



**HAL**  
open science

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

Teng Huang, Yufang Pan, Eric Maréchal, Hanhua Hu

► **To cite this version:**

Teng Huang, Yufang Pan, Eric Maréchal, Hanhua Hu. Proteomes reveal the lipid metabolic network in the complex plastid of *Phaeodactylum tricornutum*. *The Plant Journal*, 2024, 117, pp.385-403. 10.1111/tpj.16477 . hal-04216697

**HAL Id: hal-04216697**

**<https://hal.science/hal-04216697v1>**

Submitted on 25 Sep 2023

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

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



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

Teng Huang<sup>1,2</sup>, Yufang Pan<sup>1</sup>, Eric Maréchal<sup>3</sup> and Hanhua Hu<sup>1,2,\*</sup>

<sup>1</sup>Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Laboratoire de Physiologie Cellulaire Végétale, Université Grenoble Alpes, CEA, CNRS, INRA, IRIG-LPCV, 38054 Grenoble Cedex 9, France

\*For correspondence (e-mail: hanhuahu@ihb.ac.cn).

1

## SUMMARY

*Phaeodactylum tricorutum* plastid is surrounded by four membranes, and its protein composition and function remain mysterious. In this study, the *P. tricorutum* plastid-enriched fraction was obtained and 2850 proteins were identified, including ninety-two 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, 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.

### **Significance statement**

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 harbours 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.

### **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 (Kilian & Kroth 2005; Flori et al., 2016; Cavalier-Smith, 2018). 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 (Staelin, 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<sub>2</sub> 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 (Roessler & Ohlrogge, 1993; Li et al., 2018a). 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, i.e. 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 (Lupette et al., 2019; Leyland et al., 2020), 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). 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 four-layered membrane, and the integrity and purification were analyzed by observation of transformants expressing plastid marker proteins fused with a FP. A total of 2850 proteins were identified in the plastid-enriched fraction through 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 to 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 hours 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 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 detected in plastid samples. The relative abundance of three pyrenoid-localized proteins, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (rbcL), PtCA1 ( $\beta$ -carbonic anhydrase 1; Tanaka et al., 2005), and PtCA2 ( $\beta$ -CA2; Kitao et al., 2008), in plastid samples was comparable to that in whole-cell samples. Two cytosolic protein antibodies ( $\beta$ -actin and  $\beta$ -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 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 have 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 (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), 82 out of 88 lipid droplet proteins (Yoneda et al., 2016; Lupette et al., 2019), 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 (Lepetit et al., 2007; Grouneva et al., 2011; Herbstová et al., 2015) if not three different peptides (Joshi-Deo et al., 2010; Gundermann et al., 2013). 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 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; and 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 below 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 protein (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, 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 Category 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 0.5~1, 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% (1span style="font-family:'Times New Roman'">51) of the Category II & III proteins (1455) from plastid samples were predicted to be chloroplast-targeted as well (Figure S4). For the predicted 12,382 nuclear-encoded proteins in *P. tricornutum*, only 1471 and 561 proteins were considered as chloroplast-targeted by ASAFind and HECTAR 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  $\beta$ -oxidation (Jallet et al., 2020), the possibility of a plastidial FA  $\beta$ -oxidation could not be excluded based on the 16 FA  $\beta$ -oxidation related enzymes detected in the plastid samples and on the prediction (Data S14). Additional details of all identified proteins in our proteomes were showed 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 synthases, 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 4 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:  $\Delta 9$  acyl-CoA desaturase; OEM $\Delta 6$ FAD; PtFAD2; ER $\Delta 5$ FAD.1), FA elongase (ELO) ( $\Delta 6$ ELO.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- $\beta$ -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 3 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), 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-*sn*2-PG respectively to generate 16:1<sup>Δ9</sup>-ACP and 16:1<sup>Δ3trans</sup>-*sn*2-PG, which occurred in the plastid stroma based on the GFP fusion protein signals. Ordered addition of double bonds from 16:1<sup>Δ9</sup> to 16:2<sup>Δ9,12</sup>, 16:3<sup>Δ6,9,12</sup>, and 16:4<sup>Δ6,9,12,15</sup> 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 et al., 2018b). ADS, acting on 18:0-CoA to form 18:1<sup>Δ9</sup>-CoA, fused GFP fluorescence formed irregular layers and 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<sup>Δ9</sup> to 18:2<sup>Δ9,12</sup>, 18:3<sup>Δ6,9,12</sup>/18:3<sup>Δ9,12,15</sup>, and 18:4<sup>Δ6,9,12,15</sup> 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<sup>Δ6,9,12,15</sup> to 20:4<sup>Δ8,11,14,17</sup>, desaturation from 20:4<sup>Δ8,11,14,17</sup> to 20:5<sup>Δ5,8,11,14,17</sup> (eicosapentaenoic acid, EPA), elongation from EPA to 22:5<sup>Δ7,10,13,16,19</sup>, and desaturation from 22:5<sup>Δ7,10,13,16,19</sup> to 22:6<sup>Δ4,10,13,16,19</sup> (docosahexaenoic acid, DHA), also resided at the cERM.

