Check for updates

New Phytologist

Research

The genome of a nonphotosynthetic diatom provides insights into the metabolic shift to heterotrophy and constraints on the loss of photosynthesis

Anastasiia Onyshchenko, Wade R. Roberts, Elizabeth C. Ruck, Jeffrey A. Lewis and Andrew J. Alverson

Department of Biological Sciences, University of Arkansas, 1 University of Arkansas, Fayetteville, AR 72701, USA

Author for correspondence:
Andrew J. Alverson
Email: aja@uark.edu

Received: *8 February 2021*
Accepted: *3 August 2021*

New Phytologist (2021)
doi: 10.1111/nph.17673

Key words: apochloritic, β-ketoadipate pathway, Entner–Doudoroff pathway, glycolysis, lignin, mixotrophy, oxidative pentose phosphate pathway, photosynthesis.

Summary

- Although most of the tens of thousands of diatom species are photoautotrophs, a small number of heterotrophic species no longer photosynthesize. We sequenced the genome of a nonphotosynthetic diatom, *Nitzschia* Nitz4, to determine how carbon metabolism was altered in the wake of this trophic shift.
- *Nitzschia* Nitz4 has retained its plastid and plastid genome, but changes associated with the transition to heterotrophy were cellular-wide and included losses of photosynthesis-related genes from the nuclear and plastid genomes, elimination of isoprenoid biosynthesis in the plastid, and remodeling of mitochondrial glycolysis to maximize adenosine triphosphte (ATP) yield. The genome contains a β-ketoadipate pathway that may allow *Nitzschia* Nitz4 to metabolize lignin-derived compounds.
- Diatom plastids lack an oxidative pentose phosphate pathway (oPPP), leaving photosynthesis as the primary source of NADPH to support essential biosynthetic pathways in the plastid and, by extension, limiting available sources of NADPH in nonphotosynthetic plastids.
- The genome revealed similarities between nonphotosynthetic diatoms and apicomplexan parasites for provisioning NADPH in their plastids and highlighted the ancestral absence of a plastid oPPP as a potentially important constraint on loss of photosynthesis, a hypothesis supported by the higher frequency of transitions to parasitism or heterotrophy in lineages that have a plastid oPPP.

Introduction

Microbial eukaryotes (protists) capture the full breadth of eukaryotic diversity and so exhibit a wide range of trophic strategies that include parasitism, autotrophy, heterotrophy, and mixotrophy, i.e. the ability to acquire carbon and nutrients through both autotrophic and heterotrophic means (Burki *et al.*, 2020). Although the bioenergetics and genes involved in photosynthesis have been well characterized (Falkowski & Raven, 2013), the modes and mechanisms of heterotrophy and mixotrophy vary widely among lineages and, as a result, are more poorly understood (Worden *et al.*, 2015). Many protist lineages (e.g. diatoms and green algae) include a mix of autotrophic, mixotrophic, and/or heterotrophic species (Worden *et al.*, 2015; Bock *et al.*, 2021). Studies focused on evolutionary transitions between nutritional strategies in these groups provide a powerful approach to understanding the ecological drivers, metabolic properties, and genomic underpinnings of different trophic modes in eukaryotes.

Diatoms are photosynthetic algae that account for some 20% of global net primary production (Field *et al.*, 1998). Some diatoms have been shown to supplement their growth through uptake of organic carbon from the environment, making them mixotrophic (Lewin & Lewin, 1967; Hellebust, 1971; Cerón García *et al.*, 2006; Tuchman *et al.*, 2006). Like many plants and algae, some diatoms have lost their ability to photosynthesize (Lewin & Lewin, 1967), but unlike some lineages, such as flowering plants and red algae that have lost photosynthesis many times (Goff *et al.*, 1996; Barkman *et al.*, 2007), loss of photosynthesis has occurred rarely in diatoms, with just two known losses that have spawned only a few dozen species (Frankovich *et al.*, 2018; Onyshchenko *et al.*, 2019). These 'apochloritic' diatoms maintain colorless plastids with genomes that are devoid of photosynthesis-related genes (Kamikawa *et al.*, 2015; Onyshchenko *et al.*, 2019), but their plastids continue to recruit dozens of nuclear-encoded proteins to service indispensable biosynthetic pathways that remain in the plastid (Kamikawa *et al.*, 2015, 2017; Moog *et al.*, 2020). Retention of the plastid following loss of photosynthesis is not unique to diatoms but is a general pattern in photosynthetic eukaryotes (Hadariová *et al.*, 2018) – even in those rare cases where the plastid genome has been lost (Molina *et al.*, 2014). This highlights the near impossibility of restructuring carbon metabolism in a way that does not include the deeply

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

entrenched cellular compartmentalization that occurred during the establishment of the plastid as an organelle. In addition to understanding why the plastid is retained following loss of photosynthesis, it is important to understand what facilitates or constrains loss of photosynthesis to begin with, and whether such factors vary among the major groups of photosynthetic eukaryotes.

Photosynthesis is central to current models of carbon metabolism in diatoms, where metabolic pathways are highly compartmentalized and where compartmentalization varies among species (Smith *et al.*, 2012). For example, diatom plastids house the Calvin cycle, which is involved in carbon fixation and supplies triose phosphates to anabolic pathways such as biosynthesis of branched-chain amino acids, aromatic amino acids, and fatty acids (Smith *et al.*, 2012; Kamikawa *et al.*, 2017). The lower payoff phase of glycolysis, which generates pyruvate, adenosine triphosphate (ATP), and NADH, takes place in the mitochondrion (Smith *et al.*, 2012). The centric diatom, *Cyclotella nana* (formerly *Thalassiosira pseudonana*; Alverson *et al.*, 2011), has a full glycolytic pathway in the cytosol, but compartmentalization of glycolysis differs in the pennate diatom, *Phaeodactylum tricornutum*, which lacks a full glycolytic pathway in the cytosol and carries out the lower phase of glycolysis in both the mitochondrion and the plastid (Smith *et al.*, 2012). Compartmentalization helps diatoms avoid futile cycles, which occur when two metabolic pathways work in opposite directions within the same compartment (Smith *et al.*, 2012). All of this requires careful orchestration of protein targeting within the cell, as most carbon metabolism genes are encoded in the nuclear genome (Kroth *et al.*, 2008; Smith *et al.*, 2012). Unlike lineages with primary plastids (Archaeplastida), the secondary plastids of diatoms lack an oxidative pentose phosphate pathway (oPPP) in their plastids (Michels *et al.*, 2005; Gruber *et al.*, 2009), leaving photosystem I (PSI) as the primary source of endogenous NADPH for numerous biosynthetic pathways that take place in diatom plastids.

We sequenced the genome of a nonphotosynthetic diatom in the genus *Nitzschia* and used the genome to determine how central carbon metabolism was restructured in diatoms that no longer photosynthesize. Based on our findings, we hypothesize that loss of photosynthesis is an easier transition for lineages with a plastid oPPP, and the ancestral presence or absence of a plastid oPPP may play a determinant role in how frequently photosynthesis is lost in different lineages of photosynthetic eukaryotes.

Materials and Methods

Collection and culturing

We collected diatoms from Whiskey Creek, a mangrove-lined waterway in Dr Von D. Mizell-Eula Johnson State Park, FL, USA (Onyshchenko *et al.*, 2019). Colorless diatoms were isolated with a Pasteur pipette, and clonal cultures were grown in the dark at 21°C on agar plates made with L1+NPM medium (Guillard, 1960; Guillard & Hargraves, 1993) and 1% Penicillin–Streptomycin–Neomycin solution to retard bacterial growth. We also acquired a photosynthetic *Nitzschia* (strain CCMP2144) from the National Center for Marine Algae and Microbiota and grew it in L1 medium on a 12 h : 12 h, light : dark cycle.

DNA and RNA sequencing

We sequenced the genome and transcriptome of a single clonal isolate, *Nitzschia* sp. Nitz4. Cells were removed from agar plates by rinsing with L1 medium, centrifuged, then disrupted with a BeadBeater (BioSpec Products, Bartlesville, OK, USA). Total DNA was extracted with a Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). RNA was extracted from *Nitzschia* Nitz4 and *Nitzschia* CCMP2144 with a Qiagen RNeasy Kit. Paired-end, 90 bp reads were sequenced using an Illumina HiSeq2000 (Illumina, San Diego, CA, USA) for 500-bp DNA and 300-bp RNA libraries.

Genome and transcriptome assembly

A detailed summary of the genome assembly protocol is available in Supporting Information Methods S1. Briefly, we used Ace (Sheikhizadeh & de Ridder, 2015) to correct sequencing errors, Trimmomatic (Bolger *et al.*, 2014) to trim and filter reads, SPAdes (Bankevich *et al.*, 2012) to assemble the genome, and a combination of BlobTools (Laetsch & Blaxter, 2017) and the Kmer Analysis Toolkit (Mapleson *et al.*, 2017) to identify and remove assembly contaminants (Methods S1). Contigs were scaffolded using Rascaf (Song *et al.*, 2016) and Sspace (Boetzer *et al.*, 2011), followed by contig extension and gap filling with GapCloser (Luo *et al.*, 2012).

Transcriptome assembly for *Nitzschia* Nitz4 and *Nitzschia* CCMP2144 followed Parks *et al.* (2018). To facilitate comparisons with the published transcriptome of another nonphotosynthetic diatom, *Nitzschia* NIES-3581, we downloaded sequence reads from National Center for Biotechnology Information (NCBI) BioProject PRJDB5503 and followed the trimming and assembly methods of Kamikawa *et al.* (2017).

Genome annotation

A detailed summary of the genome annotation protocol is available in Methods S1. We used Maker (Cantarel *et al.*, 2008) to identify protein-coding genes in the nuclear genome, using the assembled *Nitzschia* Nitz4 transcriptome and the proteome of *Fragilariopsis cylindrus* as evidence to inform Maker gene predictions. We additionally trained and included Augustus (Stanke *et al.*, 2008) and Snap (Korf, 2004) models for *ab initio* gene prediction. In our search for genes related to carbon metabolism, we manually annotated several genes that were not in the final set of Maker predictions. Functional information was added to the predicted gene set by searching the SwissProt and UniProt Reference Proteomes databases with Ncbi-Blastp (Camacho *et al.*, 2009). Protein domains were identified through InterProScan searches (Jones *et al.*, 2014) against the Pfam (El-Gebali *et al.*, 2019) and PANTHER (Thomas *et al.*, 2003) databases. We used Tmhmm (Krogh *et al.*, 2001) to identify transmembrane helices and characterized coding and noncoding RNAs using a combination of

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

tRNAscan-SE (Chan & Lowe, 2019), RNAmmer (Lagesen *et al.*, 2007), and Infernal (Nawrocki & Eddy, 2013) against the Rfam database (Kalvari *et al.*, 2021).

