A Plastidial Glycolytic-Gluconeogenic Switch of Mitochondrial Origin Enables Diatom Adaptations to High Latitudes

Richard G. Dorrell[1,2], Yue Liang[3,4], Nolwenn Gueguen[5], Tomomi Nonoyama[1,6], Youjun Zhang[7,8], Dany Croteau[9], Mathias Penot[1,2], Sandrine Adiba[1], Benjamin Bailleul[9], Valérie Gros[5], Juan José Pierella Karlusich[1,2,10], Nathanaël Zweig[1,2], Alisdair R. Fernie[7,8], Juliette Jouhet[5], Eric Maréchal[5] 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]Center of Deep Sea Research, Institute of Oceanology, Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China.

[4]Laboratory for Marine Mineral Resources, Pilot National Laboratory for Marine Science and Technology, Qingdao 266237, China.

[5]Laboratoire de Physiologie Cellulaire et Végétale, CNRS, Univ. Grenoble Alpes, CEA, INRAE, IRIG, 17 rue des Martyrs, 38000 Grenoble, France

[6]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

[7]Center of Plant Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria

[8]Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476, Potsdam-Golm, Germany

[9]Institut de Biologie Physico-Chimique (IBPC), Université PSL, 75005 Paris, France

[10]Present address: FAS Division of Science, Harvard University, Cambridge, MA, USA

Abstract

Organic carbon fixed through the Calvin Cycle can be diverted towards different metabolic fates within and beyond the plastids of photosynthetic eukaryotes. These include export to the cytoplasm and mitochondrial respiration; gluconeogenesis of storage compounds; and the anabolic synthesis of lipids, amino acids and cofactors via the plastidial pyruvate hub. In plants, pyruvate is principally synthesised via the lower half of glycolysis-gluconeogenesis in the cytoplasm, although a secondary plastid-targeted pathway in non-photosynthetic tissue directly links glyceraldehyde-3-phosphate to the pyruvate hub. Here, we characterize a complete plastidial lower half glycolytic-gluconeogenic pathway in the photosynthetic plastids of diatoms, obligately photosynthetic eukaryotic algae that are important contributors to marine primary production. We show that the two enzymes required to complete plastidial glycolysis-gluconeogenesis, plastidial Enolase and PGAM (*bis*-phospho-glycerate mutase), originated through recent duplications of mitochondria-targeted respiratory glycolytic isoforms. Through CRISPR-Cas9 mutagenesis and integrative 'omic analyses in the diatom *Phaeodactylum tricornutum*, we present evidence that this pathway

functions to divert excess plastidial glyceraldehyde-3-phosphate into diverse fates accessed from the pyruvate hub, and may potentially also function in the gluconeogenic direction to permit more efficient management of cellular carbon. Considering meta-genomic data, we show that this pathway is of greater importance in polar and sub-polar oceans, in which diatoms dominate primary production; and considering experimental data, we show that this principally relates to the elongated photoperiods present at high latitudes. Our data provide insights into the functions of a poorly understood yet evolutionarily recurrent plastidial metabolic pathway, and a further explanation for the success of diatoms in the contemporary ocean.

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

Introduction

The photosynthetic assimilation of carbon dioxide into glyceraldehyde-3-phosphate via the Calvin Cycle is fundamental to plant and algal biology (1-3). Over 100 billion tonnes of inorganic carbon is fixed each year through this pathway, with effectively equal division of activity between terrestrial plants and aquatic cyanobacteria and eukaryotic algae, maintaining planetary climate homeostasis and supporting the entire Earth ecosystem (3, 4). Following its synthesis, glyceraldehyde-3-phosphate may be converted into larger carbon sugars (gluconeogenesis), broken down into smaller carbon sugars (via glycolysis), or used for the anabolic synthesis of other metabolites. In photosynthetic eukaryotes, which possess plastids (or chloroplasts), these metabolic pathways typically involve the export of glyceraldehyde-3-phosphate or one of its derivatives to the cytosol, or eventually the mitochondria, for subsequent anabolic and respiratory metabolism (5).

Alongside the photosynthetic production of glyceraldehyde-3-phosphate, plastids engage in multiple other carbon metabolism activities (**Fig. 1A**). In plants, these include the synthesis of starch from glucose-6-phosphate (6), produced directly in the chloroplasts of leaves via the gluconeogenic conversion of glyceraldehyde-3-phosphate into glucose-6-phosphate; or in non-photosynthetic plastids (amyloplasts) found in root and other storage tissue via the uptake of cytoplasmic hexose phosphates (**Fig. 1A**) (7, 8). Further metabolites are synthesised in the plastids from shorter-chain sugars, including: free fatty acids and lipids from acetyl-coA (9, 10); aromatic amino acids (via the shikimate pathway), aspartate and lysine from phospho-enol-pyruvate (PEP); small hydrophic amino acids (alanine, leucine, valine) from pyruvate; serine and cysteine from 3-phosphoglycerate (9, 11-13); chlorophyll (via the tetrapyrrole pathway) from glutamate and carotenoids (via the non-mevalonate pathway) from pyruvate (14, 15). The diverse plastidial anabolic reactions that utilize pyruvate, or its adjacent metabolic precursors and products (PEP, acetyl-coA) are frequently grouped together under the umbrella term “pyruvate hub” (16).

The mechanisms through which pyruvate hub intermediates are supplied to the plastid is complicated, with plants classically thought to import precursors from the cytoplasm via dedicated pyruvate and PEP transporters (**Fig. 1A**) (17). Alongside this, however, certain plant plastids may directly synthesize PEP and then pyruvate from glyceraldehyde-3-phosphate in a complete plastidial lower half glycolytic pathway. This occurs via the activities of enzymes associated with the Calvin cycle (glyceraldehyde-3-phosphate dehydrogenase and phospho-glycerate kinase) and two others: plastid-targeted enolase and phospho-glycerate mutase (henceforth referred to as ptEnolase and ptPGAM), which allow the conversion of 1,3-bis-phosphoglycerate from the Calvin cycle to PEP, and subsequent incorporation into the pyruvate hub (**Fig. 1A**). Both ptEnolase and ptPGAM have bidirectional functions due to reversible reaction kinetics, which we henceforth refer to globally as “glycolysis-gluconeogenesis”, contrasting with “glycolysis” and “gluconeogenesis” to signify metabolic activities in one direction only. In plants, ptEnolase and ptPGAM enzymes are characteristically associated with non-photosynthetic tissue such as seeds and roots (9, 18-20).

Moreover, *Arabidopsis* mutant lines for ptEnolase and ptPGAM have, respectively, limited and null single phenotypes under replete growth conditions (8, 18, 19), and the broader significance of this pathway to plant physiology remains unknown. Genes for ptEnolase and ptPGAM are believed to be absent from primary red or green algae, the closest relatives of land plants (21, 22).

Diatoms are a eukaryotic algal group, positioned within the stramenopile lineage that is distantly related to plants, with plastids originating from the secondary endosymbiosis of a red algal ancestor or endosymbiotic derivative thereof, separate from the primary endosymbiotic origin of plant plastids (23, 24). Diatoms are extraordinarily successful in the modern ocean, comprising nearly half of total algal abundance in the *Tara* Oceans dataset, and performing about one-fifth of total planetary photosynthesis (25, 26). The contributions of diatoms to primary production is even greater in specific oceanic regions, including the nutrient-rich and highly productive waters at high latitudes (i.e., the Arctic and Antarctic Southern Oceans) (27). Diatoms within polar environments are exposed to multiple photo-physiological stressors such as low environmental temperatures, limitations of key nutrients including iron, zinc, and cobalamin, and elongated photoperiods ranging from continuous illumination in the polar summer to continuous darkness in the winter (24, 26, 28, 29). Previous genomic and physiological studies of model diatoms, such as the transformable species *Phaeodactylum tricornutum*, have identified multiple strategies that allow diatoms to tolerate photo-stress, including extensive capacities for non-photochemical quenching pathways displaying different features than in plants or in primary red algae (30, 31), and complex strategies for the acquisition, storage and recycling of cellular iron (32), which may provide new strategies for metabolic engineering of enhanced photosynthetic activity in plants (33).

Comparative genomic analyses have revealed the presence of a different organization of carbon metabolism in diatoms to plants (34). These include the storage of sugars in the form of chrysolaminarin in cytosolic vacuoles, as opposed to plastidial starch (35), and the probable absence of plastidial hexose-phosphate (e.g., glucose-6-phosphate) transporters (17, 36). These features seem consistent with the accumulation of photosynthetically acquired carbon in the form of fatty acids rather than glucose in the stroma; and the export of three-carbon metabolites to the cytosol for the subsequent gluoconeogenesis of storage carbons, alongside plastidial membrane galactolipids (eg., MGDG, DGDG) and sulfoquinovosyl-diacylglycerols (SQDG) synthesised, as per in plants, from cytoplasmic UDP-galactose and UDP-glucose respectively (35-37). Moreover, the lower half of respiratory glycolysis-gluconeogenesis in diatoms (and photosynthetic and non-photosynthetic relatives of the stramenopiles) occurs in the mitochondria, as opposed to the cytosolic pathways found in plants, red and green algae (**Fig. 1A**) (34, 38).

A second, complete lower half of glycolysis-gluconeogenesis is predicted to occur in the diatom plastid, based on plastid-targeted ptEnolase and ptPGAM1A proteins identified *in silico* in multiple species (23, 34, 39, 40). Thus, diatom plastids possess independent and separate pathways for the conversion of glyceraldehyde-3-phosphate to pyruvate in plastid and for mitochondrial respiration (**Fig. 1A**). While this metabolic arrangement has also been observed in plants, as described above, it is likely to have independently originated in diatoms (e.g., given the absence of reported plastidial glycolysis pathways in primary red algae). Moreover, while the plant plastidial glycolysis pathway is principally associated with non-photosynthetic tissue, diatoms (as obligately photosynthetic and unicellular species) presumably perform photosynthesis and plastidial glycolysis in the same organelle (18). Diatom ptEnolase and ptPGAM genes have further been shown to be transcriptionally co-regulated with photosynthetic carbon assimilation, suggesting metabolically important functions linked to photosynthetic activity (41). Finally, ptPGAM proteins have recently been shown to over-accumulate under copper starvation, in the open-water diatom *Thalassiosira oceanica*, suggesting a possible connection between its function and photo-stress (40).

Here, we use combined bioinformatic profiling of large-scale sequence datasets from *P. tricornutum*, the broader algal tree of life, and the *Tara* Oceans dataset, alongside integrative functional characterization of *P. tricornutum* CRISPR-CAS9 knockout mutants, to reveal heretofore unknown functions of diatom ptEnolase and ptPGAM enzymes in adaptations to high latitudes. Using phylogenetic techniques, we demonstrate that the genes encoding these enzymes arise from a recent duplication and retargeting of the mitochondria-targeted and respiratory isoform in diatoms and their closest relatives within the stramenopiles. Using meta-genomic data from *Tara* Oceans, we further show these genes are most highly expressed in high latitudes. Considering the growth and physiological phenotypes of knockout lines lacking plastidial glycolysis-gluconeogenesis proteins, we show that this pathway has augmented importance in cells grown under continuous illumination as opposed to light-dark cycling, and present evidence that this may be due to diminished capacities to convert glyceraldehyde-3-phosphate assimilated by the Calvin cycle into plastidial amino acids and lipid pools, which may be compensated by a redirection of flux through mitochondrial glycolysis and mitochondrial: plastid amino acid shuttling. We also show that the growth phenotype of diatom plastid glycolysis mutants is counterbalanced by low temperatures, which may relate to enhanced mitochondrial respiration and impeded gluconeogenesis of plastidial phospho-*enol*-pyruvate. Overall, our data position plastidial glycolysis-gluconeogenesis as a key adaptive modulator of plastidial metabolic poise in diatoms, with particular importance to the light stresses associated with high latitudes, and reveals a new candidate pathway for the optimization of photosynthetic productivity in plants.

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, as measured with a photofluorometer, under one of four light, temperature and shaking regimes: either 19°C with 12h light: 12 dark cycling, shaken at 100 rpm (for general molecular work and transformation); 19°C with 12h light: 12 dark cycling without shaking (« LD » growth conditions and physiological analysis); 19°C with 24h continuous light and without shaking (« CL » growth conditions and physiological analysis); or 8°C with 24h continuous light and without shaking (« 8C » growth conditions and physiological analysis). Mutant *Phaeodactylum* lines were maintained on ½ ESAW 1% agarose plates, supplemented by 100µg ml^{-1} each ampicillin and streptomycin, and 30µg ml^{-1} chloramphenicol, and either 100µg ml^{-1} zeocin (single transformants), or 100µg ml^{-1} zeocin and 4µg ml^{-1} blasticidin (complementation lines), as previously described (42, 43). All functional analyses of transformant lines were performed on transformant lines grown in the absence of antibiotic selection, to avoid secondary effects of antibiotics on growth or physiology.

Phylogenetic analysis

The peptide sequences of all genes with PFAM annotations coherent with either Enolase (PF00113) or Phospho-glycerate mutase (PGAM; taking histidine phosphatase, PF00300, as a proxy) encoded in the version 3 annotation of the *P. tricornutum* genome were extracted, and searched against a library of 289 complete cryptomonad, haptophyte and stramenopile genomes and transcriptomes by reciprocal BLASTp/ BLASTp best-hit search with threshold e-value 10^{-05}, as previously described (44-46). Sequences that retrieved a reciprocal best hit against a *Phaeodactylum* enolase or PGAM gene were trimmed to the first encoded methionine, and then screened through three targeting predictors: ASAFind v 2.0, used with SignalP v 3.0 (47, 48), HECTAR, integrated into the Roscoff Galaxy platform, under the default settings (49, 50), and MitoFates, run with threshold value 0.35 (23, 51). Sequences that retrieved a plastid, mitochondrial or dual-targeting prediction (defined as

plastid-targeted with at least one programme, and mitochondria-targeted with at least one another) were annotated as such and retained.

A parallel set of reciprocal BLAST searches, with threshold e-value 10^{-05}, were performed against 85 prokaryotic and eukaryotic genomes sampled with taxonomic balance from across the remainder of the tree of life (23), and were pooled with the organelle-targeted cryptomonad, haptophyte and stramenopile proteins identified above. The pooled set of sequences were aligned first with MAFFT v 7.0 under the --auto setting, followed by the in-built alignment programme in GeneIOUS v 10.0.9, under default settings (52, 53), and incomplete and poorly aligned sequences were manually removed. The curated alignment was trimmed with trimal with setting –gt 0.5, and phylogenetic analyses were performed with MrBayes v 3.2 and RAxML v 8, integrated into the CIPRES webserver (54-56). MrBayes trees were run for 10,000,000 generations with the GTR, Jones and WAG substitution matrices, 4 starting chains and sumt and sump burnin fractions set to -0.5; all cold chains were confirmed to have reached a pvalue plateau below 0.1 prior to the start of the consensus building. RAxML trees were run with GTR, JTT and WAG substitution matrices, 350-400 ML generations, and automatic bootstopping. Alignments and individual tree topologies are provided in **Table S1**.

Co-regulation analysis

Transcriptional co-regulation of *P. tricornutum* genes to ptEnolase and ptPGAM1A were assessed using a composite dataset consisting of two meta-studies of *Phaeodactylum* gene expression: DiatomPortal, based on microarray data generated in particular from cultures exposed to different light regimes and organic and inorganic stresses (57); and PhaeoNet, constructed from RNAseq data generated from N, P and Fe-starved cultures (41, 58-60).

To reconcile both datasets, which are based on the versions 1 and 3 annotations of the *P. tricornutum* genome, respectively, each DiatomPortal version 1 gene number was remapped to a PhaeoNet version 3 gene number by reciprocal BLASTp/ BLASTp best hit search (45). Version 3 genes that were retrieved as reciprocal BLAST best hits by >1 version 1 gene (typically resulting from merged gene annotations between the two genome versions) were assigned the median of all matching DiatomPortal relative fold expression changes. The raw normalized expression abundances (PhaeoNet) and fold-expression changes (DiatomPortal) for each constituent library used within each study were converted into percentile rank values to normalize them, with 100 equating to the most highly expressed or upregulated gene, and 0 to the most lowly expressed or downregulated gene. Genes lacking expression data for a particular study were marked N/A. Finally, correlation calculations of ranked gene expression (i.e., Spearman correlations) were performed for each gene in the *P. tricornutum* version 3 genome against ptEnolase (Phatr3_J41515), ptPGAM1A (Phatr3_J17086), and ptPGAM2 (Phatr3_J37201). Finally, the correlation coefficients were identified for proteins implicated in core plastid and mitochondrial metabolism pathways as per (41), with the highest (strongest positive) value taken in the case of proteins with multiple homologues. Full data pertaining to these outputs are provided in **Table S2**, sheets 1-3.

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 (0.1M NaCl, 0.01M Tris pH8, 0.001M 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

(61), 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 late exponential phase, at counted densities of between 1 and 2 1-2 x 10^6 cells ml^{-1}. CL cultures were typically harvested eight days post-inoculation, LD cultures nine days post-inoculation, and 8C culture 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 *Phaeodactylum* (58), and per the sampling parameters used within *Tara* Oceans (62). Cell cultures were harvested at densities that were insufficient to provoke N or P limitation of the ESAW medium (63).

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 as previously described (23). 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. Precipitated RNA samples were 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 0.5M 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 ptEnolase, ptPGAM1A and ptPGAM2 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 **Table S2**, 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 (**Table S2**, sheet 4); cut, purified, and treated with 1U FastDigest DpnI (Thermo Fisher) to remove any residual plasmid DNA. The gene-specific primers for each ptEnolase and ptPGAM 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* by biolistic transformation, as previously described (42). 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 (Promega), 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.

Tara *Oceans Analysis*

The complete *Tara* Oceans and *Tara* Oceans Polar Circle libraries of meta-genome and meta-transcriptome diversity (64, 65) were searched for orthologues of diatom ptEnolase, ptPGAM1A and ptPGAM2 sequences via a phylogenetic reconciliation approach. First, a HMM (hidden Markov model) was constructed for all diatom plastidial-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}. Matching sequences were extracted, and searched by BLASTp against both the complete copy of the *P. tricornutum* genome (45) and the complete untrimmed alignment, and sequences that retrieved a best hit against (i) an Enolase or PGAM sequence and (ii) a diatom plastidial-targeted isoform in each search respectively were retained. Retained sequences were combined with the untrimmed alignment of each gene and realigned using the same MAFFT, GenelOUS and trimal pipeline as defined above. Curated alignments were finally screened by RAxML tree with the JTT substitution matrix, as above, and the monophyletic clade of *Tara* Oceans sequences and 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 correlations were calculated between relative abundances and all quantitative measured environmental variables associated with *Tara* Oceans samples as stored within the PANGAEA repository (62). To assess the possibility of non-linear correlations, ranked differences (i.e., Spearman correlation) were also tested for each variable with broadly consistent results. 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 **Table S3**.

CRISPR mutagenesis

CRISPR target sequences for ptEnolase and ptPGAM1A genes were identified using PhytoCRISP-Ex (66), 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 (67). 2 µg insertion-positive pu6:SG1 plasmids, as confirmed by Sanger sequencing and purified using a Qiagen Midiprep kit were co-introduced into wild-type *P. tricornutum* 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 0.14M NaCl, 5mM KCl, 10mM 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 (59, 67). Empty vector control lines were generated using the same protocol, except with only HA-Cas9 and pPhat plasmids, cotransformed without a CRISPR target sequence. Primers used for each mutagenesis are provided in **Table S2**, sheet 4; tabulated cleaned knockout mutants and their associated genotypes are shown in **Table S4**, sheet 1; and effective expression levels of mutated gene copies as assessed by qRT-PCR are shown in **Table S4**, sheet 2. Mutant colony genotypes were periodically confirmed (approx. once per month) by PCR and Sanger sequencing throughout the duration of all subsequent experiments, and found to remain stable.

Complementation of knockout lines

Knockout lines were complemented with pPhat:GFP vectors carrying overexpressing copies (under an FcpA promoter) of ptEnolase and ptPGAM1A synthetic constructs with all CRISPR target sequences replaced with silent mutations (Eurofins). Vectors were identical to those previously used for localization, but with a blasticidin S-deaminase gene in lieu of SHBLE (68) introduced by NEB Hi-Fi kit as before. Complementation constructs were transformed via bombardment, and cells were replated 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, and left to recover at low light for two days followed by standard growth conditions until the appearance of colonies.

For each complementation, three ptEnolase and ptPGAM1A 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 ptEnolase and ptPGAM1A 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 (**Table S4**, 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 ptEnolase:GFP and ptPGAM1:GFP, or the cytoplasm for ectopic GFP) was verified by western blot with an anti-GFP antibody (Promega) and a previously published protocol (69), 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: every 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, as defined above, 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 densites > 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 condition only), until cell lines were confirmed to have exited log phase (i.e., reached a stable 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 independent mutant line studied, providing a minimum of 24 measurements (i.e., four distinct mutant lines) per each genotype studied (ptEnolase knockout, ptPGAM1A 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 the time period over which a linear correlation (r> 0.99) was observed between both values, typically equating to densities between 10^5 and 2x 10^6 cells ml^{-1}, and covering a minimum of three and in most cases five successive measurements of each individual growth curve. Three exemplar growth curve outputs are provided in **Table S4**, sheets 3-5; and an overview of relative growth rates expressed as a function of mean empty vector control growth rates are provided in **Table S4**, 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 **Table S4**, sheet 7.

Photophysiology

Cultures for photophysiological analysis were grown in 10ml ventilated plastic flasks, without shaking, under 19C CL and 19C LD conditions as described above. Cultures were grown from a starting inoculum of 10^5 cells ml^{-1} as measured with a Malassez haemocytometer, and refreshed into fresh media when they had reached a measured density of 10^6 cells ml^{-1} 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}, as verified by Fv/Fm measurements > 0.6 across all measured lines (**Table S4**, sheet 8). Cell cultures were concentrated to between 2 and 5 × 10^7 cells ml^{-1} before photophysiology measurements. Parameters measured with the miniFIRe fluorometer (as defined below) were also quantified for cultures grown under 8C CL conditions, as the measurements were sufficiently rapid to allow the culture to be maintained at growth temperatures (70).

Steady-state light curves (SSLC) were conducted with a fluorescence CCD camera recorder (SpeedZen, JBeamBio, France). Samples were exposed to 7 increasing light steps of 3 min each (5 min

for the first light step), from 35 to 750 µmol photons m^{-2} s^{-1} (green actinic light, 532 nm). The experiment was repeated with 2-4 independent replicates for all 19 strains ($n \geq 6$ per construct under each growth condition). 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 saturating 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 non-photochemical quenching (NPQ) were calculated on the last time point of each light step as ϕPSII = $(F_M' - F_s)/F_M'$ and NPQ = $F_M/F_M' - 1$. The whole NPQ vs E curve was fitted as NPQ = $NPQ_M \times E^n/(E50NPQ^n + E^n)$, where NPQ_M is the maximal NPQ value, $E_{50}NPQ$ is the half-saturation intensity for NPQ and n is the sigmoidicity coefficient (71). 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). 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 (70). The relative electron transport (rETR) was then calculated as rETR = ϕPSII.σPSII/$(F_V/F_M).E$ and was called relative because the light qualities for ϕPSII and σPSII were different. The rETR vs E curve was fitted as rETR = $rETR_M.(1-\exp(-\alpha.E/rETR_M))$ where $rETR_M$ is the maximum rETR and α is the light-limited slope of rETR vs E. Only rETR values from 0 to 450 µmol photons m^{-2} were used for the fit because the values from 600 and 750 µmol photons m^{-2} were too noisy. The light saturation parameter E_K was calculated as $rETR_M/\alpha$ and the fitted values of the parameters were used to estimate the rETR under the growth light intensity of 60 µmol photons m^{-2} s^{-1} as $E60rETR = rETR_M.(1-\exp(-\alpha.60/rETR_M))$ (30). Measured photo-physiological values are tabulated in **Table S4**, sheets 8-11.

Gene expression analysis

RNA sequencing libraries were prepared from 200 ng DNAse-treated RNA for each mutant line and treatment condition, with a minimum of three replicates per sample. RNA 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 approximately 130-160 Gb sequence data with ≥85% of bases higher than Q30 per sample.

Output fastQ files were mapped using Nextflow's RNA sequencing assembly pipeline https://nf-co.re/rnaseq/usage, and gff3 annotations of the *P. tricornutum* version 3 genome (72). Total mapped read counts were then compared between all biological and technical replicates for (i) each mutant line sequenced, (ii) each genotype (ptEnolase knockout, ptPGAM1A 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 (73). The final dataset included 63 RNAseq libraries, including five ptEnolase and five ptPGAM1A 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 (ptEnolase, ptPGAM1A and control) considered (**Table S5**, 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 (74) (**Table S5**, sheets 2-3).

The mean transcript abundances of DEGs in knockout compared to control lines were first assessed in normalised (41) RNAseq data of N (59) and P-limited (58) *Phaeodactylum* cell lines under two and nine time-points respectively (**Table S5**, sheet 4). No significant differences were found between DEGs and other genes in the *P. tricornutum* genome (one-way ANOVA, P > 0.05; **Table S5**, 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 (**Table S5**, sheets 6-10) (41, 75). Tested traits included: PFAM and KEGG from the *P. tricornutum* version 3 genome annotation (45); GO, annotated using combined BLASTkoala and GHOSTkoala annotations (76); WGCNA coregulated gene module, as defined within the PhaeoNet meta-analysis of *P. tricornutum* RNAseq co-expression (41); and localization, predicted in silico using ASAFind v 1.0 with SignalP v 3.0 (47, 48), HECTAR under default conditions (49), and MitoFates with threshold value 0.35 (51). Functional enrichments were tested by two-tailed chi-square ($P < 0.05$) of a differentially expressed gene occurring in either one (ptEnolase v control; ptPGAM1A v control) knockout-versus-control line tests, or in both tests realised under each physiological condition. Finally, the distribution of DEGs across *Phaeodactylum* core plastid and mitochondrial metabolism pathways were mapped onto a previously defined proteomic model of each organelle (41); with the strongest DEG enrichment taken in the case of enzymes with multiple homologues (**Table S5**, 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 (77), 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 performed a minimum of three times for each amplicon: sample pair, with at least three RT- equivalents performed to subtract residual genomic DNA from each Cp value obtained, and using two housekeeping genes (Ribosomal protein S1, RPS; and TATA binding protein, TBP) previously shown to have conditionally invariant expression patterns in *P. tricornutum* used as quantification references (78). Tabulated qRT-PCR outputs are shown in **Table S5**, sheet 12.

Metabolite analysis

Culture volumes containing 1.5×10^8 exponential *P. tricornutum* cells (counted density $1\text{-}2 \times 10^6$ cells ml^{-1}), inoculated and grown in antibiotic-free ESAW under LD, CL or 8C conditions and at the time points described above for RNA library preparation, were calculated using a Malassez cytometer and harvested by centrifugation as described above. Cell pellets were washed three times in sterile growth medium, then transferred to a pre-weighed, double-pierced and double-autoclaved 1.5 ml Eppendorf tube for lyophilisation. Cell pellet masses were recorded for each sample, and samples were 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 (79). The chromatograms and mass spectra were evaluated using TagFinder software (80). Metabolite identification was manually checked by the mass spectral and retention index collection of the Golm Metabolome Database (81). 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. Normalised metabolite abundances were subsequently 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 RNA libraries

prepared from at least two genetically distinct mutant constructs for each genotype considered (**Table S6**, sheet 1).

Glycerolipid analysis

Glycerolipids were extracted from 1.5 x 10^8 exponential *P. tricornutum* cells (counted density 1-2 x 10^6 cells ml^{-1}), with cell pellets prepared and freeze-dried as per the metabolite analysis above. Pellets were suspended in 4 mL of boiling ethanol for 5 minutes to prevent lipid degradation and lipids were extracted by addition of 2 mL methanol and 8 mL chloroform at room temperature (82). 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 (83). 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 (84). 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 thylakoid (85) and hydrogenated (86). 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 $L{\cdot}min^{-1}$, sheath gas heater at 300°C, sheath gas flow at 11 $L{\cdot}min^{-1}$, nebulizer pressure at 25 psi, capillary voltage at ± 5000 V and nozzle voltage at ± 1,000 V. Nitrogen was used as collision gas. The quadrupoles Q1 and Q3 were operated at widest and unit resolution respectively.

Phosphatidylcholine (PC) and diacylglyceryl hydroxymethyltrimethyl-*β*-alanine (DGTA) analyses were carried out in positive ion 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 multiple reaction monitoring (MRM) with 50 ms dwell time with the various transitions previously recorded (63). Mass spectra were processed using the MassHunter Workstation software (Agilent) for identification and quantification of lipids. Lipid amounts (pmol) were corrected for response differences between internal standards and endogenous lipids as described previously (87).

Data deposition

RNAseq data associated with this project is deposited with NCBI BioProject with project number PRJNA788211. 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/ (88). 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.

Results

Phylogeny and localization of ptEnolase and ptPGAM sequences suggest recent recruitments of mitochondrial glycolytic enzymes to the diatom chloroplast

Previous comparative genomic studies of diatom carbon metabolism have identified multiple isoforms of core glycolysis-gluconeogenesis enzymes, targeted to both mitochondria and chloroplasts (34, 38, 39). To place these data into a phylogenetic context, we considered the evolutionary diversity and history of organelle- (plastid- and mitochondria-) targeted Enolase and PGAM sequences within the diatoms alongside related algae in the stramenopiles and two other groups, the cryptomonads and haptophytes. These algal groups all possess plastids of secondary or higher red endosymbiotic origin, surrounded by four membranes, and the plastid-targeting sequences associated with nucleus-encoded proteins in each group are sufficiently well conserved to be identified through common *in silico* prediction pipelines (23, 47). Single-gene trees were made for all organelle-targeted Enolase and PGAM sequences from 289 cryptomonad, haptophyte, and stramenopile (including diatom) genomes and transcriptomes, plus all orthologues from 85 further genomes selected from across the tree of life, based on RbH (reciprocal BLASTp best-hit) searches with all Enolase and PGAM sequences in the *P. tricornutum* version 3 genome annotation as queries (45). **Figs. 1B** and **1C** show consensus MrBayes trees realised with three substitution matrices (GTR, Jones and WAG) for selected sequences from each alignment, retaining only species with both identifiable plastid- and mitochondria- targeted orthologues of each protein. **Figs. S1-S3** provide consensus RAxML trees realised with three substitution matrices (GTR, JTT and WAG) for the complete aligned sequences for Enolase, and two non-homologous protein families annotated as PGAM (PGAM1 and PGAM2).

The obtained tree topologies revealed multiple evolutionary origins for plastidial Enolase and PGAM sequences across the cryptomonads, haptophytes and stramenopiles via the duplication of mitochondria-targeted (respiratory) enzymes, with diatom plastidial isoforms typically having recent and/or diatom-specific evolutionary origins. Diatom ptEnolase 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 (23, 75), and in turn as a sister-group to mitochondria-targeted proteins from these groups (MrBayes PP = 1.0 under all studied matrices, **Fig. 1B**), followed by mitochondria-targeted proteins from other photosynthetic (chrysophytes) and non-photosynthetic stramenopiles (oomycetes; MrBayes PP = > 0.95 under GTR and Jones matrices, **Fig. 1B**). This topology indicates a duplication and recruitment of the host-derived mitochondrial targeted protein to the plastid within a common ancestor of the diatoms, pelagophytes and dictyochophytes. The plastid-targeted isoforms identified for haptophytes positioned within the diatom/ pelagophyte/

dictyochophyte clade, consistent with a partial pelagophyte/ dictyochophyte origin of the haptophyte plastid (23), and independent duplications and plastid retargeting of mitochondria-targeted enolase proteins were identified in both the chrysophytes and cryptomonads (**Fig. S1**), suggesting recurrent origins of ptEnolase functions across the secondary red plastid tree of life.

The phylogenetic origins of diatom ptPGAM genes were more complex, typified by multiple origins and plastidial recruitments of mitochondria-targeted enzymes. Two distinct clades of diatom ptPGAM proteins were identified, with ptPGAM1A (typified by the *Phaeodactylum* protein Phatr3_J17086) closely related to mitochondrial-targeted proteins found across the stramenopiles (MrBayes PP = 1.0 under all studied matrices, **Fig. 1C**), and ptPGAM1B (typified by the *Phaeodactylum* protein Phatr3_J50414) closely related to mitochondrial-targeted proteins found at least within pelagophytes and dictyochophytes (MrBayes > = 0.85 under all studied matrices, **Fig. 1C**). More exhaustively sampled trees indicated deeper orthologies of each plastid-targeted family, respectively, to plastid-targeted proteins from the raphidophytes and chrysophytes and mitochondria-targeted oomycete proteins (ptPGAM1A, **Fig. S2**); and to plastid- and mitochondria-targeted enzymes from the chrysophytes and pelagophytes (ptPGAM1B, **Fig. S2**). This topology suggests ancestral presence and mitochondrial origins of both plastid-targeted protein families, with possible differential retention in different photosynthetic stramenopile groups. Possible duplications and plastid recruitments of mitochondria-targeted PGAM proteins were also visible in the haptophytes and cryptomonads, again positioning plastid-targeted glycolysis-gluconeogenesis as a potentially widespread innovation across the secondary red chloroplast tree of life.

A final plastid-targeted protein annotated as PGAM in the version 3 *Phaeodactylum* genome (45), hereafter termed PGAM2, was identified exclusively in diatoms, pelagophytes, and haptophytes (of inferred pelagophyte origin; **Fig. S3**), with limited homology to ptPGAM1 enzymes (BLASTp e-value > 1.0 in pairwise protein-protein searches). Only PGAM1 enzymes contain an annotated phospho-glyceromutase active site (IPR005952), while both PGAM1 and PGAM2 belong to the same PFAM superfamily (histidine phosphatase, PF03000) as inferred using InterProScan and PFAMscan respectively (46, 89). This isoform was predominantly mitochondria-targeted, with plastid- or dual-targeted isoforms only identified in *P. tricornutum* (Phatr3_J37201) and a small number of other pennate diatoms (*Attheya septentrionalis, Amphiprora coffeaformis, Entomoneis sp.* CCMP2396) and haptophytes (*Phaeocystis antarctica*, *Phaeocystis rex* sp. CCMP2000) (**Fig. S3**).

To confirm plastidial localization of *Phaeodactylum* enolase and PGAM proteins, three proteins (Phatr3_J41515, ptEnolase; Phatr3_J17086, ptPGAM1A; Phatr3_J37201, ptPGAM2) were cloned into pPhat: eGFP vector sequences and transformed into wild-type *P. tricornutum* Pt1.86 cells by biolistic transformation (**Figs. 1D, S4**). These proteins were selected based on co-regulation analysis, suggesting tight transcriptional linkage to one another and core chloroplast metabolic pathways (**Fig. S5**, *see below*). The observed fluorescence patterns were coincident with chlorophyll, confirming plastid targeting.

Physiological roles of diatom ptEnolase and ptPGAM identified from cellular and environmental sequence expression data

Next, we considered the mRNA abundances of diatom ptEnolase, ptPGAM1A and ptPGAM2 sequences, both in published microarray (57) and RNAseq data for *P. tricornutum* (41) (**Figs. S5-S7, Table S2**), and more broadly within the *Tara* Oceans dataset (64) (**Fig. 2, S8-10, Table S3**), as proxies of their physiological function. First, considering general patterns of transcriptional co-regulation across published *P. tricornutum* data, we noted strong positive correlations ($r > 0.75$) between the sole gene encoding ptEnolase (Phatr3_J41515) and two of the genes encoding plastid-targeted PGAM: ptPGAM1A (Phatr3_J17086) and ptPGAM2 (Phatr3_J37201), with weaker (< 0.5) or zero

correlations observed for genes encoding ptPGAM1B (Phatr3_J51404, Phar3_J5629) proteins (**Fig. S5**).