Six DGATs have been identified in *P. tricornutum*, and 3 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 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-days 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 sub-population formation (Jaussaud et al., 2020). Furthermore, small green dots were also observed from the isolated plastids stained with BODIPY 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)..

**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 KY, 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, USA*, **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 DW, Xie WH, Hao TB, Cai JX, Zhou TB, Balamurugan S, et al.** (2018a) 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** (2018b) 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.

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

**Millar AH, Sweetlove LJ, Giegé P & Leaver CJ** (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiology*, **127**(4), 1711–1727.

**Moog D, Rensing SA, Archibald JM, Maier UG & Ullrich KK** (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 UG** (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 AM & Gibbs SP** (1992) Immunocytochemical localization of photosystem I and the fucoxanthin-chlorophylla/c light-harvesting complex in the diatom *Phaeodactylum tricornutum*. *Protoplasma*, **166**, 208–217.
- Remmers IM, D'Adamo S, Martens DE, de Vos RCH, Mumm R, America AHP, et al.** (2018) Orchestration of transcriptome, proteome and metabolome in the diatom *Phaeodactylum tricornutum* during nitrogen limitation. *Algal Research*, **35**, 33–49.
- Roessler PG & Ohlroge JB** (1993) Cloning and characterization of the gene that encodes acetyl-coenzyme A carboxylase in the alga *Cyclotella cryptica*. *Journal of Biological Chemistry*, **268**, 19254–19259.
- Rolland N, Bouchnak I, Moyet L, Salvi D & Kuntz M** (2018) The main functions of plastids. In E Maréchal, ed., *Plastids: Methods and Protocols*. New York, USA: Springer, 73–85.
- Salvato F, Havelund JF, Chen M, Rao RSP, Rogowska-Wrzesinska A, Jensen ON, et al.** (2014) The potato tuber mitochondrial proteome. *Plant physiology*, **164**(2), 637–653.
- Schober AF, Flori S, Finazzi G, Kroth PG & 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, 189–203.
- Schober AF, Río Bártulos C, Bischoff A, Lepetit B, Gruber A & Kroth PG** (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 EB** (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 BM, 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 JA, et al.** (2021) Plastidial acyl carrier protein  $\Delta 9$ -desaturase modulates eicosapentaenoic acid

- biosynthesis and triacylglycerol accumulation in *Phaeodactylum tricornutum*. *Plant Journal*, **106**, 1247–1259.
- Smith SR, Dupont CL, McCarthy JK, Broddrick JT, 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.
- Staelin LA** (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 JM, Wilhelmi I, Zauner S & Maier UG** (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, Zybilov B, Majeran W, Friso G, Olinares PDB & van Wijk KJ** (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  $\beta$ -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 DD, 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 KA, Xie Y, Singer ER, 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 LJ, 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 PG** (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 MM** (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.

### Figure legends

**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, periplastidial 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.

**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. Abbreviations: FCP, fucoxanthin chlorophyll a/c-binding protein; TPT1, triose phosphate translocator 1; rbcL, ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit; PtCA1,  $\beta$ -carbonic anhydrase 1; PtCA2,  $\beta$ -CA2; AGO, argonaute protein; RAE1, ribonucleic acid export factor 1; ME1, NAD-dependent malic enzyme; PEPC2, phosphoenolpyruvate carboxylase 2; SHMT2, serine hydroxymethyltransferase 2; StLDP, stramenopile lipid droplet protein; COP1, coatamer subunit alpha; CNX, ER-targeted calnexin; ER, endoplasmic reticulum.

