Proteomes reveal the lipid metabolic network in the complex plastid of *Phaeodactylum tricornutum*

Teng Huang[1,2], Yufang Pan[1], Eric Maréchal[3] and Hanhua Hu[1,2,*] iD

[1]*Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China,*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China, and*
[3]*Laboratoire de Physiologie Cellulaire Végétale, Université Grenoble Alpes, CEA, CNRS, INRA, IRIG-LPCV, 38054 Grenoble Cedex 9, France*

SUMMARY

Phaeodactylum tricornutum **plastid is surrounded by four membranes, and its protein composition and function remain mysterious. In this study, the** *P. tricornutum* **plastid-enriched fraction was obtained and 2850 proteins were identified, including 92 plastid-encoded proteins, through label-free quantitative proteomic technology. Among them, 839 nuclear-encoded proteins were further determined to be plastidial proteins based on the BLAST alignments within Plant Proteome DataBase and subcellular localization prediction, in spite of the strong contamination by mitochondria-encoded proteins and putative plasma membrane proteins. According to our proteomic data, we reconstructed the metabolic pathways and highlighted the hybrid nature of this diatom plastid. Triacylglycerol (TAG) hydrolysis and glycolysis, as well as photosynthesis, glycan metabolism, and tocopherol and triterpene biosynthesis, occur in the plastid. In addition, the synthesis of long-chain acyl-CoAs, elongation, and desaturation of fatty acids (FAs), and synthesis of lipids including TAG are confined in the four-layered-membrane plastid based on the proteomic and GFP-fusion localization data. The whole process of generation of docosahexaenoic acid (22:6) from palmitic acid (16:0), via elongation and desaturation of FAs, occurs in the chloroplast endoplasmic reticulum membrane, the outermost membrane of the plastid. Desaturation that generates 16:4 from 16:0 occurs in the plastid stroma and outer envelope membrane. Quantitative analysis of glycerolipids between whole cells and isolated plastids shows similar composition, and the FA profile of TAG was not different. This study shows that the diatom plastid combines functions usually separated in photosynthetic eukaryotes, and differs from green alga and plant chloroplasts by undertaking the whole process of lipid biosynthesis.**

Keywords: diatom, lipid synthesis, metabolic pathways, *Phaeodactylum tricornutum*, **plastid proteome.**

INTRODUCTION

Diatoms, one of the largest groups of marine phytoplankton, are responsible for ~20% of the primary productivity on Earth (Field et al., 1998; Falkowski et al., 2004). Whole-genome sequencing of model species revealed a "mosaic" structure combining features of different origins from red and possibly green algae, as well as heterotrophic organisms (Armbrust et al., 2004; Bowler et al., 2008). Consistent with this mosaic genome character, a "complex" plastid has been characterized by electron microscopy and high-resolution imaging in diatoms (Cavalier-Smith, 2018; Flori et al., 2016; Kilian & Kroth, 2005). Diatom plastid is surrounded by four membranes, in contrast to chloroplasts of land plants. From the outermost to the innermost, the four membranes are the chloroplast endoplasmic reticulum

(cER) membrane (cERM), the periplastidial membrane (PPM), and the outer and inner envelope membranes (OEM and IEM) (Flori et al., 2016). It is believed that the cERM derives from the host endomembrane system and is directly connected to the outer nuclear envelope membrane and the endoplasmic reticulum (ER) (Gibbs, 1979; Gould et al., 2015). Different from the cERM, which lies closely adjacent to the chloroplast, the ER membrane is stacked and spread throughout the cytoplasm of cells thus forming a large membrane compartment (Staehelin, 1997). The PPM is presumed to originate from the endosymbiont plasma membrane (Grosche et al., 2014), and the OEM and IEM correspond to endosymbiont chloroplast envelope membranes (Archibald & Keeling, 2002). Between the PPM and OEM remains a relic of the symbiont cytoplasm,

termed the periplastidial compartment (PPC) (Gould et al., 2015). Although the PPC is a naturally minimized compartment, it is involved in some specific metabolic processes, such as the protein transport, lipid transfer and CO_2 concentrating (Moog et al., 2011). In diatom plastids, three appressed thylakoids are organized into extended bands, which traverse much of the length of the chloroplast (Pyszniak & Gibbs, 1992). Fucoxanthin chlorophyll a/c-binding proteins (FCPs) surround the cores of the two photosystems (PSs) segregated in different thylakoid domains (Flori et al., 2017; Wang et al., 2019).

Diatom plastids share many functions and metabolic pathways with primary chloroplasts from plants, green and red algae (Rolland et al., 2018). For example, the *de novo* synthesis of fatty acids (FAs) is believed to occur in the stroma of diatom plastids (Li, Xie, et al., 2018; Roessler & Ohlrogge, 1993). However, the complex structure of diatom plastids enables them to function differently from the primary endosymbiosis chloroplasts. The diatom plastid carries a complete glutamine-ornithine cycle, and the genome-wide bioinformatic analysis reveals the presence of a complete tocopherol biosynthesis pathway in it (Nonoyama et al., 2019).

In plant cells and in green and red algae, the chloroplast envelope and the ER are known to be distinct glycerolipid factories that cooperate to produce a variety of polar lipid classes composing membrane compartments (in the form of diacyl-glycerolipids) and the nonpolar storage lipid, triacylglycerol (TAG). Since the complex plastid of diatoms combines the membranes deriving from the chloroplast envelope and membranes related to the endomembrane system, the understanding of glycerolipid biosynthesis requires in-depth biochemical analyses and refined subcellular localization. In Arabidopsis, all phosphoglycerolipids, like phosphatidylcholine (PC), phosphatidylethanolamine (PE), and phosphatidylserine, are exclusively synthesized in the ER, except that phosphatidylglycerol (PG) is also synthesized in the chloroplast envelope. In contrast, phosphorus-free galactoglycerolipids, that is, mono- and digalactosyldiacylglycerol (MGDG and DGDG) and the sulfolipid sulfoquinovosyldiacylglycerol (SQDG), are synthesized in the primary chloroplast envelope (Boudière et al., 2014). Precise localization of lipid synthesis is difficult in diatoms owing to a physical coupling of the ER and cERM. It is reported that three acyl-CoA: diacylglycerol (DAG) acyltransferases (DGATs) are localized in the cERM (Zhang et al., 2021), and the lipid droplet loaded with TAG was shown to be tightly bound to the cERM surface (Leyland et al., 2020; Lupette et al., 2019), and thus TAG biosynthesis may also be carried out there. Further analyses are necessary to confirm the production sites of lipids in diatoms and address the physical connection with glycerolipid-synthesis machineries.

To achieve a more conclusive understanding of the special structure and metabolic pathways in diatom plastid, it is necessary to study its protein composition. The easiest way to possibly identify plastid proteins is bioinformatic prediction. The HECTAR algorithm was designed to predict subcellular targeting in heterokonts (Gschloessl et al., 2008), and ASAFind was more specifically developed for diatom plastid protein prediction (Gruber et al., 2015), however, obtained predictions alone may not be sufficient and should be verified by experiments (Schober et al., 2019). Experimental analysis of subcellular localization by expressing proteins fused to a fluorescent protein (FP) or preparing antibodies for immunoelectron microscopic observation is time-consuming. Proteomic detection of isolated chloroplasts is a routine method in plants, and over 2000 chloroplast proteins have been identified and compiled into the Plant Proteomics Database (Sun et al., 2009) and ChloroKB (Gloaguen et al., 2017). A similar approach has been employed in *Chlamydomonas reinhardtii* (Naumann et al., 2007) and *Thalassiosira pseudonana* (Schober et al., 2019). In this study, we obtained high-quality *P. tricornutum* plastids with a four-layered membrane, and the integrity and purification were analyzed by observation of transformants expressing plastid marker proteins fused with an FP. A total of 2850 proteins were identified in the plastid-enriched fraction through a label-free quantitative proteomics approach, and the main metabolic pathways were reconstructed based on the obtained data. It is in particular revealed that the elongation and desaturation of FAs and the synthesis of lipids including TAG, usually associated with the ER in photosynthetic eukaryotes deriving from a primary endosymbiosis, are confined in the four-layered membrane, illustrating a powerful metabolic innovation following secondary endosymbiosis.

RESULTS

Plastid purification

After density centrifugation, the plastid-enriched fraction accumulated at the interface between 10 and 20% Percoll layers (Figure 1a). This interface contained mostly intact plastids, and minor whole cells and cell debris when observed by confocal microscope. Isolated plastids exhibited the same elongated shape as that observed *in vivo*, however, they tended to condense into rounded structures 2 h after the isolation (Figure 1b). To evaluate the integrity of the four membranes in isolated plastids, we constructed transformants in which an FP was targeted to each of the four membranes and the PPC of the plastid respectively. Major intrinsic protein 1b (MIP1b) (Liu et al., 2016), symbiont Derlin1-2 (sDer1-2) (Hempel et al., 2009), heat shock protein 70 (HSP70) (Gould et al., 2006), and outer membrane protein 85 (Omp85) (Bullmann et al., 2010), which were localized to the cERM, PPM, PPC, and OEM respectively, were fused with mRFP. Triose phosphate translocator 4a (TPT4a), localized to the IEM (Moog et al., 2015), was
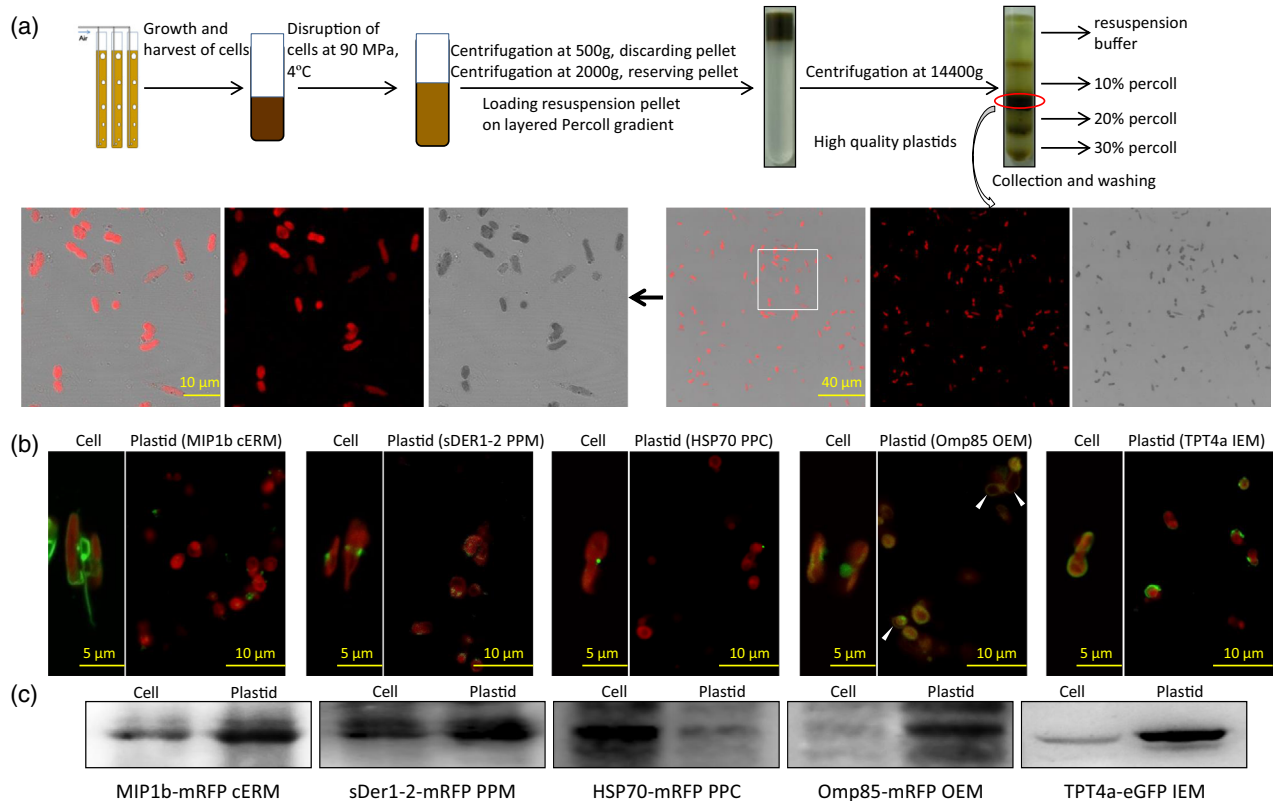
Figure 1. Integrity analysis of isolated plastids from *Phaeodactylum tricornutum*. (a) Isolation workflow, position in Percoll gradient (indicated by a red circle), and fluorescence micrographs of plastids. (b) Fluorescence micrographs of whole cells and isolated plastids from transformants expressing plastid marker proteins fused with mRFP or eGFP (from left to right: MIP1b, chloroplast endoplasmic reticulum membrane protein; sDER1-2, periplastidial membrane protein; HSP70, periplastidal compartment protein; Omp85, outer envelope membrane protein; TPT4a, inner envelope membrane protein). (c) Western blot analysis of the abundance of the fluorescence protein in whole cells and isolated plastids. Red, chlorophyll autofluorescence. Green, fluorescence protein signal. Arrowheads in (b) indicate the loop-like structure.

fused with eGFP (Figure 1b; Figures S1 and S2). In isolated plastids of transformants expressing MIP1b-mRFP, a few dotted mRFP-signals were observed. Plastids isolated from sDer1-2-mRFP transformants carried a "blob-like" structure around the red chlorophyll autofluorescence, consistent with an intact PPC vesicular network (Flori et al., 2016). There were less dotted mRFP-signals close to the isolated plastids of cells expressing HSP70-mRFP. In isolated plastids from cells expressing Omp85-mRFP, the mRFP-signals surrounded the red fluorescence, and the green mRFP fluorescence could also expand to a loop-like structure. Apparently, the OEM was much bigger than the chloroplast. An oversized OEM was observed when the cERM and PPM were ruptured during the isolation of plastid (Figure S3). A few 'circular' eGFP signals were observed in plastids isolated from transformants expressing TPT4a-eGFP, while most eGFP signals took the shape of an incomplete ring. Western blot results showed that the relative abundance of the FPs targeted to plastid membranes was much higher in the isolated plastid samples than that of whole-cell samples, suggesting that most plastid membranes remained attached to the plastid. However, HSP70 in the isolated

plastid samples was lower, which indicated that the PPC-located soluble protein was lost after plastid isolation (Figure 1c). As demonstrated in the fluorescence images, HSP70-mRFP exhibited the lowest fluorescence intensity. Overall, the fluorescence intensity of the protein targeted to plastid membranes tended to decrease gradually from the innermost to the outermost (Figure S2). This suggests that the integrity of the four membranes may vary, the inner membranes being more resistant to the purification procedure, the outermost one being evidently interrupted.