Estimation of heterozygosity

Genome-wide heterozygosity was estimated by mapping trimmed reads to the genome with Bwa-mem (Li & Durbin, 2009) and calling variants with the Genome Analysis Toolkit (Gatk) HaplotypeCaller tool (DePristo *et al.*, 2011; Poplin *et al.*, 2018). We filtered variants according to Gatk's Best Practices and by approximate read depth. Filtered variants were then evaluated using Gatk's VariantEval tool and annotated with SnpEff (Cingolani *et al.*, 2012). Additional details are provided in Methods S1.

Gene family evolution

We used OrthoFinder (Emms & Kelly, 2019) to construct clusters of orthologous protein-coding genes and infer the species phylogeny, using Mafft (Katoh & Standley, 2013) for multiple sequence alignment and FastTree (Price *et al.*, 2010) for gene tree reconstruction. Sequencing errors, allelic variants, and splice isoforms can result in transcriptome assemblies that contain many more transcripts than genes in the genome (Ono *et al.*, 2015). The large number of redundant transcripts in the transcriptomes of *Nitzschia* CCMP2144 and *Nitzschia* NIES-3581 led to artifactual patterns of gene family expansion and contraction (Supporting Information Fig. S1), so our analyses of gene family evolution only included species with sequenced genomes. To more precisely characterize gene presence/absence in *Nitzschia* Nitz4, we ran a separate OrthoFinder analysis that included transcriptomes from *Nitzschia* CCMP2144, *Nitzschia* NIES-3581, and an additional outgroup, *Bolidomonas pacifica* (Kessenich *et al.*, 2014).

We used Cafe (Han *et al.*, 2013) to characterize patterns of gene family evolution across 6092 gene families identified by OrthoFinder, excluding gene families where copy numbers varied by > 100 between species or that were not present in the last common ancestor of diatoms. Cafe uses a phylogenetic tree, gene family sizes, and a birth–death model to identify gene families and branches on the species tree that have experienced statistically significant, rather than stochastic, changes in gene family size (Hahn *et al.*, 2005). We included error models that accommodated errors in genome assembly and annotation. The birth–death rate (λ) was estimated both in a global model with a single parameter as well as a three-rate model with separate parameters for the outgroups, *Nitzschia* Nitz4, and other diatoms. The three-rate λ model provided the best fit to the data. Rapidly expanding or contracting gene families were considered significant at the default $P < 0.01$.

Characterization of carbon metabolism genes

We manually annotated and characterized genes with known roles in carbon metabolism. To do so, we began with a set of core carbon metabolic genes from diatoms (Smith *et al.*, 2012; Kamikawa *et al.*, 2017) as search terms to download related genes from NCBI's nr database (release 223.0 or 224.0), filtered to include RefSeq accessions only. We expanded our searches as necessary to characterize each gene in major carbon metabolic pathways. The Transporter Classification Database (TCDB; Saier *et al.*, 2014) was used to identify carbon transporters, and the Carbohydrate-Active Enzymes (CAZy) database (Lombard *et al.*, 2014) was used to identify genes related to saccharide degradation.

Prediction of protein localization

We used SignalP (Petersen *et al.*, 2011), ASAFind (Gruber *et al.*, 2015), ChloroP (Emanuelsson *et al.*, 1999), TargetP (Emanuelsson *et al.*, 2007), MitoProt (Claros, 1995), and Hectar (Gschloessl *et al.*, 2008) to predict whether proteins were targeted to the plastid, mitochondrion, cytoplasm, or endoplasmic reticulum (ER) for the secretory pathway. Following Traller *et al.* (2016), proteins that SignalP, ASAFind, and Hectar predicted to contain a plastid signal peptide were classified as plastid-targeted. Proteins predicted as mitochondrial-targeted by any two of TargetP, MitoProt, and Hectar were classified as localized to the mitochondrion. Proteins in the remaining set were classified as ER-targeted if they had ER signal peptides predicted by SignalP and Hectar.

Results and Discussion

Nitzschia sp. strain Nitz4

We collected and isolated *Nitzschia* Nitz4 (see micrographs in Supporting Information Fig. S2) from a shallow, warm, mangrove-lined coastal waterway in the south-eastern United States. Our sample included near-surface plankton, submerged (< 1 m) sand, and nearshore unsubmerged sand. Based on our collection, previous reports of these species in near-surface plankton (Blackburn *et al.*, 2009), and the high degree of active motility we observed in our culture, it is likely that *Nitzschia* Nitz4 is tychoplanktonic, i.e. living principally in the benthos but circulating in the shallow plankton. Previous phylogenetic analyses placed *Nitzschia* Nitz4 within a larger clade of nonphotosynthetic *Nitzschia* (Onyshchenko *et al.*, 2019). *Nitzschia* Nitz4 grew in the presence of antibiotics for approximately 1 yr, but the strain experienced gradual size diminution, which is characteristic of diatoms (Edlund & Stoermer, 1997), and ultimately died after failing to reproduce sexually and reconstitute its maximal size. This suggests that, like other members of the Bacillariales (Levialdi Ghiron *et al.*, 2008), *Nitzschia* Nitz4 was dioecious.

Plastid genome reduction

As described previously (Onyshchenko *et al.*, 2019), the plastid genome of *Nitzschia* Nitz4 experienced near-complete loss of intergenic DNA and wholesale loss of genes encoding the photosynthetic apparatus (Supporting Information Table S1). The plastid genome of *Nitzschia* Nitz4 contains just 90 of the 122

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

core genes present in photosynthetic diatoms (Ruck *et al.*, 2014). The plastid genome also experienced a substantial downward shift in GC content, from > 30% in photosynthetic species to 22.4% in *Nitzschia* Nitz4 (Table S1), a common pattern in streamlined bacterial genomes (Giovannoni *et al.*, 2014) that may reflect mutational bias (Hershberg & Petrov, 2010) or nitrogen frugality in response to its limitation (GC base pairs contain more nitrogen than AT base pairs; Dufresne *et al.*, 2005).

Nuclear genome characteristics

The genome was assembled into 1015 scaffolds totaling 27.6 Mb in length and an average coverage depth of 221 reads. The assembly was highly contiguous: the largest scaffold was 479 884 bp in length, and half the genome was contained in 88 scaffolds, each one longer than 91 206 bp (assembly N50; Table S2). These assembly statistics compare favorably to other diatom genomes, including ones sequenced with long-read technologies (Table S2). The *Nitzschia* genome is similar in GC content (48%) to other diatom genomes, indicating that the decrease in GC content experienced by the plastid genome (Table S1) did not extend to its nuclear and mitochondrial (Guillory *et al.*, 2018) genomes.

We identified 9464 gene models and 14 alternative isoforms, for a total of 9478 predicted protein genes in *Nitzschia* Nitz4 (Table S2). Most gene models (9274) were supported by transcriptome or protein similarity evidence. To estimate genome completeness, we used BUSCO to count the number of conserved protein-coding genes in the genome (Simão *et al.*, 2015). Although *Nitzschia* Nitz4 has the fewest genes of all diatom genomes sequenced to date, its BUSCO count matches or exceeds other diatom genomes (Table S2). Most of the predicted proteins (86%) had significant BLAST hits (*e*-value < 1e-6) to SwissProt or UniProt, and 66% contained significant INTERPROSCAN hits (*e*-value < 1e-5) to protein domains in the Pfam or PANTHER databases.

We characterized single nucleotide polymorphisms (SNPs) and small structural variants in the genome of *Nitzschia* Nitz4, full details of which are provided in Table S3. The estimated genome-wide heterozygosity was 0.0028 and included 13 410 synonymous and 9958 nonsynonymous SNPs in protein-coding genes (Table S3). *Nitzschia* Nitz4 had a higher proportion of nonsynonymous to synonymous substitutions (0.74) than was reported for the model diatom, *Phaeodactylum tricornutum* (0.28–0.43) (Krasovec *et al.*, 2019; Rastogi *et al.*, 2020), and the green alga, *Ostreococcus tauri* (0.2) (Blanc-Mathieu *et al.*, 2017).

Gene family evolution and signatures of heterotrophy

The complete genome sequence of *Nitzschia* Nitz4 allowed us to characterize gene family expansions, contractions, and losses related to the transition to heterotrophy (Fig. 1). Ortholog clustering showed that most genes in *Nitzschia* Nitz4 were shared with other diatoms, with 8285 (87.5%) of the *Nitzschia* Nitz4 genes assigned to orthogroups that included other diatoms. The remaining 1179 genes were unique to *Nitzschia* Nitz4, and most of these (860) were singletons. The number of Nitz4-specific genes decreased to 374 when the transcriptomes of *Nitzschia* NIES-3581 and *Nitzschia* CCMP2144 were considered. This number (374) is still probably an overestimate of the unique gene content given the greater recovery of genes in the *Nitzschia* Nitz4 genome (99% stramenopile BUSCOs) compared to the *Nitzschia* NIES-3581 transcriptome (89% stramenopile BUSCOs) (Table S2).

Although the plastid genome experienced the greatest reduction in gene content, the nuclear genome also experienced losses of genes that are otherwise highly conserved across diatoms, based on their shared presence across most or all of the other taxa in our analysis (Fig. 1). Among the 20 rapidly evolving gene families in *Nitzschia* Nitz4 identified by CAFE, most of these (15/20) were contractions, and 10 of the contractions involved gene families that were fully and uniquely lost in *Nitzschia* Nitz4 (Fig. 1b). Many of the gene families lost in *Nitzschia* Nitz4 are related directly or indirectly to photosynthesis. For example, *Nitzschia* Nitz4 is devoid of gene families associated with pigmentation and light harvesting for photosynthesis (Fig. 1b). Calcium is an important regulator of photosynthesis in plants and algae (Hochmal *et al.*, 2015), and the upregulation of plastid-localized calmodulin-dependent protein kinases at low CO_2 in *C. nana* suggested a role for calcium-dependent signaling in the regulation of photosynthesis in diatoms as well (Clement *et al.*, 2017). Loss of the calmodulin-lysine *N*-methyltransferase family of proteins in *Nitzschia* Nitz4 (Fig. 1b) suggests that its function is photosynthesis-related in diatoms and, consequently, no longer necessary in the absence of photosynthesis. *Nitzschia* Nitz4 also lost a family of protein translocases that import nuclear-encoded proteins into the plastid (Fig. 1b), as well as serine proteinase (Fig. 1b), a protein that facilitates turnover of photodamaged photosystem II (PSII) subunits in the thylakoid membrane (Nagao *et al.*, 2012). Finally, although not highlighted by the CAFE analysis because it involved loss of a single-copy gene, *Nitzschia* Nitz4 no longer has a 2-keto-3-deoxyphosphogluconate aldolase (EDA) gene, which altered mitochondrial glycolysis. Possible causes and consequences of this loss are discussed in the section entitled, 'Loss of the mitochondrial Entner–Doudoroff glycolytic pathway'.