Next, considering patterns of transcriptional coregulation to other core components of *Phaeodactylum* chloroplast and mitochondrial metabolism (41), we noted that both ptPGAM1A and ptEnolase showed consistently positive transcriptional coregulation coefficients against other components of the lower half of plastidial glycolysis-gluconeogenesis, alongside most of the genes encoding Lhcr and Lhcf proteins involved in photosynthetic light assimilation, core components of the photosynthetic electron transport chain pathway, and enzymes involved in shikimate, alanine and tetrapyrrole (chlorophyll and heme) synthesis. Both ptPGAM1A and ptEnolase genes further showed weaker positive or negative correlations to plastid fatty acid and lipid synthesis, alongside *LhcX* genes implicated in photo-protection (**Fig. S6**). Finally, considering the raw and relative abundances in both RNAseq and microarray data, we noted elevated relative abundances of ptEnolase and ptPGAM1A respectively in samples exposed to >16h and >12h continuous illumination, consistent with a photosynthesis-associated phenotype (**Fig. S7; Table S2**, sheet 1). We noted no effects of phosphorus or nitrogen starvation, which are typically associated with lipid production in *Phaeodactylum*, on the mRNA abundances of either gene (58, 59, 63).

Within *Tara* Oceans data, we noted high relative abundances of diatom ptEnolase and ptPGAM1 sequences in meta-transcriptome (metaT) data in stations from both high northern and southern latitudes, in both surface (**Fig. 2, S8**) and deep chlorophyll maximum (DCM) samples (**Fig. S9**). This accumulation was visible across all size fractions (**Figs. 2A**, **S8A, S8C, S9A, S9C**) and was notably greater than equivalent accumulations in meta-genome (metaG) data (**Figs. 2B**, **S8B, S8D, S9B, S9D**). Normalisation of metaT abundances calculated for each gene in the 0.8-2000 µm size fraction against all diatom metaT sequences (**Fig. S10A)** both in surface (ptEnolase $R^2 = 0.18$, $P < 10^{-05}$, ptPGAM1A $R^2 = 0.23$, $P < 10^{-05}$) and deep chlorophyll maximum (DCM) depths (ptEnolase $R^2 = 0.53$, $P < 10^{-05}$, ptPGAM1A $R^2 = 0.59$, $P < 10^{-05}$), suggesting that plastidial glycolysis plays a more central role in diatom cellular transcription at high latitudes. Similar, albeit weaker, positive correlations were observed when normalising the metaT abundances against metaG relative abundances calculated for each gene (**Fig. S10B**), in surface depths for ptEnolase ($R^2 = 0.10$, $P < 0.05$) and DCM for both genes (ptEnolase $R^2 = 0.28$, $P < 0.05$, ptPGAM1 $R^2 = 0.29$, $P < 0.05$), further suggesting that latitude positively impacts on the relative expression of diatom plastidial glycolysis genes. Further significant positive correlations to with 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 ptPGAM1 metaT normalised against metaG in surface 3/5-20 ($P < 10^{-05}$), 20-180 ($P < 10^{-05}$) and 180-2000 ($P < 0.05$) µm fractions (**Table S3**, sheet 10), indicating these results are relevant across the diversity of marine diatom functional groups. Finally, recurrent positive correlations in the expression of plastid glycolysis genes to day length and negative correlations to measured water temperatures and the calculated length of the ice-free period at each station, were found considering multiple normalisations, depth and size fraction combinations (**Table S3**, sheet 10), whereas weaker and inconsistent correlations were found for measured nutrients (N, P, Si, Fe) that also vary across oceanic latitudes. We observed some negative correlations between ptEnolase or ptPGAM1 transcript abundance and salinity, particularly when normalised against meta-gene abundance (**Table S3**, sheet 10), supporting recent reports of differential expression of diatom glycolysis genes across a Baltic Sea salinity gradient, although we note that the range of salinities sampled across *Tara* Oceans stations (23.0 – 39.4 per mille) are somewhat higher than those within the Baltic (i.e., 0- 25 per mille) (62, 90).

The transcriptional preference of diatom ptEnolase and ptPGAM1 for long days and low temperatures contrasted strongly with PGAM2, which showed global expression trends, including effectively equivalent relative abundance in some stations from the temperate South Pacific and Atlantic as stations from the Arctic and Southern Oceans (**Fig. S11A**; **Table S3**, 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 $R^2 = 0.37$, $P < 0.05$), no such correlation was observed for metaT data normalised against PGAM2 metaG relative abundances (surface $R^2 = 0.00062$, DCM $R^2 = 0.027$, $P > 0.05$) suggesting that latitude does not directly define PGAM2 expression (**Fig. S11B, C**). 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 (**Table S3**, sheet 10). In total, the *Tara* Oceans expression abundances, alongside *Phaeodactylum* transcriptome and microarray data, suggest diatom plastidial glycolysis to be of particular functional importance in high environmental latitudes, characterised by their extreme photoperiods and chronically low temperatures, with ptEnolase and ptPGAM1A being likely the principal drivers of this association within the lower half reactions (**Figs. 2, S7-S10**)

Growth phenotypes of ptEnolase and ptPGAM1A knockout and complementation lines reveal modular roles in adaptation to high latitudes

Next, we generated homozygous CRISPR knockout lines for both ptEnolase and ptPGAM1A, selected because of their clear transcriptional co-regulation to one another (**Fig. S5**, in contrast to ptPGAM1B) and latitudinal expression correlation in *Tara* Oceans (**Figs. 2, S10**, in contrast to ptPGAM2). Two distinct CRISPR regions with unique sequences in the *P. tricornutum* genome targeted for each gene, and multiple CRISPR knockout lines were generated in each case (ptEnolase CRISPR region 1 $n = 4$, CRISPR region 2 $n = 3$; ptPGAM1A CRISPR region 1 $n = 2$, CRISPR region 2 $n = 3$) (**Fig. S12A**). Each CRISPR line was verified by sequencing to be homozygous and to contain a frame-shift mutation sufficient to impede successful translation of the ptEnolase or ptPGAM1A gene sequence (**Fig. S12A**). We were unable to directly test the accumulation of ptEnolase and ptPGAM1A protein levels in knockout lines by western blot due to the absence of antibodies specific to each protein, but confirmed the effective knockdown of expression levels of mutated genes in knockout lines compared to empty vector control lines ($n = 4$) by qRT-PCR (measured knockout mRNA abundance 1.8-39 % mean empty vector control mRNA abundance under 19°C and 12h light: 12h dark (« 19C LD ») conditions, **Fig. S12B**), consistent with non-sense mediated decay of non-translatable mRNA sequences (91).

Next, we performed growth curves of ptEnolase and ptPGAM1A knockout lines compared to empty vector controls under three growth conditions (**Fig. 3**; **Table S4**, sheets 3-6). We chose to target changes in light and temperature, given the associations observed between these parameters and ptPGAM1A and ptEnolase expression both in *Phaeodactylum* gene expression and *Tara* Oceans data (**Fig. 2**; **Fig. S7**). The conditions were: 19°C and 12h: 12h light: dark cycling ("19C LD"), reflecting default growth conditions used in multiple other *Phaeodactylum* mutant and comparative physiology studies (e.g., (58, 59, 63)); 19°C and 24h continuous light ("19C CL") to test the effects of photoperiod; and 8°C and 24h continuous light ("8C CL") to test the effects of temperature. Growth curves were tested for seven ptEnolase knockout, five ptPGAM1A knockout and four empty vector control lines, with a minimum of two biological replicates and three independent repeats (*i.e.,* six measurements total) per line.

Under 19C LD growth conditions, plastid glycolysis-gluconeogenesis knockout lines showed an approximately 15% reduction in relative growth rate compared to empty vector controls (ptEnolase relative growth rate 0.92± 0.03 cells day^{-1}; ptPGAM1A relative growth rate 0.98± 0.04 cells day^{-1}; empty vector growth rate 1.10 ± 0.03 cells day^{-1}; **Fig. 3A, 3D; Table S4**, sheet 3; ANOVA $P < 0.05$). Under 19C CL conditions, knockout lines showed a more severe (approximately 30%) reduction in relative growth rate compared to controls (ptEnolase relative growth rate 1.02± 0.08 cells day^{-1}; ptPGAM1A relative growth rate 1.00± 0.05 cells day^{-1}; empty vector growth rate 1.45 ± 0.05 cells day-

^{1}; **Fig. 3B, 3D**; **Table S4**, sheet 4; ANOVA P < 0.01), suggesting compromised ability to capitalize on the greater photosynthetic potential afforded by continuous illumination. For lines grown under 8C CL conditions, we observed in contrast an overall reduction in the difference in growth rate between knockout and control lines, with equivalent growth observed for ptEnolase (0.54± 0.02 cells day^{-1}) and ptPGAM1A knockouts (0.56± 0.03 cells day^{-1}), compared to empty vector controls (0.56± 0.03 cells day^{-1}; **Fig. 3C, 3D; Table S4**, sheet 5).

To test the possibility of off-target activity of CRISPR constructs, we complemented mutant lines with blasticidin-resistant overexpressor lines of either ptEnolase or ptPGAM1A genes modified to remove all CRISPR target sequences (**Table S4**, sheet 2). Constructs were built using the constitutive FcpA promoter, and the genes were linked to GFP (43, 59), allowing the verification of protein expression and localization by confocal microscopy. To account for the inherent effects of (i) overexpression of complementation constructs and (ii) GFP and blasticidin deaminase gene expression in complementation lines, additional transformations were generated using CRISPR knockout lines, and placebo (cytoplasmic GFP) blasticidin-resistance vectors; and using zeocin-resistant empty vector lines transformed with each blasticidin-resistance complementation construct.

We then performed comparative growth curves of 47 complemented versus placebo transformed mutant lines (**Fig. S13A; Table S4**, sheet 7) under the 19C CL and 19C LD conditions in which the primary knockout lines showed slower growth rates than empty vector controls. We noted a general reduction in growth rates of blasticidin complemented lines compared to primary zeocin-resistant transformants, which may suggest secondary effects of the ectopic overexpression of the blasticidin deaminase and GFP-link constructs in the complemented lines, alongside some within-population variation. Nonetheless, a significantly increased growth rate was observed for complemented versus blank transformed knockout lines, as inferred by one-way ANOVA of separation between complemented and blank transformed mean growth rates. This was observed under both 19C LD (ptEnolase complemented knockout growth rate 1.07± 0.03 cells day^{-1}, blank transformed growth rate 1.00± 0.07 cells day^{-1}, ANOVA Pvalue= 7.4 x $10^{-05;}$ ptPGAM1A complemented knockout growth rate 1.12± 0.12 cells day^{-1}, blank transformed growth rate 1.02± 0.10 cells day^{-1}, P = 2 x 10^{-04}) and CL conditions (ptEnolase complemented knockout growth rate 1.36± 0.13 cells day^{-1}, blank transformed growth rate 1.29± 0.08 cells day^{-1}, P = 5 x 10^{-03}; ptPGAM1A complemented knockout growth rate 1.43± 0.15 cells day^{-1}, blank transformed growth rate 1.23± 0.14 cells day^{-1}, P = 5 x 10^{-06}). In contrast, complemented knockout line growth rates overlapped with blank transformed empty vector controls either transformed with ptEnolase (19C LD growth rate 1.07± 0.04 cells day^{-1}, 19C CL growth rate 1.37± 0.12 cells day^{-1}), ptPGAM1A (19C LD growth rate 1.04± 0.09 cells day^{-1}, 19C CL growth rate 1.28± 0.08 cells day^{-1}) or blank complementing vectors (19C LD growth rate 1.04± 0.03 cells day^{-1}, 19C CL growth rate 1.36± 0.12 cells day^{-1}), indicating effective recovery of mutant phenotypes (ANOVA P > 0.05; **Fig. S13B; Table S4**, sheet 7).

Limited photo-physiological effects of ptEnolase and ptPGAM1A knockout regardless of light condition

To understand the distinctive growth phenotypes of ptEnolase and ptPGAM1A knockouts, we performed comparative photo-physiological analysis of knockout lines in the two conditions where they presented a growth phenotype, i.e. under 19C LD and 19C CL (**Table S4**, sheets 8-11). We measured the light dependence of photosynthesis and photo-protection (see Methods) in a selected set of control lines (*n*=2), ptPGAM (*n*=3) and ptEnolase knockouts (*n*=6), as well in complemented ptEnolase (*n*=2) and ptPGAM1A (*n*=3) knockout lines in which we observed a suppression of the knockout growth defect compared to complemented control lines. We further availed of a Fluorescence Induction and Relaxation (FIRe) fluorometer to measure the absorption cross-section of photosystem II, σPSII, and Fv/Fm parameters in glycolysis mutant lines under all three growth conditions (i.e., including 8C CL) (70). 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 combination (**Table S4**, sheets 8, 9).

First, considering light-dependent changes in photophysiology, we found no significant difference between lines in rETR(II), which is the electron transfer rate through photosystem II (PSII), whatever the light irradiance (**Fig. S14A**; **Table S4**, sheet 10). We furthermore did not observe any change in the light dependency of the NPQ (non-photochemical quenching versus chlorophyll fluorescence) in each construct (Fig. **S14B**, Table S4, sheet 10). We observed a slight but significant (one-way ANOVA, $P < 0.05$) increase in the σPSII in ptEnolase (318.5± 35.3) and ptPGAM1A knockouts (307.4± 16.7) compared to empty vector control lines (292.3± 8.8) under 19C CL conditions exclusively (**Fig. S15A**; **Table S4**, sheet 11) but not in complemented lines; which may suggest that the photosystem II of plastidial glycolysis mutants has an enhanced capacity to harvest light. We further noted significantly greater maximal non-photochemical quenching rates (NPQ_m) in complemented ptEnolase mutants compared to empty vector controls under both 19C CL and 19C LD conditions (**Fig. S15B**; **Table S4**, sheet 11). Overall, our results indicate that suppression of plastidial glycolysis has limited direct effects on diatom photophysiology, even in the 19 °C CL conditions where substantial growth defects are observed.

Primary metabolic functions of ptEnolase and ptPGAM1A inferred from comparative gene expression analysis of P. tricornutum *knockout lines*

Given the limited direct effects of plastidial glycolysis absence on photosynthesis, we considered the growth rates of mutant lines were due to broader impacts in primary metabolism. To do this, we performed quantitative RNA-seq analysis using 63 RNA samples drawn from five ptEnolase and five ptPGAM1A knockout lines alongside four empty vector controls under all three conditions (**Table S5**, sheet 1; Materials and Methods). Differentially expressed genes were identified by DESeq with fold-difference threshold 2 and P-value 0.05. Complete results are provided in **Table S5**, sheets 5-11. Exemplar Volcano plots of two comparisons (ptEnolase v control lines, and ptPGAM1A v control under 19C CL conditions) are shown in **Fig. S16**, confirming under-accumulation of ptEnolase and ptPGAM1A mRNA in the corresponding knockout lines.

Genome-scale enrichment analyses of the *in silico* localizations of proteins encoded by differentially expressed genes revealed distinctive changes in glycolysis knockout organelle metabolism compared to control lines. These effects were most strongly observed in 19C CL conditions, in which of the 239 genes differentially upregulated (mean fold-change >2, P-value < 0.05) in both ptEnolase and ptPGAM1A knockout lines compared to controls, 85 (36%) were inferred to possess chloroplast targeting peptides based on ASAFind (47) and 63 (26%) were found to inferred to possess chloroplast targeting peptides based on HECTAR predictions (49), compared respectively to 13% and 4.8% of genes globally encoded across the entire genome (chi-squared $P < 10^{-05}$; **Fig. 4A**). Less dramatic changes were evident in 19C LD and 8C CL conditions, reflecting the less exaggerated growth phenotypes of glycolysis-gluconeogenesis knockout lines under each condition, although we noted that 13 of the 51 genes (25%) inferred to be downregulated in both ptEnolase and ptPGAM1A knockout lines under 8C CL conditions were inferred to encode chloroplast-targeted proteins by ASAFind, representing likewise an enrichment compared to all genes within the genome (chi-squared $P < 0.05$; **Fig. 4A**).

Considering the KEGG, PFAM, GO and transcriptional repartition of the differentially expressed genes in each knockout line (**Table S5**, sheets 5-8), we noted clear patterns of differential upregulation of genes encoding photosynthesis-related proteins in both ptEnolase and ptPGAM1A knockout lines, specifically under 19C CL conditions. These included (chi-squared $P < 0.05$) GO enrichments in light-harvesting complex (GO:0030076), photosynthesis (GO:0009765) and protein-chromophore linkage (GO:0018298) GO terms, the chlorophyll-binding (PF00504) PFAM domain. We additionally

considered the accumulation of genes that are differentially expressed in knockout lines in PhaeoNet, a WGCNA (Weighted Genome Correlation Network Analysis)-based division of the entire *Phaeodactylum* genome into thirty co-regulated gene modules (41); and noted a chi-squared $P < 0.05$ enrichment of differentially expressed genes in the lightcyan1 module, which has been linked to core processes in light harvesting and photosynthesis (41). Detailed resolution of gene expression patterns underpinning core organelle metabolism pathways (41) in RNAseq data from ptEnolase (**Fig. S17A**) and ptPGAM1A (**Fig. S17B**) knockout lines under 19C CL conditions confirmed, a concerted upregulation of genes encoding light-harvesting complexes and photosynthesis machinery and plastidial fatty acid synthesis machinery, alongside a probable upregulation of mitochondrial respiratory complex I and ATP synthase, suggesting that these transcriptional changes reflect concerted changes in both plastid and mitochondrial metabolism, and potentially plastid-mitochondrion crosstalk (92).

To gain a more precise insight into the effects of plastidial glycolysis-gluconeogenesis on *Phaeodactylum* chloroplast and mitochondria-targeted metabolism, we additionally validated by qPCR the differential expression of eleven exemplar genes encoding chloroplast- and mitochondria-targeted proteins in knockout and empty vector control lines across all three conditions (**Fig. 4B**; **Table S5**, sheet 12). These data confirmed a relatively limited difference in chloroplast metabolism under 19C LD conditions, limited to a slight (~50% downregulation, *t*-test $P < 0.05$) depression in the accumulation of *Lhcf1* (Phatr3_J18049) and chorismate mutase (Phatr3_J43277) mRNA in both ptEnolase and ptPGAM1A knockouts compared to control lines (**Fig. 4B**). Both knockout lines were found to over-accumulate (>600%; *t*-test $P < 10^{-05}$) mRNAs encoding mitochondrial phospho-glycerate mutase (Phatr3_J33839) under 19C LD conditions compared to control lines (**Fig. 4B**).

Under 19C CL conditions, we observed more significant changes in plastid metabolism, including the over-accumulation of mRNAs encoding Lhcf1 (~150%, *t*-test $P < 0.05$) and a plastid-targeted petB-type protein presumably involved in cytochrome b_6f metabolism (Phatr3_J13558, ~90%, *t*-test $P < 0.05$) in both ptEnolase and ptPGAM1A knockout lines, consistent again with the increased σPSII observed (**Fig. 4B**; **Fig. S15**). We further noted a significant over-accumulation in either ptEnolase or ptPGAM1A lines of mRNAs encoding plastidial signal processing peptidase (Phatr3_J10319), alanine transaminase (Phatr3_J34010) or copropophyrinogen oxygenase (Phatr3_J12186; **Fig. 4B**), consistent with a broader upregulation of plastid metabolic activity. We additionally noted a strong (>100%, $P < 10^{-05}$) increase in mRNA abundance of the gene encoding lysophosphatidyl acyltransferase (Phatr3_J20640), which diverts plastidial glyceraldehyde-3-phosphate into lipids via 3-phosphoglycerate, in both ptEnolase and ptPGAM1A knockout lines. Concerning mitochondrial metabolism, we noted a strong increase (>250%, $P < 10^{-05}$) in mRNA for NDH dehydrogenase subunit 1 (Phatr3_J43944), involved in mitochondrial oxidative phosphorylation, but a corresponding decrease (>40%, $P < 10^{-05}$) in mRNA for citrate synthase within the TCA cycle (Phatr3_J30145).

Finally, under 8C and CL conditions, we noted contrasting and complementary changes: upregulation (>60%; $P < 10^{-05}$) of genes encoding both the plastidial signal processing peptidase and petB-related protein, and mitochondrial PGAM and citrate synthases in both knockout lines compared to controls (**Fig. 4B**). These data suggest both enhanced flux through photosystems and through mitochondrial carbon metabolism in the absence of plastidial glycolysis-gluconeogenesis. Of note, *Lhcf1* mRNA was found to severely under-accumulate (>90%; $P < 10^{-05}$), but instead *Lhcx4* (Phatr3_J38720), encoding a dark-expressed homologue of the Lhcx1 protein implicated in photo-protection although of unknown direct function (31), was found to substantially over-accumulate in both ptEnolase and ptPGAM1A knockout lines (**Fig. 4B**).

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

Next, we considered the likely metabolic consequences for the differential gene expression patterns observed in ptEnolase and ptPGAM1A knockouts under each environmental condition studied via GC-MS profiling of 32 soluble sugar and amino acids (**Fig. 5; Fig. S18**) across 139 samples drawn from seven ptEnolase and five ptPGAM1A knockout lines alongside four empty vector controls under 19C LD, 19C CL and 8C CL conditions. Complete tabulated outputs are provided in **Table S6**, sheets 1-2. Accounting for within-line variance, we noted multiple concordances in the metabolites differentially accumulated in ptEnolase and ptPGAM1A knockout lines, with significant correlations between the P-values for metabolite accumulation in each knockout line observed under 19C CL (p= 0.448, t-test P-value < 0.01), and 8C CL conditions (p= 0.324, t-test P-value < 0.05), with an effectively zero correlation (p= 0.017, non-significant) observed in 19C LD conditions. We were unable to directly measure the accumulation of any of the products or substrates of either ptPGAM1A or ptEnolase (3-phosphoglycerate, 2-phosphoglycerate, PEP), although we detected significantly diminished (ANOVA P-value < 10^{-05}) pyruvate accumulation in ptPGAM1A knockouts under all three conditions, and in ptEnolase knockouts under 8C CL (**Fig. 5, S18**).

In all three conditions, we noted significant reductions (ANOVA P-value < 0.01 in both ptEnolase and ptPGAM1A knockout lines) in cytoplasmic sugars and sugar derivatives (glucose, sucrose, histidine, *myo*-inositol) in ptEnolase and ptPGAM1A knockouts compared to control lines (**Fig. 5**). Despite a high overall variance between samples, we noted a probable over-accumulation of phosphoric acid in all knockout lines except ptPGAM1A under 19C CL conditions (**Fig. 5, S18**). We further noted an under-accumulation in both ptEnolase and ptPGAM1A knockout lines of citric acid in all three conditions, and malic acid in 8C CL (**Fig. 5**). Finally, we identified significant (ANOVA P-value < 10^{-05}) over-accumulations of valine in ptEnolase and ptPGAM1A knockouts under 19C CL and 8C CL conditions; in methionine and ornithine in 19C CL conditions only; and an under-accumulation of arginine under 19C CL conditions only (**Fig. 5**).

We also noted specific differences in the metabolite accumulation patterns observed in ptEnolase and ptPGAM1A knockout lines (**Fig. 5; S18**). These include a significant (ANOVA P < 10^{-05}) over-accumulation of three amino acids (aspartate, leucine and phenylalanine) and one sugar phosphate (glycerol-3-phosphate) specifically in ptEnolase knockout lines under all three conditions, and in serine under 19C CL and 8C CL conditions only; which contrasts to ptPGAM1A knockouts in which no significant changes were observed, or (in the case of glycerol-3-phosphate and serine) these metabolites were found to significantly under-accumulate under all three conditions compared to controls **(Fig. 5**; **S18**). These differences may be related to a greater severity of the disruption of plastidial glycolysis in ptEnolase knockouts, for which there is only one non-redundant enzyme predicted in the *Phaeodactylum* genome, compared to redundant ptPGAM enzymes that may partly compensate for ptPGAM1A activity (**Fig. S5**). It is furthermore possible that some of these metabolites relate to the specific activities of ptEnolase and ptPGAM1A in both the glycolytic and gluconeogenic directions, notably ptEnolase whose glycolytic product and gluconeogenic substrate (PEP) is a direct precursor of both aspartate and phenylalanine synthesis (13).

Contrasting impacts of plastidial glycolysis-gluconeogenesis on lipid profiles depend on temperature

Finally, to gain a greater insight into the different effects of fluxes through ptEnolase and ptPGAM enzymes under different light and temperature conditions, we performed GC-MS (55 samples) and LC-MS (49 samples) of five ptEnolase and five ptPGAM1A knockout lines alongside four empty vector controls under 19C LD, 19C CL and 8C CL conditions (**Fig. 6; Fig. S19**). Outputs are tabulated in **Table S6**, sheets 1, 3-5. Lipid quality may be influenced by multiple metabolites impacted by plastidial glycolytic-gluconeogenic flux, including pyruvate and acyl-coA pools (in the glycolytic-gluconeogenic direction), glycerol-3-phosphate (synthesised from glyceraldehyde-3-phosphate via glycerol-3-phosphate dehydrogenase (34)) and glucosyl-1-phosphate (synthesised from exported glucosyl-1-phosphate generated in the gluconeogenic direction, via UDP-glucosyl PPiase and UDP-glucose

epimerase, as described above (35)), with compound effects on fatty acid chain length, saturation state, and lipid head group dependent on the stoichiometric availability of each metabolite (93). Lipid remodelling has also previously been implicated in diatom adaptations to light and temperature, considering crude amounts (94), fatty acid chain length (95), and relative investments in triacylglycerol to plastid membrane (galacto-) lipids (96, 97).

Growth conditions were observed to have strong effects on fatty acid profiles observed through GC-MS, with a significant (ANOVA $P < 0.05$) replacement of C20:5 side chain lipids with C16:1 side chain lipids in 19C CL compared to both 19C LD and 8C CL conditions; and over-accumulation of C16:3 side chain lipids under 19C LD, and C18:0 side chain lipids under 8C CL conditions, respectively (**Fig. S19**). The enhanced accumulation of short-chain saturated lipid side chains under continuous illumination may reflect enhanced flux into fatty acid synthesis over elongation and desaturation, e.g., for the synthesis of storage triacylglycerols, whereas the enhanced accumulation of long-chain and unsaturated lipid side chains at 8C may serve to enhance membrane fluidity in response to low temperature (58, 63, 93). We noted no substantial differences in fatty acid profiles between ptEnolase, ptPGAM1A and control lines under any conditions studied (**Fig. S19)**.

In contrast to the relatively limited effects on fatty acid profiles, we noted substantial changes in lipid class distributions in plastidial glycolysis-gluconeogenesis knockout lines compared to controls (**Fig. 6**; **Table S6,** sheet 4), as identified by LC-MS. Both ptEnolase and ptPGAM1A knockouts were found to under-accumulate triacylglycerols (TAG) compared to control lines under 19C LD (ptEnolase 2.72 ± 1.45%, ptPGAM1A 3.39 ± 1.79%, control 8.33 ± 2.36; ANOVA P separation of means between knockout and control lines < 0.001) and 19C CL conditions (ptEnolase 21.23 ± 3.58%, ptPGAM1A 20.16 ± 6.99%, control 30.35 ± 3.76%; ANOVA $P < 0.001$). We further noted that both ptEnolase and potPGAM1A knockout lines significantly over-accumulate mono-galactosyl-diacylglycerol (MGDG; ptEnolase 64.77 ± 2.52%, ptPGAM1A 62.75 ± 4.45%, control 53.92 ± 2.77%; ANOVA $P < 10^{-05}$) and under-accumulate di-galactosyl-diacylglycerol (DGDG; ptEnolase 7.62 ± 0.38%, ptPGAM1A 7.57 ± 1.12%, control 10.05 ± 0.38%; ANOVA $P < 10^{-05}$) and sulfoquinovosyl diacylglycerol (SQDG; ptEnolase 9.37 ± 0.61%, ptPGAM1A 9.22 ± 1.82%, control 11.80 ± 1.75%; ANOVA $P < 0.01$) in 19C CL conditions only (**Fig. 6; Table S6**, sheet 4).

Considering the individual profiles of lipid side chains under 19C conditions (**Table S6,** sheet 5), we noted increased levels of C16:1 *sn*-1 fatty acids in glycolysis-gluconeogenesis knockouts compared to control lines under 19C conditions (**Figs. S20, S21**). These included a conserved and significant increase (ANOVA $P < 10^{-05}$) under 19C CL in SQDG-16-1_16-0 in both ptEnolase (25.6± 1.17%) and ptPGAM1A knockouts (26.1± 1.93%) compared to control lines (20.8± 0.73%), and more broadly encompassed C16:1 *sn*-1 TAG, DGDG and PC (**Fig. S21**). In contrast, both ptEnolase and ptPGAM1A knockouts were found to have a significant ($P < 10^{-05}$) corresponding increase in 20:5 fatty acids in the *sn*-1 positions of the betaine lipid diacylglyceryl hydroxymethyltrimethyl-*β*-alanine (DGTA), implicated in lipid catabolism, recycling and TAG synthesis in *Phaeodactylum* under 19C CL conditions (**Fig. S21**) (98, 99). We additionally observed significant ($P < 0.05$) under-accumulations of C14:0, C16:0 and C16:1 *sn*-1 fatty acids in glycolysis knockout MGDG pools under both 19C CL and 19C LD conditions (**Fig. S21B; Table S6**, sheet 5). These data in total suggest recycling of 20:5 fatty acids and replacement with shorter chain equivalents, although the accumulation of C16:1 *sn*-1 SQDG may also be driven by under-accumulation of C14:0 equivalents and rerouting of short-chain DAG from MGDG into SQDG synthesis under both 19C LD and 19C CL conditions (**Figs. S20B, S21B**), (100).

Under 8C CL conditions, we observed quite different trends in fatty acid accumulation specifically in ptEnolase knockouts, compared to both ptPGAM1A knockouts and controls. Considering lipid class, these changes related principally to over-accumulations of TAG (ptEnolase 29.2 ± 3.05%, ptPGAM1A 9.19 ± 4.35%, control 5.27 ± 2.30%; ANOVA $P < 10^{-10}$) and SQDG (ptEnolase 12.0 ± 1.98%, ptPGAM1A 8.96 ± 2.05%, control 7.79 ± 0.73%; ANOVA $P < 0.01$) in lieu of MGDG (ptEnolase 31.3 ± 3.17%,

ptPGAM1A 44.3 ± 3.36%, control 51.3 ± 1.34%; ANOVA $P < 10^{-10}$; **Fig. 6**). Within the side-chain distributions of individual lipid classes, we further noted a significant (ANOVA $P < 10^{-05}$) over-accumulation of short-chain (C14:0, C16:1) and *sn*-1 fatty acids in ptEnolase knockout plastid lipid pools (i.e., MGDG, DGDG; **Fig. S22A**), although the *sn*-1 most abundant fatty acid in each lipid group remained C20:5 (**Fig. S22B**). We finally noted a probable exchange of very long-chain *sn*-2 fatty acids in SQDG pools in ptEnolase knockouts, with significant (ANOVA $P < 10^{-05}$) increases in SQDG 14-0_16-0 and SQDG-14_0-16-1 in lieu of SQDG-16-2_24-0 in in ptEnolase knockouts compared to ptPGAM1A and control lines (**Fig. S22**; **Table S6**, sheet 5).

Discussion

In this paper, we present the evolutionary history (**Figs. 1, S1-S3**), gene expression characteristics (**Figs. 2, S4-S11),** functional mutagenesis (**Figs. 3, S12-13**) and physiological analysis (**Figs. 4-6, S14-22**) of a lower half glycolytic-gluconeogenic pathway associated with diatom chloroplasts, relating specifically to two plastid-targeted proteins, ptEnolase and ptPGAM1A, in the model species *P. tricornutum*. Our data position plastid glycolysis-gluconeogenesis as arising recurrently across secondary red chloroplasts, via the duplication of genes encoding respiratory proteins (**Fig. 1**), placing this as one of many pathways of host evolutionary origin that have been recruited to support the diatom plastid (23, 101, 102). This innovation sits alongside others associated with diatom carbon metabolism, including mitochondrial targeting of respiratory glycolysis and synthesis of storage sugars in cytosolic vacuoles as chrysolaminarin (35, 38); although in contrast to these features (which are respectively shared with all stramenopiles, and with the primary red algal ancestor of the diatom chloroplast, as floridoside storage (103)) the plastidial glycolysis-gluconeogenesis isoforms associated with diatoms are uniquely shared with their closest relatives (e.g., pelagophytes, dictyochophytes) and potential endosymbiotic derivatives (haptophytes) (23).

Considering previously published transcriptomic data from *P. tricornutum*, we show that plastidial glycolysis-gluconeogenesis genes are co-expressed with those encoding core plastidial and mitochondrial carbon metabolism (**Fig. S6**), with more distant regulatory connections to other biosynthetic pathways (e.g., fatty acid synthesis; **Figs. S20-22**). We note that the role of plastidial glycolysis-gluconeogenesis is likely to not directly impact photosynthetic efficiency, given the limited photo-physiological phenotypes of mutant lines (**Figs. S14-S15**), although may indirectly impact on photosynthesis gene expression (**Fig. 4; Fig. S17**). From both *P. tricornutum* and *Tara* Oceans data, we show evidence that the functions of this pathway may be particularly pertinent at high latitudes (**Figs. 2, S7-S10**), e.g., under elongated photoperiods and low temperatures. This phenotype reflects the growth rates observed for complemented knockout mutant lines of *P. tricornutum* ptEnolase and ptPGAM1A, where we note a retardation in growth in glycolysis-gluconeogenesis knockout lines under 19C continuous light compared to 19C light-dark conditions (**Figs. 3, S12-13**). Despite clear physiological defects observed in glycolysis-gluconeogenesis knockout line under 8C continuous light conditions (e.g., upregulation of Lhcx4 and downregulation of Lhcf1 genes, and upregulation of mitochondrial TCA cycle and glycolysis genes, **Fig. 4B**; and differential accumulation of multiple metabolites and lipids; **Figs. 5-6**), we note equivalent growth rates in knockout and control lines (**Fig. 3**), suggesting additional physiological roles of plastidial glycolysis-gluconeogenesis at low temperatures may compensate for the growth defects observed due to continuous illumination.

