bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A single loss of photosynthesis in diatoms

Anastasiia Onyshchenko[1*], Elizabeth C. Ruck[1], Teofil Nakov[1], Andrew J. Alverson[1],

[1]University of Arkansas, Department of Biological Sciences, SCEN 601, Fayetteville, AR 72701 USA

***onyshchenko.ana@gmail.com**

Abstract

Loss of photosynthesis is a common and often repeated trajectory in nearly all major groups of photosynthetic eukaryotes. One small subset of 'apochloritic' diatoms in the genus *Nitzschia* have lost their ability to photosynthesize and require extracellular carbon for growth. Similar to other secondarily nonphotosynthetic taxa, apochloritic diatoms maintain colorless plastids with highly reduced plastid genomes. Although the narrow taxonomic breadth of apochloritic diatoms suggests a single loss of photosynthesis in the common ancestor of these species, previous phylogenetic analyses suggested that photosynthesis was lost multiple times. We sequenced additional phylogenetic markers from the nuclear and mitochondrial genomes for a larger set of taxa and found that the best trees for datasets representing all three genetic compartments provided low to moderate support for monophyly of apochloritic *Nitzschia*, consistent with a single loss of photosynthesis in diatoms. We sequenced the plastid genome of one apochloritic species and found that it was highly similar in all respects to the plastid genome of another apochloritic *Nitzschia* species, indicating that streamlining of the plastid genome had completed prior to the split of these two species. Finally, it is increasingly clear that some locales host relatively large numbers apochloritic *Nitzschia* species that span the phylogenetic diversity of the group, indicating that these species co-exist because of resource abundance or resource partitioning in ecologically favorable habitats. A better understanding of the phylogeny and ecology of this group, together with emerging genomic resources, will help identify the factors that have driven and maintained the loss of photosynthesis in this group, a rare event in diatoms.

Introduction

Photosynthetic eukaryotes (Archaeplastida) trace back to a single common ancestor, in which a eukaryotic host paired with a cyanobacterial endosymbiont that would eventually become the plastid, a fully integrated cellular organelle that is the site of photosynthesis (Archibald 2009, Keeling 2010). Although most archaeplastids remain photoautotrophic, loss of photosynthesis has occurred—often repeatedly—in nearly all major archaeplastid lineages (Hadariová et al. 2017). As many as five different lineages within the green algal orders Chlamydomonadales and Chlorellales have traded off photosynthesis for trophic strategies that include heterotrophy and obligate parasitism (Rumpf et al. 1996, Tartar and Boucias 2004, Yan et al. 2015). Partial or complete loss of photosynthesis has occurred in at least 11 different lineages of flowering plants, representing hundreds of species (Barkman et al. 2007); parasitic angiosperms obtain

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

extracellular carbon from the vascular tissues of host plants for all (holoparasites) or part (hemiparasites) of their life history (Těšitel 2016). Perhaps most strikingly, photosynthesis has been lost dozens or more times across the florideophycean red algae (e.g., Goff et al. 1996, Kurihara et al. 2010), which go on to parasitize a closely related photosynthetic species—a phenomenon known as adelphoparasitism (Blouin and Lane 2012). Photosynthesis has been lost in a broad range of taxa with secondary plastids as well, including euglenoids (Marin 2004), apicocomplexans (McFadden et al. 1996, Waller and McFadden 2005), ciliates (Reyes-Prieto et al. 2008), dinoflagellates (Saldarriaga et al. 2001), cryptophytes (Donaher et al. 2010, Martin-Cereceda et al. 2010), and stramenopiles (Tyler et al. 2006). Most nonphotosynthetic algae evolved from mixotrophic ancestors (Figueroa-Martinez et al. 2015), likely because they already had the means to secure extracellular carbon and because the energetic costs of mixotrophy are thought to be high (Raven 1997).

Diatoms are a lineage of stramenopile algae responsible for roughly 20% of global primary production (Field et al. 1998). They are ancestrally photosynthetic, and although the overwhelming majority of the estimated 100,000 or so diatom species remain photosynthetic, many species are mixotrophic, which allows them to use external sources of carbon for growth in fluctuating light conditions (Hellebust and Lewin 1977, Tuchman et al. 2006). A much smaller set of 20 or so mostly undescribed, colorless, free-living species in the genus *Nitzschia*, and one species in the closely related and morphologically similar genus Hantzschia, have abandoned photosynthesis altogether and rely exclusively on extracellular carbon for growth (Lewin and Lewin 1967). These "apochloritic" diatoms—the only known nonphotosynthetic diatom species—are often found in association with mangroves, and decaying seaweeds and sea grasses (Pringsheim 1956, Blackburn et al. 2009a, Kamikawa et al. 2015b).

The small number of species and narrow taxonomic range of apochloritic diatoms leads to the prediction, based on parsimony, that nonphotosynthetic diatoms are monophyletic, tracing back to a single loss of photosynthesis in their common ancestor. Despite the small number of species, however, they encompass a relatively broad range of morphological diversity (Blackburn et al. 2009b, Kamikawa et al. 2015b) and use a variety of carbon sources (Lewin and Lewin 1967, Hellebust and Lewin 1977), raising the possibility that obligate heterotrophy evolved multiple times—a hypothesis supported by a phylogenetic analysis of nuclear 28S d1–d2 sequences that separated apochloritic taxa into three separate clades (Kamikawa et al. 2015b). Monophyly of apochloritic species could not be rejected, however (Kamikawa et al. 2015b). These two competing hypotheses (one vs. multiple origins) have important implications for our understanding of the underlying phylogenetic, ecological, and genomic contexts of this radical trophic shift, which has occurred far less frequently in diatoms than it has in other groups.

Like most other lineages that have lost photosynthesis, apochloritic diatoms maintain highly reduced plastids and plastid genomes (Kamikawa et al. 2015a). Their plastids lack chlorophyll and thylakoids (Kamikawa et al. 2015b), and their plastid genomes have lost most photosynthesis-related genes, including all photosystem genes (Kamikawa et al. 2015a). The nearly complete set of ATP synthase genes in the plastid genome might function in ATP hydrolysis, creating a proton gradient that fuels protein import into the plastid (Kamikawa et al. 2015a). Carbon metabolism is highly compartmentalized in diatoms (Smith et al. 2012), and the large number of nuclear-encoded proteins targeted to the plastid point to a highly metabolically active and, as a result, indispensable (Kamikawa et al. 2017) organelle. Although comparative genomics is greatly improving our understanding of carbon metabolism in both photosynthetic (Smith et al. 2012) and nonphotosynthetic diatoms (Kamikawa et al. 2015a, 2017), the power of comparative genomics can only be fully leveraged within the framework of an accurate, densely sampled phylogenetic hypothesis.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We collected, isolated, and cultured several apochloritic *Nitzschia* species and sequenced common phylogenetic markers to test whether photosynthesis was lost one or multiple times. A combined dataset of nuclear, mitochondrial, and plastid genes support monophyly of apochloritic *Nitzschia* species, consistent with a single loss of photosynthesis in diatoms. Species were split between two major subclades that co-occur in habitats with apparently favorable, albeit poorly defined, conditions. We also sequenced and characterized the plastid genome of *Nitzschia* sp. and found it to be highly similar to that of another species in the same subclade, indicating rapid genomic streamlining following loss of photosynthesis in their common ancestor. The results presented here help frame a number of questions about the evolution, ecology, and genomic consequences of this rare, radical trophic shift in diatoms.

