The genome of the nonphotosynthetic diatom, *Nitzschia* sp.: Insights into the metabolic shift to heterotrophy and the rarity of loss of photosynthesis in diatoms

Anastasiia Pendergrass[1,2,3], Wade R. Roberts[1], Elizabeth C. Ruck[1], Jeffrey A. Lewis[1], and Andrew J. Alverson[1]

[1]University of Arkansas, Department of Biological Sciences, 1 University of Arkansas, SCEN 601, Fayetteville, AR 72701 USA, phone: +1 479 575 4886

[2]Current address: Washington University School of Medicine, 425 S. Euclid Avenue, BJCIH 11605, St. Louis, MO 63110 USA, phone: +1 314 362 3365

[3]Author for correspondence

Running head: Metabolism and redox balance in nonphotosynthetic diatoms

Key words: apicoplast, beta-ketoadipate pathway, heterotrophy, mixotrophy, NADPH, pentose phosphate pathway

Abstract

Although most of the tens of thousands of diatom species are obligate photoautotrophs, many mixotrophic species can also use extracellular organic carbon for growth, and a small number of obligate heterotrophs have lost photosynthesis entirely. We sequenced the genome of a nonphotosynthetic diatom, *Nitzschia* sp. strain Nitz4, to determine how carbon metabolism was altered in the wake of this rare and radical trophic shift. Like other groups that have lost photosynthesis, the genomic consequences were most evident in the plastid genome, which is exceptionally AT-rich and missing photosynthesis-related genes. The relatively small (27 Mb) nuclear genome did not differ dramatically in gene and intron density from photosynthetic diatoms. Genome-based models suggest that central carbon metabolism, including a central role for the plastid, remains relatively intact in the absence of photosynthesis. All diatom plastids lack an oxidative pentose phosphate pathway (PPP), leaving photosynthesis as the main source of NADPH. Consequently, nonphotosynthetic diatoms lack the primary source of NADPH required for numerous plastid-localized biosynthetic pathways. Genomic models highlighted similarities between nonphotosynthetic diatoms and apicomplexan parasites for provisioning NADPH in their plastids. The ancestral absence of a plastid PPP might constrain loss of photosynthesis in diatoms compared to archaeplastids, whose plastid PPP continues to produce reducing cofactors following loss of photosynthesis. Finally, *Nitzschia* possesses a complete β-ketoadipate pathway, which was previously known only from fungi and bacteria, that may allow mixotrophic and heterotrophic diatoms to obtain energy through the degradation of abundant plant-derived aromatic compounds.

Abbreviations

1,3biPGA, 1,3 bi phosphoglycerate; CMD, 4-carboxymuconolactone decarboxylase; CMLE, 3-carboxymuconate lactonizing enzyme; ELH, β-ketoadipate enol-lactone hydrolase; EST, expressed sequence tags; FBP, fructose 1,6-bisphosphatase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GK, glucokinase; Gly-3P, Glycerol 3-phosphate; MDH, malate dehydrogenase; ME, malic enzyme; MEP, methylerythritol 4-phosphate pathway (non-mevalonate pathway); MVA, mevalonate pathway; NADP, Nicotinamide adenine dinucleotide phosphate; NTH, NAD(P) transhydrogenase; P3,4O, protocatechuate 3,4-dioxygenase; PC, pyruvate carboxylase; PCA, protocatechuic acid / protocatechuate; PEPS, phosphoenolpyruvate synthase; PFK-1, phosphofructokinase-1; PFK-ATP, ATP-dependent 6-phosphofructokinase; PFK-PPi, pyrophosphate-dependent 6-phosphofructokinase; PGK, phosphoglycerate kinase; PK, pyruvate kinase; PPDK, pyruvate phosphate dikinase; PPP, oxidative pentose phosphate pathway; SCOT, succinyl-CoA:3-ketoacid CoA transferase; TCA cycle, tricarboxylic acid cycle; TH, 3-ketoadipyl-CoA thiolase; TR, 3-ketoadipate:succinyl-CoA transferase

Introduction

Microbial eukaryotes (protists) capture the full phylogenetic breadth of eukaryotic diversity and so exhibit a wide range of ecological, life history, and trophic strategies that include parasitism, autotrophy, heterotrophy, and mixotrophy, i.e., the ability to acquire energy through both autotrophic and heterotrophic means (Burki et al. 2020). Autotrophic and heterotrophic processes respond differently to temperature, so mixotrophs may shift toward heterotrophic metabolism as the ocean warms in response to climate change (Wilken et al. 2013). Given the contribution of mixotrophs to the burial of atmospheric carbon on the sea floor (the "biological pump"), this physiological shift could have profound impacts on the structure and function of marine ecosystems (Mitra et al. 2014; Stoecker et al. 2017). Although the bioenergetics and genomic architecture of microbial photosynthesis are fairly well characterized (Falkowski and Raven 2013), the modes, mechanisms, range of organic substrates and prey items, and the genomic underpinnings of heterotrophy and mixotrophy are highly varied across lineages and, as a result, more poorly understood (Worden et al. 2015). Many protist lineages include a heterogeneous mix of obligately autotrophic, mixotrophic, and obligately heterotrophic species. Studies focused on evolutionary transitions between nutritional strategies in these lineages provide a potentially powerful approach toward understanding the ecological drivers, metabolic properties, and genomic bases of the wide variety of trophic modes in protists.

Diatoms are unicellular algae that are widely distributed throughout marine, freshwater, and terrestrial ecosystems. They are prolific photosynthesizers responsible for some 20% of global net primary production (Field et al. 1998; Smetacek 1999), and like many other protists (Mitra et al. 2014; Godrijan et al. 2020), diatoms are mixotrophic, i.e., they are capable of supporting growth through catabolism of small dissolved organic compounds from the environment. This form of heterotrophy, sometimes described as *osmotrophy*, might include both passive diffusion (osmosis) and active transport of dissolved organic carbon compounds for catabolism. Diatoms import a variety of carbon substrates for both growth (Lewin and Lewin 1967; Tuchman et al. 2006) and osmoregulation (Liu and Hellebust 1976; Petrou and Nielsen 2018; Torstensson et al. 2019), though important questions remain about the full range of organic compounds used by diatoms and whether these vary across species or habitats. Heterotrophic growth has been characterized in just a few species (Hellebust 1971; Kitano et al. 1997; Cerón García et al. 2006), but the phylogenetic breadth of these taxa indicates that mixotrophy is a widespread and likely ancestral trait in diatoms.

Like many plant and algal lineages that are ancestrally and predominantly photoautotrophic, some diatoms have completely lost their ability to photosynthesize and are obligate heterotrophs, relying exclusively on extracellular carbon for growth (Lewin and Lewin 1967). Unlike some species-rich lineages of photoautotrophs, 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 a small number (≈ 20) of species (Frankovich et al. 2018; Onyshchenko et al. 2019). These "apochloritic" diatoms maintain colorless plastids with highly reduced, AT-rich plastid genomes that are devoid of photosynthesis-related genes (Kamikawa, Tanifuji, et al. 2015; Onyshchenko et al. 2019).

Photosynthesis is central to current models of carbon metabolism in diatoms, where carbon metabolic pathways are highly compartmentalized and, in addition, where compartmentalization varies, at least among the model species with sequenced genomes (Smith et al. 2012). For example, diatom plastids house the Calvin cycle, which is involved in carbon fixation and supplies triose phosphates for a number of anabolic pathways such as biosynthesis of branched-chain amino acid, aromatic amino acids, and fatty acids (Smith et al. 2012; Kamikawa et al. 2017). The lower payoff phase of glycolysis, which generates pyruvate, ATP, and NADH, is located in the mitochondrion (Smith et al. 2012). The centric diatom, *Cyclotella nana* (formerly *Thalassiosira pseudonana*), has a full glycolytic pathway in the cytosol, but compartmentalization of glycolysis appears to differ in the raphid pennate diatoms, such as *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). This type of compartmentalization might help diatoms avoid futile cycles, which occur when two metabolic pathways work in opposite directions within the same compartment (e.g., the oxidative pentose phosphate pathway [PPP] vs. Calvin cycle in the plastid; Smith et al. 2012). All this requires careful orchestration of protein targeting, as most carbon metabolism genes are encoded by the nuclear genome. Differences in the targeting sequences of duplicated genes ensure that nuclear-encoded enzymes end up in the correct subcellular compartment (Kroth et al. 2008; Smith et al. 2012). Unlike archaeplastids, diatom plastids lack a PPP, leaving photosystem I as the primary source of endogenous NADPH cofactors for the numerous biosynthetic pathways that occur in the diatom plastid (Michels et al. 2005).

To gain an alternative perspective on carbon metabolism in diatoms, we sequenced the genome of a nonphotosynthetic species from the genus *Nitzschia*. We used the genome to determine the extent to which central carbon metabolism has been restructured in diatoms that no longer photosynthesize. In addition, this rare transition from mixotrophy to obligate heterotrophy could offer new and unbiased insights into the trophic diversity of diatoms more generally. Our analyses revealed a previously unknown carbon metabolic pathway that may allow mixotrophic and heterotrophic diatoms to metabolize plant-derived biomass, and allowed us to formulate a hypothesis to account for why photosynthesis has been lost infrequently in diatoms.

Materials and Methods

Collection and Culturing of Nitzschia sp. Nitz4

We collected a composite sample on 10 November 2011 from Whiskey Creek, which is located in Dr. Von D. Mizell-Eula Johnson State Park (formerly John U. Lloyd State Park), Dania Beach, Florida, USA (26.081330 latitude, –80.110783 longitude). The sample consisted of near-surface plankton collected with a 10 µM mesh net, submerged sand (1 m and 2 m depth), and nearshore wet (but unsubmerged) sand. We selected for nonphotosynthetic diatoms by storing the sample in the dark at room temperature (21°C) for several days before isolating colorless diatom cells with a Pasteur pipette. Clonal cultures were grown in the dark at 21°C on agar plates made with L1+NPM medium (Guillard 1960; Guillard and Hargraves 1993) and 1% Penicillin–Streptomycin–Neomycin solution (Sigma-Aldrich P4083) to retard bacterial growth.