Considering both gene expression dynamics (**Figs. 4; S15-17**) and metabolite profiles (**Figs. 5, 6; S18-S22**) in knockout and control lines grown under 19C LD, 19C CL and 8C CL conditions, we reconstruct a map of the potential functions contributed by the lower half of plastidial glycolysis-gluconeogenesis in *P. tricornutum*, shown in **Fig. S23**. Under 19C LD conditions, we observe limited gene expression changes in ptPGAM1A and ptEnolase knockout lines, except (as inferred from qPCR) a downregulation in plastidial shikimate biosynthesis and upregulation of mitochondrial lower-half glycolysis-gluconeogenesis pathways (**Fig. 4B**). The downregulation of a shikimate biosynthesis

enzyme intermediate (which might typically consume plastidial PEP (13)) might suggest a primary function of plastid glycolysis-gluconeogenesis in the glycolytic direction for augmenting plastidial PEP for shikimate biosynthesis, whereas upregulation of mitochondrial glycolysis-gluconeogenesis may reflect an enhanced flux of plastidial glyceraldehyde-3-phosphate to the mitochondrion in the absence of a direct plastidial route for its metabolism. We additionally note diminished levels of cytoplasmic sugars and TCA cycle intermediates (i.e., citric acid) in ptEnolase and ptPGAM1A knockout lines (**Fig. 5**) (104), whereas the overexpression of mitochondrial phospho-glycerate mutase observed in qRT-PCR and RNAseq data (**Figs. 4B, S17**) does not suggest substrate limitation of the TCA cycle in glycolysis knockout lines. Thhe diminished accumulation of TCA cycle intermediates and cytoplasmic sugars might instead suggest a diversion of glycolytic intermediates into anabolic activities in the *Phaeodactylum* mitochondrion. Finally, we note in both ptEnolase and ptPGAM1A mutants an over-accumulation of phosphoric acid, which may point to decreased phosphate usage (e.g., in phospholipid or nucleic acid synthesis) and increased storage as compatible solutes. An over-accumulation of phosphoric acid may equally relate to changes in the cellular ADP/ ATP ratio (105), although we did not identify any ATP-associated GOs amongst those enriched in glycolysis-gluconeogenesis knockout DEGs under any condition (**Table S4**, sheet 6)

Under 19C CL conditions, we observed much more dramatic remodelling of chloroplast and mitochondrial metabolic architecture, consistent with reduced growth rates in knockout lines (**Figs. 3, 4A, S15, S17, S21**). These include greater overall investment in photosynthesis gene expression and carbon metabolism (including the Calvin Cycle) compared to control lines (**Figs. 4B, S17**). We did not observe consistent differences in the expression of nitrogen or phosphorus stress metabolism in knockout lines compared to controls, suggesting that these difference do not relate to difference in N or P limitation between knockout and the faster-growing control lines under 19C CL conditions (**Table S5,** sheets 4-5) (59, 104, 106). These changes are supported by the greater absorption cross-section σPSII identified in photo-physiological analyses, but do not appear to impact substantially on measured Fv/ Fm (**Figs. S13**; **S14**), and we propose that any changes to photosynthesis machinery expression in glycolysis mutant lines are likely to occur proportionally to one another and to downstream carbon metabolism pathways, without apparent changes in photosynthetic efficiency. In contrast, we observed an apparent disequilibration in the expression of respiration-associated proteins (i.e., upregulation of NDH dehydrogenase and downregulation of TCA cycle enzymes), and probable upregulation of mitochondrial and peroxisomal proteins implicated in alternative electron flow in glycolysis knockout lines (**Fig. S17**). Previous studies have noted the important role of diatom mitochondria in dissipating excess plastidial reducing potential (30, 92), and it remains to be determined to what extent the export of plastidial NADPH to the mitochondria in glycolysis-gluconeogenesis knockout lines differs compared to controls under continuous light.

Under both 19C LD and 19C CL conditions, we note probable changes in the stoichiometry of lipid classes (i.e., MGDG in lieu of DGDG, TAG and SQDG production in 19C LD; MGDG and SQDG in lieu of TAG and DGDG production in 19C CL; **Fig. 6**); and in fatty acid chain length in specific lipid pools (e.g., greater accumulation of 16:0 and 16:1 fatty acids in plastidial lipid pools, and 20:5 fatty acids in DGTA pools, Figs. **S20, S21**). These trends contrast with *Arabidopsis* ptEnolase and ptPGAM mutants, which present relatively limited lipid accumulation phenotypes (18, 19). The relative over-accumulation of TAG in control lines under 19C conditions may be an indirect product of the diminished growth rates in glycolysis knockout lines, insofar as TAG accumulation is typically associated with late- over early-exponential phase *Phaeodactylum* lines (**Fig. 4**) (63, 107). However, neither ptEnolase nor ptPGAM1A knockout lines show apparent changes in the expression of nitrogen or phosphate starvation-associated genes, which as described above are typically viewed to provoke TAG over-accumulation, compared to control lines (59, 104, 106). The relative change in the equilibrium of MGDG to TAG might suggest acyl-coA limitation of lipid synthesis, or greater investment into MGDG and SQDG synthesis from UDP-glucose produced from exported plastidial glyceraldehyde-3-phosphate (85) (**Fig. 6**). Supporting this, our RNAseq and qRT-PCR data suggest an overexpression of enzymes involved in

fatty acid incorporation (acyl carrier protein, enoyl-ACP reductase, lysophospholipid acyltransferase I; **Figs. 4, S17**) despite limited transcriptional changes to genes encoding lipid head group remodelling and cytoplasmic sugar biosynthesis (e.g., UTP-glucose-1-phosphate uridylyltransferase, Phatr3_J32708; and UDP-Glc epimerases, Phatr3_J46785, Phatr3_J44401; **Table S5**, sheet 3) (*35*(35) in glycolysis knockout lines under 19C CL conditions. The relative under-accumulation of glycose and sucrose, and overexpression of respiratory glycolysis enzymes in glycolysis knockout lines (**Figs. 4, 5**) is further coherent with a redirection of cytoplasmic triose phosphate from gluconeogenesis towards acyl-coA production, i.e. if plastidial glycolytic activity would otherwise function to directly replenish plastidial acyl-coA pools from Calvin cycle intermediates (**Fig. S23**). We note, however, from transcriptional coregulation data an apparent lack of coordination between the expression patterns of ptEnolase, ptPGAM1A and enzymes involved in fatty acid synthesis (**Fig. S6**); and it is possible that this pathway plays a secondary role to others, such as shikimate biosynthesis, for the consumption of plastidial pyruvate and PEP (**Fig. S23**).

It remains to be determined what routes enable the supply of pyruvate hub intermediates, to the *Phaeodactylum* plastid in the absence of plastidial glycolytic activity. Previous studies have noted the dual affinity of diatom triose phosphate transporters for glyceraldehyde-3-phosphate/di-hydroxyacetone phosphate and PEP (17, 36). One of the two annotated plastid triose phosphate transporters in the *Phaeodactylum* genome (Phatr3_J54017) is upregulated in both ptEnolase and ptPGAM1A knockout lines under 19C CL conditions only (**Table S5**, sheet 3), which may suggest some degree of direct PEP uptake from the cytoplasm (36). In both ptEnolase and ptPGAM1A knockout lines we further note an overexpression of the glycine cleavage system H protein (Phatr3_J32847) and underexpression of ammonia carbamoyl-phosphate synthase (Phatr3_J24195; **Fig. S16**), which may reflect an augmented importance of the mitochondrial glycine cycle in recovering photo-respired glycolate in glycolysis mutant lines.

Elsewhere our data suggest an enhanced importance of organelle amino acid metabolism in the absence of plastidial glycolysis. These include an overexpression of plastidial alanine transaminase (Phatr3_J34010; **Fig. 4B, S16**); and differential accumulation of multiple amino acids (over-accumulation of valine, ornithine and methionine and under-accumulation of arginine in both mutant lines; overaccumulation of aspartate in ptEnolase knockout lines; and overaccumulation of glutamate in ptPGAM1A knockout lines; **Fig. 5**) under 19C CL conditions. Alanine synthesised in the *Phaeodactylum* mitochondrion could, for example, be imported into the plastid and deaminated to supply pyruvate (104), and an over-accumulation of pyruvate might in turn lead to an increased synthesis of valine (92, 108). We noted an over-accumulation of mRNA for plastidial branched chain-amino acid aminotransferase (Phatr3_J10779, involved in isoleucine: valine transamination) in both ptEnolase and ptPGAM1A knockouts in 19CL RNAseq data (**Fig. S17**); which may be related to dysregulation in branched-chain amino acid synthesis or catabolism (108). Aspartate, in contrast, might be synthesised either in the *Phaeodactylum* mitochondria from oxaloacetate or peri-plastidial compartment (via PEP, from PEP carboxylase (109), and its overaccumulation outside the plastid might in turn augment flux into methionine synthesis (via homocysteine) (13). It is unlikely that aspartate would directly supply PEP to the *Phaeodactylum* plastid due to the absence of known plastidial C4 decarboxylases in this species, but its synthesis might enable the effective recycling of extra-plastidial PEP into pyruvate pools (110, 111). Ornithine, synthesised in the *Phaeodactylum* plastid at the expense of plastidial glutamate, might be used to potentiate the deamination of alanine or aspartate via a keto-glutarate shuttle (104, 112), prior to being converted into arginine (113), or dissipated via a potential urea cycle in the mitochondria (114). The depletion of both arginine and citrate in both glycolysis knockout lines, a further intermediate in the proposed mitochondrial urea cycle, and corresponding accumulation of glutamate in ptPGAM1A knockout lines, suggests that ornithine over-accumulation likely relates to enhanced plastidial ornithine synthesis as opposed to diminished mitochondrial dissipation (**Fig. 5, S23**).

Under 8C CL conditions, we note congruent and contrasting changes in mutant phenotypes to those identified at 19C (**Fig. S23**). As per 19C conditions, these include over-accumulation of mRNAs encoding photosystem subunit, plastid biogenesis and mitochondrial glycolytic proteins, and an over-accumulation of short-chain amino acids (valine) and under-accumulation of cytoplasmic sugars and amino acids (glucose, histidine) in ptEnolase and ptPGAM1A knockouts relative to empty vector controls (**Figs. 4B, 5**). Confirming the explicit significance of plastidial glycolysis for carbon metabolism under 8C CL conditions, we note under-accumulations both pyruvate and in glycerol, synthesised typically from plastidial triose-phosphate via 3-phosphoglycerate dehydrogenase in both mutant lines (**Fig. 5**). The absence of a specific growth defect under 8C conditions is therefore perhaps surprising (**Fig. 3**), although additional phenotypic differences in knockout lines unique to 8C may suggest more complicated physiological consequences for the absence of plastidial glycolysis. These include an overall enrichment in downregulated genes encoding plastid-targeted proteins (**Fig. 4A**); a specific over-accumulation of TCA cycle (citrate synthase, Phatr3_J30145) and a possible non-photochemical quenching-associated mRNA (LhcX4, Phatr3_J38720) (**Fig. 4B**) (30, 115); and an over-accumulation of TAGs and SGDQs over glucosyl-lipids (**Fig. 6**) in both ptEnolase and ptPGAM1A knockout lines at 8C only. We tentatively propose that some of the contrasting investments observed in glycolysis knockout lines under 8C (e.g., diversion of fixed carbon either into mitochondrial respiration or anabolic pathways) conditions may to some extent compensate for phenotypes (i.e., inhibited flux into pyruvate and remodelling of carbon metabolism) observed in 19C CL conditions, thus yielding equivalent growth rates between mutant and control lines (**Fig. 3; Fig. S23**).

Finally, across all three conditions, we note specific phenotypic differences between ptEnolase and ptPGAM1A knockout lines, including the ptEnolase-specific accumulation of several amino acids (aspartate, phenylalanine, serine, leucine) across all three studied conditions (**Fig. 5**; **Fig. S18**). The overaccumulation of phenylalanine is particularly striking given the apparent downregulation of mRNA of chorismate mutase, (Phatr3_J43277) in both RNAseq and qRT-PCR data (**Fig. 4B; Fig. 5; Fig. S17**) (13), suggesting strong metabolic forcing of its synthesis. We further notea predominant accumulation of short-chain *sn*-1 fatty acids in ptEnolase plastid lipid pools under 8C CL conditions (**Fig. S22**), which may act as a further electron sink in response to low temperature and continuous illumination (116). The exact reasons for these contrasting phenotypes remain to be determined, although may relate to the fact ptEnolase is projected to be the only plastid-targeted enolase protein encoded in the *Phaeodactylum* genome, whereas alternative phospho-glycerate mutases to exist that can compensate for ptPGAM1A (e.g., ptPGAM1B, ptPGAM2; **Figs. S4, S6**); and each enzyme performs different metabolic activities, with ptPGAM1A converting 3-phosphoglycerate to 2-phosphoglycerate, and ptEnolase converting 2-phosphoglycerate to PEP in the glycolytic direction, and the reverse fluxes anticipated in the gluconeogenic direction (**Fig. 1A)**. The absence of ptEnolase activity in a gluconeogenic direction, in particular, would lead to an over-accumulation of PEP in the *Phaeodactylum* plastid, which could explain the over-accumulation of aspartate (synthesised via PEP carboxylase (13, 104, 109)); phenylalanine (if excess PEP were diverted into shikimate biosynthesis) and potentially even leucine (via pyruvate) (13). In contrast, the excess of serine and short-chain *sn*-1 amino acids might relate to an absence of activity in the glycolytic direction (i.e., diminished flux of plastidial triose phosphate into acyl-coA pools from either plastidial or exported mitochondrial pyruvate (93), and increased flux into glycerol-3-phosphate synthesis (13)). The relevance of plastidial glycolysis in the gluconeogenic direction is particularly relevant under 8C CL conditions, where several of the phenotypes common to both ptEnolase and ptPGAM1A knockouts (including increased investment into respiration and neutral lipid accumulation, and diminished investment into plastid biogenesis and core plastid metabolism enzymes) might equally well be explained by a disruption of mitochondrial-to-plastid metabolic flux to support plastid primary carbon metabolism (**Fig. S23**). Ultimately, the relative importance of metabolic flux in each direction of each diatom plastid glycolysis enzyme will be best determined by kinetic studies, e.g. radio-labelled substrate tracing, in mutant lines under each growth condition.

Concluding Remarks

Overall, our data position the lower half of glycolysis-gluconeogenesis as providing a functionally redundant, but nonetheless significant component of the diatom plastid proteome, evidenced by the presence of viable mutants with observable growth and physiological phenotypes. We infer using combined growth, gene expression, photo-physiology and metabolomics analysis that plastidial glycolysis permits more efficient cycling of glyceraldehyde-3-phosphate from the Calvin cycle to the pyruvate hub, for a wide range of downstream pathways, with particular importance under continuous illumination such as the tested 19C CL condition. We additionally provide some more tentative evidence for functions of this pathway in the gluconeogenic direction, particularly in the context of the differential accumulation of metabolites specific to ptEnolase knockout and 8C CL conditions. The poise and interactions between the glycolytic and gluconeogenic functions of diatom ptEnolase and ptPGAM1A may in part contribute to the functional success of diatoms at high latitudes (27), where they must tolerate extreme annual variances in photoperiod and low temperatures (25, 26). The glycolytic functions of diatom plastidial glycolysis-gluconeogenesis may enable diatoms to dominate in the initial spring pulse of biomass that accompanies photoperiod induction (117, 118), and across the light-saturated conditions of the continuous polar summer by allowing augmented metabolic activity through the Calvin cycle (119). In contrast, a potential gluconeogenic activity may have relevance throughout the polar night (120) where photosynthesis is not possible, and might potentially induce photosynthetic activity via replenishment of the Calvin cycle from mitochondrial intermediates, in response to transient and seasonal enrichments in nutrients (e.g., in autumnal polar blooms (121)). It remains to be determined what roles plastidial glycolysis plays in other habitats where diatoms dominate phytoplankton biomass, e.g. in upwelling cycles where diatoms may preciously upregulate components of their core metabolism to rapidly avail of optimal light and nutrient conditions (119, 122)

It furthermore remains to be determined what functions a lower half of plastidial glycolysis-gluconeogenesis perform elsewhere across the tree of life. This question is pertinent not only to other eukaryotic algal groups (**Figs. 1; S1-S2**); but also in plants, where the presence of plastidial glycolysis-gluconeogenesis is to our knowledge uncharacterized in photosynthetic cells. Plastid glycolysis-gluconeogenesis in leaves has been proposed to be involved in the metabolism of imported sugars at night (123), and the synthesis of fatty acids and isoprenoids from excess glyceraldehyde-3-phosphate during the day (124), and lower half glycolysis-gluconeogenesis proteins have further been proposed to perform a secondary role as a structural component of protein contact points between plant plastids and mitochondria (125); but the eco-physiological roles of plant plastid glycolysis-gluconeogenesis remain underexplored. Notably, cytoplasmic phospho-gycerate mutase (126-128) and enolase enzymes (129-131) have been identified in proteomic surveys of plant and algal cold-responsive genes, but it remains to be determined whether plastidial versions of these enzymes possess functions relevant to high latitudes. Our data shed light into functions of a poorly understood metabolic resource that may facilitate diatom success at high latitudes, and that may serve as a candidate for metabolic engineering in crop species for enhanced photosynthetic activity under continuous light and low temperatures (33).

Acknowledgments

RGD acknowledges a CNRS Momentum Fellowship, awarded from 2019-2021, an ANR JCJC Grant (« PanArctica » ANR-21-CE02-0014-01) and an ANR t-ERC Grant ("ChloroMosaic"), awarded from 2021-2022. 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). The lipid analysis 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, 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). D.C and B.B acknowledge the support by 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.

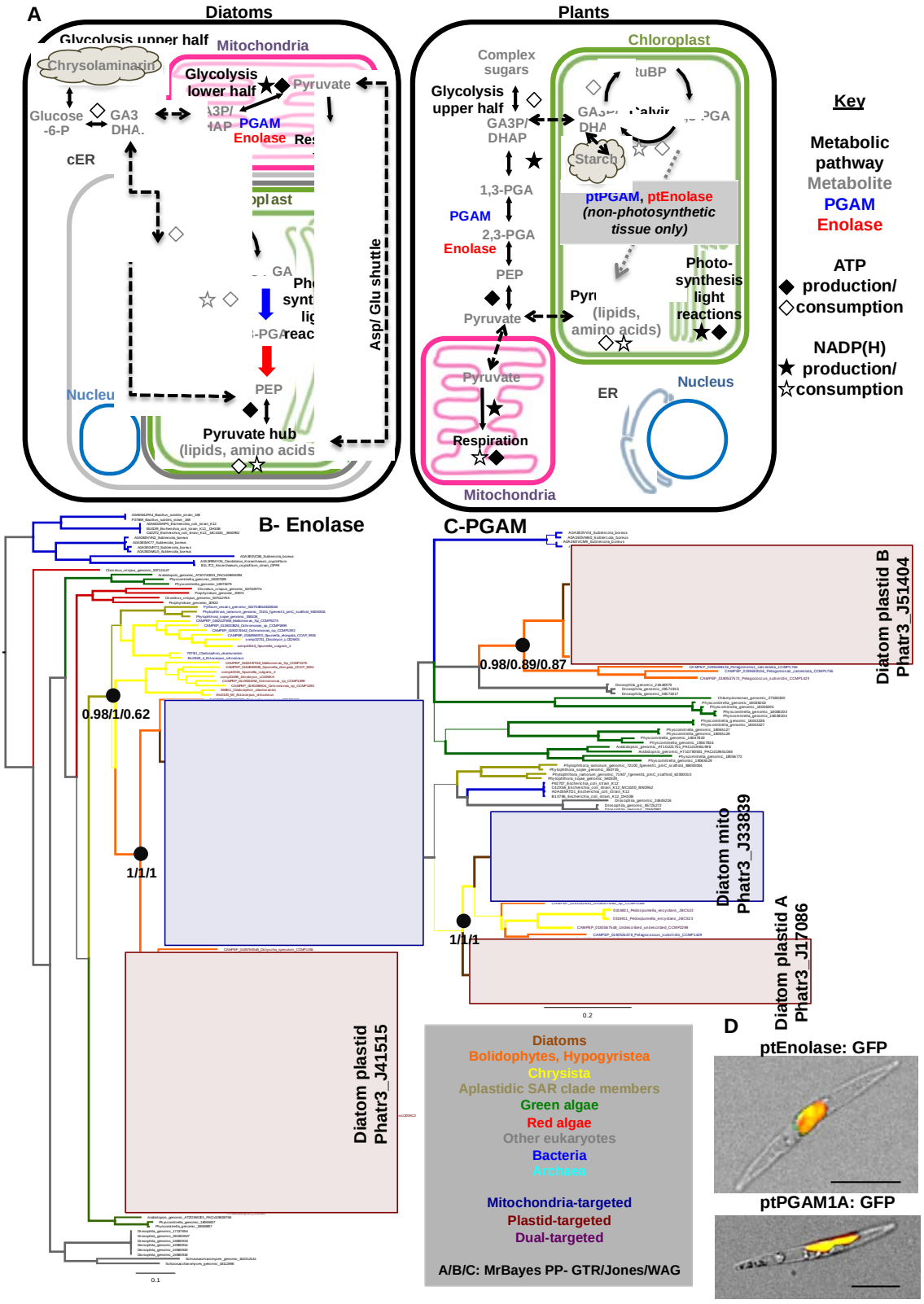


Fig. 1. Metabolic context and evolution of the lower half of diatom plastidial glycolysis-gluconeogenesis. A: schematic comparison of diatom and plant core carbon metabolism, adapted

from (34), 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. **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. Phylogenies realised with full datasets are provided in **Figs. S1-S3**. **D**: overlay images of GFP-tagged full-length ptEnolase (top) and ptPGAM1A (bottom) constructs (green), chlorophyll (red) and bright-field images of transformant *Phaeodactylum tricornutum* lines. Individual images for each channel and cytoplasmic GFP and GFP-free control lines confirming specificity and sensitivity of the excitation and emission conditions used are provided in **Fig. S4**.

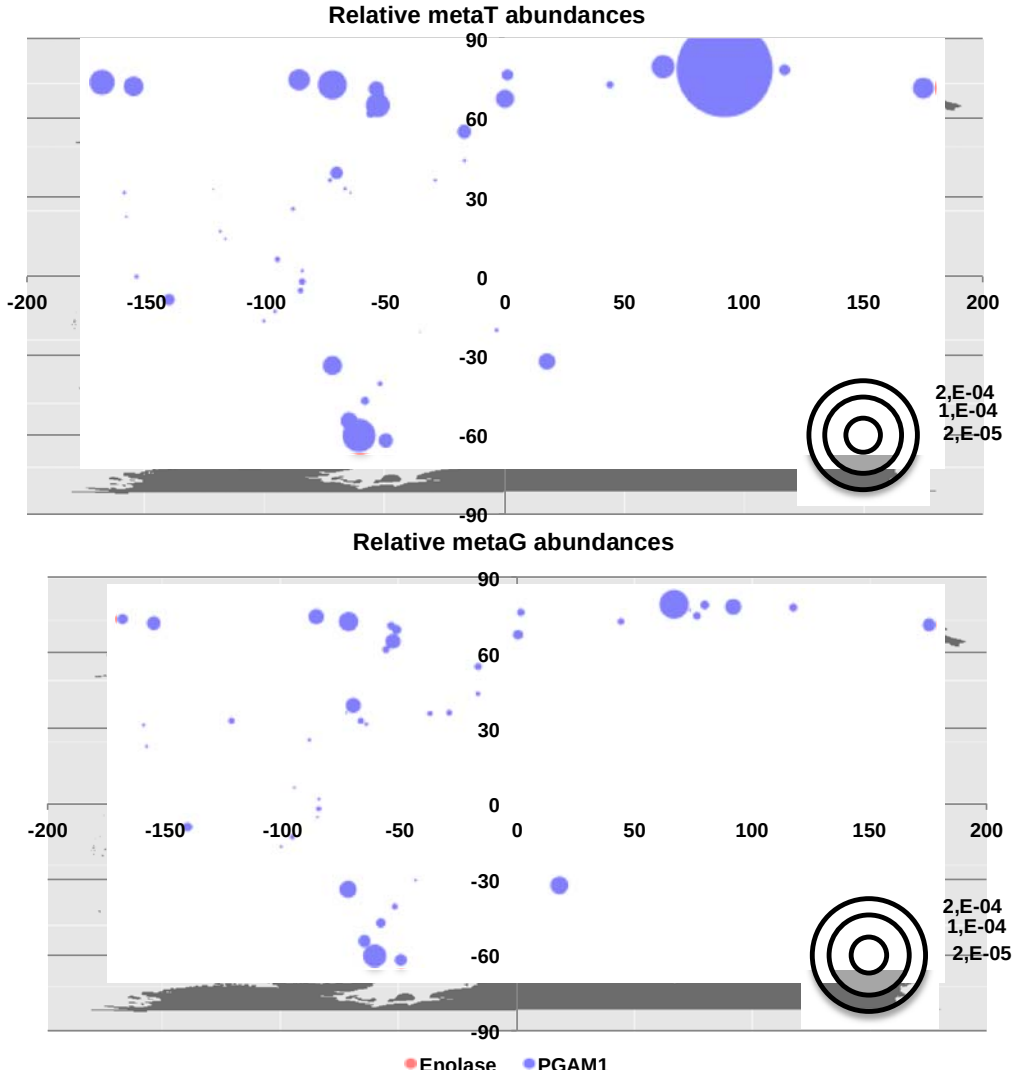


Fig. 2. Environmental distributions of diatom plastidial glycolysis-gluconeogenesis meta-genes. Total meta-transcriptome (top) and meta-genome (bottom) relative abundances for *Tara* Oceans meta-genes phylogenetically resolved to diatom ptEnolase and ptPGAM1A families (64), sampled from all size fractions and surface layer stations, and demonstrating higher meta-transcript abundance without commensurate increases in meta-gene abundance at high northern and southern latitudes. Individual abundance maps for each size fraction and depth, and the correlations of meta-

genome normalised meta-transcriptome abundance to different quantitative environmental variables, are provided in **Fig. S9-S11**.

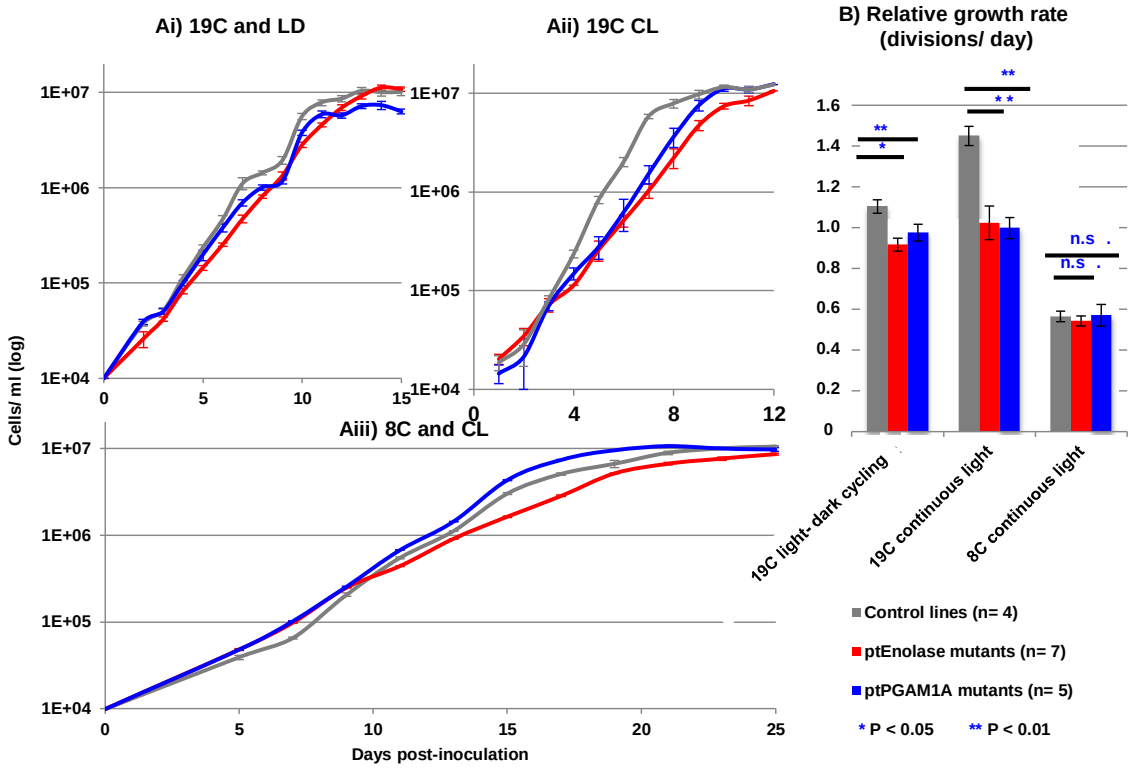


Fig. 3. Growth phenotypes of ptEnolase and ptPGAM1A CRISPR-Cas9 knockout mutant and zeocin-resistant empty vector control *P. tricornutum* lines. A: exemplar growth curves from single experiments realised for *P. tricornutum* lines in 50 µE m^{-2} s^{-1} illumination, non-shaken cultures and replete 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 »). **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). Plastid glycolysis-gluconeogenesis knockout lines show a substantial retardation in growth rate under 19C CL compared to 19C LD conditions, which is compensated by 8C CL conditions. Equivalent growth curves realised with compensated and blasticidin-resistant lines are shown in **Fig. S13**.

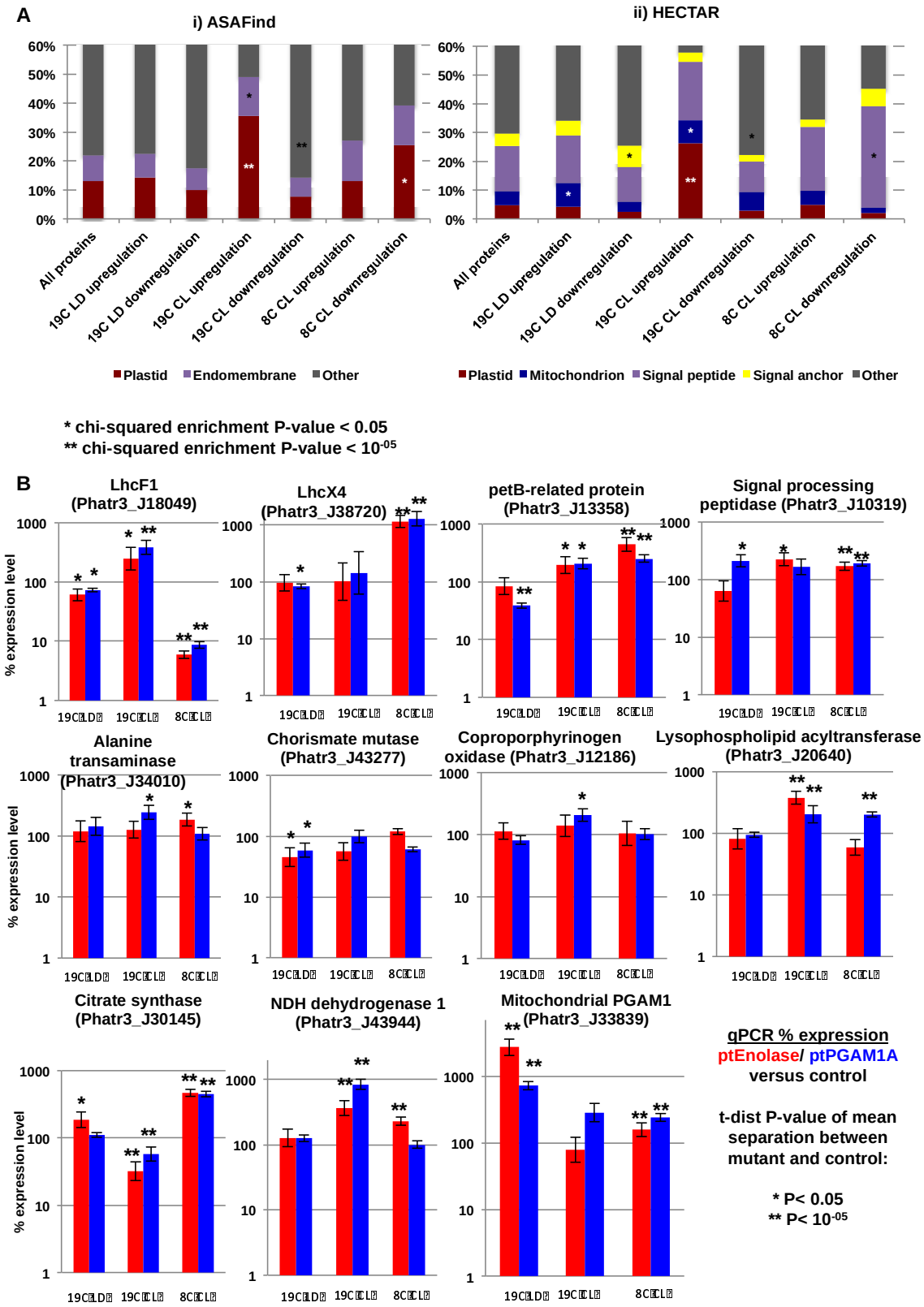


Fig. 4. Changes in plastid and mitochondrial metabolic architecture inferred from gene expression analyses. A: bar plots of the predicted localizations from ASAFind (47) and HECTAR (49) of all genes inferred (P < 0.05, fold-change expression >2) to be differentially up- or down-regulated in both

ptEnolase and ptPGAM1A knockout compared to control lines under 19C LD, 19C CL and 8C CL conditions. Localizations that are significantly enriched in gene expression categories (chi-squared $P < 0.05$), including a significant over-accumulation of genes with chloroplast- and mitochondria-targeting sequences upregulated in knockout lines under 19C CL conditions, are asterisked. Volcano plots and detailed organelle metabolism maps for genes differentially expressed under 19C CL in ptEnolase and ptPGAM1A lines are provided in **Figs. S16-S17**. **B:** relative mRNA abundances of eleven genes encoding exemplar chloroplast- and mitochondria-targeted proteins, verified by qRT-PCR. Abundances were calculated using two sets of qRT-PCR primers per gene and normalised against two housekeeping genes (Ribosomal protein S1, and TATA binding protein (78)); across five ptEnolase, seven ptPGAM1A and four empty vector control lines, with four reverse transcriptase positive and four reverse transcriptase negative replicates each. Each bar plots thus the mean and standard deviation of expression ratio of all knockout genotypes and replicates for each line. Genes differentially expressed (*t*-test, $P < 0.05$) in each condition are asterisked.

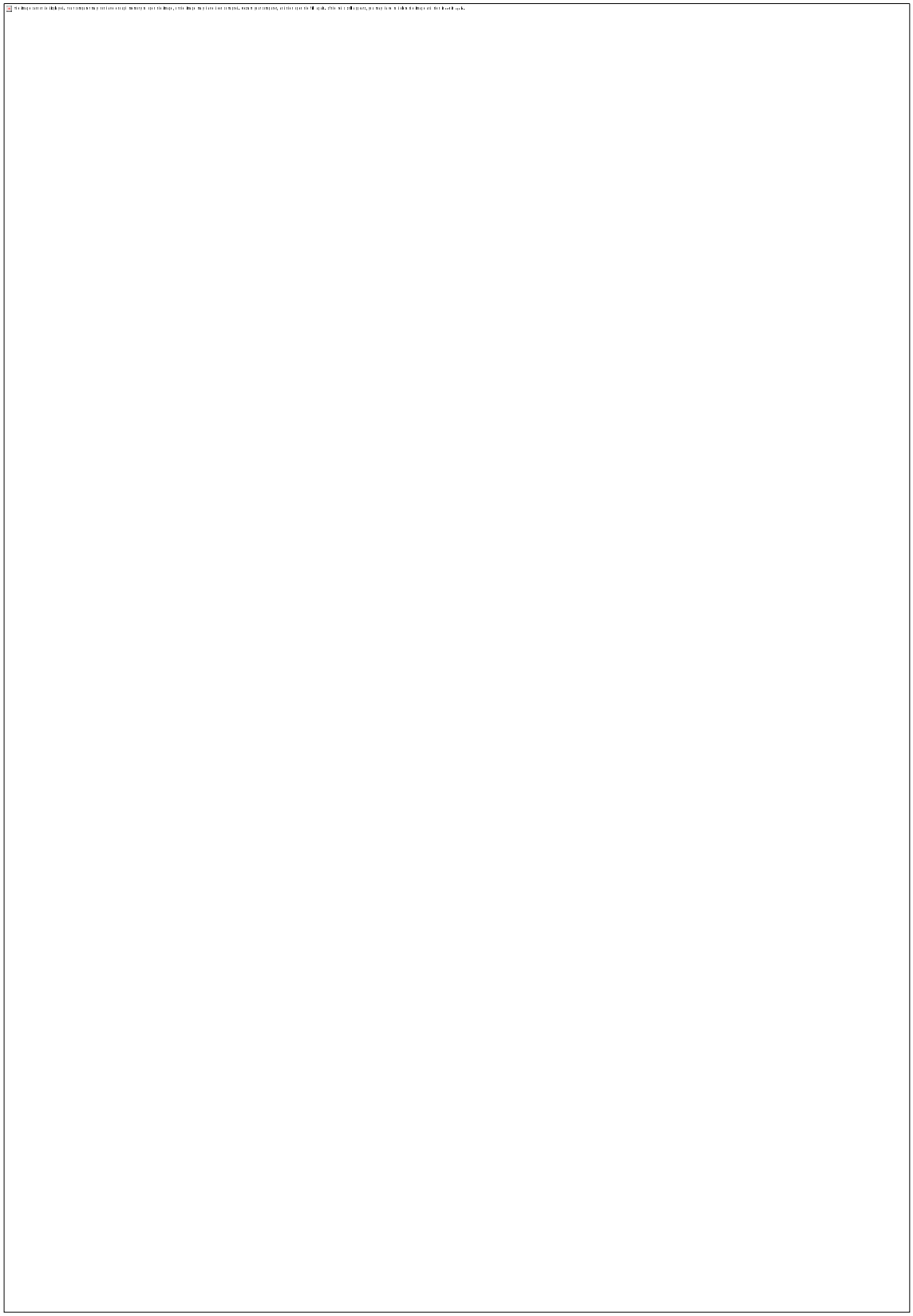


Fig. 5. Volcano plots of differentially accumulated metabolites assessed by GC-MS. Scatterplots of the $\log_2$ accumulation ratios $-\log_{10}$ P-values of difference in the mass, ribitol and quality-control-

normalised abundances of 39 sugar and amino acid metabolites in ptEnolase and ptPGAM1A 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, and metabolites that show a differential accumulation in both knockout lines in each condition are shown in black text, and five metabolites that are uniquely over-accumulated in ptEnolase knockout lines under all three conditions are shown in dark red text. Bar plots showing relative differences in metabolite abundance are shown in **Fig. S18**.

