EVOLUTIONARY BIOLOGY

Genome evolution of a nonparasitic secondary heterotroph, the diatom *Nitzschia putrida*

Ryoma Kimikawa1*, Takako Mochizuki2, Mika Sakamoto2, Yasuhiro Tanizawa2, Takuo Nakayama3†, Ryo Onuma4‡, Ugo Cenci5, Daniel Moog6,7§, Samuel Speak8, Krisztina Sarkozi8||, Andrew Toseland8, Cock van Oosterhout8, Kaori Oyama9, Misako Kato9, Keitaro Kume10, Motoki Kayama11, Tomonori Azuma11, Ken-ichiro Ishii11, Hideaki Miyashita11, Bernard Henrisssat8,12,13,14, Vincent Lombard12,13, Joe Win15, Sophien Kamoun15, Yuichiro Kashiyama16, Shigeki Mayama17, Shin-ya Miyagishima4, Goro Tanifuji18, Thomas Mock8*, Yasuakazu Nakamura2

Secondary loss of photosynthesis is observed across almost all plastid-bearing branches of the eukaryotic tree of life. However, genome-based insights into the transition from a phototroph into a secondary heterotroph have so far only been revealed for parasitic species. Free-living organisms can yield unique insights into the evolutionary consequence of the loss of photosynthesis, as the parasitic lifestyle requires specific adaptations to host environments. Here, we report on the diploid genome of the free-living diatom *Nitzschia putrida* (35 Mbp), a nonphotosynthetic osmotroph whose photosynthetic relatives contribute ca. 40% of net oceanic primary production. Comparative analyses with photosynthetic diatoms and heterotrophic algae with parasitic lifestyle revealed that a combination of gene loss, the accumulation of genes involved in organic carbon degradation, a unique secretome, and the rapid divergence of conserved gene families involved in cell wall and extracellular metabolism appear to have facilitated the lifestyle of a free-living secondary heterotroph.

INTRODUCTION
The loss of photosynthesis in photoautotrophs is successful if compensated by a competitive advantage arising from the availability of an extracellular energy source. Hence, many secondary heterotrophs evolve as parasites (1–3), relying on sufficient resources provided by their hosts. Well-studied examples are the Apicomplexa [e.g., (4)], which have lost photosynthesis secondarily. However, the loss of photosynthesis can also lead to free-living secondary heterotrophs, which are as common as parasites (2, 5–8). Despite their significance, our knowledge about the evolution of free-living secondary heterotrophs is very limited, and we therefore lack insights into evolutionary processes required for them to thrive without photosynthesis and independently of a resource-providing host. Given that a parasitic lifestyle accelerates the rate of evolution (cf. Red Queen hypothesis) (9) and of loss of conserved orthologous genes [e.g., (10)], the genome analysis of a nonparasitic secondary heterotroph can provide insights uncompromised by parasite-specific adaptations. Hence, the diatom *Nitzschia putrida*, isolated from mangrove estuaries, is the ideal model to test these hypotheses because it is an example of a free-living secondary heterotroph (5, 11) within the diverse group of largely photoautotrophic diatoms (12, 13). As several genomes of the latter have recently become available including close phylogenetic relatives (14–16), a genome-based comparative metabolic reconstruction of *N. putrida* promises to reveal fresh insights into what is required to thrive as a free-living secondary heterotroph. Thus, here, we have analyzed the draft genome sequence of *N. putrida*, which provides insights into evolutionary processes underpinning lifestyle shifts from photoautotrophy to free-living heterotrophy in the context of a coastal surface ocean ecosystem.

RESULTS
Genome assembly
K-mer–based GenomeScope analysis (17) with 150–base pair (bp)–long Illumina short reads suggested the genome of *N. putrida* (Fig. 1A) to be diploid (fig. S1A). To provide a high-quality genome with long-range contiguity, PacBio sequencing (RSII platform) was performed, resulting in ≥40-fold coverage. Because of the confirmed diploid nature of the *N. putrida* genome, we have applied the Falcon assembler and Falcon_unzip version 0.5 (18) to provide a first draft genome of...
this species. On the basis of this assembly, we estimated a genome size of 35 Mbp, including 87 scaffolds with an N50 of 860.9 kbp. The longest scaffold was 3.8 Mbp. The heterozygous regions of the genome (alternate contigs) estimated by the Falcon assembler resulted in 12 Mbp, with an N50 of 121 kbp (table S1). The Falcon assembly was error-corrected and polished by approximately 150-fold coverage of Illumina short reads, which were subsequently used for generating the final assembly with Pilon 1.2.2 (19) including manual curation.

According to the k-mer–assessed diploid nature of the *N. putrida* genome, the read coverage of the homozygous regions is approximately twofold higher than the read coverage for the heterozygous regions, suggesting the presence of diverged alleles as previously identified in the genome of the photoautotroph diatom *Fragilariopsis cylindrus* (fig. S1, A and B). Thus, most of the diverged allelic variants can be found in the heterogeneous regions characterized by the presence of alternate contigs, whereas the regions with no corresponding alternate contigs are homozygous (fig. S1B). On the basis of the analysis with Braker2 version 2.0.3 (20), the *Nitzschia* genome comprises 15,003 and 5767 inferred protein-coding loci on the primary and alternate contigs, respectively (table S1). Almost 40% of loci in the genome of *N. putrida* appear to be characterized by diverged alleles. A BUSCOv3 analysis (21) revealed the genome to be complete at a level of 90.1% based on the haploid set of genes.