DNA and RNA Extraction and Sequencing

We rinsed cells with L1 medium and removed them from agar plates by pipetting, lightly centrifuged them, and then disrupted them with a MiniBeadbeater-24 (BioSpec Products). We extracted DNA with a Qiagen DNeasy Plant Mini Kit. We sequenced total extracted DNA from a single culture strain, *Nitzschia* sp. Nitz4, with the Illumina HiSeq2000 platform housed at the Beijing Genomics Institute, with a 500-bp library and 90-bp paired-end reads. We extracted total RNA with a Qiagen RNeasy kit and sequenced a 300-bp Illumina TruSeq RNA library using the Illumina HiSeq2000 platform.

Genome and Transcriptome assembly and Annotation

A total of 15.4 GB of 100-bp paired-end DNA reads were recovered and used to assemble the nuclear, plastid, and mitochondrial genomes of *Nitzschia*. We used FastQC (ver. 0.11.5) (Andrews 2010) to check reads quality and then used ACE (Sheikhizadeh and de Ridder 2015) to correct predicted sequencing errors. We subsequently trimmed and filtered the reads with Trimmomatic (ver. 0.32) (Bolger et al. 2014) with settings "ILLUMINACLIP=<TruSeq_adapters.fa>:2:40:15, LEADING=2 TRAILING=2, SLIDINGWINDOW=4:2, MINLEN=30, TOPHRED64."

We used SPAdes (ver. 3.12.0) (Bankevich et al. 2012) with default parameter settings and k-mer sizes of 21, 33, and 45 to assemble the genome. We then used Blobtools (Laetsch and Blaxter 2017) to identify and remove putative contaminant scaffolds. Blobtools uses a combination of GC percentage, taxonomic assignment, and read coverage to identify potential contaminants. The taxonomic assignment of each scaffold was estimated from a DIAMOND BLASTX search (Buchfink et al. 2015) with settings "--max-target-seqs 1 --sensitive --evalue 1e-25" against the UniProt Reference Proteomes database

(Release 2019_06). We estimated read coverage across the genome by aligning all reads to the scaffolds with BWA-MEM (Li and Durbin 2009). We used Blobtools to identify and remove scaffolds that met the following criteria: taxonomic assignment to bacteria, archaea, or viruses; low GC percentage indicative of organellar scaffolds; scaffold length < 500 bp and without matches to the UniProt database. After removing contaminant scaffolds, all strictly paired reads mapping to the remaining scaffolds were exported and reassembled with SPAdes as described above. We also used the Kmer Analysis Toolkit (KAT; Mapleson et al. 2017) to identify and remove scaffolds whose length was > 50% aberrant k-mers, which is indicative of a source other than the *Nitzschia* nuclear genome. A total of 401 scaffolds met this criterion and of these, 340 had no hits to the UniProt database and were removed. A total of 55 of the flagged scaffolds had hits to eukaryotic sequences; these were manually searched against the National Center of Biotechnology Information (NCBI) nonredundant (nr) database collection of diatom proteins with BLASTX and removed if they did not fulfill the following criteria: scaffold length < 1000 bp, no protein domain information, and significant hits to hypothetical or predicted proteins only. This resulted in the removal of an additional 31 scaffolds from the assembly. A detailed summary of the genome assembly workflow is available in Supplementary File 1.

Due to the fragmented nature of short-read genome assemblies, we used Rascaf (Song et al. 2016) and SSPACE (Boetzer et al. 2011) to improve contiguity of the final assembly of the relatively small and gene-dense genome of *Nitzschia* Nitz4. Rascaf uses alignments of paired RNA-seq reads to identify new contig connections, using an exon block graph to represent each gene and the underlying contig relationship to determine the likely contig path. RNA-seq read alignments were generated with Minimap2 (Li 2018) and inputted to Rascaf for scaffolding (Supplementary File 1). The resulting scaffolds were further scaffolded using SSPACE with the filtered set of genomic reads (i.e., ones exported after Blobtools filtering). SSPACE aligns the read pairs and uses the orientation and position of each pair to connect contigs and place them in the correct order (Supplementary File 1). Gaps produced from scaffolding were filled with two rounds of GapCloser (Luo et al. 2012) using both the RNA-seq reads and filtered DNA reads (Supplementary File 1). Each stage of the genome assembly was evaluated with QUAST (ver. 5.0.0) (Gurevich et al. 2013) and BUSCO (ver. 3.0.2) (Simão et al. 2015). BUSCO was run in genome mode separately against both the protist_ensembl and eukaryota_odb9 datasets.

RNA extraction, sequencing, read processing, and assembly of RNA-seq reads followed Parks and Nakov et al. (2018). The *Nitzschia* Nitz4 genome and transcriptome are available through NCBI BioProject PRJNA412514. A genome browser and gene annotations are available through the Comparative Genomics (CoGe) web platform (https://genomevolution.org/coge/) under genome ID 58203.

Analysis of Heterozygosity

We mapped the trimmed paired-end reads to the genome assembly using BWA-MEM (version 0.7.17-r1188) (Li and Durbin 2009) and used Picard Tools (version 2.17.10) (https://broadinstitute.github.io/picard) to add read group information and mark duplicate read pairs. The resulting BAM file was used as input to the Genome Analysis Toolkit (GATK version 3.5.0) to call variants (SNPs and indels) using the HaplotypeCaller tool (DePristo et al. 2011; Poplin et al. 2018). SNPs and indels were separately extracted from the resulting VCF file and separately filtered to remove low quality variants. SNPs were filtered as follows: MQ > 40.0, SOR > 3.0, QD < 2.0, FS > 60.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, ReadPosRankSum > 8.0. Indels were filtered as follows: MQ < 40.0, SOR > 10.0, QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, ReadPosRankSum > 20.0. The filtered variants were combined and used as "known" variants for base recalibration. We then performed a second round of variant calling with HaplotypeCaller and the recalibrated BAM file. We extracted and filtered SNPs and indels as described above before additionally filtering both variant types by approximate read depth (DP), removing those with depth below the 5th percentile (DP < 125) and above the 95th percentile (DP > 1507). The final filtered SNPs and indels were combined, evaluated using GATK's VariantEval tool, and annotated using SnpEff (version 4.3t) (Cingolani et al. 2012) with the *Nitzschia* Nitz4 gene models.

Genome Annotation

We used the Maker pipeline (ver. 2.31.8) (Cantarel et al. 2008) to identify protein-coding genes in the nuclear genome. We used the assembled *Nitzschia* Nitz4 transcriptome (Maker's expressed sequence tag [EST] evidence) and the predicted proteome of *Fragilariopsis cylindrus* (GCA_001750085.1) (Maker's protein homology evidence) to inform the gene predictions. RepeatModeler (ver. 2.0) (Flynn et al. 2019) was used for de novo identification and compilation of transposable element families found in the *Nitzschia* Nitz4 genome. The resulting repeat library was searched using BLASTX with settings "-num_descriptions 1 -num_alignments 1 -evalue 1e-10" against the UniProt Reference Proteomes to exclude repeats that had significant protein hits using ProtExcluder (ver. 1.2) (Campbell et al. 2014). The final custom repeat library was used as input to the Maker annotation pipeline described below.

We used Augustus (ver. 3.2.2) for de novo gene prediction with settings "max_dna_len = 200,000" and "min_contig = 300." We trained Augustus with the annotated gene set for *Phaeodactylum tricornutum* (ver. 2), which we filtered as follows: (1) remove "hypothetical" and "predicted" proteins, (2) remove all but one splice variant of a gene, (3) remove genes with no introns, and (4) remove genes that overlap with neighboring gene models or the 1000 bp flanking regions of adjacent genes, as these regions

are included in the training set along with the enclosed gene. When possible, we generated UTR annotations for the selected gene models by subtracting 5' and 3'CDS coordinates from the corresponding 5' and 3' end coordinates of the associated mRNA sequence; among these, we only retained those with UTRs that extended ≥ 25 bp beyond both the 5' and 3' ends of the CDS. The final filtered training set included a total of 726 genes.

We generated a separate set of gene models for UTR training with all of the filtering criteria described above except that we retained intronless genes based on the assumption that intron presence or absence was less relevant for training the UTR annotation parameters. In addition, we retained only those genes with UTRs that were ≥ 40 bp in length. In total, our UTR training set included 531genes. We trained and optimized the Augustus gene prediction parameters (with no UTRs) on the first set of 726 genes and optimized UTR prediction parameters with the second set of 531 genes. We then used these parameters to perform six successive gene predictions within Maker. We assessed each Maker annotation for completeness using BUSCO (ver. 2.0) with the 'eukaryota_odb9' and 'protist_ensembl' databases (Simão et al. 2015). For the five subsequent runs, we followed recommendations of the Maker developers and used a different ab initio gene predictor, SNAP (Korf 2004), trained with Maker-generated gene models from the previous run. Although the five SNAP-based Maker runs discovered more genes, the number of complete BUSCOs did not increase and the number of genes with Maker annotation edit distance (AED) scores less than 0.5 decreased slightly from 0.98 to 0.97 (Supplementary Table 1). We therefore used the gene predictions from the second Maker round (the first SNAP-based round) as the final annotation (Supplemental Table 1). Finally, in our search for genes related to carbon metabolism, we found several genes that were not in the final set of Maker-based annotations. The coding regions of these genes were manually annotated.

