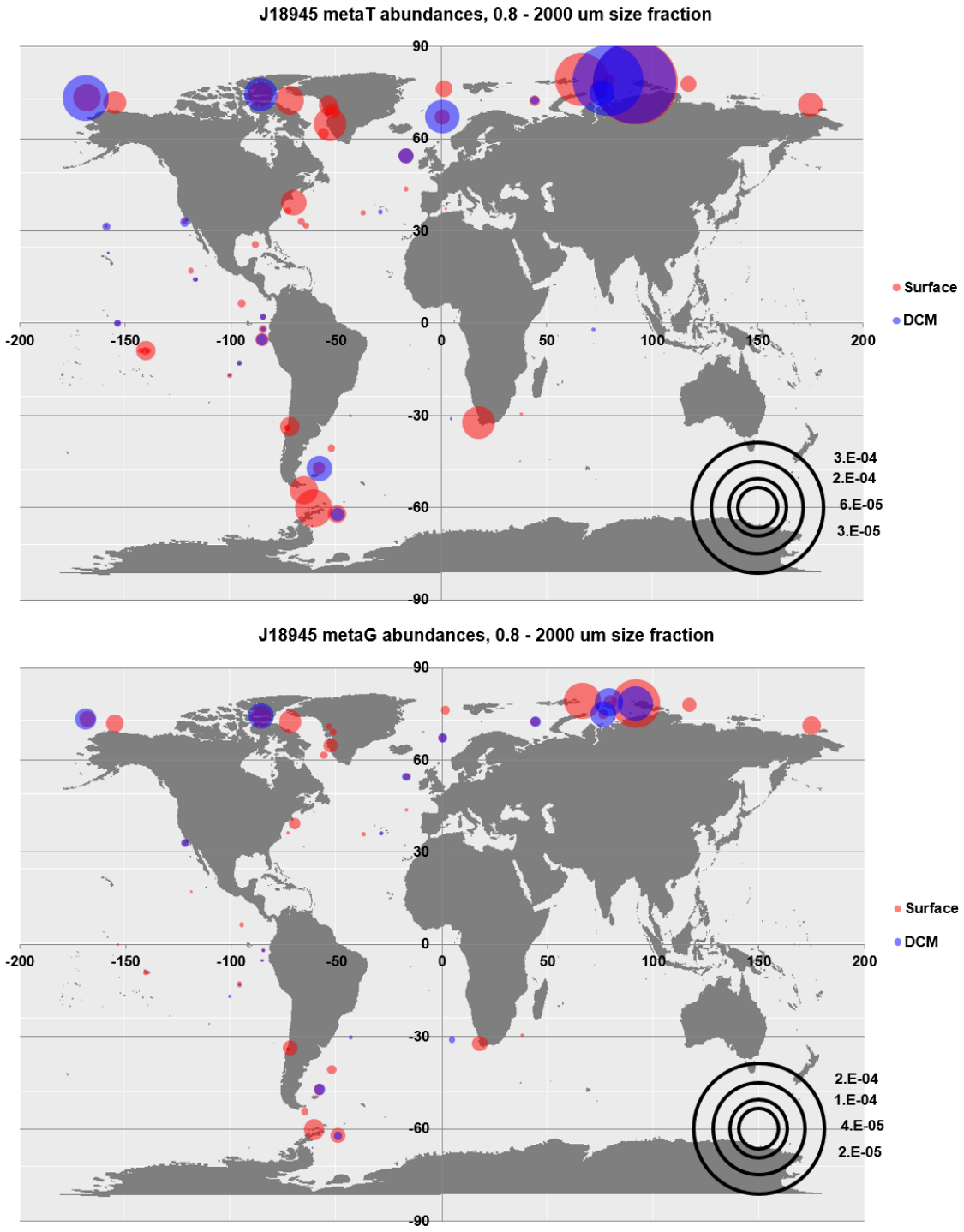

Figure 2.3 Diatom Trans3 homologue relative abundances in the global ocean. Relative abundances of phylogenetically identified diatom Trans3 meta-genes were calculated across *Tara* Oceans and *Tara* Polar Circle data across all size fractions, and both surface and DCM depths following the methodology of Annex I (Liu et al., 2022). Relative abundances are per million total sequenced meta-gene reads, with 1.0 corresponding to the total. Circle size is proportional to abundance. SUR, surface, DCM, deep chlorophyll maximum. The homologues were identified by Richard Dorrell, while I generated the maps.

Chapter 3: Localization of Trans3 in the model diatom *P. tricornutum*

Abstract

Previously, I have generated an *in silico* transcriptomic, phylogenomic and environmental atlas of predicted plastid transporters encoded by the *P. tricornutum* genome (Chapter 2). Within these data, a novel MFS superfamily transporter hereafter named Trans3 was particularly interesting, since its reciprocal BLAST best-hit (RbH) homologues were restricted to species containing secondary red plastids, and the gene encoding this transporter (Trans3) showed strong coregulation with the mitochondrial genome, based on microarray data. While the nuclear encoded Trans3 transporter protein *in silico* targeting prediction is coherent with a plastid localization (per ASAFind: an N-terminal signal peptide, followed by a phenylalanine residue in the cleavage +1 position), the precise membrane localization and orientation of Trans3 remain unknown. This chapter presents *in vivo* observations of Trans3 localization using eGFP and self-assembling GFP. Trans3 was located at the peri-plastidial membrane (PPM) in the diatom plastid implying that Trans3 participates in substrate transport between the cER and PPC. These results demonstrate how the self-assembling GFP system can determine the compartmentalization and orientation of diatom plastid transporters, providing a valuable tool for further research into the transport machinery of the diatom plastid.

3.1 Introduction

3.1.1 The diatom four membrane enclosed plastid and its protein transport system

In contrast to the double membrane-surrouded primary chloroplast, diatoms possess plastids acquired through secondary endosymbiosis, characterized by four membrane layers: the cERM, PPM, OEM, and IEM (Bhattacharya et al., 2004; Pfeil et al., 2014; Murik et al., 2019). The cERM is directly connected to the outer nuclear membrane (ONE) and endoplasmic reticulum (ER), and studded with ribosomes (Felsner et al., 2010; Solymosi, 2012; Marchand et al., 2018).

Diatom plastid proteins are mainly encoded by the nucleus and transferred to the plastid as pre-proteins (Green, 2011; Gruber et al., 2015; Dorrell et al., 2017; Gruber and Kroth, 2017). Due to the complex membrane structure of diatom plastid, the destination of plastid-localized proteins after translation requires a more intricate machinery than two membranes-enclosed plastids that found in green lineage organisms. Diatom plastid proteins possess a bipartite targeting sequence (BTS) consisting of a signal peptide (SP) and a transit peptide-like sequence (TP) at the N-terminus as shown in Figure 3.1 (Gruber et al., 2007; Felsner et al., 2010). Diatom signal peptides share some similar features to those of other eukaryotes, and it has been shown that they are interchangeable between ER proteins and plastid-targeted proteins (Nielsen et al., 1997; Kilian and Kroth et al., 2005). Most of the ER/ cER-located proteins in diatoms can be identified with reasonably high specificity by analyzing their signal peptides using the SignalP version 3.0 program (Bendtsen et al., 2004; Gruber et al., 2015). A plastid-specific targeting motif, with the consensus sequence "ASAFAP", spans the signal peptide cleavage sequences of the BTS (Figure 3.1). In particular, the presence of an aromatic amino acid (Phenylalanine: F; Tryptophan: W; Tyrosine: Y or occasionally a Leucine: L at the TP position +1 after the cleavage site) has been confirmed to be crucial for plastid protein import, it can be specifically recognized by the PPC protein transport system SELMA (Kilian and Kroth, 2005; Gruber et al., 2007). Based on this character, prediction tools such as HECTAR (Gschloessl et al., 2008) and ASAFind (Gruber et al., 2015; Gruber et al., 2023) have been developed for the prediction of plastid proteins in diatoms and other organisms with the "ASAFAP" motif in their nucleus-encoded plastid-targeted proteins. Most recently, an updated version of ASAFind has been published for the prediction of probable PPC proteins (Gruber et al., 2023).

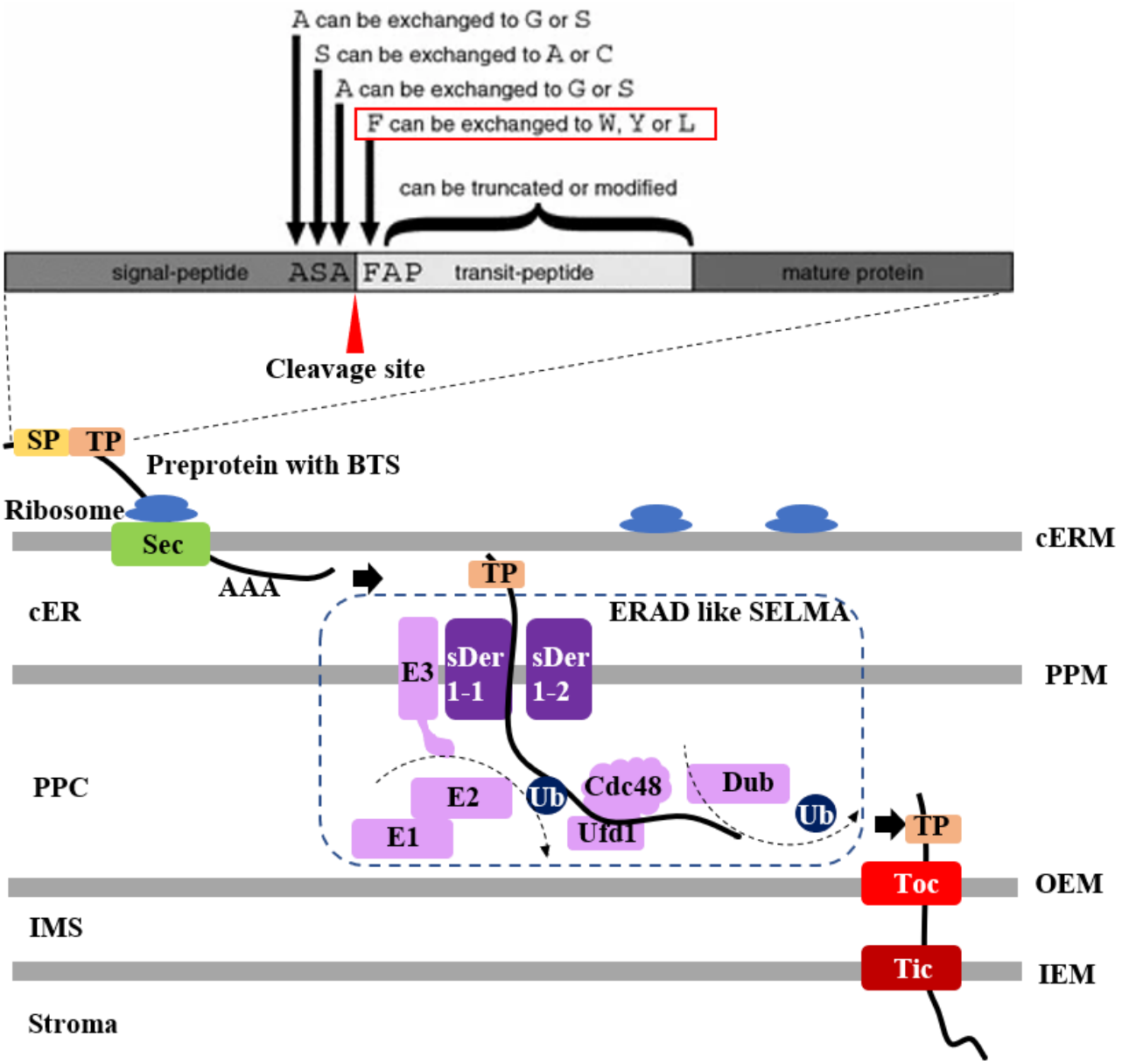

Figure 3.1 Structure of the BTS and simplified model of pre-protein import system into the diatom plastid. The BTS (SP+TP) image is adapted from Gruber et al. (2007) and slightly modified. Pre-proteins are co-translationally imported into the cER via the Sec protein translocon. The ERAD-like machinery (SELMA) is responsible for pre-protein import across the second outermost membrane PPM. The membrane specific proteins sDer1-1- and sDer1-2 form the translocation channel. Pre-proteins are ubiquitinated by the PPC-specific E1 and E2 enzymes and the ubiquitin ligase E3. Subsequently, the protein is bound by the ubiquitin-dependent ATPase Cdc48 and the cofactor Ufd1, which mediate extraction of the substrate into the PPC. Following translocation, ubiquitin is removed by the enzyme DUP. Finally, PPC-specific proteins are folded and processed, whereas stromal pre-proteins are transported across the OEM and IEM by a TOC/TIC-mediated mechanism (Huesgen et al., 2013).

A simplified process of protein transport into diatom plastids is shown in Figure 3.1. Pre-proteins are co-translationally imported into the cERM via the Sec protein translocon, and the SP in the preprotein BTS is cleaved off after import into the cER lumen (Kilian et al., 2005). After import into the cER, plastid proteins are transported into the PPC via an ERAD-derived translocon called SELMA (Hempel et al., 2010; Lau et al., 2015; Lau et al., 2016). The core components of SELMA are the symbiont-specific membrane proteins Der1 (sDer1-1 and sDer1-2), which putatively form the translocation channel (Hempel et al., 2010). A "vesicular shuttle model" has also been proposed previously for PPC protein import, however, the discovery of SELMA and the lack of putative PPC proteins acting in lipid trafficking make the mechanism for this import pathway less evident (Kilian and Kroth, 2005; Flori et al., 2017). Stromal preproteins are likely transported across the third and fourth membrane by a TOC/TIC-mediated translocon similar to that of higher plants (Bolte et al., 2009; Gruber and Kroth, 2017).

3.1.2 Diatom plastid ion and metabolic transporters localization strategy

To supply the cell and the organism with primary metabolites, a large number of substrates have to be transported across the diatom plastid by metabolite transporters present in the envelope membranes (Marchand et al., 2018). In diatom four-membrane-enclosed plastids, each membrane may require its own specific transporter to facilitate the transport of molecules across the specific membrane. For instance, a study of four potential diatom plastid triose phosphate transporters (TPT) revealed that TPT1 was located in the cER while the others were predicted to be located at the PPC and IEM, although no OEM located TPT was found (Moog et al., 2015; Lau et al., 2015). This suggests that the transport of triose phosphate may require the coordination of multiple transporters located on different membranes. The complex process of diatom plastid transport highlights the importance of plastid transporter substructural localization confirmation for understanding its function.

Although several prediction tools are available for predicting diatom plastid protein localization (ASAFind, typically used with SignalP version 3.0, or HECTAR) (Bendtsen et al., 2004; Gschloessl et al., 2008; Gruber et al., 2015; Gruber et al., 2023), the structural and topological divergence of transmembrane proteins in

organisms possessing secondary red plastids has made accurate predictions challenging (Hirakawa et al., 2009; Agrawal and Striepen., 2010). Therefore, *in vivo* experimental subcellular localization studies remain essential for studying plastid transporters.

A widely used molecular technique for diatom localization experimental research is the expression of enhanced green fluorescent protein (eGFP) in transformant cell lines (Apt et al., 2002; Kilian and Kroth, 2005). Both constitutive and inducible promoters and terminators have been adapted for GFP-tagged protein expression in diatoms (Apt et al., 1996). Moreover, some known reporter constructs have been established for multiple diatom organelles, allowing the inference of protein co-localization (Apt et al., 2002; Kilian and Kroth, 2005; Liu et al., 2016). Despite the development of eGFP as a mature localization strategy, the localizations of diatom plastid transporters remain poorly understood in comparison to those of higher plants. In the *P. tricornutum* genome, only two TPTs (J24610 and J54017) (Moog et al., 2015; Kamikawa et al., 2017; Moog et al., 2020) and NTT1 (J49533) (Ast et al., 2009; Chu et al., 2017) have been confirmed experimentally as plastid-localized transporters, consistent in both cases with ASAFind-positive targeting predictions. Two putative phosphate permeases (J23830 and J33266) have further been suggested to localize to the ER/nuclear envelope in *P. tricornutum* (Dell'Aquila et al., 2020), although ASAFind predicts a plastid localization for J23830 in Chapter 2 (Liu et al., 2022). Although the eGFP technique has provided valuable insights into protein localization determination, it has certain limitations in the study of four-membrane enclosed complex plastids. Specifically, this technique is restricted in its ability to accurately visualize the compartments within these structures and to definitively determine at which membrane the target protein should be located. In particular, different diatom plastid membranes may have similar topologies and organizations, and be difficult to tell apart by eye.

In order to gain a deeper understanding of cellular complex compartmentalization and targeting mechanisms, self-assembling GFP has been established to study protein compartmentalization, orientation and topology in complex transmembrane systems (Cabantous et al., 2005). Compared to eGFP, the utilization of self-assembling GFP enables precise labeling and tracking of proteins in sub-compartments. This technique allows the accurate

determination of a transporter localization within specific membranes and provides insights into its orientation, thereby informing our understanding of the roles of transporters in substrate import or export. Self-assembling GFP technique involves dividing the GFP protein into two non-fluorescent fragments: β-strands 1-10, which is fused to a reporter gene with a known localization, and β-strand 11, which is fused with the target protein of interest (Cabantous et al., 2005). When the reporter gene and target gene are located in the same cellular compartment, the two GFP fragments are brought into close proximity and can spontaneously reassemble to form a fluorescent GFP molecule. The resulting fluorescence can then be observed to determine the location of the target protein among different compartments within the cell (Cabantous et al., 2005).

Several standard reporter genes have been established in diatoms for self-assembling GFP (Figure 3.2). These include the mitochondria outer membrane-localized and cytoplasm-facing mitochondrial import receptor subunit protein (TOM70); the chloroplast endoplasmatic reticulum (cER)-localized binding immunoglobulin protein (Bip); the periplastidal compartment (PPC)-located heat shock protein 70 (Hsp70); the intermembrane space (IMS)-localized protein monogalactosyl diacylglycerol synthase 1 (MGD1) and the stroma-localized glyceraldehyde-3-phosphate dehydrogenase protein (GAPDH) (Kilian and Kroth 2005; Gould et al., 2006; Kroth et al., 2008; Hempel et al., 2009; Bullmann et al., 2010; Chen et al., 2019; Marter et al., 2020) (Figure 3.2). All these proteins, except GAPDH, have confirmed localizations by biochemical techniques or by transmission electron microscopy (TEM), while the plastid-targeted GAPDH isoform is expected to localize to the stroma because of its participation in the Calvin cycle (Kroth et al., 2008). The localization of MGD1 is nonetheless debated between the IMS and other cellular compartments (GueGuen, 2022). These reporters (Figure 3.2) provide specific signals for the localizations of novel plastid-associated proteins.

Self-assembling Reporter Gene	Abbreviation	Confirmed Localization	Previous usage for localization
translocase of the outer mitochondria membrane protein	Tom70	OMM located, cytoplasm-facing	Marter et al., 2020
binding immunoglobulin protein	Bip	cER	Bullmann et al., 2009
heat shock protein 70	HSP70	PPC	Chen et al., 2019
mono-galactosyldiacylglycerol synthase 1	MGD1	IMS	Chen et al., 2019
glyceraldehyde-3-phosphate dehydrogenase protein	GAPDH	Stroma	Kroth et al., 2008

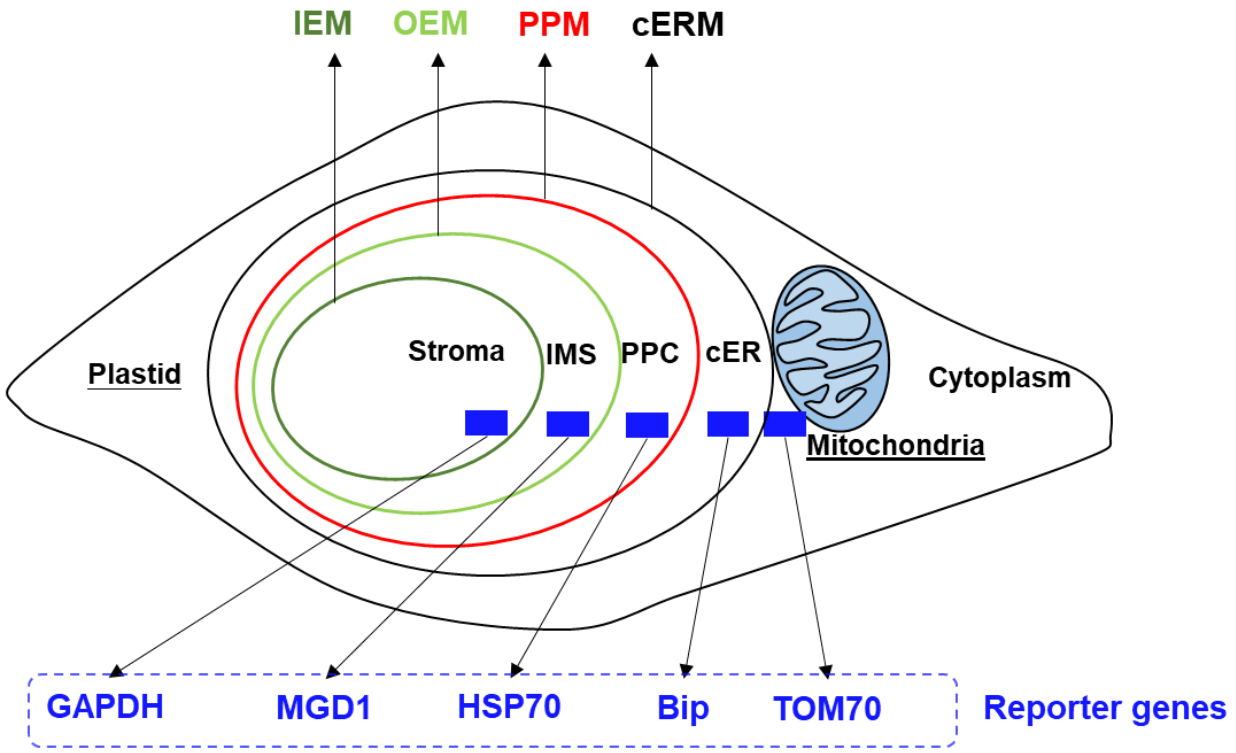

Figure 3.2 Reporter proteins used for self-assembling GFP in diatoms. The table above shows the full name, abbreviations and confirmed biochemical or TEM-based localizations of proteins that can be used as self-assembling GFP reporter genes. The diagram below indicates the localization of each reporter (blue box) in the *P. tricornutum* cell.

Self-assembling GFP in diatoms has previously been used to localize putative transport systems proteins to the second (peri-plastidial, PPM) and the third (outer envelope, OEM) membranes (Bullmann et al., 2010; Hempel et al., 2009; Hempel et al., 2010; Lau et al., 2015; Lau et al., 2016). Using self-assembling GFP, the SELMA-related subunits sDer1-1, sDer1-2 and the ubiquitin ligase ptE3P have been inferred to locate to the *P. tricornutum* PPM (Hempel et al., 2009; Hempel et al., 2010; Lau et al., 2016). Similarly, an Omp85-related transport protein has been confirmed to localize to the *P. tricornutum* plastid OEM using self-assembling GFP (Bullmann et al., 2010). Amongst other diatom plastid transporters, only triose phosphate transporters (TPT) have been studied to date by self-assembling GFP. The localizations of four potential TPT transporters in *P. tricornutum* (Moog et al., 2015) and their homologues in the non-

photosynthetic diatom *Nitzschia sp.* (Moog et al., 2020) have been verified by this technique, with TPT1 located in cER membrane and the others in the IEM (Moog et al., 2015; Moog et al., 2020). Currently, to my knowledge, only one protein topology and orientation has been tested by self-assembling GFP in algae, relating to a *Saccharina japonica* translocon Tic20 expressed heterologously in *P. tricornutum* cells (Chen et at., 2019). SjTic20 was detected by different reporter genes combined with different lengths of the Tic20 fragments, with the SjTic20 inferred to possess an "M"-type structure with N terminal domain (NTD) and -C terminal domain both located in the IEM (Chen et al., 2019).

In this chapter, I present the localization of different Trans3 CDS-eGFP and self-assembling GFP cell lines, in order to validate the probable localization, orientation and topological structure of the Trans3 protein.

3.2 Results

3.2.1 eGF P labelling suggests that Trans3 is located between the *P. tricornutum* plastids and mitochondria

First, to consider its potential subcellular localization of Trans3, the orientation and topological features of the *P. tricornutum* peptide sequence were considered *in silico*. Structural prediction of the 562 amino acid (aa) Trans3 protein sequence (shown in Methods) indicated a 27 aa N-terminal signal peptide, as identified by SignalP (Figure 3.3). Trans3 subcellular localization prediction with HECTAR only recovered the signal peptide with no downstream chloroplast targeting sequence prediction. Localization prediction with ASAFind version 2.0 recovered plastid targeting with low confidence, while the aromatic amino acid Phenylalanine (F) was found in the cleavage site (27 aa) +1 position. These targeting predictions are consistent, based on previous studies (Gruber et al., 2015), with a PPM localization of Trans3. Transmembrane topology analyses of Trans3 with TMHMM detected at least five credible transmembrane domains (TMD) with high probability from 408 aa to 550 aa, alongside a further five potential TMDs with lower probability (0.6-0.9), as shown by purple dashed box in Figure 3.3.

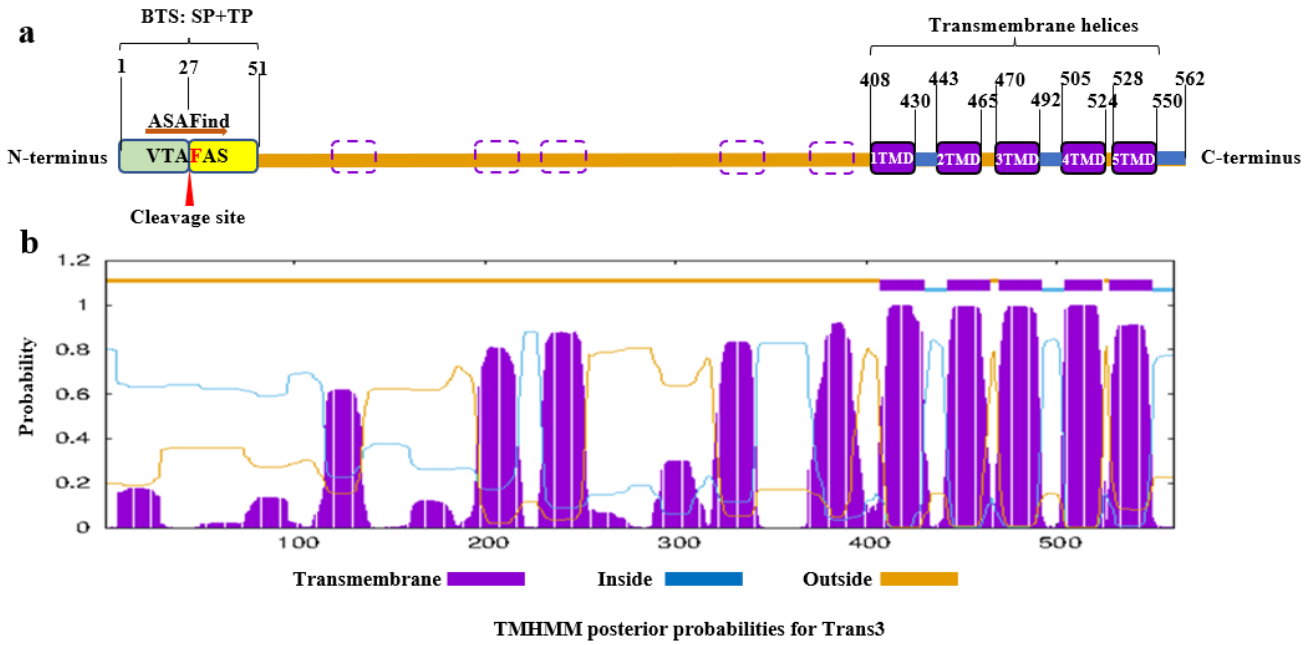

Figure 3.3 Trans3 signal peptide and transmembrane domain (TMD) predictions. a: Schematic of Trans3 and TMD prediction. The signal peptide (SP) is shown as a green box, followed by a yellow box representing the transit peptide-like sequence (TP) in the N-terminus. Credible TMDs, as predicted by TMHMM, are represented by solid purple boxes, and the five probable TMDs identified with low confidence are shown by purple dashed boxes. Numbers relate to the amino acid positions of each TMD, and the red triangle represents the cleavage site after the signal peptide. **b:** The TMHMM prediction result of Trans3 TMDs. Purple lines represent the probability that amino acids are predicted as transmembrane domains, potential TMDs are shown as purple peaks, and credible TMDs are shown as purple rectangles. Blue lines represent regions with probable localisations inside the membrane bilayer, and orange lines represent regions with probable localisations outside the membrane bilayer. The inside and outside only represent the relative sides of the located membrane, not the real orientation of the protein.

Prior to my PhD, a Trans3_full_eGFP construct was made in the host lab using the full length of the Trans3 coding sequence (CDS) with green fluorescent protein (GFP). The inferred subcellular localization of this construct is shown in Figure 3.4. Trans3 fluorescence was detected by confocal microscopy as a bright spot (shown in green) adjacent to the plastid (red), and was inferred using MitoTracker staining (false-colored blue) to localize close to the mitochondrion (Figure 3.4).

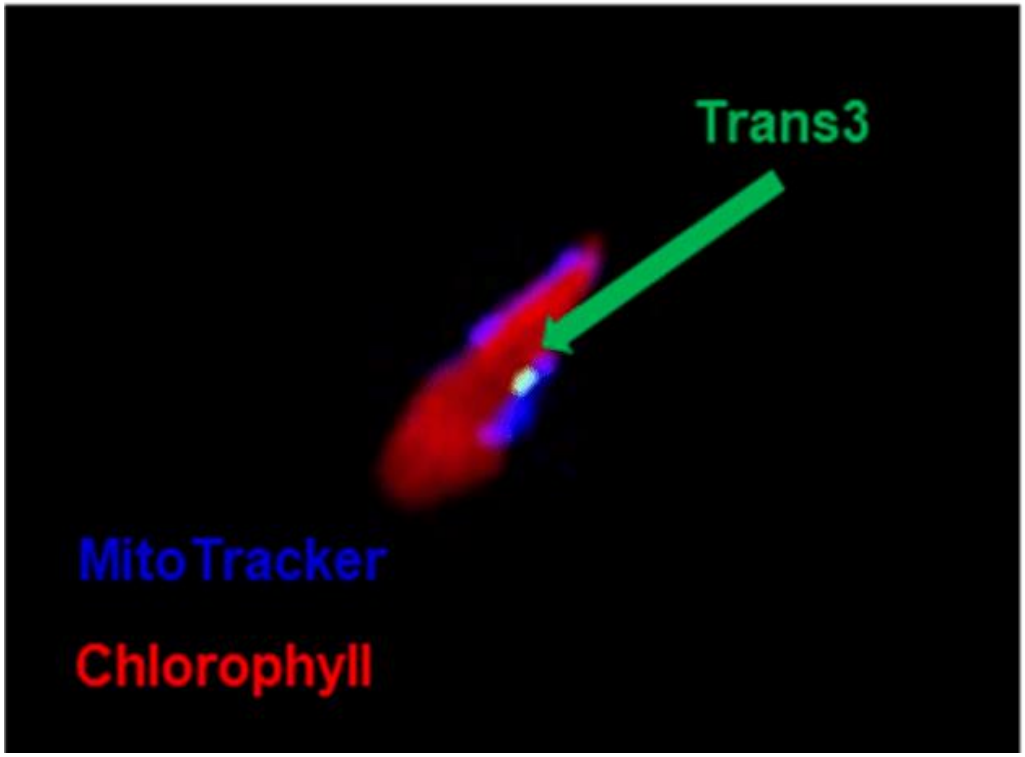

Figure 3.4 The previously generated Trans3 CDS-eGFP confocal result. MitoTracker Orange (1/5000 dilution with 15 mins incubation) is false-colored blue to allow its visualization, red represents chloroplast auto-fluorescence, and the green spot represents the eGFP-labeled Trans3 localization. This picture was taken by a cell line expressing a Trans3 full-eGFP construct previously generated by Tomomi Nonoyama and re-visualized by me.

3.2.2 Trans3 is located at the PPM by self-assembling GFP

To further determine Trans3 subcellular plastid membrane localization, orientation and topology, five proteins (Mitochondrial outer membrane located and cytoplasm facing-TOM70, cER located-Bip, PPC located-HSP70, IMS located-MGD1, stroma located-GAPDH) with different confirmed localizations in four-membranes enclosed plastid were selected as reporter genes (Figure 3.2) for generating Trans3 self-assembling GFP line.

The Trans3 self-assembling transformation vector (Figure 3.5) was generated from a template plasmid from Daniel Moog (University of Marburg), based on the pPha_DUAL_2xNR (NCBI accession no. JN180664) vector with a modified NR promoter and truncated NR terminator. Each reporter gene was fused with GFP1-10, and different lengths of the Trans3 fragment (0TMD, 1TMD, 2TMD, 3TMD, 4TMD, and full length CDS, see Methods) were fused to GFP11. The 2x

NR promoters allow the selective induction of the fluorescence signal when cell lines are transferred from ammonium (repressive) to nitrate-supplemented (inductive) source media (Hempel et al., 2009), which decreases the possibility of mislabeling or overexpression.

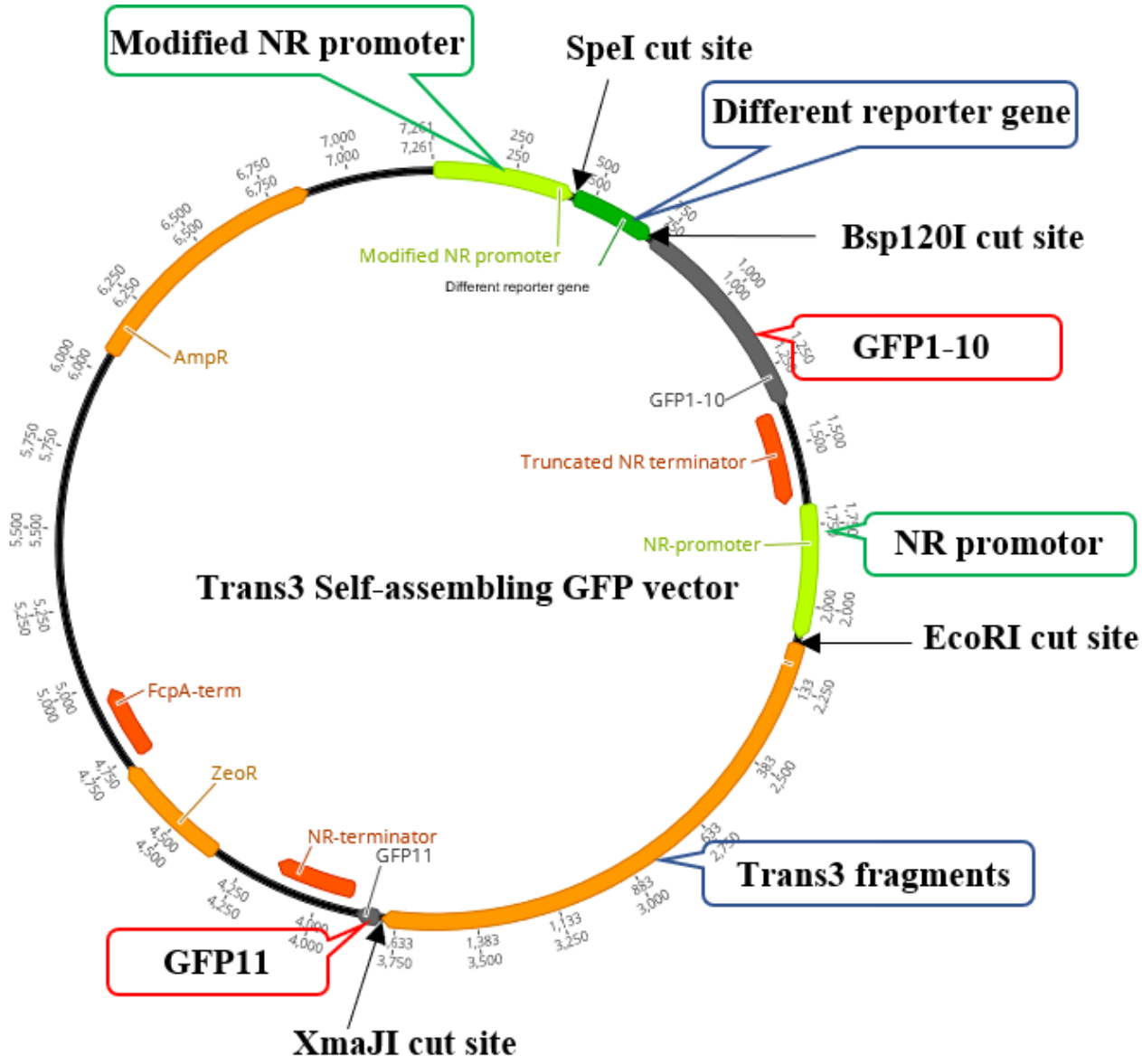

Figure 3.5 pPha 2x NR Trans3 self-assembling GFP vector structure. Reporter genes were incorporated upstream of the GFP1-GFP10 regions; and different Trans3 fragments were introduced upstream of GFP11. NR promoter and terminator sequences, resistance marker genes, and restriction cut sites used for cloning are shown.

After the synthesis of Trans3 self-assembling vectors and biolistic transformation, *Phaeodactylum* cells were checked for the presence of

fluorescent signals using confocal microscopy. First, the self-assembling GFP results of full-length Trans3 were investigated with five different reporter genes (TOM70, Bip, HSP70, MGD1, GAPDH). *Phaeodactylum* cells transformed with constructs of C-terminal full-length GFP fusions of each reporter protein, under a constitutively expressed Fcp promoter gave positive signals and the expected localisations to their corresponding organelle. No fluorescence signal was detected when Trans3 full length (562 aa) was co-expressed with TOM70, Bip, MGD1 and GAPDH self-assembling GFP (Figure 3.6). Next, the sequences of the Trans3 C-terminus fused with GFP11 and the sequences of reporter genes (Bip, GAPDH, TOM70, MGD1) C-terminus fused with GFP1-10 were amplified by Reverse Transcription (RT-) PCR of RNA samples produced from these transformed lines, using primers listed in Thesis Annex II. In each case, gave expected size bands were observed (Figure 3.7), confirming the successful transformation and transcription of self-assembling GFP plasmids in each case, despite the absence of GFP fluorescence.

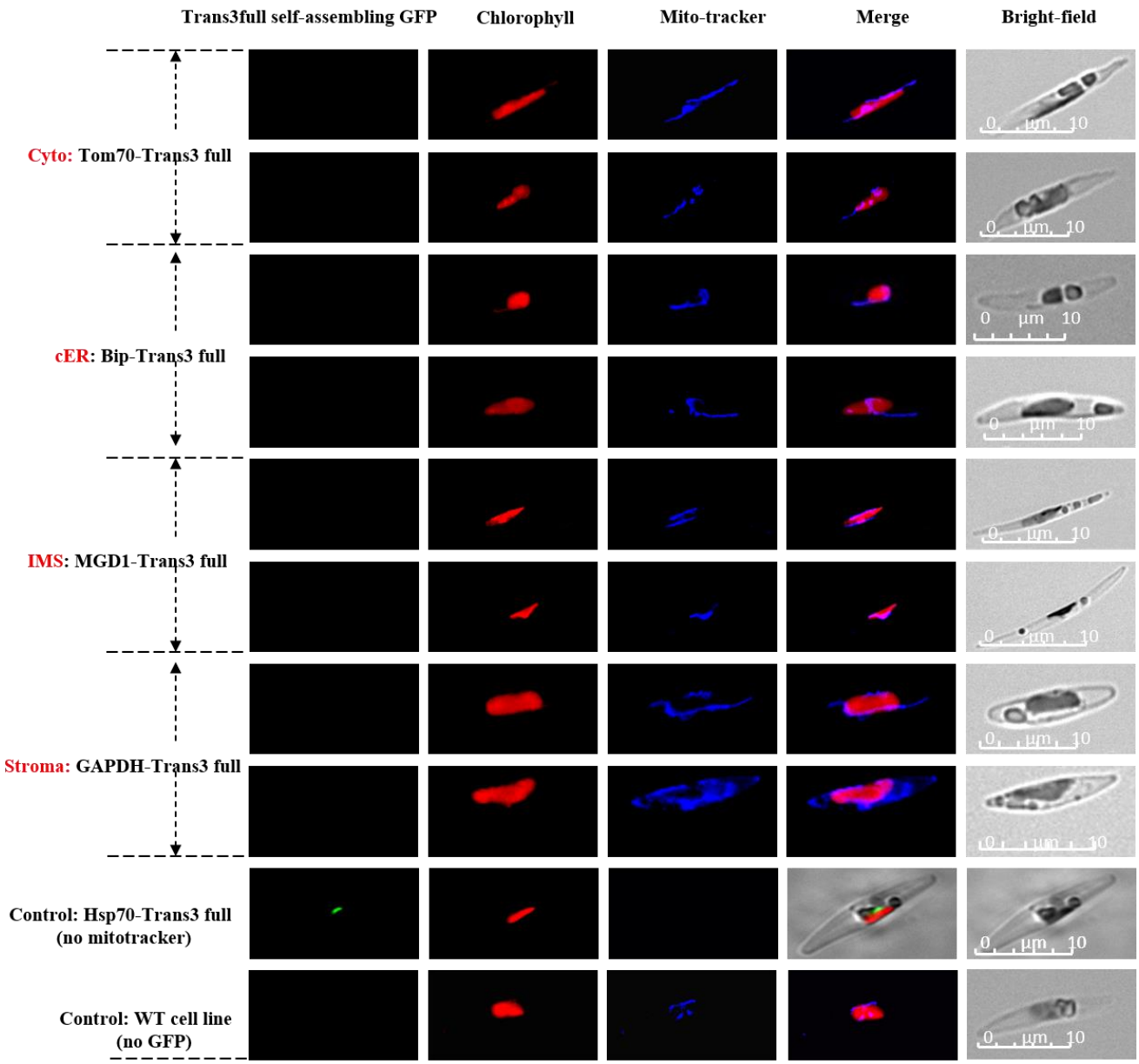

Figure 3.6 Tom70/Bip/MGD1/GAPDH- Trans3 full-length self-assembling GFP negative results. Fluorescence and detection conditions were: 485 nm excitation for GFP, with a 500-540 nm detection window; 548 nm for MitoTracker Orange (1/5000 dilution with 15 mins incubation) with detection between 575-585 nm; and between 650 and 700 nm detection for chlorophyll. MitoTracker here is false-colored blue to allow its visualization, Hsp70-Trans3 full self-assembling GFP without MitoTracker and the wild type (WT) line with MitoTracker were used as controls to maximize the positive fluorescence signal and to minimize the detection crosstalk between the GFP and MitoTracker.

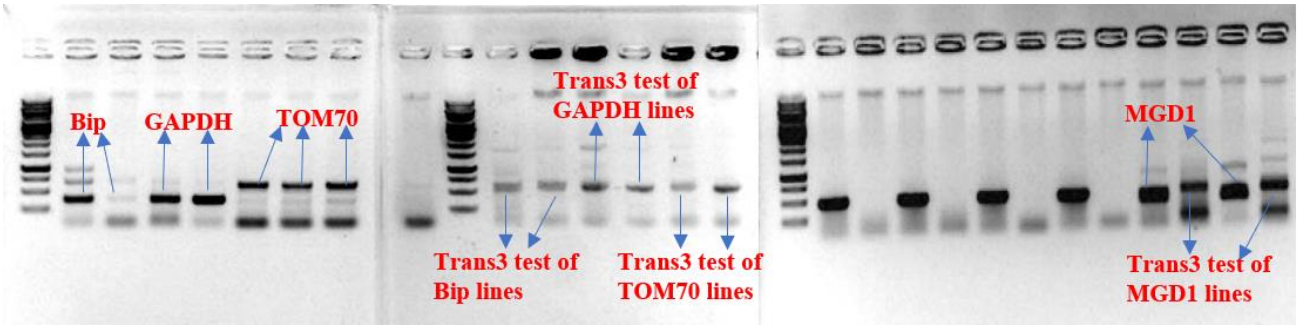

Figure 3.7 Reverse Transcription PCRs of the Trans3 fragment and reporter genes of negative fluorescence lines transformed with Bip/GAPDH/TOM70/MGD1-Trans3 full-length self-assembling GFP plasmids. The Trans3 test amplified the Trans3 C-terminus fused with GFP11. The Bip/GAPDH/TOM70/MGD1 tests amplified the respective reporter gene C-termini fused with GFP1-10. Template negative and primer negative controls are tested for each reaction.

In contrast, green fluorescent signals were detected as bright spots, resembling those of Trans3-CDS eGFP lines, when Trans3 full length protein was co-expressed with the PPC reporter Hsp70, confirming the PPC location of the Trans3 C-terminus (Figure 3.8). To confirm that the positive GFP fluorescence signal was not a result of mislabeling or overexpression, the ESAW nitrogen source was changed to ammonium which will repress the NR promoter (Hempel et al., 2009). Under ammonium, no fluorescence result was detected (Figure 3.8). Subsequently, the signal was reactivated by reverting the cell line back to the nitrate source ESAW, confirming specificity of the localization (Figure 3.8).

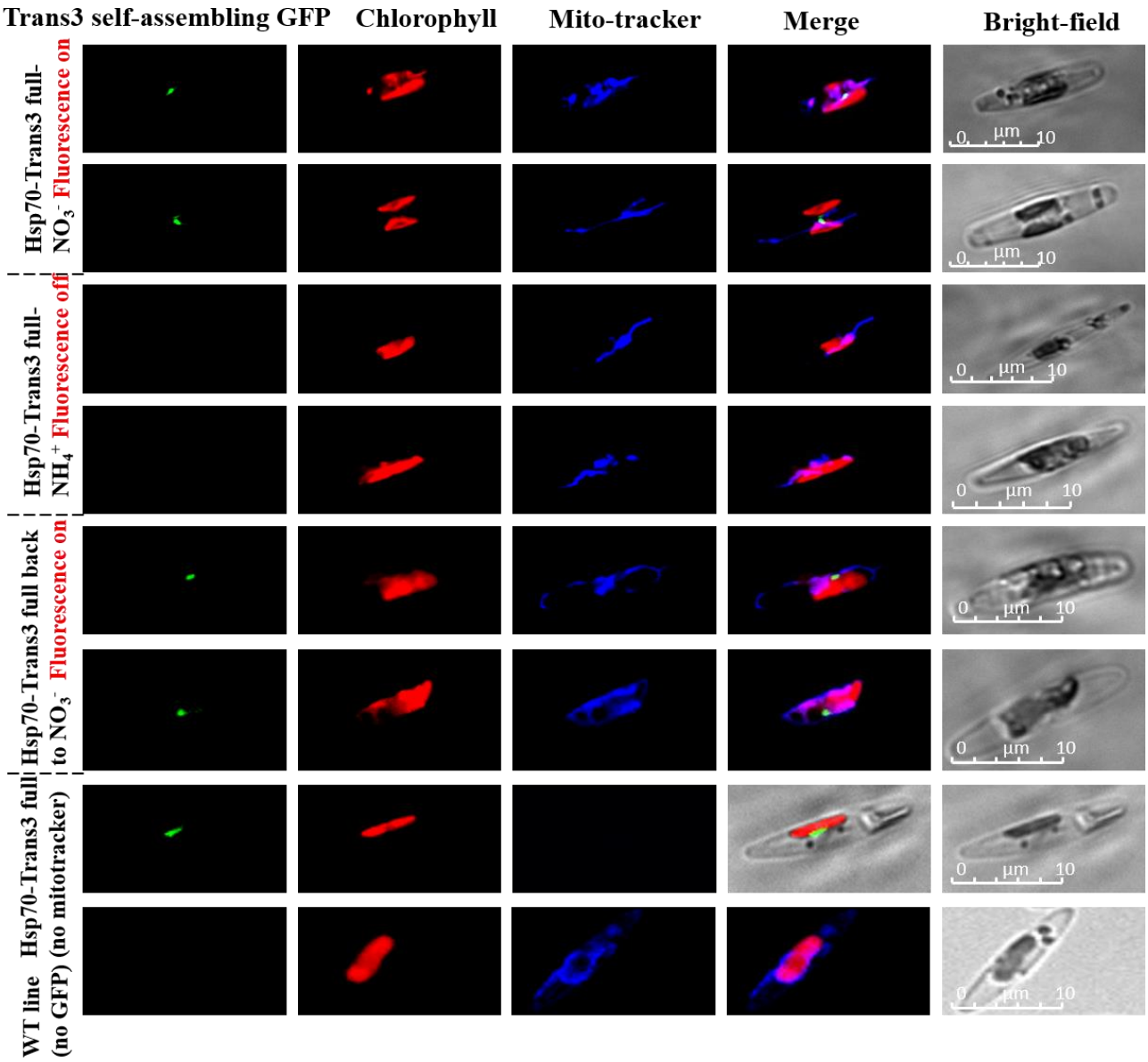

Figure 3.8 PPC localization of Trans3 C-terminus localization inferred using Hsp70-Trans3 full-length self-assembling GFP lines. Fluorescence and detection conditions are as shown in Figure 3.6. WT, wild type. NO_3^-: cell lines grown in nitrate source ESAW media under which the NR promoter will be induced; NH_4^+: cell lines incubated in ammonium source ESAW media to shut off the fluorescence signal; back to NO_3^-: cell lines initially incubated in ammonium source ESAW media, prior to washing and more than three hours incubation in nitrate media to transiently activate the fluorescence signal.

Although the HSP70-Trans3 full-length self-assembling GFP result indicated that Trans3 C-terminus located at PPC, the further TMD-truncated self-assembling cell lines need to be generated to determine the orientation and topological structure of Trans3. *Phaeodactylum* cells were then transformed with self-assembling GFP vectors containing TMD-truncated Trans3 fragments of varying

lengths with either PPC-Hsp70, or the reporter proteins associated with PPC-adjacent compartments (Bip-cER, and MGD1-IMS). Considering time limitations, along with the prediction result consistency and reliability (see Chapter 3 Discussion), we decided to use the five TMDs consistently predicted with high confidence to establish the self-assembling GFP lines, which were: NTD-110 aa, 0TMD-408 aa, 1TMD-443 aa, 2TMD-470 aa, 3TMD-505 aa and 4TMD-527/528 aa (Figure 3.3, details shown in Methods).

Positive green fluorescence signals were observed in self-assembling GFP cell lines expressing Bip-NTD and Bip-2TMD (Figure 3.9), indicating that the cER localization of the Trans3 NTD and 2TMD end. Bip-NTD and Bip-2TMD cell lines showed different fluorescence expression patterns compared with PPM-localized results, with signals spread further across the plastid surface, rather than a single spot (Figure 3.9). In contrast, positive green fluorescence signals were observed in self-assembling GFP cell lines expressing HSP-0TMD, HSP-1TMD, HSP-3TMD, and HSP-4TMD, suggesting a PPC localization of the Trans3 0TMD, 1TMD, 3TMD and 4TMD ends. HSP-1TMD, HSP-3TMD cell lines showed the same bright spot fluorescence result as the full-length Trans3-HSP70 self-assembling GFP results before (Figure 3.8; Figure 3.9). No MGD1 constructs, which have been tentatively proposed as a reporter line for the plastid intermembrane space and outer envelope membrane, gave positive fluorescence. These results confirmed that Trans3 is localized at the diatom PPM and is presumably involved in substrate transport between the *Phaeodactylum* cER and PPC.