Materials and Methods

Collection and culturing of apochloritic *Nitzschia* All cultures originated from a single composite sample, collected 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 lat, –80.110783 long). This site was previously shown to host a diverse assemblage of apochloritic *Nitzschia* species (Blackburn et al. 2009a). Our sample consisted of near-surface plankton collected with a 10 μm mesh net, submerged sand (1m and 2m depth), and nearshore wet (but unsubmerged) sand. We selected for nonphotosynthetic species 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 extraction, PCR, and DNA sequencing Cells were rinsed with L1 medium and removed from agar plates by pipetting, briefly centrifuged, then broken with MiniBeadbeater-24 (BioSpec Products). We then extracted DNA with a Qiagen DNeasy Plant Mini Kit. Nuclear SSU and partial LSU rDNA genes were PCR-amplified and sequenced using published PCR conditions and primer sequences (Alverson et al. 2007). PCR and sequencing primers for two mitochondrial markers, cytochrome b (*cob*) and NADH dehydrogenase subunit 1 (*nad1*), are listed in Table S1. PCRs for mitochondrial genes used: 1.0–5.0 μL of DNA, 6.5 μL of Failsafe Buffer E (Epicentre Technologies), 0.5 μL of each primer (20 μM stocks), 0.5 units Taq polymerase, and ddH2O to a final volume of 25 μL. In a few cases, we used a nested PCR strategy to amplify the *nad1* gene. PCR conditions for *cob* and *nad1* genes were as follows: 95 °C for 5 minutes, 36 cycles of (95 °C for 60 s, 45 °C for 60 s, 72 °C for 60 s), and a final extension at 72 °C for 5 minutes. PCR products were sequenced on an ABI 3100 capillary sequencer. Raw sequences were assembled and edited with Geneious ver. 7.1.4 (Biomatters Ltd.) and deposited in GenBank (Table 1).

Phylogenetic analyses In addition to the data generated here, we compiled previously published data for the plastid 16S and nuclear 28S rDNA genes for Bacillariales to better enable comparisons with previous studies. We downloaded all Bacillariales sequences from Genbank, checked percent identity and coverage of each sequence to a local BLAST database comprised of apochloritic *Nitzschia* sequences, and reverse-complemented sequences if necessary. We kept Bacillariales sequences that were at least 20% identical and covered at least 40% of at least one target in the database (BLAST options: identity_cutoff=20, hsp_coverage_cutoff=40). We kept only the longest

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Taxa and sources of DNA, and GenBank accessions for sequences analyzed in this study. Apochloritic taxa are in bold face.

Genus	Species	Culture.ID	cob	nad1	lsu	16s
Nitzschia	sp.	ccmp2144	MH017326	MH017300	MH017352	
Nitzschia	dippelii	AJA014-6	MH017327	MH017301	MH017353	
Hantzschia	spectabilis	UTEX FD269	MH017328	MH017302	MH017354	
Hantzschia	elongata	UTEX FD421	MH017329	MH017303	MH017355	
Hantzschia	amphioxys	AJA013-11	MH017330	MH017304	MH017356	
Fragilariopsis	cylindrus	ccmp1022	JGI16035	JGI16036		
Pseudo-nitzschia	sp.	ccmp1447	MH017331	MH017305	MH017357	
Nitzschia	frustulum	ccmp1532	MH017332	MH017306	MH017358	
Nitzschia	cf. ovalis	ccmp1118	MH017333	MH017307	MH017359	
Cylindrotheca	closterium	ccmp1855	MG271845	MG271845		
Psammodictyon	constrictum	ccmp576	MH017334	MH017308	MH017360	
Nitzschia	laevis	ccmp1092	MH017335	MH017309	MH017361	
Nitzschia	sp.	OHI12.10	MH017336	MH017310	MH017362	
Nitzschia	cf. frequens	ccmp1500	MH017337	MH017311	MH017363	
Nitzschia	pusilla	ccmp2526	MH017338	MH017312	MH017364	
Nitzschia	cf. pusilla	ccmp581	MH017339	MH017313	MH017365	
Nitzschia	sp.	ccmp2533	MH017340	MH017314	MH017366	
Nitzschia	draveiliensis	AJA010-17	MH017341	MH017315	MH017367	
Nitzschia	palea var. debilis	AJA013-2	MH017342	MH017316	MH017368	
Nitzschia	gracilis	AJA010-53	MH017343	MH017317	MH017369	
Nitzschia	capitellata	AJA012-22	MH017344	MH017318	MH017370	
Nitzschia	**alba**	**ccmp2426**	MH017345	MH017319	MH017371	
Nitzschia	**leucosigma**	**ccmp2197**	MH017346	MH017320	MH017372	MH017379
Nitzschia	**sp.**	**ccmp579**	MH017347	MH017321	MH017373	
Nitzschia	**sp.**	**Nitz2**	MH017348	MH017322	MH017377	
Nitzschia	**sp.**	**Nitz4**	MH017349	MH017323	MH017375	MH017378
Nitzschia	**sp.**	**Nitz7**	MH017350	MH017324	MH017376	
Nitzschia	**sp.**	**Nitz8**	MH017351	MH017325	MH017374	MH017380

sequence for cases in which multiple downloaded sequences had the same NCBI Taxid identifier. The total number of sequences meeting these criteria were 46 for the 16S and 116 for the 28S rDNA genes.

We used SSU-ALIGN ver. 0.1 (Nawrocki et al. 2009) to align rDNA sequences (plastid 16S and the d1–d3 region of the nuclear 28S rDNA), using SSU-ALIGN's built-in covariance models of secondary structure for bacteria for the 16S alignment and a heterokont-specific covariance model for the 28S alignment (Nakov et al. 2014). We used SSU-MASK to remove poorly aligned regions and used these more conservative alignments for downstream analyses. The protein-coding *cob* and *nad1* genes from the mitochondrial genome were aligned by hand with Mesquite ver. 3.31 (Maddison, W. P. and Maddison, D. R. 2008) after color-coding nucleotide triplets by their conceptual amino acid translations. We built phylogenetic trees from each individual gene alignment and three concatenated alignments: (1) a concatenation of the two mitochondrial genes into a *cob*+*nad1* (heretofore mito dataset), (2) a concatenation of the two mitochondrial genes and the masked 28S alignment of newly generated 28S sequences (heretofore mito-lsu dataset), and (3) a concatenation of the two mitochondrial genes and the masked alignment of all 28S sequences (newly generated and downloaded from GenBank, heretofore mito-ncbi-lsu dataset). The alignments have

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

been archived in a ZENODO online repository (https://doi.org/10.5281/zenodo.1211571).

We used RAxML ver. 8.2.4 (Stamatakis 2014) to infer phylogenetic trees from each of the three concatenated alignments. For each dataset, we performed 10 maximum likelihood searches to find the best-scoring tree and a rapid bootstrap analysis of 250 pseudoreplicates per search (Stamatakis et al. 2008), for a total of 2500 bootstrap samples per alignment. We used the general time-reversible model (GTR) of nucleotide substitution, and we used a Γ distribution to accommodate rate variation across sites within each alignment (GTR+GAMMA, the default model in RAxML).

