Environmentally-informed functional characterization of a plastid diatom metabolic bridge of mitochondrial origin

Short Title: Probing Diatom Chloroplast Lower-Half Glycolysis

Richard G. Dorrell[1,2,3], Youjun Zhang[4,5], Yue Liang[6,7], Nolwenn Gueguen[8], Tomomi Nonoyama[1,9], Dany Croteau[10], Mathias Penot[1,2,3], Sandrine Adiba[1], Benjamin Bailleul[10], Valérie Gros[8], Juan José Pierella Karlusich[1,2,11], Nathanaël Zweig[1,2], Alisdair R. Fernie[5,6], Juliette Jouhet[8], Eric Maréchal[8] and Chris Bowler[1,2]

[1]Institut de Biologie de l'ENS (IBENS), Département de Biologie, École Normale Supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France
[2]CNRS Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GOSEE, 3 rue Michel-Ange, 75016 Paris, France
[3]Laboratory of Computational and Quantitative Biology (LCQB), Institut de Biologie Paris-Seine (IBPS), CNRS, INSERM, Sorbonne Université, 75005 Paris, France
[4]Center of Plant Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria
[5]Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476, Potsdam-Golm, Germany
[6]Center of Deep Sea Research, Institute of Oceanology, Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China
[7]Laboratory for Marine Mineral Resources, Pilot National Laboratory for Marine Science and Technology, Qingdao 266237, China
[8]Laboratoire de Physiologie Cellulaire et Végétale, CNRS, Univ. Grenoble Alpes, CEA, INRAE, IRIG, 17 rue des Martyrs, 38000 Grenoble, France
[9]Division of Biotechnology and Life Science, Institute of Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan
[10]Institut de Biologie Physico-Chimique (IBPC), Université PSL, 75005 Paris, France
[11]Present address: FAS Division of Science, Harvard University, Cambridge, MA, USA

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: Richard G. Dorrell (richard.dorrell.algae@gmail.com) and Chris Bowler (cbowler@bio.ens.psl.eu)

Abstract

Organic carbon fixed in chloroplasts through the Calvin Cycle can be diverted towards different metabolic fates, including cytoplasmic and mitochondrial respiration; gluconeogenesis; and synthesis of diverse plastid metabolites via the pyruvate hub. In plants, pyruvate is principally produced via cytoplasmic glycolysis, although a plastid-targeted lower-half glycolytic pathway is known in non-photosynthetic tissue. Here, we characterize a lower-half plastid glycolytic-gluconeogenesis pathway in diatoms, ecologically important marine algae distantly related to plants. We show that two reversible enzymes required to complete plastid glycolysis-gluconeogenesis, Enolase and PGAM (*bis*-phospho-glycerate mutase), originated through duplications of mitochondria-targeted respiratory isoforms. Through CRISPR-Cas9 mutagenesis, integrative 'omic analyses, and measured kinetics of expressed enzymes in the diatom *Phaeodactylum tricornutum*, we present evidence that this pathway diverts plastid glyceraldehyde-3-phosphate into the pyruvate hub, and may also function in the gluconeogenic direction. Considering experimental and environmental data, we show that this pathway has different roles in relation to long days and low temperatures as are found in sub-polar oceans, where diatoms dominate primary production. Our data provide a further explanation for the success of diatoms in the contemporary ocean, and functional insights into a poorly understood yet evolutionarily recurrent plastid metabolic pathway.

Keywords: chloroplast; *Tara* Oceans; meta-genomics informed phenotyping; post-endosymbiotic evolution; plastid-mitochondria crosstalk; RNAseq; GC and LC-MS; photophysiology

Introduction

Each year, over 250 gigatonnes of atmospheric carbon dioxide is assimilated through photosynthesis, with effectively equal contributions from terrestrial plants and aquatic

algae (Friedlingstein, Jones et al. 2022). This activity is essential for maintaining planetary climate homeostasis and supporting the entire Earth ecosystem. Carbon assimilated through photosynthesis via the Calvin cycle is diverted into multiple metabolic fates (Raines 2003). In plants, these include gluconeogenesis directly in plastids (or chloroplasts), used in leaf tissue for starch storage (Scialdone, Mugford et al. 2013). Many additional metabolites including fatty acids and lipids, amino acids, and chlorophyll and carotenoid pigments are synthesised directly in the plastid (Tanaka and Tanaka 2007, Bromke 2013, Maréchal and Lupette 2020, Bai, Cao et al. 2022) (**Fig. 1A**). Many of these plastid metabolic reactions utilize pyruvate, or its adjacent metabolic precursor phospho-*enol*-pyruvate (or PEP), and are referred to collectively as the pyruvate hub (Shtaida, Khozin-Goldberg et al. 2015). In addition, plant photosynthate is exported from the plastids to the cytosol for subsequent glycolysis and respiration in the mitochondria (Moog, Rensing et al. 2015), or for transport to non-photosynthetic tissue (Carrera, George et al. 2021) (**Fig. 1A**).

Plants are classically thought to generate PEP and pyruvate through glycolysis in the cytoplasm, then reimport these metabolites into the plastids (**Fig. 1A**) (Moog, Nozawa et al. 2020). Alongside this, certain plants may synthesize pyruvate hub substrates directly from the Calvin cycle inside the plastid. This conversion is performed by two enzymes, a plastid-targeted enolase and phospho-glycerate mutase (henceforth referred to as cpEnolase and cpPGAM), which allow the conversion of 1,3-bis-phosphoglycerate from the Calvin cycle to PEP (**Fig. 1A**) (Raines 2003, Andriotis, Kruger et al. 2010). Both Enolase and PGAM have been shown experimentally to be fully reversible enzymes, with bidirectional functions that we henceforth refer to as glycolysis-gluconeogenesis, contrasting with glycolysis and gluconeogenesis to signify enzymatic activities in one direction only (Sutherland, Posternak et al. 1949). Documented plant cpEnolase and cpPGAM enzymes are associated with non-photosynthetic tissues such as seeds and roots (Prabhakar, Löttgert et al. 2009, Fukayama, Masumoto et al. 2015, Troncoso-Ponce, Rivoal et al. 2018). *Arabidopsis thaliana* cpEnolase and cpPGAM knockout lines have limited phenotypes under replete growth conditions (Prabhakar, Löttgert et al. 2009, Andriotis, Kruger et al. 2010, Anoman, Flores-Tornero et al. 2016), raising questions of their overall function.

Diatoms are a eukaryotic algal group that is only distantly related to plants, with over one billion years of evolutionary separation (Nonoyama, Kazamia et al. 2019, Strassert, Irisarri et al. 2021). In contrast to the primary plastids of plants, surrounded by two membranes and of bacterial origin, diatoms possess complex plastids surrounded by four membranes and derived from a eukaryotic red alga, which is likewise ancient (Nonoyama, Kazamia et al. 2019, Liu, Storti et al. 2022). Diatoms are extraordinarily successful in the modern ocean, comprising nearly half of total algal abundance e.g., in environmental sequence data from the *Tara* Oceans expedition (Malviya, Scalco et al. 2016, Behrenfeld, Halsey et al. 2021). Diatoms are particularly successful in high-latitude and temperate oceans (i.e., the North Atlantic, North Pacific and Southern Oceans) that are characterised by high primary production despite photo-physiological stresses including low temperatures and elongated photoperiods (long days in the summer, and long nights in the winter) (Gilbertson, Langan et al. 2022, Joli, Concia et al. 2023). Previous studies, particularly of the transformable species *Phaeodactylum tricornutum*, have identified multiple strategies that allow diatoms to tolerate photo-stress, including complex inter-organelle metabolite trafficking (Bailleul, Berne et al. 2015, Broddrick, Du et al. 2019, Smith, Dupont et al. 2019) and extensive photoprotective capabilities (reviewed in (Lepetit, Campbell et al. 2022). These data are further supported by extensive environmental (meta-genomic) sequence data such as those of the *Tara* Oceans mission, allowing us to study the performance of individual diatom chloroplast proteins in the wild (Kazamia, Sutak et al. 2018, Liu, Storti et al. 2022).

Diatom carbon metabolism is highly different to that of plants (Kroth, Chiovitti et al. 2008). Differences include the storage of sugars in cytoplasmic vacuoles (as chrysolaminarin) as opposed to in plastidial starch, and the synthesis of most lipid groups (e.g., galactolipids and part of triacylglycerol pathway) directly in the plastid (Zhu, Shi et al. 2016, Huang, Pan et al. 2023). Diatom plastids furthermore possess no known plastid hexose phosphate transporters, which in plants are implicated in plastidial sugar import in storage tissue, and are inferred to exchange sugars with the cytoplasm via triose phosphates only (Moog, Nozawa et al. 2020, Liu, Storti et al. 2022) (**Fig. 1A**). Strikingly, the lower half of respiratory glycolysis-gluconeogenesis in diatoms occurs in the mitochondria, as opposed to the cytoplasm (Kroth, Chiovitti et

al. 2008, Río Bártulos, Rogers et al. 2018); and a complete plastid lower half glycolysis-gluconeogenesis, including cpEnolase and cpPGAM proteins, has been inferred from sequenced diatom genomes (Kroth, Chiovitti et al. 2008, Smith, Abbriano et al. 2012) (**Fig. 1A**). As diatoms are unicellular and colonial species, plastid glycolysis presumably occurs in the same organelle as photosynthesis, contrasting with its predominantly non-photosynthetic distribution in plants (**Fig. 1A**).

Here, we use combined profiling of sequence datasets from cultivated and environmental diatoms, characterization of *P. tricornutum* CRISPR-CAS9 knockout mutants and measured kinetic activities of expressed enzymes, to reveal the probable functions of diatom cpEnolase and cpPGAM enzymes. We demonstrate that the genes encoding these enzymes arose from diatom mitochondria-targeted and respiratory isoforms, and are most highly expressed at high latitudes in environmental sequence data from *Tara* Oceans. On the basis of knockout line phenotypes, we present evidence that this pathway has augmented importance in cells grown under continuous illumination as opposed to light-dark cycling. Considering both mutant phenotypes and measured kinetic activities, we propose that the principal functions of diatom cpEnolase and cpPGAM are in the glycolytic direction from glyceraldehyde-3-phosphate into the pyruvate hub, albeit with some potential flux in the reverse gluconeogenic direction which may be influenced both by day-length and temperature. Overall, our data position lower half glycolysis-gluconeogenesis as an adaptive modulator of diatom plastid metabolic poise, providing insights into its physiological roles for photosynthetic organisms beyond plants.

Results

Phylogeny and localization of Enolase and PGAM enzymes suggest recent recruitments of mitochondrial glycolytic enzymes to the diatom plastid

To evaluate the occurrence of plastid-targeted glycolysis across the algal tree of life, we searched for plastid-targeted homologues of *Phaeodactylum tricornutum* and *Arabidopsis thaliana* enolase and PGAM enzymes in 1,673 plant and algal species, considering genomes from JGI PhycoCosm, and transcriptomes from the MMETSP

(Marine Microbial Eukaryotic Transcriptome Sequencing Project) and OneKp (One Thousand Plant Transcriptomes) initiatives (Keeling, Burki et al. 2014, Initiative 2019, Grigoriev, Hayes et al. 2021). Plastid-targeting sequences were inferred using both PFAM domain presence and the combined *in silico* predictions of HECTAR, ASAFind, TargetP and PredAlgo (Emanuelsson, Brunak et al. 2007, Gschloessl, Guermeur et al. 2008, Tardif, Atteia et al. 2012) (**Table S1,** sheet 1). Plastid lower glycolysis-gluconeogenesis was frequently inferred in diatoms, with 60/101 (59%) libraries with identified enolase and PGAM sequences possessing plastid-targeted versions of each. A lower occurrence (22/69 libraries, 32%) was found amongst close relatives in the stramenopiles (e.g., pelagophytes, dictyochophytes) and other algae with secondary red plastids (cryptomonads, haptophytes; 25/94 libraries, 27%) (**Fig. S1A**). Within primary plastid-harbouring lineages, only angiosperms were inferred to frequently possess plastid-targeted copies of both enzymes (47/537 libraries, 9%). Notably, only 4/127 (3%) occurrences were inferred in primary green algae and none in primary red algae, suggesting that diatom plastid glycolysis does not derive from the secondary red chloroplast ancestor (**Fig. S1A**). Considering collection sites, diatom species with either plastid glycolysis enzyme typically derive from higher latitudes (mean unsigned latitude 45.6°, standard deviation 13.5°, n = 81) than ones that possess neither (mean unsigned latitude 38.9°, standard deviation 24.3°, n = 10; one-way ANOVA P = 0.19; **Fig. S1B**). This difference was deemed to be significant for certain diatom groups (e.g., araphid pennate diatoms, **Dataset S1**, sheet 1; one-way ANOVA P = 0.012) but was less visible in other algal groups (e.g., green algae), where lower-half plastid glycolysis was more frequently detected in species collected from low latitude habitats (**Fig. S1B**).

Next, we explored the specific origins of *P. tricornutum* plastid Enolase and PGAM sequences from diatoms by building phylogenies of the closest orthologs obtained from other diatoms, the broader taxonomic group to which they belong, the stramenopiles, and two other algal groups, the cryptomonads and haptophytes. These lineages all possess plastids of secondary red endosymbiotic origin, surrounded by four membranes, which are likely to be closely related to one another (Strassert, Irisarri et al. 2021), but also contain non-photosynthetic members (e.g., oomycetes in stramenopiles) which only possess respiratory (i.e., mitochondria-targeted) lower half glycolytic enzymes (Río Bártulos, Rogers et al. 2018). Single-

gene trees were made for the conserved domains of all organelle-targeted Enolase and PGAM sequences from 289 cryptomonad, haptophyte and stramenopile genomes and transcriptomes, plus all orthologs from 85 further genomes selected from across the tree of life, based on a previously defined pipeline (**Supporting Dataset 1**, sheet 2-9). **Figs. 1B** and **1C** show consensus MrBayes trees realised with GTR, Jones and WAG substitution matrices for species with both identifiable plastid- and mitochondria- targeted orthologs of each protein.

The obtained topologies revealed multiple evolutionary origins for plastid Enolase and PGAM sequences from mitochondria-targeted (respiratory) enzymes, with diatom plastid isoforms typically having recent and/or diatom-specific evolutionary origins. Diatom cpEnolase sequences resolve in a well-supported clade with plastid-targeted enzymes from bolidophytes, dictyochophytes and pelagophytes, which are sisters to diatoms in the stramenopile tree (Río Bártulos, Rogers et al. 2018, Nonoyama, Kazamia et al. 2019), followed by mitochondria-targeted proteins from these groups (MrBayes PP = 1.0 under all studied matrices, **Fig. 1B**), other photosynthetic (chrysophytes) and non-photosynthetic stramenopiles (oomycetes; MrBayes PP = > 0.95 under GTR and Jones matrices, **Fig. 1B**). This indicates a duplication and recruitment of the host-derived mitochondria-targeted protein to the plastid within a common ancestor of the diatoms, pelagophytes and dictyochophytes. A broader evaluation of cpEnolase distribution suggests further duplications and plastid retargeting of mitochondria-targeted enolase proteins in both the chrysophytes and cryptomonads (**Fig. S2**).

The PGAM phylogeny revealed at least two closely-related families of plastid-targeted diatom enzymes, both likely derived from host mitochondrial isoforms. The cpPGAM1A clade (typified by the *P. tricornutum* protein Phatr3_J17086) was closely related to mitochondrial-targeted proteins found across the stramenopiles (MrBayes PP = 1.0 under all studied matrices, **Fig. 1C**), followed by plastid-targeted proteins from chrysophytes and mitochondria-targeted oomycete proteins. Similarly, the cpPGAM1B (Phatr3_J50414) clade included mitochondrial-targeted proteins from pelagophytes and dictyochophytes (MrBayes > = 0.85 under all studied matrices, **Fig. 1C**), and plastid- and mitochondria-targeted enzymes from the chrysophytes

(**Fig. S3**). Further duplications and plastid recruitments of mitochondria-targeted PGAM proteins were again visible in the haptophytes and cryptomonads (**Fig. S3**).

A final plastid-targeted protein annotated as PGAM in the version 3 *P. tricornutum* genome (Rastogi, Maheswari et al. 2018), hereafter termed PGAM2, was identified exclusively in diatoms, pelagophytes, and haptophytes (**Fig. S4**), with limited homology to PGAM1 enzymes (BLASTp e-value > 1.0 in pairwise protein-protein searches). Only PGAM1 enzymes contain an annotated phospho-glyceromutase active site (IPR005952) inferred using InterProScan, while both PGAM1 and PGAM2 belong to the same PFAM superfamily (histidine phosphatase, PF03000) per PFAMscan (Jones, Binns et al. 2014, Mistry, Chuguransky et al. 2020). PGAM2 enzymes were predominantly mitochondria-targeted, with plastid- or dual-targeted isoforms only identified in *P. tricornutum* (Phatr3_J37201) and a small number of other diatoms and haptophyte species (**Fig. S4**).

To confirm plastid localization of *P. tricornutum* cpEnolase and cpPGAM, eGFP-tagged copies of three proteins (Phatr3_J41515, cpEnolase; Phatr3_J17086, cpPGAM1A; Phatr3_J37201, cpPGAM2) were expressed in *P. tricornutum* Pt1.86 cells via biolistic transformation. The observed GFP fluorescence patterns were coincident with chlorophyll autofluorescence, confirming plastid localization (**Figs. 1D, S5**).

Physiological roles of diatom cpEnolase and cpPGAM inferred from environmental sequence expression data

Next, we considered general patterns of transcriptional co-regulation of diatom cpEnolase and cpPGAM sequences in environmental sequence data from *Tara* Oceans. First, we used a previously benchmarked pipeline, based on combined hmmer, reciprocal BLAST and phylogenetic filtration (ti et al. 2022) to identify *Tara* Oceans meta-genes that reconcile exclusively with plastid-targeted proteins from cultured diatom species, to the exclusion of non-diatom and non-plastid homologs (**Fig. S6A**). Amongst the retained meta-genes likely to be N-terminally complete (BLAST homology within the first 40 residues of a *P. tricornutum* sequence), a majority have consensus plastid-targeting sequences (enolase: 38/ 78- 49%, PGAM:

58/ 97- 60%). Only a very small number (one enolase, 10 PGAM) possess mitochondrial or endomembrane localizations, suggesting that they principally correspond to plastid-targeted environmental homologs of each protein (**Fig. S6B, Supplemental Dataset 3**, sheet 11).

Within *Tara* Oceans data, the greatest relative abundances of diatom cpEnolase and cpPGAM1 sequences were observed in meta-transcriptome (metaT) data in stations from both high northern and southern latitudes, in both surface (**Fig. 2**) and deep chlorophyll maximum (DCM) samples, and across all size fractions (**Fig. S7**). These levels were notably greater than equivalent levels in meta-genome (metaG) data (**Figs. 2, S7**). Normalization of metaT abundances calculated for each gene in the 0.8-2000 µm size fraction against all diatom metaT sequences showed positive correlations to latitude both in surface (cpEnolase $R^2 = 0.18$, $P < 10^{-05}$, cpPGAM1A $R^2 = 0.23$, $P < 10^{-05}$) and DCM depths (cpEnolase $R^2 = 0.53$, $P < 10^{-05}$, cpPGAM1A $R^2 = 0.59$, $P < 10^{-05}$) (**Supplemental Dataset 3**, sheet 10). Considering specific variables, relative abundance levels showed clearest positive correlations to daylength and negative correlations to temperature. No other parameters (e.g., nutrient concentrations) showed as clear correlations to chloroplast glycolysis metaT relative abundances (**Supplemental Dataset 3**, sheet 10).

Similar positive correlations were observed when normalising the metaT abundances obtained for diatom chloroplast glycolysis genes against the relative abundances calculated for the meta-genomic (metaG) sequences of the same genes. This was true in surface depths for cpEnolase ($R^2 = 0.10$, one-tailed *F*-test, $P < 0.05$) and DCM for both genes (cpEnolase $R^2 = 0.28$, one-tailed *F*-test $P < 0.05$, cpPGAM1 $R^2 = 0.29$, one-tailed *F*-test $P< 0.05$ (**Supplemental Dataset 3**, sheet 10). Further significant positive correlations to latitude, both considering absolute (Pearson) and ranked (Spearman) correlation values, were detected in multiple individual size fractions (0.8-5, 3/5-20, 20-180, 180-2000 µm) at each depth, including for cpPGAM1 metaT normalised against metaG in surface 3/5-20 (one-tailed *F*-test, $P < 10^{-05}$), 20-180 (one-tailed *F*-test; $P < 10^{-05}$) and 180-2000 (one-tailed *F*-test, $P < 0.05$) µm fractions (**Supplemental Dataset 3**, sheet 10).

The transcriptional preference of diatom cpEnolase and cpPGAM1 for high latitudes contrasted strongly with PGAM2, which showed equivalent relative abundance in some stations from the temperate South Pacific and Atlantic as stations from the Arctic and Southern Oceans (**Fig. S8**; **Supplemental Dataset 3**, sheet 10). Although a positive correlation between relative PGAM2 accumulation and latitude was observed in metaT data normalised against total diatom metaT abundances and 0.8-2000 µm size fractions (surface $R^2 = 0.20$, DCM $R^2 = 0.37$, one-tailed *F*-test $P < 0.05$), no such correlation was observed for metaT data normalised against PGAM2 metaG relative abundances (surface $R^2 = 0.00062$, DCM $R^2 = 0.027$, one-tailed *F*-test $P > 0.05$) suggesting that latitude does not directly influence PGAM2 expression (**Fig. S8B**). In certain size fraction and depth combinations (e.g., DCM 0.8-3, and 3/5-20 µm fractions, normalised against metaG abundances; and surface and DCM 180-2000 µm fractions normalised against all diatom metaT abundances) PGAM2 metaT abundances even demonstrated significant negative correlations to latitude (**Supplemental Dataset 3**, sheet 10).

Growth phenotypes of cpEnolase and cpPGAM1A knockout and complementation lines reveal modular roles in response to daylength

We generated homozygous CRISPR knockout lines for both cpEnolase and cpPGAM1A in the model diatom *P. tricornutum*. cpPGAM1A was selected over other PGAM isoforms because of its clear transcriptional co-regulation to cpEnolase (**Supplemental Dataset 2**, in contrast to cpPGAM1B) and latitudinal expression correlation in *Tara* Oceans (**Figs. 2, S8**, in contrast to cpPGAM2). Multiple CRISPR knockout lines were generated from two regions with unique sequences in the *P. tricornutum* genome for each gene (cpEnolase CRISPR region 1 *n*= 4, CRISPR region 2 *n*= 3; cpPGAM1A CRISPR region 1 *n*= 2, CRISPR region 2 *n*= 3) (**Fig. S9A**). Each CRISPR line was verified by sequencing to be homozygous and to contain a frame-shift mutation sufficient to impede successful translation of the encoded protein (**Fig. S9A**). Commercial antibodies against enolase and PGAM were found not to specifically label cpEnolase and cpPGAM1A in Western Blots, and so we inferred protein relative expression level by qRT-PCR using recognised *P. tricornutum* housekeeping genes (Sachse, Sturm et al. 2013, Zhang, Sampathkumar et al. 2020). The measured knockout mRNA abundance in each line was significantly

lower (1.8-39 %) than that identified in empty vector control mRNA (n = 4, one-way ANOVA, P < 0.05) under 19 °C and 12h light: 12h dark (« 19C LD ») conditions, **Fig. S9B**), consistent with effective knockdown of mutated genes (e.g., via non-sense mediated decay, (Chang, Imam et al. 2007)).

Next, we performed growth curves of cpEnolase and cpPGAM1A knockout lines compared to empty vector controls (**Fig. 3**; **Supplemental Dataset 4**, sheets 3-6). We chose to target changes in light and temperature, given that both show clear associations observed with cpPGAM1A and cpEnolase in *Tara* Oceans data (**Supplemental Dataset 3**, sheet 10). The tested conditions were: 19 °C and 12 h: 12 h light: dark cycling (19C LD), reflecting other *P. tricornutum* physiology studies (Cruz de Carvalho, Sun et al. 2016, McCarthy, Smith et al. 2017); 19 °C and 24 h continuous light (19C CL) to test the effects of photoperiod; and 8 °C and 24 h continuous light (8C CL) to test the effects of temperature.

Under 19C LD growth conditions, plastid glycolysis-gluconeogenesis knockout lines showed an approximately 10-15% reduction in relative growth rate compared to empty vector controls (cpEnolase growth rate 0.83± 0.06 cells day$^{-1;}$ cpPGAM1A growth rate 0.85± 0.07 cells day$^{-1;}$ empty vector growth rate 0.94 ± 0.05 cells day^{-1}; **Fig. 3, S10; Supplemental Dataset 4**, sheet 3; cpEnolase growth rate 87.7% control and cpPGAM1A growth rate 90.1% control, one-way ANOVA, two-tailed P < 0.05). Under 19C CL, knockout lines showed a 25-30% reduction in relative growth rate compared to controls (cpEnolase growth rate 0.99± 0.16 cells day^{-1}; cpPGAM1A growth rate 1.08± 0.04 cells day^{-1}; empty vector growth rate 1.39 ± 0.09 cells day^{-1}; **Fig. 3, S10**; **Supplemental Dataset 4**, sheet 4; cpEnolase growth rate 70.7% control and cpPGAM1A growth rate 77.5% control, one-way ANOVA, two-tailed P < 0.01). By contrast, under 8C CL we observed overlapping growth rates for knockout and empty vector control lines, albeit with a possible reduction in cpEnolase knockout growth rate (cpEnolase relative growth rate 0.49± 0.10 cells day^{-1}, cpPGAM1A growth rate 0.64± 0.02 cells day^{-1}, empty vector growth rate 0.62± 0.07 cells day^{-1}; **Fig. 3, S10; Supplemental Dataset 4**, sheet 5; cpEnolase growth rate 78.1% control and cpPGAM1A growth rate 102.9% control; one-way ANOVA, two-tailed P non-significant).