**The loss of photosynthesis**

The haploid set of genes was used to reconstruct the nuclear-encoded plastid proteome of *N. putrida* and therefore to reveal the extent of gene loss including key genes of photosynthesis. A comparative analysis of the *N. putrida* plastome (22) with its photosynthetic counterparts revealed that more than 50% of nuclear encoded plastid proteins have been lost (Fig. 1B). More than 500 orthogroups (OrthoFinder) (23) of nuclear-encoded plastid proteins, which are usually shared between photosynthetic diatoms (22), are missing in the predicted plastid proteome of *N. putrida* (Fig. 1C). The missing part of the plastid proteome included genes encoding for proteins of light-harvesting antenna including fucoxanthin–chlorophyll *a/c* protein (*fcp*), photosystem II and I (e.g., *psbA, psbC, psbO, psaA, psaB*, and *psaD*), the cytochrome *b6/f* complex (e.g., *petA*), and carbon fixation (e.g., *rbcS* and *rbcL*) in addition to genes of the Calvin cycle.
[e.g., phosphoribulokinase (prk)]. Furthermore, a substantial number of key genes were missing for the biosynthesis of chlorophyll, carotenoids, and plastoquinones (fig. S1C).

Despite the loss of some of these key photosynthesis genes, there is still a substantial number of genes left encoding common plastid metabolic pathways as known from photosynthetic diatoms, including the generation of adenosine 5′-triphosphate (ATP) by adenosine triphosphatase (ATPase) subunits, which are encoded both in the nuclear and plastid genomes (24). Almost all genes encoding for plastid enzymes to synthesize essential amino acids are still encoded in the nuclear genome of N. putrida. Furthermore, all genes of the heme pathway have been found, and N. putrida appears to be able to synthesize riboflavin. The presence of plastid-targeted transporters (25) enables the transport of phosphoenolpyruvate, 3-phosphoglycerate, and dihydroxyacetone-phosphate across the plastid membranes. In addition, our genome-based reconstruction of plastid metabolism identified the biosynthesis pathway for lipids and the ornithine cycle in N. putrida (Fig. 1D and fig. S2). The latter has been reported neither in previous transcriptome-based studies with this species (26) nor in any other secondary heterotroph ([e., (6, 7)]. As N. putrida has an osmotrophic lifestyle, it relies on dissolved organic matter and nutrients. Thus, the biosynthesis of a variety of metabolic compounds supports the osmotrophic lifestyle, lacking abilities to prey or to parasitize on other organisms.

Communication between organelles and light-dependent gene expression

The lack of CO₂ fixation in plastids of N. putrida—which reduces the amount of amino acids, lipids, and other metabolites to be synthesized—appears to be partially compensated for by the remodeling of metabolic interactions with mitochondria and peroxisomes and by retaining active recycling of nitrogen (Figs. 1 and 2 and figs. S3 and S4). It appears that the nonphotosynthetic plastid of N. putrida still exchanges glutamine and ornithine, both of which are important intermediates of the ornithine cycle. All genes for the ornithine-urea cycle have been retained in the N. putrida genome. The ornithine-urea cycle is indispensable for nitrogen recycling in photosynthetic diatoms (27, 28), and even after the loss of photosynthesis, nitrogen recycling appears to be essential in N. putrida (Fig. 2A) due to its osmotrophic lifestyle. Usually, the ornithine-urea cycle is tightly linked with tricarboxylic acid cycle and/or photorespiration in photosynthetic diatoms (27, 28). However, N. putrida is not likely to perform photorespiration (Fig. 2A). The metabolic exchange with the peroxisome through glycolate likely has ceased as phosphoglycolate phosphatase and peroxisomal glycolate oxidase are missing. Thus, photorespiration is unlikely to take place in nonphotosynthetic plastids of N. putrida due to the lack of ribulose 1,5-bisphosphate carboxylase/oxygenase and other key enzymes of the Calvin cycle (Fig. 1). Nevertheless, peroxisomes still appear to play a role in N. putrida for the production of malate or glyoxylate, which feed into respiratory pathways of the mitochondria to support ATP and NADPH (reduced form of nicotinamide adenine dinucleotide phosphate) production (Fig. 2A).