In order to determine the purity of the isolated plastids, we used specific antibodies raised against subcellular protein markers and analyzed protein extracts from whole cells and the plastid-enriched fraction by western blot. Equal amounts of protein samples were separated on gels and then stained (Figure 2a). Plastid samples showed apparently less protein bands than whole-cell samples, while a distinct band with much higher abundance was present in plastid samples. This band of about ~16 kDa corresponded to FCPs in *P. tricornutum* (Herbstová et al., 2015). Based on the western blot analysis (Figure 2b), TPT1, a cERM protein (Moog et al., 2015), was only
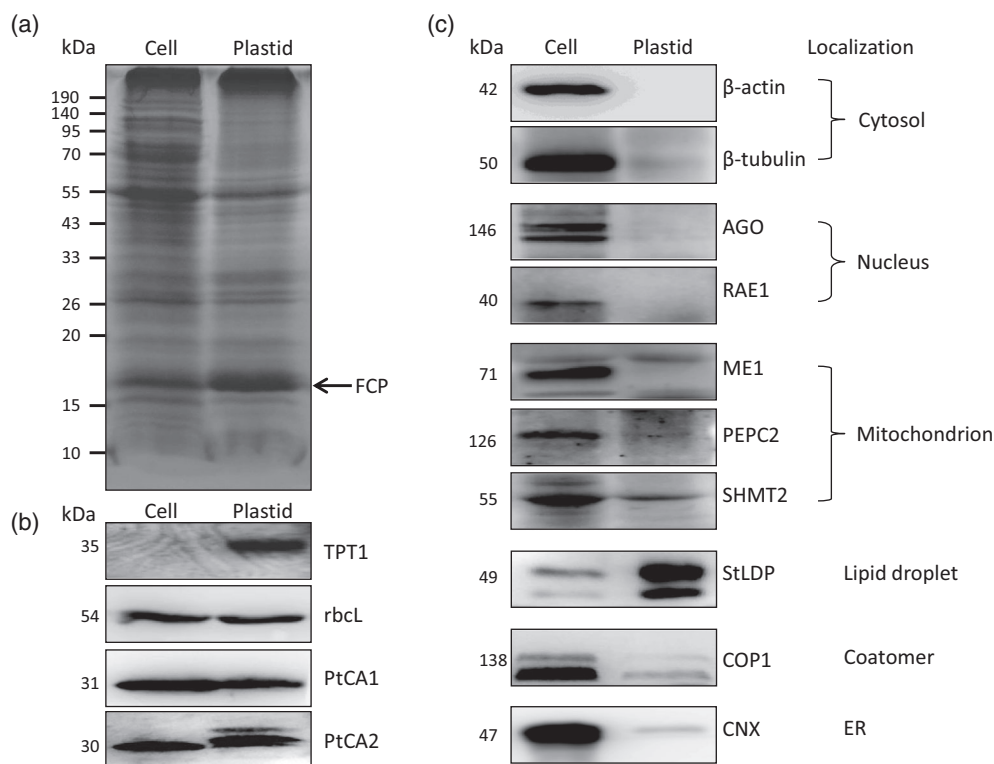
Figure 2. Western blot analysis of protein extracts from whole cells and isolated plastids. (a) Coomassie brilliant blue staining. (b) Immunoblot with antibodies against plastid proteins. (c) Immunoblot with antibodies against non-plastid proteins. AGO, argonaute protein; CNX, ER-targeted calnexin; COP1, coatomer subunit alpha; ER, endoplasmic reticulum; FCP, fucoxanthin chlorophyll a/c-binding protein; ME1, NAD-dependent malic enzyme; PEPC2, phosphoenolpyruvate carboxylase 2; PtCA1, β-carbonic anhydrase 1; PtCA2, β-CA2; RAE1, ribonucleic acid export factor 1; rbcL, ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit; SHMT2, serine hydroxymethyltransferase 2; StLDP, stramenopile lipid droplet protein; TPT1, triose phosphate translocator 1.

detected in plastid samples. The relative abundance of three pyrenoid-localized proteins, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (rbcL), PtCA1 (β-carbonic anhydrase 1; Tanaka et al., 2005), and PtCA2 (β-CA2; Kitao et al., 2008), in plastid samples was comparable to that in whole-cell samples. Two cytosolic protein antibodies (β-actin and β-tubulin) and two nuclear protein antibodies (AGO: argonaute protein; RAE1: ribonucleic acid export factor 1) did not show any cross-contamination in the plastid samples. Immunodetection with antibodies against mitochondrial proteins (ME1: NAD-dependent malic enzyme; PEPC2: phosphoenolpyruvate carboxylase 2; SHMT2: serine hydroxymethyltransferase 2) (Ewe et al., 2018) showed either no band (ME1 and PEPC2) or a faint detection (SHMT2) in plastid samples. Interestingly, the antibody raised against the stramenopile lipid droplet protein (StLDP) reacted strongly with a band in the isolated plastid sample, while a very weak band was detected using antibodies against the coatomer subunit alpha (COP1) and ER-targeted calnexin (CNX) (Figure 2c). According to the immunodetection results, the isolated plastids are unlikely to contain any residue of nuclear and coatomer compartments and might

contain a relatively small contamination by mitochondrion and ER. Considering that the cERM is contiguous with the ER and that the plastid and mitochondria of *P. tricornutum* cells are closely associated with one another (Flori et al., 2017; Uwizeye et al., 2021), such limited contamination is therefore unavoidable. Similarly, low amounts of mitochondrial or cellular contaminations (cell wall fragments, lipid bodies, or peroxisomes) occurred in the isolated plastids of *T. pseudonana* (Schober et al., 2019).

Quantitative proteomic analysis

Label-free quantitative proteomic analyses of whole cells and isolated plastids of wild-type *P. tricornutum* were performed using three independent biological replicates (named Plastid 1–3 and Cell 1–3). A total of 3837 proteins were identified (Data S1), and among them, 135 were organelle-encoded proteins (NCBI), 3680 proteins were annotated in the Ensembl Protists database, and 22 proteins (including a mitochondrial-encoded protein) were annotated by Yang et al. (2018) as "novel proteins" in *P. tricornutum*. In the three whole-cell samples, 3470 proteins were identified and among them, 987 proteins were only detected in whole-cell samples (Figure 3a). In the
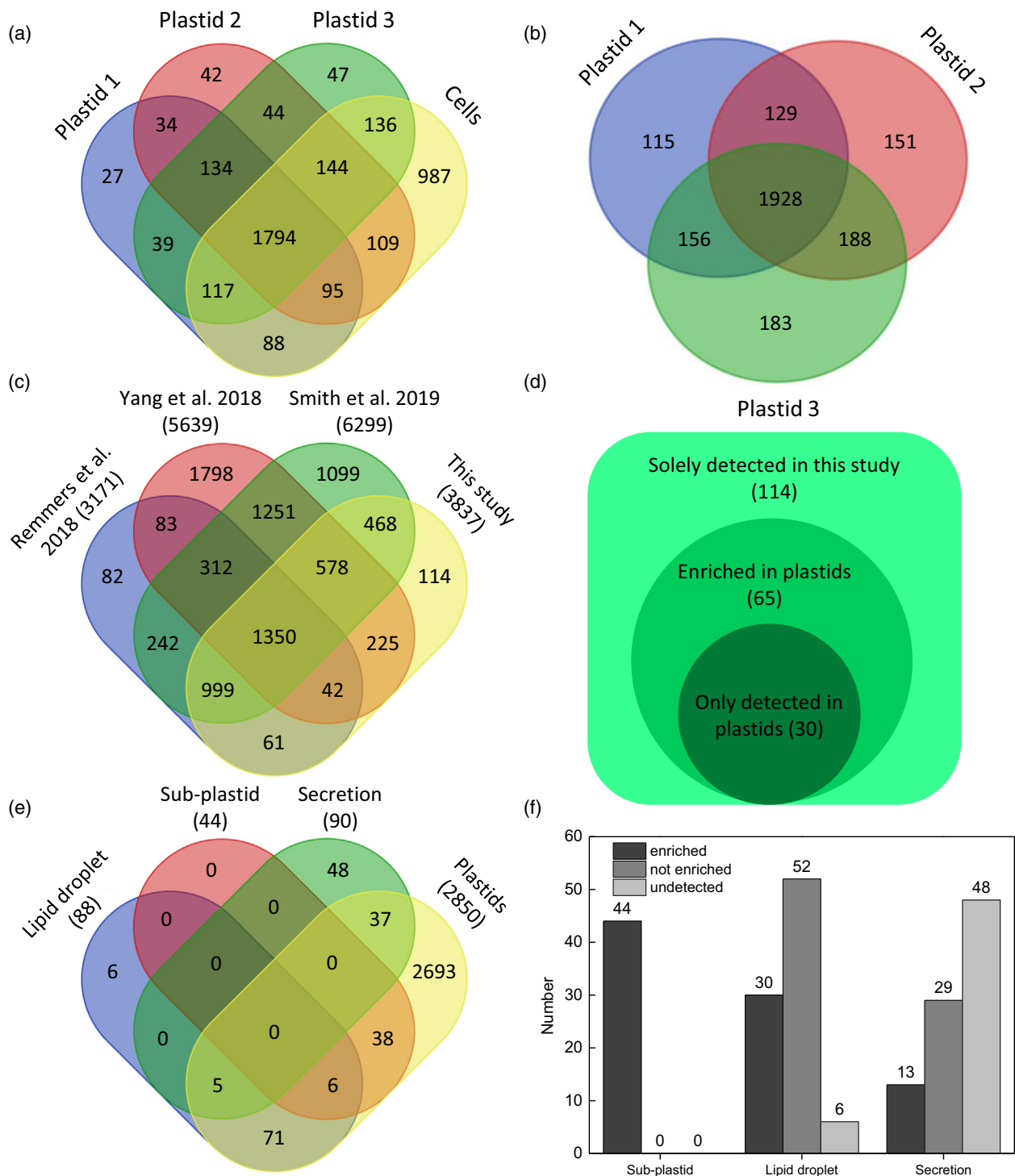
Figure 3. Comparison of proteins detected in the isolated plastids and whole cells, and proteins reported before in *Phaeodactylum tricornutum*. (a) Venn diagram representation of absolute protein numbers detected in the whole cells and three plastid samples. (b) Comparison of protein identified in the individual plastid sample. (c) Comparison of detected proteins in recent proteome studies. (d) Proteins that were solely detected in this study. (e) Comparison of proteins detected in our putative plastid proteomes with those in subcellular components reported before (Sub-plastid: Lepetit et al., 2007, Joshi-Deo et al., 2010, Grouneva et al., 2011, Gundermann et al., 2013, Herbstová et al., 2015, Levitan et al., 2019; Lipid droplet: Yoneda et al., 2016, Lupette et al., 2019; Secretion: Bruckner et al., 2011, Buhmann et al., 2016, Erdene-Ochir et al., 2019). (f) Abundance of previously reported subcellular component proteins in our putative plastid proteomes (Enriched: fold change values >1 or only detected in one plastid sample; Not enriched: fold change values <1; Undetected: not detected in our putative plastid proteomes).

three isolated plastid replicates, 2328, 2396, and 2455 proteins were identified respectively, and a total of 2850 proteins were identified, of which 1928 proteins were shared by all three samples (Figure 3b). Interestingly, 367 proteins (Data S2) were only detected in plastid samples, among which 134 proteins were detected in all three samples (Figure 3a).

Recently, 3171 proteins (no organelle-encoded proteins involved) from 14 samples grown under four conditions (Remmers et al., 2018), 5639 proteins from 45 samples grown under eight conditions (Yang et al., 2018), and 6299 proteins (no organelle-encoded proteins involved) from 15 samples under 15 conditions (Smith et al., 2019) have been identified in *P. tricornutum* by mass spectrometry. Consistent with the minimal threshold for identification mentioned in the three above reports, proteins identified by at least a single peptide were included in our study. The number of identified proteins seems to be proportional to the number of culture conditions and samples. In our study, 6 samples cultivated under a single condition were analyzed and yet we identified 114 proteins that had not been detected before (Figure 3c). Furthermore, 65 out of the 114 proteins were enriched (fold change values >1 or only detected in one plastid sample) in plastid samples (Data S3), and 30 out of the 67 proteins were only detected in plastid samples (Figure 3d).

We also compared our putative plastid proteomes with those in subcellular components reported before in *P. tricornutum*. All the 44 sub-plastid proteins previously reported (Grouneva et al., 2011; Gundermann et al., 2013; Herbstová et al., 2015; Joshi-Deo et al., 2010; Lepetit et al., 2007; Levitan et al., 2019), 82 out of 88 lipid droplet proteins (Lupette et al., 2019; Yoneda et al., 2016), and 42 out of 90 secretory pathway proteins (Bruckner et al., 2011; Buhmann et al., 2016; Erdene-Ochir et al., 2019) were detected in this study (Figure 3e; Data S4). These 44 sub-plastid proteins were identified with at least a single peptide (Grouneva et al., 2011; Herbstová et al., 2015; Lepetit et al., 2007) if not three different peptides (Gundermann et al., 2013; Joshi-Deo et al., 2010). In our study, all the reported sub-plastid proteins were enriched in the plastid samples (Figure 3f). For the lipid droplet proteomes, proteins identified with at least three different peptides were included (Lupette et al., 2019); and for the secretory pathway protein identification, at least a single peptide was matched to the protein (Buhmann et al., 2016). However, 30 lipid droplet proteins (over one-third of them were apparently plastid-targeted proteins) (Lupette et al., 2019) and 13 secretory pathway proteins (all were unknown proteins) were also enriched in our plastid samples (Figure 3f).

