Title:

Characterization of the three monogalactosyldiacylglycerol synthase isoforms in the diatom *Phaeodactylum tricornutum*

Authors:

Nolwenn Guéguen[1], Félix Cicéron[1,¶], Valérie Gros[1,#], Grégory Si Larbi[1], Denis Falconet[1], Etienne Deragon[1], Siraba D . Gueye[1], Dimitris Petroutsos[1], Hanhua Hu[2], Yangmin Gong[3,4], Morgane Michaud[1], Juliette Jouhet[1], Juliette Salvaing[1], Alberto Amato[1], Eric Maréchal[1,*].

Affiliations:

[1]Laboratoire de Physiologie Cellulaire et Végétale, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Centre National de la Recherche Scientifique, Commissariat à l'Energie Atomique et aux Energies Alternatives, Université Grenoble Alpes; IRIG, CEA-Grenoble, 17 rue des Martyrs; 38000 Grenoble, France

[2]Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

[3]Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430062, China

[4]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan 430062, China

[¶]Present address: Living Systems Institute - Exeter University Stocker Road; EX4 4QD, Exeter, UK

[#]Present address: Institut des Biomolécules Max Mousseron; Pôle Chimie Balard Recherche, 1919 route de Mende, 34293 Montpellier

***Correspondence:**

eric.marechal@cea.fr

Short Title:

MGD multigenic family in diatoms

Material distribution:

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plcell/pages/General-Instructions) are Alberto Amato (alberto.amato@cea.fr) and Eric Maréchal (eric.marechal@cea.fr).

Abstract.

Diatom cell derives from a secondary endosymbiosis, after the engulfment of a red alga and its subsequent reduction into a complex plastid. Four membranes bound this plastid: an inner and outer envelope of chloroplast origin, and peri- and epiplastidial membranes (PPM, EpM) connected to the endoplasmic reticulum (ER). A vesicular network buds from the PPM, forming a blob-like structure. The committing step for the synthesis of galactolipids, conserved in photosynthetic eukaryotes, is catalyzed by galactosyltransferases called MGDs, producing monogalactosyldiacylglycerol (MGDG). Like in angiosperms, a multigenic family of MGDs has emerged in diatoms, but via an independent evolutionary process. We characterized MGDα, MGDβ and MGDγ in *Phaeodactylum tricornutum*, combining molecular analyses, heterologous expression in yeast, and phenotypic study of overexpressing and Crispr/CAS9-edited lines. MGDα is targeted to thylakoids, MGDβ to the blob/PPM, and MGDγ to the ER/EpM and possibly the cytosol. MGDs have distinct specificities for diacylglycerol, consistent with their localization. Non-mutated MGDs could partly compensate for the loss of each isoform, providing insights on lipid homeostasis control in diatom cells. Results suggest that MGDα is required for thylakoid expansion in optimal conditions, whereas MGDβ and MGDγ have a role in extra-plastidial membrane lipid biosynthesis and in response to environmental stress.

Introduction.

Diatoms form a monophyletic phylum of photosynthetic eukaryotes, representing one of the largest groups of stramenopiles (de Vargas et al., 2015; Serôdio and Lavaud, 2020). Their ecological success is remarkable, as they dominate phytoplankton biodiversity in oceanic and freshwater ecosystems (Mann and Droop, 1996; Mann, 1999; Kooistra et al., 2007; Guiry, 2012; de Vargas et al., 2015). They account for ~40% of net primary production in oceans, and about one fifth of all the photosynthesis on Earth (Field et al., 1998). Our understanding of diatom subcellular organization and biology relies mainly on the model species *Phaeodactylum tricornutum*, for which intensive efforts have been made to sequence its genomes (Bowler et al., 2008), develop molecular tools and resources (De Martino et al., 2007; De Riso et al., 2009; Daboussi et al., 2014; Kroth et al., 2018), and characterize all membrane glycerolipids (Abida et al., 2015).

Diatoms have a cell membrane compartmentation, which is extremely complex, comprising membrane systems having no equivalent in 'more simple' models we are familiar with, like yeast, *Chlamydomonas* or *Arabidopsis*. This complex membrane architecture derives from their evolutionary origin following two endosymbiosis events. A primary endosymbiosis event consisted schematically in the engulfment of a cyanobacterium by an unknown heterotrophic cell, 1-1.5 b.y.a. (Jensen and Leister, 2014; Marechal, 2018). Eukaryotes deriving directly from this primary endosymbiosis contain a chloroplast, an organelle bounded by a two-membrane envelope, considered to be originating from the two cell membranes of the cyanobacterium (Sato and Awai, 2017; Marechal, 2018). They radiated into three main lineages, namely green algae (chlorophyta), red algae (rhodophyta) and glaucophytes. Plants derive from green algae, and their cells contain also primary chloroplasts. Appearance of diatoms is estimated to have occurred ~180-250 m.y.a. (Kooistra et al., 2003; Sorhannus, 2007), stemming from a second endosymbiosis event, following the engulfment of a red alga by a second heterotrophic eukaryote (Reyes-Prieto et al., 2007; Bowler et al., 2008; Flori et al., 2016; Benoiston et al., 2017).

In diatoms, the photosynthetic organelle differs from the primary chloroplast. It emerged from the reduction of the engulfed red alga, forming a so-called 'complex' or 'secondary plastid' bounded by four membranes (Figure 1). The origin of the two innermost membranes is not debated, corresponding to the chloroplast envelope of the engulfed red alga, namely the 'outer' and 'inner envelope membranes' (oEM and iEM, respectively; Figure 1). Different suppositions have been made about the origin of the additional bounding membranes. The

outermost, named the ‘epiplastidial membrane’ (EpM, Figure 1), could derive from the host phagosome. It is continuous with the outer membrane of the nuclear envelope (oNE) and the endoplasmic reticulum (ER) (Murakami and Hashimoto, 2009; Tanaka et al., 2015a; Flori et al., 2016). Underneath, the ‘periplastidial membrane’ (PPM, Figure 1) is usually considered to derive from the red algal (symbiont) plasma membrane (Grosche et al., 2014). Alternatively it has been hypothesized that the PPM could derive from the host ER, like the EpM (Gould et al., 2015). A ‘blob-like’ structure, or in short ‘blob’, containing a vesicular network (VN) visible in confocal and transmission electron microscopy (Figure 1) has been detected between the PPM and the oEM, but its function remains elusive (Kilian and Kroth, 2005; Flori et al., 2016). The blob was proposed to be a relic of the red algal ER (Cavalier-Smith, 2018). The EpM, PPM and blob are therefore related to the endomembrane system. To picture diatoms’ sophisticated plastid architecture completely, membrane contact sites occur between the inner membrane of the nuclear envelope (iNE) and the PPM (Flori et al., 2016) (Figure 1).

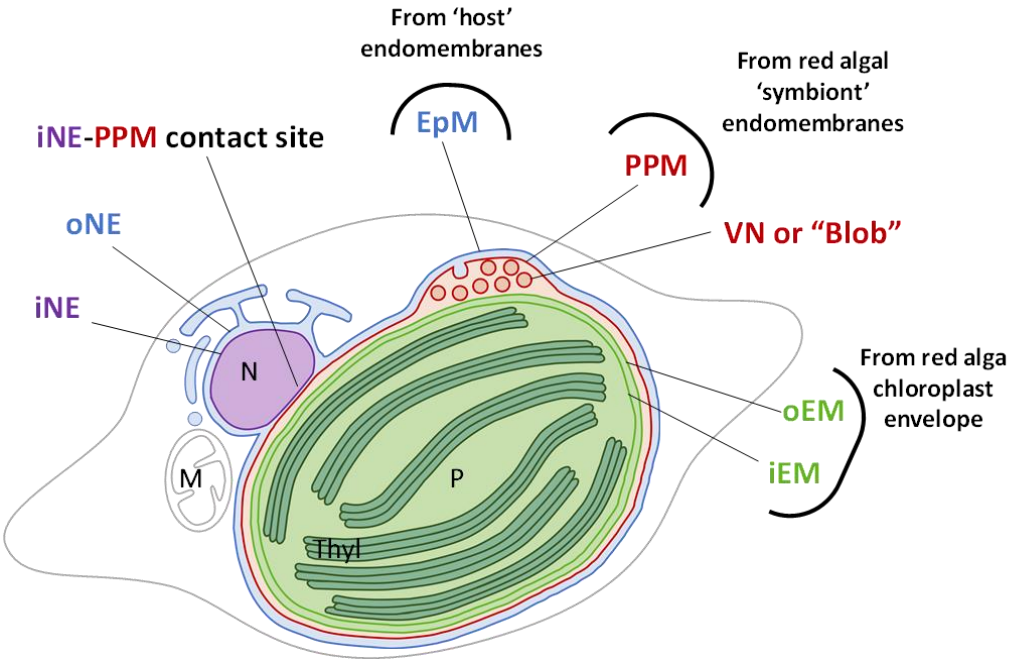

Figure 1: Chimeric origin of the secondary plastid in diatom. The scheme shows a fusiform cell of *Phaeodactylum tricornutum*. The plastid is limited by four membranes. The epiplastidial membrane (EpM), shown in blue, is continuous with the outer nuclear envelope (oNE). The periplastidial membrane (PPM), shown in red, is in tight contact with the inner nuclear envelope (iNE). The outer and inner envelope membranes (iEM and oEM), shown in light green, are tightly apposed. A specific periplastidial compartment relies on the detection of a “blob”-like structure observed by confocal and transmission electron microscopy. In confocal microscopy, the precursors of blob-residing proteins fused to GFP cross only the EpM and PPM and form fluorescent spots in the middle of the plastid (Lang et al., 1998; Kilian and Kroth, 2005). In electron microscopy, the blob is marked by the development of a vesicular network (VN) budding from the PPM (Flori et al., 2016). Each membrane surrounding the plastid derive from distinct evolutionary origin. The oEM and iEM stem from the chloroplast of the engulfed red algae. The EpM and PPM potentially derive from the endomembrane systems of the red algal symbiont and the host cell, respectively. M, mitochondrion; N, nucleus; P, plastid; Thyl: thylakoids.

Identifying the protein and lipid components of the iEM, oEM, PPM and EpM, and unravelling the biogenesis, dynamics and function of each of these membranes represent a challenge, since little knowledge can be transferred from simple eukaryotic models. The import of nuclear-encoded proteins into diatom plastid has been dissected in *P. tricornutum* and *Thalassiosira pseudonana*, giving insights on proteins participating in plastid biogenesis. Plastid protein precursors contain a bipartite N-terminal targeting peptide with a signal peptide (Sp) fused to a chloroplast transit peptide-like sequence (Ctp) and a conserved amino acid motif called 'ASAFAP' (Apt et al., 2002; Kilian and Kroth, 2005; Gruber et al., 2007; Gruber et al., 2015). Based on the bioinformatic predictions of the Sp-Ctp and the ASAFAP motif, about 8% of the proteins encoded in *P. tricornutum* nuclear genomes were predicted to be plastid localized and therefore represent a putative plastid proteome (Gruber et al., 2015).

By contrast with proteins, the lipid composition of diatom plastid membranes and the sites of their biosynthesis is unknown. Membrane glycerolipids contain two fatty acids (FA) esterified at positions *sn-1* and *sn-2* of a glycerol backbone, and harbor a polar head group at position *sn-3*. The nature of the head group defines the 'class' of glycerolipid, while FA chain-lengths and unsaturation levels define 'molecular species' in each class (Li-Beisson et al., 2010; Petroutsos et al., 2014). It was considered plausible but not demonstrated yet that the iEM and oEM may contain the glycerolipids conserved in cyanobacteria and primary chloroplast, *i.e.* two galactolipids, monogalactosyl- and digalactosyldiacylglycerol (MGDG and DGDG), a sulfolipid, sulfoquinovosyldiacylglycerol (SQDG) and a phospholipid, phosphatidylglycerol (PG) (Boudiere et al., 2014; Petroutsos et al., 2014). The EpM and PPM may be related to the ER/endomembrane lipid composition, and contain phospholipids, e.g. phosphatidylcholine (PC), phosphatidylethanolamine (PE), and betaine lipids, e.g. diacylglyceryl hydroxymethyltrimethyl-β-alanine (DGTA) (Dolch and Marechal, 2015). It is still unknown whether MGDG, DGDG or SQDG presence could expand to the PPM and EpM or any other membrane in diatoms.

Here we addressed the committing step of photosynthetic membrane galactolipid synthesis, *i.e.* the production of MGDG by MGD enzymes. MGDs catalyze the transfer of a galactosyl residue from a uridine-diphosphate-galactose (UDP-Gal) to the *sn-3* position of a diacylglycerol (DAG):

UDP-Gal + DAG —(MGD)→ MGDG + UDP

MGDs were characterized in angiosperms containing primary chloroplasts. Based on the multigenic family identified in *Arabidopsis thaliana* (AtMGD1, AtMGD2 and AtMGD3), two groups have been identified, with distinct subcellular localization and distinct roles (Awai et al., 2001).

- On the one hand, 'type A' consists of the AtMGD1 isoform, which precursor has a Ctp allowing its targeting to the iEM of the chloroplast (Awai et al., 2001). It is the major isoform, responsible for the synthesis of MGDG for thylakoid membranes and its knockout (KO) is lethal in photoautotrophic conditions (Kobayashi et al., 2007a, b).
- On the other hand, AtMGD2 and AtMGD3 belong to 'type B', lacking any N-terminal Ctp. They are dispensable under optimal growth conditions (Kobayashi et al., 2009a). Based on the analysis of GFP-fusion proteins, type B MGDs localize to the oEM of *A. thaliana* chloroplast (Awai et al., 2001). Based on expression patterns and the analysis of mutants, AtMGD2 and AtMGD3 proved to be crucial for adaptation of plant cells to nutritional stress (Kobayashi et al., 2009a). Recent work also shows a cytosolic localization of AtMGD2 in pollen tube (Billey et al., 2021a).

MGDG represents more than a third of membrane lipids in *P. tricornutum* (Abida et al., 2015). It is the main platform for the production of highly unsaturated forms of 16-carbon FAs deriving from plastid palmitoyl-ACP (16:0-ACP), in particular 16:2, 16:3 and 16:4 (Dolch and Marechal, 2015; Smith et al., 2021). *P. tricornutum* MGDG molecular species are also enriched in eicosapentaenoic acid (20:5), a very-long-chain polyunsaturated FA with twenty carbons and five desaturations. Synthesis of 20:5 follows a sequence of acyl-chain-elongations and desaturations, likely occurring at the ER (or membrane connected to the ER) on acyl chains esterified on phospholipids and betaine lipids (Moog et al., 2011; Dolch and Marechal, 2015; Dolch et al., 2017a; Billey et al., 2021b; Smith et al., 2021). The route of 20:5 from the ER to the plastid, tentatively called the "Ω-pathway" (Petroutsos et al., 2014; Dolch and Marechal, 2015; Dolch et al., 2017a) is still unresolved.

Three MGDs have been predicted previously in *P. tricornutum* based on sequence similarity (Petroutsos et al., 2014). We confirmed this prediction by a simple similarity search using AtMGD1, the main isoform in Arabidopsis, allowing the identification of three gene loci: *Phatr_J14125* (on chromosome 13), *Phatr_J43938* (on chromosome 3), and *Phatr_J9619* (on chromosome 1), that we named here *MGDα*, *MGDβ* and *MGDγ*, respectively. Four to five peptides per MGD protein were detected in a whole-cell proteome analysis of *P. tricornutum* (Lupette et al., In press), validating that all three genes are expressed.

It would be tempting to consider that a multigenic family of *MGD*s in diatoms would correspond to types A and B characterized in angiosperms, some being specialized in the synthesis of thylakoids, and some being dedicated to distinct function of MGDG molecules in response to environmental stress. However, such multigene family has not been characterized to date in red algae, which are at the origin of diatom plastids. Multiple MGDs may also be important for MGDG production in different membrane compartments, in particular some of the four limiting membranes of the secondary plastid. However, other lineages in stramenopiles also harboring a four-membrane plastid, like eustigmatophytes, contain only one MGD in their genome (Vieler et al., 2012; Corteggiani Carpinelli et al., 2014). Thus, the presence of three isoforms in *P. tricornutum* is intriguing.

In this study, we characterized the three MGD isoforms found in *P. tricornutum*, addressing the question of their subcellular localizations, nature of DAG substrates, involvement in the different MGDG and DGDG species productions, as well as their putative role under environmental stress, taking nitrogen starvation as model stress condition.

Results.

Identification of a diatom-specific multigenic family of MGD isoforms.

A midpoint rooted phylogenetic tree was generated using MGD protein sequences in taxa covering the diversity of plastid-containing eukaryotes (Figure 2), from green algae (*Chlamydomonas*, *Chlorella*, *Coccomixa*, *Monoraphidium*, *Ostreococcus*, and *Raphidocelis*), red algae (*Cyanidioschyzon*, *Cyanidiococcus*, *Gracillariopsis*, and *Porphyra*), mosses (*Physcomitrium*), lycopods (*Selaginella*), angiosperms (*Amborella*, *Arabidopsis*, *Brachypodium*, *Coffea*, *Nimphea*, *Oryza*, and *Spinacia*), stramenopiles (*Aureococcus*, *Ectocarpus*, *Nannochloropsis*, including diatoms, *i.e.* *Fistulifera*, *Fragilariopsis*, *Phaeodactylum*, *Pseudo-nitzschia*, and *Thalassiosira*), to alveolates (*Vitrella* and *Gregarina*). *Fistulifera solaris* being a diatom with an allodiploid genome structure (Tanaka et al., 2015b), only one of each homeologous gene was used in the phylogeny reconstruction. A proteobacterium (*Blastochloris viridis*) related glycosyltransferase, annotated as a prokaryotic MGDG synthase, was used as an outgroup (Supplemental Table 1).

The obtained molecular phylogeny (Figure 2) is consistent with taxonomy (Adl et al., 2019). Angiosperms MGDs form two groups, one clustering with type A AtMGD1 characterized in *A. thaliana*, and one containing the type B AtMGD2 and AtMGD3 (Awai et al., 2001), reflecting gene duplication and functional differentiation in flowering plants (Awai et al., 2001). The single MGDs encoded in bryophyta and lycopodiophyta cluster together with the angiosperms type A MGDs, consistently with the hypothesis that type A is the ancestral isoform in land plants.

MGDs from stramenopiles are closely related to rhodophyta (Figure 2), reflecting that they derived from a secondary endosymbiosis between a eukaryotic host cell and a red algal symbiont (Figure 1). With the exception of *Microchloropsis gaditana* (formerly known as *Nannochloropsis gaditana*) (Fawley et al., 2015), all selected stramenopiles contain two to three predicted MGD isoforms. The unique *M. gaditana* MGD clusters with MGDα, along with all the *Ectocarpus siliculosus* MGDs.

The multigenic family of *MGD*s in diatoms highlights three major clusters. Each MGD of *P. tricornutum* has counterparts in other diatoms, such as the pennate *Fistulifera solaris* and the centric *Thalassiosira pseudonana*. *Fragilariopsis cylindrus* seems to possess one MGDβ-like protein and one MGDγ-like protein, but apparently no MGDα-like protein, which was either undetected in the available genomic data, or lost. MGDα is in a cluster comprising MGDs

from other diatoms and other stramenopiles (eustigmatophyceae, pelagophyceae, phaeophyceae) in a sister clade to the alveolates, whereas MGDβ and MGDγ form a separate clade, sister to pelagophyceae. Since the stramenopile MGDα clade branches in a basal position to the alveolates, the red and the green lineages, MGDα, may have been vertically inherited from a MGD ancestor common to all. MGDβ and MGDγ likely stemmed from gene duplication at the base of diatoms radiation.

MGD phylogeny does not support any relation between the multigenic family found in angiosperms and that detected in diatoms. As a result, specific roles that MGDα, β and γ may have, cannot be deduced from any prior knowledge in any other photosynthetic eukaryote studied to date.

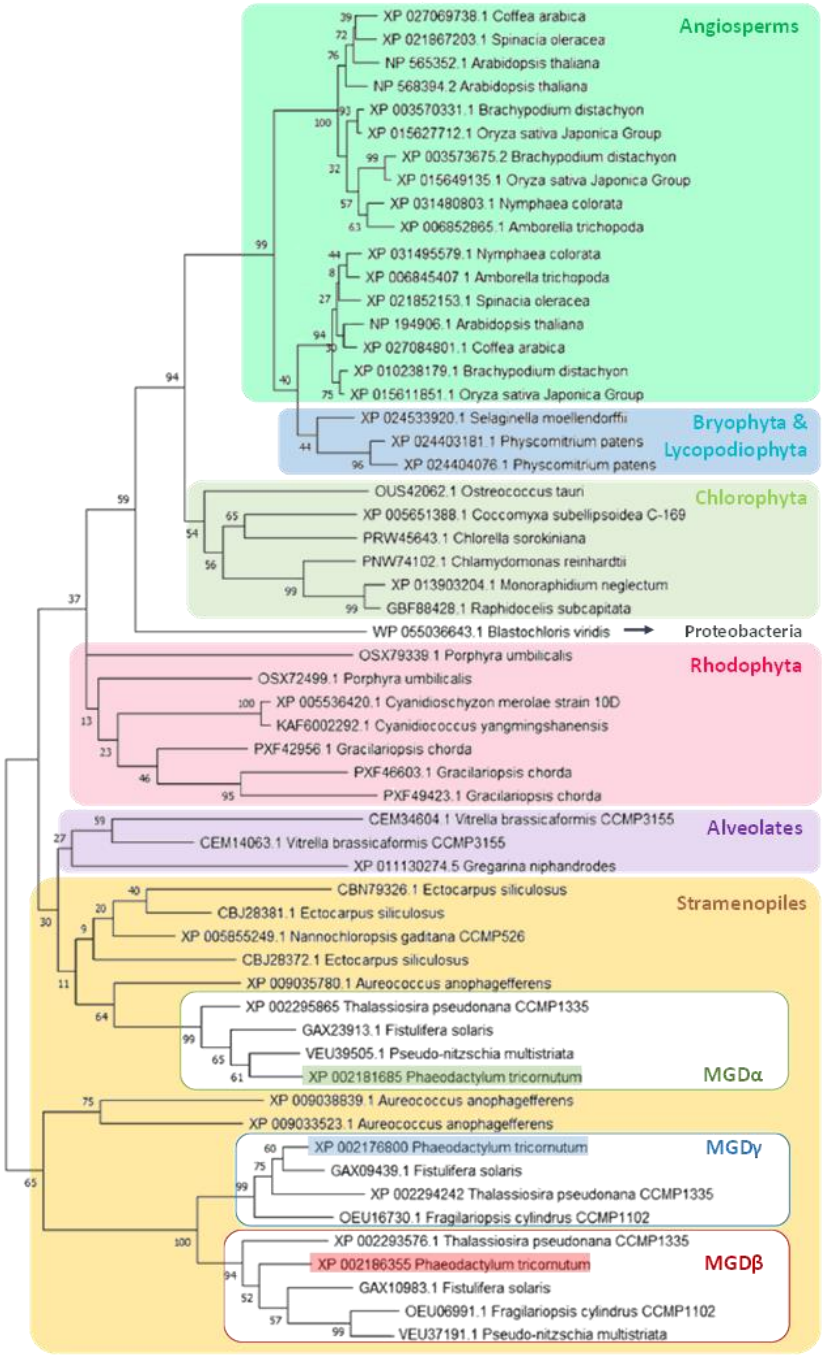

Figure 2: Phylogenetic analysis of MGDs from plastid-containing eukaryotes. MGD sequences were selected to cover the biodiversity of plastid-containing eukaryotes from green algae (*Chlamydomonas*, *Chlorella*, *Coccomixa*, *Monoraphidium*, *Ostreococcus* and *Raphidocelis* spp.), red algae (*Cyanidioschyzon*, *Cyanidiococcus*, *Gracillariopsis* and *Porphyra* spp.), mosses (*Physcomitrium*), lycopods (*Selaginella*), angiosperms (*Amborella*, *Arabidopsis*, *Brachypodium*, *Coffea*, *Nimphea*, *Oryza* and *Spinacia* spp.), stramenopiles (*Aureococcus*, *Ectocarpus*, *Nannochloropsis*, - including diatoms, *i.e. Fistulifera*, *Fragilariopsis*, *Phaeodactylum*, *Pseudo-nitzschia* and *Thalassiosira* spp.), to alveolates (*Vitrella* and *Gregarina* spp.). The tree was inferred by Maximum Likelihood, with 5,000 bootstrap pseudoreplicates, using the MEGAX software. The percentage of trees in which the associated taxa clustered together is shown next to the corresponding branches. Branch lengths are proportional to the number of substitutions per site. Main phyla are highlighted. Clusters corresponding to diatom MGDα, MGDβ and MGDγ are framed in green, red and blue, respectively. This tree highlights the conservation of a MGD multigenic family in diatoms, and the absence of evolutionary relationships with the multigenic family previously characterized in angiosperms.