Light in photosynthetic organisms not only plays a substantial role for photosynthesis generating ATP and NADPH but also regulates cell division, diel cycles, and different signaling processes unlike in many heterotrophic organisms (29–31). As a consequence, we identified the remaining photoreceptors and cell-cycle regulators such as cyclins and cyclin-dependent kinases (32). Although they were still encoded and expressed in the genome of N. putrida (Fig. 2B and fig. S5, A and B), we were unable to identify a diel cycle in cell division (Fig. 2C). This suggests that these cell-cycle regulators potentially have neo/subfunctionalized and therefore have a different regulatory role in N. putrida unrelated to the diel cycle. The loss of the transcription factor bHLH-1a (RITMO1), which has been identified as a master regulator of diel periodicity (33), corroborates our finding that N. putrida has lost the ability to perform diel cycles. In addition, most of the other photoreceptors known from photosynthetic diatoms have also been lost (Fig. 2B) such as the blue light sensing aureochromes 1a/b, both of which are transcription factors responsible for photoacclimation (34). Despite the lack of light-dependent cell-cycle regulation, a few remaining photoreceptors were identified including bHLH1b_PAS, aureochrome 1c, and cryptochrome-DASH/CPF2 (Fig. 2B) (29, 35). Basic ZIP [basic leucine zipper proteins (bZIP)] transcription factors having potentially light-sensitive Per-Arnt-Sim (PAS) domains (bZIP-PAS) (36) were also identified in the N. putrida genome such as homologs to bZIP6 and bZIP7 of Phaeodactylum tricornutum (37). The latter homolog has been duplicated and diversified in N. putrida (fig. S5C). The presence of bZIP-PAS protein in a heterotrophic eukaryote is not unprecedented as some oomycetes, nonphotosynthetic parasites, have been reported to also encode them in their genomes ([e., (38)]. Although their role in regulating gene expression remains to be investigated in N. putrida, light still appears to influence the expression of some genes in this heterotrophic species. Comparative transcriptome analyses every 4 hours during a shift from a light phase to darkness (Fig. 2D) revealed eight clusters characterized by different expression patterns. Furthermore, there was no cluster explicitly representing the light-dependent gene expression patterns as seen in photosynthetic algae (e.g., (29, 39)]. However, one of the clusters contained genes only expressed in the mid-light phase: cluster 7 containing 90 genes (0.6% total). Forty four of them were genes with known functional domains based on a KOG (EuKaryotic Orthologous Groups) analysis, and 21 of them were encoding proteins for substrate import and carbon metabolism (fig. SSD). However, the photoreceptor homologs above—bHLH1b_PAS, aureochrome 1c, and cryptochrome-DASH/CPF2—were not part of this cluster, and there was no explicit trend in their gene expression patterns with respect to changes between light and dark conditions.

The genetic toolkit for the evolution of nonparasitic secondary heterotrophy

Despite the loss of many nuclear genes and their families, the genome size of N. putrida is not significantly different to photosynthetic relatives such as F. cylindrus and P. tricornutum and the more distantly related diatom Thalassiosira pseudonana (table S1). This is distinct from evolutionary trends observed in parasitic eukaryotes that have lost photosynthesis as they have smaller genomes encoding smaller gene families compared to their photosynthetic relatives (fig. S6). By comparing KOGs of paralogous proteins, there was no significant difference in the number of unique KOG IDs between these four diatom species (fig. S7, A and B). However, when we compared the number of paralogous proteins assigned to each KOG ID, there were several KOG categories for which N. putrida had a higher number of paralogous proteins compared to the other diatom species: nucleotide transport (F), transcription (K), signal transduction (T), intracellular trafficking, secretion, vesicular transport (U), and cytoskeleton (Z) (fig. S7C). Even after normalization by total gene numbers, nucleotide transport (F), signal transduction (T),
Fig. 2. Loss of genes for the plastid-peroxisome metabolic flow and photoreceptors. (A) Metabolic interactions between a mitochondrion and a nonphotosynthetic plastid and between a mitochondrion and a peroxisome. Black, orange, and blue arrows show presence of responsible protein sequences for the reactions in a plastid, a mitochondrion, and a peroxisome, respectively, while light gray arrows show absence of responsible protein sequences. Dashed arrows show possible interorganellar metabolic flows. Abbreviations are described in the Supplementary Materials. (B) Photoreceptor and cell-cycle genes in the N. putrida genome. The other genes are shown in fig. S5. Light green and light gray boxes show the presence and absence of corresponding genes, respectively. (C) Growth of the heterotrophic diatom under the different light conditions. Closed boxes show growth in the continuous dark condition, while open boxes show growth in the light-dark condition. Shaded in gray are the dark periods in the light-dark cultivation conditions. (D) Left: Heatmap showing the reproducible expression patterns of genes (Pearson’s correlation coefficient < 0.9). k-means clustering was calculated for each gene based on reads per kilobase of transcript per million mapped reads (RPKM) + 1 values, which were transformed to log2 and centered by median values. Yellow and blue indicate up-regulation and down-regulation of the gene, respectively. Right: The line graphs showing expression pattern of genes in each cluster. The colored line indicates the average value of the expression patterns. LogFC, log fold change.
and cytoskeleton (Z) genes were more abundant in the N. putrida genome (fig. S8A). This observation was corroborated by N. putrida–specific enrichment of Pfam domains such as adenylate/guanulate cyclase and cyclic nucleotide esterase, leucine-rich repeat (LRR), and glycosyl/galactosyl transferase domains (fig. S8B).

A microbial heterotroph acquires nutrients either by phagotrophy, the preferred nutrition of many parasites, or by osmotrophy. The latter requires uptake of dissolved organic compounds by osmosis as realized by bacteria and fungi, for instance (40, 41). As N. putrida grows well under axenic conditions (5, 42), it is likely an osmotroph, dependent on the uptake of dissolved organic compounds across the silicified cell wall and the plasma membrane. As realized by osmotrophic fungi, N. putrida may even be able to degrade higher–molecular weight compounds extracellularly to be subsequently taken up as individual molecules by specific transporters or even osmosis (40, 41). Thus, it is likely that cell wall, membrane, and secreted proteins were diversified in N. putrida compared to photosynthetic diatoms to facilitate osmotrophy.

