Algal Research 1 (2012) 2–16

Contents lists available at SciVerse ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways

Sarah R. Smith, Raffaela M. Abbriano, Mark Hildebrand *

Marine Biology Research Division, Scripps Institution of Oceanography, UCSD, La Jolla, CA, United States

ARTICLE INFO

Article history:
Received 24 February 2012
Accepted 16 April 2012

Keywords:
Diatom
Glycolysis
Gluconeogenesis
Carbon metabolism
Evolution
Compartmentation

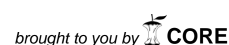
ABSTRACT

A major challenge in the development of microalgal strains for large-scale production is the optimization of biomass accumulation and production of fuel-relevant molecules such as triacylglycerol. Selecting targets for genetic manipulation approaches will require a fundamental understanding of the organization and regulation of carbon metabolic pathways in these organisms. Functional genomic and metabolomics data is becoming easier to obtain and process, however interpreting the significance of these data in a physiological context is challenging since the metabolic framework of all microalgae remains poorly understood. Owing to a complex evolutionary history, diatoms differ substantially from many other photosynthetic organisms in their intracellular compartmentation and the organization of their carbon partitioning pathways. A comparative analysis of the genes involved in carbon partitioning metabolism from *Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, and *Fragilariopsis cylindrus* revealed that diatoms have conserved the lower half of glycolysis in the mitochondria, the upper half of glycolysis (including key regulatory enzymes) in the cytosol, and several mitochondrial carbon partitioning enzymes. However, some substantial differences exist between the three diatoms investigated, including the translocation of metabolic pathways to different compartments, selective maintenance and horizontal acquisition of genes, and differential gene family expansions. A key finding is that metabolite transport between intracellular compartments is likely to play a substantial role in the regulation of carbon flux. Analysis of the carbon partitioning components in the mitochondria suggests an important role of this organelle as a carbon flux regulator in diatoms. Differences between the analyzed species are specific examples of how diatoms may have modified their carbon partitioning pathways to adapt to environmental niches during the diversification of the group. This comparative analysis highlights how even core central pathways can be modified considerably within a single algal group, and enables the identification of suitable targets for genetic engineering to enhance biofuel precursor production.

© 2012 Elsevier B.V. Open access under CC BY-NC-ND license.

d similar papers at core.ac.uk

brought to you by



provided by Elsevier - Publisher Connector

Algal biofuels research is gaining momentum, fueled in part by the ease of obtaining genomic and transcriptomic information. A goal of many researchers is to mine this data to identify gene targets for manipulation that will improve growth and lipid, or more specifically, triacylglycerol (TAG), accumulation characteristics that will drive down the cost of production. There are several approaches that can be taken to enhance the TAG content in microalgae including over-expressing fatty acid or TAG biosynthesis genes, inhibiting lipid catabolism, and inhibiting metabolic pathways that compete with lipid biosynthesis for carbon intermediates such as the synthesis of storage carbohydrates [1]. Many carbon metabolic pathways are extensively studied and consequently it may be assumed that these pathways and their genes are well understood in microalgae. However, algae have an evolutionary history that is quite divergent from many model organisms, and the biochemistry of these pathways in algae is in general poorly characterized. Characterization of these pathways and their regulation in microalgae is

Abbreviations: ACCase, acetyl-CoA carboxylase; ATP-PFK, ATP-dependent phosphofructokinase; CCM, carbon concentrating mechanism; DNRA, dissimilatory nitrate reduction to ammonia; ED, Entner–Dourdoff glycolysis; EGT, endosymbiotic gene transfer; EMP, Embden–Meyerhof–Parnas glycolysis; ENO, enolase; FA, fatty acid; FBA, fructose-bisphosphate aldolase; FBP, fructose 1,6 bisphosphatase; Fru 2,6 bisP, fructose 2,6 bisphosphate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GLK, glucokinase; GPI, glucose-6-phosphate isomerase; HGT, horizontal gene transfer; MDH, malate dehydrogenase; ME, malic enzyme; OAA, oxaloacetate; PC, pyruvate carboxylase; PDRP, pyruvate phosphate dikinase regulatory protein; PEP, phosphoenolpyruvate; PEPC, phosphoenolpyruvate carboxylase; PEPCK, phosphoenolpyruvate carboxykinase; PEPS, phosphoenolpyruvate synthase; PFK, phosphofructokinase; PF2K/F26BP, bifunctional 6-phosphofructo-2-kinase/fructose 2,6 bisphosphatase; PGAM, phosphoglycerate mutase; PGK, phosphoglycerate kinase; PK, pyruvate kinase; PPDK, pyruvate phosphate dikinase; PP_i-PFK, pyrophosphate-dependent phophofructokinase; TAG, triacylglycerol; TCA, tricarboxylic acid; TPI, triose phosphate isomerase.

* Corresponding author at: 9500 Gilman Dr., La Jolla, CA 92093-0202, United States. Tel.: +1 858 822 0167.

E-mail address: mhildebrand@ucsd.edu (M. Hildebrand).

2211-9264 © 2012 Elsevier B.V. Open access under CC BY-NC-ND license.
doi:10.1016/j.algal.2012.04.003

essential to identify the appropriate gene targets to modify for increased productivity.

In early work to modify algal strains for improved fuel precursor molecule production, acetyl-CoA carboxylase (ACCase), the first committed step of fatty acid biosynthesis, was successfully overexpressed in the diatom *Cyclotella cryptica* [2]. Despite an increased enzyme activity, there was no resulting increase in TAG content, suggesting that there are other factors that regulate the flux of carbon into fatty acid biosynthesis. Generally, microalgae do not accumulate TAG during growth, and only develop lipid bodies during the stationary growth phase or when nutrient-limited [3,4]. Under silicon-limited TAG accumulation conditions in *C. cryptica*, it was also demonstrated that the flux of carbon was repartitioned from storage carbohydrates into lipid over the course of TAG induction [5,6]. Blocking storage carbohydrate synthesis, as demonstrated in starchless mutants of the green alga *Chlamydomonas reinhardtii*, also enhances TAG accumulation [7]. From these data, it is reasonable to expect that enhancing carbon flux towards TAG by reducing flux to competing pathways is a viable approach to improve de novo TAG biosynthesis and accumulation in algae.

Central carbon metabolic pathways are reasonable targets for the modification of intracellular carbon flux in algae. In all known photosynthetic organisms, the primary pathways involved in the partitioning of fixed carbon into either storage carbohydrates or TAG are glycolysis, gluconeogenesis, and pyruvate metabolism (Fig. 1). Glycolysis, or the catabolism of hexoses to produce pyruvate and ATP, provides the cell with energy and metabolic intermediates required to supply either the TCA cycle or fatty acid biosynthesis. Gluconeogenesis is essentially the reverse of the glycolysis pathway in that pyruvate is converted to hexoses that supply storage carbohydrate biosynthetic pathways. In this way, these pathways and their subsequent branch points act as partitioning regulators of intracellular carbon flux.

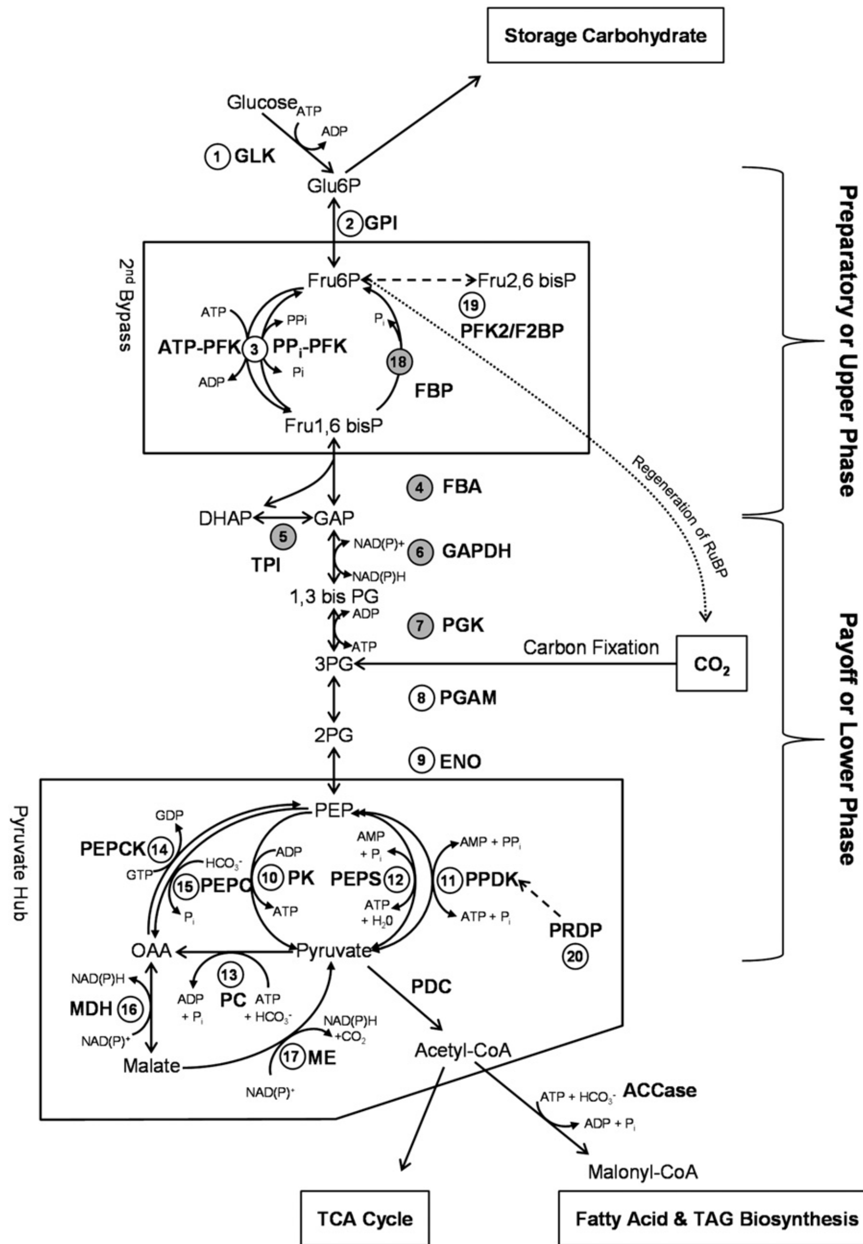


Fig. 1. Schematic of carbon partitioning enzymes involved in glycolysis, gluconeogenesis and pyruvate metabolism. Circled numbers adjacent to enzyme abbreviations correspond to those in Table 1. Shaded circles denote enzymes that are also involved in the Calvin–Benson cycle. Arrows indicate the directionality of the reaction catalyzed by each enzyme. Brackets indicate the two phases of glycolysis; the upper phase (GLK to TPI) requires an initial energy investment, while the lower phase (GAPDH to PFK) produces ATP and reducing equivalents.

From decades of research on terrestrial plants and other model organisms, it is well known that the direction of carbon flux through these core and conserved pathways is regulated by mass action (the concentration of reactants relative to products) and by dedicated glycolysis or gluconeogenesis enzymes at a few key regulatory points [8,9]. The regulatory points are known as glycolysis bypass points (glycolysis reactions must be "bypassed" by dedicated gluconeogenic enzymes; Fig. 1). The directionality and magnitude of metabolite flux through these bypass points is tightly regulated by a combination of fine (enzyme phosphorylation, allosteric effectors) and coarse (transcription, translation, and protein turnover) controls [9]. Glycolysis bypass points are appropriate targets for genetic engineering since they are natural points of regulation, and control the flux of carbon in one favorable direction. Another important mechanism of carbon flux regulation in eukaryotes is their ability to adjust the concentration of metabolic intermediates in sub-cellular compartments, thereby affecting the supply of these intermediates into various pathways. Compartmentation makes it possible for eukaryotic cells to run competing pathways simultaneously, regulate cellular and organellar energetics, and control the rate of intermediate supply to other compartments via specialized transporters [10–12]. Considering the latter point, the transport of metabolites between compartments can be important regulatory steps.

Compartmentation in diatoms is significantly different from green algae and terrestrial plants due to differences in their respective evolutionary histories [13,14]. Photosynthetic eukaryotes arose via a primary endosymbiosis event where chloroplasts were derived from a cyanobacterial endosymbiont (Fig. 2). The progenitor plant cell gave rise to the glaucophytes, green algae (and eventually plants derived from this lineage) and red algae (Fig. 2). Diatoms and other chromalveolates arose through a secondary endosymbiotic event, in which a red algal endosymbiont was enslaved as a plastid, followed by the eventual loss of the red algal nucleus and mitochondria (Fig. 2). One consequence of the complex evolutionary history of secondary endosymbionts is the distinct chloroplast membrane organization [15,16]. In addition to the chloroplast inner and outer membranes typical of all photosynthetic eukaryotes, diatom (and other chromalveolate) plastids are surrounded by a periplastid membrane (the relic plasma membrane of the red algal symbiont), which defines the periplastid compartment (PPC), and the endoplasmic reticulum, commonly called the "chloroplast ER" (Fig. 2, [17]). This additional complexity adds to the challenge of reconstructing algal metabolic networks.