Organelle-encoded proteins in the plastid-enriched fraction

In our proteome data, 107 plastid-encoded and 29 mitochondrial-encoded proteins were identified, and 92 out of the 107 and 20 out of the 29 proteins were present in the plastid samples (Figure 4a; Data S1). The absence of the 15 plastid-encoded proteins that have no transmembrane
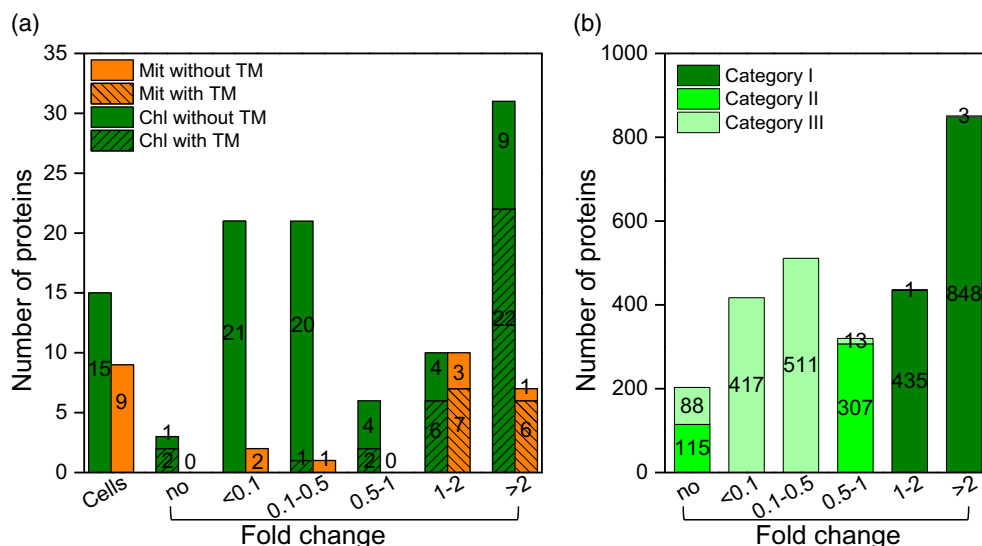

Figure 4. Fold change values (plastid/cell) of detected organelle-encoded proteins and plastid samples' nuclear-encoded proteins. (a) Organelle-encoded proteins detected in all samples. (b) Nuclear-encoded proteins detected in plastid samples. Fold change: The ratio of protein relative abundances of plastid versus those of whole cells in proteomes. Cells: organelle-encoded proteins only detected in whole cell samples; no: proteins detected in only one plastid sample or detected in one plastid and one whole-cell samples. Mit with/without TM: Mitochondrial-encoded protein with/without transmembrane domains (TM); Chl with/without TM: plastid-encoded proteins with/without TM. Category I: detected in at least two plastid samples with fold change values >1; Category II: only detected in one plastid sample or detected in at least two plastid samples with fold change values ranging from 0.5 ~ 1; Category III: all the rest proteins present in plastid samples.

domains indicated that soluble proteins in the plastid leaked out easily during the plastid isolation. Among the 92 plastid-encoded proteins, 42 proteins were enriched in plastid samples; in particular, 29 out of the 42 enriched proteins had transmembrane domains, and the other 13 included 5 components of the ATP synthase complex, 3 PS I proteins (Psa), and 5 other known plastid proteins. The fold change values for the relative abundance of 42 proteins from the 92 proteins were blew 0.5, and only one in the 42 proteins contained transmembrane domains, which further supported the loss of soluble proteins during plastid isolation. Thirteen proteins in the 20 mitochondrial-encoded proteins present in our putative plastid proteome had transmembrane domains, indicating a close association between plastid and mitochondria.

Filtering nuclear-encoded plastid proteins

Since the cell fractionation process cannot eliminate all non-plastid fractions and mass spectrometry is a highly sensitive technique that can easily detect low-abundance proteins (Elias et al., 2005), we cannot conclude that all 2738 nuclear-encoded proteins from our putative plastid proteomes (Data S5) are exclusively plastid-targeted proteins. To further estimate the quality of putative plastid proteome, protein sequences from all detected proteins were blasted against Arabidopsis proteome database in PPDB with an *E*-value cutoff of 10^{-10}, and a total of 1673 detected proteins in our proteomes could be matched to the database and were assigned the localization (Data S1). Based on the protein intensities in our proteome data, the enrichment factors were calculated by dividing the percentage of matched proteins in plastids by that of whole cells (Data S6 and S7). Only assigned plastid proteins (including plastid-encoded proteins) have the enrichment factor of above 1 (1.353), and non-plastid proteins have the enrichment factor of 0.471. The assigned mitochondrial proteins (including mitochondrial-encoded proteins) have the enrichment factor of 0.996. Considering the assigned localization might not be so accurate, we compared the assigned localization with the experimentally determined localization in *P. tricornutum* (Data S8). Among the 146 assigned nuclear-encoded mitochondrial proteins (Data S7), 12 are of known localization (Data S8): 5 are reportedly mitochondria-localized, 5 are plastid-localized, 1 is cytosol-localized, and 1 is vacuolar membrane-localized. Therefore the actual enrichment factor of mitochondrial proteins should be lower. We calculated the enrichment factor of all proteins with known localization, and the enrichment factor for plastid proteins and non-plastid proteins was 1.211 and 0.252 respectively, while for the mitochondrial proteins, it was 0.133 (Data S6 and S7). These experimentally determined mitochondrial proteins are all soluble proteins (Data S9), and considering the typical diatom ultrastructure (mitochondria sticking to plastids), it is

the mitochondrial membrane proteins that were co-enriched with the isolated plastids, indicating a strong contamination of mitochondrial membrane proteins, especially the mitochondrial-encoded ones.

The enrichment factor of the assigned plasma membrane proteins is 0.929 (Data S6 and S7), however, the assigned plasma membrane localization by blasting against Arabidopsis proteome database in PPDB might not be accurate any more in diatoms, the secondary endosymbionts. Diatoms contain a secondary plastid that derives from a red algal symbiont, and the plastid is surrounded by four membranes. The second outermost plastid membrane, PPM, is derived from the plasma membrane of the former endosymbiont (Grosche et al., 2014), and some of the assigned plasma membrane proteins might therefore become plastid proteins after the secondary endosymbiosis in diatoms. In fact, among the 95 proteins (Data S1 and S7) assigned with plasma membrane localization by PPDB, 55 proteins contain no transmembrane domains any more. Furthermore, among the 95 assigned plasma membrane proteins, 6 are of known localization (Data S8 and S9), and only J20755 was dual-targeted to the plasma membrane and cERM (the outermost membrane of plastid) (Matsui et al., 2018) but was found in the cERM by Liu et al. (2016). In addition, J44871 was localized in the PPM and J9617 in the vacuolar membrane. The rest three are not membrane proteins and were localized in the nucleus, vesicles, and cytosol. Therefore, though the enrichment factor of the assigned plasma membrane proteins is high, we cannot conclude that *P. tricornutum* plasma membrane proteins were also co-enriched with the isolated plastid.

To discriminate potential plastid-targeted proteins, all identified nuclear-encoded proteins from putative plastid proteomes were firstly classified according to fold change values (plastid/cell) (Figure 4b) to Categories I, II, and III (Data S5). The fold change values (plastid/cell) were calculated based on the normalized protein intensity by the total intensity in each sample. Except proteins detected in only one plastid sample or detected in one plastid and one whole-cell samples (without the fold change values), 47% of the proteins (1287) had the fold change values higher than 1, indicating that almost half of the detected nuclear-encoded proteins (2738) were enriched efficiently in the plastid samples. Based on the fold change values, Category I includes 1283 proteins detected in at least two plastid samples with fold change values >1, Category II includes 422 proteins only detected in one plastid sample or detected in at least two plastid samples with fold change values ranging from 0.5 to 1, and Category III includes all the rest 1033 proteins present in plastid samples. Prediction of the Category I proteins by ASAFind and HECTAR showed that 25.1% (322) and 12.7% (163) proteins were chloroplast-targeted, while 19.0% (277) and 10.4% (151) of the Categories II and III proteins (1455) from plastid

samples were predicted to be chloroplast-targeted as well (Figure S4). For the predicted 12382 nuclear-encoded proteins in *P. tricornutum*, only 1471 and 561 proteins were considered as chloroplast-targeted by ASAFind and HEC-TAR methods (Figure S5). Apparently the number of predicted chloroplast-targeted proteins is underestimated considering the complexity of the secondary plastid in diatoms relative to the primary plastid, and about 3000 (Tardif et al., 2012) and over 2245 (Wang et al., 2023) were predicted to be chloroplast proteins respectively in model land plant *A. thaliana* and model green alga *C. reinhardtii*. Relatively speaking, the prediction tool ASAFind was more specifically for diatom plastid protein prediction (Gruber et al., 2015), and enriched proteins in the isolated plastid samples have a higher possibility of being chloroplast-targeted. The enrichment factor of predicted plastid proteins by ASAFind and plastid-encoded proteins is 1.468 (Data S6 and S7). Therefore, to identify the potential plastidial proteins in our proteomes ASAFind prediction and BLAST alignments within PPDB were used, and 1121 nuclear-encoded proteins were chloroplast-targeted, among which 839 were detected in our putative plastid proteomes (Data S10).

In addition, in our putative plastid proteomes, 863 proteins (including 32 plastid-encoded, 9 mitochondrial-encoded, and 822 nuclear-encoded proteins) with fold change values above 1.5 were found in all the three plastid samples and identified by at least two peptides (Data S11). Prediction of the 822 nuclear-encoded proteins by ASAFind showed that 29.4% (242) proteins were chloroplast-targeted. If statistical significance ($P < 0.05$) was taken into account, 594 proteins (including 30 plastid-encoded, 7 mitochondrial-encoded, and 557 nuclear-encoded proteins) were enriched in plastid samples (Figure S6; Data S12), and 35.9% (200) of the 557 proteins were predicted to be chloroplast-targeted by ASAFind; meanwhile 1159 nuclear-encoded proteins were depleted, among which 19.7% (228) proteins were predicted to be chloroplast-targeted.

Metabolic pathways in diatom plastid compared to known primary chloroplasts

Based on KEGG pathways generated from the putative plastid proteomes (Figure S7; Data S13) and bioinformatic predictions (Data S14), *P. tricornutum* plastid hosted more metabolic processes than the essential ones identified in plant chloroplasts. As expected and like plant chloroplasts, the *P. tricornutum* plastid carried out photosynthesis, chlorophyll and carotenoid biosynthesis, glycan metabolism and *de novo* FA synthesis, etc. However, biosynthesis of branched-chain amino acids (BCAAs), *de novo* NAD, and nucleotide (except ATP) were not detected in the plastid samples (Data S14). Surprisingly, the whole glycolysis pathway, the ancient Entner–Doudoroff glycolytic pathway, all steps of FA desaturation and elongation, lipid

biosynthesis, and TAG hydrolysis might also take place in the plastid. In addition, though mitochondrion is the main place for FA β-oxidation (Jallet et al., 2020), the possibility of a plastidial FA β-oxidation could not be excluded based on the 16 FA β-oxidation-related enzymes detected in the plastid samples and on the prediction (Data S14). Additional details of all identified proteins in our proteomes were shown in Appendix S1 and Data S14, and the chloroplast-endomembrane "hybrid" metabolism in diatom complex plastid was discussed in Appendix S2.

Identified plastid proteins involved in FA and lipid biosynthesis

De novo FA synthesis is largely similar between diatoms and plants, and it occurs in the stroma of plastids by the concerted action of acetyl-CoA carboxylase (ACCase) with a dissociated complex of enzymes forming the FA synthase of type II. Acetyl-CoA, the precursor of FA synthesis, can be generated by pyruvate dehydrogenase complex (PDC) or acetyl-CoA synthetase in the plastid, and all components of the former were detected in the putative plastid proteomes while the latter was only detected in whole cells. Two ACCases, 1 malonyl-CoA:ACP transacylase (MCAT), 1 ketoacyl-ACP synthase, 2 ketoacyl-ACP reductases, 1 hydroxyacyl-ACP dehydrase, and 1 enoyl-ACP reductase were present in the plastid samples. Acyl-ACP thioesterase (ptTES1) catalyzing the removal of the acyl group from ACP, a terminal step of FA synthesis, was also detected in the putative plastid proteomes. Long-chain acyl-CoA synthetase (ptACSL) catalyzes the production of acyl-CoA pool using free FAs from acyl-ACP hydrolysis as substrates, and four ptACSLs were enriched in the plastid samples (Data S14).

Except palmitoyl (16:0)-ACP desaturase (PAD), a soluble FA desaturase, which was only detected in the whole-cell samples, all detected FA desaturases (FADs) (ADS: Δ9 acyl-CoA desaturase; OEMΔ6FAD; PtFAD2; ERΔ5FAD.1), FA elongase (ELO) (Δ6ELO.2), and enzymes involved in Kennedy-like pathway (GPAT: glycerol-3-phosphate acyltransferase; LPAAT: lysophosphatidic acid acyltransferase, EG02461; ATS2a: plastidic LPAAT) were present in the plastid samples. In addition, two enzymes (DGAT2B and phospholipid:DAG acyltransferase, PDAT) of the last step for TAG biosynthesis were enriched in the plastid samples. Enzymes of the final steps leading to the formation of PG (PGPP: phosphatidylglycerolphosphate phosphatase), MGDG (MGDGS: MGDG synthase), DGDG (DGDGS: DGDG synthase, J43116), SQDG (SQD2: sulfoquinovosyltransferase), and diacylglycerylhydroxymethyltrimethyl-β-alanine (DGTA) (BTA: betaine lipid synthase) were also found in the plastid samples. Furthermore, an acyl-editing enzyme acyl-CoA:lysophospholipid acyltransferase (LPLAT: J20460) and three triacylglycerol lipases (TGL, one of the 3 was assigned plastid protein by BLAST alignments within PPDB) were

present in the plastid samples, suggesting that the plastid could also be involved in FA remobilization and remodeling.