To address this hypothesis, we analyzed the enrichment of paralog proteins and differences in nutrient transporters involved in the uptake of dissolved organic compounds such as solute carriers. A comparison to photosynthetic diatoms and parasitic nonphotosynthetic algal species [Prototheca and Helicosporidium (green algae), and the apicomplexans Plasmodium and Toxoplasma] has revealed that N. putrida has a unique composition of genes encoding transporters, which is therefore different to photosynthetic algae and parasitic nonphotosynthetic algal species (Fig. 3, A to C). For instance, the number of genes encoding silicon transporters (SITs), solute symporters, and the resistance-nodulation–cell division superfamily was more than twice as abundant in N. putrida compared to photosynthetic diatom species (Fig. 3D and fig. S9A). However, in contrast to the difference between N. putrida and photosynthetic diatom species, there is no enrichment of particular transporters in parasitic algal species when compared to their photosynthetic relatives (fig. S9, B and C).

Expansion of those gene families may, at least partly, have been achieved by recent tandem duplications (Fig. 3E). To gain insight into when the expansion had occurred, we performed a coalescence analysis, which revealed that SITs in N. putrida began to expand around 3.3 million years (Ma) ago (1.2 to 6.6, 95% confidence interval [CI]), while divergence from another nonphotosynthetic diatom N. alba is estimated to have occurred around 6.67 Ma ago (2.5 to 11.5, 95% CI; fig. S10). The split between F. cylindrus and P. multiseries, N. putrida and P. tricornutum, and T. pseudonana, respectively, which corresponds to between 5 and 7% of the total number of genes in their genomes (fig. S12A). Nevertheless, there were significant differences when we compared the diversity of proteins between these four diatom species (Fig. 4, A and B); N. putrida, on average, had a significantly higher number of proteins per tribe than any of the other diatom species (two-sided Wilcoxon signed-rank test: P < 0.01; Fig. 4C). In particular, proteins involved in heterotrophy such as organic matter degradation/modification including CAZymes and peptidases were more abundant in N. putrida than in the photosynthetic diatom genomes (188 in N. putrida, 142 in F. cylindrus, 118 in P. tricornutum, and...
Fig. 3. Diversity of transporters and carbohydrate active enzymes in *N. putrida*. (A) Distribution of the number of transporters in each transporter family of diatoms. Differences in the distributions among species were tested by the Wilcoxon signed-rank test corrected with the Benjamini-Hochberg procedure (*P* < 0.05), but there is no significant difference. Outliers were omitted in the boxplot. Nonphotosynthetic species are highlighted in gray. (B) Distribution of the number of transporters in each transporter family of Alveolata. Details are described in (A). (C) Distribution of the number of transporters in each transporter family of green algae (Trebouxiophyceae). Details are described in (A). (D) The gene number of transporters in the 12 most abundant transporter families of *N. putrida*. (E) Silicon transporter (SIT) genes tandemly located in the contig 000000F. SIT genes are highlighted in light red with the gene IDs. (F) Glycoside hydrolase (GH) families from the Carbohydrate Active enZyme (CAZy) database focused on diatoms. The diagram shows a heatmap of CAZyme prevalence in each taxon (number of a particular CAZyme family divided by the total number of CAZyme genes in the organism); the white to blue color scheme indicates low to high prevalence, respectively. Dendrograms (left and top) show respectively the relative taxa proximity with respect co-occurrence of CAZyme families and the co-occurrence of CAZyme families with one another within genomes. (G) GH114 genes tandemly located in the contig 000022F. GH114 genes are highlighted in light green with the gene IDs.
Fig. 4. Secretome of nonphotosynthetic algae. (A) The number of secretome tribes of diatoms, including at least four sequences, clustered by TribeMCL (45). Different colors represent tribe categories as follows: 1, species specific tribes; 2 to 4, tribes shared by two to four species, respectively. OTUs, operational taxonomic unit. (B) Proportion of each tribe category in diatoms. Details are described in (A). (C) Distribution of the number of protein sequences in each secretome tribe in diatoms. Outliers were omitted in the boxplot. The Wilcoxon signed-rank test corrected with the Benjamini-Hochberg procedure was used for tests of statistical significance. (D) The number of secretome tribes in green algae (trebouxiophytes), including at least four sequences, clustered by TribeMCL (45). Different colors represent tribe categories as follows: 1, species specific tribes; 2 to 5, tribes shared by two to five species, respectively. (E) Proportion of each tribe category in green algae. Details are described in (D). (F) Distribution of the number of protein sequences in each secretome tribe in green algae. Details are described in (C). (G) Expression of the 10 largest tribes in N. putrida during the 25 hours of cultivation. Genes in the tribes could be divided into four clusters. Details are described in Fig. 2D.
from homologs of a photosynthetic ancestor. genes of potential lateral origin (figs. S14 and S15). Thus, the origin of secretome genes in N. putrida for attaching to disintegrating mangrove leaves (5, 42, 45). Endopeptidases and aromatic compound degradation may facilitate the utilization of their complex carbon compounds.