As secondary endosymbionts, the nuclear genomes of diatoms are a combination of genes from several evolutionarily distinct organisms, and analysis of diatom genomes has revealed both plant-like and animal-like features [18,19]. The genomes of both non-photosynthetic and photosynthetic protists, including diatoms, are known to have been shaped by the acquisition of genes from endosymbiotic/horizontal gene transfer events and duplication events, as well as by the selective deletion of certain genes or gene families [20,21]. Furthermore, there is evidence that some organisms have re-targeted nuclear-encoded genes to distinct organelles, and in some cases have single isozymes that can be dually targeted to more than one subcellular location, adding to the flexibility of their pathway compartmentation significantly [22]. Additionally, diatoms seem to possess several distinct genes encoding isozymes of conserved metabolic pathways, indicating that they have acquired genes from several different sources and that these isozymes may be functionally differentiated [23]. Analysis of the genome of the first diatom sequence available, *Thalassiosira pseudonana*, first identified enzymes for the complete cytosolic glycolysis and gluconeogenesis pathways [18]. Later work investigating the genome of *Phaeodactylum tricornutum* suggested that all the reactions of glycolysis might also occur in diatom plastids [23]. Additionally, Kroth et al. showed that there are isozymes for the complete lower half of glycolysis (TPI to PK) predicted to be mitochondrially located, a finding that supported earlier work demonstrating the presence of a triose phosphate isomerase (TPI)/glyceraldehyde 3P dehydrogenase (GAPDH) fusion protein in diatom mitochondria [23,24]. The evolutionary divergence of the centric and pennate classes of diatoms occurred at least 90 mya (based on the fossil record), yet the genome of *T. pseudonana* (centric) and *P. tricornutum* (pennate) differ to the same extent as those of fish and mammals, which diverged approximately 550 mya [25]. Diatoms evolved over a time of substantial environmental change on the planet, particularly with regards to CO_2 and O_2 levels [26,27], and it is reasonable to assume that some of the differences distinguishing different classes of diatoms are adaptations for optimal productivity under the particular conditions during which they arose. There have been substantial differences described for carbon concentrating mechanisms in different classes of diatoms [28,29], and recently, a phosphoketolase pathway was identified in *P. tricornutum*, but not in *T. pseudonana* [30] consistent with evolutionarily based alterations in fundamental aspects of carbon metabolism. Despite previous valuable insights, a comprehensive comparative analysis of the homology and targeting of the carbon partitioning enzymes and the extent to which diatoms differ from one another has not been conducted. The extent to which the core carbon partitioning proteome has been modified by the selective maintenance or deletion of genes, duplications, re-targeting, or acquisition via horizontal gene transfer is not well-understood but should

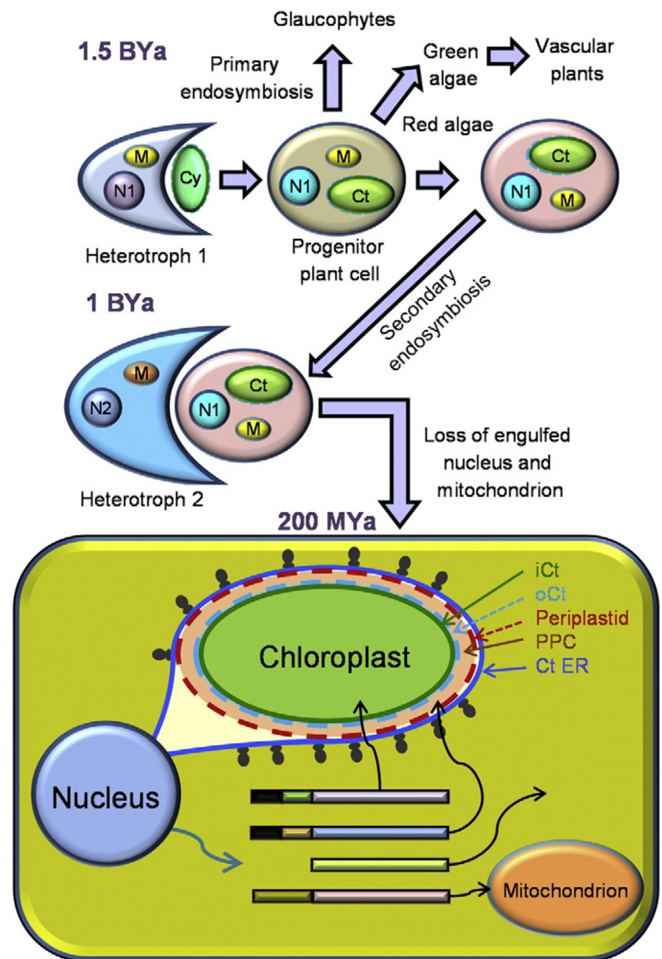


Fig. 2. The primary and secondary endosymbiotic events that gave rise to modern diatoms. The upper panel shows the evolutionary progression leading to the heterokonts. Organelles are colored to denote different origins and labeled as such (N = nucleus, M = mitochondria, Ct = chloroplast, Cy = cyanobacterium). The lower panel is a diagram of diatom intracellular compartmentation, with an emphasis on the chloroplast and associated extra membranes relative to the progenitor plant cell. Colored bars represent proteins with different leader sequences targeting them to different intracellular locations. The inner chloroplast membrane (iCt), outer chloroplast membrane (oCt), periplastid membrane, periplastid compartment (PPC), and chloroplast ER (ct ER) are labeled.

provide insight into unique adaptive metabolic capabilities characteristic of a group or species.

The primary aim of this study was to utilize the available genome sequences for the centric *T. pseudonana*, the raphid pennate *P. tricornutum*, and the assembly scaffolds for the psychrophilic raphid pennate diatom *Fragilariopsis cylindrus*, to perform a comprehensive in silico evaluation of the carbon partitioning proteome (Fig. 3). The analysis resulted in the identification of a rich variety of features, including annotation of absolutely core metabolic genes (likely to have an indispensable metabolic function), characterization of gene duplication events, description of unique genes arising from selective deletion, evidence of genes acquired by horizontal gene transfer, and differential intracellular targeting, and enabled the reconstruction of metabolic networks, with an emphasis on intracellular compartmentation. The analysis highlights distinctions between representatives of the centric and pennate diatom lineages, and enables generalizations to be made about diatom carbon metabolism that distinguishes them from known model organisms and other algae. The information gained from this type of analysis is essential to inform metabolic engineering approaches to improve precursor molecule production for biofuel applications.

2. Materials and methods

2.1. Sequence screening and functional annotation

KEGG Pathway Database and Gene Ontology annotations were used to identify protein sequences for all genes of interest from the genomes of *T. pseudonana* v3.0 (http://genome.jgi-psf.org/Thaps3/Thaps3.home.html), *P. tricornutum* v2.0 (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html), and *F. cylindrus* v1.0 (http://genome.jgi-psf.org/Fracy1/Fracy1.home.html). It is common that gene models in these genomes are not full-length, therefore in all cases, protein models were checked to ensure they were extended to the first in-frame methionine, and occasionally EST data was used to manually obtain the full ORF. These manual modifications are noted in the supplemental table (Table S1). To verify that no sequences were missed, any proteins identified from the annotations were queried against the other diatom genomes using BLAST [31]. In many cases, predicted gene models with significant homology but lacking annotations were identified (Table S1). When appropriate, functional domains that were used as criteria to identify predicted proteins were included in the supplemental table (Table S1). Some enzymes with weak annotated functions did not meet our criteria as the functional protein of interest and are not included in Table 1 but are included in the supplemental table (Table S1).

2.2. Bioinformatics-based targeting analysis

Several bioinformatics software programs were used for intracellular targeting prediction. For general localization predictions, the programs HECTAR, Predotar, and TargetP were used [32–34]. The presence of mitochondrial target peptides was determined using Mitoprot [35]. Diatom plastid-targeted peptides must cross four membranes, so their plastid pre-sequences are distinct from green algae and plants [36,37]. Identification of plastid-targeted proteins based on sequence data can

Table 1
Number of genes encoding enzymes involved in carbon partitioning metabolism (and their putative metabolic pathways) in the genomes of *T. pseudonana*, *P. tricornutum*, *F. cylindrus*, and *Chlamydomonas reinhardtii*. Gray shadowing indicates the occurrence of more than three isoenzymes in any of the diatom genomes. Gly = glycolysis, GNG = gluconeogenesis, Pyr = pyruvate metabolism, CB = Calvin–Benson cycle, TCA = tricarboxylic acid cycle, C4 = C4-type photosynthesis.

#	Enzyme name	EC number	Putative roles in metabolism						Total #	# of Isoenzymes/genome			
			Gly	GNG	Pyr	CB	TCA	C4	genes	T. pseudo	P. tricorn	F. cylin	C. reinh.
1	HXK	2.7.1.1	x	--	--	--	--	--	0	0	0	0	1
1	GLK	2.7.1.2	x	--	--	--	--	--	1	1	1	1	1
2	GPI	5.3.1.9	x	x	--	--	--	--	4	4	4	4	1
3	ATP -PFK	2.7.1.11	x	--	--	--	--	--	3	1	3	1	1
3	PP$_i$-PFK	2.7.1.90	x	x	--	--	--	--	1	1	1	1	2
4	FBA	4.1.2.13	x	x	--	x	--	--	6	4	5	6	4
5	TPI	5.3.1.1	x	x	--	x	--	--	5	3	3	5	1
5/6	TPI/GAPDH	5.3.1.1/1.2.1.12	x	x	--	--	--	--	1	1	1	1	0
6	GAPDH	1.2.1.12/1.2.1.13	x	x	--	x	--	--	5	4	4	4	4
7	PGK	2.7.2.3	x	x	--	x	--	--	6	5	3	3	2
8	PGAM	5.4.2.1	x	x	--	--	--	--	10	6	7	7	1
9	ENO	4.2.1.11	x	x	--	--	--	--	3	2	2	3	1
10	PK	2.7.1.40	x		x	--	--	--	8	5	7	5	5
18	FBP	3.1.3.11	--	x	--	x	--	--	6	3	5	7	1
11	PPDK	2.7.9.1	x	x	x	--	--	x	1	1	1	1	2
12	PEPS	2.7.9.2	x	x	x	--	--	--	3	1	0	2	0
13	PC	6.4.1.1	--	x	x	--	x	--	4	3	2	2	1
14	PEPCK	4.1.1.32	--	x	x	--	--	x	1	1	1	1	1
15	PEPC	4.1.1.31	--	--	x	--	x	x	2	2	2	2	1
16	NAD(P)-MDH	1.1.1.37/1.1.1.82	--	--	x	--	x	x	2	2	1	1	4
17	NAD-ME	1.1.1.39	--	--	x	--	--	x	1	1	1	1	1
17	NADP-ME	1.1.1.40	--	--	x	--	--	x	1	0	1	0	4
19	P2FK/F2BP	3.1.3.46/2.7.1.105	x	x	--	--	--	--	2	2	2	2	2
20	PRDP	N/A	x	x	x	--	--	--	2	0	1	1	0
								Total:	74	51	55	58	N/A
								Conserved in all diatoms:		40, 78%	40, 73%	40, 69%	N/A
								Unique to genome:		7, 14%	8, 15%	10, 17%	N/A
								Found in 1 other genomes:		5, 8%	7, 13%	8, 14%	N/A

be challenging in green algae [38], however, the requirement for an ER signal peptide facilitates these predictions in diatoms. Additionally, a motif has been detected for targeting proteins to the periplastid compartment in diatoms [39]. To detect plastid and periplastid compartment targeting, the programs SignalP v4.0 and SignalP v3.0 were used along with manual inspection of the predicted cleavage site [39–41]. The program ChloroP was used to detect the presence of predicted chloroplast transit domains [42]. In cases where individual targeting predictions did not agree with one another, the localization was predicted based on weighing the relative predictions. The results of all the predictions are summarized in Table S1. In diatoms, several predictions using the bioinformatics approaches we outline have been confirmed experimentally [24,37,39,43,44]. In some cases there were high predictions for mitochondria and also for the presence of a signal peptide and/or chloroplast targeting. A signal peptide would be a default for chloroplast targeting, however, it is possible that different transcription starts or splicing could eliminate the signal peptide and/or chloroplast targeting sequences [45,46]. Because the mitochondrial predictions were so high, these enzymes were stated to be potentially dually targeted.

2.3. Phylogenetic analysis and criteria for homology and monophyly

Homologs (enzymes that catalyze the same reaction but are not necessarily evolutionarily related) which were identified in all three diatom genomes were aligned to determine sequence similarities and to assess phylogenetic relationships. The web-based ClustalW2 at the European Bioinformatics Institute (www.ebi.ac.uk/Tools/msa/clustalw2/) was used to generate pairwise output scores (Table S1). The slow pairwise alignment option and otherwise default multiple sequence alignment options were selected. Sequences homologous to the diatom enzymes were identified by searching the nonredundant GenBank CDS database (nr). Sequences were aligned with ClustalX 2.1, and RAxML-HPC BlackBox (7.2.8, Cipres Science Gateway, www.phylo.org) was used to generate protein maximum likelihood trees. Trees were used to identify monophyletic clusters of diatom genes. Monophyletic clusters of diatom genes found in all three diatom genomes were considered orthologs (genes that are derived from an enzyme found in the last common ancestor of diatoms). Genes that arose through an apparent duplication event are considered paralogs. When the duplication occurred prior to the evolutionary divergence of the different diatoms, these duplications were considered out-paralogs. Duplications that occurred after the divergence (and are only found in one diatom for example) are considered in-paralogs.

2.4. Nomenclature

Monophyletic clusters of diatom enzymes were arbitrarily numbered because at the outset it was not possible to assign functional differences or their relative importance in metabolism. In Table S1, protein ID numbers (PIDs) are given for each gene cluster. The PIDs are prefixed with Tp, Pt, or Fc to indicate *T. pseudonana*, *P. tricornutum*, and *F. cylindrus* respectively. In the case of FBA, the nomenclature from previous studies was maintained [23,44].