To test the possibility of off-target effects of the CRISPR constructs, we complemented mutant lines with blasticidin resistance genes linked to either cpEnolase-GFP or cpPGAM1A-GFP modified to remove all CRISPR target sequences (**Supplemental Dataset 4**, sheet 2) (McCarthy, Smith et al. 2017, Buck, Río Bártulos et al. 2018). Despite an overall lower growth rate in all blasticidin-resistant lines compared to primary transformants, and within-line variation, comparative growth curves of 47 complemented versus placebo transformed mutant lines revealed increased growth rates in complemented cpEnolase and cpPGAM1A versus blank transformed knockout lines under 19C CL and 19C LD (**Supplemental Dataset 4**, sheet 7; one-way one-way ANOVA, two-tailed P, < 0.05). By contrast, complemented knockout line growth rates overlapped with empty vector controls either transformed with cpEnolase or blank complementing vectors, indicating effective recovery of mutant phenotypes (**Supplemental Dataset 4**, sheet 7).

Finally, we performed comparative photophysiological measurements of knockout lines in the two conditions (19C LD and 19C CL) where they presented a growth phenotype (see Methods). Our data indicate that the presence/ absence of these enzymes does not significantly impact photosynthetic performance. The light dependencies of either electron transfer rate through photosystem II (PSII) (rETR(II)) or photoprotection (non-photochemical quenching, NPQ) were very similar between control and knock-out lines (**Fig. S11A**; **Supplemental Dataset 4**, sheets 8-11). A slight but significant increase in the functional absorption cross-section of photosystem II (σPSII) was found under 19C CL in both cpEnolase (319.3± 22.5) and cpPGAM1A knockouts (306.6± 11.6) compared to controls (292.3± 8.2; one-way ANOVA, $P < 0.05$) (Gorbunov, Shirsin et al. 2020). This elevation was suppressed in both complemented lines (**Fig. S11B**; **Supplemental Dataset 4**, sheet 11).

Primary metabolic functions of cpEnolase and cpPGAM1A inferred from comparative gene expression analysis of P. tricornutum *knockout lines*

Next, we investigated the impacts of disruption of plastid glycolysis on diatom metabolism beyond photosynthesis. First, we performed quantitative RNA-seq analysis using 63 RNA samples drawn from multiple knockout and empty vector lines under all three physiological conditions (19C LD, 19C CL, and 8C CL; **Supplemental**

406 **Dataset 5**, sheet 1; Materials and Methods). Complete results are provided in
407 **Supplemental Dataset 5**, sheets 5-11. Both cpEnolase and cpPGAM1A mRNA were
408 found to significantly under-accumulate in the corresponding knockout lines,
409 consistent with qRT-PCR analysis (**Fig. S9B**) and suggesting maintenance of the
410 mutant genotypes throughout RNA sequencing.
411
412 Genome-scale enrichment analyses of the *in silico* localizations of proteins encoded
413 by differentially expressed genes revealed distinctive changes in glycolysis knockout
414 organelle metabolism. These effects were most evident in 19C CL, in which 90/239
415 (38%) of the genes differentially upregulated (mean fold-change >2, P-value < 0.05)
416 in both cpEnolase and cpPGAM1A knockout lines compared to controls were
417 predicted to possess chloroplast targeting peptides based on ASAFind (Gruber,
418 Rocap et al. 2015) or HECTAR (Gschloessl, Guermeur et al. 2008). This was
419 significantly greater than the proportion of genes (1,585/11,514, 14%) across the
420 entire genome predicted to encoding chloroplast-targeted proteins that were detected
421 in RNAseq data (one-tailed chi-squared $P < 10^{-05}$; **Fig. 4A; Supplemental Dataset 5**,
422 sheet 10). These results were supported by domain enrichment analyses, indicating
423 significant (one-tailed chi-squared P < 0.05) enrichments in light-harvesting complex
424 (GO:0030076), photosynthesis (GO:0009765) and protein-chromophore linkage
425 (GO:0018298) GO terms. A more detailed resolution of gene expression patterns
426 underpinning core organelle metabolism pathways (Ait-Mohamed, Novák Vanclová et
427 al. 2020) suggested concerted upregulation of genes encoding light-harvesting
428 complexes and photosynthesis machinery and plastid fatty acid synthesis machinery,
429 alongside a probable upregulation of mitochondrial respiratory complex I and ATP
430 synthase (**Supplemental Dataset S5**, sheets 10-11). Less dramatic changes were
431 evident in 19C LD and 8C CL, although 13 of the 51 genes (25%) inferred to be
432 downregulated in both cpEnolase and cpPGAM1A knockout lines under 8C CL were
433 inferred to encode chloroplast-targeted proteins by either ASAFind or HECTAR,
434 representing likewise an enrichment compared to all genes identified within the
435 RNAseq data (one-tailed chi-squared P < 0.05; **Fig. 4A**).
436
437 To gain a more precise insight into the effects of plastid glycolysis-gluconeogenesis
438 on *P. tricornutum* metabolism, we additionally validated the differential expression of
439 eleven exemplar genes encoding chloroplast- and mitochondria-targeted proteins by

qPCR in knockout and empty vector control lines across all three conditions (**Fig. 4B**; **Supplemental Dataset 5**, sheet 12). These genes showed relatively limited differences under 19C LD, limited to a slight depression in the accumulation of *Lhcf1* (Phatr3_J18049) and chorismate mutase (Phatr3_J43277) mRNA in both cpEnolase and cpPGAM1A knockouts compared to control lines (~50% downregulation, two-tailed *t*-test $P < 0.05$; **Fig. 4B**). Both knockout lines over-accumulated (>600%; two-tailed *t*-test $P < 10^{-05}$) mRNAs encoding mitochondrial phospho-glycerate mutase (Phatr3_J33839) under 19C LD compared to control lines (**Fig. 4B**).

Under 19C CL, we observed more substantial changes in plastid metabolism, including the significant (two-tailed *t*-test $P < 0.05$) over-accumulation of mRNAs encoding Lhcf1 (~150%), a plastid-targeted petB-type protein presumably involved in cytochrome b_6f metabolism (Phatr3_J13558, ~90%), and a particularly strong over-accumulation of plastid lysophosphatidyl acyltransferase, involved in plastid lipid synthesis (Phatr3_J20640, ~100%, two-tailed *t*-test $P < 10^{-05}$) in both knockout lines (**Fig. 4B**). Significant over-accumulations were also observed of mRNAs encoding plastid signal processing peptidase (Phatr3_J10319, 60-120%), alanine transaminase (Phatr3_J34010) and coporphyrinogen oxygenase (Phatr3_J12186), in either cpEnolase or cpPGAM1A knockout lines (**Fig. 4B**). Concerning mitochondrial metabolism, a strong increase (>250%, two-tailed *t*-test $P < 10^{-05}$) was observed in mRNA for NDH dehydrogenase subunit 1 (Phatr3_J43944), involved in oxidative phosphorylation, but a corresponding decrease (>40%, two-tailed *t*-test $P < 10^{-05}$) in mRNA for citrate synthase within the TCA cycle (Phatr3_J30145).

Finally, under 8C CL, contrasting and complementary changes were observed: up-regulation (>60%; two-tailed *t*-test $P < 10^{-05}$) of genes encoding both the plastid signal processing peptidase and petB-related protein, and mitochondrial PGAM and citrate synthases in both knockout lines compared to controls (**Fig. 4B**). Both knockout lines were found to under-accumulate *Lhcf1* mRNA (>90%; two-tailed *t*-test $P < 10^{-05}$), while *Lhcx4* (Phatr3_J38720), encoding a dark-expressed gene of unknown direct function but homologous to the Lhcx1 protein implicated in photoprotection (Buck, Sherman et al. 2019), was found to substantially over-accumulate in both cpEnolase and cpPGAM1A knockout lines (**Fig. 4B**).

Metabolite profiling indicates potential systemic outputs of disrupted plastid glycolysis-gluconeogenesis

Next, we considered the compound effects of cpEnolase and cpPGAM1A knockout on global metabolite accumulation under each environmental condition via GC-MS profiling of 32 sugars and amino acids (**Fig. 5; Fig. S13**), across 139 samples drawn from multiple knockout and control lines under 19C LD, 19C CL and 8C CL. These samples were obtained from cell pellets collected from mid-exponential phase cultures, and thus correspond to the long-term impacts on metabolite accumulation in plastid glycolysis knockout lines. Complete outputs are tabulated in **Supplemental Dataset 6**, sheets 1-2.

We were unable to directly measure the accumulation of any of the products or substrates of either cpPGAM1A or cpEnolase (3-phosphoglycerate, 2-phosphoglycerate, PEP), although we detected significantly diminished (one-way ANOVA two-tailed P-value < 10^{-05}) pyruvate accumulation in cpPGAM1A knockouts under all three conditions, and in cpEnolase knockouts under 8C CL (**Fig. 5, S13**). In all three conditions, significant reductions (one-way ANOVA two-tailed P-value < 0.01 in both cpEnolase and cpPGAM1A knockout lines) were observed in cytoplasmic sugars and sugar derivatives (glucose, sucrose, histidine, *myo*-inositol) in cpEnolase and cpPGAM1A knockouts compared to control lines (**Fig. 5**). cpEnolase and cpPGAM1A knockout lines further under-accumulated citric acid in all three conditions, and malic acid in 8C CL (**Fig. 5**).

A probable over-accumulation of phosphoric acid was observed in all knockout lines except cpPGAM1A under 19C CL (**Fig. 5; S13**). Significant (one-way ANOVA two-tailed P-value < 10^{-05}) over-accumulations were identified for valine in cpEnolase and cpPGAM1A knockouts under 19C CL and 8C CL; for methionine and ornithine in 19C CL only; and an under-accumulation for arginine under 19C CL only (**Fig. 5**).

Finally, specific differences were observed in the metabolite accumulation patterns observed in cpEnolase and cpPGAM1A knockout lines (**Fig. 5; S13**). These include a significant (one-way ANOVA two-tailed P-value < 10^{-05}) over-accumulation of three amino acids (aspartate, leucine and phenylalanine) and one sugar phosphate

(glycerol-3-phosphate) specifically in cpEnolase knockout lines under all three conditions, and in serine under 19C CL and 8C CL only. These differences contrast to cpPGAM1A knockouts in which no significant changes were observed, or (in the case of glycerol-3-phosphate and serine) these metabolites significantly under-accumulate under all three conditions compared to controls **(Fig. 5**; **S13**).

Temperature-specific impacts of plastid glycolysis-gluconeogenesis on lipids

Next, we performed GC-MS (55 samples) and LC-MS (49 samples) of lipid profiles in multiple knockout and control lines under 19C LD, 19C CL and 8C CL. Outputs are tabulated in **Supplemental Dataset 6**, sheets 1, 3-5. While the GC-MS data project significant (one-way ANOVA two-tailed P-value < 0.05) impacts of growth condition on fatty acid profiles (e.g., a decrease of C20:5 side chain lipids balanced by an increase of C16:1 side chain lipids in 19C CL, and an over-accumulation of C16:3 side chain lipids under 19C LD, and of C18:0 side chain lipids under 8C CL), no substantial differences were observed between cpEnolase, cpPGAM1A and control lines under any conditions studied (**Supplemental Dataset 6,** sheet 3).

In contrast to the relatively limited effects on total fatty acid profiles, LC-MS analyses of lipid class distributions revealed substantial changes in lipid class distribution in plastid glycolysis-gluconeogenesis knockout lines (**Fig. 6**; **Supplemental Dataset 6,** sheet 4). Even accounting for within-line variation, both cpEnolase and cpPGAM1A knockouts were found to significantly under-accumulate triacylglycerols (TAG) (cpEnolase 3.98 ± 1.94%, cpPGAM1A 3.60 ± 1.72%, control 12.18 ± 7.26%; one-way ANOVA, two-tailed P separation of means between knockout and control lines < 0.05) and over-accumulate monogalactosyldiacylglycerols (MGDG; cpEnolase 63.83 ± 4.33%, cpPGAM1A 60.89 ± 5.64%, control 49.68 ± 8.88%; one-way ANOVA, two-tailed P< 0.05) under 19C LD (**Fig. 6A**). Further significant (P < 0.05) under-accumulations were detected in knockout lines for diacylglycerols (DAG) and sulfoquinovosyl-diacylcerols (SQDG) under 19C LD. Similar tradeoffs were observed under 19C CL, albeit with an over-accumulation, rather than under-accumulation of DAG, and an additional under-accumulation of digalactosyldiacylglycerols (DGDG), in glycolysis knockouts compared to control lines (**Fig; 6B**).

Detailed analyses of the individual fatty-acid side-chains associated with different lipid classes in glycolysis knockout lines under 19C indicated increased relative contributions of C16:1 fatty acids to plastid membrane lipid *sn*-1 positions (**Supplemental Dataset 6,** sheet 5). These included conserved ($P < 0.01$) over-accumulations of DGDG-16-1_16-2 under 19C LD **(Fig. S14)**; and SQDG 16-1_16-0, MGDG-16-1_16-2, MGDG-16-1_16-3 and DGDG-16-1_16-1, in both cpEnolase and cpPGAM1A knockout lines under 19C CL (**Fig. S15**). A significant over-accumulation of 16-1_16-1 side chains and under-accumulation 20-5_18-4 was also observed for diacylglyceryl hydroxymethyltrimethyl-*β*-alanine (DGTA), a betaine lipid known to act as a platform for the biosynthesis of 20:5 fatty acids, in both cpEnolase and cpPGAM1A knockout lines under 19C LD (**Fig. S14**) (Dolch and Maréchal 2015, Popko, Herrfurth et al. 2016).

Under 8C CL, quite different trends were observed in fatty acid accumulation in cpEnolase knockouts compared to cpPGAM1A knockouts and controls. These correlated principally with an over-accumulation of TAG (cpEnolase 20.88 ± 12.21%, cpPGAM1A 9.62 ± 6.31%, control 8.15 ± 3.95%; one-way ANOVA, two-tailed $P < 0.05$) in lieu of MGDG (cpEnolase 34.20 ± 6.74%, cpPGAM1A 42.94 ± 6.01%, control 46.61.3 ± 6.25%; one-way ANOVA, two-tailed $P < 0.5$; **Fig. 6A**). An over-accumulation of SQDG was observed in both cpEnolase and cpPGAM1A knockouts compared to controls, albeit with greater severity in cpEnolase knockouts (**Fig. 6B**). Considering side-chain distributions of individual lipid classes, a significant (one-way ANOVA two-tailed P-value < 0.01) over-accumulation of short-chain (C14:0, C16:1) and *sn*-1 and *sn-2* fatty acids was observed in cpEnolase knockouts (**Fig. S16A**). A probable exchange of very long-chain *sn*-2 fatty acids in SQDG pools was further observed in cpEnolase knockouts, with significant (one-way ANOVA two-tailed P-value < 0.01) increases in SQDG 14-0_16-0 and SQDG-14_0-16-1 in lieu of SQDG-16-2_24-0 in cpEnolase knockouts compared to cpPGAM1A and control lines (**Fig. S16B**; **Supplemental Dataset 6**, sheet 5).

Lower half diatom plastid glycolysis-gluconeogenesis enzymes are reversible, but with greater activity in the glycolytic direction

Finally, we assessed the kinetics of cpPGAM and cpEnolase in both glycolytic and gluconeogenic directions using a previously defined assay (Sutherland, Posternak et al. 1949, Zhang, Sampathkumar et al. 2020) with modified versions of each protein (codon-optimised, and lacking signal peptides) expressed in *E. coli,* alongside measured NADH consumption coupled to either lactate dehydrogenase (glycolysis) or glyceraldehyde-3-phosphate dehydrogenase (**Fig. S16**). Both enzymes were inferred to possess reversible reaction kinetics, metabolizing NADH when supplied both with 3-PGA (in the glycolytic direction) and PEP (in the gluconeogenic direction; **Fig. 7A, Fig. S16**). Nonetheless, the measured reaction rates were almost 1.5 times greater in the glycolytic than gluconeogenic direction (e.g., 34.4 versus 23.4 units of relative NADH consumption when supplied with 9 mM 3PGA or 9 mM PEP; **Fig. 7A**) suggesting an innate preference for glycolytic activity.

Discussion

We characterise a lower half glycolytic-gluconeogenic pathway associated with diatom plastids, relating specifically to two plastid-targeted proteins, cpEnolase and cpPGAM1A, and focusing on the model species *P. tricornutum*. Our data position plastid glycolysis-gluconeogenesis as arising in a recent ancestor of diatoms and their closest relatives (e.g., pelagophytes, dictyochophytes) and potential endosymbiotic derivatives (haptophytes) (Nonoyama, Kazamia et al. 2019). We further show that lower half plastidial glycolysis-gluconeogensis has a limited distribution across the algal tree of life, with no detectable examples in primary red or green algae (**Fig. S5**), although it is possible that the occurrence of organelle-targeted isoforms of these enzymes is underestimated, e.g., due to lower sensitivity of diatom and plant-trained targeting predictors on other algal groups (Fuss, Liegmann et al. 2013, Gruber, Rocap et al. 2015). We propose that diatom plastid glycolysis most likely originated through the duplication and retargeting of mitochondrial respiratory enzymes (**Fig. 1**).

Using meta-genomic data from *Tara* Oceans we demonstrate that diatom plastid glycolysis is likely highly expressed at high latitudes (**Figs. 2, S6-S8**), which are subject to extreme photoperiods and low temperature. These data are further supported by collection sites of cultured species, with no occurrences of species

lacking plastid-targeted PGAM enzymes beyond 50°N (**Fig. S1B**). These enrichments appear to be specific to diatoms, with polar circle haptophytes, cryptomonads and other ochrophytes lacking apparent plastidial glycolysis found further than 60°N and 70°S, and a restriction of plastid glycolysis to low latitude green algae. These results will be better substantiated by considering meta-genome assembled genomes from uncultivated diatoms with known spatial distributions, or by competition assays of closely related diatom species possessing and lacking plastid glycolysis enzymes under light and temperature stress (Siegel, Baker et al. 2020, Delmont, Gaia et al. 2022).

From growth analysis of *P. tricornutum* knockout lines, we infer that photoperiod is the principal driver of diatom plastid glycolytic activity, with more intense growth defects observed under continuous illumination than in light:dark cycles (**Figs. 3, S9**). It has been proposed that the effective management of primary carbon metabolism may enable diatoms to outcompete other phytoplankton groups under nutrient-rich and light-saturated conditions (Behrenfeld, Halsey et al. 2021, Lampe, Hernandez et al. 2021). Under low temperatures, no difference was observed in the growth rate of glycolysis knockouts showed to control lines (**Fig. 3**). This may reflect previous studies that have projected low rates of diatom photo-assimilation of carbon dioxide in low temperature habitats at high latitudes (Lacour, Larivière et al. 2017). The relevance of plastidial glycolysis to diatoms in other environmental conditions where they predominate (e.g., coastal and upwelling cycles) remains to be determined (Lampe, Hernandez et al. 2021).

Considering the observed phenotypes of knockout and control lines (**Figs. 3-6; S10-S15**) and the reversible kinetics of expressed enzymes, we reconstruct potential functions contributed by the lower half of plastid glycolysis-gluconeogenesis in diatoms under 19C LD, 19C CL and 8C CL (**Fig. 7B)**. Overall, these data seem predominantly in favour of metabolic flux in the glycolytic direction, perhaps reflecting the greater relative abundance of triose phosphate than pyruvate in the plastid under illuminated and photosynthetically active conditions, alongside the greater kinetic activity observed in the glycolytic than gluconeogenic activity of expressed enzymes (**Fig. 7A, Fig. S16**). These results are inferential based on the long-term accumulation patterns of stable metabolites and the expression of implicated

643 metabolic genes, and would be more effectively validated via direct flux
644 measurements, e.g., comparative ^{13}C-glycerol or -glucose labelling of glycolysis
645 knockout and control lines (Zheng, Quinn et al. 2013, Huang, Liu et al. 2015).
646
647 Under 19C LD, we observe limited gene expression changes in cpPGAM1A and
648 cpEnolase knockout lines, except (as inferred from qPCR) a downregulation in
649 plastid chorismate mutase and upregulation of mitochondrial PGAM (**Fig. 4B**).
650 Chorismate mutase forms part of the plastid shikimate pathway, which typically
651 consumes PEP (Bromke 2013) and may form a primary acceptor of glycolytic
652 products, whereas upregulation of mitochondrial glycolysis-gluconeogenesis may
653 reflect an enhanced export of plastid glyceraldehyde-3-phosphate to the
654 mitochondrion in the absence of plastid metabolism. We also note diminished levels
655 of cytoplasmic sugars and TCA cycle intermediates (i.e., citric acid) in cpEnolase and
656 cpPGAM1A knockout lines (**Fig. 5**) (Smith, Dupont et al. 2019), which might suggest
657 a diversion of cellular carbon away from gluconeogenesis and/or respiration.
658
659 We also note some evidence for lipid remodelling in glycolysis mutant lines, e.g., a
660 relative over-accumulation of galactolipids in lieu of TAGs, and short-chain fatty acids
661 in lieu of longer equivalents (**Figs. 6, S14**). Previous studies have noted the
662 importance of lipid metabolism in diatom stress responses (Zulu, Zienkiewicz et al.
663 2018), and that most or all diatom lipid synthesis occurs directly in the plastid
664 (Huang, Pan et al. 2023). Many of the metabolic reactions required for lipid synthesis,
665 including acyl-coA synthesis from pyruvate (Maréchal and Lupette 2020), glycerol-3-
666 phosphate from glyceraldehyde-3-phosphate (Kroth, Chiovitti et al. 2008), and
667 glucosyl-1-phosphate from cytoplasmic glucosyl-1-phosphate (Zhu, Shi et al. 2016),
668 are likely to be Impacted by plastid carbon metabolism. We note that there is no
669 evidence for transcriptional modifications to fatty acid synthesis in glycolysis mutant
670 lines, suggesting that these changes are driven by substrate limitation. As our
671 metabolomics data project under-accumulations of both pyruvate and hexose sugars
672 in knockout lines (**Fig. 5**), we propose that the diminished abundances of TAGs and
673 shorter chain fatty acids are due to a diminished supply of plastid pyruvate for acyl-
674 coA synthesis, rather than greater gluconeogenesis of galactosyl-phosphates from
675 exported plastid glyceraldehyde-3-phosphate (Demé, Cataye et al. 2014).
676

Under 19C CL, we observed much more dramatic remodelling of cellular transcription in knockout lines compared to controls (**Fig. 4A**). These include greater overall photosynthesis gene expression, e.g., *Lhcf1* (**Fig. 4B**), which was corroborated in photo-physiological analyses by larger PSII antenna size, i.e., a larger functional cross-section (σPSII) (**Fig. S11**). It should be noted that the increase in PSII antenna size does not necessarily change the quantum yield of individual PSII reaction centres, and therefore the increased σPSII is independent of the Fv/ Fm measured, which remains equivalent between knockout and control lines (**Fig. S11**). We did not observe consistent differences in the expression of nitrogen or phosphorus stress metabolism, or in the expression of the *P. tricornutum* biophysical carbon concentration mechanisms of knockout lines, suggesting that these differences were not caused by N, P or CO_2 limitation in the control lines (**Supplemental Dataset 5,** sheets 4-5) (McCarthy, Smith et al. 2017, Nawaly, Matsui et al. 2023). We further did not measure differences in photosynthetic performance (electron transport), or an upregulation of genes encoding proteins involved in photoprotection, e.g., LhcX family or xanthophyll cycle enzymes in knockout lines under 19C CL (**Fig. 4B**; **Fig. S11**; **Supplemental Dataset 4,** sheet 12) (Buck, Sherman et al. 2019, Bai, Cao et al. 2022), suggesting that the differential expression of photosynthesis genes in the knockout lines does not directly influence photosynthesis. In contrast, we observed an upregulation of mitochondrial NDH dehydrogenase and downregulation of TCA cycle enzymes in glycolysis knockout lines (**Fig. 4B**). Previous studies have noted the important role of diatom mitochondria in dissipating excess plastid reducing potential (Bailleul, Berne et al. 2015, Broddrick, Du et al. 2019), and it remains to be determined if the formal export of plastid NADPH to the mitochondria in knockout lines is impeded under continuous light.