Central carbon metabolism

Carbon metabolism in heterotrophs Classically, the major pathways for carbon metabolism in heterotrophic organisms include glycolysis, the oPPP, and the tricarboxylic acid (TCA, or Krebs) cycle. Glycolysis uses glucose (or other sugars that can be converted to glycolytic intermediates) to generate pyruvate and ATP. Pyruvate serves as a key 'hub' metabolite that can be used for amino acid biosynthesis, anaerobically fermented to lactate or ethanol, or converted into another hub metabolite, acetyl- coenzyme A (acetyl-CoA). Acetyl-CoA can then be used for fatty acid biosynthesis or can be further oxidized via the Krebs cycle to generate ATP, reducing equivalents (FADH2 and NADH), and certain amino acid precursors. The oPPP runs parallel to glycolysis, but unlike glycolysis, is largely anabolic, generating ribose-5-phosphate (a nucleotide precursor) and the key reducing agent,

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

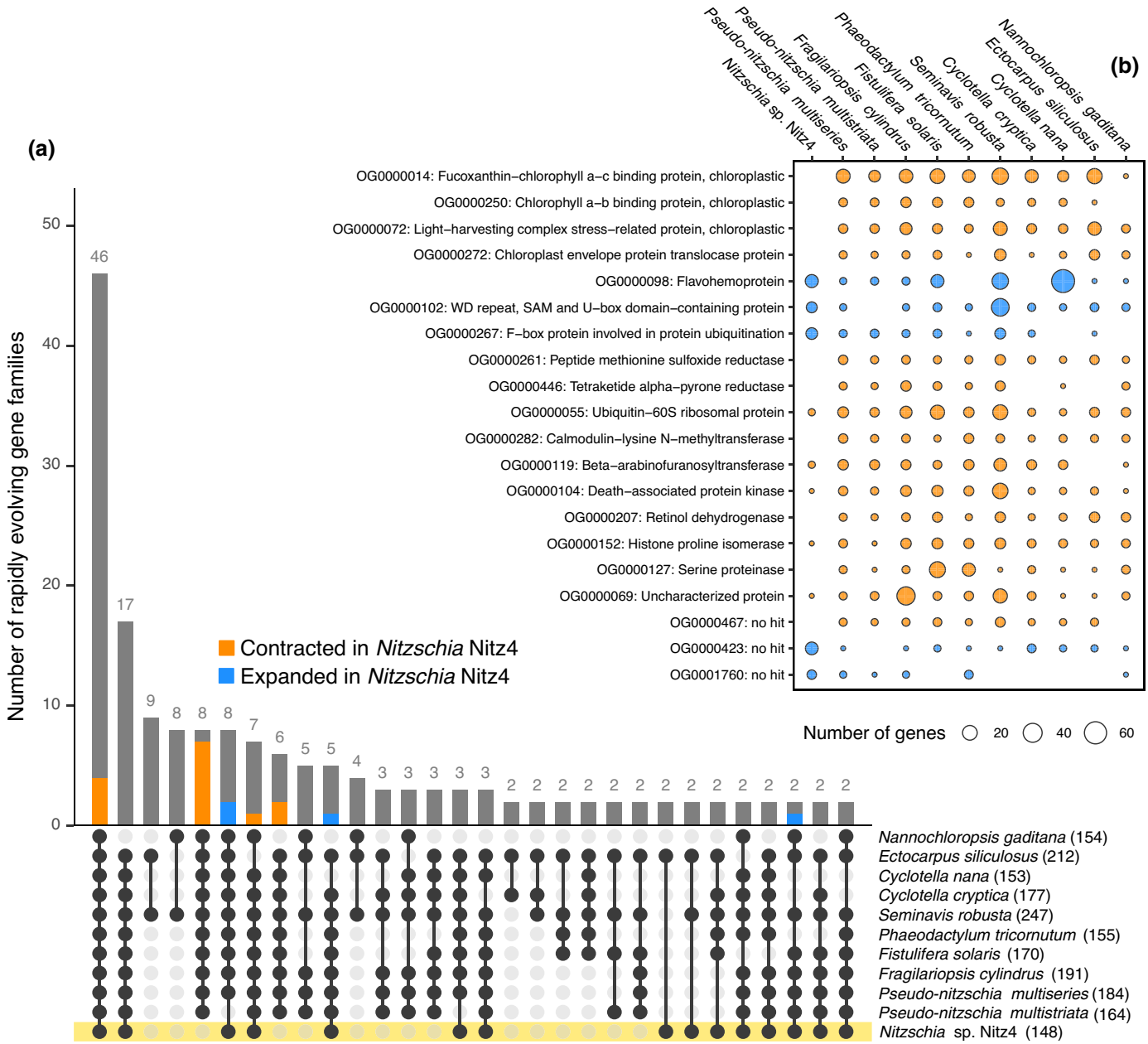

Fig. 1 Rapidly evolving gene families in *Nitzschia* Nitz4. (a) The histogram shows the total count of gene families for each combination of species; only those combinations/intersections containing ≥ 2 gene families are shown. Each row in the upset plot (bottom) represents a diatom or outgroup species, and the total number of intersecting gene families follows the species name. Vertical black lines connecting rows show which species contain the rapidly evolving gene families in a given intersection (i.e. category in the histogram). As an example, the far-left column shows 46 rapidly evolving gene families, and the upset plot below shows that those 46 families are shared by all 11 species. Intersections containing gene families that have contracted or expanded significantly in *Nitzschia* Nitz4 are colored orange or blue, respectively. (b) Each column represents a diatom or outgroup species, and each row a rapidly evolving gene family in *Nitzschia* Nitz4, identified by its orthogroup identifier and annotation. Circle size is proportional to gene family size, whereas the color of the circle shows whether the gene family was significantly expanded (blue) or contracted (orange) in *Nitzschia* Nitz4.

NADPH. When the carbon and energy demands of a cell have been met, gluconeogenesis converts pyruvate back into glucose, which can be converted into storage carbohydrates (e.g. chrysolaminarin and trehalose; Beattie *et al.*, 1961; Villanova *et al.*, 2017) as well as glycolytic intermediates for anabolism (Smith *et al.*, 2012; Tymoczko *et al.*, 2013). Because gluconeogenesis is essentially the reverse of glycolysis and shares many of the same enzymes, flux through each pathway is regulated through the control of specific enzymes driving irreversible reactions (GK, PFK, PK, and FBP).

Oxidative pentose phosphate pathway In diatoms, flux through central metabolism is also controlled through compartmentalization of different portions of key metabolic pathways. For

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

example, in diatoms the oxidative branch of the PPP, which reduces NADP+ to NADPH, is only present in the cytoplasm, whereas an incomplete nonoxidative PPP is present in the plastid, thereby limiting the available sources of NADPH in the plastid to PSI, sugar phosphates imported from the cytosol (Michels *et al.*, 2005; Moog *et al.*, 2020), and as detailed later, possibly a plastid NAD(P) transhydrogenase (NTH) that interconverts NADH to NADPH. As a result, nonphotosynthetic diatoms lack two principal sources of plastid NADPH – PSI and the oPPP – and therefore require alternative ways to reduce NADP+ in their plastids, which continue to house essential NADPH-consuming biosynthetic pathways (e.g. fatty acid and amino acid biosynthesis; Fig. 2). Moreover, diatom plastids export reducing equivalents to the mitochondria in exchange for ATP (Bailleul *et al.*, 2015), potentially further depleting NADPH levels in the plastid. Photosynthetic diatoms exposed to low or no light likely face a similar NADPH deficit (Grouneva *et al.*, 2009; Gruber *et al.*, 2009; Mekhalfi *et al.*, 2014; Kim *et al.*, 2016; Gruber & Kroth, 2017), which suggests that the ability to continue reducing NADP+ in their plastids in the absence of photosynthesis may be a general challenge for diatoms.

Glycolysis in *Nitzschia* Nitz4 We developed a model of central carbon metabolism to determine whether the switch to heterotrophy in *Nitzschia* Nitz4 included changes beyond the loss of photosynthesis-related genes. Cytosolic glycolysis in *Nitzschia* Nitz4 is similar to photosynthetic diatoms (Table S4). Like other raphid pennate diatoms, *Nitzschia* Nitz4 is missing cytosolic enolase, which catalyzes the penultimate step of glycolysis, suggesting that *Nitzschia* Nitz4 lacks a full cytosolic glycolysis pathway (Fig. 2; Table S4). *Nitzschia* Nitz4 does have plastid- and mitochondrial-targeted enolase genes, indicating that later stages of glycolysis occur in these compartments (Figs 2, 3; Table S4). *Nitzschia* Nitz4 has two copies of the gene encoding phosphoglycerate kinase (PGK), a key enzyme that catalyzes a reversible reaction used in both glycolysis and gluconeogenesis. Enzymes involved in the latter half of glycolysis are localized to the mitochondrion in most diatoms (Kroth *et al.*, 2008; Smith *et al.*, 2012), but neither PGK gene in *Nitzschia* Nitz4 contains a mitochondrial targeting signal – instead, one copy is predicted to be plastid-targeted and the other cytosolic (Fig. 2; Table S4). As discussed later, while plastid targeting of PGK may be important for reducing NADP+, it is likely that one or both of the plastid and cytosolic PGK genes are dual-targeted to the mitochondrion to complete the latter half of the pathway. Dual-targeting of proteins to the plastid and mitochondrion has been demonstrated in other diatoms (Gile *et al.*, 2015).