Furthermore, transcriptional dynamics of the predicted secretome over a diel cycle (Fig. 2) revealed the presence of four different clusters. Genes in cluster 1 were transcribed at the beginning of the first light phase and genes in cluster 2 at the end of the dark phase and into the second light phase (Fig. 4G). Genes of cluster 3 were most strongly expressed in the middle and end of the first light phase, whereas genes in cluster 4 were relatively weakly expressed throughout day and night. These results suggest that stimuli including light and/or nutrients play a role in the regulation of these genes, which might either be a relic from the photosynthetic ancestor or a response to diel cycles of organic substances in the aquatic system occupied by N. putrida. There is only weak evidence of lateral transfer of secretome genes in N. putrida (figs. S14 and S15) with five genes of potential lateral origin (figs. S14 and S15). Thus, the origin of most secretome proteins in N. putrida likely was derived vertically from homologs of a photosynthetic ancestor.

**DISCUSSION**

N. putrida experienced a series of genetic adaptations toward a heterotrophic lifestyle. This diatom species took a step backward in one of the major evolutionary transitions, from photoautotrophs to heterotrophs, potentially relaxing selection on some of the now redundant gene networks and their functions. As expected, more than 50% of nuclear encoded plastid proteins have been lost in the N. putrida plastid proteome in comparison to its photosynthetic counterparts (22). However, the total number of genes (~15,000) fell within the range of photosynthetic microalgae, and we found no evidence of pseudogene formation, genome streamlining [e.g., (46)], gene family contraction (cf. birth-and-death hypothesis) (47), or reductive genome evolution (Black Queen hypothesis) (48). The relatively large genome size is not unexpected given that N. putrida is a free-living osmotroph. This free-living lifestyle in a complex and highly variable coastal marine environment likely is the reason why a substantial number of genes including some photoreceptors, cell cycle regulators, and common plastid metabolic pathways usually present in photosynthetic diatoms have remained. Although some of the latter genes were still expressed, N. putrida appears to lack a diel growth cycle, which suggests that these cell-cycle regulators have neo/subfunctionalized. However, as a certain number of genes still appear to be regulated by light, osmotrophy potentially benefits from diel fluctuations of resources such as dissolved organic carbon in aquatic environments (49–51). For photoautotrophs, it is important to regulate the cell cycle in accordance with diel cycles for optimizing photosynthesis and therefore cell proliferation (29–31, 52, 53). Without being reliant on light as its primary energy source, the osmotroph N. putrida no longer requires coordinating its cell cycle with diel cycles. Thus, after the loss of photosynthesis, the strict light-dependent regulation of gene expression might have become less important and gene expression therefore may have become predominantly regulated by other stimuli. Many photoreceptors are missing, but duplication of genes for bZIP transcription factors with PAS domains and genes for signal transduction and cellular regulatory roles such as adenyl/guanyl cyclase and cyclic nucleotide esterase domains was enriched in the N. putrida genome. Furthermore, the peroxisome-plastid interaction is no longer required after the loss of photosynthesis, giving rise to loss of carbon fixation in the context of glycolate recycling. In contrast, the ornithine-urea cycle likely remains to be functional to facilitate nitrogen recycling.

Gene family expansions and neo/subfunctionalizations appear to have played a prominent role in the adaptation to its different lifestyle given that many proteins predicted to be secreted have diversified in N. putrida, possibly to facilitate osmotrophy. Together, the marked change of lifestyle associated with the “devolution” did not result in reductive genome evolution as known from non-photosynthetic plastid-bearing parasites.

**METHODS**

**Cultivation, DNA and RNA extraction, and sequencing**

N. putrida NIES-4239 was cultivated in the Daigo’s IMK medium (Wako) including 1% Luria-Bertani medium based on the artificial seawater made with MARINE ART SF-1 (Osaka Yakken Co.) at 20°C under the 12-hour light and 12-hour dark conditions: 50 μmol photons/m² per second with plant cultivation light-emitting diode light (BC-BML3, Biomedical Science). DNA was extracted with the Extrup Soil DNA Kit Plus version 2 (Nippon Steel). Total DNA was subjected to library construction with TruSeq DNA PCR-Free (350; Illumina) and to 151-bp paired-end sequencing by HiSeqX, resulting in 660 million paired-end reads, and to PacBio RSII, with SMRT cell 8Pac v3 and DNA Polymerase Binding Kit P6 v2, in Macrogen, resulting in 1.3-Gb subreads. Total RNA was extracted with TRIzol (Sigma-Aldrich) according to the manufacturer’s instruction and was subjected to library construction with TruSeq RNA Sample Prep Kit v2 (Illumina) and 101-bp paired-end sequencing by HiSeq 2500, resulting in 107.5 million paired-end reads.

**Genome assembly and construction of gene models**

PacBio reads were assembled into contigs using Falcon (version 0.7.0) (18) with a length cutoff of 7000 bp for seed reads and an estimated genome size of 33 Mbp. Genome size estimation was performed on the GenomeScope web server (http://qb.cshl.edu/genomescope/) based on the k-mer frequency distribution of Illumina reads calculated by JellyFish version 2.2.6 with a k-mer size of 21. The resultant primary and associate contigs were then subjected to Falcon_unzipped
cies were obtained from the JGI Genome Portal: Reference proteome datasets for three photosynthetic diatom species studied in this work were deposited to DNA Data Bank of Japan (http://getentry.ddbj.nig.ac.jp/) under the accession numbers BLYE0100001 to BLYE01000234 for the nuclear genome. Assembled genomes were deposited to DNA Data Bank of Japan (https://ddbj.nig.ac.jp/resource/bioproject/PRJDB11016 and https://ddbj.nig.ac.jp/resource/bioproject/PRJDB12553).