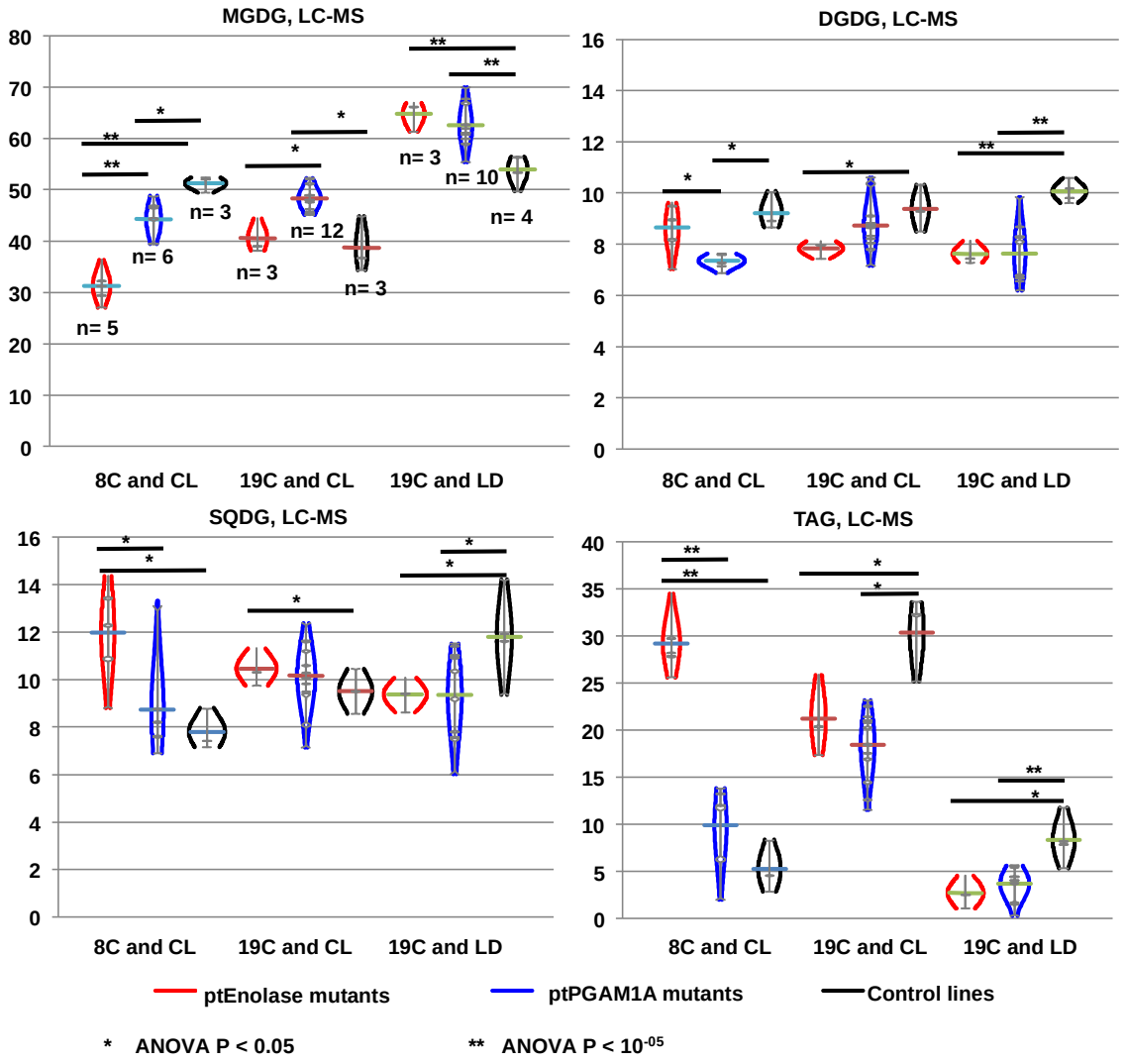


Fig. 6. Violin plots of LC-MS lipid distributions in glycolysis-gluconeogenesis mutant lines. This figure shows plots of the observed frequencies (short bars), mean (long bars) and inferred normal distribution of the % of lipids found to belong to one of four lipid categories in ptEnolase and ptPGAM1A knockout and empty vector control lines under 19C LD, 19C CL and 8C CL conditions. Significant differences between knockout and control lines (ANOVA, $P < 0.05$) are asterisked. Equivalent plots of GC-MS fatty acid distributions, showing limited differences between knockout and control lines are provided in **Fig. S19**. ANOVA P-value distributions of *sn-1* and *sn-2* fatty acids associated with different lipid classes are shown in **Figs. S20-S22**.

References

1. Raines CA. The Calvin cycle revisited. *Photosynth Res*. **2003**;75(1):1-10. doi: 10.1023/A:1022421515027.
2. Martin W, Schnarrenberger C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr Genet*. **1997**;32(1):1-18. doi: 10.1007/s002940050241.
3. Hügler M, Sievert SM. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Ann Rev Mar Sci*. **2011**;3:261-89. doi: 10.1146/annurev-marine-120709-142712.
4. Malhi Y, Grace J. Tropical forests and atmospheric carbon dioxide. *Trends Ecol Evol*. **2000**;15(8):332-7. doi: 10.1016/s0169-5347(00)01906-6 .
5. Moog D, Rensing SA, Archibald JM, Maier UG, Ullrich KK. Localization and evolution of putative triose phosphate translocators in the diatom *Phaeodactylum tricornutum*. *Genom Biol Evol*. **2015**;7(11):2955-69. doi: 10.1093/gbe/evv190.
6. Scialdone A, Mugford ST, Feike D, Skeffington A, Borrill P, Graf A, et al. *Arabidopsis* plants perform arithmetic division to prevent starvation at night. *Elife*. **2013**;2:e00669. doi: 10.7554/eLife.00669.
7. Carrera D, George GM, Fischer-Stettler M, Galbier F, Eicke S, Truernit E, et al. Distinct plastid fructose bisphosphate aldolases function in photosynthetic and non-photosynthetic metabolism in *Arabidopsis*. *J Exp Bot*. **2021**;72(10):3739-55. doi: 10.1093/jxb/erab099.
8. Anoman AD, Flores-Tornero M, Rosa-Telléz S, Muñoz-Bertomeu J, Segura J, Ros R. The specific role of plastidial glycolysis in photosynthetic and heterotrophic cells under scrutiny through the study of glyceraldehyde-3-phosphate dehydrogenase. *Plant Signal Behav*. **2016**;11(3):e1128614. doi: 10.1080/15592324.2015.1128614.
9. Troncoso-Ponce MA, Rivoal J, Dorion S, Sánchez R, Venegas-Calerón M, Moreno-Pérez AJ, et al. Molecular and biochemical characterization of the sunflower (*Helianthus annuus* L.) cytosolic and plastidial enolases in relation to seed development. *Plant Sci*. **2018**;272:117-30. doi: 10.1016/j.plantsci.2018.04.007.
10. Flores-Tornero M, Anoman AD, Rosa-Téllez S, Toujani W, Weber AP, Eisenhut M, et al. Overexpression of the triose phosphate translocator (TPT) complements the abnormal metabolism and development of plastidial glycolytic glyceraldehyde-3-phosphate dehydrogenase mutants. *Plant J*. **2017**;89(6):1146-58. doi: 10.1111/tpj.13452.
11. Igamberdiev AU, Kleczkowski LA. The glycerate and phosphorylated pathways of serine synthesis in plants: The branches of plant glycolysis linking carbon and nitrogen metabolism. *Front Plant Sci*. **2018**;9:318. doi: 10.3389/fpls.2018.00318.
12. Muñoz-Bertomeu J, Cascales-Miñana B, Alaiz M, Segura J, Ros R. A critical role of plastidial glycolytic glyceraldehyde-3-phosphate dehydrogenase in the control of plant metabolism and development. *Plant Signal Behav*. **2010**;5(1):67-9. doi: 10.4161/psb.5.1.10200.
13. Bromke MA. Amino acid biosynthesis pathways in diatoms. *Metabolites*. **2013**;3(2):294-311. doi: 10.3390/metabo3020294.
14. Coesel S, Obornik M, Varela J, Falciatore A, Bowler C. Evolutionary origins and functions of the carotenoid biosynthetic pathway in marine diatoms. *PLoS One*. **2008**;3(8). doi: 10.1371/journal.pone.0002896.
15. Tanaka R, Tanaka A. Tetrapyrrole biosynthesis in higher plants. *Ann. Rev. Plant Biol*. **2007**;58:321-46. doi: 10.1146/annurev.arplant.57.032905.105448.

16. Shtaida N, Khozin-Goldberg I, Boussiba S. The role of pyruvate hub enzymes in supplying carbon precursors for fatty acid synthesis in photosynthetic microalgae. *Photosynth Res*. **2015**;125(3):407-22. doi: 10.1007/s11120-015-0136-7.
17. Moog D, Nozawa A, Tozawa Y, Kamikawa R. Substrate specificity of plastid phosphate transporters in a non-photosynthetic diatom and its implication in evolution of red alga-derived complex plastids. *Sci Rep*. **2020**;10(1):1167. doi: 10.1038/s41598-020-58082-8.
18. Andriotis VM, Kruger NJ, Pike MJ, Smith AM. Plastidial glycolysis in developing *Arabidopsis* embryos. *New Phytol*. **2010**;185(3):649-62. doi: 10.1111/j.1469-8137.2009.03113.x.
19. Prabhakar V, Löttgert T, Gigolashvili T, Bell K, Flügge UI, Häusler RE. Molecular and functional characterization of the plastid-localized Phosphoenolpyruvate enolase (ENO1) from *Arabidopsis thaliana*. *FEBS Lett*. **2009**;583(6):983-91. doi: 10.1016/j.febslet.2009.02.017.
20. Fukayama H, Masumoto C, Taniguchi Y, Baba-Kasai A, Katoh Y, Ohkawa H, et al. Characterization and expression analyses of two plastidic enolase genes in rice. *Biosci Biotechnol Biochem*. **2015**;79(3):402-9. doi: 10.1080/09168451.2014.980219.
21. Polle JE, Neofotis P, Huang A, Chang W, Sury K, Wiech EM. Carbon partitioning in green algae (chlorophyta) and the enolase enzyme. *Metabolites*. **2014**;4(3):612-28. doi: 10.3390/metabo4030612.
22. Moriyama T, Sakurai K, Sekine K, Sato N. Subcellular distribution of central carbohydrate metabolism pathways in the red alga *Cyanidioschyzon merolae*. *Planta*. **2014**;240(3):585-98. doi: 10.1007/s00425-014-2108-0.
23. Dorrell RG, Gile G, McCallum G, Méheust R, Bapteste EP, Klinger CM, et al. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *Elife*. **2017**;6: 23717. doi: 10.7554/eLife.23717.
24. Bowler C, Vardi A, Allen AE. Oceanographic and biogeochemical insights from diatom genomes. *Ann. Rev. Mar. Sci*. **2010**;2:333-65. doi: 10.1146/annurev-marine-120308-081051.
25. de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, et al. Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science*. **2015**;348(6237):1261605. doi: 10.1126/science.1261605.
26. Malviya S, Scalco E, Audic S, Vincent F, Veluchamy A, Poulain J, et al. Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci USA*. **2016**;113(11):E1516-25. doi: 10.1073/pnas.1509523113.
27. Gilbertson R, Langan E, Mock T. Diatoms and their microbiomes in complex and changing polar oceans. *Front Microbiol*. **2022**;13:786764. doi: 10.3389/fmicb.2022.786764.
28. Ye N, Han W, Toseland A, Wang Y, Fan X, Xu D, et al. The role of zinc in the adaptive evolution of polar phytoplankton. *Nat Ecol Evol*. **2022**;6(7):965-78. E doi: 10.1038/s41559-022-01750-x.
29. Browning TJ, Achterberg EP, Rapp I, Engel A, Bertrand EM, Tagliabue A, et al. Nutrient co-limitation at the boundary of an oceanic gyre. *Nature*. **2017**;551(7679):242-6. doi: 10.1038/nature24063.
30. Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, et al. Energetic coupling between plastids and mitochondria drives CO2 assimilation in diatoms. *Nature*. **2015**;524(7565):366-U267. doi: 10.1038/nature14599.
31. Buck JM, Sherman J, Bártulos CR, Serif M, Halder M, Henkel J, et al. Lhcx proteins provide photoprotection via thermal dissipation of absorbed light in the diatom

Phaeodactylum tricornutum. *Nat Commun*. **2019**;10(1):4167. doi: 10.1038/s41467-019-12043-6.
32. Gao X, Bowler C, Kazamia E. Iron metabolism strategies in diatoms. *J Exp Bot*. **2021**;72(6):2165-80. doi: 10.1093/jxb/eraa575.
33. Cardona T, Shao S, Nixon PJ. Enhancing photosynthesis in plants: the light reactions. *Essays Biochem*. **2018**;62(1):85-94. doi: 10.1042/EBC20170015.
34. Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, et al. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One*. **2008**;3(1):e1426. doi: 10.1371/journal.pone.0001426.
35. Zhu B-H, Shi H-P, Yang G-P, Lv N-N, Yang M, Pan K-H. Silencing UDP-glucose pyrophosphorylase gene in *Phaeodactylum tricornutum* affects carbon allocation. *New Biotechnol*. **2016**;33(1):237-44. doi: 10.1016/j.nbt.2015.06.003.
36. Liu S, Storti M, Bowler C, Finazzi G, Dorrell RG. An integrative environmental atlas of diatom chloroplast transporters. *Frontiers Plant Sci* **2022**; in press. doi: 10.3389/fpls.2022.950467.
37. Norambuena L, Marchant L, Berninsone P, Hirschberg CB, Silva H, Orellana A. Transport of UDP-galactose in plants. Identification and functional characterization of AtUTr1, an *Arabidopsis thaliana* UDP-galactos/UDP-glucose transporter. *J Biol Chem*. **2002**;277(36):32923-9. doi: 10.1074/jbc.M204081200.
38. Río Bártulos C, Rogers MB, Williams TA, Gentekaki E, Brinkmann H, Cerff R, et al. Mitochondrial glycolysis in a major lineage of eukaryotes. *Genom Biol Evol*. **2018**;10(9):2310-25. doi: 10.1093/gbe/evy164.
39. Smith SR, Abbriano RM, Hildebrand M. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Res*. **2012**;1(1):2-16. doi: 10.1016/j.algal.2012.04.003.
40. Hippmann AA, Schuback N, Moon K-M, McCrow JP, Allen AE, Foster LF, et al. Proteomic analysis of metabolic pathways supports chloroplast–mitochondria cross-talk in a Cu-limited diatom. *Plant Direct*. **2022**;6(1):e376. doi: 10.1002/pld3.376.
41. Ait-Mohamed O, Novák Vanclová AMG, Joli N, Liang Y, Zhao X, Genovesio A, et al. PhaeoNet: a holistic RNAseq-based portrait of transcriptional coordination in the model diatom *Phaeodactylum tricornutum*. *Frontiers Plant Sci*. **2020**;11. doi: 10.3389/fpls.2020.590949.
42. Falciatore A, Casotti R, Leblanc C, Abrescia C, Bowler C. Transformation of nonselectable reporter genes in marine diatoms. *Mar Biotechnol* (NY). **1999**;1(3):239-51. doi: 10.1007/PL00011773.
43. Buck JM, Río Bártulos C, Gruber A, Kroth PG. Blasticidin-S deaminase, a new selection marker for genetic transformation of the diatom. *PeerJ*. **2018**;6:e5884. doi: 10.7717/peerj.5884.
44. Nonoyama T, Kazamia E, Nawaly H, Gao X, Tsuji Y, Matsuda Y, et al. Metabolic innovations underpinning the origin and diversification of the diatom chloroplast. *Biomolecules*. **2019**;9(8). doi: 10.3390/biom9080322.
45. Rastogi A, Maheswari U, Dorrell RG, Vieira FRJ, Maumus F, Kustka A, et al. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of diatoms. *Sci Rep*. **2018**;8(1):4834. doi: 10.1038/s41598-018-23106-x.
46. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res*. **2020**. doi: 10.1093/nar/gkaa913.

47. Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J.* **2015**;81(3):519-28. doi: 10.1111/tpj.12734. P
48. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* **2004**;340(4):783-95. doi: 10.1016/j.jmb.2004.05.028.
49. Gschloessl B, Guermeur Y, Cock JM. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformat.* **2008**;9:393. doi: 10.1186/1471-2105-9-393.
50. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucl Acids Res.* **2016**;44(W1):W3-W10. doi: 10.1093/nar/gkw343.
51. Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. MitoFates: Improved prediction of mitochondrial targeting sequences and their cleavage Sites. *Mol Cell Proteom.* **2015**;14(4):1113-26. doi: 10.1074/mcp.M114.043083.
52. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* **2012**;28(12):1647-9. doi: 10.1093/bioinformatics/bts199.
53. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **2017**. doi: 10.1093/bib/bbx108.
54. Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, et al. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol Bioinform Online.* **2015**;11:43-8. doi: 10.4137/EBO.S21501.
55. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformat.* **2014**;30(9):1312-3. doi: 10.1093/bioinformatics/btu033.
56. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformat.* **2009**;25(15):1972-3. doi: 10.1093/bioinformatics/btp348.
57. Ashworth J, Turkarslan S, Harris M, Orellana MV, Baliga NS. Pan-transcriptomic analysis identifies coordinated and orthologous functional modules in the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. *Mar Genom.* **2016**;26:21-8. doi: 10.1016/j.margen.2015.10.011.
58. Cruz de Carvalho MH, Sun HX, Bowler C, Chua NH. Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytol.* **2016**;210(2):497-510. doi: 10.1111/nph.13787.
59. McCarthy JK, Smith SR, McCrow JP, Tan M, Zheng H, Beeri K, et al. Nitrate reductase knockout uncouples nitrate transport from nitrate assimilation and drives repartitioning of carbon flux in a model pennate diatom. *Plant Cell.* **2017**;29(8):2047-70. doi: 10.1105/tpc.16.00910.
60. Smith SR, Gillard JT, Kustka AB, McCrow JP, Badger JH, Zheng H, et al. Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLoS Genet.* **2016**;12(12):e1006490. doi: 10.1371/journal.pgen.1006490.
61. Nash EA, Barbrook AC, Edwards-Stuart RK, Bernhardt K, Howe CJ, Nisbet RER. Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. *Mol Biol Evol.* **2007**;24(7):1528-36. doi: 10.1093/molbev/msm074.

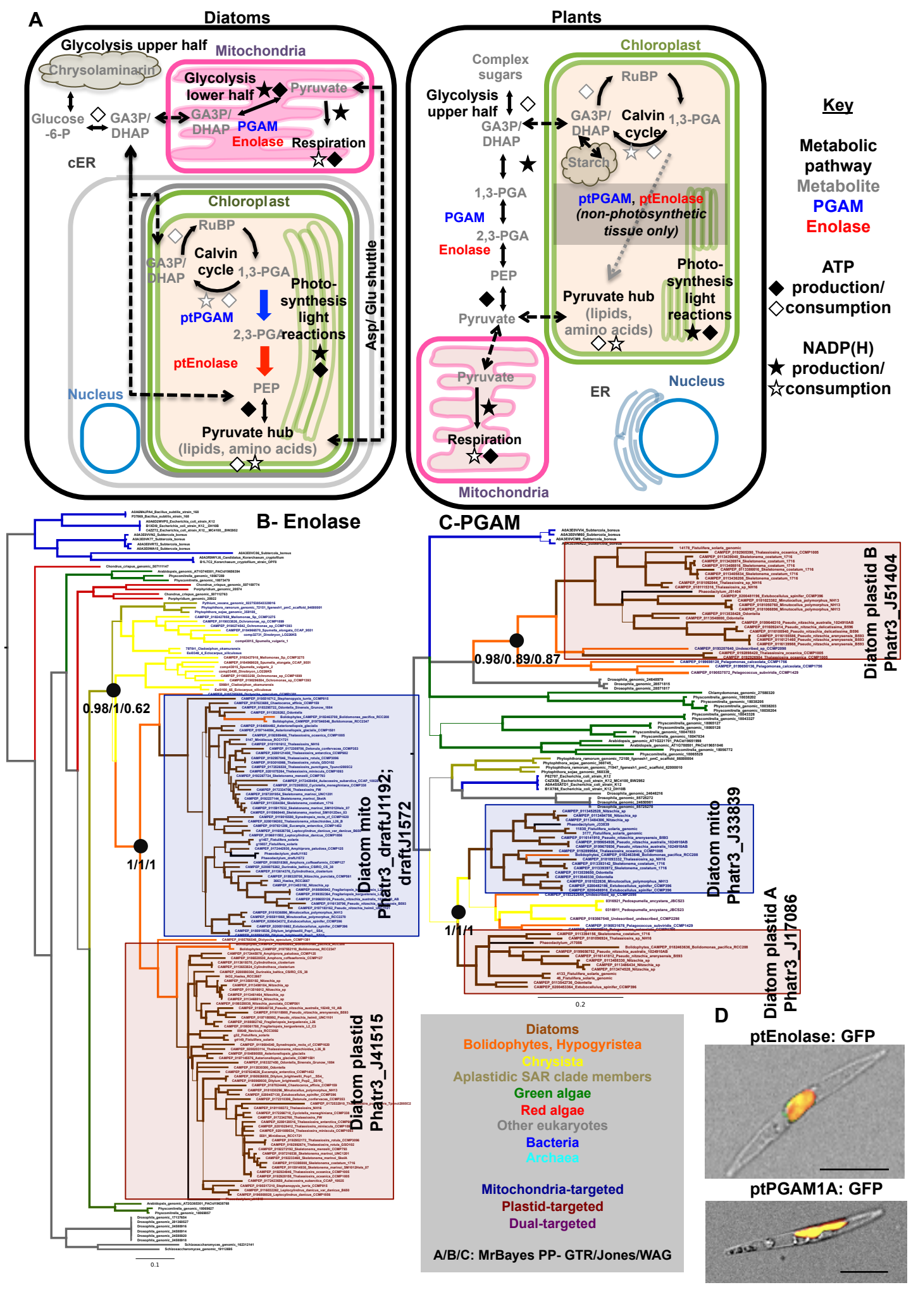
62. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, et al. Open science resources for the discovery and analysis of *Tara* Oceans data. *Scientific Data*. **2015**;2. doi: 10.1038/sdata.2015.23.
63. Abida H, Dolch LJ, Meï C, Villanova V, Conte M, Block MA, et al. Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in *Phaeodactylum tricornutum*. *Plant Physiol*. **2015**;167(1):118-36. doi: 10.1104/pp.114.252395.
64. Carradec Q, Pelletier E, Da Silva C, Alberti A, Seeleuthner Y, Blanc-Mathieu R, et al. A global ocean atlas of eukaryotic genes. *Nat Commun*. **2018**;9(1):373. doi: 10.1038/s41467-017-02342-1.
65. Royo-Llonch M, Sánchez P, Ruiz-González C, Salazar G, Pedrós-Alió C, Sebástian M, et al. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol*. **2021**;6:1561–1574. doi: 10.1038/s41564-021-00979-9.
66. Rastogi A, Murik O, Bowler C, Tirichine L. PhytoCRISP-Ex: a web-based and stand-alone application to find specific target sequences for CRISPR/CAS editing. *BMC Bioinformatics*. **2016**;17(1):261. doi: 10.1186/s12859-016-1143-1.
67. Nymark M, Sharma AK, Sparstad T, Bones AM, Winge P. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci Rep.* **2016**;6:24951. doi: 10.1038/srep24951.
68. Karas BJ, Diner RE, Lefebvre SC, McQuaid J, Phillips AP, Noddings CM, et al. Designer diatom episomes delivered by bacterial conjugation. *Nat Commun*. **2015**;6:6925. doi: 10.1038/ncomms7925.
69. Erdene-Ochir E, Shin BK, Kwon B, Jung C, Pan CH. Identification and characterisation of the novel endogenous promoter HASP1 and its signal peptide from *Phaeodactylum tricornutum*. *Sci Rep*. **2019**;9(1):9941. doi: 10.1038/s41598-019-45786-9.
70. Gorbunov MY, Shirsin E, Nikonova E, Fadeev VV, Falkowski PG. A multi-spectral fluorescence induction and relaxation (FIRe) technique for physiological and taxonomic analysis of phytoplankton communities. *Mar Ecol Progr Ser*. **2020**;644:1-13. doi: 10.3354/meps13358.
71. Serôdio J, Lavaud J. A model for describing the light response of the nonphotochemical quenching of chlorophyll fluorescence. *Photosynth Res*. **2011**;108(1):61-76. doi: 10.1007/s11120-011-9654-0.
72. Lataretu M, Hölzer M. RNAflow: An effective and simple RNA-Seq differential gene expression pipeline using Nextflow. *Genes* (Basel). **2020**;11(12). doi: 10.3390/genes11121487.
73. Kassambara A, Mundt F. Package 'factoextra'. Extract and visualize the results of multivariate data analyses. **2017**;76(2). https://CRAN.R-project.org/package=factoextra
74. Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three differential expression analysis methods for RNA sequencing: limma, EdgeR, DESeq2. *J Vis Exp*. **2021**(175). doi: 10.3791/62528.
75. Dorrell RG, Villain A, Perez-Lamarque B, Audren de Kerdrel G, McCallum G, Watson AK, et al. Phylogenomic fingerprinting of tempo and functions of horizontal gene transfer within ochrophytes. *Proc Natl Acad Sci USA*. **2021**;118(4). doi: 10.1073/pnas.2009974118.
76. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*. **2016**;428(4):726-31. doi: 10.1016/j.jmb.2015.11.006.

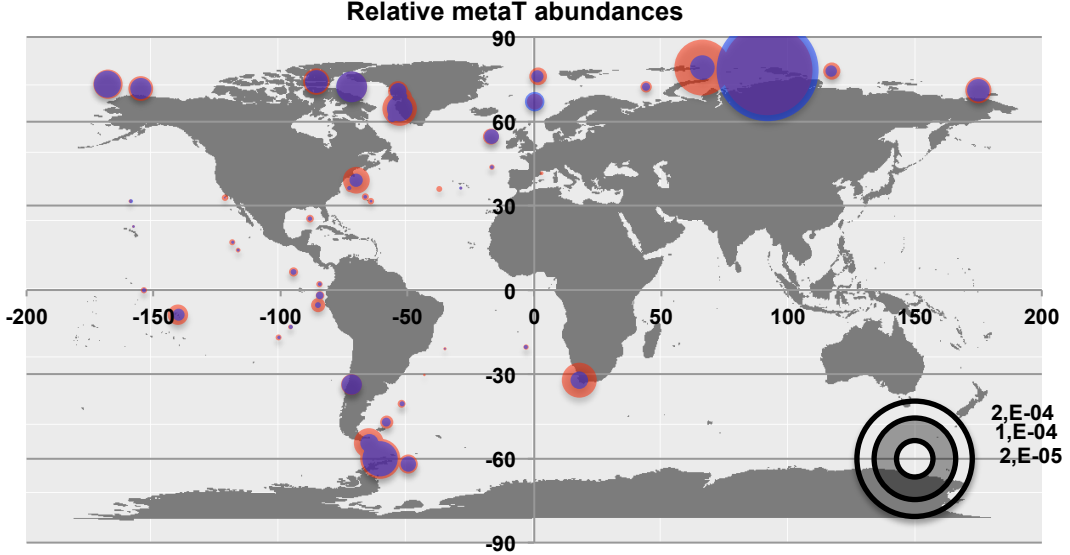
77. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. **2012**;13:134. doi: 10.1186/1471-2105-13-134.
78. Sachse M, Sturm S, Gruber A, Kroth P. Identification and evaluation of 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*. **2013**;24:7.
79. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry–based metabolite profiling in plants. *Nat Protocol*. **2006**;1(1):387-96. doi: 10.1038/nprot.2006.59.
80. Luedemann A, von Malotky L, Erban A, Kopka J. TagFinder: preprocessing software for the fingerprinting and the profiling of gas chromatography-mass spectrometry based metabolome analyses. *Methods Mol Biol*. **2012**;860:255-86. doi: 10.1007/978-1-61779-594-7_16.
81. Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, et al. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformat* **2005**;21(8):1635-8. doi: 10.1093/bioinformatics/bti236.
82. Folch J, Lees M, Stanley GHS. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem*. **1957**;226(1):497-509. doi: 10.1016/S0021-9258(18)64849-5.
83. Jouhet J, Maréchal E, Bligny R, Joyard J, Block MA. Transient increase of phosphatidylcholine in plant cells in response to phosphate deprivation. *FEBS Lett.* **2003**;544(1):63-8. doi: 10.1016/S0014-5793(03)00477-0.
84. Rainteau D, Humbert L, Delage E, Vergnolle C, Cantrel C, Maubert M-A, et al. Acyl chains of Phospholipase D transphosphatidylation products in *Arabidopsis* cells: a study using multiple reaction monitoring mass spectrometry. *PLoS One*. **2012**;7(7):e41985. doi: 10.1371/journal.pone.0041985.
85. Demé B, Cataye C, Block MA, Maréchal E, Jouhet J. Contribution of galactoglycerolipids to the 3-dimensional architecture of thylakoids. *FASEB J*. **2014**;28(8):3373-83. doi: 10.1096/fj.13-247395.
86. Buseman CM, Tamura P, Sparks AA, Baughman EJ, Maatta S, Zhao J, et al. Wounding stimulates the accumulation of glycerolipids containing oxophytodienoic acid and dinor-oxophytodienoic acid in *Arabidopsis* leaves. *Plant Physiol*. **2006**;142(1):28-39. doi: 10.1104/pp.106.082115.
87. Jouhet J, Lupette J, Clerc O, Magneschi L, Bedhomme M, Collin S, et al. 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* **2017**;12(8):e0182423. doi: 10.1371/journal.pone.0182423.
88. Dorrell RG, Novak Vanclova AMG, Penot M, Pierella Karlusich JJ, Bowler C, Liu S, et al. Functional physiology of novel diatom chloroplast proteins. Open Science Foundation **2022** https://osf.io/89vm3/.
89. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformat*. **2014**;30(9):1236-40. doi: 10.1093/bioinformatics/btu031.
90. Pinseel E, Nakov T, Van den Berge K, Downey KM, Judy KJ, Kourtchenko O, et al. Strain-specific transcriptional responses overshadow salinity effects in a marine diatom sampled along the Baltic Sea salinity cline. *ISME J*. **2022**;16(7):1776-87. doi: 10.1038/s41396-022-01230-x.

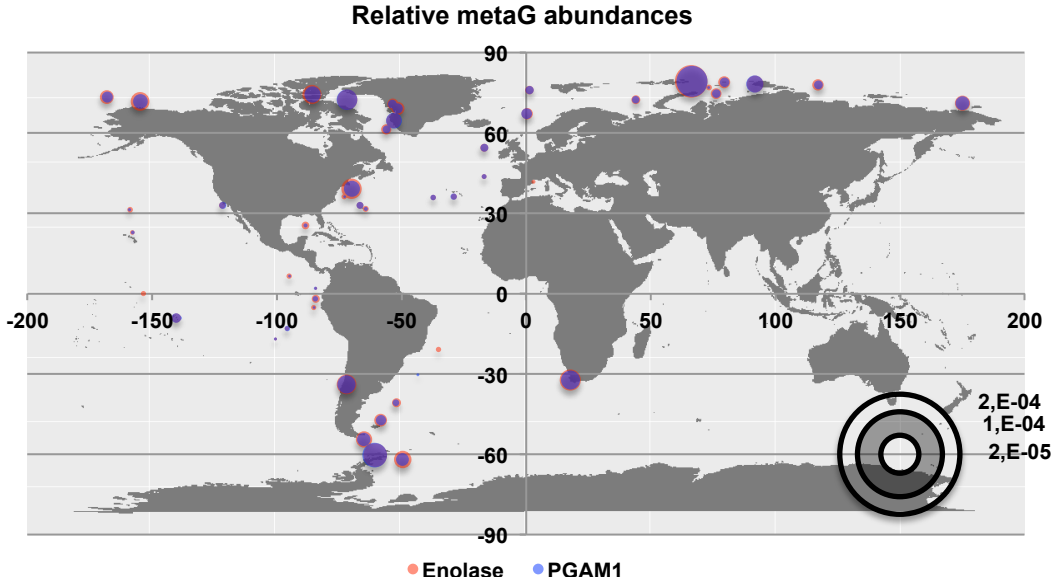
91. Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem*. **2007**;76:51-74. doi: 10.1146/annurev.biochem.76.050106.093909.
92. Broddrick JT, Du N, Smith SR, Tsuji Y, Jallet D, Ware MA, et al. Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom Phaeodactylum tricornutum. *New Phytol*. **2019**. doi: 10.1111/nph.15685.
93. Maréchal E, Lupette J. Relationship between acyl-lipid and sterol metabolisms in diatoms. *Biochimie*. **2020**;169:3-11. doi: 10.1016/j.biochi.2019.07.005.
94. Araújo S, Garcia V. Growth and biochemical composition of the diatom *Chaetoceros* cf. *wighamii* Brightwell under different temperature, salinity and carbon dioxide levels. I. Protein, carbohydrates and lipids. *Aquaculture*. **2005**;246:405-12. doi: 10.1016/j.aquaculture.2005.02.051.
95. Rousch JM, Bingham SE, Sommerfeld MR. Changes in fatty acid profiles of thermo-intolerant and thermo-tolerant marine diatoms during temperature stress. *J Exp Mar Biol Ecol*. **2003**;295(2):145-56. doi: 10.1016/S0022-0981(03)00293-4.
96. Sicko-Goad L, Andresen N. Effect of growth and light/ dark cycles on diatom lipid content and composition. *J Phycol*. **1991**;27:9. doi: 10.1111/j.0022-3646.1991.00710.x
97. Lepetit B, Goss R, Jakob T, Wilhelm C. Molecular dynamics of the diatom thylakoid membrane under different light conditions. *Photosynth Res*. **2012**;111(1):245-57. doi: 10.1007/s11120-011-9633-5.
98. Popko J, Herrfurth C, Feussner K, Ischebeck T, Iven T, Haslam R, et al. Metabolome analysis reveals betaine lipids as major source for triglyceride formation, and the accumulation of sedoheptulose during nitrogen-starvation of *Phaeodactylum tricornutum*. *PLoS One*. **2016**;11(10):e0164673. doi: 10.1371/journal.pone.0164673.
99. Dolch LJ, Maréchal E. Inventory of fatty acid desaturases in the pennate diatom *Phaeodactylum tricornutum*. *Mar Drugs*. **2015**;13(3):1317-39. doi: 10.3390/md13031317.
100. Bishop DG, Sparace SA, Mudd JB. Biosynthesis of sulfoquinovosyldiacylglycerol in higher plants: the origin of the diacylglycerol moiety. *Archive Biochem Biophys*. **1985**;240(2):851-8. doi: 10.1016/0003-9861(85)90095-5.
101. Dorrell RG, Azuma T, Nomura M, Audren de Kerdrel G, Paoli L, Yang S, et al. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc Natl Acad Sci USA*. **2019**;116(14):6914-23. doi: 10.1073/pnas.1819976116.
102. Gile GH, Moog D, Slamovits CH, Maier UG, Archibald JM. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome Biol Evol*. **2015**;7(6):1728-42. doi: 10.1093/gbe/evv095.
103. Viola R, Nyvall P, Pedersén M. The unique features of starch metabolism in red algae. *Proc Biol Sci*. **2001**;268(1474):1417-22. doi: 10.1098/rspb.2001.1644.
104. Smith SR, Dupont CL, McCarthy JK, Broddrick JT, Oborník M, Horák A, et al. Evolution and regulation of nitrogen flux through compartmentalized metabolic networks in a marine diatom. *Nat Commun*. **2019**;10(1):4552. doi: 10.1038/s41467-019-12407-y.
105. Walker JE. The ATP synthase: the understood, the uncertain and the unknown. *Biochem Soc Transact*. **2013**;41:1-16. doi: 10.1042/bst20110773.
106. Zhang K, Zhou Z, Li J, Wang J, Yu L, Lin S. SPX-related genes regulate phosphorus homeostasis in the marine phytoplankton, *Phaeodactylum tricornutum*. *Commun Biology*. **2021**;4(1):797. doi: 10.1038/s42003-021-02284-x.