Pyruvate hub Because pyruvate is a key central metabolite that can be used for both catabolism and anabolism, we characterized the presence and localization of 'pyruvate hub enzymes' (Smith *et al.*, 2012) in *Nitzschia* Nitz4 (Table S4). *Nitzschia* Nitz4 has lost several pyruvate conversion genes, most dramatically in the plastid. For example, the pennate diatom *Phaeodactylum tricornutum* uses pyruvate phosphate dikinase (PPDK) to perform the first step of gluconeogenesis in the plastid, whereas the multipolar centric diatom *C. nana* uses phosphoenolpyruvate synthase (PEPS) (Smith *et al.*, 2012). *Nitzschia* Nitz4 lacks plastid-localized copies of both enzymes (Table S4). Additionally, and in contrast to photosynthetic diatoms, *Nitzschia* Nitz4 lacks a plastid-localized pyruvate carboxylase. Together, these results suggest that *Nitzschia* Nitz4 has lost the ability to initiate gluconeogenesis in the plastid and instead relies on mitochondrial-localized enzymes to determine the fate of pyruvate (Figs 2, 3; Table S4). The mitochondrial-localized pyruvate hub enzymes of photosynthetic diatoms are mostly present in *Nitzschia* Nitz4 (Fig. 3), with the sole exception being malate decarboxylase (or malic enzyme (ME)), which converts malate to pyruvate and CO_2 (Kroth *et al.*, 2008; Smith *et al.*, 2012). In C4 plants, ME is used to increase carbon fixation at low CO_2 concentrations (Sage, 2004). If ME performs a similar function in diatoms, then it may be dispensable in nonphotosynthetic species.

Loss of isoprenoid biosynthesis in the plastid Many photosynthesis-related molecules such as chlorophylls, carotenoids, and plastoquinones are derivatives of isoprenoids that are produced by the mevalonate (MVA) pathway in the cytosol or, in plastids, by the methylerythritol 4-phosphate (MEP) pathway, also known as the nonmevalonate pathway. The absence of an MVA pathway in apicomplexans, chlorophytes, and red algae suggests that it is not necessary to maintain both the MVA and MEP pathways for isoprenoid biosynthesis (Rohmer, 1999). The genome of *Nitzschia* Nitz4 has lost its plastid MEP pathway (Fig. 2), indicating an inability to synthesize carbon- and NADPH-consuming isoprenoids, the precursors of photosynthetic pigments, in the plastid. The transcriptome of *Nitzschia* NIES-3581 is missing a plastid MEP pathway as well (Kamikawa *et al.*, 2017).

Loss of the mitochondrial Entner–Doudoroff glycolytic pathway The loss of photosynthesis in *Nitzschia* Nitz4 underscored the energetic link between the plastids and mitochondria of diatoms (Bailleul *et al.*, 2015). Diatoms use a prokaryotic-like Entner–Doudoroff (ED) pathway for mitochondrial glycolysis (Fabris *et al.*, 2012). The ED glycolytic pathway is characterized by two key reactions, the first of which is catalyzed by 6-phosphogluconolactone dehydratase (EDD). Although *Nitzschia* Nitz4 lacks a mitochondrial-targeted gene for EDD (Fig. 3), this was not different from other diatoms, most of which have an EDD with a plastid, not mitochondrial, targeting prediction. Although the first reaction in the ED pathway was previously predicted to occur in the mitochondrion (Fabris *et al.*, 2012), our findings suggest that in diatoms EDD is either dual-targeted to the plastid and mitochondrion, or that the first reaction in the ED pathway occurs in the cytosol.

The EDA enzyme catalyzes the second reaction in the ED pathway and was present in all of the photosynthetic diatoms in our analysis, including a photosynthetic species from the genus *Nitzschia* (CCMP2144). The EDA gene was missing, however, from both nonphotosynthetic diatoms – *Nitzschia* Nitz4 (Fig. 3) and *Nitzschia* NIES-3581 – indicating that its loss was recent and likely related to the loss of photosynthesis. Loss of EDA

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

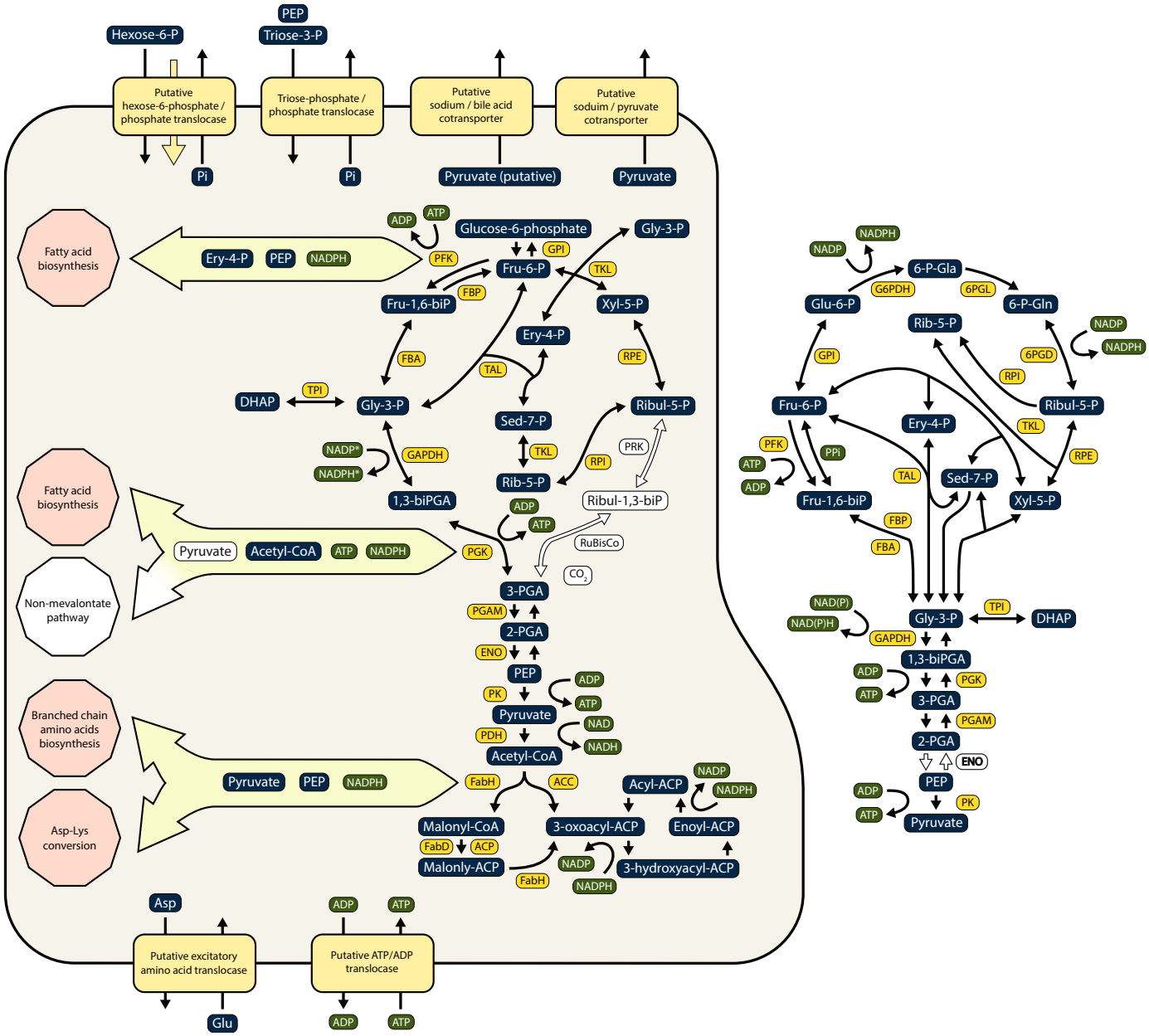

Fig. 2 Genome-based model of carbon metabolism in the plastid (left) and cytosol (right) of *Nitzschia* Nitz4. Chemical substrates and products are shown in blue, enzymes encoded by genes present in *Nitzschia* Nitz4 are shown in dark yellow, and cofactors are shown in green. Black and yellow arrows show predicted reactions, pathways, or direction of transport of chemical substrates across membranes. Polygons show presence (pink) or absence (white) of complete pathways. Missing genes and pathways are shown in white.

remodeled mitochondrial glycolysis in these species, transitioning them from the ancestral and highly conserved ED pathway to the more classic Embden–Meyerhof–Parnas (EMP) glycolytic pathway (Fig. 3).

The ED glycolytic pathway generates half the ATP as the EMP pathway but requires production of much less protein to achieve the same flux (Flamholz *et al.*, 2013). As a result, these two strategies reflect a trade-off between energy yield and production costs of glycolytic enzymes (Flamholz *et al.*, 2013). Organisms such as phototrophs that have abundant nonglycolytic sources of ATP but face limiting nitrogen commonly possess the ED pathway (Flamholz *et al.*, 2013). We hypothesize that the loss of photosynthesis exerts strong selection to shift from mitochondrial ED to EMP glycolysis. The presence of an ED pathway in many autotrophic, but not heterotrophic, stramenopiles (Río Bártulos *et al.*, 2018) supports this hypothesis, and suggests that the transition from an ED to EMP pathway in secondary heterotrophs may be a general phenomenon.

Diatoms plastids likely export and shuttle reducing equivalents to the mitochondria in exchange for ATP (Bailleul *et al.*, 2015), so the switch to a mitochondrial EMP pathway would not only increase ATP yield in the mitochondrion, but it might also increase ATP flux into plastids where ATP is probably in short supply in nonphotosynthetic diatoms. NADPH might also be scarce in the