**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, Leitan 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).

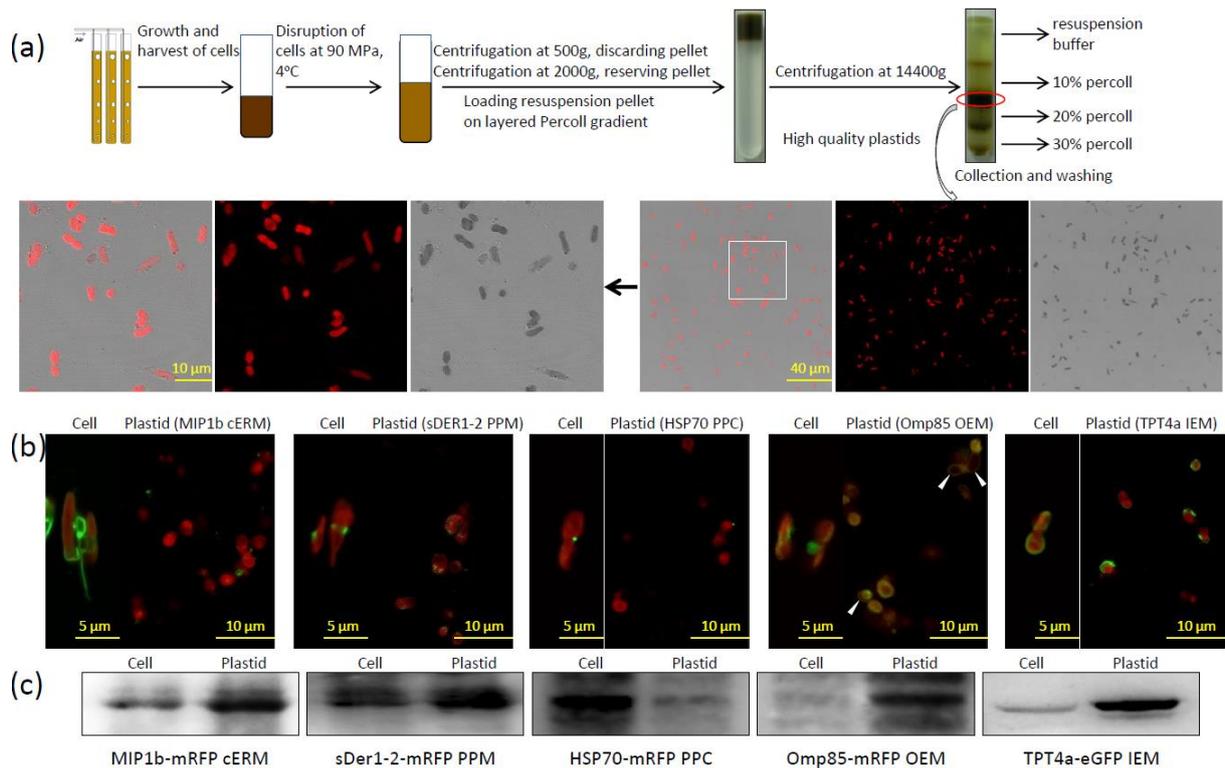
**Figure 4.** Fold change values (plastid/cell) of detected organelle-encoded proteins and plastid samples' nuclear-encoded proteins. **(a)** Organelle-encoded proteins detected in all samples. **(b)** Nuclear-encoded proteins detected in plastid samples. Fold change: The ratio of protein relative abundances of plastid versus those of whole cell 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.

**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 image of chlorophyll and GFP fluorescence from left to right. Abbreviations: MCAT, malonyl-CoA:ACP transacylase; PAD, palmitoyl-ACP desaturase; FAD, fatty acid desaturase; ADS,  $\Delta^9$  acyl-CoA desaturase; ELO, fatty acid elongase. Scale bar = 5  $\mu\text{m}$ .