3. Results and discussion

3.1. Overview of glycolysis, gluconeogenesis, and pyruvate metabolism

Although glycolysis, gluconeogenesis, and pyruvate metabolism may be familiar core metabolic pathways, the details of their regulation are not as familiar; therefore we present an overview of the pathways and regulatory steps here to set the stage for subsequent analyses (Fig. 1). Most steps of glycolysis are carried out by enzymes that function bi-directionally in the pathway. These steps tend to be regulated by mass action, allostery, or post-translational modification. Steps that play key regulatory roles generally involve enzymes catalyzing unidirectional reactions.

Overall, glycolysis is an energy-generating pathway, but the pathway is divided into two main phases (Fig. 1). The preparatory, or upper phase of glycolysis (GLK to TPI) involves an initial investment of ATP in order to ultimately generate ATP and NADH in the payoff, or lower phase (GAPDH to PK). Gluconeogenesis is not divided into preparatory or payoff phases since all of the reactions in the pathway are either at equilibrium (freely reversible) or require energy, but can be divided into upper and lower phases. The production of metabolites (such as pyruvate for fatty acid biosynthesis) not only involves carbon flux, but also is intimately connected with cellular energetics. In order to prevent the wasteful consumption of ATP, cells must achieve a balance between the supply of biosynthetic precursors and their energy budget through careful regulation of the competing pathways of glycolysis and gluconeogenesis. The key regulatory steps in glycolysis (Fig. 1) are catalyzed by glucokinase (GLK), phosphofructokinase (ATP-PFK, PP_i-PFK), and pyruvate kinase (PK), which are reversed by the committed enzymes of gluconeogenesis. The first committed reaction of gluconeogenesis is the conversion of pyruvate to phosphoenolpyruvate (PEP, Fig. 1). This bypass can be accomplished in several ways in different organisms. In mammals, gluconeogenesis is initiated from pyruvate by the combined activities of pyruvate carboxylase (PC) and phosphoenolpyruvate carboxykinase (PEPCK). In plants, gluconeogenesis at the first bypass is initiated with the intermediate oxaloacetate (during seed germination) and therefore requires only PEPCK [47]. Alternatively, plants use a pyruvate phosphate dikinase (PPDK) to directly initiate gluconeogenesis [47]. Finally, bacteria and archaea are known to use both PPDK and phosphoenolpyruvate synthase (PEPS) in gluconeogenesis [48]. Clearly, there is considerable flexibility in the organization of the first bypass across the domains of life. In addition to some regulatory capacity over the direction of the glycolysis/gluconeogenesis pathway, these enzymes may govern flux into other pathways since they exist at the intersection of

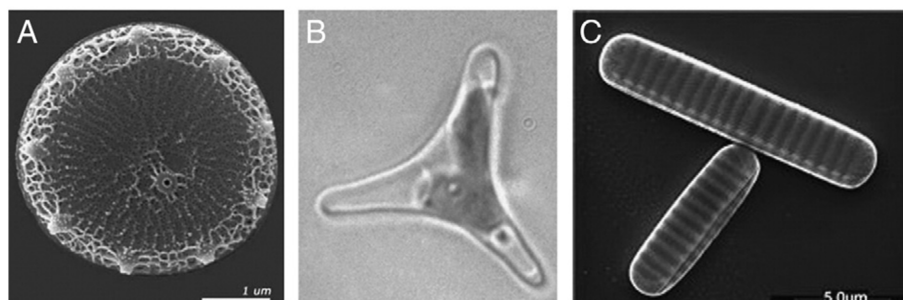


Fig. 3. Diatom species with sequenced genomes. A) The multipolar centric diatom *Thalassiosira pseudonana* has a cosmopolitan distribution and a relatively small genome (34Mb). B) The raphid pennate diatom *Phaeodactylum tricornutum* (30Mb) is not naturally abundant, but is a good model organism for laboratory studies in diatom physiology, morphogenesis, and silicification. C) The raphid pennate diatom *Fragilariopsis cylindrus* is abundant in polar regions and is a model organism for studying tolerance to extreme environmental conditions. Images A&B are provided courtesy of the DOE Joint Genome Institute and image C is provided courtesy of Henrik lange and Gerhard Dieckmann, Alfred-Wegener Institute for Polar and Marine Research, Germany.

glycolysis/gluconeogenesis with the TCA cycle, fatty acid biosynthesis, and amino acid metabolism. Since they may have a variety of different and variable roles in metabolism, defining them as glycolysis bypass enzymes could be misleading. Therefore, the direct conversion enzymes, along with PEPC, PEPCK, PC, MDH, and ME, which also involve the interconversion of 3-carbon and 4-carbon intermediates, will here be referred to as the pyruvate hub instead of the first bypass (Fig. 1).

The second bypass step (Fig. 1) in the glycolysis direction (fructose 6 phosphate to fructose 1, 6 bisphosphate) can be catalyzed by either an ATP-dependent phosphofructokinase (ATP-PFK) or a pyrophosphate-dependent phosphofructokinase (PP_i-PFK). The gluconeogenic reaction (Fig. 1) is catalyzed by fructose 1,6 bisphosphatase (FBP). In eukaryotes, the second bypass is reciprocally regulated by a potent allosteric effector molecule, fructose 2,6 bisphosphate (Fru 2,6 bisP). Typically, Fru 2,6 bisP activates glycolytic ATP-PFK while acting as a competitive inhibitor of gluconeogenic FBP [49]. Interestingly, plant ATP-PFK is insensitive to Fru 2,6 bisP, and PP_i-PFK is responsive instead [50]. In diatoms, it is unknown whether Fru 2,6 bisP activates ATP-PFK or PP_i-PFK. Fru 2,6 bisP is produced and degraded from the glycolytic intermediate fructose 6-phosphate by the activity of a bi-functional 6-phosphofructo-2-kinase/fructose 2,6 bisphosphatase (PF2K/F2BP). The only known function of Fru 2,6 bisP is to act as an allosteric effector of the second bypass, and therefore the concentration of Fru 2,6 bisP is an important determinant of the direction of carbon flux into either glucose oxidation or carbohydrate synthesis.

The final unidirectional glycolysis step is catalyzed by glucokinase (GLK). However, a glucose-6-phosphatase, the gluconeogenic enzyme which would reverse the activity of GLK, has not been identified in diatoms [23]. Thus, there is no third bypass in diatoms, presumably since glucose in its non-phosphorylated form is not an important metabolite in these organisms [23].

3.2. The conservation of diatom carbon metabolism genes

Examination of the three diatom genomes identified a total of 164 genes encoding enzymes with putative roles in carbon metabolism pathways, with 51 of these genes found in *T. pseudonana*, 55 in *P. tricornutum*, and 58 in *F. cylindrus* (Table 1). Genes were clustered according to sequence similarity to identify homologs (Table S1), and phylogenies were constructed to validate monophyletic clusters of diatom genes. Monophyletic clusters were given a gene name, and were considered to be orthologs (inherited from the most recent common ancestor) if present in at least two diatom genomes. There were 40 orthologs identified in all three diatom genomes which comprised the core carbon partitioning proteome (Fig. 4). The majority of the carbon partitioning enzymes in a given diatom genome belong to this core proteome with 78%, 73%, and 69% in *T. pseudonana*, *P. tricornutum*, and *F. cylindrus* respectively. Most of the enzymes that belong to the core proteome (33/40, or 83%) were predicted to be targeted to the same sub-cellular location, a finding that lends some support to the accuracy of the localization predictions.

In the core proteome, 7 of the 40 orthologs (17%) were found to not share targeting predictions (Fig. 4). In most of these cases, it is not clear whether this was because the enzymes are truly differently targeted, or whether there were inaccuracies in the prediction software or incorrect gene models. For example, the glycolysis enzyme phosphoglycerate mutase 5 (PGAM5) is predicted to be targeted to the chloroplast in both *T. pseudonana* and *P. tricornutum*, whereas it does not have a predicted chloroplast localization in *F. cylindrus* (Table S1). Both N-termini of the *T. pseudonana* and *P. tricornutum* models are supported by EST data, however the *F. cylindrus* model is likely erroneously predicted owing to the presence of a 666 nucleotide intron directly following the predicted start codon. Despite the possibility that the correct ORFs were not used to make the targeting predictions, there are two cases where gene models that are well-supported by EST data in *T. pseudonana* and *P. tricornutum* give different targeting predictions. These include one of the enolase orthologs (ENO2) and pyruvate phosphate dikinase (PPDK), which are both predicted to be chloroplast-localized in the pennate diatoms but not in *T. pseudonana*. Since these enzymes catalyze two sequential reactions of the glycolysis/gluconeogenesis pathway, they may represent a translocation of a portion of this metabolic pathway following the divergence of centric and pennate diatoms, and are an example of how core metabolism has been rearranged in diatoms. This metabolic rearrangement will be discussed more in later sections.

Carbon partitioning genes that are not common to all diatom genomes are specific examples of enzymatic steps that have been modified following the diversification of diatoms and may confer some metabolic adaptation specialized for a given species or class. These modifications include species or class-specific duplications (in-paralogs), the selective maintenance or deletion of orthologs, and variable horizontal acquisition of foreign genes. A significant portion (23–31%) of genes in any given genome is either unique to that species or found only in one other diatom genome, indicating that carbon metabolic pathways are not static in diatoms and that there have been several adjustments throughout the evolution of modern species (Table 1, Fig. 4). The next several sections will explore both the common features of the organization and regulation of diatom carbon partitioning metabolic pathways as well as the features of these pathways that appear to be more flexible within the realm of diatom diversity represented by the sequenced genomes.

3.3. Enzymes of carbon metabolism exist as several isozymes

Many of the enzymes involved in carbon metabolism exist as a number of isozymes in the three diatom genomes (Table 1). Diatoms have on average twice as many enzymes involved in glycolysis/gluconeogenesis than were found in the *C. reinhardtii* genome, and additional analysis suggests that this feature is conserved in other green algal genomes (Table 1, and unpublished observations). A possible explanation for this is that unlike green algae, diatoms are secondary endosymbionts, and could have acquired many of these additional isozymes from endosymbiotic gene transfer (EGT). However it is also possible that the additional isozymes found in both the core and accessory proteomes of diatoms have arisen through duplications or through horizontal gene transfer events (HGT). Regardless of the origin of these additional enzymes, many seem to have been maintained in diatoms at least since the divergence of centrics and pennates, indicating that they are useful and may be functionally differentiated (for example GPI). In multicellular organisms, variations in isozyme form and function may be useful at different times during development or in differentiated tissues [51]. In unicellular organisms, different isozyme forms may be useful under certain environmental conditions or may be targeted to different sub-cellular location or organelles. The high number of isozymes found in these pathways in diatoms suggests flexibility or optimization in regulating carbon metabolism.

3.4. The organization of carbon partitioning in diatoms

Isozymes of the glycolysis/gluconeogenesis pathways and the pyruvate hub are predicted to be distributed across several sub-cellular compartments, including the plastid, periplastid compartment, cytosol, and mitochondria. There are several conserved features of this organization that enable generalizations regarding the organization of carbon metabolism in diatoms (Fig. 5). First, all three diatoms have conserved orthologs of the lower half of the glycolysis pathway (TPI to PK) that are predicted to be targeted to the mitochondria [22,23]. Second, only a partial cytosolic glycolysis pathway is conserved in all three diatoms (Fig. 5). Specifically, orthologs could be identified for the upper half and final step of the cytosolic glycolysis pathway, but not for the mid-payoff phase (reactions catalyzed by PGK, PGAM, and ENO). Several

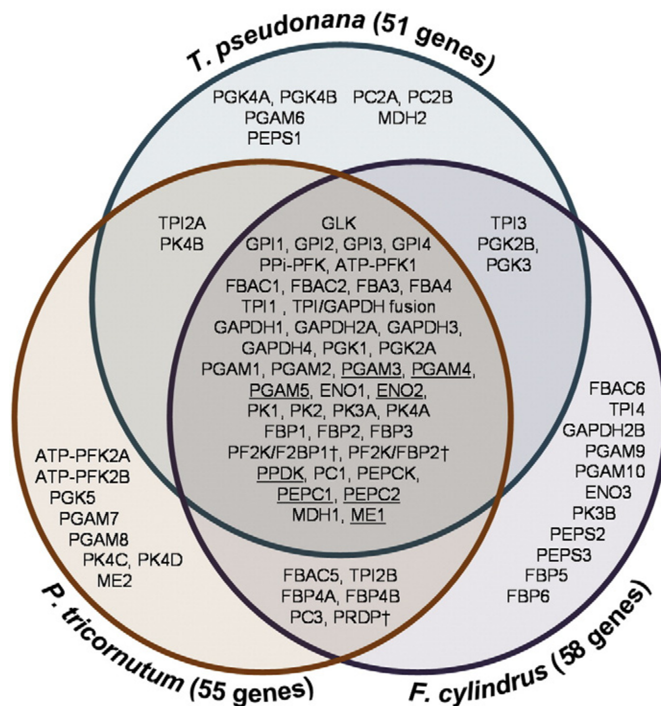


Fig. 4. Venn diagram of glycolysis, gluconeogenesis, or pyruvate hub enzymes showing the proportion of shared genes (the core proteome), genes found in two genomes, and unique genes in the genomes of *T. pseudonana*, *P. tricornutum*, and *F. cylindrus*. See Table 1 for enzyme identifications. † Regulatory enzyme, not included in the tallies for carbon partitioning enzymes. Underlined enzymes are orthologs for which the targeting predictions are not shared. See text and Table S1 for details.