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

New Phytologist

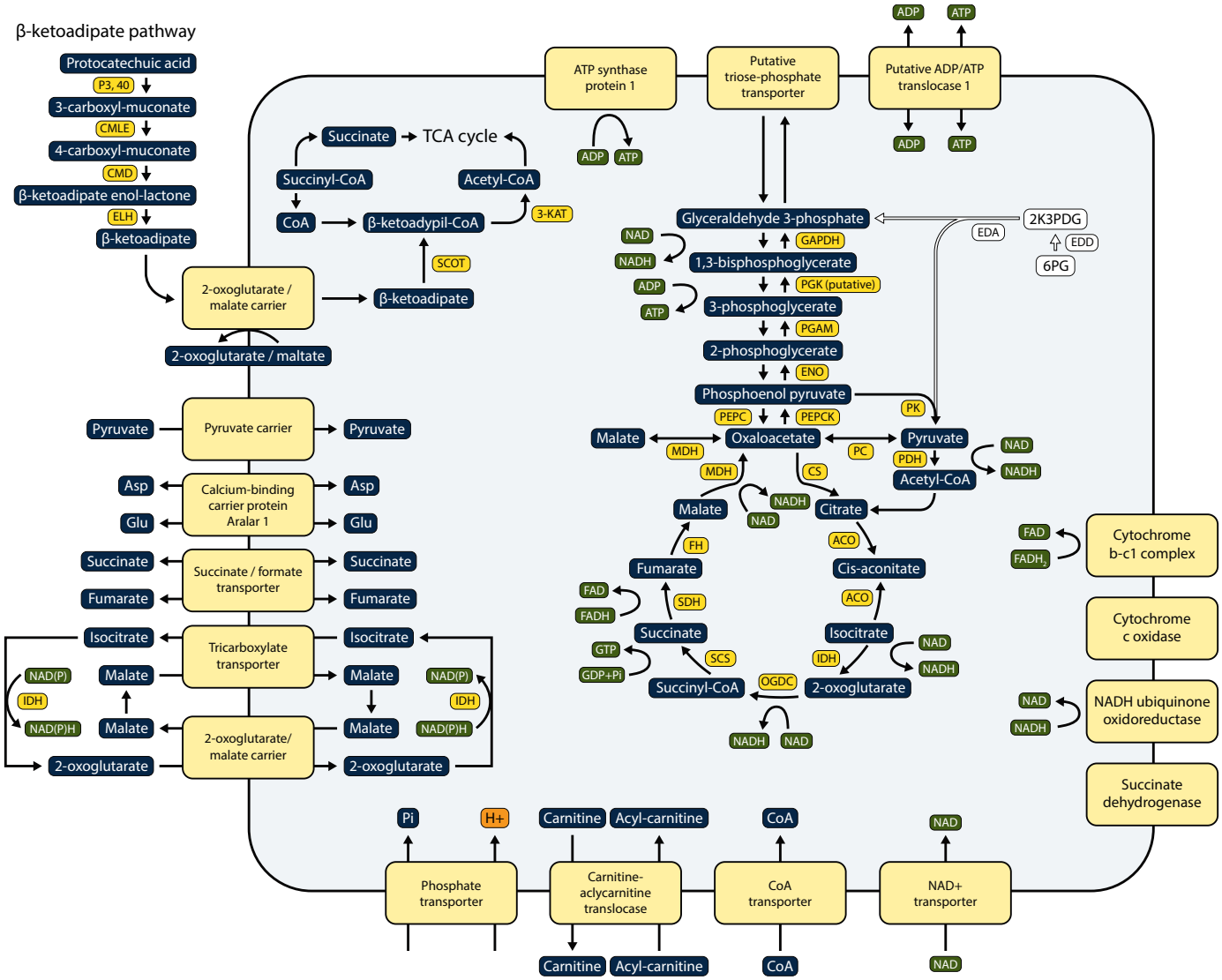

Fig. 3 Genome-based model of carbon metabolism in the mitochondria of *Nitzschia* Nitz4. Membrane proteins are shown in light yellow, chemical substrates and products are shown in blue, enzymes encoded by genes present in *Nitzschia* Nitz4 are shown in dark yellow, cofactors are shown in green, and protons in orange. Arrows show predicted reactions, pathways, or direction of transport of chemical substrates across membranes. The missing Entner–Doudoroff glycolytic pathway is shown in white.

plastids of nonphotosynthetic diatoms, which need to preserve an optimal ATP : NADPH ratio through decreases in demand (e.g. by loss of the MEP) and modulation of ATP production in the mitochondrion. In addition to generating half the ATP, the ED and EMP pathways also differ in their redox cofactors. In diatoms, the ED pathway yields one NADPH in the cytosol and one NADH in the mitochondrion (Fabris *et al.*, 2012), whereas the EMP pathway yields two NADH in the mitochondrion. As a result, the transition from an ED to EMP pathway eliminated a source of cytosolic NADPH, which is noteworthy given the potential importance of NADPH limitation in the plastid (see section entitled 'NADPH limitation and constraints on the loss of photosynthesis').

Carbon active enzymes The transition to heterotrophy in *Nitzschia* Nitz4 did not appear to be facilitated by acquisitions of novel carbon transporters or carbohydrate-cleaving enzymes, or by significant expansions of these gene families (Fig. 1; Tables S5, S6).

A β-ketoadipate pathway in diatoms

One of our primary goals was to develop a comprehensive model of central carbon metabolism for nonphotosynthetic diatoms. A systematic search for carbon-active enzymes revealed a β-ketoadipate pathway, whose primary function is degradation of plant-derived aromatic substrates (Harwood & Parales, 1996). The β-ketoadipate pathway is typically associated with soil-dwelling fungi and eubacteria and consists of two main branches, one of which converts protocatechuic acid (protocatechuate, or PCA) and the other catechol, into β-ketoadipate (Harwood & Parales, 1996). Diatoms possess the PCA branch of the pathway

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

(Fig. 3; Table S7). Intermediate steps of the PCA branch differ between fungi and bacteria (Harwood & Parales, 1996), and diatoms possess the bacterial-like PCA branch. In the final steps of the pathway, β-ketoadipate is converted to the TCA intermediates succinyl-CoA and acetyl-CoA (Fig. 3). The β-ketoadipate pathway is not unique to *Nitzschia* Nitz4 as most or all of the requisite genes are present in other sequenced diatom genomes and transcriptomes, which suggests that the pathway is an ancestral feature of diatom genomes (Table S8).

We initially discovered eight genes in *Nitzschia* Nitz4 with matches to bacterial-like proteins annotated as intradiol ring-cleavage dioxygenase or protocatechuate 3,4-dioxygenase (P3,4O). This is the first enzyme in the protocatechuate branch of the β-ketoadipate pathway and is involved in the degradation of aromatic compounds through intradiol ring cleavage of PCA (Harwood & Parales, 1996). Fission of the highly stable aromatic ring structure of PCA is the most difficult step of the pathway (Fuchs *et al.*, 2011), so it is noteworthy that both nonophotosynthetic *Nitzschia* species are enriched for this gene – each possessing at least eight copies – whereas other diatoms have four or fewer copies (Table S8). Expansion of the P3,4O gene family in *Nitzschia* Nitz4 and NIES-3581 suggests that non-photosynthetic diatoms might be more efficient at carrying out the putative bottleneck reaction of the β-ketoadipate pathway.

Targeted searches of the annotated genes and scaffolds revealed genes for subsequent steps of the pathway, including 3-carboxymuconate lactonizing enzyme (CMLE), 4-carboxymuconolactone decarboxylase (CMD), and β-ketoadipate enol-lactone hydrolase (ELH) (Fig. 3; Table S7). Enzymes encoding the final two steps, 3-ketoadipate:succinyl-CoA transferase (TR) and 3-ketoadipyl-CoA thiolase (TH), are missing from *Nitzschia* Nitz4. However, these enzymes are similar in both sequence and predicted function – attachment and removal of CoA – to two mitochondrial-targeted genes that are present in *Nitzschia* Nitz4, succinyl-CoA:3-ketoacid CoA transferase (SCOT) and 3-ketoacyl-CoA thiolase (3-KAT), respectively (Parales & Harwood, 1992; Kaschabek *et al.*, 2002). Finally, all sequenced diatom genomes, including *Nitzschia* Nitz4, have a putative 4-hydroxybenzoate transporter (PcaK; Table S8) that reportedly can import PCA into the cell (Nichols & Harwood, 1997).

Nonphotosynthetic *Nitzschia* are strictly marine and found on coastlines, and they are especially common in mangrove habitats (Ishii & Kamikawa, 2017; Onyshchenko *et al.*, 2019). Vascular plants, including mangroves, are a major source of dissolved organic carbon (DOC) in coastal ocean ecosystems (Moran *et al.*, 1991b; Moran & Hodson, 1994b; Hedges *et al.*, 1997). Plant-derived DOC includes lignin and a suite of lignin-derived phenols that support secondary production of diverse communities of marine bacteria (Moran & Hodson, 1994a; Mou *et al.*, 2008). Vanillin is one of the principal phenolic subunits of lignin and comprises a large fraction of the total lignin phenols in some systems (Kattner *et al.*, 1999; Cotrim da Cunha *et al.*, 2001). One of the products of vanillin degradation, PCA, is the predicted substrate for the β-ketoadipate pathway (Fig. 3), which serves as a 'funnel' for numerous lignin-derived PCA or catechol precursors (Harwood & Parales, 1996; Fuchs *et al.*, 2011). Although bacteria drive biodegradation of lignin phenols in the ocean, the presence of a β-ketoadipate pathway in diatoms suggests they may play a role in this part of the ocean carbon cycle as well.

Although *Nitzschia* Nitz4 and other diatoms possess a full β-ketoadipate pathway, the genomic prediction – that diatoms with this pathway metabolize naturally occurring aromatic compounds into products (succinyl- and acetyl-CoA) that feed into the TCA cycle – requires experimental corroboration. It will be important to show whether diatoms import PCA, whether it augments growth, and whether the terminal products are converted into cellular biomass or, alternatively, whether the pathway is for catabolism of PCA or a related compound. The piecemeal β-ketoadipate pathway in some diatoms raises questions about the functions of these genes and whether the complete pathway is restricted to diatoms in coastal areas where lignin-derived phenols are most abundant (Moran *et al.*, 1991a; Opsahl & Benner, 1997).

NADPH limitation and constraints on the loss of photosynthesis

Plastids are the site of essential biosynthetic pathways and, consequently, remain biochemically active following loss of their primary function, photosynthesis (Neuhaus & Emes, 2000). Most biosynthetic pathways in the plastid require the reducing cofactor, NADPH, which is generated principally through the light-dependent reactions of photosynthesis and the oPPP. The primary-plastid-containing Archaeplastida – including glaucophytes (Fester & Schnenk, 1997), green algae (Weber & Linka, 2011), land plants (Schnarrenberger *et al.*, 1995), and red algae (Moriyama *et al.*, 2014) – have both cytosolic- and plastid-localized oPPPs. In plants, the plastid oPPP generates NADPH reductants in nonphotosynthetic embryos (Andriotis & Smith, 2019) and root tissues (Bowsher *et al.*, 1992), and in Archaeplastida, the plastid oPPP continues to reduce NADP+ when photosynthesis is lost – something that has occurred at least five times in green algae (Rumpf *et al.*, 1996; Tartar & Boucias, 2004), nearly a dozen times in angiosperms (Barkman *et al.*, 2007), and some 100 times in red algae (Blouin & Lane, 2012). Chrysophyte algae, which also have a plastid oPPP, have lost photosynthesis at least seven times (Dorrell *et al.*, 2019). Despite their age (200 Myr) and estimates of species richness on par with angiosperms and greatly exceeding green and red algae, diatoms are known to have lost photosynthesis only twice (Frankovich *et al.*, 2018; Onyshchenko *et al.*, 2019).