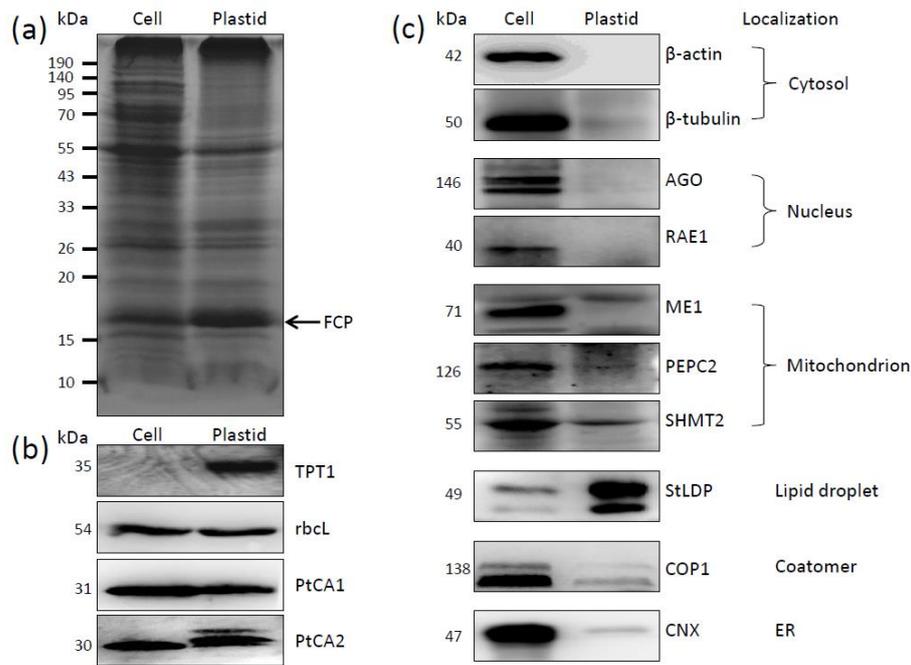
**Figure 6.** Subcellular localization of enzymes involved in lipid metabolism. Panels show microscopical images of transmitted light, chlorophyll autofluorescence (red), GFP fluorescence (green), merged image of chlorophyll and GFP fluorescence from left to right. Abbreviations: DGAT, acyl CoA:diacylglycerol acyltransferase; PGPP, phosphatidylglycerolphosphate phosphatase; SQD1, UDP-sulfoquinovose synthase; DGK, diacylglycerol kinase; EPT/CPT, ethanolamine-/choline-phosphotransferase; LPCAT, lysophosphatidylcholine acyltransferase; PLA2, phospholipase A2; Tgl1, SDP1-patatin like lipase. Scale bar = 5  $\mu\text{m}$ .

**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  $\pm$  SD ( $n = 3$ ). Asterisks above the bars indicate statistically significant differences from whole cells at  $P < 0.05$  level by  $t$  test.

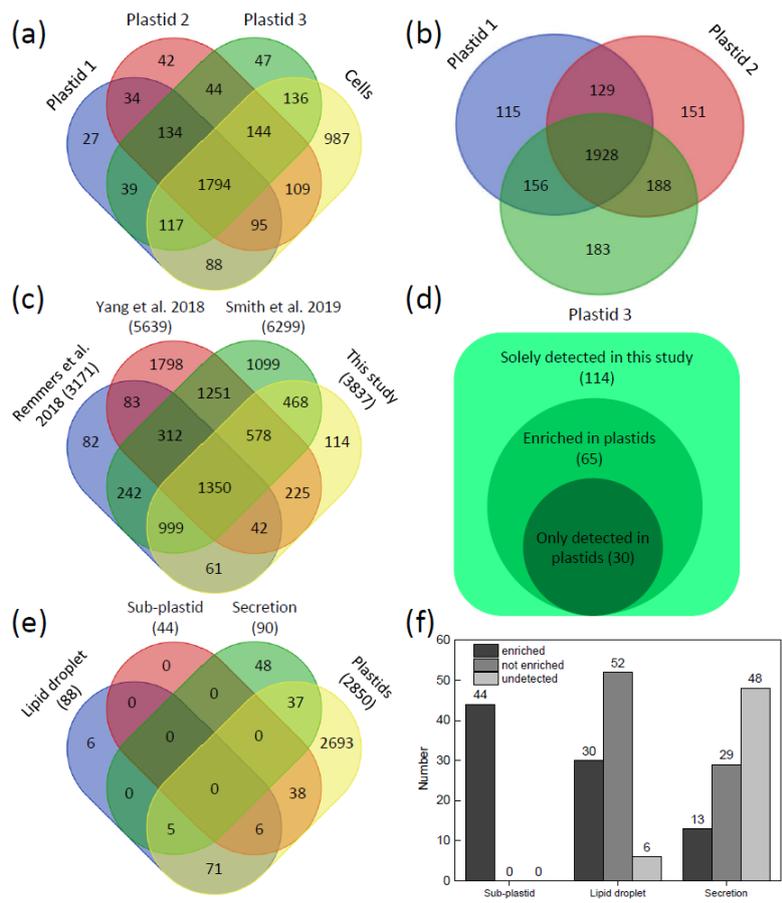
**Figure 8.** A model illustrating the fatty acid biosynthesis and lipid metabolism in *Phaeodactylum tricornutum*. Proteins present in the putative plastid proteomes were in red, protein localizations determined by GFP were shown in green box, and proteins with transmembrane domains were underlined. Abbreviation: 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- $\beta$ -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; MDGDS, 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.