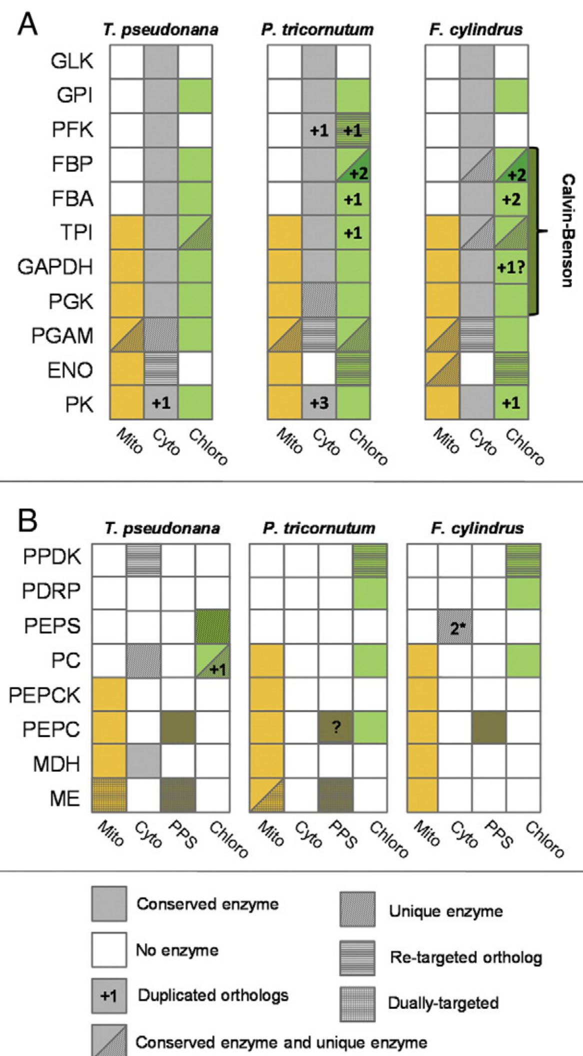


Fig. 5. The compartmentation of A) glycolysis/gluconeogenesis and B) pyruvate hub enzymes in diatoms. Filled boxes indicate the presence of an enzyme targeted to the mitochondria (orange), cytosol (grey), chloroplast (green), or periplastid compartment (brown). White boxes designate the absence of an enzyme. All enzymes are referenced in Table S1. Orthologs are indicated by solid colors while unique enzymes (found only in a single diatom genome) are indicated by diagonal stripes. Horizontal stripes show enzymes with orthologs that have been re-targeted in all three diatom genomes. Cross-hatching denotes possible targeting of an enzyme to two different intracellular locations (dual targeting). The numbers indicate the number of additional enzymes identified per compartment, when the number is on a solid box it indicates a duplication event. When the number is on a horizontal striped box it indicates an enzyme that was either horizontally acquired or selectively maintained. The 2* in *F. cylindrus* PEPS indicates two unique enzymes.

gluconeogenesis and regulatory enzymes are conserved in diatoms, particularly at the second bypass. Finally, there are several unifying features in the organization of pyruvate metabolism that illustrate the role of the mitochondria as a regulatory hub for the distribution of intracellular organic carbon, but there are many features that suggest that the pyruvate hub is a metabolic network that has been subject to modification as diatoms diversified.

3.4.1. Metabolism in the mitochondria

Mitochondrial glycolysis is currently believed to occur only in diatoms and non-photosynthetic oomycetes, which remain relatively understudied with respect to metabolism as compared to mitochondria in conventional model organisms [22]. Consequently, there is little known about the origin or function of this pathway [23,24]. The first enzyme of the diatom Embden–Meyerhoff–Parnas (EMP) mitochondrial glycolysis pathway is a TPI-GAPDH fusion protein, which has been experimentally localized to mitochondria in *P. tricornutum* with immunogold labeling [24]. Orthologs of this TPI-GAPDH fusion, along with enzymes that catalyze the complete lower phase of glycolysis (PGK to PK) predicted to be mitochondrially targeted were identified in all three diatoms with a remarkable degree of conservation (Fig. 5). There have been very few duplication events, substitutions, or variable acquisitions of enzymes involved in mitochondrial glycolysis.

Interestingly, slightly upstream of the TPI-GAPDH fusion, on the opposite strand in the reverse orientation, a duplicated mitochondrially targeted GAPDH (GAPDH1) is found in an arrangement that is strikingly well conserved in all three diatom genomes. The conservation of this gene order is consistent with their coordinated regulation using a bidirectional promoter in the intergenic region [52]. This arrangement is apparently found only in diatoms, since a TPI-GAPDH fusion protein was not identified in the genome of the related photosynthetic heterokont *Aureococcus anophagefferens*, and though the TPI-GAPDH fusion protein is also found in oomycetes like *Phytophthora sojae* (PID: 285408, JGI v3.0), there was no adjacent GAPDH found in an inverted orientation. Taken together, it seems likely diatoms have a unique way to coordinately regulate mitochondrial TPI-GAPDH and GAPDH, and that both of the triose phosphate metabolic intermediates DHAP and GAP can be a starting point for mitochondrial glycolysis in diatoms.

The next steps of glycolysis are catalyzed by phosphoglycerate kinase and phosphoglycerate mutase. A single mitochondrially targeted isoform of phosphoglycerate kinase (PGK1) was identified in all three diatom genomes and is phylogenetically related to PGKs from other heterokonts (Fig. 6). In contrast, there are several phosphoglycerate mutases (PGAM) that are putatively targeted to the mitochondria, however, only one (PGAM1) is conserved in all three diatoms and is of the expected length (approx. 288aa or 32 kDa). PGAM1 sequences are most closely related to the chloroplast-localized PGAM2. PGAM1 and PGAM2 are apparently distantly related to other diatom PGAM

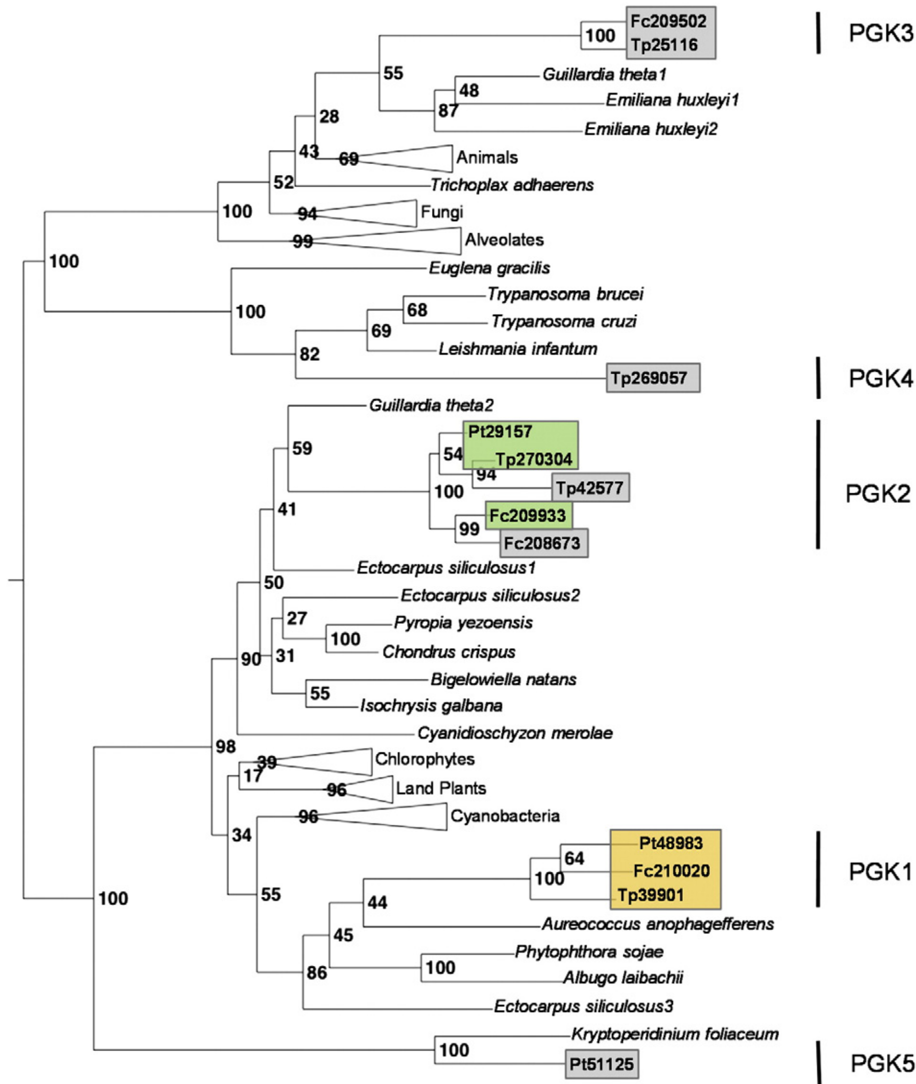


Fig. 6. Maximum likelihood tree of PGK isozymes. The phylogenetic tree was generated using RAxML-HPC BlackBox (CIPRES Science Gateway) from 64 PGK sequences. The tree has been mid-point rooted. Diatom sequences discussed in this study are shown boxed and in bold font. Filled boxes indicate the presence of an enzyme targeted to the mitochondria (orange), cytosol (grey), chloroplast (green). Bootstrap values are labeled at each node and are presented in %. Sequence accession numbers are listed in Table S2.

orthologs (PGAM3–PGAM10), some of which have predicted mitochondrial-targeting (Table S1). Since several of the gene models for the remaining PGAM predicted proteins appear to be partial or incorrect, and they are larger than the typical PGAM subunit, these ORFs may just be part of proteins of unknown function that include the catalytic activity of PGAM but may not be directly involved in the glycolysis pathway (Table S1).

The final two steps of the glycolysis pathway are catalyzed by enolase (ENO) and pyruvate kinase (PK). A mitochondrial ortholog of enolase (ENO1) was identified in all three diatom genomes. ENO1 is similar to sequences both from other heterokonts and the chromalveolate *Emiliania huxleyi* and is similar to the other orthologous enolase (ENO2), which is predicted to be cytosolic in *T. pseudonana* but chloroplast-targeted in the pennates. A second mitochondrially targeted enolase (ENO3) was identified only in the genome of *F. cylindrus*, and this enzyme is more highly divergent from ENO1 and ENO2 (Table S1). The final step of glycolysis, catalyzed by pyruvate kinase (PK) is the only unidirectional reaction of the mitochondrial glycolysis pathway. Pyruvate kinase is a key enzyme in the regulation of carbon metabolism and is known to possess a wide range of kinetic and regulatory properties [53]. Two mitochondrial PKs (PK1, PK2) were found to be conserved and orthologous in all three diatom genomes. Both diatom mitochondrial PKs belong to a family of PKs that are conserved within heterokonts. This heterokont group of PKs is most closely related to enzymes from metazoans than the other PKs found in diatom genomes suggesting that they may be ancient eukaryotic enzymes.

Major questions remain concerning both the origin and function of mitochondrial glycolysis. It is currently unknown whether mitochondrial glycolysis is an ancient trait or a trait derived from gene transfer after the acquisition of a plastid [22]. The finding that non-photosynthetic oomycetes also have mitochondrial glycolysis seems to support the idea that mitochondrial glycolysis is an ancient eukaryotic trait. However, since oomycetes have several genes that suggest they may have harbored a transient plastid, it's possible that enzymes acquired via EGT were targeted to the mitochondria prior to plastid loss and the diversification of diatoms [14]. Therefore, without more rigorous phylogenetic analysis, the origin of mitochondrial glycolysis remains unclear.

There are several possible reasons for and consequences of a complete mitochondrial glycolysis payoff phase in diatoms. In plants, cytoplasmic glycolysis enzymes have been reported to be physically associated with mitochondria under conditions of high respiration to facilitate the channeling of pyruvate to the TCA cycle [54,55]. By internalizing the lower glycolysis pathway into the mitochondria,

diatoms would not need to transport pyruvate to supply the TCA cycle since pyruvate would be produced within the target organelle. Intermediates for the mitochondrial glycolysis pathway would need to be supplied from the cytosol by a triose phosphate transporter that has yet to be identified. Alternatively, the supply of GAP for mitochondrial glycolysis may be predominantly from a recently discovered mitochondrial Entner–Doudoroff (ED) pathway [30]. Neither intracellular transporters nor diatom ED glycolysis are well-characterized, and it is presently unclear what the relative importance of each supply route would be. Another implication of an internalized mitochondrial glycolysis pathway is that NADH is produced within the organelle rather than in the cytosol. Typically, NADH must be imported into the mitochondria via the malate–aspartate shuttle so that the reducing power can be used to drive oxidative phosphorylation [56]. By translocating the production of NADH into the mitochondria, diatoms would no longer require these enzymes for this purpose. Finally, glycolate produced from photorespiration can be enzymatically converted to glycerate in diatom mitochondria where it is believed to then be imported to the chloroplast, phosphorylated, and can re-enter the Calvin–Benson cycle [23]. Recently, a chloroplast-localized glycerate kinase was identified [30]. However, if a mitochondrial glycerate kinase could be identified, it's possible that this photorespiration by-product could then directly enter mitochondrial glycolysis and the TCA cycle and be used to either drive energy generation or to replenish TCA cycle intermediates.

The malate–aspartate shuttle involves specialized enzymes and transporters, but the key enzyme is malate dehydrogenase (MDH), which is part of the pyruvate hub (Fig. 1). Mitochondrial MDH is an essential TCA cycle enzyme, and not surprisingly, orthologs for mitochondrial MDH were conserved in diatom genomes (MDH1). However, the malate–aspartate shuttle requires a cytosolic MDH, which could be identified in the *T. pseudonana* genome (MDH2) but was not found in the genome of either pennate diatom. Therefore, it appears that pennate diatoms are unable to import reducing equivalents into the mitochondria from the cytosol in this manner while the centric *T. pseudonana* may possess this capacity. This finding supports the idea that pennate diatoms do not need to import reducing equivalents (NADH) since they are produced in the mitochondrial matrix.