To verify the subcellular localization of enzymes involved in FA (Figure 5) and lipid biosynthesis (Figure 6), the expression of GFP fusion proteins in *P. tricornutum* were visualized by confocal microscopy. MCAT, encoded by a single copy gene, was apparently localized on the thylakoid membrane according to the MCAT-eGFP fluorescence signals present as spots in the plastid (Liu et al., 2016). After de novo FA synthesis, PAD (Smith et al., 2021)

and FAD4 act on 16:0-ACP and 16:0-sn2-PG respectively to generate 16:1$^{\Delta9}$-ACP and 16:1$^{\Delta3trans}$-sn2-PG, which occurred in the plastid stroma based on the GFP fusion protein signals. Ordered addition of double bonds from 16:1$^{\Delta9}$ to 16:2$^{\Delta9,12}$, 16:3$^{\Delta6,9,12}$, and 16:4$^{\Delta6,9,12,15}$ were catalyzed by PtFAD6 (Domergue et al., 2003), OEMΔ6FAD, and ω3FAD respectively, and the three enzyme-eGFP fusion signals distributed all over the plastid with a high accumulation in the middle of the plastid, indicating the OEM localization (Li, Pan, et al., 2018). ADS, acting on 18:0-CoA to form 18:1$^{\Delta9}$-CoA, fused GFP fluorescence formed irregular layers and

Figure 5. Subcellular localization of enzymes involved in fatty acid biosynthesis. Panels show microscopical images of transmitted light, chlorophyll autofluorescence (red), GFP fluorescence (green), merged images of chlorophyll and GFP fluorescence from left to right. ADS, Δ9 acyl-CoA desaturase; ELO, fatty acid elongase; FAD, fatty acid desaturase; MCAT, malonyl-CoA:ACP transacylase; PAD, palmitoyl-ACP desaturase. Scale bar = 5 μm.
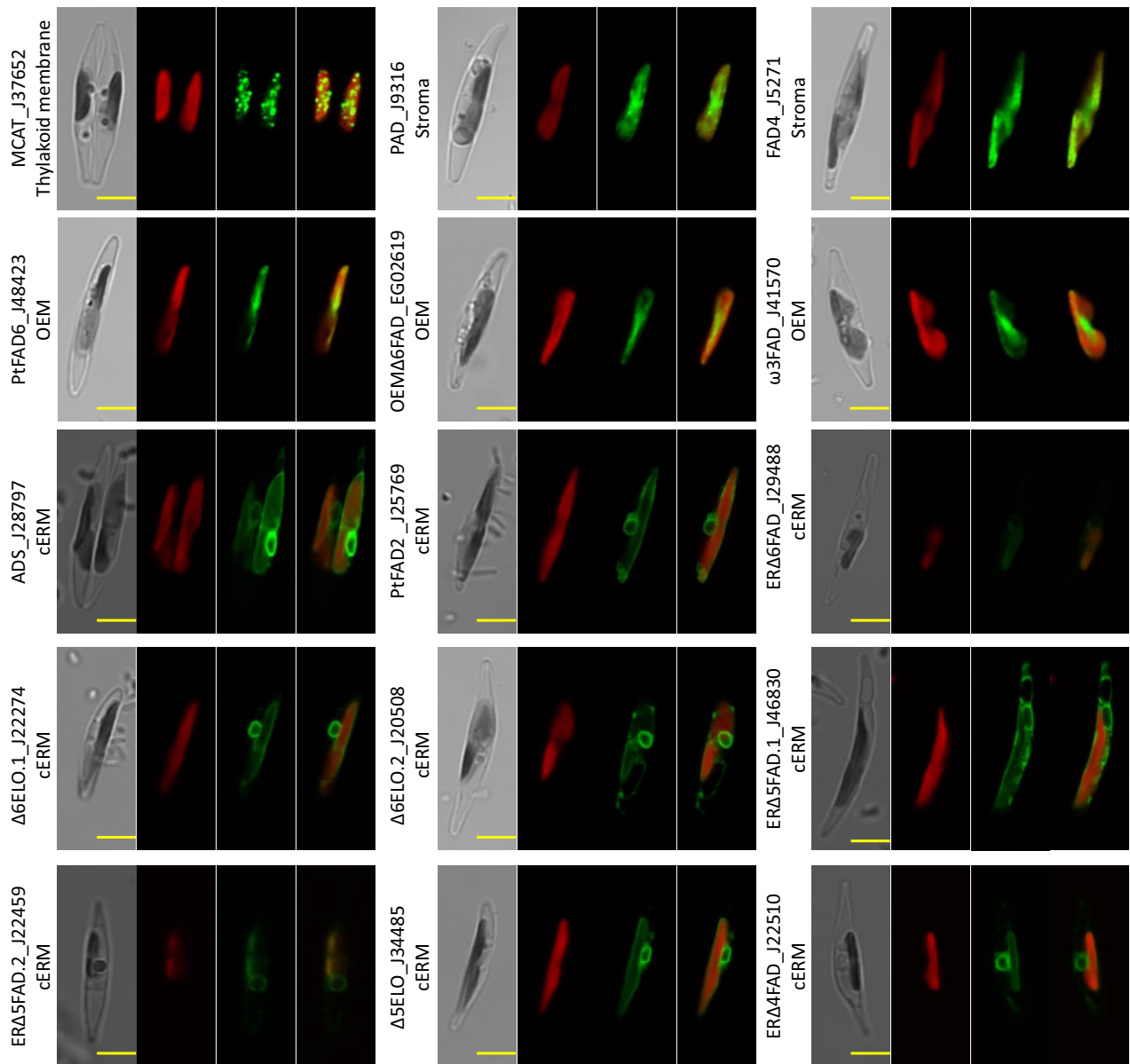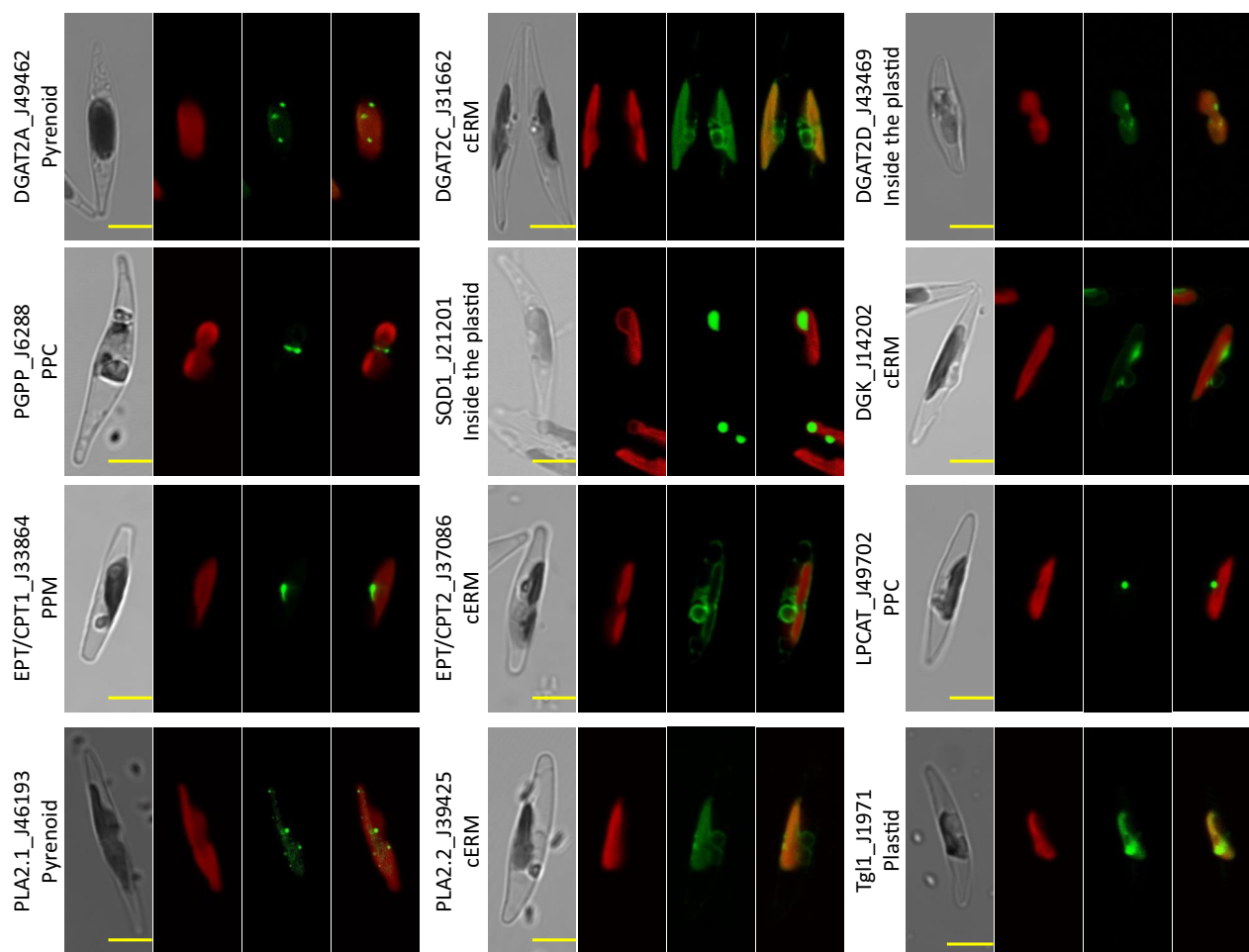
Figure 6. Subcellular localization of enzymes involved in lipid metabolism. Panels show microscopical images of transmitted light, chlorophyll autofluorescence (red), GFP fluorescence (green), merged images of chlorophyll and GFP fluorescence from left to right. DGAT, acyl-CoA:diacylglycerol acyltransferase; DGK, diacylglycerol kinase; EPT/CPT, ethanolamine−/choline-phosphotransferase; LPCAT, lysophosphatidylcholine acyltransferase; PGPP, phosphatidylglycerolphosphate phosphatase; PLA2, phospholipase A2; SQD1, UDP-sulfoquinovose synthase; Tgl1, SDP1-patatin like lipase. Scale bar = 5 μm.

surrounded the plastid and nucleus (Figure S8), resembling the localization pattern of the cERM-localized protein (Hempel et al., 2009). Ordered addition of double bonds from $18:1^{\Delta9}$ to $18:2^{\Delta9,12}$, $18:3^{\Delta6,9,12}$/$18:3^{\Delta9,12,15}$, and $18:4^{\Delta6,9,12,15}$ were catalyzed by PtFAD2 (Domergue et al., 2003), ERΔ6FAD (previously described as PtD6, Domergue et al., 2002)/ ω3FAD, and ERΔ6FAD respectively, and these three enzyme-eGFP fusion fluorescence signals showed the same pattern with ADS-eGFP, indicating a cERM localization. In addition, 2 Δ6ELO, 2 ERΔ5FAD (ERΔ5FAD.1 previously described as PtD5, Domergue et al., 2002), 1 Δ5ELO, and 1 ERΔ4FAD involved in elongation from $18:4^{\Delta6,9,12,15}$ to $20:4^{\Delta8,11,14,17}$, desaturation from $20:4^{\Delta8,11,14,17}$ to $20:5^{\Delta5,8,11,14,17}$ (eicosapentaenoic acid, EPA), elongation from EPA to $22:5^{\Delta7,10,13,16,19}$, and desaturation from $22:5^{\Delta7,10,13,16,19}$ to $22:6^{\Delta4,10,13,16,19}$ (docosahexaenoic acid, DHA), also resided at the cERM.

Six DGATs have been identified in *P. tricornutum*, and three of them (DGAT1, DGAT2B, and DGAT3) were reported to reside at the cERM (Zhang et al., 2021). DGAT2C was also localized at the cERM, while DGAT2A and DGAT2D were apparently localized in the plastid (Figure 6), and the former appeared as punctate spots resembling the pyrenoid-localized fructose 1,6-bisphosphate aldolases (FBAC1 and FBAC5) (Allen et al., 2012). Four chloroplast lipids (MGDG, SQDG, DGDG, and PG) together with PC are the main lipids in *P. tricornutum* (Abida et al., 2015). PGPP and SQD1 involved in the biosynthesis of PG and SQDG, respectively, were chosen to clarify the lipid synthetic site. PGPP-eGFP showed a typical dot structure in the middle of the plastid ("blob-like" structure), indicating a PPC localization (Kilian & Kroth, 2005), and SQD1-eGFP fluorescence appeared as big spots (usually only one or two) within the red fluorescence. Two ethanolamine-/choline-phosphotransferases EPT/CPT1 and EPT/CPT2 (the bifunctional enzyme responsible for the last biosynthetic step of PE and PC) were localized at the PPM

and cERM, respectively. Lands cycle enzymes acyl-CoA: lysophosphatidylcholine acyltransferase (LPCAT) and phospholipase A2 2 (PLA2.2) resided at the PPC and cERM respectively, and PLA2.1-eGFP demonstrated a similar fluorescence pattern to DGAT2A with a pyrenoid localization. Diacylglycerol kinase (DGK), catalyzing DAG conversion into phosphatidic acid, was localized at the cERM. Although the precise localization of Tgl1, a SDP1-patatin-like lipase (Barka et al., 2016), in the plastid was uncertain, Tgl1-eGFP signal appeared inside the plastid (Figure 6).

Lipidomics of whole cells and plastids

About 15 FAs were detected in *P. tricornutum*, and the FA compositions of total lipids and TAGs were similar in the whole cells and isolated plastids with C14:0, C16:0, C16:1, and EPA as the main FAs (Figure 7). No significant difference in the relative abundance of C14:0, C16:2, C16:3, C16:4, C18:3, and C20:4 was found between the whole-cell and plastid samples ($P > 0.05$), while in the total lipid of plastid samples, the abundance of C16:0, C18:0, C18:4, and EPA were higher and that of C16:1, C18:1, C18:2, and C22:6 were lower compared with whole cells ($P < 0.05$). In TAGs, only C16:0, C18:1, and C24:0 showed a slight difference ($P < 0.05$) in the abundance. These minor differences might simply reflect that the plastid is quantitatively the major membrane compartment within the cell, representing the

main source of polar lipids and that it is also the main biogenic platform for TAG.