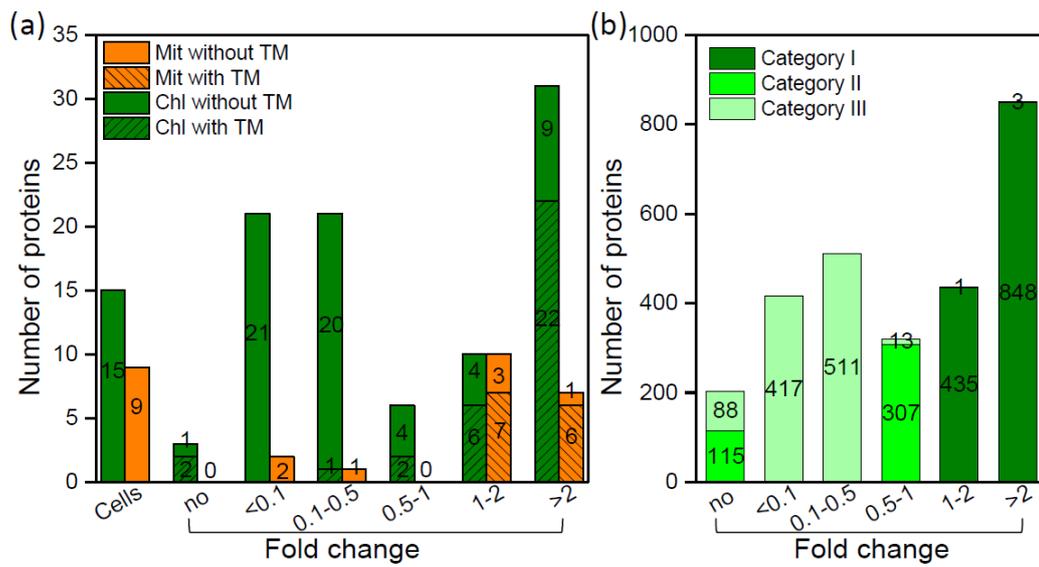
**Figure 1.** Integrity analysis of isolated plastids from *Phaeodactylum tricoratum*. **(a)** Isolation workflow, position in Percoll gradient (indicated by a red circle) and fluorescence micrographs of plastids. **(b)** Fluorescence micrographs of whole cells and isolated plastids from transformants 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, periplastidial 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.