Plastid genome sequencing and analysis We sequenced the plastid genome of *Nitzschia* sp. (Nitz4) using the Illumina HiSeq2000 platform, with a 300-bp library and 90-bp paired-end reads. We removed adapter sequences and trimmed raw reads with Trimmomatic ver. 0.32 and settings 'ILLUMINACLIP=<TruSeq_adapters.fasta>:2:30:10, TRAILING=5, SLIDINGWINDOW=6:18, HEADCROP=9, MINLEN=50' (Bolger et al. 2014). We assembled trimmed reads using Ray ver. 2.3.1 with default settings and k-mer length of 45 (Boisvert et al. 2012). We assessed the assembly quality with QUAST ver. 2.3 (Gurevich et al. 2013), confirmed high genome-wide read coverage by mapping the trimmed reads to the assembly with Bowtie ver. 0.12.8 (Langmead et al. 2009), and evaluated these results with SAMtools. We used DOGMA (Wyman et al. 2004) and ARAGORN (Laslett and Canback 2004) to annotate the genome. The annotated genome has been archived in GenBank under accession MG273660. We used Easyfig (Sullivan et al. 2011) to perform a BLASTN-based synteny comparison of the plastid genomes of two apochloritic *Nitzschia* taxa, strains Nitz4 (this study) and NIES-3581 (Kamikawa et al. 2015a).

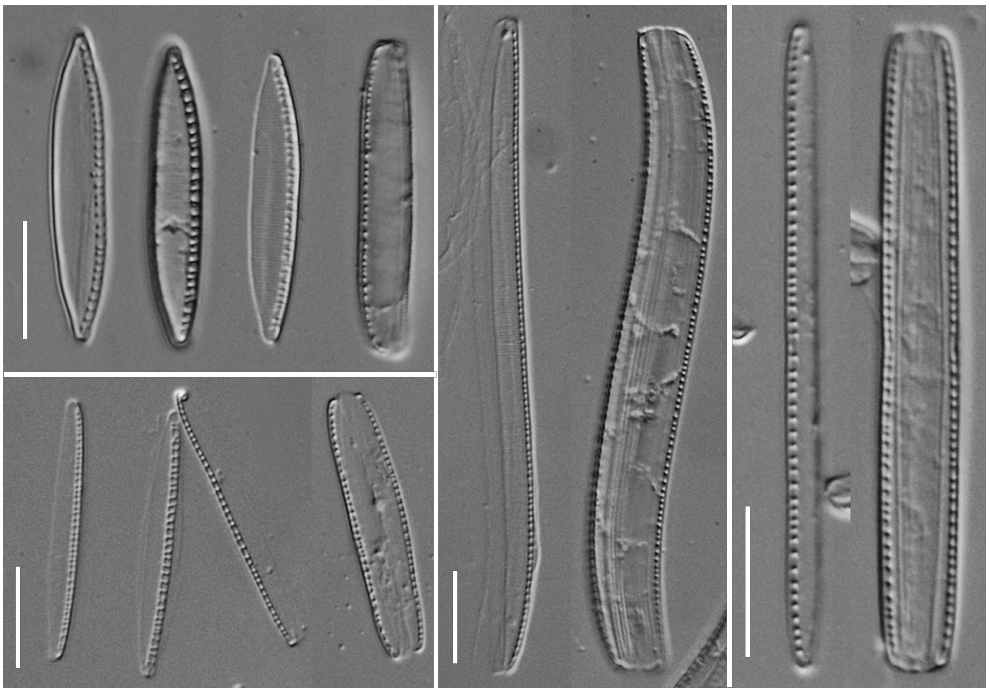


Figure 1. Light micrographs of newly sequenced nonphotosynthetic *Nitzschia* species. Clockwise from top: *Nitzschia* sp. nitz2, *Nitzschia* sp. nitz7, *Nitzschia* sp. nitz8, *Nitzschia* sp. nitz4.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Phylogeny of apochloritic *Nitzschia* We isolated and cultured four strains (Nitz2, Nitz4, Nitz7, and Nitz8) of apochloritic *Nitzschia* from Whiskey Creek (Florida, USA), including both linear and undulate forms (Fig. 1). Although cells grew well in culture, we never observed sexual reproduction and, unlike some other apochloritic *Nitzschia* strains that have been maintained in culture for decades, all of our strains experienced substantial size reduction and eventually died off.

We sequenced the nuclear 28S rDNA and mitochondrial *cob* and *nad1* genes for 26 Bacillariales taxa, including seven apochloritic strains (Table 1). We combined these data with GenBank sequences to reconstruct phylogenetic trees for individual and concatenated alignments. The plastid 16S rDNA gene supported monophyly of apochloritic *Nitzschia* species (Fig. 2A, Bootstrap proportion (BS)=100), though the exceptionally long branch lengths–driven by decreased GC content–raises the possibility that this grouping simply reflects shared nucleotide composition (Steel et al. 1993, Galtier and Gouy 1995, Kamikawa et al. 2015b). One way to reduce the influence of GC bias in tree inference is to transform the alignment into purine/pyrimidine (R/Y) coding. By treating all A and G nucleotides as R and all C and T nucleotides as Y, the GC bias present in some sequences is masked, the base frequencies are normalized, and the majority of phylogenetic signal originates from transversions (i.e., R $\leftrightarrow$ Y). The phylogeny resulting from the R/Y-transformed alignment should better reflect the history of the lineage rather than the nucleotide composition bias. Performing this transformation for the plastid 16S alignment, we again found strong support for monophyly of the apochloritic *Nitzschia* (bootstrap support = 95%; Fig 2B). This result suggests that monophyly of apochloritic *Nitzschia* species reconstructed with the plastid-encoded 16S gene might not necessarily be an artifact of the decreased GC content in the plastid genomes of nonphotosynthetic taxa.

A previous analysis of nuclear 28S d1–d2 rDNA sequences weakly supported non-monophyly of apochloritic taxa, though monophyly could not be rejected (Kamikawa et al. 2015b). Expanding the analysis to include the large number of Bacillariales 28S d1–d2 sequences from GenBank and newly sequenced taxa from this study (Table 1) returned a monophyletic grouping of apochloritic *Nitzschia* (Fig. 2C). We also sequenced two mitochondrial genes, *cob* and *nad1*, and although neither of them individually supported monophyly of apochloritic taxa (Fig. 2D, E), analysis of a concatenated mitochondrial *cob*+*nad1* alignment did return a clade of apochloritic *Nitzschia* (Fig. 2F). Concatenated alignments of nuclear 28S d1–d2 and the two mitochondrial genes supported monophyly of apochloritic *Nitzschia* as well, both for an analysis restricted to just newly sequenced taxa (Fig. 2G) and one that included both newly sequenced and GenBank taxa (Fig. 2H). In all cases, branch support for monophyly of apochloritic *Nitzschia* with nuclear and mitochondrial markers, analyzed individually or in combination, was less than 70% (Fig. 2). The fully labeled trees are available in Fig. S1.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