Construction and Analysis of Orthologous Clusters

We used Orthofinder (ver. 1.1.4) with default parameters to build orthologous clusters from the complete set of predicted proteins from the genomes of *Nitzschia* Nitz4, *Fragilariopsis cylindrus*, *Phaeodactylum tricornutum*, *Cyclotella nana*, and the transcriptome of a photosynthetic *Nitzschia* species (strain CCMP2144). We characterized species-specific *Nitzschia* Nitz4 genes by using BLASTP to search orthogroups consisting exclusively of *Nitzschia* Nitz4 genes against NCBI's nr database (GenBank release 223.0 or 224.0).

Characterization of Carbon Metabolism Genes

We performed a thorough manual annotation of primary carbon metabolism genes in the *Nitzschia* Nitz4 genome. We started with a set of core carbon metabolic genes from diatoms (Smith et al.

2012; Kamikawa et al. 2017) and used these annotations as search terms to download all related genes, from both diatoms and non-diatoms, from NCBI's nr database (release 223.0 or 224.0). We then filtered these sets to include only RefSeq accessions. As we proceeded with our characterizations of carbon metabolic pathways, we expanded our searches as necessary to include genes that were predicted to be involved in those pathways. We constructed local BLAST databases of *Nitzschia* Nitz4 genome scaffolds, predicted proteins, and assembled transcripts separately searched the set of sequences for a given carbon metabolism gene against these three *Nitzschia* Nitz4 databases. In most cases, each query had thousands of annotated sequences on GenBank, so if the GenBank sequences for a given gene were, in fact, homologous and accurately annotated, then we expected our searches to return thousands of strong query–subject matches, and this was often the case. We did not consider further any putative carbon genes with weak subject matches (e.g., tens of weak hits vs. the thousands of strong hits in verified, correctly annotated sequences). We extracted the *Nitzschia* Nitz4 subject matches and used BLASTX or BLASTP to search the transcript or predicted protein sequence against NCBI's nr protein database to verify the annotation. *Nitzschia* Nitz4 subject matches in scaffold regions were checked for overlap with annotated genes, and if the subject match overlapped with an annotated gene, the largest annotated CDS in this region was extracted and searched against the nr protein database with BLASTX to verify the annotation. As necessary, we repeated this procedure for other diatom species that were included in the OrthoFinder analysis.

To identify carbon transporters in *Nitzschia* Nitz4, we searched for genes using the Transporter Classification Database (TCDB; Saier et al. 2014) for the following annotations: Channels/Pores, Electrochemical Potential-Driven Transporters, Primary Active Transporters, and Group Translocators. We used the Carbohydrate-Active enZYmes database (CAZy; Lombard et al. 2014) to identify potential saccharide degradation genes.

Prediction of Protein Localization

We used the software programs 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. Our approach was slightly modified from that of Traller et al. (2016). Proteins with a SignalP-, ASAFind, and HECTAR-predicted plastid signal peptide were classified as plastid targeted. We placed less weight on ChloroP target predictions, which are optimized for targeting to plastids bound by two membranes rather than the four-membrane plastids found in diatoms. Proteins predicted as mitochondrion-targeted by any two of the TargetP, MitoProt, and HECTAR programs 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. Most of the remaining proteins were cytoplasmic.

Results

To understand the genomic signatures of the transition from mixotrophy to heterotrophy, we collected and isolated nonphotosynthetic *Nitzschia* (Onyshchenko et al. 2019) from a locale known to harbor several species in moderate abundance (Blackburn et al. 2009). We generated an annotated genome for one strain, *Nitzschia* sp. strain Nitz4. *Nitzschia* Nitz4 grew well in near-axenic culture for approximately one year before dying due to size diminution after failing to reproduce sexually. Based on both morphological (Fig. 1) and molecular data, we were not able to assign a name to this species, but phylogenetic analyses placed it within a clade of Bacillariales consisting exclusively of apochloritic *Nitzschia* (Onyshchenko et al. 2019).

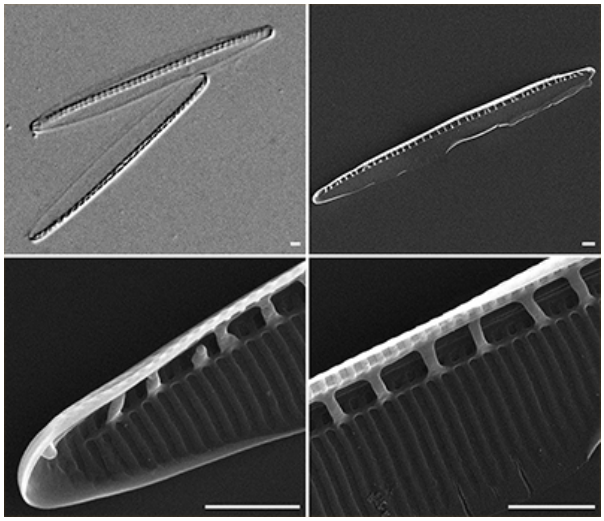

Figure 1. Light and scanning electron micrographs of *Nitzschia* sp. strain Nitz4. Scale bar = 1 micrometer.

Plastid Genome Reduction

We examined the fully sequenced plastid genomes of *Nitzschia* Nitz4 (Onyshchenko et al. 2019) and another apochloritic *Nitzschia* (NIES-3581; Kamikawa, Tanifuji, et al. 2015) to understand how loss of photosynthesis impacted the plastid genome. Both plastid genomes revealed a substantial reduction in the amount of intergenic DNA and wholesale loss of genes encoding the photosynthetic apparatus, including those involved in carbon fixation, photosystems I and II, the cytochrome b6f complex, and porphyrin and thiamine metabolism (Table 1). Compared to the plastid genomes of photosynthetic diatoms, which contain a core set of 122 protein and 30 RNA genes (Ruck et al. 2014), the *Nitzschia* Nitz4 plastid genome contains just 95 genes that mostly serve housekeeping functions (e.g., rRNA, tRNA,

and ribosomal protein genes). The plastid genome also experienced a substantial downward shift in GC content, from >30% in photosynthetic species to just 22.4% in *Nitzschia* Nitz4 (Table 1), a pattern that is common in streamlined bacterial genomes (Giovannoni et al. 2014) and may reflect mutational bias (Hershberg and Petrov 2010) or nitrogen frugality (GC base pairs contain more nitrogen than AT base pairs; Dufresne et al. 2005).

Table 1. Plastid genome characteristics for four photosynthetic diatoms and the nonphotosynthetic diatom, *Nitzschia* sp. Nitz4.

	Cyclotella nana	*Phaeodactylum tricornutum*	*Fragilariopsis cylindrus*	*Nitzschia palea*	*Nitzschia* sp. (Nitz4)
Genome size	128,814	117,369	123,275	119,447	67,764
GC content (%)	30.7	32.6	30.8	32.7	22.4
Total gene count[1]	159	162	141	160	95
Gene count (photosystem I)	10	10	9	10	0
Gene count (photosystem II)	18	18	18	18	0
Reference	Oudot-Le Secq et al. (2007)	Oudot-Le Secq et al. (2007)	Zheng et al. (2019)	Crowell et al. (2019)	Onyshchenko et al. (2019)

[1]Gene count includes rRNAs, tRNAs, and ORFs of unknown function.

Nuclear Genome Characteristics

After removal of putative contaminant sequences, a total of 58,141,034 100-bp paired-end sequencing reads were assembled into 992 scaffolds totaling 27.2 Mb in length (Table 2). The genome was sequenced to an average coverage depth of 221 reads. Half of the *Nitzschia* genome was contained in 86 scaffolds, each one longer than 91,515 bp (assembly N50). The average scaffold length was 27,448 bp, and the largest scaffold was 479,884 bp in length.

The *Nitzschia* genome is similar in size to other small diatom genomes (Table 2) but is the smallest so far sequenced from Bacillariales, the lineage that includes *Fragilariopsis cylindrus* (61 Mb) and *Pseudo-nitzschia multistriata* (59 Mb) (Table 2). The genomic GC content (48%) is similar to other diatom nuclear genomes (Table 2) and indicates that the exceptional decrease in GC content experienced by the plastid genome (Onyshchenko et al. 2019) did not similarly affect the nuclear (Table 2) or mitochondrial genomes (Guillory et al. 2018).

Table 2. Nuclear genome characteristics for two phylogenetically diverse model diatom species (*C. nana* and *P. tricornutum*) and three species from the order Bacillariales (*F. cylindrus*, *P. multistriata*, and *Nitzschia* Nitz4).

	Cyclotella nana	*Phaeodactylum tricornutum*	*Fragilariopsis cylindrus*	*Pseudo-nitzschia multistriata*	*Nitzschia* sp. Nitz4
Genome size (Mb)	32.4	27.4	61.1	55.1	27.2
GC content (%)	47	48.8	38.8	46.4	48
Protein-coding gene count	11,673	10,408	18,111	not annotated	9,359
Complete/Duplicated BUSCO count[1]	233/10	240/10	202/47	not annotated	248/6
Average gene size (bp)	992	1621	1575	not annotated	2083
Gene density (per Mb)	360	380	296	not annotated	343
Intron density (per Mb)	552	298	373	not annotated	340
	GCA_000149405.2	GCA_000150955.2	GCA_001750085.1	GCA_900005105.1	PRJNA412514
Reference	Armbrust et al. (2004)	Bowler et al. (2008)	Mock et al. (2017)	Basu et al. (2017)	This study

[1]BUSCO analyses performed in protein mode against the eukaryota_odb9 database.

We identified a total of 9,359 protein gene models in the *Nitzschia* Nitz4 nuclear genome. This number increased slightly, to 9,373 coding sequences, when alternative isoforms were included. The vast majority of gene models (9,153/9,359) were supported by transcriptome or protein evidence based on alignments of assembled transcripts (*Nitzschia* Nitz4) and/or proteins (*F. cylindrus*) to the genome. To determine completeness of the genome, we used the Benchmarking Universal Single-Copy Orthologs (BUSCO) pipeline, which characterizes the number of conserved protein-coding genes in a genome (Simão et al. 2015). The *Nitzschia* Nitz4 gene models included a total of 254 of 303 eukaryote BUSCOs (Table 2). Although *Nitzschia* Nitz4 has the fewest genes of all diatoms sequenced to date, it has a slightly greater BUSCO count than other diatom genomes (Table 2), indicating that the smaller gene number is not due to incomplete sequencing of the genome. Additional sequencing of Bacillariales genomes is necessary to determine whether the smaller genome size of *Nitzschia* Nitz4 is atypical for the lineage.