MGDα, MGDβ and *MGDγ* gene structures and coding sequences

MGD sequences were recovered from genomic data of *P. tricornutum* CCAP 1055/1 via the EnsemblProtist web portal (Yates et al., 2022), verified experimentally and compared to entries in NCBI/Uniprot (Supplemental Table 2). As a source of confusion, the *Phatr3_J14125*/*Phatr3_EG02525* gene was previously numbered as either *MGD1* (Bullmann et al., 2010) or *MGD3* (Dolch et al., 2017b; Shang et al., 2022). We named genes with greek alphabet letters. The determination of *MGDα* (*Phatr3_J14125*/*Phatr3_EG02525*), *MGDβ* (*Phatr3_J54168*) and *MGDγ* (*Phatr3_J9619*) coding regions (with corrected exons and introns) in *P. tricornutum*, with 1,615, 1,805 and 2,099 bp lengths respectively, is summarized in Supplemental Figure 1. The presence and length of introns was confirmed by cDNA sequencing. The length of 5'-UTRs was estimated using series of primers designed to anneal upstream of the first predicted start codon on cDNA sequences, in combination with reverse primers binding in the first exons. For *MGDα*, we could obtain PCR products with a forward primer located 804 nucleotides upstream of the corrected start codon (second ATG predicted for *Phatr3_J14125*), suggesting that the 5'-UTR region was at least ~800 bp-long. For *MGDβ*, the cDNA sequence could be amplified 219, but not 460 bp upstream of the start codon, suggesting a 5'-UTR between 220 and 460 bp. For *MGDγ*, the cDNA could be amplified 49 but not 247 bp upstream of the start codon of *Phatr3_J9619*, suggesting a shorter 5'-UTR between 50 and 250 bp long. *MGDα* possesses therefore the longest 5'-UTR.

Prediction of distinct N-terminal targeting sequences in MGDα, MGDβ and MGDγ

The presence of targeting sequences in encoded MGD proteins was investigated in all isoforms. For this purpose, sequence alignment with AtMGD1 highlighted N-terminal variable portions, where putative targeting sequences could be sought (Figure 3E, N-terminus shown in grey). AtMGD1 contains a cTP allowing its localization to the chloroplast iEM (Awai et al., 2001). The N-terminal parts of varying length in *P. tricornutum* MGDs might therefore correspond to targeting sequences as well. In *P. tricornutum*, nuclear-encoded plastid proteins harboring a Sp-Ctp bipartite pre-sequence, marked by the presence of an ASAFAP motif, are driven across the four bounding membranes of the plastid (Apt et al., 2002). The function of the ASAFAP motif can be tuned by changing some residues, arresting the transport of pre-proteins after crossing the two outermost membranes, leading to their targeting to the blob at the PPM and the so-called periplastidial compartment (PPC) (Apt et al., 2002; Kilian and Kroth, 2005).

In MGDα, a Sp-Ctp could not be identified, when considering the first putative start codon predicted in the *Phatr3_J14125* CDS. However, it was evident when considering the only other methionine located upstream of the first conserved region. A Sp could clearly be identified using SignalP (Emanuelsson et al., 2007; Almagro Armenteros et al., 2019), and a cTP was then predicted downstream the presumptive cleavage site using WoLF PSORT (Horton et al., 2007) software, with default parameters. This prediction was confirmed using the HECTAR software specifically designed to detect Sp-Ctp in stramenopiles (Gschloessl et al., 2008). An 'ASAFAP'-like motif was tentatively identified in MGDα Sp-Ctp, with a SAAFSP sequence (Supplemental Figure 2). In MGDβ, four possible methionine start residues were detected upstream of the first conserved region shown in the alignment. However, only the first methionine led to a prediction of a Sp with SignalP and HECTAR. Then, a cTP was predicted starting from the cleavage site using both ChloroP and WoLF PSORT tools, but it was not detected using HECTAR. We deduced that this first residue was the actual start methionine of MGDβ. A low similarity ASAFAP-like motif was tentatively identified with a GSGFVL sequence (Supplemental Figure 2). In MGDγ, we validated the first methionine as the start, based on the splicing of an intron placed upstream of other possible methionine residues. However, it should be noted that between the intron and the first conserved region of the protein, as many as four methionine residues are present in the long N-terminal part of MGDγ. No prediction of a bipartite presequence could be obtained using any of these methionine codons as a start. Thus, MGDγ protein does not appear to contain any predictable targeting sequence, in support of a possible cytosolic localization.

Conservation of active site residues in MGDα, MGDβ and MGDγ structural models.

We used the structure of AtMGD1 as a template to compare *P. tricornutum* MGDs (Rocha et al., 2016; Makshakova et al., 2020). Specific AtMGD1 domains are involved in the binding to the UDP-Gal sugar donor, the DAG acceptor, and the lipid activator phosphatidylglycerol (PG). Supplemental Figure 2 shows the sequence alignment of the three AtMGDs and the three *P. tricornutum* MGDs and the conservation of these domains and key residues. Two residues are particularly critical for catalysis in AtMGD1 (Botte et al., 2005; Dubots et al., 2010; Rocha et al., 2016). Firstly, a histidine (H) in position 159 of the alignment shown in Supplemental Figure 2 is involved in the deprotonation of the nucleophile hydroxyl (OH) group of DAG. This residue is in the first highly conserved region of the proteins, which includes residues 148-170 of the alignment. Secondly, a lysine (K) in position 576 of the

alignment binds to the sugar donor, stabilizing the phosphates of the departing UDP during the reaction. K576 is also detected in a conserved region of *A. thaliana* and *P. tricornutum* proteins.

The presence of anionic lipids such as PA, PG and SQDG is critical for AtMGD1 activity. A PG-H catalytic dyad has been proposed to form an acid/base relay system (Makshakova et al., 2020; Nitenberg et al., 2020). Such a catalytic dyad would facilitate the deprotonation of the acceptor substrate by H159. It was shown that PG develops interactions with the catalytic residue H159 and the neighboring arginine R160, thus placing this lipid close to the active site (Makshakova et al., 2020). R160 is conserved in all *P. tricornutum* sequences further supporting the presence of a functional PG-H catalytic dyad (Supplemental Figure 2).

MGDs belong to the Glycosyltransferase B (GT-B) superfamily (Rocha et al., 2016). The conserved GT-B fold consists of two Rossmann-type α/β/α domains of comparable size, named the N- and C-domains, separated by a large cleft, and stabilized by two long C-terminal α-helices. The N-domain of AtMGD1 corresponds to positions 141-338 and 655-672, the C-domain to positions 362-654. They are connected by a long loop corresponding to positions 339-361, in the alignment (Supplemental Figure 2). Secondary and tertiary structures of *P. tricornutum* MGDs were predicted using Phyre² (Kelley et al., 2015), based on the first 3D-structure generated from the crystal of AtMGD1 in complex with UDP (template 4X1T) (Rocha et al., 2016) and the AlphaFold *ab initio* method (Jumper et al., 2021; Jumper and Hassabis, 2022; Mirdita et al., 2022). All obtained *P. tricornutum* MGD protein structures, starting approximately at position 145 of the alignment in Supplemental Figure 2 and ending around position 670, appear to be typical GT-B fold structures, highly similar to AtMGD1 (Figure 3A-D).

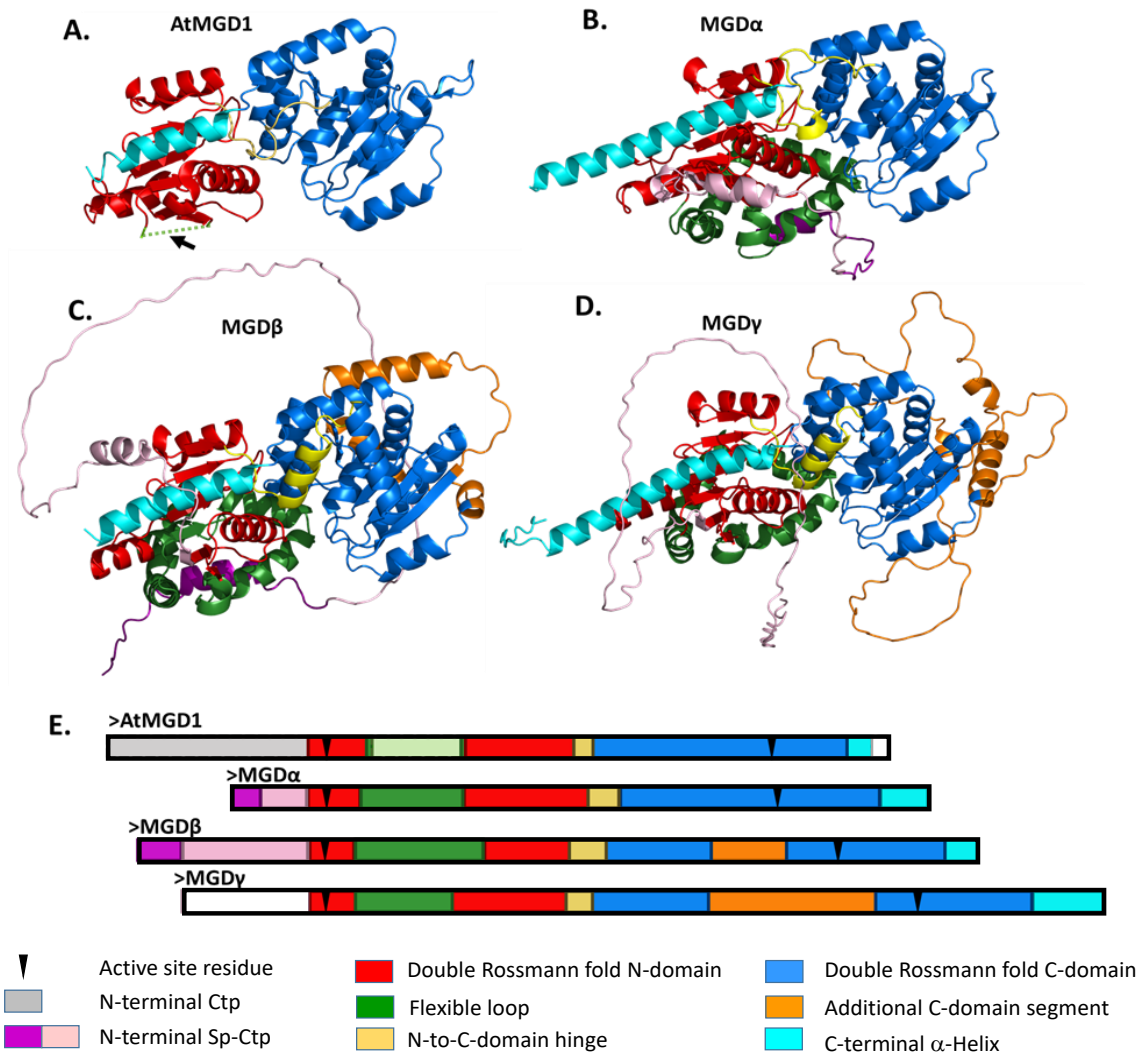

Figure 3: Protein models of *P. tricornutum* MGDs. (A) Protein structure of AtMGD1 from *A. thaliana*. The AtMGD1 protein model (4X1T) (Rocha et al., 2016) was retrieved from the Protein Data Bank (wwPDB consortium, 2019). (B-D) Protein models of MGDα, MGDβ and MGDγ from *P. tricornutum*. Models were obtained using AlphaFold. Structures are viewed with PyMOL software in the same orientation. (E) Linear representation of conserved structural domains. The identified protein segments are shown as follow. In the N-terminal end, AtMGD1 Ctp is shown in grey; putative MGDα and MGDβ Sp-Ctp in purple; other non-conserved N-terminal sequences in pink. In the double Rossmann fold N-domain, conserved sequences are shown in red, the AtMGD1 loop with unresolved structure in light green, and alpha helices at the level of this loop in dark green. The hinge connecting the N- and C-domains of the double Rossmann fold is shown in yellow. The double Rossmann fold C-domain is shown in blue; and additional segments in MGDβ and MGDγ in orange. The C-terminal alpha helix that folds in the direction of the N-domain is shown in light blue. Black triangles represent the localization of conserved amino acids involved in the enzymatic reaction. AlphaFold attributes a high accuracy index (90 – 100) for all modeled portions showing homology with AtMGD1 and a low one (50 – 70) for the non-conserved N-terminal ends, as well as for the second half of the C-terminal closing helix for MGDβ and MGDγ.

The *ab initio* models obtained using AlphaFold (Jumper et al., 2021; Jumper and Hassabis, 2022) include N-terminal sequences starting from the putative Sp-Ctp for MGDα and MGDβ and the first start methionine for MGDγ (Figure 3 B-D). The MGDα C-domain has a similar size to that of AtMGD1 (~170 amino acids), whereas it is longer in MGDβ (~223 amino

acids) and MGDγ (~290 amino acids). The hinge connecting N- and C-domains is longer in *P. tricornutum*, possibly providing some flexibility in the structure. Additional segments present in the C-domain of MGDβ and MGDγ, shown in orange in Figure 3, at the level of an unresolved loop in AtMGD1 structure, were predicted with a lower confidence level (50% and 70%, respectively) than the rest of the structure (100%). The last amino acids of the C-terminal stabilizing helix were also modeled with a low degree of confidence for MGDα and MGDγ (50% confidence).

Overall, these analyses show the prediction of the canonical double-Rossmann fold structure defined in AtMGD1 in all *P. tricornutum* MGDs, with conserved domains and residues that are critical for enzymatic activity, and highlight some additional features that may provide specific functionalities. These models have also allowed designing gRNA located in structured and conserved regions, or within the enzymes active sites for the CRISPR/Cas9 editing experiments.

Functional analysis of MGDG synthase activity by heterologous expression in yeast.

Sequences encoding the three mature MGDs were cloned under the control of the galactose-inducible promoter pGAL1 and introduced into *Saccharomyces ceresiviae* for heterologous expression. A transformed yeast expressing AtMGD2 was used as a positive control. The lipid profile of three independent expressing lines were analysed per transgene. The heterologous expression of all *P. tricornutum* MGD isoforms in yeast led to the biosynthesis of MGDG (Supplemental Figure 3), a lipid class that do not occur in wild type yeast. It is important to note that DAG substrates produced in yeast are distinct from those encountered by MGD isoforms in their native environment in *P. tricornutum*. In particular, yeast lack 20:5, which is abundant in diatom MGDG. Therefore, the observed MGDG profiles do not reflect the enzymatic specificity *per se*. Nevertheless, the pattern of MGDG species produced by the different expressing lines highlighted some divergence between isoforms. MGDα-expressing yeast showed the most distinct MGDG species pattern compared to yeast expressing AtMGD2, with higher proportions of MGDG-36:1 and MGDG-36:2 compared to any other line. MGDβ-expressing yeast showed significantly higher proportions of MGDG-32:1 and 32:2, while no traces of MGDG 36:1 and MGDG-36:2 could be detected. Finally, MGDγ was the only isoform able to produce MGDG-28:0 and MGDG-30:0, *i.e.* with saturated acyl species. Although caution is needed when interpreting heterologous expression, results

presented here suggest different substrate specificities in the three MGD isoforms of *P. tricornutum*.

***Phaeodactylum tricornutum* MGDα, MGDβ and MGDγ subcellular localization.**

P. tricornutum strains overexpressing *MGD* genes fused to *eGFP* were generated. For each isoform, two independent overexpressing strains were analysed and gave identical results, namely MGDα-eGFP-A and -B, MGDβ-eGFP-A and -B, and MGDγ-eGFP-A and -B.

We first performed a western blot analysis on total protein extracts from each strain using anti-GFP-HR antibodies to assess the size and expression level of each construct (Supplemental Figure 4). The molecular weight expected for each MGD protein prior cleavage of targeting sequences are 51, 61, and 66 kDa for MGDα, β and γ, respectively. The eGFP protein weighs 27 kDa. Therefore, we expected to detect the chimeric proteins around 78, 89 and 93 kDa for MGDα-eGFP, MGDβ-eGFP and MGDγ-eGFP, respectively. Western blot analysis revealed a band at ~70 and ~80 kDa for MGDα-eGFP and MGDβ-eGFP, respectively, which are in the expected range (Supplemental Figure 4), coherent with the cleavage of a targeting presequence in these proteins. Two bands were detected for MGDγ-eGFP, one at the predicted molecular weight, ~90 kDa, and one at ~100 kDa. The 100 kDa band may correspond to post-translational modifications of MGDγ. These analyses indicate that the transformed lines expressed MGD-eGFP fusions, and could be used for subcellular localization studies as well as phenotypic analyses of MGD overexpression.

Observation of overexpressing lines by confocal microscopy revealed distinct subcellular localizations of MGD isoforms (Figure 4). The chlorophyll auto-fluorescence could be observed in the WT strain since the eGFP signal was within the range of detection using a laser power of 5%. However, eGFP signal is clearly distinguishable from chlorophyll auto fluorescence in the overexpressing strains.

With MGDα-eGFP, we obtained a clear intraplastidial localization signal visible as dots inside the plastid (Figure 4). These dots are reminiscent of the proteo-lipidic domains observed *in vitro* with AtMGD1 bound to biomimetic membranes (Sarkis et al., 2014). They correspond to MGDα localization in thylakoids. This localization is consistent with the predicted bipartite Sp-Ctp, assuming that the detected SAAFSP sequence has the same targeting properties as the ASAFAP motif (Apt et al., 2002).

MGDβ-eGFP fluorescence was detected in the blob-like structure (Figure 4). The vesicular network inside the blob shows continuity with the PPM but not the oEM (Flori et al., 2016). Therefore, MGDβ localization might extend from the periplastidial compartment to the PPM. This localization is consistent with the predicted Sp-Ctp, and suggests that the GSGFVL motif may be a specific alteration of the ASAFAP motif leading to a retention in the blob and PPM (Apt et al., 2002).

MGDγ-eGFP signal appeared clearly in the ER, and to a lower extent at the periphery of the plastid (EpM) and in the cytosol (Figure 4). A presence in soluble form in the cytosol is intriguing since MGDs are usually membrane bound enzymes. Nevertheless, in *A. thaliana* cells, AtMGD2-GFP fusion protein could be partly observed in the oEM of chloroplasts and in the cytosol as well (Awai et al., 2001). AtMGD1 proved to be partially soluble *in vitro* (Rocha et al., 2016). Moreover, AtMGD2 was recently found to localize to the cytosol of elongating pollen tubes (Billey et al., 2021a). MGDγ could therefore have a peripheral localization, bound to the ER and the EpM, and partly mobile in the cytosol. The extra amino acid sequence in the C-domain of MGDγ is particularly rich in hydrophilic residues (Figure 3E), potentially facilitating its solubility in the cytosol. This localization is consistent with the absence of any predicted targeting sequence in MGDγ.