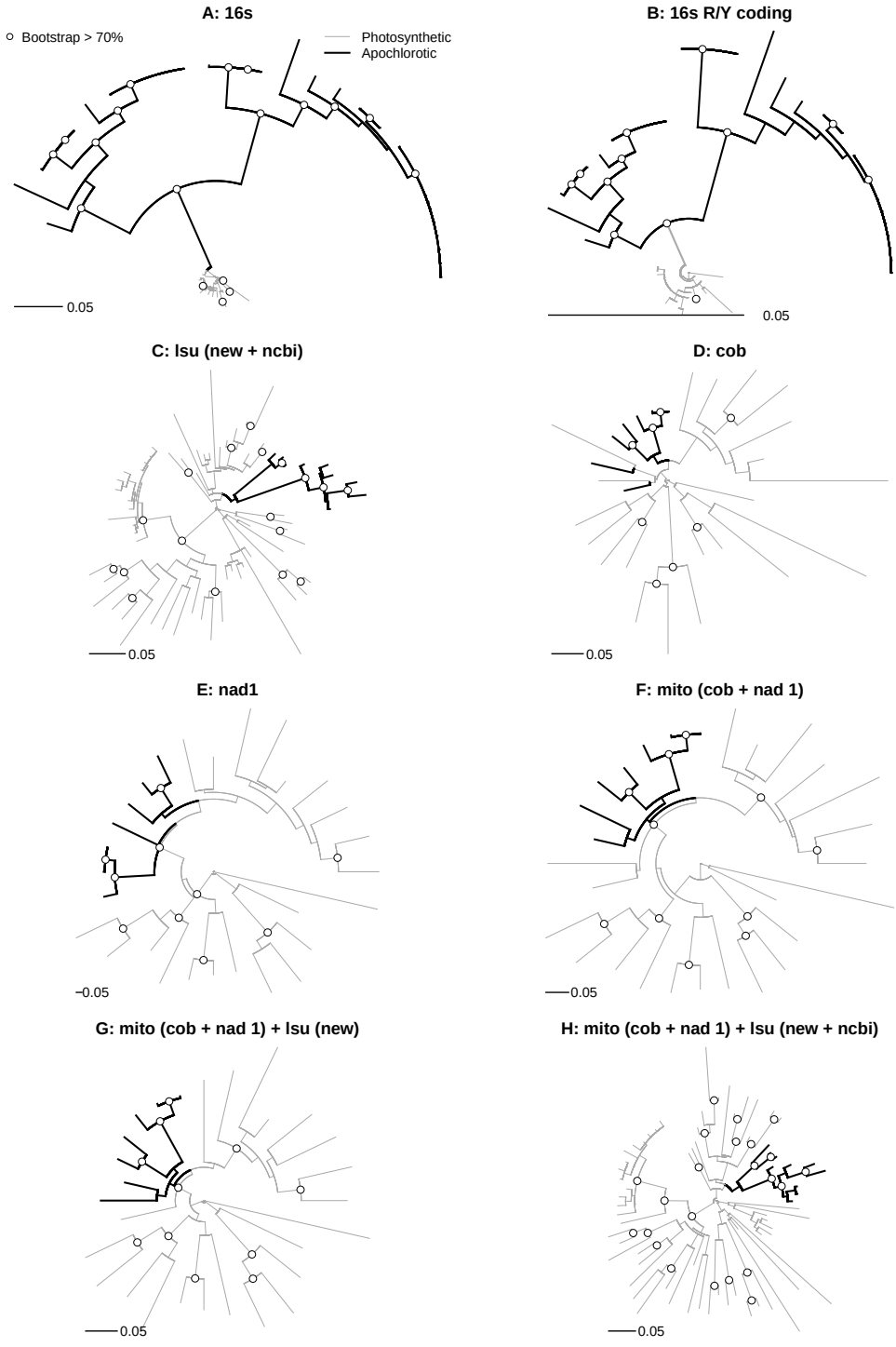


Figure 2. Phylogenetic trees inferred from plastid 16S (A), plastid 16S transformed into purine/pyrimidine (R/Y) coding (B), nuclear 28S d1–d2 genes for all newly sequenced taxa and data from GenBank ["lsu (new + ncbi)", C], mitochondrial *cob* (D), mitochondrial *nad1* (E), concatenated *cob* and *nad1* genes ["mito (*cob* + *nad1*)", F], concatenated *cob*, *nad1*, and nuclear 28S d1–d2 genes for all newly sequenced taxa ["mito (*cob* + *nad1*) + lsu (new)", G], and a concatenated alignment of *cob*, *nad1*, and nuclear 28S d1–d2 genes for taxa from this study and GenBank ["mito (*cob* + *nad1*) + lsu (new + ncbi)", H]. For the phylogenies in panels C and H, we removed branches shorter than 0.00001 units for clarity. The full phylogenies are available in Supplementary Figure 1. Thicker black branches correspond to apochlorotic taxa. White points identify nodes with bootstrap support >70%.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Plastid genome sequencing and analysis The plastid genome of *Nitzschia* sp. Nitz4 maps as a circular chromosome and has the conserved quadripartite architecture typical of most plastid genomes, with two inverted repeat (IR) regions and small and large single copy regions (SSC and LSC, respectively). At 67,764 bp in length, the genome is roughly half the size of plastid genomes from photosynthetic diatoms (Ruck et al. 2014, Yu et al. 2018). The genome contains minimal intergenic DNA, with 12 genes that overlap by anywhere from 1–61 bp in length and another two genes that are immediately adjacent to one another. (Table S2). The genome consists mainly of genes with conserved housekeeping functions, including 32 tRNA and six rRNA genes. All three rRNA and six of the tRNA genes are present twice in the genome because of their location in the IR region. More than half (45) of the 73 protein-coding genes encode ribosomal proteins or subunits of RNA polymerase, with ATP synthase genes and ORFs constituting most of the remaining genes. Some of the ORFs include ones (ycf41, ycf89 and ycf90) that are highly conserved among diatoms (Ruck et al. 2014, Yu et al. 2018). The genome also contains Sec-independent protein translocator TatC and and SecA subunit of Sec-mediated transport system, iron-sulfur clusters (SufB and SufC), molecular chaperones dnaK, and a protease (ClpC). The genome is highly AT-rich (77.6% A+T).

Similar to the plastid genome of another apochloritic diatom, *Nitzschia* sp. NIES-3581 (aka IriIs04) (Kamikawa et al. 2015a), the genome of *Nitzschia* sp. Nitz4 is missing all genes encoding subunits of photosystem I and II, proteins of cytochrome b6f complex, carbon fixation system, porphyrin and thiamine metabolism. As well Nitz4 lacks several conserved ORFs and membrane translocators subunits, dnaB helicase, chaperonine groEL and ftsH cell division protein. The genomes of Nitz4 and NIES-3581 are highly similar in size, gene content, gene order, nucleotide composition, and sequence (Fig. 3). The loss of rps20 in Nitz4, which appears to have been replaced by a unique ORF (orf122), is among the few minor differences between the two genomes (Fig. 3). Two gene fusions, orf96-atpB and orf122-rpoB, present in Nitz4 are separated in NIES-3581. Likewise, the overlapping dnaK-tRNA-Arg genes in NIES-3581 are adjacent but non-overlapping in Nitz4.