Gene models from *C. nana*, *F. cylindrus*, *Nitzschia* sp. CCMP2144, and *Nitzschia* Nitz4 clustered into 10,954 multigene (≥ 2 gene models) orthogroups (Supplemental Table 2). A total of 8,078 of the *Nitzschia* Nitz4 genes were assigned to orthogroups, with 13 genes falling into a total of five Nitz4-specific orthogroups (Supplemental Table 2). A total of 1,295 genes did not cluster with sequences from other species (Supplemental Table 2)—these likely represent a combination of genes that are specific to *Nitzschia* Nitz4 or are shared with closely related but unsequenced *Nitzschia* species.

Finally, we characterized single nucleotide polymorphisms and small structural variants in the genome of *Nitzschia* Nitz4 (Supplemental Table 3). A total of 77,115 high quality variants were called across the genome, including 62,528 SNPs, 8,393 insertions, and 6,272 deletions (Supplemental Table 3). From these variants, the estimated genome-wide heterozygosity was 0.0028 (1 variant per 353 bp), and included 13,446 synonymous and 10,001 nonsynonymous SNPs in protein-coding genes (Supplemental Table 3). *Nitzschia* Nitz4 had a higher proportion of nonsynonymous to synonymous substitutions (π_N/π_S = 0.74) than was reported for the diatom, *P. 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).

Central Carbon Metabolism

A key question related to the transition from autotrophy to heterotrophy is how metabolism changes to support fully heterotrophic growth. To address this, we carefully examined and characterized central carbon metabolism in *Nitzschia* Nitz4 to identify genomic changes associated with the switch to heterotrophy. Classically, the major pathways for central carbon utilization in heterotrophs include glycolysis, the oxidative pentose phosphate pathway (PPP), and the Krebs cycle [or tricarboxylic acid (TCA) 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-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 PPP runs parallel to glycolysis, but unlike glycolysis, is largely anabolic, generating ribose-5-phosphates (a nucleotide precursor) and the key reducing agent 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 (Berg et al. 2002; Smith et al. 2012). 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-1, PK, FBP).

In diatoms, flux through central metabolism is also controlled by compartmentalization of different portions of metabolic pathways (Smith et al. 2012). For example, in photosynthetic diatoms the oxidative branch of the PPP is only present in the cytoplasm, whereas an incomplete nonoxidative PPP is present in the plastid. As a result, without the NADPH produced by photosystem I, the source of NADPH for biosynthetic pathways that remain localized to the plastid in nonphotosynthetic diatoms is unclear. As a first step in understanding heterotrophic metabolism in *Nitzschia* Nitz4, we examined the genome for differences in the presence, absence, copy number, or enzyme compartmentalization of genes involved in carbon metabolism. We searched the nuclear genome for genes involved in biosynthesis of the light-harvesting pigments, fucoxanthin and diatoxanthin (Kuczynska et al. 2015). We did this by querying the predicted proteins and genomic scaffolds of *Nitzschia* Nitz4 against protein databases for lycopene beta cyclase, violaxanthin de-epoxidase, and zeaxanthin epoxidase sequences from *C. nana*, *P. tricornutum*, and *F. cylindrus* as described in Kuczynska et al. (2015). We found no evidence of any genes or pseudogenes related to xanthophyll biosynthesis in *Nitzschia* Nitz4—an unpigmented, colorless diatom—indicating that loss of photosynthesis-related genes was not restricted to the plastid genome.

In general, cytosolic glycolysis in *Nitzschia* Nitz4 does not differ dramatically from other diatoms. For example, despite wholesale loss of plastid genes involved directly in photosynthesis, we detected no major changes in the presence or absence of nuclear genes involved in central carbon metabolism (Supplementary Table 4). Like other other raphid pennate diatoms, *Nitzschia* Nitz4 is missing cytosolic enolase, which catalyzes the penultimate step of glycolysis, indicating that *Nitzschia* Nitz4 lacks a full cytosolic glycolysis pathway (Fig. 2 and Supplementary Table 4). *Nitzschia* Nitz4 does have both plastid- and mitochondrial-targeted enolase genes, indicating that the later stages of glycolysis occur in these compartments (Figs. 2 and 3; Supplementary Table 4). *Nitzschia* Nitz4 has two copies of the genes encoding cytosolic ATP-dependent phosphofructokinase (PFK-ATP) and phosphoglycerate kinase (PGK), two genes which copy numbers vary across diatoms (Smith et al. 2012). PFK-ATP performs the irreversible phosphorylation reaction that is considered the first "committed" step of glycolysis. In contrast, PFK-PPi is a reversible enzyme that, depending on the concentration of reactants, can participate in either glycolysis or gluconeogenesis. Thus, modulation of PFK-ATP activity is one of the major ways cells control flux through glycolysis (Kroth et al. 2008; Smith et al. 2012; Traller et al. 2016). As a result, duplication of PFK-ATP in *Nitzschia* Nitz4 might increase glycolytic flux during heterotrophic growth. PGK is a reversible enzyme used in both glycolysis and gluconeogenesis. Interestingly, while the enzymes involved in the latter half of glycolysis are localized to the mitochondria in most diatoms (Kroth et al. 2008; Smith et al. 2012), PGK is the sole exception. Unlike other diatoms, however, neither PGK gene in *Nitzschia* Nitz4 contains a mitochondrial targeting signal—instead, one copy is predicted to be plastid-targeted and the other is cytosolic (Fig. 2 and Supplementary Table 4). We verified the absence of

a mitochondrial PGK thorough careful manual searching of both the scaffolds and the transcriptome. Lack of a mitochondrial PGK suggests that *Nitzschia* Nitz4 either lacks half of the mitochondrial glycolysis pathway or that one or both of the plastid and cytosolic PGK genes are, in fact, dual-targeted to the mitochondria. However, plastid targeting of PGK may be important for NADPH production in the plastid (see Discussion).

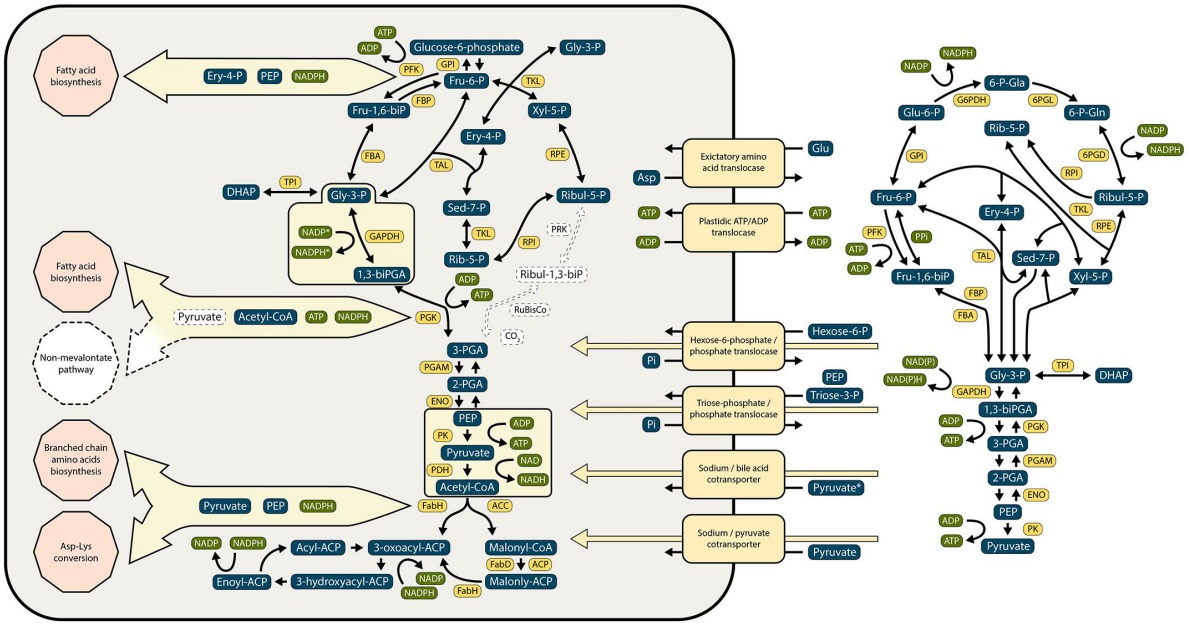

Figure 2. Genome-based model of carbon metabolism in the plastid (green square on left) and cytosol (right) of the nonphotosynthetic diatom, *Nitzschia* sp. Nitz4. Enzymes present in the genome are shown by yellow boxes, and missing enzymes or pathways (arrows) are shown in white. Chemical substrates and products are shown in blue boxes, enzymes encoded by genes present in *Nitzschia* Nitz4 are shown in dark yellow, and reactions that produce or use energy or reducing cofactors are shown in green. Missing genes or pathways are shown in white.

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 (Supplementary Table 4). Interestingly, *Nitzschia* Nitz4 appears to have lost several pyruvate conversion genes, most dramatically in the plastid. For example, the raphid pennate diatoms, *P. tricornutum* and *F. cylindrus*, use pyruvate phosphate dikinase (PPDK) to perform the first step of gluconeogenesis in the plastid, whereas *C. nana* uses a phosphoenolpyruvate synthase (PEPS) that may have been acquired by horizontal gene transfer (Smith et al. 2012). *Nitzschia* Nitz4 lacks plastid-localized

copies of both enzymes, however. Additionally, in contrast to photosynthetic diatoms, *Nitzschia* Nitz4 lacks a plastid-localized pyruvate carboxylase (PC). 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 (Fig. 3). Indeed, 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 taxa.