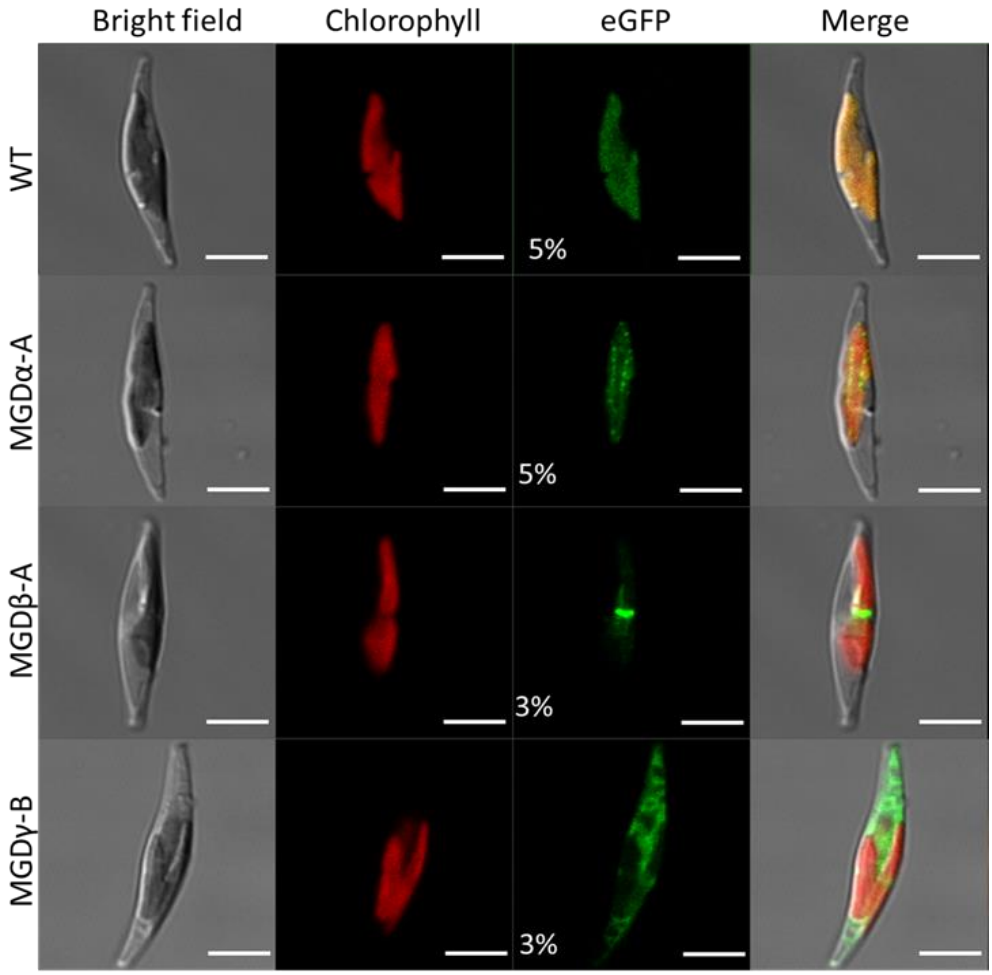

Figure 4: Subcellular localization of MGD isoforms fused to eGFP in *P. tricornutum*. Detection of MGD-eGFP expressed in transgenic *P. tricornutum* strains with wild type as a negative control. For each cell, bright field, chlorophyll fluorescence and eGFP signal are shown, with a merge of all images. eGFP excitation was achieved at either 5 % or 3 % of laser power, as indicated. An intraplastidial localization was obtained when MGDα-eGFP is expressed. With MGDβ-eGFP, the eGFP signal is in the blob-like structure, indicating a localization of MGDβ at the level of the periplasitidal compartment, and possibly the PPM. MGDγ-eGFP is detected mainly in the ER, and at lower level in the EpM, and the cytosol. Scale bar: 5 µm.

Interestingly, the MGDβ-eGFP signal always appeared at a location corresponding to the center of plastids (Figure 4 and Figure 5A). The blob-associated fluorescence was most clearly observed in dividing plastids, which could suggest a role of MGDβ during plastid division. In one instance of a dividing plastid observed by confocal microscopy, the blob was localized at the interface between the two forming plastids (Figure 5A). Then, the blob-like structure appeared as a link between the two generated plastids inside the dividing cell. This observation was supported by scanning transmission electron microscopy (STEM) observation (Figure 5B and C). After separation of the daughter cells, blob-like structures appeared facing each other. When MGDβ-eGFP was observed in a plastid that is not in division, the signal was much more diffuse. It might be either that the blob-like structure gets more exposed, when plastids undertake divisions -and hence the signal is more visible, or that it gets more compacted

at this specific time. Thus, it seems that the blob-like structure may play an important role during cytokinesis, possibly as a glycerolipid synthesis machinery, and that this machinery might be scattered in non-dividing cells.

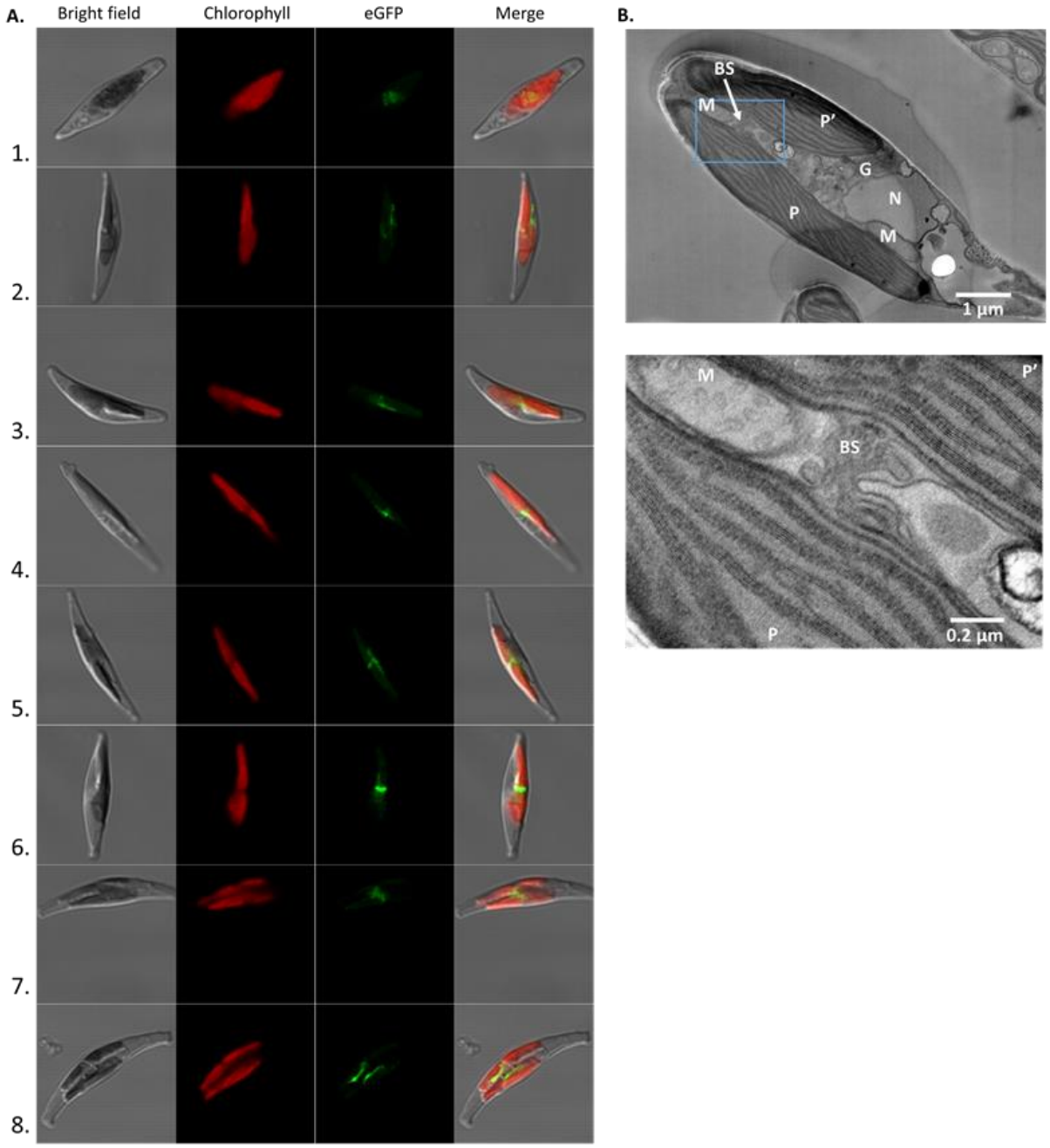

Figure 5: Subcellular localization of the blob-like structure, associated to MGDβ-eGFP fluroescence. Observation of blob-like structure localization by confocal microscopy (A) and transmission electron microscopy (B). *P. tricornutum* transgenic strains overexpressing MGDβ-eGFP gene were used to monitor the appearance of blob-like structures in the cell (A). Cells 1 and 2 show a scattered eGFP signal, cells 3-6 show a blob-like structure at the central constriction of dividing plastids, cell 7 shows a blob-like structure extending from one plastid to the other inside a dividing cell, and cell 8 shows two blob-like structures facing each other on distinct plastids during cytokinesis. Observation of a wild type *P. tricornutum* cell containing two plastids following plastid division inside the cell (B). Cleavage furrows are visible at both end of the cell. The blob-like structure was observed connecting the two plastids. BS, Blob-like structure; G, Golgi apparatus; M, mitochondrion; N, nucleus; P and P', plastids.

Generation of MGDα, MGDβ and MGDγ edited lines.

We generated KO lines by CRISPR-Cas9 editing, using biolistic to transform cells with plasmid-encoded Cas9 and gRNAs. Since random genomic insertions and off-target mutations can lead to phenotypes unrelated to the targeted edition, we selected several independent mutant strains for each gene, obtained from different sgRNAs (distinct potential off-targets), and not originating from the same initial transformed colony (distinct vector insertions in the genome) (Primers and gRNAs are shown in Supplemental Table 3).

For *MGDα*, we selected three pure KO lines with frameshift disruptions *i.e.* one KO obtained with the sgRNA 125i (*mgdαi1*), and two independent KO lines obtained with the sgRNA 125j (*mgdαj1* and *mgdαj3*) (Supplemental Figure 5).

For *MGDβ*, we selected one KO line obtained with the sgRNA 168b (*mgdβb1*), and three mutants obtained with the sgRNA 168c (*mgdβc1*, *mgdβc2* and *mgdβc3*) (Supplemental Figure 5). The three lines *mgdβc1*, *mgdβc2* and *mgdβc3* derive from the same initially transformed cell. They harbor the same vector insertion, but different editions at the targeted locus. In particular, *mgdβc2* mutant has a deletion of 6 nucleotides (DEL6), which does not generate a frameshift in the sequence (Supplemental Figure 5). This *mgdβc2* mutant lacks two amino acids between the predicted signal peptide and the first N-terminal conserved region. This deletion was not expected to impact the enzyme function and thus *mgdβc2* was kept as a control for *mgdβc1* and *mgdβc3* KO lines.

For *MGDγ*, we selected two KO lines obtained with the sgRNA 619h (*mgdγh1* and *mgdγh2*) deriving from the same initially transformed cell, and one obtained with the sgRNA 619i (*mgdγi1*) (Supplemental Figure 5). For this gene, the generation of mutant colonies was laborious, as transformed lines were greatly affected in their ability to form colonies. Only one mosaic colony was obtained after transformations with 619h and 619i sgRNAs. We managed to isolate three pure colonies from the mosaic colony obtained with 619h sgRNA. For the mosaic colony obtained with 619i sgRNA, the mutant profile revealed that there was no trace of WT sequence, and that all indels, although mixed, led to either a frameshift in the sequence or the deletion of crucial amino acids in the active site region (Supplemental Figure 5). We used this third heterogeneous KO line, without further purification. The growth phenotype observed on plates may indicate some important role in stressful environmental conditions. Nevertheless,

no such drastic phenotype could be observed for any of the three *MGDγ* KO lines in liquid cultivation conditions.

Impact of MGD knockout and overexpression on *Phaeodactylum* growth and photosynthesis efficiency in optimal growth conditions.

A first key observation on the *MGD* mutant strains is that none of the *MGD* isoform KO was lethal. The growth of each KO line was monitored in optimal condition, using a rich medium (Supplemental Figure 6A-F). Most of the KO mutants had a slightly slower growth than the WT strain. Overexpression of MGDα and MGDγ but not MGDβ led to a slight increase in growth (Supplemental Figure 6G). The moderate impact of MGDs KO on growth in liquid medium suggests possible mutual compensating effects between isoforms.

As MGDG plays an important role in the stabilization and function of the photosystem, as well as in the xanthophyll cycle (Azadi-Chegeni et al., 2022; Garab et al., 2022), we sought whether MGDs KO had any effect on photosynthesis and induction of photoprotection. We measured the photosynthetic efficiency probed as the effective photochemical quantum yield of photosystem II (Ψ_{II}) and non-photochemical quenching (NPQ) during two different kinds of light stress: a high light stress consisted in exposing dark-adapted cells to high light (700 µmol photons.m^{-2}.s^{-1}) (Supplemental Figure 7A-C), while a moderate light stress consisted in exposing dark-adapted cells to a 2-steps increase in light intensity (55 then 335 µmol.photons m^{-2}.s^{-1}) (Supplemental Figure 7D-F). Under both types of stresses, Ψ_{II} recovery and NPQ relaxation were monitored at low light intensity (20 µmol photons.m^{-2}.s^{-1}) (Ruban et al., 2004; Goss and Jakob, 2010). In both tested conditions, no striking effect on photosynthetic efficiency and non-photochemical quenching could be related to any of the mutations (Supplemental Figure 7), again suggesting mutual compensations between isoforms.

Impact of MGD knockout and overexpression on *Phaeodactylum* cell morphology and plastid architecture.

Cell morphology of MGD KO mutants was examined by epifluorescence microscopy (Supplemental Figure 8). All mutant cells were fusiform with a normally sized plastid compared to the WT. No change in cell size was detected. In MGD overexpressing strains, no changes in cell morphology were recorded either.

Potential impact of MGDs KO on plastid membranes was checked by STEM (Figure 6), with cells from *mgdαj3*, *mgdβc3* and *mgdγh2* strains. No modifications could be observed

in plastids as in any other parts of the cell in any MGD KO line. In MGDβ null mutants, blob-like structures were still present (Figure 6). Likewise, no change in cell ultrastructure could be noticed in overexpressing lines (Supplemental Figure 9).

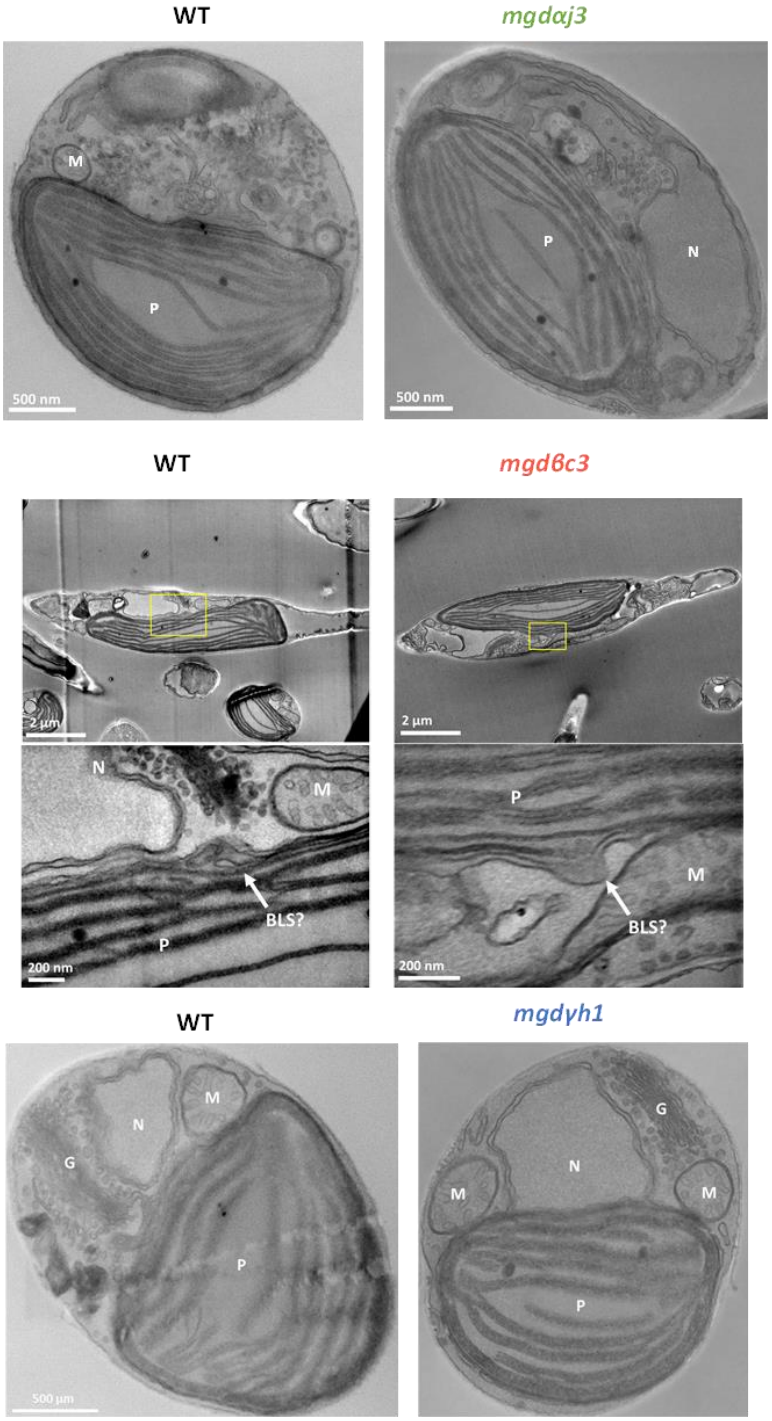

Figure 6: *P. tricornutum* cell ultrastructure in MGD KO lines determined by scanning transmission electron microscopy A WT and KO *mgdαj3* and *mgdγh2* lines were cultured in parallel. In a separate experiment, a WT and a *mgdβc3* mutant were cultured in parallel. Cell ultrastructure is shown in each mutant with corresponding WT control on the left. No impact of *MGD*α, *MGD*β or *MGD*γ KO could be observed at the level of membrane compartments, including plastids. BLS, blob-like structure; G, Golgi, M, mitochondria; N, nucleus, P, plastid.

Impact of MGD knockout and overexpression on *Phaeodactylum* total fatty acid and glycerolipid class profiles.

We cultured WT and MGD KO and overexpressing lines in parallel in a nutrient-replete medium (10N10P). The total amount of fatty acids (FA) in all MGD edited strains was similar to the WT analysed in parallel (Supplemental Figure 10A and 11A) and FA distributions were similar to those already reported in the literature in the same conditions (Abida et al., 2015). As observed previously, *P. tricornutum* is rich in 20:5 (about 25-30 % of total FAs) and C16 molecular species (mainly 16:0, 16:1, 16:3), while poor in C18 FA.

The distribution of the different lipid classes in MGD KO and overexpressing lines was analysed by LC-MS/MS (Supplemental Figure 10B and 11B). The glycerolipid profiles were dominated by MGDG, PC and SQDG, which accounted for about 70 % of total glycerolipid content, as previously reported (Abida et al., 2015).

The glycerolipid profiles of MGD KO lines were moderately changed compared to the WT (Supplemental Figure 10B). Most importantly, the distribution of the different lipid classes indicated that the editions had little impact on galactolipids levels, when compared to the WT cultured in parallel. Concerning the other lipid classes, MGDα KO strain showed slightly lower levels of SQDG (16.0% in the edited strains *vs* 19.6% in the WT) and more PC (17.8% in the edited strains *vs* 14.10% in the WT) (Supplemental Figure 10B). In MGDβ KO strains, the proportion of PC was higher (14.10%) compared to both controls, the WT and the silent edition *mgdβc*2 (12.7%) (Supplemental Figure 10B), consistently with an opposite decrease in PC in the MGDβ overexpressing strains (Supplemental Figure 11B). All MGDγ KO strains had slightly lower proportions of PG (5.1% *vs* 6.3% in the WT) and slightly higher proportions of DAG (1.23% *vs* 1.05% in the WT) (Supplemental Figure 10B). Consistently, DAG tended to decrease in MGDγ overexpressing strains (Supplemental Figure 11B). Besides these minor changes, lipid class profiles in KO and overexpressing lines were close to that of WT.

Impact of MGD knockout and overexpression on *Phaeodactylum* acyl profiles in each glycerolipid class.

Since KO or overexpression of MGD isoforms had only minor impacts on cell division, photosynthesis, morphology, total FAs and glycerolipid class profiles, one may wonder whether the three isoforms might be fully similar and interchangeable. However, expression of the different MGD isoforms in yeast suggested that they have distinct substrate specificities

(Supplemental Figure 3). We thus looked in detail at the variation of each lipid molecular species in each lipid class (Abida et al., 2015; Jouhet et al., 2017).

We based our analysis on the following assumptions, (1) considering that molecular species decreasing in a MGD KO line were determined by the knocked-out enzyme, and (2) verifying whether any consistencies could be detected in the corresponding MGD over-expressors by an opposite trend (shown in Supplemental Figures 12, 13). An additional principle was to assume that (3) an increase of a molecular species in a KO line might be determined by a compensation by one or both un-edited isoforms. Finally, (4) we focused our analysis on the four glycerolipids known to be conserved in photosynthetic membranes, *i.e.* MGDG, DGDG, SQDG and PG (Boudiere et al., 2014) (Figures 7-9), and on other phospho- and betaine lipids, assumed to be related to the extraplastidial membranes and acting as molecular platforms for the synthesis of 20:5.

In MGDα KO lines, all molecular species containing a C16 in position *sn-1* decreased in MGDG (Figure 7), with the notable exceptions of 16:1-16:1, 16:3-16:3, and 16:4-16:3. Consistently, MGDG 16:1-16:2, 16:1-16:3 and 16:2-16:3 species increased in MGDα-eGFP-B overexpressing strain (Supplemental Figure 12). By contrast, the proportion of MGDG molecular species containing a 20:5 FA increased, except the 20:5-16:4. This is coherent with a decrease in MGDG 20:5-16:3 in MGDα-eGFP-B overexpressor. The proportion of all molecular species containing a 14:0 FA also decreased in MGDG in the KO lines but did not change in the overexpressing lines. Concerning DGDG, the change in molecular species profile did not reflect that of MGDG. An increase in species with a 16:1 in position *sn-1* was observed in the KO lines, while species containing a 20:5 decreased, except for 20:5-16:0 and 20:5-16:4 molecular species (Figure 7). In MGDα-eGFP-B overexpressing line, DGDG 16:1-16:0 and 16:1-16:1 decreased, while 20:5-16:2 species increased (Supplemental Figure 12). Concerning SQDG, in the KO lines, the *sn-1* C16-SQDG tended to increase, while species containing a 20:5 FA tended to decrease. SQDG 14:0-16:0 and 14:0-16:1 species also decreased. Eventually, in PG, the proportion of 16:1-16:0 molecular species decreased while that of 16:1-16:1 species increased (Figure 7).

In addition, MGDα KO lines had many non-plastidial glycerolipid molecular species affected as well (Supplemental Figure 14). All 14:0-containing molecular species decreased, as seen in plastid glycerolipids. Consistently, these species increased in the non-plastidial lipids of MGDα-eGFP-B strain (Supplemental Figure 13). All C18-containing species decreased,

with the exception of 18:2-18:1 and 18:2-18:2 in PC, and 20:5-18:4 in PE. All species containing a C16 in positions *sn-1* and *sn-2* increased in PC and DGTA. However, 16:0-16:0 and 16:0-16:1 molecular species decreased in DAG while 16:1-16:1 increased. Species containing a 20:5 in position *sn-1* and a C16 in position *sn-2* increased or tended to increase in PC, DGTA and PE. In MGDα KO, the 20:5-22:6 molecular species increased in PC (Supplemental Figure 14), while it consistently decreased in MGDα-eGFP-B (Supplemental Figure 13).