Functional annotation

The predicted protein coding genes were annotated using InterProScan, and RPS-BLAST search was performed against KEGG orthology database (63, 64). KO identifiers for Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways were assigned using KEGG Automatic Annotation Server (65). Transporter proteins were annotated with TransportTP (66) followed by manual curation. Reference proteome datasets for three photosynthetic diatom species were obtained from the JGI Genome Portal: *P. tricornutum* CCAP 1055/1 v2.0 (Phatr2_bd_unmapped_GeneModels_FilteredModels1_aa.fasta and Phatr2_chromosomes_geneModels_FilteredModels2_aa.fasta, 10,402 protein sequences in total), *T. pseudonana* CCMP 1335 (Thaps3_bd_unmapped_GeneModels_FilteredModels1_aa.fasta and Thaps3_chromosomes_geneModels_FilteredModels2_aa.fasta, 11,776 sequences), and *F. cylindrus* CCMP 1102 (Fracy1_GeneModels_FilteredModels3_aa.fasta, 21,066 sequences). KEGG and KO annotation was performed with them in the same manner as NIES-4239. Other details for annotation of CAZyme, cyclins, cyclin-dependent kinases, bZIP transcription factors, photoreceptor proteins, mitochondrial proteins, plastid proteins, and secretome proteins are described in the Supplementary Materials. Evolutionary analyses, comparative transcriptome analyses under the 12-hour light and 12-hour dark condition, those in different carbon sources, and biochemical experiments for lipids, fatty acids, and quinones are also described in the Supplementary Materials. Transcriptome data obtained in this study were deposited to DNA Data Bank of Japan (https://ddbj.nig.ac.jp/resource/bioproject/PRJDB11016 and https://ddbj.nig.ac.jp/resource/bioproject/PRJDB12553).

REFERENCES AND NOTES

1. J. M. Freese, C. E. Lane, Parasitism finds many solutions to the same problems in red algae (Florideophyceae, Rhodophyta). *Mol. Biochem. Parasitol.* 214, 105–111 (2017).
2. L. Hadarova, M. Vesteg, V. Hampl, J. Krajcovic, Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protozoa. *Cur. Genet.* 64, 365–387 (2018).
3. J. Janoušková, G. G. Paskerova, T. S. Minolubova, K. V. Mikhailov, T. Birey, V. V. Aleoshin, T. G. Simdyanov, Apicomplexan-like parasites are polyphyletic and widely but selectively dependent on cryptic plastid organelles. *eLife* 8, e49662 (2019).
4. J. C. Kissing, B. P. Brunk, J. Crabtree, M. J. Fraunholz, B. Gajria, A. J. Milgram, D. S. Pearson, J. Schug, A. Bahl, S. J. Diskin, H. Ginsburg, G. R. Grant, D. Gupta, P. Labo, L. Li, M. D. Mailman, S. K. McWeeney, P. Whetzel, C. J. Stoeckert Jr., D. S. Roos, The Plasmidome genome database. *Nature* 419, 490–492 (2002).
5. R. Kamikawa, N. Yubuki, M. Yoshida, M. Taira, N. Nakamura, K. Ishida, B. S. Leander, H. Miyashita, T. Hashimoto, S. Mayama, Y. Inagaki, Multiple losses of photosynthesis in *Nitzschia* (Bacillariophyceae). *Phycol. Res.* 63, 19–28 (2015).
6. R. G. Dorrell, T. Azuma, K. Nomura, G. A. de Kerdrel, L. Paoli, S. Yang, C. Bowler, K. Ishii, H. Miyashita, G. H. Gile, R. Kamikawa, Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 6914–6923 (2019).
7. M. Kayama, J. F. Chen, T. Nakada, Y. Nishimura, T. Shikakai, T. Azuma, H. Miyashita, S. Takaichi, Y. Kashiwaya, R. Kamikawa, A non-photosynthetic green alga illuminates the reductive evolution of plastid electron transport systems. *MBB* 18, 126 (2020).
8. M. Kayama, K. Maciszewski, A. Yabuki, H. Miyashita, A. Karnkowska, R. Kamikawa, Highly reduced plastid genomes of the non-photosynthetic dictyochophyceans *Preiodanomias* spp. (Ochrophyta, SAR) are retained for RNA-Glu-based organellar heme biosynthesis. *Front. Plant Sci.* 11, 602455 (2020).
9. L. Van Valen, Molecular evolution as predicted by natural selection. *J. Mol. Evol.* 3, 89–101 (1974).
10. G. Sun, Y. Xu, H. Liu, T. Sun, J. Zhang, C. Hettenhausen, G. Shen, J. Qi, Y. Qin, J. Li, L. Wang, W. Chang, Z. Guo, I. T. Baldwin, J. Wu, Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat. Commun.* 9, 2683 (2018).
11. C. W. Li, B. E. Volcani, Four new apochlorotic diatoms. *Br. Phycol. J.* 22, 375–382 (1987).
12. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281, 237–240 (1998).
13. D. G. Mann, The species concept in diatoms. *Phycologia* 38, 437–495 (1999).
14. E. V. Armburst, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, K. B. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Dettter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kröger, W. W. Y. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Monsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pozaur, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, D. S. Rokhsar, The genome...
15. C. Bowler, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, A. Martens, F. Maumus, R. P. Ottillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, E. Verret, J. A. Berges, C. Brownlee, J.-P. Cadoret, A. Chiotti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, M. Lommer, V. Martin-Jézéquel, P. M. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, J. K. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: Unsupervised microbial variant detection and genome assembly improvement. PLOS ONE 9, e112963 (2014).