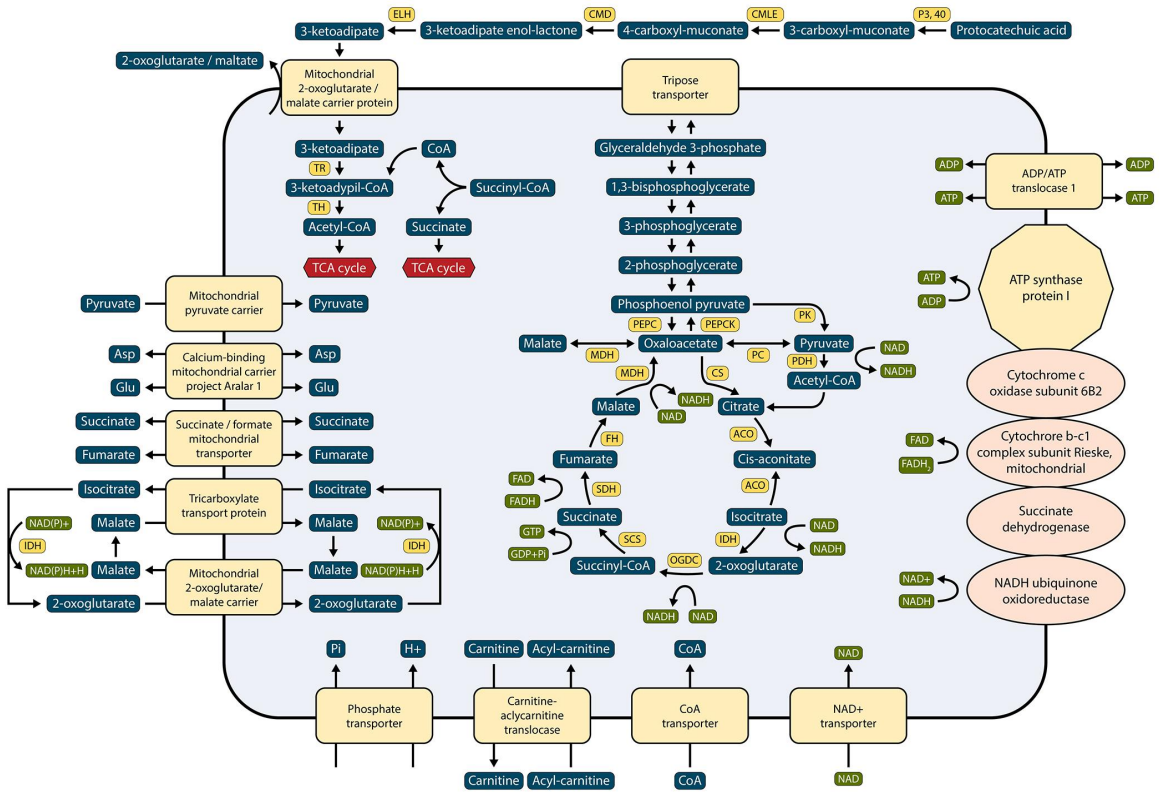

Figure 3. Genome-based model of carbon metabolism in the mitochondria of the nonphotosynthetic diatom, *Nitzschia* sp. Nitz4. Chemical substrates and products are shown in blue boxes, enzymes encoded by genes present in *Nitzschia* Nitz4 are shown in dark yellow, and reactions that produce or use energy or reducing cofactors are shown in green.

Many photosynthesis-related molecules—such as chlorophylls, carotenoids, and plastoquinones—are derivatives of isoprenoids that are produced by the mevalonate pathway (MVA) in

the cytosol or, in plastids, by the methylerythritol 4-phosphate pathway (MEP), also known as the non-mevalonate pathway. The absence of MEP pathway in groups such as apicomplexans, chlorophytes, and most red algae suggests that it is not necessary to maintain both MVA and MEP pathways for isoprenoid biosynthesis (Rohmer 1999). The genome of *Nitzschia* sp. Nitz4 has lost its plastid-localized MEP (Fig. 2), indicating an inability to synthesize carbon- and NADPH-consuming isoprenoids—the precursors of photosynthetic pigments—in the plastid. A previous transcriptome study of a closely related nonphotosynthetic *Nitzschia* found a cytosolic MVA but no plastid-localized MEP as well (Kamikawa et al. 2017)

Finally, we also annotated and characterized carbon transporters and carbohydrate-cleaving enzymes in *Nitzschia* Nitz4 (Supplementary File 2 and Supplementary Table 5). These analyses suggested that the switch to heterotrophy in *Nitzschia* Nitz4 was not accompanied by expansion and diversification of organic carbon transporters and polysaccharide-digesting enzymes (Supplementary File 2).

β-ketoadipate Pathway

One goal of our study was to determine whether apochloritic diatoms, as obligate heterotrophs, have genes that allow them to exploit carbon sources not used by their photosynthetic relatives. To do this, we focused our search on genes and orthogroups specific to *Nitzschia* Nitz4. Although we could not identify or ascribe functions to most of these genes, five of them returned strong but low similarity matches to bacterial-like genes annotated as intradiol ring-cleavage dioxygenase or protocatechuate 3,4-dioxygenase (P3,4O), an enzyme involved in the degradation of aromatic compounds through intradiol ring cleavage of protocatechuic acid (protocatechuate, or PCA), a hydroxylated derivative of benzoate (Harwood and Parales 1996). Additional searches of the nuclear scaffolds revealed three additional P3,4O genes not annotated by Maker (Supplementary Table 6). All eight of the P3,4O genes are located on relatively long scaffolds (14–188 kb) without any indication of a chimeric- or mis-assembly based on both read depth and consistent read-pair depth. Moreover, the scaffolds also contain genes with strong matches to other diatom genomes (see CoGe genome browser). Taken together, all of the available evidence suggests that these genes are diatom in origin and not the product of bacterial contamination in our conservatively filtered assembly. Later searches revealed homologs for some of the P3,4O genes in other diatom species (see below), providing further support that these are bona fide diatom genes.

In addition to the P3,4O genes, targeted searches of the nuclear scaffolds revealed that *Nitzschia* Nitz4 has a nearly complete set of enzymes for the β-ketoadipate pathway, including single copies of the 3-carboxymuconate lactonizing enzyme (CMLE), 4-carboxymuconolactone decarboxylase (CMD), and β-ketoadipate enol-lactone hydrolase (ELH) (Fig. 3 and Supplementary Table 6). All of the genes were transcribed. P3,4O is the first enzyme in the protocatechuate branch of the β-ketoadipate pathway and is

responsible for ortho-cleavage of the activated benzene ring between two hydroxyl groups, resulting in β-carboxymuconic acid (Harwood and Parales 1996). Following fission of the benzene ring, the CMLE, CMD, and ELH enzymes catalyze subsequent steps in the pathway, and in the final two steps, β-ketoadipate is converted to the TCA intermediates succinyl-Coenzyme A and acetyl-Coenzyme A (Fig. 4; Harwood and Parales 1996). The enzymes encoding these last two steps—3-ketoadipate:succinyl-CoA transferase (TR) and 3-ketoadipyl-CoA thiolase (TH)—are missing from the *Nitzschia* genome. 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 the *Nitzschia* genome, Succinyl-CoA:3-ketoacid CoA transferase (SCOT) and 3-ketoacyl-CoA thiolase (3-KAT), respectively (Parales and Harwood 1992; Kaschabek et al. 2002).

Although we found evidence of parts of the β-ketoadipate pathway in other sequenced diatom genomes, only the two *Nitzschia* species—one photosynthetic (CCMP2144) and one not (Nitz4)—possessed a complete pathway (Supplemental Table 7). If the β-ketoadipate pathway serves an especially important role in energy production in *Nitzschia* Nitz4, then we hypothesized that its genome might contain additional gene copies for one or more parts of the pathway. *Nitzschia* Nitz4 has eight copies of the P3,4O gene, compared to four or fewer copies in other diatoms.

Discussion

The Genomic Signatures of Loss of Photosynthesis in Diatoms

Since their origin some 200 Mya, diatoms have become an integral part of marine, freshwater, and terrestrial ecosystems worldwide. Diatoms alone account for 40% of primary productivity in the oceans, making them cornerstones of the "biological pump," which describes the burial of fixed carbon in the sea floor. Photosynthetic output by diatoms is likely a product of their sheer numerical abundance, species richness, and an apparent competitive advantage in areas of upwelling and high turbulence (Falkowski et al. 2004).

Like the genomes of the model diatoms *C. nana* (Armbrust et al. 2004) and *P. tricornutum* (Bowler et al. 2008)—both of which are photosynthetic—the *Nitzschia* Nitz4 genome is relatively small and gene-dense. Repetitive sequences such as transposable elements are present but in low abundance. Levels of genome-wide heterozygosity are generally <1% in sequenced diatom genomes, ranging from 0.2% in *Pseudo-nitzschia multistriata* (Basu et al. 2017) to nearly 1% in *C. nana* (Armbrust et al. 2004), *P. tricornutum* (Bowler et al. 2008), and *F. cylindrus* (Mock et al. 2017). The relatively low observed heterozygosity (<0.2%) and high π_N/π_S (0.74) in *Nitzschia* Nitz4 might reflect a small population size and/or low rate of sexual reproduction (Tsai et al. 2008), hypotheses that cannot be tested with a single genome sequence. Given the small size of the genome, the ability to withstand antibiotic treatments for

removal of bacterial DNA contamination, and the abundance of apochloritic *Nitzschia* species in certain known locales worldwide (Kamikawa, Yubuki, et al. 2015; Onyshchenko et al. 2019), population-level genome resequencing should provide a feasible path forward for understanding the frequency of sexual reproduction, the effective population sizes, and the extent to which their genomes have been shaped by selection or drift.