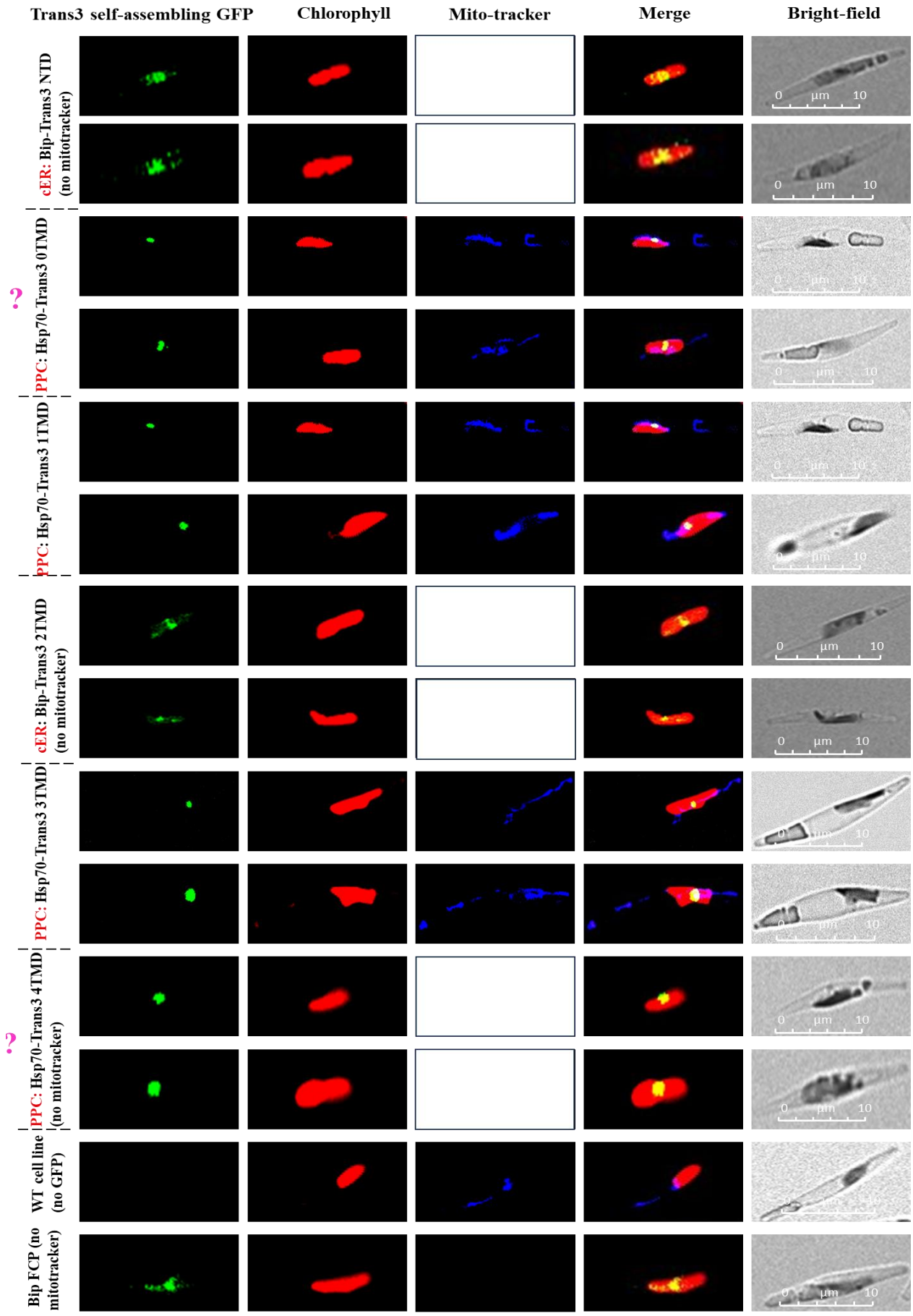
Trans3 self-assembling GFP
Chlorophyll
Mito-tracker
Merge
Bright-field
cER: Bip-Trans3 NTD (no mitotracker)
PPC: Hsp70-Trans3 0TMD
PPC: Hsp70-Trans3 1TMD
cER: Bip-Trans3 2TMD (no mitotracker)
PPC: Hsp70-Trans3 3TMD
PPC: Hsp70-Trans3 4TMD (no mitotracker)
WT cell line (no GFP)
Bip FCP (no mitotracker)
?
?
0 µm 10

Figure 3.9 Trans3 orientation and topology tests using self-assembling GFP. Fluorescence and detection conditions are shown as per Figure 3.6. Some cell lines (Hsp70-Trans3 0TMD; Hsp70-Trans3 1TMD; Hsp70-Trans3 3TMD) were further stained with Mitotracker to visualize the mitochondria, false-colored blue; while others (Bip-Trans3 NTD; Bip-Trans3 2TMD; Hsp70-Trans3 4TMD) were not, with Mitotracker images in each case depicted as white boxes. Question marks represent fluorescence results not consistent with the TMHMM prediction.

3.3 Discussion

In this chapter, I investigate the localization, orientation, and topology of a novel transporter protein, Trans3, in *P. tricornutum* cells using both eGFP and self-assembling GFP systems. I observed a bright spot fluorescence pattern, also called a "blob-like structure" (Kilian and Kroth 2005), which has been previously documented in studies on self-assembling GFP by Lau et al. (2016) and Moog et al. (2015). This pattern has been reported as a typical result for PPM-located proteins, even though PPM should be distributed throughout the entire plastid in the cell. I hypothesize that the observed pattern may be due to the tested protein's specific localization within the PPM adjacent to the blob-like structure, rather than being uniformly distributed across the entire membrane. Alternatively, the fluorescence results may reflect that adding the GFP fragment at the C-terminus results in a larger protein size, leading to protein clustering within the blob-like structure, generating a characteristic bright spot. Finally, it is likely that the PPC self-assembling GFP results are confined to the blob-like structure due to the likely concentration of the Hsp70 reporter protein in the PPC, as shown in previous studies (Liu et al., 2016; Dorrell et al., 2019).

This study represents an experimental confirmation of a diatom PPM metabolite transporter, with preliminary insights into its orientation, using self-assembling GFP. While previous studies have focused on diatom plastid protein import and cross-membrane systems (Bullmann et al., 2010; Hempel et al., 2009; Hempel et al., 2010; Lau et al., 2015; Lau et al., 2016; Chen et al., 2019), the self-assembling GFP system has been underutilized in the study of plastid transporters related to ion and metabolite transport. Most studies to date have only tested C-terminal localizations of complete peptide sequences, neglecting the localization of C-terminal fusions of N-terminal fragments, which allow insights into the orientation and *in vivo* topology of the protein (Bullmann et al., 2010;

Hempel et al., 2009; Hempel et al., 2010; Lau et al., 2015; Moog et al., 2015; Lau et al., 2016; Moog et al., 2020).

Previous studies by Moog et al. (2015; 2020) have attempted to localize the N-terminal regions of diatom plastid TPTs by using self-assembling GFP. PPM transporter compartmentalization was more found to be more challenging to localize using this technique compared to other membrane-located proteins due to the signal peptide at the beginning of the N-terminus. In certain cases, e.g., with a predicted PPM-localized TPT transporter from *Phaeodactylum* and its homologue in the non-photosynthetic diatom *Nitzschia sp.*, both the cER and PPC reporter genes showed fluorescence with the target gene (Moog et al., 2015; Moog et al., 2020), which may suggest that some PPM transporters exist in multiple orientations.

In this chapter, a C-terminal fusion of an N-terminal fragment gave a visualization of the localization of the part of the interior of the Trans3 peptide sequence. These data indicate that the specific visualization of PPM transporters is possible by self-assembling GFP, and that insights into the localization of the protein N-terminus testing is feasible by using a C-terminal fusion of a target gene N-terminal fragment (Figure 3.9). N-terminal localization testing by self-assembling GFP has previously been reported on membrane protein import system related *Saccharina. japonica* Tic20 and *P. tricornutum* Om85 with the stroma-localized ATPase gamma subunit (AtpC) and PPC-localized HSP70 (Chen et al., 2019; Bullmann et al., 2010).

For topology testing using self-assembling GFP, it is thus possible to cut the transporter protein according to its TMDs, and use the fluorescence results of truncated sequences with different reporter genes to infer protein structure and orientation. However, the challenge with this approach is the precision of protein TMD prediction by different tools, which can yield different prediction results. For Trans3, five TMDs predicted by TMHMM and other tools were used to establish self-assembling GFP lines (Figure 3.3). As noted above, the fluorescence results were not entirely consistent with the TMD prediction, as Trans3-0TMD_HSP70 and Trans3-4TMD_HSP70 showed positive fluorescence results that were inconsistent with the TMD prediction (Figure 3.9; Figure 3.10).

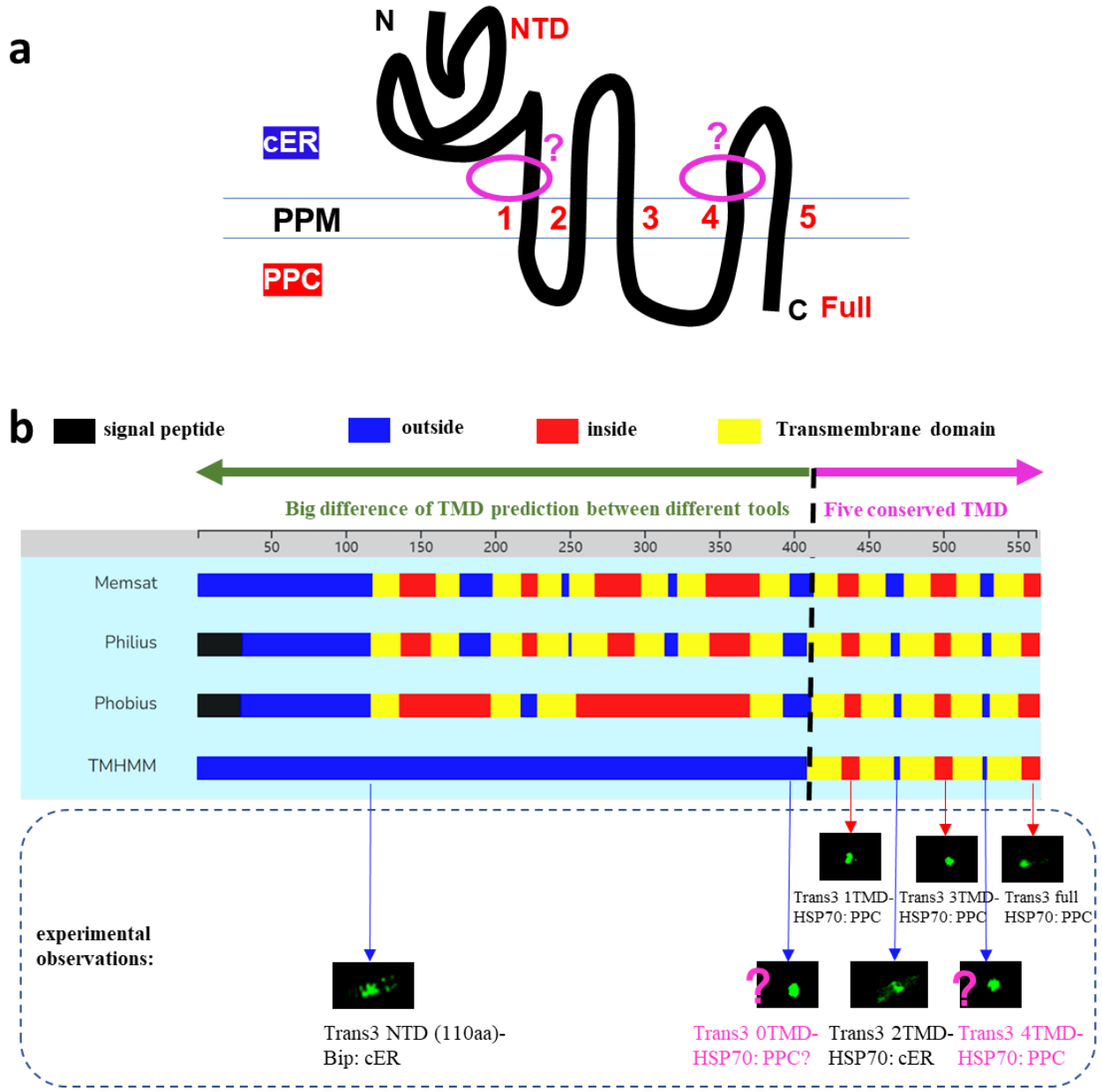

Figure 3.10 Trans3 topology and TMD prediction results by different tools and self-assembling experimental observations. a, Consensus Trans3 topology conclusion inferred from 5 confidently-predicted TMDs (red number) and the self-assembling GFP results. **b,** TMD prediction results obtained using different tools (Memsat, Philius, Phobius and TMHMM) and shown according to the Consensus Constrained TOPology prediction web server (CCTOP, Dobson et. al 2015). Experimental observations are shown underneath the TMHMM prediction. The pink question marks represent inconsistent results of the TMHMM predictions and the self-assembling GFP results.

In general, the self-assembling GFP results suggest a more complex Trans3 protein architecture than that inferred from TMHMM predictions (Figure 3.3;

Figure 3.9; Figure 3.10). The cER localization of the NTD construct (amino acids 1-110), despite the presence of a BTS between amino acids 1-52, would either suggest that the N-terminus of the Trans3 protein is reinserted across the PPM, or an undocumented TMD domain exists between amino acids 45 and 110. Similarly, the PPC localizations of the 0-TMD and 4-TMD constructs (which should produce cER localizations according to TMHMM) suggest an imprecise prediction of the Trans3 TMDs by TMHMM (Figure 3.9). The very short outside distance projected between certain TMDs adds to this uncertainty: for example, only three amino acids separate the fourth (505-524 aa) and final TMDs (528-550 aa) (Figure 3.3). The region between amino acids 110 and 400 may further possess TMDs, with weakly predicted domains from TMHMM (Figure 3.3) and non-uniform results produced by other TMD prediction tools (Figure 3.10). The complex transmembrane structure of the Trans3 protein renders the full reconstruction of its topology *in silico* complicated, and the exact number of associated membrane domains awaits further study (e.g., via a crystal structure).

Despite the challenges encountered in topology testing, in this chapter, I was able to establish a self-assembling GFP system with five reporter genes located between four diatom plastid membranes. I have confirmed Trans3 to be a PPM localization transporter, with the first (110aa-NTD) construct localizing outside (in the cER) and the C-terminal (full length) constructs localizing inside (in the PPM). These data show that using self-assembling GFP to test subcellular localization and orientation of plastid transporters in diatoms is possible, and can provide important insights into the fundamental biology of the diatom plastid.

3.4 Material and methods

3.4.1 Trans3 transmembrane domain prediction

Signal peptides were analyzed by SignalP 3.0 (Bendtsen et al., 2004) and Phobius in InterPro (Madeira et al., 2022), respectively. Protein subcellular localization was predicted using ASAFind version 2.0 (Gruber et al., 2015) and HECTAR (Gschloessl et al., 2008) under default conditions. The online TMHMM 2.0 server (https://services.healthtech.dtu.dk/services/TMHMM-2.0/) (Krogh et al., 2001) and consensus constrained topology prediction web server CCTOP

(https://cctop.ttk.hu/job/submit) (Dobson et al., 2015) tool were used to identify the Trans3 transmembrane domain (TMD) and the topology, run under default conditions.

3.4.2 *Phaeodactylum* cell culture

Wild-type *P. tricornutum* v 1.86 cells were grown in ESAW medium, under a 19℃ constant, 12h 50uE light: 12h dark cycling, and shaking at 100 rpm. The growth medium was enriched artificial seawater, ESAW (Falciatore et al., 2000). Salts and microelements required for media preparation listed in Table 3.1 (Note, the silica was not added for normal condition culturing. Autoclaved ESAW common solution (N-, P-, Fe-, antibiotic-, trace metal-) were unified preparation by the laboratory, then 4°C filtered 1000x N/P/Fe/Trace-metal stocks were added to the ESAW common solution according to Figure before utilization.

Table 3.1 ESAW media protocol and final concentration

ESAW	Composition	Concentration (moles/L)
common	Ca.Cl2	9.14E-03
	K.Cl	8.03E-03
	Mg.Cl2·6H2O	4.72E-02
	Na2.SO4	2.50E-02
	Na.Cl	3.63E-01
N source	Na.NO3/NH4Cl	5.49E-04
P source	NaH2.PO4	2.58E-05
C source	NaHCO3	2.07E-03
chelators	Na2.EDTA	8.30E-06
halogens	K.Br	7.25E-04
	Sr.Cl2	8.18E-05
	NaF	6.67E-05
	H3.B.O4	3.72E-04
Fe	ferric EDTA	8.69E-06
vitamins	cobalamin (B12)	7.38E-07
	biotin (B7)	4.09E-06
	thiamine (B1)	5.93E-04
trace metal	Mn.Cl2	2.73E-06
	Zn.Cl2	2.54E-07
	Na2.MoO4	6.12E-09
	Co.Cl2	6.72E-08
	Ni.SO4	6.27E-09
	Na2.SeO3	1.00E-09
	Cu.SO4	3.92E-08

Transformant cell lines were typically maintained on 1% agar plates containing a half-strength dilution of ESAW, alongside ampicillin (100 ug/mL), streptomycin (100 ug/mL) and chloramphenicol (30 ug/mL) to maintain axenicity, and zeocin (100 ug/mL) to maintain selection for transformation constructs. Plated colonies were inoculated in liquid medium and grown for 2-4 weeks prior to microscopy.

For NR promoter induction (Hempel et al., 2009), cell lines were grown in modified medium containing 5.49×10^{-4} moles/L NH_4Cl, as detailed below: Cell lines were centrifuged and washed with 1× marine Phosphate-Buffered Saline (mPBS) buffer (http://cshprotocols.cshlp.org/content/2006/1/pdb.rec8303) three times, then transferred to nitrate-free ESAW and incubated three hours to eliminate any residual cellular nitrate reserves. Cell lines were centrifuged and washed again, then incubated in $NaNO_3$-supplemented ESAW medium again more than three hours before confocal microscopy to allow transient and specific induction of the self-assembling GFP construct. This transient induction was performed to minimize non-specific localization of the self-assembling GFP construct.

3.4.3 Trans3 eGFP and self-assembling GFP plasmid generation

A previously generated eGFP (mNeonGreen) C-terminal fusion *Phaeodactylum* transformant line was maintained on zeocin selective medium, as above. This line was generated by biolistic transformation of wild-type cells, with a construct consisting the 3'-end of the complete Trans3 CDS without a terminal codon, then integrated into the pPhat vector, which contains antibiotic resistance cassettes for *E. coli* selection, a diatom cassette for diatom selection on phleomycin (Dorrell et al., 2017; Falciatore et al., 1999). Vector were constructed with the Gibson HiFi DNA Assembly Master Mix Kit (New England Biolabs, E2621).

The self-assembling GFP method (Cabantous et al., 2005) was utilized for subcellular localization, orientation and topology analyses for plastid transporter in this chapter. The *P. tricornutum* Trans3 self-assembling transformation vector was generated from a self-assembling plasmid from University of Marburg. This plasmid was based on pPha_DUAL_2xNR (NCBI accession no. JN180664) with a modified NR promoter and truncated NR

terminator (Figure 3.5). The first multiple-cloning site generated by SpeI and SacII restriction sites allowed for the insertion of five different reporter genes with 3'-ends combined with the large GFP fragment (GFP1-10). The five reporter genes used were: Cytoplasm-localized mitochondrial import receptor subunit protein (TOM70) (Marter et al., 2020), cER-localized binding immunoglobulin protein (Bip) (Bullmann et al., 2010), PPC-localized heat shock protein 70 (Hsp70) (Chen et al., 2019), IMS-localized protein mono-galactosyldiacylglycerol synthase 1 (MGD1) (Chen et al., 2019) and stroma-localized glyceraldehyde-3-phosphate dehydrogenase protein (GAPDH) (Kroth et al., 2008) (Figure 3.2). The second multiple-cloning site between EcoRI and HindIII restriction sites allowed for the insertion of different length of Trans3 sequence (110aa, 0TMD, 1TMD, 2TMD, 3TMD, 4TMD and Full length) (Figure 3.5) with the 3'-end fused to the small GFP fragments (GFP11).

Trans3 fragment and reporter gene amplification were amplified from *P. tricornutum* wild type line cDNA. Primers used are listed in Thesis Annex II. Phusion High-Fidelity DNA Polymerase was used for the amplification (ThermoFisher F-530XL). In the aim of making the vector construction easier, an additional restriction cut site of Bsp120I was inserted between the reporter gene and GFP1-10, and another restriction cut site of XmaJI was inserted between the Trans3 region and GFP11 (Figure 3.5). The primers used for generating the Trans3 region and establishing vectors for localization are listed in Thesis Annex II. Some vectors were constructed via traditional restriction ligation with a T4 DNA ligase Kit (New England Biolabs, M0202) according to the protocol, while others were constructed using the Gibson HiFi DNA Assembly Master Mix Kit (New England Biolabs, E2621) as detailed above. Vector sequences were verified by Sanger sequencing (Eurofins) before use.

Trans3 Fragments	Length	Self-assembling plasmid establishing				
		Tom70	Bip	HSP70	MGD1	GAPDH
Trans3 NTD	110 aa	-	+	+	+	-
Trans3 0TMD	408 aa	-	+	+	+	-
Trans3 1TMD	443 aa	-	+	+	+	-
Trans3 2TMD	470 aa	-	+	+	+	-
Trans3 3TMD	505 aa	-	+	+	+	-
Trans3 4TMD	527/528 aa	-	+	+	+	-
Trans3 Full length	562 aa	+	+	+	+	+

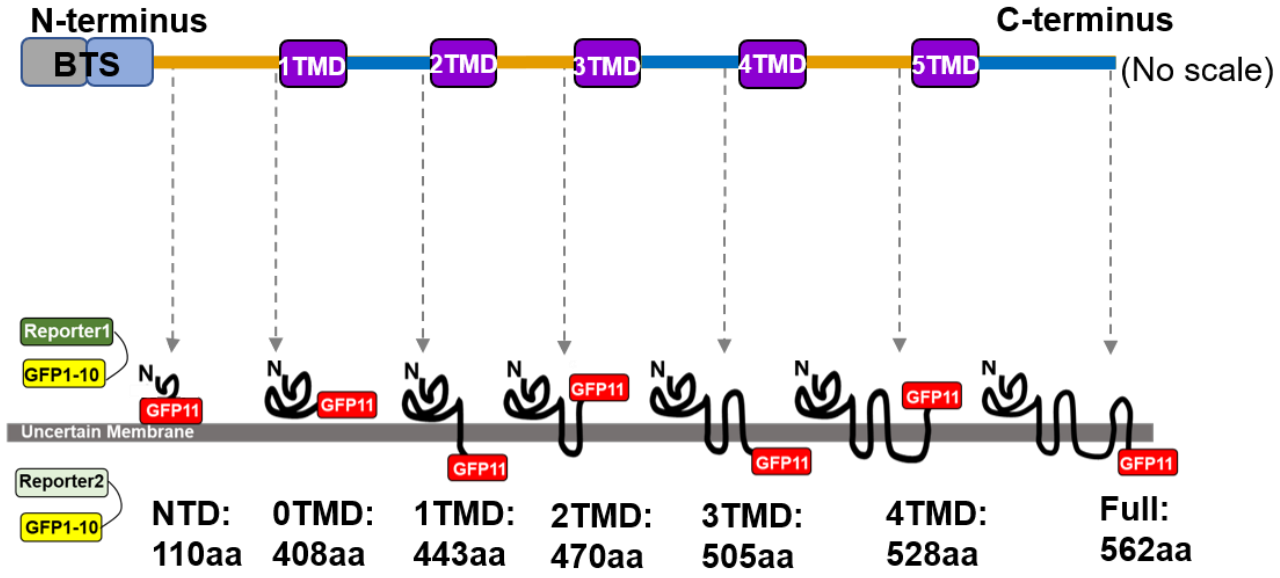

Figure 3.11 Trans3 self-assembling GFP plasmids and sketch map. a, Overview of Trans3 self-assembling GFP plasmids and the Trans3 fragment lengths. Plasmids synthesised for Trans3 localization and orientation detection are indicated with a (+), and plasmids that were not established (due to preceding results indicating a localisation of the full length protein adjacent to the PPC) are indicated with a (-). **b** Sketch map describing the different Trans3 TMD regions explored using self-assembling GFP.

3.4.4 Bacterial transfection and biolistic transformation

Vectors were introduced into *E.coli* competent Top10 cells by heat shock. Single colonies were picked and cultured overnight, and then plasmids were purified by Midiprep (QIAGEN Plasmid Midi Kit cat.nos. 12143). All plasmids were transformed into *P. tricornutum* cells with the Biolistic PDS-1000/He particle delivery system (Bio-Rad, Hercules, CA, USA), according to the protocol of (Falciatore et al., 1999).

3.4.5 Confocal Microscopy

Images were acquired using an inverted SP8 confocal imaging microscope (Leica Microsystems, Germany). This microscope was equipped with 40x/1.25 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 3608x 3608, 8-line accumulations and pixel development time of 100 ns per frame. Scanning was sequential between frames (Tanaka et al., 2015a).

Mito-Tracker incubated samples were prepared following the protocol as below: 2 µL Mito-Tracker (Molecular Probes™ MitoTracker™ Orange, ThermoFisher) stock was diluted into 2 mL 1x mPBS. Next, 1 mL of mid-exponential phase cells were sampled under sterile conditions, and centrifuged at 15000 g for 2 mins to remove the supernatant. 800 µL 1x mPBS were added to resuspend the cell pellet. 200 mL diluted Mito-tracker were then added into the resuspended sample with the final dilution as 1/ 5000. Samples were incubated on a shaker for 15 mins, followed by centrifugation at 15000 g for 2 mins to remove the supernatant. Stained cells were resuspended in 50 µL mPBS and detected by confocal microscopy.

The Leica SP8 was set up as below for GFP fluorescence with Mito-Tracker: cells were excited with a first laser 475 nm (light intensity: 10% power as default) with detection window at 500-540 nm for GFP fluorescence, and with a econd excitation laser set at 548 nm (light intensity: 10% power as default) with detection window at 575 nm-585 nm for Mito-Tracker fluorescence. These fluorescence spectra are compatible with one another, and with chlorophyll autofluorescence visualised through a 650 nm-700 nm detection window. The GFP and chlorophyll autofluorescence were measured using a photomultiplier detector (PMT). Each detection window was set with 700 mV-900 mV smart gain. Wild type cell lines stained with Mito-Tracker, and Mito-Tracker negative cell lines with GFP signal were used as controls to modify the detection values appropriately. Each laser excitation was also tested sequentially to exclude the possibility of false positive fluorescence signal and minimize the crosstalk between GFP, Mito-Tracker and chlorophyll channels.

3.4.6 RNA extraction, cDNA synthesis and Reverse Transcription PCR tests

Self-assembling GFP cell lines were grown into exponential phase and then centrifuged for 20 min at 4000 rpm to obtain pellets. Cell pellets were washed three times with 1 mL 1× mPBS buffer, aliquoted into 1.5 mL Eppendorf tubes, pelleted for 2 min at 11,000 rpm, flash-frozen in liquid nitrogen and stored at −80°C before proceeding with RNA extraction. Total RNA was extracted from around 50 mg cell pellets using 1 mL of TRIzol reagent (T9424, Sigma-Aldrich) and precipitation by isopropanol. RNA was treated with RQ1 DNase (Promega Corporation: M6101) according to the manufacturers' protocol to remove DNA contamination. RNA integrity and concentration confirmed respectively by electrophoresis on 1% agarose gels and with a Nano-Photometer spectrophotometer (DeNovix, USA).

cDNA was synthesized from 200 ng of each sample, as quantified by nanodrop spectrometry, using a Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher: K1672) with Recombinant RNasin Ribonuclease Inhibitor (Promega Corporation: N2511) following the manufacturer's instructions.

Reverse Transcription PCR was then used to confirm successful transformation and expression of vector sequence fluorescence-negative fluorescence self-assembling GFP cell lines. The cDNA served as a template for exponential amplification using Phusion High-Fidelity DNA Polymerase (ThermoFisher F-530XL). Both Trans3 fragments (Trans3 C-terminus fused with GFP11 part) and reporter genes (reporter genes C-terminus fused with GFP1-10 part) sequences were amplified by PCR and checked by sequencing the purified PCR products (NucleoSpin™ Gel and PCR Clean-up Kit: Macherey-Nagel Nº Cat. 740610.20). Primers for amplification were listed Thesis Annex II, and template-negative and primer-negative controls were performed for each reaction.

Chapter 4: Trans3 knockout mutant line generation and phenotyping by growth, photo-physiological and transcriptome measurements

Abstract

Previously, I have shown that the novel MFS transporter Trans3 is localized to the diatom PPM. In this chapter, I investigate the role of Trans3 in *P. tricornutum* using three presumed heterozygous knockout (henceforth termed KO) mutants generated by CRISPR-Cas9, with multiple attempts to separate mutant alleles by different methods. Quantitative polymerase chain reaction (qPCR) analysis revealed a 30-90% decrease in Trans3 mRNA relative abundances in Trans3 KO lines compared to empty vector controls. Measurements of Trans3 KO versus control line growth rates were performed under iron (Fe) repletion, Fe depletion and different nitrogen (N) source conditions, considering the environmental distribution of Trans3 homologues from *Tara* Oceans data. The partial knockout lines showed an average 20% reduction in exponential phase growth rate in Fe-replete, nitrate (NO_3^-) -supplemented media. In contrast, when grown under Fe depletion stress or in ammonium supplemented media, Trans3 KO and control lines showed equivalent growth rates.

Next, I present the measured photophysiology of Trans3 KO lines. Under Fe and nitrate-replete conditions, Trans3 KO lines showed diminished photosynthetic capacity, particularly considering photosystem I (PSI) P700 content. The defective photophysiology of Trans3 KO lines was dependent on Fe availability, with limited differences in measured P700 content between KO and control lines following three days Fe depletion. Consistent with the growth rate phenotype, control line PSI seemed to be more sensitive to Fe depletion, with a stronger reduction in P700 following three days Fe depletion than that measured in KO lines.

Finally, I performed RNAseq on KO and empty vector control lines, over four progressive Fe depletion conditions defined from photo-physiology and growth measurements. KO lines showed different expression trends across multiple plastid-associated proteins, with greater expression of genes involved in

pigment (chlorophyll and terpenoid) synthesis, photosystem and light-harvesting pathways. I propose that the greater expression of these genes in KO lines may relate to lower photosynthetic efficiency, or efficiency of use of photosynthetic metabolites, compared to controls. Comparatively limited impacts were observed on the expression of genes encoding mitochondria-targeted proteins in KO lines, except increased relative expression levels of the arginine/urea biosynthesis pathway detected in KO lines under long-time Fe depletion. The RNA seq data further indicated potential interaction of Trans3 KO and Fe depletion status on LHC and photosystem gene expression.

Overall, these findings suggest that Trans3 is essential for cell growth and photosystem efficiency in diatoms. The function of Trans3 is presumed to interact with cellular Fe status and require the presence of both NO_3^- and Fe in the growth media. Knockout lines may compensate for lower light harvesting efficiency by upregulating photosynthetic and primary plastid metabolism pathways, which may mitigate the negative influence caused by Trans3 KO.

4.1 Introduction

4.1.1 The use of diatom knockout mutants for protein functional investigation

Although the physiological functions of plastids depend on the transport of multiple ion and metabolites across plastid membranes, the specific functions of most algal plastid transporters remain unidentified (Marchand et al., 2018). In the previous chapter, a novel MFS family plastid transporter, Trans3, was confirmed to localize to PPM, and here its potential functions are explored by the phenotyping of partial KO mutant lines.

The phenotypic characterization of algal mutant lines, particularly if coupled with mutant complementation, allows for the detailed investigation of protein function and physiology (Bailleul et al., 2015; Murik et al., 2019; Yang et al., 2016; Haimovich-Dayan et al., 2013; Ma et al., 2014). Several collections of algal mutants are available, such as the *Chlamydomonas* Resource Center (http://chlamycollection.org/) and the *P. tricornutum* mutant collection and genetic resources at the French European Marine Biological Resource Centre

(https://www.embrc-france.fr/fr).

The CRISPR/Cas9 system has emerged as a powerful tool for genome editing and can be employed to generate knockout and knockdown mutants for algal genes (Nymark et al., 2017). For instance, a KO line of the H^+/K^+ antiporter KEA3 was generated in the model diatom *P. tricornutum*, which has been identified as a key regulator affecting photosynthetic electron transport and proton motive force (PET/PMF) coupling, as well as non-photochemical quenching (NPQ) responses (Seydoux et al., 2022). Using CRISPR/Cas9 mediated-knockout and overexpression analyses, a trypsin in *P. tricornutum* (PtTryp2) was revealed as a coordinate regulator of N:P stoichiometric homeostasis (You et al., 2022). Bai et al. generated violaxanthin de-epoxidase (VDL2) and zeaxanthin epoxidase-like (ZEP1) CRISPR/Cas9 mediated-KO, indicating their essential roles for fucoxanthin formation in diatoms (Bai et al., 2022). Finally, through ptEnolase and ptPGAM1A CRISPR-Cas9 KO generation and complementation, alongside with integrative 'omic analyses, Dorrell et al. proposed that diatoms may use a plastidial lower half glycolytic-gluconeogenic pathway to redirect excess plastidial glyceraldehyde-3-phosphate to diverse fates via the pyruvate hub, potentially optimizing cellular carbon management (Dorrell et al., 2022).

Beyond CRISPR/Cas9, a type of engineered DNA-cutting enzymes transcription activator-like effector nucleases (TALENs)-mediated method has also been used for diatom gene editing and KO generation (Weyman et al., 2015; Serif et al., 2017). With the usage of TALEN-mediated KO and mutant complementation, Buck et al. investigated the important roles of *P. tricornutum* light-harvesting genes Lhcx1 and Lhcx2 in dissipation of excess absorbed energy (Buck et al., 2019). TALEN-generated knockout lines were utilized to explore the significance of Fe acquisition-related genes ISIP2 (McQuaid et al., 2018). Notably, the deletion of ISIP2 was found to disrupt high-affinity iron uptake in *P. tricornutum*, and restoration was achieved through mutant complementation with human transferrin (McQuaid et al., 2018).

4.1.2 The roles of physiolog ical phenotyping and transcriptomics for identifying the functions of novel diatom proteins

To investigate the role of novel protein functions in diatoms, a combination of comprehensive physiological measurements and ‘omic studies can be employed.

Mutant growth rate phenotyping serves as a starting point and fundamental strategy that provides insights into the physiological significance of specific genes of interest (Görlich et al., 2019). The model diatom *P. tricornutum* is a single-celled species and can be easily grown in lab conditions (Thomas et al., 1984), and its growth rate can be calculated from cell densities measured using e.g., a flow cytometer (FCM). However, as cell growth is a complex outcome influenced by numerous interconnected fluxes, a more detailed and specific comprehension necessitates the measurement of additional parameters such as photosynthesis and primary metabolism.

Mutations in some plastid transporters have been found to significantly impact on microalgal photosynthesis (Seydoux et al., 2022). Considering transporters are vital components of the ion and metabolite exchange pathway across plastid membranes (Marchand et al., 2018), it is likely that they occupy key positions that affect the efficiency of photosynthesis. Hence, photosynthetic parameter measurements can provide important insights into the functions of plastid-located transporters. The process of absorption, transfer and conversion of light energy in algae primarily occurs on the thylakoid membrane, which comprises of four macro-complexes, including Photosystem II (PSII), cytochrome b6f (Cyt_{b6f}), Photosystem I (PSI), and ATP synthase (Nelson et al., 2015). Various chlorophyll fluorescence parameters can be used to measure the efficiency of photosynthetic photosystem. For PSII, Fv/Fm measures the maximum photochemical efficiency of PSII, while ΦII determines the actual photosynthetic quantum yield of PSII under light. NPQ is the non-photochemical quenching coefficient, and ETR(II) indicates the electron transfer rate of PSII (Nelson et al., 2015). These parameters can help us better understand the efficiency of photosystem PSII. Photosystem I (PSI) is another key protein complex in the process of light-dependent reactions, which plays a critical role in electron transfer during photosynthesis (Nelson et al., 2015). The reaction center of PSI is composed of chlorophyll and binding proteins, named P700 due to its absorption peak at 700 nm. P700 is in a reduced state when it gains electrons and in an oxidized state when it loses electrons (Nelson et al., 2015). There is slight difference in the light absorption of P700 between two states, which can be detected as a signal by a highly sensitive P700 measuring instrument with high sample concentrations.

To further explore the molecular mechanisms underlying the mutant phenotypes in physiological measurements, additional 'omic studies, especially transcriptomic analyses of RNAseq data, can be carried out as a complement. Transcriptome sequencing provides a global view of gene expression patterns and regulatory networks. For example, Ait-Mohamed et al. (2020) used large scale transcriptome datasets assembled from multiple previous studies to identify co-expressed modules of genes within the *Phaeodactylum* genome, elucidating probable metabolic links between the diatom mitochondria and chloroplast. Beyond this, differentially expressed genes (DEGs) associated with specific biological processes can be identified by comparing the transcriptomes of mutants and control lines, which allow researchers to gain insights into the transcriptional changes associated with the manipulated genes and their potential roles in cellular processes. For instance, by DEG pathway analysis of nitrate reductase KO and control lines, McCarthy et al. (2017) found nitrate reductase activity disruption can uncouple nitrate transport from assimilation and alters the partitioning of carbon flux in diatoms. Finally, RNAseq analysis can also give insights into gene expression patterns in wild-type or control cell lines by comparing specific physiological stresses to optimal growth conditions, e.g., phosphorus (Cruz de Carvalho et al., 2016), iron (Smith et al., 2016) and nitrogen (McCarthy et al., 2017) limitation.

4.1.3 Iron and nitrogen: two important ion factors for diatom cell growth and photosynthesis

Iron (Fe) is an essential cofactor for the electron transfer chain, catalytic processes such as chlorophyll biosynthesis and protein import, and is mostly utilised in the plastid (Marchand et al., 2018). Many transporters associated with the plastid, including Trans3, have a strong negative correlation with Fe abundance according to the *Tara* Oceans dataset (Chapter 2, Liu et al., 2022). Fe is the important limiting element in high nitrate, low chlorophyll regions (HNLC) such as the South Ocean and sub-polar Arctic where diatoms form the most abundant component of marine phytoplankton (Brownlee et al., 2022), and understanding the roles of plastid-targeted proteins upregulated under Fe stress conditions is thus particularly important for understanding diatom ecological success.

Nitrogen (N) is another critical limiting factor for phytoplankton growth and photosynthesis (Brownlee et al., 2022), providing the reduced amino groups necessary for the synthesis of nucleic acids, proteins (e.g., Rubisco, light harvesting complexes) and tetrapyrrole pigments (Turpin, 1991; McCarthy et al., 2017; Nunes-Nesi et al., 2010). Diatoms have a superior competitive ability for dissolved nitrate compared to other algal groups, with a faster response to nitrogen enrichment than other algal groups (Lampe et al., 2021). Diatoms can further adapt and utilize different nitrogen sources such as inorganic NO_3^-, ammonium (NH_4^+) and organic nitrogen (N) compounds, as reflected by the diverse range of N transporters encoded in diatom genomes (McCarthy et al., 2017; Brownlee et al., 2022).

Overall, the combination of mutant physiological measurements and transcriptomic analysis provide powerful approaches to determine protein functions. In this chapter, I endeavored to create knockout mutant lines of Trans3 and examined their phenotypic differences to empty vector controls. Growth rates were measured under different Fe depletion and nitrogen source conditions, in view of the roles of Fe and N as master regulators of phytoplankton growth and photosynthetic activity (Gao et al., 2021; Brownlee et al., 2022). These measurements were complemented through comparative photophysiological and transcriptomic analyses of mutant and empty vector control cell lines over progressive Fe starvation. Considering these data, I propose that Trans3 plays important roles in diatom cell growth and plastid metabolism, with diminished growth rate, photosynthetic efficiency and compensatory transcriptomic changes observed in mutants. These functions depend on Fe availability, and may have further interactions with NO_3^-.

4.2 Results

4.2.1 Trans3 knockout mutant genotype and growth rate measurement

Generation of three presumed heterozygous KO lines by CRISPR-Cas9

With the goal of investigating the function of the novel plastid transporter Trans3, CRISPR-Cas9 and biolistic bombardment transformation techniques were used to generate potential Trans3 knockout (KO) mutant lines. Six CRISPR regions were designed specifically to target the exon of the Trans3 gene in *P.*

tricornutum as shown in Figure 4.1.

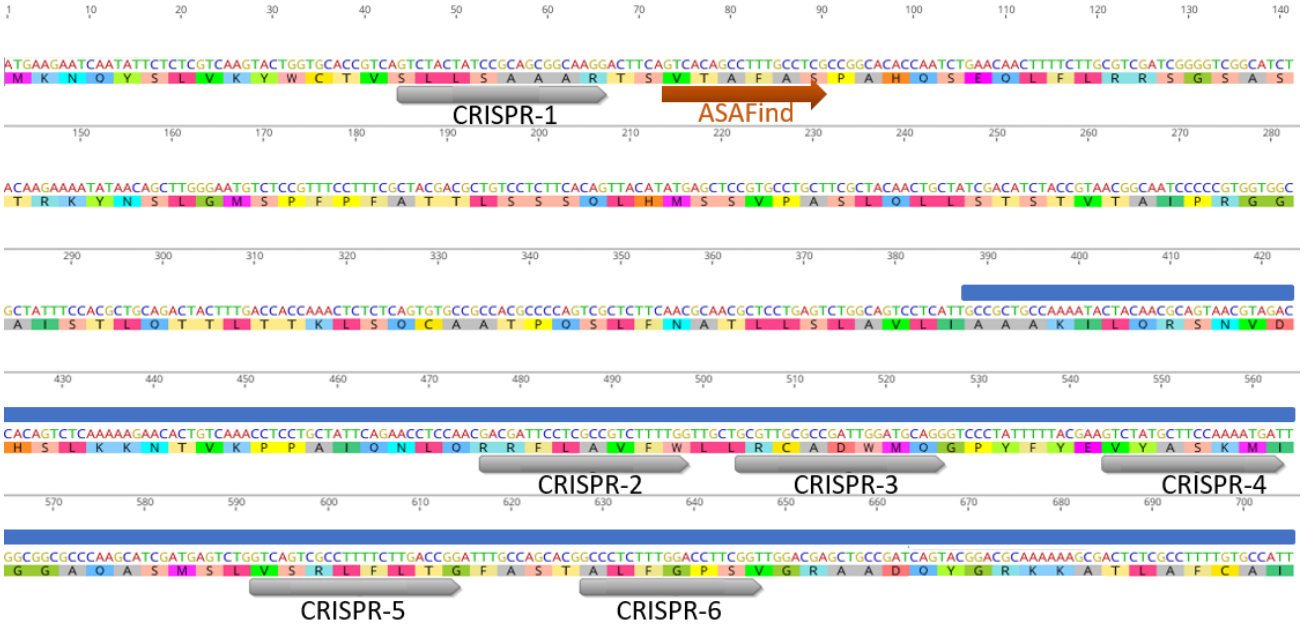

Figure 4.1 CRISPR regions designed in the Trans3 cDNA sequence (partial). Orange arrow represented "ASAFAP" sequence motif in bipartite topogenic signal (BTS) by ASAFind. CRISPR region (1-6) were designed phytoCRISPEX: https://www.phytocrispex.bio.ens.psl.eu/CRISP-Ex/, and were shown by gray color arrows. Blue line represented MFS domain found in Trans3 protein.

Out of the six CRISPR regions (1-6) designed to target the Trans3 gene, only CRISPR region 1 and CRISPR region 4 gave rise to potential KO lines. The first line with an apparent CRISPR mutation (against CRISPR region 4), henceforth termed KO1, was generated by Tomomi Nonoyama, a previous PhD student of the Bowler group (Nonoyama et al., 2019). Using colony PCR and PCR result sequencing, I eventually obtained other six potential Trans3 mutant lines from CRISPR region 1 (CRISPR1-a to f) (Figure 4.2). The Sanger sequences obtained from these lines by colony PCR showed a reduction in peak quality across the CRISPR region, consistent with heterozygous or mixed wild-type/mutant genotypes, as illustrated in the Figure 4.2. Unfortunately, no clean KO results were observed across the hundreds of colonies (>200) obtained from bombardment. The only clean KO line obtained displayed a six bp gap, which would not cause a frameshift mutation of Trans3 and is thus unlikely to disrupt its function.

Figure 4.2 CRISPR region 1 potential Trans3 KO lines colony PCR sequencing results. Potential KO mutant lines exhibited colony PCR results as overlap peaks following the CRISPR region. CRISPR1-c and CRISPR1-f showed the most obvious gene gaps in CRISPR region.

Next, the six potential KO cell lines (CRISPR1-a to f) (Figure 4.2) were separated using various methods, to potentially purify homozygous knockout (KO) cells from mixed colonies containing wild-type (WT) and KO cells. Particular effort was made to separate pure lines of CRISPR1-c and CRISPR1-f that showed obvious CRISPR region 1 gene gaps from the crude colony PCR alignment (Figure 4.2). Separations were attempted by streaking (serial dilution) and transfer with plating beads on solid plate media, and single-cell dilution of liquid cultures on microtiter plates.

Using these separation methods, I finally separated two further presumed heterozygous knockout lines (termed KO2 and KO3) that isolated from original CRISPR1-c and CRISPR1-f mutant lines, as in both cases no exclusively WT sequence colonies were observed through separation of these lines, suggesting that they consist uniformly of mutants rather than a mixture of segregated mutant and WT cells. Random PCR was used to validate the individual allele sequences of each KO line. The results of the random PCR showed that the

percentage of KO and WT genotypes was approximately 50%:50% (genotypes were tested though 20 randomly selected positive colonies and repeated three times). Each cell line exhibited only one type of KO genotype: KO1 (the old one generated by Tomomi Nonoyama from CRISPR region 4) displayed a one bp insert in CRISPR region 4, while KO2 and KO3 exhibited 23 bp and 19 bp deletions in CRISPR region 1, respectively (Figure 4.3). Based on these results, the three KO lines of Trans3 (KO1, KO2 and KO3) were presumed to be heterozygous KO lines, rather than a mixture of WT and KO cells, and were finally selected for further phenotyping. The presence of gaps in CRISPR region 1 (KO2 and KO3) leads to frameshift mutations starting from the BTS N-terminus, while gaps in CRISPR region 4 (KO1) result in frameshift mutations occurring within the middle of the MFS domain. These frameshift mutations are likely to prevent the translation of a functional Trans3 sequence.

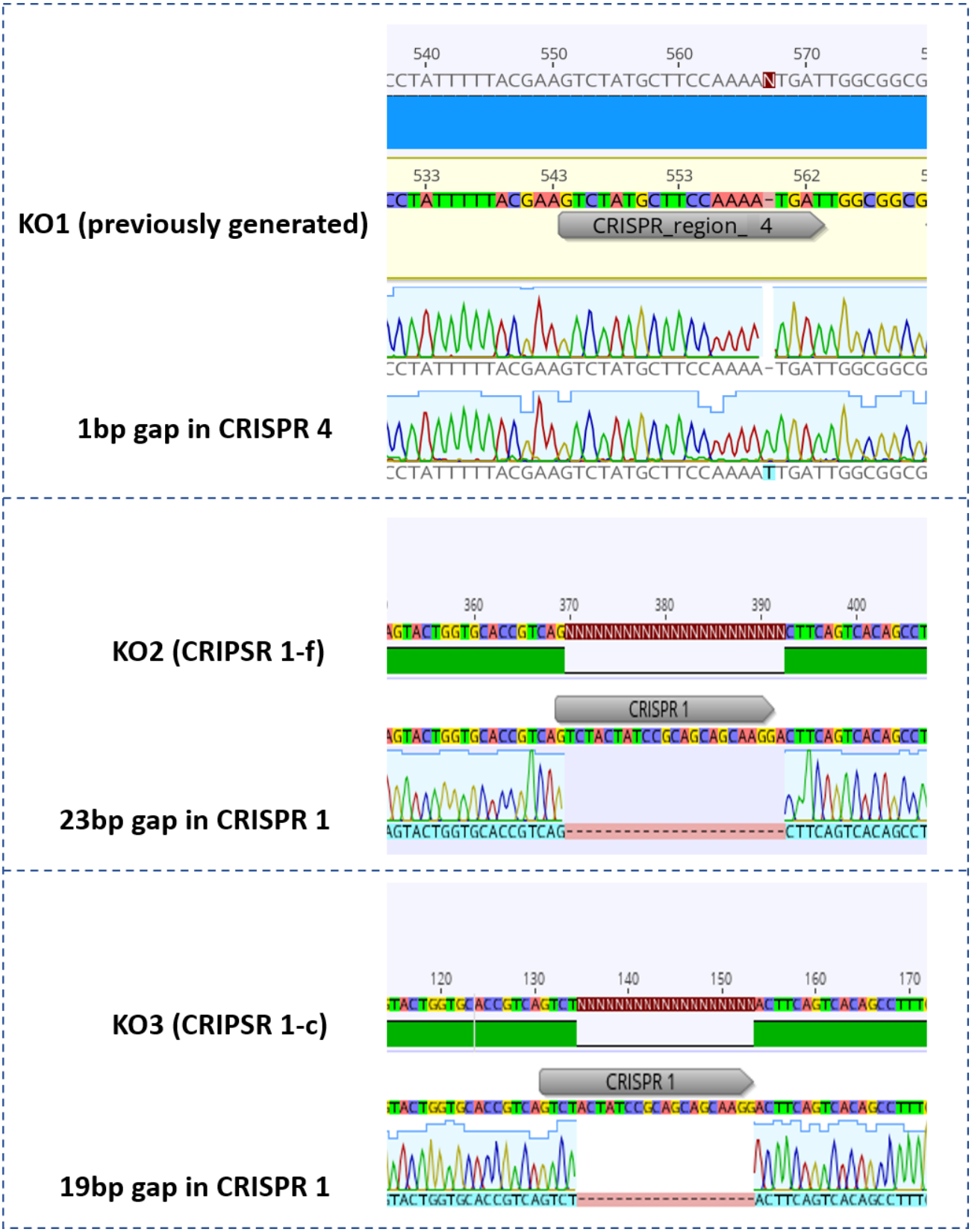

Figure 4.3 Genotypes of heterozygous KO lines as determined via random PCR sequencing

qPCR validation of Trans3 expression in KO lines

To determine whether the Trans3 heterozygous KO lines were capable of influencing Trans3 expression and function, I measured the relative mRNA abundances of Trans3 in KO lines (KO1, KO2, and KO3) compared to empty vector transformed WT cell lines (termed HA1 and HA2) using qPCR under normal (Fe replete, NO_3-supplemented) ESAW and 19℃, light: dark cycling cultivation conditions 12:12. Ribosomal protein S1 RPS (protein ID 44451) and TATA-box binding protein TBP (protein ID: 10199) were selected as endogenous control genes for normalization (Siaut et al., 2007, Sachse et al., 2013). The qPCR results showed that all three KO lines exhibited lower relative Trans3 mRNA abundances when compared to the empty vector controls (Figure 4.4). Among them, KO1, which had only one inserted nucleotide in CRISPR region 4, exhibited the smallest decrease of approximately 30% compared to the control lines (Figure 4.4). KO2 and KO3, which exhibited long deletions in CRISPR region 1, showed more substantial decreases (70-90%) (Figure 4.4). These qPCR results suggest that Trans3 expression was suppressed to different extents in the partial KO lines, but they nonetheless have potential for further phenotyping. Considering our potential heterozygous Trans3 KO lines were generated by CRISPR-Cas9, not RNAi or antisense RNA, we prefer continue to use abbreviation KO to present the mutants we generated for Trans3 study, although the functional of these mutants were more likely to knockdown based on qPCR results.