It remains to be determined what routes enable the supply of pyruvate hub intermediates to the *P. tricornutum* plastid in the absence of plastid glycolytic activity. Previous studies have noted that diatom plastid triose phosphate transporters may be able to transport PEP directly from the cytoplasm, and one of these (Phatr3_J54017) is indeed upregulated in both cpEnolase and cpPGAM1A knockout lines under 19C CL (**Supplemental Dataset 5**, sheet 3) (Moog, Nozawa et al. 2020, Liu, Storti et al. 2022). Elsewhere our data suggest that amino acids may also contribute to resupplying the *P. tricornutum* plastid with PEP and/or pyruvate. The

overexpression of plastid alanine transaminase in cpPGAM1A knockouts (**Fig. 4B**) might allow alanine imported into the plastid to be deaminated to supply pyruvate (Smith, Dupont et al. 2019). The overaccumulation of small amino acids synthesised either from pyruvate (valine), PEP (aspartate via PEP carboxylase) and its derivatives (methionine from aspartate via the homocysteine pathway) in knockout lines may relate to this (**Fig. 5**) (Bromke 2013, Yu, Nakajima et al. 2022). Strikingly, the knockout lines over-accumulate ornithine and glutamate, implicated in diatom plastid amino acid recycling via the ornithine cycle (Levering, Broddrick et al. 2016, Smith, Dupont et al. 2019), but under-accumulate arginine and citrate produced via mitochondrial ornithine catabolism in the urea cycle (Allen, Dupont et al. 2011, Dolch, Lupette et al. 2017). These data broadly suggest dysregulation in plastid-to-mitochondria amino acid shuttling in glycolysis knockout lines under 19C CL (**Fig. 5**).

Under 8C CL, we identify an over-accumulation of mRNAs encoding plastid biogenesis and mitochondrial glycolytic proteins, an over-accumulation of short-chain amino acids (valine) and an under-accumulation of cytoplasmic sugars and amino acids (glucose, histidine) in cpEnolase and cpPGAM1A knockouts relative to controls (**Figs. 4B, 5**). We further note under-accumulations of pyruvate in both knockout lines (**Fig. 5**). These reflect the phenotypes identified at 19C, and the absence of a specific growth defect under 8C conditions is therefore perhaps surprising (**Fig. 3**). Knockout lines under 8C conditions, however, have additional phenotypes not observed at 19C. These include an overall enrichment in down-regulated genes encoding plastid-targeted proteins (**Fig. 4A**); a specific over-accumulation of TCA cycle (citrate synthase) and a possible non-photochemical quenching-associated mRNA (*LhcX4*) (**Fig. 4B**) (Bailleul, Berne et al. 2015, Murik, Tirichine et al. 2019). It is possible that these additional phenotypes have compensatory effects on knockout line growth rates.

One possible explanation for the more complex phenotypes observed under 8C CL would be a gluconeogenic activity of plastidial lower-half glycolysis-gluconeogenesis. The reversibility of the cpPGAM1A and cpEnolase reaction is confirmed by enzymatic data (**Fig. 7A**), and in particular is suggested by the cpEnolase knockout phenotypes, which could relate to the accumulation of PEP on the acceptor side of the gluconeogenic reaction. These include an overaccumulation of TAGs and

SGDQs over glucosyl-lipids, which might relate to a greater accumulation of acyl-coA synthesised from plastidial pyruvate in cpEnolase knockouts (**Fig. 6**). These also include the over-accumulation of aspartate and phenylalanine, synthesised from PEP, which is further visible in cpEnolase knockouts at 19C conditions (**Figs. 5, S13**). The differences in cpEnolase and cpPGAM1A phenotypes may relate to the fact cpEnolase is the only plastid-targeted enolase protein encoded in the *P. tricornutum* genome, whereas alternative phospho-glycerate mutases exist that may compensate for cpPGAM1A (e.g., cpPGAM1B, cpPGAM2; **Figs. S4, S7**). Establishing the relative fluxes that occur in each direction of cpEnolase and cpPGAM enzymes will depend on the construction of multigene knockouts for different *P. tricornutum* Enolase and PGAM isoforms (Yin and Hu 2023).

The complex phenotypes for diatom plastid glycolysis inferred from environmental data and experimentally visualised in *P. tricornutum* knockout lines contrast with those for plant plastid glycolysis, with (for example) *A. thaliana* cpEnolase and cpPGAM mutants presenting relatively limited phenotypes (Prabhakar, Löttgert et al. 2009, Andriotis, Kruger et al. 2010). These contrast with cytoplasmic and respiratory plant Enolase and PGAM1 isoforms which, alongside having predominant impacts on plant carbon flux, may have important moonlighting roles in plant development, immune responses and even in the structural coordination of plastids and mitochondria (Zhao and Assmann 2011, Zhang, Sampathkumar et al. 2020, Yang, Wang et al. 2022). We similarly anticipate that further surprises will be identified for the functions of diatom plastid glycolysis, and for this still poorly understood pathway in the photosynthetic tree of life.

Materials and Methods

Culture conditions

Phaeodactylum tricornutum strain Pt1.86 was grown in enhanced seawater (ESAW) medium supplemented with vitamins, but without silicon or added antibiotics, in 50 µE m^{-2} s^{-1} white light. Light profiles were measured with a SpectraPen photofluorometer (Photon Systems Instruments, Czech Republic); and are provided in **Supplemental Dataset 4**, sheet 13. Cultures were grown under one of four light, temperature and

779 shaking regimes. For general molecular work and transformation, cultures were
780 grown under 19 °C with 12h light: 12 dark cycling, shaken at 100 rpm (for general
781 molecular work and transformation), following the established methodology of
782 Falciatore et al. (Falciatore, Casotti et al. 1999). For comparative physiology work,
783 we were unable to replicate shaking conditions at low temperatures, and therefore
784 chose to use conditions without shaking: 19 °C with 12 h light: 12 h dark cycling
785 (« LD » growth conditions and physiological analysis); 19 °C with 24 h continuous
786 light and without shaking (« CL » growth conditions and physiological analysis); or
787 8°C with 24h continuous light and without shaking (« 8C » growth conditions and
788 physiological analysis). All cultures achieved measured mid exponential Fv/Fm
789 values of > 0.6, suggesting that the absence of shaking did not impact on
790 photosynthetic efficiency (**Supplemental Dataset 5**, sheet 8).
791
792 Batch culturing of *P. tricornutum* for genetic manipulation was performed under
793 fluorescent lamps. Physiological experiments were principally performed at 19°C in
794 an AlgaeTron AG230 (Photon Systems Instruments) with cool white LED (WIR)
795 illumination, and technical specifications described in https://growth-
796 chambers.com/data/algaetron-ag-230/download/AlgaeTron_AG_230_Manual2021-
797 finalweb.pdf . Growth experiments were performed at 8°C using a low-temperature
798 adapted cool white LED (WIR, ECCLIM). Details of all three spectra used, as
799 measured with a SpectraPen (PSI), are provided in **Table S4**, sheet 13.
800
801 Mutant *P. tricornutum* lines were maintained on ½ ESAW 1% agarose plates,
802 supplemented by 100 µg ml^{-1} each ampicillin and streptomycin, and 30 µg ml^{-1}
803 chloramphenicol, and either 100 µg ml^{-1} zeocin (single transformants), or 100 µg ml^{-1}
804 zeocin and 4 µg ml^{-1} blasticidin (complementation lines), as previously described
805 (Falciatore, Casotti et al. 1999, Buck, Río Bártulos et al. 2018). All functional
806 analyses of transformant lines were performed on transformant lines grown in the
807 absence of antibiotic selection, to avoid secondary effects on growth or physiology.
808
809 *Phylogenetic identification of plastid lower half glycolysis-gluconeogenesis enzymes*
810
811 Plastid-targeted glycolysis lower half enzymes were searched across 1,673 plant and
812 algal genomes and transcriptomes (**Dataset S1**, sheet 1). Briefly, this involved all

annotated *P. tricornutum* PGAM (Phatr3_J17086, Phatr3_J51404, Phatr3_J5605, Phatr3_J5629, Phatr3_J8982, Phatr3_J37201, Phatr3_J47096) and enolase (Phatr3_draftJ1192, Phatr3_draftJ1572, Phatr3_J41515) peptide sequences with BLASTp and a threshold e-value of 10^{-05}, and a reciprocal BLASTp with criteria -max_target_seqs 1 to retrieve the best homologues against the entire *P. tricornutum* genome. For PGAM, where *P. tricornutum* queries failed to retrieve homologues in >50% searched libraries, a second BLASTp was performed with query peptide sequences from *A. thaliana* (AT2G17280, AT1G09780, AT3G05170, AT3G08590, AT3G50520, AT5G04120, AT5G64460), and a reciprocal BLASTp was performed with the *P. tricornutum* genome supplemented with these sequences.

The domain content of each potential homologue was identified using hmmscan and the version 33.1 Pfam database (Mistry, Chuguransky et al. 2020). Only Enolase sequences that contained >90% predicted domain coverage to both Enolase_N and Enolase_C domains; and PGAM sequences that contained >50% domain coverage to the His_Phos domain (based on the corresponding coverage observed in *P. tricornutum* sequences) were viewed as being complete. Sequences for which the N-terminus of the region homologous to the PFAM domain was located within the first 20 aa of the predicted sequence (i.e., less than the length of a typical plastid-targeting sequence) (Emanuelsson, Brunak et al. 2007) were viewed as lacking credible targeting sequences. All remaining proteins were scanned, considering both complete proteins and sequences trimmed to the first encoded N-terminal methionine, using targetp (using a plant scoring matrix) (Emanuelsson, Brunak et al. 2007), PredAlgo (Tardif, Atteia et al. 2012), HECTAR (Gschloessl, Guermeur et al. 2008) and ASAFind (with SignalP 5.0) (Gruber, Rocap et al. 2015, Almagro Armenteros, Tsirigos et al. 2019). Sequences from primary plastid-containing organisms (plants, green and red algae, glaucophytes) that were inferred to possess a plastid-targeting sequence either with TargetP or PredAlgo, and sequences from secondary plastid-containing organisms that were inferred to possess a plastid-targeting sequence with either HECTAR or ASAFind, considering both complete and N-trimmed sequence models, were annotated as putatively plastid-targeted.

A more detailed phylogenetic analysis was performed using Enolase and PGAM homologues obtained from a subset of 289 complete cryptomonad, haptophyte and

847 stramenopile genomes and transcriptomes in the above library, alongside
848 homologues identified from a further 85 prokaryotic and eukaryotic genomes
849 sampled with taxonomic balance from across the remainder of the tree of life (Liu,
850 Storti et al. 2022). Sequences were also screened for mitochondrial presequences
851 using HECTAR (Gschloessl, Guermeur et al. 2008), and MitoFates, run with
852 threshold value 0.35 (Fukasawa, Tsuji et al. 2015).
853
854 The pooled set of sequences were aligned first with MAFFT v 7.0 under the --auto
855 setting, followed by the in-built alignment programme in GeneIOUS v 10.0.9, under
856 default settings (Kearse, Moir et al. 2012, Katoh, Rozewicki et al. 2017). Incomplete
857 and poorly aligned sequences, alongside taxonomically uninformative N- and C-
858 terminal regions were removed from the alignment manually, followed by trimal with
859 setting –gt 0.5 (Capella-Gutiérrez, Silla-Martínez et al. 2009). Phylogenetic analyses
860 were performed with MrBayes v 3.2 and rAxML v 8, integrated into the CIPRES
861 webserver (Stamatakis 2014, Miller, Schwartz et al. 2015). MrBayes trees were run
862 for 10,000,000 generations with the GTR, Jones and WAG substitution matrices, 4
863 starting chains and sumt and sump burnin fractions set to -0.5; all cold chains were
864 confirmed to have reached a pvalue plateau below 0.1 prior to the start of the
865 consensus building. rAxML trees were run with GTR, JTT and WAG substitution
866 matrices, 350-400 ML generations, and automatic bootstopping. Phylogenies were
867 either rooted between bacterial and eukaryotic sequences (Enolase); or on the mid-
868 point (PGAM1, PGAM2) due to the absence of a single monophyletic bacterial
869 outgroup. A summary of these analyses is provided in **Supplemental Dataset 1**.
870
871 Tara *Oceans Analysis*
872
873 The complete *Tara* Oceans and *Tara* Oceans Polar Circle libraries of meta-genome
874 and meta-transcriptome diversity (Carradec, Pelletier et al. 2018, Royo-Llonch,
875 Sánchez et al. 2020) were searched for orthologues of diatom cpEnolase,
876 cpPGAM1A and PGAM2 sequences via a phylogenetic reconciliation approach
877 benchmarked in previous studies (Kazamia, Sutak et al. 2018, Liu, Storti et al. 2022).
878 This approach uses the combined outputs of hmmer, BLAST best-hit, and single-
879 gene tree topologies to only retain *Tara* Oceans meta-genes that reconcile as

monophyletic with a defined query set, in these case plastid-targeted diatom isoforms of each enzyme. Exemplar tree topologies are shown in **Fig. S6.**

First, a HMM (hidden Markov model) was constructed for all diatom plastid-targeted sequences in the untrimmed alignments for each phylogeny, as detailed above, and searched into the complete *Tara* Oceans catalog by hmmer (http://hmmer.org) with evalue 10^{-10} to identify putative meta-gene homologues of each protein. Matching sequences were extracted, and searched by BLASTp against the complete copy of the *P. tricornutum* genome (Rastogi, Maheswari et al. 2018). Only sequences that retrieved a best hit against an Enolase or PGAM sequence (and therefore likely correspond to homologues of each protein) were retained. Next, the retained sequences were similarly searched by BLASTp against the complete untrimmed alignment of cultured Enolase and PGAM sequences. Only sequences that retrieved a diatom plastid-targeted isoform were retained, allowing the elimination of non-diatom and homologues of diatom non-plastid sequences. Finally, sequences were combined with the untrimmed alignment of cultured sequences from each gene and realigned using the same MAFFT, GeneIOUS and trimal pipeline as defined above. Curated alignments were screened by rAxML tree with the JTT substitution matrix, as above. Only *Tara* Oceans sequences that resolved within a monophyletic clade with diatom plastid-targeted proteins, defined as all sequences that position closer on a midpoint rooting of the tree to diatom plastid-targeted proteins than to any non-diatom or non-plastid targeted sequences, was extracted for further analyses.

Relative abundances were calculated for the total occurrence of all phylogenetically verified diatom plastid-targeted proteins in both meta-transcriptome and meta-genome data. Relative expression levels of each gene were estimated by reconciling the calculated meta-transcriptome abundances either to total diatom meta-transcriptome sequences using the formula $10^{\wedge 6}(\Sigma_{metaT}/ \Sigma_{DiatomT})$, i.e., expressed per million reconciled diatom reads, or to calculated meta-genome abundances, using the formula and $\log_{10}(1+ \Sigma_{metaT}) - \log_{10}(1+ \Sigma_{metaG})$, to allow inclusion of null values. Pearson and Spearman correlations were calculated between relative abundances and all quantitative measured environmental variables associated with *Tara* Oceans samples as stored within the PANGAEA repository (Pesant, Not et al. 2015). All calculations were repeated independently for each depth (surface, or deep

chlorophyll maximum/ DCM) and size fraction (0.8- 2000 µm, 0.8- 5 µm, 3/5- 20 µm, 20- 180 µm, and 180- 2000 µm), with 3 and 5 µm filters viewed as equivalent to allow reconciliation of Arctic and non-Arctic data, respectively. All *Tara* Oceans meta-gene assignations, alongside individual and total abundance calculations are provided in **Supplemental Dataset 3**.

Nucleic acid isolation

For DNA isolation, 150 ml early stationary phase *P. tricornutum* culture, grown under 19°C with 12h light: 12h dark cycling, and shaken at 100 rpm as described above, was centrifuged at 4000 rpm for 10 minutes. The resulting cell pellet was washed in sterile growth medium three times, and incubated for four hours in 5 ml TEN buffer (0.1 M NaCl, 0.01 M Tris pH8, 0.001 M EDTA) supplemented with 2% volume: volume SDS, and 1U μl^{-1} proteinase K (Fisher Scientific). Lysed cell fractions were used for phenol: chloroform precipitation of cellular DNA, as previously described (Nash, Barbrook et al. 2007), prior to dissolution in 50 µl nuclease-free water, and quantification with a nanodrop photospectrometer.

For RNA isolations, 10^5 stationary phase *P. tricornutum* cells, as calculated with cell densities counted from a Malassez haemocytometer were inoculated in a 250 ml conical Erlenmeyer containing 80 ml ESAW without antibiotics. Cell cultures were harvested in mid-exponential phase, at counted densities of between 1 and 2 x 10^6 cells ml^{-1}. 19C CL cultures were typically harvested eight days post-inoculation, 19C LD cultures nine days post-inoculation, and 8C CL cultures seventeen days post-inoculation, in agreement with growth curve dynamics. Cells were harvested at the mid-point of the light-induction phase of the LD growth condition (15:00 CET), per previous gene expression studies in *P. tricornutum* (Cruz de Carvalho, Sun et al. 2016).

RNA was isolated from 10^8 cells from each culture, pelleted and washed as before, and snap-frozen in liquid nitrogen. Frozen cell suspensions were re-equilibrated with 1 ml Trizol reagent (Invivogen) and 200 µl chloroform (Honeywell), prior to phenol: chloroform extraction. An additional separation step was performed in 500 µl pure chloroform to remove any residual phenol traces from the aqueous phase, and

purified nucleic acids were precipitated overnight in 500 µl isopropanol at -20°C. RNA was collected by centrifugation at 10,000 rpm for 30 minutes, washed with 900 µl 100% ethanol, and resupended in 50 µl RNAse-free water (Qiagen).

2 µg RNA, as quantified with a nanodrop photospectrometer, was subsequently treated with 2U RNAse-free DNAse (Promega) for 30 minutes at 37 °C, with the reaction arrested with 1 µl 0.5 M EDTA. The digested RNA sample was reprecipitated in isopropanol for one hour at -20 °C, washed in ethanol, and resuspended in 20 µl RNAse-free water. Purified RNA sample concentrations were quantified with a nanodrop spectrometer, and a 3 µl aliquot was migrated on an RNAse-free 1% agarose gel stained with 0.2 µg ml^{-1} ethidium bromide to confirm RNA integrity prior to all downstream applications.

GFP localization

Full length mRNA sequences of cpEnolase, cpPGAM1A and cpPGAM2 were amplified from *P. tricornutum* RNA libraries grown under 19 °C, light: dark cycling and replete nutrient conditions as described above, by reverse transcription with RT Maxima First Strand synthesis kit (Thermo Fisher) from 200 ng template RNA, following the manufacturer's instructions; and gene-specific primers as shown in **Supplemental Dataset 2**, sheet 4. PCRs were performed using Pfu high-fidelity DNA polymerase, in 50 µl total reaction volume, including 1 µl cDNA template and 2 µl each specific primer, following the manufacturer's protocol. Typical PCR conditions were: 10 minutes at 95 °C; followed by 35 cycles of 45 seconds at 95 °C, 45 seconds at 55 °C, and 2 minutes at 72 °C; followed by a terminal elongation phase of 5 minutes at 72 °C. Amplified products were migrated on a 1% agarose gel stained with ethidium bromide, cut out, and purified using a MinElute PCR purification kit (Qiagen).

Purified products were cloned into linearised versions of pPhat vectors containing eGFP and a zeocin resistance gene (SHBLE). These products were amplified using an analogous PCR protocol as above, with 1 ng purified plasmid DNA, and outward-directed PCR primers extending from the start of the fluorescence protein gene sequence to the end of the FcpA promoter region (**Supplemental Dataset 2**, sheet

4); cut, purified, and treated with 1U FastDigest DpnI (Thermo Fisher) to remove any residual plasmid DNA. The gene-specific primers for each cpEnolase and cpPGAM construct were modified with 15 5' nucleotides complementary to the terminal regions of the FcpA and GFP sequences, allowing cloning of complete vector sequences using a HiFi DNA assembly kit (NEB), following the manufacturer's instructions. Assembled vectors were transformed into chemically competent Top10 *E. coli*, and positive clones (as identified by Sanger sequencing of positive colony PCR products) were used to generate purified plasmid DNA with a Plasmid Midi Kit (Qiagen).

Subcellular localization constructs were transformed into wild type *P. tricornutum* Pt186 by biolistic transformation, as previously described (Falciatore, Casotti et al. 1999). 5 x 10^7 mid-exponential phase cells were plated on a ½ ESAW- 1% agarose plate, and left to recover for two days, prior to bombardment with 10 mg 1 µm tungsten beads treated with 5 µg purified plasmid DNA in a Helios gene gun (BioRad) at 1,550 PSI. Cells were left to recover for two days, prior to replating on ½ ESAW- 1% agarose plates supplemented with 100 µg ml^{-1} ampicillin, 100 µg ml^{-1} streptomycin, 30 µg ml^{-1} chloramphenicol and 100 µg ml^{-1} zeocin. Plates post-bombardment and for the first two days post-selection were maintained in a low light environment (10 µE m^{-2} s^{-1}) prior to return to standard growth conditions.

Positive transformant colonies, as verified by Western Blot with a mouse anti-GFP antibody (Thermo Fisher), were visualised using a SP8 inverted spinning disc confocal microscopy (Leica) under 400 x magnification, with excitation wavelength 485 nm and emission wavelength filters 500-550 nm. GFP-negative colonies were used to confirm detection specificity, and empty-vector GFP (with cytoplasmic localizations) were used as fluorescence positive controls. A minimum of 12 GFP expressing clones were visualised for each construct with consistent localization.

CRISPR mutagenesis

CRISPR target sequences for cpEnolase and cpPGAM1A genes were identified using PhytoCRISP-Ex (Rastogi, Murik et al. 2016), privileging positions in the N-terminal region of the conserved domain to minimize the probability of enzyme functionality in knockout lines, and uniqueness across the entire *P. tricornutum*

genome within the final 11 bp of the target sequence to minimize off-target effects. Primers were designed for each target sequence, and introduced into a pu6:SG1 CRISPR target sequence plasmid by PCR, as previously described (Nymark, Sharma et al. 2016). 2 µg insertion-positive pu6:SG1 plasmids, as confirmed by Sanger sequencing were co-introduced into wild type *P. tricornutum* Pt186 cells by bombardment along with 2 µg HA-tagged Cas9 and pPhat vectors, as described above. Mutant colonies were genotyped using a DNA lysis buffer containing 0.14 M NaCl, 5 mM KCl, 10 mM Tris-HCl pH 7.5, 1% v/v NP40 to generate crude DNA extracts, followed by PCR amplification across the CRISPR target sequences with DreamTaq PCR reagent (Promega) and Sanger sequencing (Eurofins genomics). Mixed mutant: wild-type colonies were segregated via repeated dilution on ESAW: zeocin plates until only homozygous mutant genotypes were detected (Nymark, Sharma et al. 2016, McCarthy, Smith et al. 2017). Empty vector control lines were generated using the same protocol, except with only HA-Cas9 and pPhat plasmids, cotransformed without a CRISPR target sequence.

Tabulated cleaned knockout mutants, their associated genotypes and the expression levels of mutated gene copies are shown in **Supplemental Dataset 4**, sheets 1-2. Mutant colony genotypes were periodically confirmed (approx. once per month) by PCR and Sanger sequencing throughout the duration of all subsequent experiments, and the CRISPR-induced gene modifications were found to remain stable. *P. tricornutum* Enolase proteins were determined by Western blot to be crossreactive to an anti-*Arabidopsis thaliana* Enolase-2 antibody (Agrisera), and thus knockout line protein expression was confirmed by qRT-PCR, as described below.

Complementation of knockout lines

Knockout lines were complemented with pPhat:GFP vectors carrying overexpressing copies (under an FcpA promoter) of cpEnolase and cpPGAM1A synthetic constructs, with all CRISPR target sequences replaced with silent mutations (Eurofins). Genes were fused to C-terminal GFP, allowing the verification of protein expression and localization. Vectors were identical to those previously used for localization, but with a blasticidin S-deaminase gene in lieu of SHBLE (Buck, Río Bártulos et al. 2018) introduced by NEB Hi-Fi kit as before. Complementation constructs were transformed

via bombardment, and cotransformed colonies were selected on ½ ESAW- 1% agarose plates supplemented with 100 µg ml^{-1} ampicillin, 100 µg ml^{-1} streptomycin, 30 µg ml^{-1} chloramphenicol, 100 µg ml^{-1} zeocin, 4 µg ml^{-1} blasticidin.

For each complementation, three cpEnolase and cpPGAM1A knockout lines (including at least one for each CRISPR target sequence) were complemented both with the conjugate complementation construct, and an empty blasticidin resistance vector as a placebo; and two empty vector lines were further complemented with both cpEnolase and cpPGAM1A overexpressor constructs, plus an empty blasticidin resistance vector, to exclude possible effects from ectopic overexpression of each gene on cell physiology. A total of 47 colonies, with a minimum of 6 colonies for each knockout: complementation combination, including lines complemented from at least two distinct primary knockout mutant genotypes, were selected for subsequent study (**Supplemental Dataset 4**, sheet 7). The retention of the primary knockout mutant genotype in each complemented line was verified by colony PCR and sequencing as above, and the overexpression and correct localization of the complementing protein sequence (i.e., to the chloroplast for cpEnolase:GFP and cpPGAM1:GFP, or the cytoplasm for ectopic GFP) was verified by western blot with an anti-GFP antibody (Thermo Fisher) (Erdene-Ochir, Shin et al. 2019) and confocal microscopy.