Like most pyruvate hub enzymes, there are several possible metabolic roles for MDH in addition to its essential role as a TCA cycle intermediate, and in the malate–aspartate shuttle. The OAA produced by MDH1 in mitochondria could be used to supply mitochondrial gluconeogenesis, the next step of which would be catalyzed by PEPCK. In mammals, PEPCK is known to be the rate-controlling step of gluconeogenesis and is under transcriptional control [57]. Alternatively, in C4 plants, the decarboxylating activity of PEPCK is also part of a biochemical carbon concentrating mechanism [56]. The exact role of PEPCK in diatoms is not clear; however it exists as a single-copy ortholog in all three diatom genomes and is strongly predicted to be mitochondrial (Table S1). Based on its mitochondrial localization, PEPCK is most likely a gluconeogenesis enzyme (rather than a C4-type enzyme) in diatoms. If PEPCK is a gluconeogenesis enzyme, it could use the OAA produced either by MDH1, or by mitochondrial pyruvate carboxylase (PC). Though there are several isozymes of PC in diatom genomes, only one mitochondrial isoform was found (PC3), but was absent in the genome of *T. pseudonana*. Therefore, pennate diatoms have the capacity to initiate gluconeogenesis from pyruvate through the combined activities of PC and PEPCK in the mitochondria, while centric diatoms may be restricted to initiating gluconeogenesis from OAA in this organelle.

Under normal growth conditions, the metabolic intermediates used to initiate gluconeogenesis in the mitochondria are typically drawn away from the TCA cycle for biosynthetic reactions, such as the synthesis of amino acids. These intermediates can be replenished by activity of anapleurotic enzymes, including mitochondrial PC and PEPC. All three diatoms have two PEPC, and one isozyme (PEPC2) is strongly predicted to be mitochondrially localized, indicating its most likely role is anapleurotic (Table S1). The other isozyme, PEPC1, has high scores predicting mitochondrial localization, but is also predicted to be localized to the periplastid compartment (Table S1). In both PEPC1 and PEPC2, there is some ambiguity about whether the correct gene models are predicted, and it may be possible that there are several transcript isoforms originating from the same locus (Table S1). Although it is tempting to speculate on the role of PEPC1 and PEPC2 in diatom metabolism, any significance cannot be ascribed without more confidence in the sub-cellular localization of these enzymes.

The final enzyme conserved in the mitochondria of diatoms is malic enzyme (ME1). The role of diatom ME is currently unknown, but it could participate in both pyruvate metabolism and C4-type photosynthesis. In all three diatoms, ME1 has high predicted mitochondrial localization; however, a signal peptide that targets ME1 to the periplastid compartment was detected in *T. pseudonana* (Table S1). *P. tricornutum* has an additional ME (ME2) for which there is no apparent ortholog in the other diatoms though there are closely related enzymes found in oomycetes and *Ectocarpus*, suggesting it is at least partially conserved in heterokonts. The additional *P. tricornutum* ME (ME2), like *T. pseudonana* ME1 has predicted dual targeting to the mitochondria and the periplastid compartment.

Although there is compelling bioinformatics evidence as well as experimental evidence to support the existence of a mitochondrial glycolysis pathway in diatoms, these predictions should be validated experimentally [23,24]. However, if it is assumed that the mitochondrial glycolysis pathway will be demonstrated in diatoms, more questions will remain to be answered. For example, what is the magnitude of carbon flux through this pathway and what is the nature of the carbon intermediates and transporters that supply this pathway? Also, what proportion of the pyruvate produced in this pathway is catabolized completely to make ATP or to make 4C skeletons for other biosynthetic processes? Since the enzymes of the mitochondrial glycolysis pathway catalyze reversible reactions, it is possible that this pathway is used to run gluconeogenesis, allowing the mitochondria to export triose phosphates. The complement of pyruvate hub enzymes found in diatom mitochondria includes those capable of both anapleurotic and key glycolysis bypass (gluconeogenic) reactions, so this is theoretically possible. Running this pathway in the gluconeogenic direction would require carbon intermediates, which could theoretically be supplied from mitochondrial fatty acid oxidation or photorespiration [18]. In sum, if the manipulation of carbon flux in diatoms is proposed, it must be done with a better idea of the importance of this compartmentalized pathway relative to other metabolically compartmentalized pathways in contributing to the production, destruction, or recycling of fatty acids.

3.4.2. Distribution of carbon partitioning enzymes between the cytosol and chloroplast

Cytosolic orthologs were identified for all steps in the upper phase of the glycolysis/gluconeogenesis pathways (GLK to TPI), including those involved in the second bypass (PFK and FBP). All three diatoms possess an orthologous copy of a cytosol-localized ATP-PFK. Interestingly, the *P. tricornutum* genome contains two additional copies of ATP-PFK (ATP-PFK2A, ATP-PFK2B), which are arranged adjacent to one another on chromosome 29 and share a high degree of sequence similarity (89%), suggesting that they arose through a relatively recent in-paralogous duplication. ATP-PFK2B from *P. tricornutum* is putatively targeted to the chloroplast (Table S1), whereas the other two diatoms apparently lack a chloroplast-localized PFK despite possessing near-complete plastidic glycolysis (Fig. 4). This suggests that *P. tricornutum* has an increased capacity for hexose metabolism in the chloroplast relative to *T. pseudonana* and *F. cylindrus*.

In a deviation from classical EMP glycolysis, all diatom genomes also encode an alternative enzyme to ATP-PFK that utilizes pyrophosphate as the phosphoryl donor (PP_i-PFK; Table 1, Table S1). This glycolysis variant is thought to be energetically advantageous, since ATP is not

consumed in the preparatory phase [58]. This strategy may be particularly important in anaerobic conditions, when the generation of ATP from the combined actions of the TCA cycle and oxidative phosphorylation ceases [58–60]; for example, amitochondriate protists and anaerobic bacteria both lack ATP-PFK and instead rely the PP_i-dependent enzyme [9]. The presence of PP_i-PFK is widespread in photosynthetic organisms [61]; in terrestrial plants PP_i-PFK activity is more responsive to regulatory controls than ATP-PFK and therefore is thought to be a method of adjusting plant carbon metabolism under changing environmental conditions [62]. However, diatom PP_i-PFK enzymes contain the bacterial PRK06555 domain (ProtClustDB ID 417822) and share similarity to several bacterial sequences (Table S1), suggesting that PP_i-PFK may have been acquired horizontally from bacteria at some point prior to the divergence of diatoms. The high conservation of PP_i-PFK amino acid sequence among the diatoms is consistent with this enzyme playing an important role as a regulatory step in central carbon metabolism. As one of the key regulatory steps in glycolysis, PFK is a potentially promising target for genetic manipulation. However, the relative importance of ATP-PFK and PP_i-PFK in regulating diatom intracellular carbon flux is currently unknown.

FBP, at the second bypass step, is an important site of regulation to supply carbon for carbohydrate synthesis. Previous FBP phylogenies in diatoms suggest a bacterial origin for cytosolic FBP [63,64], while chloroplast-localized FBPs are thought to have arisen from an ancient duplication of a cytosolic FBP that was redirected to the plastid [43,64]. In the chloroplast, these enzymes have been co-opted to function in carbon fixation and are involved in the regeneration of ribulose 1,5 bisphosphate in the Calvin–Benson cycle [65]. Additional chloroplast-localized FBP genes were acquired by diatoms via the red algal endosymbiont [66].

The cytosolic form of FBP is well conserved among diatoms; each diatom genome contains a single orthologous copy (FBP3; Fig. 4). The *F. cylindrus* genome contains an additional cytosolic FBP (FBP6) that is highly divergent from other diatom FBPs (<30% similar), suggesting that it is evolving rapidly. Multiple chloroplast-localized FBPs were identified in each diatom genome; two orthologous groups were conserved among the three diatoms (FBP1, 2). Localization studies using GFP fusions to FBP N-terminal bipartite pre-sequences in *P. tricornutum* confirmed the plastid localization of these enzymes [43]. *P. tricornutum* and *F. cylindrus* have two additional orthologous FBPs (FBP4) that are absent in *T. pseudonana*, and *F. cylindrus* has another unique FBP (FBP5). Interestingly, in both the *P. tricornutum* and *F. cylindrus* genome, the FBP4 genes have high sequence similarity (>60%) and neighbor one other on the same chromosome/scaffold, suggesting that perhaps they are a result of a duplication event early in pennate evolution.

The widespread duplication and diversification of chloroplast targeted FBPs observed in pennates (especially in *F. cylindrus*) may confer an advantage to this lineage by allowing the cell to quickly adjust its rate of carbon fixation to accommodate its metabolic demands. Alternatively, the multiple copies may have different activities or substrate affinities, and therefore be optimal only under certain conditions. For example, FBP catalyzes a similar reaction to the enzyme sedoheptulose-1,7-bisphosphatase (SBP). A plastid-targeted SBP is apparently absent in diatom chloroplasts, and it is possible that a plastidic FBP could serve as a substitute for SBP in the reductive pentose phosphate pathway [43]. An additional possibility is that certain FBPs function as dedicated Calvin–Benson enzymes associated with a sub-chloroplast location (e.g. the pyrenoid [67]), while other FBPs are located elsewhere in the chloroplast where they may participate in plastidal gluconeogenesis.

In addition to FBP, the enzymes FBA, TPI, GAPDH, and PGK could potentially play dual roles in the Calvin–Benson cycle if targeted to the plastid. Localization studies have demonstrated pyrenoid targeting for FBA isozymes in *P. tricornutum* [44]. Although all diatom genomes have multiple copies of TPI targeted to the chloroplast, they only have a single chloroplast-targeted ortholog of GAPDH. Therefore, this GAPDH likely has an essential role in the Calvin–Benson cycle and is probably not involved in plastidial glycolysis under photosynthetic conditions. Although diatoms have a nearly complete complement of genes that would permit the upper phase of glycolysis/gluconeogenesis in the plastid, several of those genes may be partially or exclusively involved in carbon fixation.

In contrast to the upper phase glycolysis/gluconeogenesis pathway, in which orthologs are remarkably well-conserved, the enzymes that catalyze the 7th–9th steps of cytosolic glycolysis or the "mid-payoff phase" seem to have been subjected to deletion and retargeting throughout the diversification of diatoms. First, though all diatoms seem to have homologs of PGAM and PGK predicted to be localized to the cytosol, there are no conserved orthologs found in all three diatom genomes in this space. Furthermore, of the three diatoms, only *T. pseudonana* appears to have a full cytosolic glycolysis pathway while both *P. tricornutum* and *F. cylindrus* apparently lack cytosolic isoforms of enolase (ENO).

A more careful analysis reveals that there are several versions of cytosolic PGK and PGAM found in diatom genomes. Only mitochondrial and chloroplast versions of PGAM were identified with confidence and no obvious candidates for cytosolic PGAMs were found (See Section 3.4.1). While it is possible that diatoms have cytosolic PGAM activity, they have not maintained conserved orthologs of this enzyme in the cytoplasm.

There is a surprising amount of diversity within diatom PGK, and the acquisition of several PGK genes from a variety of sources was revealed through phylogenetic analysis (Fig. 6). The distribution of PGK among the diatoms genomes highlights the ability of these organisms to selectively duplicate, re-target and/or delete acquired genes, presumably in order to optimize metabolic processes. For example, although each diatom genome encodes for a single plastid-localized PGK (PGK2A), *T. pseudonana* and *F. cylindrus* both contain an additional PGK with high sequence similarity but without chloroplast targeting information (PGK2B). No PGK2B was found in *P. tricornutum*. The PGK2 group clusters with sequences from red algae and secondary endosymbionts from the red algal lineage, indicating that theses enzymes likely originated via EGT from the acquisition of a red algal plastid. Curiously, a sequence from the chlorachniophyte *Bigelowiella natans* (a green algae-derived secondary endosymbiont) also clusters with this group, the significance of which is difficult to interpret without a more comprehensive understanding of the evolutionary origin of the *B. natans* PGK. Additional cytosolic PGKs were found in *T. pseudonana* and *F. cylindrus* (PGK3) that belong to clade comprised of eukaryotic organisms, including chromalveolates and opisthokonts (Fig. 6). These two groups are distantly related, and it is possible that this enzyme is an ancient isoform, perhaps a remnant from the secondary exosymbiont, which has been lost in *P. tricornutum*. Interestingly, *P. tricornutum* does have a unique version of PGK (PGK5), which is most similar to a sequence from the dinoflagellate *Kryptoperidinium foliaceum*. This dinoflagellate is known to maintain a tertiary plastid derived from a diatom endosymbiont, and since neither *T. pseudonana* nor *F. cylindrus* has a PGK5, the diatom endosymbiont found in *K. foliaceum* is likely related to a group of diatoms including *P. tricornutum* [68]. Finally, *T. pseudonana* has a unique isoform of PGK (PGK4) that is not found in the other diatom genomes. PGK4 clades with a well-supported cluster with sequences from excavates, an unrelated group to diatoms or any of their endosymbionts. It's possible that *T. pseudonana* acquired this gene horizontally or that it is a metabolic relic from deep in eukaryotic evolutionary history. Despite the inability to determine the absolute origins of these distinct enzyme types, it is clear that cytosolic PGKs, unlike their highly conserved mitochondrial homologs, have not been conserved throughout the diversification of diatoms (Fig. 6). Whether these isozymes are functional equivalents, or have kinetic or regulatory differences that confer an adaptive advantage to one diatom or another is not clear without functional characterization.