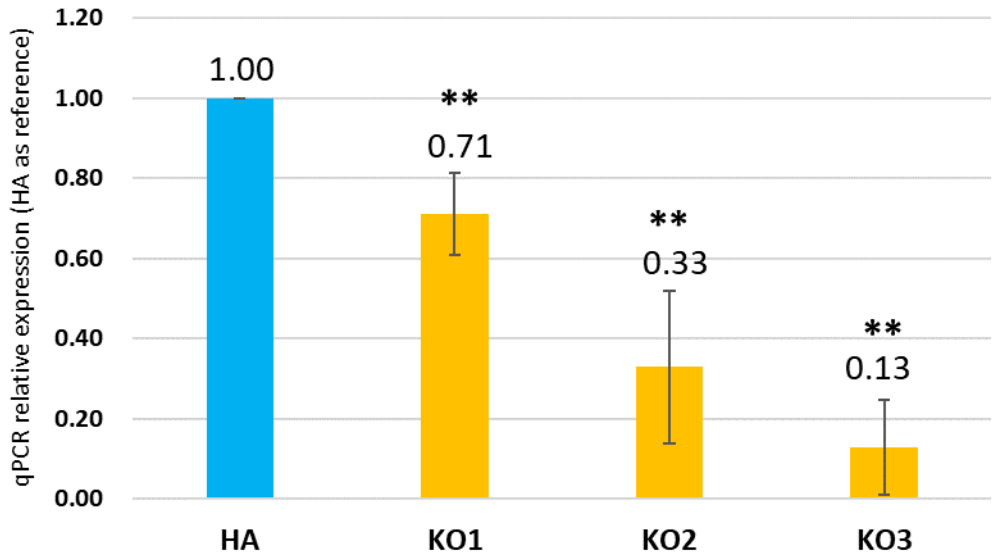

Figure 4.4 Relative expression levels of Trans3 in partial KO lines inferred by qRT-PCR.

empty vector control lines HA (HA1, HA2) generated by co-transformation of *P. tricornutum* cells with HA-Cas9 and pPhat vectors without guide RNA (Stukenberg et al., 2018) are used as a normalization reference. Relative expression fold changes compared with HA lines shown by the numbers (HA = 1.0). ** represents significant differences ($P < 0.01$, one-way ANOVA), with standard deviation values shown as error bars. Two Trans3 qPCR regions were used for each sample, and RPS (ribosomal protein S1) and TBP (TATA binding protein) were used as the control genes.

Growth rate measurements under different iron and nitrogen conditions

First, mid-exponential phase growth rates were calculated for Trans3 KO and empty vector control lines in nutrient replete ESAW, under a constant 19℃ and 100 rpm shaking environment, with 12h 50 uE light: 12h dark cycling. Growth curves were established with a starting cell density of 10,000 cells/mL. Growth rates were measured in exponential phase cell lines with cell density 100,000 – 2,000,000 cells/mL (Figure 4.5) to avoid effects caused by cell shading in excessive dense cell density. The successive time points used for growth rate calculation showed clear linear correlations between the cell density Log_2 value and time ($r > 0.95$), confirming the cells were in the exponential phase when the measurements were taken. A minimum of 4 successive time points were used for each growth rate calculation.

Considering cell lines growth rates from a minimum of 3 biological replicates per line measured 3 times, Trans3 KO exhibited a significant decrease in relative growth rate (average: 20%) compared to empty control HA lines under control conditions (Figure 4.6). The total exponential phase data usage for growth rate calculation was also tested (6 or 7 successive time points in exponential phase data) to assess its impact on the comparison of growth rates between KO and control lines. However, the results confirmed that the growth rate of KO lines still exhibited a 20% reduction compared to the control lines, indicating that the observed trend remained unchanged.

Next, taking into account the important role of iron in diatom growth, photosynthesis and plastid primarily metabolism, the growth rate of Trans3 KO under control (Fe+) and Fe limitation (Fe-) conditions. Fe-free media preparation details are described in the methods. To eliminate the influence of Fe storage in *P. tricornutum* cells, all cell lines were subjected to Fe- treatment for at least one

week before the measurement of growth curve. When the Fe source was removed from the media (Fe-), all cell lines showed significant reduced relative growth rates in Fe- conditions compared to replete conditions, and empty vector controls showed a greater decrease (40%, HA_Fe- vs HA_Fe+) than KO lines (25% decrease, KO_Fe- vs KO_Fe+). No significant differences were observed between KO and control lines in the Fe- condition (Figure 4.6), suggesting a suppression of the growth defect observed in replete media.

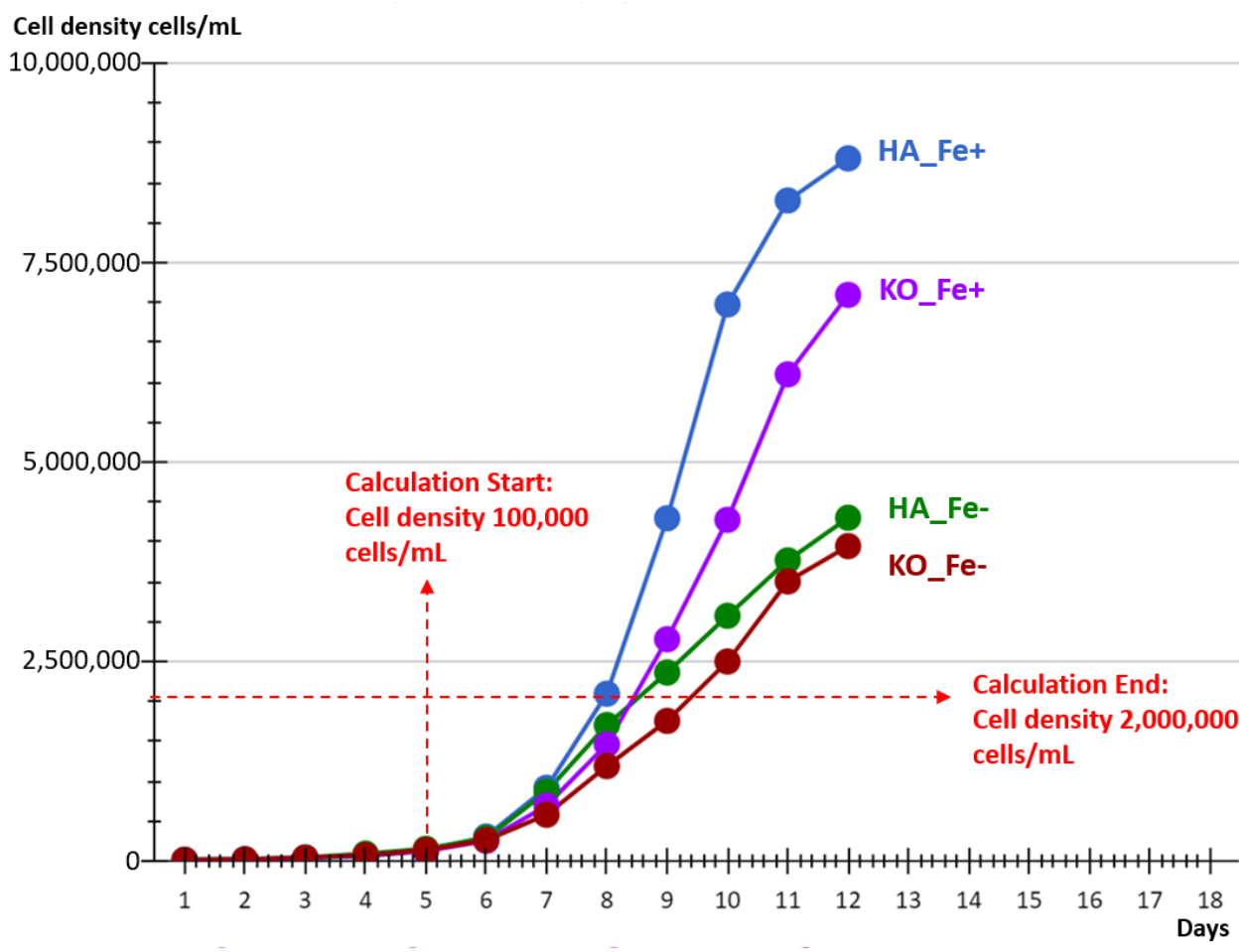

Figure 4.5 An exemplar growth curve performed with KO and control lines under Fe+ and Fe- condition. Cell lines (control: average value of HA1 and HA2 empty vector lines; KO: average value of KO1, KO2 and KO3) growth curves are represented by different colors. Growth rates were measured in exponential phase cell lines with cell density 100,000 – 2,000,000 cells/mL as measured by FCM with more than 4 points (days) to do the calculation, with the aim of avoiding the effect caused by cell shading due to excessive cell density.

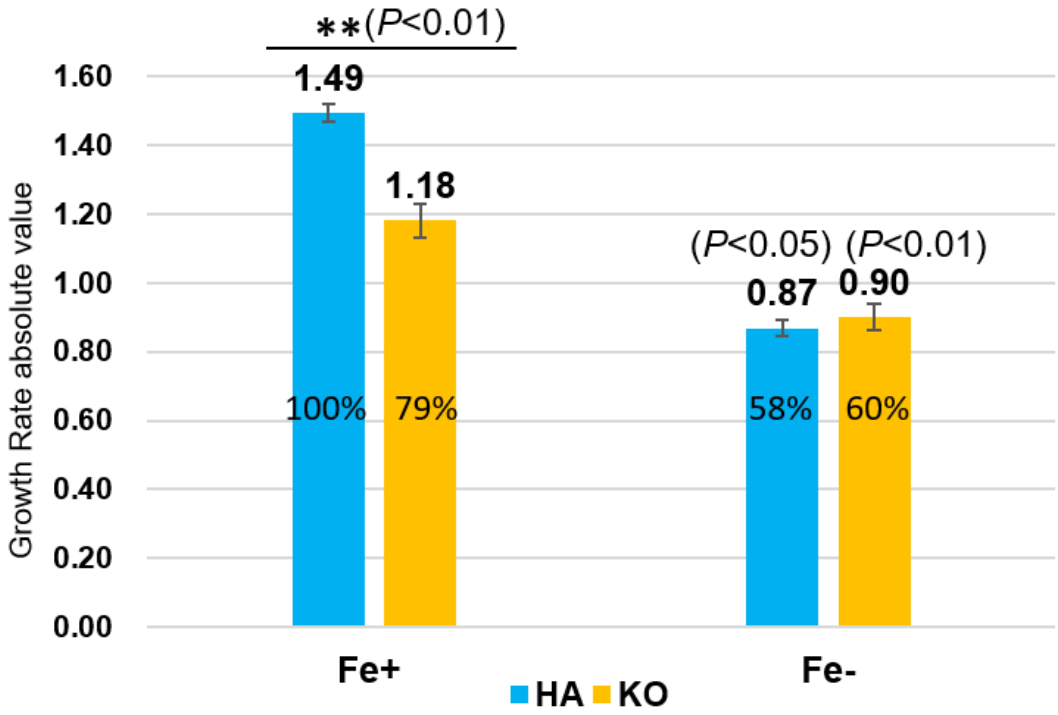

Figure 4.6 Growth rate of KO and control lines under Fe+ and Fe- condition. This histogram is generated based on three growth curve measurements, with at least two biological repetitions for each cell line (HA: HA1 and HA2, KO: KO1, KO2 and KO3) and three technical replicates as described in Figure 4.5. The error bars represent the relative growth rate percentage standard deviation between three times growth curve measurement. ** indicates significant differences (P < 0.01, one-way ANOVA) between HA and KO lines. *P values* calculation between HA_{Fe-} and HA_{Fe+}; KO_{Fe-} and KO_{Fe+} were also shown in the figure. % values (relative growth rate expressed as a function of HA_{Fe+}) are shown above each bar.

Growth rate phenotypes of KO lines under different nitrogen sources and concentrations

Next, the impacts of nitrogen (N) source on Trans3 KO line growth rates were investigated. The KO lines still showed an average 20% decrease in growth rate compared to empty vector controls under 0.4× ESAW NO_3^- concentrations (2.20×10^{-4} moles/L) in the media, which was almost the same as the replete condition (Figure 4.7). It should be noted that the nitrate concentration under this condition is still well in excess of that of seawater (2.82×10^{-5} moles/L-2.15×10^{-8} moles/L) (Pierella Karlusich et al., 2022), and while it may constitute nitrogen limitation is unlikely to constitute starvation. When the medium NO_3^- concentration was reduced to 0.2× in ESAW (1.10×10^{-4} moles/L), although the

concentration still higher than nitrogen concentration in seawater, empty vector control lines (HA) had a measured growth rate of 1.30 divisions/day, i.e. a 13% reduction of that observed in control lines under 1× NO_3^- (1.49 divisions/ day), while the mean KO line growth rate was 1.11 divisions/ day, showed a 5% reduction comparing to its own growth rate under 1× NO_3^- (1.18 divisions/day) (Figure 4.7). The KO lines still showed a slightly slower growth rate than controls under 0.4× /0.2× NO_3^- treatments, but the differences between them were not significant (Figure 4.7).

Finally, when the nitrogen source was changed from NO_3^- to NH_4^+ (5.49 x 10^{-4} M) in the media, all tested lines exhibited a lower growth rate of around 1.0 divisions per day. This reflects a 33% decrease for control and a 13% decrease for KO lines compared to the 1× NO_3^- control condition. No observable difference in growth rate however was noted between KO and control lines (Figure 4.7), suggesting a probable suppression of the mutant growth phenotype.

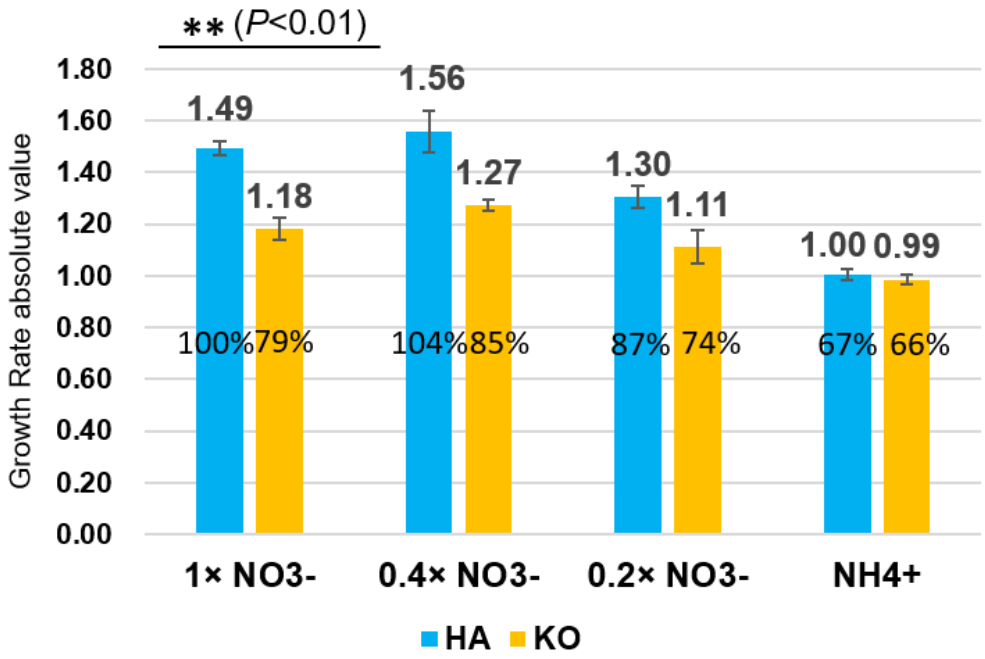

Figure 4.7 Growth rates of KO and control lines under different N conditions. Growth rates were measured with three biological repetitions with three technical replicates for each cell line (HA1, HA2, KO1, KO2 and KO3), with three technical replicates. Error bars represent the relative growth rate standard deviation of cell line duplicates. ** indicates significant differences (P < 0.01, one-way ANOVA) between HA and KO lines. % values of growth rate compared to HA $_{1\times\ NO3-}$ are provided above each bar.

4.2.2 Photo-physiological phenotypes of Trans3 KO lines

As the Trans3 KO genotype showed the greatest difference and interaction with Fe in growth rate measurements, I aimed to determine the photosynthetic response of the Trans3 KO line to Fe starvation. Photophysiological measurements were performed in collaboration with Giovanni Finazzi and Mattia Storti at the CEA Grenoble, Cell & Plant Physiology Laboratory (https://www.lpcv.fr/en/Pages/Giovanni-Finazzi.aspx), considering indicators associated with both photosystems I (PSI) and photosystems II (PSII) in KO and control lines, in Fe-replete conditions, and short-time Fe limitation (after 3 days Fe- treatment).

Fv/Fm and ETR in Fe depletion suggest limited impacts of Trans3 on PSII

First, the photosynthetic activity of PSII was evaluated in KO and control lines using Fv/Fm and Electron Transport Rate (ETR) measurements. In nutrient-replete ESAW conditions, KO Fv/Fm (average: 0.62) was slightly lower than control line HA (average: 0.63), (Figure 4.8 a). Under short-time (three days) Fe depletion, Fv/Fm was diminished in both KO (average: 0.55) and control lines (average: 0.54), and the relative Fv/Fm value is 0.85 (15% reduction) for HA control lines and 0.90 (10% reduction) for KO lines. In all comparisons no significant difference was observed between KO and control lines (Figure 4.8 a).

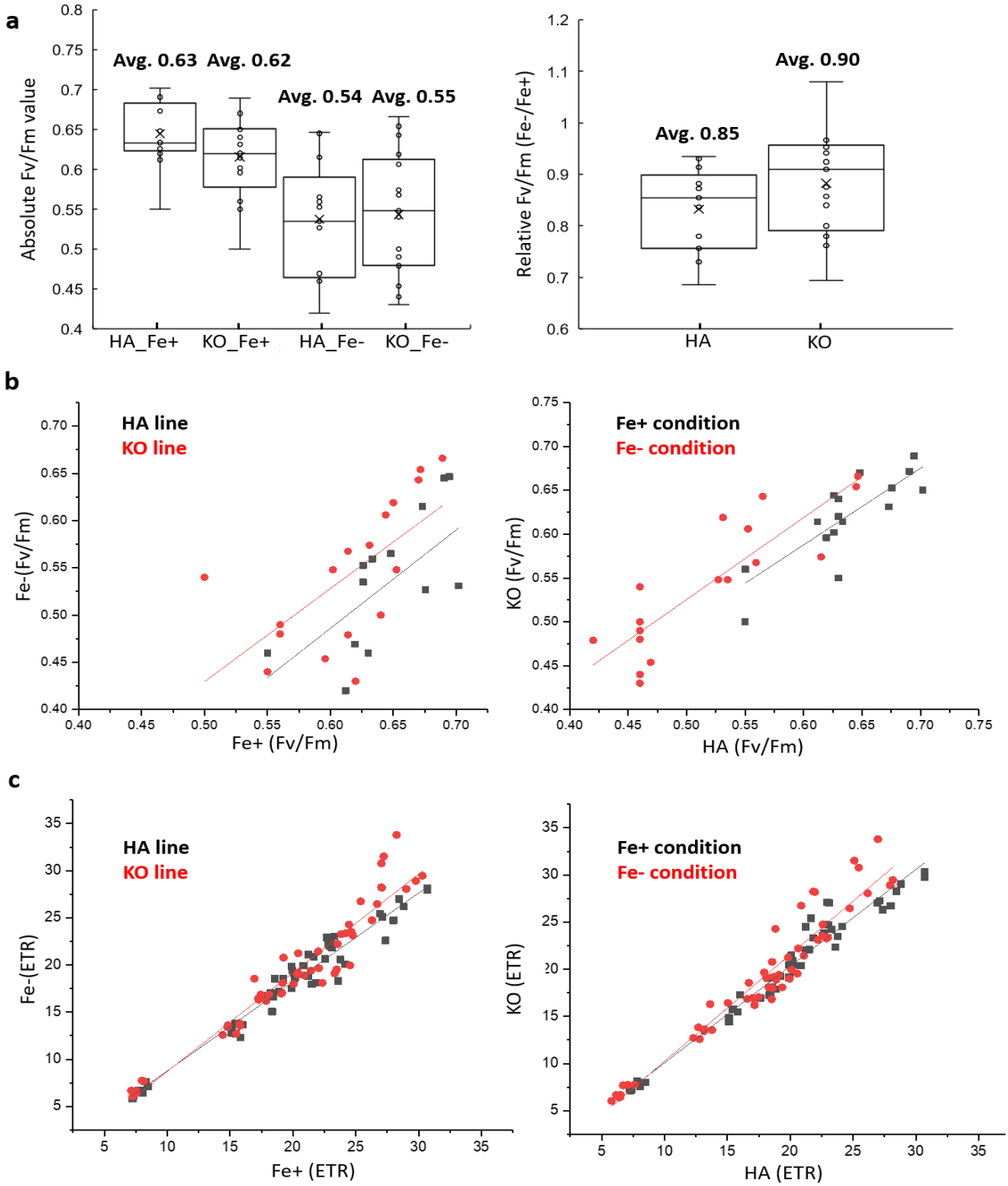

Figure 4.8 KO and HA PSII measurements of Fv/Fm and ETR. a, boxplots of absolute Fv/Fm values of cell lines, shown on the left, and the relative Fv/Fm content (Fe-/Fe+), shown on the right. Cell lines are HA: HA1 and HA2, and KO: KO1, KO2 and KO3. Significant differences ($P < 0.05$, one-way ANOVA) between KO and control (HA) lines are represented by *. **b,** scatterplots showing the linear relations of Fv/Fm absolute values. Fe+ and Fe- values across different genotypes are shown on the left; and HA and KO Fv/Fm values across different conditions are shown on the right. **c,** scatterplots of ETR (electron transport rate) of algae adapted for 5 min at different light intensities (25, 58, 105, 178, 315, 540 µmol photons m^{-2} s^{-1}). Fe+ Fe- ETR absolute values in different genotypes are shown on the left; and HA and KO ETR absolute values in different conditions are shown on the right.

Next, to test if there was a combined impact of KO and Fe depletion on PSII, graphs (Figure 4.8 b) of linear correlation coefficients were generated for Fv/Fm values either between genotypes or between treatments. Briefly, linear relationships were plotted between Fv/Fm values for each genotype in three days Fe-depletion (horizontal axis) and Fe-replete (vertical axis) conditions, and the slope of the linear regression between treatments was calculated for each genotype (Figure 4.8 b, left). Similarly, a linear relationship was plotted for Fv/Fm values for control (horizontal) and KO lines (vertical) under both Fe-replete and Fe-depletion conditions, and the slope of the linear regression between genotypes was calculated for each treatment (Figure 4.8 b, right).

The slopes of the linear correlation either between two cell lines (HA control and KO mutants) or between two conditions (Fe- and Fe+) were almost the same (Figure 4.8 b), indicating that knockout of the Trans3 gene resulted in a slight negative influence on PSII Fv/Fm, but with no obvious interaction with Fe depletion (Figure 4.8 b). ETR is indicative of photosynthetic rate and can be calculated from PSII relative yield (ΦPSII) as function of a given light intensity (see Methods). No significant difference in relative electron transport ETR between control and KO lines was observed under either Fe- or Fe+ conditions (Figure 4.8 c), confirming limited impacts of Trans3 KO on PSII. Longer durations of Fe starvation showed essentially the same of Fv/Fm values as those observed after 3 days, and are not presented for simplicity.

PSI content measurements show distinct responses of Trans3 KO lines to Fe depletion

In addition to the Fv/Fm and ETR measurements detailed above, PSI photosynthetic content (P700) was also tested in KO and control lines (HA) under Fe-replete and Fe-starvation conditions. PSI reaction center (P700) absorption was measured in presence of the PSII inhibitor (DCMU) allowing light dependent oxidation estimation of PSI content. Under Fe-replete conditions, KO lines showed a lower absolute P700 content (789/million cells) than control lines (944/million cells). In contrast, after three days Fe depletion, both the relative and absolute P700 content of KO lines was significantly (P<0.01) higher than that of control lines (Figure 4.9 a). These data together imply that Trans3 KO show limited differences in PSI P700 under Fe depletion and Fe-replete conditions, contrasting with control lines in which Fe depletion provokes a substantial decrease in P700 content (Figure 4.9 a). Over longer times (> 7 days) Fe depletion, P700 content in both KO and control lines decreased to very low levels that were difficult to measure (data not shown).

Next, to examine if Fe stress had different impacts on KO versus control line P700 content, linear regressions of P700 across Fe condition and genotype were calculated as per the PSII Fv/Fm and ETR measurements (Figure 4.9 b). The slope calculated for the linear regression between Fe-replete and 3-day Fe depleted PSI P700 in control lines (0.5±0.15) revealed a substantial response to Fe depletion. The slope calculated between Fe-replete and 3-days Fe depleted PSI P700 in KO lines (Figure 4.9 b, left) was nearly 1.07±0.13, suggesting a limited sensitivity to Fe depletion. Similarly, a regression of the measured PSI P700 between KO and control lines under Fe-replete conditions had a slope value of 0.87±0.19, suggesting limited differences in P700 content, whereas a regression of the measured PSI P700 between KO and control lines after 3-days Fe-depletion conditions had a slope value of 1.49±0.19, suggesting Trans3 KO genotype had different influence in P700 content under 3-days Fe- and Fe+ treatment (Figure 4.9 b, right). These linear regressions indicated that KO genotype and Fe supply had potential interactions in PSI P700 content.

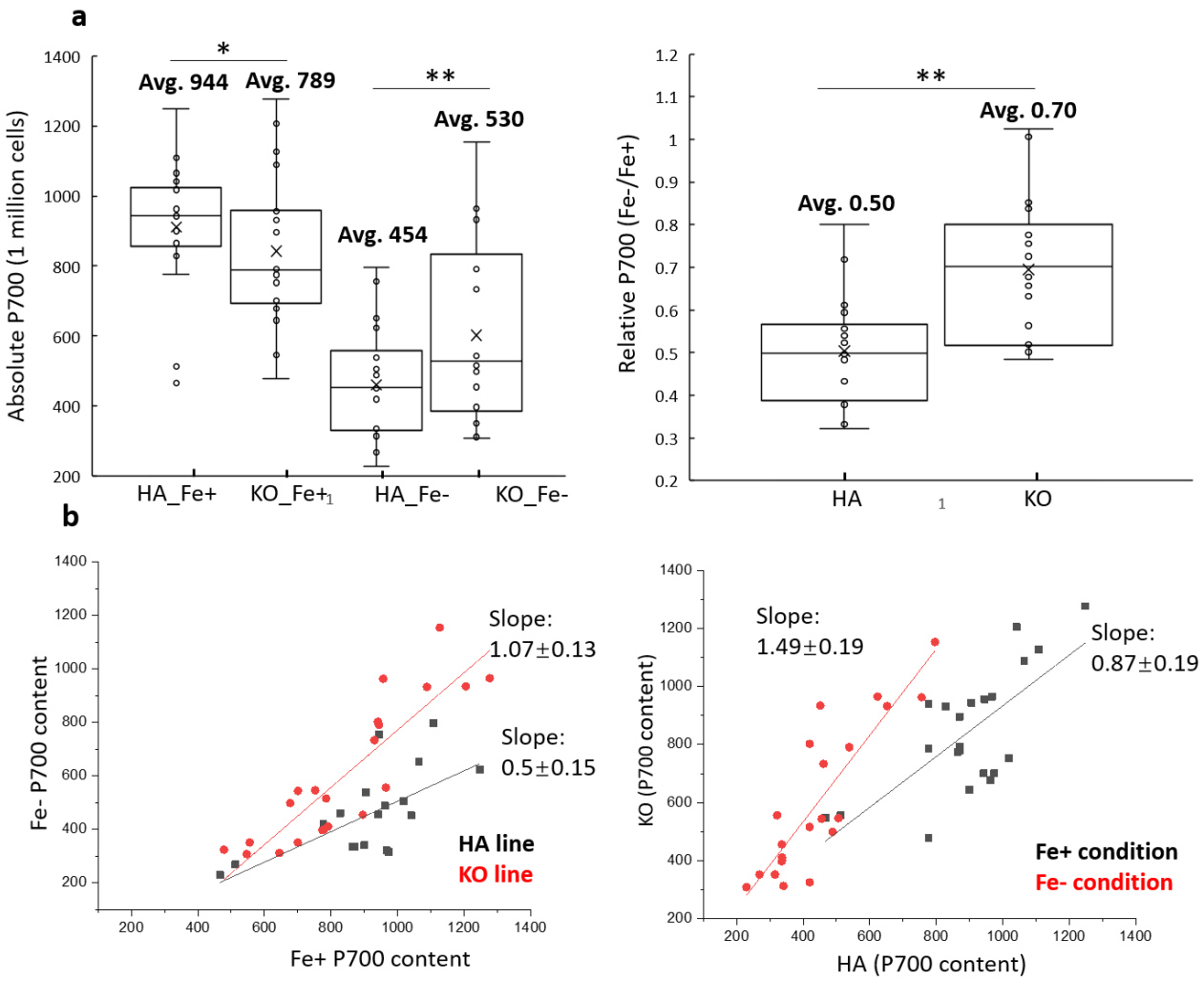

Figure 4.9 Trans3 KO and control P700 content comparisons under Fe-replete and Fe-deplete (3 days) conditions. a, boxplots of absolute P700 content values of cell lines shown on the left, and relative P700 content (comparing Fe- and Fe+ conditions) shown on the right. Cell lines are denoted as HA: HA1 and HA2, KO: KO1, KO2 and KO3. Significant differences between KO and control lines are represented by * ($P < 0.05$, one -way ANOVA) and ** ($P < 0.01$, one -way ANOVA). **b**, Scatterplots showing linear regressions of Fe+ and Fe- P700 absolute values for different genotypes (left); and for control and KO P700 absolute values for different conditions are shown (right).

4.2.3 RNAseq analysis under different Fe limitation conditions

Next, according to the observed phenotypes of photosynthesis and growth rates of Trans3 KO lines, I prepared RNA sequencing libraries on two empty vector transformed WT lines (HA1 and HA2) and three Trans3 KO lines (KO1, KO2, and KO3) under Fe replete conditions (Fe+) and three different lengths of Fe depletion: three days as the short-time Fe- treatment (FeS); one week as the medium time Fe- treatment (FeM); and two weeks as the long-time Fe- treatment (FeL) (Figure 4.10). The durations of Fe starvation were selected according to photophysiological (FeS) and growth measurements (FeM, FeL).

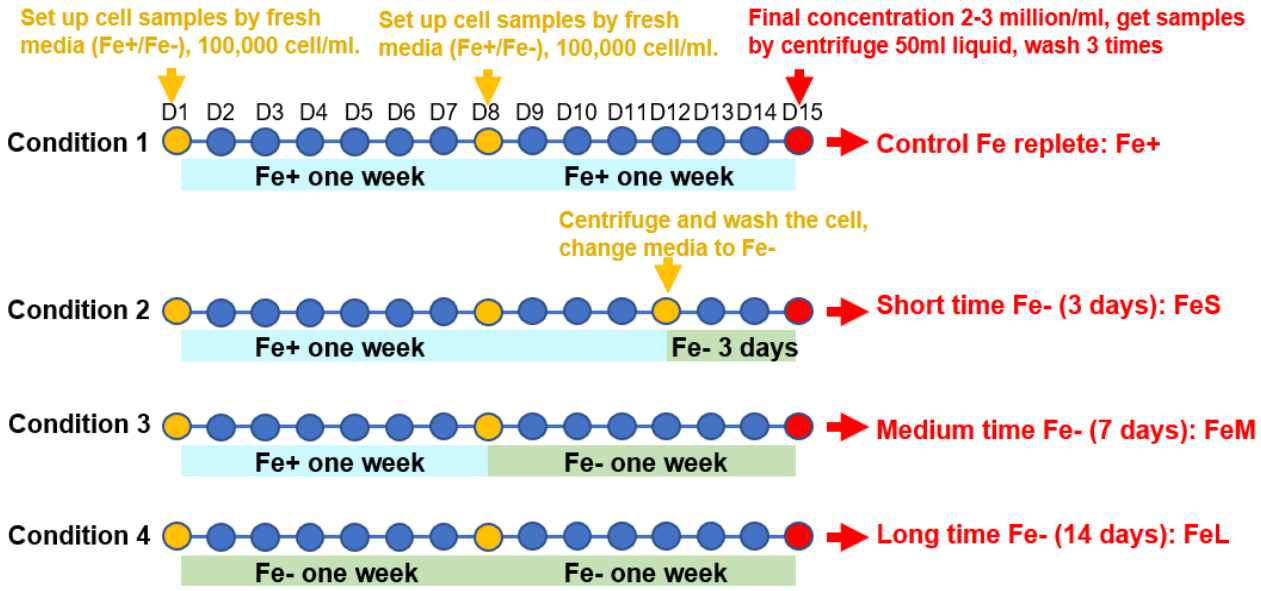

Figure 4.10 Experimental design for Trans3 RNA sequencing under different Fe conditions. Each dot represents a day number. Yellow dots represent days on which cultures were refreshed, and red dots days on which cultures were sampled. Fe+ treatments are shown with blue bars; and Fe- treatments are shown with green bars. Four conditions were designed (Fe+, FeS, FeM and FeL) for RNAseq, 5 cell lines (HA1, HA2, KO1, KO2 and KO3) with three biological replicates set up with starting concentrations of 100,000 cells/mL. All cell lines were refreshed with fresh ESAW media (Fe+ or Fe- according to different designed conditions) at the same cell concentration after seven days incubation (D8) to avoid other nutrient-depletion stresses. For the FeS condition, cell lines were centrifuged at Day 12 (D12) and transferred to Fe- media prior to a short-time (three day) incubation. All cell lines were finally sampled at Day 15 (D15) with final concentrations about 2-3 million cells/mL.

Overview of transcriptome sequencing information

In total, 60 cDNA libraries were generated for transcriptome sequencing. After filtering the raw reads and performing FastQC tests (see Materials and Methods), 0.03 million to 12.15 million clean reads were obtained per library, with a Q20% value around 98% and GC% around 51%. The GC content value in the transcriptome data is similar with previously reported GC contents in *P. tricornutum* ESTs (53.7%) and the whole genome (GC = 48.5%), and no significant bacterial sequence contamination was detected in the data (Scala et al., 2002). Cleaned reads were mapped onto the 12,312 genes annotated in the version 3 *P. tricornutum* genome (Rastogi et al., 2018) using Nextflow with previous defined conditions, and average TPM values were calculated for each gene in each library (Di Tommaso et al., 2017; Dorrell et al., 2022).

PCA analysis and hierarchical clustering

Principal component analysis (PCA), hierarchical clustering and differentially expression genes (DEGs) were calculated from mapped transcriptome data using the iDEP-DiatOmicBase (https://idep.diatomicsbase.bio.ens.psl.eu/). Gene raw read counts were uploaded, with pre-processing to filter lowly expressed genes (keep genes with minimal counts per million (CPM) >0.5 in at least 1 libraries). mRNA relative abundances were transformed as $\log_2$(CPM+1) values by EdgeR for downstream analyses.

Principal Component Analysis (PCA) was performed to investigate the reproducibility of the RNAseq data and examine transcriptional changes in response to different treatments. As shown in Figure 4.11, biological replicates within each cell line exhibited good reproducibility, and all samples were clustered into three groups based on their respective treatments (Figure 4.11). In the Fe+ control condition, cell lines were grouped together and separated from other treatments, while the FeM and FeL treatment groups were clustered as one big group, indicating that the cells exhibited similar responses and influence after seven days of Fe limitation. Interestingly, the FeS cluster was completely separated from the FeM, FeL and Fe+ cell lines, suggesting that diatom cells may have a rapid response to short-time Fe limitation that differs from their response to medium-time and long-time Fe- stress (Figure 12).

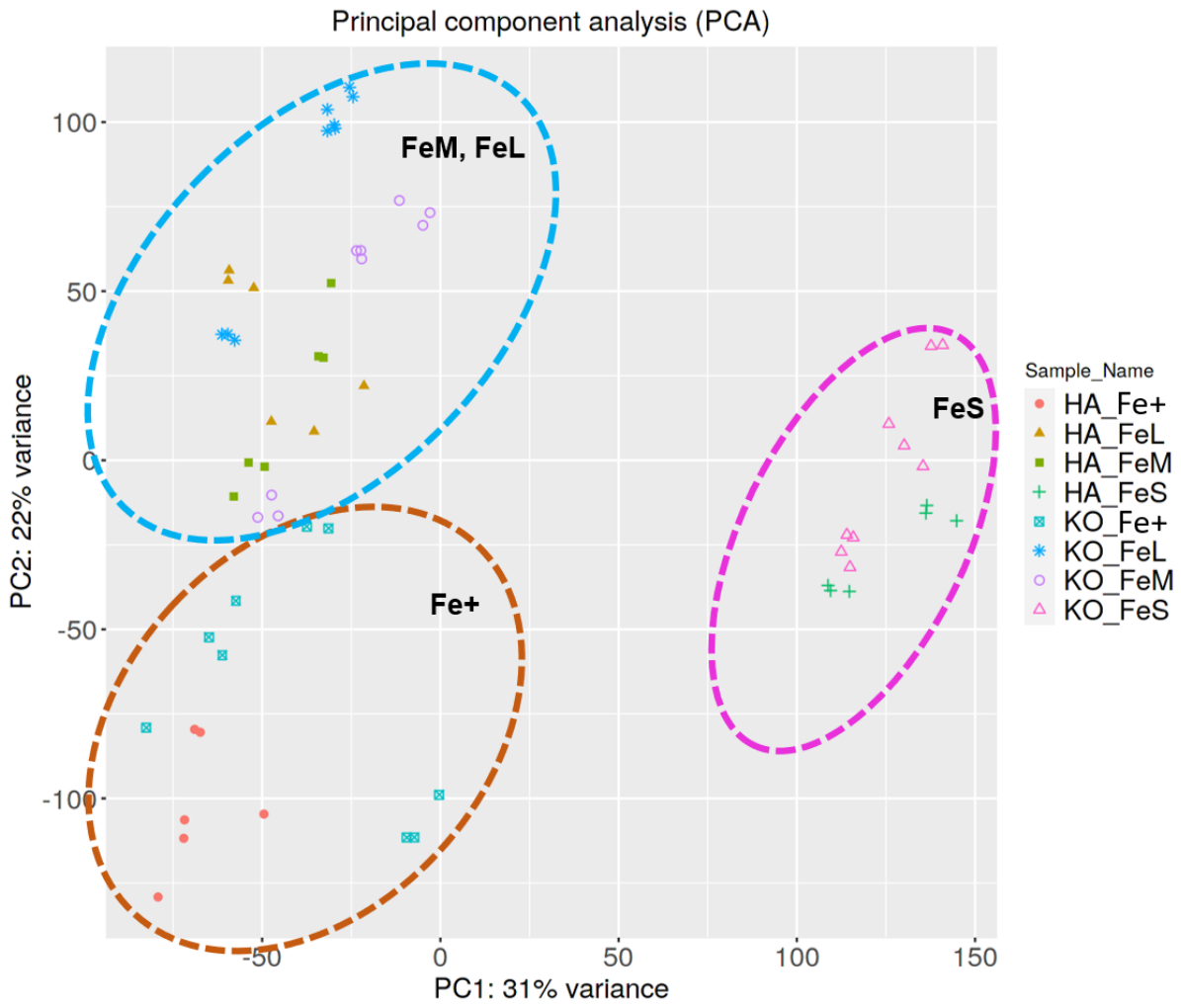

Figure 4.11 Principal Component Analysis (PCA) of RNAseq data. Relative mRNA abundances transformed as $\log_2$(CPM+1) values were used for PCA. All samples were grouped into three clusters by treatment condition (Fe+, FeS, FeM and FeL) and shown by different color circles.

Furthermore, hierarchical clustering was performed as shown in Figure 4.12. These data showed similar results to the PCA (Figure 4.11), with most of the samples being grouped together based on their respective treatments (Figure 4.12). KO2 and KO3 lines were more similar to each other and typically separated from the control (HA) lines, the difference was less apparent for KO1 when compared to controls according to both PCA and hierarchical clustering results (Figure 4.12). Previous qPCR results showed that Trans3 mRNA abundance in the presumed heterozygote KO1 line was not as severely diminished as either KO2 and KO3 (Figure 4.4), w hich may explain the less dramatic RNAseq phenotype observed.

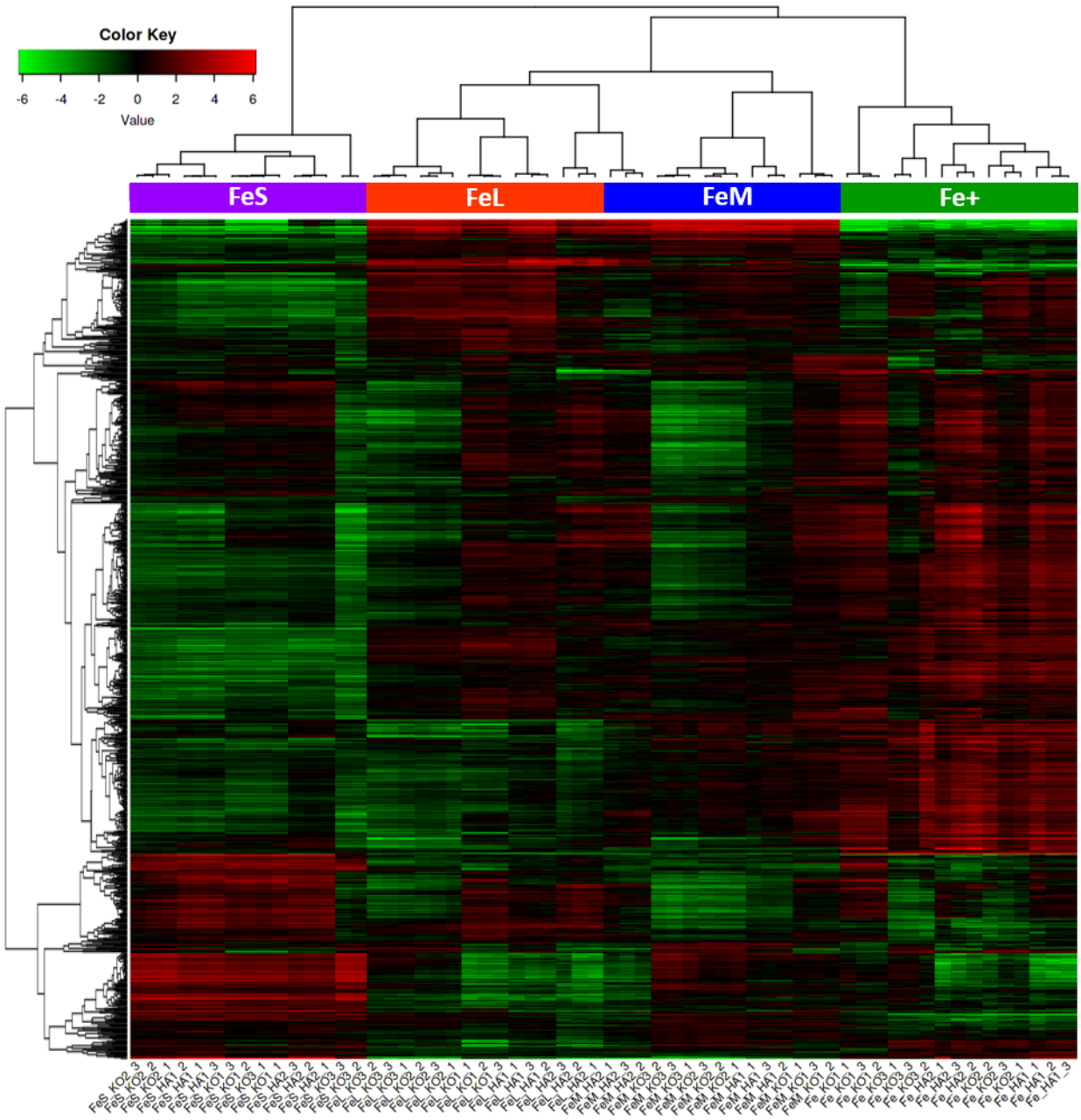

Figure 4. 12 Hierarchical cluster heatmap results of the genes with the 1000 highest standard deviations. Relative mRNA abundance data transformed as $\log_2(CPM+1)$ was used and ranked by standard deviation across all samples. By default, the top 1000 genes were used in hierarchical clustering using the heatmap.2 function in iDEP-DiatOmicBase. The data is centered by subtracting the average expression level for each gene. The distance matrix is 1-*r*, where *r* is Pearson's correlation coefficient. The average linkage is used. The cluster presents both the treatment condition information, shown in different colors above (FeS: purple, FeL: red; FeM: blue and Fe+: green), and different cell line names, listed at the bottom.

Before conducting further analysis of the differentially expressed genes and related pathways between the Trans3 partial knockout (KO) and control lines, I examined the expression trends of known Fe-stress associated genes (Gao et al., 2019; Ait-Mohamed et al., 2020) in the control lines (HA). This preliminary check aimed to determine if a cellular Fe limitation response could be observed after just 3 days of Fe depletion.

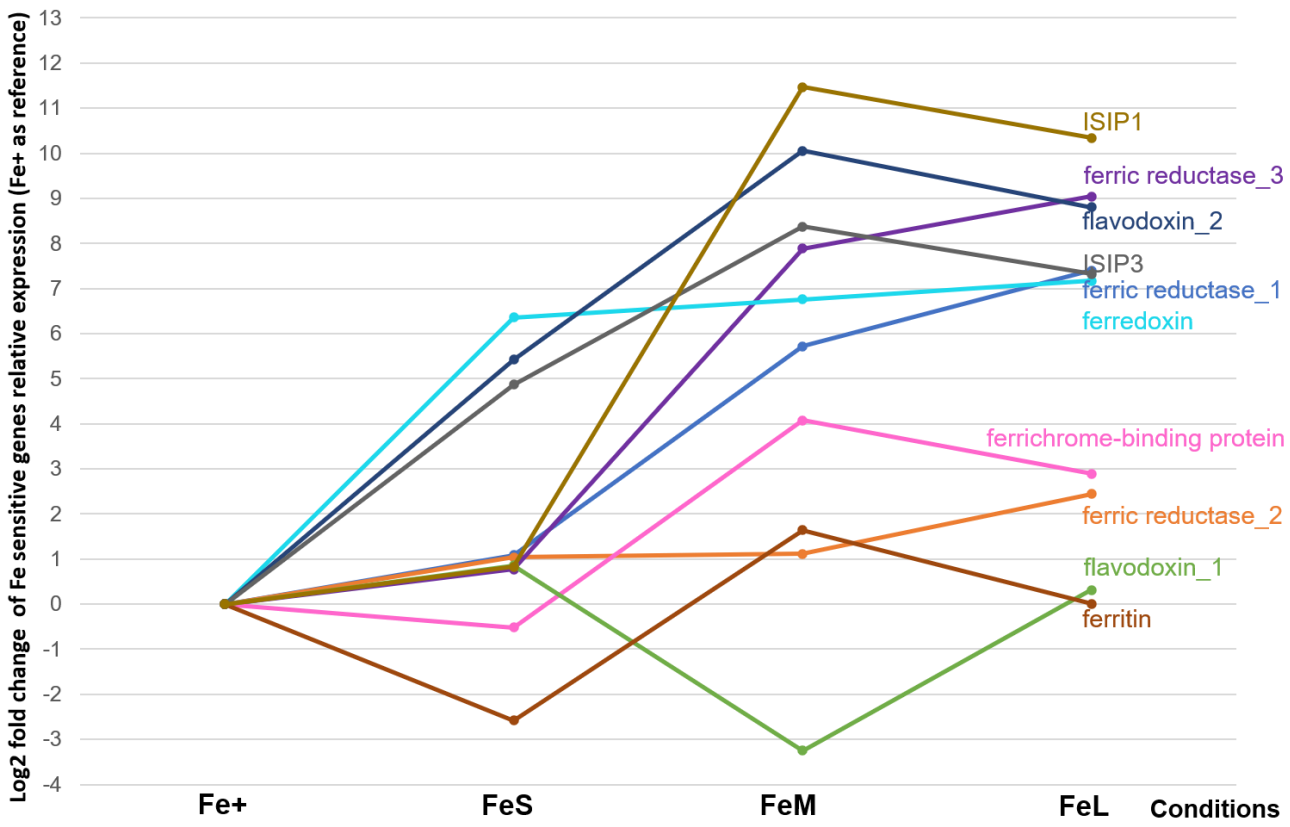

Figure 4.13 mRNA relative expression trends of genes encoding known Fe-stress associated proteins in control lines under different Fe conditions. Fe sensitive genes are selected based on Gao et al. (2019), Ait-Mo hamed et al. (2020) and the version 3 *Phaeodactylum* genome annotation. TPM values are shown relative to those observed in Fe+ lines: $Log_2((TPM+1)_{Fe-\_treatment}/(TPM+1)_{HA\_Fe+})$. Different genes are represented by different colors.

Figure 4.13 illustrates the results, indicating that compared to the Fe+ condition, several Fe-sensitive genes including ferredoxin, flavodoxin 2 and ISIP3 exhibited a more than 2-fold change induction under 3 days of Fe depletion stress (FeS). This finding suggests that the cell lines were already under stress after 3 days of

Fe depletion. Additionally, genes such as ferrichrome-binding protein (J46929) and ferritin (J16343) were upregulated under FeM conditions, which might be indicative of a response to longer periods of Fe depletion as 7 days.

Differentially expression genes (DEGs) in KO and control lines, assessed by pairwise comparison and by interaction between genotype and treatment

DEGs between KO and control lines (HA) under different Fe treatments were identified from RNA sequencing data using raw read counts, following pre-processing for lowly expressed genes, PCA and hierarchical clustering. DEGs were identified using raw read counts with the DiatOmicBase DESeq2 package (Love et al., 2017). Pairwise comparisons of relative mRNA abundances in KO compared with control (HA) cell lines under four different Fe conditions (KO_Fe+ vs HA_Fe+, KO_FeS vs HA_FeS, KO_FeM vs HA_FeM and KO_FeL vs HA_FeL) were tested to determine the effects of Trans3 knockout (Figure 4.14). To determine whether there were differential responses to Fe depletion among the two genotypes, three interaction terms (Figure 4.14) between genotype (HA, KO) and treatments (FeS/FeM/FeL) were also calculated by fitting the negative binomial generalized linear model (GLM) with the command in DESeq2 (Love et al., 2014).

Total DEGs identified in each comparison are shown in Figure 4.14. In general, many DEGs were found in pairwise comparisons of KO with HA cells under each treatment condition, with the fewest (928 DEGs) in KO_FeS vs HA_FeS (named as KO-HA_for_Fe+ in Figure 4.14) and the most (1873 DEGs) found in KO_FeL vs HA_FeL (named as KO-HA_for_FeL in Figure 4.14). Typically, more DEGs were found to have greater relative mRNA abundances (consistent with upregulation) in KO compared to control lines than lower relative mRNA abundances (consistent with downregulation) in KO compared to control lines. This was especially true in the Fe+ and FeS condition, while a greater number of potentially downregulated genes were observed in the FeM and FeL conditions. Fewer DEGs were observed considering the interaction terms between genotypes and Fe depletion treatments than were observed for the pairwise comparisons between KO and control lines in each condition (Figure 4.14).