Growth rate measurements

A starting density of 10^4 ml^{-1} stationary phase *P. tricornutum* cells of a given culture line, as verified with a Malassez haemocytometer, were inoculated into a 15 ml volume antibiotic-free ESAW within a sterile, ventilated cap plastic culture flask (Celltreat) and grown under LD, CL, or 8C culture conditions as described. Cell densities were recorded: very day from one day post-inoculation (CL); every day from two days post-inoculation (LD); or every two days from five days post-inoculation (8C) at the mid-point of the LD light induction phase using a counting CyFlow Cube 8 cytometer (ParTec).

Typically, 15 µl cell culture, passed at 0.5 µl s^{-1}, were used for each measurement, with three technical replicates performed for each culture of which the first (enriched in non-cellular particles) was excluded from downstream calculations. Cytometer

particle counts were correlated to actual cell densities using a calibration curve realised from haemocytometer counted densities of wild-type cell culture, and cultures with observed densities > 2 x 10^6 cells ml^{-1} were diluted ten-fold in blank growth media to avoid saturation of the cytometer.

Cell densities were measured daily from one day post-inoculation (19C conditions) and every second day from five days post-inoculation (8C CL only), until cell lines were confirmed to have exited log phase (i.e., reached a stationary phase plateau). Primary knockout mutant growth curves were repeated a minimum of six times (three biological replicates per-inoculation, with two independent repetitions) for each mutant line. Growth curves were tested for seven cpEnolase knockout, five cpPGAM1A knockout and four empty vector control lines, providing a minimum of 24 measurements (i.e., four distinct mutant lines) per each genotype studied (cpEnolase knockout, cpPGAM1A knockout and empty vector control lines).

To avoid artifacts based on the proximity of the seed cell culture to exponential phase at the time of inoculation (which may impact on lag phase length) or the relative diameter of each cell in culture (which may impact on carrying capacity), cell growth rates were measured exclusively from the log-phase relative division rate. This was calculated via considering $\Delta\log_2$ (cell density) / $\Delta\log_2$ (time) for a time period corresponding to 5 x 10^4 to 4 x 10^6 cells/ ml, covering in most cases six successive measurements of each individual growth curve. To confirm that the cells were measured in exponential phase, and were neither influenced by particular contamination of the cytometer or cell exhaustion of the growth medium, the linear correlation was calculated between the log value, with most calculated correlations (129/ 132) showing linearity (r> 0.95). Three exemplar growth curve outputs are provided in **Supplemental Dataset 4**, sheets 3-5; and an overview of relative growth rates expressed as a function of mean empty vector control growth rates are provided in **Supplemental Dataset 4**, sheet 6.

Complementation growth curves were repeated with at least two independent repetitions for each cell line, with five timepoints taken to project growth rates, and therefore a minimum of sixty independent measurements for each mutant: complementation genotype under each growth condition, with the average of the two

fastest growth rates of each culture calculated as estimates for the growth rate. A heatmap of all estimated complementation line growth rates is provided in **Supplemental Dataset 4**, sheet 7.

Photophysiology

Cultures for photophysiological analysis were grown in 10ml ventilated plastic flasks, without shaking, under 19C CL and 19C LD as described above. Cultures were grown from a starting inoculum of 10^5 cells ml^{-1} as measured with a Malassez haemocytometer. Cell cultures that had reached a measured density of 10^6 cells ml^{-1} were then refreshed into fresh media at the initial starting concentration of 10^5 cells ml^{-1} to allow a prolonged adaptation to each photophysiological condition under a continuous exponential phase. Cells from refreshed culture lines were harvested in exponential phase (between 1 and 3 × 10^6 cells ml^{-1}, and good physiology was verified by Fv/Fm measurements > 0.6 across all measured lines (**Supplemental Dataset 4**, sheet 8).

Steady-state light curves (SSLC) were conducted with a fluorescence CCD camera recorder (SpeedZen, jBeamBio, France) in a selected set of control lines (*n*=2), cpPGAM (*n*=3) and cpEnolase knockouts (*n*=6), as well in complemented cpEnolase (*n*=2) and cpPGAM1A (*n*=3) knockout lines in which we observed a suppression of the knockout growth defect compared to complemented control lines. Measurements were repeated a minimum of two and in most cases four times per line and treatment condition, with a minimum of six unique measurements performed for each genotype and treatment. Curves were measured on cell cultures concentrated to between 2 and 5 × 10^7 cells ml^{-1}. Samples were exposed to an initial 5 min illumination of 35 µmol photons m^{-2} s^{-1} green actinic light (532 nm), followed by a 6 steps of 3 min each of increasing intensity to 750 µmol photons m^{-2} s^{-1}.

Minimum (F_0) and maximum (F_M) fluorescence were measured in dark-adapted (at least 1 min) samples, before and at the end of a 250 ms saturating (multiple turnover) pulse of light (532 nm, 5000 µmol photons m^{-2} s^{-1}) and the maximum quantum yield of PSII in the dark was calculated as $F_V/F_M = (F_M - F_0)/F_M$. Every minute of light exposure, steady-state fluorescence (F_S) and maximum fluorescence under Light

(F_M') were measured. PSII quantum yield (φPSII) and nonphotochemical quenching (NPQ) were calculated on the last time point of each light step as φPSII = (F_M'-F_s)/F_M' and NPQ = F_M/F_M'-1, and rETR at PSII as rETR = φPSII.E.

The whole rETR vs E curve was fitted as rETR = $rETR_M$.(1-exp(-α.E/$rETR_M$)) where $rETR_M$ is the maximum rETR and α is the light-limited slope of rETR vs E (Jassby and Platt 1976). Only rETR values from 0 to 450 µmol photons m^{-2} were used for the fit because values from 600 and 750 µmol photons m^{-2} were too noisy. The light saturation parameter E_K was calculated as $rETR_M$/α and the fitted values of the parameters were used to estimate φPSII under the growth light intensity of 50 µmol photons m^{-2} s^{-1} as $φPSII_{50µE}$= ($rETR_M$.(1-exp(-α.50/$rETR_M$)))/ 50. The NPQ vs E curve was fitted as NPQ = NPQ_M× E^n/ ($E_{50}NPQ^n$+E^n), where NPQ_M is the maximal NPQ, E_{50}NPQ is the half- saturation intensity for NPQ and n is the sigmoidicity coefficient (Serôdio and Lavaud 2011).

The PSII functional absorption cross-section, σPSII, was calculated from the fluorescence induction upon a single turnover flash of blue light (100 µs, 455 nm, 60 nm bandwidth) on non-concentrated cell culture. The induction curve was measured on 20 min dark-acclimated samples before centrifugation (average of 2-4 independent replicates) with a Fluorescence Induction and Relaxation (miniFIRe) fluorometer (Gorbunov, Shirsin et al. 2020), which also measures single turnover F_V/F_M and PSII connectivity. Parameters measured with the miniFIRe fluorometer (as defined below) were also quantified for cultures grown under 8C CL, as the measurements were sufficiently rapid to allow the culture to be maintained at growth temperatures (Gorbunov, Shirsin et al. 2020). Measured photophysiological values are tabulated in **Supplemental Dataset 4**, sheet 8. Violin plots of photophysiological parameters were generated with BoxPlotR (Spitzer, Wildenhain et al. 2014).

Gene expression analysis

Libraries were prepared from 200 ng DNAse-treated RNA for each mutant line and treatment condition, with at least three replicates per sample. Sequencing was performed by Fasteris (Plan-les-Ouates, Switzerland). After initial quality control checks, stranded Illumina mRNA libraries were prepared with a Novaseq V1.5 kit and

sequenced with an SP-flow cell with 2x 100 bp over 200 cycles, yielding circa 130-160 Gb sequence data per sample with ≥85% of bases higher than Q30.

FastQ files were mapped using Nextflow's RNA sequencing assembly pipeline https://nf-co.re/rnaseq/usage, to gff3 annotations of the *P. tricornutum* version 3 genome (Rastogi, Maheswari et al. 2018, Lataretu and Hölzer 2020). Total mapped read counts were then compared between all biological and technical replicates for (i) each mutant line sequenced, (ii) each genotype (cpEnolase knockout, cpPGAM1A knockout, control), and (iii) each treatment condition performed (LD, CL, 8C) by principal component analysis (PCA) using the R package factoextra, with highly variant libraries removed (Kassambara and Mundt 2017). The final dataset included 63 RNAseq libraries, including five cpEnolase and five cpPGAM1A knockout lines and four empty vector controls, and a minimum of four RNA libraries prepared from at least two genetically distinct mutant constructs for each genotype (cpEnolase, cpPGAM1A and control) considered (**Supplemental Dataset 5**, sheets 1-2)., Differentially expressed genes (DEGs) were then calculated between each genotype for each treatment condition using DESeq2 with cutoff fold-change 2 and P-value 0.05 (Liu, Wang et al. 2021) (**Supplemental Dataset 5**, sheets 2-3).

The mean transcript abundances of DEGs in knockout compared to control lines were first assessed in RNAseq data of N and P-limited *P. tricornutum* cell lines under two and nine time-points respectively (**Supplemental Dataset 5**, sheet 4) (Cruz de Carvalho, Sun et al. 2016, McCarthy, Smith et al. 2017). No significant differences were found between DEGs and other genes in the *P. tricornutum* genome (one-way ANOVA, $P > 0.05$; **Supplemental Dataset 5**, sheet 5), confirming that the RNAseq samples were not generated from N- or P-limited cultures. Next, functional enrichments in DEGs from previously tabulated values for the entire *P. tricornutum* genome (**Supplemental Dataset 5**, sheets 6-10) (Rastogi, Maheswari et al. 2018, Ait-Mohamed, Novák Vanclová et al. 2020). Functional enrichments were tested by two-tailed chi-square ($P < 0.05$) of a differentially expressed gene occurring in either one (cpEnolase v control; cpPGAM1A v control) knockout-versus-control line tests, or in both tests realised under each physiological condition. Finally, the distribution of DEGs across *P. tricornutum* core plastid and mitochondrial metabolism pathways were mapped onto a previously defined proteomic model of each organelle (Ait-

Mohamed, Novák Vanclová et al. 2020); with the strongest DEG enrichment taken in the case of enzymes with multiple homologues (**Supplemental Dataset 5**, sheet 11).

Quantitative RT-PCR (qRT-PCR) validations were performed using cDNA synthesised from 5 ng dNase-treated RNA (per qRT-PCR reaction) and a RT Maxima First Strand synthesis kit (Thermo Fisher), following the manufacturer's instruction; using a 384-well Lightcycler (Roche) and Takyon™ No ROX SYBR 2X MasterMix (Eurogentec), following the manufacturers' protocols. Typical amplification conditions were: 10 minutes at 95°C, followed by 40 cycles of 30 seconds at 95°C, 30 seconds at 55°C, and 30 seconds at 72°C. Primer pairs for qRT-PCR amplifications were designed using NCBI Primer-BLAST (Ye, Coulouris et al. 2012), privileging unique amplification targets within the genomic sequence, an amplicon size of 100 to 150 base pairs, primer positions at different regions of the gene studied, and a 3' terminal G or C on each primer. Primer efficiencies were tested by qPCR with serial dilutions of *P. tricornutum* gDNA, with only primer pairs that yielded a Cp increment of between 0.9 and 1.1 per half dilution of DNA retained for qRT-PCR analysis. qRT-PCRs were at least three times for each amplicon: sample pair. RT-equivalents were performed to subtract residual genomic DNA from each Cp value obtained, and two housekeeping genes (Ribosomal protein S1, RPS; and TATA binding protein, TBP) previously shown to have conditionally invariant expression patterns in *P. tricornutum* were used as quantification references (Sachse, Sturm et al. 2013). Tabulated qRT-PCR outputs are shown in **Supplemental Dataset 5**, sheet 13; and sample information and reaction conditions per MIQE guidelines (Bustin, Benes et al. 2009) are tabulated in **Supplemental Dataset 5**, sheet 14.

Metabolite analysis

Cell pellets were taken from exponential-phase *P. tricornutum* culture (counted density 1-2 x 10^6 cells ml^{-1}, 1.5 x 10^8 cells per sample) for metabolite and lipid analysis. Cell pellets were collected without washing to minimise impacts on metabolite turnover, then transferred to a pre-weighed, double-pierced and double-autoclaved 1.5 ml Eppendorf tube for lyophilization. Cell pellet masses were recorded, and samples were immediately snap-frozen in liquid nitrogen and stored at -80 °C for subsequent analysis.

Metabolite profiling was carried out by gas chromatography–mass spectrometry (ChromaTOF software, Pegasus driver 1.61; LECO) as described previously (Lisec, Schauer et al. 2006). The chromatograms and mass spectra were evaluated using TagFinder software (Luedemann, von Malotky et al. 2012). Metabolite identification was manually checked by the mass spectral and retention index collection of the Golm Metabolome Database (Kopka, Schauer et al. 2005). Peak heights of the mass fragments were normalized successively on the basis of the fresh weight of the sample, the added amount of an internal standard (ribitol) and values obtained for loading column controls obtained from the same experiment.

Glycerolipid analysis

Glycerolipids were extracted by suspending cell pellets in 4 mL of boiling ethanol for 5 minutes to prevent lipid degradation. Lipids were extracted by addition of 2 mL methanol and 8 mL chloroform at room temperature (Folch, Lees et al. 1957). The mixture was then saturated with argon and stirred for 1 hour at room temperature. After filtration through glass wool, cell remains were rinsed with 3 mL chloroform/methanol 2:1, v/v and 5 mL of NaCl 1% was added to the filtrate to initiate biphase formation. The chloroform phase was dried under argon and stored at -20 °C. The lipid extract was resuspended in pure chloroform when needed.

Total glycerolipids were quantified from their fatty acids: in an aliquot fraction, 5 µg of 15:0 was added and the fatty acids present were converted to methyl esters (FAME) by a 1-hour incubation in 3 mL 2.5% H_2SO_4 in pure methanol at 100 °C (Jouhet, Maréchal et al. 2003). The reaction was stopped by addition of 3 mL water and 3 mL hexane. The hexane phase was analyzed by a gas chromatography-flame ionization detector (GC-FID) (Perkin Elmer) on a BPX70 (SGE) column. FAMEs were identified by comparison of their retention times with those of standards (Sigma) and quantified using 15:0 for calibration.

Glycerolipids were further analyzed by high pressure liquid chromatography-tandem mass spectrometry (HPLC-MS/MS), based on a previously described procedure (Rainteau, Humbert et al. 2012). The lipid extracts corresponding to 25 nmol of total

fatty acids 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 thylakoids (Demé, Cataye et al. 2014) and hydrogenated (Buseman, Tamura et al. 2006). Lipid classes were separated using an Agilent 1200 HPLC system using a 150 mm × 3 mm (length × 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/v/v)] (A) and isopropanol/ water/ 1 M ammonium acetate, pH 5.3 [850/149/1, (v/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 the 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 modes 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 -56eV. PE, phosphatidylinositol (PI), phosphatidylglycerol (PG), monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG) measurements were performed in positive ion modes 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

MRM with 50 ms dwell time with the various transitions previously recorded (Abida, Dolch et al. 2015). Mass spectra were processed using the MassHunter Workstation software (Agilent) for lipid identification and quantification. Lipid amounts (pmol) were corrected for response differences between internal standards and endogenous lipids as described previously (Jouhet, Lupette et al. 2017).

Normalised metabolite and lipid abundances were screened by PCA, as per the RNAseq analysis above, and outliers and biologically non-representative samples were removed. The final datasets consist of 139 libraries (metabolite GC-MS), 55 libraries (lipid GC-MS) and 49 libraries (lipid LC-MS), with a minimum of three libraries prepared from at least two genetically distinct mutant constructs for each genotype considered (**Supplemental Dataset 6**, sheet 1). Violin plots of differentially accumulated lipids were generated with BoxPlotR (Spitzer, Wildenhain et al. 2014).

Expressed enzyme reaction kinetics

Measurements of cpEnolase and cpPGAM1A reaction rates were performed following a previously defined protocol (Zhang, Sampathkumar et al. 2020) (**Fig. S16**). First, codon-optimised constructs for *E. coli* expression were synthesized from cpEnolase and cpPGAM1A mRNA sequences (Eurofins). Constructs were cloned into a Gateway pDest-CTDHis vector via a pDONR intermediate and BP /LR clonase (all reagents Thermo Fisher) following the manufacturer's instructions (Hartley, Temple et al. 2000). To enable optimal expression in *E. coli*, multiple N-terminal length variants were synthesized from each gene, with those corresponding to the full gene length minus the predicted N-terminal signal peptide domain as inferred with SignalP (Almagro Armenteros, Tsirigos et al. 2019). Complete constructs and primers tested are provided in **Dataset S6**, sheet 7.

cpEnolase and cpPGAM1A –CTDHis vectors were cloned into Rosetta (DE3) strain *E. coli* (Novagen) and coselected on ampicillin (100 µg /ml) and chloramphenicol (34 µg /ml). Proteins were induced in overnight cultures at 28°C, purified on a His-Trap column (GE Healthcare) following the manufacturers' instructions, and eluted in a buffer consisting of 125 mM NaCl, 250 mM Imidiazol (Sigma) and protease inhibitors. Eluted proteins were desalted using a Q10/ PD10 column (GE Healthcare) and

quantified using a Bradford. Protein integrity and quantity were assessed routinely throughout the purification using SDS-PAGE.

Reaction rates were measured on purified 100 µg cpPGAM1A and 100 µg cpEnolase, as quantified with a nanodrop spectrometer. Rates were measured separately for glycolytic and gluconeogenic activity. Briefly, to measure glycolytic reaction rates, both enzymes were combined with 10 units pyruvate kinase and 10 units lactate dehydrogenase (both Sigma-Aldrich) at 25°C, alongside varying concentrations 9 mM D(-)3-Phosphoglyceric Acid, 25 mM Adenosine ''-Diphosphate, and 25 mM reduced ß-Nicotinamide Adenine Dinucleotide (NADH). Enzymatic activity was measured by considering 340 nm colorimetry as a proxy for NADH consumption following a previously defined protocol (Sigma protocols EC 5.4.2.1) (Sutherland, Posternak et al. 1949). To measure gluconeogenic reaction rates, a similar reaction was performed with both enzymes combined with 10 units phosphoglycerate kinase and 10 units glyceraldehyde-3-phosphate dehydrogenase (both Sigma-Aldrich), alongside 9 mM phospho-enol-pyruvate, 25 mM Adenosine 5'-Diphosphate, and 25 mM reduced ß-Nicotinamide Adenine Dinucleotide (NADH). Enzymatic activity was similarly measured by 340 nm colorimetry. A schematic of the measured reactions is provided in **Fig. S16**. Complete measured reaction rates over all technical replicates are provided in **Dataset** S6, sheet 8.

Accession Numbers

RNAseq data associated with this project is deposited with NCBI BioProject with project number PRJNA788211.

Materials Distribution Statement

The author(s) 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: Richard G. Dorrell (richard.dorrell.algae@gmail.com) and Chris Bowler (cbowler@bio.ens.psl.eu).

List of Supporting Files

Supplemental Dataset S1. Phylogenetic diversity of Enolase and PGAM sequences from across the tree of life.

Supplemental Dataset S2. Transcriptional and localization patterns of cpPGAM and cpEnolase genes in *Phaeodactylum tricornutum*.

Supplemental Dataset S3. *Tara* Oceans analysis of diatom plastid glycolysis.

Supplemental Dataset S4. Genotyping, growth dynamics and photophysiology in *P. tricornutum* plastid glycolysis mutant lines.

Supplemental Dataset S5. Differentially and conditionally expressed genes in *P. tricornutum* plastid glycolysis mutants.

Supplemental Dataset S6. Lipid and metabolite profiles of *P. tricornutum* plastid glycolysis mutant lines revealed by GC- and LC-mass spectrometry, and measured reaction kinetics of expressed enzymes.

All remaining supporting data not provided directly in paper supporting tables are provided in the linked Open Science Foundation Supporting database https://osf.io/89vm3/ (Dorrell, Novak Vanclova et al. 2022). Project contents are ordered hierarchically by theme, with an overview of all contents provided on the site wiki page. A dedicated README file in each project folder explains the data presented and provides detailed methodology for each analysis.

Author Contributions

RGD designed the research, with critical input from YZ, YZ, DC, BB, ARF, JJ, EM and CB. RGD, YZ, NG, TN, DC, MP, and VG performed the research. RGD, YZ, YL, DC and MP analysed the data. SA provided new analytical tools for cell growth measurements, and JJPK and NZ provided new computational tools for meta-genomic and RNA-seq analysis. RGD wrote the paper, with critical input from YL,

TN, MP, SA, BB, JJPK, JJ, EM and CB. All coauthors read and approved the manuscript draft prior to final submission.

Acknowledgments

RGD acknowledges a CNRS Momentum Fellowship, awarded from 2019-2021, an ANR JCJC Grant (« PanArctica » ANR-21-CE02-0014-01) awarded 2021-2022 and an ERC Starting Grant Grant (ChloroMosaic, 101039760), awarded from 2023-2027. CB acknowledges support from FFEM - French Facility for Global Environment, French Government 'Investissements d'Avenir' programs OCEANOMICS (ANR-11-BTBR-0008), FRANCE GENOMIQUE (ANR-10-INBS-09-08), MEMO LIFE (ANR-10-LABX-54), and PSL Research University (ANR-11-IDEX-0001-02), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Diatomic; grant agreement No. 835067), and from the ANR 'BrownCut' project (ANR-19-CE20-0020). CB, EM and JJ were supported by ANR 'DIM' (ANR-21-CE02-0021) and PEPR AlgAdvance (22-PEBB-0002). The lipid analyses were performed at the LIPANG (Lipid analysis in Grenoble) platform hosted by the LPCV (UMR 5168 CNRS-CEA-INRAE-UGA) and supported by the Rhône-Alpes Region, the European Regional Development Fund (ERDF), Institut Carnot 3BCAR and Labex GRAL (10-LABX-0049), financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). YZ and ARF acknowledge funding from the European Union's Horizon 2020 research and innovation program, project PlantaSYST (SGA-CSA No. 739582 under FPA No. 66462; and the BG05M2OP001-1.003-001-C01 project, financed by the European Regional Development Fund through the Bulgarian 'Science and Education for Smart Growth' Operational Programme. D.C and B.B acknowledge the support of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (PhotoPHYTOMIX project, grant agreement No. 715579). The authors acknowledge Giselle McCallum, Elena Kazamia and Xia Gao (IBENS) for assistance with the optimization of cell cytometry protocols; Amandine Baylet (Lycée ENCPB-Pierre Gilles de Gennes), Quentin Caris (Lycée de la Vallée de Chevreuse) and Yonna Lauruol (Ecole Sup'BioTech) for assistance with growth measurements, Frédy Barneche and Clara Richet-Bourbousse (IBENS) for the use of culture facilities; Nathalie Joli (IBENS) for

aid with RNAseq library production; Catherine Cantrel and Priscillia Pierre-Elies (IBENS), Pauline Clémente (Lycée ENCPB-Pierre Gilles de Gennes) and Shun-Min Yang (University of South Bohemia) for the preparation of media substrates and aid with biolistic transformations; Max Gorbunov (Rutgers University, NJ, USA) for the provision of the miniFIRE fluorometer used for photophysiological assays; and Mattia Storti (CEA Grenoble) for cell outlines used in the production of **Fig. 1A**. This paper is contribution XXX of *Tara* Oceans.

References

Abida, H., L. J. Dolch, C. Meï, V. Villanova, M. Conte, M. A. Block, G. Finazzi, O. Bastien, L. Tirichine, C. Bowler, F. Rébeillé, D. Petroutsos, J. Jouhet and E. Maréchal (2015). Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricornutum*. Plant Physiol 167(1): 118-136.

Ait-Mohamed, O., A. M. G. Novák Vanclová, N. Joli, Y. Liang, X. Zhao, A. Genovesio, L. Tirichine, C. Bowler and R. G. Dorrell (2020). PhaeoNet: a holistic RNAseq-based portrait of transcriptional coordination in the model diatom *Phaeodactylum tricornutum*. Frontiers Plant Sci 11: 590949.

Allen, A. E., C. L. Dupont, M. Obornik, A. Horak, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie and C. Bowler (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473(7346): 203-207.

Almagro Armenteros, J. J., K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne and H. Nielsen (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnol 37(4): 420-423.

Andriotis, V. M., N. J. Kruger, M. J. Pike and A. M. Smith (2010). Plastidial glycolysis in developing *Arabidopsis* embryos. New Phytol 185(3): 649-662.

Anoman, A. D., M. Flores-Tornero, S. Rosa-Telléz, J. Muñoz-Bertomeu, J. Segura and R. Ros (2016). The specific role of plastidial glycolysis in photosynthetic and heterotrophic cells under scrutiny through the study of glyceraldehyde-3-phosphate dehydrogenase. Plant Signal Behav 11(3): 1128614.