Finally, the only diatom with a predicted cytosolic enolase is *T. pseudonana* (ENO2). In pennates, a signal peptide and chloroplast

transit peptide were detected for this protein. The predicted ENO2 model in *T. pseudonana* was validated with 5′ RACE (unpublished results) to eliminate the possibility that the model was missing an upstream exon with a targeting motif. The localization of ENO2 in the chloroplast should be validated experimentally in the pennate diatoms. If confirmed, the presence of the penultimate glycolysis step either in the cytoplasm of centrics or in the chloroplast of pennates represents a significant difference in the organization of metabolism between the two major diatom lineages.

The lack of conservation in the mid-payoff phase (Fig. 5A) suggests that cytosolic glycolysis is not an essential energy-producing pathway in diatoms. Though the pennates are apparently missing a full cytosolic glycolysis pathway, they have a more complete plastid-localized pathway that *T. pseudonana* seems to lack and this variability in pathway organization has major implications for how the regulation of the direction of carbon flux may be different in these major groups.

Diatoms have several isozymes of PK, which catalyzes the final and unidirectional step of the payoff phase of glycolysis (Table 1, Fig. 1). The mitochondrial isoforms have already been discussed (PK1, PK2, Section 3.4.1), however there are additional PKs found in both the chloroplast (PK3) and the cytosol (PK4). Both *T. pseudonana* and *P. tricornutum* have a single chloroplast PK3 ortholog, while *F. cylindrus* has two (Table S1). Each diatom has a different number of cytosolic PK4s, *T. pseudonana* has two, *P. tricornutum* has 4, and *F. cylindrus* has a single copy. It is unclear why the number of cytosolic PKs varies among species, but one explanation is that PK4 may have been duplicated and deleted independently several times during diatom evolution. Supporting this idea is evidence for a recent duplication event in *P. tricornutum* that gave rise to two in-paralogs (homologs arising from a within-species duplication) on chromosome 8 (data not shown).

The pyruvate produced by PK3 in the chloroplast or PK4 in the cytosol enters the pyruvate metabolic hub. Compared with the mitochondria, there are relatively few enzymes of the pyruvate hub found in the chloroplast and cytosol of diatoms. Only a single chloroplast enzyme (PC1) was found in all three diatom genomes. Though *T. pseudonana* is apparently missing a mitochondrial PC3, there are two apparently recently duplicated copies of a PC2 found only in the centric diatom. One of the PC2 copies has predicted chloroplast localization, while the other copy is apparently cytosolic (Table S1). Typically, PC is a mitochondrial or cytosolic enzyme involved in gluconeogenesis or an anapleurotic role, and little is known about its role in chloroplasts. Since PEPCK is only predicted to be mitochondrial (Section 3.4.1) it is unlikely that plastid PC is functioning to initiate gluconeogenesis, since there are no other known enzymes capable of converting OAA to PEP found in this compartment. It is more likely that the OAA produced in chloroplasts is used in other biosynthetic pathways. In *E. huxleyi*, the transcription of chloroplast PC has been shown to be light-regulated and has been proposed to be involved in the production of OAA as a precursor for the biosynthesis of amino acids [69]. OAA can also be produced by PEPC, and as discussed previously (Section 3.4.1), all three diatoms have a PEPC2 with some predicted periplastid compartment/chloroplast localization. In C4 plants, PEPC has a role in a biochemical carbon concentrating mechanism to facilitate the delivery of CO_2 to RuBisCO in photosynthesis [70]. If PEPC1 is in fact periplastid localized in diatoms, it could only function in this capacity provided there is an enzyme near RuBisCO that can decarboxylate OAA. Though there is some experimental evidence to suggest that diatoms may have a biochemical carbon concentrating mechanism, the localization of a decarboxylating enzyme near RuBisCO has not been demonstrated conclusively [23,70].

The only non-mitochondrial MDH that could be identified in diatoms was the cytosolic MDH2 from *T. pseudonana*, indicating that diatoms apparently lack the ability to convert chloroplast OAA to malate in either the plastid or the periplastid compartment. However, as mentioned previously (Section 3.4.1), both *T. pseudonana* and *P. tricornutum* have isozymes of ME (ME1, and ME2 respectively) that are possibly targeted to the periplastid compartment, suggesting that malate may be either imported or produced by still unknown enzymes in this space. The decarboxylation of malate by ME in this compartment would produce CO_2, pyruvate, and NAD(P)H, and if the malate is imported from the mitochondria this reaction could effectively serve both as a biochemical CCM and a conduit for the precursors of fatty acid biosynthesis. However, ME1 in *T. pseudonana* and ME2 in *P. tricornutum* also have high predicted targeting to the mitochondria and they may be incorrectly predicted to the periplastid compartment. Because of the potentially important role in intracellular carbon flux, validating the sub-cellular localization of these enzymes should be a priority for the field.

Another enzyme of the pyruvate hub, PPDK, is found as a single copy in all three diatom genomes. Though diatom PPDK was determined to be monophyletic, the sequence from *F. cylindrus* appears to be more highly divergent (lower sequence similarity) to the sequences from *T. pseudonana* and *P. tricornutum*, suggesting it may be experiencing a higher rate of mutation. PPDK is phylogenetically limited to prokaryotes, protists (including diatoms), some fungi, and green plants, but its function can be quite variable [71]. PPDK is capable of acting bidirectionally, by replacing the activity of PK in the final step of glycolysis or by initiating gluconeogenesis [72]. Anaerobic bacteria utilize a pyrophosphate-dependent variant of glycolysis in which PPDK acts as an alternative to PK in the final glycolysis step. In C4 plants the role of PPDK is specialized, as it resupplies PEP in the stroma of leaf-mesophyll cell chloroplasts, however it is believed that PPDK first became functionally seated in C3 plants, where it has a role in balancing the flux of carbon through glycolysis or gluconeogenesis, and was only slightly modified to achieve a new function in C4 plants [71]. Interestingly, just like the ENO2, PPDK is predicted to be localized to the plastid in both pennate diatoms, while the PPDK from *T. pseudonana* is cytosolic. Whether the differential localization of PPDK between centrics and pennates indicates some functional distinction, or is connected to the translocation of ENO2, remains unclear

The final enzyme of the pyruvate hub is PEPS. PEPS is found mostly in prokaryotes where it is believed to catalyze the conversion of pyruvate to PEP in the gluconeogenic direction [48,73,74]. PEPS was identified in the *T. pseudonana* and *F. cylindrus* genomes but appeared to be absent from the *P. tricornutum* genome. The *T. pseudonana* PEPS is chloroplast-localized and is most similar to a sequence from *Cyanothece* sp. PCC 7424 (ZP_01910935.1). Both the *F. cylindrus* PEPS genes are apparently cytosolic and are most similar (but with low similarity scores) to PEPS from the heterotrophic bacteria *Plesiocystis pacifica* (ZP_01910935.1,) and *Bacillus pumilis* (ZP_03053952.1). Since none of these enzymes are found in any other heterokonts or related eukaryotes, the most parsimonious explanation for the occurrence of PEPS in these diatom genomes is that they have been acquired horizontally following the evolutionary divergence of all three diatoms investigated.

The specific functions of PPDK and PEPS in diatoms remain unclear, but they are undoubtedly important enzymes in the distribution of carbon intermediates in the pyruvate hub. Very little is known about the role of PEPS in eukaryotes though its role in bacteria as a regulatory enzyme is more well characterized [48]. It has been proposed that plastidic PPDK in C3 plants may have an important role in supplying PEP for the biosynthesis of aromatic amino acids [75]. If it has this role in diatoms, it may be an important target for down-regulation since it represents a sink for pyruvate (a fatty acid biosynthesis precursor) though there may be detrimental effects on growth by inhibiting aromatic amino acid production.

Overall, there are some generalizations about diatom carbon partitioning in the chloroplast and cytosol that emerge from this analysis that have implications for the way intracellular carbon flux is regulated. First, diatoms do not share a conserved upper half or preparatory phase of glycolysis in chloroplasts, which means that diatom plastids (with the exception of *P. tricornutum*) are not metabolically equipped

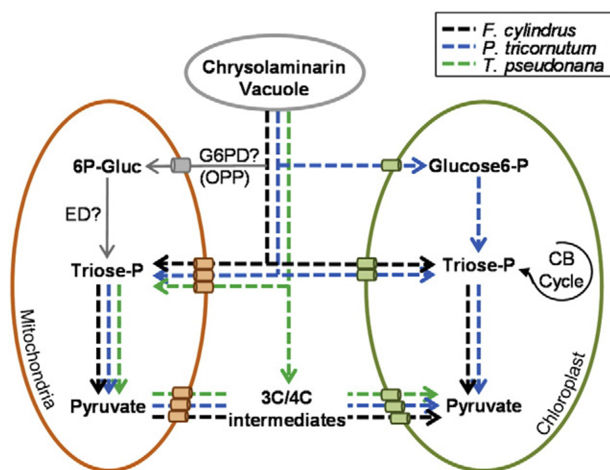


Fig. 7. Schematic diagram of the compartmentation of glycolytic flux in *T. pseudonana*, *P. tricornutum*, and *F. cylindrus*. Arrows follow the direction of glycolytic flux from cytosolic Glucose 6-P (produced by degradation of chrysolaminarin) to oxidized 3C or 4C metabolic intermediates. The pathway organization for each diatom is indicated by a different color (see key). All diatoms have a cytosolic preparatory phase and a mitochondrial payoff phase of classical Emden–Meyerhoff–Parnas glycolysis. Only *T. pseudonana* has a complete cytosolic pathway and only the pennate diatoms have a chloroplast-localized payoff phase. All compartments are shown with unknown transporters. Gray arrows indicate a possible pathway from cytosolic Glucose 6-P into mitochondrial glycolysis via the activity of the first enzyme of the oxidative pentose phosphate pathway (OPP), which is G6PD (glucose 6-P dehydrogenase, EC 1.1.1.49). To enter the mitochondrial Entner–Doudoroff glycolysis pathway, the 6-phosphoglucono-δ-lactone (6P-Gluc) produced in the cytosol would have to be imported by an unknown transporter.

for the catabolism of hexoses (Fig. 7). This is not surprising, since unlike terrestrial plants and green algae, which store and break down starch inside the plastid, diatom chrysolaminarin vacuoles (where diatoms store carbohydrate) are extra-plastidial. This is supported by the conservation of the upper half of glycolysis in the cytosol of diatoms. In diatoms, fatty acids are synthesized de novo in the chloroplast by a type II fatty acyl synthase, meaning that pyruvate must be either produced in or imported to this organelle [18]. Since pennate diatoms have isozymes for the lower half of glycolysis targeted to the plastid, they are theoretically capable of producing pyruvate in the chloroplast from triose phosphates that are either produced during photosynthesis or imported from the cytosol. In contrast, *T. pseudonana* apparently does not have a chloroplast-localized lower half of glycolysis meaning that in order for organic carbon fixed during photosynthesis to be incorporated into fatty acids it must first be exported from the chloroplast. Therefore, in *T. pseudonana* (and possibly other centric diatoms), the transporters that export triose phosphates and import organic carbon precursors like oxaloacetate may be an important site of carbon flux regulation into other fatty acid supply pathways (Fig. 7). In pennates the regulation may also include a fine level of regulatory control within the plastid, at the import of glucose 6-P (in *P. tricornutum*) or triose phosphates. As a result of these analyses, it is now clear that there are some significant differences between the centric and pennate diatom lineages with respect to the organization of the carbon partitioning pathways in the cytosol and the chloroplast, and these differences have implications for the way intracellular carbon flux is regulated (Fig. 7).

The most notable difference with respect to the organization of the pyruvate metabolic hub between centrics and pennates is that the capacity to initiate gluconeogenesis (by bypassing PK and catalyzing the pyruvate → PEP reaction) has not been conserved in the plastid or cytosol in diatoms. This may indicate that 1) gluconeogenesis is mostly initiated from pyruvate (or OAA) in the mitochondria, 2) there has been significant adaptability of this carbon flux regulatory node during the diversification of diatoms, or 3) a combination of both. The adaptability of this carbon flux distribution node is supported by finding that the orthologous PPDK has been re-targeted in centrics and pennates and that PEPS seems to have been acquired horizontally independently in *T. pseudonana*, and *P. tricornutum*. The extent to which these enzymes affect carbon flux distribution in a primary or accessory role cannot be predicted bioinformatically.

3.4.3. Pyrophosphate-dependent glycolysis and other EMP variants

Both PP_i-PFK and PPDK are found in the genomes of all three diatoms investigated suggesting that these organisms may utilize the PP_i-dependent variant of EMP glycolysis. As was mentioned previously, PP_i-dependent glycolysis functions in anaerobic bacteria and amitochondriate protists to confer an energetic advantage when oxidative phosphorylation cannot be used to generate cellular ATP. Though this glycolysis variant generates less ATP than the combined activities of EMP glycolysis and oxidative phosphorylation, it may be sufficient to sustain cellular energetic demands temporarily or for long periods of time when cells are in a resting stage. Diatoms are known to survive long periods (months to decades) in dark and anoxic sediment layers [76]. Recently, the respiration of intracellular nitrate stores under anoxic conditions in diatoms (dissimilatory nitrate reduction to ammonia, DNRA) has been implicated as a potential survival mechanism under these conditions as they transition to a resting state [77]. While DNRA could be occurring in diatoms, it's also possible that PP_i-dependent glycolysis may also serve to sustain cellular energetic demands during dark anoxic conditions and may be a longer-term strategy for sustaining metabolic demands throughout the resting stage.