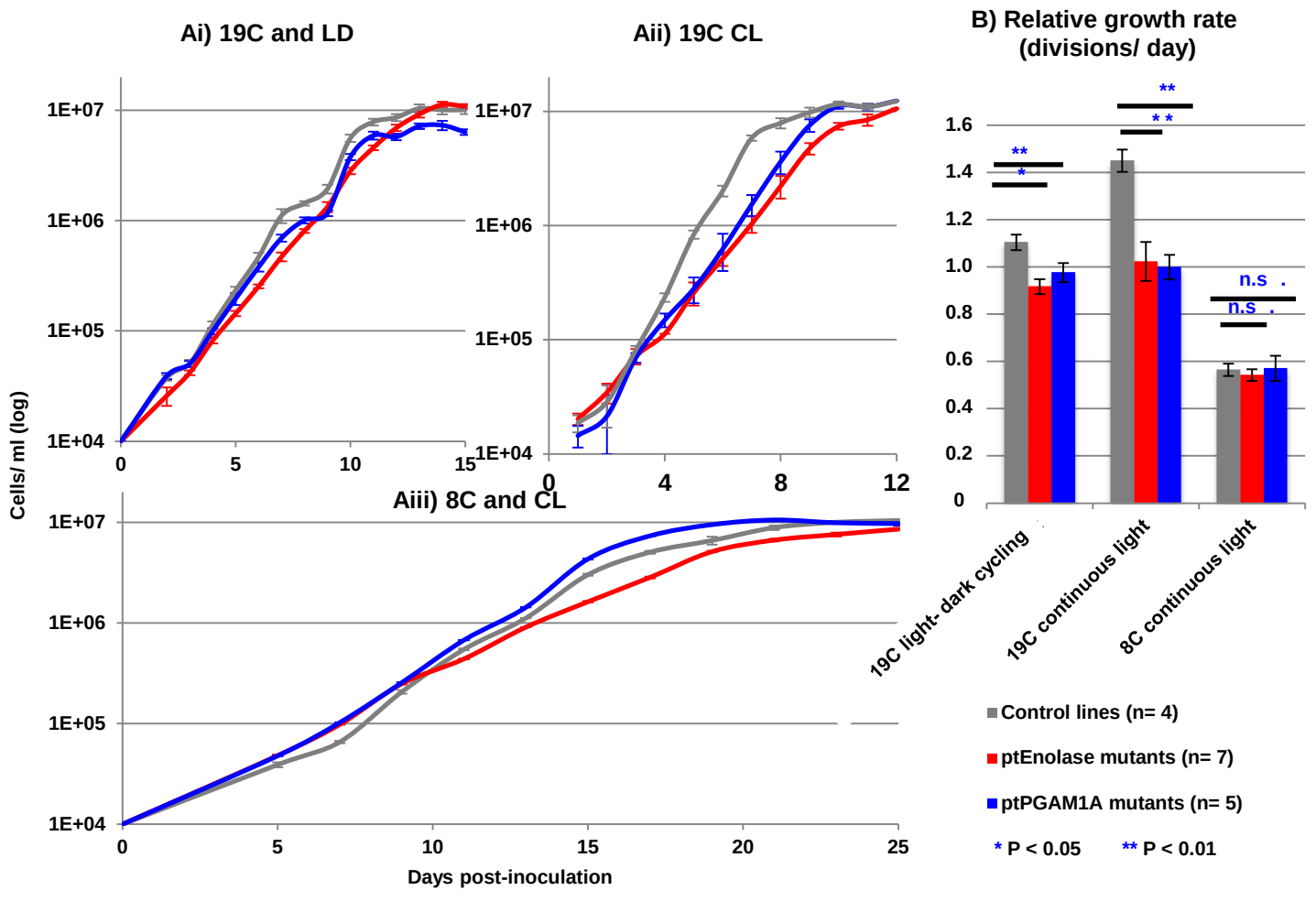
107. Matsumoto M, Nojima D, Nonoyama T, Ikeda K, Maeda Y, Yoshino T, et al. Outdoor cultivation of marine diatoms for year-round production of biofuels. *Mar Drugs*. **2017**;15(4). doi: 10.3390/md15040094.
108. Ge F, Huang W, Chen Z, Zhang C, Xiong Q, Bowler C, et al. Methylcrotonyl-CoA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum*. *Plant Cell*. **2014**;26(4):1681-97. doi: 10.1105/tpc.114.124982.
109. Yu G, Nakajima K, Gruber A, Rio Bartulos C, Schober AF, Lepetit B, et al. Mitochondrial phosphoenolpyruvate carboxylase contributes to carbon fixation in the diatom *Phaeodactylum tricornutum* at low inorganic carbon concentrations. *New Phytol*. **2022**; in press. doi: 10.1111/nph.18268.
110. Ewe D, Tachibana M, Kikutani S, Gruber A, Río Bártulos C, Konert G, et al. The intracellular distribution of inorganic carbon fixing enzymes does not support the presence of a C4 pathway in the diatom *Phaeodactylum tricornutum*. *Photosynthesis research*. **2018**;137(2):263-80. doi: 10.1007/s11120-018-0500-5.
111. Davis A, Abbriano R, Smith SR, Hildebrand M. Clarification of photorespiratory processes and the role of Malic Enzyme in diatoms. *Protist*. **2017**;168(1):134-53. doi: 10.1016/j.protis.2016.10.005.
112. Levering J, Broddrick J, Dupont CL, Peers G, Beeri K, Mayers J, et al. Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One*. **2016**;11(5). doi: 10.1371/journal.pone.0155038.
113. Dolch LJ, Lupette J, Tourcier G, Bedhomme M, Collin S, Magneschi L, et al. Nitric oxide mediates nitrite-sensing and acclimation and triggers a remodeling of lipids. *Plant Physiol*. **2017**;175(3):1407-23. doi: 10.1104/pp.17.01042.
114. Allen AE, Dupont CL, Oborník M, Horák A, Nunes-Nesi A, McCrow JP, et al. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature*. **2011**;473(7346):203-7. doi: 10.1038/nature10074.
115. Murik O, Tirichine L, Prihoda J, Thomas Y, Araújo WL, Allen AE, et al. Downregulation of mitochondrial alternative oxidase affects chloroplast function, redox status and stress response in a marine diatom. *New Phytol*. **2019**;221(3):1303-16. Edoi: 10.1111/nph.15479.
116. Liang Y, Koester JA, Liefer JD, Irwin AJ, Finkel ZV. Molecular mechanisms of temperature acclimation and adaptation in marine diatoms. *ISME J*. **2019**;13(10):2415-25. doi: 10.1038/s41396-019-0441-9.
117. Terrado R, Medrinal E, Dasilva C, Thaler M, Vincent WF, Lovejoy C. Protist community composition during spring in an Arctic flaw lead polynya. *Polar Biology*. **2011**;34(12):1901-14. doi: 10.1007/s00300-011-1039-5
118. Lovejoy C, Massana R, Pedrós-Alió C. Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl Environ Microbiol*. **2006**;72(5):3085-95. doi: 10.1128/aem.72.5.3085-3095.2006.
119. Behrenfeld MJ, Halsey KH, Boss E, Karp-Boss L, Milligan AJ, Peers G. Thoughts on the evolution and ecological niche of diatoms. *Ecological Monographs*. **2021;** 91(3):e01457. doi: 10.1002/ecm.1457.
120. Freyria NJ, Joli N, Lovejoy C. A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels. *Sci Rep*. **2021**;11(1):8413. doi: 10.1038/s41598-021-87906-4.
121. Ardyna M, Babin M, Gosselin M, Devred E, Rainville L, Tremblay J-É. Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. *Geophys Res Lett*. **2014**;41(17):6207-12. doi: 10.1002/2014GL061047.

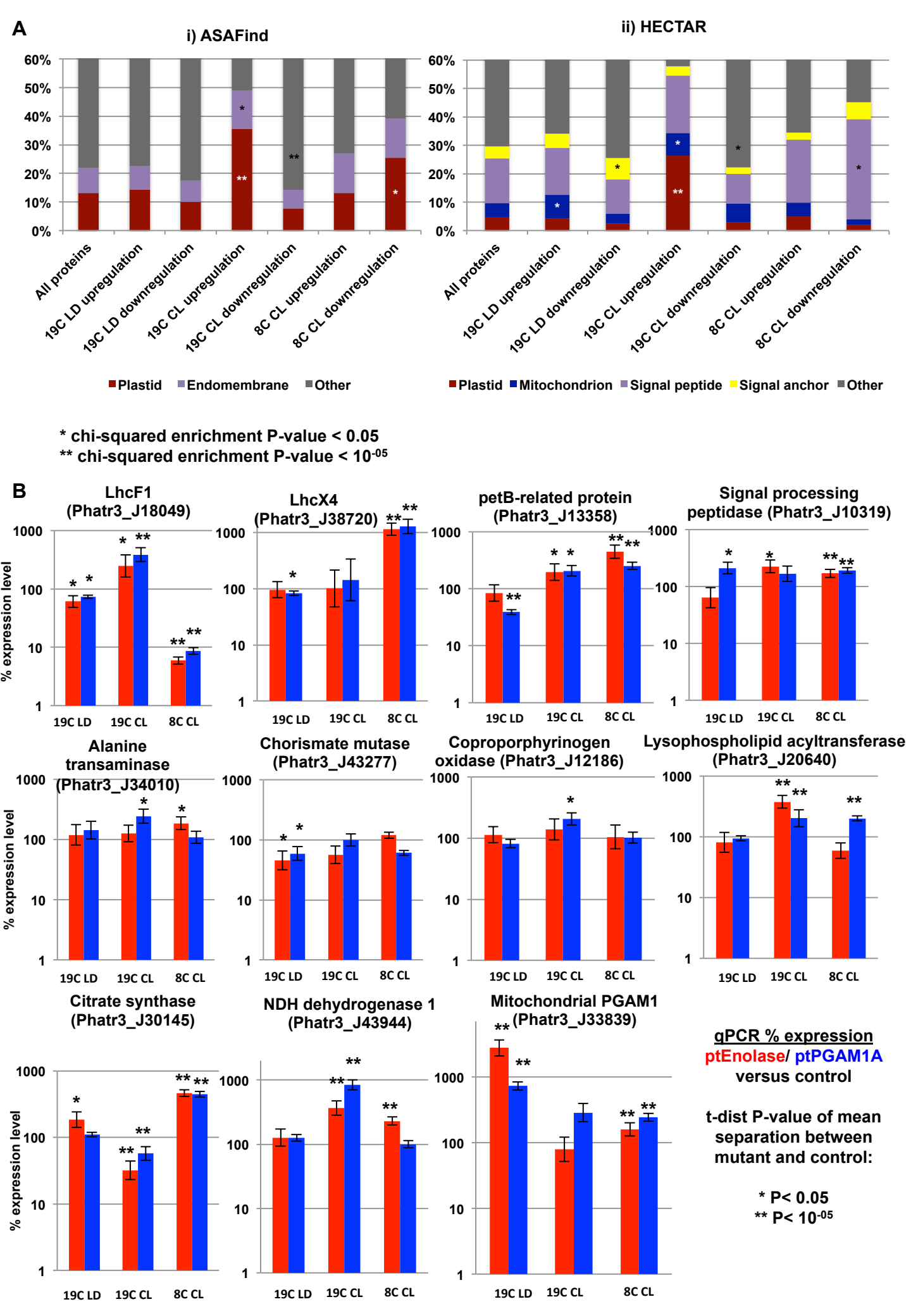
122. Lampe RH, Hernandez G, Lin YY, Marchetti A. Representative diatom and coccolithophore species exhibit divergent responses throughout simulated upwelling cycles. *mSystems*. **2021**;6(2). doi: 10.1128/mSystems.00188-21.
123. Häusler RE, Heinrichs L, Schmitz J, Flügge UI. How sugars might coordinate chloroplast and nuclear gene expression during acclimation to high light intensities. M*ol Plant*. **2014**;7(7):1121-37. doi: 10.1093/mp/ssu064.
124. Schulze-Siebert D, Heintze A, Schultz G. Substrate flow from photosynthetic carbon metabolism to chloroplast isoprenoid synthesis in spinach. Evidence for a plastidic phosphoglycerate mutase. *Zeitschrift für Naturforschung C*. **1987**;42(5):570-80. doi: 10.1515/znc-1987-0513.
125. Zhang Y, Sampathkumar A, Kerber SM-L, Swart C, Hille C, Seerangan K, et al. A moonlighting role for enzymes of glycolysis in the co-localization of mitochondria and chloroplasts. *Nat Commun.* **2020**;11(1):4509. doi: 10.1038/s41467-020-18234-w.
126. Gao F, Zhou Y, Zhu W, Li X, Fan L, Zhang G. Proteomic analysis of cold stress-responsive proteins in *Thellungiella* rosette leaves. *Planta*. **2009**;230(5):1033-46. Edoi: 10.1007/s00425-009-1003-6.
127. Amme S, Matros A, Schlesier B, Mock HP. Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. *J Exp Bot*. **2006**;57(7):1537-46. doi: 10.1093/jxb/erj129.
128. Peng X, Teng L, Yan X, Zhao M, Shen S. The cold responsive mechanism of the paper mulberry: decreased photosynthesis capacity and increased starch accumulation. *BMC Genomics*. **2015**;16(1):898. doi: 10.1186/s12864-015-2047-6.
129. Li R, Wang C, Chen T, Chen P. Quantitative proteomic analysis of cold-responsive proteins in *Abelmoschus moschatus*. *J Animal Plant Sci*. **2012**;14(3):18.
130. Vítámvás P, Prášil IT, Kosová K, Planchon S, Renaut J. Analysis of proteome and frost tolerance in chromosome 5A and 5B reciprocal substitution lines between two winter wheats during long-term cold acclimation. *Proteomics*. **2012**;12(1):68-85. doi: 10.1002/pmic.201000779.
131. Hlaváčková I, Vítámvás P, Santrůček J, Kosová K, Zelenková S, Prášil IT, et al. Proteins involved in distinct phases of cold hardening process in frost resistant winter barley (*Hordeum vulgare* L.) cv Luxor. *Int J Mol Sci*. **2013**;14(4):8000-24. doi: 10.3390/ijms14048000.

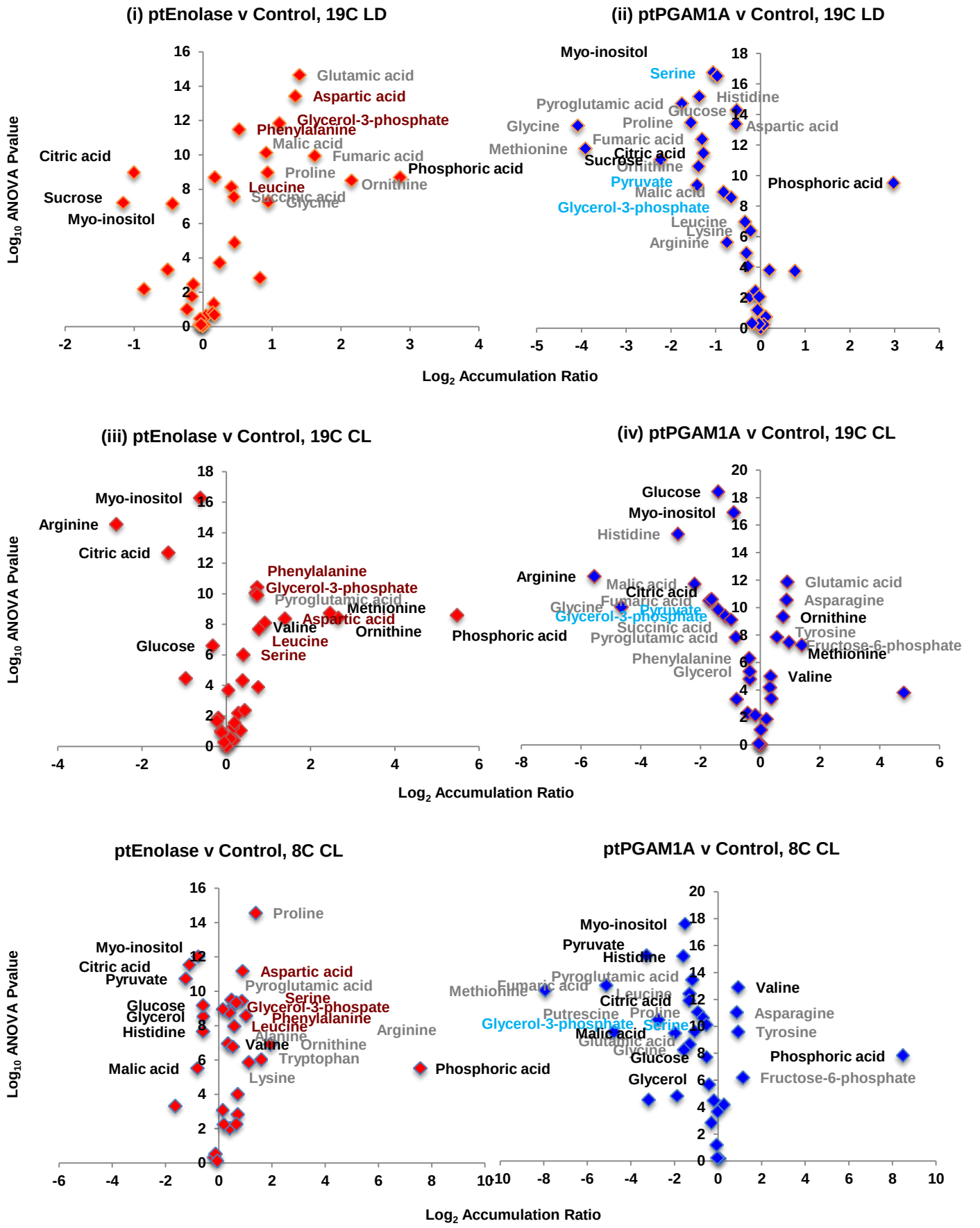

A
Diatoms
Glycolysis upper half
Mitochondria
Chrysolaminarin
Glycolysis lower half
Pyruvate
Glucose -6-P
GA3P/ DHAP
GA3P/ DHAP
PGAM
Enolase
Respiration
cER
Chloroplast
RuBP
Calvin cycle
GA3P/ DHAP
1,3-PGA
Photo-synthesis light reactions
ptPGAM
2,3-PGA
ptEnolase
PEP
Asp/ Glu shuttle
Nucleus
Pyruvate hub (lipids, amino acids)
Plants
Chloroplast
Complex sugars
RuBP
Glycolysis upper half
GA3P/ DHAP
Calvin cycle
1,3-PGA
GA3P/ DHAP
Starch
1,3-PGA
ptPGAM, ptEnolase (non-photosynthetic tissue only)
PGAM
2,3-PGA
Enolase
PEP
Photo-synthesis light reactions
Pyruvate hub (lipids, amino acids)
Pyruvate
Pyruvate
Nucleus
ER
Respiration
Mitochondria
Key
Metabolic pathway
Metabolite
PGAM
Enolase
ATP
◆ production/
◇consumption
NADP(H)
★ production/
☆consumption
B- Enolase
0.98/1/0.62
1/1/1
Diatom mito Phatr3_draftJ1192; draftJ1572
Diatom plastid Phatr3_J41515
0.1
C-PGAM
0.98/0.89/0.87
Diatom plastid B Phatr3_J51404
Diatom mito Phatr3_J33839
1/1/1
Diatom plastid A Phatr3_J17086
0.2
D
Diatoms
Bolidophytes, Hypogyristea
Chrysista
Aplastidic SAR clade members
Green algae
Red algae
Other eukaryotes
Bacteria
Archaea
Mitochondria-targeted
Plastid-targeted
Dual-targeted
A/B/C: MrBayes PP- GTR/Jones/WAG
ptEnolase: GFP
ptPGAM1A: GFP

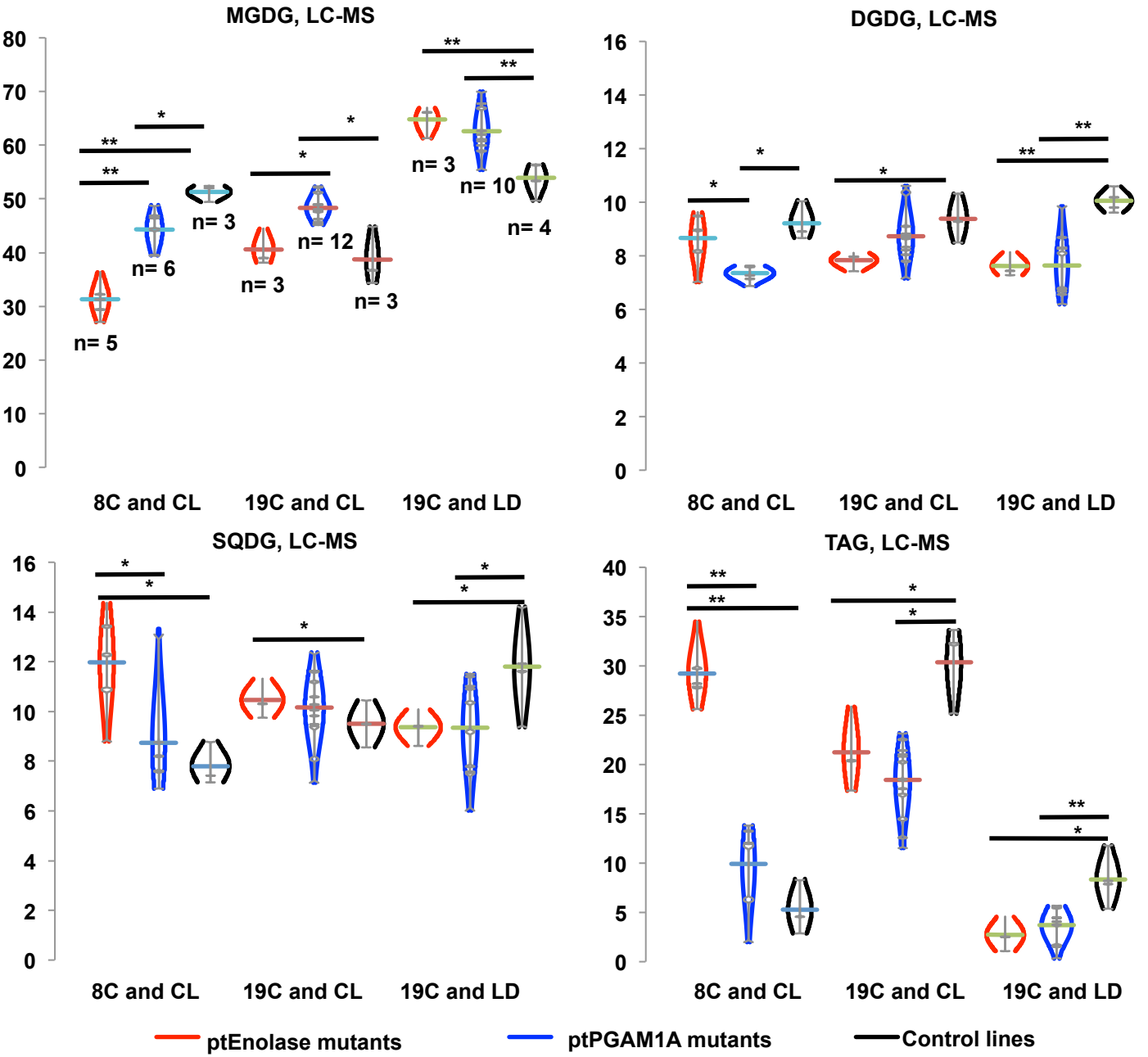

Relative metaT abundances
90
60
30
0
-30
-60
-90
-200
-150
-100
-50
0
50
100
150
200
2,E-04
1,E-04
2,E-05


Relative metaG abundances
90
60
30
0
-30
-60
-90
-200
-150
-100
-50
0
50
100
150
200
2,E-04
1,E-04
2,E-05
Enolase
PGAM1


Ai) 19C and LD
Aii) 19C CL
Aiii) 8C and CL
B) Relative growth rate (divisions/ day)
Cells/ ml (log)
1E+07
1E+06
1E+05
1E+04
0
5
10
15
4
8
12
20
25
Days post-inoculation
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
**
*
n.s
19C light- dark cycling
19C continuous light
8C continuous light
Control lines (n= 4)
ptEnolase mutants (n= 7)
ptPGAM1A mutants (n= 5)
* P < 0.05
** P < 0.01


A
i) ASAFind
ii) HECTAR
60%
50%
40%
30%
20%
10%
0%
All proteins
19C LD upregulation
19C LD downregulation
19C CL upregulation
19C CL downregulation
8C CL upregulation
8C CL downregulation
Plastid
Endomembrane
Other
Mitochondrion
Signal peptide
Signal anchor
* chi-squared enrichment P-value < 0.05
** chi-squared enrichment P-value < 10-05
B
LhcF1 (Phatr3_J18049)
LhcX4 (Phatr3_J38720)
petB-related protein (Phatr3_J13358)
Signal processing peptidase (Phatr3_J10319)
Alanine transaminase (Phatr3_J34010)
Chorismate mutase (Phatr3_J43277)
Coproporphyrinogen oxidase (Phatr3_J12186)
Lysophospholipid acyltransferase (Phatr3_J20640)
Citrate synthase (Phatr3_J30145)
NDH dehydrogenase 1 (Phatr3_J43944)
Mitochondrial PGAM1 (Phatr3_J33839)
% expression level
1000
100
10
1
19C LD
19C CL
8C CL
qPCR % expression
ptEnolase/ ptPGAM1A
versus control
t-dist P-value of mean separation between mutant and control:
* P< 0.05
** P< 10-05


(i) ptEnolase v Control, 19C LD
Log10 ANOVA Pvalue
Log2 Accumulation Ratio
Glutamic acid
Aspartic acid
Glycerol-3-phosphate
Phenylalanine
Malic acid
Fumaric acid
Proline
Phosphoric acid
Ornithine
Leucine
Succinic acid
Glycine
Citric acid
Sucrose
Myo-inositol
(ii) ptPGAM1A v Control, 19C LD
Myo-inositol
Serine
Histidine
Pyroglutamic acid
Glucose
Glycine
Proline
Aspartic acid
Fumaric acid
Methionine
Citric acid
Sucrose
Ornithine
Pyruvate
Malic acid
Phosphoric acid
Glycerol-3-phosphate
Leucine
Lysine
Arginine
(iii) ptEnolase v Control, 19C CL
Myo-inositol
Arginine
Citric acid
Phenylalanine
Glycerol-3-phosphate
Pyroglutamic acid
Methionine
Aspartic acid
Valine
Ornithine
Leucine
Glucose
Serine
Phosphoric acid
(iv) ptPGAM1A v Control, 19C CL
Glucose
Myo-inositol
Histidine
Arginine
Malic acid
Citric acid
Fumaric acid
Glycine
Pyruvate
Glycerol-3-phosphate
Succinic acid
Pyroglutamic acid
Phenylalanine
Glycerol
Glutamic acid
Asparagine
Ornithine
Tyrosine
Fructose-6-phosphate
Methionine
Valine
ptEnolase v Control, 8C CL
Proline
Myo-inositol
Citric acid
Pyruvate
Aspartic acid
Pyroglutamic acid
Serine
Glucose
Glycerol-3-phospate
Glycerol
Phenylalanine
Histidine
Leucine
Alanine
Arginine
Valine
Ornithine
Tryptophan
Malic acid
Lysine
Phosphoric acid
ptPGAM1A v Control, 8C CL
Myo-inositol
Pyruvate
Histidine
Pyroglutamic acid
Methionine
Fumaric acid
Leucine
Citric acid
Putrescine
Proline
Glycerol-3-phosphate
Serine
Malic acid
Glutamic acid
Glycine
Glucose
Glycerol
Valine
Asparagine
Tyrosine
Phosphoric acid
Fructose-6-phosphate


MGDG, LC-MS
DGDG, LC-MS
SQDG, LC-MS
TAG, LC-MS
8C and CL
19C and CL
19C and LD
n= 5
n= 6
n= 3
n= 3
n= 12
n= 3
n= 3
n= 10
n= 4
ptEnolase mutants
ptPGAM1A mutants
Control lines
* ANOVA P < 0.05
** ANOVA P < 10-05

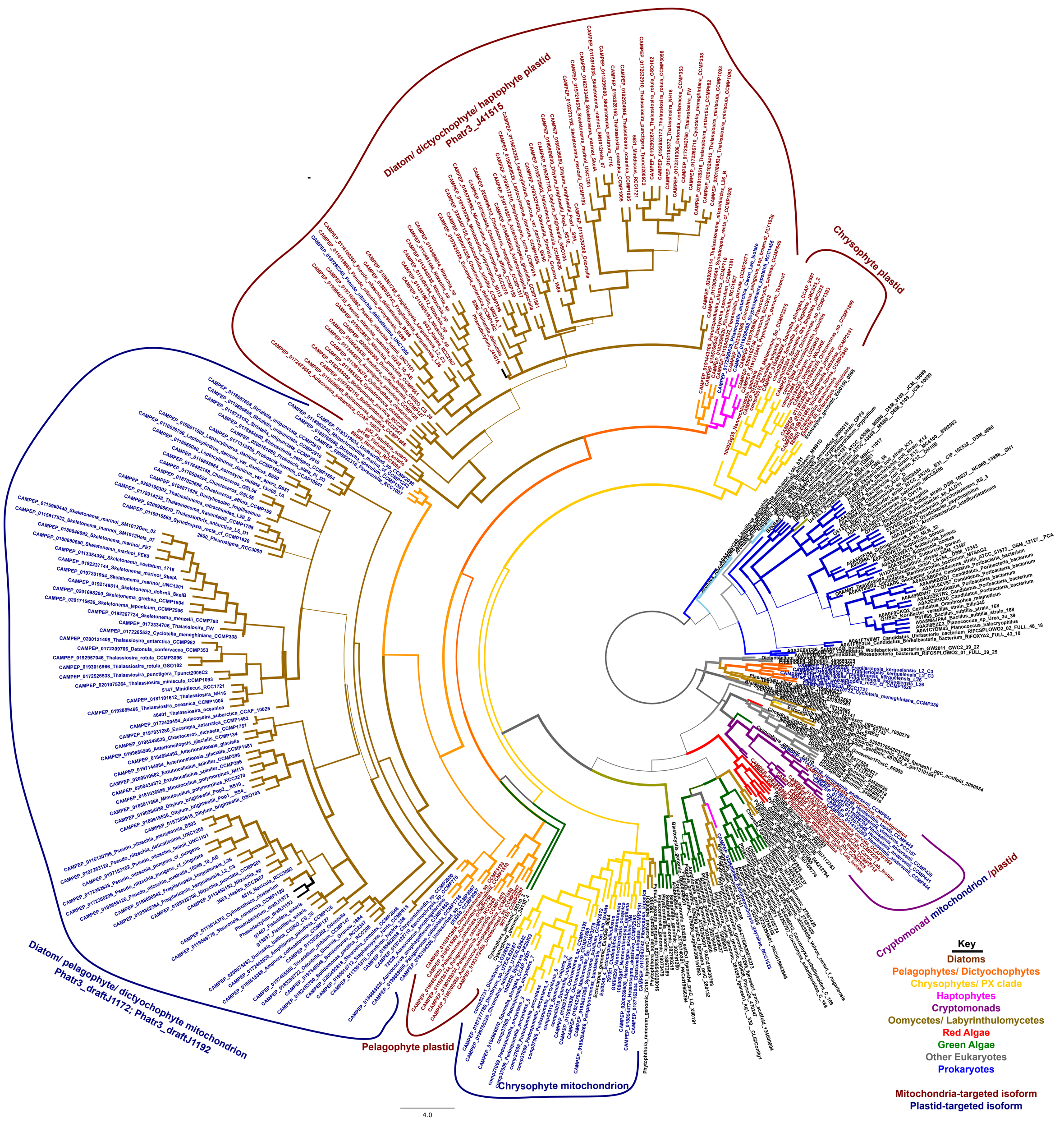


Fig. S1. Consensus topology of a 380 taxa x 413 aa alignment of Enolase sequences, sampling all organelle-targeted isoforms from cryptomonads, haptophytes and ochrophytes and representatives from a densely-sampled dataset of 151 taxonomic groups $across the tree of life (Dorrell et al., 2021). The tree 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 topology recovery in the consensus of consenus trees; branches are coloured by taxonomic affiliation; and tips (cryptomonads, haptophytes and ochrophytes only) by predicted *in silico* localisation. Individual clades (considering both taxonomic origins and inferred localisation) of organelle-targeted enolase isoform are labelled with coloured brackets.