17. G. W. Vurture, F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, A. Gruber, R. Hipkin, M. G. Janech, P. G. Kroth, F. Leese, E. A. Lindquist, B. R. Lyon, J. Martin, C. Mayer, M. Parker, H. Quesneville, J. A. Raymond, C. Uihle, R. E. Valas, K. U. Valentin, A. Z. Warden, E. V. Armbrust, M. D. Clark, C. Evolution, V. Green, V. Morisgn, C. Van Dosterhout, I. V. Grigoriev, Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. Nature 454, 536–540 (2017).

19. S. Coesel, M. Mangogna, T. Ishikawa, M. Heijde, A. Rogato, G. Finazzi, T. Todo, C. Bowler, Diatom PCCF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep. 10, 655–661 (2009).

21. A. Gruber, M. Jaubert, G. Enomoto, J. Bouly, R. Raniello, M. Thaler, M. Malviya, J. S. Bernardes, F. Rappaport, B. Gentili, M. J. J. Huysman, A. Carbone, C. Bowler, M. Ribera d’Alcalà, M. Ikeuchi, A. Falciatore, Diatom phycophyton phophates the existence of far-red-light-based sensing in the ocean. Plant Cell 28, 616–628 (2016).

23. A. E. Fortunato, M. Jaubert, G. Enomoto, J. Bouly, R. Raniello, M. Thaler, M. Malviya, J. S. Bernardes, F. Rappaport, B. Gentili, M. J. J. Huysman, A. Carbone, C. Bowler, M. Ribera d’Alcalà, M. Ikeuchi, A. Falciatore, Diatom phycophyton phophates the existence of far-red-light-based sensing in the ocean. Plant Cell 28, 616–628 (2016).

25. G. Kong, Y. Chen, Y. Deng, F. Feng, L. Jiang, L. Wan, M. Li, Z. Jiang, P. Xi, The basic leucine zipper transcription factor PBZP2 associated with the oxidative stress response is critical for pathogenicity of the lychee downy blight oomycete Peronosphaeria litchi. mSphere 5, e00621-20 (2020).

27. T. Fujisawa, S. Hirooka, O. Ohbayashi, R. Onuma, S. Miyagishima, Relationship between cell cycle and diel transcriptomic changes in metabolism in a unicellular red alga. Plant Physiol. 183, 1484–1501 (2020).

29. T. A. Richards, N. J. Talbot, Horizontal gene transfer in oomycetes: Playing with public goods. Nat. Rev. Microbiol. 11, 720–737 (2013).

31. S. Coesel, M. Mangogna, T. Ishikawa, M. Heijde, A. Rogato, G. Finazzi, T. Todo, C. Bowler, Diatom PCCF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. EMBO Rep. 10, 655–661 (2009).

33. A. Fernie, Z. Nikoloski, B. O. Palsson, A. E. Allen, Evolution and regulation of nitrogen flux during diel growth state transitions in the diatom Thalassiosira pseudonana. Proc. Natl. Acad. Sci. U.S.A. 110, 110–114 (2013).

35. S. Coesel, M. Mangogna, T. Ishikawa, M. Heijde, A. Rogato, G. Finazzi, T. Todo, C. Bowler, Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. New Phytol. 188, 52–66 (2010).

37. Y. I. Wolf, E. V. Koonin, Genome reduction as the dominant mode of evolution. Nat. Rev. Genet. 8, 73–83 (2007).

39. M. Nei, S. Kumar, Molecular Evolution and Phylogenetics (Oxford Univ. Press, New York, 2000), pp. 17–203.

41. T. A. Richards, N. J. Talbot, Horizontal gene transfer in oomycetes: Playing with public goods. Nat. Rev. Microbiol. 11, 720–737 (2013).

43. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584 (2002).

45. Y. Belkhadir, L. Yang, J. Heitzel, J. L. Dangl, J. Chory, The growth–defense pivot: Crisis management in plants mediated by LRR-RK surface receptors. Trends Biochem. Sci. 39, 447–456 (2014).

47. M. Nei, S. Kumar, Molecular Evolution and Phylogenetics (Oxford Univ. Press, New York, 2000), pp. 17–203.

51. K. R. Frischkorn, S. T. Haley, S. T. Dyhrman, Coordinated gene expression between phytoplankton and diatoms: A comprehensive analysis of transcriptome changes across four diatom species. Nat. Commun. 10, 4552 (2019).

53. J. Ashworth, S. Coesel, A. Lee, E. V. Armbrust, M. V. Orellana, N. S. Baliga, Genome-wide diel growth state transitions in the diatom Thalassiosira pseudonana. Proc. Natl. Acad. Sci. U.S.A. 110, 7518–7523 (2013).