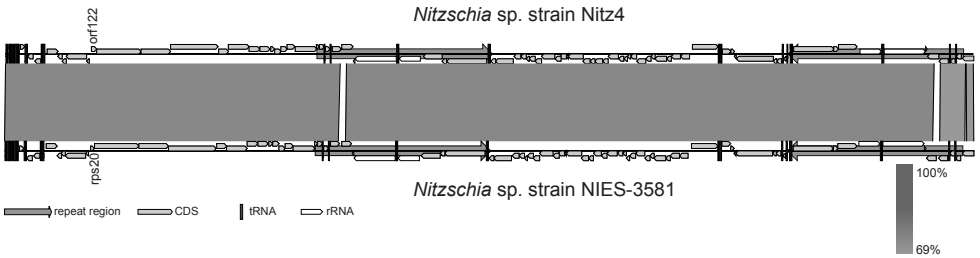


Figure 3. Conserved synteny, gene content, and sequence in the plastid genomes of two non-photosynthetic diatoms in the genus *Nitzschia*, Nitz4 (this study) and NIES-3581 (Kamikawa et al. 2015a). Unlabeled genes are shared between the two species.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Monophyly of apochloritic diatoms The small number of species (roughly 20) and narrow taxonomic range (all Bacillariales, mostly all *Nitzschia*) of apochloritic diatoms suggests that these species—the only known nonphotosynthetic diatoms—are monophyletic, reflecting a single loss of photosynthesis and transition to obligate heterotrophy in their common ancestor. Previous phylogenetic tests of this hypothesis were equivocal, however, with the plastid 16S rDNA sequences supporting monophyly of apochloritic species and the nuclear 28S d1–d2 gene splitting them, albeit with low bootstrap support, into three separate clades (Kamikawa et al. 2015b). Further underscoring the uncertainty in their relationships, the non-normalized, highly AT-rich plastid 16S sequences may have resulted in long-branch artifacts (Steel et al. 1993, Galtier and Gouy 1995), and the nuclear 28S dataset could not reject monophyly of apochloritic species (Kamikawa et al. 2015b), pointing to insufficient signal in this relatively short sequence fragment to address this question. As efforts to understand the causes and genomic consequences of loss of photosynthesis in diatoms accelerate (Kamikawa et al. 2015a, 2015b, 2017), it is important to determine if the shift occurred one or multiple times. The number, pattern, and timing of shifts has important implications for understanding the ease of transition(s) to obligate heterotrophy, the ecological setting of the transition(s), and, in practical terms, the power of comparative approaches to resolve these questions—as evolutionary replication, represented in this case by multiple losses of photosynthesis—maximizes the power of comparative methods to reveal possible mechanisms underlying these kinds of evolutionary transitions (e.g., Maddison and FitzJohn 2015). A single transition would, by contrast, greatly limit the power of comparative methods to uncover the underlying genomic and ecological drivers of the switch away from auto- or mixotrophy to obligate heterotrophy.

To address this problem, we increased both the number of taxa and genes available for phylogenetic analyses. We also normalized the plastid 16S rDNA sequences to guard against artefactual grouping of apochloritic species with highly AT-rich plastid genomes. Although support was generally low, the best trees inferred from genes representing all three genetic compartments uniformly supported monophyly of apochloritic *Nitzschia* (Fig. 2). In short, the best available data support a single loss of photosynthesis in *Nitzschia* and, by extension, diatoms as a whole. In light of this, future research efforts can focus attention away from questions that naturally arise for characters that evolve multiple times (e.g., the role of convergent evolution) and focus more on firmly placing the apochloritic clade within the broader Bacillariales phylogeny to help identify, for example, the ecological and mixotrophic characteristics of their closest relatives. Considerable divergence and phylogenetic structure exists within the apochloritic clade, with the largest combined-data tree (mito-lsu-new-ncbi) splitting taxa into two major subclades (Figs. 2 and S1). The modest level of species diversity, in turn, might make it feasible to sample each subclade more-or-less exhaustively and apply comparative approaches that will be useful for understanding the genomic and metabolic consequences of the switch to heterotrophy, including, for example, the rate of decay of the photosynthetic apparatus. All of this requires a robust, time-calibrated phylogeny of Bacillariales, which is one of the largest and most species-rich taxonomic orders of diatoms (Kociolek et al. 2018). This and other phylogenetic studies have begun resolving some parts of the Bacillariales tree using small numbers of commonly used phylogenetic markers (e.g., Lundholm et al. 2002, Rimet et al. 2011, Smida et al. 2014), but many relationships remain unresolved. With most of the traditional markers now more-or-less exhausted, effort should turn to much deeper sampling of the nuclear genome, which appears to hold great promise for resolving relationships across many phylogenetic scales within diatoms (Parks et al. 2018).

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The ecology and biogeography of apochloritic diatoms Although relatively few apochloritic *Nitzschia* species have been formally described and named, the amount of phylogenetic diversity described in this (Fig. 2) and other studies (Kamikawa et al. 2015b), suggests that this clade could easily comprise on the order of 20 species. This seems even more probable when considering the relatively modest historical efforts to collect and characterize apochloritic diatoms. One emerging, and quite striking, theme among the small number of studies that focused on characterizing species diversity in this group is the large number and diversity of apochloritic *Nitzschia* that co-occur over very small spatial scales (e.g., within a sample from a single site). The diversity is apparent from both morphological and molecular phylogenetic evidence alike (Figs. 1 and 2; Blackburn et al. 2009a, Kamikawa et al. 2015b). Although some of the apochloritic taxa in our trees came from culture collections (Table 1 and Fig. S1), most of the taxa derive from a small number of mangrove-dominated habitats in Japan (Kamikawa et al. 2015b) and the United States (Figs. 1 and 2; Blackburn et al. 2009a). Both of these sites host multiple (apparently undescribed) apochloritic *Nitzschia* species that span the full phylogenetic breadth of the clade (Fig. S1). Focused efforts to collect and culture apochloritic *Nitzschia* from amenable habitats (e.g., seagrasses and mangroves) worldwide will show whether this anecdotal trend holds. If it is upheld, the next natural step will be to determine whether species co-occurrence is made possible by the sheer abundance of local resources in these habitats or rather by resource partitioning among species, either in fine-scale microhabitats or through specialization on different carbon sources. Of course, both of these alternatives require knowledge of the organic carbon species used by these species. Nevertheless, the phylogeny clearly shows species are not clustered geographically (Fig. S1), which suggests that many or most apochloritic diatom species have broad, presumably worldwide, geographic distributions and that the local species pool at any one site is the product of dispersal, not in situ speciation.

Plastid genome reduction in apochloritic diatoms Although probably different species, the two *Nitzschia* strains with sequenced plastid genomes belong to the same subclade (Fig. S1A, bottom) and so, in the context of the entire apochloritic clade, are very close relatives. Their close relationship was also evident in their plastid genomes, which were highly similar in nearly all respects. Additional sampling across the entire apochloritic clade will show whether all species share the same fundamental, highly reduced plastid genome—indicative of rapid genomic streamlining following the loss of photosynthesis in their common ancestor—or whether decay of the plastid genome is ongoing in some taxa. Similar to some green algae (Knauf and Hachtel 2002), diatoms have retained a nearly full set of ATP synthase genes whose products, instead of functioning in ATP hydrolysis, generate a proton gradient for tat-dependent protein translocation across the thylakoid membrane (Kamikawa et al. 2015a). This model underscores both the compartmentalized nature of carbon metabolism in diatoms (Smith et al. 2012, Kamikawa et al. 2017) and, consequently, the indispensable nonphotosynthetic plastids in diatoms.