The frequency with which EMP glycolysis variants occur in diatoms is just beginning to be appreciated. Recently, a phosphoketolase pathway was identified in *P. tricornutum*, but not in *T. pseudonana* [30]. Additionally, a mitochondrial Entner–Dourdoff (ED) glycolysis pathway was discovered in both *P. tricornutum* and *T. pseudonana* [30]. The relative importance of these glycolysis variants in diatoms is currently unknown, as is the effect that running each pathway would have on cellular energetics. What is clear is that diatoms are complex organisms and much remains to be learned about the organization and regulation of their most conserved, core metabolic pathways.

3.5. Regulation of carbon partitioning pathways

Carbon flux through glycolysis or gluconeogenesis in the cytosol must be precisely regulated to prevent futile cycling between these opposing pathways. Reciprocal regulation coordinates the simultaneous activation of one pathway and suppression of the other. The reciprocal regulation of the second bypass is accomplished by the synthesis and breakdown of the allosteric effector Fru 2,6 bisP. Fru 2,6 bisP is synthesized from the glycolytic intermediate fructose 6-phosphate by the activity of 6-phosphofructo-2-kinase (PF2K) and converted back to fructose 6-phosphate by fructose 2,6 bisphosphatase (F2BP). In animals, these reactions are catalyzed by a single polypeptide, with a kinase domain on the C-terminal end and a phosphatase domain on the N-terminal end; this bifunctional enzyme arose early in eukaryotic evolution by the fusion of these two functional units [78]. Each of the diatom genomes examined encode for two PF2K/F2BP isozymes, all of which lack signal peptide sequences and are localized to the cytosol. The PF2K/F2BP enzymes form two orthologous clusters (PF2K/F2BP 1,2); the PF2K/F2BP 2 group was not previously annotated and was identified using BLAST.

The function of these two PF2K/F2BP isozymes in respect to the control of carbon metabolism in diatoms is currently unknown, although in other eukaryotes it has been shown that different PF2K/F2BP isozymes are expressed during a change in environmental condition or developmental stage [49]. In mammals, four different isozymes are known and are differentially expressed in various tissues to maintain glucose homeostasis. In contrast, higher plants typically encode for only one PFK-2/F2BP [50]. There is evidence for the

presence of monofunctional enzymes in yeast (and in some plants), where the phosphatase domain has been inactivated due to a critical amino acid substitution in a universally conserved “RHG motif” [78]. The polypeptide still contains a vestigial phosphatase domain, but lacks the histidine amino acid required to accept a phosphate group during catalysis. Alignment of PF2K/F2BP 1 and 2 sequences with those of other eukaryotes reveals that the PF2K/F2BP 1 group contains the critical histidine amino acid, while the PF2K/F2BP 2 group has replaced it either with the non-polar amino acid alanine (*T. pseudonana*) or proline (*P. tricornutum* and *F. cylindrus*, Fig. 8). This suggests that the phosphatase domain in the PF2K/F2BP 2 homologues may not be active. Furthermore, this inactivation may have arisen independently in centric and pennate diatoms, given the difference in the interfering amino acid.

Another bifunctional kinase/phosphatase regulatory protein (PDRP) is known to govern the activity of the PPDK enzyme via light-mediated phosphorylation of the PPDK active site [79]. PDRP was discovered and initially described in C4 plants, where it works to coordinate PPDK activity with photosynthesis [75]. Chloroplast-localized PDRPs were putatively identified in the genomes of both *P. tricornutum* and *F. cylindrus* (Table S1) based on the presence of a PRK05339 domain that is also found in the PRDP from *Zea mays*. Therefore, in pennates, both PPDK and its regulatory protein appear to occur exclusively in the chloroplast. In contrast, no PDRP could be identified for *T. pseudonana*, which is consistent with the absence of a plastidal PPDK (Fig. 5). This arrangement suggests that pennate and centric diatoms may utilize these enzymes in functionally distinct capacities. As discussed in the previous section, the metabolic role of PPDK in diatoms is not entirely clear. In other photosynthetic organisms, plastid-localized PPDK plays a specialized role in C4 photosynthesis, although PPDK is also found in C3 organisms where it may work to supplement the PEP requirement for aromatic amino acid biosynthesis [75]. If PPDK performs the latter role in diatoms, it may be an important target for down-regulation since it represents a sink for pyruvate (a fatty acid biosynthesis precursor).

“RHG” Motif

Homo sapiens	RTIYLCRHGESEFNLLG
Mus musculus	RTIYLCRHGESEFNLLG
Gallus gallus	RTIYLCRHGESEYNLLG
Danio rerio	HSIYLCRHGESQHNVQG
Drosophila melanogaster	RTIYLTRHGESEYNLSG
Saccharomyces cerevisiae	RQIWITRSGESEDNVSG
Neurospora crassa	RTVWLSRHGESMLNLEG
Arabidopsis thaliana	RPILLTRHGESMDNVRG
Spinacia oleracea	RPILLTRHGESQDNVRA
Zea mays	RPILLTRHGESLHNVRG
Chlamydomonas reinhardtii	RKIFLTRHGESQYNQKG
Pt PF2K/F2BP 1	RTFYLTRHGQSEYNLLG
Fc PF2K/F2BP 1	RTFYFTRHGQSEYNLLG
Tp PF2K/F2BP 1	REFYLTRHGQSEYNLSG
Pt PF2K/F2BP 2	RPVFLCRPGQTISGILT
Fc PF2K/F2BP 2	RPILLCRPGQTLSDIIT
Tp PF2K/F2BP 2	RPIFICRAGQTMADSDR
	: . : * *:: .

Fig. 8. Alignment of PFK2/F2BP bifunctional regulatory proteins. Shows the conserved “RHG” motif in PFK2/F2BP bifunctional proteins among diverse taxa including animals, fungi, higher plants, green algae and diatoms. Both diatom PFK2/F2BP group 2 sequences and a 860 known monofunctional PFK2/F2BP sequence from *Saccharomyces cerevisiae* (with an inactivated phosphatase domain) have substituted a different amino acid at the place of a critical histidine residue, as indicated by boxes. The accession numbers for sequences used in the alignment are as follows: *Homo sapiens* (GenBank ID: NP_006203.2), *Mus musculus* (GenBank ID: NP_032851.2), *Gallus gallus* (GenBank ID: XP_417979.2), *Danio rerio* (GenBank ID: NP_957302.1), *Drosophila melanogaster* (UniProt ID: Q9Y1W3_DROME), *Saccharomyces cerevisiae* (GenBank ID: AAA34858.1), *Neurospora crassa* (GenBank ID: XP_958926.1), *Arabidopsis thaliana* (GenBank ID: AEE28077.1), *Spinacia oleracea* (UniProt ID: O64983_SPIOL), *Zea mays* (GenBank ID: AAL09471.1), *Chlamydomonas reinhardtii* (UniProt ID: A8JAE8_CHLRE). Diatom PIDs are listed in Table S1.

With an ultimate goal of identifying gene targets for modification to improve fuel precursor molecule production, it is important to consider what sorts of regulatory processes are most amenable to manipulation. Eukaryotic cells employ several mechanisms to regulate the synthesis and breakdown of carbon intermediates. Transcriptional regulation enables a short-term response and large-scale control; for example, the activity of transcription factors allows for up- or down-regulation of entire pathways of related function in response to environmental cues. The bypass points and secondary regulatory molecules previously discussed are unidirectional control points, which would be especially attractive to alter the direction of carbon flux. Most of the steps in glycolysis are bidirectional, and are regulated by mass action, allostery, or post-translational modification. Modification of these steps to improve carbon flux would be more difficult than choosing a unidirectional control point, or genes that are largely regulated by transcription. Since changes in transcription are commonly major indicators of changes in cellular metabolism, determining transcript levels for each of the steps involved in carbon metabolism would greatly contribute to our understanding of carbon flux in diatoms and other algae and would aide in the identification of interesting targets for genetic manipulation.

4. Concluding remarks

The decreasing cost and increased output of high-throughput sequencing have meant that algal genome and transcriptome sequences can now be obtained easily, and the characterization of metabolomes and proteomes is becoming commonplace. This systems biology approach is incredibly powerful, but the challenge remains to distill these data into useful information that gives insight into the functional organization of a cell or organism [80]. With the aim of developing algal strains for biofuel production, such information is essential to characterize pathways that govern the flux of carbon from the fixation of CO_2 to the de novo biosynthesis of fatty acids and TAG. For the first time, the availability of genome sequence data from three diatoms representing both major diatom lineages has allowed for a comparative analysis of the organizing principles of one of the most fundamental and conserved metabolic pathways. Through this comparative analysis, several core features of diatom metabolism emerged that distinguish diatoms from other model organisms. First, it is clear that metabolic pathways are organized into various sub-cellular compartments in diatoms, and the conservation of the cytosolic preparatory phase and mitochondrial payoff phase of glycolysis is a unifying feature. The mitochondrial glycolysis pathway is likely to be the primary supply pathway for energy generation in the TCA cycle since a cytosolic pathway is not conserved. Consequently, an important point of carbon flux regulation should be the selective transport of metabolites between the cytosol and organelles (such as the export of organic carbon in the plastid, import of organic carbon in the mitochondria). Second, several isozymes of the pyruvate hub (PEPCK, PEPC, MDH) are conserved and targeted to diatom mitochondria. The presence or absence of isozymes in each compartment determines the possible fates for fixed carbon, which has associated metabolic and energetic consequences, so it can be inferred that mitochondrial PEPCK, PEPC, and MDH have indispensable roles in diatom metabolism.

These examples are in contrast to the significant proportion of the enzymes of the core carbon partitioning pathways that is not strictly conserved in diatom genomes. The differences between lineages and species are examples of how diatoms may have modified their carbon partitioning pathways to adapt to specific environmental niches during the diversification of the group. These modifications include

the translocation of metabolic pathways (such as the payoff phase of glycolysis, which is plastid-localized in both pennates and cytosolic in *T. pseudonana*), selective maintenance and horizontal acquisition of gene families for enzymes like PGK and PEPS, and novel gene family expansions (as in the case of pennate FBPs). Further experimental work is required to determine what the physiological implications of these metabolic variations are, and additional genome sequences that span a greater diversity within the diatom family will be important to determine how significant some of these species-specific or lineage-specific modifications are in an evolutionary context. However, documenting these differences is an important first step in identifying which isozymes may confer adaptive flexibility within this algal group and should help in the interpretation and annotation of functional-omics datasets.

In summary, this analysis has characterized both conserved and variable features of central carbon metabolism pathways within a single class of algae. Characterizing the conserved features establishes a blueprint for the metabolic organization of diatoms over which functional data can be overlaid and interpreted in an ecological, evolutionary, or physiological context. Additionally, it facilitates the identification of enzymes with important roles in regulating carbon flux that are suitable targets for metabolic engineering. The variability in the organization of metabolic pathways in diatoms that was documented here illustrates how even core central pathways can be modified considerably within a single algal group. Overall, we have aimed to generate a framework using available genomic information to highlight gaps in our understanding, identify important areas for clarification, and to facilitate the interpretation of future functional-omics studies.

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.algal.2012.04.003.

Acknowledgments

These sequence data were produced by the US Department of Energy Joint Genome Institute http://www.jgi.doe.gov/ in collaboration with the user community. Diatom biofuels research in the Hildebrand lab is supported by AFOSR grant FA9550-08-1-0178, DOE grant DE-EE0001222, and NSF grant CBET-0903712. SRS was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. RA was supported by the San Diego Fellowship awarded through the UCSD Temporal Dynamics of Learning Center. The authors would like to thank Thomas Mock (University of East Anglia, UK), Ansgar Gruber, and Peter Kroth (University of Konstanz, Germany) for permission to use data acquired from the *F. cylindrus* genome sequencing project. The authors would also like to thank Juan Ugalde and Eric Allen for assistance with the construction of phylogenetic trees.