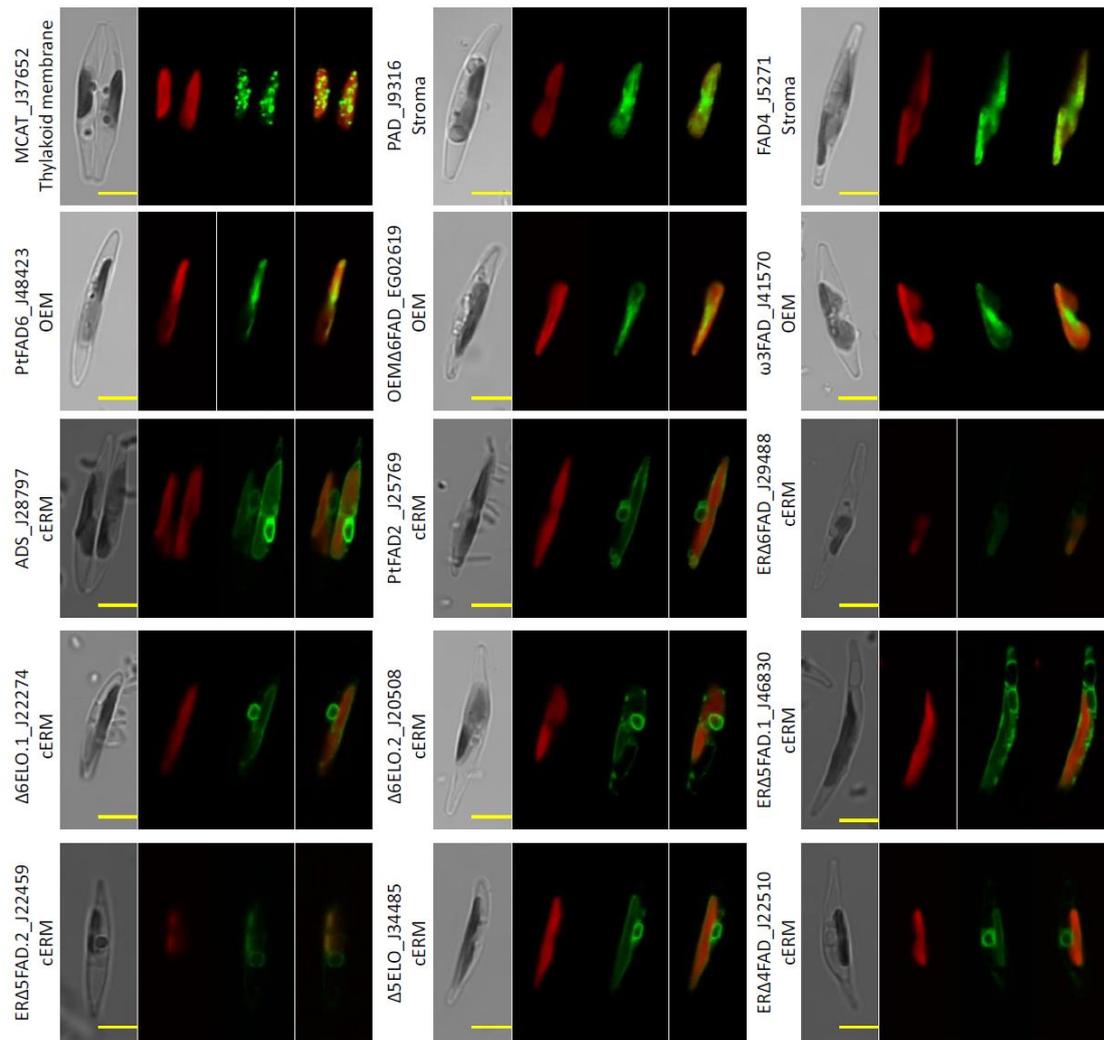
**Figure 2.** Western blot analysis of protein extracts from whole cells and isolated plastids. **(a)** Coomassie brilliant blue staining. **(b)** Immunoblot with antibodies against plastid proteins. **(c)** Immunoblot with antibodies against non-plastid proteins. Abbreviations: FCP, fucoxanthin chlorophyll a/c-binding protein; TPT1, triose phosphate translocator 1; rbcL, ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit; PtCA1,  $\beta$ -carbonic anhydrase 1; PtCA2,  $\beta$ -CA2; AGO, argonaute protein; RAE1, ribonucleic acid export factor 1; ME1, NAD-dependent malic enzyme; PEPC2, phosphoenolpyruvate carboxylase 2; SHMT2, serine hydroxymethyltransferase 2; StLDP, stramenopile lipid droplet protein; COP1, coatamer subunit alpha; CNX, ER-targeted calnexin; ER, endoplasmic reticulum.