Compare	KO-HA_Fe+		KO-HA_FeS		KO-HA_FeM		KO-HA_FeL		Inter_FeS		Inter_ FeM		Inter_ FeL	
	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up	Down
Other	498	359	386	186	448	535	682	567	134	110	154	317	120	220
Chloro/Mito/Endo	381	116	270	86	374	231	397	227	60	76	82	163	44	117
Total	879	475	656	272	822	766	1079	794	194	186	236	480	164	337
Plastid located P vaule	0.00	1.00	0.00	0.20	0.00	0.74	0.00	0.96	0.50	0.00	0.03	0.00	0.86	0.06

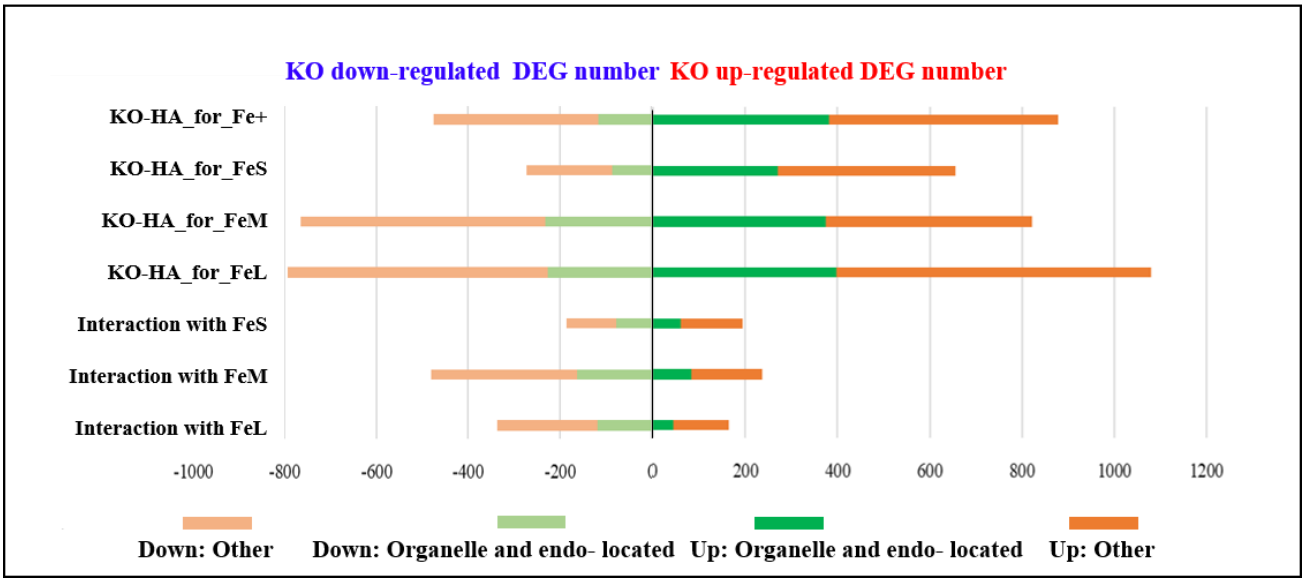

Figure 4.14 Total DEG distributions across pairwise comparison and interaction terms analysis. The total number of DEGs identified in pairwise comparisons of KO and HA lines under different conditions (Fe+, FeS, FeM, FeL), and interaction terms of genotype with treatments are listed above. DEGs for each comparison are separated into those that are upregulated (higher relative mRNA abundance) and downregulated (lower relative mRNA abundance) in KO versus control lines. The number of DEGs encoding proteins inferred to possess different consensus localizations is shown in the table at the top and the color bar plot below, following previous studies (Rastogi et al., 2018; Ait-Mohamed et al., 2020 and Dorrell et al., 2022). Chloro/ Mito/ Endo refer to the predicted localization of the protein encoded by each DEG: chloroplast/ mitochondria/ endoplasmic reticulum. Plastid located *P value*: (red: $P<0.05$; orange $P<0.01$) calculated by hypergeometric test are shown for the total number of DEGs inferred to encode plastid-targeted proteins from consensus ASAFind/ HECTAR predictions, compared to the total occurrence of these genes across the entire *Phaeodactylum* genome. Smaller *P values* correspond to a greater significance of enrichment.

Functional impacts of Trans3 KO on plastid-associated proteins

As a preliminary indication of the impacts of Trans3 KO on cellular physiology, the inferred localization of the proteins encoded by each DEG (considering the consensus of SignalP, ASAFind, HECTAR, and MitoFates) were compared to the entire *Phaeodactylum* genome, following previous studies (Rastogi et al., 2018; Ait-Mohamed et al., 2020 and Dorrell et al., 2022). As shown in Figure 4.14, many

of the DEGs encode proteins inferred to localize to organelles, including the plastid, mitochondria and endoplasmic reticulum, and the DEG lists were significantly enriched (hypergeometric test, $P<0.05$) in genes encoding suspected plastid-targeted proteins. These data, which are consistent with the previous experimental localization of the Trans3 protein to the plastid membranes (Chapter 3), reveal the plastid as the primary site of physiological difference between KO and control lines.

Functional enrichment analysis of organelle related DEGs

Next, organelle-associated DEGs were profiled between KO and control lines under Fe repletion and different depletion conditions, to determine the different physiological impacts of Trans3 KO in response to Fe status. Significantly enriched (hypergeometric test, $P<0.05$) organelle-related pathways amongst DEGs for each condition, relative to all genes in the *Phaeodactylum* genome, are presented as a Venn diagram in Figure 4.15. In general, comparing the KO genotype with control lines, I found a relative upregulation of numerous functional pathways. However, when exposed to prolonged Fe deficiency (FeM, FeL), I observed a more pronounced downregulation of pathways in the KO genotype than control lines.

Over all treatment conditions, the DEGs inferred to be overexpressed in KO relative to control lines were significantly enriched in chlorophyll and terpenoid biosynthesis (Figure 4.15). Significant enrichments were also detected in genes encoding light harvesting complex (LHC) and photosystem proteins in Fe-replete, medium-time and long-time Fe-depletion treatments (Fe+, FeM and FeL). Genes encoding amino acid metabolism (including part of glycolysis) and fatty acid biosynthesis enzymes were enriched (hypergeometric test, $P<0.05$) amongst the upregulated DEGs under most of the conditions except FeL. Genes related to cysteine/ serine/ threonine metabolism (Cys/Ser/Thr); aspartate/ lysine metabolism (Asp/Lys), aminoacyl-tRNA synthesis, iron stress metabolism, and glycerolipid synthesis were enriched amongst the downregulated DEGs in more than one treatment condition (Figure 4.15).

Other pathways were only enriched amongst the DEGs calculated between KO and control lines under one single treatment condition. For example, genes encoding LHCs were only significantly enriched amongst the DEGs

downregulated in KO compared to control lines under short-time Fe depletion (FeS) (Figure 4.15). In contrast, genes associated with fatty acid degradation, carbonic anhydrases and carotenoid biosynthesis were enriched in upregulated DEGs only under Fe-replete conditions; glycolysis and cofactor metabolism were enriched in downregulated DEGs only under FeS; Asp/ Lys synthesis genes were enriched in upregulated DEGs only under FeM (although were enriched in downregulated DEGs in both FeM and FeL conditions); and branched chain amino acid synthesis and the exclusively mitochondria-related arginine/ urea pathway were enriched in DEGs found in FeL conditions only (Figure 4.15).

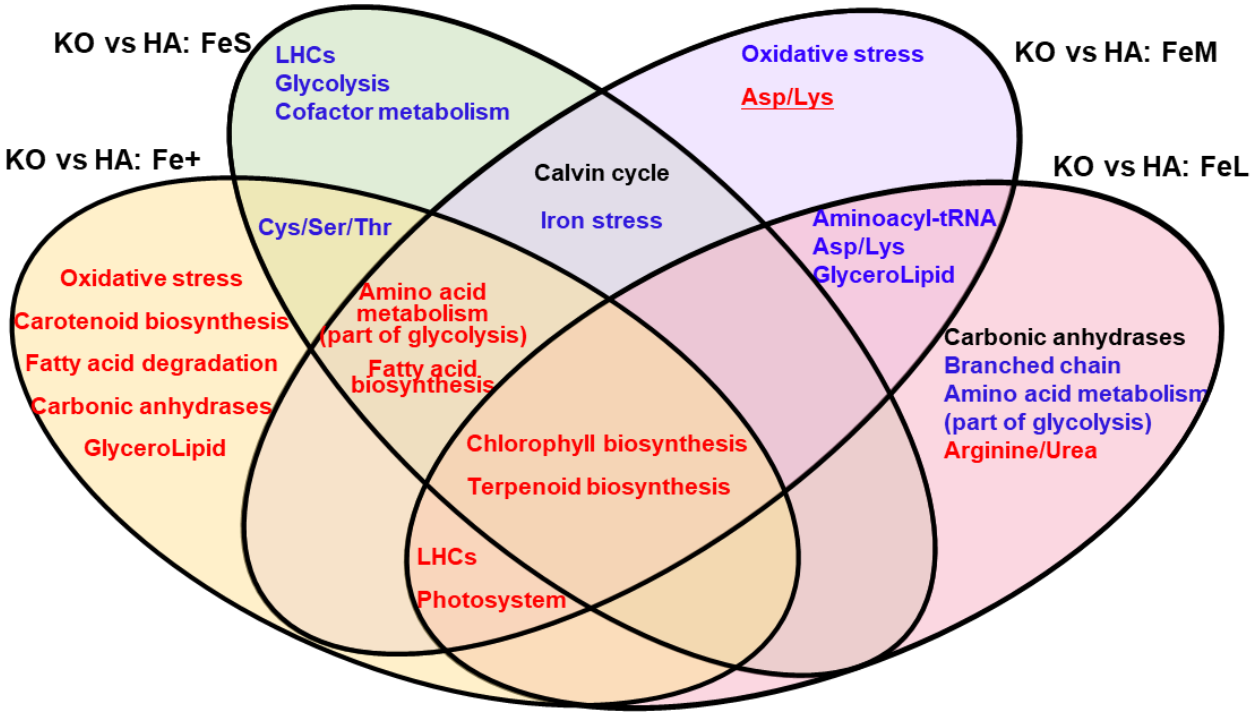

Figure 4.15. KO DEGs significant enrichment pathway in four conditions. *P values* were calculated by hypergeometric test, with significantly enriched pathways defined by $P < 0.05$. Red represents a pathway significantly enriched in the DEGs inferred to be upregulated in KO compared to control lines under each condition; and blue represents a pathway significantly enriched in the DEGs inferred to be downregulated in KO compared to control lines under each condition. Two pathways found to be enriched in both up- and down-regulated DEGs (the Calvin cycle, under FeS and FeM conditions; and carbonic anhydrases, under FeL conditions) are shown in black. Asp/ Lys synthesis is underlined because this pathway is enriched in down-regulated DEGs in both FeM and FeL, but also enriched in up-regulated DEGs in FeM conditions.

Finally, to understand the specific responses of KO lines to Fe depletion, a Venn diagram was made of pathways that were significantly enriched amongst DEGs detected between KO and control lines and that were also detected in three genotype-Fe depletion treatment interactions. As shown in Figure 4.16, the most frequently observed DEGs amongst the interaction terms relate to photosystem and LHC proteins and also the Cys/ Ser/ Thr pathway. Genes encoding LHCs and photosystem proteins were typically inferred to be less affected in KO than control lines tby Fe depletion (Figure 4.16).

This may suggest that these pathways are downregulated in control lines under Fe depletion, and the relative upregulation of these pathways in KO lines relates to a less pronounced downregulation in the KO genotype compared to the control HA genotype. In contrast, the Cys/ Ser/ Thr was typically found to show a greater transcriptional response under Fe depletion in KO than control genotypes (Figure 4.16). The enrichment of the Cys/ Ser/ Thr pathway amongst the DEGs upregulated in KO compared to control lines therefore suggests an exaggerated response in KO lines to Fe deprivation.

Finally, some of the LHC genes were more strongly affected in KO lines compared to control lines under FeL, while others were downregulated less affected. This may reflect that different Lhcs perform different physiological functions (e.g. LhcX genes implicated in photoprotection) and may have different responses to long-time Fe starvation in both KO and control lines (Figure 4.16). Other pathways show stronger transcriptional responses in KO than control lines under medium- and long-time Fe depletion conditions (FeM or FeL) interactions, whereas none were unique to FeS (Figure 4.16).

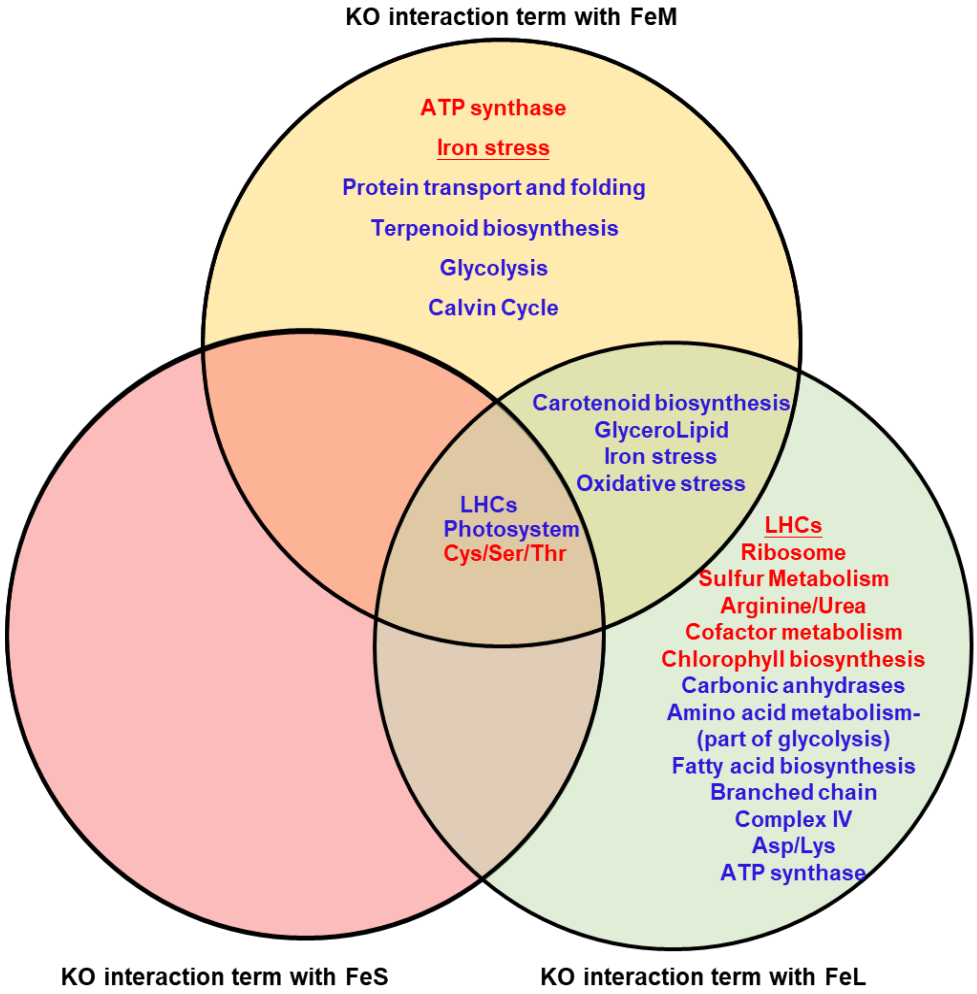

Figure 4.16. Significantly enriched pathways in KO and Fe DEG interaction terms. *P values* were calculated by a hypergeometric distribution test with threshold *P value* 0.05. Red color represents relative genes that show greater transcriptional changes in KO compared to control lines in response to each Fe stress condition; blue represents genes that show less transcriptional change in KO compared to control lines in response to each Fe stress condition. Pathways significantly enriched (iron stress and LHCs) in genes with both stronger and weaker transcriptional responses in KO than control lines under FeM condition and FeL condition are underlined.

Pigment, photosystem and LHC related DEG expression trends

In total, chlorophyll and terpenoid biosynthesis, LHC and photosystem proteins showed the most consistent differences in relative mRNA abundances between KO and control lines across all four Fe conditions. To understand the detailed physiological consequences of these expression trends, heatmap (Figure 4.17) and box plots (Figure 4.18) were produced to explore which genes in each pathway were differentially expressed in KO lines across each condition and their common expression trends across different Fe depletion treatments.

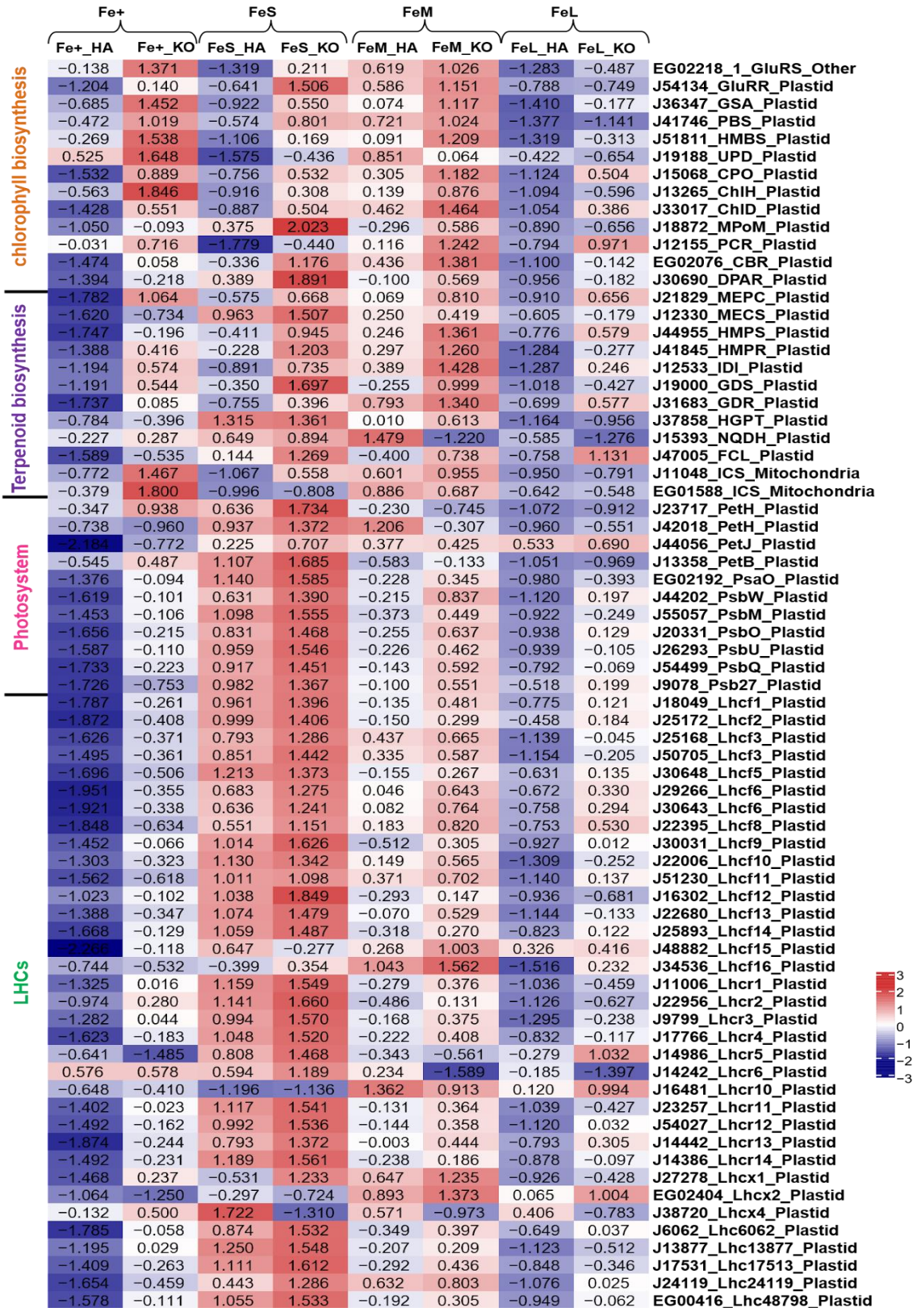

Figure 4.17 Chlorophyll, terpenoid, photosystem and LHCs related DEGs relative expression heatmap. This heatmap was drawn based on relative fold-changes represented by Log_2 (TPM+1) value. Gene ID, common short name and predicted localization are shown in the right column. The heat map was normalized by the Z-score.

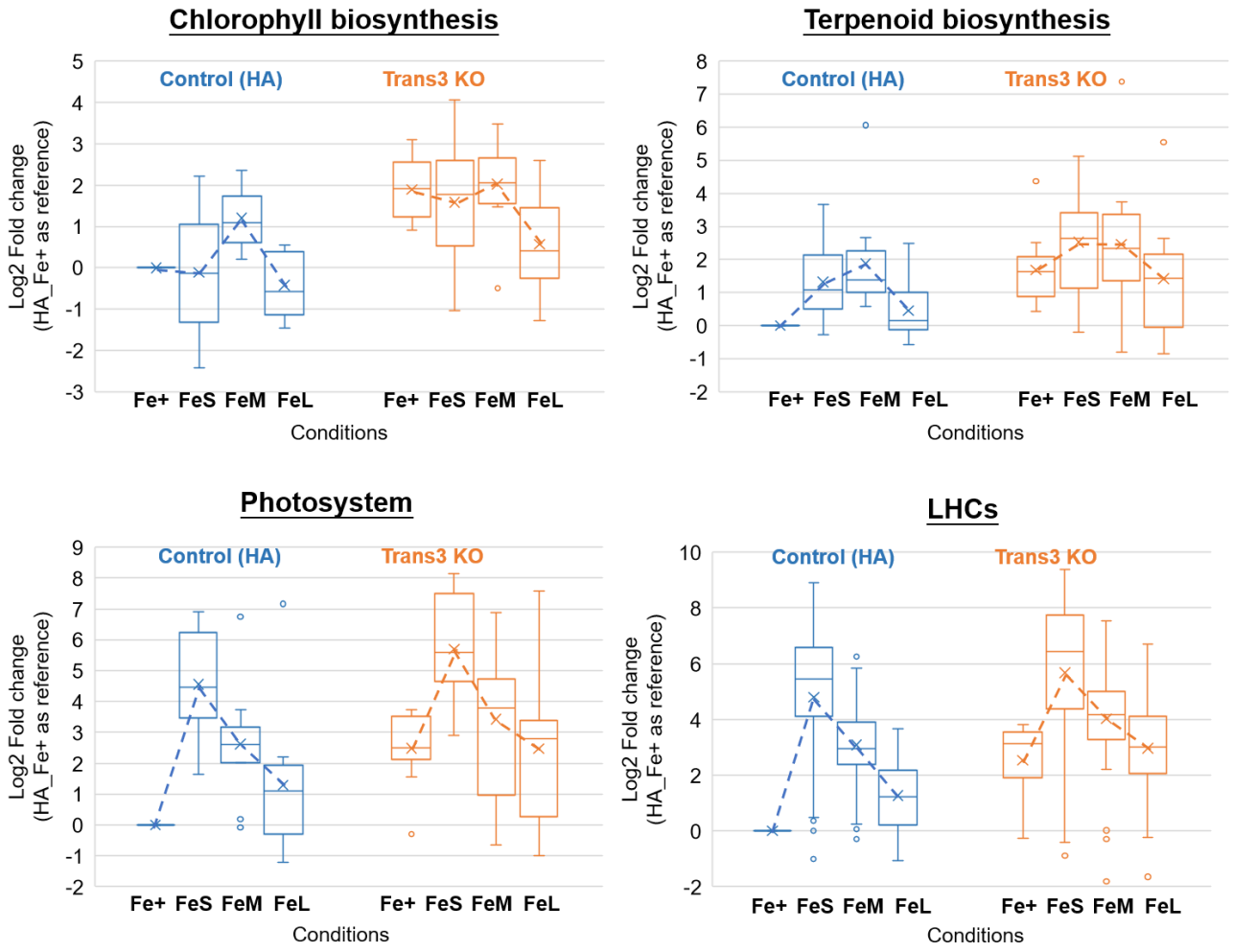

Figure 4.18 Chlorophyll, terpenoid, photosystem and LHC- related DEG relative mRNA abundance and expression trends. In order to show DEG expression trends across Fe stress, relative mRNA were calculated using HA_Fe+ as a normalized reference, and box plots were plotted by the Log_2fold change of relative TPM change: $Log_2((TPM+1)_{treatment}/(TPM+1)_{HA\_Fe+})$. Both average lines and medians are shown. Dotted are drawn between the median values observed for specific pathways across each Fe treatment to represent common expression trends.

As shown in Figure 4.17 and Figure 4.18, in general, DEGs involved in chlorophyll synthesis always showed greater relative mRNA abundances in KO than HA lines across all conditions. The only exception was uroporphyrinogen decarboxylase (Phatr3_J19188, UPD), which shows lower relative mRNA abundances in KO lines in both FeM and FeL conditions. Compared with control lines, most of the chlorophyll biosynthesis genes showed 2-8 folds-change upregulation in KO

lines under Fe-replete conditions (Figure 4.18). Most chlorophyll biosynthesis DEGs showed lower relative mRNA abundances in both KO and control lines under FeS compared with Fe+ conditions (Figure 4.18). In response to FeM stress, most chlorophyll biosynthesis genes were upregulated in both KO and control lines, and in response to FeL stress most chlorophyll biosynthesis genes were downregulated in both lines, showing even lower relative mRNA abundances than under Fe+ conditions (Figure 4.18).

Many terpenoid biosynthesis genes showed higher relative expression value in KO compared to control lines in each condition (Figure 4.17 and Figure 4.18). Under Fe+ conditions, many genes showed greater than 2-fold relative abundance in KO compared to control lines (Figure 4.18). Reflecting the enrichments observed in interaction terms analysis, less dramatic fold-changes were observed in KO than control lines under each Fe depletion condition (Figure 4.18).

LHC and photosystem-associated genes had similar expression trends to one another under Fe-depletion conditions. Under short-time Fe-depletion (FeS) conditions, the relative mRNA abundances of these genes were increased in both KO and control lines compared to Fe-replete conditions (Fe+). Typically, the fold-changes observed in control lines were greater than those observed in KO lines (Figure 4.17 and Figure 4.18), consistent with the interaction terms that suggest that these two pathways are less impacted by Fe- in KO than control lines (Figure 4.16). The relative mRNA abundances of these genes were decreased by FeM and FeL conditions in both control and KO lines, although the final relative abundances under FeL conditions were mostly still higher than those found in Fe+ conditions (Figure 4.17 and Figure 4.18). Some other LHC genes had different expression trends, including Lhcr5 (Phatr3_J14986), Lhcr6 (Phatr3_J14242), Lhcr10 (Phatr3_J16481), Lhcx1 (Phatr3_J27278) and Lhcx4 (Phatr3_J38720). For example, Lhcr5, Lhcr6 and Lhcx4 showed no obvious changes with Fe-depletion; while Lhcx1 and Lhcr10 showed greater relative mRNA abundances under FeM treatment. Lhcx family genes have previously been reported to have photoprotection functions, although the role of Lhcx4 is unclear (Taddei et al., 2016; Buck et al., 2019). Most of the Lhc genes showed greater relative expression in KO than control lines across all conditions (except Lhcr5, Lhcr6, Lhcx4 and Lhcx2), although these two pathways were not

significantly enriched in the DEGs associated with KO lines under FeS conditions (Figure 4.15).

4.3 Discussion

In this chapter, I investigated the physiological significance of Trans3 in *P. tricornutum* cells using three mutants generated by CRISPR-Cas9. These mutants are presumed to be heterozygous, given their genotypes, and the failure to generate clean homozygotic mutants for each line. I have chosen to name these "heterozygous knockout" rather than knockdown lines, reflecting that they represent the stable deletion of one allele rather than the suppression of expression of wild-type alleles in the *P. tricornutum* genome. Taking into account the number of attempts made to generate homozygous Trans3 KO and colonies screened, and the functional centrality of Trans3 to other plastid metabolic pathways (Chapter 2), I propose that Trans3 gene function may be of significant importance to cell viability, and have a mutant-lethal homozygous KO phenotype.

Nonetheless, the phenotypes observed within the partial knockout lines render them a credible system for exploring the functional impacts of Trans3 depletion. The knockout lines showed a 30-90% decrease in Trans3 relative mRNA abundance, as confirmed by qRT-PCR, presumably due to nonsense-mediated decay of translationally non-functional Trans3 transcripts. A variation in the severity of the Trans3 knockdown between each line, and in particular the >50% reduction in mRNA relative abundance in two of the mutant lines, may relate to allele-specific expression biases previously observed in the diploid *P. tricornutum* genome (Hoguin et al., 2021).

Partial suppression of Trans3 expression reduced the log-phase growth rates of the heterozygous knockout lines by an average of 20% in Fe -replete, nitrate-enriched media (Figure 4.6), suggesting that Trans3 function is important to cell physiology, and that KO lines have diminished ability to capitalize on abundant media resources. In contrast, in media lacking supplemented Fe or with ammonium as the nitrogen source, in which control lines had a lower growth rate than observed in Fe- and nitrate-supplemented media, no obvious difference can be observed between Trans3 KO and control lines (Figure 4.6;

Figure 4.7). This may indicate that Trans3 function relates both to nitrate metabolism (although may not be influenced by its concentration), and Fe availability in diatom cells.

Furthermore, based on photo-physiological parameters, it appears that Trans3 KO has a significant negative impact on the ability of diatom plastids to carry out photosynthesis. The strongest impacts were observed on PSI, considering measured P700 content, while PSII seems to be less affected, with only a slight albeit significant reduction of KO line Fv/Fm ratios (Figure 4.8; Figure 4.9). The limited impacts of Trans3 on ETR may reflect that both PSI and PSII complexes were proportionately downregulated in Trans3 knockout lines, with the relative electron transport rate between these two complexes remaining the same. The diminished P700 content of PSI identified in Trans3 KO lines appears to be influenced by Fe depletion, while the Fv/Fm of PSII was not (Figure 4.8; Figure 4.9). This may relate to the greater requirement of Fe for the synthesis of PSI than PSII (Gao et al., 2019; Allen et al., 2011). PSI has been identified as central to the diatom Fe stress in response, e.g., via the use of flavodoxin instead of ferredoxin, and possibly plastocyanin in place of cytochrome c6 (Peers et al., 2006). Interestingly, and consistent with growth rate measurements, the degree of the impact of Fe depletion on KO PSI P700 was less severe than on HA control lines (Figure 4.6; Figure 4.9). It is possible that KO mutant lines regulate certain pathways to compensate for the reduced abundance of Trans3 protein, which may mitigate sensitivity to Fe depletion. Alternatively, the diminished photosynthetic activity about PSI observed in KO lines under Fe-replete conditions may render them less sensitive to stress conditions such as Fe depletion.

RNAseq facilitates the discovery of novel protein functions by providing a global view of gene expression patterns and the regulatory networks involved. Among the DEGs identified between control and KO lines under different Fe conditions, many encoded organelle-targeted proteins (Figure 4.14), consistent with the localization of Trans3 as a plastid membrane transporter. Across each Fe abundance and starvation condition tested, KO lines showed consistently greater relative expression of genes encoding pigment synthesis, photosystem, and light harvesting complex proteins. The greater expression of these genes may compensate partially for the diminished PSI photosynthetic content

observed in Trans3 KO lines, e.g., by increasing antenna size or potentially increasing the *de novo* synthesis of photosynthesis complexes in response to photodamage, using intracellular Fe storage reserves (Gao et al., 2021). Other pathways (e.g., fatty acid degradation in Fe+; glycolysis in FeS; organelle division in FeM; and branched chain amino acid synthesis in FeL conditions) were significantly impacted in KO line RNAseq data under only one Fe condition (Figure 4.15). Moreover, even within individual pathways, some DEGs were inferred to have different transcriptional responses to the treatment, such as certain LHC family proteins (Lhcr5, Phatr3_J14986; Lhcr6, Phatr3_J14242; and Lhcx4, Phatr3_J38720) that showed no difference in expression between KO and control lines (Figure 4.17 and Figure 4.18)

The interaction term analysis gives insights into how the effect of Fe depletion on gene expression differs between genotypes (Love et al., 2014). Considering the interaction between KO and different Fe depletion conditions, LHC and photosystem genes were inferred to be downregulated by Fe stress, but with less sensitive responses in KO lines. The upregulation of these genes in KO compared to control lines might reflect a diminished ability to downregulate photosynthetic metabolism in the absence of Fe, consistent with the relative insensitivity of KO P700 content to Fe limitation (Figure 4.9). The relative abundances of LHC and photosystem gene mRNAs were up to 4-16 fold greater in KO than control lines under Fe+ conditions (Figure 4.18). It is possible that the KO genotype therefore invests in a higher baseline expression of LHC genes using mechanisms that are not responsive to Fe depletion. In contrast, Cys/ Ser/ Thr metabolism was inferred to be upregulated by Fe stress in both lines, with stronger effects observed in KO than control cells (Figure 4.16).

Finally, considering the total gene expression phenotype of knockout lines, limited evidence was observed for a direct function of Trans3 in Fe transport. While many of the genes known to function in Fe stress metabolism (e.g., ISIPs, ferritin and flavodoxin) (Gao et al., 2019) were upregulated in cell lines from three days Fe limitation onwards, little difference was observed between KO and control lines (Figure 4.15). Instead, the observed DEGs suggest a role of Trans3 on plastid metabolic pathways that impact on, and are sensitive to cellular Fe status, such as LHCs and photosynthesis.

Finally, as described in Chapter 2, the Trans3 gene shows a strong positive correlation of expression with genes encoded in the *Phaeodactylum* mitochondrial genome (Ashworth et al., 2016) related to electron transfer and ATP synthesis (Complex I, III, IV, ATP synthase, ribosome and translocase complexes). However, among the RNAseq data, only one exclusively mitochondrial metabolic pathway (arginine/ urea metabolism) under one condition (FeL) was significantly enriched in DEGs between KO and control lines (Figure 4.15). That said, many of the pathways enriched in DEGs require the cooperation of both organelles, e.g., glycolysis, aminoacyl-tRNA synthesis, Cys/ Ser/ Thr and Asp/ Lys metabolism pathways (Figure 4.15). A previous Weighted Gene Co-expression Network Analysis (WGCNA) of large-scale *Phaeodactylum* RNAseq data placed Trans3 in a coregulated gene module (Blue) that unifies proteins involved in metabolic cooperation between the plastids and mitochondria (Ait-Mohamed et al., 2020), such as carbon fixation, fatty acid biosynthesis and glycolysis enzymes, and including aminoacyl-tRNA synthetases, which are typically dual-targeted to both organelles in diatoms (Gile et al., 2015, Dorrell et al., 2017, Dorrell et al., 2019). Overall, my localization, photo-physiological and RNAseq data strongly suggest that the activity of Trans3 impacts primarily on the plastid, but this may include metabolic pathways that indirectly impact on the mitochondria.

In conclusion, the observations in this chapter provide new insights into the physiological role of Trans3 in *P. tricornutum*. I highlight the crucial role of Trans3 in diatom growth and photosynthesis, which relate in particular to pigment synthesis, photosystem and LHC function, and potential interactions with cellular Fe status, along with the requirement of nitrate as a nitrogen source.

4.4 Materials and methods

4.3.1 Trans3 partial knockout mutant line generation by CRISPR-Cas9

P. tricornutum Trans3 mutants were generated by CRISPR-Cas9 as previously described (Nymark et al., 2017). Trans3 CRISPR-Cas9 specific target regions for guide RNA were designed using the online tool phytoCRISPEX: https://www.phytocrispex.bio.ens.psl.eu/CRISP-Ex/ (Rastogi et al., 2016). CRISPR target regions were selected according to the following requirements: a) a

consensus sequence G/A(N)19NGG; b) unique occurrence of the first 12 bases before the NGG within the *P. tricornutum* genome to minimize off-target effects; and c) proximity to the CDD 5′ end. In particular, CRISPR target regions were selected within Exon 1 of the Trans3 gene sequence that were unlikely to introduce cryptic ATG codons, to maximize the probability of disruption of the Trans3 protein. Primer pairs were designed to replace the fragment of the target region within the pU6-sgRNA vector (Figure 4.19), consisting of the target region and adjacent vector sequences.

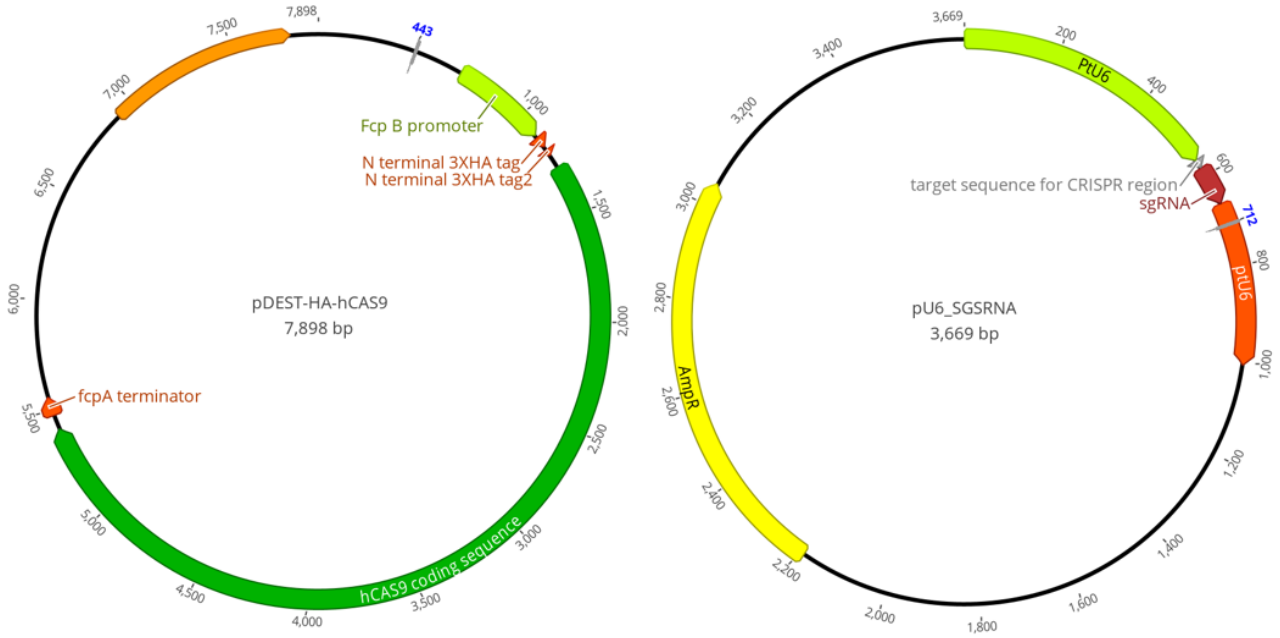

Figure 4.19 pU6 sgRNA and pDEST-hCas9 Vector maps for CRISPR-Cas9 Trans3 knockout mutant generation. Vector maps were generated from Geneious version 10.0.9 (Kearse et al., 2012)

Plasmids were generated by PCR amplification of the linearized complete vector sequence, bordered by terminal direct repeats of the guide RNA sequence, with Phusion polymerase prior to transformation and circularization in chemically competent *E. coli*. The plasmid sequence was then confirmed by Sanger sequencing of the introduced guide RNA. The pU6-target-sgRNA that contained CRISPR region and sgRNA, pDEST-HA-hCas9 that contained hCas9 and pPhaT1 vector that contained zeocin for selection were co-transformed into wild type cells by bombardment. Empty vector controls were generated by bombardment using only hCas9 and pPhaT1 vectors. Recovered colony sequences were genotyped using primer pairs spanning the CRISPR target region. All primers for

KO mutant generation and testing are listed in Thesis Annex II.

Single colonies separation from potential heterozygous mutant line

For potential heterozygous mutant lines that showed "ambiguities" or "disordered overlaps" in the Sanger sequence after the target region, multiple further separation methods were utilized to determine if the cell lines were a mixture of KO cells and WT cells, and if the KO cells were homozygous or included heterozygous KO lines. These (Figure 4.20) included serial dilution by streaking on agar plates, resuspension with agar plating beads, and single-cell dilution on liquid medium in 96-well microtitre plates. Streaking was done using a sterile inoculation loop to drag a single diatom colony across the surface of an ESAW agar plate (containing 100 µg/mL zeocin) back and forth in a zigzag motion until approximately 30% of the plate was covered. The streaking procedure was repeated several times while rotating the plate to produce single clonal colonies or diatom cells. Resuspension with agar plating beads were performed by diluting KO cell culture to 50 cells/mL concentration, and spread 1 mL diluted cell culture on a zeocin/ESAW agar plate as above using Colirollers plating beads (Novagen, 71013-1PKG). Single-cell dilutions were performed by counting cell densities using Malassez cell counting, and diluting cultures to a density of 10 cells/mL, prior to inoculation of 100 µL diluted culture into each plate well. Presumed heterozygous Trans3 KO lines and empty vector control HA lines were cryopreserved in 12% DMSO in -80℃ according to a previously defined protocol (McLellan et al., 1989).

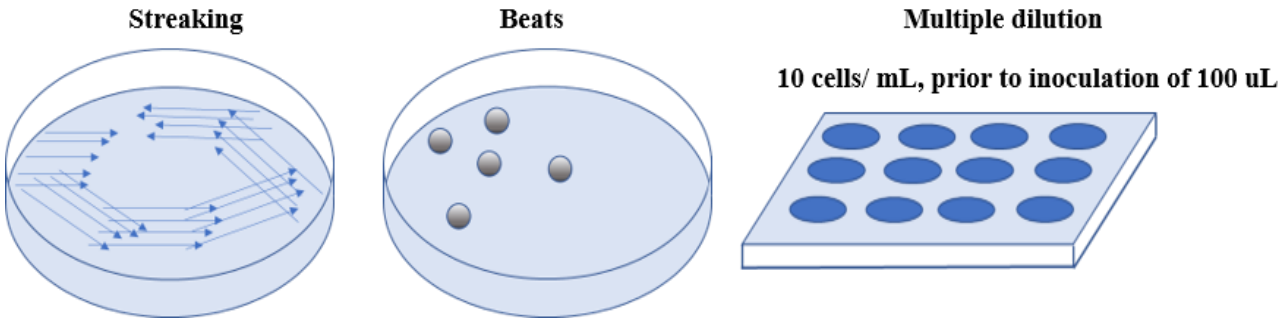

Figure 4.20. Diagram of methods for single colony (pure strain) separation from potential mixed or heterozygous KO colonies

Random PCR

Despite multiple rounds of dilution by the above methods, no homozygous mutant lines were obtained. Random cloning was used to identify the genotype and allele ratios from potential heterozygous cell line colony PCR products. Crude colony PCR products were purified using a NucleoSpin™ Gel and PCR Clean-up Kit (Macherey-Nagel Nº Cat. 740610.20), and ligated into a pGEM-T vector (Promega) via the following protocol:

2x Rapid ligation buffer: 5 µL
pGEM-T vector (50ng): 1 µL
PCR product (≈50 ng/µL)/Control insert DNA: 2-3 µL
T4 DNA ligase (3 weiss units/µL): 1µL
Deionized water to final volume 10 µL

Ligations were performed at 4 ℃ overnight, and introduced into chemically competent *E.coli* (Top10) cells as before. 200 µL transformed cells were plated on LB+ampicillin agar plates with 20 µL X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) solution with concentration 20 µg/mL and 20 µL IPTG (isopropyl β-D-1-thiogalactopyranoside) solution with concentration 100 mM, and incubated overnight. The solutions were made as below:

20 µg/mL X-gal solution: 0.006 g X-gal + 300 µL DMSO
100 mM IPTG solution: 0.006g IPTG + 250 µL water

The pGEM vector sequence encodes a β-galactosidase, which will cleave X-gal and produce a blue color on induction with IPTG (Figure 4.21). Plasmid sequences in which a PCR product was successfully cloned in contrast have an interrupted β-galactosidase gene, which will not produce a blue color. Colonies that showed a white color indicated that the pGEM vector was interrupted by a PCR product, 20 of them were therefore randomly selected in one plate and sequencing to test the KO genotype and its percentage. For each cell line, colony random selections were repeated three times in three different plates.

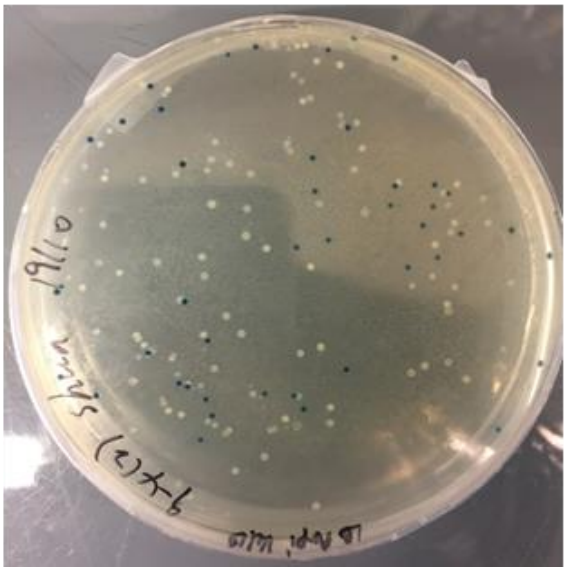

Figure 4.21. An example of blue-white selection plate for random PCR

4.3.2 qPCR analysis

DNase-treated RNA and cDNA, generated from three biological replicates of each culture line (HA1, HA2, KO1, KO2, KO3) was generated as previously described in Chapter 3. Trans3 qPCR primers were designed using NCBI primer blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) with amplicon size 100-200 bp and melting temperature 58-62 ℃, and were tested using a serial DNA dilution to have a more than 90% amplification efficiency. Primers for Trans3 qPCR were also listed at the end of the thesis.

Trans3 mRNA relative abundances were calculated using a LightCycler® 480 High Resolution Melting Master (Roche: 04909631001) with the following cycling conditions: 95 °C for 30 s, followed by 40 cycles of 95 °C for 10 s, 55 °C for 10 s and 72 °C for 20 s. Reactions were performed in 10 µL volumes containing 5 µL of 2 × LightCycler® 480 High Resolution Melting Master, 1 µL diluted cDNA sample (corresponding to approximately 4 ng template RNA), 1 µL nuclease-free water, 3 µL diluted primer mix of each pair of forward and reverse primer (1.5 µM concentration of each pair of primers). Triplicate technical replicates were performed for duplicate cultures. Ribosomal protein S1 RPS (Phatr3_J44451) and TATA-box binding protein TBP (Phatr3_J10199) were used as the endogenous control genes for normalizing expression of the target gene (Siaut et al., 2007, Sachse et al., 2013). The $2^{-\triangle\triangle Ct}$ method was used to

calculate relative gene expression values. *P values* of difference between the expression levels of Trans3 in knockout versus control lines were calculated by one-way ANOVA, considering both selected Trans3 qPCR amplicons, and both normalization genes.

4.3.3 Different cultivation conditions for phenotyping

Fe limitation (Fe-)

Culturing was performed as previously described in Chapter 3. Fe-replete (Fe+) and Fe-depleted (Fe-) ESAW media were both produced using iron-free reagents based on a protocol from Xia Gao (Gao et al., 2019). For Fe- condition preparation, all reagents were handled only using plasticware in a flow hood as a trace metal clean environment to minimize iron contamination, because glass bottles can absorb heavy metals and medium can be contaminated by metal tools, like steel spatulas (Kazamia et al., 2018; Gao et al., 2019). Salts and microelements required for media preparation were dissolved as concentrated stock solutions according to the concentration protocol in Chapter 3 Table 3.1 and kept in Falcon tubes. All stocks were stored in a 4°C fridge avoiding light. Fresh media was always mixed from the stocks and filter-sterilized before experiments. Filtration units (NalgeneTM Rapid-FlowTM Disponsible Sterile Filtration Units with PES Membrane) were used to sterilize the medium instead of autoclaving, since the autoclave process is also known to contaminate media solutions with iron (Kazamia et al., 2018; Gao et al., 2019).

For growth curve comparisons of Fe-, cells were pre-cultivated with different treatments at least one week before being used as original cells in growth curve experiments.

Different Nitrogen source and concentration conditions

Trans3 KO line phenotypes were examined in different nitrogen sources and concentrations. $NaNO_3$ was used as a control nitrogen source in ESAW with final concentration 5.49×10^{-4} mol/L. 20% and 40% of this NO_3^- concentration were used as low nitrate conditions. $NaNO_3$ was exchanged for $NaNH_4Cl$ in ESAW media with the same concentration (5.49×10^{-4} mol/L) to test the phenotypic impact of ammonium as an N source.

4.3.4 Growth curve and flow cytometry set up

Growth curve cell line set up and growth rate calculation

Cell lines without contamination were counted using a Malassez counting chamber under the microscope. Growth curves were established with a starting counted cell density of 10^4 cells/mL. Each cell line (two empty vector-transformed WT cell lines, named HA1 and HA2; and three Trans3 KO lines, named KO1, KO2 and KO3) were screened with three technical replicates. Colony PCR tests were repeated before each growth curve measurement to ensure all KO lines retained mutant genotypes.

To characterize the growth rate phenotypes in different condition, cell numbers were counted using flow cytometry (FCM) at the same time every day over a two-week period. The Flow Cytometer (FCM) used is the 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. The Process was set up as: 2 uL/sec, 40 uL, Maximum Acquisition Speed 25,000 events/sec.

Relative growth rates were measured in cell lines inferred to be in exponential phase (1×10^4 - 2×10^5 cells/mL by FCM) (Figure 4.5). A minimum of three and up to six successive time points were used for each calculation. In each case a strong correlation was observed between the log value of cell concentration and time ($r > 0.99$) confirming the cells were in exponential phase, and not impacted by cell shading or other nutrient stresses. For cell cultures with densities > 2×10^6 cell/mL (FCM) (for which counted events were judged to exceed the maximum acquisition speed), cell lines were diluted 10 times to measure cell concentration and finish the growth curve. Values were represented as mean ± SD. Statistical significances were calculated by one-way ANOVA.

4.3.5 Photosynthesis parameters measurements

PSII parameter test

To image chlorophyll fluorescence, 200 µl of cell suspension were placed on wells of a 96-well plate. Cells were dark-adapted for 10 min prior to measurement. Wells were imaged for fluorescence emission using a SpeedzenDeepGreen III fluorescence imaging setup (Figure 4.22) (IRIG,

Grenoble, France). The maximum fluorescence (Fm) and Fm′ were measured using saturating light (2180 µE m^{-2} s^{-1}, duration 280 ms). The maximum PSII efficiency, Fv/Fm, was calculated as a function of basal fluorescence (F0) and maximum fluorescence (Fm) of dark-adapted algae (Fv/Fm=(F0-Fm)/Fm). ETR derives from the relative PSII yield (ΦPSII) in the dark as (Fm′ – Fs)/Fm′, where Fm′ and Ft are maximum fluorescence (during saturating pulse) and steady state fluorescence of light exposed algae. The electron transport rate (ETR) was calculated as follows = (ΦPSII*PAR)/2 (Maxwell and Johnson, 2000) (Figure 4.22). To measure the ETR algae were exposed to a red actinic light ramping at different intensities following 5 minutes intervals (PAR: 25, 58, 105, 178, 315, 540 µmol photons m^{-2} s^{-1}).