References

[1] R. Radakovits, R.E. Jinkerson, A. Darzins, M.C. Posewitz, Genetic engineering of algae for enhanced biofuel production, Eukaryotic Cell 9 (2010) 486–501.
[2] T.G. Dunahay, E.E. Jarvis, S.S. Dais, P.G. Roessler, Manipulation of microalgal lipid production using genetic engineering, Applied Biochemistry and Biotechnology 57–58 (1996) 223–231.
[3] E.T. Yu, F.J. Zendejas, P.D. Lane, S. Gaucher, B.A. Simmons, T.W. Lane, Triacylglycerol accumulation and profiling in the model diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Baccilariophyceae) during starvation, Journal of Applied Phycology 21 (2009) 669–681.
[4] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, et al., Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, The Plant Journal 54 (2008) 621–639.
[5] P. Roessler, Effects of silicon deficiency on lipid composition and metabolism in the diatom *Cyclotella cryptica*, Journal of Phycology 24 (1988) 394–400.
[6] P. Roessler, Changes in the activities of various lipid and carbohydrate biosynthetic enzymes in the diatom *Cyclotella cryptica* in response to silicon deficiency, Archives of Biochemistry and Biophysics 267 (1988) 521–528.
[7] Y. Li, D. Han, G. Hu, M. Sommerfeld, Q. Hu, Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*, Biotechnology and Bioengineering 107 (2010) 258–268.
[8] C. Givan, Evolving concepts in plant glycolysis: two centuries of progress, Biological Reviews 74 (1999) 277–309.
[9] W. Plaxton, The organization and regulation of plant glycolysis, Annual Review of Plant Physiology 47 (1996) 185–214.
[10] N. Linka, A.P.M. Weber, Intracellular metabolite transporters in plants, Molecular Plant 3 (2010) 21–53.
[11] U.I. Flügge, R.E. Häusler, F. Ludewig, M. Gierth, The role of transporters in supplying energy to plant plastids, Journal of Experimental Botany 62 (2011) 2381–2392.
[12] J.E. Lunn, Compartmentation in plant metabolism, Journal of Experimental Botany 58 (2007) 35–47.
[13] J.M. Archibald, The puzzle of plastid evolution, Current Biology 19 (2009) R81–R88.
[14] P.J. Keeling, The endosymbiotic origin, diversification and fate of plastids, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 365 (2010) 729–748.
[15] G. McFadden, Primary and secondary endosymbiosis and the origin of plastids, Journal of Phycology 37 (2001) 951–959.
[16] B. Stoebe, U.G. Maier, One, two, three: nature's tool box for building plastids, Protoplasma 219 (2002) 123–130.
[17] K. Bolte, L. Bullmann, F. Hempel, A. Bozarth, S. Zauner, U.G. Maier, Protein targeting into secondary plastids, Journal of Eukaryotic Microbiology 56 (2009) 9–15.
[18] E.V. Armbrust, J.A. Berges, C. Bowler, B.R. Green, D. Martinez, N.H. Putnam, et al., The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism, Science 306 (2004) 79–86.
[19] P.J. Lopez, J. Desclés, A.E. Allen, C. Bowler, Prospects in diatom research, Current Opinion in Biotechnology 16 (2005) 180–186.
[20] J.N. Timmis, M.A. Ayliffe, C.Y. Huang, W. Martin, Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes, Nature Reviews Genetics 5 (2004) 123–135.
[21] P.J. Keeling, J.D. Palmer, Horizontal gene transfer in eukaryotic evolution, Nature Reviews Genetics 9 (2008) 605–618.
[22] M.L. Ginger, G.I. McFadden, P.A.M. Michels, Rewiring and regulation of cross-compartmentalized metabolism in protists, Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 365 (2010) 831–845.
[23] P.G. Kroth, A. Chiovitti, A. Gruber, V. Martin-Jezequel, T. Mock, M.S. Parker, et al., A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis, PloS One 3 (2008) e1426.
[24] M. Liaud, C. Lichtl, K. Apt, W. Martin, R. Cerff, Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway, Molecular and Biological Evolution 17 (2000) 213–223.
[25] C. Bowler, A.E. Allen, J.H. Badger, J. Grimwood, K. Jabbari, A. Kuo, et al., The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes, Nature 456 (2008) 239–244.
[26] E.V. Armbrust, The life of diatoms in the world's oceans, Nature 459 (2009) 185–192.
[27] P.G. Falkowski, M.E. Katz, A.H. Knoll, A. Quigg, J.A. Raven, O. Schofield, et al., The evolution of modern eukaryotic phytoplankton, Science 305 (2004) 354–360.
[28] K. Roberts, E. Granum, R.C. Leegood, J.A. Raven, Carbon acquisition by diatoms, Photosynthesis Research 93 (2007) 79–88.
[29] M. Tachibana, A.E. Allen, S. Kikutani, Y. Endo, C. Bowler, Y. Matsuda, Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, Photosynthesis Research 109 (2011) 205–221.
[30] M. Fabris, M. Matthijs, S. Rombauts, W. Vyverman, A. Goossens, G.J.E. Baart, The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner–Doudoroff glycolytic pathway, The Plant Journal (2012).
[31] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Research 25 (1997) 3389–3402.
[32] B. Gschloessl, Y. Guermeur, J.M. Cock, HECTAR: a method to predict subcellular targeting in heterokonts, BMC Bioinformatics 9 (2008) 393.
[33] I. Small, N. Peeters, F. Legeai, C. Lurin, Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences, Proteomics 4 (2004) 1581–1590.
[34] O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, Locating proteins in the cell using TargetP, SignalP and related tools, Nature Protocols 2 (2007) 953–971.
[35] M.G. Claros, P. Vincens, Computational method to predict mitochondrially imported proteins and their targeting sequences, European Journal of Biochemistry 241 (1996) 779–786.
[36] K.E. Apt, L. Zaslavkaia, J.C. Lippmeier, M. Lang, O. Kilian, R. Wetherbee, et al., In vivo characterization of diatom multipartite plastid targeting signals, Journal of Cell Science 115 (2002) 4061–4069.
[37] O. Kilian, P. Kroth, Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids, The Plant Journal 41 (2005) 175–183.
[38] L. Franzen, J. Rochaix, G. von Heijne, Chloroplast transit peptides from the green alga *Chlamydomonas reinhardtii* share features with both mitochondrial and higher plant chloroplast presequences, FEBS Letters 260 (1990) 165–168.
[39] A. Gruber, S. Vugrinec, F. Hempel, S.B. Gould, U.G. Maier, P.G. Kroth, Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif, Plant Molecular Biology 64 (2007) 519–530.
[40] T.N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nature Methods 8 (2011) 785–786.
[41] J. Bendtsen, H. Nielsen, G. von Heijne, S. Brunak, Improved prediction of signal peptides: SignalP 3.0, Journal of Molecular Biology 340 (2004) 783–795.
[42] O. Emanuelsson, H. Nielsen, G. von Heijne, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites, Protein Science 8 (1999) 978–984.

[43] A. Gruber, T. Weber, C.R. Bártulos, S. Vugrinec, P.G. Kroth, Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms, Journal of Basic Microbiology 49 (2009) 58–72.
[44] A.E. Allen, A. Moustafa, A. Montstat, A. Eckert, P. Kroth, C. Bowler, Evolution and Functional Diversification of Fructose Bisphosphate Aldolase Genes in Photosynthetic Marine Diatoms, , 2011, pp. 1–48.
[45] N. Peeters, I. Small, Dual targeting to mitochondria and chloroplasts, Biochimica et Biophysica Acta (BBA) — Molecular Cell Research 1541 (2001) 54–63.
[46] C. Carrie, E. Giraud, J. Whelan, Protein transport in organelles: dual targeting of proteins to mitochondria and chloroplasts, FEBS Journal 276 (2009) 1187–1195.
[47] S.S. Sung, D. Xu, C.M. Galloway, C.C. Black, A reassessment of glycolysis and gluconeogenesis in higher plants, Physiologia Plantarum 72 (1988) 650–654.
[48] U. Sauer, B.J. Eikmanns, The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria, FEMS Microbiology Reviews 29 (2005) 765–794.
[49] D.A. Okar, A.J. Lange, Fructose-2,6-bisphosphate and control of carbohydrate metabolism in eukaryotes, Biofactors 10 (1999) 1–14.
[50] T.H. Nielsen, J.H. Rung, D. Villadsen, Fructose-2,6-bisphosphate: a traffic signal in plant metabolism, Trends in Plant Science 9 (2004) 556–563.
[51] L.D. Gottlieb, Conservation and duplication of isozymes in plants, Science 216 (1982) 373–380.
[52] L.D. Hurst, C. Pál, M.J. Lercher, The evolutionary dynamics of eukaryotic gene order, Nature Reviews Genetics 5 (2004) 299–310.
[53] M. Munoz, E. Ponce, Pyruvate kinase: current status of regulatory and functional properties, Comparative Biochemistry and Physiology. B 135 (2003) 197–218.
[54] P. Giege, J.L. Heazlewood, U. Roessner-Tunali, A. Millar, A.R. Fernie, C.J. Leaver, et al., Enzymes of glycolysis are functionally associated with the mitochondrion in *Arabidopsis* cells, The Plant Cell 15 (2003) 2140–2151.
[55] J.W.A. Graham, Mitochondrial Glycolysis in Plants, University of Oxford, Oxford, 2007.
[56] D. Nelson, M.M. Cox, Lehninger Principles of Biochemistry, fifth ed. W.H. Freeman, 2008.
[57] K. Chakravarty, H. Cassuto, L. Reshef, R.W. Hanson, Factors that control the tissue-specific transcription of the gene for phosphoenolpyruvate carboxykinase-C, Critical Reviews in Biochemistry and Molecular Biology 40 (2005) 129–154.
[58] E. Mertens, Pyrophosphate-dependent phosphofructokinase, an anaerobic glycolytic enzyme? FEBS Letters 285 (1991) 1–5.
[59] E. Mertens, ATP versus pyrophosphate — glycolysis revisited in parasitic protists, Parasitology Today 9 (1993) 122–126.
[60] C. Slamovits, P. Keeling, Pyruvate-phosphate dikinase of oxymonads and parabasalia and the evolution of pyrophosphate-dependent glycolysis in anaerobic eukaryotes, Eukaryotic Cell 5 (2006) 148–154.
[61] N. Carnal, C.W. Black, Phosphofructokinase activities in photosynthetic organisms — the occurrence of pyrophosphate-dependent 6-phosphofructokinase in plants and algae, Plant Physiology 71 (1983) 150–155.
[62] C. Black, L. Mustardy, S. Sung, P. Kormanik, D. Xu, N. Paz, Regulation and roles for alternative pathways of hexose metabolism in plants, Physiologia Plantarum 69 (1987) 387–394.
[63] M. Rogers, P. Keeling, Lateral transfer and recompartmentalization of Calvin cycle enzymes of plants and algae, Journal of Molecular Evolution 58 (2004) 367–375.
[64] W. Martin, A. Mustafa, K. Henze, C. Schnarrenberger, Higher-plant chloroplast and cytosolic fructose-1,6-bisphophosphatase isoenzymes: origins via duplication rather than prokaryote–eukaryote divergence, Plant Molecular Biology 32 (1996) 485–491.
[65] C.A. Raines, The Calvin cycle revisited, Photosynthesis Research 75 (2003) 1–10.
[66] R. Teich, S. Zauner, D. Baurain, H. Brinkmann, J. Petersen, Origin and distribution of Calvin cycle fructose and sedoheptulose bisphosphatases in plantae and complex algae: a single secondary origin of complex red plastids and subsequent propagation via tertiary endosymbioses, Protist 158 (2007) 263–276.
[67] K. Suss, I. Prokhorenko, K. Adler, In-situ association of Calvin cycle enzymes, ribulose-1,5-bisphosphate carboxylase oxygenase activase, ferredoxin-NADP (+) reductase, and nitrite reductase with thylakoid and pyrenoid membranes of *Chlamydomonas reinhardtii* chloroplasts as revealed by immunoelectron microscopy, Plant Physiology 107 (1995) 1387–1397.
[68] B. Imanian, P.J. Keeling, The dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping, BMC Evolutionary Biology 7 (2007).
[69] Y. Tsuji, I. Suzuki, Y. Shiraiwa, Photosynthetic carbon assimilation in the coccolithophorid *Emiliania huxleyi* (Haptophyta): evidence for the predominant operation of the C3 cycle and the contribution of B-carboxylases to the active anaplerotic reaction, Plant & Cell Physiology 50 (2009) 318–329.
[70] P.J. McGinn, F.M.M. Morel, Expression and inhibition of the carboxylating and decarboxylating enzymes in the photosynthetic C4 pathway of marine diatoms, Plant Physiology 146 (2008) 300–309.
[71] C. Chastain, C. Failing, L. Manandhar, M.A. Zimmerman, M.M. Lakner, T.H.T. Nguyen, Functional evolution of C4 pyruvate, orthophosphate dikinase, Journal of Experimental Botany 62 (2011) 3083–3091.
[72] B. Tjaden, A. Plagens, C. Dorr, B. Siebers, R. Hensel, Phosphoenolpyruvate synthetase and pyruvate, phosphate dikinase of *Thermoproteus tenax*: key pieces in the puzzle of archaeal carbohydrate metabolism, Molecular Microbiology 60 (2006) 287–298.
[73] A. Hutchins, J. Holden, M. Adams, Phosphoenolpyruvate synthetase from the hyperthermophilic Archaeon *Pyrococcus furiosus*, Journal of Bacteriology 183 (2001) 709–715.
[74] H. Imanaka, A. Yamatsu, T. Fukui, H. Atomi, T. Imanaka, Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden–Meyerhof pathway in *Thermococcus kodakarensis*, Molecular Microbiology 61 (2006) 898–909.
[75] C.J. Chastain, R. Chollet, Regulation of pyruvate, orthophosphate dikinase by ADP-/Pi-dependent reversible phosphorylation in C3 and C4 plants, Plant Physiology and Biochemistry 41 (2003) 523–532.
[76] J. Lewis, A. Harris, K. Jones, R. Edmonds, Long-term survival of marine planktonic diatoms and dinoflagellates in stored sediment samples, Journal of Plankton Research 21 (1999) 343–354.
[77] A. Kamp, D. de Beer, J.L. Nitsch, G. Lavik, P. Stief, Diatoms respire nitrate to survive dark and anoxic conditions, Proceedings of the National Academy of Sciences of the United States of America 108 (2011) 5649–5654.
[78] M. Rider, L. Bertrand, D. Vertommen, P. Michels, G. Rousseau, L. Hue, 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: head-to-head with a bifunctional enzyme that controls glycolysis, Biochemistry Journal 381 (2004) 561–579.
[79] C.J. Chastain, J.P. Fries, J.A. Vogel, C.L. Randklev, A.P. Vossen, S.K. Dittmer, et al., Pyruvate, orthophosphate dikinase in leaves and chloroplasts of C3 plants undergoes light-/dark-induced reversible phosphorylation, Plant Physiology 128 (2002) 1368–1378.
[80] F.J. Bruggeman, H.V. Westerhoff, The nature of systems biology, Trends in Microbiology 15 (2007) 45–50.