Bai, Y., T. Cao, O. Dautermann, P. Buschbeck, M. B. Cantrell, Y. Chen, C. D. Lein, X. Shi, M. A. Ware, F. Yang, H. Zhang, L. Zhang, G. Peers, X. Li and M. Lohr (2022). Green diatom mutants reveal an intricate biosynthetic pathway of fucoxanthin. Proc Natl Acad Sci USA 119(38): 2203708119.

Bailleul, B., N. Berne, O. Murik, D. Petroutsos, J. Prihoda, A. Tanaka, V. Villanova, R. Bligny, S. Flori, D. Falconet, A. Krieger-Liszkay, S. Santabarbara, F. Rappaport, P. Joliot, L. Tirichine, P. G. Falkowski, P. Cardol, C. Bowler and G. Finazzi (2015). Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. Nature 524(7565): 366-U267.

Behrenfeld, M. J., K. H. Halsey, E. Boss, L. Karp-Boss, A. J. Milligan and G. Peers (2021). Thoughts on the evolution and ecological niche of diatoms. Ecol Monographs 91(3): 01457.

Broddrick, J. T., N. Du, S. R. Smith, Y. Tsuji, D. Jallet, M. A. Ware, G. Peers, Y. Matsuda, C. L. Dupont, B. G. Mitchell, B. O. Palsson and A. E. Allen (2019). Cross-compartment

metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricornutum*. New Phytol 222: 1364-1379.
Bromke, M. A. (2013). Amino Acid biosynthesis pathways in diatoms. Metabolites 3(2): 294-311.
Buck, J. M., C. Río Bártulos, A. Gruber and P. G. Kroth (2018). Blasticidin-S deaminase, a new selection marker for genetic transformation of the diatom. PeerJ 6: 5884.
Buck, J. M., J. Sherman, C. R. Bártulos, M. Serif, M. Halder, J. Henkel, A. Falciatore, J. Lavaud, M. Y. Gorbunov, P. G. Kroth, P. G. Falkowski and B. Lepetit (2019). Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom *Phaeodactylum tricornutum*. Nat Commun 10(1): 4167.
Buseman, C. M., P. Tamura, A. A. Sparks, E. J. Baughman, S. Maatta, J. Zhao, M. R. Roth, S. W. Esch, J. Shah, T. D. Williams and R. Welti (2006). Wounding stimulates the accumulation of glycerolipids containing Oxophytodienoic Acid and Dinor-Oxophytodienoic Acid in *Arabidopsis* leaves. Plant Physiol 142(1): 28-39.
Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer (2009). The MIQE Guidelines: Minimum Information for publication of Quantitative real-time PCR Experiments. Clinical Chemistry 55(4): 611-622.
Capella-Gutiérrez, S., J. M. Silla-Martínez and T. Gabaldón (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15): 1972-1973.
Carradec, Q., E. Pelletier, C. Da Silva, A. Alberti, Y. Seeleuthner, R. Blanc-Mathieu, G. Lima-Mendez, F. Rocha, L. Tirichine, K. Labadie, A. Kirilovsky, A. Bertrand, S. Engelen, M. A. Madoui, R. Méheust, J. Poulain, S. Romac, D. J. Richter, G. Yoshikawa, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, O. Jaillon, J. M. Aury, E. Karsenti, M. B. Sullivan, S. Sunagawa, P. Bork, F. Not, P. Hingamp, J. Raes, L. Guidi, H. Ogata, C. de Vargas, D. Iudicone, C. Bowler, P. Wincker and T. O. Coordinators (2018). A global ocean atlas of eukaryotic genes. Nat Commun 9(1): 373.
Carrera, D., G. M. George, M. Fischer-Stettler, F. Galbier, S. Eicke, E. Truernit, S. Streb and S. C. Zeeman (2021). Distinct plastid fructose bisphosphate aldolases function in photosynthetic and non-photosynthetic metabolism in *Arabidopsis*. J Exp Bot 72(10): 3739-3755.
Chang, Y. F., J. S. Imam and M. F. Wilkinson (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76: 51-74.
Cruz de Carvalho, M. H., H. X. Sun, C. Bowler and N. H. Chua (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. New Phytol 210(2): 497-510.
Delmont, T. O., M. Gaia, D. D. Hinsinger, P. Frémont, C. Vanni, A. Fernandez-Guerra, A. M. Eren, A. Kourlaiev, L. d'Agata, Q. Clayssen, E. Villar, K. Labadie, C. Cruaud, J. Poulain, C. Da Silva, M. Wessner, B. Noel, J.-M. Aury, S. Sunagawa, S. G. Acinas, P. Bork, E. Karsenti, C. Bowler, C. Sardet, L. Stemmann, C. de Vargas, P. Wincker, M. Lescot, M. Babin, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, O. Jaillon, S. Kandels, D. Iudicone, H. Ogata, S. Pesant, M. B. Sullivan, F. Not, K.-B. Lee, E. Boss, G. Cochrane, M. Follows, N. Poulton, J. Raes, M. Sieracki, S. Speich and E. Pelletier (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genom 2(5): 100123.
Demé, B., C. Cataye, M. A. Block, E. Maréchal and J. Jouhet (2014). Contribution of galactoglycerolipids to the 3-dimensional architecture of thylakoids. FASEB J 28(8): 3373-3383.

Dolch, L. J., J. Lupette, G. Tourcier, M. Bedhomme, S. Collin, L. Magneschi, M. Conte, K. Seddiki, C. Richard, E. Corre, L. Fourage, F. Laeuffer, R. Richards, M. Reith, F. Rébeillé, J. Jouhet, P. McGinn and E. Maréchal (2017). Nitric oxide mediates nitrite-sensing and acclimation and triggers a remodeling of lipids. Plant Physiol 175(3): 1407-1423.

Dolch, L. J. and E. Maréchal (2015). Inventory of fatty acid desaturases in the pennate diatom *Phaeodactylum tricornutum*. Mar Drugs 13(3): 1317-1339.

Dorrell, R. G., A. M. G. Novak Vanclova, M. Penot, J. J. Pierella Karlusich, C. Bowler, S. Liu, E. Maréchal, J. Jouhet, B. Bailleul and D. Croteau (2022). Functional physiology of novel diatom chloroplast proteins. Open Science Foundation. https://osf.io/89vm3/.

Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocol 2(4): 953-971.

Erdene-Ochir, E., B. K. Shin, B. Kwon, C. Jung and C. H. Pan (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. Sci Rep 9(1): 9941.

Falciatore, A., R. Casotti, C. Leblanc, C. Abrescia and C. Bowler (1999). Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol 1(3): 239-251.

Folch, J., M. Lees and G. H. S. Stanley (1957). A somple method for the isolation and purification of total lipids from animal tissues. J Biol Chem 226(1): 497-509.

Friedlingstein, P., M. W. Jones, M. O'Sullivan, R. M. Andrew, D. C. Bakker, J. Hauck, C. Le Quéré, G. P. Peters, W. Peters and J. Pongratz (2022). Global carbon budget 2021. Earth System Science Data 14(4): 1917-2005.

Fukasawa, Y., J. Tsuji, S. C. Fu, K. Tomii, P. Horton and K. Imai (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Mol Cell Proteom 14(4): 1113-1126.

Fukayama, H., C. Masumoto, Y. Taniguchi, A. Baba-Kasai, Y. Katoh, H. Ohkawa and M. Miyao (2015). Characterization and expression analyses of two plastidic enolase genes in rice. Biosci Biotechnol Biochem 79(3): 402-409.

Fuss, J., O. Liegmann, K. Krause and S. A. Rensing (2013). Green Targeting Predictor and Ambiguous Targeting Predictor 2: the pitfalls of plant protein targeting prediction and of transient protein expression in heterologous systems. New Phytol 200: 222-233.

Gilbertson, R., E. Langan and T. Mock (2022). Diatoms and their microbiomes in complex and changing polar oceans. Front Microbiol 13: 786764.

Gorbunov, M. Y., E. Shirsin, E. Nikonova, V. V. Fadeev and P. G. Falkowski (2020). A multi-spectral fluorescence induction and relaxation (FIRe) technique for physiological and taxonomic analysis of phytoplankton communities. Mar Ecol Progress Series 644: 1-13.

Grigoriev, I. V., R. D. Hayes, S. Calhoun, B. Kamel, A. Wang, S. Ahrendt, S. Dusheyko, R. Nikitin, S. J. Mondo, A. Salamov, I. Shabalov and A. Kuo (2021). PhycoCosm, a comparative algal genomics resource. Nucl Acids Res 49(D1): D1004-D1011.

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

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

Hartley, J. L., G. F. Temple and M. A. Brasch (2000). DNA cloning using in vitro site-specific recombination. Genome Res 10(11): 1788-1795.

Huang, A., L. Liu, C. Yang and G. Wang (2015). *Phaeodactylum tricornutum* photorespiration takes part in glycerol metabolism and is important for nitrogen-limited response. Biotechnol Biofuels 8(1): 1-16.

Huang, T., Y. Pan, E. Maréchal and H. Hu (2023). Proteomes reveal the lipid metabolic network in the complex plastid of *Phaeodactylum tricornutum*. Plant J in press: 10.1111/tpj.16477.
Initiative, O. T. P. T. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574(7780): 679-685.
Jassby, A. D. and T. Platt (1976). Mathematical formulation of the relationship between photosynthesis and light for phytoplankton. Limnol Oceanography 21(4): 540-547.
Joli, N., L. Concia, K. Mocaer, J. Guterman, J. Laude, S. Guerin, T. Sciandra, F. Bruyant, O. Ait-Mohamed, M. Beguin, M.-H. Forget, C. Bourbousse, T. Lacour, B. Bailleul, C. Nef, M. Savoie, J.-E. Tremblay, D. A. Campbell, J. Lavaud, Y. Schwab, M. Babin and C. Bowler (2023). Hypometabolism to survive the long polar night and subsequent successful return to light in the diatom *Fragilariopsis cylindrus*. New Phytol in press: 10.1111/nph.19387.
Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez and S. Hunter (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9): 1236-1240.
Jouhet, J., J. Lupette, O. Clerc, L. Magneschi, M. Bedhomme, S. Collin, S. Roy, E. Maréchal and F. Rébeillé (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(8): 0182423.
Jouhet, J., E. Maréchal, R. Bligny, J. Joyard and M. A. Block (2003). Transient increase of phosphatidylcholine in plant cells in response to phosphate deprivation. FEBS Lett 544(1): 63-68.
Kassambara, A. and F. Mundt (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses 76(2).
Katoh, K., J. Rozewicki and K. D. Yamada (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform.
Kazamia, E., R. Sutak, J. Paz-Yepes, R. G. Dorrell, F. R. J. Vieira, J. Mach, J. Morrissey, S. Leon, F. Lam, E. Pelletier, J. M. Camadro, C. Bowler and E. Lesuisse (2018). Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. Sci Adv 4(5): 4536.
Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647-1649.
Keeling, P. J., F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone and A. Z. Worden (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional

diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12(6): 1001889.
Kopka, J., N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmüller, P. Dörmann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie and D. Steinhauser (2005). GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21(8): 1635-1638.
Kroth, P. G., A. Chiovitti, A. Gruber, V. Martin-Jezequel, T. Mock, M. S. Parker, M. S. Stanley, A. Kaplan, L. Caron, T. Weber, U. Maheswari, E. V. Armbrust and C. Bowler (2008). A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. PLoS One 3(1): 1426.
Lacour, T., J. Larivière and M. Babin (2017). Growth, Chl a content, photosynthesis, and elemental composition in polar and temperate microalgae. Limnol Oceanog 62(1): 43-58.
Lampe, R. H., G. Hernandez, Y. Y. Lin and A. Marchetti (2021). Representative diatom and coccolithophore species exhibit divergent responses throughout simulated upwelling cycles. mSystems 6(2): 10.1128/msystems.00188-21.
Lataretu, M. and M. Hölzer (2020). RNAflow: an effective and simple RNA-Seq differential gene expression pipeline using Nextflow. Genes (Basel) 11(12).
Lepetit, B., D. A. Campbell, J. Lavaud, C. Büchel, R. Goss and B. Bailleul (2022). Photosynthetic Light Reactions in diatoms. II. The dynamic regulation of the various light reactions. The Molecular Life of Diatoms. A. Falciatore and T. Mock. Cham, Springer International Publishing: 423-464.
Levering, J., J. Broddrick, C. L. Dupont, G. Peers, K. Beeri, J. Mayers, A. A. Gallina, A. E. Allen, B. O. Palsson and K. Zengler (2016). Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PloS One 11(5): 0155038.
Lisec, J., N. Schauer, J. Kopka, L. Willmitzer and A. R. Fernie (2006). Gas chromatography mass spectrometry–based metabolite profiling in plants. Nature Protocols 1(1): 387-396.
Liu, S., M. Storti, C. Bowler, G. Finazzi and R. G. Dorrell (2022). An integrative environmental atlas of diatom chloroplast transporters. Front Plant Sci 13: 950467.
Liu, S., Z. Wang, R. Zhu, F. Wang, Y. Cheng and Y. Liu (2021). Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. J Vis Exp(175).
Luedemann, A., L. von Malotky, A. Erban and J. Kopka (2012). TagFinder: preprocessing software for the fingerprinting and the profiling of gas chromatography-mass spectrometry based metabolome analyses. Methods Mol Biol 860: 255-286.
Malviya, S., E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. de Vargas, L. Bittner, A. Zingone and C. Bowler (2016). Insights into global diatom distribution and diversity in the world's ocean. Proc Natl Acad Sci USA 113(11): 1516-1525.
Maréchal, E. and J. Lupette (2020). Relationship between acyl-lipid and sterol metabolisms in diatoms. Biochimie 169: 3-11.
McCarthy, J. K., S. R. Smith, J. P. McCrow, M. Tan, H. Zheng, K. Beeri, R. Roth, C. Lichtle, U. Goodenough, C. P. Bowler, C. L. Dupont and A. E. Allen (2017). Nitrate Reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. Plant Cell 29(8): 2047-2070.
Miller, M. A., T. Schwartz, B. E. Pickett, S. He, E. B. Klem, R. H. Scheuermann, M. Passarotti, S. Kaufman and M. A. O'Leary (2015). A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. Evol Bioinform Online 11: 43-48.

Mistry, J., S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn and A. Bateman (2020). Pfam: The protein families database in 2021. Nucleic Acids Res 49: D412-D419.

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

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

Murik, O., L. Tirichine, J. Prihoda, Y. Thomas, W. L. Araújo, A. E. Allen, A. R. Fernie and C. Bowler (2019). Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. New Phytol 221(3): 1303-1316.

Nash, E. A., A. C. Barbrook, R. K. Edwards-Stuart, K. Bernhardt, C. J. Howe and R. E. R. Nisbet (2007). Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. Mol Biol Evol 24(7): 1528-1536.

Nawaly, H., H. Matsui, Y. Tsuji, K. Iwayama, H. Ohashi, K. Nakajima and Y. Matsuda (2023). Multiple plasma membrane SLC4s contribute to external HCO3– acquisition during CO2 starvation in the marine diatom *Phaeodactylum tricornutum*. J Exp Bot 74(1): 296-307.

Nonoyama, T., E. Kazamia, H. Nawaly, X. Gao, Y. Tsuji, Y. Matsuda, C. Bowler, T. Tanaka and R. G. Dorrell (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. Biomolecules 9(8): 464.

Nymark, M., A. K. Sharma, T. Sparstad, A. M. Bones and P. Winge (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci Rep 6: 24951.

Pesant, S., F. Not, M. Picheral, S. Kandels-Lewis, N. Le Bescot, G. Gorsky, D. Iudicone, E. Karsenti, S. Speich, R. Trouble, C. Dimier, S. Searson and *Tara* Oceans Coordinators (2015). Open science resources for the discovery and analysis of *Tara* Oceans data. Scientific Data 2.

Popko, J., C. Herrfurth, K. Feussner, T. Ischebeck, T. Iven, R. Haslam, M. Hamilton, O. Sayanova, J. Napier, I. Khozin-Goldberg and I. Feussner (2016). Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of *Phaeodactylum tricornutum*. PLoS One 11(10): 0164673.

Prabhakar, V., T. Löttgert, T. Gigolashvili, K. Bell, U. I. Flügge and R. E. Häusler (2009). Molecular and functional characterization of the plastid-localized Phosphoenolpyruvate enolase (ENO1) from *Arabidopsis thaliana*. FEBS Lett 583(6): 983-991.

Raines, C. A. (2003). The Calvin Cycle revisited. Photosynth Res 75(1): 1-10.

Rainteau, D., L. Humbert, E. Delage, C. Vergnolle, C. Cantrel, M.-A. Maubert, S. Lanfranchi, R. Maldiney, S. Collin, C. Wolf, A. Zachowski and E. Ruelland (2012). Acyl chains of phospholipase D transphosphatidylation products in *Arabidopsis* cells: a study using multiple reaction monitoring mass spectrometry. PLoS One 7(7): 41985.

Rastogi, A., U. Maheswari, R. G. Dorrell, F. R. J. Vieira, F. Maumus, A. Kustka, J. McCarthy, A. E. Allen, P. Kersey, C. Bowler and L. Tirichine (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of *Phaeodactylum tricornutum* genome and evolutionary origin of diatoms. Sci Rep 8(1): 4834.

Rastogi, A., O. Murik, C. Bowler and L. Tirichine (2016). PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. BMC Bioinformatics 17(1): 261.

1744 Royo-Llonch, M., P. Sánchez, C. Ruiz-González, G. Salazar, C. Pedrós-Alió, K. Labadie, L.
1745 Paoli, *Tara* Oceans Coordinators, S. Chaffron, D. Eveillard, E. Karsenti, S. Sunagawa, P.
1746 Wincker, L. Karp-Boss, C. Bowler and S. G. Acinas (2021). Compendium of 530
1747 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean.
1748 Nat Microbiol 6: 1561-1574.
1749 Río Bártulos, C., M. B. Rogers, T. A. Williams, E. Gentekaki, H. Brinkmann, R. Cerff, M. F.
1750 Liaud, A. B. Hehl, N. R. Yarlett, A. Gruber, P. G. Kroth and M. van der Giezen (2018).
1751 Mitochondrial glycolysis in a major lineage of eukaryotes. Genom Biol Evol 10(9): 2310-
1752 2325.
1753 Sachse, M., S. Sturm, A. Gruber and P. Kroth (2013). Identification and evaluation of
1754 endogenous reference genes for steady state transcript quantification by qPCR in the
1755 diatom *Phaeodactylum tricornutum* with constitutive expression independent from time
1756 and light. Endocytobiosis Cell Res 24: 7.
1757 Scialdone, A., S. T. Mugford, D. Feike, A. Skeffington, P. Borrill, A. Graf, A. M. Smith and M.
1758 Howard (2013). *Arabidopsis* plants perform arithmetic division to prevent starvation at
1759 night. Elife 2: 00669.
1760 Serôdio, J. and J. Lavaud (2011). A model for describing the light response of the
1761 nonphotochemical quenching of chlorophyll fluorescence. Photosynth Res 108(1): 61-
1762 76.
1763 Shtaida, N., I. Khozin-Goldberg and S. Boussiba (2015). The role of pyruvate hub
1764 enzymes in supplying carbon precursors for fatty acid synthesis in photosynthetic
1765 microalgae. Photosynth Res 125(3): 407-422.
1766 Siegel, P., K. G. Baker, E. Low-Décarie and R. J. Geider (2020). High predictability of direct
1767 competition between marine diatoms under different temperatures and nutrient states.
1768 Ecol Evol 10(14): 7276-7290.
1769 Smith, S. R., R. M. Abbriano and M. Hildebrand (2012). Comparative analysis of diatom
1770 genomes reveals substantial differences in the organization of carbon partitioning
1771 pathways. Algal Res Biomass Biofuels Bioproducts 1(1): 2-16.
1772 Smith, S. R., C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J.
1773 Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A.
1774 Fernie, Z. Nikoloski, B. O. Palsson and A. E. Allen (2019). Evolution and regulation of
1775 nitrogen flux through compartmentalized metabolic networks in a marine diatom. Nat
1776 Commun 10(1): 4552.
1777 Spitzer, M., J. Wildenhain, J. Rappsilber and M. Tyers (2014). BoxPlotR: a web tool for
1778 generation of box plots. Nature Methods 11(2): 121-122.
1779 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-
1780 analysis of large phylogenies. Bioinformatics 30(9): 1312-1313.
1781 Strassert, J. F. H., I. Irisarri, T. A. Williams and F. Burki (2021). A molecular timescale for
1782 eukaryote evolution with implications for the origin of red algal-derived plastids. Nat
1783 Commun 12(1): 1879.
1784 Sutherland, E. W., T. Posternak and C. F. Cori (1949). Mechanism of the phosphoglyceric
1785 mutase reaction. J Biol Chem 181(1): 153-159.
1786 Tanaka, R. and A. Tanaka (2007). Tetrapyrrole biosynthesis in higher plants. Ann Rev
1787 Plant Biol 58: 321-346.
1788 Tardif, M., A. Atteia, M. Specht, G. Cogne, N. Rolland, S. Brugière, M. Hippler, M. Ferro, C.
1789 Bruley, G. Peltier, O. Vallon and L. Cournac (2012). PredAlgo: a new subcellular
1790 localization prediction tool dedicated to green algae. Mol Biol Evol 29(12): 3625-3639.
1791 Troncoso-Ponce, M. A., J. Rivoal, S. Dorion, R. Sánchez, M. Venegas-Calerón, A. J. Moreno-
1792 Pérez, S. Baud, R. Garcés and E. Martínez-Force (2018). Molecular and biochemical

characterization of the sunflower (*Helianthus annuus* L.) cytosolic and plastidial enolases in relation to seed development. Plant Sci 272: 117-130.
Yang, L., Z. Wang, A. Zhang, R. Bhawal, C. Li, S. Zhang, L. Cheng and J. Hua (2022). Reduction of the canonical function of a glycolytic enzyme enolase triggers immune responses that further affect metabolism and growth in *Arabidopsis*. Plant Cell 34(5): 1745-1767.
Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13: 134.
Yin, W. and H. Hu (2023). CRISPR/Cas9-mediated genome editing via homologous recombination in a centric diatom *Chaetoceros muelleri*. ACS Synth Biol 12(4): 1287-1296.
Yu, G., K. Nakajima, A. Gruber, C. Rio Bartulos, A. F. Schober, B. Lepetit, E. Yohannes, Y. Matsuda and P. G. Kroth (2022). Mitochondrial phosphoenolpyruvate carboxylase contributes to carbon fixation in the diatom *Phaeodactylum tricornutum* at low inorganic carbon concentrations. New Phytol 235: 1379-1393.
Zhang, Y., A. Sampathkumar, S. M.-L. Kerber, C. Swart, C. Hille, K. Seerangan, A. Graf, L. Sweetlove and A. R. Fernie (2020). A moonlighting role for enzymes of glycolysis in the co-localization of mitochondria and chloroplasts. Nature Commun 11(1): 4509.
Zhao, Z. and S. M. Assmann (2011). The glycolytic enzyme, phosphoglycerate mutase, has critical roles in stomatal movement, vegetative growth, and pollen production in *Arabidopsis thaliana*. J Exp Bot 62(14): 5179-5189.
Zheng, Y. T., A. H. Quinn and G. Sriram (2013). Experimental evidence and isotopomer analysis of mixotrophic glucose metabolism in the marine diatom *Phaeodactylum tricornutum*. Microbial Cell Factories 12: 16.
Zhu, B.-H., H.-P. Shi, G.-P. Yang, N.-N. Lv, M. Yang and K.-H. Pan (2016). Silencing UDP-glucose pyrophosphorylase gene in *Phaeodactylum tricornutum* affects carbon allocation. New Biotechnol 33(1): 237-244.
Zulu, N. N., K. Zienkiewicz, K. Vollheyde and I. Feussner (2018). Current trends to comprehend lipid metabolism in diatoms. Progress Lipid Res 70: 1-16.

Figures

Fig. 1. Metabolic context and evolution of the lower half of diatom plastid glycolysis-gluconeogenesis. A: schematic comparison of diatom and plant core carbon metabolism, highlighting the localization and functions of two enzymes in the lower half of glycolysis-gluconeogenesis (phospho-glycerate mutase, and enolase) whose localization to the chloroplast can connect endogenous enzymes in the Calvin cycle and pyruvate hub to create a complete glycolytic-gluconeogenic-gluconeogenic pathway. Abbreviations: GA3P- glyceraldehyde-3-phosphate; 1,3-PGA and 2,3-PGA- 1,3 and 2,3 bis-phosphoglycerate; Glu-6-P- glucose-6-phosphate; PEP- phospho-enol-pyruvate; RuBP- ribulose *bis*-phosphate; PGAM- phospho-glycerate mutase; cER- chloroplast: ndoplasmic reticulum. **B, C:** consensus MrBayes topologies

realised with three substitution matrices (GTR, Jones, WAG) of a 163 taxa x 413 aa alignment of organelle-targeted enolase and 105 taxa x 220 aa alignment of selected organelle-targeted PGAM1 enzymes from diatoms and their closest relatives, identifying recent duplications and recruitments of respiratory glycolytic-gluconeogenic enzymes from the mitochondria to plastid in diatoms and their closest relatives. **D:** overlay images of GFP-tagged full-length cpEnolase (top) and cpPGAM1A (bottom) constructs (green), chlorophyll (red) and bright-field images of transformant *Phaeodactylum tricornutum* lines. Scale bar: 10 µm.