In addition to understanding the specific consequences for the plastid genome to following the loss of photosynthesis, a fuller understanding of these plastid genomes can shed light on photosynthetic diatom plastids as well. For example, the retention several conserved ORFs in these nonphotosynthetic genomes—in the context of near wholesale loss of the photosynthetic apparatus—strongly suggests these conserved, diatom-specific ORFs have functions that are unrelated, or only peripherally related, to photosynthesis. These ORFs include ycf41, ycf89, and ycf90. Finally, the loss of photosynthesis can have cascading effects on rates of sequence evolution all three genomes (Nickrent et al. 1998). Branch lengths based on plastid, mitochondrial, and nuclear genes showed that the

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

AT-driven rate acceleration in the plastid genomes of apochloritic species appears to be restricted to the plastid genome alone (Fig. 2). This does not rule out that there have been other effects on the mitochondrial and nuclear genomes including, for example, gene content and whether some of the missing plastid genes have been transferred to the nucleus.

Conclusions The discovery that apochloritic *Nitzschia* are monophyletic represents an important step forward in understanding the loss of photosynthesis and switch to obligate heterotrophy in diatoms—a transition that has occurred just once in a lineage of some 100,000 species of photoautotrophs. A clearer view of this relationship highlights new research avenues and priorities, including more intensive taxon sampling within Bacillariales to identify the closest relatives of the apochloritic clade, which almost certainly evolved from a mixotrophic ancestor (Hellebust and Lewin 1977, Figueroa-Martinez et al. 2015). The exact carbon sources, modes of carbon uptake and utilization, and the degree of carbon specialization within and across species remains unclear for photosynthetic and nonphotosynethetic *Nitzschia* alike. Bacillariales features a diverse set of taxa with fascinating biology, motivating the development of excellent genomic resources for this group (Basu et al. 2017, Kamikawa et al. 2017, Mock et al. 2017). A nuclear genome sequence for a nonphotosynthetic *Nitzschia* could help address several of these outstanding questions, leading to hypotheses that can be directly tested in a group of diatoms that has proven highly amenable to experimental manipulation (e.g., Lewin and Lewin 1967, Azam and Volcani 1974, McGinnis and Sommerfeld 2000).

Acknowledgments

We thank Sarah Hamsher, Nina Lundholm, and Pat Kociolek for early discussions on the phylogeny of Bacillariales. This work was supported by National Science Foundation (NSF) (grant number DEB-1353131) and multiple awards from the Arkansas Biosciences Institute to AJA. 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.

References

1. A. J. Alverson, R. K. Jansen, and E. C. Theriot. Bridging the rubicon: Phylogenetic analysis reveals repeated colonizations of marine and fresh waters by thalassiosiroid diatoms. *Mol. Phylogenet. Evol.*, 45(1):193–210, 2007.

2. J. M. Archibald. The puzzle of plastid evolution. *Curr. Biol.*, 19(2):R81–8, Jan. 2009.

3. F. Azam and B. E. Volcani. Role of silicon in diatom metabolism. *Arch. Microbiol.*, 101(1):1–8, Jan. 1974.

4. A. C. Barbrook, C. J. Howe, and S. Purton. Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci.*, 11(2):101–108, Feb. 2006.

5. T. J. Barkman, J. R. McNeal, S. H. Lim, G. Coat, H. B. Croom, N. D. Young, and C. W. DePamphilis. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol. Biol.*, 7(1):248, Dec. 2007.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6. S. Basu, S. Patil, D. Mapleson, M. T. Russo, L. Vitale, C. Fevola, F. Maumus, R. Casotti, T. Mock, M. Caccamo, M. Montresor, R. Sanges, and M. I. Ferrante. Finding a partner in the ocean: Molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytol.*, 215(1):140–156, 2017.

7. M. V. Blackburn, F. Hannah, and A. Rogerson. First account of apochlorotic diatoms from intertidal sand of a south florida beach. *Estuar. Coast. Shelf Sci.*, 84(4):519–526, Oct. 2009.

8. M. V. Blackburn, F. Hannah, and A. Rogerson. First account of apochlorotic diatoms from mangrove waters in florida. *J. Eukaryot. Microbiol.*, 56(2):194–200, Mar. 2009.

9. N. A. Blouin and C. E. Lane. Red algal parasites: Models for a life history evolution that leaves photosynthesis behind again and again. *Bioessays*, 34(3):226–235, Mar. 2012.

10. N. A. Blouin and C. E. Lane. Red algae provide fertile ground for exploring parasite evolution. *Perspectives in Phycology*, 3(1):11–19, Apr. 2016.

11. S. Boisvert, F. Raymond, E. Godzaridis, F. Laviolette, and J. Corbeil. Ray meta: Scalable *de novo* metagenome assembly and profiling. *Genome Biol.*, 13(12):R122, Dec. 2012.

12. A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, Aug. 2014.

13. T. Borza, C. E. Popescu, and R. W. Lee. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga prototheca wickerhamii. *Eukaryot. Cell*, 4(2):253–261, Feb. 2005.

14. L. Bromham, P. F. Cowman, and R. Lanfear. Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evol. Biol.*, 13(1):126, June 2013.

15. R. A. Bungard. Photosynthetic evolution in parasitic plants: Insight from the chloroplast genome. *Bioessays*, 26(3):235–247, Mar. 2004.

16. J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Müller, N. Pande, Z. Shang, N. Yu, and R. R. Gutell. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3:2, 2002.

17. T. Cavalier-Smith. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.*, 46(4):347–366, July 1999.

18. T. Cavalier-Smith. Evolution and relationships of algae: Major branches of the tree of life. In J. Brodie and J. Lewis, editors, *Unravelling the Algae: The Past, Present, and Future of Algal Systematics*, pages 21–55. CRC Press, Boca Raton, 2007.

19. T. Cavalier-Smith and E. E. Y. Chao. Phylogeny and megasystematics of phagotrophic heterokonts (kingdom chromista). *J. Mol. Evol.*, 62(4):388–420, Apr. 2006.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. N. Donaher, G. Tanifuji, N. T. Onodera, S. A. Malfatti, P. S. G. Chain, Y. Hara, and J. M. Archibald. The complete plastid genome sequence of the secondarily nonphotosynthetic alga *Cryptomonas paramecium*: Reduction, compaction, and accelerated evolutionary rate. *Genome Biol. Evol.*, 1(0):439–448, Nov. 2010.

21. N. M. Fast, L. Xue, S. Bingham, and P. J. Keeling. Re-examining alveolate evolution using multiple protein molecular phylogenies. *J. Eukaryot. Microbiol.*, 49(1):30–37, Jan. 2002.

22. C. B. Field, M. J. Behrenfeld, J. T. Randerson, and P. Falkowski. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*, 281(5374):237–240, July 1998.

23. F. Figueroa-Martinez, A. M. Nedelcu, D. R. Smith, and A. Reyes-Prieto. When the lights go out: The evolutionary fate of free-living colorless green algae. *New Phytol.*, 206(3):972–982, May 2015.