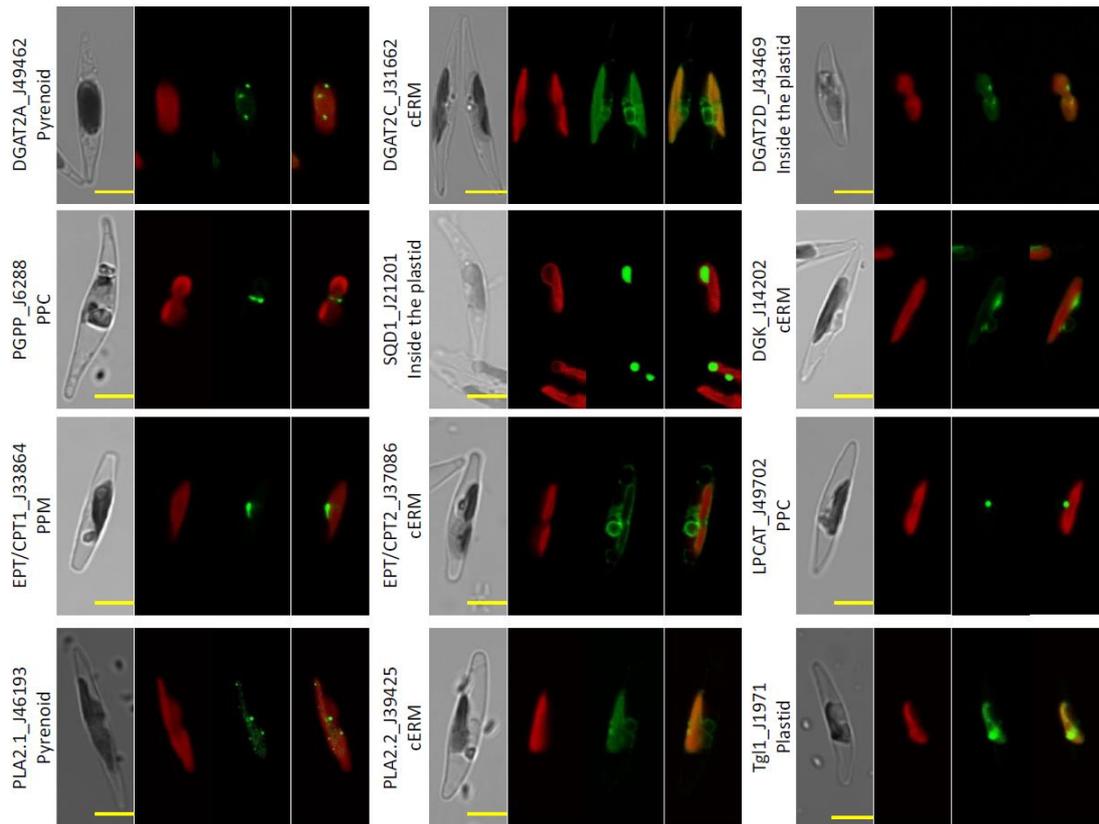
**Figure 3.** Comparison of proteins detected in the isolated plastids and whole cells, and proteins reported before in *Phaeodactylum tricornutum*. **(a)** Venn diagram representation of absolute protein numbers detected in the whole cells and three plastid samples. **(b)** Comparison of protein identified 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).



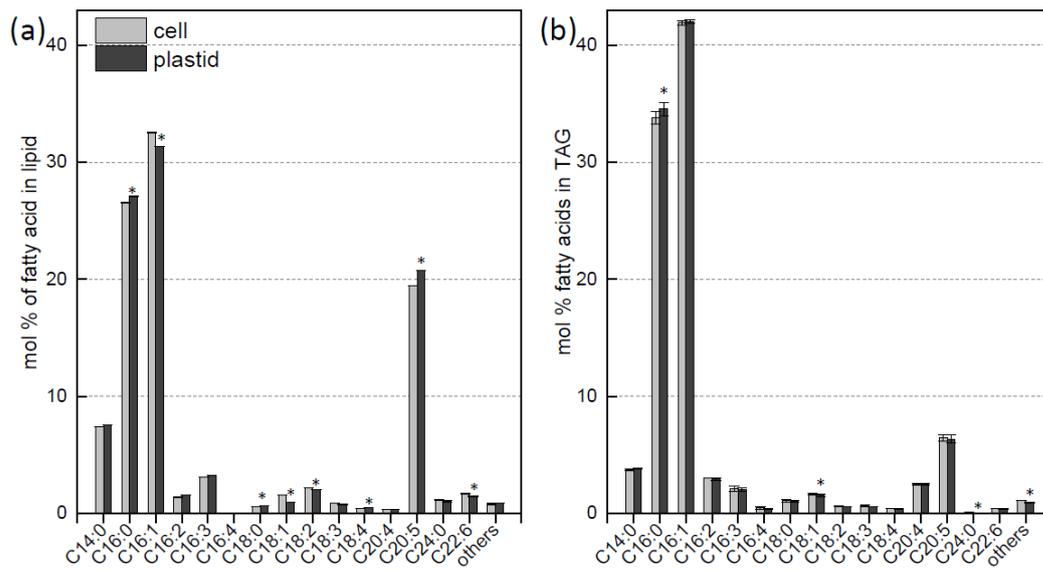
**Figure 4.** Fold change values (plastid/cell) of detected organelle-encoded proteins and plastid samples' nuclear-encoded proteins. **(a)** Organelle-encoded proteins detected in all samples. **(b)** Nuclear-encoded proteins detected in plastid samples. Fold change: The ratio of protein relative abundances of plastid versus those of whole cell 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.