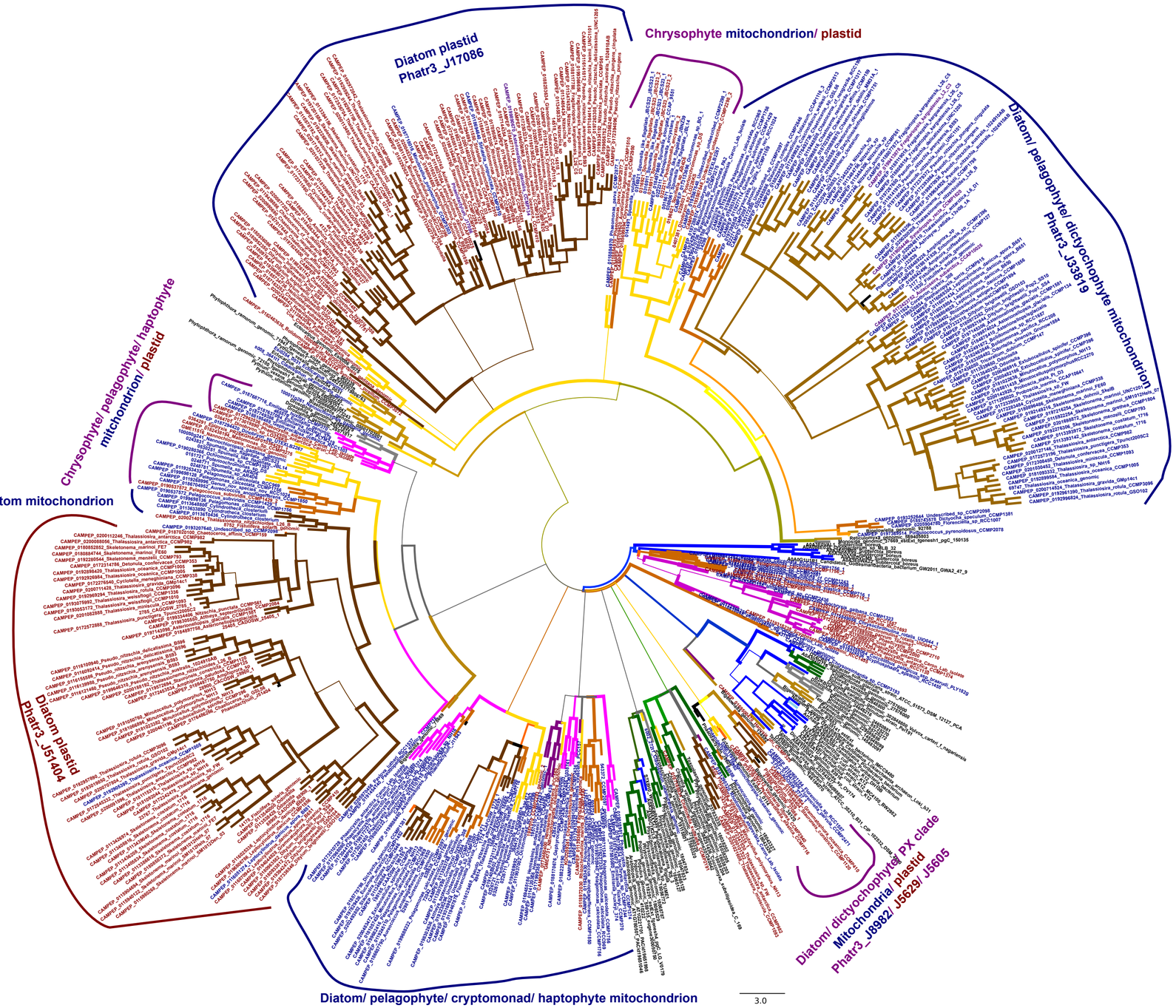


Fig. S2. Consensus topology of a 220 aa x 560 taxa alignment of PGAM isoform 1 sequences, shown as per Fig. **S1.**

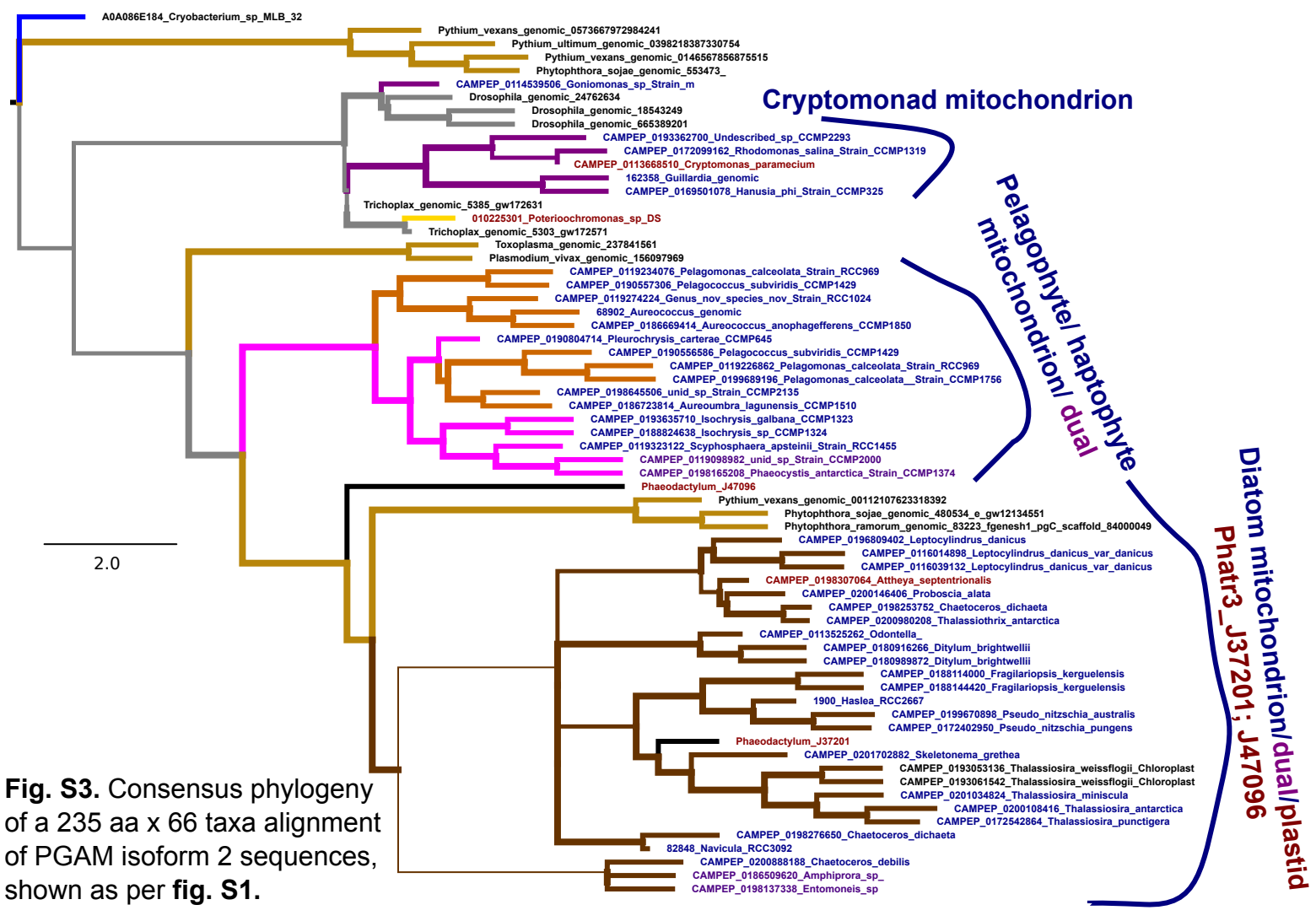


Fig. S3. Consensus phylogeny of a 235 aa x 66 taxa alignment of PGAM isoform 2 sequences, shown as per **fig. S1.**

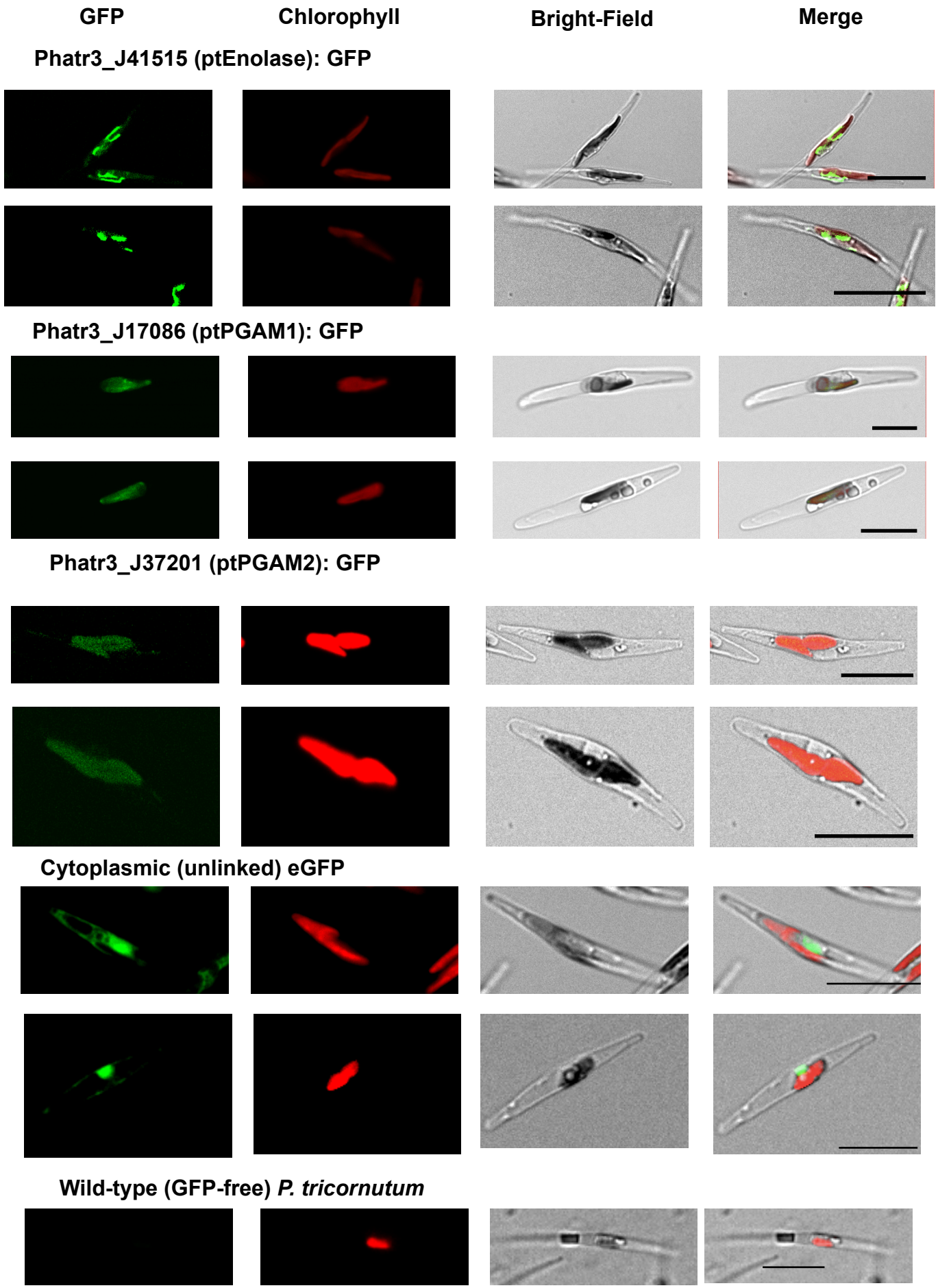


Fig. S4. Individual fluorescence channel and control confocal microscopy images for *P. tricornutum* plastis glycolysis proteins.

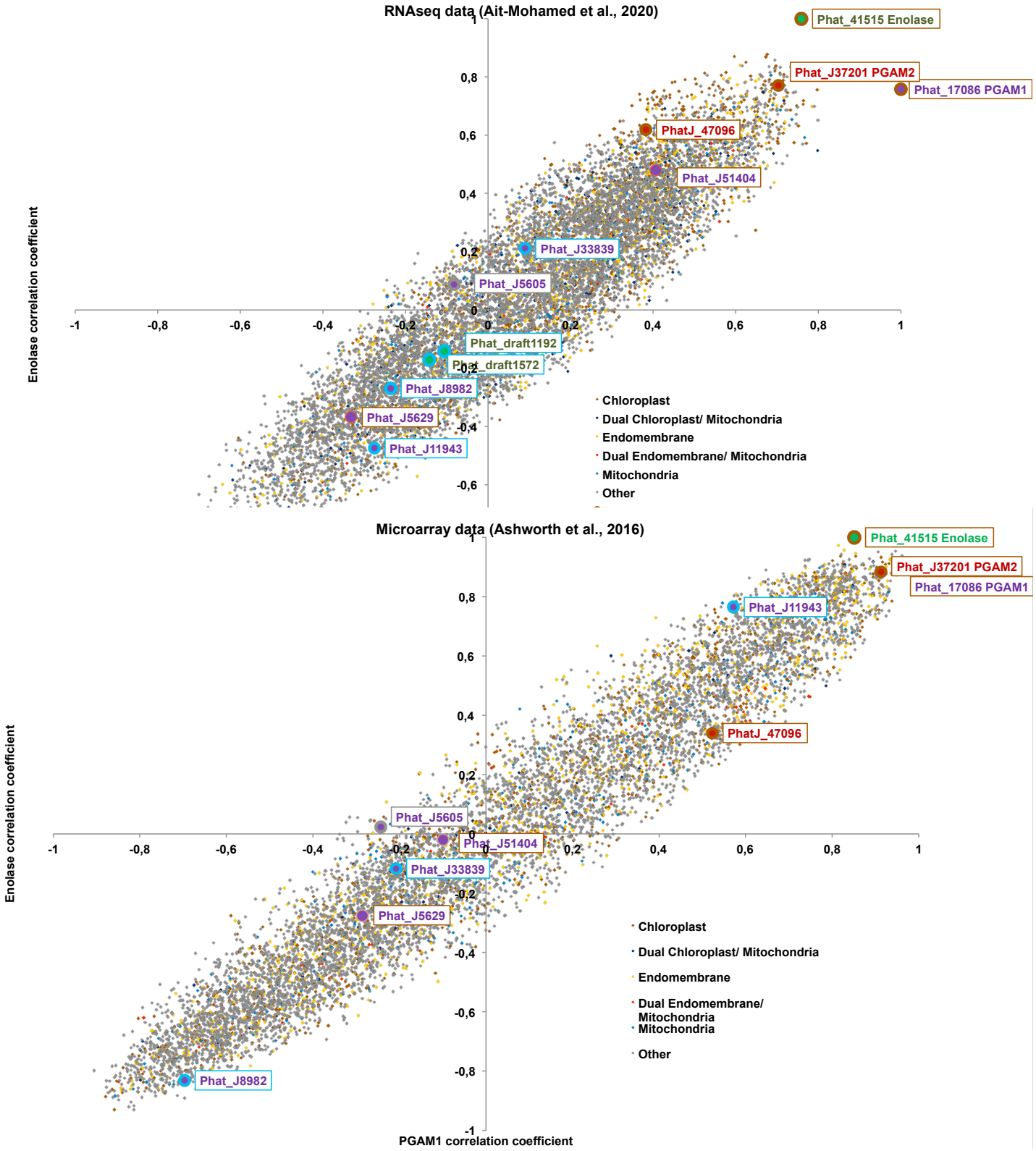
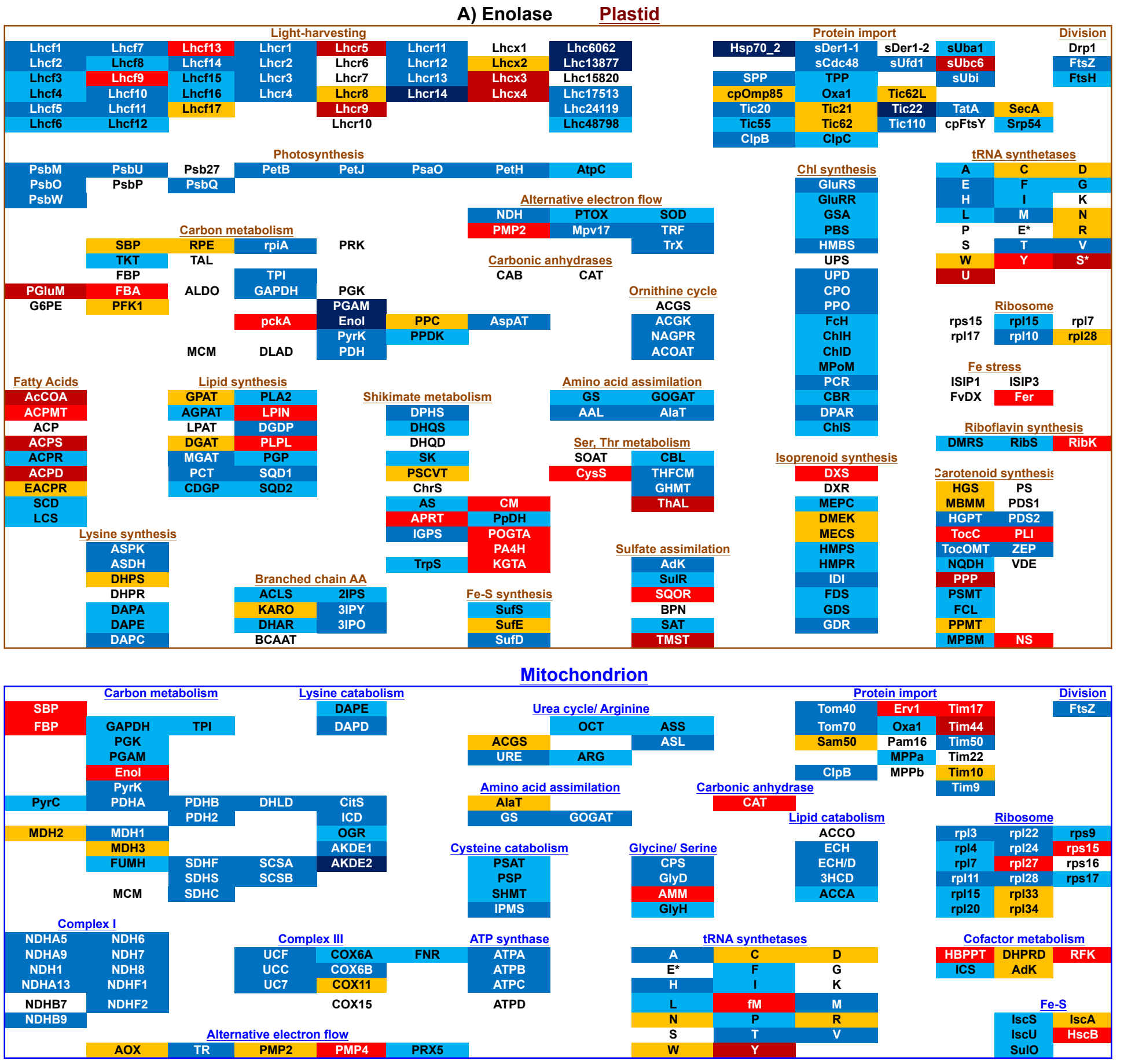


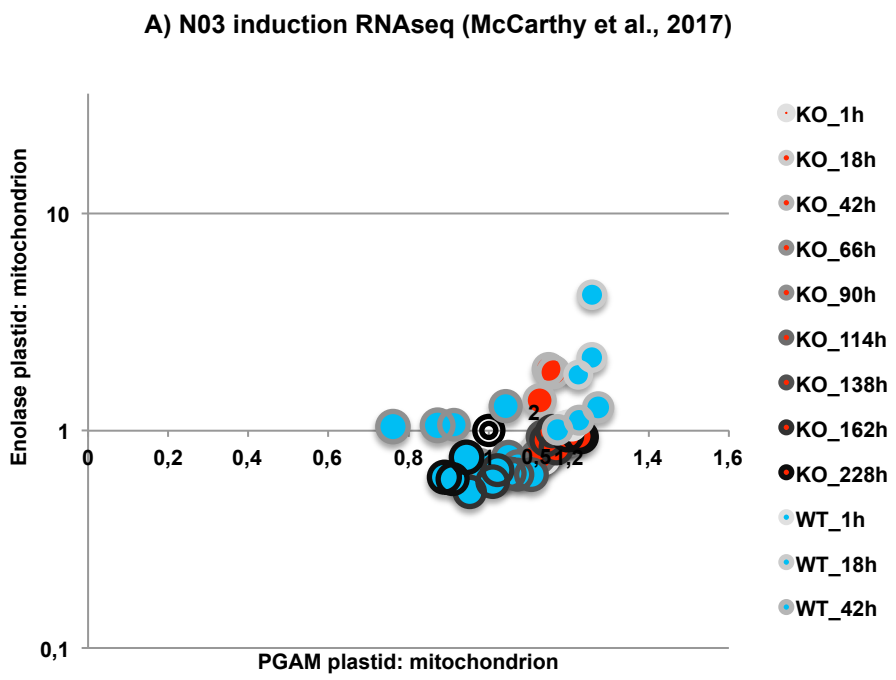
Fig. S5. Transcriptional coordination of plastidial enolase, PGAM1 and PGAM2 from published gene expression data. This figure shows Spearman coefficients of all genes across the version 3 annotation of the *Phaeodactylum tricornutum* genome against two query genes: Phatr3_J41515, plastid-targeted enolase; and Phatr3_J17086, plastid-targeted PGAM1; calculated from ranked mRNA abundances within normalised RNAseq (PhaeoNet, Ait-Mohamed et al., 2020) and microarray (DiatomPortal, Ashworth et al., 2016) meta-datasets. Genes are coloured by inferred in silico localisation; and other annotated enolase and PGAM enzymes in the *Phaeodactylum* genome are shown as large, labelled points.

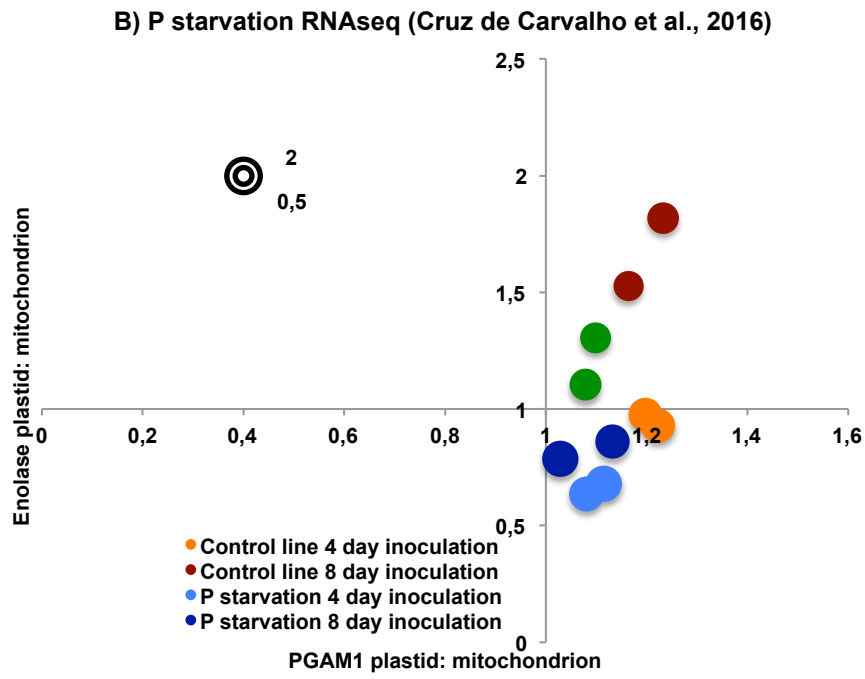


Scale

>0.8
0.6-0,6
0.4-0,6
0.2-0,4
0-0,2
< -0.2

Fig. S6. Transcriptional coregulation of plastid-targeted Enolase and PGAM1 to other core organelle metabolism pathways. These heatmaps show the Spearman correlation coefficients calculated between ptEnolase and ptPGAM1 against genes encoding other core *Phaeodactylum* plastid- and mitochondrial-targeted proteins, as annotated in Ait-Mohamed et al. 2020, considering a merged and ranked RNAseq and microarray dataset. Where more than one homologue of a particular enzyme exist, the highest correlation coefficient is shown. Complete gene lists, including abbreviations, are provided in Table S2.





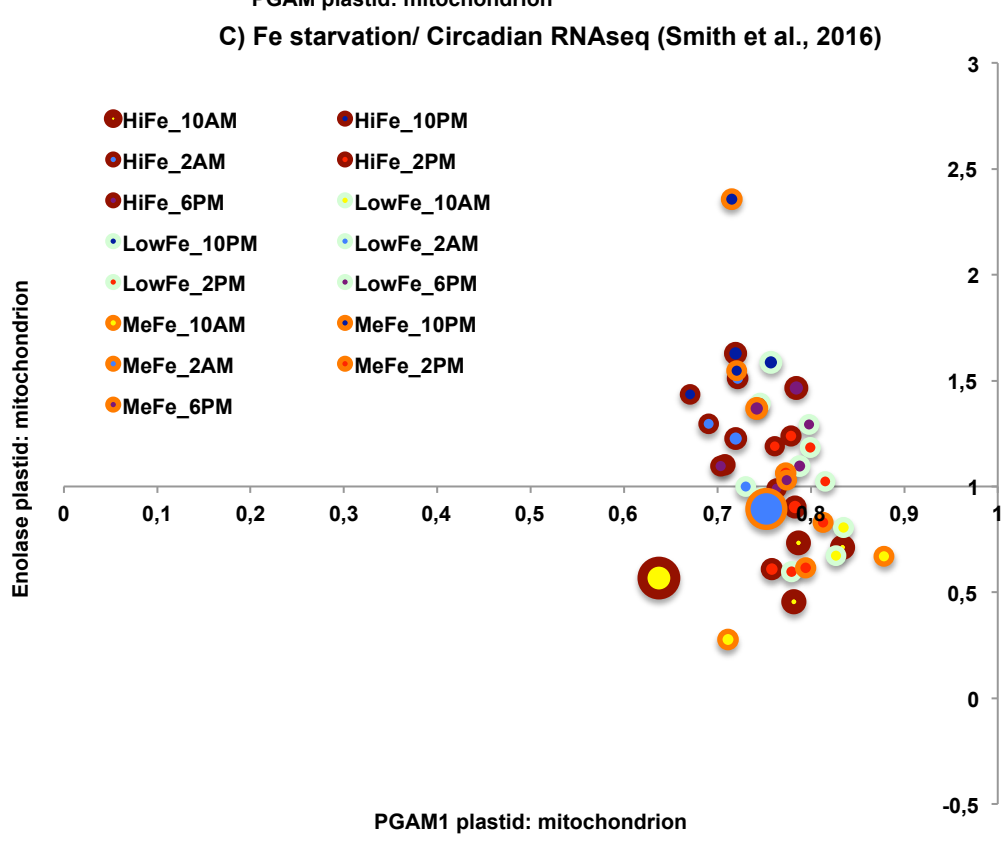
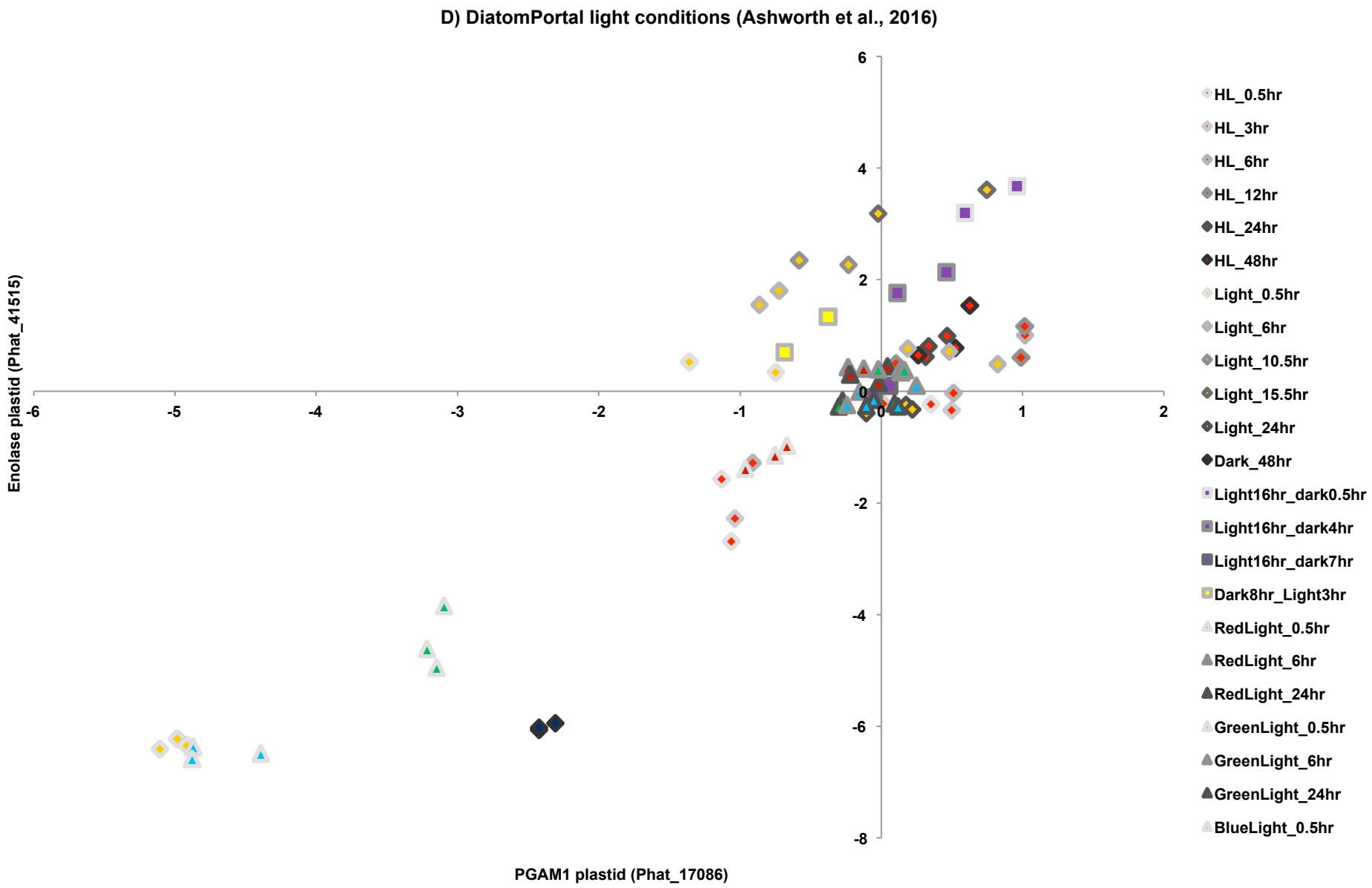


Fig. S8. Relative transcriptional regulation of plastidial enolase and PGAM under nutrient and light stress conditions. A-C: the relative ratio of plastid to mitochondria-targeted gene isoform expression for enolase (vertical axis), PGAM1 (horizontal axis), and PGAM2 (bubble size) in published *Phaeodactylum* RNAseq datasets subject to nitrate limitation (McCarthy et al., 2017), phosphate starvation and addition (Cruz de Carvalho et al., 2016) and Circadian Fe enrichment and limitation (Smith et al., 2016), respectively. While limited transcriptional responses are identifiable in response to changing N, P or Fe availability, the relative transcription of plastidial to mitchondrial-targeted enolase is substantially greater in RNA sampled in long day (> 12h post-illumination) conditions in **C.** D: relative fold changes in plastidial enolase and PGAM1 expression in published microarray data assembled in Ashworth et al., 2016, under different illumination conditions. Both Enolase and PGAM1 show substantial downregulation in dark-incubated (48h) and short post-illumination-incubated (< 0.5h) cultures, and show the greatest positive fold expression changes respectively > 16h and > 12h post-illumination.



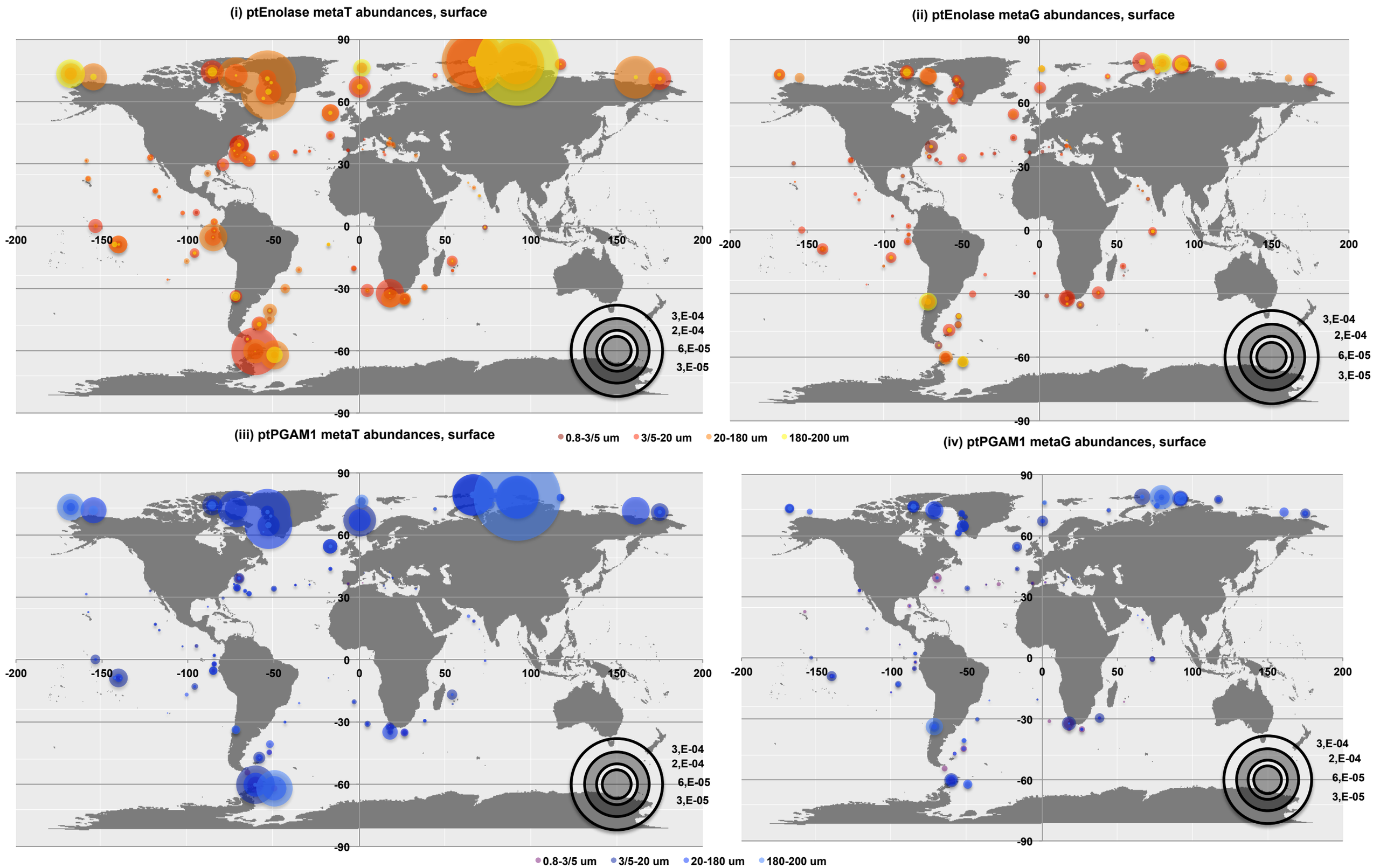


Fig. S8. Relative abundances of Tara Oceans meta-genes that resolve phylogenetically with (top) diatom plastidial enolase and (bottom) diatom plastidial PGAM1 sequences over all size fractions of surface meta-transcriptome (left) and meta-genome (right) data, shown as per Fig. 2.