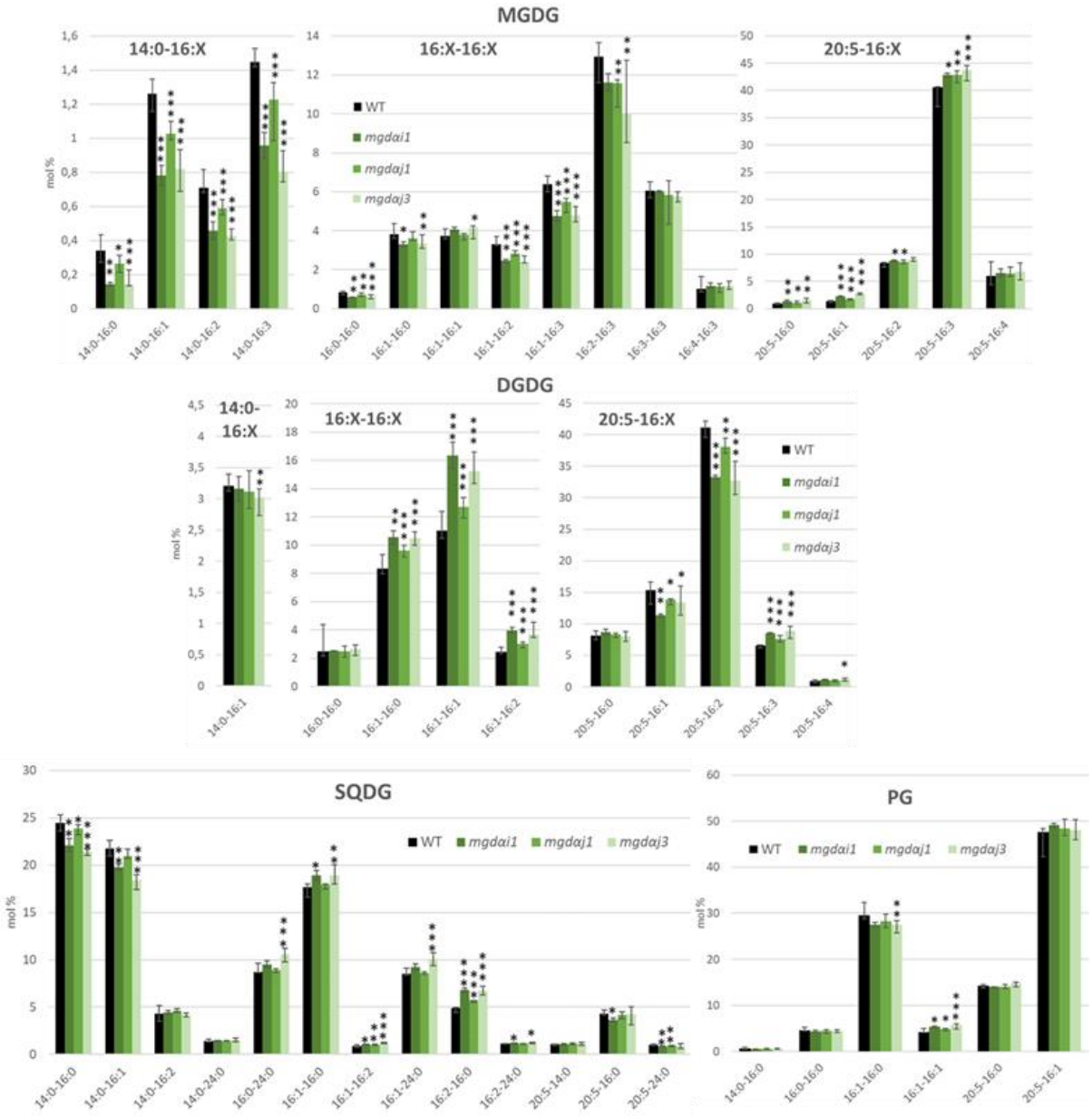

Figure 7: Impact of MGDα mutations on the molecular species constituting the main plastid glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences compared to the WT are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GraphPad Prism software.

For the analysis of the impact of MGDβ KO edition on glycerolipids, we used the silent *mgdβc2* edited strain as a negative control in addition to the WT. MGDG molecular species containing a 20:5 in position *sn-1* decreased (Figure 8), with the exception of 20:5-16:0 and 20:5-16:4. MGDG 16:1-16:0 and MGDG 16:1-16:1 species decreased, while other MGDG species containing a C16 FA in positions *sn-1* and *sn-2* and two or more unsaturations on a C16 increased. The opposite trend was observed in the MGDβ overexpressing strains (Supplemental Figure 12). DGDG molecular species were little affected in the MGDβ mutants as compared to the striking impact on MGDG species (Figure 8). 14:0-16:1 tended to decrease and 16:1-16:1 strongly decreased, consistently with the impact on MGDG, while 16:1-16:2 strongly decreased as well. DGDG species containing 20:5 were almost not affected, with only a slight increase in 20:5-16:0 compared to the controls. Only a few changes were observed in non-plastidial glycerolipids: for the PC, DGTA and PE, 16:1-16:2, 16:1-16:3, and 20:5-16:2 species either increased or tended to increase in the mutants while 20:5-18:1, 20:5-18:2 and 20:5-18:3 species tended to decrease (Supplemental Figure 15). PE 20:5-20:5 molecular species increased in MGDβ KO, while it decreased consistently in the MGDβ overexpressing strains (Supplemental Figure 13).

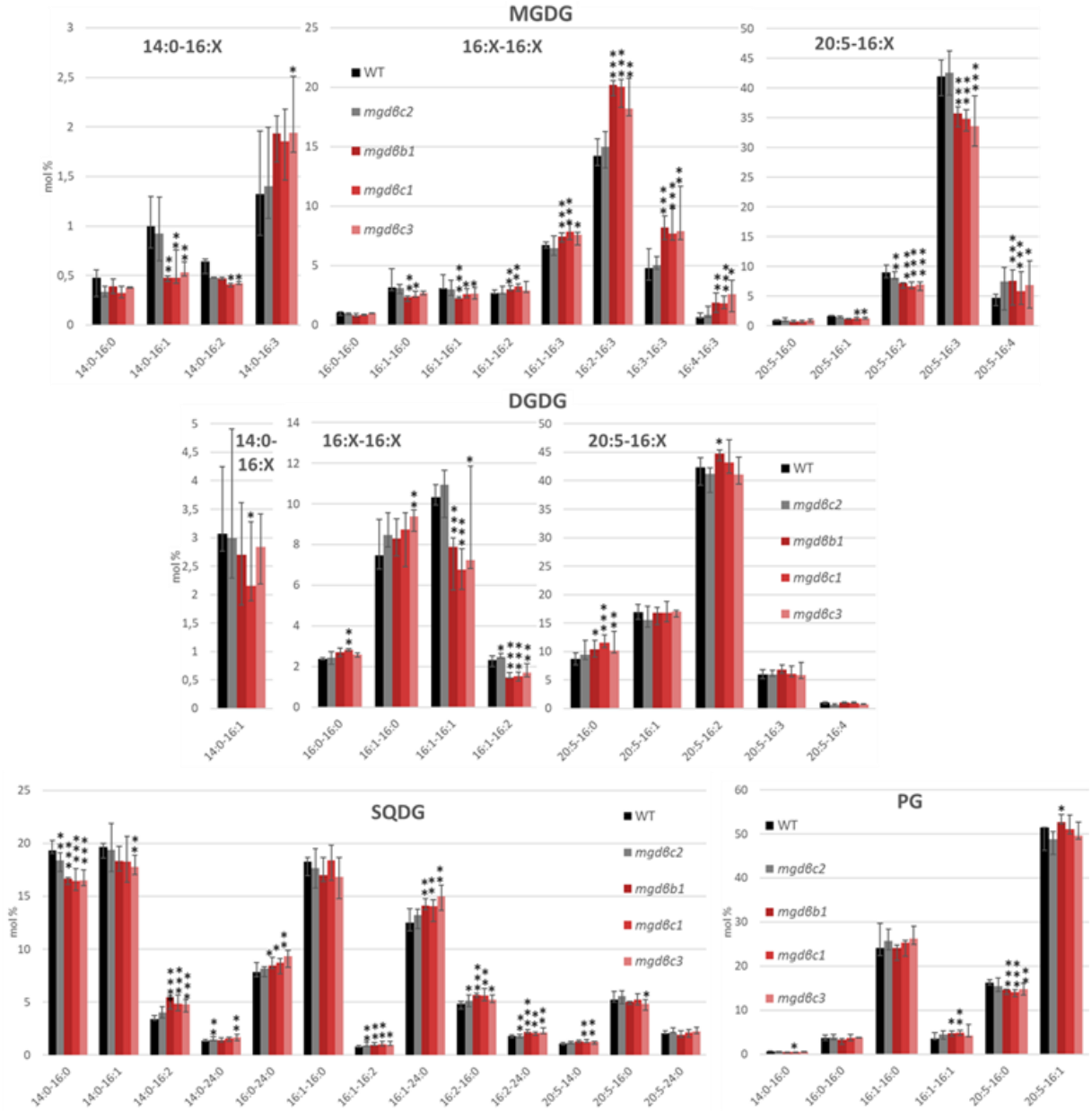

Figure 8: Impact of MGDβ mutations on the molecular species constituting the main plastidial glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. The *mgdβc2* mutant contains a silent mutation and was used as a control. Each result is the median of six biological replicates ± min and max values. Significant differences with the WT are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GraphPad Prism software.

In MGDγ KO lines, levels of 16:0-16:0, 16:1-16:0 and 16:1-16:1 were strongly decreased in galactolipids (Figure 9), while species with the same composition were stable in SQDG and PG. In MGDγ overexpressing lines, these same galactolipid species increased, particularly in MGDG (Supplemental Figure 12). The proportion of all 14:0 FA-containing molecular species in plastidial glycerolipids were decreased compared to the WT. For several molecular species, the *mgdγh1* and *mgdγh2* mutants behaved differently compared to *mgdγi1* mutant. Therefore, we cannot conclude on the impact of MGDγ mutation on the proportion of

16:1-16:2, 16:1-16:3, 16:2-16:3 and 16:3-16:3 species (Figure 9). However, 16:1-16:3, 16:2-16:3 and 16:3-16:3 species clearly decreased in the overexpressors (Supplemental Figure 12). No change was observed in 20:5-containing MGDG species, except a decrease in 20:5-16:4 mirrored by an opposite trend in the overexpressing lines. In DGDG and SQDG, 20:5-containing species tended to increase except for DGDG 20:5-16:4. PG molecular species distribution was globally unaffected by MGDγ mutations (Figure 9).

In non-plastidial membrane glycerolipids of MGDγ mutants, many changes in species proportions were observed (Supplemental Figure 16). 16:0-16:1, 16:1-16:0, 16:1-16:1, 16:1-18:1, and 16:1-18:2 molecular species in PC and DGTA increased, while DGTA 16:1-18:3 and PC 16:1-18:4 decreased. There was a clear decrease in 20:5-20:4 and 20:5-20:5 molecular species of PC, DGTA and PE. Compared to MGDα and MGDβ mutations, 14:0-containing species were not impacted in MGDγ mutants. In DAG, changes were limited to 16:0-16:1 and 20:5-16:0, which tended to increase and decrease, respectively.

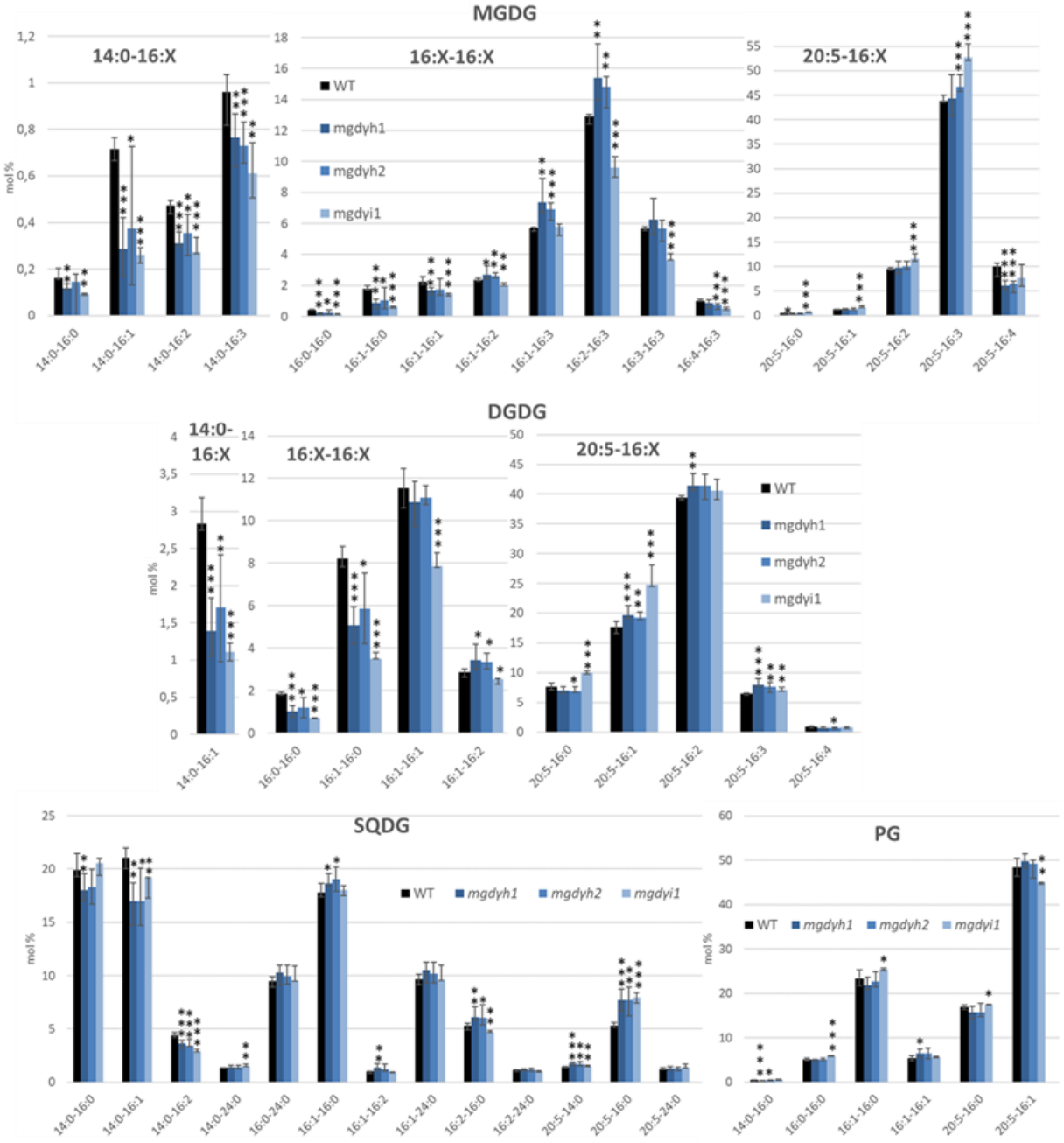

Figure 9: Impact of MGDγ mutations on the molecular species constituting the main plastidial glycerolipids in *P. tricornutum*. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

Impact of each MGD knockout mutation on the expression level of other MGD isoforms

We performed qPCR analyses on MGD edited and WT strains cultured in parallel to check for a possible compensation effect at the transcriptional level among *MGD* genes. A first observation was that in replete conditions, the WT, *MGDβ* and *MGDγ* were expressed at similar levels, and *MGDα* was ~4 times more expressed (Supplemental Figure 17). No activation of

transcript steady state level could be detected in the mutants of MGDα and MGDβ. In MGDγ mutants, an increase of *MGDα* and *MGDβ* gene expression by 1.37 and 1.69-fold on average, respectively, was detected. This indicates that there might be a compensation mechanism at transcriptional level when *MGDγ* is absent. We also observed a decrease in *MGDβ* gene expression in MGDβ KO mutants by 2.25-fold on average (p-value for *mgdβb1* was 0.05028), which was not observed in the silent mutant *mgdβc2* (Supplemental Figure 17).

Taking together all results, MGDα appeared to be the main MGD isoform under optimal condition in the WT. Only MGDγ KO seemed to trigger a moderate compensation effect at the transcriptional level of the two other MGDs. If a compensational effect existed in the absence of MGDα and MGDβ, it might be at the post-translational level.

Impact of *MGDα*, *MGDβ* and *MGDγ* mutations on the response of *P. tricornutum* to a nitrogen starvation

The diversification of the MGD isoforms in diatoms might play a role for environmental stress adaptation. As nitrogen deprivation is a common stress model applied to microalgae, and since a decrease in some MGD isoforms had been detected previously following a removal of nitrogen in the culture medium (Lupette et al., In press), we used this condition to test the effect of each MGD mutation on *P. tricornutum* phenotype. In a whole-cell proteome analysis performed under nitrogen starvation (Lupette et al., In press), the protein levels of MGDα and MGDγ were reduced to 6.1% and 26.3% of control level grown in replete conditions, respectively. Protein levels of MGDβ showed a slight increase to 127.0% of its level in control culture.

Two KO mutants per MGD were used for the analysis of the impact of nitrogen deprivation: *mgdαi1*, *mgdαj3*, *mgdβb1*, *mgdβc1*, *mgdγh2*, and *mgdγi1*. Cells grown in nutrient-replete medium (enriched ESAW 10N10P) were shifted to nitrogen-deprived medium (00N10P) at a starting cell concentration of 3 to 3.5x10^6 cells.mL^{-1}. Growth rate was monitored after 3, 4 and 5 days of nitrogen deprivation (Supplemental Figure 18). Nitrogen deprivation condition led to a slowing down of cell division. Cultures kept growing between day 3 and day 5, with WT cells reaching a concentration of around 9x10^6 cells.mL^{-1} at day 5, which is about half of the concentration that would be expected in 10N10P condition. By comparison with WT, the growth rate was visibly altered in MGDα and MGDβ KO lines, whereas growth was affected in both MGDγ KO lines (Supplemental Figure 18).

The TAG accumulation in response to nitrogen deprivation was first monitored by Nile Red staining at days 3 and 4 following deprivation. We observed that TAG accumulation was significantly faster in MGDβ KO lines compared to the WT (Supplemental Figure 19). Cells stained with Nile Red at day 5 of nitrogen deprivation were observed by epifluorescence microscopy (Supplemental Figure 20). All strains presented a fusiform morphotype and showed large lipid droplets. Plastids appeared shrank compared to cells grown in 10N10P condition (Supplemental Figure 8), consistent with the senescence of plastid membranes previously reported during nitrogen starvation (Abida et al., 2015). Besides a few lipid droplets detected in the medium in the *mgdγi1* KO mutant, possibly reflecting a higher fragility of *mgdγi1* cells under these conditions, no other striking difference in cell, plastid and lipid droplet size could be noticed between WT and KO lines (Supplemental Figure 20).

After 5 days of nitrogen deprivation, cells were harvested for glycerolipid analysis. The response of *P. tricornutum* to nitrogen shortage was consistent with past reports (Abida et al., 2015) with a total FA profile becoming close to the composition of TAG molecular species (Supplemental Figure 21A), due to the increase in TAG, reaching 4.10 nmol.10^{-6} cells after 5 days of nitrogen deprivation, and representing ~60% of total lipid content (Supplemental Figure 21B). Consistently with a very low MGDα protein level in the WT in low nitrogen (Lupette et al., In press), no changes could be detected in the proportions of membrane glycerolipids in MGDα KO lines compared to the WT. For MGDβ KO lines, a slight increase in PE and PG, and a very small decrease in PI were observed.

The changes of membrane glycerolipid distribution in MGDγ KO lines were more striking. There were lower proportions in DGDG (~22% decrease) and DAG (~10% decrease), and an increase in PC (~27 % increase) and DGTA (~57% increase) compared to the WT (Supplemental Figure 21B). Quantitatively, the decrease in DGDG was roughly equivalent to the increase in PC.

Concerning MGDG acyl composition, the impact was barely detected in MGDα KO mutants (Figure 10), in line with an absence of any role of MGDα in nitrogen shortage. By contrast, MGDG molecular species composition was severely affected in MGDβ and MGDγ KO lines (Figure 10) suggesting a role for those two isoforms upon nitrogen stress. In MGDβ KO lines, a strong decrease in 14:0-16:2 and 14:0-16:3 was observed (Figure 10). Although MGDG 14:0-16:1 seemed stable, a decrease in DGDG 14:0-16:1 was observed. MGDG 16:0-16:0, 16:1-16:1, 16:1-16:2, 16:1-16:3 and 16:2-16:3 strongly decreased while 16:3-16:3 and

16:4-16:3 strongly increased. All 20:5-containing MGDG species were decreased, except for 20:5-16:4 MGDG that strongly increased. In DGDG, 16:0-16:0, 16:1-16:1, 16:1-16:2, like in MGDG, and only 20:5-16:2 and 20:5-16:4 species showed a significant increase, while other 20:5-containing species were stable. Changes in galactolipid species distribution were the most spectacular in MGDγ mutants (Figure 10): MGDG and DGDG species containing a 16:0 or a 16:1 in position *sn-2* and either a 14:0 or a C16 in position *sn-1* were severely decreased. On the contrary, species with a 16:2 or 16:3 in position *sn-2* were increased. Most 20:5-containing galactolipid species were mainly increased.

In other plastidial lipids, the impact of MGDα mutation was mild (Figure 10). In MGDβ KO lines, the proportion of 14:0-16:0, 14:0-16:1, 20:5-16:0 and 20:5-16:1 in both SQDG and PG decreased. The proportion of 16:1-16:0, 16:1-16:2, 16:2-16:0 and 16:2-24:0 increased in SQDG, and that of 16:0-16:0, 16:1-16:0 and 16:1-16:1 increased in PG. MGDγ mutation had no effect on SQDG, whereas the proportion of PG species containing a C16 FA in position *sn-1* decreased significantly while 20:5-containing species increased. It seems that PG species were affected similarly as the galactolipid species with the same FA composition, suggesting that they shared some substrates in their biosynthetic pathways.

Under nitrogen shortage, non-plastidial glycerolipid were mostly unaffected by MGDα KO, whereas both MGDβ and MGDγ KO influenced non-plastidial molecular species distributions, with a particularly strong impact of MGDγ KO on a couple of species in DGTA and TAG (Supplemental Figure 22). The MGDγ KO lines were the only ones to show a striking impact on TAG upon nitrogen shortage, verified by opposite effects in overexpressing lines. MGDγ inactivation led in particular to an increase of 14:0-16:1-16:1, 16:1:16-1:16-1 and 20:5-16:1-16:1 TAG species that were consistent with an opposite trend in the overexpressing lines in optimal conditions (Supplemental Figures 13 and 22). Additionally, KO of MGDγ in nitrogen starvation led to a decrease of 14:0-16:0-16:0, 16:0-16:0-16:0, 16:0-16:0-16:1, 16:0-18:1-16:0, 16:1-18:1-16:0, and 16:1-18:1-16:1 TAG species, consistent with an increase of these species in the MGDγ overexpressing lines (Supplemental Figures 11 and 22).