The molecular species in lipids of *P. tricornutum* shared a similar profile in whole-cell and plastid samples (Figure S9; Data S15). MGDG is the most abundant lipid in *P. tricornutum* with 20:5/16:3 as the most abundant molecular species, and relative contents of this molecular together with most of the main species 16:1/16:2, 16:1/16:3, 16:3/16:3, and 20:5/16:4 showed no difference in whole-cell and plastid samples ($P > 0.05$). DGDG, synthesized from MGDG, contained two main species 20:5/16:1 and 20:5/16:2 whose relative contents showed no difference between whole cells and plastid ($P > 0.05$). Two main molecular species 16:1/16:0 and 16:1/18:1 of PC were comparable ($P > 0.05$), while the content of main PE species 20:5/22:6 and SQDG species 16:1/16:0 were only a little higher in plastid samples ($P < 0.05$). In addition, the content of two main PG species 20:5/16:1 and 20:5/16:0 were also higher in plastid samples ($P < 0.05$), while that of PG 16:1/16:0 was much lower ($P < 0.05$).

In photosynthetic eukaryotes, TAGs are usually accumulated and stored in cytoplasmic oil droplets (Ischebeck et al., 2020). *P. tricornutum* cells from 4-day culture stained with BODIPY presented clearly small green oil droplets bound to plastid (Figure S10), consistent with past studies on the dynamics of lipid droplet subpopulation formation
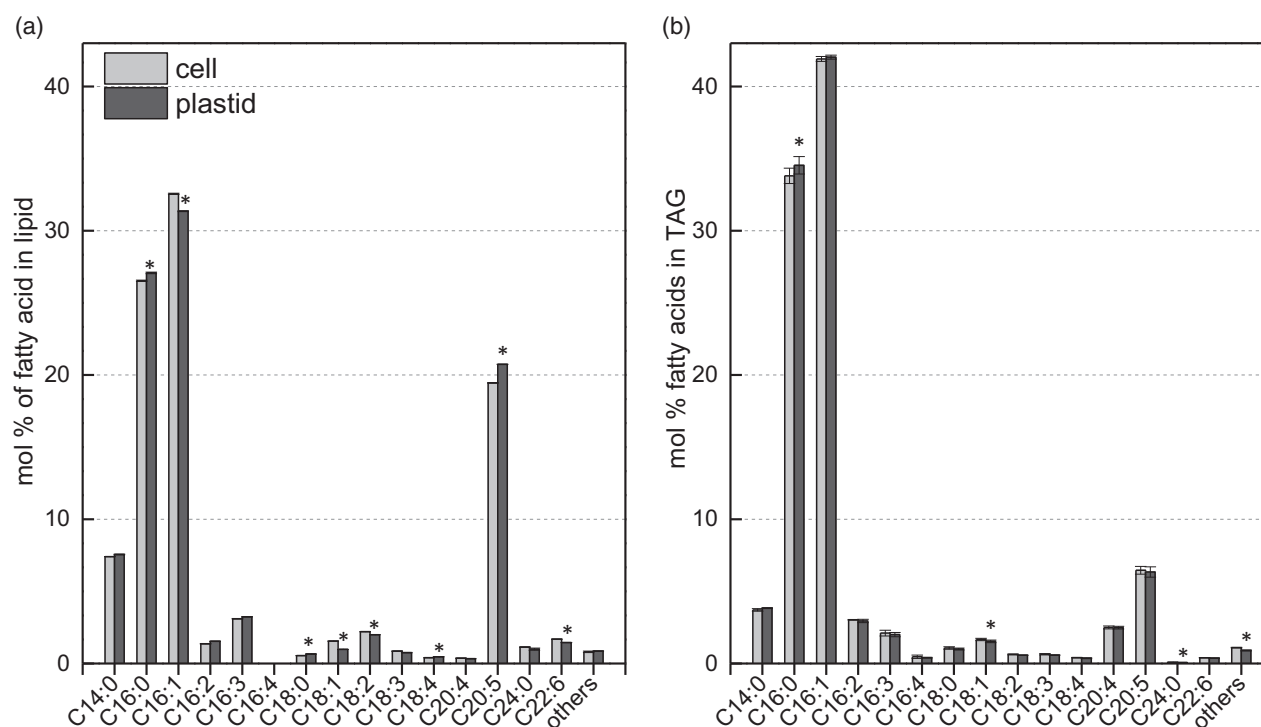

Figure 7. Relative fatty acid (FA) abundance in whole cells and isolated plastids. (a) Relative FA abundance of total FAs. (b) Relative FA abundance of triacylglycerol (TAG). Data represent mean ± SD ($n = 3$). Asterisks above the bars indicate statistically significant differences from whole cells at $P < 0.05$ level by *t*-test.

(Jaussaud et al., 2020). Furthermore, small green dots were also observed from the isolated plastids stained with BOD-IPY and a 3D visualization verified a tight connection between the green (nonpolar lipids) and red (chlorophyll) fluorescence. Electrospray ionization MS analysis revealed that no difference was observed in relative contents of the six main TAG species 16:0/16:0/16:1, 16:1/16:1/16:0, 16:0/16:1/16:2, 20:5/16:0/16:1, 14:0/16:1/16:0, and 20:5/16:1/16:1, while the two main DAG species 16:0/16:1 and 16:1/16:1 had the higher content in whole cells (Figure S9; Data S15).

DISCUSSION

Integrity and purity of isolated plastid

The cERM of isolated plastid was damaged and even the IEM was not perfectly intact. The breakage of the cERM, expected as the separation of the secondary plastid to the nucleus, relies on the breakage of the hinge connecting the cERM and the outer nuclear envelope. Consequently, some cER lumen-, PPC- and stroma-localized proteins are expected to leak through this opened area during the isolation process, as indicated by the weaker signal of HSP70 in the plastid immunoblot and by 12 soluble proteins (including HSP70) reported to be localized in the plastid but detected only in whole-cell proteomes (Data S9).

A total of 2850 proteins were detected in isolated plastids, which is over 10 times higher than the 217 proteins detected in a diatom *T. pseudonana* plastid fraction reported recently (Schober et al., 2019). 672 (52%) Category I (fold change values >1) and 259 (61%) Category II (fold change values = 0.5 ∼ 1) proteins in the putative plastid proteomes were predicted to have no transmembrane domains (Figure S11). Furthermore, among the detected 3701 nuclear-encoded proteins, about 152 proteins are of known localization (Data S9), and 66 proteins present in our putative plastid proteomes have been reported to be localized in the plastid. Over 66% (40 proteins) of the 66 proteins were predicted to be non-transmembrane proteins including 18 PPC-localized and 6 cER lumen-localized proteins, indicating that a large proportion of soluble proteins could still be preserved during the isolation. For example, PtsUBX (Stork et al., 2012), localized in the PPC, was only detected in the plastid samples with a very high fold change value of 38.6. In fact, a total of 34 proteins (20 with transmembrane domains) from the 66 proteins showed a fold change value higher than 1, indicating that these proteins were enriched in the plastid samples. The fold change values of 28 proteins (24 without transmembrane domains) were under 0.5, out of which 19 proteins exhibited a fold change value lower than 0.1 (the lowest is 0.0025 for the PPC-localized thioredoxin h, Weber et al., 2009). Therefore, the assessment of a specific protein localization in the putative plastid proteomes cannot be too

dependent on the fold change values due to the co-enrichment of mitochondrial membrane proteins and leakage of soluble plastidial proteins in the isolated plastid samples.

Accurate localization of some reported GFP-fused proteins with fluorescent signals close to autofluorescence could hardly be defined. For instance, CGHR protein Pt43232 with a reported localization in the space between chloroplast and vacuole (Kikutani et al., 2016) was detected only in our plastid samples and showed the highest fold change value (100.5) in all known localization proteins (Data S9). Pt43232 was not a predicted plastid protein by ASAFind, and we infer that this protein, localized at the contact site between chloroplast and vacuole, remains bound to the plastid surface. Several proteins of known localization at peroxisome (Gonzalez et al., 2011; Hao et al., 2022) or outside of the plastid (Moog et al., 2015; Shao et al., 2019) showed fluorescent foci close to autofluorescence and might also localize at the contact site between chloroplast and other organelles, and most of these proteins were enriched in the plastid samples (Data S9).

Our proteomes showed that 17 non-plastid proteins with experimentally confirmed subcellular localization were only detected in the whole cells (Data S9). In the plastid samples 11 cytoplasmic or/and nuclear proteins (the full-length cDNA was not used for the fluorescent fusion protein construction of four proteins) with known localization were detected, indicating that contamination of whole cells and nucleus cannot be completely eliminated. In addition, among the 17 known mitochondrion-localized proteins detected in the plastid samples, cDNA of 7 proteins (6 with fold change values lower than 0.12, only one with the fold change value 2.6) were not of full length in their studies and the rest 10 (9 with fold change values lower than 0.2, only one with the fold change value 1.3) might be the debris of mitochondria in the plastid samples (Data S9). However, FP-based targeting analyses were not always accurate. Chu et al. (2017) demonstrated that N - and C-terminal GFP constructs caused different fluorescence patterns in *P. tricornutum* nucleotide transporter (NTT) isoform 5 (*Pt*NTT5). Usually, GFP fusion to the C-terminal is performed for the subcellular localization investigations, while GFP fused C-terminally to *Pt*NTT5 gave rise to ambiguous fluorescence patterns.

Surprisingly, all detected proteins involved in putative mitochondrial electron transport complexes and mitochondrial membrane ATP synthase, including 17 mitochondrial-encoded proteins, were enriched except the cytochrome *c* oxidase subunit VIb (fold change values is 0.87) in plastid samples (Data S16; Appendix S3), reflecting that plastid and mitochondrion membranes are physical adjacent and even attached to each other, as reported before, which is advantageous to channel both electrons and ATP

between the two organelles (Bailleul et al., 2015; Uwizeye et al., 2021). Similarly, between the chloroplast and mitochondria in plants, there exists a very high molecular weight membrane contact domain containing components of the respiratory complexes (Michaud et al., 2016). Mitochondria are much smaller than plastids and are easily removed under differential centrifugation after cell disruption. The density of mitochondria is much higher than that of plastids. It is generally believed that mitochondria can easily pass through 20% Percoll after centrifugation (Millar et al., 2001; Salvato et al., 2014; Wang et al., 2015). It has also been confirmed that mitochondria of *T. pseudonana* can pass through 28% Percoll (Schober et al., 2019), whereas our collected plastids were at the interface between 10 and 20% Percoll layers, so a co-enrichment of plastids and intact mitochondria in this gradient is less likely. To verify it, *P. tricornutum* mitochondria were isolated from transformants expressing ME1-eGFP using 20% Percoll. As shown in Figure S12, the brown plastid-enriched fraction was in the upper layer of the tube, and a cream-colored band of mitochondria was present at the bottom of the tube, indicating that *P. tricornutum* mitochondria passed through the 20% Percoll layer. Immunodetection with two antibodies against mitochondrial proteins (ME1 and SHMT2) found that proteins from Fraction 4 had a strong band, and a strong band was also present there in immunodetection with antibody against GFP. Immunodetection with the antibody against a pyrenoid-localized protein PtCA1 found a band in the four fractions. Therefore, after Percoll-density centrifugation the plastid-enriched fraction is unlikely to contain the intact mitochondria at least. Altogether, as mentioned above, co-enrichment of the plastid with mitochondrial membrane and a little debris in the isolated plastid samples is unavoidable owing to numerous physical contacts between plastids and mitochondria (Bailleul et al., 2015; Uwizeye et al., 2021).

Chloroplast-endomembrane "unified" acyl-lipid machinery produces all lipids in diatom plastid

Although the precise localization of *P. tricornutum* ACCase (PtACC2) in the plastid is not clear, the expression of the PtACC2-eGFP fusion protein showed an overlaid signal with the chlorophyll autofluorescence (Li, Xie, et al., 2018). MCAT, catalyzing the second step of FA biosynthesis, showed a classical thylakoid membrane localization according to our GFP fusion experiments. Furthermore, eight FA biosynthesis enzymes (only one was not assigned plastid protein by BLAST alignments within PPDB) and nine PDC components (four were assigned plastid protein by BLAST alignments within PPDB) were detected in our plastid samples. However, among the 17 enzymes only 3, 6, and 10 were predicted to localize to the plastid (chloroplast) by TargetP, HECTAR, and ASAFind, respectively (Data S14). Both HECTAR and ASAFind predicted that

MCAT was a plastidial protein, while for PtACC2, none of the three prediction methods suggested a plastid localization, indicating an apparent underestimation of plastid proteins by the prediction tools (Figure S5). Furthermore, ptTES1 (Hao et al., 2018) and ptACSL1 (Chen et al., 2018; Hao et al., 2022) present in our plastid samples were known to localize in the plastid stroma and cER respectively. According to GFP fusion protein signals, ptACSL2, and ptACSL4 displayed plastid localization (Hao et al., 2022), most probably at the cER. However, ptTES1, ptACSL2, and ptACSL4 were predicted not to localize in the plastid by ASAFind.

Up till now, five *P. tricornutum* FA desaturases (ERΔ6-FAD, ERΔ5FAD.1, PtFAD2, PtFAD6, and PAD) were experimentally studied and only one of them was of known localization in the plastid (Domergue et al., 2002, 2003; Smith et al., 2021). PtFAD6-eGFP fluorescence colocalized with the chlorophyll fluorescence was not evenly distributed within the plastid (Domergue et al., 2003), and our results showed that the plastidial PtFAD6 together with OEMΔ6FAD and ω3FAD were all at the OEM, forming a desaturation machinery capable of generating 16:4 from 16:1. Microsomal desaturase PtFAD2 was localized in the cytosol (Domergue et al., 2003), whereas Dolch and Maréchal (2015) considered the desaturase together with two additional functionally characterized microsomal desaturases PtD6 (ERΔ6FAD) and PtD5 (ERΔ5FAD.1) to be ER-localized. The present study showed that PtFAD2 was localized in the cERM, similar to FAD2 desaturases in the ER from plants (Shanklin & Cahoon, 1998). In fact, except ω3 desaturation, all desaturation and elongation for generating EPA and DHA from the stearic acid (18:0) take place in the cERM according to the localization of corresponding enzymes (Figures 5 and 8). Since ω3FAD localizes at the OEM, the introduction of an ω-3 double bond into C18:2n6 or C18:3n6 possibly depends on a PC or PE transfer between the cER and OEM, based on the detection of molecular species in PC and PE enriched in C18:2, C18:3, and C18:4. Two soluble enzymes, PAD (Smith et al., 2021) and FAD4, act on 16:0-ACP and 16:0-PG (Dolch et al., 2017), respectively, to produce 16:1n7 and 16:1n13, and were observed in the chloroplast stroma. Trace 24:0 FA was detected in *P. tricornutum*, and the elongation from 16:0-CoA to 18:0-CoA was presumably catalyzed by the saturated ER FA elongases (J16376 and J49867) (Smith et al., 2021), also known as Δ0ELOs (Dolch et al., 2017), which would possibly function for the further elongation into 24:0.