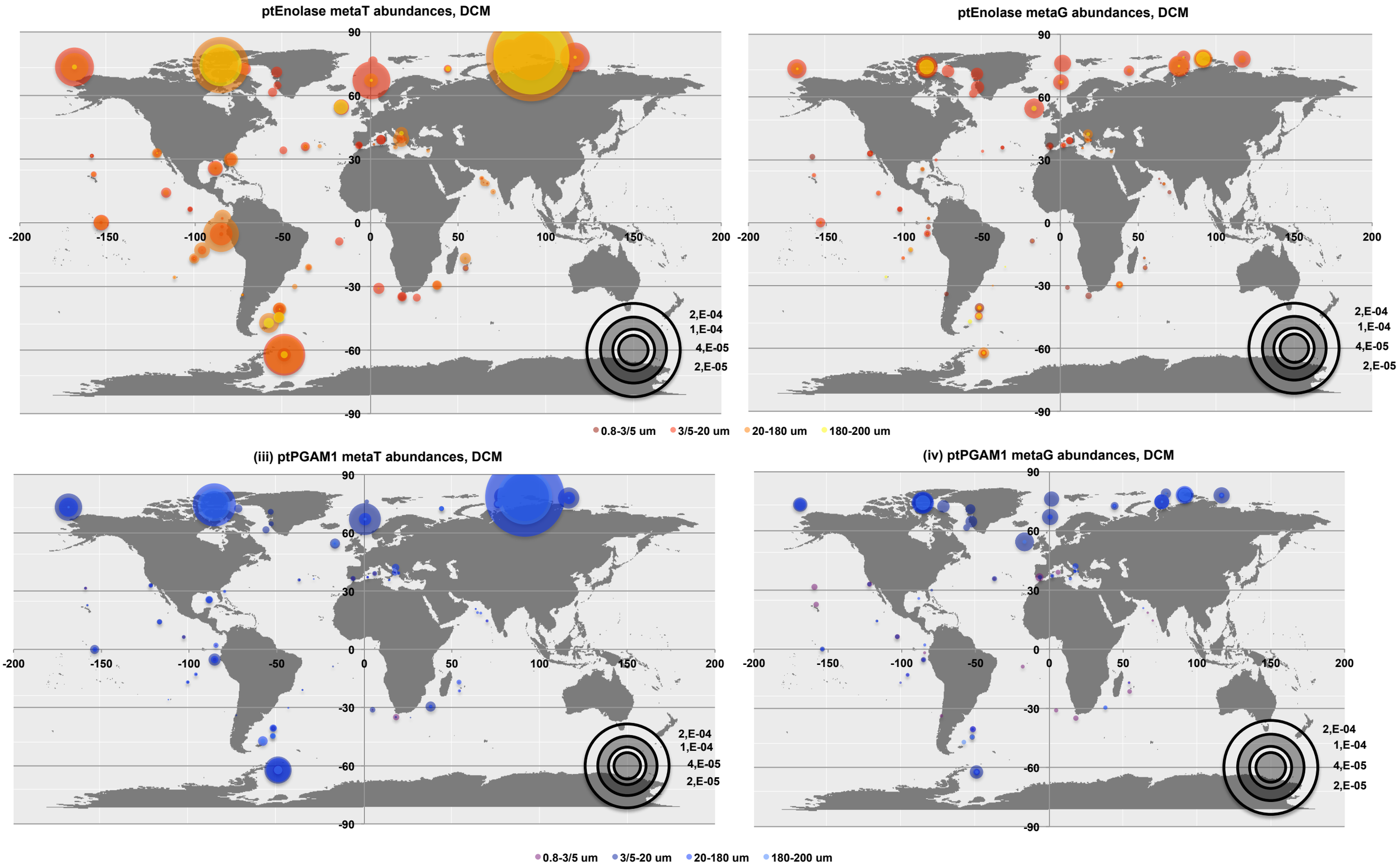


Fig. S9. Relative abundances of Tara Oceans meta-genes that resolve phylogenetically with (top) diatom plastidial enolase and (bottom) diatom plastidial PGAM1 sequences over all size fractions of DCM meta-transcriptome (left) and meta-genome (right) data, shown as per Fig. 2.

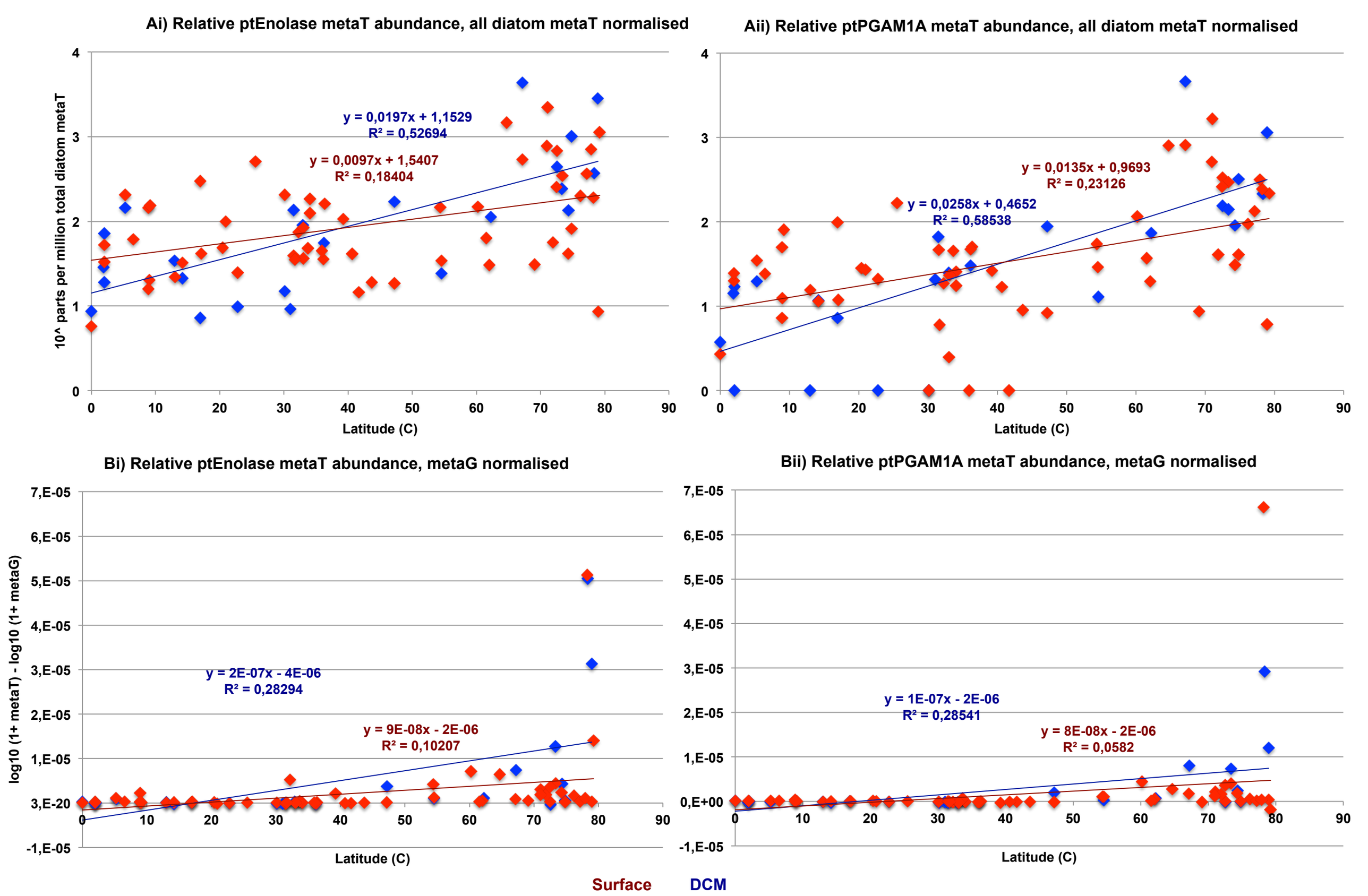


Fig. S10. Scatterplots of *Tara* Ocean expression patterns of sequences assigned phylogenetically to diatom ptEnolase and ptPGAM1A against station latitude. Abundances are shown for 0.8-2000 µm surface and DCM sample meta-transcriptome data, and are normalised relative to **(A)** total diatom metaT abundances at each station and **(B)** the corresponding metaG abundances for diatom ptEnolase and ptPGAM1A. In each case, a significant positive correlation between latitude and relative expression is observed.

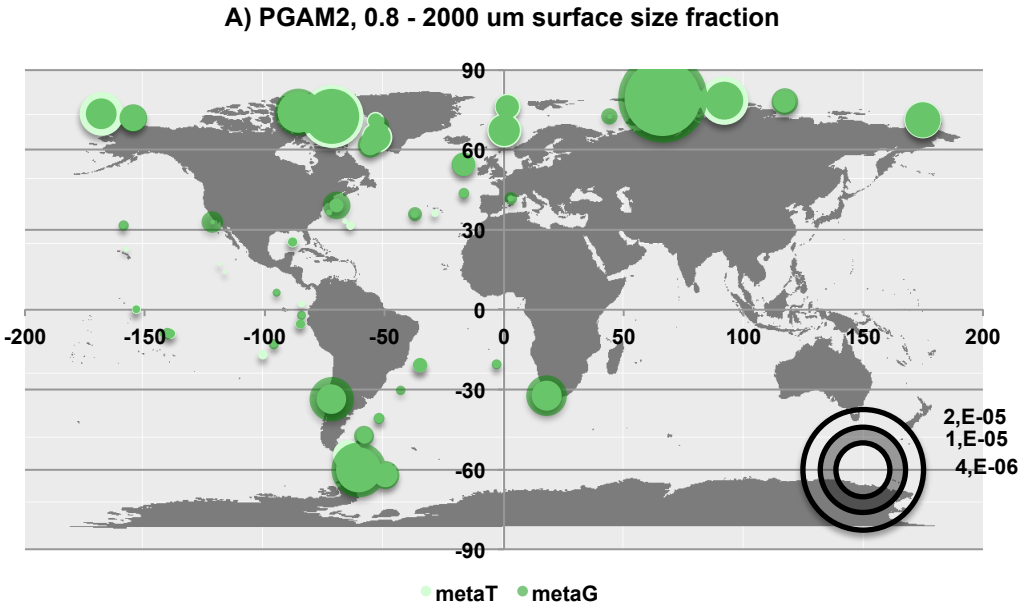
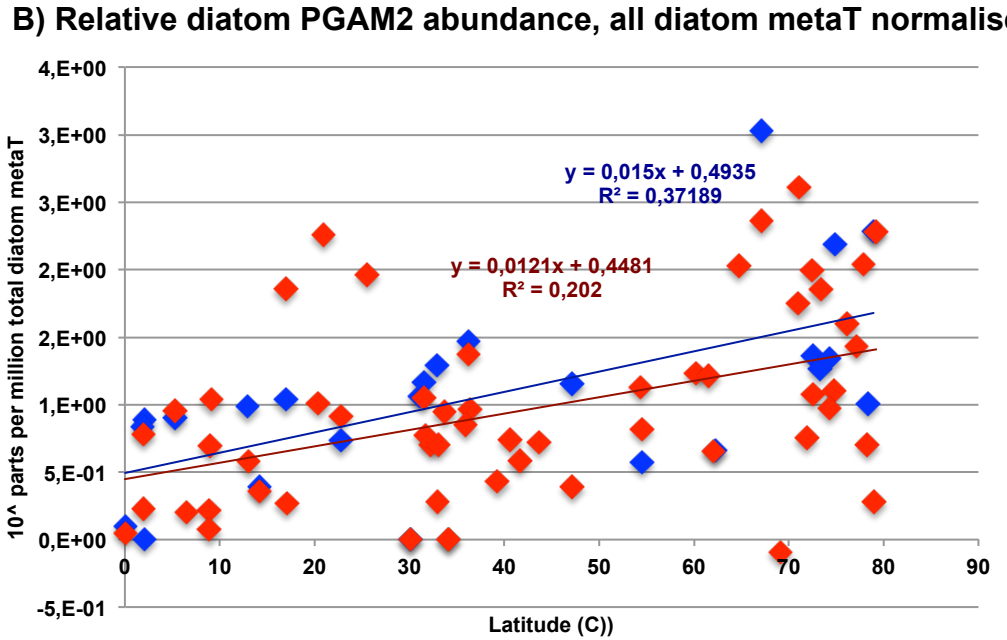
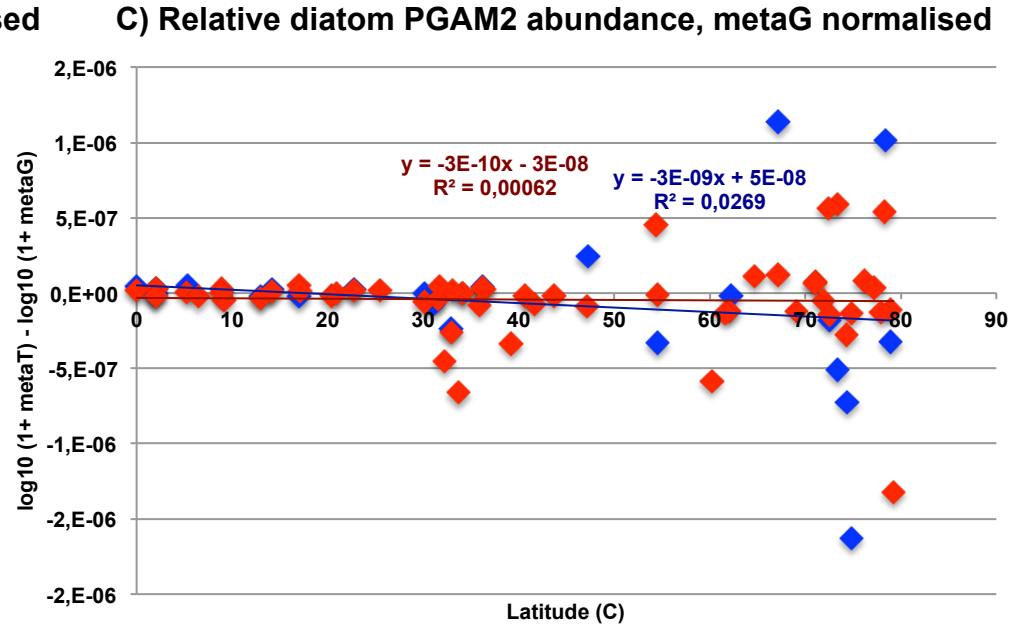


Fig. S11. *Tara* Ocean expression patterns of diatom PGAM2. **A**: total relative abundances of meta-genes phylogenetically reconciled to diatom PGAM2 in 0.8-2000 µm surface sample meta-transcriptome and meta-genome data, showing effective congruence between both. **B**, **C:** scatterplot as per Fig. S10 of phylogenetically reconciled diatom PGAM2 metaT abundances, normalised against all diatom metaT abundances and against PGAM2 metaG abundances respectively, for both surface and DCM stations in the 0.8-2000 µm size fraction, showing effectively no correlation between metaT abundance and latitude.





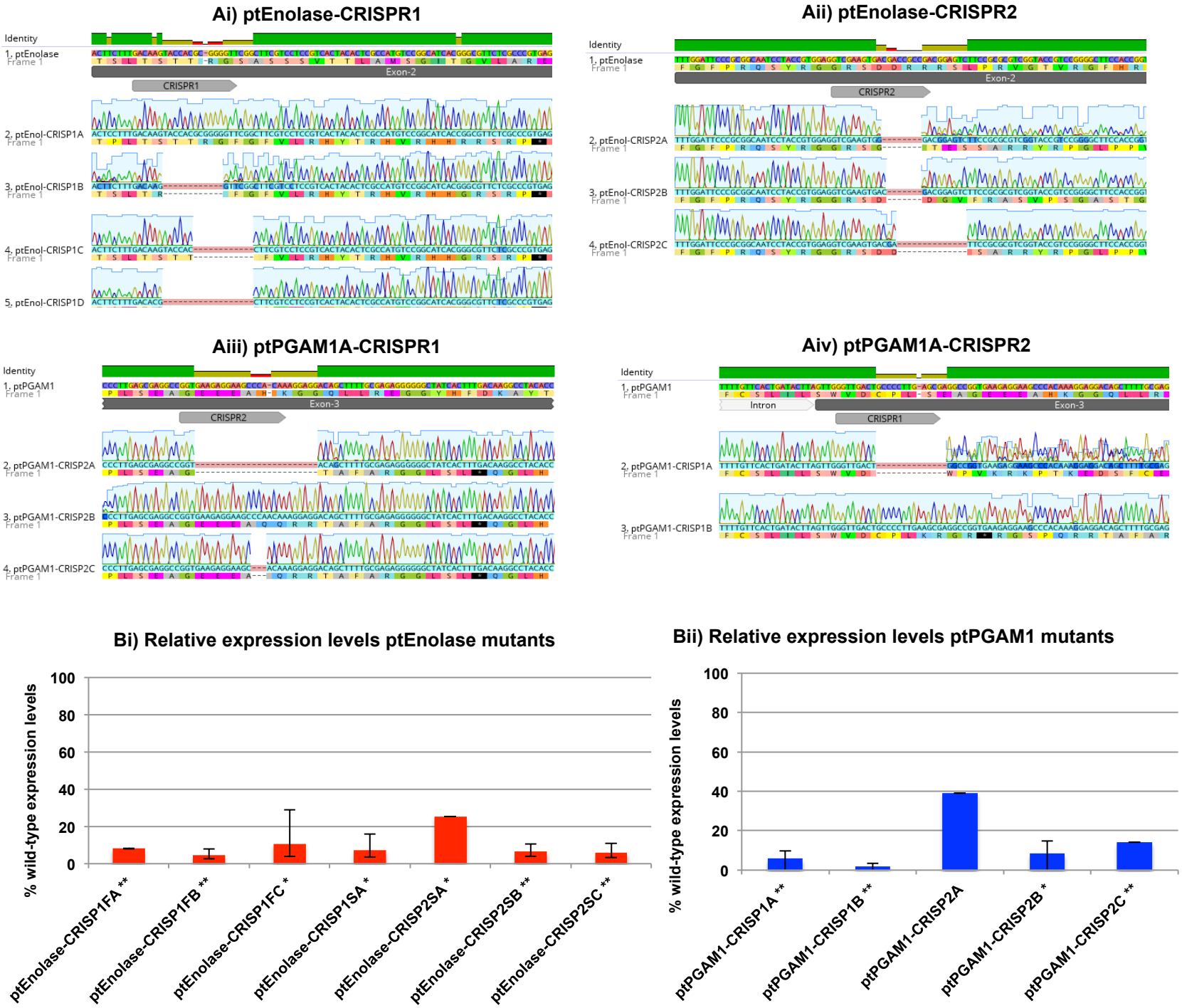
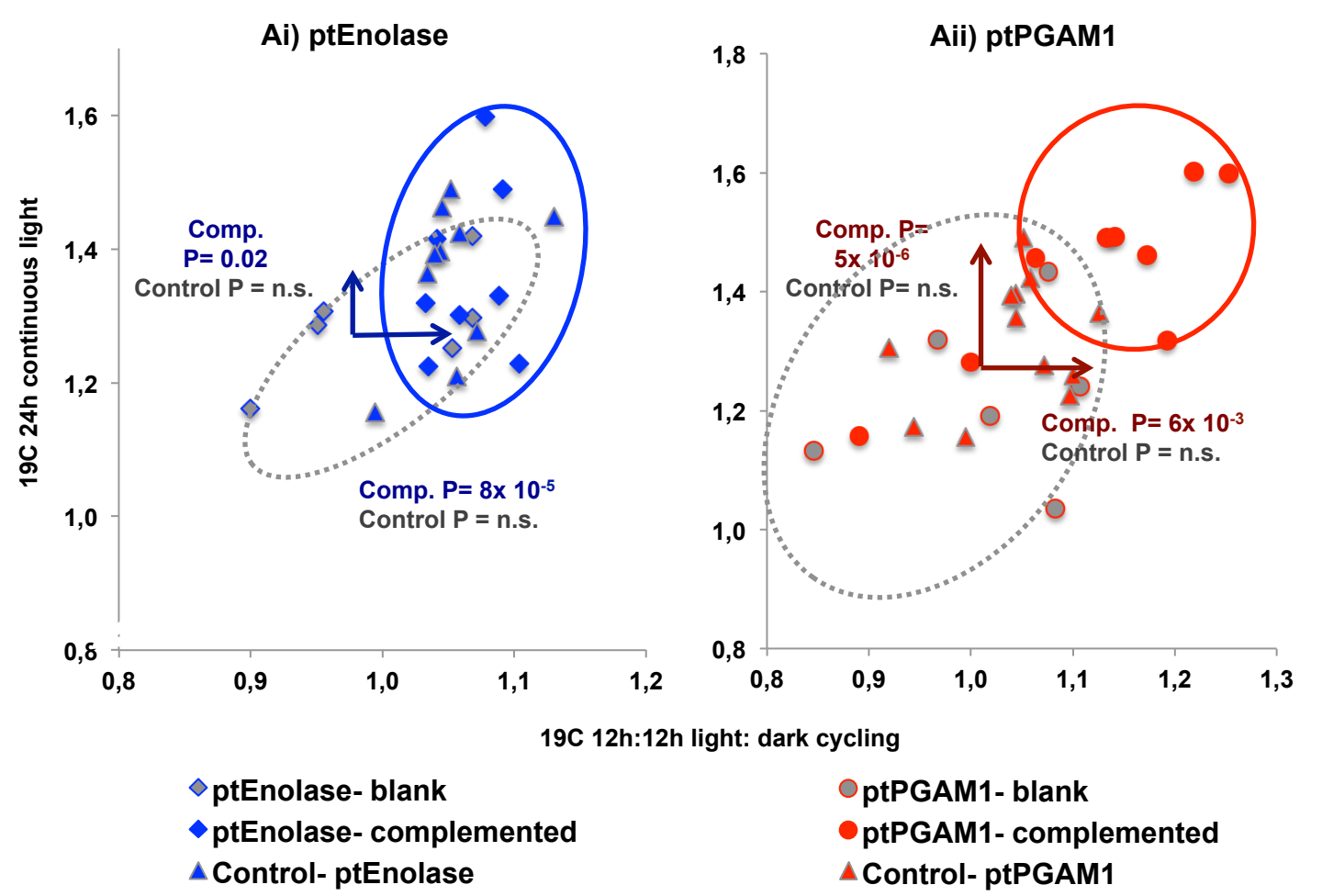


Fig. S12. Genotypes of plastid glycolysis gene mutants. A: alignments of the two CRISPR regions targeted for mutagenesis of ptEnolase (Phatr3_J41515) and ptPGAM1A (Phatr3_J17086), and the genotypes obtained from Sanger sequences of homozygous CRISPR mutants obtained for each gene. **B:** average relative 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 mutant line compared to the empty vector controls are provided.

*** Significant to P < 0.05**
**** Significant to P < 0.01**



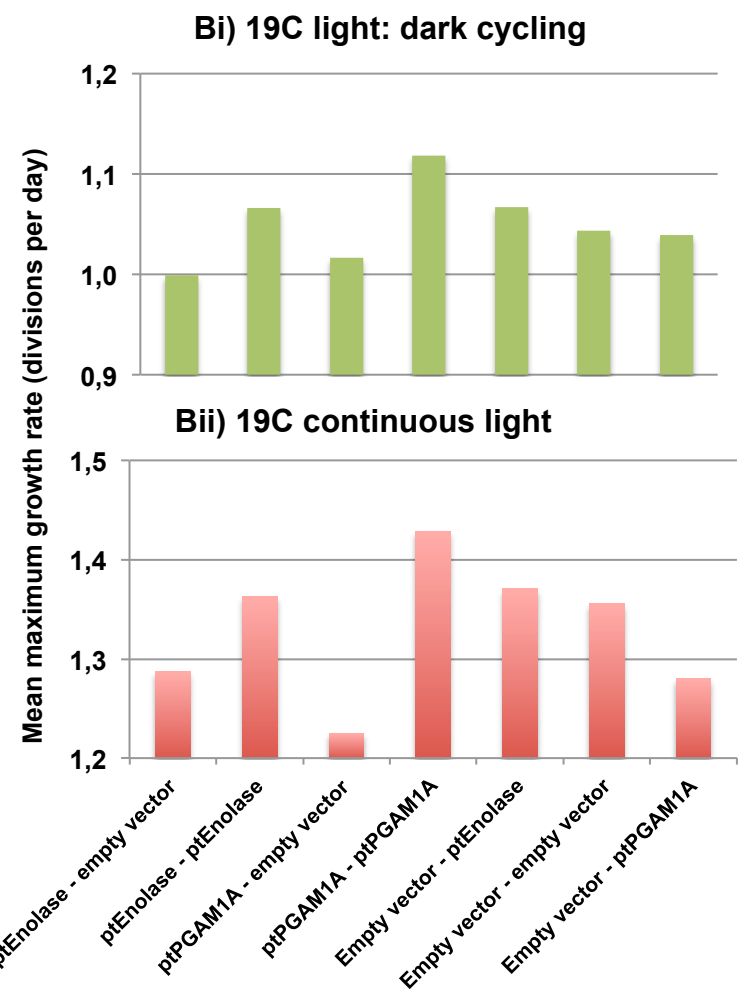


Fig. S13. Growth phenotypes of complemented glycolysis mutants. A: scatterplot of maximum growth rates, calculated from two replicates, under 19C LD cycling (horizontal) and 19C CL conditions (vertical axis) for 47 total ptEnolase **(i)** and ptPGAM1A mutants **(ii)** complemented with blasticidin-resistant and GFP-linked ptEnolase, ptPGAM1 or blank constructs, compared to empty vector zeocin-resistant control lines transformed with the same constructs. The distribution of individual growth rates within complemented and blank transformed populations are shown using solid coloured, and grey dashed line circles respectively. Vectors plot the difference between the mean growth rates calculated for blank transformed versus complemented mutant lines, alongside ANOVA P-values of separation. **B:** mean growth rates calculated for blank transformed and complemented mutant lines, alongside empty vector primary transformants complemented either with blank, ptEnolase or ptPGAM1A constructs, showing effective overlap between complemented mutant and empty vector primary transformant line growth rates.

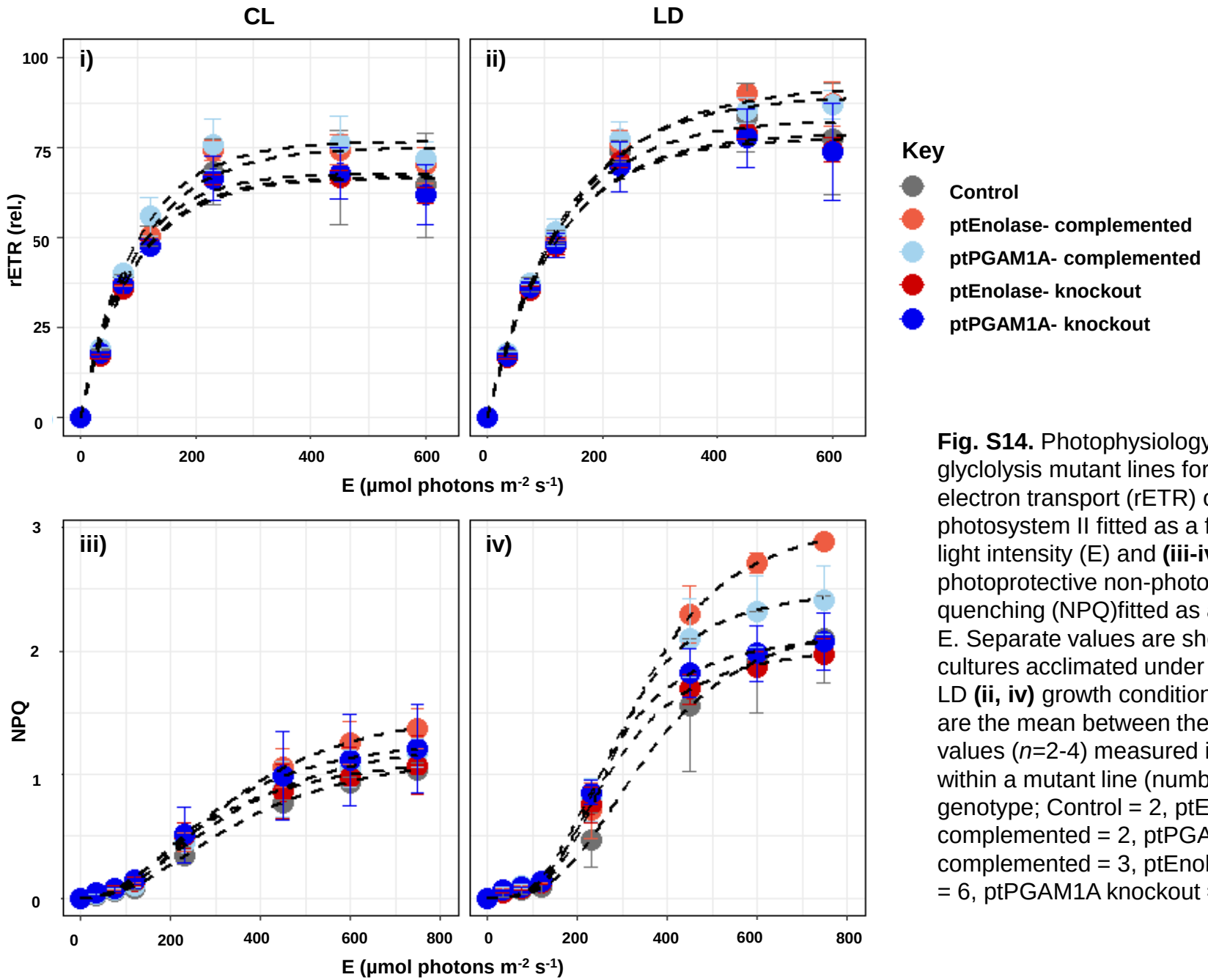


Fig. S14. Photophysiology curves in glyclolysis mutant lines for **(i-ii)** relative electron transport (rETR) of photosystem II fitted as a function of light intensity (E) and **(iii-iv)** photoprotective non-photochemical quenching (NPQ)fitted as a function of E. Separate values are shown for cultures acclimated under 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 a mutant line (number of lines per genotype; Control = 2, ptEnolase complemented = 2, ptPGAM1A complemented = 3, ptEnolase knockout = 6, ptPGAM1A knockout = 3).

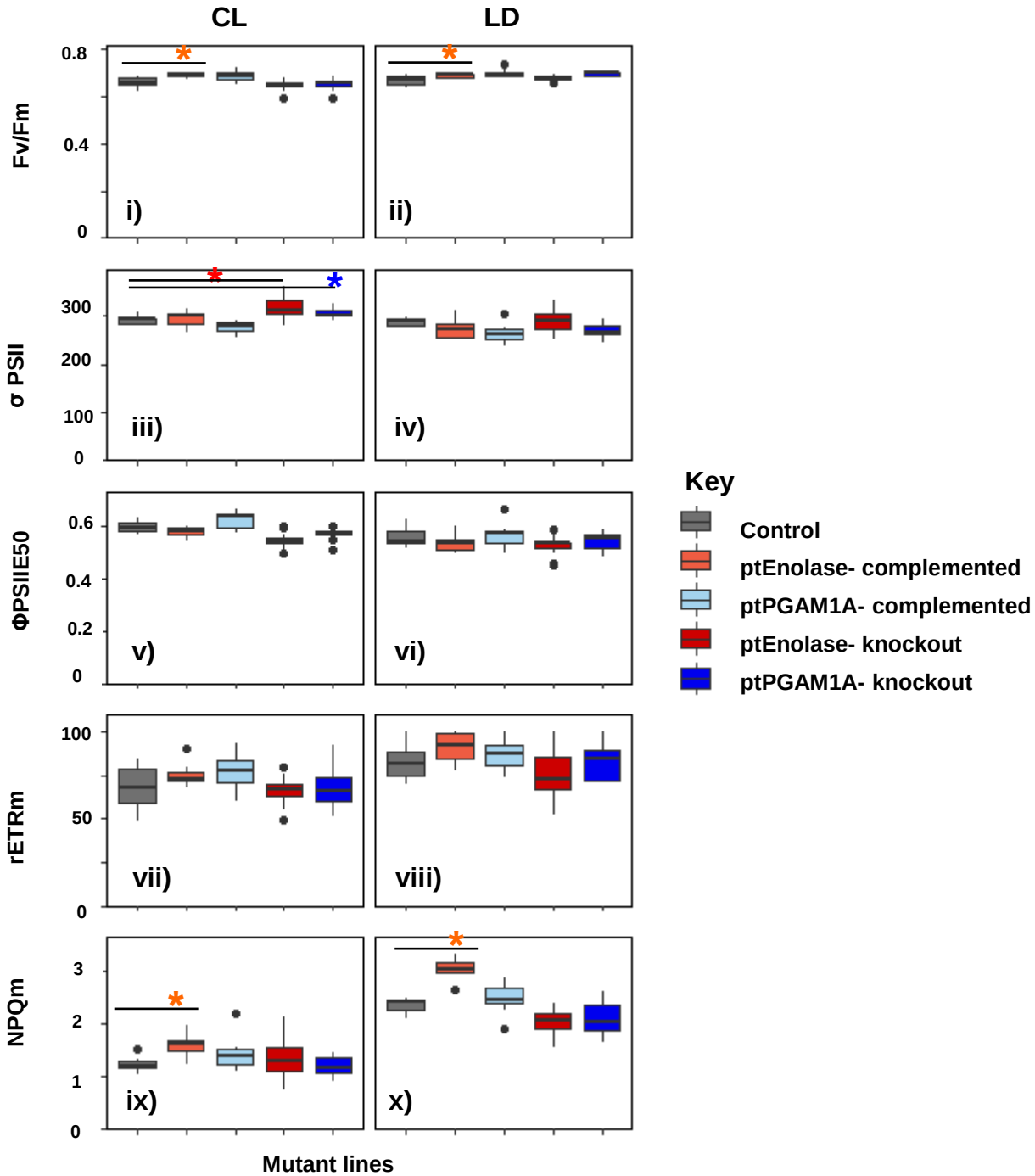


Fig. S15. Boxplots of measured photophysiological parameters in glycolysis mutants lines for cultures acclimated under 19C CL (left panels) and LD (right panels) growth conditions: (**i-ii**) maximum quantum yield of photosystem (PSII) in the dark (Fv/Fm) (**iii-iv**) PSII functional absorbtion cross-section (σPSII), (**v-vi**) PSII quantum yield under the growth light of 50 µmol photons m^{-2} s^{-1} (ΦPSIIE50) (**vii-viii**) maximal relative electron transport at PSII and (**ix-x**) maximum non-photochemical quenching (NPQm). Significantly different values observed for knkockout and complementation mutants relative to control lines (one-way ANOVA, $P < 0.05$) are asterisked, with asterisk colour corresponding to the mutant line considered. Each boxplot includes all measured/fitted values for each strain within a mutant line (number of lines per genotype; Control = 2, ptEnolase complemented = 2, ptPGAM1A complemented = 3, ptEnolase knockout = 6, ptPGAM1A knockout = 3).

.