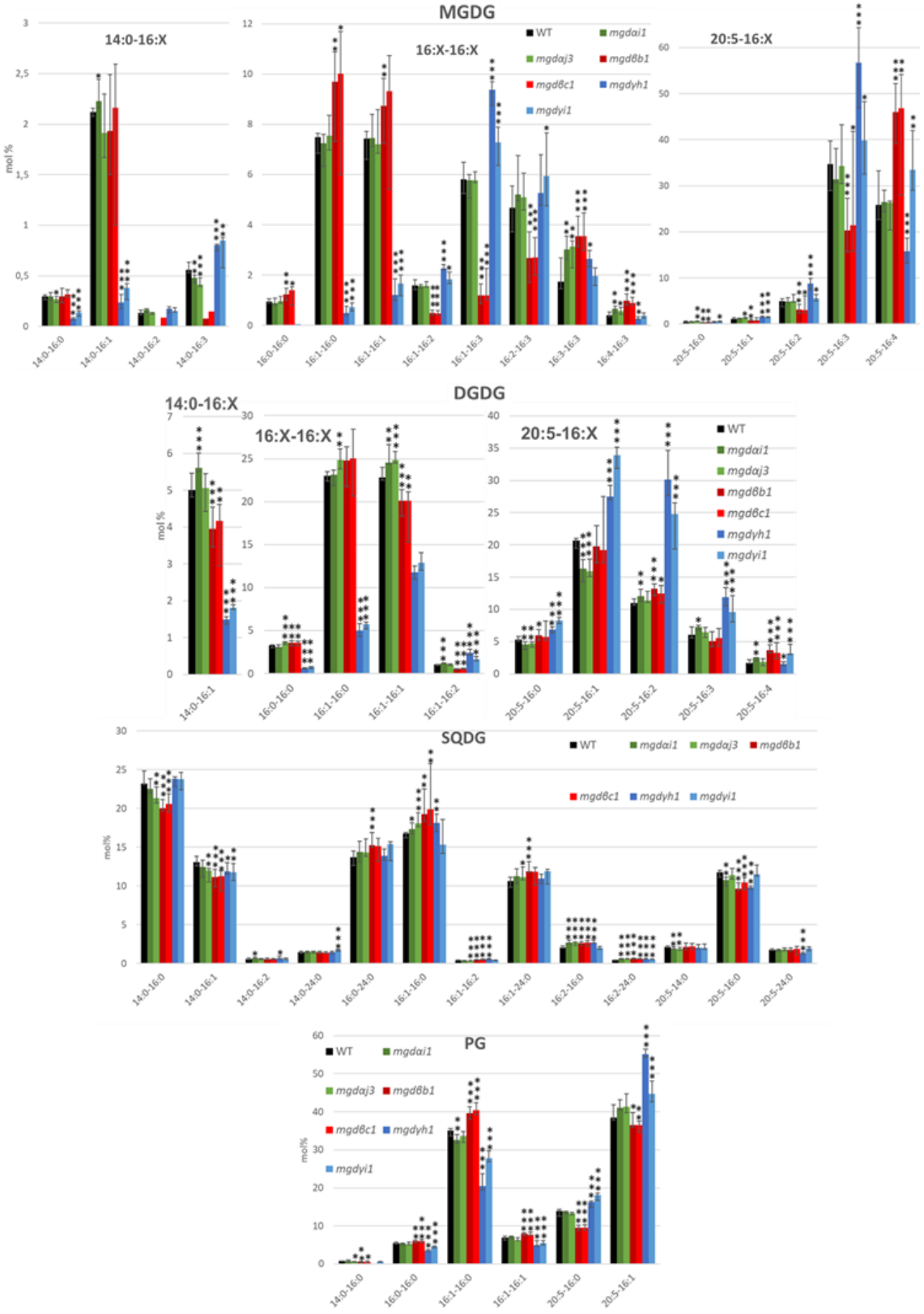

Figure 10: Impact of MGD mutations on the molecular species constituting the main plastidial glycerolipids in nitrogen-deprived *P. tricornutum* cells. Lipid molecular species in MGDG, DGDG, SQDG and PG were analysed by LC-MS/MS. Each result is the median of six biological replicates ± min and max values. Significant differences are shown by one asterisk for 0.05 > p-value > 0.01, two asterisks for 0.01 > p-value > 0.001, and three asterisks for 0.001 > p-value, and were calculated by an unpaired multiple t test using GrahPad Prism software.

Altogether, the impact of MGDα mutation during nitrogen deprivation was very weak, consistently with the very low MGDα protein level in such nutrient stress (Lupette et al., In press). Considering that MGDα is nearly absent in these conditions, the KO of either MGDβ or MGDγ led to a stronger phenotype than in replete medium. MGDβ looked important to keep a high level of polyunsaturated galactolipids in the absence of MGDα. MGDγ seemed to play a specific role in response to nitrogen shortage, in particular on galactolipids with a low unsaturation level. MGDγ KO had a more important effect on non-plastidial lipids than MGDβ KO. The increase in DGTA 16:0-16:1 and 16:1-16:1 in MGDγ KO lines was particularly striking and highlights a role for this isoform outside of the plastid, consistent with its ER/EpM/cytosolic localization. Eventually, modification of TAG species distribution suggests a possible coupling between MGDγ products and TAG synthesis.

Discussion.

The emergence of a multigenic family of MGDs likely occurred before the radiation of diatoms and pelagophyceae from other photosynthetic stramenopiles

Phylogenetic analyses of MGD proteins in photosynthetic eukaryotes (Figure 2) revealed the conservation of three major isoforms of MGDs in diatoms, originating from gene duplications and specialization, similar but distinct to the process that led to the emergence of type A and type B in angiosperms. Based on parsimonious assumptions, an ancestral MGDα type might have been vertically inherited from a red alga during the secondary endosymbiosis event at the base of all stramenopiles. Then, a first gene duplication occurred, tentatively dated before the radiation of diatoms and phaeophyceae from other stramenopiles. This first gene duplication is likely to have led to the emergence of a common ancestor to MGDβ and MGDγ isoforms. A second gene duplication potentially followed the separation of pelagophyceae from diatoms. Gene or genome duplications are crucial drivers in the evolution of stramenopiles (Martens et al., 2008; Parks et al., 2018). In pennate diatoms for instance, up to six whole genome duplications (WGD) have been identified (Parks et al., 2018). It is important to stress that the genomes of some non-diatom stramenopiles contain only one *MGD* gene -such as the eustigmatophytes *Nannochloropsis* or *Microchloropsis* (Vieler et al., 2012; Corteggiani Carpinelli et al., 2014). A single isoform seems therefore sufficient for the biogenesis of a secondary plastid bounded by four membranes. This adds to the puzzle of the emergence of a multigenic family of MGDs and its conservation in diatoms.

***P. tricornutum* MGDα, MGDβ and MGDγ are active MGDG synthases and have distinct substrate specificities.**

After assessing the accurate sequences of MGD proteins encoded by the genome of *P. tricornutum*, structural modelling highlighted a typical double Rossmann fold (Figure 3) characterizing the three-dimensional structure of MGD determined in angiosperms (Rocha et al., 2016). The critical residues for activity are conserved, as well as amino acids involved in a PG-H catalytic dyad, supporting that diatom proteins were functional MGDs. A noticeable distinctive feature lied in the relative size of the N- and C-domains among isoforms. Overall, the hinge separating the N- and C-Rossmann folds was longer in *P. tricornutum* proteins than in plant sequences. Compared to AtMGD1 and MGDα, the C-domain of MGDβ and MGDγ contains additional stretches of 50 and 110 amino acids, respectively. These two regions appeared to be essentially hydrophilic and could influence protein solubility. In addition, the

N-domain of MGDβ and MGDγ contains 82 and 81 additional amino acids, respectively. MGDβ and MGDγ mature proteins are therefore bigger than MGDα; the role of the detected protein extensions is unknown. The enzymatic function was confirmed by (1) heterologous expression of MGD isoforms in yeast and (2) lipid changes in *P. tricornutum* MGD KO lines and overexpressors. These two sets of experiments showed that substrate specificity was distinct for each isoform, with consistent results.

P. tricornutum MGDα, MGDβ and MGDγ localize to distinct subcellular compartments inside and outside of the plastid

MGDα is predicted to contain a bipartite Sp-Ctp presequence suggesting its localization inside the plastid. A putative cleavage site is predicted at the level of a SAAFSP sequence, matching partly with the profile of the 'ASAFAP motif' characterize in nuclear-encoded plastid proteins (Kilian and Kroth, 2005). In *P. tricornutum* expressing MGDα-eGFP fusion proteins, MGDα localization was consistently observed in thylakoid membranes (Figure 4). The pattern observed in the form of spots is quite unusual and could mean that the protein localizes to subdomains in the thylakoids, or that overexpression led to such formation inside the plastid.

In MGDβ, a Sp and a GSGFVL sequence with weak similarity to the ASAFAP motif were predicted. MGDβ-eGFP fusion proteins localize to the blob-like structure (Figures 4 and 5). Altogether, these results indicate that the precursor of MGDβ is targeted to the secondary plastid, successfully crosses the EpM, but is not transported further through the PPM, oEM and iEM. Arrest of MGDβ in the blob-like structure might be determined by specific changes in the ASAFAP motif, corresponding to a different amino acid pattern than previously characterized (Apt et al., 2002; Kilian and Kroth, 2005). This may explain why MGDβ was not detected by bioinformatics methods developed to predict the secondary plastid proteome in *P. tricornutum* (Gruber et al., 2015). Additionally, MGDβ localization could extend to the PPM, as the vesicular network present in the blob-like structure emerges from the PPM (Flori et al., 2016). Since the oEM and iEM correspond to the chloroplast envelope in primary plastids, in rigorous terms, the suborganellar localization of MGDβ represents a unique case of MGD outside a membrane system vertically related to the primary chloroplast. This evolution may have been determined by important mutations of the ASAFAP motif of an MGDα-like ancestral protein, which failed to reach its final destination by crossing the PPM, being retained in the blob-like structure, where it gained a new function.

Finally, no targeting sequence could be predicted in MGDγ, which was consistent with an MGDγ-eGFP fluorescence detected outside of the secondary plastid, in the ER and possibly the EpM and the cytosol (Figure 4). A recent report showing MGDγ-GFP expression in *P. tricornutum* shows a fluorescence within the cells, partly overlapping that of chlorophyll, and in the cytosol (Shang et al., 2022). Authors supposed that such pattern was likely that of an intra plastidial localization, whereas presented images rather show a localization at the periphery of the plastid consistent with results presented here. Localization of MGDγ in the ER and EpM is consistent with its detection in two replicates of the proteome of cytosolic lipid droplets isolated from *P. tricornutum* (Leyland et al., 2020). Overall, we do not exclude that MGDγ is distributed between the ER, the EpM and the cytosol. In *Arabidopsis*, although MGDs are classically considered bound to chloroplast envelope membranes, the initial study of AtMGD2 and AtMGD3 fused to GFP showed some labelling in the cytosol (Awai et al., 2001) and AtMGD2 was shown to localize to the cytosol of elongating pollen tubes (Billey et al., 2021a). Based on this localization, MGDγ represents an evolutionary innovation, performing MGDG synthesis outside of the plastid, for membrane lipid synthesis involving other cell compartments. This is in line with the phenotype we observed at the level of non-plastidial membrane lipids (Supplemental Figures 13-16).

Altogether, *P. tricornutum* MGD proteins localize to distinct subcellular compartments, indicating that these isoforms are likely non-redundant. None of them seem to localize to the iEM or the oEM, as in primary chloroplasts of plants and green algae (Petroutsos et al., 2014). The concentric localization (MGDα / thylakoid; MGDβ / PPM; MGDγ / ER-EpM-Cytosol) suggests possible roles in each of the compartment hosting MGD isoforms, and some kind of sophisticated interactions among glycerolipid pathways occurring inside and outside the secondary plastid. This localization also contradicts the assumption that MGDG, and its derivative DGDG, would be restricted to thylakoid, iEM and oEM membranes. Multiple pools of MGDG are likely to occur in diatom cell, an important feature to take into account in the interpretation of the lipid phenotypes in MGD KO and overexpressing lines.

The blob-like-structure, a lipid-synthesis machinery playing a role during cytokinesis?

Observation of MGDβ-eGFP in the blob-like structure between the PPM and oEM also gave insightful information about the positioning of this structure, particularly during plastid division (Figure 5). The MGDβ-overexpressing system allowed the visualization of the blob with a similar positioning when eGFP protein was fused to a targeting sequence to this

suborganellar compartment (Kilian and Kroth, 2005). Our observations of MGDβ-eGFP fluorescence by confocal microscopy and of the WT by STEM (Figure 5) suggest that the VN might converge and take a blob-like structure mostly during cytokinesis, from the moment the plastid starts dividing until the whole cell division. When the plastid does not show any sign of division, the MGDβ-eGFP fluorescence signal was more diffuse, either suggesting that the vesicular network itself loses the blob structure and become more scattered, or that MGDβ protein relocalizes in the PPM near the blob. During plastid division, the blob-like structure was located at the central constriction site of the dividing plastid. Therefore, MGDβ could be involved during this step for the provision of MGDG to the growing and dividing plastid membranes. Following division, the two generated plastids seem to remain attached by the blob-like structure, meaning that the PPM and EpM are not fully divided, probably until the cleavage furrow comes through the blob-like structure. This is particularly important as it supports that the oNE-EpM continuum is never broken during plastid division, a point that was not addressed yet in studies of diatoms' plastid division (Tanaka et al., 2015a).

Efficient mutual compensation of MGDs in the event of a defective isoform

Two independent MGD-eGFP lines were analysed to evaluate the impact of overexpression of each MGD isoform. Based on western blot analysis of MGD-eGFP expression (Supplemental Figure 3), expression levels of MGDα-eGFP were slightly lower compared to those of MGDβ-eGFP and MGDγ-eGFP, but sufficient to obtain consistent phenotypes. For phenotypic analyses of knocked-out MGDs, a minimum of two CRISPR-Cas9 KO mutants, each obtained with a different sgRNA, were used. None of the MGD isoforms proved strictly necessary for cell survival, as KO mutants could be successfully obtained for each of them. In spite of careful scrutiny, neither cell growth and morphology, nor membrane integrity, nor photosynthetic activity were affected (Supplemental Figures 6 and 7), suggesting either a dispensable role of the targeted isoforms or, more likely, a performant complementation by the action of intact isoforms.

Under optimal culture condition, moderate differences could be observed in the relative proportions of lipid classes of all mutant strains. Again, the level of these changes indicates that in the mutants the intact MGDs were able to compensate for the inactivation of the mutated isoform. No change in transcript level was observed in MGDα and MGDβ KO lines, suggesting that knocking out those two MGD isoforms do not significantly change the gene expression profile. By contrast, in MGDγ mutants, an increase in *MGDα* and *MGDβ* transcripts supports

a possible compensation mechanism at the transcriptional level. Control of MGD protein stability, enzymatic regulation (*i.e.* activation of remaining isoforms *via* metabolic regulatory loops), and an intense relocation of MGDG between membranes are likely required to establish appropriate MGDG levels at the different subcellular compartments.

Specialization of MGDα in thylakoid membrane synthesis

Expression analysis of MGDs in *P. tricornutum* WT strain showed that *MGDα* was the most expressed MGD under our optimal laboratory conditions with relative expressions ~4 times higher than *MGDβ* and *MGDγ*, (Supplemental Figure 17). A relatively higher gene expression level could mean that MGDα is the main isoform under optimal conditions, consistent with its higher similarity to the MGDs from other phylogenetic lineages, responsible for the massive demand of lipids for photosynthetic membranes. However, this is only looking at the gene expression pattern and it may not reflect the protein level in the cell. Consistently with a main role of MGDα for thylakoid expansion, the relative protein level of MGDα was shown to drop to 6 % in a whole-proteome study of nitrogen starved *P. tricornutum* (Lupette et al., In press), coinciding with a decrease of thylakoid membranes observed in this stress condition (Abida et al., 2015). It also reveals that MGDβ and MGDγ might be sufficient to sustain the cell in MGDG during nitrogen starvation. The complementary role of MGDβ and MGDγ for an accurate biogenesis of thylakoids is further demonstrated in MGDα KO lines, showing only a moderate phenotype, with apparently well-structured plastids and healthy photosynthesis.

MGDα uses a DAG pool deriving from plastid 16:0- and 16:1-ACP and produces MGDG with 16 carbon-FA at position *sn*-1, unsaturated up to 16:4 by plastid acyl-desaturases

Through combined observations of MGDα, MGDβ and MGDγ KO effects in optimal and nutrient-depleted conditions, we identified the main DAG substrates used by each isoform and the fate of their products (Figure 11). It is important to note that this reasoning is based on whole cell lipidomic changes, considering both the location of an MGD isoforms and the DAG substrate at its vicinity. Substrate specificity was further confirmed for medium chain fatty acids (C16) by the analysis of yeass after heterologous expression.

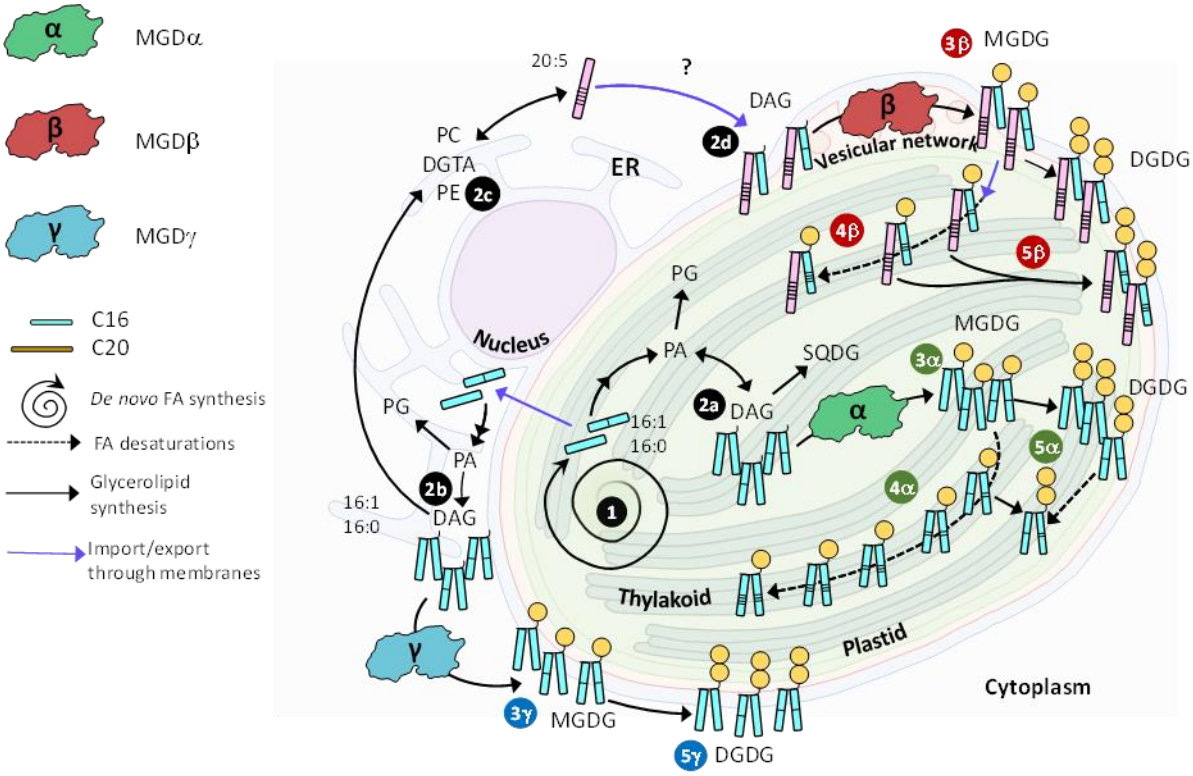

Figure 11. Function of MGDα, MGDβ and MGDγ in *P. tricornutum* cell. The model summarizes all compartmentalization, structural and functional data obtained in this study. FA are synthesized in the stroma as 16:0- and 16:1-ACP (1), where they can be used for the synthesis of DAG (2a). Alternatively, FA are exported to extraplastidial membranes where they can serve for the synthesis of DAG (2b) and phospholipids such as PC and PE, or the betaine lipid DGTA (2c). PC and DGTA serve as platforms to produce 20:5 (2c), used inside the plastid (at an unknown membrane site) for the production of 20:5-containing DAG (2d). MGDα is localized at the level of thylakoids, and functional studies reported here are consistent with a utilization of DAG available inside the plastid (3α), rapidly desaturated (16:1 → 16:4) by plastid desaturases (4α). Based on obtained phenotype of KO and overexpressing lines, MGDG produced by MGDα is likely used by a specific DGD enzyme (5α) to form DGDG. MGDβ is localized at the level of the blob-like structure at the PPM, and functional studies reported here are consistent with a utilization of DAG either produced inside the plastid or with 20:5 acyls imported from the ER/cytosol by a still uncharacterized Ω-pathway (3β). MGDG produced can be desaturated (16:1 → 16:4) by plastid desaturases (4β). Based on obtained phenotype of KO and overexpressing lines, MGDG produced by MGDβ is likely used by a specific DGD enzyme (5β) to form DGDG. MGDγ is localized at the periphery of the plastid, possibly partitioned between the ER, the EpM and/or the cytosol. Functional studies reported here are consistent with a utilization of DAG containing FA freshly exported from the plastid (3γ), producing MGDG at a location far from plastid desaturases. Based on obtained phenotype of KO and overexpressing lines, MGDG produced by MGDγ is also likely associated to a specific DGD enzyme (5γ). In nitrogen shortage, MGDα protein level decreases drastically, and MGDG production relies on MGDβ and MGDγ. When one MGD is knocked-out, compensation mechanisms rely on intact isoforms located in distinct membrane systems, suggesting that inward and outward trafficking of glycerolipids including MGDG must occur through the four membranes limiting the plastid.