Fig. 2. Environmental distributions of diatom plastid lower half glycolysis-gluconeogenesis meta-genes. Total transcriptome **(top)** and genome **(bottom)** relative abundances for *Tara* Oceans meta-genes phylogenetically resolved to diatom cpEnolase and cpPGAM1, sampled from all size fractions and surface layer stations, demonstrating higher meta-transcript abundance without commensurate increases in meta-gene abundance at high northern and southern latitudes.

Fig. 3. Growth phenotypes of cpEnolase and cpPGAM1A CRISPR-Cas9 knockout mutant and zeocin-resistant empty vector control *P. tricornutum* lines. A: xemplar growth curves from single experiments realised for *P. tricornutum* lines in 50 µE m^{-2} s^{-1} illumination, non-shaken cultures and replete ESAW media, under three conditions- (i) 19°C and 12h light: 12h dark Circadian cycles (« 19C LD »); (ii) 19°C and 24h continuous light (« 19C CL »); and (iii) 8°C and 24h continuous light (« 8C CL »). Hashed black lines show the approximative concentrations (between 5 x10^4 and 4 x 10^6 cells ml^{-1}) over which growth rates were calculated). **B:** mean relative log phase growth rates of each genotype under each condition, measured through a minimum of three biological replicates and two technical repetitions (six measurements per line, minimum 24 measurements per genotype), over five time-points with linear (r^2 > 0.95 relationship between log cell density and time). Asterisks indicate significant differences as inferred by one-way ANOVA. An alternative version of this figure showing absolute growth rates of individual cell lines is provided in **Fig. S10.**

Fig. 4. Changes in plastid and mitochondrial metabolic architecture inferred from gene expression analyses. A: predicted consensus localizations (either:

chloroplast, or non-chloroplast) from ASAFind (Gruber, Rocap et al. 2015) and HECTAR (Gschloessl, Guermeur et al. 2008) of all genes inferred ($P < 0.05$, fold-change expression >2) to be up- or down-regulated in both cpEnolase and cpPGAM1A knockout compared to control lines under 19C LD, 19C CL and 8C CL. Significantly enriched localizations (two-tailed chi-squared test) are asterisked. **B:** relative mRNA abundances of eleven genes encoding exemplar chloroplast- and mitochondria-targeted proteins, verified by qRT-PCR. Genes differentially expressed (*t*-test, $P < 0.05$) in each condition are asterisked.

Fig. 5. Volcano plots of differentially accumulated metabolites assessed by GC-MS. Scatterplots of the $\log_2$ accumulation ratios and $-\log_{10}$ P-values of difference in the mass, ribitol and quality-control-normalised abundances of 39 sugars and amino acid metabolites in cpEnolase and cpPGAM1A knockout compared to empty vector control lines, measured by GC-MS in all three experimental conditions tested. Metabolites that show a differential accumulation in each plot ($P < 10^{-05}$) are labelled, with metabolites that show a differential accumulation in both knockout lines in each condition shown in black text, and five metabolites that are uniquely over-accumulated in cpEnolase knockout lines under all three conditions shown in dark red text.

Fig. 6. LC-MS lipid distributions in glycolysis-gluconeogenesis mutant lines. A: scatterplots of relative proportions of MGDG and TAG in total lipid LC-MS samples in cpEnolase and cpPGAM1A knockout lines and empty vector controls under each growth condition, showing increased MGDG: TAG in glycolysis knockout lines under 19C, and the inverse relationship in cpEnolase knockout lines only under 8C. **B**: violin plots of relative abundances of three further lipid categories inferred to differentially accumulate in glycolysis knockout lines under different growth conditions. Significant differences between knockout and control lines (one-way ANOVA) are asterisked.

Fig. 7. Inferred kinetic activities of *P. tricornutum* plastid lower-half glycolysis-gluconeogenesis. A: Measured reaction rates of purified cpEnolase and cPGAM1A supplemented with 3-PGA (glycolytic direction) or PEP (gluconeogenic direction). **B:**

schematic diagram showing the inferred fluxes through lower half diatom plastid glycolysis-gluconeogenesis in each environmental condition tested.

Fig. S1. Distribution of lower-half plastid glycolysis-glucneogenesis across photosynthetic eukaryotes. A: Occurrence of plastid-targeted enolase and PGAM enzymes across 1,673 plant and algal genomes and transcriptomes, inferred using reciprocal BLAST best hit of *P. tricornutum* query enzymes as per **Fig. 1B** and **1C,** PFAM domain annotations, and in silico targeting predictions with TargetP and PredAlgo (primary chloroplast bearing lineages) and HECTAR and ASAFind (secondary lineages). **B:** scatterplots of collection site latitude for (**i)** diatoms, (ii) other stramenopiles, **(iii)** cryptomonads and haptophytes and **(iv)** green algae with detectable enolase and PGAM enzymes, divided by presence of inferred plastid-targeted isoforms. Notably, diatoms lacking both plastid-targeted glycolysis enzymes do not occur outside of low and intermediate latitudes (50°N in the northern and 60°S in the southern hemisphere) compared to other groups, which either show no significant association between plastid glycolysis and latitude (cryptomonads, haptophytes, other stramenopiles) or even an association with low latitudes (green algae). The data in this figure were subselected for the phylogenies shown in **Fig. 1** and support the latitudinal correlations revealed by Tara analysis of **Fig. 2.**

Fig. S2. Consensus topology of a 380 taxa x 413 aa alignment of Enolase sequences. Sequences represent a sample of all organelle-targeted isoforms from cryptomonads, haptophytes and stramenopiles and representatives from a densely-sampled dataset of 151 taxonomic groups across the tree of life (Dorrell et al., 2021). The tree topology shown is the consensus of the best-scoring rAxML trees identified using three substitution matrices: GTR, JTT, and WAG. Branch thickness corresponds to frequency of consensus tree topology recovery in individual trees; branches are coloured by taxonomic affiliation; and tips (cryptomonads, haptophytes and stramenopiles only) by predicted *in silico* localization. Individual clades (considering both taxonomic origins and inferred localization) of organelle-targeted enolase isoform are labelled with coloured brackets. This figure extends on the topology shown in **Fig. 1B**.

Fig. S3. Consensus topology of a 220 aa x 560 taxa alignment of PGAM isoform 1 sequences. Data are shown as per **Fig. S1**, extending on the topology of **Fig. 1C.**

Fig. S4. Consensus phylogeny of a 235 aa x 66 taxa alignment of PGAM isoform 2 sequences. Data are shown as per **Fig. S1,** forming a complement to the PGAM1 topologies shown in **Fig. 1C** and **S3.**

Fig. S5. Individual fluorescence channel and control confocal microscopy images for *P. tricornutum* plastid glycolysis proteins. Images complement the overlays shown in **Fig. 1D.**

Fig. S6. Identification of *Tara* Oceans homologs of diatom plastid-targeted enolase and PGAM enzymes. A: Consensus rAxML JTT topologies of the phylogenetically verified *Tara* Oceans homologs of diatom plastidial enolase and PGAM enzymes and cultured species sequences, demonstrating reconciliation of retained homologs within monophyletic clades containing exclusively diatom plastidial isoforms amongst cultured species. **B:** in silico targeting predictions of all retrieved homologs inferred by BLAST alignment to be probably N-terminally complete, showing a strong enrichment in homologs with predicted plastid-targeting sequences. Sequences shown in this figure are analysed globally in **Fig. 2**.

Fig. S7. Relative abundances of *Tara* Oceans diatom plastid glycolysis meta-genes. Plots show relative abundances of meta-genes that group with **(i, iii)** cpEnolase and **(ii, iv)** cpPGAM1 sequences over individual size fractions of **(i, ii)** surface and **(iii, iv)** DCM meta-transcriptome (left) and -genome (right) data, supporting data from all (unfiltered) size fractions and surface layers shown in **Fig. 2**.

Fig. S8. Total relative abundances of meta-genes phylogenetically reconciled to diatom PGAM2 in unfiltered surface samples. Plots showing **(A)** meta-transcriptome and **(B)** meta-genome data, showing effective congruence between both, in contrast to the high latitudinal abundance specific to meta-transcriptome data for diatom cpEnolase and cpPGAM1 as per **Fig. 2**.

Fig. S9. Genotypes of *P. tricornutum* glycolysis knockout lines. A: alignments of the two CRISPR regions targeted for mutagenesis of cpEnolase (Phatr3_J41515) and cpPGAM1A (Phatr3_J17086), and the genotypes obtained from Sanger

sequences of homozygous CRISPR knockouts obtained for each gene. **B:** average relative expression level of each mutated gene, assessed by quantitative RT-PCR with two primer combinations and normalised against two housekeeping genes (RNA polymerase II and TATA binding protein), expressed as a % of the relative expression levels calculated in two empty vector expression controls. One-way *t*-test significance levels of the knockdown of gene expression in each knockout line compared to the empty vector controls are provided. Knockout lines shown in this figure were used for growth and integrative 'omic analyses as per **Figs. 3-6**.

Fig. S10. Absolute and individual growth phenotypes of cpEnolase and cpPGAM1A CRISPR-Cas9 knockout mutant lines. A: growth curves of knockout lines, shown as per **Fig. 3**, but with absolute as opposed to logarithmic cell concentrations B: scatterplot showing the average and standard deviation relative growth rates for each cell line studied under 19C CL (vertical) and 19C LD (horizontal axis). Each point corresponds to an individual line, with genotype indicated by point colours, and standard deviations of growth rates by error bars. Despite individual variances in growth rate between lines, knockout lines show consistently slower growth than empty vector controls under both conditions, particularly 19C CL.

Fig. S11. Measured photo-physiology of glycolysis knockout lines A: Curves for **(i-ii)** relative electron transport (rETR) of photosystem II fitted as a function of light intensity) and **(iii-iv)** photoprotective non-photochemical quenching (NPQ) fitted as a function of E. Separate values are shown for cultures in CL **(i, iii)** and LD **(ii, iv)** growth conditions. Data points are the mean between the average values (n=2-4) measured in each strain within each genotype (Control = 2, cpEnolase complemented = 2, cpPGAM1A complemented = 3, cpEnolase knockout = 6, cpPGAM1A knockout = 3). **B:** Violin plots of PSII functional absorption cross-section (σPSII), measured with a MINIFIRe spectrometer for glycolysis mutant versus control lines under each growth conditions. Significantly different values observed for knockout and complementation mutants relative to control lines (one-way ANOVA, $P < 0.05$) are asterisked, with asterisk colour corresponding to the line considered. Each boxplot includes all measured/ fitted values for each strain within a mutant line. The absence of clear photosynthetic defects contrasts with the diminished growth of knockout lines, as per **Fig. 3.**

Fig. S12: Bar plots of the mean and standard deviation of the ratios of 39 metabolites assessed by GC-MS in plastid glycolysis mutant lines under the three tested experimental conditions. Data support the Volcano Plots shown in **Fig. 5**. Metabolites are sorted in ranked decreasing accumulation in mutant lines over all three conditions. Metabolites inferred to be differentially accumulated (one-way ANOVA) in each mutant line and condition are asterisked.

Fig. S13. Lipid accumulation profiles under 19C LD conditions. **A:** Volcano Plots showing (horizontal axis) $\log_2$ accumulation ratios and (vertical axis) $-\log_{10}$ one-way ANOVA, two-tailed Pvalues for separation of mean proportions of specific fatty acids, across all fatty acids observed in a specific lipid class in glycolysis mutants versus control lines, supporting the global scatter- and violin plots shown in **Fig. 6**. Specific lipids that show extreme ($P < 0.01$) differences in accumulation between both mutant genotypes and control lines are labelled, and coloured by lipid class. **B:** Bar plots showing total DGTA lipid class distributions in all three lines under these conditions. These data suggest limited changes in glycolysis mutant lipid architecture, barring a probable over-accumulation of *sn*-1 C16 in gycolysis mutant lipid pools, and corresponding under-accumulation of *sn*-1 C20 in mutant DGTA pools.

Fig. S14. Lipid accumulation profiles under 19C CL conditions. A: Volcano plots showing (horizontal axis) $\log_2$ accumulation ratios and (vertical axis) $-\log_{10}$ one-way ANOVA, two-tailed Pvalues for separation of mean proportions of specific fatty acids, across all fatty acids observed in a specific lipid class in glycolysis mutants versus control lines, and **B: bar plots** of SQDG and DGTA accumulation in lines harvested under **19C CL.** Data are shown as per **Fig. S13** and support global scatter- and violin plots shown in **Fig. 6.** These data suggest similar changes in glycolysis mutant lipid architecture to **19C LD**, including probable over-accumulations of *sn*-1 C16 in lieu of C20 and C14 in cpEnolase and cpPGAM1A mutant SQDG and MGDG pools.

Fig. S15. Lipid accumulaiton profiles under 8C CL conditions. Volcano plots showing (horizontal axis) $\log_2$ accumulation ratios and (vertical axis) $-\log_{10}$ ANOVA Pvalues for separation of mean proportions of specific fatty acids, across all fatty acids observed in a specific lipid class in cpEnolase mutants versus control lines, and

cpEnolase mutants versus cpPGAM1A mutants harvested under 8C CL conditions. Data are shown as per **Fig. S13** and support global scatter- and violin plots shown in **Fig. 6.** No sigificantly differentially accumulated ($P < 10^{-05}$) lipids were observed in corresponding comparisons of cpPGAM1A mutants and control lines. These data suggest specific overaccumulations in short-chain *sn*-1 MGDG, and *sn*-2 SQDG, and C20 *sn-1* DGTA in cpEnolase mutants compared to other lines.

Fig. S16. Schematic diagram of the reaction kinetics measured for *P. tricornutum* cpEnolase and cPGAM1A enzymes. The measured activities of this assay are shown in **Fig. 7A.** Common enzymes are shown in green, enzymes unique to the glycolytic assay in blue, and enzymes unique to the gluconeogenic assay in red. Reaction intermediates and reversible substrates are shown in yellow and gray respectively

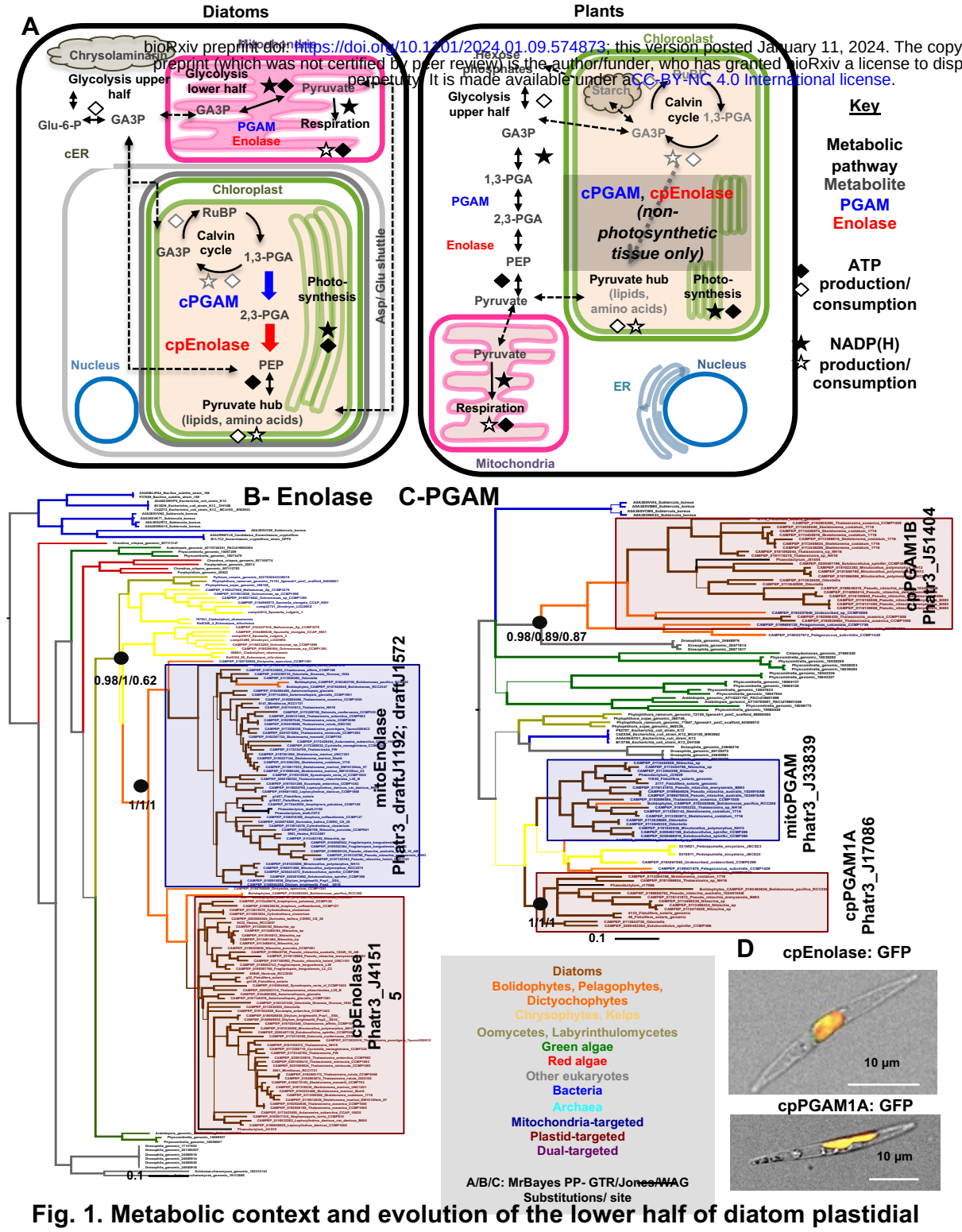


Fig. 1. Metabolic context and evolution of the lower half of diatom plastidial glycolysis-gluconeogenesis. A: schematic comparison of diatom and plant core carbon metabolism, highlighting the localization and functions of two enzymes in the lower half of glycolysis-gluconeogenesis (phospho-glycerate mutase, and enolase) whose localization to the chloroplast can connect endogenous enzymes in the Calvin cycle and pyruvate hub to create a complete glycolytic-gluconeogenic-gluconeogenic pathway. Abbreviations: GA3P- glyceraldehyde-3-phosphate; 1,3-PGA and 2,3-PGA- 1,3 and 2,3 bis-phosphoglycerate; Glu-6-P- glucose-6-phosphate; PEP- phospho-enol-pyruvate; RuBP- ribulose *bis*-phosphate; PGAM- phospho-glycerate mutase; cER- chloroplast: endoplasmic reticulum. **B, C:** consensus MrBayes topologies realised with three substitution matrices (GTR, Jones, WAG) of a 163 taxa x 413 aa alignment of organelle-targeted enolase and 105 taxa x 220 aa alignment of selected organelle-targeted PGAM1 enzymes from diatoms and their closest relatives, identifying recent duplications and recruitments of respiratory glycolytic-gluconeogenic enzymes from the mitochondria to plastid in diatoms and their closest relatives. **D:** overlay images of GFP-tagged full-length cpEnolase (top) and cpPGAM1A (bottom) constructs (green), chlorophyll (red) and bright-field images of transformant *Phaeodactylum tricornutum* lines. Scale bar: 10 µm.

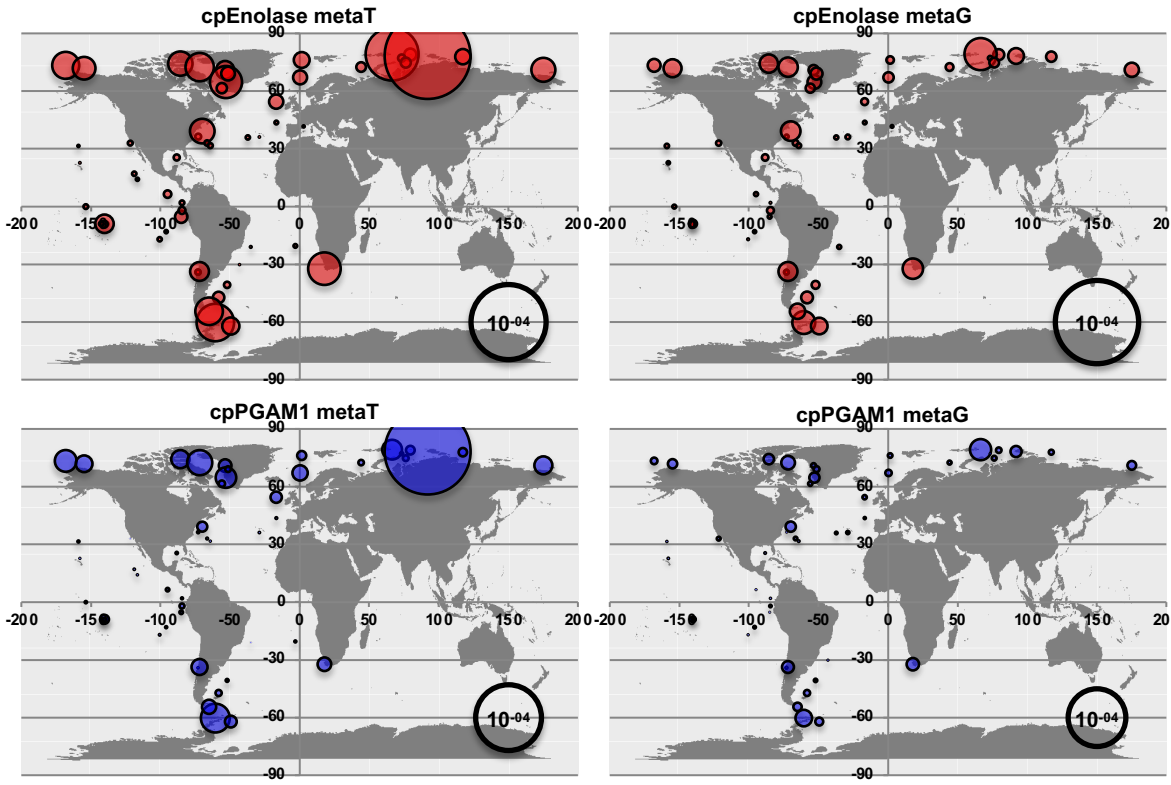


Fig. 2. Environmental distributions of diatom plastidial lower half glycolysis-gluconeogenesis meta-genes. Total transcriptome (top) and genome (bottom) relative abundances for *Tara* Oceans meta-genes phylogenetically resolved to diatom cpEnolase and cpPGAM1, sampled from all size fractions and surface layer stations, demonstrating higher meta-transcript abundance without commensurate increases in meta-gene abundance at high northern and southern latitudes.

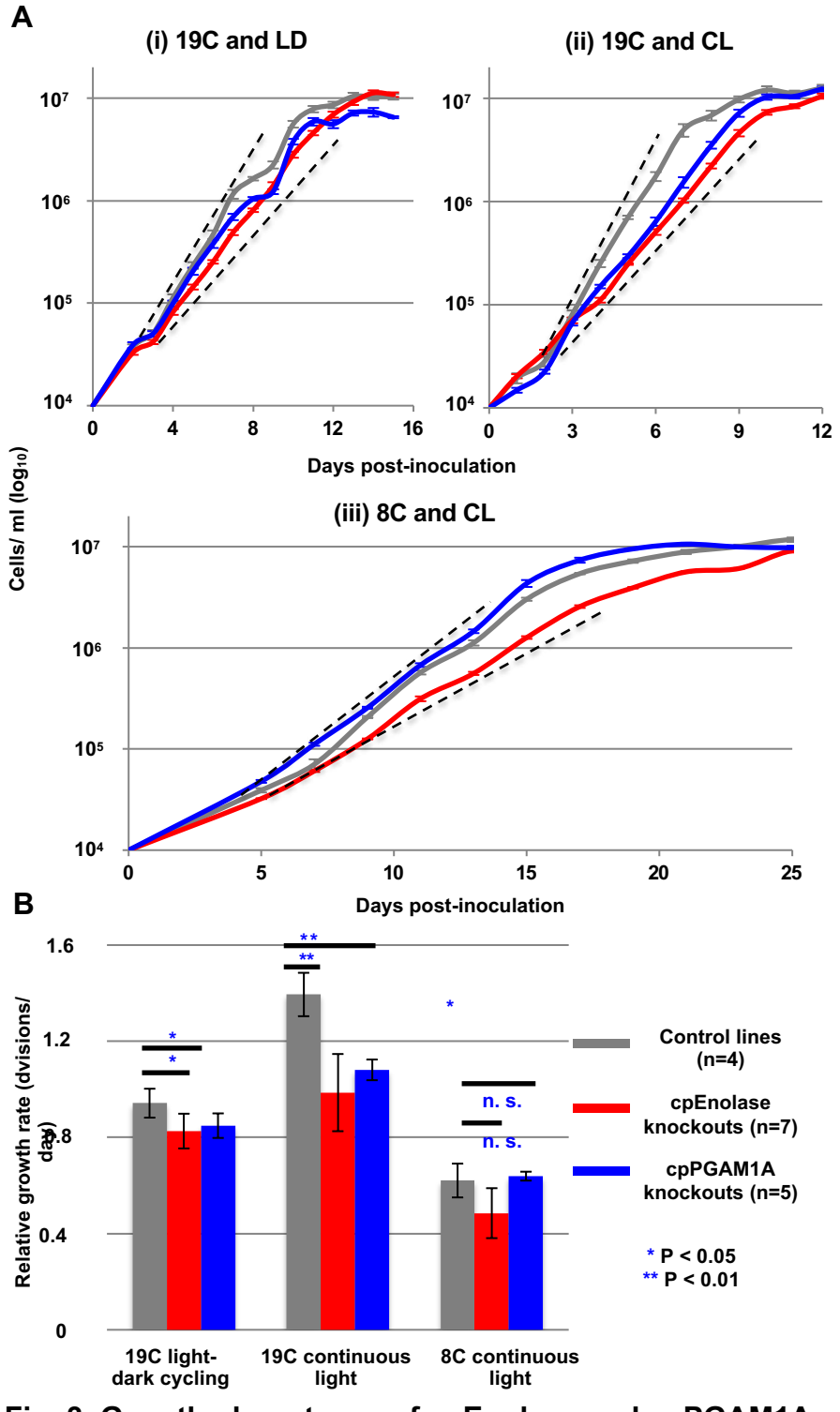


Fig. 3. Growth phenotypes of cpEnolase and cpPGAM1A CRISPR-Cas9 knockout mutant and zeocin-resistant empty vector control *P. tricornutum* lines. A: exemplar growth curves from single experiments realised for *P. tricornutum* lines in 50 µE m^{-2} s^{-1} illumination, non-shaken cultures and replete ESAW media, under three conditions- **(i)** 19°C and 12h light: 12h dark Circadian cycles (« 19C LD »); **(ii)** 19°C and 24h continuous light (« 19C CL »); and **(iii)** 8°C and 24h continuous light (« 8C CL »). Hashed black lines show the approximative concentrations (between 5 x10^4 and 4 x 10^6 cells ml^{-1}) over which growth rates were calculated). **B:** mean relative log phase growth rates of each genotype under each condition, measured through a minimum of three biological replicates and two technical repetitions (six measurements per line, minimum 24 measurements per genotype), over five time-points with linear (r^2 > 0.95 relationship between log cell density and time). Asterisks indicate significant differences as inferred by one-way ANOVA. An alternative version of this figure showing absolute growth rates of individual cell lines is provided in **Fig. S10**.