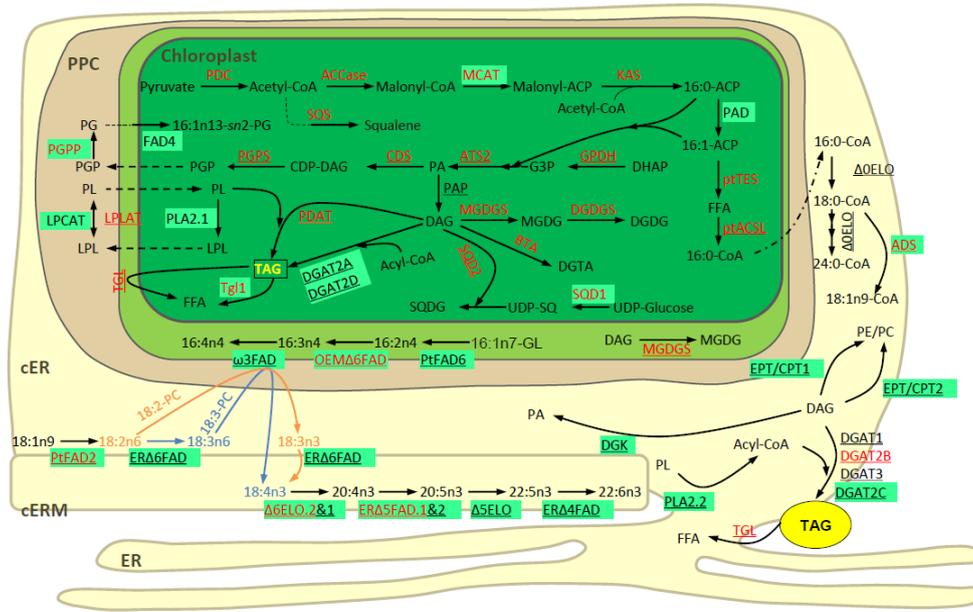
**Figure 5.** Subcellular localization of enzymes involved in fatty acid biosynthesis. Panels show microscopical images of transmitted light, chlorophyll autofluorescence (red), GFP fluorescence (green), merged image of chlorophyll and GFP fluorescence from left to right. Abbreviations: MCAT, malonyl-CoA:ACP transacylase; PAD, palmitoyl-ACP desaturase; FAD, fatty acid desaturase; ADS,  $\Delta 9$  acyl-CoA desaturase; ELO, fatty acid elongase. Scale bar = 5  $\mu\text{m}$ .



**Figure 6.** Subcellular localization of enzymes involved in lipid metabolism. Panels show microscopical images of transmitted light, chlorophyll autofluorescence (red), GFP fluorescence (green), merged image of chlorophyll and GFP fluorescence from left to right. Abbreviations: DGAT, acyl CoA:diacylglycerol acyltransferase; PGPP, phosphatidylglycerolphosphate phosphatase; SQD1, UDP-sulfoquinovose synthase; DGK, diacylglycerol kinase; EPT/CPT, ethanolamine-/choline-phosphotransferase; LPCAT, lysophosphatidylcholine acyltransferase; PLA2, phospholipase A2; Tgl1, SDPI-patatin like lipase. Scale bar = 5  $\mu$ m.



**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  $\pm$  SD ( $n = 3$ ). Asterisks above the bars indicate statistically significant differences from whole cells at  $P < 0.05$  level by  $t$  test.



**Figure 8.** A model illustrating the fatty acid biosynthesis and lipid metabolism in *Phaeodactylum tricorutum*. 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. Abbreviation: 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; 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, diacylglycerol hydroxymethyltrimethyl- $\beta$ -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.