The interpretation of MGDα substrates identity and products fate is consistent with MGDα being localized close to *de novo* FA synthesis as well as to the plastid desaturases. Under optimal condition, MGDα mutations most strongly affected MGDG species containing a C16-FA with two or more unsaturations, except those containing a C20:5 in *sn-1* position. Consistently, these species are highly reduced in nitrogen deprivation, when MGDα is no more important. In the total FA distribution in MGDα mutants, a small increase in 16:1 was observed,

while 16:2, 16:3, and 16:4 FAs tended to decrease. Given that in *P. tricornutum* 16:0-ACP is produced by the fatty acid synthase in the stroma of the plastid, desaturated into 16:1-ACP by a plastid palmitoyl-ACP-desaturase (Smith et al., 2021), the pool of DAG used by MGDα are likely generated inside the plastid. In addition, C16-desaturations are performed by desaturases acting mostly on MGDG, which were predicted to localize to the stromal side of the plastid issued from the detection of Sp-Ctp and ASAFAP motifs (Dolch and Marechal, 2015). Based on observed phenotypes, MGDα could mainly use DAG with the following compositions: 14:0-16:0, 14:0-16:1, 16:0-16:0, 16:1-16:0 and 16:1-16:1 (the main DAG species detected by LC-MS), corresponding to a *de novo* production of FA (Figure 11). The heterologous expression of MGDα in yeast confirms this capacity to use DAG with C16 unsaturated substrates, but MGDα appears to be also able to use C18 species in this non-native context (Supplemental Figure 3): the subcellular localization of the enzyme is therefore also determinant in the main molecular species of MGDG it produces. Once the corresponding MGDG products are formed, they do not accumulate, as plastid desaturases rapidly generate MGDG species with higher unsaturation levels, increasing the whole-cell level of 16:2, 16:3, and 16:4 FAs (Figure 11 and Table 1). Although the two MGDα-eGFP constructs were detected with less intensity on immunoblotting than the other MGD-eGFP fusions, the MGDα-eGFP-B line exhibited a strong difference in galactolipids level. This change was consistent with that observed in KO lines, *i.e.* with more MGDG species containing a C16 with two or more unsaturations, along with those containing a 20:5.

Table 1: Contribution of each MGD isoform to the accumulation of different pools of MGDG species. The level of contribution of each MGD to the accumulation of each MGDG species is indicated in three different intensities of grey. Diacyl species corresponding to DAG substrates deriving from FAs synthesized *de novo* in the stroma of the plastid are shown in green. Diacyl species corresponding to DAG substrates containing eicosapentaenoic acid (20:5) imported from the cytosol/ER are shown in light brown. Diacyl species corresponding to MGDG molecular species desaturated inside the plastid are shown in dark green and dark brown. This table summarizes information we deduced from (1) the lipidome analysis of *P. tricornutum* KO and overexpressing lines, and (2) the range of medium chain MGDG FAs species obtained after heterologous expression of MGDα, MGDβ and MGDγ in yeast.

	C14/C16-MGDG			
	14:0-16:0	**14:0-16:1**	**14:0-16:2**	**14:0-16:3**
MGDα				
MGDβ				
MGDγ				

	C16/C16-MGDG							
	16:0-16:0	**16:1-16:0**	**16:1-16:1**	**16:1-16:2**	**16:1-16:3**	**16:2-16:3**	**16:3-16:3**	**16:4-16:3**
MGDα								
MGDβ								
MGDγ								

	C20:5/C16-MGDG				
	20:5-16:0	**20:5-16:1**	**20:5-16:2**	**20:5-16:3**	**20:5-16:4**
MGDα					
MGDβ					
MGDγ					

MGD contribution	
	Low
	Medium
	High

The impact of MGDα mutation on DGDG molecular species, leads to an additional information on the fate of the MGDG pool produced by MGDα. We observed a decrease in DGDG species containing a 20:5 in position *sn-1*, and an increase in DGDG species containing a 16:1 in position *sn-1*, opposite to that seen on MGDG species. It is well established in *A. thaliana* that AtMGD1 and AtDGD1 operate together for the bulk of thylakoid galactolipids, whereas AtMGD2/3 and AtDGD2 operate together when galactolipids contribute to whole-cell membrane lipid remodelling processes (Benning and Ohta, 2005). A reasonable hypothesis is that MGDα might operate in close association with a specific DGDG synthase (MGDα-associated DGD), very active on any 20:5-16:0 or 20:5-16:1 MGDG produced. In the absence of MGDα, the unused DAG pool may be available for the other MGD isoforms, remotely from this MGDα-associated DGD. The MGDG thus produced close to MGDβ and/or MGDγ would fuel the syntheses of distinct DGDG pools (Figure 11).

MGDβ uses a DAG pool enriched in 20:5 imported from the cytosol, and contributes to the production of highly unsaturated MGDG species in nitrogen deprivation

Under optimal conditions, MGDβ mutation strongly decreased the accumulation of 20:5-containing MGDG species, except MGDG 20:5-16:4. Interestingly, MGDG 14:0-16:1,

16:1-16:0 and 16:1-16:1 species also decreased. This firstly suggests that under normal conditions, MGDβ would mostly use DAG 20:5-16:0 and 20:5-16:1, and that the corresponding MGDG products would be quickly desaturated, leading to the accumulation of 20:5-containing MGDG species with a higher unsaturation level of the C16 at position *sn-2* (Figure 11 and Table 1).

MGDβ is localized in the blob-like structure and potentially the PPM, and its direct MGDG products, assumed to be the less desaturated forms, do not seem to accumulate much in the cell considering that unsaturation occurs in the stromal side of the plastid. Therefore, 20:5-containing MGDG produced by MGDβ must be quickly available to the desaturases. This means either that desaturases are present in the oEM and iEM and that the 20:5-containing MGDG species are quickly transported to these two membranes, or that the transport of these species is fast enough to reach thylakoid desaturases without accumulating (Figure 11). In either case, the 20:5-containing species are unambiguously transported to the thylakoid membranes where the bulk of MGDG is requested for photosynthetic membranes. The role of MGDβ in the production of 20:5-containing MGDG species is also consistent with its subcellular localization, since 20:5 FA is produced outside of the plastid following elongation and desaturation steps (Figure 11). For the massive incorporation of 20:5 into the plastid, it needs to cross the four limiting membranes, including the PPM where MGDβ is localized.

The decrease of MGDG 14:0-16:1, 16:1-16:0 and 16:1-16:1 species suggests that these three species are also direct products of MGDβ, but that most of them are not directly available for desaturation. Consistently, the heterologous expression of MGDβ in yeast highlights clearly 32:1 and 32:2 MGDG as major products (Supplemental Figure 3). Therefore, these saturated/monounsaturated MGDG species might derive from a pool of DAG using *de novo* synthesized FAs for the production of MGDG that may accumulate in membranes where desaturases are not present, *e.g.* potentially in the PPM and EpM. Consistently, MGDG 16:1-16:0 and 16:1-16:1 proportions increased in the MGDβ overexpressing lines, while the MGDG 16:2-16:3 and 16:3-16:3 decreased. By this mean, MGDβ may be able to compensate for the loss of MGDγ.

On the one hand, MGDβ appears able to use a DAG pool deriving from the *de novo* synthesis of FA in the stroma, but forming MGDG accumulating in non-photosynthetic membranes. On the other hand, MGDβ uses a DAG pool enriched in 20:5 imported from the cytosol. This isoform in the blob-like structure seems therefore to facilitate bidirectional fluxes

of MGDG molecular species deriving from internal and external DAG pools. Alternatively, the cytosolic origin of the 20:5 could be questioned. Indeed, we could hypothesize that this FA is also produced in close vicinity to the blob, on betaine lipids and phospholipids present in the endomembrane-derived PPM. This would explain how 20:5-containing DAG species, barely detectable in LC-MSMS compared to 16:0-16:0, 16:1-16:0 and 16:1-16:1 DAG species, are the main substrates of MGDβ.

Again, changes in the DGDG profile highlight a possible association of MGDβ isoform with a specific DGD enzyme. Under normal conditions, only DGDG 16:1-16:1 and 16:1-16:2 proportions significantly decreased in MGDβ KO lines, while DGDG 14:0-16:1 tended to decrease. This would be consistent with the role of MGDβ in the production of MGDG 14:0-16:1 and MGDG 16:1-16:1 for the formation of a specific pool of DGDG. The decrease in DGDG 16:1-16:2 is likely related to that of DGDG 16:1-16:1 from which it can be produced through desaturation by the plastid FAD6 desaturase (Dolch and Marechal, 2015). The stable proportion of 20:5-containing DGDG species might be due to compensation effects by MGDα and MGDγ. In line with this hypothesis, 20:5-containing MGDG are still produced, and could be used in priority to maintain 20:5-containing DGDG levels, most importantly by the MGDα-associated DGD pointed above.

Under nitrogen deprivation, MGDβ mutation led to a decrease in 20:5-containing MGDG species, except for the 20:5-16:4 molecular species. This is consistent with the role of MGDβ in the production of 20:5-containing MGDG species, and the priority given to the production of 20:5-containing DGDG species. We also observed a decrease of MGDG 14:0-16:2, 14:0-16:3, 16:1-16:2, 16:1-16:3, and 16:2-16:3. These species are produced by desaturation after a large proportion of MGDG produced by MGDβ is transported toward the stromal side of the plastid, where desaturases operate. In nitrogen deprivation, the DGDG species in the MGDβ mutants is again decreased in 14:0-16:1, 16:1-16:1, and 16:1-16:2, similar to the phenotype observed in replete conditions. This specific context means that MGDγ could largely compensate the lack of MGDG substrates for DGDG species following a defect of both MGDα- and MGDβ-dependent mechanisms.

MGDγ uses a DAG pool enriched in 16:0 and 16:1 exported to the EpM/ER, for the production of MGDG species with low unsaturation level in non-platidial membranes.

Under optimal conditions, MGDγ mutation led to a strong decrease in all 14:0-containing MGDG species and in MGDG 16:0-16:0, 16:1-16:0, 16:1-16:1 and 20:5-16:4. Therefore, the main DAG substrates for MGDγ under normal conditions are likely those directly deriving from *de novo* FA synthesis in the stroma, *i.e.* 14:0-16:0, 14:0-16:1, 16:0-16:0, 16:0-16:1 and 16:1-16:1, requiring that these DAG are exported at least to the EpM (Figure 11). Heterologous expression in yeast is consistent with this substrate preference, as MGDγ led to the synthesis of 32:0, 32:1 and 32:2-MGDG (Supplemental Figure 3). As for MGDα, in yeast MGDγ can also accommodate DAG substrates with C18-FAs. Given that MGDγ is located in the ER and at the outermost membrane of the plastid, its MGDG products are not directly available for plastid desaturases (Table 1). Coherently, MGDγ mutation did not have any strong impact on MGDG species containing a C16 with two or more unsaturations, except for 14:0-16:2 and 14:0-16:3 species. The strong impact on MGDG 14:0-16:0, 14:0-16:1, 16:0-16:0, 16:1-16:0, 16:1-16:1 species indicates that MGDγ is responsible for their production and accumulation (Figure 11 and Table 1).

Under nitrogen deprivation condition, MGDγ also led to a very strong decrease of the same molecular species. Coherently, MGDG 14:0-16:1, 16:0-16:0, 16:1-16:0 and 16:1-16:1 species were increased in MGDγ overexpressing lines, while MGDG 16:1-16:3, 16:2-16:3 and 16:3-16:3 were decreased. The products of MGDγ are therefore able to accumulate in membranes where plastid desaturases are not found.

The nature of the membranes where these MGDG species accumulate are intriguing, because this would be the first time that such synthesis in a non-plastidial membrane system would be observed in a photosynthetic eukaryote. The MGDγ subcellular localization suggests that MGDG might be synthesized in the EpM and the ER. By extension, this pool of MGDG could virtually be transferred to any other compartment of the endomembrane system, directly or indirectly connected to the ER via membrane continuity (like the outer nuclear envelope) or via vesicular trafficking. We have no element to comprehend a role of MGDG in extraplastidial membranes, and this represents therefore a critical question to address in diatoms. The role of MGDG in membranes, a lipid class usually characterized by the high unsaturation level of its esterified FAs, is often associated to its capacity to form Hexagonal II phase. However, it is uncertain whether the highly saturated products of MGDγ would form such structures or not. Therefore, both the localization and the distinct nature of MGDγ products should be kept in mind for further studies.

Consistently with its impact on MGDG species, MGDγ led to a decrease of DGDG 14:0-16:1, 16:0-16:0, 16:1-16:0, and 16:1-16:1, and to a proportional increase in 20:5-containing DGDG species. These decreases were moderate in optimal conditions, and very strong in nitrogen deprivation condition. Therefore, we can speculate that a particular DGD enzyme might be highly dependent on MGDγ for the production of DGDG species with a 14:0 or a C16 in position *sn-1*. This hypothetical DGD enzyme should localize at or close to the EpM. Nevertheless, it must be noted that in the specific condition consisting of 1) nitrogen deprivation, 2) the absence of MGDβ following its KO and 3) a low abondance MGDα as observed in proteomic data, MGDγ was able to produce 20:5-containing MGDG species. It must be noted however that these species did not reach the same proportions as in the WT and led to the additional production of 20:5-containing DGDG species. Therefore, MGDγ appears to have some kind of flexibility in its roles, able to also use DAG 20:5-16:0 and 20:5-16:1 for the production of MGDG species (Figure 11). Nevertheless, it seems that MGDγ support to 20:5-containing MGDG is very low when other MGDs are active.

MGD isoforms and extraplastidial lipid homeostasis

Mutation and overexpression of the three MGD isoforms led to changes in non-plastidial lipids, showing a high level of integration between plastidial and non-plastidial membrane glycerolipid syntheses. The observed phenotypes support the existence of a flux of 16:0/16:1-rich DAG going outward from the stromal side of the plastid to the EpM and ER, where this pool is available for extraplastidial lipids formation by MGDγ. They also highlighted that 20:5-DAG species could be imported inward or produced at the PPM, forming a DAG pool available for MGDβ, and further transported to the stroma, where it joined the 16:0/16:1-rich DAG pool available for plastidial lipid syntheses by MGDα. The compiled results presented here show that the impact of MGDα and MGDβ on extraplastidial lipids rather reflect the establishment of a novel equilibrium compensating alterations in DAG pools. By contrast, alterations of MGDγ reveal more spectacular changes, supporting a specific role of this isoform in whole-cell lipid remodeling occurring in response to environmental stress conditions.

MGDγ mutation had a stronger impact on non-plastidial lipid composition. In particular, 16:1 and 16:0 containing PC and DGTA increased both under optimal and nitrogen deprived conditions. MGDγ being located on the outside of the plastid, it is possible that it sits near to the PC and DGTA synthesis sites. Therefore, it is coherent that the DAG species not used by MGDγ to produce MGDG are used as substrate for the synthesis of PC and DGTA (Figure 11).

Following MGDγ KO, 16:0 amounts available for elongation into 18:0 by Δ0-elongases (Dolch et al., 2017a), the committing step for further desaturations up to 18:4 and 20:5 in phospholipids and DGTA increased. Consistently, the amounts of PC, PE and DGTA containing an 18:1 or an 18:2 increased. It seems however that the rates of 18:2-to-18:3 and 18:3-to-18:4 desaturations and 18:4-to-20:4 elongation were slower than the 16:0 export rate.

In *P. tricornutum* and other photosynthetic stramenopiles, an excess of 16:0 or 16:1 in the cytosol is known to be 'absorbed' by an extra-production of the storage lipid TAG (Dolch et al., 2017a; Smith et al., 2021). In MGDγ mutants grown under optimal growth conditions, MGDγ KO led to spectacular differences in the profile of several TAG species, with a decrease in 16:0-16:0-16:0 and 16:0-16:0-16:1, and an increase in 16:1-16:1-16:1. This impact on TAG was enhanced when cells were subjected to a nitrogen starvation. This coupling between MGDγ activity and TAG production highlights a contribution of this MGD isoform to the intense membrane and storage glycerolipid remodeling occurring in response to a variety of environmental factors. The control of a MGDG/TAG balance is known to occur in other algal systems, such as *Chlamydomonas reindhardtii* containing only one *MGD* gene (Li et al., 2012; Gu et al., 2021; Iwai et al., 2021), however diatoms seem to exhibit an isoform of MGD in their extraplastidial membranes specialized in controlling part of this equilibrium.

Comparison of the emergence of specialized MGD isoforms in diatoms and in angiosperms.

The emergence of specialized MGD isoforms in diatoms and angiosperm is a fascinating example of evolutionary convergence, as important similarities can be drawn. In both cases, MGD isoforms derived from an ancestral form dedicated to the production of the bulk of MGDG for photosynthetic membranes, following gene duplications and speciation occurring before the main radiation of the phylum. Whereas in angiosperms, type B MGD could gain function distinct from type A bound to the iEM, following a relocation to the oEM after a loss of the Ctp, MGDβ/γ could gain additional functions following a similar topological relocation. A plausible mechanism lies in the evolution of the original Sp-Ctp bipartite sequence characterizing MGDα, with putative mutations of the ASAFAP motif arresting the import of the protein precursor into the PPM (MGDβ), followed by a loss of the targeting sequence, leading to a localization in the EpM/ER (MGDγ). In both angiosperms and diatoms, the topological separation of MGD isoforms correlated with the use of DAG pools with distinct acyl-profiles, the production of MGDG with distinct acyl-profiles, and the specific association

with DGD enzymes, also producing DGDG with distinct acyl-profiles. In both cases, MGDG dedicated to photosynthetic membrane highlighted the highest level of unsaturations (16:3 and 18:3 in angiosperms, up to 16:4 and 20:5 in diatoms).

In both cases, 'novel' MGD isoforms (type B MGDs in angiosperm and MGDβ/γ in diatoms) strengthened the ties with extraplastidial membranes and implied intense trafficking of lipid precursors. On the one hand, they allow the incorporation of DAG molecular species deriving from extraplastidial precursors in both systems, via the so-called eukaryotic pathway in angiosperms (Browse et al., 1986), and the still unresolved Ω-pathway in diatoms (Petroutsos et al., 2014). The nature of the imported precursor may differ, as well as the mechanism involved. On the other hand, without sparing MGDG required for photosynthesis, novel MGD isoforms could be involved in the specific production of galactolipids, for novel function, like lipid remodeling occurring under environmental stress. An example is the role of AtMGD2/3 in *Arabidopsis* upon phosphate shortage (Kobayashi et al., 2009b) and that of MGDγ and to a lesser extent, MGDβ, suggested here under a nitrogen shortage. In both cases, the emergence of specialized MGD isoforms correlated with a remarkable ecological success of each phylum populating a variety of terrestrial and marine habitats, marked by constant environmental changes and stress conditions.

Among the differences, the roles played by type A and type B do not overlap in angiosperms: a KO of type A is lethal and cannot be compensated by type B MGDs (Kobayashi et al., 2007a, b). By contrast, the multigenic family of MGDs in diatoms showed strong overlapping activities, preventing as much as possible the loss of integrity of photosynthetic membranes. Any defect of one isoform is nearly completely compensated by the two others. Even under nitrogen shortage, when MGDα level is drastically reduced, MGDγ compensates for the loss of MGDβ and *vice versa*. This does not make the dissection of respective roles easy, but this highlights the powerful flexibility of membrane glycerolipid metabolism in diatoms, despite their very sophisticated membrane compartmentation.

Eventually, diatoms introduced an important milestone in the evolution of MGD proteins, unreported to date in any other phylum, marked by the targeting of a MGDG synthesis system in non-plastidial membranes, *i.e.* in the PPM (MGDβ) supposed to derive from the red alga endosymbiont plasma membrane, and the EpM and ER (MGDγ). MGDG has been assumed to be the hallmark of chloroplastic membranes, and in textbooks it is referred to as such. We do not have yet the conceptual framework to address MGDG role outside of the

plastid, but this work provided genetic tools to address this fascinating question. It is important to note, that whereas the biogenesis of the iEM and oEM could be supposed to follow that of the primary chloroplast envelope in angiosperms and green algae, and whereas the biogenesis of the EpM could be considered to derive from the ER, no lipid-synthesis protein could be located to the PPM yet. This work provided therefore the first clues on the biogenesis of all membrane systems of a secondary plastid, with biological processes unsuspected until now, paving the way for future works.

Methods.

Primary sequence analyses.

P. tricornutum MGD genes were identified based on protein sequence homology using BLAST, with *Arabidopsis thaliana* main isoform, *i.e.* AtMGD1 (NP_194906), as query. The identified locus tags of *P. tricornutum* were named *MGDα* (Phatrdraft_14125, Uniprot; XM_002181649.1, RefSeq; J14125, EnsemblProtists), *MGDβ* (Phatr_43938; XM_002186319; J54168) and *MGDγ* (Phatrdraft_9619; XM_002176764.1; J9619) respectively. Target sequence predictions were performed considering multiple START/Methionine codons present in the N-terminal part of putative reading frames. Signal peptides (Sp) were predicted with the SignalP-6.0 tool, with default parameters (Emanuelsson et al., 2007; Almagro Armenteros et al., 2019) Chloroplast transit peptide (Ctp) were predicted using WolF PSORT (Horton et al., 2007) and ChloroP - 1.1 (Emanuelsson et al., 2007), with default parameters. Specific bipartite pre-sequence were predicted using the HECTAR software specifically designed to detect Sp-Ctp in stramenopiles (Gschloessl et al., 2008).

Protein structure predictions.

The software Phyre^2 (Protein Homology/analogY Recognition Engine V 2.0) (Kelley et al., 2015), and AlphaFold (Jumper et al., 2021; Jumper and Hassabis, 2022) run from ChimeraX (Mirdita et al., 2022) were used to model *P. tricornutum* MGD sequences, which were then viewed on PyMol (Janson et al., 2017). AtMGD1 structure (4X1T- wwPDB consortium, 2019) (Rocha et al., 2016) was used to build Phyre^2 PtMGDs models. For each *P. tricornutum* MGD, the peptide sequences used as input were those deduced after cDNA sequencing and signal peptide prediction, as represented in Figure 3E. These correspond to the following sequences: for *MGDα*, the sequence retrieved from Phatr3_J14125.t1 starting from the second methionine (MCKL to TRTS, 460 amino acid); for *MGDβ*, the whole sequence from Phatr3_J54168.t1 (MVWS to LNNK, 559 amino acid); for *MGDγ*, the whole sequence from Phatr3_J9619.t1 (MATG to EPSR, 618 amino acids).

MGD phylogenetic analysis.

All MGD protein sequences used for phylogenetic analysis were retrieved from the National Centre for Biotechnology Information (NCBI) (Supplemental Table 1) after a protein BLAST search using MGDα (XM_002181649.1), MGDβ (XM_002186319) and MGDγ (XM_002176764.1) as queries. Sequences from different phyla covering the biodiversity of

plastid-containing eukaryotes, including Archaeplastida and Stramenopiles were retained, with at least two species per phylum. The dataset was manually curated to validate the presence of full-length sequences with conserved MGD domains including residues at the active site. Alignment was performed with a customised pipeline in NGphylogeny.fr (Lemoine et al., 2019) using the MUSCLE v3.8.1551 software (Edgar, 2004). The ambiguously aligned regions were curated using the Block Mapping and Gathering with Entropy (BMGE v1.12_1) software (Criscuolo and Gribaldo, 2010) implemented in NGphylogeny.fr using default settings. Preliminary phylogenetic trees were inferred using FastME with 2,500 bootstrap pseudoreplicates. MEGA X v10.0.5 software (Kumar et al., 2018) was fed with the aligned and curated data set. The best evolutionary model was evaluated, and a Maximum Likelihood phylogenetic analysis was performed. To define the best evolutionary model, MEGA X was used to compare the 56 models implemented. The LG+G model (Le et al., 2008) was chosen (lowest Bayesian Information Criterion (BIC) score). The phylogeny was inferred by Maximum Likelihood with 5,000 bootstrap pseudoreplicates. The tree with the highest log likelihood (-12414.63) was retained. Initial trees for the heuristic search were obtained by applying the BioNJ method (Gascuel, 1997) to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (five categories (+G, parameter = 0.7081)). The best phylogenetic tree was then retrieved in Newick format and visualised by iToL (v5) (Letunic and Bork, 2021).