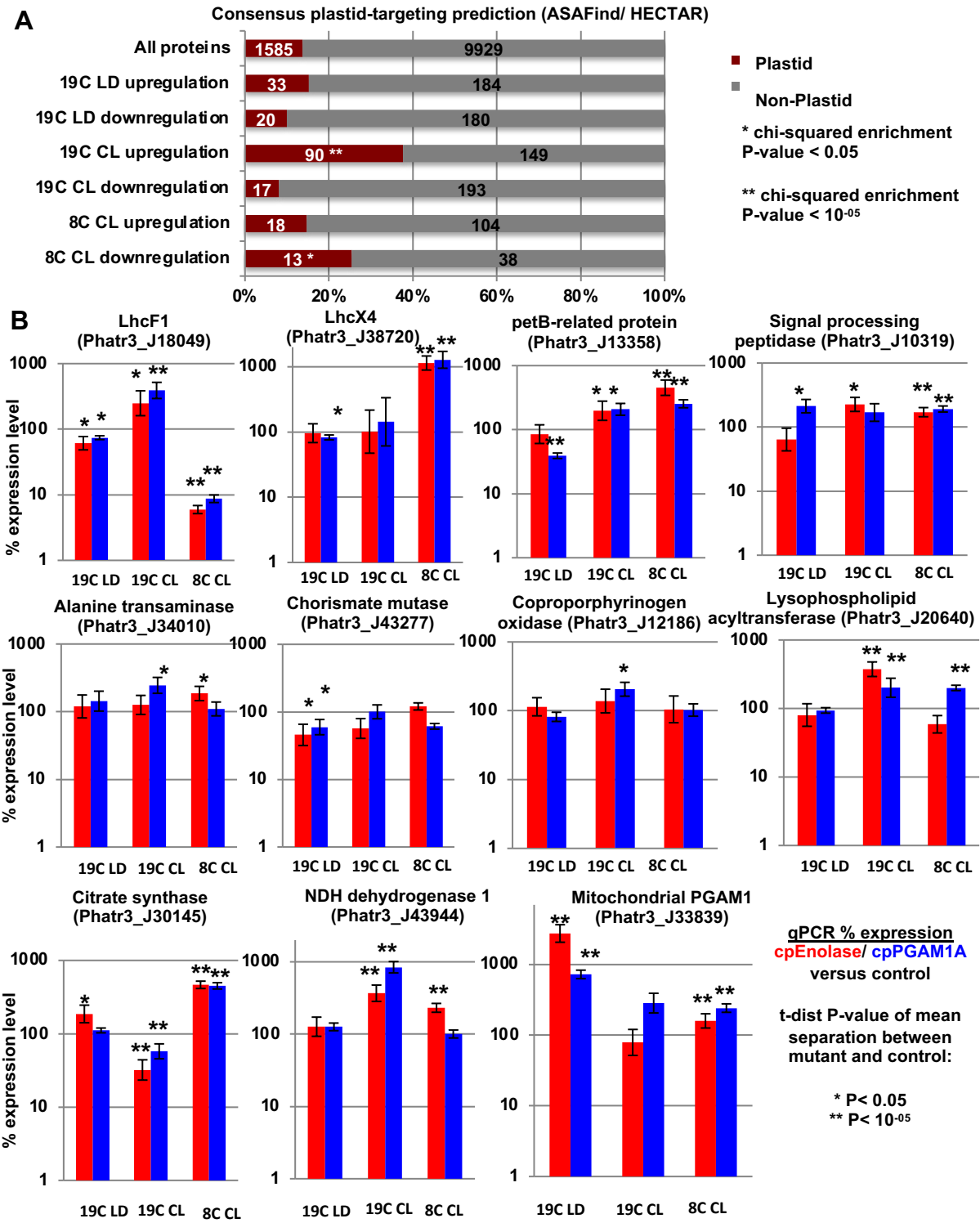


Fig. 4. Changes in plastid and mitochondrial metabolic architecture inferred from gene expression analyses. A: predicted consensus localizations (either: chloroplast, or non-chloroplast) from ASAFind (Gruber, Rocap et al. 2015) and HECTAR (Gschloessl, Guermeur et al. 2008) of all genes inferred ($P < 0.05$, fold-change expression >2) to be up- or down-regulated in both cpEnolase and cpPGAM1A knockout compared to control lines under 19C LD, 19C CL and 8C CL conditions. Significantly enriched localisations (two-tailed chi-squared test) are asterisked. **B:** relative mRNA abundances of eleven genes encoding exemplar chloroplast- and mitochondria-targeted proteins, verified by qRT-PCR. Genes differentially expressed (*t*-test, $P < 0.05$) in each condition are asterisked.

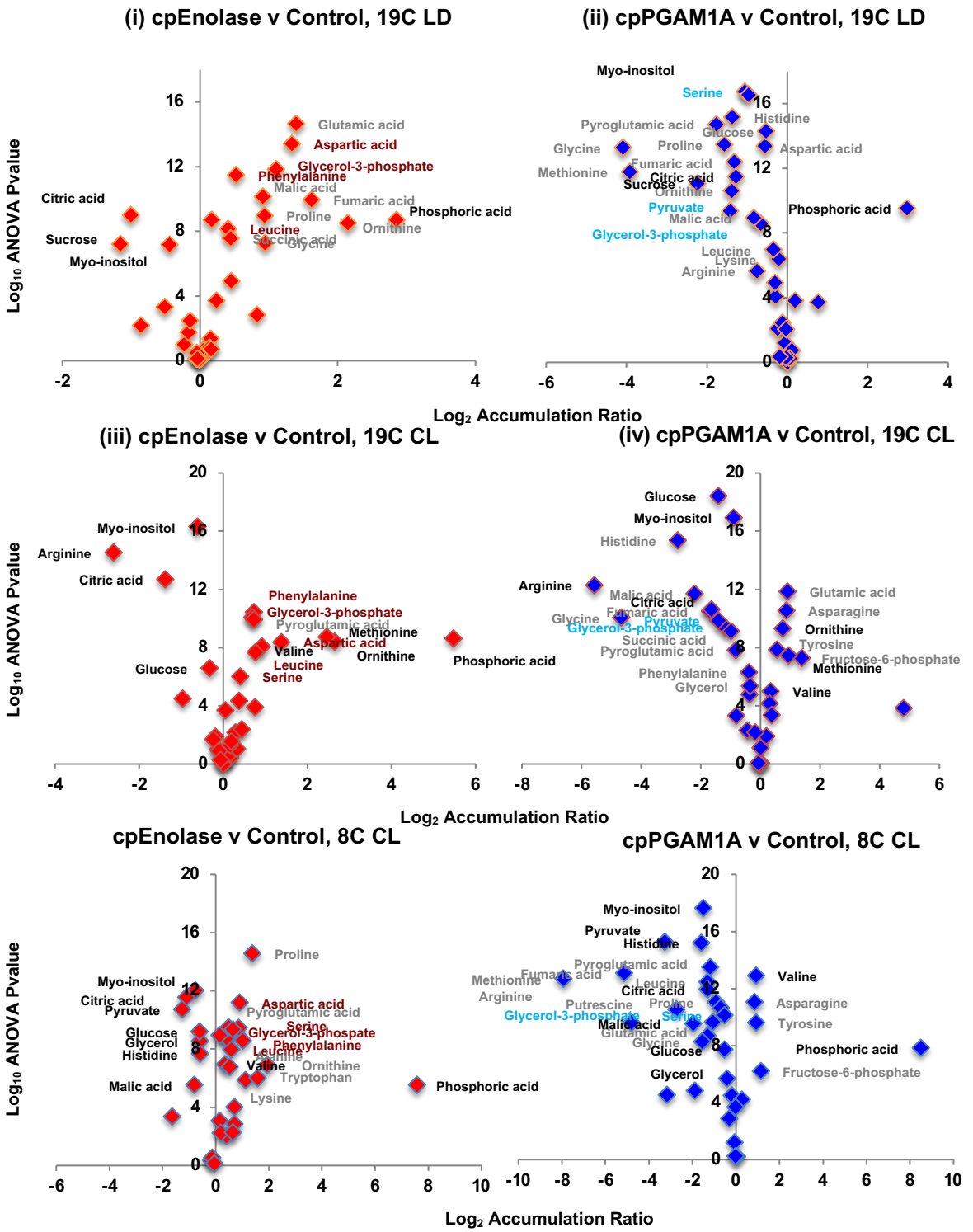
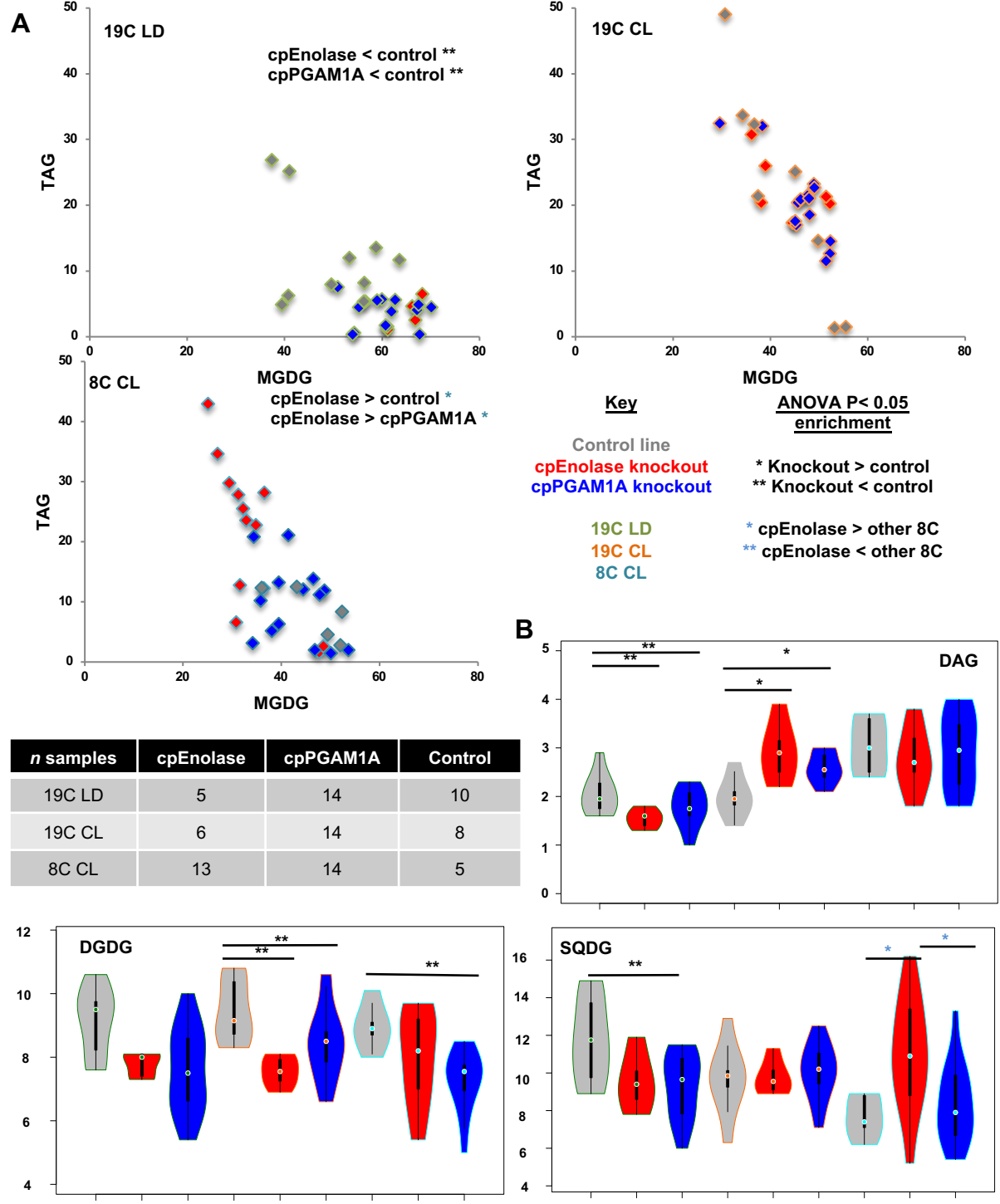


Fig. 5. Volcano plots of differentially accumulated metabolites assessed by GC-MS. Scatterplots of the $\log_2$ accumulation ratios and $-\log_{10}$ P-values of difference in the mass, ribitol and quality-control-normalised abundances of 39 sugar and amino acid metabolites in cpEnolase and cpPGAM1A knockout compared to empty vector control lines, measured by GC-MS in all three experimental conditions tested. Metabolites that show a differential accumulation in each plot ($P < 10^{-05}$) are labelled, with metabolites that show a differential accumulation in both knockout lines in each condition shown in black text, and five metabolites that are uniquely over-accumulated in cpEnolase knockout lines under all three conditions shown in dark red text.



n samples	cpEnolase	cpPGAM1A	Control
19C LD	5	14	10
19C CL	6	14	8
8C CL	13	14	5

Fig. 6. LC-MS lipid distributions in glycolysis-gluconeogenesis mutant lines. **A:** scatterplots of relative proportions of MGDG and TAG in total lipid LC-MS samples in cpEnolase and cpPGAM1A knockout lines and empty vector controls under each growth condition, showing increased MGDG: TAG in glycolysis knockout lines under 19C, and the inverse relationship in cpEnolase knockout lines only under 8C. **B**: violin plots of relative abundances of three further lipid categories inferred to differentially accumulate in glycolysis knockout lines under different growth conditions. Significant differences between knockout and control lines (one-way ANOVA) are asterisked.

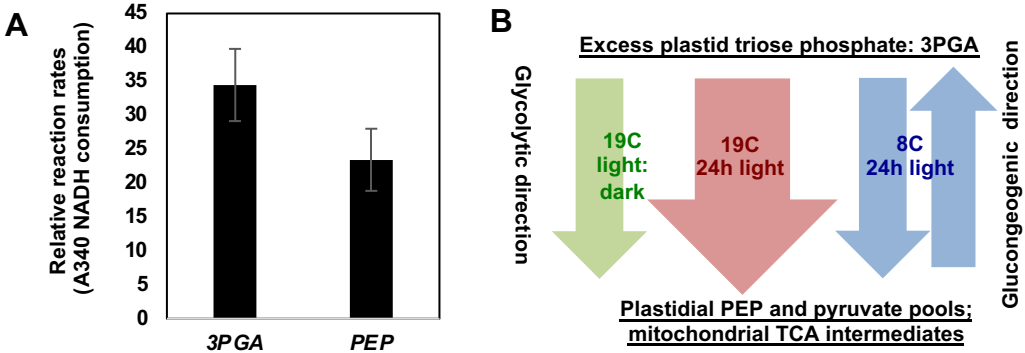


Fig. 7. Inferred kinetic activities of *P. tricornutum* plastid lower-half glycolysis-gluconeogenesis. A : Measured reaction rates of purified cpEnolase and cPGAM1A supplemented with 3-PGA (glycolytic direction) or PEP (gluconeogenic direction). **B :** schematic diagram showing the inferred fluxes through lower half diatom plastid glycolysis-gluconeogenesis in each environmental condition tested.

Parsed Citations

Abida, H., L. J. Dolch, C. Meï, V. Villanova, M. Conte, M. A. Block, G. Finazzi, O. Bastien, L. Tirichine, C. Bowler, F. Rébeillé, D. Petroutsos, J. Jouhet and E. Maréchal (2015). Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in Phaeodactylum tricornutum. Plant Physiol 167(1): 118-136.
Google Scholar: Author Only Title Only Author and Title

Ait-Mohamed, O., A. M. G. Novák Vanclová, N. Joli, Y. Liang, X. Zhao, A. Genovesio, L. Tirichine, C. Bowler and R. G. Dorrell (2020). PhaeoNet: a holistic RNAseq-based portrait of transcriptional coordination in the model diatom Phaeodactylum tricornutum. Frontiers Plant Sci 11: 590949.
Google Scholar: Author Only Title Only Author and Title

Allen, A. E., C. L. Dupont, M. Obornik, A. Horak, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie and C. Bowler (2011). Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473(7346): 203-207.
Google Scholar: Author Only Title Only Author and Title

Almagro Armenteros, J. J., K. D. Tsirigos, C. K. Sønderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne and H. Nielsen (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnol 37(4): 420-423.
Google Scholar: Author Only Title Only Author and Title

Andriotis, V. M., N. J. Kruger, M. J. Pike and A. M. Smith (2010). Plastidial glycolysis in developing Arabidopsis embryos. New Phytol 185(3): 649-662.
Google Scholar: Author Only Title Only Author and Title

Anoman, A. D., M. Flores-Tornero, S. Rosa-Telléz, J. Muñoz-Bertomeu, J. Segura and R. Ros (2016). The specific role of plastidial glycolysis in photosynthetic and heterotrophic cells under scrutiny through the study of glyceraldehyde-3-phosphate dehydrogenase. Plant Signal Behav 11(3): 1128614.
Google Scholar: Author Only Title Only Author and Title

Bai, Y., T. Cao, O. Dautermann, P. Buschbeck, M. B. Cantrell, Y. Chen, C. D. Lein, X. Shi, M. A. Ware, F. Yang, H. Zhang, L. Zhang, G. Peers, X. Li and M. Lohr (2022). Green diatom mutants reveal an intricate biosynthetic pathway of fucoxanthin. Proc Natl Acad Sci USA 119(38): 2203708119.
Google Scholar: Author Only Title Only Author and Title

Bailleul, B., N. Berne, O. Murik, D. Petroutsos, J. Prihoda, A. Tanaka, V. Villanova, R. Bligny, S. Flori, D. Falconet, A. Krieger-Liszkay, S. Santabarbara, F. Rappaport, P. Joliot, L. Tirichine, P. G. Falkowski, P. Cardol, C. Bowler and G. Finazzi (2015). Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. Nature 524(7565): 366-U267.
Google Scholar: Author Only Title Only Author and Title

Behrenfeld, M. J., K. H. Halsey, E. Boss, L. Karp-Boss, A. J. Milligan and G. Peers (2021). Thoughts on the evolution and ecological niche of diatoms. Ecol Monographs 91(3): 01457.
Google Scholar: Author Only Title Only Author and Title

Broddrick, J. T., N. Du, S. R. Smith, Y. Tsuji, D. Jallet, M. A. Ware, G. Peers, Y. Matsuda, C. L. Dupont, B. G. Mitchell, B. O. Palsson and A. E. Allen (2019). Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom Phaeodactylum tricornutum. New Phytol 222: 1364-1379.
Google Scholar: Author Only Title Only Author and Title

Bromke, M. A. (2013). Amino Acid biosynthesis pathways in diatoms. Metabolites 3(2): 294-311.
Google Scholar: Author Only Title Only Author and Title

Buck, J. M., C. Río Bártulos, A. Gruber and P. G. Kroth (2018). Blasticidin-S deaminase, a new selection marker for genetic transformation of the diatom. PeerJ 6: 5884.
Google Scholar: Author Only Title Only Author and Title

Buck, J. M., J. Sherman, C. R. Bártulos, M. Serif, M. Halder, J. Henkel, A. Falciatore, J. Lavaud, M. Y. Gorbunov, P. G. Kroth, P. G. Falkowski and B. Lepetit (2019). Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom Phaeodactylum tricornutum. Nat Commun 10(1): 4167.
Google Scholar: Author Only Title Only Author and Title

Buseman, C. M., P. Tamura, A. A. Sparks, E. J. Baughman, S. Maatta, J. Zhao, M. R. Roth, S. W. Esch, J. Shah, T. D. Williams and R. Welti (2006). Wounding stimulates the accumulation of glycerolipids containing Oxophytodienoic Acid and Dinor-Oxophytodienoic Acid in Arabidopsis leaves. Plant Physiol 142(1): 28-39.
Google Scholar: Author Only Title Only Author and Title

Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer (2009). The MIQE Guidelines: Minimum Information for publication of Quantitative real-time PCR Experiments. Clinical Chemistry 55(4): 611-622.
Google Scholar: Author Only Title Only Author and Title

Capella-Gutiérrez, S., J. M. Silla-Martínez and T. Gabaldón (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15): 1972-1973.
Google Scholar: Author Only Title Only Author and Title

Carradec, Q., E. Pelletier, C. Da Silva, A. Alberti, Y. Seeleuthner, R. Blanc-Mathieu, G. Lima-Mendez, F. Rocha, L. Tirichine, K. Labadie, A. Kirilovsky, A. Bertrand, S. Engelen, M. A. Madoui, R. Méheust, J. Poulain, S. Romac, D. J. Richter, G. Yoshikawa, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, O. Jaillon, J. M. Aury, E. Karsenti, M. B. Sullivan, S. Sunagawa, P. Bork, F. Not, P. Hingamp, J. Raes, L. Guidi, H. Ogata, C. de Vargas, D. Iudicone, C. Bowler, P. Wincker and T. O. Coordinators (2018). A global ocean atlas of eukaryotic genes. Nat Commun 9(1): 373.
Google Scholar: Author Only Title Only Author and Title

Carrera, D., G. M. George, M. Fischer-Stettler, F. Galbier, S. Eicke, E. Truernit, S. Streb and S. C. Zeeman (2021). Distinct plastid fructose bisphosphate aldolases function in photosynthetic and non-photosynthetic metabolism in Arabidopsis. J Exp Bot 72(10): 3739-3755.
Google Scholar: Author Only Title Only Author and Title

Chang, Y. F., J. S. Imam and M. F. Wilkinson (2007). The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76: 51-74.
Google Scholar: Author Only Title Only Author and Title

Cruz de Carvalho, M. H., H. X. Sun, C. Bowler and N. H. Chua (2016). Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. New Phytol 210(2): 497-510.
Google Scholar: Author Only Title Only Author and Title

Delmont, T. O., M. Gaia, D. D. Hinsinger, P. Frémont, C. Vanni, A. Fernandez-Guerra, A. M. Eren, A. Kourlaiev, L. d'Agata, Q. Clayssen, E. Villar, K. Labadie, C. Cruaud, J. Poulain, C. Da Silva, M. Wessner, B. Noel, J.-M. Aury, S. Sunagawa, S. G. Acinas, P. Bork, E. Karsenti, C. Bowler, C. Sardet, L. Stemmann, C. de Vargas, P. Wincker, M. Lescot, M. Babin, G. Gorsky, N. Grimsley, L. Guidi, P. Hingamp, O. Jaillon, S. Kandels, D. Iudicone, H. Ogata, S. Pesant, M. B. Sullivan, F. Not, K.-B. Lee, E. Boss, G. Cochrane, M. Follows, N. Poulton, J. Raes, M. Sieracki, S. Speich and E. Pelletier (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. Cell Genom 2(5): 100123.
Google Scholar: Author Only Title Only Author and Title

Demé, B., C. Cataye, M. A. Block, E. Maréchal and J. Jouhet (2014). Contribution of galactoglycerolipids to the 3-dimensional architecture of thylakoids. FASEB J 28(8): 3373-3383.
Google Scholar: Author Only Title Only Author and Title