55. M. S. Chauton, P. Winge, T. Brembu, O. Vadebon, A. M. Bones, Gene regulation of carbon fixation, storage, and utilization in the diatom Phaeodactylum tricornutum acclimated to light/dark cycles. Plant Physiol. 161, 1034–1048 (2013).

57. S. R. Smith, J. T. F. Gillard, A. B. Kustka, J. P. McCrow, J. H. Badger, H. Zheng, A. M. New, C. L. Dupont, T. Obata, R. A. Fernie, A. E. Allen, Correction: Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. PLOS Genet. 13, e1006668 (2017).
A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10, 210 (2010).

B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).

R. Ihaka, R. Gentleman, R: A language for data analysis and graphics. J. Comp. Graph. Stat. 5, 299–314 (1996).

M. J. L. de Hoon, S. Imoto, J. Nolan, S. Miyano, Open source clustering software. Bioinformatics 20, 1453–1454 (2004).

A. J. Saldanha, Java Treeview - Extensible visualization of microarray data. Bioinformatics 20, 3246–3248 (2004).

A. Löytynoja, Phylogeny-aware alignment with PRANK. Methods Mol. Biol. 1079, 155–170 (2014).

S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).

G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577 (2007).

R. Bouckaert, T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kuhnert, N. De Maio, M. Matschiner, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 15, e1006650 (2019).

S. Höhna, M. R. May, B. R. Moore, TESS: An R package for efficiently simulating phylogenetic trees and performing Bayesian inference of lineage diversification rates. Bioinformatics 32, 789–791 (2016).

Z. Yang, W. S. W. Wong, R. Nielsen, Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118 (2005).

Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591 (2007).

J.-F. Pombert, N. A. Blouin, C. Lane, D. Boucias, P. J. Keeling, A lack of parasitic reduction in the obligate parasitic green alga Helicosporidium. PLOS Genet. 10, e1004355 (2014).

S. Suzuki, R. Endoh, R. Manabe, M. Ohkuma, Y. Hirakawa, Multiple losses of photosynthesis and convergent reductive genome evolution in the colourless green algae Prototheca. Sci. Rep. 8, 940 (2018).

C. A. Durkin, J. A. Koester, S. J. Bender, E. V. Armbrust, The evolution of silicon transporters in diatoms. J. Phycol. 52, 716–731 (2016).

Acknowledgments: We thank F. Hayashi for the manual correction of the genome assembly and the gene annotation for the secretome. Computational work was, in part, performed at the SuperComputer System, Institute for Chemical Research, Kyoto University, and the NIG supercomputer at the National Institute of Genetics. Funding: This work was supported by JSPS Grants-in-Aid for Scientific Research (B) (19H03274 to R.K., 20H03305 to T.N., and 17H03723 to G.T.); JSPS Grants-in-Aid for Early-Career Scientists (20K15783 to Y.T.); JSPS Grant-in-Aid for Challenging Research (Exploratory) (21K19303 to G.T.); JSPS Grants-in-Aid for Scientific Research (A) (18H03743 to Y.K.); JSPS Grants-in-Aid for Scientific Research on Innovative Areas, Platform for Advanced Genome Science (16H06279 to R.K., T. Mock, M.S., Y.T., and Y.N.); NERC (NE/R000883/1 to T. Mock, C.v.O., A.T., K.S., and S.S.); School of Environmental Sciences at the University of East Anglia, UK (to T. Mock, C.v.O., A.T., K.S., and S.S.); The Gatsby Charitable Foundation and Biotechnology and Biological Sciences Research Council (BBSRC) (to S.K. and J.W.); and NIG-JOINT (7A2017, 6A2018, and 30A2019 to R.K.).

Author contributions: Conceptualization: R.K., G.T., T. Mock, and Y.N. Methodology: R.K., T. Mock, and Y.N. Investigation: J.W.; and NIG-JOINT (7A2017, 6A2018, and 30A2019 to R.K.).

Funding: This work was supported by JSPS Grants-in-Aid for Scientific Research (B) (19H03274 to R.K., 20H03305 to T.N., and 17H03723 to G.T.); JSPS Grants-in-Aid for Early-Career Scientists (20K15783 to Y.T.); JSPS Grant-in-Aid for Challenging Research (Exploratory) (21K19303 to G.T.); JSPS Grants-in-Aid for Scientific Research (A) (18H03743 to Y.K.); JSPS Grants-in-Aid for Scientific Research on Innovative Areas, Platform for Advanced Genome Science (16H06279 to R.K., T. Mock, M.S., Y.T., and Y.N.); NERC (NE/R000883/1 to T. Mock, C.v.O., A.T., K.S., and S.S.); School of Environmental Sciences at the University of East Anglia, UK (to T. Mock, C.v.O., A.T., K.S., and S.S.); The Gatsby Charitable Foundation and Biotechnology and Biological Sciences Research Council (BBSRC) (to S.K. and J.W.); and NIG-JOINT (7A2017, 6A2018, and 30A2019 to R.K.).

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Additional data related to this paper are available in DDBJ and Dryad as described in Methods.

Submitted 15 March 2021
Accepted 16 March 2022
Published 29 April 2022
10.1126/sciadv.aba5075