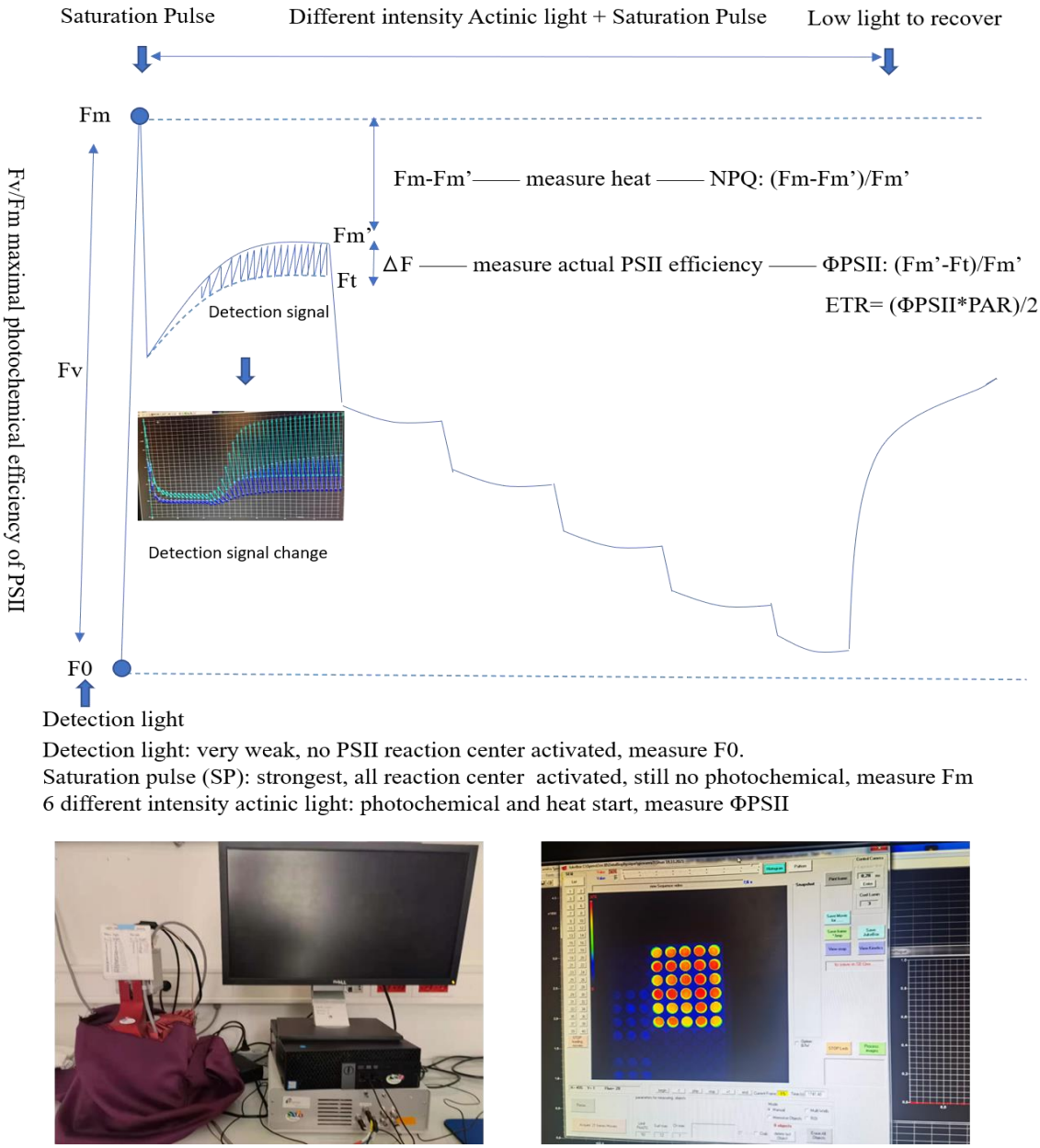

Figure 4.22 Schematic diagram of PSII parameters and the SpeedzenDeepGreen device

P700 content measurement for PSI

DCMU, 3-(3, 4-dichlorophenyl)-1,1-dimethylurea), is a commonly used PSII inhibitor to study ring electron transport activity, which can greatly reduce or block linear electrons from the upstream PSI, resulting in a complete P700 oxidation allowing so estimation of PSI content (Alric et al., 2010). DCMU-sensitized P700 (20 µM) was measured in 1 mL cell cultures with a concentration of 1×10^6 cells/mL using a JTS-10 spectrophotometer (Biologic, France) as previously described (Bailey et al., 2008).

Cells were exposed for 5s to 2000 µmol photons m^{-2} s^{-1} to induce PSI oxidation. Light-induced difference spectra of the reduced-minus-oxidized forms of the P700 reaction center of PSI and the amplitude of the absorbance change, at 705 nm following photo-oxidation of P700 was used to determine functional reaction center concentration: P700- (fully reduced, absorbance in the dark) minus P700+ (completely oxidized, DCMU +strong light).

4.3.6 RNAseq analysis

Experimental treatment design and sampling

Transcriptome sequencing experiments were designed according to the growth rate and photo-physiological values observed in Fe limitation conditions. Three different Fe limitation conditions were designed (Figure 4.10): three days Fe limitation as a short-time Fe- treatment (FeS), corresponding to the condition used for photosynthetic parameters; and one week Fe limitation as a medium-time Fe- treatment (FeM) and two weeks Fe limitation as a long-time Fe-treatment (FeL), corresponding to the conditions used for growth curves. Fe-replete conditions (Fe+) were used as a control, following both photophysiology and growth conditions. Five cell lines were tested for each condition, corresponding to two empty control lines (HA1 and HA2) corresponding to *P. tricornutum* wild-type cell lines transformed with pPhat and HA-Cas9 vectors without guide RNAs, and three presumed heterozygous Trans3 KO lines (KO1, KO2 and KO3). Three biological replicates were performed for each cell line and condition. Cell lines were set up with starting concentration of 1×10^5 cells mL^{-1}, as 100 mL volumes in 225 mL plastic flasks with filtered tops.

Cell lines were refreshed with fresh media at Day 8 to make sure the cell lines were in the same growth period as the exponential phase when sampled, and

no other nutrient stress than Fe limitation influenced the results. Fast Fv/Fm (with 10% FP, 70% SP) tests were performed for all cell lines using a PAM (PAR-FluorPen FP 110, Photon Systems Instruments) prior to sampling, to confirm that harvested cell line Fv/Fm values were stable around 0.65, i.e. were in a healthy state with no apparent nitrogen/ phosphate exhaustion of growth media.

Sample collection and RNA extraction

For each condition, 50 mL of each cell lines with final concentration around 2-3×10^6 cell/mL were harvested by centrifugation for 20 min at 4000 rpm for RNA extraction, as described in Chapter 3.

Transcriptome sequencing and assembly

60 DNAse-treated RNA libraries were sequenced on a DNBseq Illumina platform (BGI Genomics Co., Ltd, Hongkong, China) with 100 bp paired-end sequencing. Raw reads were filtered by removing adaptor sequences, contamination and low-quality reads (reads containing over 40% bases with Q value < 20%) to obtain clean reads. Clean reads were mapped to the version 3 annotation of the *P. tricornutum* genome (Rastogi et al., 2018) using Nextflow with previous defined conditions, and average TPM values for each gene in each library were calculated following previous studies (Di Tommaso et al., 2017; Dorrell et al., 2022).

PCA analysis and hierarchical clustering

Principal component analysis (PCA) and hierarchical clustering were processed using DiatOmicBase, integrated into the iDEP.91 platform (https://idep.diatomicsbase.bio.ens.psl.eu/) to allow comparative genomics with diatom reference genomes. After filtering lowly expressed genes, defined as those with a count per million (CPM) < 0.5 in all libraries, gene expressions were transformed as $\log_2$(CPM+1) for PCA and hierarchical clustering. Genes were ranked by standard deviation across all samples. By default, the top 1000 genes were used in hierarchical clustering using the heatmap.2 function in iDEP-DiatOmicBase. The data was centered by subtracting the average expression level for each gene. The distance matrix is 1-*r*, where *r* was the Pearson correlation coefficient. The average linkage was used.

Differentially Expressed Gene (DEG) analysis and pathway enrichment tests

As described above, genes with CPM higher than 0.5 in at least one library were kept for quantitative analysis. Gene raw read counts were used for differentially expressed gene (DEGs) analysis though DESeq2 in iDEP.91. Genes that were differentially expressed in knockout compared to empty vector control lines, and interaction terms of genotype with treatments likewise were detected using the DESeq2 package on the DiatOmicBase platform. Genes with fold-changes higher than 2 and FDR values lower than 0.05 were defined as DEGs.

Interaction term analysis:

KO = 0 —— Treatment effect on KO

EV = 0 —— Treatment effect on EV

KO > EV?

Organelle related DEGs Pathway enrichment

The localization predictions of proteins encoded by DEGs were based on consensus ASAFind, HECTAR and MitoFates predictions, and organelle-related pathways were defined according to gene function annotations and a previously obtained metabolic map of the *Phaeodactylum* plastid and mitochondrial proteomes (Ait-Mohamed et al., 2020; Dorrell et al., 2022). Enrichments in annotations associated with upregulated or downregulated genes (treated separately) in each condition were calculated using a hypergeometric test with threshold *P value* 0.05.

Chapter 5. Structural and biochemical exploration of Trans3 as a potential molybdate transporter

Abstract

Previously, I have presented the growth, photophysiological and transcriptome phenotypes of three potentialy heterozygous Trans3 knockout (KO) lines. These combined data suggest that the PPM-located transporter Trans3 plays essential roles in *Phaeodactylum* cells principally related to pigment biosynthesis, photosystem and light harvesting complexes, and its physiological functions may relate to iron availability and nitrate source. Here, I consider structural and biochemical data to identify a potential substrate of Trans3. According to conserved domain analysis, Trans3 possesses similar motifs to the *Chlamydomonas reinhardtii* molybdate transporter 2 (MOT2). Possible molybdate-related functions of Trans3 were tested by qPCR, by experimental measurements of molybdate uptake, and by transcriptomic analysis of genes implicated in molybdate metabolism in Trans3 knockout lines. Trans3 mRNA abundances were significantly increased in empty vector control lines in Fe-replete, nitrate-containing media when supplemented with high molybdate (91.8 nM or 918 nM), but were not induced in media containing ammonium as a primary nitrate source. Moreover, molybdate uptake rates in KO lines were approximately 1.5-fold higher than empty vector controls in molybdate-rich media. RNA-seq comparisons further demonstrated that higher relative mRNA abundances in Trans3 KO than empty vector control lines of several genes associated with molybdate. Considering its experimental localization, measured biochemical activity and RNAseq data, I propose that Trans3 may act as a plastid molybdate transporter and allow molybdate (MoO_4^{-2}) to cross the peri-plastid membrane.

5.1 Introduction

5.1.1 Molybdenum (Mo) is an important element for living organisms

Molybdenum (Mo) is a transition element present in relatively small amounts in the environment (Wang et al., 2009; Tejada-Jiménez et al., 2013). Measured Mo concentrations range from 0.2–36 mg/kg in soils (Broadley et al., 2012), and are approximately 100 nM in seawater (Wang et al., 2009; Tejada-Jiménez et al.,

2013). Mo plays essential and irreplaceable roles for almost all living organisms except some thermophilic anaerobes, and can be taken up by organisms in the form of molybdate oxyanions (MoO_4^{-2}) (Hille, 2002), termed MO as an abbreviation for convenience. The measured MO content in *Chlamydomonas reinhardtii* is about 100000 atoms per cell, and tissue MO concentrations in *Arabidopsis thaliana* vary between 0.1 and 1 µg/ 1 g dry weight (Merchant et al., 2006; Llamas et al., 2011).

Despite its limited abundance in the environment and in living cells (Merchant et al., 2006; Wang et al., 2009; Llamas et al., 2011; Tejada-Jiménez et al., 2013), Mo plays significant roles in determining plant growth (Kaiser et al., 2005). Pioneering obs ervations of Arnon and Stout (1939) demonstrated that Mo depletion led to mottling lesions and whiptail in tomato leaves, and subsequent laboratory studies have indicated that Mo deficiency can lead to necrotic leaves, impairment in flower development, and a decrease in overall plant growth (Kaiser et al., 2005; Schwarz and Mendel, 2006; Mendel and Bittner, 2006; Mendel, 2011; Tejada-Jimenez et al., 2013). In phytoplankton, several studies have indicated that Mo depletion potentially restricts growth (Howarth and Cole, 1985; Ter Steeg et al., 1986), alongside nitrogen assimilation and photosynthesis (Cartier, 1973; Wallen and Cartier et al., 1975; Rueter and Petersen, 1987; Baldisserotto et al., 2013). It should be noted that Mo deficiencies are generally considered to be infrequent in nature (Kaiser et al., 2005), although most of these studies have focused on physical parameter measurements of Mo under environmental or low concentrations. Moreover, it is currently uncertain whether elevated Mo concentrations are toxic to photosynthetic organisms, and whether phytoplankton possess any tolerance strategies in response to such conditions.

Finally, previous reports have indicated significant crosstalk between molybdate and Fe (Bittner, 2014). The interplay between the uptake mechanisms for molybdate and iron is evident as (i) most Mo-dependent enzymes rely on iron-containing redox groups, such as iron-sulfur clusters or heme, (ii) molybdenum metabolism employs mechanisms typically associated with iron-sulfur cluster synthesis, and (iii) both molybdenum cofactor synthesis and extra-mitochondrial iron-sulfur proteins require the function of a specific mitochondrial ABC-type transporter.

5.1.2 Molybdenum cofactor (MoCo) and associated synthesis pathway

Molybdate itself is biologically inactive and must be incorporated into metal cofactors by a complex biosynthetic machinery to allow its functional activity in molybdenum-requiring enzymes (molybdoenzymes). There are two types of Mo-related cofactors in nature, namely a pterin-based molybdenum cofactor (MoCo) and an iron–sulfur-cluster-based iron–molybdenum cofactor (FeMoCo). MoCo is the normal form of cofactor found in most molybdoenzymes, while FeMoCo is only found in the bacterial enzyme nitrogenase and is involved in symbiotic nitrogen fixation in legumes (Schwarz and Mendel, 2006; Schwarz et al., 2009, Gil-Díez et al., 2018). This may explain why the Mo requirement in legumes is substantially higher than other dicotyledons (Schwarz et al., 2009).

The MoCo production process (Figure 5.1) has been well studied in higher plants and is believed to be conserved across all domains of life. In eukaryotes, MoCo complexes are synthesized in two cellular compartments: the mitochondria and cytoplasm, and can be divided into four consecutive steps according to the biosynthetic intermediates of each step (Llamas et al., 2011). The first step is guanosine triphosphate (GTP) circularization, which takes place in the mitochondria, where GTP is converted to cyclic pyranopterin monophosphate (cPMP) by the S-adenosine methionine (SAM) dependent enzyme Cnx2 and the hexamer protein Cnx3 (Llamas et al., 2011). Next, cPMP is exported to the cytosol through the involvement of a mitochondrial inner membrane ABC-transporter ATM3 (Llamas et al., 2011). Cytoplasmic cPMP is then converted into molybdopterin (MPT) by the formation of a dithiolene group, catalyzed by a hetero-tetrameric MPT–synthase complex, consisting of two larger Cnx6 and two smaller Cnx7 subunits (Llamas et al., 2011). Following its enzymatic activity, the Cnx7 subunits of MPT synthase must be replenished with sulfur by the MPT synthase sulfurase Cnx5. MPT activation and Mo insertion are the last two steps, which are catalyzed by the two different domains of the actin-associated molybdenum insertase Cnx1 (Llamas et al., 2011). The Cnx1 G-domain (Cnx1G) first adenylates MPT to form activated MPT-AMP, and then the MoCo complex is formed by the insertion of Mo and cleavage of AMP by the Cnx1 E-domain (Cnx1E), finally generating a complete MoCo (Llamas et al., 2011).

MoCo is a highly oxygen-sensitive complex. In the green algae *Chlamydomonas*

reinhardtii and *Volvox carteri*, and the cyanobacterium *Rippkaea orientalis*, a MoCo-binding protein that lacked direct enzymatic activity, named MCP (MoCo C carrier protein), has been suggested to protect MoCo from cellular oxygen in aerobic species (Kruse, 2022; Krausze et al., 2020; Hercher et al., 2020). However, whether MCP participates in the transfer of MoCo to molybdenum-containing enzymes, or merely acts as a cellular reservoir to buffer the supply of MoCo is still unclear (Kruse, 2022; Krausze et al., 2020; Hercher et al., 2020). In higher plants, bimolecular fluorescence complementation experiments in *A. thaliana* suggest that Cnx1 interacts directly with target Mo-containing enzymes to insert MoCo (Kaufholdt et al., 2016). A secondary protein family, previously identified as MoCo binding proteins (MoBPs) in *A. thaliana*, has subsequently been shown to function as a cytoplasmic riboside 5'-monophosphate phosphoribohydrolase, and its direct functions in MoCo insertion remain debated (Kruse et al., 2010; Kuroha et al., 2009).

FeMoCo is a distinct molecule from MoCo and has only been found to date the bacterial enzyme nitrogenase. Due to the scope of this thesis, I will not delve into its detailed synthesis. In general, it is composed of two partial cubanes ($MoFe_3S_3$ and Fe_4S_3) connected by three bridging sulfurs, which are bound by homocitrate, with a bridging central carbon atom identified by recent studies (Schwarz and Mendel, 2006; Schwarz et al., 2009).

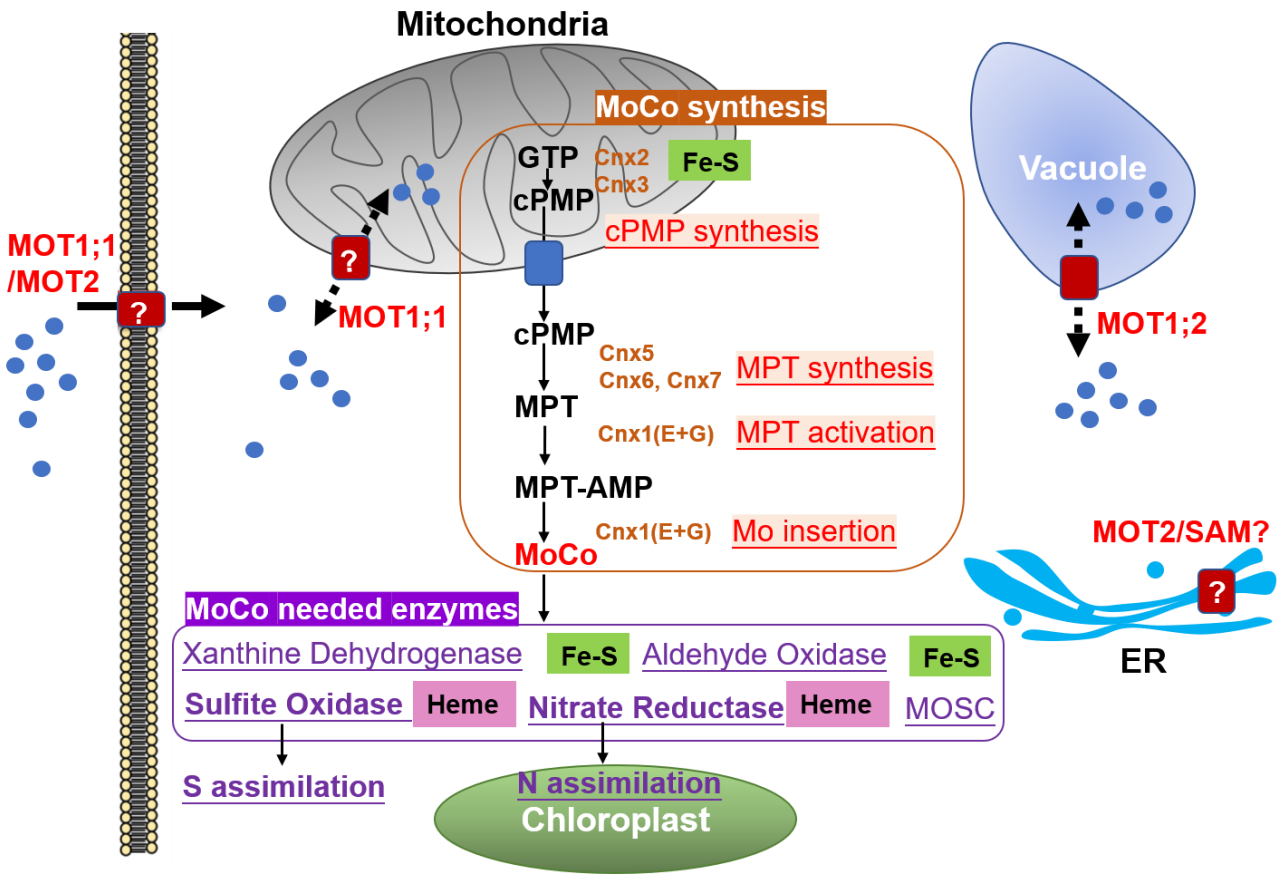

Figure 5.1 Overview of MO transport, MoCo synthesis and usage in the cell. This figure illustrates the four steps involved in MoCo biosynthesis, starting from GTP in mitochondria and leading to the formation of MoCo in cytoplasm. Proteins and enzymes catalyzing individual steps are depicted in the figure. The figure further presents the MoCo-containing enzymes (purple), alongside the presence of either Fe–S (green), and or heme (pink) cofactors in each protein. MOSC stands for: the C-terminal domain of eukaryotic MoCo sulfurases. Nitrate reductase is linked to N (nitrogen) assimilation in the chloroplast and the sulfite oxidase is linked to the following S (sulfur) assimilation. Potential MOT distributions are also shown as red box according to proteins identified in higher plants and *Chlamydomonas reinhardtii* (Tejada-Jimenez et al., 2007; Tejada-Jimenez et al., 2013; Huang et al., 2022).

5.1.3 Molybdoenzymes and their functions

After their synthesis, MoCo complexes are incorporated into multiple downstream enzymes. Molybdoenzymes exist in nearly all living organisms except *Saccharomyces* and *Thermophilic anaerobe* (Zhang and Gladyshev, 2008). More than 40 different molybdoenzymes has been recorded, but most of them are found in prokaryotes, and only five of them to date have been identified in

plants (Llamas et al., 2011; Tejada-Jime´nez et al., 2013). These enzymes are: (a) nitrate reductase (NR), which reduces nitrate to nitrite, a key step in inorganic nitrogen assimilation. NR is one of the most important molybdoenzymes, which can link MO uptake and transport to N assimilation and the downstream Glu-Gln cycle; (b) sulfite oxidase (SO) , which oxidizes sulfite to sulfate producing hydrogen peroxide and thus has a role in radical oxygen species (ROS) production; (c) aldehyde oxidase (AO), which is related to the production of abscisic acid and auxin; (d) xanthine dehydrogenase (XDH), which catalyzes the hydroxylation of aldehydes and aromatic heterocycles in the purine degradation metabolic pathway; and (e) the MOSC (MoCo sulfurase C-terminal) domain-containing amidoxime reducing component (ARC), which catalyzes the reduction of N-hydroxylated products (Tejada-Jime´nez et al., 2013; Huang et al., 2022). These Mo-dependent enzymes participate in crucial biological processes such as nitrate assimilation, phytohormone biosynthesis, purine metabolism, and sulfite detoxification.

5.1.4 Cellular molybdate transporters (MOT)

Living organisms that synthesize molybdoenzymes take up molybdate from the external environment through an energy-dependent process. A bacterial ABC-type transporter, encoded by the modABC gene cluster, was the first identified high-affinity molybdate (MO) transport system; however, this molybdate transport complex may not be present in eukaryotes (Grunden and Shanmugam, 1997; Hollenstein, 2007).

MO transport in plants and algae is more complicated, and the full diversity of eukaryotic MO transport activities in these groups are poorly understood. At least three molybdate transport pathways in plants have been reported. First, molybdate has been reported to taken up via the non-specific activity of the sulfate transport system of higher plants. The sulfate transporter SHST1 of the legume *Stylosanthes hamata* is able to mediate molybdate uptake when expressed in yeast cells (Fitzpatrick et al., 2008). This may relate to the similar biochemical properties of molybdate to sulfate, as both are divalent anions with similar geometry and hydrogen-bonding properties (Dudev and Lim et al., 2004). However, this non-specific molybdate uptake competes with sulfate and is largely dependent on environmental sulfur concentration, i.e. is increased under

sulfate deficiency but decreased by sulfur repletion (Shinmachi et al., 2010; Maillard et al., 2016).

Alongside this non-specific molybdate uptake pathway, two molybdate specific transporters, MOT1 and MOT2 (Huang et al., 2022), have been identified in eukaryotes. The MOT1 gene is present only in the genomes of algae, mosses, and higher plants, but not in animals, while MOT2 is believed to present in algae, plants, and animals including humans (Tejada-Jimenez et al., 2007; Tejada-Jimenez et al., 2011). MOT1 and MOT2 are both predicted to have 13 transmembrane domains by AlphaFold three-dimensional (3D) predictions, but in the case of MOT1 the last two domains of the C-terminus are not fully integrated into the membrane (Huang et al., 2022). These two kinds of MOT belong to different superfamilies with not much sequence similarity to one another: MOT1 belongs to Group V of the sulfate transporter family with limited similarity (13%), while MOT2 belongs to the Major Facilitator Superfamily (MFS).

The first eukaryotic high-affinity molybdate transporter, CrMOT1, was identified in the green alga *Chlamydomonas reinhardtii* (Tejada-Jimenez et al., 2007). Soon after that, homologues of CrMOT1 were identified in higher plants including *Arabidopsis thaliana*, *Oryza sativa*, *Lotus japonica* and *Medicago truncatula* (Tomatsu et al., 2007; Baxter et al., 2008; Huang et al., 2019). MOT1 enzymes in higher plants exist in multiple copies (MOT1;1 and MOT1;2), which may have different localizations and functions in MO transport, although not all of these copies have confirmed MOT transport functions as inferred by expression in yeast cells. CrMOT1 from *C. reinhardtii*, AtMOT1;1 from *A. thaliana* and OsMOT1;1 from *Oryza sativa* expressed in yeast cells possess specific molybdate transport activity without sulfate transport activity, likely due to the lack of a STAS (sulfate transporter and anti-sigma factor antagonist) domain at the C-terminus (Tejada-Jimenez et al., 2007; Tomatsu et al., 2007; Huang et al., 2019). The subcellular localizations of AtMOT1;1 and OsMOT1;1 are still controversial, with two different reported localizations at the plasma membrane and mitochondria (Tomatsu et al., 2007, Baxter et al., 2008; Huang et al., 2019; Ishikawa et al. 2021). A second vacuolar membrane located transporter in *Arabidopsis*, AtMOT1;2, has been reported to supply the MoCo biosynthesis complex with molybdate by direct interaction with molybdenum insertase Cnx1 (Gasber et al., 2011, Minner-Meinen et al., 2022). The activity of AtMOT1;2 may

allow the controlled distribution of MO, which is otherwise sequestered in the vacuole, across different tissues and organs in *Arabidopsis*. Similarly, a second vacuolar membrane located OsMOT1;2 has also been shown to be associated with the distribution of MO between rice organs (Ishikawa et al. 2021). Finally, a MOT1 homologue named NcMOT-1 in the fungus *Neurospora crassa* has been identified to mediate molybdate export out of the vacuole into the cytoplasm (Oliphant et al., 2022).

Compared with MOT1, the MOT2 of plants has been studied in less detail. Although homologues of the MOT2 gene exist in higher plant genomes, their MO transport activity has not yet been confirmed experimentally. Knockout mutants of *Oryza sativa* OsMOT2;1 and OsMOT2;2 do not show a molybdate related phenotype, but instead show diminished cadmium accumulation, suggesting potential functions as a Cd transporter (Hibara et al., 2013; Yan et al., 2019). There are three copies of AtMOT2 in the *Arabidopsis* genome, although their MO transport activities and functions have not been confirmed (Huang et al., 2022). Recently, two of the AtMOT2 homologues have been reported to localize to the Golgi, and show putative S-Adenosyl methionine transport functions (Temple et al., 2022).

In algae, different MO transport arrangements may exist. *C. reinhardtii* CrMOT1 and CrMOT2 both act as molybdate transporters and have been shown experimentally to have essential roles in maintaining MO steady state in the cell with different kinetic properties (Tejada-Jimenez et al., 2007; Tomatsu et al., 2007; Tejada-Jimenez et al., 2011; Tejada-Jimenez et al., 2013). CrMOT2 is a specific molybdate transporter with no transport activities for sulfate as are observed in CrMOT1. CrMOT2 has a much higher Km value of 550 nM than CrMOT1 (Km 7 nM) and lower capacity (Tejada-Jimenez et al., 2013). In contrast to CrMOT1, whose expression remains unchanged under different Mo conditions and induced by NO_3^-, CrMOT2 transcription is induced by molybdate deficiency (Tejada-Jimenez et al., 2007; Tomatsu et al., 2007; Tejada-Jimenez et al., 2011). Therefore, CrMOT1 and CrMOT2 may play distinct roles in controlling Mo homeostasis in *C. reinhardtii* (Tejada-Jimenez et al., 2007; Tomatsu et al., 2007; Tejada-Jimenez et al., 2011).

In previous chapters, I have studied the growth and photophysiological

phenotypes of heterozygous Trans3 knockout (KO) lines, which suggest physiological functions related to Fe availability and nitrogen source (e.g., lower sensitivity to Fe depletion in nitrate-supplemented culture). Considering that Mo crosstalk impacts on iron and nitrogen metabolism (Bittner, 2014), and that the measured impacts of Mo deficiency on algal growth, photosynthesis, N and Fe assimilation resemble in some ways the Trans3 mutant phenotype (Chapter 4), I considered the possibility of Trans3 as a potential molybdate transporter.

5.2 Results

5.2.1 Trans3 shows structural similarity to MOT2

Previous studies have reported that MOT2 family have four conserved motifs (Tejada-Jimenez et al., 2011). Substantial similarities were observed between domains within the Trans3 protein sequence and the conserved motifs of identified MOT2 transporter proteins, although some amino acid variations exist (Figure 5.2 a).

The Trans3 protein three dimensional (3D) structure, as predicted by AlphaFold (Jumper et al., 2021; Varadi et al., 2021), was compared to that of the confirmed MOT2 transporter of *Chlamydomonas reinhardtii* (CrMOT2). The first 100 amino acid positions of Trans3 were predicted with less confidence than the conserved domain according to the Predicted Aligned Error (PAE) plot (Figure 5.2 b), which likely reflect the N-terminal plastid-targeting sequence (signal and transit peptide) in the Trans3 protein sequence (Gruber et al.,2015). The Trans3 protein 3D structure consists of a long random coil with flexible and disordered regions at the N-terminus, followed by 13 confidently predicted (confidence measurement predicted local distance difference test pLDDT > 70) α-helices in Trans3 (Figure 5.2 b), which is similar to that of CrMOT (Figure 5.2 b) and among MOT2 family proteins at large (Huang et al., 2022). Given their structural similarity, I propose that Trans3 is likely to correspond to a *P. tricornutum* homologue of MOT2.

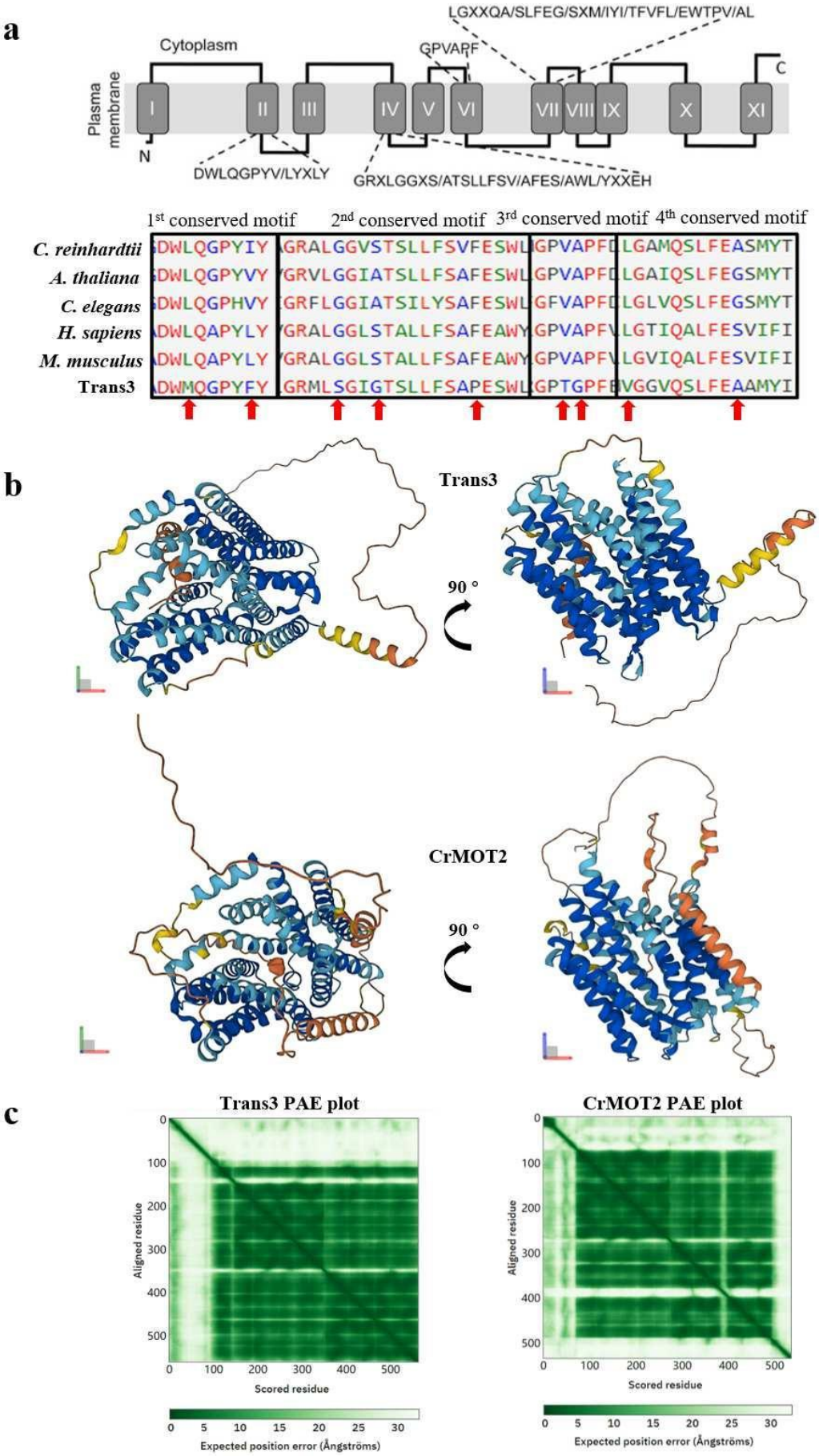
a
LGXXQA/SLFEG/SXM/IYI/TFVFL/EWTPV/AL
Cytoplasm
GPVARF
Plasma membrane
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
C
N
DWLQGPYV/LYXLY
GRXLGGXS/ATSLLFSV/AFES/AWL/YXXEH
1st conserved motif
2nd conserved motif
3rd conserved motif
4th conserved motif
C. reinhardtii
A. thaliana
C. elegans
H. sapiens
M. musculus
Trans3
b
Trans3
90 °
CrMOT2
90 °
c
Trans3 PAE plot
CrMOT2 PAE plot
Aligned residue
Scored residue
Expected position error (Ångströms)

Figure 5.2 Trans3 similarity with MOT2. a, Top: a diagram of MOT2 conserved motifs, adapted from (Tejada-Jimenez et al., 2011). Bottom: an alignment of the four conserved MOT2 motifs in Trans3 and MOT2 proteins from other species. MOT2 homologue sequences used for the alignment were: *Chlamydomonas reinhardtii*, XP_001693567.2; *Arabidopsis thaliana*, NP_567786.1; *Caenorhabditis elegans*, NP_500274; *Homo sapiens*, BAC11137.1; *Mus musculus*, NP_598861.1. **b**, 3D structure prediction comparisons of Trans3 and CrMOT2, as determined by AlphaFold (https://alphafold.ebi.ac.uk/). In the protein 3D predicted structure, residues with different confidence estimates by pLDDT (predicted local distance difference test) are represented by different colors: very high confidence (pLDDT > 90) is shown with dark blue; medium confidence (90 > pLDDT > 70) is shown with light blue; low confidence (70 > pLDDT > 50) is shown with yellow; and very low confidence (pLDDT < 50) is shown with orange. **c**, Predicted aligned error (PAE) heatmaps for each species. Heatmaps display inter-domain accuracy, with the shade of green indicating expected distance error. Dark green represents low error, i.e. well-defined relative positions and orientations; and light green/white represents high error, i.e. uncertain relative positions or orientations.

5.2.2 qPCR tests of Trans3 relative expression under different nitrogen sources and MO concentrations

Considering the structural similarity of Trans3 with MOT2 and the growth and photo-physiological phenotypes observed in different Fe depletion conditions and nitrogen sources, I next tested the relative expression of the *Phaeodactylum* Trans3 gene in empty vector control (HA) lines under different combinations of Fe, N and MO conditions using qRT-PCR. First, Trans3 mRNA relative abundances were tested under the same Fe conditions used for transcriptome sequencing, that is to say with progressive (replete, 3-day, 7-day and 14-day) Fe condition in ESAW media with normal concentrations of nitrate (5.49×10^{-4} M) and Na_2MoO_4 (6.12 nM) (Falciatore et al., 2000). Consistent with the RNAseq data (Chapter 4), Trans3 was induced > 1.5-fold under three days short-time Fe depletion (FeS) compared with Fe-replete conditions (Fe+), but was downregulated from 7 days Fe limitation (FeM) onwards (Figure 5.3 a).

Next, Trans3 relative expression levels were tested under Fe-replete conditions, varying nitrogen source (nitrate and ammonium) and MO concentration (MO-; 1×MO: 6.12 nM; 15×MO: 98.1 nM; 150×MO: 981nM). As shown in Figure 5.3 b, Trans3 relative mRNA abundances were increased 2-2.5 fold by high MO conditions (15×MO: 98.1 nM and 150×MO: 981nM), in nitrate-supplemented

and Fe-replete media. In contrast, in media supplemented with ammonium, Trans3 expression was not induced by high MO (Figure 5.3). No increase was observed in Trans3 relative mRNA abundances in cell lines incubated without supplemented nitrogen. Broadly, these data suggest that Trans3 expressed is induced by high MO concentrations, and that this induction is dependent on both Fe and NO_3^- availability. These data are broadly consistent with the observed growth rate measurements in Chapter 4 (Figure 4.6; Figure 4.7).

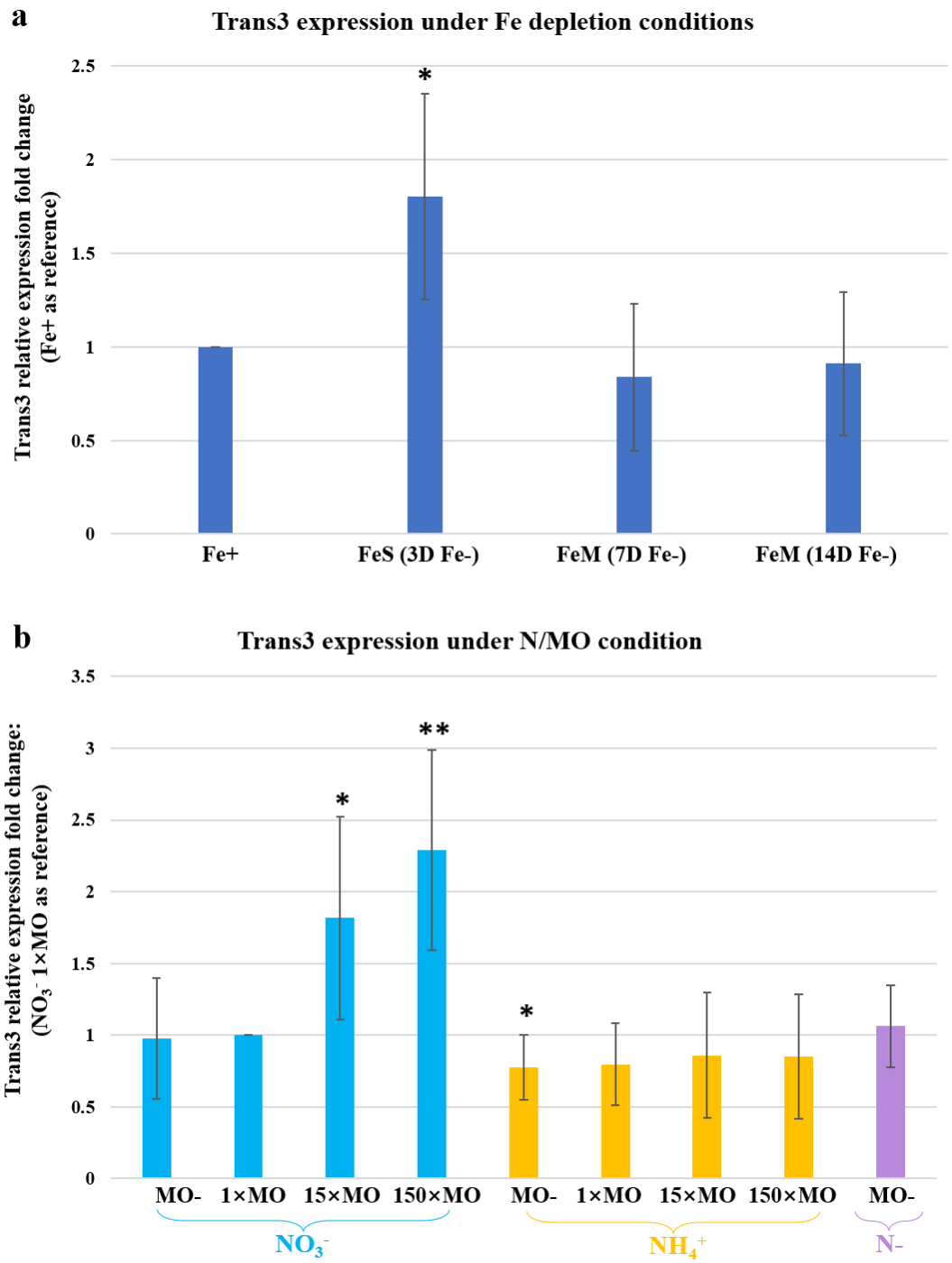

Figure 5.3 qPCR result of empty vector control line Trans3 expression under different Fe depletion/ N source/ MO concentration conditions. a, Trans3 relative expression fold-change under different Fe depletion conditions, with Fe+ expression levels as a reference. **b,** Trans3 relative expression fold-change under different N source/ MO concentration conditions, with NO_3^- 1×MO as a reference. RPS (protein ID Phatr3_J44451) and TBP (protein ID: Phatr3_J10199) were used as housekeeping genes. The $2^{-\Delta\Delta Ct}$ method was used to calculate relative gene expression values. ** *P values* ＜0.01; * *P values* ＜0.05.

5.2.3 MO uptake measurement of Trans3 KO and WT

To confirm a potential MO transport activity of Trans3, I next tested the MO uptake rates of KO lines and empty vector controls under high MO concentrations (150×MO: 981nM) in which the greatest Trans3 relative mRNA abundances were observed (Figure 5.3). A serial-dilution gradient (Figure 5.4) containing different MO concentrations was utilized as a standard to generate a standard curve of the absorbance value of the greenish-yellow Mo-dithiol complex and MO concentration in the media. The correlation coefficient R^2 of the standard curve linear equation was 0.9994, which was deemed acceptable for further analysis (Figure 5.4). MO and final diatom cell concentrations were then measured after 24 hours incubation in high molybdate media (details described in Methods) as a proxy for MO uptake rate. Interestingly, KO showed significantly and consistently greater MO uptake rates (> 1.5-fold change) than control HA lines (Figure 5.5).

Gradient-diluted media	MO-	1×MO	5×MO	15×MO	25×MO	125×MO	150×MO	625×MO
MO concentration (nM)	0	6.12	30.6	91.8	153	765	918	3825
A680 average	0.00	0.01	0.02	0.03	0.05	0.19	0.24	1.05

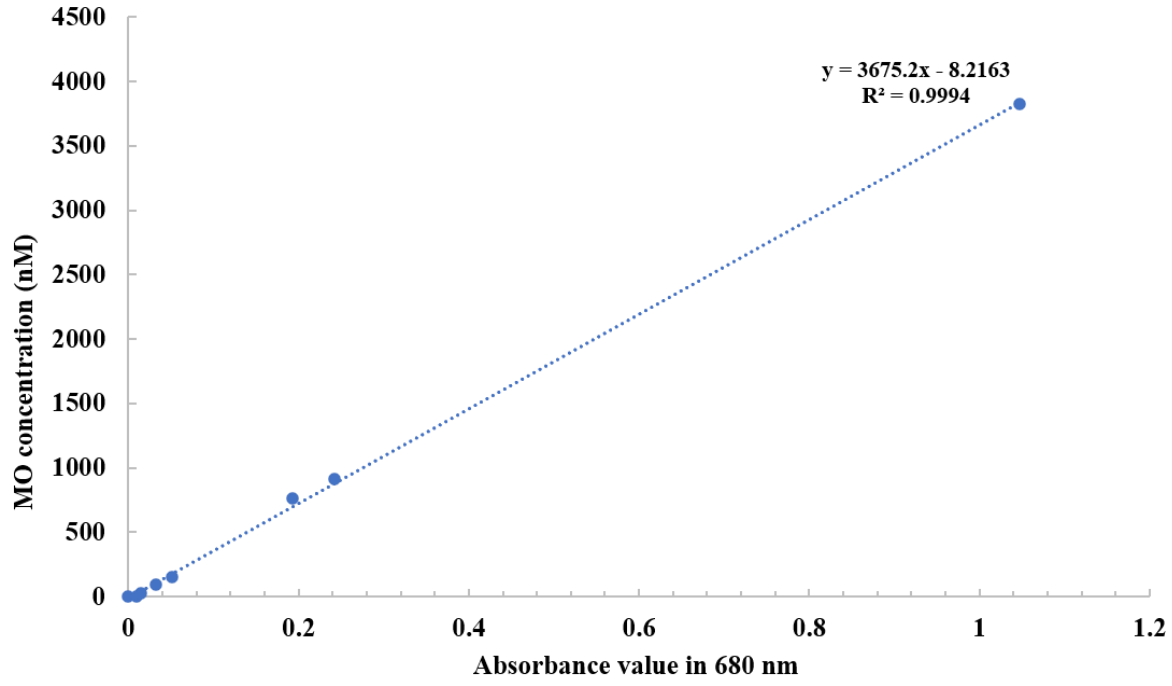

Figure 5.4 MO concentration of gradient-diluted standard solution and the standard curve for MO uptake rate measurement. The serial-dilution gradient standard solutions MO concentration for standard curve are listed above.

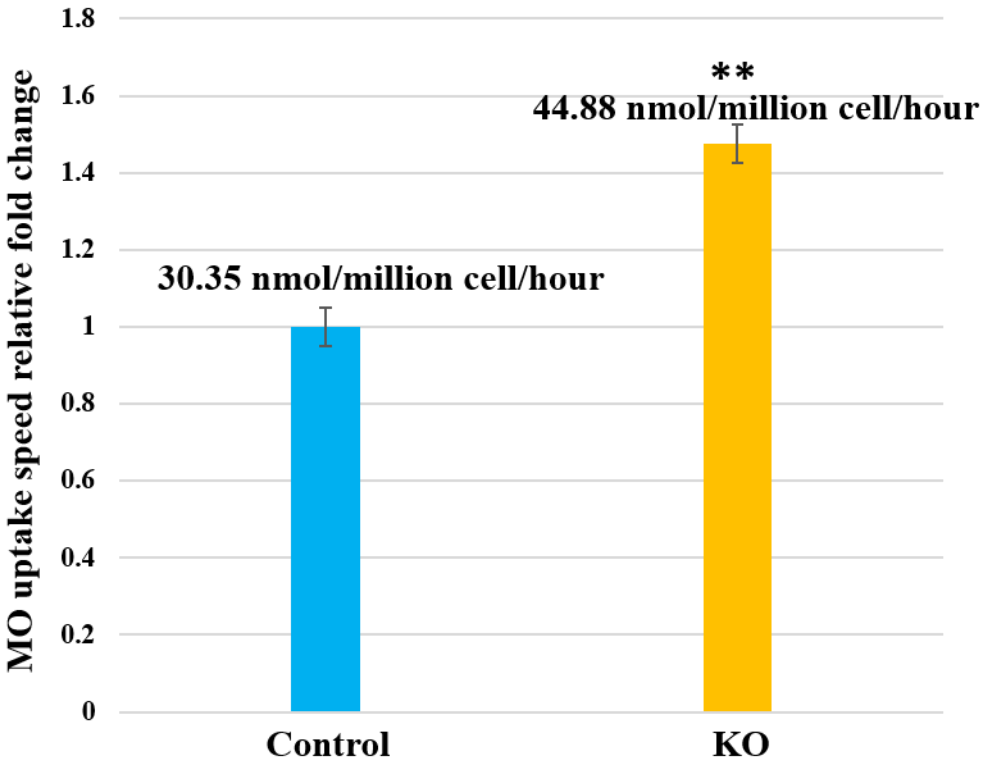

Figure 5.5 MO uptake rate comparison between Trans3 KO and control HA lines. Cell lines were incubated in Fe (8.69×10^{-6} M), NO_3^- (5.49×10^{-4} M) and high MO (918 nM) ESAW for 24 hours. MO uptake rates were measured based on the remaining MO concentration in the media after 24 hours incubation, and calculated as nmols MO uptake/ million cells/ hour. Cell concentrations were measured by flow cytometry (FCM). MO concentrations in media after incubation were measured according to the greenish-yellow Mo-dithiol complex absorbance value under 680 nm. Measurement details and a flow chart of the experiment are described in the Chapter 5 Methods.

5.2.4 Bioinformatic identification of further MOTs and Mo related pathway genes in the *Phaeodactylum* genome

Next, I searched the *P. tricornutum* genome for other potential molybdate transport and utilization genes beyond Trans3. As shown in Table 5.1, two other potential MOT2 genes beyond Trans3 were found, which were Phatr3_J16421 and Phatr3_J42677. Phatr3_J16421 and Trans3 resemble one another, and Phatr3_J16421 also possesses an inferred plastid targeting sequence, (Table 5.1). Given the defective growth, photophysiological and transcriptional phenotypes of Trans3 KO lines that nonetheless retain a complete Phatr3_J16421, it is likely to play a different role, e.g., transport of a different substrate, or transport of

MO transport across a different plastid membrane in a different direction. Phatr3_J42677 was inferred to localize *in silico* to the plasma membrane. Finally, one potential MOT1 homologue, Phatr3_J38899, was also found in the genome, likewise with predicted plasma membrane localization. These two transporters may be responsible for MO import into the cytoplasm from the outside environment (Table 5.1).

Reflecting its broad conservation across aerobic eukaryotes, a complete MoCo synthesis pathway was detected in the *P. tricornutum* genome, from GTP cyclization by Cnx2 (Phatr3_J15625) in the mitochondria to the final MO insertion and MoCo sulfuration by potential Cnx1 (Phatr3_J13395; J44961) and ABA homologues (Phatr3_J46641). The *Phaeodactylum* Cnx3 homologue (Phatr3_J10532), which should be responsible for generating cPMP in the mitochondria, lacks a predicted mitochondrial transit peptide, although this localization needs further experimental testing. Two potential homologues of *Phaeodactylum* Cnx5 (Phatr3_J20318, Phatr3_J34373), which can function as molybdopterin synthase sulfurases were identified, one of which (Phatr3_J20318) is predicted to be dual-targeted to the mitochondria and plastid (Table 5.1). For the final MO insertion step of the MoCo synthesis pathway, a probable merged Cnx1 homologue (Phatr3_J44961) containing both Cnx1G+ Cnx1E domains was found, as the situation in other eukaryotes. Additionally, another gene (Phatr3_J13395) containing only a single Cnx1G domain was also found in the genome (Table 5.1), which is rarely observed in other eukaryotes (Schwarz et al., 2000).

Putative homologues of the downstream MoCo dependent enzymes nitrate reductase (NR, Phatr3_J54983), xanthine dehydrogenase (XDH, Phatr3_J15968) and a predicted endomembrane- located sulfite oxidase (SO, Phatr3_J44482) were also found, although no credible aldehyde oxidase (AO) homologues were detected in the *P. tricornutum* genome (Table 5.1). Consistent with the absence of an AO homologue, no endogenous synthesis of auxin and abscisic acid has been identified in diatoms (Shibl et al., 2020; Chu et al., 2019; Fierli et al., 2022) (Table 5.1). Four proteins were identified that contain predicted MOSC domain, similar to the newly discovered MoCo dependent enzyme amidoxime reducing component (ARC), which catalyzes the reduction of N-hydroxylated products (Jakobs et al., 2014; Chamizo-Ampudia et al., 2016). One of these, Phatr3_J49807,

is potentially targeted to the plastid, as inferred using ASAFind and HECTAR (Table 5.1).