P. tricornutum MGDGS1 (previously described as MGD1) is an intermembrane space localized enzyme between the innermost and second innermost plastid membranes (Bullmann et al., 2010), and was only detected in our putative plastid proteomes. Another MGDGS (J9619), one DGDGS (J43116), SQD2, and BTA are all
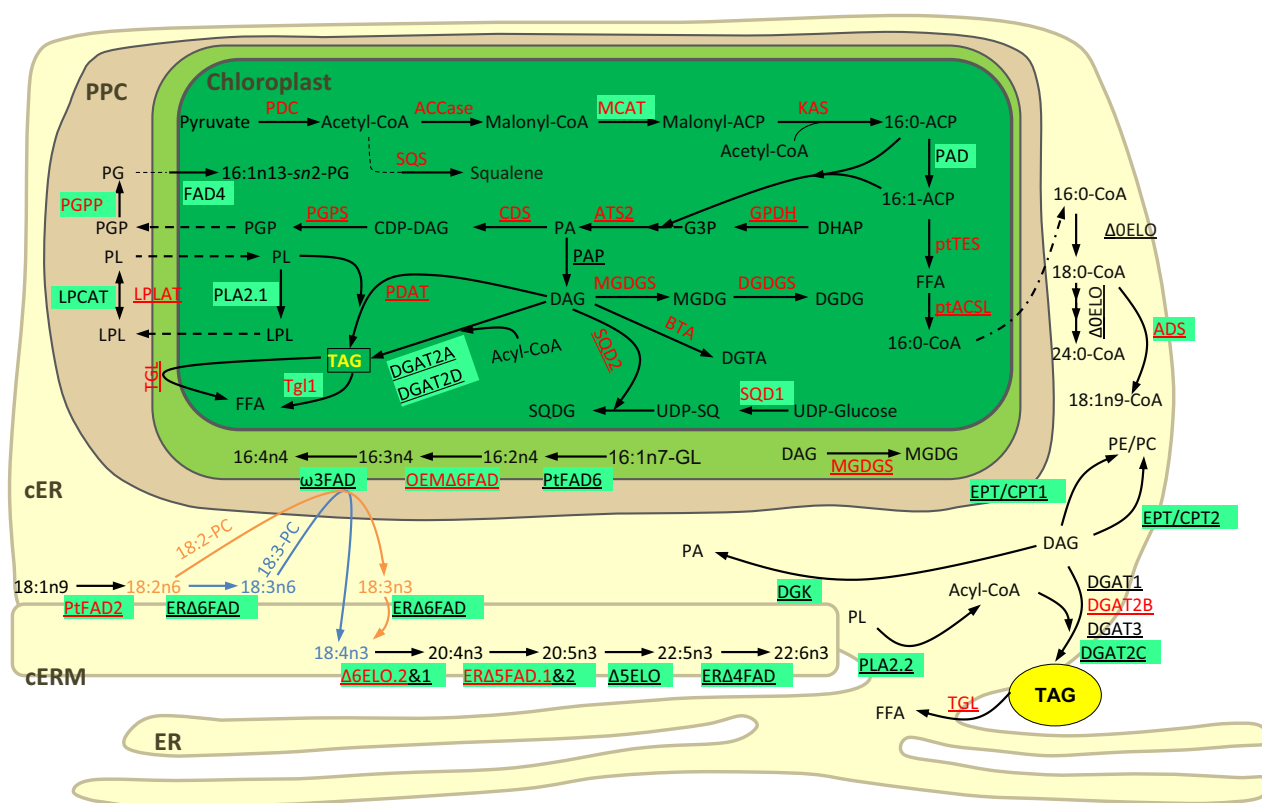
Figure 8. A model illustrating the fatty acid biosynthesis and lipid metabolism in *Phaeodactylum tricornutum*. Proteins present in the putative plastid proteomes were in red, protein localizations determined by GFP were shown in green box, and proteins with transmembrane domains were underlined. ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; ADS, $\Delta 9$ acyl-CoA desaturase; ATS2, plastidic lysophosphatidic acid transferase; BTA, betaine lipid synthase; CDS, cytidine diphosphate (CDP)-diacylglycerol synthase; CDP-DAG, cytidine diphosphate-diacylglycerol; cER, chloroplast endoplasmic reticulum; cERM, chloroplast endoplasmic reticulum membrane; CoA, coenzyme A; DAG, diacylglycerol; DGAT, acyl-CoA:diacylglycerol acyltransferase; DGDG, digalactosyldiacylglycerol; DGDGS, digalactosyldiacylglycerol synthase; DGK, diacylglycerol kinase; DGTA, diacylglycerylhydroxymethyltrimethyl-β-alanine; DHAP, dihydroxyacetone phosphate; ELO, fatty acid elongase; EPT/CPT, ethanolamine–/choline-phosphotransferase; ER, endoplasmic reticulum; FAD, fatty acid desaturase; FFA, free fatty acid; G3P, glycerol-3-phosphate; GL, glycolipid; GPDH, glycerol-3-phosphate-dehydrogenase; KAS, 3-ketoacyl-ACP synthase; LPCAT, lysophosphatidylcholine acyltransferase; LPL, lysophospholipid; LPLAT, lysophospholipid acyltransferase; MCAT, malonyl-CoA:ACP transacylase; MGDG, monogalactosyldiacylglycerol; MGDGS, monogalactosyldiacylglycerol synthase; PA, phosphatidic acid; PAD, palmitoyl-ACP desaturase; PAP, phosphatidic acid phosphatase; PC, phosphatidylcholine; PDAT, phospholipid:diacylglycerol acyltransferase; PDC, pyruvate dehydrogenase complex; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PGP, phosphatidylglycerol phosphate; PGPP, phosphatidylglycerophosphate phosphatase; PGPS, phosphatidylglycerol phosphate synthase; PL, phospholipid; PLA2, phospholipase A2; PPC, periplastidial compartment; ptACSL, long chain acyl-CoA synthetase; ptTES, fatty acyl-ACP thioesterase; SQD1, UDP-sulfoquinovose synthase; SQD2, sulfoquinovosyltransferase; SQDG, sulfoquinovosyldiacylglycerol; SQS, squalene synthase; TAG, triacylglycerol; TGL, triacylglycerol lipase; Tgl1, SDP1-patatin like lipase; UDP-SQ, UDP-sulfoquinovose.

enriched in the plastid samples (only BTA was not assigned plastid protein by BLAST alignments within PPDB), indicating that glycolipids (galactolipids and the sulfolipid) as well as the betaine lipid DGTA are synthesized in the plastid. According to the localization of PGPP and EPT/CPT (Figure 6), syntheses of 3 phospholipids PG, PE, and PC could occur in the PPC, and the last two might also take place in the cER.

An LPAAT (EG02461, previously described as PtGPAT) (Niu et al., 2016) and a plastidic LPAAT (ATS2b, previously described as AGPAT1) (Balamurugan et al., 2017) were reported to localize to the plastid and plastid membranes respectively, and the former was enriched in our plastid samples. Another 2 enzymes of the Kennedy pathway, GPAT, and ATS2a (an assigned plastid protein by BLAST

alignments within PPDB), were also enriched in the putative plastid proteomes. It is indicated that the Kennedy pathway takes place in the plastid.

According to the localization experiments, six DGATs were all plastid-localized (cERM and intraplastid). In addition, PDAT was only detected in the plastid samples. These data support a production of TAG both at the surface of the cERM and inside the complex plastid, confirmed by the lipidomic analyses of the isolated organelle. In addition to the fact that many plastid proteins have been identified in the proteomes of lipid droplets (Lupette et al., 2019; Yoneda et al., 2016), TAG biosynthesis observed in our study and components of the lipid droplet membrane (SQDG, DGTA and minor component PC) (Lupette et al., 2019) all originated from the plastid. In addition, TAG hydrolysis could

also take place in the plastid of *P. tricornutum* based on the detected TAG lipases in putative plastid proteomes and GFP-fusion localization experiments in our present (Figure 6) and previous (Li, Pan, et al., 2018) studies.

At present, the protein composition and biological function of secondary plastids are still poorly understood, and here we present the novelty of the secondary plastid. Diatom plastid harbors unique pathways, like the glutamine-ornithine cycle, which are absent in other photosynthetic eukaryotes. Most importantly, the diatom plastid centralizes pathways that are split in red algae, making them much more efficient, especially for acyl-glycerolipid metabolism. In the future, it will be interesting to verify more PPC-localized proteins (Data S17; Appendix S2) to reveal the specific role of the PPC in the lipid metabolism.

EXPERIMENTAL PROCEDURES

Algal culture and plastid isolation

Axenic cultures of *Phaeodactylum tricornutum* Bohlin (CCMP2561) (Bigelow Laboratory for Ocean Sciences, USA) and transformants expressing plastid marker proteins were grown in artificial seawater enriched with f/2 (nitrate concentration was increased to 3.53 mM) (Guillard, 1975) at 22°C under continuous illumination of 70 μmol photons m^{-2} s^{-1} in 5-cm diameter glass culture tubes (60-cm long) bubbled with filtered air. Cultivation was performed for 5 days to a concentration of 2–3 $\times$ 10^7 cells ml^{-1} with the initial cell density of 2.5 $\times$ 10^5 ml^{-1}, then about 4–6 $\times$ 10^9 cells were collected by centrifugation at 2000 g for 10 min at 4°C. Plastids from the harvested algal cells were isolated by Percoll-density centrifugation as described by Schober et al. (2018) with minor modifications, and all the solutions for the isolation were prepared according to Schober et al. (2018). Algal pellets were washed twice with 20 mM HEPES-KOH (pH 7.2), and then thoroughly suspended in ice-cold isolation buffer. Algal cells were subsequently broken up using a French Press at 80 MPa. The suspension of broken cells was centrifuged at 300 g for 8 min at 4°C, and the supernatant was collected and centrifuged at 2000 g for 5 min at 4°C. Subsequently, the supernatant was discarded and the pellet was cautiously resuspended with washing buffer (with BSA). The resuspended solution was loaded on a three-layered Percoll gradient (10%, 20%, and 30%) and then centrifuged at 14 400 g for 1 h at 4°C. The plastid-enriched fraction accumulated at the interface between the 10 and 20% Percoll layers. The plastid-enriched fraction was transferred into a new tube and washed with washing buffer (without BSA) by centrifugation at 4000 $\times$ g for 10 min at 4 °C. The pellet was collected for protein extraction or observation by microscope.

Mitochondrion isolation

Cell pellets of *P. tricornutum* transformants expressing ME1-eGFP were collected and washed as mentioned above. The pellets were resuspended in breaking buffer (50 mM HEPES-KOH, pH 7.2, 5 mM EDTA, 0.25 M sorbitol, 4 mM Cys, 0.5% PVP 40, 0.1% BSA) to a concentration of 0.5 g ml^{-1}. The suspension of 5 ml was mixed with 10 ml of glass beads in a 50 ml tube and vortexed for 1 min. The cell homogenate was rinsed with breaking buffer and then centrifuged for 5 min at 2000 g. The pellet was discarded and the supernatant was further centrifuged for 10 min at 5000 g. The resulting pellets (Fraction 1) were mixed with 30 ml of 20% Percoll (20% Percoll, 0.25 M sorbitol, 10 mM MOPS-KOH, pH 7.2,

1 mM EDTA, 0.1% BSA, 0.5% PVP 40) and centrifuged for 40 min at 20 000 g. In the tube, four different colored fractions can be seen. Fractions 2–4 were collected and diluted with 40 ml washing buffer (10 mM potassium phosphate buffer, pH 7.2, 0.1% BSA, 0.25 M sorbitol, 1 mM EDTA), and then centrifuged at 10 000 g for 10 min (Eriksson et al., 1995). The resulting pellets of each fraction were used for western blot analysis after protein extraction.

Vector construction and transformation

P. tricornutum total RNA was isolated and cDNA was generated following previously described methods (Ge et al., 2014). Full-length coding sequences were amplified from the *P. tricornutum* cDNA and inserted into the multiple cloning sites of pPhaT1-eGFP or pPhaT1-mRFP according to Li, Pan, et al. (2018). All the full-length coding sequences and primers used in the study are listed in Data S18. The expression vectors were introduced into *P. tricornutum* by electroporation and transformants were subsequently screened on a solid medium containing zeocin (Zhang & Hu, 2014). Positive transformants and isolated plastids were observed using a Leica TCS SP8 laser scanning confocal microscope (Li, Pan, et al. 2018). For the observation of lipid droplets, cells, and isolated plastids were stained with BODIPY 505/515 and detected based on fluorescence (excitation wavelength, 488 nm; detection at 490–535 nm).

Protein preparation and western blot analysis

For protein extraction, whole cells, and isolated plastids were resuspended in cell lysis buffer (Beyotime Biotechnology, China) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and sonicated for 10 min on ice (Ge et al., 2014). After centrifuge (12 000 g for 10 min at 4°C), protein concentration in the supernatant was quantified using the Beyotime BCA Protein Assay Kit (China). Denatured protein samples were separated on 10% SDS-PAGE gels and then either stained with Coomassie brilliant blue (R250, Bio-Rad) or transferred to a polyvinylidene fluoride membrane. The membrane was blocked for 1 h in 5% bovine serum albumin and then incubated with a 1:500 dilution of primary antibody (information of the antibodies is shown in Data S18) for 1.5 h, followed by 2 h incubation with a horseradish peroxidase-labeled secondary antibody (1:10 000 dilution) (Kehbio, China). Immune-labeled membranes were visualized using the DAB Horseradish Peroxidase Color Development Kit and scanned by an ImageQuant LAS 4000 Luminescent Transmitted UV imaging Image Analyzer (GE, USA).