24. N. Galtier and M. Gouy. Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Natl. Acad. Sci. U. S. A.*, 92(24):11317–11321, Nov. 1995.

25. G. Gockel and W. Hachtel. Complete gene map of the plastid genome of the nonphotosynthetic euglenoid flagellate astasia longa. *Protist*, 151(4):347–351, Dec. 2000.

26. L. J. Goff, D. A. Moon, P. Nyvall, B. Stache, K. Mangin, and G. Zuccarello. The evolution of parasitism in the red algae: Molecular comparisons of adelphoparasites and their hosts. *J. Phycol.*, 32(2):297–312, 1996.

27. R. R. L. Guillard. A mutant of *Chlamydomonas moewusii* lacking contractile vacuoles. *J. Eukaryot. Microbiol.*, 7(3):262–268, 1960.

28. R. R. L. Guillard and P. E. Hargraves. *Stichochrysis immobilis* is a diatom, not a chrysophyte. *Phycologia*, 32(3):234–236, May 1993.

29. A. Gurevich, V. Saveliev, N. Vyahhi, and G. Tesler. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8):1072–1075, Apr. 2013.

30. L. Hadariová, M. Vesteg, V. Hampl, and J. Krajčovič. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. *Curr. Genet.*, Oct. 2017.

31. H. S. Heide-Jørgensen. *Parasitic flowering plants*. Brill, Leiden, 2008.

32. J. A. Hellebust and J. Lewin. Heterotrophic nutrition. In D. Werner, editor, *The Biology of Diatoms*, volume 13, pages 169–197. University of California Press, 1977.

33. K. A. Hildebrand, S. L.-Y. Woo, D. W. Smith, C. R. Allen, M. Deie, B. J. Taylor, and C. C. Schmidt. The effects of Platelet-Derived growth Factor-BB on healing of the rabbit medial collateral ligament. *Am. J. Sports Med.*, 26(4):549–554, July 1998.

34. R. Kamikawa, D. Moog, S. Zauner, G. Tanifuji, K.-I. Ishida, H. Miyashita, S. Mayama, T. Hashimoto, U. G. Maier, J. M. Archibald, and Y. Inagaki. A non-photosynthetic diatom reveals early steps of reductive evolution in plastids. *Mol. Biol. Evol.*, 34(9):2355–2366, Sept. 2017.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

35. R. Kamikawa, G. Tanifuji, S. A. Ishikawa, K. I. Ishii, Y. Matsuno, N. T. Onodera, K. I. Ishida, T. Hashimoto, H. Miyashita, S. Mayama, and Y. Inagaki. Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Mol. Biol. Evol.*, 32(10):2598–2604, Oct. 2015.

36. R. Kamikawa, N. Yubuki, M. Yoshida, M. Taira, N. Nakamura, K. I. Ishida, B. S. Leander, H. Miyashita, T. Hashimoto, S. Mayama, and Y. Inagaki. Multiple losses of photosynthesis in *Nitzschia* (bacillariophyceae). *Phycological Res.*, 63(1):19–28, Jan. 2015.

37. K. Katoh and D. M. Standley. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.*, 30(4):772–780, Apr. 2013.

38. P. J. Keeling. The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365(1541):729–748, Mar. 2010.

39. U. Knauf and W. Hachtel. The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Mol. Genet. Genomics*, 267(4):492–497, June 2002.

40. J. P. Kociolek, K. Balasubramanian, S. Blanco, M. Coste, L. Ector, Y. Liu, M. Kulikovskiy, N. Lundholm, T. Ludwig, M. Potapova, F. Rimet, K. Sabbe, S. Sala, E. Sar, J. Taylor, B. Van de Vijver, C. E. Wetzel, D. M. Williams, A. Witkowski, and J. Witkowski. DiatomBase. http://www.diatombase.org/, 2018. Accessed: 2018-4-5.

41. K. Krause. From chloroplasts to "cryptic" plastids: Evolution of plastid genomes in parasitic plants. *Curr. Genet.*, 54(3):111–121, Sept. 2008.

42. A. Kurihara, T. Abe, M. Tani, and A. R. Sherwood. Molecular phylogeny and evolution of red algal parasites: a case study of *Benzaitenia*, *Janczewskia*, and *Ululania* (ceramiales). *J. Phycol.*, 46(3):580–590, 2010.

43. B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, 10(3):R25, Mar. 2009.

44. D. Laslett and B. Canback. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.*, 32(1):11–16, Jan. 2004.

45. J. R. Leake. Plants parasitic on fungi: Unearthing the fungi in myco-heterotrophs and debunking the 'saprophytic' plant myth. *Mycologist*, 19(3):113–122, Aug. 2005.

46. C. A. Leander. *Phylogeny of the Labyrinthulomycota*. PhD thesis, The University of Georgia, 2001.

47. B. Lemaire, S. Huysmans, E. Smets, and V. Merckx. Rate accelerations in nuclear 18S rDNA of mycoheterotrophic and parasitic angiosperms. *J. Plant Res.*, 124(5):561–576, Sept. 2011.

48. J. Lewin and R. A. Lewin. Culture and nutrition of some apochlorotic diatoms of the genus *Nitzschia*. *Microbiology*, 46(3):361–367, Mar. 1967.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

49. N. Lundholm, N. Daugbjerg, and Ø. Moestrup. Phylogeny of the bacillariaceae with emphasis on the genus pseudo-nitzschia (bacillariophyceae) based on partial LSU rDNA. *Eur. J. Phycol.*, 37(1):115–134, Mar. 2002.

50. W. P. Maddison and R. G. FitzJohn. The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.*, 64(1):127–136, Jan. 2015.

51. Maddison, W. P. and Maddison, D. R. Mesquite: A modular system for evolutionary analysis. *Evolution*, 62:1103–1118, 2008.

52. D. G. Mann. The species concept in diatoms. *Phycologia*, 38(6):437–495, Nov. 1999.

53. B. Marin. Origin and fate of chloroplasts in the euglenoida. *Protist*, 155(1):13–14, Mar. 2004.

54. M. Martin-Cereceda, E. C. Roberts, E. C. Wootton, E. Bonaccorso, P. Dyal, A. Guinea, D. Rogers, C. J. Wright, and G. Novarino. Morphology, ultrastructure, and small subunit rDNA phylogeny of the marine heterotrophic flagellate *Goniomonas* aff. *amphinema*. *J. Eukaryot. Microbiol.*, 57(2):159–170, Mar. 2010.

55. G. I. McFadden, M. E. Reith, J. Munholland, and N. Lang-Unnasch. Plastid in human parasites. *Nature*, 381(6582):482, June 1996.

56. K. M. McGinnis and M. R. Sommerfeld. PLASTID FATTY ACID BIOSYNTHESIS IN THE DIATOMS NITZSCHIA ALBA AND NITZSCHIA LAEVIS. *J. Phycol.*, 36(s3):46–46, Dec. 2000.