Table 5.1 potential MOT, MoCo synthesis and MoCo dependent enzymes

Catergory		ID in genome	Short Name	Localization
MO Transporter	MOT2	**Trans3**	**MOT2_1/Trans3**	Plastid
		Phatr3_J16421	MOT2_2	Plastid
		Phatr3_J42677	MOT2_3	Plasma membrane
	MOT1	Phatr3_J38899	MOT1	Plasma membrane
MoCo synthesis	cPMP synthesis	Phatr3_J15625	Cnx2	Mitochondria
		Phatr3_J10532	Cnx3	Other
	cPMP transport	Phatr3_J45404	ATM3	Mitochondria
	MPT synthesis	Phatr3_J9137	Cnx6	Cytoplasm
		Phatr3_EG00664	Cnx7	Other
		Phatr3_J20318	Cnx5_1	Mitochondria/Plastid
		Phatr3_J34373	Cnx5_2	Cytoplasm
	MPT activation	Phatr3_J13395	Cnx1G	Cytoplasm
	MO insertion	Phatr3_J44961	Cnx1(E+G)	Other
	MoCo sulphuration	Phatr3_J46641	ABA3	Endomembrane
MoCo dependent enzyme and MOSC conatined protein		Phatr3_J15968	XDH	Plasma membrane
		Phatr3_J54983	NR	Cytoplasm
		Phatr3_J44482	SO	Endomembrane
		Phatr3_EG01365	MOSC_1	Endomembrane
		Phatr3_J48037	MOSC_2	Endomembrane
		Phatr3_J49807	MOSC_3	Plastid
		Phatr3_J40382	MOSC_4	Cytoplasm/Nucleus

5.2.5 Mo related genes transcriptional expression in RNAseq data

Finally, I explored the relative expression levels of genes encoding putative MO uptake and molybdate-dependent proteins, along with genes implicated in nitrogen and sulfate assimilation pathways in previously generated RNAseq data (Chapter 4). I specifically looked for differences between Trans3 KO and empty vector control lines, and the impacts of Fe depletion on each searched gene.

According to the relative mRNA abundances measured by RNAseq, all of the MOT2 transporter family genes identified (Trans3; Phatr3_J16421; Phatr3_J42677) had similar expression trends, with greater relative expression levels identified under short-time Fe depletion (FeS) than Fe-replete conditions (Fe+) (Figure 5.6; Figure 5.7 a). Trans3 expression in KO lines under short-time Fe depletion surprisingly showed higher mRNA abundance than control lines

(Figure 5.6; Figure 5.7 a), although this may relate to an upregulation of the non-mutant allele in each heterozygotic mutant line. The greatest fold-changes in relative mRNA abundances (2-) were observed for the plastid-targeted MOT2-2 (Phatr3_J16421) and plasma membrane-targeted MOT2-3 (Phatr3_J42677) in KO_FeS compared to KO_Fe+ conditions. This may suggest a compensatory transcriptional induction of these genes in response to lower relative expression of Trans3 under short-time iron deprivation.

In contrast to the short-time Fe stress induction observed for MOT2, the MOT1 transporter gene Phatr3_J38899 was induced by long-time Fe depletion (FeL), with a greater increase (＞2 fold-change) in relative mRNA abundance in KO than control lines (Figure 5.6; Figure 5.7 a). MOSC-containing proteins showed limited transcriptional responses to Fe depletion and limited transcriptional differences between KO and control lines. One exception was MOSC-4, which has higher relative mRNA abundances in KO than control lines across all Fe treatment conditions. Furthemore, the potentially plastid-target MOSC-3 showed less pronounced mRNA abundance changes in KO lines when facing Fe depletion than those observed for control lines (Figure 5.6; Figure 5.7 a).

Some MoCo synthesis-related genes showed slightly higher relative mRNA abundances in KO than control lines under Fe+ and FeL conditions. These differences were augmented by short-time Fe depletion (Figure 5.6; Figure 5.7 b). Under FeS conditions, the potentially dual-targeted Cnx5-1 (Phatr3_J20318) showed greater relative mRNA abundances in KO than control lines; while ATM3 (Phatr3_J45404), the protein responsible for cPMP transport from mitochondria, showed two-fold lower relative mRNA abundances in KO compared to control lines (Figure 5.6; Figure 5.7 b). In contrast to other MoCo synthesis genes that were induced by FeS, Cnx1 (Phatr3_J44961) showed the greatest relative mRNA abundances under longer Fe depletion times, and always showed higher mRNA relative abundances in KO than control lines (fold change > 2 in most conditions) (Figure 5.6; Figure 5.7 b).

Because nitrate reductase is one of the most important MoCo-dependent enzymes, I also looked at the relative expression of genes involved in nitrogen assimilation and the downstream Glu-Gln cycle in the RNAseq data. Most of these genes showed the highest relative expression in Fe+ conditions, with

lower relative mRNA abundances according to the duration of the Fe depletion treatment, except for carbamoyl-phosphate synthase (CPSIII: Phatr3_J24195) that showed greater relative mRNA abundances (an almost 4-fold change) in all cell lines under FeS than Fe+ (Figure 5.6; Figure 5.7 b). Different relative mRNA abundances were observed between KO and control lines under Fe+ and FeS conditions, although most of the fold-changes were lower than 2 (Figure 5.6; Figure 5.7 b). Nitrogen reductase (NR: Phatr3_J54983) showed higher relative expression in KO than control lines in Fe+, but with the Fe depletion treatments, the difference between cell lines was not obvious. Interestingly, two mitochondria-targeted genes CPSIII (Phatr3_J24195) and glutamine synthetase type II (GSII-m) showed strong differential expression trends between KO and control lines. KO had higher relative expression than control lines under all of the conditions, with more than 2-fold change abundance under Fe+ (Figure 5.6; Figure 5.7 b).

I also checked the relative expression of genes involved in sulfate assimilation, considering that sulfite oxidase is another important MoCo-dependent enzyme, and that MO may be uptaken non-specifically by plasma membrane-located sulfate transporters in higher plants. The sulfate-assimilation related genes also showed greater relative mRNA abundances in KO than empty vector control lines under Fe+ conditions, but only two (sulfite oxidase-SO, Phatr3_J44482; and ferredoxin/sulfite reductase-SiR; Phatr3_J9538) showed greater than 2-fold change between KO and control lines (Figure 5.6; Figure 5.7 b).

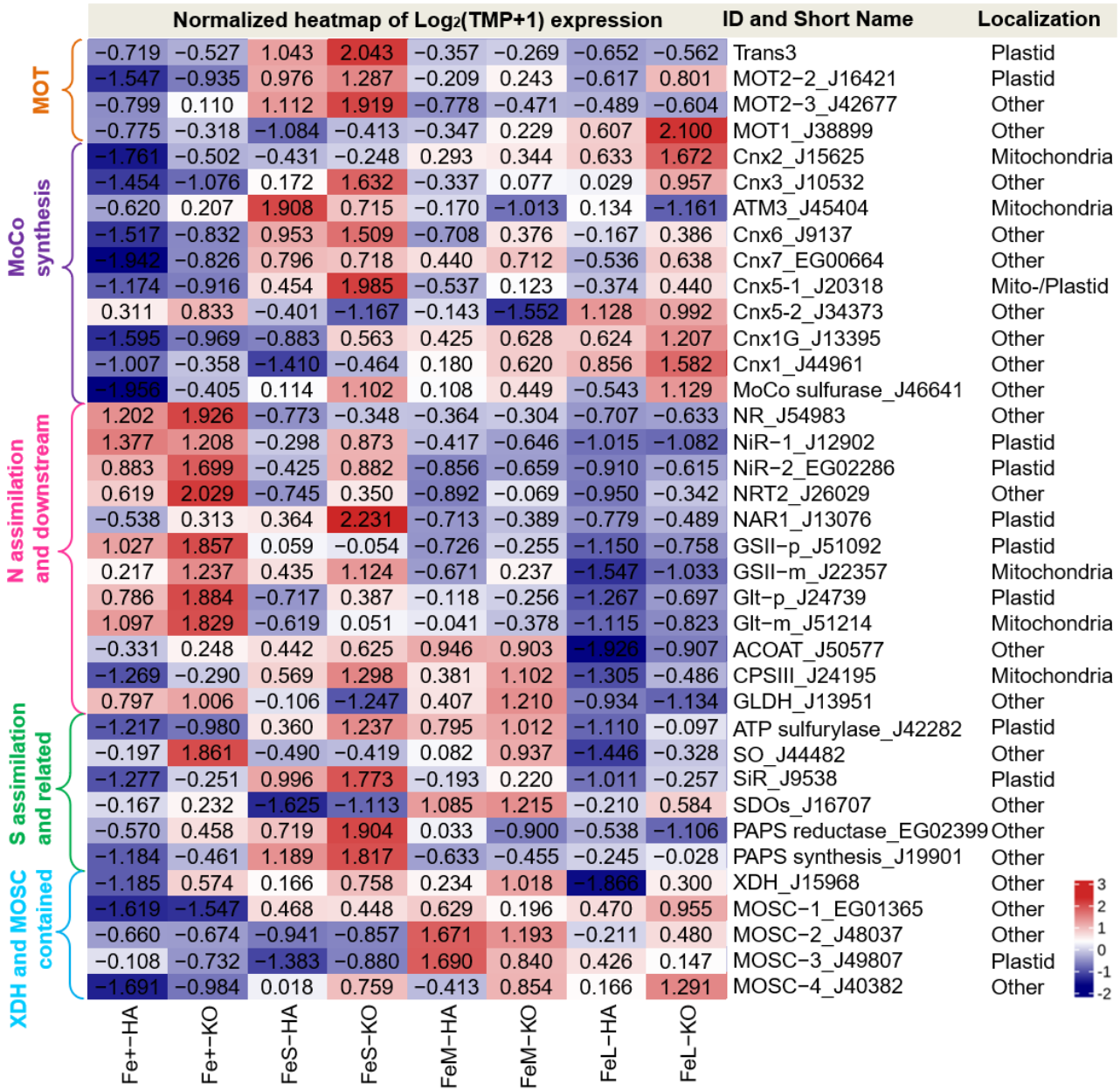

Figure 5.6 Relative mRNA abundance heatmap of Mo-related and -regulated pathway genes in each cell line and condition. N and S assimilation genes were identified from previous studies (McCarthy et al., 2017; Ide et al., 2011; Levitan et al., 2015; Van Tol et al., 2021), and other Mo-related genes were added as per Table 5.1. Relative mRNA abundances were calculated as Log_2(TPM+1) in RNAseq data from Chapter 4. Z-score method has been used for heatmap raw normalized.

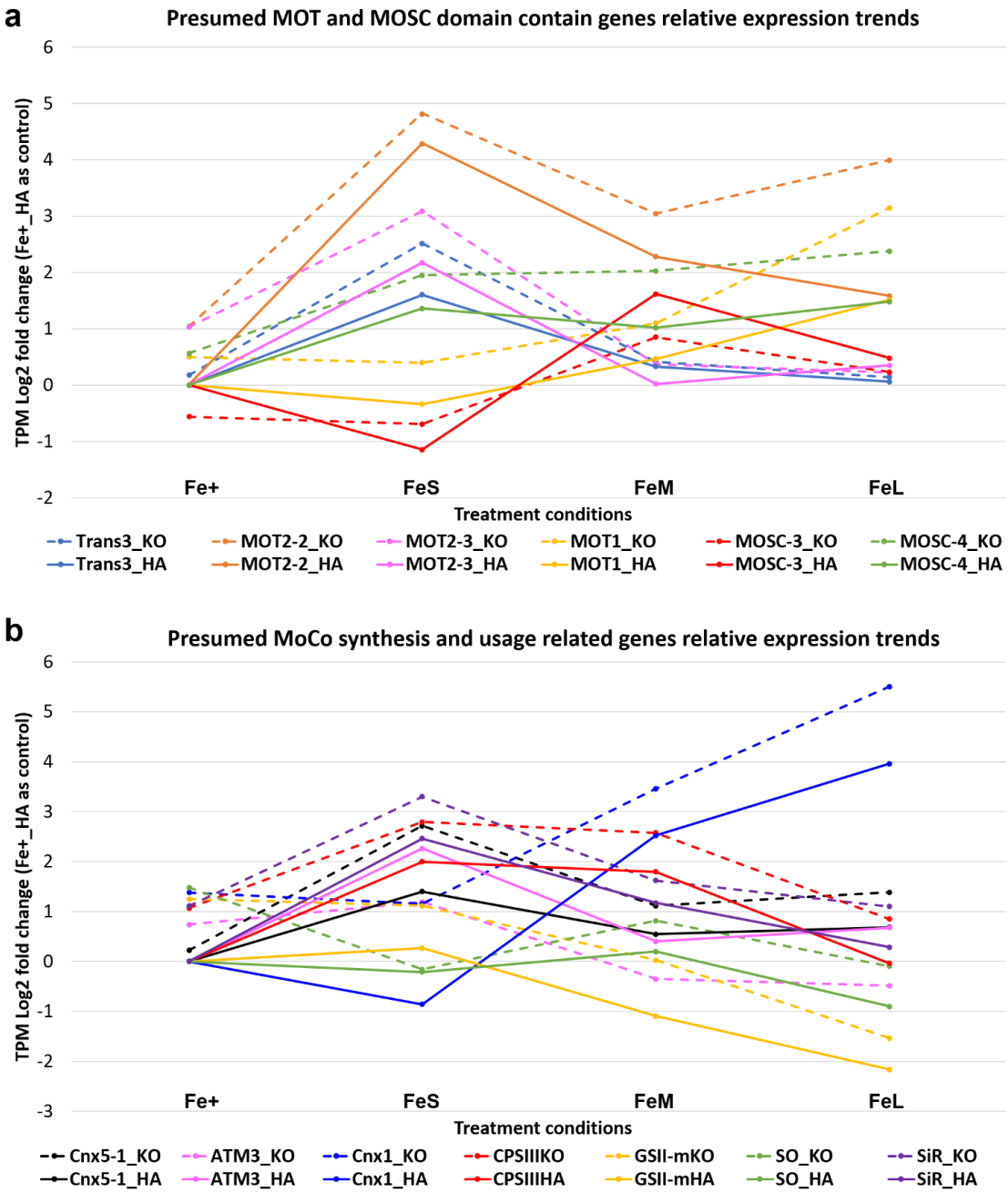

Figure 5.7 Mo-related gene relative mRNA abundance trends. In order to clearly represent gene expression trends under different Fe stress conditions, relative mRNA abundances in HA_Fe+ conditions were considered as a normalized reference. Line charts are presented by the Log_2Fold changes of the TPM value: $Log_2((TPM+1)_{Fe-\_treatment}/(TPM+1)_{HA\_Fe+})$. Different genes are represented by different colors; KO values are drawn as dotted lines and HA values are drawn as solid lines. MOT: Molybdate transporter; MOSC: MOSC domain-containing protein; ATM3: cPMP transport protein ATM3; Cnx1: molybdenum cofactor biosynthetic enzyme Cnx1; CPSIII: carbamoyl-phosphate synthase; GSII-m: glutamine synthetase type II; SO: sulfite oxidase; SiR: ferredoxin/sulfite reductase

5.3 Discussion

In this chapter, I present structural and biochemical exploration of Trans3 as a potential molybdate transporter, along with transcriptomic analysis of Mo-related genes in KO and empty vector control lines under progressive Fe starvation conditions. Based on the above characterization, I propose Trans3 functions as a potential MOT2 transporter within the plastid membranes. This would to my knowledge be the first time that a potential MOT has been associated with this organelle across the tree of life. A plastid-targeted MOT may facilitate the subcellular distribution of MO transport and utilization in diatom cells. Diatom plastids have intrinsic roles in Fe storage and buffering (Gao et al., 2021), and it is possible that they likewise function in cellular MO storage. Plastid-targeted reciprocal BLAST best hit (RbH) homologues of Trans3 were also detected in other lineages with secondary red plastids, although the activity of these homologues awaits experimental characterization.

The plastid-targeted localization of Trans3, and potentially a further MOT2 homologue (Phatr3_J16421) suggests important differences in MO transport and utilization between diatoms, higher plants and green algae. According to previous studies, high-affinity MO transport in plants can be performed by MOT1 (MOT1;1 and MOT1;2), however MOT2 activity was still undetermined through heterologous expression in yeast cells testing, although some Mo related phenotypes have been observed in MOT2 mutants. Moreover, the *Chlamydomonas* MOT2 homologue is inferred to localize to the plasma membrane and not the chloroplast, and is strongly induced by changes in MO concentration, suggesting an important accessory role in MO uptake from extracellular space (Tejada-Jimenez et al., 2007; Tejada-Jimenez et al., 2013) but not in intracellular partitioning. Within the cell, it has previously been shown that a MOT1 homologue localized to the vacuole of both higher plants and bacteria is responsible for molybdate storage and the regulation (Gasber et al., 2011, Ishikawa et al. 2021, Minner-Meinen et al., 2022, Oliphant et al., 2022). Diatoms also have a vacuole that may act as a phosphate and polysaccharide storage compartment (Huang et al., 2018; Dell'Aquila et al., 2020), and influence lipid production though a vacuolar-located H^+-ATPase (Zhang et al., 2016), but little is known about its implication in metal homeostasis.

Bioinformatic analysis suggests the presence of a complete MoCo synthesis pathway in the *P. tricornutum* cell, with the participation of both the cytoplasm and mitochondria. It appears that the active MO form (MoCo) is synthesized in the cytoplasm. However, an additional predicted molybdopterin synthase sulfurase Cnx5 (Phatr3_J20318) shows a predicted dual mitochondria/plastid localization, which to my knowledge has not previously been reported in other species. While most of the classic Mo-dependent enzymes, e.g., NR, SO, and XDH, have predicted cytoplasmic localizations, some MOSC (MOCO sulfurase C-terminal) domain-containing proteins were found potentially to localize in the plastid and endomembrane, opening up the possibility of internal membrane system usage of MO.

From qRT-PCR data, a greater relative abundance of Trans3 mRNA was identified in cells incubated in high MO and NO_3^- supplemented media, and under short-time Fe depletion. A similar transcriptional induction of the Trans3 gene was not observed in NH_4^+ media regardless of MO and Fe availability. This trend is slightly different to the expression dynamic inferred for *C. reinhardtii* MOT homologues. CrMOT1 expression remains unchanged under different Mo conditions, and is inferred to be induced by the presence of nitrate, while the transcription of CrMOT2 was induced by molybdate deficiency (Tejada-Jimenez et al., 2007; Tejada-Jimenez et al., 2013). The different transcriptional dynamics for Trans3 might reflect that it is plastid-localized, and not targeted to the plasma membrane as per CrMOT1 and CrMOT2. The induction of Trans3 by high MO concentrations may be attributed to the presence of undiscovered Mo-dependent proteins and metabolic pathways within the plastid. These mechanisms could be activated in response to elevated MO levels, necessitating the presence of plastid-localized molybdate transporters (MOT) to facilitate increased MO transport into the plastid. Alternatively, the plastid might serve as a buffer for MO, modulating Trans3 expression to maintain a balance between cytoplasmic and plastid MO concentrations.

When incubated in Fe-replete and nitrate-supplemented media, KO lines showed approximately 1.5-fold higher Mo uptake rates than empty vector controls. If Trans3 is indeed a plastidial Mo or Mo-related transporter, a decrease in its relative expression could impact on Mo uptake into the plastid. Defective MO incorporation into plastidial proteins might plausibly lead to a

compensatory induction of other proteins involved in MO uptake such as the other two presumed MOT genes located in plasma membrane (Phatr3_J42677 and Phatr3_J38899), and an overaccumulation of MO in the cytoplasm.

Some important genes involved in MOT, MoCo synthesis, N and S assimilation such as Cnx1 and Cnx5-1, GSII and CPSIII, SO and SiR showed higher relative mRNA abundances in KO lines than empty vector controls under many treatment conditions in the RNAseq data (Figure 5.7). It has been reported that the Mo depletion and NR KO may have similar influences on cell physiology because of the significant impacts of Mo scarcity on nitrogen assimilation (Ide et al., 2011). Interestingly, in a study of *P. tricornutum* NR KO cell lines (McCarthy et al., 2017), Trans3 expression showed higher relative transcriptional expression in NR KO compared with control cell lines following >48h incubation in NO_3^- -supplemented media. The transcriptional regulation of the nitrogen assimilation pathway seems to be sensitive to Fe depletion, reflecting previous studies that have shown an induction of iron stress induced proteins (ISIPs) in diatoms under N limitation conditions (McCarthy et al., 2017).

The transcriptional sensitivity of N and S metabolism to Trans3 knockout may reflect that NR and SO are molybdenum-containing enzymes, i.e. may be impacted by changes in cellular MO status. The photophysiological and growth defects observed in KO lines might likewise relate to defects in plastidial nitrogen metabolism. Inorganic N assimilation requires reducing agents (NADPH) and organic carbon skeletons (α-ketoglutarate, via glutamate) produced by photosynthesis and the tricarboxylic acid (TCA) cycle. In turn, it provides reduced amino groups required for the synthesis of proteins (especially Rubisco), light capture complexes, and photosynthetic pigments (Nunes-Nesi et al., 2010; Turpin, 1991; Falkowski and Stone, 1975). This may explain why Trans3 knockouts had lower measured photosynthetic efficiency, which might also indirectly lead to the transcriptional upregulation of pigment synthesis, LHC proteins, and photosystem subunits compared to empty vector controls.

These phenotypes notwithstanding, important questions about the cellular roles of MO in diatoms remain. Diatom MoCo synthesis takes place in the cytoplasm, and only one potential plastid MoCo-needed protein (MOSC-containing protein Phatr3_J49807) is encoded in the *P. tricornutum* genome. A potential use of MO

in the plastid is unknown, and the specific activity of Trans3 as a MO transporter remains to be confirmed experimentally. Heterologous expression of the Trans3 peptide sequence in *Saccharomyces cerevisiae*, which lacks native molybdenum uptake capabilities, or conducting mutant complementation tests (Li et al., 2016; Cheng et al., 2017) using *C. reinhardtii* MOT2 knockout lines, could serve as valuable approaches to further confirm the MO transport function of Trans3. These experimental strategies would provide additional evidence and insights into the role of Trans3 in diatom molybdenum transport.

Finally, it is important not to dismiss the possibility that Trans3 does not directly transport MO, but has another transport activity that indirectly impacts on Mo uptake in diatom cells. The recent report that a Golgi-located AtMOT2 gene functions as a S-adenosylmethionine (SAM) transporter (Temple et al., 2022), provides another direction for further Trans3 substrate confirmation. SAM is one of the most widely used enzyme substrates (Cantoni, 1975), and the MoCo synthesis enzyme Cnx2 is dependent on SAM (Llamas et al., 2011). SAM is primarily synthesized in cytoplasm, however, it plays a crucial role in chloroplasts. SAM functions as both a methyl donor, contributing to the biosynthesis of small molecules like prenyl lipids, as well as a regulator of aspartate-derived amino acid synthesis; and the methylation of Rubisco and the activity of the methyltransferases involved in tocopherol and plastoquinone synthesis also need SAM (Bouvier et al., 2006; Palmieri et al., 2006). A plastid-located SAM transporter has been reported in *Arabidopsis*, and its impaired function leads to decreased accumulation of prenyl lipids and negatively impacted on chlorophyll synthesis (Bouvier et al., 2006). Similar results were observed in the RNAseq data of KO line pigment and LHC synthesis pathways (Chapter 4). Ultimately, the exact *in situ* functions of Trans3 will be further verified by the biochemical validation of a substrate, and the identification of interacting proteins, e.g., via yeast two-hybrid analysis (Huysman et al., 2013; Huysman et al., 2015) or proximity proteomics (Turnšek et al., 2021).

5.4 Materials and methods

5.3.1 Trans3 conserved domain similarity with MOT2 and 3D prediction

The MOT2 conserved domain was defined as described in a previous study

(Tejada-Jimenez et al., 2011). First, the MOT2 domain was observed within the Trans3 gene model as recorded in EnsemblProtists (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index) and annotated by InterPro (https://www.ebi.ac.uk/interpro/), and was confirmed BLASTp within NCBI (https://www.ncbi.nlm.nih.gov/) and homologue alignment. The previous reported MOT2 homologue sequences used for the alignment were: *Chlamydomonas reinhardtii* (CrMOT2): XP_001693567.2; *Arabidopsis thaliana* (AtMOT2;1): NP_567786.1; *Caenorhabditis elegans* (CeMOT2), NP_500274; *Homo sapiens* (HsMOT2): BAC11137.1; *Mus musculus* (MmMOT2): NP_598861.1, and alignments were performed with GeneIOUS alignment (Kearse et al., 2012). Trans3 and CrMOT2 three-dimensional structure predictions were performed using AlphaFold: https://alphafold.ebi.ac.uk/ (Jumper et al., 2021; Varadi et al., 2021), and the MOT2 common structure reported previously was also considered (Huang et al., 2022).

AlphaFold generates a confidence estimate for each residue by predicted local distance difference test (pLDDT), on a scale ranging from 0 to 100. This measure quantifies the model's confidence in the predicted structure. Regions with pLDDT > 90 are anticipated to exhibit high accuracy in the model. Additionally, regions with pLDDT between 70 and 90 are expected to be well-modeled. AlphaFold produces the PAE (Predicted Aligned Error) output, which is visualized using a heatmap, and enables the assessment of inter-domain accuracy.

5.3.2 qRT-PCR of Trans3 relative mRNA abundances under different iron, molybdate and nitrate sources

Quantitative RT-PCR (qPCR) was performed using two primer pairs specific to Trans3 and two housekeeping controls (RPS, TBP) previously described in Chapter 4, and total cellular RNA from two human antigen Cas9 and zeocin-resistant co-transformed empty vector control lines (HA1, HA2) previously used for growth and physiological analysis as described in Chapter 4. qRT-PCR was performed with RNA samples from different Fe depletion conditions (Fe+, FeS, FeM, FeL) with NO_3^- (5.49 x 10^{-4} M) and MO (6.12 nM) obtained via phenol-chloroform separation and DNase treatment as previously described in Chapter 4.

HA lines (HA1 and HA2) were also treated with different combinations of nitrogen source (NO_3^- or NH_4^+) and Mo concentration (MO-: 0 nM, 1×MO: 6.12 nM, 15×MO: 91.8 nM, 150×MO: 918 nM). HA and Trans3 KO cell lines were set up with starting concentrations of 100,000 cells/mL, and cultured to exponential phase in ESAW media supplemented with NH_4^+ as nitrogen source, Fe+, and MO-, in order to repress the expression of potential MOT1 and MOT2 transport systems (Tejada-Jimenez et al., 2011). The concentrations of harvested cell lines were around 1-2 million cells mL^{-1}. Cell lines were incubated in ESAW media without N, Fe, or MO for 48 hours to remove the possibility that Fe, MO or different N sources impact on measured uptake rates. Different nitrate sources (NO_3^- or NH_4^+, 5.49×10^{-4} M), Fe (8.69×10^{-6} M) and Na_2MoO_4 (0/ 6.12/ 91.8/ 918 nM) were added to cell cultures. Cell lines were incubated for 3 hours, a duration of time previously inferred not to lead to the induction of MOT1 in *C. reinhardtii* (Tejada-Jimenez et al., 2011), to specifically test the relative expression of Trans3 mRNA. 50 mL Cell lines were collected for qPCR, and washed three times by mPBS, with three biological replicates. Cell line culture temperature, light, and shaking conditions were previously described in Chapter 4. RNA extraction and qPCR methods were the same as described in Chapters 3 and Chapter 4.

5.3.3 MO uptake incubate conditions and measurement

A flow diagram for MO uptake rate measurements is shown in Figure 5.8. Cell lines were cultured as described in Section 5.4.1 before 48 hours incubation in ESAW media without N, Fe, or MO. After that, different nitrate sources (NO_3^- or NH_4^+), Fe (8.69×10^{-6} M) and high Na_2MoO_4 concentrations (918 nM) were added to 50 mL cell cultures, and incubated 24 hours in 50 mL falcon tubes on a 100 rpm shaker. Final cell concentrations for each line were measured via flow cytometry (FCM). Cell pellets were collected by centrifugation at 4000 rpm for 30 mins. After centrifugation, the supernatant was passed through using a sterile 0.45 μm filter to remove any remaining cells and stored at -20 ℃ prior to measurement of molybdate concentrations. Cell pellets were washed three times by 1×mPBS , snap-frozen in liquid nitrogen and stored at -80 ℃ for future study.

Molybdate (MO) uptake measurements were based on the theory that greenish-yellow molybdenum-dithiol complexes can be generated in strong sulfuric acid

(8 N) in the cold. Complexes were generated and extracted into isoamyl acetate using a modified version of a previously published protocol (Cardenas and Mortenson, 1974; Tejada-Jiménez et al., 2011 and Minner-Meinen et al., 2022).

MO test reagent was prepared using: 2% (w/v) sodium hydroxide, 2 g/L 1,2-Dimercapto-4-methylbenzene (toluene-3,4-dithiol) and 16 mL/L thioglycolic acid, in a 100 mL volume. The mixture was stirred for 1 hour. The solution of dithiol was transferred to polyethylene bottles and stored in -4 °C. Under these conditions the reagent is stable for 1 month, if a precipitate appeared upon standing the solution was filtered before use.

MO uptake rate flow diagram was shown in Figure 5.8. 10 mL samples of filtered supernatants were transferred into 50 mL polyethylene tubes. To avoid interference in final measurements, 10 mL samples were pre-extracted with 1 mL of isoamyl acetate and carefully removed to avoid any final measurement interference. Sulfuric acid was then added to make the final concentration 8 N (including the volume of the dithiol reagent, adding 3.5 mL sulfuric acid for 10 mL sample). After cooling samples on ice, 3 mL MO test reagent was added. The mixture was well mixed for 10 seconds on a vortex shaker, and allowed to stand at room temperature for 15 min to permit the complete formation of the greenish-yellow molybdenum-dithiol complex precipitate. The complex was extracted with 2 mL isoamyl acetate by shaking for 10 seconds on a vortex. Organic and aqueous phases were separated by standing at room temperature for 15 min. The isoamyl acetate layer (top layer) was carefully removed, and the absorbance was measured at 680 nm against a blank of isoamyl acetate. By using a standard curve with known MO concentrations, we calculated the MO concentration in the remaining media, and from this the MO uptake rate was calculated as: (MO starting concentration-final MO concentration)/(cell concentration*incubation time).

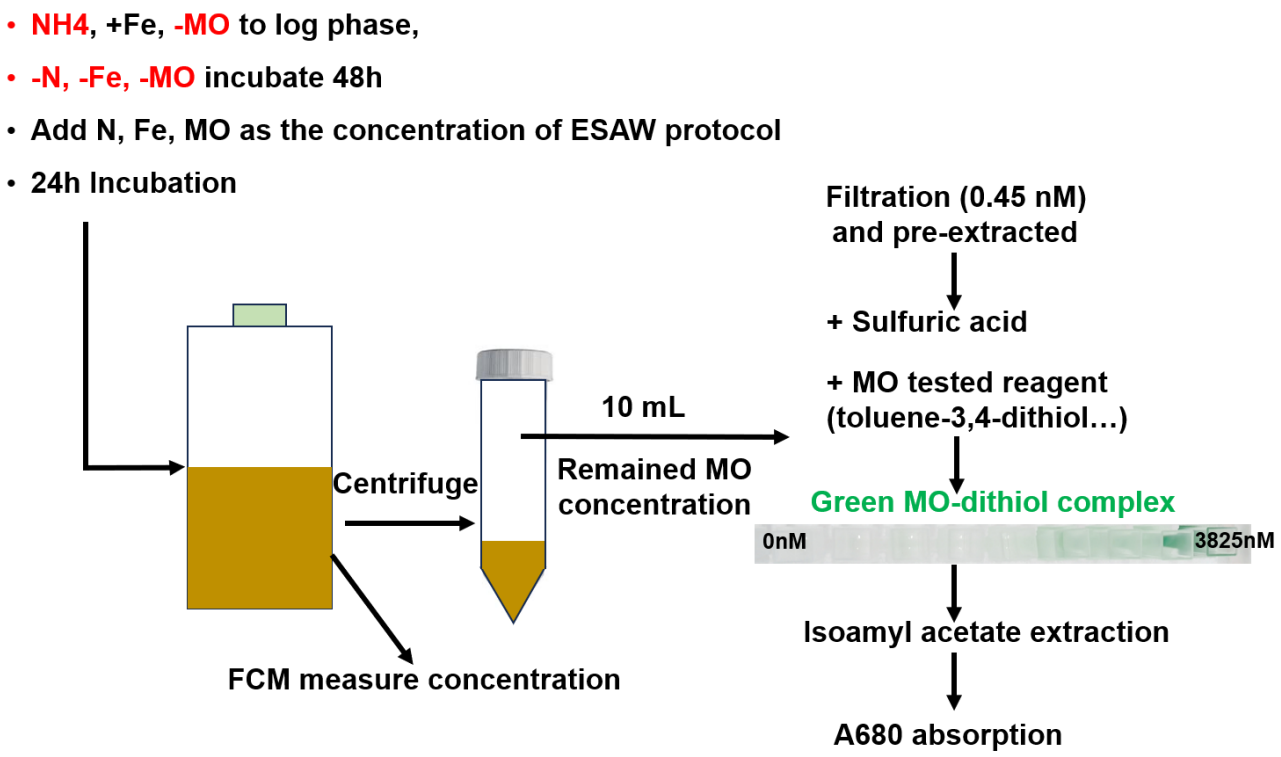

Figure 5.8 Flow diagram of MO uptake rate measurement.

5.3.4 Relative expression of Mo-related genes in the *P. tricornutum* genome

Homologues of MoCo synthesis pathway-related genes, MoCo dependent enzymes genes and MOSC domain-containing genes were annotated based on: genes in the version 3 annotation of the *P. tricornutum* genome (Rastogi et al., 2018); conserved domains and other information from EnsemblProtists (http://protists.ensembl.org/Phaeodactylum_tricornutum/Info/Index), InterPro (https://www.ebi.ac.uk/interpro/), Unipro (https://www.uniprot.org/), and BLASTp by NCBI (https://www.ncbi.nlm.nih.gov/). Mo-influenced pathways, including nitrogen assimilation, the downstream Glu-Gln cycle, and sulfate assimilation pathway genes were compiled from previous studies of nitrogen and sulfur metabolism in diatoms and higher plants (McCarthy et al., 2017; Ide et al., 2011; Levitan et al., 2015; Van Tol et al., 2021). mRNA relative expression levels were calculated from RNAseq data described in Chapter 4.

Chapter 6: Conclusion and Perspectives

6.1 Overview of thesis conclusions

Diatoms play a pivotal role in marine photosynthesis on a planetary scale. Diatoms are distantly related to plants, and possess a plastid of secondary red origin surrounded by a complex membrane structure, and with a poorly understood associated metabolism (Chapter 1). In my PhD, I have used bioinformatic and experimental approaches to better understand the unusual transporters associated with diatom plastids. This first involved using transcriptomic, phylogenomic and environmental data to construct an *in silico* atlas of predicted plastid transporters, identified by genome-wide searches of the model diatom species *Phaeodactylum tricornutum* (Chapter 2.1).

This global survey of plastid transporter diversity inspired a more detailed molecular and functional characterization of a novel MFS superfamily transporter, Trans3. Trans3 was inferred to be distributed across and unique to species containing secondary red plastids (Dorrell et al., 2021), and the underlying gene showed probable transcriptional co-regulation with genes encoded in the *Phaeodactylum* mitochondrial genome (Ashworth et al., 2016) (Chapter 2.2).

Based on self-assembling GFP localization (Chapter 3), Trans3 localizes at the PPM (peri-plastid membrane), with the most N-terminal constructs investigated oriented outside (facing the cER) and the C-terminus oriented inside (facing the PPC). Trans3 is thus likely to transport a substrate into the plastid. The Trans3 PPM localization is consistent with its evolutionary restriction to algae with secondary red plastids, as the PPM is the one unique membrane found in secondary plastids but missing from primary chloroplasts. These results show that a PPM localization, which was previously believed to be difficult to study, is possible to confirm using self-assembling GFP studies (Moog et al., 2015; Moog et al., 2020). The complete topology of Trans3 across the PPM remains to be confirmed given its multiple complex transmembrane structures and the inconsistent TMD prediction results by different tools (TMHMM, Phobius, Philius and Memsat). Nonetheless, these data reinforce previous studies (Moog et al., 2015; Moog et al., 2020; Hempel et al., 2009; Hempel et al., 2010) that suggest

self-assembling GFP is a powerful and operable strategy for identifying transporter localization and orientation across the complex diatom plastid.

Next, based on three potential heterozygous KO mutant lines generated by CRISPR-Cas9, partial suppression of Trans3 expression was found to lead to an average 20% reduction in cell growth in replete iron and nitrate-supplemented media (Chapter 4). This w as supported by a significant reduction in KO Photosystem I P700 content compared to control lines in iron replete media (Figure 4.9, Chapter 4). Considering RNAseq data (Chapter 4), KO lines further showed greater mRNA abundances of chlorophyll and terpenoid biosynthesis (mainly the plastid methylerythritol phosphate pathway related to isoprenoid precursors), photosystem and light harvesting complex-encoding genes. This implies that Trans3 plays important roles for diatom photosynthetic efficiency and cell growth.

Interestingly, when changing media from Fe-replete to Fe-depleted conditions, partial Trans3 KO lines seem to be less sensitive to Fe stress than empty vector control lines, with a smaller decrease in relative growth rate and PSI P700 content (Chapter 4). From transcriptomic data, this phenotype may relate to a relative upregulation expression of pigment synthesis and light-harvesting complex-related pathways that may mitigate the photo-sensitivity of KO lines to Fe depletion. Additionally, KO and control lines showed equivalent growth rates when grown in ammonium-supplemented media, suggesting that Trans3 function relates to the supply of nitrate as a cellular nitrogen source.

Finally, structural and biochemical data indicate a potential function of Trans3 as a molybdate transporter (Chapter 5). This relates to its structural and sequence similarity to the conserved domain of the functionally confirmed *C. reinhardtii* MOT2 gene (Tejada-Jimenez et al., 2011; Huang et al., 2022). Transcriptional expression of the Trans3 gene was induced by high molybdate concentrations and that KO lines had increased molybdate uptake compared to empty vector controls (Chapter 5). Assuming that Trans3 functions in the *Phaeodactylum* plastid, a diminished accumulation of protein levels in KO lines might diminish the efficiency of internal MO allocation in the *Phaeodactylum* cell, which could be compensated by increased MO uptake from the media.

6.2 Hypothesis: Trans3 acts as a plastid molybdate transporter

Taking all of my data together, I propose that Trans3 may act as a novel molybdate transporter in the diatom PPM. To my knowledge, this is the first time that an MOT domain-containing transporter has been located to a plastid across the tree of life. Previous localizations of plant, fungus and *C. reinhardtii* MOT1 and MOT2 position respectively within the plasma membrane, mitochondria (e.g., AtMOT1-1, albeit with inconsistent localization) and vacuole (e.g., AtMOT1-2; *Neurospora crassa* MOT), which may be responsible for the storage and regulation of cytoplasmic MO (Gasber et al., 2011, Ishikawa et al. 2021, Minner-Meinen et al., 2022, Oliphant et al., 2022). The localization of transporter proteins in primary plastid-bearing algae is notoriously difficult to predict (Tardif et al., 2012; Fuss et al., 2013), however, and an exhaustive search of plastid-targeted MOT orthologues across primary red or green algae remains to be performed.

The global phenotype of the Trans3 partial KO lines may suggest some roles in molybdate transport. Molybdenum limitation is associated with significant defects in higher plant photosynthesis, chloroplast configuration, and chlorophyll biosynthesis (Kaiser et al., 2007; Bambara and Ndakidemi et al., 2009; Ide et al., 2011; Imran et al., 2019, Imran et al., 2021). A further study of molybdenum supplementation in *Triticum aestivum* indicated that the application of molybdenum significantly improved chlorophyll content (Imran et al., 2019). These results are similar to the photo-physiological and transcriptomic patterns (e.g., upregulated chlorophyll biosynthesis) observed in Trans3 KO lines under iron and molybdate replete conditions (Chapter 4).

Some molybdate-related genes also showed higher mRNA abundance in KO than control lines as shown in Chapter 5. Genes whose activity would be impacted by defective MoCo supply, such as Cnx1, Cnx5-1 and AMT3 were differentially expressed in Trans3 KO and control lines. Both Trans3 KO and the *Arabidopsis* MOT1 KO lines show greater relative expression of genes related to the N assimilation and downstream Glu-Gln cycle, such as mitochondria-targeted carbamoyl-phosphate synthase (CPIII: Phatr3_J24195) and glutamine synthetase type II (GSII: Phatr3_J22357). Cellular glutamate levels impact on chlorophyll and heme synthesis via tetrapyrrole synthesis (McCarthy et al., 2017;

Nunes-Nesi et al., 2010; Ide et al., 2011), which might again relate to the photo-physiological phenotypes observed. Finally, sulfite oxidase (SO: Phatr3_J44482) also showed higher relative mRNA abundances in Trans3 KO lines.

Several genes implicated in MO transport and/or MoCo synthesize were also upregulated under FeS starvation conditions. Both MoCo biosynthesis (Cnx2) and the functioning of the majority of Mo enzymes are strictly dependent on Fe metabolism for providing Fe-S clusters and heme groups (Bittner, 2014). A potential crosstalk between Mo and Fe may explain the phenotypes observed under Fe depletion conditions.

Nonetheless, the specific function and substrate of Trans3 requires further experimental validation (see Discussion below). I do not rule out the possibility that Trans3 may transport other substrates (sugars or amino acids, given its other annotated features as an organic acid transporter in the MFS family) whose activity indirectly impacts on cellular MO uptake. Recently, the discovery of a Golgi-located AtMOT2 gene possessed function as a S-adenosyl methionine (SAM) transporter (Temple et al., 2022) also gives another direction for further Trans3 substrate discovery.

Moreover, the functional implication of Trans3 as a plastid molybdate transporter requires careful consideration, as no plastid-targeted molybdo-proteins are yet known in plants or algae. I was able to detect a conserved MoCo synthesis pathway in the *P. tricornutum* genome, and a potential MOSC domain-containing protein (Phatr3_J49807) is predicted to localize to the plastid (Table 5.1). It remains to be determined if the diatom plastid participates as a storage and buffering organelle for MO, or if further novel Mo-dependent proteins in the diatom plastid are essential for its metabolic activity. To date, of the 40 Mo-dependent enzymes recorded, but only five have been documented in plants (Huang et al., 2022), and further uses of molybdenum in diatoms have not to my knowledge been documented.

6.3 Perspectives and future research directions

First of all, since Trans3 topology is still too complex to be fully confirmed by self-assembling GFP, understanding of its biochemistry might be improved in

the future with a crystal structure. To determine the crystal structure of a protein, purified and crystallized the protein is subjected to X-ray diffraction analysis, producing a three-dimensional model of the protein structure, that includes the positions of the atoms, bond lengths, and angles between the bonds (Smyth et al., 2000). X-ray structures may provide valuable insights into protein folding, function, and interactions with other molecules, and have previously been used successfully to characterize novel diatom proteins (Wang et al., 2020), algal chloroplast enzymes (Le Moigne et al., 2022), and plant chloroplast transporter subunits (Tsai et al., 2013). Alongside crystal structure, immuno-electron microscopy (immuno-EM) (Herman et al., 2000) or cryogenic electron microscopy (cryo-EM) (Bai et al., 2015) may be also good choices for future studies of Trans3 localization and topology.

Considering the Trans3 KO lines obtained are probably heterozygous, i.e. partial knockouts of Trans3 function, which may functional similar with knockdown mutant, we are now trying to augment the partial knockdown observed in KO lines with a further, conditional RNAi plasmid (using an NR promoter). In particular, we are interested to isolate mutants with conditional lethality, in this case determined by an ability to grow on ammonium source media when the RNAi construct is not expressed, but non-viability on media containing a nitrate source. We intend to use these lines to more precisely test the cell growth, photosynthesis and molybdate uptake rate phenotypes in response to stricter suppression of the Trans3 gene.

If Trans3 is a potential molybdate transporter, we may be able to confirm this function by expressing the Trans3 gene in MOT lacking *Saccharomyces cerevisiae* cells, to see if its expression allows Mo uptake from growth media. We may also complement a commercially available *C. reinhardtii* MOT2 KO mutant (from the CLiP library, Li et al., 2019) with the Trans3 sequence, to see if its expression can complement any previously observed phenotypes (e.g., the deficiency of molybdate uptake activity in CrMOT2 RNAi mutants, Tejada-Jiménez et al., 2011). As these experiments depend on heterologous expression, the encoded constructs should be codon-optimised for expression in the recipient organisms, and the N-terminal targeting signal may need to be further modified to allow localization to the plasma membrane. The molybdate uptake of complemented, or empty vector transformed yeast and *Chlamydomonas* cells

could be tested directly, or via nitrate reductase activity assays.

If Trans3 is found not to transport molybdate by either approach, further work should involve the identification of alternative substrates. This may involve *in silico* approaches, e.g., identifying protein-protein interactions using STRINGdb (Szklarczyk et al., 2019) (https://string-db.org/), using existing co-expression (WGCNA) data such as PhaeoNet (Ait-Mohamed et al., 2020; Langfelder et al., 2008), or using the RNAseq data from the existing KO lines (Chapter 4). Experimental identification of potential substrates can be achieved through various approaches. For instance, the identification of differentially accumulated metabolites in knockout (KO) lines, as demonstrated by Murik et al. (2019) and Dorrell et al. (2022), can offer valuable insights into protein function. Additionally, proximity proteomics techniques (Turnšek et al., 2021) enable large-scale identification of interaction proteins, and more precise experimental evidence can be obtained through yeast two-hybrid assays (Fields et al., 1994), co-immunoprecipitation (Lau et al., 2015) or the Bacterial Adenylate Cyclase-based Two-Hybrid System (BACTH) (Olson et al., 2018). Finally, phenotype microArrays (Greetham, 2014) (https://www.biolog.com/products-portfolio-overview/phenotype-microarrays-for-microbial-cells/) represents a powerful tool to evaluate nearly 2000 phenotypes of a microbial cell, including metabolite uptake, in a single experiment. By comprehensively and precisely quantifying phenotypes, we can explore other potential alternative substrates of Trans3 beyond its predicted role as a Mo transporter.

Finally, as shown in Chapter 5, another potential *Phaeodactylum* homologue of the MOT2 transporter Phatr3_J16421 also has an inferred plastid-targeting sequence, along with Phatr3_J4267 7 inferred plasma membrane located. It remains to be determined, considering eGFP and self-assembling GFP, to what membranes these transporters localize, and whether they show similar expression trends under different MO/Fe concentrations and nitrogen sources as Trans3. This will help us to understand the potential molybdate transport system in the diatom cell more completely. The absence of known plastid-targeted molybdoproteins in plants or algae, combined with a limited understanding of the composition of the diatom PPC (Flori et al., 2016; Moog et al., 2011), underlines the need to identify potential interacting partners of Trans3. A particularly important question is if there are Mo-containing proteins in the

Phaeodactylum plastid, which would validate the role of Trans3 as a putative MO transporter, and shed light on the relevance of Trans3 to diatom cells.

In the end, I cannot help but acknowledge the significant growth and development I have experienced throughout the past years. This PhD journey is challenging, especially considering the challenges of studying in a foreign country. Nevertheless, these obstacles have served to fuel my determination. This doctoral experience has imparted upon me a profound understanding of the essential traits required for a young researcher: organization, innovation, independence, and curiosity. The realization that experiments may not always unfold exactly as originally envisioned is also important, however, in these unpredictable possibilities the allure of science lies. The task of conducting functional tests for novel proteins has further reinforced to me the significance of collaboration across complementary technologies and fields. It is my aspiration to delve deeper into unraveling the underlying diatom plastid transport mechanisms, and to contribute towards effecting positive change.

References

Agrawal, S., & Striepen, B. 2010. More membranes, more proteins: complex protein import mechanisms into secondary plastids. ***Protist***, 161(5): 672-687.

Ait-Mohamed, O., Novák Vanclová, A. M. G., Joli, N., et al. 2020. PhaeoNet: A Holistic RNAseq-Based Portrait of Transcriptional Coordination in the Model Diatom *Phaeodactylum tricornutum*. ***Frontiers in Plant Science***, 11: 590949.

Allen, A. E., Dupont, C. L., Oborník, M.,et al. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. ***Nature***, 473(7346): 203-207.

Alric, J. 2010. Cyclic electron flow around photosystem I in unicellular green algae. ***Photosynthesis Research***, 106: 47-56.

Andersen, R. A. 2013. The microalgal cell. ***Handbook of Microalgal Culture: Applied Phycology and Biotechnology***, 1-20.

Andersen, R. A. 1992. Diversity of eukaryotic algae. ***Biodiversity & Conservation***, 1: 267-292.

Apt, K. E., Grossman, A., Kroth-Pancic, P. J. M., et al. 1996. Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. ***Molecular and General Genetics***, 252: 572-579.

Apt, K. E., Zaslavkaia, L., Lippmeier, J. C., et al. 2002. In vivo characterization of diatom multipartite plastid targeting signals. ***Journal of Cell Science***, 115(21): 4061-4069.

Armbrust, E. V., Berges, J. A., Bowler, C., et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. ***Science***, 306(5693): 79-86.

Arnon, D. I., & Stout, P. R. 1939. Molybdenum as an essential element for higher plants. ***Plant Physiology***, 14(3), 599.

Ashworth, J., Turkarslan, S., Harris, M., et al. 2016. Pan-transcriptomic analysis identifies coordinated and orthologous functional modules in the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. ***Marine Genomics***, 26: 21-28.

Ast, M., Gruber, A., Schmitz-Esser, S., et al. 2009. Diatom plastids depend on nucleotide import from the cytosol. ***Proceedings of the National Academy of Sciences***, 106(9): 3621-3626.

Bai, X. C., McMullan, G., & Scheres, S. H. 2015. How cryo-EM is revolutionizing structural biology. ***Trends in Biochemical Sciences***, 40(1), 49-57.

Bai, Y., Cao, T., Dautermann, O., et al. 2022. Green diatom mutants reveal an intricate biosynthetic pathway of fucoxanthin. ***National Academy of Sciences***, 119(38): e2203708119.

Bailey, S., Melis, A., Mackey, K. R., et al. 2008. Alternative photosynthetic electron flow to oxygen in marine *Synechococcus*. ***Biochimica et Biophysica Acta (BBA)-Bioenergetics***, 1777(3): 269-276.

Bailleul, B., Berne, N., Murik, O., et al. 2015. Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. ***Nature***, 524(7565): 366-369.

Baldisserotto, C., Ferroni, L., Pantaleoni, L.,et al. 2013. Comparison of photosynthesis recovery dynamics in floating leaves of Trapa natans after inhibition by manganese or molybdenum: effects on photosystem II. ***Plant Physiology and Biochemistry***, 70: 387-395.

Bambara, S. K., & Ndakidemi, P. A. 2009. Effects of Rhizobium inoculation, lime and molybdenum on photosynthesis and chlorophyll content of *Phaeseolus vulgaris L*. ***African Journal of Microbiology Research***, 3, 791–798.

Basu, S., Patil, S., Mapleson, D., et al. 2017. Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. ***New Phytologist***, 215(1): 140-156.

Baxter, I., Muthukumar, B., Park, H. C., et al. 2008. Variation in molybdenum content across broadly distributed populations of *Arabidopsis thaliana* is controlled by a mitochondrial molybdenum transporter (MOT1). ***PLoS Genetics***, 4(2): e1000004.

Bendtsen, J. D., Nielsen, H., Von Heijne, G., et al. 2004. Improved prediction of signal peptides: SignalP 3.0. ***Journal of Molecular Biology***, 340(4): 783-795.