Cultivation of *Phaeodactylum tricornutum* cells

Phaeodactylum tricornutum (Pt1) Strain 8.6 CCAP 1055/1 (CCMP2561- Culture Collection of Marine Phytoplankton, now known as NCMA: National Center for Marine Algae and Microbiota) was used to generate overexpressing and KO out lines. Cells were maintained and grown in a modified ESAW (Enriched Seawater, Artificial Water) medium (NaCl 362.7 mM; Na_2SO_4 25 mM; KCl 8.03 mM; $NaHCO_3$ 2.067 mM; KBr 0.725 mM; H_3BO_3 0.372 mM; NaF 0.0657 mM; $MgCl_2$ 47.18 mM; $CaCl_2$ 9.134 mM; $SrCl_2$ 0.082 mM; Na_2-glycerophosphate 21.8 µM; Na_2SiO_3 105.6 µM; disodium ethylenediaminetetraacetate dehydrate (Na_2EDTA) 14.86 µM; $Fe(NH_4)_2(SO_4)_2$ 5.97 µM; $FeCl_3$ 0.592 µM; $MnSO_4$ 2.42 µM; $ZnSO_4$ 0.254 µM; $CoSO_4$ 0.0569 µM; Na_2MoO_4 0.52 µM; H_3BO_3 61.46 µM; Na_2SeO_3 10 nM; biotin (vitamin H) 8.18 nM; cobalamin (vitamin B_{12}) 2.94 nM; thiamine (vitamin B_1) 0.594 µM) (Falciatore et al., 2000) using ten times enriched nitrogen and phosphate sources (10N10P containing 5.49 mM $NaNO_3$ and 0.224 mM NaH_3PO_4) (Abida et al., 2015). Culture in liquid media was performed at 20°C, on a 12:12 light (40 µmol photons.m^{-2}.sec^{-1}) dark cycle, under gentle agitation, in an

INFORS HT Multitron Pro incubator. Cell concentration was kept between $0.5x10^6$ and $10x10^6$ cell.mL^{-1}. For nitrogen starvation, *P. tricornutum* cells were cultivated in triplicates in 50 mL of ESAW 10N10P in 250-mL flasks until a concentration of $3\text{-}4x10^6$ cell.mL^{-1}. After centrifugation at 1,500 g for 10 min and removal of supernatant, cells were washed with 50 mL of ESAW 0N10P medium, and transferred into to 250-mL flaks containing 50 mL of ESAW 0N10P, at a concentration of $2.5\text{-}3x10^6$ cell·mL^{-1}. Cultures on solid medium were performed with ESAW 10N10P complemented with Agar (1% m/v), at 20°C under continuous light (50 µmol photons.m^{-2}.s^{-1}), in a MLR-352-PE Climate Chamber. To guarantee axenic conditions, all media were supplemented with carbenicillin disodium salt (0.237 µM). For culture of mutant lines transformed with CRISPR-Cas9 vectors and of overexpressing lines, zeocin (Invitrogen; 0.07 µM) was added to the media. Growth was evaluated by cell counting using a TECAN infinite M1000Pro plate reader and determined following the equation $y = 0.01834\ x + 0.03758$ with y the absorbance at 730 nm and x the number of cells in million per milliliter (Conte et al., 2018).

DNA and RNA extractions.

For extraction of genomic DNA from *P. tricornutum*, a cell pellet corresponding to $100x10^6$ cells was resuspended in 500 µL of lysis buffer (250 mM Tris pH 8.2, 100 mM EDTA, 2% SDS, 350 mM NaCl) and transferred to a 1.5-mL tube. The sample was incubated 15 s at 60°C under agitation (450 rpm) using a thermobloc. 500 µL of cold phenol:chloroform:isoamyl alcohol (25:24:1) was added to the sample and homogenized gently. The sample was centrifuged at 13,000 g at 4 °C for 10 minutes. 300 µL of the upper aqueous phase was collected. A second extraction was performed by addition of one volume of chloroform:isoamyl alcohol (24:1), and 200 µL of the upper aqueous phase was collected. For DNA precipitation, 30 µL of sodium acetate 3 M pH 5 and three volumes of ethanol were added. After a 20-min incubation at -20°C and a 10-min centrifugation at 13,000 g at 4°C the precipitated DNA was washed with 700 µL of ethanol 70%, centrifuged at 13,000 g at 4°C for 5 min and air dried at room temperature. DNA was resuspended in DNAse-free water. Concentration and purity were evaluated using a NanoDrop2000 (ThermoFisher Scientific). For RNA extraction from *P. tricornutum* cells, 1.5 mL of TRI Reagent (Sigma-Aldrich) waere added to a frozen cell pellet corresponding to $150x10^6$ cells. The sample was first vigorously mixed and incubated 5 minutes at 60°C, a process repeated twice for a thorough cell lysis. 300 µL of chloroform were added to the sample then vigorously mixed by tube inversion. The sample was incubated at room temperature for 15 minutes before centrifugation at 11,000 g at 4°C for 15 minutes for phase

separation. About 600 µL of the upper aqueous phase was transferred to a Phasemaker tube (ThermoFisher Scientific). A volume of 1-bromo-3-chloropropane corresponding to one fifth of the transferred aqueous phase was added and mixed vigorously. The sample was incubated 3 minutes at room temperature before centrifugation at 16,000 g at 4°C for 5 minutes. After centrifugation, 1 mL of the aqueous phase was collected. RNA was precipitated by addition of one volume of isopropanol and a 30-min incubation at room temperature. RNA was pelleted by centrifugation at 11,000 g at 4°C for 10 minutes and further washed in 1 mL of ethanol 75%, followed by a centrifugation at 7,500 g at 4°C for 10 minutes. The pellet was left to air dry for 30 minutes at room temperature. RNA was resuspended in 35 µL of RNA-free water. A first DNAse treatment was applied using the Invitrogen Ambion TURBO DNA-free kit following manufacturer's instructions. Cleaning and purification of the sample was then achieved using the RNeasy MinElute Cleanup kit (Qiagen) following manufacturer's instructions. RNA concentration and quality was assessed using a NanoDrop2000 (ThermoFisher Scientific) and by electrophoresis. RNA was stored at - 80°C until use.

***MGD* gene expression analysis by quantitative polymerase chain reaction (qPCR).**

This method is detailed in supplemental material.

Heterologous expression of MGD isoforms in yeast

This method is detailed in supplemental material.

Construction of MGD overexpressing lines.

Full-length coding sequence of *MGDα*, *MGDβ* and *MGDγ* genes were amplified by PCR using cDNA derived from *P. tricornutum* as a template and primers detailed in Supplemental Table 4. The PCR products were cloned into pEASY-T1 simple (TransGen, Beijing) vector for DNA sequencing. Sequence-confirmed *MGDα*, *MGDβ* and *MGDγ* genes were excised from T-cloning vector with BamHI and XhoI in the case of *MGDα*, BamHI and SalI for *MGDβ*, and NheI and SalI for *MGDγ*. Each gene was inserted upstream an in-frame with the sequence coding for eGFP used as reporter gene in the multi cloning site of pPha-CG vector (GenBank AF219942) (Zaslavskaia et al., 2000). The pPha-CG vectors harboring *MGDα-eGFP*, *MGDβ-eGFP* and *MGDγ-eGFP* were linearized with ScaI and introduced in *P. tricornutum* cells by electroporation as previously described (Zhang and Hu, 2014).

Construction of CRISPR-Cas9 knocked-out lines.

Single guide RNAs (sgRNA) were designed using the PhytoCrispex online tool (Rastogi et al., 2016), choosing NGG as Protospacer Adjacent Motif (PAM) sequence, and allowing CRISPR targets to start with any nucleotide. sgRNA were selected based on their proximity to START codons and active site residues. sgRNA forward and reverse sequences preceded with the short nucleotide sequences TCGA and AAAC respectively are shown in Supplemental Table 3. The sgRNA sequences were inserted in the pKSdiaCas9_sgRNA-zeo vector, kindly provided by Cécile Giustini (LPCV, Grenoble), derived from pKS diaCas9_sgRNA vector (Addgene). *P. tricornutum* cells were transformed with corresponding vectors by particle bombardment as described previously (Falciatore et al., 1999; Dolch et al., 2017b). Colony PCR (Phire Plant Direct PCR Master Mix, F160, ThermoFisher Scientific) were conducted to analyse the mutation profile by sequencing. Primers were designed for each sgRNA to amplify a region of about 500 nucleotides surrounding the cutting site of the Cas9 protein (Supplemental Table 3). In some rare cases, no amplification was obtained and primers amplifying a larger region were used to reveal a possible very large deletion. Three different kinds of edition profiles were usually obtained: either no edition (pure WT colony), several different editions (mosaic colony), or one edition only (pure mutant colony). For the analysis of mosaic colonies, TIDE (Tracking of Indels by Decomposition v3.3.0) (Brinkman et al., 2014) and ICE (Inference of CRISPR Edits v2, Synthego.com) online software were used to decompose the chromatograms and estimate the edition profile of the colony. Whenever a fraction of cells within the colony presented insertions/deletions (INDELs), a strain isolation was attempted. The equivalent of 200 cells from a mosaic colony were spread on a plate with carbenicillin and zeocin. After three to six weeks, the same steps of PCR and sequencing analyses were conducted. If pure edited colonies were obtained, the screening process was stopped. Otherwise, subsequent isolation attempts were performed.

Glycerolipid analyses.

Glycerolipids were extracted according to the Folch method (Folch et al., 1957) as described previously (Abida et al., 2015). One tenth of total glycerolipid extract was used for an evaluation of total FA content, after methanolysis. Lipids were solubilised with 1 mL of chloroform and a 100-µL aliquot was transferred to a 10-mL crim cap vial (Gerstel). Methanolysis was performed by a MultiPurpose Sampler (MPS, Gerstel). Briefly, 5 µL of 15:0 FA (a FA with a 15-carbon chain length) at 1 $mg.mL^{-1}$ was added as internal standard and 3 mL of methanolysis buffer (methanol/sulphuric acid; 40:1 v/v) were added, vigorously mixed, and incubated at 80°C for one hour, yielding FAs methyl esters (FAME). Reaction was stopped

with the addition of 3 mL of distilled water and 3 mL of hexane. FAME retrieved from the upper phase were suspended in hexane and analysed by Gas Chromatography coupled to Flame Ionization Detection (GC-FID, Perkin Elmer Clarus 580 equipped with a 30-m long cyanopropyl polysilphenesiloxane column, 0.22 mm diameter). FAME were identified by comparison of their retention times with those of standards (Sigma) and quantified by the surface peak method using 15:0 for calibration. Extraction and quantification were performed with three biological replicates. Glycerolipids were then analysed and quantified by high-pressure liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) as previously described (Dolch et al., 2017b), with appropriate standard lipids (Jouhet et al., 2017). Lipid extracts corresponding to 25 nmol of total FAs were dissolved in 100 µL of chloroform/methanol [2/1, v/v] containing 125 pmol of each internal standard. Internal standards used were phosphatidylethanolamine (PE) 18:0-18:0 and diacylglycerol (DAG) 18:0-22:6 from Avanti Polar Lipid, and sulfoquinovosyldiacylglycerol (SQDG) 16:0-18:0 extracted from spinach thylakoid (Deme et al., 2014) and hydrogenated (Buseman et al., 2006). Lipids were then separated by HPLC and quantified by MS/MS. Lipid classes were separated using an Agilent 1200 HPLC system using a 150 mm x 3 mm (length x internal diameter) 5 µm diol column (Macherey-Nagel), at 40°C. The mobile phases consisted of hexane/isopropanol/water/1 M ammonium acetate, pH 5.3 [625/350/24/1, v/v] (A) and isopropanol/water/1 M ammonium acetate, pH 5.3 [850/149/1, v/v] (B). The injection volume was 20 µL. After 5 min, the percentage of B was increased linearly from 0 to 100% in 30 min and kept at 100% for 15 min. This elution sequence was followed by a return to 100% A in 5 min and an equilibration for 20 min with 100% A before the next injection, leading to a total runtime of 70 min. The flow rate of the mobile phase was 200 µL.min^{-1}. The distinct glycerophospholipid classes were eluted successively as a function of the polar head group. Mass spectrometric analysis was performed on a 6460 triple quadrupole mass spectrometer (Agilent) equipped with a Jet stream electrospray ion source under following settings: drying gas heater at 260°C, drying gas flow at 13 L.min^{-1}, sheath gas heater at 300°C, sheath gas flow at 11 L.min^{-1}, nebulizer pressure at 25 psi, capillary voltage at ± 5000 V and nozzle voltage at ± 1,000 V. Nitrogen was used as collision gas. The quadrupoles Q1 and Q3 were operated at widest and unit resolution respectively. Phosphatidylcholine (PC) and diacylglyceryl hydroxymethyltrimethyl-β-alanine (DGTA) analyses were carried out in positive ion mode by scanning for precursors of m/z 184 and 236 respectively at a collision energy (CE) of 34 and 52 eV. SQDG analysis was carried out in negative ion mode by scanning for precursors of m/z

-225 at a CE of -56 eV. Phosphatidylethanolamine (PE), phosphatidylinositol (PI), phosphatidylglycerol (PG), monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG) measurements were performed in positive ion mode by scanning for neutral losses of 141 Da, 277 Da, 189 Da, 179 Da and 341 Da at CEs of 20 eV, 12 eV, 16 eV, 8 eV and 8 eV, respectively. DAG and triacylglycerol (TAG) species were identified and quantified by multiple reaction monitoring (MRM) as singly charged ions [M+NH4]+ at a CE of 16 and 22 eV respectively. Quantification was done for each lipid species by multiple reaction monitoring (MRM) with 50 ms dwell time with the various transitions previously recorded (Abida et al., 2015). Mass spectra were processed using the MassHunter Workstation software (Agilent) for identification and quantification of lipids. Lipid amounts (pmol) were corrected for response differences between internal standards and endogenous lipids and by comparison with a qualified control (Jouhet et al., 2017).

Epifluorescence microscopy.

Cells were observed using an epifluorescence microscope (Zeiss AxioScope A1) equipped with a Zeiss AxioCam MRc. Images were captured using a a Zeiss EC Plan Neofluar 100x/1.3 oil immersion objective. Chlorophyll autofluorescence and Nile Red fluorescence in lipid droplets were visualized with Zeiss filter set 16 (BP 485/20, FT510, LP515) as described previously (Dolch et al., 2017b).

Laser scanning confocal microscopy.

Confocal microscopy was performed with a microscope Zeiss LSM880 equipped with a 63x/1.4 oil-immersed Plan-Apochromat objective, running Zen 2.3 SP1 acquisition software (Platform µLife, IRIG, LPCV). Chlorophyll autofluorescence and eGFP fluorescence were excited using the 488 nm ray of an Argon Multiline laser, and were detected at 600-730 nm and 500-531 nm, respectively, using a GaAsP detector. Laser intensity was set at 5% for visualisation of chlorophyll, and at 2%, 3%, and 5% for eGFP visualisation according to the eGFP signal intensity in each cell. ‘Pseudo brightfield’ images were acquired in parallel by differential interference contrast (DIC), using the 488 nm laser ray at 0.6% as light source and a photomultiplier tube detector for transmitted light (T-PMT). Z-stacks containing consecutive images with a distance of 0.47 nm were obtained for each cell.

Transmission electron microscopy.

Samples were prepared as previously described (Flori et al., 2018). Ultrathin sections (50-70 nm) were prepared with a diamond knife on a PowerTome ultramicrotome (RMC

products, Tucson, AZ, USA) and collected on 200-µm carbon-coated gold grids. Samples were visualised by scanning transmission electron microscopy (STEM) using a MERLIN microscope (Zeiss, Oberkochen, Germany) set up at 30 kV and 240 pA or using a FEI tecnai OSIRIS microscope (Hillsboro, Oregon, USA) set up at 200 kV and ~300 pA.

Accession numbers.

Species and associated MGD protein sequences used for phylogeny (NCBI accession numbers): *Vitrella brassicaformis* (CEM14063.1, CEM34604.1); *Gregarina niphandrodes* (XP_011130274.5); *Pseudo-nitzschia multistriata* (VEU37191.1, VEU39505.1); *Fragilariopsis cylindrus* (OEU16730.1, OEU06991.1); *Fistulifera solaris* (GAX23913.1, GAX10983.1, GAX09439.1); *Phaeodactylum tricornutum* (XP_002186355, XP_002176800, XP_002181685), *Thalassiosira pseudonana* (XP_002295865, XP_002293576.1, XP_002294242); *Ectocarpus siliculosus* (CBJ28381.1, CBJ28372.1, CBN79326.1); *Aureococcus anophagefferens* (XP_009033523.1, XP_009038839.1, XP_009035780.1); Microchloropsis/Nannochloropsis gaditana (XP_005855249.1); *Coccomyxa subellipsoidea* (XP_005651388.1); *Chlorella sorokiniana* (PRW45643.1); *Ostreococcus tauri* (OUS42062.1); *Chlamydomonas reinhardtii* (PNW74102.1); *Monoraphidium neglectum* (XP_013903204.1); *Raphidocelis subcapitata* (GBF88428.1); *Arabidopsis thaliana* (NP_194906.1, NP_565352.1, NP_568394.2); *Coffea arabica* (XP_027069738.1, XP_027084801.1); *Spinacia oleracea* (XP_021867203.1, XP_021852153.1); *Nymphaea colorata* (XP_031495579.1, XP_031480803.1); *Amborella trichopoda* (XP_006852865.1, XP_006845407.1); *Brachypodium distachyon* (XP_010238179.1, XP_003573675.2, XP_003570331.1); *Oryza sativa* (XP_015611851.1, XP_015649135.1, XP_015627712.1); *Physcomitrium patens* (XP_024403181.1, XP_024404076.1); *Selaginella moellendorffii* (XP_024533920.1); *Cyanidioschyzon merolae* (XP_005536420.1); *Cyanidiococcus yangmingshanensis* (KAF6002292.1); *Gracilariopsis chorda* (PXF49423.1, PXF42956.1, PXF46603.1); *Blastochloris viridis* (WP_055036643.1). *Phaeodactylum tricornutum* MGD gene accession numbers: *MGDα* (PHATRDRAFT_14125, NCBI; Phatr3_J14125, Ensembl); *MGDβ* (PHATR_43938, NCBI; Phatr3_J54168, Ensembl); *MGDγ* (PHATRDRAFT_9619, NCBI; Phatr3_J9619, Ensembl).

Acknowledgements.

N.G. was supported by a PhD grant from INRAE. A.A., F.C., E.D., N.G., S.D.G., V.G, J.J., D.P., G.S.L. and E.M were supported by the French National Research Agency (GRAL Labex ANR-10-LABEX-04, EUR CBS ANR-17-EURE-0003) and IDEX Université Grenoble-Alpes (Glyco@Alps Cross-Disciplinary Program; Grant ANR-15-IDEX-02). H.H. and Y.G. were supported by the National Natural Science Foundation of China (41976119, 91751117). A.A., E.M., H.H. and Y.G. were supported by a CEA-CAS bilateral program. Imaging was performed at the µLife Platform of IRIG. Authors wish to thank Guillaume Allorent, Baptiste Doussot and Cécile Giustini for technical assistance.

Author contributions.

N.G. has performed most experimental works. F.C. has provided guidance on Crispr-CAS9 strategy. V.G. and G.S.L. have provided technical assistance on lipidomic analyses and STEM, respectively. D.P. has contributed to the design of photosynthetic measurements and their analyses. A.A., E.D., S.D.G. and M.M. have provided assistance on heterologous expression in yeast. H.H. and Y.G. have developed overexpressing lines. J.J. has provided guidance in the analysis of lipidomic profiles. J.S., A.A. and E.M. have contributed to the design of experiments and their analyses. All authors have contributed to the writing of the manuscript.

References.

Abida, H., Dolch, L.J., Mei, C., Villanova, V., Conte, M., Block, M.A., Finazzi, G., Bastien, O., Tirichine, L., Bowler, C., Rebeille, F., Petroutsos, D., Jouhet, J., and Marechal, E. (2015). Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in Phaeodactylum tricornutum. Plant Physiol 167, 118-136.

Adl, S.M., Bass, D., Lane, C.E., Lukes, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., Cardenas, P., Cepicka, I., Chistyakova, L., Del Campo, J., Dunthorn, M., Edvardsen, B., Eglit, Y., Guillou, L., Hampl, V., Heiss, A.A., Hoppenrath, M., James, T.Y., Karnkowska, A., Karpov, S., Kim, E., Kolisko, M., Kudryavtsev, A., Lahr, D.J.G., Lara, E., Le Gall, L., Lynn, D.H., Mann, D.G., Massana, R., Mitchell, E.A.D., Morrow, C., Park, J.S., Pawlowski, J.W., Powell, M.J., Richter, D.J., Rueckert, S., Shadwick, L., Shimano, S., Spiegel, F.W., Torruella, G., Youssef, N., Zlatogursky, V., and Zhang, Q. (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J Eukaryot Microbiol 66, 4-119.

Almagro Armenteros, J.J., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37, 420-423.

Apt, K.E., Zaslavkaia, L., Lippmeier, J.C., Lang, M., Kilian, O., Wetherbee, R., Grossman, A.R., and Kroth, P.G. (2002). In vivo characterization of diatom multipartite plastid targeting signals. J Cell Sci 115, 4061-4069.

Awai, K., Marechal, E., Block, M.A., Brun, D., Masuda, T., Shimada, H., Takamiya, K., Ohta, H., and Joyard, J. (2001). Two types of MGDG synthase genes, found widely in both 16:3 and 18:3 plants, differentially mediate galactolipid syntheses in photosynthetic and nonphotosynthetic tissues in Arabidopsis thaliana. Proc Natl Acad Sci U S A 98, 10960-10965.

Azadi-Chegeni, F., Thallmair, S., Ward, M.E., Perin, G., Marrink, S.J., Baldus, M., Morosinotto, T., and Pandit, A. (2022). Protein dynamics and lipid affinity of monomeric, zeaxanthin-binding LHCII in thylakoid membranes. Biophysical journal 121, 396-409.