Unlike Archaeplastida, diatoms lack a plastid oPPP and therefore rely primarily on photosynthesis for the reduction of NADP+ in their plastids (Michels *et al.*, 2005; Gruber *et al.*, 2009). Nevertheless, the plastids of nonphotosynthetic diatoms continue to be the site of numerous NADPH-consuming reactions (Fig. 2; Kamikawa *et al.*, 2017). Without an oPPP, it is unclear how diatom plastids meet their NADPH demands during periods of prolonged darkness or, by extension, in species that have lost photosynthesis. Possible enzymatic sources include an NADP-dependent isoform of the glycolytic enzyme GAPDH or conversion of malate to pyruvate by ME (Smith *et al.*, 2012). In

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

Nitzschia Nitz4, NADP-GAPDH is predicted to be targeted to the periplastid compartment, but ME appears to have been lost.

In addition to enzymatic production, plastids might acquire NADPH via transport from the cytosol, but transport of large polar molecules like NADPH through the inner plastid membrane might be prohibitive (Taniguchi & Miyake, 2012). *Arabidopsis* has a plastid NAD(H) transporter (Palmieri *et al.*, 2009), but to our knowledge no NADP(H) transporters have been identified in diatoms. Cells commonly use 'shuttle' systems to exchange redox equivalents across intracellular membranes (Noctor *et al.*, 2011). Although these systems generally shuttle excess reductants from the plastid to the cytosol (Taniguchi & Miyake, 2012), the triose phosphate/3-phosphoglycerate shuttle is hypothesized to run in reverse to generate NADPH in the plastids of nonphotosynthetic diatoms (Moog *et al.*, 2020) as well as the apicoplast of the malaria parasite, *Plasmodium falciparum* (Lim & McFadden, 2010). In this model, triose phosphates are transferred from the cytosol to the plastid in exchange for 3-phosphoglyceric acid (PGA), resulting in the generation of ATP and NADPH in the plastid by PGK and NADP-GAPDH (Fig. 4; Moog *et al.*, 2020). *Nitzschia* Nitz4 contains a plastid-targeted NADP-GAPDH and two copies of PGK (one cytosolic and one plastid-localized). This differs from photosynthetic diatoms, where at least one PGK is localized to the mitochondrion (Smith *et al.*, 2012).

An apicoplast membrane-targeted NTH was recently discovered in *Plasmodium* (Saeed *et al.*, 2020). Normally localized to the mitochondrial membrane in eukaryotes, NTH catalyzes, via hydride transfer, the interconversion of NADH and NADPH (Fig. 4; Jackson *et al.*, 2015). The apicoplast-targeted NTH might provide a novel and necessary source of NADPH to that organelle (Lim & McFadden, 2010). *Nitzschia* Nitz4 has two NTH genes, one targeted to the mitochondrion (NITZ4_001957) and a second (NITZ4_000738) to the plastid. Although optimized for Viridiplantae, SChloro (Savojardo *et al.*, 2017) predicted that NITZ4_000738 is localized to the thylakoid membrane (Fig. 4). In addition to facilitating translocation of proteins into the thylakoids (Kamikawa *et al.*, 2015), the proton gradient created through ATP hydrolysis might also reduce NADP+ in the plastid. If NTH is indeed involved in the reduction of NADP+, this highlights another way in which nonphotosynthetic diatoms and apicomplexans converged upon similar solutions to NADPH limitation in their plastids.

Indispensable pathways in diatom plastids include fatty acid biosynthesis, branched chain and aromatic amino acid biosynthesis, and Asp-Lys conversion – all of which consume carbon and NADPH (Fig. 2). In the absence of photosynthesis for carbon fixation and reduction of NADP+, and without an oPPP to reduce NADP+ in the plastid, essential building blocks and reducing cofactors for biosynthetic pathways in the plastid are probably in short supply in nonphotosynthetic diatoms. The need to conserve carbon and NADPH likely drove loss of the isoprenoid-producing MEP from the plastid in nonphotosynthetic diatoms (Kamikawa *et al.*, 2017), chrysophytes (Dorrell *et al.*, 2019), and euglenophytes (Kim *et al.*, 2004). Despite lack of an oPPP in the apicoplast (Ralph *et al.*, 2004), *Plasmodium falciparum* has retained its MEP, most likely because it lacks a cytosolic MVA (Imlay & Odom, 2014), an alternate source of isoprenoid production. Taken together, whether loss of photosynthesis leads to loss of the MEP likely depends on several interrelated factors, including whether the functionally overlapping MVA and MEP pathways (Liao *et al.*, 2016) were both present in the photosynthetic ancestor and whether there is a dedicated plastid oPPP. Given the challenges of shedding the plastid entirely, the limited number of requisite supply lines to plastid-localized biosynthetic pathways is likely an important constraint on loss of photosynthesis in groups without a plastid oPPP. In the absence of both PSI and an oPPP, the ease with which nonphotosynthetic diatoms meet their NADPH demands, either by eliminating dispensable pathways and increasing flux of plastid NADPH through a limited number of known pathways (Fig. 4), or

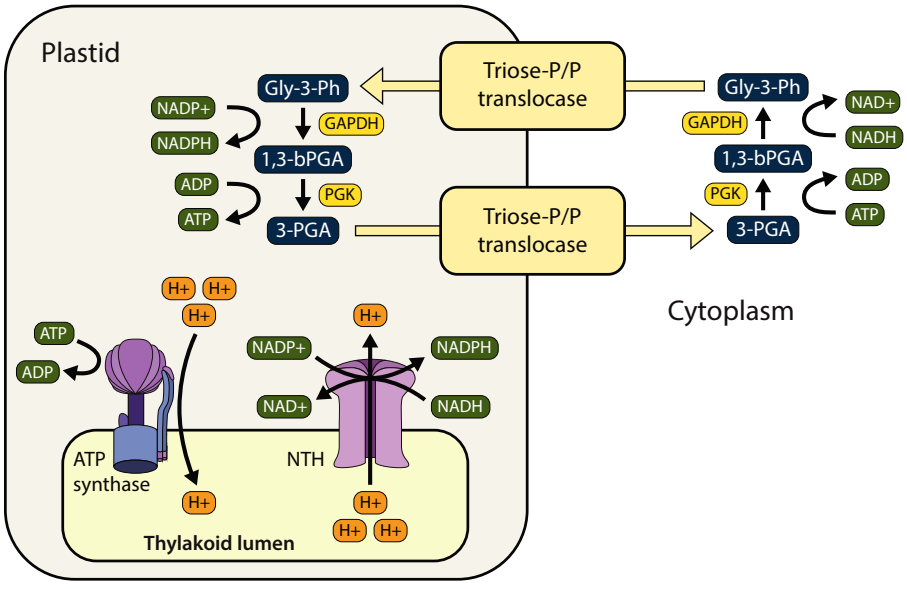

Fig. 4 Possible sources of NADPH in the plastid of *Nitzschia* Nitz4. The 'reverse' triose phosphate/3-phosphoglycerate shuttle would exchange cytoplasmic NADH for NADPH (Moog *et al.*, 2020), and NAD(P) transhydrogenase (NTH) would use the proton gradient generated by adenosine triphosphate (ATP) hydrolysis, via ATP synthase (Kamikawa *et al.*, 2015), to catalyze interconversion of NADH to NADPH. Chemical substrates and products are shown in blue, protons in orange, and cofactors are shown in green. Arrows show predicted reactions or direction of transport of protons or chemical substrates across membranes. In addition to ATP synthase and NTH, other enzymes encoded by genes present in *Nitzschia* Nitz4 are shown in dark yellow.

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

whether they possess novel mechanisms for provisioning NADPH in their plastids, remains unclear.

Conclusions

Since their origin some 200 Myr ago, diatoms have become an integral part of marine and freshwater ecosystems worldwide. Photosynthesis by diatoms accounts for 20% of global primary production, making them cornerstones of the 'biological pump' that fixes and then buries atmospheric carbon on the seafloor. We sequenced the genome of a nonphotosynthetic diatom to understand how metabolism has been restructured in heterotrophic diatoms and, more broadly, how core aspects of diatom metabolism manifest at macroevolutionary scales, such as the frequency of transition between trophic modes.

Metabolic and genomic changes associated with the switch to heterotrophy in *Nitzschia* Nitz4 were cellular-wide and included losses of photosynthesis-related genes from the nuclear and plastid genomes, loss of the plastid MEP pathway, and remodeling of an ancient and conserved mode of mitochondrial glycolysis – a switch that may help compensate for the loss of autotrophic sources of ATP. Lack of a dedicated plastid oPPP is an ancestral feature of diatoms. Considering that groups which have lost photosynthesis more frequently tend to have a plastid oPPP, the comparative rarity of loss of photosynthesis in diatoms may reflect the fewer avenues available to diatoms to compensate for the loss of photosynthesis-derived NADPH.

The presence of a β-ketoadipate pathway underscores the genomic complexity of diatoms and the ability of genome sequencing to provide a first glimpse into the metabolism of marine microbes. As for other diatoms, we could not ascribe functions to many of the genes in *Nitzschia* Nitz4, potentially masking important aspects of its metabolism and highlighting a persistent challenge in comparative genomics.

Acknowledgements

This research was carried out by AO in partial fulfillment of a Master of Science degree, which was supported by the Fulbright Graduate Student Exchange Program (Ukraine). This research was supported by the National Science Foundation (NSF; grant no. DEB-1353131 to AJA) and multiple awards from the Arkansas Biosciences Institute. This research used computational resources available through the Arkansas High Performance Computing Center, which was funded through multiple NSF grants and the Arkansas Economic Development Commission. The authors thank Simon Tye for help designing and creating figures, Jelena Godrijan for insights regarding the functional classification of heterotrophs, and Ben Lambert and two anonymous reviewers for comments on an earlier draft of the manuscript.

Author contributions

AJA, JAL, AO and WRR designed the research. AJA, AO, ECR and WRR performed the research. AJA, JAL, AO and WRR wrote the article with input from all co-authors.

ORCID

Andrew J. Alverson https://orcid.org/0000-0003-1241-2654
Jeffrey A. Lewis https://orcid.org/0000-0003-2952-9701
Anastasiia Onyshchenko https://orcid.org/0000-0001-7948-5925
Wade R. Roberts https://orcid.org/0000-0002-5100-7558
Elizabeth C. Ruck https://orcid.org/0000-0002-9288-1989

Data availability

Sequencing reads are available through NCBI BioProject PRJNA412514, and the genome assembly is available through NCBI accession WXVQ00000000. Transcriptome assemblies for *Nitzschia* Nitz4 and *Nitzschia* Nitz2144 are available from NCBI accessions GIQR00000000 and GIQQ00000000, respectively. A genome browser with gene annotations is available through the Comparative Genomics (CoGe) web platform (https://genomevolution.org/coge/) under genome ID 60130. Transcriptome assemblies, OrthoFinder and Cafe results, and genome annotations are available from an online Zenodo repository: 10.5281/zenodo.4479681.