Benoiston, A. S., Ibarbalz, F. M., Bittner, L., et al. 2017. The evolution of diatoms and their biogeochemical functions. ***Philosophical Transactions of the Royal Society B: Biological Sciences***, 372(1728).

Bhatta, H., Kong, T. K., & Rosengarten, G. 2009. Diffusion through diatom nanopores. ***Journal of Nano Research***, 7, 69-74.

Bhattacharya, D., Yoon, H. S., & Hackett, J. D. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. ***Bioessays***, 26(1): 50-60.

Biggins, J., & Bruce, D. 1985. Mechanism of the light state transition in photosynthesis. III. Kinetics of the state transition in *Porphyridium cruentum*. ***Biochimica et Biophysica Acta (BBA)-Bioenergetics***, 806(2): 230-236.

Bittner, F. 2014. Molybdenum metabolism in plants and crosstalk to iron. ***Frontiers in Plant Science***, 5: 28.

Bolte, K., Bullmann, L., Hempel, F., et al. 2009. Protein Targeting into Secondary Plastids. ***Journal of Eukaryotic Microbiology***, 56(1): 9-15.

Bouvier, F., Linka, N., Isner, J. C., et al. 2006. *Arabidopsis* SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. ***The Plant Cell***, 18(11): 3088-3105.

Bowler, C., Allen, A. E., Badger, J. H., 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. ***Nature***, 456(7219): 239-244.

Bowler, C., & Falciatore, A. 2019. *Phaeodactylum tricornutum*. ***Trend in Genetics***, 35(9): 706-707.

Boyd, P. W., Jickells, T., Law, C., et al. 2007. Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. ***Science***, 315(5812): 612-617.

Broadley, M., Brown, P., Cakmak, I., et al. 2012. Function of nutrients: micronutrients, ***Marschner's mineral nutrition of higher plants***, 191-248.

Broddrick, J. T., Du, N., Smith, S. R., et al. 2019. Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricornutum*. ***New Phytologist***, 222(3): 1364-1379.

Brownlee, C., Helliwell, K. E., Meeda, Y., et al. 2022. Regulation and integration of membrane transport in marine diatoms. ***Cell & Developmental Biology***. 134: 79-89.

Buck, J. M., Sherman, J., Bártulos, C. R., et al. 2019. Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom *Phaeodactylum tricornutum*. ***Nature Communications***, 10(1): 4167.

Bullmann, L., Haarmann, R., Mirus, O., et al. 2010. Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. ***Journal of Biological Chemistry***, 285(9): 6848-6856.

Cabantous, S., Terwilliger, T. C., & Waldo, G. S. 2005. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. ***Nature Biotechnology***, 23(1): 102-107.

Cantoni, G. L. 1975. Biological methylation: selected aspects. ***Annual Review of Biochemistry***, 44: 435-451.

Caputi, L., Carradec, Q., Eveillard, D., et al. 2019. Community‐level responses to iron availability in open ocean plankton ecosystems. ***Global Biogeochemical Cycles***, 33(3): 391-419.

Cardenas, J., & Mortenson, L. E. 1974. Determination of molybdenum and tungsten in biological materials. ***Analytical Biochemistry***, 60(2): 372-381.

Cartier, L. D. 1973. The effect of molybdenum on growth rates, photosynthetic rates and nitrogen assimilation in freshwater plankton algae. ***Electronic Theses and Dissertations***, 3889

Chamizo-Ampudia, A., Sanz-Luque, E., Llamas, Á., et al. 2016. A dual system formed by the ARC and NR molybdoenzymes mediates nitrite-dependent NO production in *Chlamydomonas*. ***Plant, Cell & Environment***, 39(10): 2097-2107.

Chapman, R. L. 2013. Algae: the world's most important "plants"—an introduction. ***Mitigation and Adaptation Strategies for Global Change***, 18: 5-12.

Chen, Z., Wang, X., Li, S., et al. 2019. Verification of the *Saccharina japonica* translocon Tic20 and its localization in the chloroplast membrane in diatoms. ***International Journal of Molecular Sciences***, 20(16): 4000.

Cheng, X., Liu, G., Ke, W., et al. 2017. Building a multipurpose insertional mutant library for forward and reverse genetics in *Chlamydomonas*. ***Plant Methods***, 13: 1-16.

Chu, J., Li, Y., Cui, Y., et al. 2019. The influences of phytohormones on triacylglycerol accumulation in an oleaginous marine diatom *Phaeodactylum tricornutum*. ***Journal of Applied Phycology***, 31: 1009-1019.

Chu, L., Gruber, A., Ast, M., et al. 2017. Shuttling of (deoxy‐) purine nucleotides between compartments of the

diatom *Phaeodactylum tricornutum*. ***New Phytologist***, 213(1): 193-205.
Cruz de Carvalho, M. H., Sun, H. X., Bowler, C., et al. 2016. Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. ***New Phytologist***, 210(2): 497-510.
De Martino, A., Bartual, A., Willis, A., et al. 2011. Physiological and molecular evidence that environmental changes elicit morphological interconversion in the model diatom *Phaeodactylum tricornutum*. ***Protist***, 162(3): 462-481.
De Riso, V., Raniello, R., Maumus, F., et al. 2009. Gene silencing in the marine diatom *Phaeodactylum tricornutum*. ***Nucleic Acids Research***, 37(14): e96-e96.
Dell'Aquila, G., Zauner, S., Heimerl, T., et al. 2020. Mobilization and cellular distribution of phosphate in the diatom *Phaeodactylum tricornutum*. ***Frontiers in Plant Science***, 11: 579.
Di Tommaso, P., Chatzou, M., Floden, E. W., et al. 2017. Nextflow enables reproducible computational workflows. ***Nature Biotechnology***, 35(4): 316-319.
Dobson, L., Reményi, I., & Tusnády, G. 2015. CCTOP: a Consensus Constrained TOPology prediction web server. ***Nucleic Acids Research***, 43(W1): W408-W412.
Dorrell, R. G., Azuma, T., Nomura, M., et al. 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. ***Proceedings of the National Academy of Sciences***, 116(14): 6914-6923.
Dorrell, R. G., & Bowler, C. 2017. Secondary Plastids of Stramenopiles. ***Advances in Botanical Research***, 84: 57-103
Dorrell, R. G., Gile, G., McCallum, G., et al. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. ***Elife***, 6.
Dorrell, R. G., Liang, Y., Gueguen, N., et al. 2022. A Plastidial Glycolytic-Gluconeogenic Switch of Mitochondrial Origin Enables Diatom Adaptations to High Latitudes. ***BioRxiv***, 2022-09
Dorrell, R. G., & Smith, A. G. 2011. Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. ***Eukaryotic Cell***, 10(7): 856-868.
Dorrell, R. G., 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(4).
Dudev, T., & Lim, C. 2004. Oxyanion selectivity in sulfate and molybdate transport proteins: an ab initio/CDM study. ***Journal of the American Chemical Society***, 126(33): 10296-10305.
Falciatore, A., Casotti, R., Leblanc, C., et al. 1999. Transformation of nonselectable reporter genes in marine diatoms. ***Marine Biotechnology***, 1: 239-251.
Falciatore, A., d'Alcalà, M. R., Croot, P., et al. 2000. Perception of environmental signals by a marine diatom. ***Science***, 288(5475): 2363-2366.
Falciatore, A., Jaubert, M., Bouly, J. P., et al. 2020. Diatom molecular research comes of age: model species for studying phytoplankton biology and diversity. ***Plant Cell***, 32(3): 547-572.
Falkowski, P., & Stone, D. 1975. Nitrate uptake in marine phytoplankton: energy sources and the interaction with carbon fixation. ***Marine Biology***, 32: 77-84.
Felsner, G., Sommer, M. S., & Maier, U. 2010. The physical and functional borders of transit peptide-like sequences in secondary endosymbionts. ***BMC Plant Biology***, 10(1): 1-10.
Field, C. B., Behrenfeld, M. J., Randerson, J. T., et al. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. ***Science***, 281(5374): 237-240.
Fields, S., & Sternglanz, R. 1994. The two-hybrid system: an assay for protein-protein interactions. ***Trends in Genetics***, 10(8): 286-292.
Fierli, D., Barone, M. E., Graceffa, V., et al. 2022. Cold stress combined with salt or abscisic acid supplementation enhances lipogenesis and carotenogenesis in *Phaeodactylum tricornutum* (Bacillariophyceae). ***Bioprocess and Biosystems Engineering***, 1-11.
Fitzpatrick, K. L., Tyerman, S. D., & Kaiser, B. 2008. Molybdate transport through the plant sulfate transporter SHST1. ***FEBS Letters***, 582(10): 1508-1513.
Flori, S., Jouneau, P. H., Bailleul, B., et al. 2017. Plastid thylakoid architecture optimizes photosynthesis in diatoms. ***Nature Communications***, 8: 15885.

Fuss, J., Liegmann, O., Krause, K., et al. 2013. Green targeting predictor and ambiguous targeting predictor 2: the pitfalls of plant protein targeting prediction and of transient protein expression in heterologous systems. ***New Phytologist***, 200(4): 1022-1033.

Galachyants, Y. P., Zakharova, Y. R., 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

Gao, X., Bowler, C., & Kazamia, E. 2021. Iron metabolism strategies in diatoms. ***Journal of Experimental Botany***, 72(6): 2165-2180.

Gasber, A., Klaumann, S., Trentmann, O., et al. 2011. Identification of an *Arabidopsis* solute carrier critical for intracellular transport and inter-organ allocation of molybdate. ***Plant Biology***, 13(5): 710-718.

Gil‐Díez, P., Tejada‐Jiménez, M., León‐Mediavilla, J., et al. 2019. MtMOT1. 2 is responsible for molybdate supply to Medicago truncatula nodules. ***Plant, Cell & Environment,*** 42(1): 310-320.

Gile, G. H., Moog, D., Slamovits, C. H., et al. 2015. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. ***Genome Biology and Evolution***, 7(6): 1728-1742.

Gould, S. B., Sommer, M. S., Kroth, P. G., et al. 2006. Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. ***Molecular Biology and Evolution***, 23(12):2413-2422.

Görlich, S., Pawolski, D., Zlotnikov, I., et al. 2019. Control of biosilica morphology and mechanical performance by the conserved diatom gene *Silicanin-1*. ***Communications Biology,*** 2(1): 245.

Green, B. R. 2011. Chloroplast genomes of photosynthetic eukaryotes. ***The Plant Journal***, 66(1): 34-44.

Greetham, D. 2014. Phenotype microarray technology and its application in industrial biotechnology. ***Biotechnology Letters***, 36: 1153-1160.

Grouneva, I., Gollan, P. J., Kangasjärvi, S., et al. 2013. Phylogenetic viewpoints on regulation of light harvesting and electron transport in eukaryotic photosynthetic organisms. ***Planta***, 237(2): 399-412.

Gruber, A., & Kroth, P. G. 2017. Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research. ***Philosophical Transactions of the Royal Society B: Biological Sciences***, 372(1728).

Gruber, A., McKay, C., Rocap, G., et al. 2023. Comparison of different versions of SignalP and TargetP for diatom plastid protein predictions with ASAFind. ***ArXiv Preprint ArXiv***, 2303.02509.

Gruber, A., Rocap, G., Kroth, P. G., et al. 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. ***The Plant Journal***, 81(3): 519-528.

Gruber, A., Vugrinec, S., Hempel, F., et al. 2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. ***Plant Molecular Biology***, 64: 519-530.

Grunden, A. M., & Shanmugam, K. J. 1997. Molybdate transport and regulation in bacteria. ***Archives of Microbiology***, 168: 345-354.

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

Gundermann, K., & Büchel, C. 2014. Structure and functional heterogeneity of fucoxanthin-chlorophyll proteins in diatoms. ***The Structural Basis of Biological Energy Generation***, 21-37.

Gueguen, N. 2022. Characterisation of monogalactosyldiacylglycerol synthases in the model diatom *Phaeodactylum tricornutum*. ***Doctoral dissertation, Université Grenoble Alpes***.

Haimovich‐Dayan, M., Garfinkel, N., Ewe, D., et al. 2013. The role of C 4 metabolism in the marine diatom *Phaeodactylum tricornutum*. ***New Phytologist***, 197(1): 177-185.

Hempel, F., Bullmann, L., Lau, J., et al. 2009. ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. ***Molecular Biology and Evolution***, 26(8): 1781-1790.

Hempel, F., Felsner, G., & Maier, U. 2010. New mechanistic insights into pre‐protein transport across the second outermost plastid membrane of diatoms. ***Molecular Microbiology***, 76(3): 793-801.

Hercher, T. W., Krausze, J., Yang, J., et al. 2020. Identification and characterisation of the volvox carteri Moco carrier protein. ***Bioscience Reports***, 40(11).

Herman, E. M. 2000. Electron microscopic immunogold localization. ***Methods in Plant Electron Microscopy and Cytochemistry***, 247-262.

Hille, R. 2002. Molybdenum and tungsten in biology. ***Trends in Biochemical Sciences***, 27(7): 360-367.
Hirakawa, Y., Nagamune, K., & Ishida, K. 2009. Protein targeting into secondary plastids of chlorarachniophytes. ***Proceedings of the National Academy of Sciences***, 106(31): 12820-12825.
Hoguin, A., Rastogi, A., Bowler, C., et al. 2021. Genome-wide analysis of allele-specific expression of genes in the model diatom *Phaeodactylum tricornutum*. ***Scientific Reports***, 11(1): 2954.
Hollenstein, K. 2007. Structural studies of an archaeal ABC transporter specific for tungstate/molybdate. ***Doctoral dissertation***, ***ETH Zurich***.
Howarth, R. W., & Cole, J. J. 1985. Molybdenum availability, nitrogen limitation, and phytoplankton growth in natural waters. ***Science***, 229(4714): 653-655.
Huang, W., Haferkamp, I., Lepetit, B.,et al. 2018. Reduced vacuolar β-1, 3-glucan synthesis affects carbohydrate metabolism as well as plastid homeostasis and structure in *Phaeodactylum tricornutum*. ***Proceedings of the National Academy of Sciences***, 115(18): 4791-4796.
Huang, X. Y., Hu, D. W., & Zhao, F. J. 2022. Molybdenum: Mo re than an essential element. ***Journal of Experimental Botany***, 73(6): 1766-1774.
Huang, X. Y., Liu, H., Zhu, Y. F.,et al. 2019. Natural variation in a molybdate transporter controls grain molybdenum concentration in rice. ***New Phytologist***, 221(4): 1983-1997.
Huysman, M. J., Fortunato, A. E., Matthijs, M., et al. 2013. AUREOCHROME1a-mediated induction of the diatom-specific cyclin dsCYC2 controls the onset of cell division in diatoms (*Phaeodactylum tricornutum*). ***The Plant Cell***, 25(1): 215-228.
Huysman, M. J., Tanaka, A., Bowler, C., et al. 2015. Functional characterization of the diatom cyclin-dependent kinase A2 as a mitotic regulator reveals plant-like properties in a non-green lineage. ***BMC Plant Biology***, 15: 1-11.
Ide, Y., Kusano, M., Oikawa, A., et al. 2011. Effects of molybdenum deficiency and defects in molybdate transporter MOT1 on transcript accumulation and nitrogen/sulphur metabolism in *Arabidopsis thaliana*. ***Journal of Experimental Botany***, 62(4): 1483-1497.
Imran, M., Hu, C., Hussain, S., et al. 2019. Molybdenum-induced effects on photosynthetic efficacy of winter wheat (*Triticum aestivum L.*) under different nitrogen sources are associated with nitrogen assimilation. *Plant* ***Physiology and Biochemistry***, 141: 154-163.
Imran, M., Hussain, S., Rana, M. S., et al. 2021. Molybdenum improves 2-acetyl-1-pyrroline, grain quality traits and yield attributes in fragrant rice through efficient nitrogen assimilation under cadmium toxicity. ***Ecotoxicology and Environmental Safety***, 211: 111911.
Ishikawa, S., Hayashi, S., Tanikawa, H., et al. 2021. Tonoplast-localized OsMOT1; 2 participates in interorgan molybdate distribution in rice. ***Plant and Cell Physiology***, 62(5): 913-921.
Jakobs, H. H., Froriep, D., Havemeyer, A., et al. 2014. The mitochondrial amidoxime reducing component (mARC): involvement in metabolic reduction of N-oxides, oximes and N-hydroxyamidinohydrazones. ***ChemMedChem***, 9(10): 2381-2387.
Jarvis, P., & Robinson, C. J. 2004. Mechanisms of protein import and routing in chloroplasts. ***Current Biology***, 14(24): R1064-R1077.
Jensen, E. L., Clement, R., Kosta, A., et al. 2019. A new widespread subclass of carbonic anhydrase in marine phytoplankton. ***The ISME Journal***, 13(8): 2094-2106.
Jumper, J., Evans, R., Pritzel, A., et al. 2021. Highly accurate protein structure prediction with AlphaFold. ***Nature***, 596(7873): 583-589.
Kaiser, B. N., Gridley, K. L., Ngaire Brady, J., et al. 2005. The role of molybdenum in agricultural plant production. ***Annals of Botany***, 96(5): 745-754.
Kamikawa, R., Mochizuki, T., Sakamoto, M., et al. 2022. Genome evolution of a nonparasitic secondary heterotroph, the diatom *Nitzschia putrida*. ***Science Advances***, 8(17): eabi5075.
Kamikawa, R., Moog, D., Zauner, S., et al. 2017. A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. ***Molecular Biology and Evolution***, 34(9): 2355-2366.
Kang, L. K., & Rynearson, T. 2019. Identification and expression analyses of the nitrate transporter gene (NRT 2) family among *Skeletonema* species (*Bacillariophyceae*). ***Journal of Phycology***, 55(5): 1115-1125.
Karas, B. J., Diner, R. E., Lefebvre, S. C., et al. 2015. Designer diatom episomes delivered by bacterial

conjugation. ***Nature Communications***, 6: 6925.
Karlusich, J. J. P., Nef, C., Bowler, C., & Dorrell, R. G. 2022. Biogeographical Patterns and Genomes of Aquatic Photoautotrophs. ***Blue Planet, Red and Green Photosynthesis: Productivity and Carbon Cycling in Aquatic Ecosystems***.
Kaufholdt, D., Baillie, C. K., Bikker, R., et al. 2016. The molybdenum cofactor biosynthesis complex interacts with actin filaments via molybdenum insertase Cnx1 as anchor protein in *Arabidopsis thaliana*. ***Plant Science***, 244: 8-18.
Kazamia, E., Sutak, R., Paz-Yepes, J., et al. 2018. Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. ***Science Advance***, 4(5): eaar4536.
Kearse, M., Moir, R., Wilson, A., et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. ***Bioinformatics***, 28(12): 1647-1649.
Keeling, P. J., Burki, F., Wilcox, H. M., 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(6): e1001889.
Kikutani, S., Nakajima, K., Nagasato, C., et al. 2016. Thylakoid luminal θ-carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. ***Proceedings of the National Academy of Sciences***, 113(35): 9828-9833.
Kilian, O., & Kroth, P. G. 2005. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. ***The Plant Journal***, 41(2): 175-183.
Krausze, J., Hercher, T. W., Archna, A., et al. 2020. The structure of the Moco carrier protein from *Rippkaea orientalis*. ***Structural Biology Communications***, 76(9): 453-463.
Krogh, A., Larsson, B., Von Heijne, G., et al. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. ***Journal of Molecular Biology***, 305(3): 567-580.
Kroth, P. G., Chiovitti, A., Gruber, A., et al. 2008. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. ***PloS One***, 3(1): e1426.
Kruse, T., Gehl, C., Geisler, M., et al. 2010. Identification and biochemical characterization of molybdenum cofactor-binding proteins from *Arabidopsis thaliana*. ***Journal of Biological Chemistry***, 285(9): 6623-6635.
Kruse, T. 2022. Moco Carrier and Binding Proteins. ***Molecules***, 27(19): 6571.
Kuczynska, P., Jemiola-Rzeminska, M., & Strzalka, K. 2015. Photosynthetic Pigments in Diatoms. ***Marine Drugs***, 13(9): 5847-5881.
Kuroha, T., Tokunaga, H., Kojima, M., et al. 2009. Functional analyses of LONELY GUY cytokinin-activating enzymes reveal the importance of the direct activation pathway in *Arabidopsis*. ***The Plant Cell***, 21(10): 3152-3169.
Lampe, R. H., Hernandez, G., Lin, Y. Y., et al. 2021. Representative diatom and coccolithophore species exhibit divergent responses throughout simulated upwelling cycles. ***MSystems***, 6(2).
Langfelder, P., & Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. ***BMC Bioinformatics***, 9(1): 1-13.
Lau, J. B., Stork, S., Moog, D., et al. 2016. Protein–protein interactions indicate composition of a 480 kDa SELMA complex in the second outermost membrane of diatom complex plastids. ***Molecular Microbiology***, 100(1): 76-89.
Lau, J. B., Stork, S., Moog, D., et al. 2015. N‐terminal lysines are essential for protein translocation via a modified ERAD system in complex plastids. ***Molecular Microbiology***, 96(3): 609-620.
Le Moigne, T., Sarti, E., Nourisson, A., et al. 2022. Crystal structure of chloroplast fructose-1, 6-bisphosphate aldolase from the green alga *Chlamydomonas reinhardtii*. ***Structural Biology***, 214(3): 107873.
Leblanc, K., Queguiner, B., Diaz, F., et al. 2018. Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and carbon export. ***Nature Communications***, 9(1): 953.
Levitan, O., Dinamarca, J., Zelzion, E., et al. 2015. An RNA interference knock-down of nitrate reductase enhances lipid biosynthesis in the diatom *Phaeodactylum tricornutum*. ***The Plant Journal***, 84(5): 963-973.

Li, X., Patena, W., Fauser, F., et al., 2019. A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. ***Nature genetics***, 51(4): 627-635.

Li, X., Zhang, R., Patena, W., et al. 2016. An indexed, mapped mutant library enables reverse genetics studies of biological processes in *Chlamydomonas reinhardtii*. ***The Plant Cell*,** 28(2): 367-387.

Liu, S., Storti, M., Finazzi, G., et al. 2022. A metabolic, phylogenomic and environmental atlas of diatom plastid transporters from the model species *Phaeodactylum*. ***Frontiers in Plant Science***, 3640.

Liu, X., Hempel, F., Stork, S., et al. 2016. Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. ***Algal Research***, 20: 249-257.

Liu, Y., Kociolek, J. P., & Wang, Q. 2013. Six New Species of Gomphonema Ehrenberg (Bacillariophyceae) Species from the Great Xing'an Mountains, Northeastern China. ***Cryptogamie Algologie***, 34(4): 301-324.

Llamas, Á., Tejada-Jiménez, M., Fernández, E., et al. 2011. Molybdenum metabolism in the alga *Chlamydomonas* stands at the crossroad of those in *Arabidopsis* and humans. ***Metallomics***, 3(6): 578-590.

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(7): 1-21.

Love, M., Anders, S., & Huber, W. 2014. Differential analysis of count data–the DESeq2 package. ***Genome Biology***, 5(550), 10-1186.

Ma, Y. H., Wang, X., Niu, Y. F. et al. 2014. Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom *Phaeodactylum tricornutum*.***Microbial Cell Factories***, 13: 1-9.

Madeira, F., Pearce, M., Tivey, A. R., et al. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. ***Nucleic Acids Research***, 50(W1): W276-W279.

Maillard, A., Etienne, P., Diquélou, S., et al. 2016. Nutrient deficiencies modify the ionomic composition of plant tissues: a focus on cross-talk between molybdenum and other nutrients in *Brassica napus*. ***Journal of Experimental Botany***, 67(19): 5631-5641.

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***, 113(11): E1516-1525.

Marchand, J., Heydarizadeh, P., Schoefs, B., et al. 2018. Ion and metabolite transport in the chloroplast of algae: lessons from land plants. ***Cellular and Molecular Life Sciences***, 75(12): 2153-2176.

Marchand, J., Heydarizadeh, P., Schoefs, B., et al. 2020. Chloroplast ion and metabolite transport in algae. ***Photosynthesis in Algae: Biochemical and Physiological Mechanisms***, 107-139.

Marter, P., Schmidt, S., Kiontke, S., et al. 2020. Optimized mRuby3 is a suitable fluorescent protein for in vivo co-localization studies with GFP in the diatom *Phaeodactylum tricornutum*. ***Protist***, 171(1): 125715.

Maxwell, K., & Johnson, G. N. 2000. Chlorophyll fluorescence--a practical guide. ***Journal of Experimental Botany***, 51(345): 659-668.

McCarthy, J. K., Smith, S. R., McCrow, J. P., et al. 2017. Nitrate Reductase Knockout Uncouples Nitrate Transport from Nitrate Assimilation and Drives Repartitioning of Carbon Flux in a Model Pennate Diatom. ***The Plant Cell***, 29(8): 2047-2070.

McQuaid, J. B., Kustka, A. B., Oborník, M., et al. 2018. Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. ***Nature***, 555(7697): 534-537.

Mendel, R. R., & Bittner, F. 2006. Cell biology of molybdenum. ***Biochimica et Biophysica Acta (BBA)-Molecular Cell Research***, 1763(7): 621-635.

Mendel, R. R. 2011. Cell biology of molybdenum in plants. ***Biochimica et Biophysica Acta (BBA)-Molecular Cell Research***, 30: 1787-1797.

Merchant, S. S., Allen, M. D., Kropat, J., et al. 2006. Between a rock and a hard place: trace element nutrition in *Chlamydomonas*. ***Biochimica et Biophysica Acta (BBA)-Molecular Cell Research***, 1763(7): 578-594.

Minner-Meinen, R., Weber, J. N., Kistner, S., et al. 2022. Physiological importance of Molybdate Transporter Family 1 in feeding the molybdenum cofactor biosynthesis pathway in *Arabidopsis thaliana*. ***Molecules***, 27(10).

Mix, A. K., Cenci, U., Heimerl, T., et al. 2018. Identification and localization of peroxisomal biogenesis proteins

indicates the presence of peroxisomes in the cryptophyte *Guillardia theta* and other "chromalveolates". ***Genome Biology and Evolution***, 10(10): 2834-2852.

Mock, T., Otillar, R. P., Strauss, J., McMullan, M., Paajanen, P., Schmutz, J., Salamov, A., Sanges, R., Toseland, A., & Ward, B. 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. ***Nature***, 541(7638): 536-540.

Moog, D., Nozawa, A., Tozawa, Y., et al. 2020. Substrate specificity of plastid phosphate transporters in a non-photosynthetic diatom and its implication in evolution of red alga-derived complex plastids. ***Scientific Reports***, 10(1): 1-12.

Moog, D., Rensing, S. A., Archibald, J. M., et al. 2015. Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. ***Genome Biology and Evolution***, 7(11): 2955-2969.

Murik, O., Tirichine, L., Prihoda, J., et al. 2019. Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. ***New Phytologist***, 221(3): 1303-1316.

Nakajima, K., Tanaka, A., & Matsuda, Y. 2013. SLC4 family transporters in a marine diatom directly pump bicarbonate from seawater. ***Proceedings of the National Academy of Sciences***, 110(5): 1767-1772.

Nakov, T., Beaulieu, J. M., & Alverson, A. 2018. Insights into global planktonic diatom diversity: the importance of comparisons between phylogenetically equivalent units that account for time. ***ISME Journal***, 12(11): 2807-2810.

Nelson, D. M., Tréguer, P., Brzezinski, M. A., et al. 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): 359-372.

Nelson, N., & Junge, W. 2015. Structure and energy transfer in photosystems of oxygenic photosynthesis. ***Annual Review of Chemistry***, 84: 659-683.

Nielsen, H., Engelbrecht, J., Brunak, S., et al. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. ***Protein Engineering***, 10(1): 1-6.

Niño-González, M., Novo-Uzal, E., Richardson, D. N., et al. 2019. More transporters, more substrates: the Arabidopsis major facilitator superfamily revisited. ***Molecular Plant***, 12(9), 1182-1202.

Nonoyama, T., Kazamia, E., Nawaly, H., et al. 2019. Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. ***Biomolecules***, 9(8).

Nunes-Nesi, A., Fernie, A. R., & Stitt, M. 2010. Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions. ***Molecular Plant***, 3(6): 973-996.

Nymark, M., Sharma, A. K., Hafskjold, M. C. G., et al. 2017. CRISPR/Cas9 gene editing in the marine diatom *Phaeodactylum tricornutum*. ***Bio-Protocal***, 7(15): e2442.

Nymark, M., Valle, K. C., Brembu, T., et al. 2009. An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. ***PLoS One***, 4(11): e7743.

Nymark, M., Valle, K. C., Hancke, K., et al. 2013. Molecular and photosynthetic responses to prolonged darkness and subsequent acclimation to re-illumination in the diatom *Phaeodactylum tricornutum*. ***PLoS One***, 8(3): e58722.

Ogura, A., Akizuki, Y., Imoda, H., et al. 2018. Comparative genome and transcriptome analysis of diatom, *Skeletonema costatum*, reveals evolution of genes for harmful algal bloom. ***BMC Genomics***, 19(1): 765.

Oliphant, K. D., Rabenow, M., Hohtanz, L. et al. 2022. The *Neurospora crassa* molybdate transporter: characterizing a novel transporter homologous to the plant MOT1 family. ***Fungal Genetics and Biology***, 163: 103745.

Olson, M. G., Goldammer, M., Gauliard, E., et al. 2018. A bacterial adenylate cyclase-based two-hybrid system compatible with Gateway® cloning. ***Two-hybrid systems: methods and protocols***, 75-96.

Oudot-Le Secq, M. P., & Green, B. R. 2011. Complex repeat structures and novel features in the mitochondrial genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. ***Gene***, 476(1-2): 20-26.

Oudot-Le Secq, M. P., Grimwood, J., Shapiro, H., et al. 2007. Chloroplast genomes of the diatoms

Phaeodactylum tricornutum and *Thalassiosira pseudonana*: comparison with other plastid genomes of the red lineage. ***Molecular Genetics and Genomics***, 277(4): 427-439.

Palmieri, L., Arrigoni, R., Blanco, E., et al. 2006. Molecular identification of an *Arabidopsis* S-adenosylmethionine transporter. Analysis of organ distribution, bacterial expression, reconstitution into liposomes, and functional characterization. ***Plant Physiology***, 142(3): 855-865.

Pankowski, A., & McMinn, A. 2009. Iron availability regulates growth, photosynthesis, and production of ferredoxin and flavodoxin in Antarctic sea ice diatoms. ***Aquatic Biology***, 4(3): 273-288.

Pao, S. S., Paulsen, I. T., & Saier Jr, M. H. 1998. Major facilitator superfamily. ***Microbiology and Molecular Biology Reviews***, 62(1), 1-34.

Peers, G., & Price, N. M. 2006. Copper-containing plastocyanin used for electron transport by an oceanic diatom. ***Nature***, 441(7091): 341-344.

Peschke, M., Moog, D., Klingl, A., et al. 2013. Evidence for glycoprotein transport into complex plastids. ***Proceedings of the National Academy of Sciences,*** 110(26): 10860-10865.

Pfeil, B. E., Schoefs, B., & Spetea, C. 2014. Function and evolution of channels and transporters in photosynthetic membranes. ***Cellular and Molecular Life Sciences***, 71(6): 979-998.

Pi, X., Zhao, S., Wang, W., et al. 2019. The pigment-protein network of a diatom photosystem II–light-harvesting antenna supercomplex. ***Science***, 365(6452): eaax4406.

Prihoda, J., Tanaka, A., de Paula, W. B., et al. 2012. Chloroplast-mitochondria cross-talk in diatoms. ***Journal of Experimental Botany***, 63(4): 1543-1557.

Pyszniak, A., & Gibbs, S. P. 1992. Immunocytochemical localization of photosystem I and the fucoxanthin-chlorophyll a/c light-harvesting complex in the diatom *Phaeodactylum tricornutum*. ***Protoplasma***, 166: 208-217.

Rastogi, A. 2016. *Phaeodactylum tricornutum* genome and epigenome: characterization of natural variants. Doctoral Thesis, ***Université Paris Sciences et Lettres***.

Rastogi, A., Maheswari, U., Dorrell, R. G., 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(1): 4834.

Reinfelder, J. R., Kraepiel, A. M., & Morel, F. M. 2000. Unicellular C4 photosynthesis in a marine diatom. ***Nature***, 407(6807): 996-999.

Roberts, K., Granum, E., Leegood, R. C., et al. 2007. Carbon acquisition by diatoms. ***Photosynthesis Research***, 93: 79-88.

Rosenblad, M. A., Träger, C., Schünemann, D. et al. 2013. Structural diversity of signal recognition particle RNAs in plastids. ***Plant Signaling and Behaviour,*** 8(10): e26848.

Rueter, J. G. & Petersen, R. R. 1987. Micronutrient effects on cyanobacterial growth and physiology. ***New Zealand Journal of Marine and Freshwater Research***, 21(3): 435-445.

Sachse, M., Sturm, S., Gruber, A., et al. 2013. Identification and evaluation of endogenous reference genes for steady state transcript quantification by qPCR in the diatom *Phaeodactylum tricornutum* with constitutive expression independent from time and light. ***Journal of Endocytobiosis and Cell Research*** 24: 1-7.

Scala, S., Carels, N., Falciatore, A., et al. 2002. Genome properties of the diatom *Phaeodactylum tricornutum*. ***Plant Physiology***, 129(3): 993-1002.

Schober, A. F., Río Bártulos, C., Bischoff, A., et al. 2019. Organelle studies and proteome analyses of mitochondria and plastids fractions from the diatom *Thalassiosira pseudonana*. ***Plant and Cell Physiology***, 60(8): 1811-1828.

Schwarz, G., Mendel, R. R., & Ribbe, M. W. 2009. Molybdenum cofactors, enzymes and pathways. ***Nature***, 460(7257): 839-847.

Schwarz, G., & Mendel, R. R. 2006. Molybdenum cofactor biosynthesis and molybdenum enzymes. ***Annual Review of Plant Biology***, 57: 623-647.

Schwarz, G., Schulze, J., Bittner, F., et al. 2000. The molybdenum cofactor biosynthetic protein Cnx1 complements molybdate-repairable mutants, transfers molybdenum to the metal binding pterin, and is associated with the cytoskeleton. ***Plant Cell***, Dec;12(12):2455-2472.

Serif, M., Lepetit, B., Weißert, K., et al. 2017. A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornutum*. ***Algal Research***, 23: 186-195.

Seydoux, C., Storti, M., Giovagnetti, V., et al. 2022. Impaired photoprotection in *Phaeodactylum tricornutum* KEA3 mutants reveals the proton regulatory circuit of diatoms light acclimation. ***New Phytologist***, 234(2): 578-591.

Shai, N., Schuldiner, M., & Zalckvar, E. 2016. No peroxisome is an island - peroxisome contact sites. ***Biochimica et Biophysica Acta (BBA)-Molecular Cell Research***, 1863(5): 1061-1069.

Shibl, A. A., Isaac, A., Ochsenkuhn, M. A., et al. 2020. Diatom modulation of select bacteria through use of two unique secondary metabolites. ***Proceedings of the National Academy of Sciences***, 117(44): 27445-27455.

Shinmachi, F., Buchner, P., Stroud, J. L., et al. 2010. Influence of sulfur deficiency on the expression of specific sulfate transporters and the distribution of sulfur, selenium, and molybdenum in wheat. ***Plant Physiology***, 153(1): 327-336.

Siaut, M., Heijde, M., Mangogna, M., et al. 2007. Molecular toolbox for studying diatom biology in *Phaeodactylum tricornutum*. ***Gene***, 406(1-2): 23-35.

Sims, P. A., Mann, D. G., & Medlin, L. K. 2006. Evolution of the diatoms: insights from fossil, biological and molecular data. ***Phycological Reviews***, 45(4): 361-402.

Singer, D., Seppey, C. V., Lentendu, G., et al. 2021. Protist taxonomic and functional diversity in soil, freshwater and marine ecosystems. ***Environment International***, 146: 106262.

Smith, S. R., Dupont, C. L., McCarthy, J. K., et al. 2019. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. ***Nature Communications***, 10(1): 4552.

Smith, S. R., Gillard, J. T., Kustka, A. B., et al. 2016. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. ***PLoS Genetics***, 12(12): e1006490.

Smol, J. P. 2012. Climate Change: a planet in flux. ***Nature***, 483(7387): S12-S15.

Smyth, M. S., & Martin, J. H. 2000. X-R ay crystallography. ***Molecular Pathology***, 53(1): 8-14.

Solymosi, K. 2012. Plastid Structure, Diversification and Interconversions I. Algae. ***Current Chemical Biology***, 6(3): 167 -186.

Stiller, J. W 2014. Toward an empirical framework for interpreting plastid evolution. ***Journal of Phycology***, 50(3): 462-471.

Strassert, J. F., Irisarri, I., Williams, T. A., et al. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. ***Nature Communications***, 12(1): 1879.

Stukenberg, D., Zauner, S., Dell'Aquila, G., et al. 2018. Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. ***Frontiers in Plant Science***, 9: 740.

Sunagawa, S., Acinas, S. G., Bork, P., et al. 2020. *Tara* Oceans: towards global ocean ecosystems biology. ***Nature Reviews in Microbiology***, 18(8): 428-445.

Szklarczyk, D., Gable, A. L., Lyon, D., et al. 2019. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. ***Nucleic Acids Research,*** 47(D1): D607-D613.

Taddei, L., Stella, G. R., Rogato, A., et al. 2016. Multisignal control of expression of the LHCX protein family in the marine diatom *Phaeodactylum tricornutum*. ***Journal of Experimental Botany***, 67(13): 3939-3951.

Tanaka, A., De Martino, A., Amato, A., et al. 2015a. Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom Phaeodactylum tricornutum. ***Protist***, 166(5): 506-521.

Tanaka, T., Maeda, Y., Veluchamy, A., et al. 2015b. Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. ***Plant Cell***, 27(1): 162-176.

Tardif, M., Atteia, A., Specht, M., et al. 2012. PredAlgo: a new subcellular localization prediction tool dedicated to green algae. ***Molecular Biology and Evolution***, 29(12): 3625-3639.

Tejada-Jiménez, M., Chamizo-Ampudia, A., Galván, A., et al. 2013. Molybdenum metabolism in plants. ***Metallomics***, 5(9): 1191-1203.

Tejada-Jiménez, M., Galván, A., & Fernández, E.. 2011. Algae and humans share a molybdate transporter. ***Proceedings of the National Academy of Sciences***, 108(16): 6420-6425.

Tejada-Jiménez, M., Llamas, Á., Sanz-Luque, E., et al. 2007. A high-affinity molybdate transporter in eukaryotes. ***Proceedings of the National Academy of Sciences***, 104(50): 20126-20130.

Temple, H., Phyo, P., Yang, W., et al. 2022. Golgi-localized putative S-adenosyl methionine transporters required for plant cell wall polysaccharide methylation. ***Nature Plants***, 8(6): 656-669.

Ter Steeg, P. F., Hanson, P. J., & Paerl, H. W. 1986. Growth-limiting quantities and accumulation of molybdenum in *Anabaena oscillarioides* (Cyanobacteria). ***Hydrobiologia***, 140: 143-147.

Thomas, W., Seibert, D., Alden, M., et al. 1984. Yields, photosynthetic efficiencies and proximate composition of dense marine microalgal cultures. I. Introduction and *Phaeodactylum tricornutum* experiments. ***Biomass***, 5(3): 181-209.

Tomatsu, H., Takano, J., Takahashi, H., et al. 2007. An *Arabidopsis thaliana* high-affinity molybdate transporter required for efficient uptake of molybdate from soil. ***Proceedings of the National Academy of Sciences***, 104(47): 18807-18812.

Traller, J. C., Cokus, S. J., Lopez, D. A., et al. 2016. Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. ***Biotechnol Biofuels***, 9: 258.

Tréguer, P., Bowler, C., Moriceau, B., et al. 2018. Influence of diatom diversity on the ocean biological carbon pump. ***Nature Geoscience***, 11(1): 27-37.

Tsai, J. Y., Chu, C. C., Yeh, Y. H., et al. 2013. Structural characterizations of the chloroplast translocon protein Tic110. ***The Plant Journal***, 75(5): 847-857.

Turnšek, J., Brunson, J. K., Viedma, M., et al. 2021. Proximity proteomics in a marine diatom reveals a putative cell surface-to-chloroplast iron trafficking pathway. ***Elife***, 10, 52770.

Turpin, D. H. 1991. Effects of inorganic N availability on algal photosynthesis and carbon metabolism. ***Journal of Phycology***, 27(1): 14-20.

Ustick, L. J., Larkin, A. A., Garcia, C. A., et al. 2021. Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. ***Science***, 372(6539): 287-291.

Uwizeye, C., Decelle, J., Jouneau, P. H., et al. 2020. Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. ***Nature Communications***, 12: 1049.

Valle, K. C., Nymark, M., Aamot, I., et al. 2014. System responses to equal doses of photosynthetically usable radiation of blue, green, and red light in the marine diatom *Phaeodactylum tricornutum*. ***PLoS One***, 9(12): e114211.

Van Tol, H. M., & Armbrust, E. V. 2021. Genome-scale metabolic model of the diatom *Thalassiosira pseudonana* highlights the importance of nitrogen and sulfur metabolism in redox balance. ***PLoS One***, 16(3): e0241960.

Varadi, M., Anyango, S., Deshpande, M., et al. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. ***Nucleic Acids Research***, 50(D1): D439-D444.

Vincent, F., & Bowler, C. 2020. Diatoms are selective segregators in global ocean planktonic communities. ***mSystems***, 5(1).

Wallen, D., & Cartier, L. 1975. Molybdenum dependence, nitrate uptake and photosynthesis of freshwater plankton algae. ***Journal of Phycology***, 11(3): 345-349.

Wang, D., Aller, R. C., & Sañudo-Wilhelmy, S. A. 2009. A new method for the quantification of different redox-species of molybdenum (V and VI) in seawater. ***Marine Chemistry***, 113(3-4): 250-256.

Wang, W., Yu, L. J., Xu, C., et al. 2019. Structural basis for blue-green light harvesting and energy dissipation in diatoms. ***Science***, 363(6427): eaav0365.

Wang, W., Zhao, S., Pi, X., et al. 2020. Structural features of the diatom photosystem II–light‐harvesting antenna complex. ***FEBS Journal***, 287(11): 2191-2200.

Weyman, P. D., Beeri, K., Lefebvre, S. C., et al. 2015. Inactivation of *Phaeodactylum tricornutum urease* gene using transcription activator‐like effector nuclease‐based targeted mutagenesis. ***Plant Biotechnology Journal***, 13(4): 460-470.

Yang, J., Pan, Y., Bowler, C., et al. 2016. Knockdown of phosphoenolpyruvate carboxykinase increases carbon flux to lipid synthesis in *Phaeodactylum tricornutum*. ***Algal Research***, 15: 50-58.

Yoon, H. S., Hackett, J. D., Ciniglia, C., et al. 2004. A molecular timeline for the origin of photosynthetic

eukaryotes. ***Molecular Biology and Evolution***, 21(5): 809-818.
You, Y., Sun, X., Ma, M., et al. 2022. Trypsin is a coordinate regulator of N and P nutrients in marine phytoplankton. ***Nature Communications***, 13(1): 4022.
Yu, G., Nakajima, K., Gruber, A., et al. 2022. Mitochondrial phosphoenolpyruvate carboxylase contributes to carbon fixation in the diatom *Phaeodactylum tricornutum* at low inorganic carbon concentrations. ***New Phytologist***, 235(4): 1379-1393.
Zaslavskaia, L. A., Lippmeier, J. C., Kroth, P. G., et al. 2000. Transformation of the diatom *Phaeodactylum tricornutum* (Bacillariophyceae) with a variety of selectable marker and reporter genes. ***Journal of Phycology***, 36(2): 379-386.
Zhang, H., Zeng, R., Chen, D., et al. 2016. A pivotal role of vacuolar H^+-ATPase in regulation of lipid production in *Phaeodactylum tricornutum*. ***Scientific Reports***, 6: 31319.
Zhang, Y., & Gladyshev, V. N. 2008. Molybdoproteomes and evolution of molybdenum utilization. ***Journal of Molecular Biology***, 379(4): 881-899.

Annexes

Annex II: Primer list

Colony PCR and qPCR	Trans3_seqF	ATGAAGAATCAATATTCTCTCGTCAAG
	Trans3_seqR	GCAAACTGGTACCAATCCCCG
	Trans3-full-R	CGACGTCGCCGCAGCTTCGAG
	qPCR_F1	CGTCAGTCTACTATCCGCAGC
	qPCR_R1	TTCTTGTAGATGCCGACCCC
	qPCR_F2	TCTCCACCAAAAGTCCGCTC
	qPCR_R2	ACCAACTAACCACGACTCGG
	qPCR_F3	TCCAGATTCACACCGTTCCG
	qPCR_R3	GCCCTGTAGTCCCAGCTTTT
	RPS-F	AATTCCTCGAAGTCAACCAGG
	RPS-R	GTGCAAGAGACCGGACATAC
	TBP-F	ATCGATTTGTCAATCCACGAG
	TBP-R	ATACAGATTCTGTGTCCACGG
self-assembling	GB-MGD1-F	CACTTGTGCGAACGACTAGTATGGACAGACCAAAG
	GB-MGD1-R	ACTGGTGCCACCCATGGGCCCAAGGAAAAACAGTTT
	GB-Bip-F	CACTTGTGCGAACGACTAGTATGATATTCATGAGA
	GB-Bip-R	ACTGGTGCCACCCATGGGCCCCACGGTACCCATCTT
	GB-HSP70-F	CCACTTGTGCGAACGACTAGTATGGTGCATCTTCCA
	GB-HSP70-R	ACTGGTGCCACCCATGGGCCCGGATCCCATTTTGTT
	GB-GAPDH-F	AAGTCACCACTTGTGCGAACGATGAAGTTCTCTGCCGCCACTTTT
	GB-GAPDH-R	CTCCTTTGCTACTGGTGCCACCCATGGCCTTGATCTTGGCGTCGACG
	GB-BSP-GFP10-F	GGGCCCATGGGTGGCACCAG
	GB-Tom70-R	CTGGTGCCACCCATGGGCCCCAATTCGGGCTGGTT
	GB-110aa-F	CACTTGTGCGAACGGAATTCATGAAGAATCAATAT
	GB-110aa-R	ACCAGACCCTCCATCCCTAGGACACTGAGAGAGTTT
	GB-Tran3-0TMD-F	CGTGCCTTTGGTGCGACAGCTCCTAGGGATGGAGGGTCTGG
	GB-Tran3-0TMD-R	AGCTGTCGCACCAAAGGCACG
	GB-Tran3-1TMD-F	ACCGAAAACTCACCTAGGGATGGAGGGTCTGGT
	GB-Tran3-1TMD-R	CCTAGGTGAGTTTTCGGT
	GB-Tran3-2TMD-F	TCCATCGTTACCGCCTAGGGATGGAGGGTCTGGT
	GB-Tran3-2TMD-R	CCTAGGCGGTAACGATGGA
	GB-Tran3-3TMD-F	TCACACCGTTCCGTCCCTAGGGATGGAGGGTCTGGT
	GB-Tran3-3TMD-R	CCTAGGGACGGAACGGTGTGA
	GB-Tran3-4TMD-F	TCCATTCAAAAGCTGCCTAGGGATGGAGGGTCTGGT
	GB-Tran3-4TMD-R	CCTAGGCAGCTTTTGAATGGA
	GFP11-F	GATGGAGGGTCTGGTGGC
	GFP11-Tran3full-R	CCACCAGACCCTCCATCCGACGTCGCCGCAGCTTCGAG
	pPhat-intF	CTTCGTGGCCGAGGAGCAG
	pPhat-intR	CTGCTCCTCGGCCACGAAG
	pPhat-intF2	CCCACGCTCACCGGCTCCAGATTTATC
	pPhat-intR2	GATAAATCTGGAGCCGGTGAGCGTGGG
	GFP-5end -F	ATGGGTGGCACCAGTAGC
	GFP-5end -HSP70-R	GCTACTGGTGCCACCCATGGATCCCATTTTGTTGTCTG
	NR-prom-Trans3-F	GTCACCACTTGTGCGAACGGAATTCATGAAGAATCAATATTCTCTCGTC
	NR-prom-R	CGTTCGCACAAGTGGTG
	Testing-F	ACTTTGTATTTTGGTCGGGT
	Testing-R	CCCTCTCCTCTGACAGAAAA
CRISPR	CRISPR-1-F	GTCTACTATCCGCAGCGGCAAGGGTTTTAGAGCTAGAAATAGC
	CRISPR-1-R	CCTTGCCGCTGCGGATAGTAGACCGACTTTGAAGGTGTTTTTTG
	CRISPR-2-F	GACGATTCCTCGCCGTCTTTTGGGTTTTAGAGCTAGAAATAGC
	CRISPR-2-R	CCAAAAGACGGCGAGGAATCGTCCGACTTTGAAGGTGTTTTTTG
	CRISPR-3-F	GCGTTGCGCCGATTGGATGCAGGGTTTTAGAGCTAGAAATAGC
	CRISPR-3-R	CCTGCATCCAATCGGCGCAACGCCGACTTTGAAGGTGTTTTTTG
	CRISPR-4-F	GTCTATGCTTCCAAAATGATGTTTTAGAGCTAGAAATAGC
	CRISPR-4-R	ATCATTTTGGAAGCATAGACCGACTTTGAAGGTGTTTTTTG
	CRISPR-5-F	GTCAGTCGCCTTTTCTTGACCGGGTTTTAGAGCTAGAAATAGC
	CRISPR-5-R	CCGGTCAAGAAAAGGCGACTGACCGACTTTGAAGGTGTTTTTTG
	CRISPR-6-F	CACGGCCCTCTTTGGACCTTCGGGTTTTAGAGCTAGAAATAGC
	CRISPR-6-R	CCGAAGGTCCAAAGAGGGCCGTGCGACTTTGAAGGTGTTTTTTG

Annex III: Dynamic Cell Imaging: application to the diatom *Phaeodactylum tricornutum* under environmental stresses

Bey, H., Charton, F., Cruz de Carvalho, H., Liu, S., Dorrell, R. G., Bowler, C., ... & Boccara, M. (2022). Dynamic cell imaging: application to the diatom phaeodactylum tricornutum under environmental stresses. European Journal of Phycology, 1-11.

My contributions:

Fe depletion cell line preparation: 10%

EUROPEAN JOURNAL OF PHYCOLOGY, 2022
https://doi.org/10.1080/09670262.2022.2081732

Dynamic Cell Imaging: application to the diatom *Phaeodactylum tricornutum* under environmental stresses

Houda Bey[a], Florent Charton[b], Helena Cruz de Carvalho[b,c], Shun Liu[b], Richard G. Dorrell[b], Chris Bowler[b], Claude Boccara[a] and Martine Boccara[a,b]

[a]Institut Langevin, ESPCI Paris, PSL Research University, CNRS UMR 7587, 1 rue Jussieu, 75005 Paris, France; [b]Ecole Normale Supérieure, PSL Research University, Institut de Biologie de l'Ecole Normale Supérieure (IBENS), CNRS UMR 8197, INSERM U1024, 46 rue d'Ulm, F-75005 Paris, France; [c]Université Paris Est-Créteil (UPEC), Faculté des Sciences et Technologie, 61, Avenue du Général De Gaulle, 94000 Créteil, France

ABSTRACT

The dynamic movement of cell organelles is an important and poorly understood component of cellular organization and metabolism. In this work we present a non-invasive non-destructive method (Dynamic Cell Imaging, DCI) based on light scattering and interferometry to monitor dynamic events within photosynthetic cells using the diatom *Phaeodactylum tricornutum* as a model system. For this monitoring we acquire for a few seconds movies of the signals that are related to the motion of dynamic structures within the cell (denoted scatterers), followed by a statistical analysis of each pixel time series. Illuminating *P. tricornutum* with LEDs of different wavelengths associated with short pulsed or continuous-wave modes of illumination revealed that dynamic movements depend on chloroplast activity, in agreement with the reduction in the number of pixels with dynamic behaviour after addition of photosystem II inhibitors. We studied *P. tricornutum* under two environmentally relevant stresses, iron and phosphate deficiency. The major dynamic sites were located within lipid droplets and chloroplast envelope membranes. By comparing standard deviation and cumulative sum analyses of the time series, we showed that within the droplets two types of scatterer movement could be observed: random motion (Brownian type) but also anomalous movements corresponding to a drift which may relate to molecular fluxes within a cell. The method appears to be valuable for studying the effects of various environments on microalgae in the laboratory as well as in natural aquatic environments.