Label-free proteomic analysis

Protein extraction from whole cells and isolated plastids was performed by supplementing samples with four volumes of cold acetone. After cell lysis, protein extraction, and digestion, resulted peptides were analyzed using a Q Exactive HF system (Thermo Scientific) coupled to a Nano Acuity HPLC (Waters). LC–MS/MS raw data were processed using MaxQuant (version 1.6.5.0) (Cox et al., 2014; Cox & Mann, 2008). Mass spectrometry spectra were searched against the *P. tricornutum* protein database from the Ensembl Protists, the organelle-encoded protein database of NCBI and Uniprot (166 entries), and the "novel protein" database identified by Yang et al. (2018). All *P. tricornutum* organelle-encoded proteins and "novel protein" are listed in Data S19. The relative abundance of proteins was evaluated by intensity after normalization by total intensity. See Methods S1 for additional information.

Total lipid and fatty acid profile analysis

Total lipids were extracted from 150 mg of freeze-dried whole cells and isolated plastids (Lupette et al., 2019), and lipid extracts

were analyzed using a Triple TOF 6600 mass spectrometer (Applied Biosystems/MDS Sciex, Canada) equipped with an electrospray ionization source to quantify the major glycerolipids (Xie et al., 2020). The total lipid extracts were methylated by 1 mol L^{-1} sulfuric acid–methanol for 1 h at 100°C and then methylated FAs were analyzed by gas chromatography (Agilent 7890A, USA) (Ge et al., 2014). For the fatty acid compositions of TAG, the content of a specific fatty acid was the sum of the number of moles in TAG species determined by mass spectrometer.

Bioinformatics analysis

For the KEGG annotations, all identified protein sequences were aligned using blast method (BLASTP) with *P. tricornutum* protein sequences in the KEGG database. For the unannotated proteins in the KEGG database, BLASTP (threshold, e-value <10^{-5}) was performed with all the proteins from five closely related species and a mixed library of 32 species, besides, the HMM profile-based KO annotation and KEGG mapping method were used (Aramaki et al., 2020). All the species used in KEGG annotation were listed in Data S13. Subcellular targeting predictions of all proteins were obtained using TargetP (http://www.cbs.dtu.dk/services/TargetP/), HECTAR (https://webtools.sb-roscoff.fr/) and ASAFind (https://rocaplab.ocean.washington.edu/tools/asafind/) (signal peptides prediction by SignalP-4.1). Prediction of transmembrane helices was performed using TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0).

ACCESSION NUMBERS

MS-based identification information of all proteins identified in this study was listed in Data S1. Mass spectrometric proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD038187.

AUTHOR CONTRIBUTIONS

HH conceived and designed the research. TH and YP performed the experiments. TH, YP, EM and HH analyzed the data and drafted the manuscript. HH made critical revision. All authors read and approved the final article. HH agrees to serve as the author responsible for contact and ensures communication.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Fluorescence micrographs of whole cells from transformants expressing plastid marker proteins fused with mRFP or eGFP.

Figure S2. Overall view of isolated transformants' plastids.

Figure S3. Microscopic images of isolated plastids at 4°C for 5 h.

Figure S4. Prediction of protein subcellular localization by TargetP, HECTAR, and ASAFind.

Figure S5. Prediction of plastid proteins via different bioinformatic tools on all 12 382 nuclear-encoded proteins in *Phaeodactylum tricornutum*.

Figure S6. Volcano plots showing the statistically significant (*P* < 0.05) enriched proteins (red) in putative plastid proteomes.

Figure S7. Metabolic pathway diagram based on the putative plastid proteomic data.

Figure S8. Fluorescent microscopy of *Phaeodactylum tricornutum* cells harboring ADS-eGFP after stained with Hoechst 33342.

Figure S9. Compared analysis of lipid molecular species between whole cells and isolated plastids.

Figure S10. *Phaeodactylum tricornutum* cells and isolated plastids stained with Bodipy.

Figure S11. Transmembrane domain prediction of all detected nuclear-encoded proteins (3701) in our proteome.

Figure S12. *Phaeodactylum tricornutum* mitochondrion isolation and western blot analysis using two antibodies against mitochondrial proteins (ME1 and SHMT2), an antibody against GFP, and an antibody against a pyrenoid-localized protein PtCA1.

Methods S1. Label-free proteomic analysis.

Appendix S1. Details on the identified proteins involved in metabolic pathways based on our proteomes.

Appendix S2. Details on chloroplast-endomembrane "hybrid" metabolism in diatom complex plastid.

Appendix S3. Association between mitochondria and plastids.

Data S1. List of all identified proteins.

Data S2. Proteins only detected in plastid samples.

Data S3. Proteins solely detected in this study.

Data S4. Comparison of proteins detected in our putative plastid proteomes with those in subcellular components reported before.

Data S5. Classification of proteins detected in plastid samples.

Data S6. Estimation of organelle enrichment factors based on the assigned localization blasted on PPDB and localization experiments using the relative protein intensity.

Data S7. List of proteins used for estimation of organelle enrichment factors in Data S6.

Data S8. Comparison of the assigned localization by blast on PPDB with the experimentally determined intracellular localization in *P. tricornutum*.

Data S9. Comparison of proteins detected in our proteomes with those with experimentally determined intracellular localization.

Data S10. Potential plastidial proteins detected in our proteomes were determined by ASAFind prediction or BLAST alignments within PPDB.

Data S11. Proteins with fold change values >1.5 detected in all three plastid samples and identified by at least two peptides.

Data S12. List of statistically significant enriched [>Log$_2$(1.5)] and depleted [<−Log$_2$(1.5)] proteins in plastid samples.

Data S13. KEGG Orthology (KO) numbers of all identified proteins.

Data S14. All proteins involved in metabolic pathways identified by our putative plastid proteomic data.

Data S15. Lipidomic data of whole cells (cell-1 ∼ 3) and plastid-enriched fractions (plastid-1 ∼ 3).

Data S16. Mitochondrial proteins of electron transport and ATP synthesis present in the plastid samples.

Data S17. Bioinformatical identification (Moog et al., 2011) of *Phaeodactylum tricornutum* PPC-localized proteins present in our plastid samples.

Data S18. Primary antibodies, primers, and gene sequences used in this study.

Data S19. All organelle-encoded proteins in NCBI and Uniprot, and the "novel proteins" identified by Yang et al. (2018).

REFERENCES

Abida, H., **Dolch, L.J.**, **Meï, C.**, **Villanova, V.**, **Conte, M.**, **Block, M.A.** *et al.* (2015) Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricornutum. Plant Physiology*, **167**, 118–136.

Allen, A.E., **Moustafa, A.**, **Montsant, A.**, **Eckert, A.**, **Kroth, P.G.** & **Bowler, C.** (2012) Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. *Molecular Biology and Evolution*, **29**, 367–379.

Aramaki, T., **Blanc-Mathieu, R.**, **Endo, H.**, **Ohkubo, K.**, **Kanehisa, M.**, **Goto, S.** *et al.* (2020) KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, **36**, 2251–2252.

Archibald, J.M. & **Keeling, P.** (2002) Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends in Genetics*, **18**, 577–584.

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

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**, 366–369.

Balamurugan, S., **Wang, X.**, **Wang, H.-L.**, **An, C.-J.**, **Li, H.**, **Li, D.-W.** *et al.* (2017) Occurrence of plastidial triacylglycerol synthesis and the potential regulatory role of AGPAT in the model diatom *Phaeodactylum tricornutum. Biotechnology for Biofuels*, **10**, 97.

Barka, F., **Angstenberger, M.**, **Ahrendt, T.**, **Lorenzen, W.**, **Bode, H.B.** & **Büchel, C.** (2016) Identification of a triacylglycerol lipase in the diatom *Phaeodactylum tricornutum. Biochimica et Biophysica Acta—Molecular and Cell Biology of Lipids*, **1861**, 239–248.

Boudière, L., **Michaud, M.**, **Petroutsos, D.**, **Rébeillé, F.**, **Falconet, D.**, **Bastien, O.** *et al.* (2014) Glycerolipids in photosynthesis: composition, synthesis and trafficking. *Biochimica et Biophysica Acta – Bioenergetics*, **1837**, 470–480.

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

Bruckner, C.G., **Rehm, C.**, **Grossart, H.P.** & **Kroth, P.G.** (2011) Growth and release of extracellular organic compounds by benthic diatoms depend on interactions with bacteria. *Environmental Microbiology*, **13**, 1052–1063.

Buhmann, M.T., **Schulze, B.**, **Förderer, A.**, **Schleheck, D.** & **Kroth, P.G.** (2016) Bacteria may induce the secretion of mucin-like proteins by the diatom *Phaeodactylum tricornutum. Journal of Phycology*, **52**, 463–474.

Bullmann, L., **Haarmann, R.**, **Mirus, O.**, **Bredemeier, R.**, **Hempel, F.**, **Maier, U.G.** *et al.* (2010) Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. *Journal of Biological Chemistry*, **285**, 6848–6856.

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

Chen, Z., **Luo, L.**, **Chen, R.**, **Hu, H.**, **Pan, Y.**, **Jiang, H.** *et al.* (2018) Acetylome profiling reveals extensive lysine acetylation of the fatty acid metabolism pathway in the diatom *Phaeodactylum tricornutum. Molecular & Cellular Proteomics*, **17**, 399–412.

Chu, L., **Gruber, A.**, **Ast, M.**, **Schmitz-Esser, S.**, **Altensell, J.**, **Neuhaus, H.E.** *et al.* (2017) Shuttling of (deoxy-) purine nucleotides between compartments of the diatom *Phaeodactylum tricornutum. New Phytologist*, **213**, 193–205.

Cox, J., **Hein, M.Y.**, **Luber, C.A.**, **Paron, I.**, **Nagaraj, N.** & **Mann, M.** (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics*, **13**, 2513–2526.

Cox, J. & **Mann, M.** (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, **26**, 1367–1372.

Dolch, L.J. & **Maréchal, E.** (2015) Inventory of fatty acid desaturases in the pennate diatom *Phaeodactylum tricornutum. Marine Drugs*, **13**, 1317–1339.

Dolch, L.J., **Rak, C.**, **Perin, G.**, **Tourcier, G.**, **Broughton, R.**, **Leterrier, M.** *et al.* (2017) A palmitic acid elongase affects eicosapentaenoic acid and plastidial monogalactosyldiacylglycerol levels in *Nannochloropsis. Plant Physiology*, **173**, 742–759.

Domergue, F., **Lerchl, J.**, **Zahringer, U.** & **Heinz, E.** (2002) Cloning and functional characterization of *Phaeodactylum tricornutum* front-end desaturases involved in eicosapentaenoic acid biosynthesis. *European Journal of Biochemistry*, **269**, 4105–4113.

Domergue, F., **Spiekermann, P.**, **Lerchl, J.**, **Beckmann, C.**, **Kilian, O.**, **Kroth, P.G.** *et al.* (2003) New insight into *Phaeodactylum tricornutum* fatty acid metabolism. Cloning and functional characterization of plastidial and microsomal Δ12-fatty acid desaturases. *Plant Physiology*, **131**, 1648–1660.

Elias, J.E., **Haas, W.**, **Faherty, B.K.** & **Gygi, S.P.** (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods*, **2**, 667–675.

Erdene-Ochir, E., **Shin, B.K.**, **Kwon, B.**, **Jung, C.** & **Pan, C.H.** (2019) Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum. Scientific Reports*, **9**, 1–10.

Eriksson, M., **Gardestrom, P.** & **Samuelsson, G.** (1995) Isolation, purification, and characterization of mitochondria from *Chlamydomonas reinhardtii. Plant Physiology*, **107**(2), 479–483.

Ewe, D., **Tachibana, M.**, **Kikutani, S.**, **Gruber, A.**, **Río Bártulos, C.**, **Konert, G.** *et al.* (2018) The intracellular distribution of inorganic carbon fixing enzymes does not support the presence of a C4 pathway in the diatom *Phaeodactylum tricornutum. Photosynthesis Research*, **137**, 263–280.

Falkowski, P.G., **Katz, M.E.**, **Knoll, A.H.**, **Quigg, A.**, **Raven, J.A.**, **Schofield, O.** *et al.* (2004) The evolution of modern eukaryotic phytoplankton. *Science*, **305**, 354–360.

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

Flori, S., **Jouneau, P.H.**, **Bailleul, B.**, **Gallet, B.**, **Estrozi, L.F.**, **Moriscot, C.** *et al.* (2017) Plastid thylakoid architecture optimizes photosynthesis in diatoms. *Nature Communications*, **8**, 15885.

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.

Ge, F., **Huang, W.**, **Chen, Z.**, **Zhang, C.**, **Xiong, Q.**, **Bowler, C.** *et al.* (2014) Methylcrotonyl-CoA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum. Plant Cell*, **26**, 1681–1697.

Gibbs, S.P. (1979) The route of entry of cytoplasmically synthesized proteins into chloroplasts of algae possessing chloroplast ER. *Journal of Cell Science*, **35**, 253–266.

Gloaguen, P., **Bournais, S.**, **Alban, C.**, **Ravanel, S.**, **Seigneurin-Berny, D.**, **Matringe, M.** *et al.* (2017) ChloroKB: a web application for the integration of knowledge related to chloroplast metabolic network. *Plant Physiology*, **174**, 922–934.

Gonzalez, N.H., **Felsner, G.**, **Schramm, F.D.**, **Klingl, A.**, **Maier, U.G.** & **Bolte, K.** (2011) A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS One*, **6**, e25316.

Gould, S.B., **Maier, U.G.** & **Martin, W.F.** (2015) Protein import and the origin of red complex plastids. *Current Biology*, **25**, R515–R521.

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 Evolution*, **23**, 2413–2422.

Grosche, C., **Hempel, F.**, **Bolte, K.**, **Zauner, S.** & **Maier, U.G.** (2014) The periplastidial compartment: a naturally minimized eukaryotic cytoplasm. *Current Opinion in Microbiology*, **22**, 88–93.