Benning, C., and Ohta, H. (2005). Three enzyme systems for galactoglycerolipid biosynthesis are coordinately regulated in plants. Journal of Biological Chemistry 280, 2397-2400.

Benoiston, A.S., Ibarbalz, F.M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., and Bowler, C. (2017). The evolution of diatoms and their biogeochemical functions. Philos Trans R Soc Lond B Biol Sci 372.

Billey, E., Hafidh, S., Cruz-Gallardo, I., Litholdo, C.G., Jean, V., Carpentier, M.C., Picart, C., Kumar, V., Kulichova, K., Marechal, E., Honys, D., Conte, M.R., Deragon, J.M., and Bousquet-Antonelli, C. (2021a). LARP6C orchestrates posttranscriptional reprogramming of gene expression during hydration to promote pollen tube guidance. Plant Cell.

Billey, E., Magneschi, L., Leterme, S., Bedhomme, M., Andres-Robin, A., Poulet, L., Michaud, M., Finazzi, G., Dumas, R., Crouzy, S., Laueffer, F., Fourage, L., Rebeille, F., Amato, A., Collin, S., Jouhet, J., and Marechal, E. (2021b). Characterization of the Bubblegum acyl-CoA synthetase of Microchloropsis gaditana. Plant Physiol 185, 815-835.

Botte, C., Jeanneau, C., Snajdrova, L., Bastien, O., Imberty, A., Breton, C., and Marechal, E. (2005). Molecular modeling and site-directed mutagenesis of plant chloroplast monogalactosyldiacylglycerol synthase reveal critical residues for activity. J Biol Chem 280, 34691-34701.

Boudiere, L., Michaud, M., Petroutsos, D., Rebeille, F., Falconet, D., Bastien, O., Roy, S., Finazzi, G., Rolland, N., Jouhet, J., Block, M.A., and Marechal, E. (2014). Glycerolipids

in photosynthesis: composition, synthesis and trafficking. Biochim Biophys Acta 1837, 470-480.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R.P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J.A., Brownlee, C., Cadoret, J.P., Chiovitti, A., Choi, C.J., Coesel, S., De Martino, A., Detter, J.C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M.J., Jenkins, B.D., Jiroutova, K., Jorgensen, R.E., Joubert, Y., Kaplan, A., Kroger, N., Kroth, P.G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jezequel, V., Lopez, P.J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L.K., Montsant, A., Oudot-Le Secq, M.P., Napoli, C., Obornik, M., Parker, M.S., Petit, J.L., Porcel, B.M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T.A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M.R., Taylor, A.R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L.S., Rokhsar, D.S., Weissenbach, J., Armbrust, E.V., Green, B.R., Van de Peer, Y., and Grigoriev, I.V. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature 456, 239-244.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res 42, e168.

Browse, J., Warwick, N., Somerville, C.R., and Slack, C.R. (1986). Fluxes through the prokaryotic and eukaryotic pathways of lipid synthesis in the 16-3 Plant Arabidopsis thaliana. Biochemical Journal 235, 25-31.

Bullmann, L., Haarmann, R., Mirus, O., Bredemeier, R., Hempel, F., Maier, U.G., and Schleiff, E. (2010). Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. J Biol Chem 285, 6848-6856.

Buseman, C.M., Tamura, P., Sparks, A.A., Baughman, E.J., Maatta, S., Zhao, J., Roth, M.R., Esch, S.W., Shah, J., Williams, T.D., and Welti, R. (2006). Wounding Stimulates the Accumulation of Glycerolipids Containing Oxophytodienoic Acid and Dinor-Oxophytodienoic Acid in Arabidopsis Leaves. Plant physiology 142, 28.

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.

Conte, M., Lupette, J., Seddiki, K., Meï, C., Dolch, L.-J., Gros, V., Barette, C., Rébeillé, F., Jouhet, J., and Maréchal, E. (2018). Screening for Biologically Annotated Drugs That Trigger Triacylglycerol Accumulation in the Diatom Phaeodactylum. Plant physiology 177, 532.

Corteggiani Carpinelli, E., Telatin, A., Vitulo, N., Forcato, C., D'Angelo, M., Schiavon, R., Vezzi, A., Giacometti, G.M., Morosinotto, T., and Valle, G. (2014). Chromosome scale genome assembly and transcriptome profiling of Nannochloropsis gaditana in nitrogen depletion. Molecular plant 7, 323-335.

Criscuolo, A., and Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC evolutionary biology 10, 210.

Daboussi, F., Leduc, S., Marechal, A., Dubois, G., Guyot, V., Perez-Michaut, C., Amato, A., Falciatore, A., Juillerat, A., Beurdeley, M., Voytas, D.F., Cavarec, L., and Duchateau, P. (2014). Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. Nat Commun 5, 3831.

De Martino, A., Meichenin, A., Shi, J., Pan, K., and Bowler, C. (2007). Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions1. J Phycol 43, 992-1009.

De Riso, V., Raniello, R., Maumus, F., Rogato, A., Bowler, C., and Falciatore, A. (2009). Gene silencing in the marine diatom Phaeodactylum tricornutum. Nucleic Acids Res 37, e96.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J.M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., Flegontova, O., Guidi, L., Horak, A., Jaillon, O., Lima-Mendez, G., Lukes, J., Malviya, S., Morard, R., Mulot, M., Scalco, E., Siano, R., Vincent, F., Zingone, A., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Tara Oceans, C., Acinas, S.G., Bork, P., Bowler, C., Gorsky, G., Grimsley, N., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Raes, J., Sieracki, M.E., Speich, S., Stemmann, L., Sunagawa, S., Weissenbach, J., Wincker, P., and Karsenti, E. (2015). Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. Science 348, 1261605.

Deme, B., Cataye, C., Block, M.A., Marechal, E., and Jouhet, J. (2014). Contribution of galactoglycerolipids to the 3-dimensional architecture of thylakoids. FASEB J 28, 3373-3383.

Dolch, L.J., and Marechal, 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., Morosinotto, T., Tellier, F., Faure, J.D., Falconet, D., Jouhet, J., Sayanova, O., Beaudoin, F., and Marechal, E. (2017a). A Palmitic Acid Elongase Affects Eicosapentaenoic Acid and Plastidial Monogalactosyldiacylglycerol Levels in Nannochloropsis. Plant Physiol 173, 742-759.

Dolch, L.J., Lupette, J., Tourcier, G., Bedhomme, M., Collin, S., Magneschi, L., Conte, M., Seddiki, K., Richard, C., Corre, E., Fourage, L., Laeuffer, F., Richards, R., Reith, M., Rebeille, F., Jouhet, J., McGinn, P., and Marechal, E. (2017b). Nitric Oxide Mediates Nitrite-Sensing and Acclimation and Triggers a Remodeling of Lipids. Plant Physiol 175, 1407-1423.

Dubots, E., Audry, M., Yamaryo, Y., Bastien, O., Ohta, H., Breton, C., Marechal, E., and Block, M.A. (2010). Activation of the chloroplast monogalactosyldiacylglycerol synthase MGD1 by phosphatidic acid and phosphatidylglycerol. J Biol Chem 285, 6003-6011.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature protocols 2, 953-971.

Falciatore, A., d'Alcalà, M.R., Croot, P., and Bowler, C. (2000). Perception of Environmental Signals by a Marine Diatom. Science 288, 2363-2366.

Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C., and Bowler, C. (1999). Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol (NY) 1, 239-251.

Fawley, M.W., Jameson, I., and Fawley, K.P. (2015). The phylogeny of the genus Nannochloropsis (Monodopsidaceae, Eustigmatophyceae), with descriptions of N. australis sp. nov. and Microchloropsis gen. nov. Phycologia 54, 545-552.

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

Flori, S., Jouneau, P.H., Finazzi, G., Marechal, E., and Falconet, D. (2016). Ultrastructure of the Periplastidial Compartment of the Diatom Phaeodactylum tricornutum. Protist 167, 254-267.

Flori, S., Jouneau, P.H., Gallet, B., Estrozi, L.F., Moriscot, C., Schoehn, G., Finazzi, G., and Falconet, D. (2018). Imaging Plastids in 2D and 3D: Confocal and Electron Microscopy. Methods Mol Biol 1829, 113-122.

Folch, J., Lees, M., and Sloane Stanley, G.H. (1957). A simple method for the isolation and purification of total lipides from animal tissues. J Biol Chem 226, 497-509.
Garab, G., Yaguzhinsky, L.S., Dlouhy, O., Nesterov, S.V., Spunda, V., and Gasanoff, E.S. (2022). Structural and functional roles of non-bilayer lipid phases of chloroplast thylakoid membranes and mitochondrial inner membranes. Prog Lipid Res 86, 101163.
Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol Biol Evol 14, 685-695.
Goss, R., and Jakob, T. (2010). Regulation and function of xanthophyll cycle-dependent photoprotection in algae. Photosynth Res 106, 103-122.
Gould, S.B., Maier, U.G., and Martin, W.F. (2015). Protein Import and the Origin of Red Complex Plastids. Curr Biol 25, R515-R521.
Grosche, C., Hempel, F., Bolte, K., Zauner, S., and Maier, U.G. (2014). The periplastidal compartment: a naturally minimized eukaryotic cytoplasm. Curr Opin Microbiol 22, 88-93.
Gruber, A., Rocap, G., Kroth, P.G., Armbrust, E.V., and Mock, T. (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J 81, 519-528.
Gruber, A., Vugrinec, S., Hempel, F., Gould, S.B., Maier, U.G., and Kroth, P.G. (2007). Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol 64, 519-530.
Gschloessl, B., Guermeur, Y., and Cock, J.M. (2008). HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics 9, 393.
Gu, X., Cao, L., Wu, X., Li, Y., Hu, Q., and Han, D. (2021). A Lipid Bodies-Associated Galactosyl Hydrolase Is Involved in Triacylglycerol Biosynthesis and Galactolipid Turnover in the Unicellular Green Alga Chlamydomonas reinhardtii. Plants (Basel) 10.
Guiry, M.D. (2012). How Many Species of Algae Are There? J Phycol 48, 1057-1063.
Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res 35, W585-587.
Iwai, M., Yamada-Oshima, Y., Asami, K., Kanamori, T., Yuasa, H., Shimojima, M., and Ohta, H. (2021). Recycling of the major thylakoid lipid MGDG and its role in lipid homeostasis in Chlamydomonas reinhardtii. Plant Physiol 187, 1341-1356.
Janson, G., Zhang, C., Prado, M.G., and Paiardini, A. (2017). PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. Bioinformatics 33, 444-446.
Jensen, P.E., and Leister, D. (2014). Chloroplast evolution, structure and functions. F1000Prime Rep 6, 40.
Jouhet, J., Lupette, J., Clerc, O., Magneschi, L., Bedhomme, M., Collin, S., Roy, S., Maréchal, E., and Rébeillé, F. (2017). LC-MS/MS versus TLC plus GC methods: Consistency of glycerolipid and fatty acid profiles in microalgae and higher plant cells and effect of a nitrogen starvation. PLOS ONE 12, e0182423.
Jumper, J., and Hassabis, D. (2022). Protein structure predictions to atomic accuracy with AlphaFold. Nat Methods 19, 11-12.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols 10, 845-858.
Kilian, O., and Kroth, P.G. (2005). Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J 41, 175-183.
Kobayashi, K., Nakamura, Y., and Ohta, H. (2009a). Type A and type B monogalactosyldiacylglycerol synthases are spatially and functionally separated in the plastids of higher plants. Plant Physiology and Biochemistry 47, 518-525.
Kobayashi, K., Kondo, M., Fukuda, H., Nishimura, M., and Ohta, H. (2007a). Galactolipid biosynthesis is essential for proper chloroplast biogenesis and embryogenesis. Photosynthesis Research 91, 216-216.
Kobayashi, K., Kondo, M., Fukuda, H., Nishimura, M., and Ohta, H. (2007b). Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. Proc Natl Acad Sci U S A 104, 17216-17221.
Kobayashi, K., Awai, K., Nakamura, M., Nagatani, A., Masuda, T., and Ohta, H. (2009b). Type-B monogalactosyldiacylglycerol synthases are involved in phosphate starvation-induced lipid remodeling, and are crucial for low-phosphate adaptation. Plant Journal 57, 322-331.
Kooistra, W.H., De Stefano, M., Mann, D.G., and Medlin, L.K. (2003). The phylogeny of the diatoms. Prog Mol Subcell Biol 33, 59-97.
Kooistra, W.H., Gersonde, R., Medlin, L.K., and Mann, D.G. (2007). The Origin and Evolution of the Diatoms: Their Adaptation to a Planktonic Existence. In Evolution of Primary Producers in the Sea P.G. Falkowski and A.H. Knoll, eds (Cambridge, U.S.A.: Academic Press), pp. 207-249.
Kroth, P.G., Bones, A.M., Daboussi, F., Ferrante, M.I., Jaubert, M., Kolot, M., Nymark, M., Rio Bartulos, C., Ritter, A., Russo, M.T., Serif, M., Winge, P., and Falciatore, A. (2018). Genome editing in diatoms: achievements and goals. Plant Cell Rep 37, 1401-1408.
Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35, 1547-1549.
Lang, M., Apt, K.E., and Kroth, P.G. (1998). Protein transport into "complex" diatom plastids utilizes two different targeting signals. J Biol Chem 273, 30973-30978.
Le, S.Q., Lartillot, N., and Gascuel, O. (2008). Phylogenetic mixture models for proteins. Philos Trans R Soc Lond B Biol Sci 363, 3965-3976.
Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., and Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res 47, W260-W265.
Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49, W293-W296.
Leyland, B., Zarka, A., Didi-Cohen, S., Boussiba, S., and 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. J Phycol.
Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., Debono, A., Durrett, T.P., Franke, R.B., Graham, I.A., Katayama, K., Kelly, A.A., Larson, T., Markham, J.E., Miquel, M., Molina, I., Nishida, I., Rowland, O., Samuels, L., Schmid, K.M., Wada, H., Welti, R., Xu, C., Zallot, R., and Ohlrogge, J. (2010). Acyl-lipid metabolism. In The Arabidopsis Book, pp. e0133.
Li, X., Moellering, E.R., Liu, B., Johnny, C., Fedewa, M., Sears, B.B., Kuo, M.H., and Benning, C. (2012). A galactoglycerolipid lipase is required for triacylglycerol accumulation and

survival following nitrogen deprivation in Chlamydomonas reinhardtii. Plant Cell 24, 4670-4686.
Lupette, J., Tardif, M., Brugiere, S., Coute, Y., Salvaing, J., and Marechal, E. (In press). Quantitative proteomic analyses reveal the impact of nitrogen starvation on the proteome of the model diatom Phaeodactylum tricornutum. Proteomics.
Makshakova, O., Breton, C., and Perez, S. (2020). Unraveling the complex enzymatic machinery making a key galactolipid in chloroplast membrane: a multiscale computer simulation. Scientific reports 10, 13514.
Mann, D.G. (1999). The species concept in diatoms. Phycologia 38, 437-495.
Mann, D.G., and Droop, S.J.M. (1996). Biodiversity, biogeography and conservation of diatoms. . Hydrobiologia, 19–32.
Marechal, E. (2018). Primary Endosymbiosis: Emergence of the Primary Chloroplast and the Chromatophore, Two Independent Events. Methods Mol Biol 1829, 3-16.
Martens, C., Vandepoele, K., and Van de Peer, Y. (2008). Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species. Proc Natl Acad Sci U S A 105, 3427-3432.
Mirdita, M., Schutze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nature Methods 19, 679-+.
Moog, D., Stork, S., Zauner, S., and Maier, U.G. (2011). In silico and in vivo investigations of proteins of a minimized eukaryotic cytoplasm. Genome Biol Evol 3, 375-382.
Murakami, R., and Hashimoto, H. (2009). Unusual nuclear division in Nannochloropsis oculata (Eustigmatophyceae, Heterokonta) which may ensure faithful transmission of secondary plastids. Protist 160, 41-49.
Nitenberg, M., Makshakova, O., Rocha, J., Perez, S., Marechal, E., Block, M.A., Girard-Egrot, A., and Breton, C. (2020). Mechanism of activation of plant monogalactosyldiacylglycerol synthase 1 (MGD1) by phosphatidylglycerol. Glycobiology 30, 396-406.
Parks, M.B., Nakov, T., Ruck, E.C., Wickett, N.J., and Alverson, A.J. (2018). Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). Am J Bot 105, 330-347.
Petroutsos, D., Amiar, S., Abida, H., Dolch, L.J., Bastien, O., Rebeille, F., Jouhet, J., Falconet, D., Block, M.A., McFadden, G.I., Bowler, C., Botte, C., and Marechal, E. (2014). Evolution of galactoglycerolipid biosynthetic pathways--from cyanobacteria to primary plastids and from primary to secondary plastids. Prog Lipid Res 54, 68-85.
Rastogi, A., Murik, O., Bowler, C., and Tirichine, L. (2016). PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. BMC Bioinformatics 17, 261.
Reyes-Prieto, A., Weber, A.P., and Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. Annu Rev Genet 41, 147-168.
Rocha, J., Sarkis, J., Thomas, A., Pitou, L., Radzimanowski, J., Audry, M., Chazalet, V., de Sanctis, D., Palcic, M.M., Block, M.A., Girard-Egrot, A., Marechal, E., and Breton, C. (2016). Structural insights and membrane binding properties of MGD1, the major galactolipid synthase in plants. Plant J 85, 622-633.
Ruban, A., Lavaud, J., Rousseau, B., Guglielmi, G., Horton, P., and Etienne, A.L. (2004). The super-excess energy dissipation in diatom algae: comparative analysis with higher plants. Photosynth Res 82, 165-175.
Sarkis, J., Rocha, J., Maniti, O., Jouhet, J., Vie, V., Block, M.A., Breton, C., Marechal, E., and Girard-Egrot, A. (2014). The influence of lipids on MGD1 membrane binding highlights novel mechanisms for galactolipid biosynthesis regulation in chloroplasts. FASEB J 28, 3114-3123.

Sato, N., and Awai, K. (2017). "Prokaryotic Pathway" Is Not Prokaryotic: Noncyanobacterial Origin of the Chloroplast Lipid Biosynthetic Pathway Revealed by Comprehensive Phylogenomic Analysis. Genome Biol Evol 9, 3162-3178.

Serôdio, J., and Lavaud, J. (2020). Diatoms and Their Ecological Importance. In Life Below Water, W. Leal Filho, A.M. Azul, L. Brandli, A. Lange Salvia, and T. Wall, eds (Cham: Springer International Publishing), pp. 1-9.

Shang, S., Liu, R., Luo, L., Li, X., Zhang, S., Zhang, Y., Zheng, P., Chen, Z., and Wang, B. (2022). Functional Characterization of the Monogalactosyldiacylglycerol Synthase Gene ptMGD2 in the Diatom Phaeodactylum tricornutum. Front. Mar. Sci. 9, e874448.

Smith, R., Jouhet, J., Gandini, C., Nekrasov, V., Marechal, E., Napier, J.A., and Sayanova, O. (2021). Plastidial acyl carrier protein Delta9-desaturase modulates eicosapentaenoic acid biosynthesis and triacylglycerol accumulation in Phaeodactylum tricornutum. Plant J 106, 1247-1259.

Sorhannus, U. (2007). A nuclear-encoded small-subunit ribosomal RNA timescale for diatom evolution. Mar Micropaleontol 65, 1-12.

Tanaka, A., De Martino, A., Amato, A., Montsant, A., Mathieu, B., Rostaing, P., Tirichine, L., and Bowler, C. (2015a). Ultrastructure and Membrane Traffic During Cell Division in the Marine Pennate Diatom Phaeodactylum tricornutum. Protist 166, 506-521.

Tanaka, T., Maeda, Y., Veluchamy, A., Tanaka, M., Abida, H., Marechal, E., Bowler, C., Muto, M., Sunaga, Y., Tanaka, M., Yoshino, T., Taniguchi, T., Fukuda, Y., Nemoto, M., Matsumoto, M., Wong, P.S., Aburatani, S., and Fujibuchi, W. (2015b). Oil accumulation by the oleaginous diatom Fistulifera solaris as revealed by the genome and transcriptome. Plant Cell 27, 162-176.

Vieler, A., Wu, G., Tsai, C.H., Bullard, B., Cornish, A.J., Harvey, C., Reca, I.B., Thornburg, C., Achawanantakun, R., Buehl, C.J., Campbell, M.S., Cavalier, D., Childs, K.L., Clark, T.J., Deshpande, R., Erickson, E., Armenia Ferguson, A., Handee, W., Kong, Q., Li, X., Liu, B., Lundback, S., Peng, C., Roston, R.L., Sanjaya, Simpson, J.P., Terbush, A., Warakanont, J., Zauner, S., Farre, E.M., Hegg, E.L., Jiang, N., Kuo, M.H., Lu, Y., Niyogi, K.K., Ohlrogge, J., Osteryoung, K.W., Shachar-Hill, Y., Sears, B.B., Sun, Y., Takahashi, H., Yandell, M., Shiu, S.H., and Benning, C. (2012). Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga Nannochloropsis oceanica CCMP1779. PLoS Genet 8, e1003064.

Yates, A.D., Allen, J., Amode, R.M., Azov, A.G., Barba, M., Becerra, A., Bhai, J., Campbell, L.I., Martinez, M.C., Chakiachvili, M., Chougule, K., Christensen, M., Contreras-Moreira, B., Cuzick, A., Fioretto, L.D., Davis, P., De Silva, N.H., Diamantakis, S., Dyer, S., Elser, J., Filippi, C.V., Gall, A., Grigoriadis, D., Guijarro-Clarke, C., Gupta, P., Hammond-Kosack, K.E., Howe, K.L., Jaiswal, P., Kaikala, V., Kumar, V., Kumari, S., Langridge, N., Le, T., Luypaert, M., Maslen, G.L., Maurel, T., Moore, B., Muffato, M., Mushtaq, A., Naamati, G., Naithani, S., Olson, A., Parker, A., Paulini, M., Pedro, H., Perry, E., Preece, J., Quinton-Tulloch, M., Rodgers, F., Rosello, M., Ruffier, M., Seager, J., Sitnik, V., Szpak, M., Tate, J., Tello-Ruiz, M.K., Trevanion, S.J., Urban, M., Ware, D., Wei, S., Williams, G., Winterbottom, A., Zarowiecki, M., Finn, R.D., and Flicek, P. (2022). Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Research 50, D996-D1003.

Zaslavskaia, L.A., Lippmeier, J.C., Kroth, P.G., Grossman, A.R., and Apt, K.E. (2000). Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a variety of selectable marker and reporter genes. J Phycol 36, 379-386.

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