HIGHLIGHTS

- Light scattering is an alternative to fluorescence to rapidly evidence dynamic processes.
- Lipid droplets are the major metabolic active sites under stress.
- A non-destructive visualization method suitable for laboratory microalgae and aquatic samples.

ARTICLE HISTORY Received 8 November 2021; Revised 14 February 2022; Accepted 17 April 2022

KEYWORDS Chloroplast; endoplasmic reticulum; iron limitation; light scattering; lipid droplets; organelle movement; phosphate limitation

Introduction

Diatoms are a diverse group of photosynthetic microorganisms which account for up to 40% of ocean primary production (Bowler *et al.*, 2010). They are distributed worldwide, from tropical and subtropical regions to polar ecosystems in oceans and fresh waters, and thus have an exceptional ability to adapt to highly dynamic aquatic environments (Falciatore *et al.*, 2020; Pierella Karlusich *et al.*, 2020). These include stressful environmental fluctuations and chronic scarcities of key nutrients including nitrogen, phosphorus and iron (Abida *et al.*, 2015; Alipanah *et al.*, 2015; Cruz de Carvalho *et al.*, 2016; Kazamia *et al.*, 2018; Gao *et al.*, 2021), extreme temperatures ranging from continuously near-freezing in ocean polar regions to hot currents in equatorial oceans (Yao *et al.*, 2012; Liang *et al.*, 2019) or highlight illumination levels (Domingues *et al.*, 2012; Alboresi *et al.*, 2016).

Diatoms have a complex evolutionary origin, which involved a secondary symbiotic event where a heterotrophic eukaryote engulfed a microalga of the red lineage (Cavalier-Smith, 1999). This has led to the presence of four membranes around the plastids of diatoms, with the outermost chloroplast membrane joined to the endoplasmic reticulum (ER) and several internal chloroplast compartments including the periplastid compartment (located between the second and third innermost membranes). To identify the metabolically active organelles in diatom cells under stress, methods involving organelle isolation, biochemical methods, genomic, and cell products characterization have been very valuable (Abida *et al.*, 2015; Lupette *et al.*, 2019). Fluorescence imaging techniques with high spatial and temporal resolution have previously been used to further investigate dynamic processes in stressed diatoms (Lupette *et al.*, 2019; Jaussaud *et al.*, 2020). Indeed, cytological methods involving the use of specific stains or of fluorescent proteins are widely used in microscopy. Green Fluorescent Protein (GFP) tagged proteins are used to visualize and track in real time proteins of interest as well as cell organelles (Apt *et al.*, 2002; Kazamia *et al.*, 2018).

CONTACT Claude Boccara claude.boccara@espci.fr; Martine Boccara mboccara@bio.ens.psl.eu

Published online 19 Jul 2022

Among many GFP applications, Fluorescence Recovery After Photobleaching (FRAP) has been used to characterize the transport of integral membrane proteins from the endoplasmic reticulum to lipid droplets in yeast (Jacquier *et al.*, 2011).

Studies of biological materials by Dynamic Cell Imaging (DCI), also called Dynamic Full field Optical Transmission Tomography (D-FFOTT), are based on the detection of light back-scattered by movements of subcellular organelles (denoted scatterers) and do not need elaborate molecular manipulations, unlike GFP. This method is non-invasive and non-destructive (Apelian *et al.*, 2016; Thouvenin *et al.*, 2017*a*, 2017*b*; Scholler *et al.*, 2019). A more recent variation of DCI, working in transmission and forward scattering, that we use here is described in Thouvenin *et al.* (2021). DCI was shown to be associated with metabolic activity as it disappeared when cells were fixed or treated with the metabolic inhibitor deoxy-glucose (Apelian *et al.*, 2016). DCI appeared to be a promising technique for studying cell substructural dynamics in eukaryotic microorganisms that cannot be genetically modified. Thus we first studied the model unicellular photosynthetic microorganism *Phaeodactylum tricornutum* in responding to nutrient stress conditions.

We studied two major nutrient stresses: iron and phosphate depletion. Iron (Fe) is an essential micronutrient for all living cells since it is crucial for photosynthesis and respiration, and for phytoplankton it is a cofactor of proteins involved in a number of metabolic reactions. Iron deficiency leads to a decrease in photosynthetic efficiency (Kolber *et al.*, 1994) by inducing partial blocking of the electron transfer between photosystem II (PSII) and photosystem I (PSI) (Roncel *et al.*, 2016). Iron is an important growth-limiting or co-limiting nutrient in many regions of the world's ocean in which diatoms are important primary producers (Ustick *et al.*, 2021).

Phosphate is a fundamental element of all living cells as it is a constituent of membranes, nucleic acids and other biomolecules. Phosphate limitation and co-limitation are frequently observed in low northern latitudes (Ustick *et al.*, 2021). Deficiency of its inorganic form (Pi) is a common form of stress that strongly limits ocean primary production. Indeed, inorganic phosphate (Pi) deficiency is known to induce a shift in diatom metabolism leading to arrested cell growth and production of lipid vesicles (Cruz de Carvalho *et al.*, 2016). It has recently been shown that these lipid vesicles (also known as lipid bodies or lipid droplets) accumulate in the vicinity of the chloroplasts (Alipanah *et al.*, 2018; Lupette *et al.*, 2019). Indeed limitation of nutrients such as nitrate and phosphate leads to remodelling of cellular membranes, with degradation of membrane phospholipids and fatty acids which are redirected to lipid bodies (Goss *et al.*, 2020).

Various possible kinds of organellar movements can be observed in cells from a pure Brownian type to drifting at a defined velocity (Saunter *et al.*, 2009). Here we characterize the small amplitude movements of scatterers within *P. tricornutum* by analysing the time series associated with each pixel of the field of view in order to get a map of the transmission, the standard deviation, the cumulative sum and the average frequency. Through this computational analysis, we were able to highlight random and non-random movements suggesting that this method can spot *in vivo* molecular fluxes.

Materials and methods

Diatom culture conditions

Phaeodactylum tricornutum CCMP632 was cultured in artificial seawater (40 g l^{-1}, Sigma) supplemented with f/2 nutrients and vitamins (F/2 Media Kit, Bigelow National Center for Marine Algae and Microbiota, USA) under continuous shaking (100 rpm) at 19°C under cool white fluorescent lights (30 µmol m^{-2} s^{-1}) with a 12 h photoperiod. For phosphate depletion studies equal aliquots of 4-day-old cultures from the same batch culture were inoculated in parallel in fresh f/2 media (control conditions) and in 250 ml fresh f/2 media without phosphate supplement (Pi depleted) and cultured in the same conditions as described above for 8 days.

For iron depletion, cultures were grown in enhanced seawater medium with or without iron-EDTA (supplemented concentration 86.9 µmol l^{-1}), and analysed after two subcultures (with starting inoculum of 100 000 cells ml^{-1}) for 8 days. All cell lines were grown in ventilator-capped plastic culture flasks (Sigma) to minimize contamination into the growth media by residual iron or phosphorus from the culture flask. Cells were collected by centrifugation and mixed with melted 1% agar (42°C) in culture medium and immediately observed after jellification (20°C).

For photosystem II inhibition, we used 3-(3,4-dichlorophenyl)- 1,1-dimethylurea (DCMU; 40 µM) and hydroxylamine (HA; 2 mM) solutions in water. Cells were incubated for 10 minutes at room temperature with inhibitors, embedded in agar and immediately observed.

DCI: Data acquisition and treatment

To follow the movement of internal structures within diatoms at the pixel level we used the time variation of optical tomographic images, the principle of which is described in detail in Apelian

et al. (2016), Thouvenin *et al.* (2017*a*, 2017*b*) and Scholler *et al.* (2019), but using a new experimental approach, called Dynamic Full Field Optical Transmission Tomography (D-FFOTT) that is described in detail in Thouvenin *et al.* (2021). In brief, the sample is illuminated in transmission by the incoherent light emitted by a LED; this beam is partially scattered by the sample structure and partially transmitted. Both beams propagate along the same path and interfere with a phase shift (known as the Gouy phase shift) that depends on the relative position of the scatterer and of the focus of the objective. The subcellular organelle movements induce a time variation of the signal detected on each pixel of the camera. We use a Photonfocus A1024B camera working at 140 Hz with a CMOS image sensor with a 0.2 Me^{-} full well capacity that ensures measurements with a good signal-to-noise ratio. The light source (LED + stabilized power supply) is very stable and so the main source of noise relates to photoelectron statistics associated with each camera pixel, known as 'shot noise'. In our experiment each pixel of the camera stores N = 160 000 electrons (we work below the saturation of 0.2 Me^{-}), so the expected shot noise is $\sqrt{N} = 400$, i.e. allowing the detection of a variation in transmission of 1/400 the average value.

To represent the evolution of signal over time we displayed a number of characteristic images extracted from a movie of typically 100–200 images lasting a few seconds (frequency 140 images s^{-1}), as follows: (1) The transmission image is analysed to check the stability of the cells and the lack of bleaching. (2) Steps 1 and 2 are run with the open source platform Fiji (Schindelin *et al.*, 2012). (3) The average cumulative sum of one tenth of the movie (Scholler, 2019) is compared with the standard deviation of each pixel. Unlike pure Brownian motion, a biased Brownian motion is hyper-diffusive. Organelle movements in cells are either a drift induced for instance by a flow, or a random Brownian-like motion. Often, the time series of the signals linked to the organelle movements exhibit a combination of these two movements referred to as 'biased Brownian motion'. Using the cumulative sum instead of the standard deviation is a direct way to distinguish the two types of motion. Indeed summing Brownian steps will give a trajectory that stays close to zero whereas if there is a small bias added to the Brownian random steps (Supplementary fig. S1), it will be summed for every sample and could be revealed by the cumulative sum; a more detailed description of this approach can be found in Scholler (2019). We used the cumsum function of Matlab which, when a small bias is present, is more adapted to reveal anomalous Brownian motion and, in addition, increase the signal-to-noise ratio of the DCI. The normalization used the shot noise normal distribution in the background. All experiments were repeated three times.

Fluorescence microscopy

We used GFP (green fluorescent protein) constructs to characterize the dynamic location of several *P. tricornutum* organelles, favouring previously established reporter proteins for each organelle (Liu *et al.*, 2016). C-terminal GFP constructs were assembled by Gibson assembly into pPhat vectors, and introduced into wild-type *P. tricornutum* by biolistic transformation, followed by selection on 100 µg ml^{-1} zeocin-supplemented ESAW-agar plates, as previously described (Falciatore *et al.*, 1999). The selected markers were: GFP with no targeting signal in which the signal accumulates in the cytoplasm; GFP fused to BiP (Binding immunoglobulin protein), localized in the chloroplast endoplasmic reticulum (Apt *et al.*, 2002); and GFP fused to the N-terminal extension comprising a signal peptide and a transit peptide of cpHsp70 (chloroplastic heat shock protein 70), expressed in the periplastid compartment (Gould *et al.*, 2006). Full construct vector sequences are provided in Supplementary table S1. For all GFP experiments transgenic diatoms were analysed using a standard inverted epifluorescence microscope (Nikon Ti-E) equipped with an oil immersion objective (90×, 1.4 NA) and an EMCCD camera (Andor Ixon Ultra X-10381). GFP was excited at 488 nm and the emitted fluorescence detection window was 500–550 nm. The sensitivity of the camera used for bright field recording (10 times lower full well capacity than the Photon Focus CMOS camera) did not allow merging of the two types of signal (fluorescence and interferometry) using the same field of view.

We also used multimodal imaging combining interferometry and fluorescence in a single optical setup (described in Thouvenin *et al.*, 2017*b*) to compare the fluorescent labelling with the distribution of DCI signals within lipid bodies. Diatoms were incubated with BODYPY 505/515 (dissolved in DMSO; 0.1 µg ml^{-1} final concentration) for 30 minutes and observed by DCI and fluorescence. In this system we used a LED source centred at 470 nm (M470L4, Thorlabs, Newton, NJ, USA) for the excitation and filtered with a bandpass filter centred on 475 nm with a bandwidth of 50 nm (Semrock FF02-475/50-25). The interferometry signal was recorded on a custom camera (Quartz

2A750, Adimec). The emitted fluorescence was filtered with another bandpass filter centred on 540 nm (Semrock FF01-540/50-25) then imaged using a sCMOS camera (Hamamatsu Photonics). A dichroic mirror at 506 nm (Semrock FF506-Di03-25) separates excitation and fluorescence wavelengths. The experiment was repeated twice with similar results.

Results

DCI reveals the dynamics of diatom organelles

Diatoms grown in supplemented artificial seawater were embedded into agar in the same medium and immediately observed with the DCI set up (or D-FFOTT; Thouvenin *et al.*, 2021). The preparation was illuminated for a few seconds (pulsed light, 1 ms, LED blue 455 nm) at a few tens of Hz (140 frames s^{-1}) and a film recorded synchronously at the same frequency. We used pulsed illumination to freeze movements within cells and to avoid light saturation of photosynthesis. When observing the film of the transmission, no overall cell movement was detected and the diatoms looked still (Supplementary movie S1). When we computed the standard deviation of each pixel for all the frames of the film we observed two sites of movement within the cells: vesicles or droplets and the chloroplast membrane. The figures display the cells' metabolic activity (Apelian *et al.*, 2016); the colour map code (ImageJ, 16 colours) is such that red corresponds to the highest activity and dark blue to the lowest activity that is linked to the speed of the scatterers' movements. In the vicinity of the chloroplast, vesicles or lipid droplets of different sizes appeared green, yellow and red with the most dynamic regions located at edges of the droplet (Fig. 1., red arrow). The other highly dynamic location corresponded to the chloroplast membranes and often looked punctuated (Fig. 1., white arrows). Interestingly, these results were observed independently of the three morphotypes that were present in the cultures (fusiform, triradiate or oval) (De Martino *et al.*, 2007). *P. tricornutum* cells were also illuminated with a green LED (505 nm) or a deep red LED (735 nm, a spectral region where no *P. tricornutum* pigments absorb and no photosystems are excited). The green LED dynamic region surrounding the chloroplast and the strong signal at the edges of the vesicles were very similar to what was detected with the 455 nm LED (Fig. 2., Supplementary fig. S2A). Histogram profiles of the standard deviation of signal intensity of each pixel were very similar between blue and 505 nm LEDs with a shoulder corresponding to the strongest signals (Supplementary fig. S2B). When we used the 735 nm LED the signal was significantly reduced to one-third (Fig. 2., Supplementary figs S2A and S2B). We tested two inhibitors at concentrations inhibiting photosystem II: 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU; 40 µM) or hydroxylamine (HA; 2mM) and measured the number of pixels with dynamic behaviour. We observed a significant reduction to one-third in the presence of both inhibitors suggesting that the dynamic structures we described depend on chloroplast metabolic activity (Fig. 2.). To further investigate this observation, we compared image acquisition between continuously and pulsed LED lighting (Fig. 1.). No obvious dynamic structures within the cell were highlighted when illumination was continuous, unlike with pulsed light, although no difference in transmission was observed meaning that no photobleaching was involved (Fig. 1., C, B, D). The continuous illumination (1000 W m^{-2}) with the 455 nm LED corresponded to the high illumination level used to block chloroplast activity while the pulsed illumination (pulses of 0.25 ms at 140 Hz) corresponded to a low illumination level (30 W m^{-2} or 120 µmol photons m^{-2} s^{-1}). This suggested that the high continuous light photon budget prevents chloroplast function unlike the pulsed mode level of irradiation (average power 30 times lower, Fig. 1.) and is in agreement with the photon budget given by Prins *et al.* (2020). Our observations suggest that during continuous exposure to light, chloroplasts are saturated with photons, which results in the low metabolic activity recorded (Bailleul *et al.*, 2011). We reasoned that the elimination of harmful excess energy could be dissipated through heat, however using a microthermocouple immersed in the sample cuvette (volume 5 µl) we did not detect an overall measurable increase in temperature (< 0.1°C) under the same illumination level. This low temperature increase was confirmed by an estimation of the power absorbed by the algae in the sample volume which was found to be of a few thousandths of a degree Celsius.

Fluorescence studies indicate dynamic movements in the chloroplast: endoplasmic reticulum and lipid bodies

To better characterize the cell compartments in which dynamic structures are localized, C-terminal GFP fusions of proteins targeted to various compartments within *P. tricornutum* cells were observed with a fluorescent microscope (same samples analysed with DCI). GFP with no targeting signal was localized in the cytosol (Supplementary fig. S3, CYT). GFP

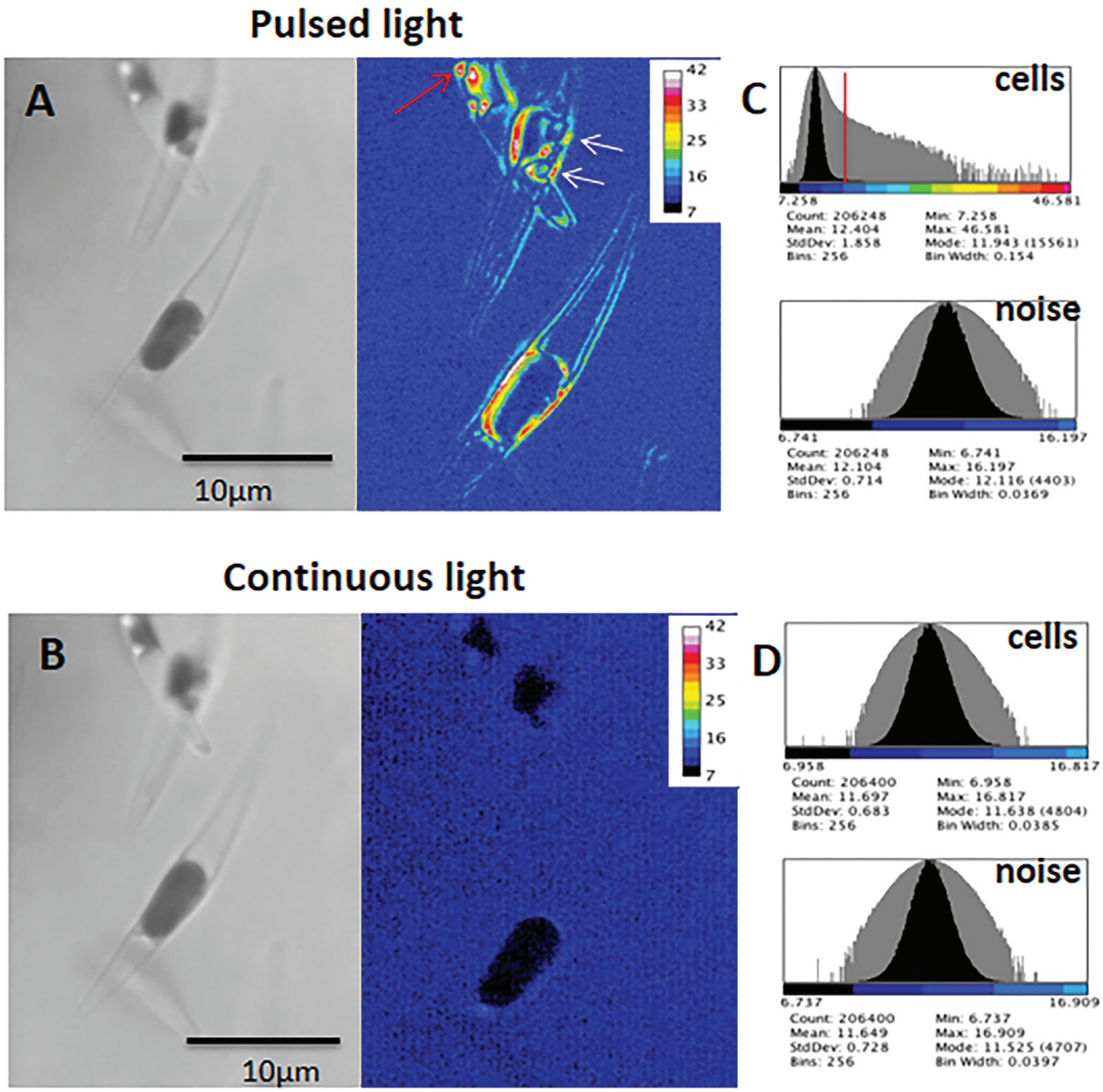

Fig. 1. Effect of continuous or pulsed lighting on intracellular movements in *Phaeodactylum tricornutum*. Cells were successively illuminated with the LED_{455}, with either continuous or pulsed light and a film recorded for a few seconds. Left image: transmission (single capture from the 100 frames movie); right image: standard deviation (scale on right) of each pixel of the same field over the 100 frames and in successive acquisition (pulsed light followed by continuous lighting). **A**: pulsed lighting; red arrow indicates lipid droplet and white arrows correspond to punctuations in chloroplast membranes. **B**: continuous lighting.

fused to a bipartite targeting signal (signal sequence and transit peptide) of HSP70, enabling this protein to cross the two chloroplast membranes, did not label the vesicles suggesting that they were not associated with the periplastid compartment (PPC) (Supplementary fig. S3, PPC). In contrast, the signal for the BIP protein fused with GFP, resulting in fluorescence in the endoplasmic reticulum (ER), labelled the dynamic vesicles (Supplementary fig. S3, ER). These observations suggested that vesicles, possibly lipid droplets (see below), originate from the endoplasmic reticulum as suggested by Jaussaud *et al.* (2020).

We took advantage of a setup which combined interferometry and fluorescence (Thouvenin *et al.*, 2017*b*) to compare the location of dynamic signals with those identifying lipid bodies. For this purpose, we labelled lipid bodies with BODIPY505/515 and as shown in Fig. 3. we observed co-localization of dynamic signals and lipid labelling. In addition

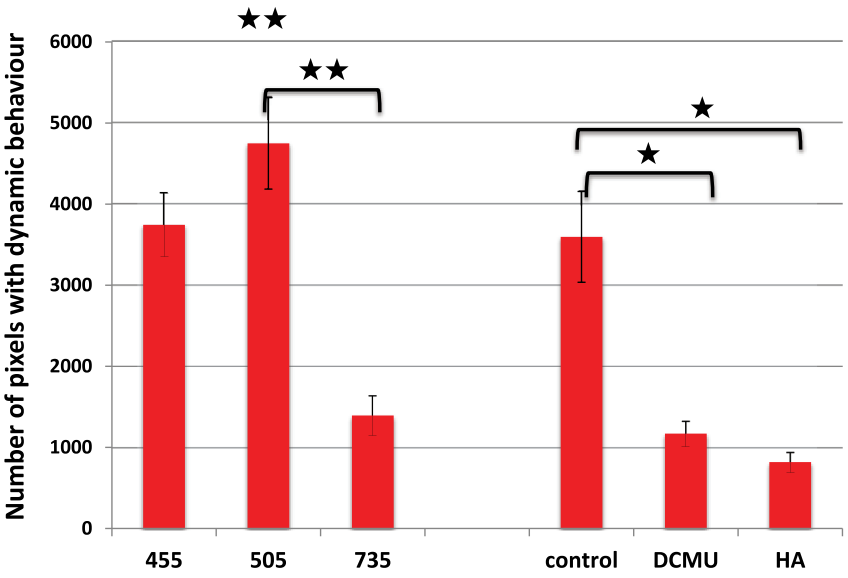

Fig. 2. Bar chart of pixels with dynamic behaviour: effects of wavelength illumination and photosystem II inhibitors. Average values plus and minus standard deviation are presented. The different wavelengths tested are: 455 nm (blue), 505 nm (green) and 735 nm (deep red). Two photosystem II inhibitors were used: DCMU (3,4-dichlorophenyl)-1,1-dimethyl-urea; 40 µM) and HA (hydroxylamine; 2 mM). 20 to 30 cells were analysed. A Wilcoxon test was carried out; the comparison of the signal observed between 455 nm (blue) and 735 nm (deep red) and between 505 nm (green) and 735 nm (deep red) gave a p-value of 0.0079 at a significance level of 0.05. The same test was carried out with photosystem II inhibitors: the signal observed between control and DCMU or HA gave a p-value of 0.03 at a significance level of 0.05.

we showed in other cells that dynamic signals were associated with more structures than the specific signal observed by fluorescence (Supplementary fig. S4).

Dynamics of organelles in diatoms grown in iron-limited conditions

Kazamia *et al.* (2018) demonstrated that iron uptake in *Phaeodactylum* under iron-limited conditions involved an endocytosis step and a delivery step in the vicinity of the chloroplast. This study suggested dynamic events in *Phaeodactylum* cells when iron is limited. We thus visualized the effect of iron limitation on *P. tricornutum* intracellular dynamics after one week and two weeks of culture in iron-depleted growth media. After one week of culture in iron-depleted medium, a slight decrease in growth was observed, suggesting a growth slow-down and after two weeks a 10-fold growth reduction was observed (Table 1). Samples were analysed using DCI after one (T1) or two (T2) transfers of culture with iron-replete or iron-depleted media. In the presence of iron in the culture medium, we did not observe much difference between the two subcultures with dynamic pixels localized mostly in lipid droplets. We computed the number of pixels after removing the background noise of the same stack and found no differences between the two measurements (Table 1, Supplementary fig. S4). After one week of culture in iron-limited medium we observed a slight reduction (by about 10%) in the number of pixels with dynamic behaviour. After two weeks of culture in iron-limited medium the number of pixels with dynamic behaviour was 75% lower (Table 1, Supplementary fig. S4). The pixels with dynamic behaviour were localized in the lipid droplets, however on average there was no significant difference in the droplet diameters between iron-replete and iron-depleted conditions (1.1 ± 0.5 µm and 0.9 ± 0.2 µm respectively). All these results suggest that in the iron-depleted medium the accumulation of lipids in droplets is arrested after prolonged starvation, which may be due to the shutdown of photosynthetic carbon assimilation and/or lipid synthesis.

Dynamics of lipid bodies in diatoms subjected to phosphate deprivation

It has been previously reported that after 8 days of culture in full-nutrient media small lipid droplets start forming in *P. tricornutum* cells, suggesting a developing nutrient stress (Cruz de Carvalho *et al.*, 2016). After 8 days in full culture media we observed on average two lipid droplets per *P. tricornutum* cell and their diameters were computed (Table 2, Fig. 4., Supplementary fig. S5). The analysis of the standard deviation of signal intensity of each pixel showed that 30–50% of the droplets exhibited a dynamic behaviour (Table 2, Fig. 4., Supplementary fig. S5). In the case of *P. tricornutum* grown for 8 days in phosphate-depleted medium there was a net reduction in cell growth (2–3

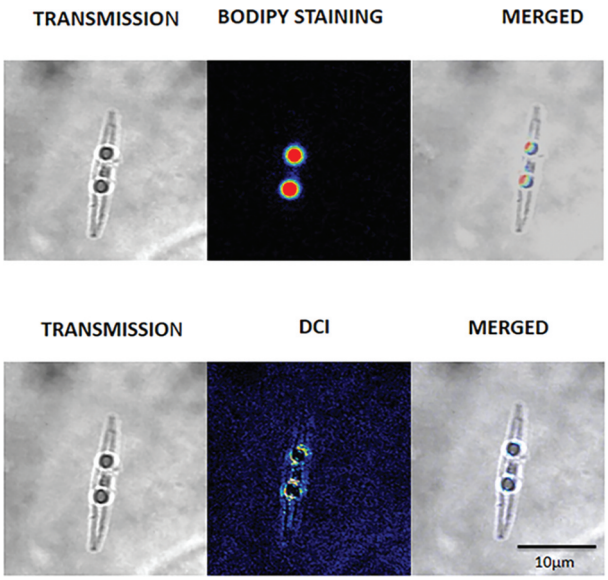

Fig. 3. Co-localization of dynamic droplets from *Phaeodactylum tricornutum* and BODIPY© labelled lipid droplets. Left panel: transmission microscopy, central panel: fluorescence (grey scale camera, green colouration), right panel merge of the two previous images. Left panel: transmission microscopy, central panel: standard deviation (grey scale camera, 16 colours), right panel merge of the two previous images. Scale bar values were deduced from the format of the image (90 × 90 µm).

times) (Cruz de Carvalho *et al.*, 2016) and a net increase in the number of droplets (5–10), which were significantly larger than when grown in Pi-replete medium (Table 2). Furthermore, the majority of the lipid droplets showed dynamic behaviour (Table 2, Fig. 4., Supplementary fig. S5). Generally, the droplets were in close association with the chloroplast but sometimes they were found free in the cytoplasm and were able to fuse, giving rise to very large structures (red arrow, Supplementary fig. S5). Histogram profiles showed a net increase (about 5-fold) in the number of pixels above the background noise between the conditions with and without phosphate (Table 3, Fig. 4., Supplementary fig. S5).

To get more insights into the dynamics of lipid droplets we analysed the time series of each pixel in order to get a map of the transmission, the standard deviation, the cumulative sum (Matalb: cumsum function) and the average frequency. Despite a noticeable difference in signal level (standard deviation and cumulative sum, see below) between the cells grown with and without Pi, we did not observe any significant differences with regard to the central frequencies, which were around 15 ± 5 Hz. We observed a 3–5 times increase in the signal amplitude (standard deviation and cumulative sum respectively) in *P. tricornutum* grown without phosphate (Table 3). The analysis of the signal value also pointed out some differences in the signal distribution between standard deviation and cumulative sum computations. The computation of value variation of a pixel, based on the cumulative sum of the signals, highlighted a non-Brownian displacement within the lipid droplets (Table 3). Following the method developed in Scholler (2019) we found that the normalized cumulative sum is about twice that of the normalized standard deviation in culture with phosphate while it is slightly greater (×1.2) in culture without phosphate (Table 3). This result suggests that the random movement of scatterers within the droplets is likely to be associated with a drift that makes it hyper-diffusive, rather than being a pure Brownian motion and could correspond to the flux of lipids and proteins filling lipid bodies (Olzmann & Carvalho, 2019; Nettebrock & Bohnert, 2020).

Table 1. Effect of iron limitation after one (T1) and two (T2) passages on dynamic movements in *Phaeodactylum tricornutum*

	Number of cells × 10^6 ml^{-1} ± SD	Average number of pixels with dynamic behaviour ± SD[a]	Number of analysed cells
+ Fe T1	106 ± 0.8	1312 ± 280	39
+ Fe T2	127 ± 2	1307 ± 343	36
- Fe T1	89 ± 4	1166 ± 311[b]	40
- Fe T2	10 ± 0.8	279 ± 131[c]	46

[a]90 000 pixels were selected for each analysed field;
[b]Not significantly different from +Fe T1; p-value: 0.56 (significance level of 0.05);
[c]Significantly different from +Fe T2; p-value: 1.2 10^{-8} (significance level of 0.05).

Discussion

Here, we present a non-invasive (no genetic manipulation) and non-destructive (no fixation step) method, which allows the detection and characterization of internal movements within photosynthetic cells. In particular, through image analysis we show evidence that these dynamic processes are correlated with the functioning of the chloroplasts. This detection was possible by relying on the scattering properties of the cell organelles and of the high frame rate of our camera recordings. We were able to identify and quantify movements within lipid droplets, specific organelles, which accumulate in nutrient-stressed diatoms. In cells deprived of phosphate, in addition to the random movement of little scatterers, which may correspond to the movement of membrane proteins from ER (Jacquier *et al.*, 2011) or storage of misfolded proteins in lipid droplets as suggested by Lupette *et al.* (2019), we described a drift movement of scatterers, which might correspond to molecular fluxes of proteins and/or lipids. The distribution of dynamic pixels which were often observed in an eggcup shape (Supplementary fig. S6) could be in agreement with the ER surrounding lipid droplets (Jacquier

Table 2. Diameter and dynamic behaviour of droplets in full or phosphate-depleted medium.

	Average droplet diameter (µm ± SD) (number of analysed droplets)	Number of droplets with dynamic behaviour (percentage)
+ Pi: exp1	0.56 ± 0.22 (42)	22 (52%)
− Pi: exp1	0.97 ± 0.33 (58)	55 (94%)
+ Pi: exp2	0.43 ± 0.11 (18)	6 (33%)
− Pi: exp2	0.96 ± 0.32 (26)	21 (80%)
+ Pi: exp3	0.51 ± 0.21 (26)	12 (46%)
− Pi: exp3	0.88 ± 0.22 (86)	61 (71%)

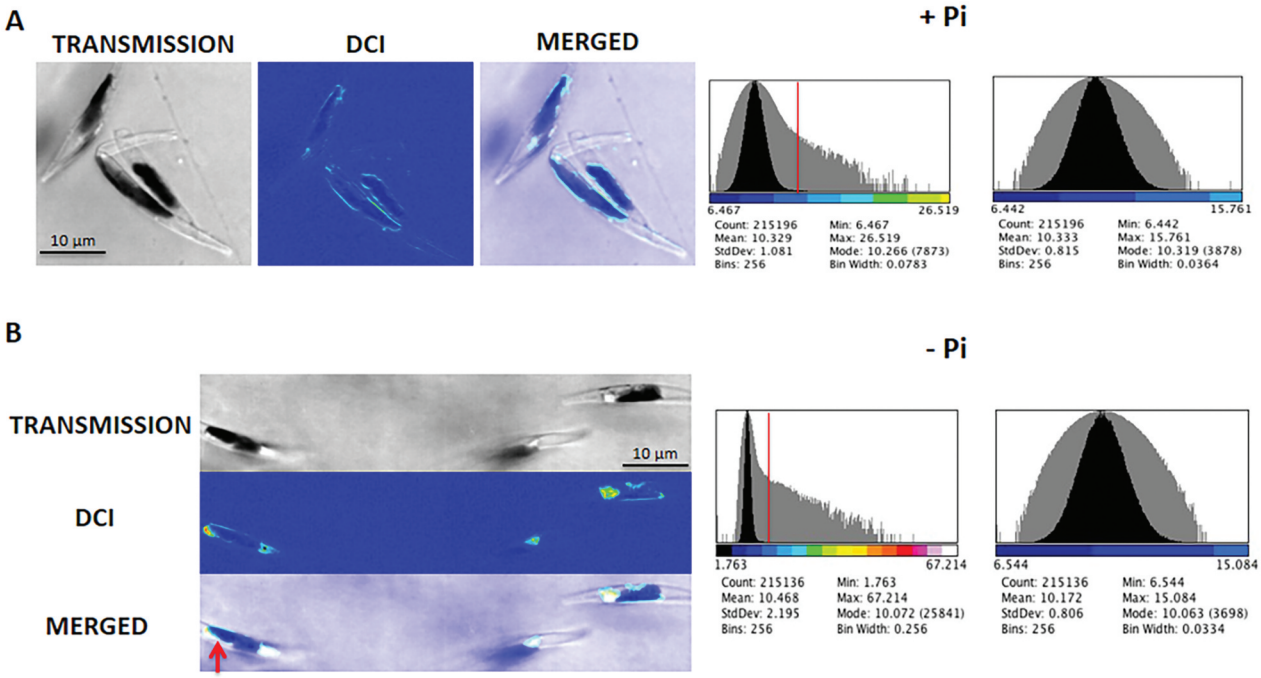

Fig. 4. Increase in lipid droplets size and numbers in *Phaeodactylum tricornutum* cells grown for 8 days in phosphate-depleted medium. From left to right panel (top to bottom) are shown the transmission (single capture from the 100 frames movie), the standard deviation of each pixel of the same field over the 100 frames and the merged image (artificial blue colour). Red arrow indicates lipid droplet in phosphate-depleted medium. Histograms of the standard deviation images using image J (Fiji) were recorded and the number of pixels above noise level were computed (indicated by a red bar on the histogram). Histograms of the same image size without cells were recorded to determine noise levels. Scale bar values were deduced from the format of the image (60 × 60 µm).

Table 3. Comparison of movements of *Phaeodactylum* scatterers, computing their standard deviation or cumulative sum in media with or without phosphate

	Number[a] of dynamic pixels ±SD (method: standard deviation)	Number of dynamic pixels ±SD (method: Cumulative sum)	Ratio Cumulative sum/ Standard deviation
+Pi	586 ± 167	1174 ± 196	2
-Pi	2844 ± 155	3311 ± 236	1.16

[a]Three stacks were analysed in each condition. A total of 33 000 pixels was selected for each analysed field.

et al., 2011) and more precisely in the specific case of diatoms the outermost membrane of the chloroplast (Flori *et al.*, 2016; Jaussaud *et al.*, 2020; Leyland *et al.*, 2020).

Cytoplasmic streaming generated by myosin in the alga *Chara* or in plants has long been known (reviewed by Tominaga & Ito, 2015). Indeed the speed of cytoplasmic streaming generated by *Chara* myosin observed by optical trap nanometry is in the range of 50 µm s^{-1} (Tominaga & Ito, 2015). With DCI we observed live imaging of low speed movements (about 0.1 µm s^{-1}) of undetectable scatterers within *P.tricornutum*organelles, corresponding to their metabolic activity. The DCI method can be extended to any microalgae even though transformation is not available for many of them. For example, we have studied the polar diatom *Fragilariopsis cylindrus* grown under light or in dark, alongside polar, temperate and tropical isolates of the haptophyte genus *Pavlova* grown under multiple environmental temperatures, identifying differences in dynamic cell structure in each case (unpublished data). The method is very sensitive and allows the study of slight changes due to environmental fluctuations such

as variation in CO_2 concentration, light quality responses, or nutrient abundance. In addition, most of the image analyses have been run in Fiji (ImageJ), a public domain application (Schindelin *et al.*, 2012).

Cell exploration by microscopic techniques has recently made tremendous progress (Kwon, 2021) however most of these require skill and expensive equipment. Here we used a very simple set up (Thouvenin *et al.*, 2021) to rapidly analyse diatom cells under stress. The DCI method can easily be used as a non-intrusive first step before in depth investigations of intracellular movements via FRAP, single-cell -resolution secondary ion mass spectroscopy (SIMS) or other techniques.

Acknowledgements

We wish to thank our colleagues: Ignacio Izzedine (Institut Langevin) for the use of the fluorescence microscope, Olivier Thouvenin (Institut Langevin) for his help in the dynamic signal analysis, and Benjamin Bailleuil (IBPC) for providing photosystem II inhibitors.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Houda Bey, Claude Boccara: DIM ELICIT (INFLUERE IN VISCERA).Florent Charton: Q-life (Q-life ANR.-17-CONV-0005). Helena Cruz de Carvalho: Agence Nationale de la Recherche (ANR DiaLincs 19-CE43-0011-01 585). Richard Dorrell: CNRS (CNRS Momentum Fellowship). Chris Bowler: ERC (European Research Council, Diatomic project); Agence Nationale de la Recherche (ANR Browncut).

Supplementary material

The following supplementary material is accessible via the Supplementary Content tab on the article's online page at https://doi.org/10.1080/09670262.2022.2081732

Supplementary table S1: list of primers for GFP constructs.

Supplementary movie S1: Transmission film of diatoms illuminated at 455 nm.

Supplementary fig. S1: Simulation of a pure Brownian motion (blue) and of a biased movement (red) where drift is added to the Brownian motion. The y-axis of the graph shows the dynamic movement of a particle measured over time (x-axis). For this specific example the standard deviation was found to be 1.16 (Brownian) and 1.37 (Brownian biased) whereas the cumsum (cumulative sum) was found to be 12.30 (Brownian) and 180.64 (Brownian biased), demonstrating a very large enhancement of the signal when a drift is present.

Supplementary fig. S2: Effect of wavelength illumination on intracellular movements in Phaeodactylum tricornutum. Cells were successively illuminated with pulsed light from LED735 (red), LED505 (green) and LED455 (blue) and a film (100 or 200 frames) was recorded for a few seconds. The same field of view in the successive acquisition was selected; upper image: transmission; lower image standard deviation (scale on bottom left) (Supplementary fig. S2A). Histograms of the standard deviation images using image J (Fiji) were recorded and the number of pixels above noise level were computed (indicated by a red bar on the histogram) in Supplementary fig. S2B. Histograms of the same image for each wavelength, but without cells, were recorded to determine noise levels. Scale bar was deduced from the format of the image (60x60 µm).

Supplementary fig. S3: Co-localization of dynamic lipid droplets from Phaeodactylum tricornutum and labelled endoplasmic reticulum. Left panel: light microscopy, central panel: GFP fluorescence (grey scale camera, green coloration), right panel merge of the two upper images. Cytosolic (CYT) expressed GFP accumulates in the entire cell. The signal peptide of BIP protein fused to GFP leads to endoplasmic localization of GFP (ER). The dual topogenic signal of Hsp70 fused to GFP targets GFP to the periplastid compartment (PPC). Scale bar value was deduced from the format of the image (60x60 µm).

Supplementary fig. S4: Dynamic signal in Phaeodactylum is not limited to lipid bodies Upper panel: transmission microscopy, central panel: Standard deviation (grey scale camera, 16 colours), lower panel is merger of the two previous images. Scale bar value was deduced from the format of the image (90 x 90 µm).

Supplementary fig. S5: Decrease in dynamic signal in Phaeodactylum tricornutum grown in iron depleted medium. Representative cells grown in medium with iron (+Fe) or depleted of iron (-Fe). From left to right is shown the transmission, the standard deviation of each pixel and the merged image (artificial blue colour). T1 and T2 correspond to 8 days and 16 days growth in medium with iron or depleted iron. Histograms of the standard deviation images using image J (Fiji) were recorded and the number of pixels above noise level were computed (indicated by a red bar on the histogram) in Table 1. Histograms of the same image size without cells were recorded to determine noise levels Scale bar value was deduced from the format of the image (60x60 µm).

Supplementary fig. S6: Diatoms grown with and without Phosphate: Increase in lipid droplets size and numbers in medium depleted in Phosphate.

Author contributions

H. Bey: compared interferometry and fluorescence, discussed the results and commented on the manuscript; F. Charton: compared interferometry and fluorescence, cultured P. tricornutum and studied its growth under stress, discussed the results and commented on the manuscript; H.C. de Carvalho: contributes to the writing of the manuscript, discussed the results and commented on the manuscript; S. Liu: cultured P. tricornutum and studied its growth under stress, discussed the results and commented on the manuscript; R.G. Dorrell: built the GFP fusions, contributes to the writing of the manuscript, discussed the results and commented on the manuscript; C. Bowler: contributes to the writing of the manuscript, discussed the results and commented on the manuscript; C. Boccara: built, characterized the optical setup and analysed the movies and wrote the manuscript, discussed the results and commented on the manuscript; M. Boccara: performed and analysed

the biological experiments and wrote the manuscript, discussed the results and commented on the manuscript.

ORCID

Richard G. Dorrell http://orcid.org/0000-0001-6263-9115

References

Abida, H., Dolch, L.J., Mei, C., Villanova, V., Conte, M., Block, M.A., Finazzi, G., Bastien, O., Tirichine, L., Bowler, C., Rebeille, F., Petroutsos, D., Jouhet, J. & Marechal, E. (2015). Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricornutum*. *Plant Physiology*, **167**: 118–136.

Alboresi, A., Perin, G., Vitulo, N., Diretto, G., Block, M., Jouhet, J., Meneghesso, A., Valle, G., Giuliano, G., Marechal, E. & Morosinotto, T. (2016). Light remodels lipid biosynthesis in *Nannochloropsis gaditana* by modulating carbon partitioning between organelles. *Plant Physiology*, **171**: 2468–2482.

Alipanah, L., Rohloff, J., Winge, P., Bones, A.M. & Brembu, T. (2015). Whole-cell response to nitrogen deprivation in the diatom *Phaeodactylum tricornutum*. *Journal of Experimental Botany*, **66**: 6281–6296.

Alipanah, L., Winge, P., Rohloff, J., Najafi, J., Brembu, T. & Bones, A.M. (2018). Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom *Phaeodactylum tricornutum*. *PLoS ONE*, **13**: e0193335.

Apelian, C., Harms, F., Thouvenin, O. & Boccara, C. (2016). Dynamic full field optical coherence tomography: subcellular metabolic contrast revealed in tissues by interferometric signals temporal analysis. *Biomedical Optics Express*, **7**: 1511.

Apt, K.E., Zaslavkaia, L., Lippmeier, J.C., Lang, M., Kilian, O., Wetherbee, R., Grossman, A.R. & Kroth, P.G. (2002). *In vivo* characterization of diatom multipartite plastid targeting signals. *Journal of Cell Science*, **115**: 4061–406

Bailleul, B., Cardol, P., Breyton, C. & Finazzi, G. (2011). Electrochromism: a useful probe to study algal photosynthesis. *Photosynthesis Research*, **106**: 179–189. Erratum in: *Photosynthesis Research*, **10**: 151–152.

Bowler, C., Vardi, A. & Allen, A.E. (2010). Oceanographic and biogeochemical insights from diatom genomes. *Annual Review of Marine Science*, **2**: 333–365.

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.

Cruz de Carvalho, H., Sun, H.X., Bowler, C. & Chua N.H. (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytologist*, **210**: 497–510.

De Martino, A., Meichenin, A., Shi, J., Pan, K. & Bowler C. (2007). Genetic and phenotypic characterization of *Phaeodactylum tricornutum* (Bacillariophyceae) accessions. *Journal of Phycology*, **43**: 992–1009.

Domingues, N., Matos, A.R., Marques da Silva, J. & Cartaxana, P. (2012). Response of the diatom *Phaeodactylum tricornutum* to photooxidative stress resulting from high light exposure. *PLoS ONE*, **7**: e38162.

Falciatore, A, Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. (1999). Transformation of nonselectable reporter genes in marine diatoms. *Marine Biotechnology (NY)*, **1**: 239–251.

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.

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.

Gao, X., Bowler, C. & Kazamia, E. (2021). Iron metabolism strategies in diatoms. *Journal of Experimental Botany*, **72**: 2165–2180.

Goss, R., Wilhelm, C. & Jakob, T. (2020). Photosynthesis in diatoms. In *Handbook of Algal Science, Technology and Medicine* (Konur, O., editor), 217–229. Academic Press, London.

Gould, S.B., Sommer M.S., Kroth P.G., Gile G.H., Keeling P.J. & Maier U.G. (2006). Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Molecular Biology and Evol*ution, **23**: 2413–2422.

Jacquier, N., Choudhary, V., Mari, M., Toulmay, A., Reggiori, F. & Schneiter, R. (2011). Lipid droplets are functionally connected to the endoplasmic reticulum in *Saccharomyces cerevisiae*. *Journal of Cell Science*, **124**: 2424–2437.

Jaussaud, A., Lupette, J., Salvaing, J., Jouhet, J., Bastien, O., Gromova, M. & Maréchal E. (2020). Stepwise biogenesis of subpopulations of lipid droplets in nitrogen starved *Phaeodactylum tricornutum* cells. *Frontiers in Plant Sci*ence, **11**: 48.

Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R.G., Rocha Jimenez Vieira, F., Mach, J., Morrissey, J., Leon, S., Lam, F., Pelletier, E., Camadro, J-M., Bowler, C. & Lesuisse, E. (2018). Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. *Science Advances*, **4**: eaar4536.

Kolber, Z., Barber, R., Coale, K. et al. (1994). Iron limitation of phytoplankton photosynthesis in the equatorial Pacific Ocean. *Nature*, **371**: 145–149.

Kwon, D. (2021). Proteins in their natural habitats *Nature*, **598**: 558–560

Leyland, B., Boussiba, S. & Khozin-Goldberg, I. (2020). A review of diatom lipid droplets. *Biology*, **9**: 38.

Liang, Y., Koester, J.A., Liefer, J.D, Irwin, A.J. & Finkel, Z. V. (2019). Molecular mechanisms of temperature acclimation and adaptation in marine diatoms. *ISME Journal*, **13**: 2415–2425.

Liu, X., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T., Maier, U.G. & Zauner, S. (2016). Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker protein. *Algal Research*, **20**: 249–257.

Lupette, J., Jaussaud, A., Seddiki, K., Morabito, C., Brugière, S., et al. (2019). The architecture of lipid droplets in the diatom *Phaeodactylum tricornutum*. *Algal Research – Biomass, Biofuels and Bioproducts*, **38**: 101415.

Nettebrock, N.T. & Bohnert, M. (2020). Born this way – biogenesis of lipid droplets from specialized ER subdomains. *BBA – Molecular and Cell Biology of Lipids*, **1865**: 158448.

Olzmann, J.A. & Carvalho, P. (2019). Dynamics and functions of lipid droplets. *Nature Review of Molecular Cell Biology*, **20**: 137–155.

Pierella Karlusich, J.J., Ibarbalz, F.M. & Bowler, C. (2020). Phytoplankton in the Tara Ocean. *Annual Review of Marine Science*, **12**: 233–265.

Prins, A., Deleris, P., Hubas, C. & Jesus, B. (2020). Effect of light intensity and light quality on diatom behavioral and physiological photoprotection. *Frontiers in Marine Science*, **7**: 203.

Roncel, M., González-Rodríguez, A.A., Naranjo, B., Bernal-Bayard, P., Lindahl, A.M., Hervás, M., Navarro, J.A. & Ortega, J.M. (2016). Iron deficiency induces a partial inhibition of the photosynthetic electron transport and a high sensitivity to light in the diatom *Phaeodactylum tricornutum*. *Frontiers in Plant Science*, **7**: 1050.

Saunter, C.D., Perng, M., Love, G. & Quinlan, R. (2009). Stochastically determined directed movement explains the dominant small-scale mitochondrial movements within non-neuronal tissue culture cells. *FEBS Letters*, **583**: 1267–1273.

Scholler, J. (2019). Motion artifact removal and signal enhancement to achieve in vivo dynamic full field OCT. *Optics Express*, **27**: 19562–19572.

Scholler, J., Mazlin, V., Thouvenin, O., Groux, K., Xiao, P., Sahel, J.A., Fink, M., Boccara, C. & Grieve, K. (2019). Probing dynamic processes in the eye at multiple spatial and temporal scales with multimodal full field OCT. *Biomedical Optics Express*, **10**: 731–746.

Schindelin, J., Arganda-Carreras, I., Frise, E., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, **9**: 676–682.

Thouvenin, O., Boccara, C., Fink, M., Sahel, J.A., Paques, M. & Grieve, K. (2017a). Cell motility as contrast agent in retinal explant imaging with full-field optical coherence tomography. *Investigative Ophthalmology & Visual Science*, **58**: 4605.

Thouvenin, O., Fink, M. & Boccara, C. (2017b). Dynamic multimodal full-field optical coherence tomography and fluorescence structured illumination microscopy. *Journal of Biomedical Optics*, **22**(**2**): 026004.

Thouvenin, O., Alhaddad, S., Mazlin, V., Boccara, M., & Boccara, C. (2021). Label free optical transmission tomography for biosystems: Intracellular structures and dynamics. *arXiv*, 3991089.

Tominaga, M. & Ito, K. (2015). The molecular mechanism and physiological role of cytoplasmic streaming. *Current Opinion in Plant Biology*, **27**: 104–110.

Ustick, L.J., Larkin, A.A., Garcia, C.A., Garcia, N.S., Brock, M.L., Lee, J.A., Wiseman, N.A., Moore, J.K. & Martiny, A.C. (2021). Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. *Science*, **372**: 287–291.

Yao, S., Brandt, A., Egsgaard, H. and Gjermansen, C. (2012). Neutral lipid accumulation at elevated temperature in conditional mutants of two microalgae species. *Plant Physiology and Biochemistry*, **61**: 71–79.