References

Alverson AJ, Beszteri B, Julius ML, Theriot EC. 2011. The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evolutionary Biology* **11**: 125.

Andriotis VME, Smith AM. 2019. The plastidial pentose phosphate pathway is essential for postglobular embryo development in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **116**: 15297–15306.

Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, Villanova V, Bligny R, Flori S, Falconet D *et al.* 2015. Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. *Nature* **524**: 366–369.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Barkman TJ, McNeal JR, Lim S-H, Coat G, Croom HB, Young ND, dePamphilis CW. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* **7**: 248.

Beattie A, Hirst EL, Percival E. 1961. Studies on the metabolism of the Chrysophyceae. Comparative structural investigations on leucosin (chrysolaminarin) separated from diatoms and laminarin from the brown algae. *Biochemical Journal* **79**: 531–537.

Blackburn MV, Hannah F, Rogerson A. 2009. First account of apochlorotic diatoms from mangrove waters in Florida. *Journal of Eukaryotic Microbiology* **56**: 194–200.

Blanc-Mathieu R, Krasovec M, Hebrard M, Yau S, Desgranges E, Martin J, Schackwitz W, Kuo A, Salin G, Donnadieu C *et al.* 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Science Advances* **3**: e1700239.

Blouin NA, Lane CE. 2012. Red algal parasites: models for a life history evolution that leaves photosynthesis behind again and again. *BioEssays* **34**: 226–235.

Bock NA, Charvet S, Burns J, Gyaltshen Y, Rozenberg A, Duhamel S, Kim E. 2021. Experimental identification and in silico prediction of bacterivory in green algae. *ISME Journal* **15**: 1987–2000.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

Bowsher CG, Boulton EL, Rose J, Nayagam S. 1992. Reductant for glutamate synthase is generated by the oxidative pentose phosphate pathway in non-photosynthetic root plastids. *The Plant Journal* **2**: 893–898.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. *Trends in Ecology & Evolution* **35**: 43–55.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188–196.

Cerón García MC, Camacho FG, Mirón AS, Fernández Sevilla JM, Chisti Y, Molina GE. 2006. Mixotrophic production of marine microalga *Phaeodactylum tricornutum* on various carbon sources. *Journal of Microbiology and Biotechnology* **16**: 689–694.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* **1962**: 1–14.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**: 80–92.

Claros MG. 1995. MitoProt, a Macintosh application for studying mitochondrial proteins. *Computer Applications in the Biosciences* **11**: 441–447.

Clement R, Lignon S, Mansuelle P, Jensen E, Pophillat M, Lebrun R, Denis Y, Puppo C, Maberly SC, Gontero B. 2017. Responses of the marine diatom *Thalassiosira pseudonana* to changes in CO_2 concentration: a proteomic approach. *Scientific Reports* 7: 42333.

Cotrim da Cunha L, Serve L, Gadel F, Blazi J-L. 2001. Lignin-derived phenolic compounds in the particulate organic matter of a French Mediterranean river: seasonal and spatial variations. *Organic Geochemistry* **32**: 305–320.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**: 491–498.

Dorrell RG, Azuma T, Nomura M, Audren de Kerdrel G, Paoli L, Yang S, Bowler C, Ishii K-I, Miyashita H, Gile GH *et al.* 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proceedings of the National Academy of Sciences, USA* **116**: 6914–6923.

Dufresne A, Garczarek L, Partensky F. 2005. Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biology* **6**: R14.

Edlund MB, Stoermer EF. 1997. Ecological, evolutionary, and systematic significance of diatom life histories. *Journal of Phycology* **33**: 897–918.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A *et al.* 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* **47**: D427–D432.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**: 953–971.

Emanuelsson O, Nielsen H, von Heijne G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Science* **8**: 978–984.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.

Fabris M, Matthijs M, Rombauts S, Vyverman W, Goossens A, Baart GJE. 2012. The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner–Doudoroff glycolytic pathway. *The Plant Journal* **70**: 1004–1014.

Falkowski PG, Raven JA. 2013. *Aquatic photosynthesis, 2nd edn.* Princeton, NJ, USA: Princeton University Press.

Fester T, Schnenk HEA. 1997. Glucose-6-Phosphate dehydrogenase isoenzymes from *Cyanophora paradoxa*: Examination of their metabolic integration within the meta-endocytobiotic system. In: Schnenk HEA, Herrmann RG, Jeon KW, Müller NE, Schwemmler W, eds. *Eukaryotism and symbiosis: intertaxonic combination versus symbiotic adaptation.* Berlin Heidelberg, Germany: Springer, 243–251.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237–240.

Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. 2013. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences, USA* **110**: 10039–10044.

Frankovich TA, Ashworth MP, Sullivan MJ, Theriot EC, Stacy NI. 2018. Epizoic and apochlorotic *Tursiocola* species (Bacillariophyta) from the skin of Florida manatees (*Trichechus manatus latirostris*). *Protist* **169**: 539–568.

Fuchs G, Boll M, Heider J. 2011. Microbial degradation of aromatic compounds – from one strategy to four. *Nature Reviews Microbiology* **9**: 803–816.

Gile GH, Moog D, Slamovits CH, Maier U-G, Archibald JM. 2015. Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. *Genome Biology and Evolution* **7**: 1728–1742.

Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME Journal* **8**: 1553–1565.

Goff LJ, Moon DA, Nyvall P, Stache B, Mangin K, Zuccarello G. 1996. The evolution of parasitism in the red algae: molecular comparisons of adelphoparasites and their hosts. *Journal of Phycology* **32**: 297–312.

Grouneva I, Jakob T, Wilhelm C, Goss R. 2009. The regulation of xanthophyll cycle activity and of non-photochemical fluorescence quenching by two alternative electron flows in the diatoms *Phaeodactylum tricornutum* and *Cyclotella meneghiniana*. *Biochimica et Biophysica Acta* **1787**: 929–938.

Gruber A, Kroth PG. 2017. Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **372**: 20160402.

Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. 2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *The Plant Journal* **81**: 519–528.

Gruber A, Weber T, Bártulos CR, Vugrinec S, Kroth PG. 2009. Intracellular distribution of the reductive and oxidative pentose phosphate pathways in two diatoms. *Journal of Basic Microbiology* **49**: 58–72.

Gschloessl B, Guermeur Y, Cock JM. 2008. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinformatics* **9**: 393.

Guillard RRL. 1960. A mutant of *Chlamydomonas moewusii* lacking contractile vacuoles. *Journal of Eukaryotic Microbiology* 7: 262–268.

Guillard RRL, Hargraves PE. 1993. *Stichochrysis immobilis* is a diatom, not a chrysophyte. *Phycologia* **32**: 234–236.

Guillory WX, Onyshchenko A, Ruck EC, Parks M, Nakov T, Wickett NJ, Alverson AJ. 2018. Recurrent loss, horizontal transfer, and the obscure origins of mitochondrial introns in diatoms (Bacillariophyta). *Genome Biology and Evolution* **10**: 1504–1515.

Hadariová L, Vesteg M, Hampl V, Krajčovič J. 2018. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Current Genetics* **64**: 365–387.

Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. 2005. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Research* **15**: 1153–1160.

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution* **30**: 1987–1997.

Harwood CS, Parales RE. 1996. The beta-ketoadipate pathway and the biology of self-identity. *Annual Review of Microbiology* **50**: 553–590.

Hedges JI, Keil RG, Benner R. 1997. What happens to terrestrial organic matter in the ocean? *Organic Geochemistry* **27**: 195–212.

Hellebust JA. 1971. Glucose uptake by *Cyclotella cryptica*: dark induction and light inactivation of transport system. *Journal of Phycology* 7: 345–349.

Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genetics* **6**: e1001115.

Hochmal AK, Schulze S, Trompelt K, Hippler M. 2015. Calcium-dependent regulation of photosynthesis. *Biochimica et Biophysica Acta* **1847**: 993–1003.

Imlay L, Odom AR. 2014. Isoprenoid metabolism in apicomplexan parasites. *Current Clinical Microbiology Reports* **1**: 37–50.

Ishii K-I, Kamikawa R. 2017. Growth characterization of non-photosynthetic diatoms, *Nitzschia* spp., inhabiting estuarine mangrove forests of Ishigaki Island, Japan. *Plankton and Benthos Research* **12**: 164–170.

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

Jackson JB, Leung JH, Stout CD, Schurig-Briccio LA, Gennis RB. 2015. Review and hypothesis. New insights into the reaction mechanism of transhydrogenase: Swivelling the dIII component may gate the proton channel. *FEBS Letters* **589**: 2027–2033.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240.

Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z *et al.* 2021. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* **49**: D192–D200.

Kamikawa R, Moog D, Zauner S, Tanifuji G, Ishida K-I, Miyashita H, Mayama S, Hashimoto T, Maier UG, Archibald JM *et al.* 2017. A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. *Molecular Biology and Evolution* **34**: 2355–2366.

Kamikawa R, Tanifuji G, Ishikawa SA, Ishii K-I, Matsuno Y, Onodera NT, Ishida K-I, Hashimoto T, Miyashita H, Mayama S *et al.* 2015. Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Molecular Biology and Evolution* **32**: 2598–2604.

Kaschabek SR, Kuhn B, Müller D, Schmidt E, Reineke W. 2002. Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: purification and characterization of 3-oxoadipate: succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase. *Journal of Bacteriology* **184**: 207–215.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kattner G, Lobbes JM, Fitznar HP, Engbrodt R, Nöthig E-M, Lara RJ. 1999. Tracing dissolved organic substances and nutrients from the Lena River through Laptev Sea (Arctic). *Marine Chemistry* **65**: 25–39.

Kessenich CR, Ruck EC, Schurko AM, Wickett NJ, Alverson AJ. 2014. Transcriptomic insights into the life history of Bolidophytes, the sister lineage to diatoms. *Journal of Phycology* **50**: 977–983.

Kim D, Filtz MR, Proteau PJ. 2004. The methylerythritol phosphate pathway contributes to carotenoid but not phytol biosynthesis in *Euglena gracilis*. *Journal of Natural Products* **67**: 1067–1069.

Kim J, Fabris M, Baart G, Kim MK, Goossens A, Vyverman W, Falkowski PG, Lun DS. 2016. Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. *The Plant Journal* **85**: 161–176.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59.

Krasovec M, Sanchez-Brosseau S, Piganeau G. 2019. First estimation of the spontaneous mutation rate in diatoms. *Genome Biology and Evolution* **11**: 1829–1837.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Transmembrane helices predicted using TMHMM: predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**: 567–580.

Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T *et al.* 2008. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* **3**: e1426.

Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Research* **6**: 1287.

Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* **35**: 3100–3108.

Levialdi Ghiron JH, Amato A, Montresor M, Kooistra WHCF. 2008. Plastid inheritance in the planktonic raphid pennate diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae). *Protist* **159**: 91–98.

Lewin J, Lewin RA. 1967. Culture and nutrition of some apochlorotic diatoms of the genus *Nitzschia*. *Microbiology* **46**: 361–367.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Liao P, Hemmerlin A, Bach TJ, Chye M-L. 2016. The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnology Advances* **34**: 697–713.

Lim L, McFadden GI. 2010. The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **365**: 749–763.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490–D495.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**: 18.

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**: 574–576.

Mekhalfi M, Puppo C, Avilan L, Lebrun R, Mansuelle P, Maberly SC, Gontero B. 2014. Glyceraldehyde-3-phosphate dehydrogenase is regulated by ferredoxin-NADP reductase in the diatom *Asterionella formosa*. *New Phytologist* **203**: 414–423.

Michels AK, Wedel N, Kroth PG. 2005. Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. *Plant Physiology* **137**: 911–920.

Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelser P, Barcelona J, Inovejas SA *et al.* 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution* **31**: 793–803.

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

Moran MA, Hodson RE. 1994a. Support of bacterioplankton production by dissolved humic substances from three marine environments. *Marine Ecology Progress Series* **110**: 241.

Moran MA, Hodson RE. 1994b. Dissolved humic substances of vascular plant origin in a coastal marine environment. *Limnology and Oceanography* **39**: 762–771.

Moran MA, Pomeroy LR, Sheppard ES, Atkinson LP, Hodson RE. 1991a. Distribution of terrestrially derived dissolved organic matter on the southeastern US continental shelf. *Limnology and Oceanography* **36**: 1134–1149.

Moran MA, Wicks RJ, Hodson RE. 1991b. Export of dissolved organic matter from a mangrove swamp ecosystem: evidence from natural fluorescence, dissolved lignin phenols, and bacterial secondary production. *Marine Ecology Progress Series* **76**: 175–184.

Moriyama T, Sakurai K, Sekine K, Sato N. 2014. Subcellular distribution of central carbohydrate metabolism pathways in the red alga *Cyanidioschyzon merolae*. *Planta* **240**: 585–598.

Mou X, Sun S, Edwards RA, Hodson RE, Moran MA. 2008. Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**: 708–711.

Nagao R, Tomo T, Noguchi E, Suzuki T, Okumura A, Narikawa R, Enami I, Ikeuchi M. 2012. Proteases are associated with a minor fucoxanthin chlorophyll a/c-binding protein from the diatom, *Chaetoceros gracilis*. *Biochimica et Biophysica Acta* **1817**: 2110–2117.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**: 2933–2935.

Neuhaus HE, Emes MJ. 2000. Nonphotosynthetic metabolism in plastids. *Annual Review of Plant Physiology and Plant Molecular Biology* **51**: 111–140.

Nichols NN, Harwood CS. 1997. PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*. *Journal of Bacteriology* **179**: 5056–5061.

Noctor G, Hager J, Li S. 2011. Biosynthesis of NAD and its manipulation in plants. In: Rébeillé F, Douce R, eds. *Advances in botanical research. Biosynthesis of vitamins in plants part A: vitamins A, B1, B2, B3, B5*. London, UK: Academic Press, 153–201.

Ono H, Ishii K, Kozaki T, Ogiwara I, Kanekatsu M, Yamada T. 2015. Removal of redundant contigs from de novo RNA-Seq assemblies via homology search improves accurate detection of differentially expressed genes. *BMC Genomics* **16**: 1031.

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

Onyshchenko A, Ruck EC, Nakov T, Alverson AJ. 2019. A single loss of photosynthesis in the diatom order Bacillariales (Bacillariophyta). *American Journal of Botany* **106**: 560–572.

Opsahl S, Benner R. 1997. Distribution and cycling of terrigenous dissolved organic matter in the ocean. *Nature* **386**: 480–482.

Palmieri F, Rieder B, Ventrella A, Blanco E, Do PT, Nunes-Nesi A, Trauth AU, Fiermonte G, Tjaden J, Agrimi G *et al.* 2009. Molecular identification and functional characterization of *Arabidopsis thaliana* mitochondrial and chloroplastic NAD+ carrier proteins. *Journal of Biological Chemistry* **284**: 31249–31259.

Parales RE, Harwood CS. 1992. Characterization of the genes encoding beta-ketoadipate: succinyl-coenzyme A transferase in *Pseudomonas putida*. *Journal of Bacteriology* **174**: 4657–4666.

Parks MB, Nakov T, Ruck EC, Wickett NJ, Alverson AJ. 2018. Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *American Journal of Botany* **105**: 330–347.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**: 785–786.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D *et al.* 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*: 201178.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Ralph SA, van Dooren GG, Waller RF, Crawford MJ, Fraunholz MJ, Foth BJ, Tonkin CJ, Roos DS, McFadden GI. 2004. Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nature Reviews Microbiology* **2**: 203–216.

Rastogi A, Vieira FRJ, Deton-Cabanillas A-F, Veluchamy A, Cantrel C, Wang G, Vanormelingen P, Bowler C, Piganeau G, Hu H *et al.* 2020. A genomics approach reveals the global genetic polymorphism, structure, and functional diversity of ten accessions of the marine model diatom *Phaeodactylum tricornutum*. *ISME Journal* **14**: 347–363.

Río Bártulos C, Rogers MB, Williams TA, Gentekaki E, Brinkmann H, Cerff R, Liaud M-F, Hehl AB, Yarlett NR, Gruber A *et al.* 2018. Mitochondrial glycolysis in a major lineage of eukaryotes. *Genome Biology and Evolution* **10**: 2310–2325.

Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports* **16**: 565–574.

Ruck EC, Nakov T, Jansen RK, Theriot EC, Alverson AJ. 2014. Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. *Genome Biology and Evolution* **6**: 644–654.

Rumpf R, Vernon D, Schreiber D, Birky CW. 1996. Evolutionary consequences of the loss of photosynthesis in Chlamydomonadaceae: phylogenetic analysis of *rrn18* (18S rDNA) in 13 *Polytoma* strains (Chlorophyta). *Journal of Phycology* **32**: 119–126.

Saeed S, Tremp AZ, Sharma V, Lasonder E, Dessens JT. 2020. NAD(P) transhydrogenase has vital non-mitochondrial functions in malaria parasite transmission. *EMBO Reports* **21**: e47832.

Sage RF. 2004. The evolution of C4 photosynthesis. *New Phytologist* **161**: 341–370.

Saier MH Jr, Reddy VS, Tamang DG, Västermark A. 2014. The transporter classification database. *Nucleic Acids Research* **42**: D251–D258.

Savojardo C, Martelli PL, Fariselli P, Casadio R. 2017. SChloro: directing Viridiplantae proteins to six chloroplastic sub-compartments. *Bioinformatics* **33**: 347–353.

Schnarrenberger C, Flechner A, Martin W. 1995. Enzymatic evidence for a complete oxidative pentose phosphate pathway in chloroplasts and an incomplete pathway in the cytosol of spinach leaves. *Plant Physiology* **108**: 609–614.

Sheikhizadeh S, de Ridder D. 2015. ACE: accurate correction of errors using K-mer tries. *Bioinformatics* **31**: 3216–3218.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Smith SR, Abbriano RM, Hildebrand M. 2012. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Research* **1**: 2–16.

Song L, Shankar DS, Florea L. 2016. Rascaf: improving genome assembly with RNA sequencing data. *Plant Genome* **9**: 1–12.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644.

Taniguchi M, Miyake H. 2012. Redox-shuttling between chloroplast and cytosol: integration of intra-chloroplast and extra-chloroplast metabolism. *Current Opinion in Plant Biology* **15**: 252–260.

Tartar A, Boucias DG. 2004. The non-photosynthetic, pathogenic green alga *Helicosporidium* sp. has retained a modified, functional plastid genome. *FEMS Microbiology Letters* **233**: 153–157.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Research* **13**: 2129–2141.

Traller JC, Cokus SJ, Lopez DA, Gaidarenko O, Smith SR, McCrow JP, Gallaher SD, Podell S, Thompson M, Cook O *et al.* 2016. Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnology for Biofuels* **9**: 258.

Tuchman NC, Schollett MA, Rier ST, Geddes P. 2006. Differential heterotrophic utilization of organic compounds by diatoms and bacteria under light and dark conditions. *Hydrobiologia* **561**: 167–177.

Tymoczko JL, Berg JM, Stryer L. 2013. Gluconeogenesis. In: Tymoczko JL, Berg JM, Stryer L, eds. *Biochemistry: a short course*. New York, NY, USA: W.H. Freeman and Company, 299–314.

Villanova V, Fortunato AE, Singh D, Bo DD, Conte M, Obata T, Jouhet J, Fernie AR, Marechal E, Falciatore A *et al.* 2017. Investigating mixotrophic metabolism in the model diatom, *Phaeodactylum tricornutum*. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **372**: 20160404.

Weber APM, Linka N. 2011. Connecting the plastid: Transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. *Annual Review of Plant Biology* **62**: 53–77.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ. 2015. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Contrasting inferences of gene family expansion and contraction in nonphotosynthetic diatoms based on analyses of protein-coding genes from genomes only or genomes and transcriptomes.

Fig. S2 Light and scanning electron micrographs of *Nitzschia* sp. strain Nitz4.

Methods S1 Genome assembly, contaminant removal, annotation, and analysis of heterozygosity in the nuclear genome of *Nitzschia* sp. Nitz4.

Table S1 Plastid genome characteristics of five phylogenetically diverse diatoms.

Table S2 Genome assembly characteristics for *Nitzschia* sp. Nitz4 and other diatoms.

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation

Table S3 Summary of variant calling for the nuclear genome of *Nitzschia* sp. Nitz4.

Table S4 Annotation and localization of carbon metabolism genes in the genome of *Nitzschia* sp. Nitz4.

Table S5 Annotation of carbon transporters in the genome of *Nitzschia* sp. Nitz4.

Table S6 Annotation of carbohydrate-active enzymes in the genome of *Nitzschia* sp. Nitz4.

Table S7 Annotation information for genes in the β-ketoadipate pathway of *Nitzschia* sp. Nitz4.

Table S8 β-Ketoadipate pathway genes in the sequenced genomes and transcriptomes of photosynthetic and nonphotosynthetic diatoms.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Foundation, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and and Tansley insights.
- Regular papers, Letters, Viewpoints, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <26 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**

© 2021 The Authors
New Phytologist © 2021 New Phytologist Foundation