Experimental data have shown that mixotrophic diatoms can use a range of carbon sources for growth (Hellebust and Lewin 1977; Tuchman et al. 2006), though the exact sources of carbon in strictly heterotrophic, nonphotosynthetic diatoms is unclear, however, and identifying the carbon species used by these diatoms was a principal goal of this study. The *Nitzschia* Nitz4 genome sequence allowed us to compile the gene models and make such predictions based on the presence of specific carbon-related genes and pathways or, alternatively, based on the expansions of gene families related to specific modes of carbon metabolism relative to photosynthetic species. Other than wholesale loss of genes involved in photosynthesis and pigmentation, other genomic signatures of loss of photosynthesis were more subtle. For example, we detected no major changes in the presence or absence of genes involved in central carbon metabolism and limited evidence for expansions of gene families associated with carbon acquisition or carbon metabolism. One exception was the presence in *Nitzschia* Nitz4 of two copies of cytosolic PFK-ATP, which is the "committed" step of glycolysis and one of the major ways cells control flux through glycolysis. Duplication of PFK-ATP in *Nitzschia* Nitz4 might increase glycolytic flux and, consequently, may be linked to the loss of carbon fixation in nonphotosynthetic diatoms. By examining possible alternative pathways by which *Nitzschia* Nitz4 acquires carbon from the environment, we did find evidence that *Nitzschia* Nitz4 possesses a complete β-ketoadipate pathway that may allow it to catabolize plant-derived aromatic compounds. As discussed below, however, all or parts of this pathway were found in photosynthetic species as well.

Species in the pennate diatom clade (to which *Nitzschia* Nitz4 belongs) are mixotrophic, and mixotrophy is naturally thought to be the intermediate trophic strategy between autotrophy and heterotrophy (Figueroa-Martinez et al. 2015), so genome comparisons to other raphid pennate diatoms might conceal obvious features related to obligate versus facultative heterotrophy. For example, differences between mixotrophs and strict heterotrophs could involve other kinds of changes not investigated here including, for example, amino acid substitutions, transcriptional regulation, or post-translational modifications of key enzymes. Detailed flux analysis of growth of most closely related photosynthetic and nonphotosynthetic diatoms under a variety of conditions is a logical next step for connecting genomic changes to metabolism.

A β-ketoadipate Pathway in Diatoms

We used the *Nitzschia* genome sequence to develop a comprehensive model of central carbon metabolism for nonphotosynthetic diatoms with the goal of understanding how they meet their energy demands in the absence of photosynthesis. We systematically searched for known transporters of organic compounds and other carbon-active enzymes to determine whether *Nitzschia* was enriched for such genes, either through acquisition of novel genes or duplication of existing ones. By widening our search to include *Nitzschia*-specific genes and orthogroups that might contain novel or unexpected carbon-related genes, we found one gene—P3,4O—that appeared to be both unique and highly duplicated in *Nitzschia*. The major role of P3,4O is degradation of aromatic substrates through the β-ketoadipate pathway. Additional searches showed that the photosynthetic and nonphotosynthetic *Nitzschia* species investigated here both possess a complete β-ketoadipate pathway, which is typically associated with soil-dwelling fungi and eubacteria (Stanier and Ornston 1973; Harwood and Parales 1996; Fuchs et al. 2011). The pathway consists of two main branches, one of which converts protocatechuate (PCA) and the other catechol, into β-ketoadipate (Harwood and Parales 1996). Diatoms possess the PCA branch of the pathway. Intermediate steps of the PCA branch differ between fungi and bacteria, and diatoms possess a bacterial-like PCA branch (Harwood and Parales 1996).

The β-ketoadipate pathway serves as a funneling pathway for the metabolism of numerous naturally occurring PCA or catechol precursors, all of which feed into the β-ketoadipate pathway through either PCA or catechol (Harwood and Parales 1996; Fuchs et al. 2011). The precursor molecules are derivatives of aromatic environmental pollutants or natural substrates such as plant-derived lignin (Harwood and Parales 1996; Masai et al. 2007; Fuchs et al. 2011; Wells and Ragauskas 2012). As a key component of plant cell walls, lignin is of particular interest because it is one of the most abundant biopolymers on Earth. Lignin from degraded plant material is a major source of naturally occurring aromatic compounds that feed directly into numerous bacterial and fungal aromatic degradation pathways, including the β-ketoadipate pathway.

The first gene in the β-ketoadipate pathway, P3,4O, catalyzes aromatic ring fission which is presumably the most difficult step because of the high resonance energy and enhanced chemical stability of the carbon ring structure of PCA (Fuchs et al. 2011). It is noteworthy that *Nitzschia* Nitz4 possesses eight copies of the P3,4O gene, whereas other diatoms have four or fewer copies. Expansion of the P3,4O gene family in *Nitzschia* Nitz4 suggests that it might be more efficient than photosynthetic diatoms in cleaving the aromatic ring structure of PCA—the putative bottleneck reaction of this pathway.

The β-ketoadipate pathway is typically found in fungi and bacteria (Harwood and Parales 1996), so its presence in *Nitzschia* Nitz4 was unexpected. Follow-up analyses revealed a complete pathway in the transcriptome of a photosynthetic *Nitzschia* (CCMP2144) and parts of the pathway in other sequenced

diatom genomes. These findings show that: (1) even though it was discovered in *Nitzschia* Nitz4, the β-ketoadipate pathway is not restricted to nonphotosynthetic diatoms, and (2) the pathway might be an ancestral feature of diatom genomes. That diatoms have a bacterial rather than fungal form of the pathway raises important questions about the origin and spread of this pathway across the tree of life. These questions can best be addressed through comparative phylogenetic analyses of a broader range of diatom, stramenopile, and other pro- and eukaryotic genomes.

Although both the photosynthetic and nonphotosynthetic *Nitzschia* species analyzed here possess a full and transcribed set of genes for a β-ketoadipate pathway, the genomic prediction—that diatoms with this pathway metabolize an abundant, naturally occurring aromatic compound into products (Succinyl- and Acetyl-CoA) that feed directly into the TCA cycle for energy—requires experimental validation. The similarly surprising discovery of genes for a complete urea cycle in the genome of *Cyclotella nana* (Armbrust et al. 2004) was later validated experimentally (Allen et al. 2011), but in contrast and despite some effort, the genomic evidence in support of C4 photosynthesis in diatoms has not been fully validated (Kustka et al. 2014; Clement et al. 2016; Ewe et al. 2018). It will be important to show, for example, whether diatoms can import PCA and whether the terminal products of the pathway are converted into cellular biomass or, alternatively, whether the pathway is for catabolism of PCA or related compounds. If it is indeed for cellular catabolism, it will be important to determine how much it contributes to the overall carbon budget of photosynthetic and nonphotosynthetic diatoms alike. The piecemeal β-ketoadipate pathways in some diatom genomes raise important questions about the functions of these genes, and possibly the pathway overall, across diatoms.

NADPH Limitation and Constraints on the Loss of Photosynthesis in Diatoms

Plastids, in all their forms across the eukaryotic tree of life, are the site of essential biosynthetic pathways and, consequently, remain both present and biochemically active following loss of their principal function, photosynthesis (Neuhaus and Emes 2000). Most essential biosynthetic pathways in the plastid require the reducing agent, NADPH, which is generated through the light-dependent reactions of photosynthesis or the oxidative PPP. The primary-plastid-containing Archaeplastida—including glaucophytes (Fester and Schenk 1997), green algae (Weber and Linka 2011; Johnson and Alric 2013), land plants (von Schaewen et al. 1995; Schnarrenberger et al. 1995), and red algae (Oesterhelt et al. 2007; Moriyama et al. 2014)—have both cytosolic- and plastid-localized PPPs. In plants, the plastid PPP can provide NADPH reductants in nonphotosynthetic cells such as embryos (Hutchings et al. 2005; Pleite et al. 2005; Andriotis and Smith 2019) and root tissues (Bowsher et al. 1992). Although loss of photosynthesis has occurred many times in eukaryotes, concomitant loss of the plastid is rare, underscoring its entrenched and indispensable role in carrying out NADPH-dependent biosynthetic

reactions in the absence of photosynthesis (Emes and Neuhaus 1997). In archaeplastids, the plastid PPP continues to supply NADPH when photosynthesis is lost—something that has occurred at least five times in green algae (e.g., Rumpf et al. 1996; Tartar and Boucias 2004), nearly a dozen times in angiosperms (Barkman et al. 2007), and possibly on the order of 100 times in red algae (Goff et al. 1996; Blouin and Lane 2012). Despite their age (ca. 200 My) and estimates of species richness on par with angiosperms and greatly exceeding green and red algae, diatoms are only known to have lost photosynthesis twice (Frankovich et al. 2018; Onyshchenko et al. 2019). Although additional losses may be discovered, the switch from photosynthesis to obligate heterotrophy is a rare transition for diatoms.

Unlike archaeplastids, diatoms lack a plastid PPP and thus predominantly rely on photosynthesis for plastid NADPH production (Michels et al. 2005). Nevertheless, genomic data from this and other studies (Kamikawa et al. 2016; Kamikawa et al. 2017) suggest that the plastids of nonphotosynthetic diatoms continue to be the site of many NADPH-dependent reactions (Fig 2). Without a plastid PPP, however, it is unclear how diatom plastids meet their NADPH demands during periods of prolonged darkness or in species such as *Nitzschia* Nitz4 that have permanently lost photosynthesis.

In the absence of photosynthesis and a PPP, what are the possible sources of NADPH in the plastid of *Nitzschia* Nitz4? Possible enzymatic sources include an NADP-dependent isoform of the glycolytic enzyme GAPDH or conversion of malate to pyruvate by malic enzyme (ME) (Smith et al. 2012). While NADP-GAPDH is predicted to be targeted to the periplastid compartment in *Nitzschia* Nitz4, ME appears to have been lost. It is also possible that ancestrally NADH-producing enzymes have evolved instead to produce NADPH in the plastid, which has likely happened in yeast (Grabowska and Chelstowska 2003).