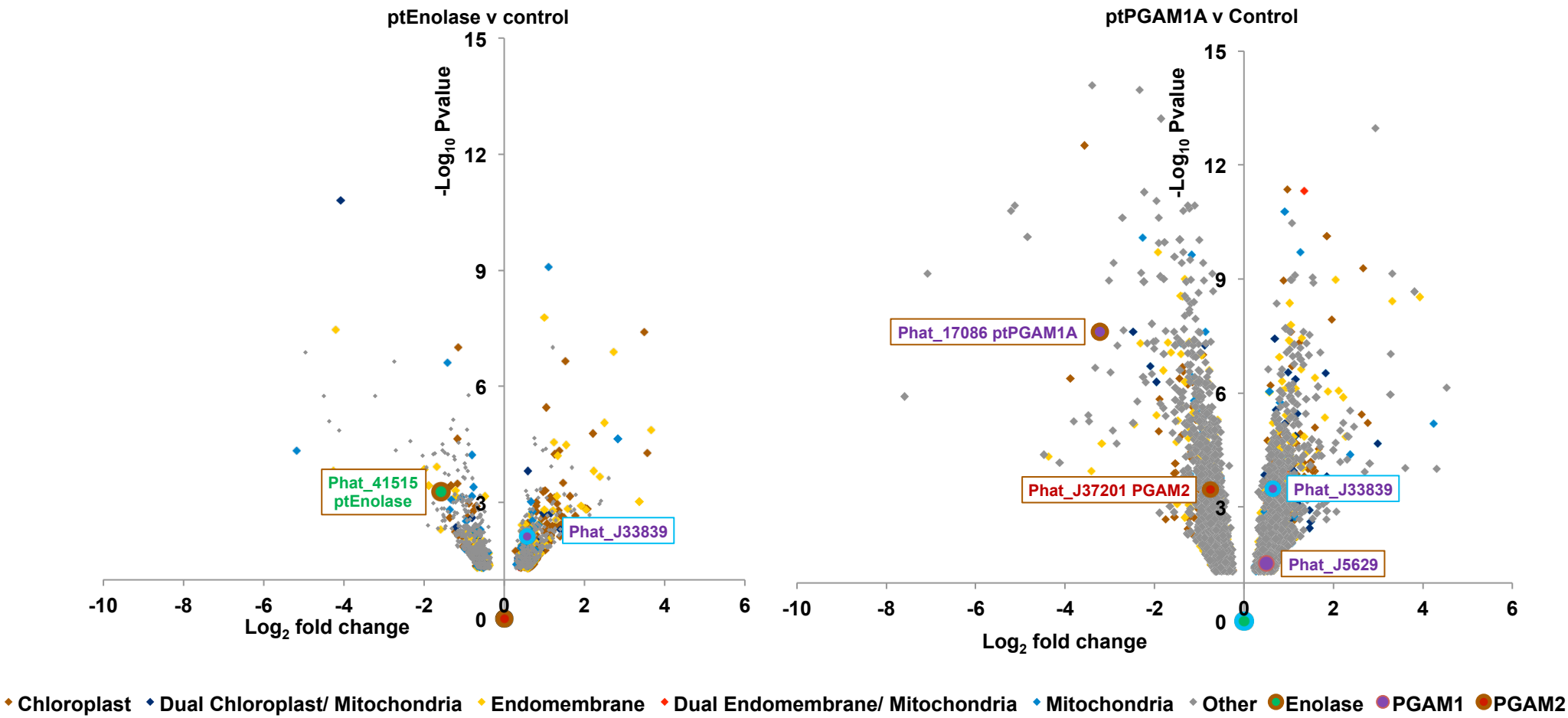
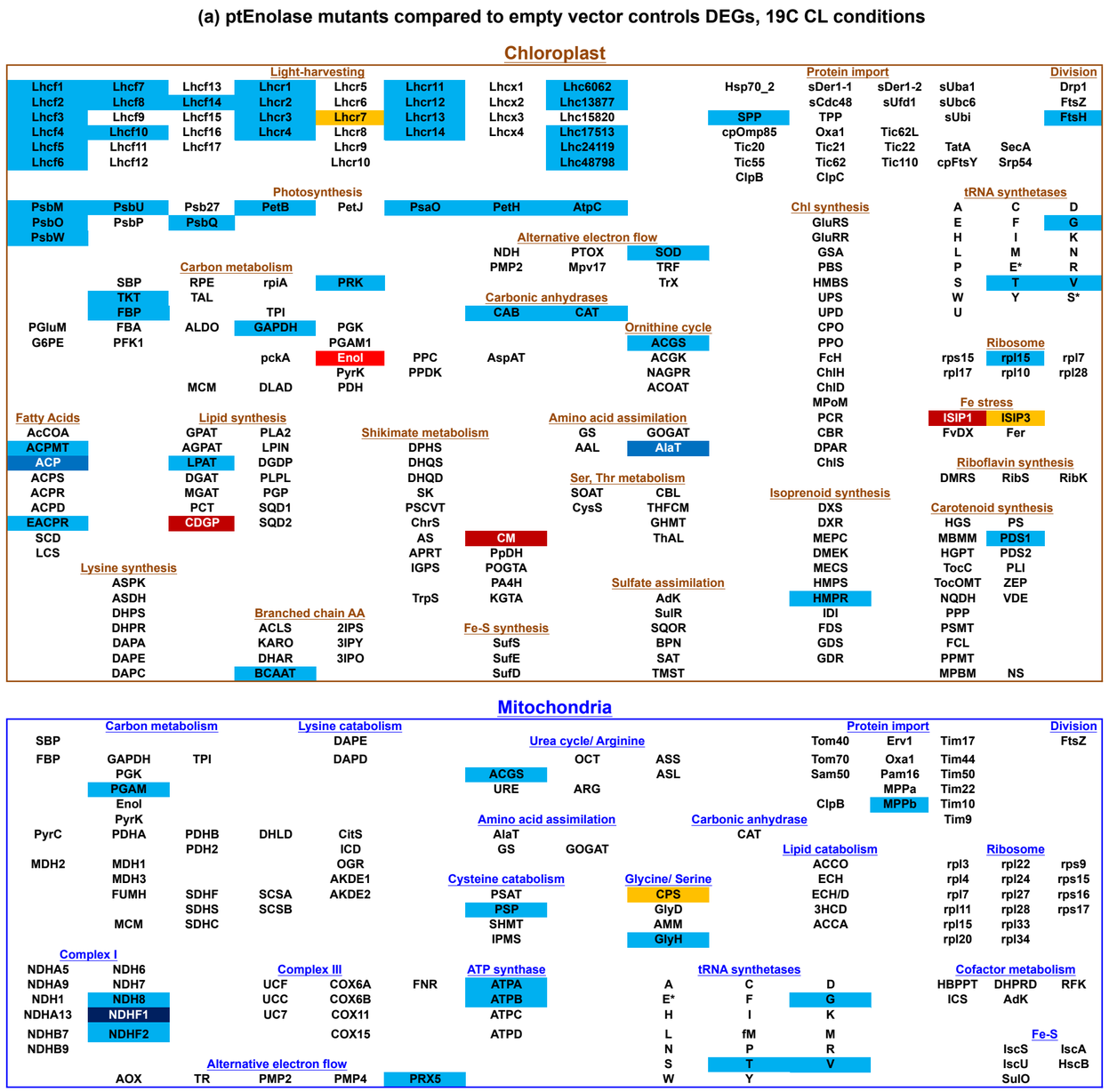
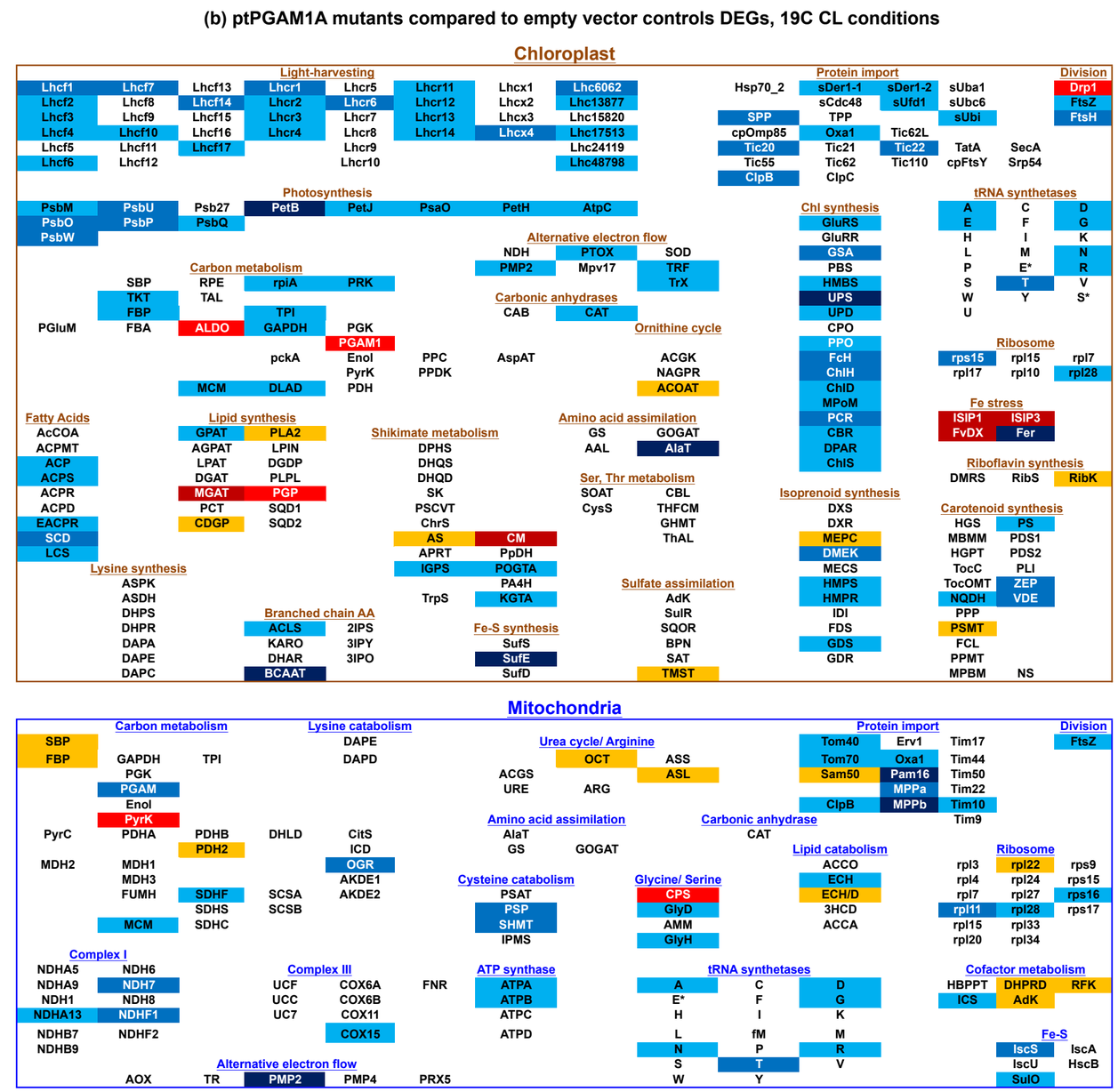


Fig. S16. Volcano plots of differentially expressed genes in ptEnolase (left) or ptPGAM1A mutants (ight) compared to empty vector control lines under 19C CL conditions. Genes are labelled by consensus localisation prediction, and differentially expressed Enolase and PGAM isoforms are enlarged as per **Fig. S6**. Of note, ptEnolase (PhatrJ_41515) and ptPGAM1A (PhatrJ_17086) are significantly less expressed in the corresponding mutant lines than the controls, consistent with effective transcriptional suppression of the mutated genes.





Scale

- P < E-5
- P < 0,001
- P < 0,05
- n.s.
- P < 0,05
- P < 0,001
- P < E-5

Fig. S17 Differentially expressed components of core plastid and mitochondrial metabolism in **(a)** ptEnolase and **(b)** ptPGAM1A mutants compared to empty vector control lines under 19C CL conditions, shown per **Fig. S7**. Cells are shaded based on DESeq P-value of gene expression difference, with blue cells indicating upregulation and red cells downregulation compared to control lines. Where more than one gene exists for an encoded function in the *Phaeodactylum* genome, only the most strongly differentially expressed isoform is shown.

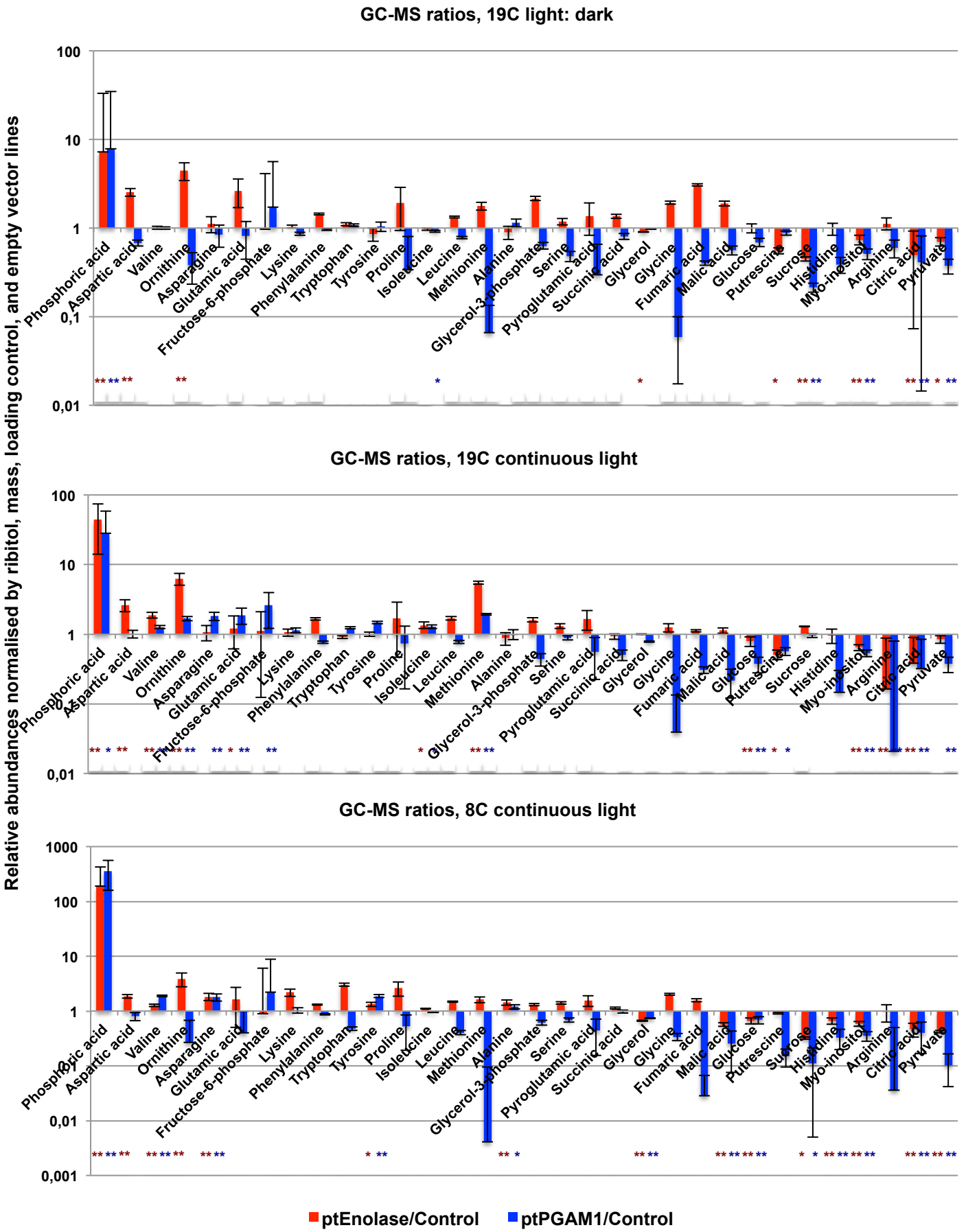


Fig. S18. Differentially accumulated metabolites in glycolysis mutant lines. This figure shows bar plots of the mean and standard deviation of the ratios of 39 metabolites assessed by GC-MS in plastid glycolysis mutant lines under the three tested experimental conditions. Metabolites are sorted in ranked decreasing accumulation in mutant lines over all three conditions. Metabolites inferred to be differentially accumulated in each mutant line and condition are asterisked: ** denotesANOVA $P < 10^{-05}$; * denotes $P < 0.01$.

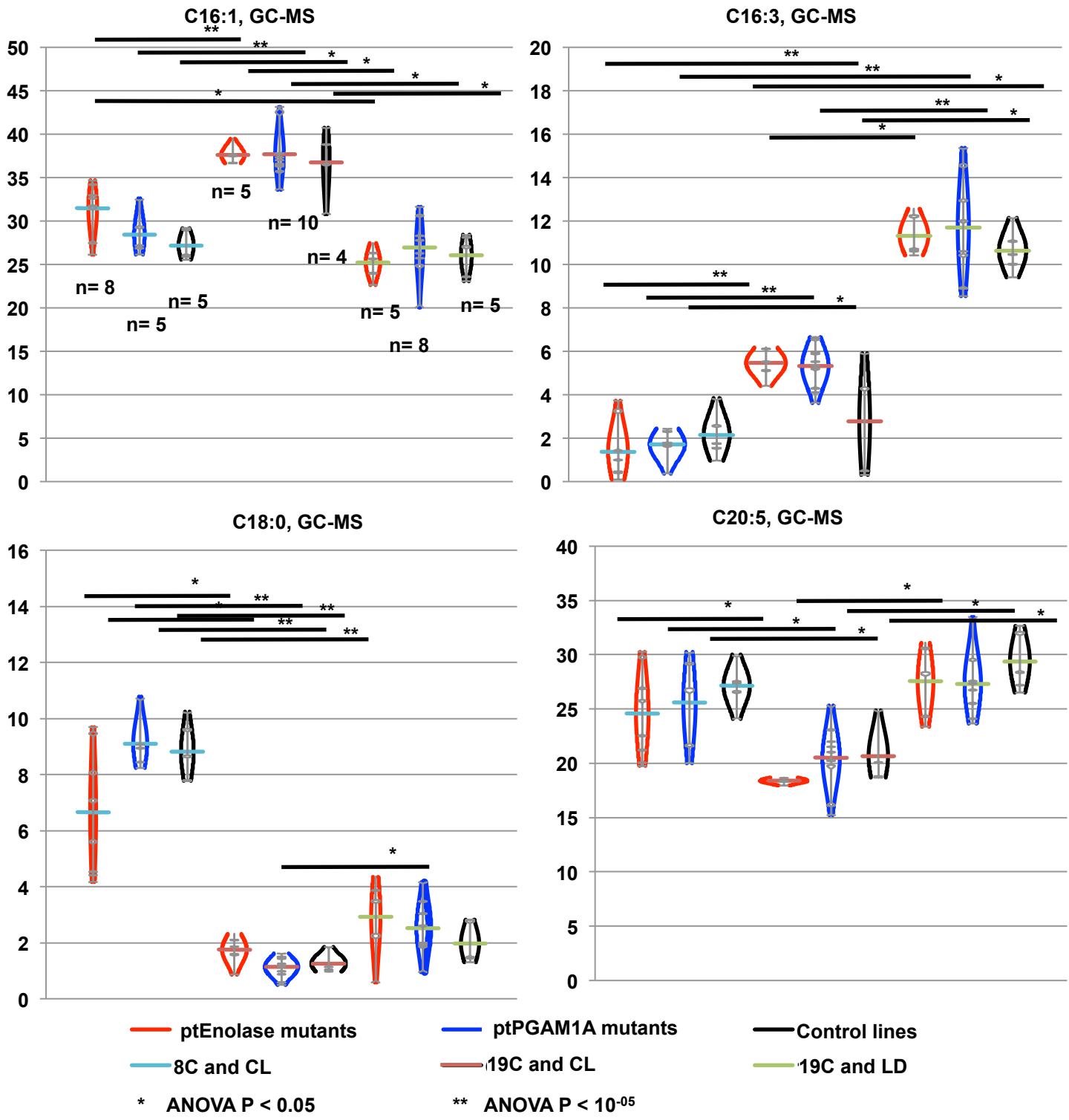


Fig. S19. Differential accumulation of fatty acids across experimental conditions. This figure shows violin plots as per **Fig. 6** of the frequency, mean and normal distributions of the relative proportions of four fatty acids assessed by GC mass spectrometry in glycolysis mutant and control lines across three experimental conditions. Despite substantial variation in fatty acid profile across experimental treatments (e.g., overaccumulation of C16:1 in lieu of C20:5 under 19C CL; C16:3 overaccumulation specifically in 19C LD; and C18:0 overaccumulation specifically in 8C CL), little difference is observed between glycolysis mutant and control lines under each condition.

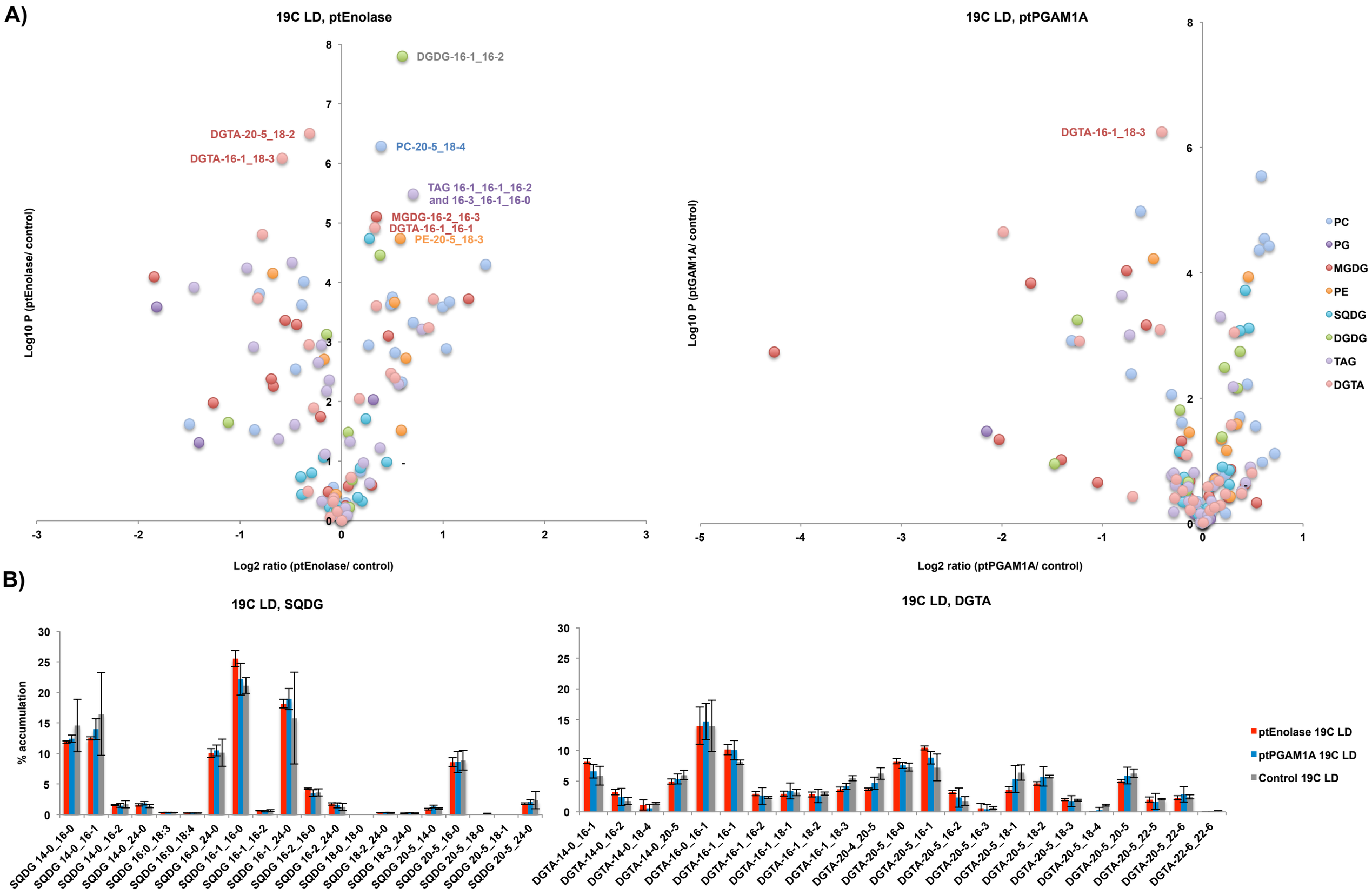


Fig. S20. Differentially accumulated lipids in 19C LD condition. A: Volcano plots shows (horizontal axis) $\log_2$ accumulation ratios and (vertical axis) $-\log_{10}$ ANOVA Pvalues for separation of mean proportions of specific fatty acids, across all fatty acids observed in a specific lipid class in glycolysis mutants versus control lines, harvested under 19C LD conditions. Specific lipids that show extreme ($P < 10^{-05}$) 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 overaccumulation of *sn*-1 C16 in ptEnolase mutant MGDG and DGDG pools, a corresponding underaccumulation of *sn*-1 C20 in ptEnolase mutant DGTA pools, and a conserved underaccumulation of DGTA-16-1_18-3 in both ptEnolase and ptPGAM1A mutants.

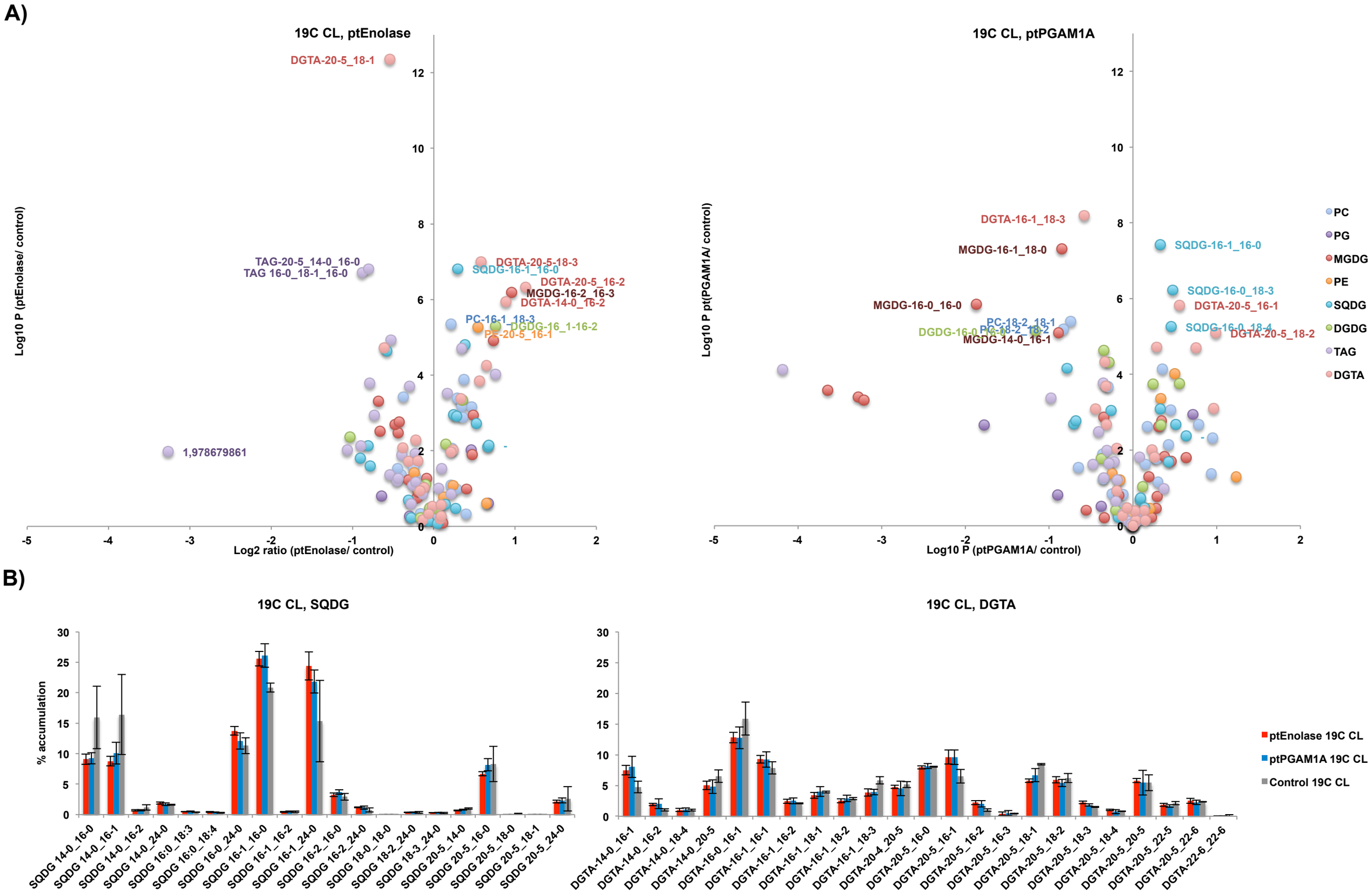


Fig. S21. Differentially accumulated lipids in 19C CL condition. A: Volcano shows (horizontal axis) $\log_2$ accumulation ratios and (vertical axis) $-\log_{10}$ ANOVA Pvalues for separation of mean proportions of specific fatty acids, across all fatty acids observed in a specific lipid class in glycolysis mutants versus control lines, and **B:** bar plots of SQDG and DGTA accumulation in lines harvested under 19C CL conditions, shown as per **Fig. S19**. These data suggest greater changes in glycolysis mutant lipid architecture than under 19C LD conditions, including conserved overaccumulations of *sn*-1 C16 in ptEnolase and ptPGAM1A mutant SQDG and *sn*-1 C20 in ptEnolase and ptPGAM1A mutant DGTA pools compared to control lines.

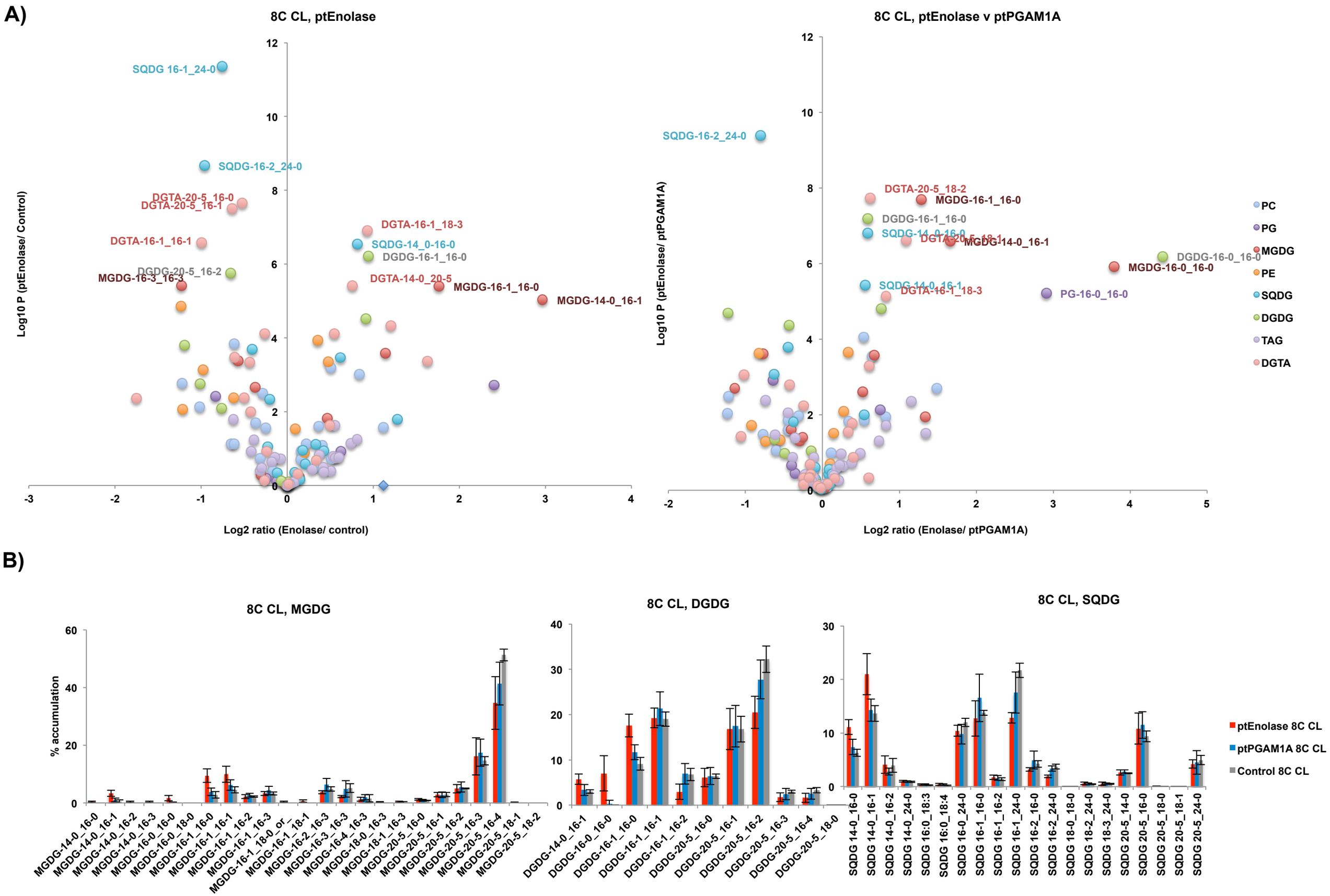


Fig. S22. Volcano plots of specific differentially accumulated lipids in ptEnolaqse mutants under 8C CL conditions. These plots shows (horizontal axis) $\log_2$ accumulation ratios and (vertical axis) $-\log_{10}$ ANOVA Pvalues for separation of mean proportions of specific fatty acids, across all fatty acids observed in a specific lipid class in ptEnolase mutants versus control lines, and ptEnolase mutants versus ptPGAM1A mutants harvested under 8C CL conditions, shown as per **Fig. S19**. No sigificantly differentially accumulated ($P < 10^{-05}$) lipids were observed in corresponding comparisons of ptPGAM1A mutants and control lines. These data suggest specific overaccumulations in short-chain *sn*-1 MGDG, DGDG, and *sn*-2 SQDG-in ptEnolase mutants compared to other lines.

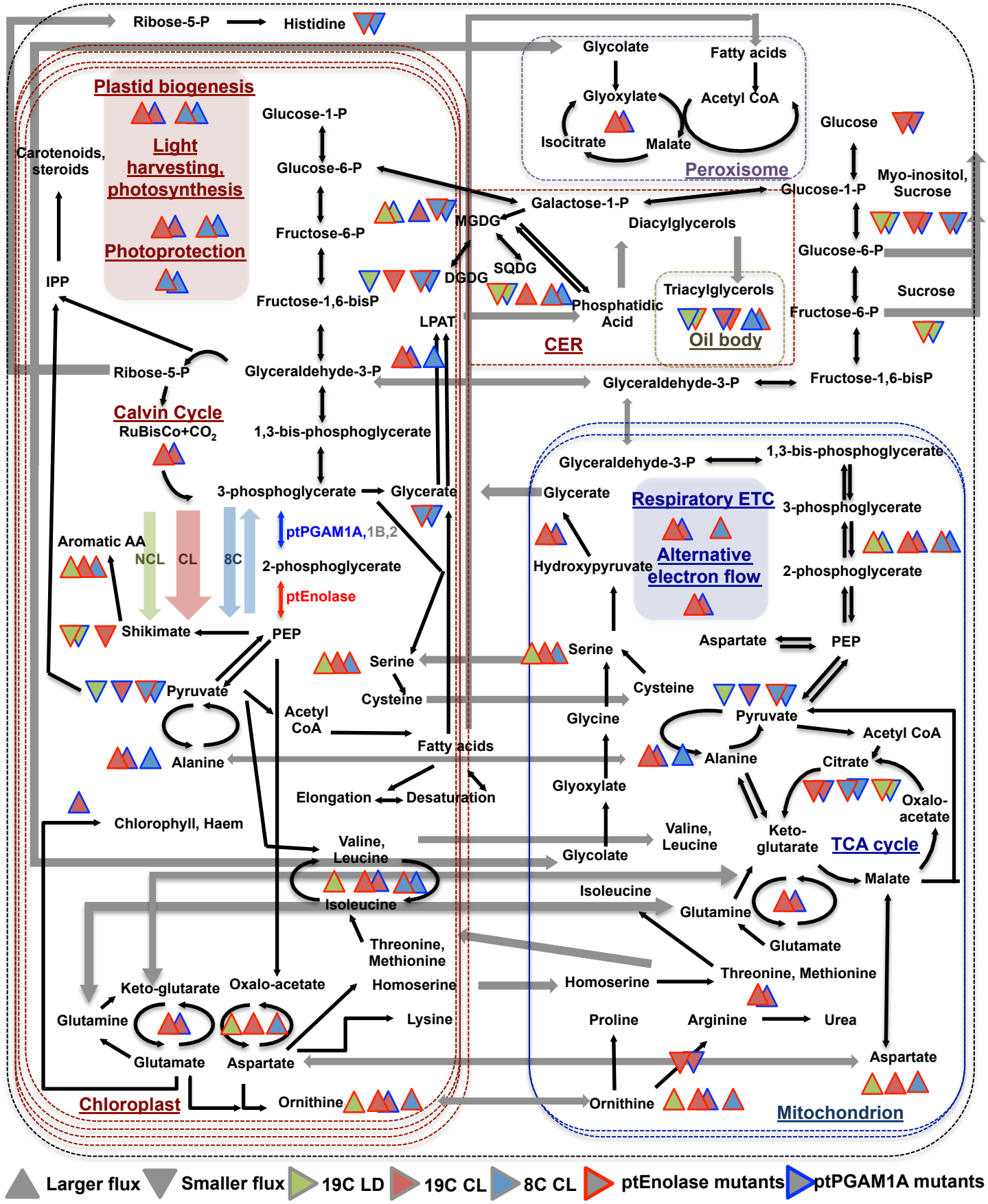


Fig. S23. Systemic effects of plastidial glycolysis on *Phaeodactylum* organelle metabolism. This figure schematic diagrams of core *Phaeodactylum* chloroplast, mitochondrial and cytoplasmic metabolism pathways, and the proposed functions of plastidial glycolysis under 19C LD, 19C CL and 8C CL conditions based on combined transcriptomic, qPCR, metabolomic and lipid profiles.