Dolch, L. J., J. Lupette, G. Tourcier, M. Bedhomme, S. Collin, L. Magneschi, M. Conte, K. Seddiki, C. Richard, E. Corre, L. Fourage, F. Laeuffer, R. Richards, M. Reith, F. Rébeillé, J. Jouhet, P. McGinn and E. Maréchal (2017). Nitric oxide mediates nitrite-sensing and acclimation and triggers a remodeling of lipids. Plant Physiol 175(3): 1407-1423.
Google Scholar: Author Only Title Only Author and Title

Dolch, L. J. and E. Maréchal (2015). Inventory of fatty acid desaturases in the pennate diatom Phaeodactylum tricornutum. Mar Drugs 13(3): 1317-1339.
Google Scholar: Author Only Title Only Author and Title

Dorrell, R. G., A. M. G. Novak Vanclova, M. Penot, J. J. Pierella Karlusich, C. Bowler, S. Liu, E. Maréchal, J. Jouhet, B. Bailleul and D. Croteau (2022). Functional physiology of novel diatom chloroplast proteins. Open Science Foundation. https://osf.io/89vm3/.
Google Scholar: Author Only Title Only Author and Title

Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocol 2(4): 953-971.
Google Scholar: Author Only Title Only Author and Title

Erdene-Ochir, E., B. K. Shin, B. Kwon, C. Jung and C. H. Pan (2019). Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from Phaeodactylum tricornutum. Sci Rep 9(1): 9941.
Google Scholar: Author Only Title Only Author and Title

Falciatore, A., R. Casotti, C. Leblanc, C. Abrescia and C. Bowler (1999). Transformation of Nonselectable Reporter Genes in Marine Diatoms. Mar Biotechnol 1(3): 239-251.
Google Scholar: Author Only Title Only Author and Title

Folch, J., M. Lees and G. H. S. Stanley (1957). A somple method for the isolation and purification of total lipids from animal tissues. J Biol Chem 226(1): 497-509.
Google Scholar: Author Only Title Only Author and Title

Friedlingstein, P., M. W. Jones, M. O'Sullivan, R. M. Andrew, D. C. Bakker, J. Hauck, C. Le Quéré, G. P. Peters, W. Peters and J. Pongratz (2022). Global carbon budget 2021. Earth System Science Data 14(4): 1917-2005.
Google Scholar: Author Only Title Only Author and Title

Fukasawa, Y., J. Tsuji, S. C. Fu, K. Tomii, P. Horton and K. Imai (2015). MitoFates: improved prediction of mitochondrial targeting

sequences and their cleavage sites. Mol Cell Proteom 14(4): 1113-1126.
Google Scholar: Author Only Title Only Author and Title

Fukayama, H., C. Masumoto, Y. Taniguchi, A. Baba-Kasai, Y. Katoh, H. Ohkawa and M. Miyao (2015). Characterization and expression analyses of two plastidic enolase genes in rice. Biosci Biotechnol Biochem 79(3): 402-409.
Google Scholar: Author Only Title Only Author and Title

Fuss, J., O. Liegmann, K. Krause and S. A. Rensing (2013). Green Targeting Predictor and Ambiguous Targeting Predictor 2: the pitfalls of plant protein targeting prediction and of transient protein expression in heterologous systems. New Phytol 200: 222-233.
Google Scholar: Author Only Title Only Author and Title

Gilbertson, R., E. Langan and T. Mock (2022). Diatoms and their microbiomes in complex and changing polar oceans. Front Microbiol 13: 786764.
Google Scholar: Author Only Title Only Author and Title

Gorbunov, M. Y., E. Shirsin, E. Nikonova, V. V. Fadeev and P. G. Falkowski (2020). A multi-spectral fluorescence induction and relaxation (FIRe) technique for physiological and taxonomic analysis of phytoplankton communities. Mar Ecol Progress Series 644: 1-13.
Google Scholar: Author Only Title Only Author and Title

Grigoriev, I. V., R. D. Hayes, S. Calhoun, B. Kamel, A. Wang, S. Ahrendt, S. Dusheyko, R. Nikitin, S. J. Mondo, A. Salamov, I. Shabalov and A. Kuo (2021). PhycoCosm, a comparative algal genomics resource. Nucl Acids Res 49(D1): D1004-D1011.
Google Scholar: Author Only Title Only Author and Title

Gruber, A., G. Rocap, P. G. Kroth, E. V. Armbrust and T. Mock (2015). Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. Plant J 81(3): 519-528.
Google Scholar: Author Only Title Only Author and Title

Gschloessl, B., Y. Guermeur and J. M. Cock (2008). HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinformatics 9: 393.
Google Scholar: Author Only Title Only Author and Title

Hartley, J. L., G. F. Temple and M. A. Brasch (2000). DNA cloning using in vitro site-specific recombination. Genome Res 10(11): 1788-1795.
Google Scholar: Author Only Title Only Author and Title

Huang, A., L. Liu, C. Yang and G. Wang (2015). Phaeodactylum tricornutum photorespiration takes part in glycerol metabolism and is important for nitrogen-limited response. Biotechnol Biofuels 8(1): 1-16.
Google Scholar: Author Only Title Only Author and Title

Huang, T., Y. Pan, E. Maréchal and H. Hu (2023). Proteomes reveal the lipid metabolic network in the complex plastid of Phaeodactylum tricornutum. Plant J in press: 10.1111/tpj.16477.
Google Scholar: Author Only Title Only Author and Title

Initiative, O. T. P. T. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574(7780): 679-685.
Google Scholar: Author Only Title Only Author and Title

Jassby, A. D. and T. Platt (1976). Mathematical formulation of the relationship between photosynthesis and light for phytoplankton. Limnol Oceanography 21(4): 540-547.
Google Scholar: Author Only Title Only Author and Title

Joli, N., L. Concia, K. Mocaer, J. Guterman, J. Laude, S. Guerin, T. Sciandra, F. Bruyant, O. Ait-Mohamed, M. Beguin, M.-H. Forget, C. Bourbousse, T. Lacour, B. Bailleul, C. Nef, M. Savoie, J.-E. Tremblay, D. A. Campbell, J. Lavaud, Y. Schwab, M. Babin and C. Bowler (2023). Hypometabolism to survive the long polar night and subsequent successful return to light in the diatom Fragilariopsis cylindrus. New Phytol in press: 10.1111/nph.19387.
Google Scholar: Author Only Title Only Author and Title

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez and S. Hunter (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9): 1236-1240.
Google Scholar: Author Only Title Only Author and Title

Jouhet, J., J. Lupette, O. Clerc, L. Magneschi, M. Bedhomme, S. Collin, S. Roy, E. Maréchal and F. Rébeillé (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(8): 0182423.
Google Scholar: Author Only Title Only Author and Title

Jouhet, J., E. Maréchal, R. Bligny, J. Joyard and M. A. Block (2003). Transient increase of phosphatidylcholine in plant cells in

response to phosphate deprivation. FEBS Lett 544(1): 63-68.
Google Scholar: Author Only Title Only Author and Title

Kassambara, A. and F. Mundt (2017). Package 'factoextra'. Extract and visualize the results of multivariate data analyses 76(2).
Google Scholar: Author Only Title Only Author and Title

Katoh, K., J. Rozewicki and K. D. Yamada (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform.
Google Scholar: Author Only Title Only Author and Title

Kazamia, E., R. Sutak, J. Paz-Yepes, R. G. Dorrell, F. R. J. Vieira, J. Mach, J. Morrissey, S. Leon, F. Lam, E. Pelletier, J. M. Camadro, C. Bowler and E. Lesuisse (2018). Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms. Sci Adv 4(5): 4536.
Google Scholar: Author Only Title Only Author and Title

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647-1649.
Google Scholar: Author Only Title Only Author and Title

Keeling, P. J., F. Burki, H. M. Wilcox, B. Allam, E. E. Allen, L. A. Amaral-Zettler, E. V. Armbrust, J. M. Archibald, A. K. Bharti, C. J. Bell, B. Beszteri, K. D. Bidle, C. T. Cameron, L. Campbell, D. A. Caron, R. A. Cattolico, J. L. Collier, K. Coyne, S. K. Davy, P. Deschamps, S. T. Dyhrman, B. Edvardsen, R. D. Gates, C. J. Gobler, S. J. Greenwood, S. M. Guida, J. L. Jacobi, K. S. Jakobsen, E. R. James, B. Jenkins, U. John, M. D. Johnson, A. R. Juhl, A. Kamp, L. A. Katz, R. Kiene, A. Kudryavtsev, B. S. Leander, S. Lin, C. Lovejoy, D. Lynn, A. Marchetti, G. McManus, A. M. Nedelcu, S. Menden-Deuer, C. Miceli, T. Mock, M. Montresor, M. A. Moran, S. Murray, G. Nadathur, S. Nagai, P. B. Ngam, B. Palenik, J. Pawlowski, G. Petroni, G. Piganeau, M. C. Posewitz, K. Rengefors, G. Romano, M. E. Rumpho, T. Rynearson, K. B. Schilling, D. C. Schroeder, A. G. Simpson, C. H. Slamovits, D. R. Smith, G. J. Smith, S. R. Smith, H. M. Sosik, P. Stief, E. Theriot, S. N. Twary, P. E. Umale, D. Vaulot, B. Wawrik, G. L. Wheeler, W. H. Wilson, Y. Xu, A. Zingone and A. Z. Worden (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12(6): 1001889.
Google Scholar: Author Only Title Only Author and Title

Kopka, J., N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmüller, P. Dörmann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie and D. Steinhauser (2005). GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21(8): 1635-1638.
Google Scholar: Author Only Title Only Author and Title

Kroth, P. G., A. Chiovitti, A. Gruber, V. Martin-Jezequel, T. Mock, M. S. Parker, M. S. Stanley, A. Kaplan, L. Caron, T. Weber, U. Maheswari, E. V. Armbrust and C. Bowler (2008). A model for carbohydrate metabolism in the diatom Phaeodactylum tricornutum deduced from comparative whole genome analysis. PLoS One 3(1): 1426.
Google Scholar: Author Only Title Only Author and Title

Lacour, T., J. Larivière and M. Babin (2017). Growth, Chl a content, photosynthesis, and elemental composition in polar and temperate microalgae. Limnol Oceanog 62(1): 43-58.
Google Scholar: Author Only Title Only Author and Title

Lampe, R. H., G. Hernandez, Y. Y. Lin and A. Marchetti (2021). Representative diatom and coccolithophore species exhibit divergent responses throughout simulated upwelling cycles. mSystems 6(2): 10.1128/msystems.00188-21.
Google Scholar: Author Only Title Only Author and Title

Lataretu, M. and M. Hölzer (2020). RNAflow: an effective and simple RNA-Seq differential gene expression pipeline using Nextflow. Genes (Basel) 11(12).
Google Scholar: Author Only Title Only Author and Title

Lepetit, B., D. A. Campbell, J. Lavaud, C. Büchel, R. Goss and B. Bailleul (2022). Photosynthetic Light Reactions in diatoms. II. The dynamic regulation of the various light reactions. The Molecular Life of Diatoms. A. Falciatore and T. Mock. Cham, Springer International Publishing: 423-464.
Google Scholar: Author Only Title Only Author and Title

Levering, J., J. Broddrick, C. L. Dupont, G. Peers, K. Beeri, J. Mayers, A. A. Gallina, A. E. Allen, B. O. Palsson and K. Zengler (2016). Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. PloS One 11(5): 0155038.
Google Scholar: Author Only Title Only Author and Title

Lisec, J., N. Schauer, J. Kopka, L. Willmitzer and A. R. Fernie (2006). Gas chromatography mass spectrometry–based metabolite profiling in plants. Nature Protocols 1(1): 387-396.
Google Scholar: Author Only Title Only Author and Title

Liu, S., M. Storti, C. Bowler, G. Finazzi and R. G. Dorrell (2022). An integrative environmental atlas of diatom chloroplast transporters. Front Plant Sci 13: 950467.
Google Scholar: Author Only Title Only Author and Title

Liu, S., Z. Wang, R. Zhu, F. Wang, Y. Cheng and Y. Liu (2021). Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. J Vis Exp(175).
Google Scholar: Author Only Title Only Author and Title

Luedemann, A., L. von Malotky, A. Erban and J. Kopka (2012). TagFinder: preprocessing software for the fingerprinting and the profiling of gas chromatography-mass spectrometry based metabolome analyses. Methods Mol Biol 860: 255-286.
Google Scholar: Author Only Title Only Author and Title

Malviya, S., E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. de Vargas, L. Bittner, A. Zingone and C. Bowler (2016). Insights into global diatom distribution and diversity in the world's ocean. Proc Natl Acad Sci USA 113(11): 1516-1525.
Google Scholar: Author Only Title Only Author and Title

Maréchal, E. and J. Lupette (2020). Relationship between acyl-lipid and sterol metabolisms in diatoms. Biochimie 169: 3-11.
Google Scholar: Author Only Title Only Author and Title

McCarthy, J. K., S. R. Smith, J. P. McCrow, M. Tan, H. Zheng, K. Beeri, R. Roth, C. Lichtle, U. Goodenough, C. P. Bowler, C. L. Dupont and A. E. Allen (2017). Nitrate Reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. Plant Cell 29(8): 2047-2070.
Google Scholar: Author Only Title Only Author and Title

Miller, M. A., T. Schwartz, B. E. Pickett, S. He, E. B. Klem, R. H. Scheuermann, M. Passarotti, S. Kaufman and M. A. O'Leary (2015). A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. Evol Bioinform Online 11: 43-48.
Google Scholar: Author Only Title Only Author and Title

Mistry, J., S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn and A. Bateman (2020). Pfam: The protein families database in 2021. Nucleic Acids Res 49: D412-D419.
Google Scholar: Author Only Title Only Author and Title

Moog, D., A. Nozawa, Y. Tozawa and R. Kamikawa (2020). Substrate specificity of plastid phosphate transporters in a non-photosynthetic diatom and its implication in evolution of red alga-derived complex plastids. Sci Rep 10(1): 1167.
Google Scholar: Author Only Title Only Author and Title

Moog, D., S. A. Rensing, J. M. Archibald, U. G. Maier and K. K. Ullrich (2015). Localization and evolution of putative triose phosphate translocators in the diatom Phaeodactylum tricornutum. Genom Biol Evol 7(11): 2955-2969.
Google Scholar: Author Only Title Only Author and Title

Murik, O., L. Tirichine, J. Prihoda, Y. Thomas, W. L. Araújo, A. E. Allen, A. R. Fernie and C. Bowler (2019). Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. New Phytol 221(3): 1303-1316.
Google Scholar: Author Only Title Only Author and Title

Nash, E. A., A. C. Barbrook, R. K. Edwards-Stuart, K. Bernhardt, C. J. Howe and R. E. R. Nisbet (2007). Organization of the mitochondrial genome in the dinoflagellate Amphidinium carterae. Mol Biol Evol 24(7): 1528-1536.
Google Scholar: Author Only Title Only Author and Title

Nawaly, H., H. Matsui, Y. Tsuji, K. Iwayama, H. Ohashi, K. Nakajima and Y. Matsuda (2023). Multiple plasma membrane SLC4s contribute to external HCO_3^- acquisition during CO_2 starvation in the marine diatom Phaeodactylum tricornutum. J Exp Bot 74(1): 296-307.
Google Scholar: Author Only Title Only Author and Title

Nonoyama, T., E. Kazamia, H. Nawaly, X. Gao, Y. Tsuji, Y. Matsuda, C. Bowler, T. Tanaka and R. G. Dorrell (2019). Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. Biomolecules 9(8): 464.
Google Scholar: Author Only Title Only Author and Title

Nymark, M., A. K. Sharma, T. Sparstad, A. M. Bones and P. Winge (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. Sci Rep 6: 24951.
Google Scholar: Author Only Title Only Author and Title

Pesant, S., F. Not, M. Picheral, S. Kandels-Lewis, N. Le Bescot, G. Gorsky, D. Iudicone, E. Karsenti, S. Speich, R. Trouble, C. Dimier, S. Searson and Tara Oceans Coordinators (2015). Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data 2.
Google Scholar: Author Only Title Only Author and Title

Popko, J., C. Herrfurth, K. Feussner, T. Ischebeck, T. Iven, R. Haslam, M. Hamilton, O. Sayanova, J. Napier, I. Khozin-Goldberg and I. Feussner (2016). Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of Phaeodactylum tricornutum. PLoS One 11(10): 0164673.
Google Scholar: Author Only Title Only Author and Title

Prabhakar, V., T. Löttgert, T. Gigolashvili, K. Bell, U. I. Flügge and R. E. Häusler (2009). Molecular and functional characterization

of the plastid-localized Phosphoenolpyruvate enolase (ENO1) from Arabidopsis thaliana. FEBS Lett 583(6): 983-991.
Google Scholar: Author Only Title Only Author and Title

Raines, C. A. (2003). The Calvin Cycle revisited. Photosynth Res 75(1): 1-10.
Google Scholar: Author Only Title Only Author and Title

Rainteau, D., L. Humbert, E. Delage, C. Vergnolle, C. Cantrel, M.-A. Maubert, S. Lanfranchi, R. Maldiney, S. Collin, C. Wolf, A. Zachowski and E. Ruelland (2012). Acyl chains of phospholipase D transphosphatidylation products in Arabidopsis cells: a study using multiple reaction monitoring mass spectrometry. PLoS One 7(7): 41985.
Google Scholar: Author Only Title Only Author and Title

Rastogi, A., U. Maheswari, R. G. Dorrell, F. R. J. Vieira, F. Maumus, A. Kustka, J. McCarthy, A. E. Allen, P. Kersey, C. Bowler and L. Tirichine (2018). Integrative analysis of large scale transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of diatoms. Sci Rep 8(1): 4834.
Google Scholar: Author Only Title Only Author and Title

Rastogi, A., O. Murik, C. Bowler and L. Tirichine (2016). PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. BMC Bioinformatics 17(1): 261.
Google Scholar: Author Only Title Only Author and Title

Royo-Llonch, M., P. Sánchez, C. Ruiz-González, G. Salazar, C. Pedrós-Alió, K. Labadie, L. Paoli, Tara Oceans Coordinators, S. Chaffron, D. Eveillard, E. Karsenti, S. Sunagawa, P. Wincker, L. Karp-Boss, C. Bowler and S. G. Acinas (2021). Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. Nat Microbiol 6: 1561-1574.
Google Scholar: Author Only Title Only Author and Title

Río Bártulos, C., M. B. Rogers, T. A. Williams, E. Gentekaki, H. Brinkmann, R. Cerff, M. F. Liaud, A. B. Hehl, N. R. Yarlett, A. Gruber, P. G. Kroth and M. van der Giezen (2018). Mitochondrial glycolysis in a major lineage of eukaryotes. Genom Biol Evol 10(9): 2310-2325.
Google Scholar: Author Only Title Only Author and Title

Sachse, M., S. Sturm, A. Gruber and P. Kroth (2013). Identification and evaluation of endogenous reference genes for steady state transcript quantification by qPCR in the diatom Phaeodactylum tricornutum with constitutive expression independent from time and light. Endocytobiosis Cell Res 24: 7.
Google Scholar: Author Only Title Only Author and Title

Scialdone, A., S. T. Mugford, D. Feike, A. Skeffington, P. Borrill, A. Graf, A. M. Smith and M. Howard (2013). Arabidopsis plants perform arithmetic division to prevent starvation at night. Elife 2: 00669.
Google Scholar: Author Only Title Only Author and Title

Serôdio, J. and J. Lavaud (2011). A model for describing the light response of the nonphotochemical quenching of chlorophyll fluorescence. Photosynth Res 108(1): 61-76.
Google Scholar: Author Only Title Only Author and Title

Shtaida, N., I. Khozin-Goldberg and S. Boussiba (2015). The role of pyruvate hub enzymes in supplying carbon precursors for fatty acid synthesis in photosynthetic microalgae. Photosynth Res 125(3): 407-422.
Google Scholar: Author Only Title Only Author and Title

Siegel, P., K. G. Baker, E. Low-Décarie and R. J. Geider (2020). High predictability of direct competition between marine diatoms under different temperatures and nutrient states. Ecol Evol 10(14): 7276-7290.
Google Scholar: Author Only Title Only Author and Title

Smith, S. R., R. M. Abbriano and M. Hildebrand (2012). Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. Algal Res Biomass Biofuels Bioproducts 1(1): 2-16.
Google Scholar: Author Only Title Only Author and Title

Smith, S. R., C. L. Dupont, J. K. McCarthy, J. T. Broddrick, M. Oborník, A. Horák, Z. Füssy, J. Cihlář, S. Kleessen, H. Zheng, J. P. McCrow, K. K. Hixson, W. L. Araújo, A. Nunes-Nesi, A. Fernie, Z. Nikoloski, B. O. Palsson and A. E. Allen (2019). Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. Nat Commun 10(1): 4552.
Google Scholar: Author Only Title Only Author and Title

Spitzer, M., J. Wildenhain, J. Rappsilber and M. Tyers (2014). BoxPlotR: a web tool for generation of box plots. Nature Methods 11(2): 121-122.
Google Scholar: Author Only Title Only Author and Title

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9): 1312-1313.
Google Scholar: Author Only Title Only Author and Title

Strassert, J. F. H., I. Irisarri, T. A. Williams and F. Burki (2021). A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat Commun 12(1): 1879.

Google Scholar: Author Only Title Only Author and Title

Sutherland, E. W., T. Posternak and C. F. Cori (1949). Mechanism of the phosphoglyceric mutase reaction. J Biol Chem 181(1): 153-159.
Google Scholar: Author Only Title Only Author and Title

Tanaka, R. and A. Tanaka (2007). Tetrapyrrole biosynthesis in higher plants. Ann Rev Plant Biol 58: 321-346.
Google Scholar: Author Only Title Only Author and Title

Tardif, M., A. Atteia, M. Specht, G. Cogne, N. Rolland, S. Brugière, M. Hippler, M. Ferro, C. Bruley, G. Peltier, O. Vallon and L. Cournac (2012). PredAlgo: a new subcellular localization prediction tool dedicated to green algae. Mol Biol Evol 29(12): 3625-3639.
Google Scholar: Author Only Title Only Author and Title

Troncoso-Ponce, M. A., J. Rivoal, S. Dorion, R. Sánchez, M. Venegas-Calerón, A. J. Moreno-Pérez, S. Baud, R. Garcés and E. Martínez-Force (2018). Molecular and biochemical characterization of the sunflower (Helianthus annuus L.) cytosolic and plastidial enolases in relation to seed development. Plant Sci 272: 117-130.
Google Scholar: Author Only Title Only Author and Title

Yang, L., Z. Wang, A. Zhang, R. Bhawal, C. Li, S. Zhang, L. Cheng and J. Hua (2022). Reduction of the canonical function of a glycolytic enzyme enolase triggers immune responses that further affect metabolism and growth in Arabidopsis. Plant Cell 34(5): 1745-1767.
Google Scholar: Author Only Title Only Author and Title

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13: 134.
Google Scholar: Author Only Title Only Author and Title

Yin, W. and H. Hu (2023). CRISPR/Cas9-mediated genome editing via homologous recombination in a centric diatom Chaetoceros muelleri. ACS Synth Biol 12(4): 1287-1296.
Google Scholar: Author Only Title Only Author and Title

Yu, G., K. Nakajima, A. Gruber, C. Rio Bartulos, A. F. Schober, B. Lepetit, E. Yohannes, Y. Matsuda and P. G. Kroth (2022). Mitochondrial phosphoenolpyruvate carboxylase contributes to carbon fixation in the diatom Phaeodactylum tricornutum at low inorganic carbon concentrations. New Phytol 235: 1379-1393.
Google Scholar: Author Only Title Only Author and Title

Zhang, Y., A. Sampathkumar, S. M.-L. Kerber, C. Swart, C. Hille, K. Seerangan, A. Graf, L. Sweetlove and A. R. Fernie (2020). A moonlighting role for enzymes of glycolysis in the co-localization of mitochondria and chloroplasts. Nature Commun 11(1): 4509.
Google Scholar: Author Only Title Only Author and Title

Zhao, Z. and S. M. Assmann (2011). The glycolytic enzyme, phosphoglycerate mutase, has critical roles in stomatal movement, vegetative growth, and pollen production in Arabidopsis thaliana. J Exp Bot 62(14): 5179-5189.
Google Scholar: Author Only Title Only Author and Title

Zheng, Y. T., A. H. Quinn and G. Sriram (2013). Experimental evidence and isotopomer analysis of mixotrophic glucose metabolism in the marine diatom Phaeodactylum tricornutum. Microbial Cell Factories 12: 16.
Google Scholar: Author Only Title Only Author and Title

Zhu, B.-H., H.-P. Shi, G.-P. Yang, N.-N. Lv, M. Yang and K.-H. Pan (2016). Silencing UDP-glucose pyrophosphorylase gene in Phaeodactylum tricornutum affects carbon allocation. New Biotechnol 33(1): 237-244.
Google Scholar: Author Only Title Only Author and Title

Zulu, N. N., K. Zienkiewicz, K. Vollheyde and I. Feussner (2018). Current trends to comprehend lipid metabolism in diatoms. Progress Lipid Res 70: 1-16.
Google Scholar: Author Only Title Only Author and Title