In addition to enzymatic production, plastids might acquire NADPH via transport from the cytosol, but transport of large polar molecules, like NADPH and other nucleotides, through the inner plastid membrane may be prohibitive (Taniguchi and 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 and Miyake 2012), the triose phosphate/3-phosphoglycerate shuttle is hypothesized to run in reverse to generate NADPH in the relic plastid (apicoplast) of the malaria parasite, *Plasmodium falciparum* (Lim and McFadden 2010). In *Plasmodium*, triose phosphates would be transferred from the cytosol to the apicoplast in exchange for PGA, resulting in the generation of ATP and NADPH in the apicoplast by PGK and glyceraldehyde-3-phosphate dehydrogenase (NADP-GAPDH). Notably, *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

isoform is generally localized to the mitochondrion (Smith et al. 2012). Thus, it is possible that nonphotosynthetic diatoms also employ a reverse triose phosphate/PGA shuttle to generate plastid NADPH, similar to what has been proposed for apicomplexans. Finally, an apicoplast membrane-targeted NAD(P) transhydrogenase (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 (Jackson et al. 2015). The apicoplast-targeted NTH might provide a novel and necessary source of NADPH to that organelle (Ralph et al. 2004). Diatoms, including *Nitzschia* Nitz4, have an additional NTH with predicted localization to the plastid membrane, suggesting another example in which nonphotosynthetic diatoms and apicomplexans may have converged upon the same solution 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 have high demands for both carbon and NADPH (Fig. 2). In the absence of photosynthesis for carbon fixation and generation of NADPH, and without an ancillary PPP for generation of NADPH, the essential building blocks and reducing cofactors for biosynthetic pathways in the plastid are probably in short supply in *Nitzschia* Nitz4. The need to conserve carbon and NADPH likely drove loss of the isoprenoid-producing MEP from the plastid, which has also been linked to loss of photosynthesis in chrysophytes (Dorrell et al. 2019) and euglenophytes (Kim et al. 2004). Despite lack of a PPP in the apicoplast (Ralph et al. 2004), *Plasmodium* has not lost its MEP, presumably because it lacks a cytosolic MVA (Imlay and 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, as in archaeplastids, there is a dedicated plastid PPP. The need to retain expensive pathways, in the absence of requisite supply lines, may represent an important constraint on loss of photosynthesis in lineages such as diatoms and apicomplexans.

The correlated loss of photosynthesis and the MEP in algae should be tested formally through comparative genomics, whereas the functional hypothesis of limiting carbon and NADPH concentrations in nonphotosynthetic plastids requires comparative biochemistry studies. Whether nonphotosynthetic diatoms have other undescribed mechanisms for provisioning NADPH in their plastids, or whether this is being accomplished by conventional means, remains unclear. Regardless, nonphotosynthetic diatoms offer a promising system for studying these and other outstanding questions related to the evolution of trophic shifts in microbial eukaryotes.

Conclusions

The goal of our study was to characterize central carbon metabolism in the nonphotosynthetic diatom, *Nitzschia* sp., one of just a few obligate heterotrophs in a diverse lineage of photoautotrophs that make vast contributions to the global cycling of carbon and oxygen (Falkowski et al. 1998). Aside from loss of the Calvin Cycle and plastid MEP, the genome revealed that central carbon metabolic pathways remained relatively intact following loss of photosynthesis. Moreover, we found no evidence that loss of photosynthesis was accompanied by compensatory gains of novel genes or expansions of genes families associated with carbon acquisition. The genome sequence of *Nitzschia* sp. Nitz4 suggests that nonphotosynthetic diatoms rely on the same set of heterotrophic mechanisms present in their mixotrophic ancestors. Taken together, these findings suggest that elimination of photosynthesis and the switch from mixotrophy to obligate heterotrophy should not be a difficult transition, a hypothesis supported by the experimental conversion of an obligate photosynthetic diatom to a heterotroph with the introduction of a single additional glucose transporter (Zaslavskaia et al. 2001). Nevertheless, given their age (200 My) and species richness (tens of thousands of species), photosynthesis appears to have been lost less frequently in diatoms compared to other photosynthetic lineages (e.g., angiosperms and other Archaeplastida)—a pattern that may reflect diatoms' nearly exclusive dependency on photosynthesis for generation of NADPH in the plastid, which is required to support numerous biosynthetic pathways that persist in the absence of photosynthesis (Fig. 2).

The presence of a β-ketoadipate pathway in diatoms expands its known distribution across the tree of life and raises important questions about its origin and function in photosynthetic and nonphotosynthetic diatoms alike. If the genomic predictions are upheld, a β-ketoadipate pathway would allow diatoms to exploit an abundant source of plant-derived organic carbon. The surprising discovery of this pathway in diatoms further underscores the genomic complexity of diatoms and, decades into the genomic era, the ability of genome sequencing to provide an important first glimpse into previously unknown metabolic traits of marine microbes. As for other diatom genomes, however, we could not ascribe functions to most of the predicted genes in *Nitzschia* Nitz4, potentially masking important aspects of cellular metabolism and highlighting more broadly one of the most difficult challenges in comparative genomics.

Acknowledgements

This work was supported by the National Science Foundation (NSF) (Grant No. DEB-1353131 to A.J.A.) and multiple awards from the Arkansas Biosciences Institute to A.J.A. and J.A.L. This research was carried out by A.P. in partial fulfillment of a Master of Science degree, which was supported by the Fulbright Graduate Student Exchange Program (Ukraine). This research used computational resources

available through the Arkansas High Performance Computing Center, which were funded through multiple NSF grants and the Arkansas Economic Development Commission. We thank Simon Tye for help designing and creating figures.

Literature Cited

Allen AE, Dupont CL, Obornik M, Horak A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu HH, Fernie AR, et al. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature 473:203–207.

Andrews S. 2010. FastQC: A quality control tool for high throughput sequence data.

Andriotis VME, Smith AM. 2019. The plastidial pentose phosphate pathway is essential for postglobular embryo development in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 116:15297–15306.

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. Science 306:79–86.

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. J. Comput. Biol. 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 Evol. Biol. 7:248.

Basu S, Patil S, Mapleson D, Russo MT, Vitale L, Fevola C, Maumus F, Casotti R, Mock T, Caccamo M, et al. 2017. Finding a partner in the ocean: Molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. New Phytol. 215:140–156.

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. Biochem. J 79:531–537.

Berg JM, Tymoczko JL, Stryer L. 2002. Glycolysis and Gluconeogenesis. In: Biochemistry: A Short Course (2nd edition). New York: W.H. Freeman and Company.

Blackburn MV, Hannah F, Rogerson A. 2009. First account of apochlorotic diatoms from mangrove waters in Florida. J. Eukaryot. Microbiol. 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. Sci Adv 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.

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.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244.

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. Plant J 2:893–898.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12:59–60.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of eukaryotes. Trends Ecol. Evol. 35:43–55.

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al. 2014. MAKER-P: A tool kit for the rapid creation, management, and quality control of plant genome annotations. Plant Physiol. 164:513–524.

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 Res. 18:188–196.

Cerón García MC, Camacho FG, Mirón AS, Fernández Sevilla JM, Chisti Y, Molina Grima E. 2006. Mixotrophic production of marine microalga *Phaeodactylum tricornutum* on various carbon sources. J. Microbiol. Biotechnol. 16:689–694.

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. Comput. Appl. Biosci. 11:441–447.

Clement R, Dimnet L, Maberly SC, Gontero B. 2016. The nature of the CO2-concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. New Phytol. 209:1417–1427.

Crowell RM, Nienow JA, Cahoon AB. 2019. The complete chloroplast and mitochondrial genomes of the diatom *Nitzschia palea* (Bacillariophyceae) demonstrate high sequence similarity to the endosymbiont organelles of the dinotom *Durinskia baltica*. J. Phycol. 55:352–364.

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. Nat. Genet. 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. Proc. Natl. Acad. Sci. U.S.A. 116:6914–6923.

Dufresne A, Garczarek L, Partensky F. 2005. Accelerated evolution associated with genome reduction in

a free-living prokaryote. Genome Biol. 6:R14.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 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 Sci. 8:978–984.

Emes MJ, Neuhaus HE. 1997. Metabolism and transport in non-photosynthetic plastids. J. Exp. Bot. 48:1995–2005.

Ewe D, Tachibana M, Kikutani S, Gruber A, Río Bártulos C, Konert G, Kaplan A, Matsuda Y, Kroth PG. 2018. The intracellular distribution of inorganic carbon fixing enzymes does not support the presence of a C4 pathway in the diatom *Phaeodactylum tricornutum*. Photosynth. Res. 137:263–280.

Falkowski PG, Barber RT, Smetacek V V. 1998. Biogeochemical controls and feedbacks on ocean primary production. Science 281:200–207.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O, Taylor FJR. 2004. The evolution of modern eukaryotic phytoplankton. Science 305:354–360.

Falkowski PG, Raven JA. 2013. Aquatic Photosynthesis: Second Edition. Princeton University Press

Fester T, Schenk HEA. 1997. Glucose-6-Phosphate dehydrogenase isoenzymes from *Cyanophora paradoxa*: Examination of their metabolic integration within the meta-endocytobiotic system. In: Eukaryotism and Symbiosis. Springer Berlin Heidelberg. p. 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.

Figueroa-Martinez F, Nedelcu AM, Smith DR, Adrian R-P. 2015. When the lights go out: The evolutionary fate of free-living colorless green algae. New Phytol. 206:972–982.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2019. RepeatModeler2: Automated genomic discovery of transposable element families. bioRxiv [Internet]:856591. Available from: https://www.biorxiv.org/content/10.1101/856591v1.abstract

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. Nat. Rev. Microbiol. 9:803–816.

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

Godrijan J, Drapeau D, Balch WM. 2020. Mixotrophic uptake of organic compounds by coccolithophores. Limnol. Oceanogr. 17:1.