Grouneva, I., **Rokka, A.** & **Aro, E.M.** (2011) The thylakoid membrane proteome of two marine diatoms outlines both diatom-specific and species-specific features of the photosynthetic machinery. *Journal of Proteome Research*, **10**, 5338–5353.

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

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

Guillard, R.R.L. (1975) Culture of phytoplankton for feeding marine invertebrates. In: Smith, W.L. & Canley, M.H. (Eds.) *Culture of Marine Invertebrate Animals*. Boston, MA, USA: Springer, pp. 29–60.

Gundermann, K., Schmidt, M., Weisheit, W., Mittag, M. & Büchel, C. (2013) Identification of several sub-populations in the pool of light harvesting proteins in the pennate diatom *Phaeodactylum tricornutum*. *Biochimica et Biophysica Acta – Bioenergetics*, **1827**, 303–310.

Hao, X., Chen, W., Amato, A., Jouhet, J., Maréchal, E., Moog, D. et al. (2022) Multiplex CRISPR/Cas9 editing of the long-chain acyl-CoA synthetase family in the diatom *Phaeodactylum tricornutum* reveals that mitochondrial ptACSL3 is involved in the synthesis of storage lipids. *New Phytologist*, **233**, 1797–1812.

Hao, X., Luo, L., Jouhet, J., Rébeillé, F., Maréchal, E., Hu, H. et al. (2018) Enhanced triacylglycerol production in the diatom *Phaeodactylum tricornutum* by inactivation of a hotdog-fold thioesterase gene using TALEN-based targeted mutagenesis. *Biotechnology for Biofuels*, **11**, 312.

Hempel, F., Bullmann, L., Lau, J., Zauner, S. & Maier, U.G. (2009) ERAD-derived preprotein transport across the second outermost plastid membrane of diatoms. *Molecular Biology and Evolution*, **26**, 1781–1790.

Herbstová, M., Bína, D., Koník, P., Gardian, Z., Vácha, F. & Litvín, R. (2015) Molecular basis of chromatic adaptation in pennate diatom *Phaeodactylum tricornutum*. *Biochimica et Biophysica Acta – Bioenergetics*, **1847**, 534–543.

Ischebeck, T., Krawczyk, H.E., Mullen, R.T., Dyer, J.M. & Chapman, K.D. (2020) Lipid droplets in plants and algae: distribution, formation, turnover and function. *Seminars in Cell and Developmental Biology*, **l108**, 82–93.

Jallet, D., Xing, D., Hughes, A., Moosburner, M., Simmons, M.P., Allen, A.E. et al. (2020) Mitochondrial fatty acid β-oxidation is required for storage-lipid catabolism in a marine diatom. *New Phytologist*, **228**, 946–958.

Jaussaud, A., Lupette, J., Salvaing, J., Jouhet, J., Bastien, O., Gromova, M. et al. (2020) Stepwise biogenesis of subpopulations of lipid droplets in nitrogen starved *Phaeodactylum tricornutum* cells. *Frontiers in Plant Science*, **11**, 48.

Joshi-Deo, J., Schmidt, M., Gruber, A., Weisheit, W., Mittag, M., Kroth, P.G. et al. (2010) Characterization of a trimeric light-harvesting complex in the diatom *Phaeodactylum tricornutum* built of FcpA and FcpE proteins. *Journal of Experimental Botany*, **61**, 3079–3087.

Kikutani, S., Nakajima, K., Nagasato, C., Tsuji, Y., Miyatake, A. & Matsuda, Y. (2016) Thylakoid luminal θ-carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. *Proceedings of the National Academy of Sciences, USA*, **113**, 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. *Plant Journal*, **41**, 175–183.

Kitao, Y., Harada, H. & Matsuda, Y. (2008) Localization and targeting mechanisms of two chloroplastic β-carbonic anhydrases in the marine diatom *Phaeodactylum tricornutum*. *Physiologia Plantarum*, **133**, 68–77.

Lepetit, B., Volke, D., Szabó, M., Hoffmann, R., Garab, G., Wilhelm, C. et al. (2007) Spectroscopic and molecular characterization of the oligomeric antenna of the diatom *Phaeodactylum tricornutum*. *Biochemistry*, **46**, 9813–9822.

Levitan, O., Chen, M., Kuang, X., Cheong, K.Y., Jiang, J., Banal, M. et al. (2019) Structural and functional analyses of photosystem II in the marine diatom *Phaeodactylum tricornutum*. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 17316–17322.

Leyland, B., Zarka, A., Didi-Cohen, S., Boussiba, S. & Khozin-Goldberg, I. (2020) High resolution proteome of lipid droplets isolated from the pennate diatom *Phaeodactylum tricornutum* (Bacillariophyceae) strain pt4 provides mechanistic insights into complex intracellular coordination during nitrogen deprivation. *Journal of Phycology*, **56**, 1642–1663.

Li, D.W., Xie, W.H., Hao, T.B., Cai, J.X., Zhou, T.B., Balamurugan, S. et al. (2018) Constitutive and chloroplast targeted expression of acetyl-CoA carboxylase in oleaginous microalgae elevates fatty acid biosynthesis. *Marine Biotechnology*, **20**, 566–572.

Li, X., Pan, Y. & Hu, H. (2018) Identification of the triacylglycerol lipase in the chloroplast envelope of the diatom *Phaeodactylum tricornutum*. *Algal Research*, **33**, 440–447.

Liu, X., Hempel, F., Stork, S., Bolte, K., Moog, D., Heimerl, T. et al. (2016) Addressing various compartments of the diatom model organism *Phaeodactylum tricornutum* via sub-cellular marker proteins. *Algal Research*, **20**, 249–257.

Lupette, J., Jaussaud, A., Seddiki, K., Morabito, C., Brugière, S., Schaller, H. et al. (2019) The architecture of lipid droplets in the diatom *Phaeodactylum tricornutum*. *Algal Research*, **38**, 101415.

Matsui, H., Hopkinson, B.M., Nakajima, K. & Matsuda, Y. (2018) Plasma membrane-type aquaporins from marine diatoms function as CO_2/NH_3 channels and provide photoprotection. *Plant Physiology*, **178**, 345–357.

Michaud, M., Gros, V., Tardif, M., Brugière, S., Ferro, M., Prinz, W.A. et al. (2016) AtMic60 is involved in plant mitochondria lipid trafficking and is part of a large complex. *Current Biology*, **26**, 627–639.

Millar, A.H., Sweetlove, L.J., Giegé, P. & Leaver, C.J. (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiology*, **127**(4), 1711–1727.

Moog, D., Rensing, S.A., Archibald, J.M., Maier, U.G. & Ullrich, K.K. (2015) Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. *Genome Biology and Evolution*, **7**, 2955–2969.

Moog, D., Stork, S., Zauner, S. & Maier, U.G. (2011) In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. *Genome Biology and Evolution*, **3**, 375–382.

Naumann, B., Busch, A., Allmer, J., Ostendorf, E., Zeller, M., Kirchhoff, H. et al. (2007) Comparative quantitative proteomics to investigate the remodeling of bioenergetic pathways under iron deficiency in *Chlamydomonas reinhardtii*. *Proteomics*, **7**, 3964–3979.

Niu, Y.-F., Wang, X., Hu, D.-X., Balamurugan, S., Li, D.-W., Yang, W.-D. et al. (2016) Molecular characterization of a glycerol-3-phosphate acyltransferase reveals key features essential for triacylglycerol production in *Phaeodactylum tricornutum*. *Biotechnology for Biofuels*, **9**, 60.

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**, 322.

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

Remmers, I.M., D'Adamo, S., Martens, D.E., de Vos, R.C.H., Mumm, R., America, A.H.P. et al. (2018) Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal Research*, **35**, 33–49.

Roessler, P.G. & Ohlrogge, J.B. (1993) Cloning and characterization of the gene that encodes acetyl-coenzyme a carboxylase in the alga *Cyclotella cryptic*. *Journal of Biological Chemistry*, **268**, 19254–19259.

Rolland, N., Bouchnak, I., Moyet, L., Salvi, D. & Kuntz, M. (2018) The main functions of plastids. In: Maréchal, E. (Ed.) *Plastids: Methods and Protocols*. New York, USA: Springer, pp. 73–85.

Salvato, F., Havelund, J.F., Chen, M., Rao, R.S.P., Rogowska-Wrzesinska, A., Jensen, O.N. et al. (2014) The potato tuber mitochondrial proteome. *Plant Physiology*, **164**(2), 637–653.

Schober, A.F., Flori, S., Finazzi, G., Kroth, P.G. & Río Bártulos, C. (2018) Isolation of plastid fractions from the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. In: Maréchal, E. (Ed.) *Plastids: Methods and Protocols*. New York, USA: Springer, pp. 189–203.

Schober, A.F., Río Bártulos, C., Bischoff, A., Lepetit, B., Gruber, A. & Kroth, P.G. (2019) Organelle studies and proteome analyses of mitochondria and plastids fractions from the diatom *Thalassiosira pseudonana*. *Plant and Cell Physiology*, **60**, 1811–1828.

Shanklin, J. & Cahoon, E.B. (1998) Desaturation and related modifications of fatty acids. *Annual Review of Plant Physiology and Plant Molecular Biology*, **49**, 611–641.

Shao, Z., Thomas, Y., Hembach, L., Xing, X., Duan, D., Moerschbacher, B.M. et al. (2019) Comparative characterization of putative chitin deacetylases from *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* highlights the potential for distinct chitin-based metabolic processes in diatoms. *New Phytologist*, **221**, 1890–1905.

Smith, R., **Jouhet, J.**, **Gandini, C.**, **Nekrasov, V.**, **Maréchal, E.**, **Napier, J.A.** **et al.** (2021) Plastidial acyl carrier protein Δ9-desaturase modulates eicosapentaenoic acid biosynthesis and triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Plant Journal*, **106**, 1247–1259.

Smith, S.R., **Dupont, C.L.**, **McCarthy, J.K.**, **Broddrick, J.T.**, **Oborník, M.**, **Horák, A.** **et al.** (2019) Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nature Communications*, **10**, 1–14.

Staehelin, L.A. (1997) The plant ER: a dynamic organelle composed of a large number of discrete functional domains. *The Plant Journal*, **11**, 1151–1165.

Stork, S., **Moog, D.**, **Przyborski, J.M.**, **Wilhelmi, I.**, **Zauner, S.** & **Maier, U.G.** (2012) Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. *Eukaryotic Cell*, **11**, 1472–1481.

Sun, Q., **Zybailov, B.**, **Majeran, W.**, **Friso, G.**, **Olinares, P.D.B.** & **van Wijk, K.J.** (2009) PPDB, the plant proteomics database at Cornell. *Nucleic Acids Research*, **37**(suppl_1), D969–D974.

Tanaka, Y., **Nakatsuma, D.**, **Harada, H.**, **Ishida, M.** & **Matsuda, Y.** (2005) Localization of soluble β-carbonic anhydrase in the marine diatom *Phaeodactylum tricornutum*. Sorting to the chloroplast and cluster formation on the girdle lamellae. *Plant Physiology*, **138**, 207–217.

Tardif, M., **Atteia, A.**, **Specht, M.**, **Cogne, G.**, **Rolland, N.**, **Brugière, S.** **et al.** (2012) PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Molecular Biology and Evolution*, **29**, 3625–3639.

Uwizeye, C., **Decelle Johan Jouneau, P.-H.**, **Flori, S.**, **Gallet, B.**, **Keck, J.-B.**, **Bo, D.D.** **et al.** (2021) Morphological bases of phytoplankton energy management and physiological responses unveiled by 3D subcellular imaging. *Nature Communications*, **12**, 1049.

Wang, L., **Patena, W.**, **Van Baalen, K.A.**, **Xie, Y.**, **Singer, E.R.**, **Gavrilenko, S.** **et al.** (2023) A chloroplast protein atlas reveals punctate structures and spatial organization of biosynthetic pathways. *Cell*, **186**, 3499–3518.

Wang, S., **Zhang, G.**, **Zhang, Y.**, **Song, Q.**, **Chen, Z.**, **Wang, J.** **et al.** (2015) Comparative studies of mitochondrial proteomics reveal an intimate protein network of male sterility in wheat (*Triticum aestivum* L.). *Journal of Experimental Botany*, **66**(20), 6191–6203.

Wang, W., **Yu, L.J.**, **Xu, C.**, **Tomizaki, T.**, **Zhao, S.**, **Umena, Y.** **et al.** (2019) Structural basis for blue-green light harvesting and energy dissipation in diatoms. *Science*, **363**, eaav0365.

Weber, T., **Gruber, A.** & **Kroth, P.G.** (2009) The presence and localization of thioredoxins in diatoms, unicellular algae of secondary endosymbiotic origin. *Molecular Plant*, **2**, 468–477.

Xie, Y., **Wu, B.**, **Wu, Z.**, **Tu, X.**, **Xu, S.**, **Lv, X.** **et al.** (2020) Ultrasound-assisted one-phase solvent extraction coupled with liquid chromatography-quadrupole time-of-flight mass spectrometry for efficient profiling of egg yolk lipids. *Food Chemistry*, **319**, 126547.

Yang, M., **Lin, X.**, **Liu, X.**, **Zhang, J.** & **Ge, F.** (2018) Genome annotation of a model diatom *Phaeodactylum tricornutum* using an integrated proteogenomic pipeline. *Molecular Plant*, **11**, 1292–1307.

Yoneda, K., **Yoshida, M.**, **Suzuki, I.** & **Watanabe, M.M.** (2016) Identification of a major lipid droplet protein in a marine diatom *Phaeodactylum tricornutum*. *Plant and Cell Physiology*, **57**, 397–406.

Zhang, C. & **Hu, H.** (2014) High-efficiency nuclear transformation of the diatom *Phaeodactylum tricornutum* by electroporation. *Marine Genomics*, **16**, 63–66.

Zhang, Y., **Pan, Y.**, **Ding, W.**, **Hu, H.** & **Liu, J.** (2021) Lipid production is more than doubled by manipulating a diacylglycerol acyltransferase in algae. *GCB Bioenergy*, **13**, 185–200.