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. J. Phycol. 32:297–312.

Grabowska D, Chelstowska A. 2003. The ALD6 gene product is indispensable for providing NADPH in yeast cells lacking glucose-6-phosphate dehydrogenase activity. J. Biol. Chem. 278:13984–13988.

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. Plant J. 81:519–528.

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. J. Eukaryot. Microbiol. 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 Biol. Evol. 10:1504–1515.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.

Harwood CS, Parales RE. 1996. The beta-ketoadipate pathway and the biology of self-identity. Annu. Rev. Microbiol. 50:553–590.

Hellebust JA. 1971. Glucose uptake by *Cyclotella cryptica*: Dark induction and light inactivation of transport system. J. Phycol. 7:345–349.

Hellebust JA, Lewin J. 1977. Heterotrophic nutrition. In: Werner D, editor. The Biology of Diatoms. Vol. 13. University of California Press. p. 169–197.

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

Hutchings D, Rawsthorne S, Emes MJ. 2005. Fatty acid synthesis and the oxidative pentose phosphate pathway in developing embryos of oilseed rape (*Brassica napus* L.). J. Exp. Bot. 56:577–585.

Imlay L, Odom AR. 2014. Isoprenoid metabolism in apicomplexan parasites. Curr Clin Microbiol Rep 1:37–50.

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 Lett. 589:2027–2033.

Johnson X, Alric J. 2013. Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: Metabolic constraints for carbon partitioning between oil and starch. Eukaryot. Cell 12:776–793.

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. Mol. Biol. Evol. 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. Mol. Biol. Evol. 32:2598–2604.

Kamikawa R, Tanifuji G, Ishikawa SA, Ishii K-I, Matsuno Y, Onodera NT, Ishida K-I, Hashimoto T, Miyashita H, Mayama S, et al. 2016. Proposal of a twin aarginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. Mol. Biol. Evol. 33:303.

Kamikawa R, Yubuki N, Yoshida M, Taira M, Nakamura N, Ishida K-I, Leander BS, Miyashita H, Hashimoto T, Mayama S, et al. 2015. Multiple losses of photosynthesis in *Nitzschia* (Bacillariophyceae). Phycological Res. 63:19–28.

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. J. Bacteriol. 184:207–215.

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

Kitano M, Matsukawa R, Karube I. 1997. Changes in eicosapentaenoic acid content of *Navicula saprophila*, *Rhodomonas salina* and *Nitzschia* sp. under mixotrophic conditions. J. Appl. Phycol. 9:559–563.

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 Biol. Evol. 11:1829–1837.

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.

Kuczynska P, Jemiola-Rzeminska M, Strzalka K. 2015. Photosynthetic pigments in diatoms. Mar. Drugs 13:5847–5881.

Kustka AB, Milligan AJ, Zheng H, New AM, Gates C, Bidle KD, Reinfelder JR. 2014. Low CO2 results in a rearrangement of carbon metabolism to support C4 photosynthetic carbon assimilation in *Thalassiosira pseudonana*. New Phytol. 204:507–520.

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. F1000Res. 6:1287.

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

Liao P, Hemmerlin A, Bach TJ, Chye M-L. 2016. The potential of the mevalonate pathway for enhanced isoprenoid production. Biotechnology Advances [Internet] 34:697–713. Available from: http://dx.doi.org/10.1016/j.biotechadv.2016.03.005

Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.

Bioinformatics 25:1754–1760.

Lim L, McFadden GI. 2010. The evolution, metabolism and functions of the apicoplast. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365:749–763.

Liu MS, Hellebust JA. 1976. Regulation of proline metabolism in the marine centric diatom *Cyclotella cryptica*. Can. J. Bot. 54:949–959.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 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.

Masai E, Katayama Y, Fukuda M. 2007. Genetic and biochemical investigations on bacterial catabolic pathways for lignin-derived aromatic compounds. Biosci. Biotechnol. Biochem. 71:1–15.

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

Mitra A, Flynn KJ, Burkholder JM, Berge T, Calbet A, Raven JA, Granéli E, Glibert PM, Hansen PJ, Stoecker DK, et al. 2014. The role of mixotrophic protists in the biological carbon pump. Biogeosciences 11:995–1005.

Mock T, Otillar RP, Strauss J, McMullan M, Paajanen P, Schmutz J, Salamov A, Sanges R, Toseland A, Ward BJ, et al. 2017. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. Nature 541:536–540.

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.

Neuhaus HE, Emes MJ. 2000. Nonphotosynthetic metabolism in plastids. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:111–140.

Noctor G, Hager J, Li S. 2011. Biosynthesis of NAD and its manipulation in plants. In: Rébeillé F, Douce R, editors. Advances in Botanical Research. Vol. 58. Academic Press. p. 153–201.

Oesterhelt C, Klocke S, Holtgrefe S, Linke V, Weber APM, Scheibe R. 2007. Redox regulation of chloroplast enzymes in *Galdieria sulphuraria* in view of eukaryotic evolution. Plant Cell Physiol. 48:1359–1373.

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

Oudot-Le Secq M-P, Grimwood J, Shapiro H, Armbrust EV, Bowler C, Green BR. 2007. Chloroplast genomes of the diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*: Comparison with other plastid genomes of the red lineage. Mol. Genet. Genomics 277:427–439.

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. J. Biol. Chem. 284:31249–31259.

Parales RE, Harwood CS. 1992. Characterization of the genes encoding beta-ketoadipate: Succinyl-coenzyme A transferase in *Pseudomonas putida*. J. Bacteriol. 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). Am. J. Bot. 105:330–347.

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

Petrou K, Nielsen DA. 2018. Uptake of dimethylsulphoniopropionate (DMSP) by the diatom *Thalassiosira weissflogii*: A model to investigate the cellular function of DMSP. Biogeochemistry 141:265–271.

Pleite R, Pike MJ, Garcés R, Martínez-Force E, Rawsthorne S. 2005. The sources of carbon and reducing power for fatty acid synthesis in the heterotrophic plastids of developing sunflower (*Helianthus annuus* L.) embryos. J. Exp. Bot. 56:1297–1303.

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 [Internet]:201178. Available from: https://www.biorxiv.org/content/10.1101/201178v3.abstract

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. Nat. Rev. Microbiol. 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 J. 14:347–363.

Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat. Prod. Rep. 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 Biol. Evol. 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). J. Phycol. 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 Rep. 21:e47832.

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

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

von Schaewen A, Langenkämper G, Graeve K, Wenderoth I, Scheibe R. 1995. Molecular characterization

of the plastidic glucose-6-phosphate dehydrogenase from potato in comparison to its cytosolic counterpart. Plant Physiol. 109:1327–1335.

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 Physiol. 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.

Smetacek V. 1999. Diatoms and the ocean carbon cycle. Protist 150:25–32.

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

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

Stanier RY, Ornston LN. 1973. The beta-ketoadipate pathway. Adv. Microb. Physiol. 9:89–151.

Stoecker DK, Hansen PJ, Caron DA, Mitra A. 2017. Mixotrophy in the marine plankton. Ann. Rev. Mar. Sci. 9:311–335.

Taniguchi M, Miyake H. 2012. Redox-shuttling between chloroplast and cytosol: Integration of intra-chloroplast and extra-chloroplast metabolism. Curr. Opin. Plant Biol. 15:252–260.

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

Torstensson A, Young JN, Carlson LT, Ingalls AE, Deming JW. 2019. Use of exogenous glycine betaine and its precursor choline as osmoprotectants in Antarctic sea-ice diatoms. J. Phycol. 55:663–675.

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. Biotechnol. Biofuels 9:258.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. Proc. Natl. Acad. Sci. U. S. A. 105:4957–4962.

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.

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*. Philos. Trans. R. Soc. Lond. B 372:20160404.

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

Wells T Jr, Ragauskas AJ. 2012. Biotechnological opportunities with the β-ketoadipate pathway. Trends Biotechnol. 30:627–637.

Wilken S, Huisman J, Naus-Wiezer S, Van Donk E. 2013. Mixotrophic organisms become more heterotrophic with rising temperature. Ecol. Lett. 16:225–233.

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.

Zaslavskaia LA, Lippmeier JC, Shih C, Ehrhardt D, Grossman AR, Apt KE. 2001. Trophic conversion of an obligate photoautotrophic organism through metabolic engineering. Science 292:2073–2075.

Zheng Z, Chen H, Du N. 2019. Characterization of the complete plastid genome of *Fragilariopsis cylindrus*. Mitochondrial DNA B 4:1138–1139.

SUPPLEMENTARY MATERIALS

Supplementary File 1. Assembly and filtering of the *Nitzschia* sp. Nitz4 nuclear genome.

Supplementary File 2. Summary of carbon transporters and carbohydrate-cleaving enzymes in the *Nitzschia* sp. Nitz4 nuclear genome.

Supplementary Table 1. Results of iterative of the genome annotation procedure used to identify protein-coding genes in the genome of Nitzschia sp. Nitz4.

Supplementary Table 2. Summary of orthogroup statistics for the four genomes (*C. nana*, *P. tricornutum*, *F. cylindrus*, and *Nitzschia* sp. Nitz4) and one transcriptome (*Nitzschia* CCMP2144) used in the OrthoFinder analysis.

Supplementary Table 3. Summary of variant calling for the nuclear genome of *Nitzschia* sp. Nitz4.

Supplementary Table 4. Annotation and localization of carbon metabolism genes in *Nitzschia* sp. Nitz4.

Supplementary Table 5. Gene family sizes for carbohydrate-active enzymes.

Supplementary Table 6. B-ketoadipate annotations

Supplementary Table 7. B-ketoadipate pathway genes in the sequenced genomes (*C. nana*, *F. cylindrus*, *P. tricornutum*, and *P. multiseries*) and transcriptome (*Nitzschia* sp. CCMP2144) of several diatoms.