57. T. Mock, R. P. Otillar, J. Strauss, M. McMullan, P. Paajanen, J. Schmutz, A. Salamov, R. Sanges, A. Toseland, B. J. Ward, A. E. Allen, C. L. Dupont, S. Frickenhaus, F. Maumus, A. Veluchamy, T. Wu, K. W. Barry, A. Falciatore, M. I. Ferrante, A. E. Fortunato, G. Glöckner, 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. Uhlig, R. E. Valas, K. U. Valentin, A. Z. Worden, E. V. Armbrust, M. D. Clark, C. Bowler, B. R. Green, V. Moulton, C. van Oosterhout, and I. V. Grigoriev. Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature*, 541(7638):536–540, Jan. 2017.

58. C. W. Morden, K. H. Wolfe, C. W. dePamphilis, and J. D. Palmer. Plastid translation and transcription genes in a non-photosynthetic plant: Intact, missing and pseudo genes. *EMBO J.*, 10(11):3281–3288, 1991.

59. B. R. Morton. Chloroplast DNA codon use: evidence for selection at the psb a locus based on tRNA availability. *J. Mol. Evol.*, 37(3):273–280, Sept. 1993.

60. B. R. Morton. Selection on the codon bias of chlamydomonas reinhardtii chloroplast genes and the plant psba gene. *J. Mol. Evol.*, 43(1):28–31, July 1996.

61. B. R. Morton. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *J. Mol. Evol.*, 46(4):449–459, Apr. 1998.

62. B. R. Morton and B. G. So. Codon usage in plastid genes is correlated with context, position within the gene, and amino acid content. *J. Mol. Evol.*, 50(2):184–193, Feb. 2000.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

63. T. Nakov, E. C. Ruck, Y. Galachyants, S. A. Spaulding, and E. C. Theriot. Molecular phylogeny of the Cymbellales (Bacillariophyceae, Heterokontophyta) with a comparison of models for accommodating rate variation across sites. *Phycologia*, 53(4):359–373, June 2014.

64. E. P. Nawrocki, D. L. Kolbe, and S. R. Eddy. Infernal 1.0: Inference of RNA alignments. *Bioinformatics*, 25(10):1335–1337, May 2009.

65. H. E. Neuhaus and M. J. Emes. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 51(1):111–140, June 2000.

66. D. L. Nickrent, R. J. Duff, a. E. Colwell, a. D. Wolfe, N. D. Young, K. E. Steiner, and C. W. DePamphilis. Molecular phylogenetic and evolutionary studies of parasitic plants. In *Molecular Systematics of Plants II. DNA Sequencing.*, pages 211–241. Springer US, Boston, MA, 1998.

67. J. Ohlrogge. Lipid biosynthesis. *the Plant Cell Online*, 7(7):957–970, July 1995.

68. M. B. Parks, N. J. Wickett, and A. J. Alverson. Signal, uncertainty, and conflict in phylogenomic data for a diverse lineage of microbial eukaryotes (diatoms, bacillariophyta). *Mol. Biol. Evol.*, 35(1):80–93, Jan. 2018.

69. M. C. Press and A. L. Gurney. Plant eats plant: Sap-feeding witchweeds and other parasitic angiosperms. *Biologist*, 47(4):189–193, Sept. 2000.

70. E. G. Pringsheim. Micro-organisms from decaying seaweed. *Nature*, 178(4531):480–481, Sept. 1956.

71. J. A. Raven. Phagotrophy in phototrophs. *Limnol. Oceanogr.*, 42(1):198–205, Jan. 1997.

72. A. Reyes-Prieto, A. Moustafa, and D. Bhattacharya. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. *Curr. Biol.*, 18(13):956–962, July 2008.

73. F. Rimet, L. Kermarrec, A. Bouchez, L. Hoffmann, L. Ector, and L. K. Medlin. Molecular phylogeny of the family bacillariaceae based on 18S rDNA sequences: focus on freshwater nitzschia of the section lanceolatae. *Diatom Res.*, 26(3):273–291, Sept. 2011.

74. F. E. Round, R. M. Crawford, and D. G. Mann. *Diatoms: Biology and Morphology of the Genera*. Cambridge University Press, May 1990.

75. E. C. Ruck, T. Nakov, R. K. Jansen, E. C. Theriot, and A. J. Alverson. Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. *Genome Biol. Evol.*, 6(3):644–654, Mar. 2014.

76. R. Rumpf, D. Vernon, D. Schreiber, and C. W. Birky. Evolutionary consequences of the loss of photosynthesis in chlamydomonadaceae: Phylogenetic analysis of *Rrn18* (18s rDNA) in 13 *Polytoma* strains (chlorophyta). *J. Phycol.*, 32(1):119–126, Feb. 1996.

77. J. F. Saldarriaga, F. J. R. Taylor, P. J. Keeling, and T. Cavalier-Smith. Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J. Mol. Evol.*, 53(3):204–213, Sept. 2001.

78. E. D. Salomaki and C. E. Lane. Are all red algal parasites cut from the same cloth? *Acta Soc. Bot. Pol.*, 83(4):369–375, Dec. 2014.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

79. J. D. Silberman, M. L. Sogin, D. D. Leipe, and C. G. Clark. Human parasite finds taxonomic home. *Nature*, 380(6573):398, Apr. 1996.

80. D. B. Smida, N. Lundholm, W. H. C. F. Kooistra, I. Sahraoui, M. V. Ruggiero, Y. Kotaki, M. Ellegaard, C. Lambert, H. H. Mabrouk, and A. S. Hlaili. Morphology and molecular phylogeny of nitzschia bizertensis sp. nov.—a new domoic acid-producer. *Harmful Algae*, 32:49–63, Feb. 2014.

81. S. R. Smith, R. M. Abbriano, and M. Hildebrand. Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Research*, 1(1):2–16, May 2012.

82. A. Stamatakis. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9):1312–1313, May 2014.

83. A. Stamatakis, P. Hoover, and J. Rougemont. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.*, 57(5):758–771, Oct. 2008.

84. M. A. Steel, P. J. Lockhart, and D. Penny. Confidence in evolutionary trees from biological sequence data. *Nature*, 364(6436):440–442, July 1993.

85. M. J. Sullivan, N. K. Petty, and S. A. Beatson. Easyfig: A genome comparison visualizer. *Bioinformatics*, 27(7):1009, Apr. 2011.

86. A. Tartar and D. G. Boucias. The non-photosynthetic, pathogenic green alga *Helicosporidium* sp. has retained a modified, functional plastid genome. *FEMS Microbiol. Lett.*, 233(1):153–157, Apr. 2004.

87. J. Těšitel. Functional biology of parasitic plants: A review. *Plant Ecol. Evol.*, 149(1):5–20, Mar. 2016.

88. N. C. Tuchman, M. A. Schollett, S. T. Rier, and P. Geddes. Differential heterotrophic utilization of organic compounds by diatoms and bacteria under light and dark conditions. *Hydrobiologia*, 561(1):167–177, May 2006.

89. B. M. Tyler, S. Tripathy, X. Zhang, P. Dehal, R. H. Y. Jiang, A. Aerts, F. D. Arredondo, L. Baxter, D. Bensasson, J. L. Beynon, J. Chapman, C. M. B. Damasceno, A. E. Dorrance, D. Dou, A. W. Dickerman, I. L. Dubchak, M. Garbelotto, M. Gijzen, S. G. Gordon, F. Govers, N. J. Grunwald, W. Huang, K. L. Ivors, R. W. Jones, S. Kamoun, K. Krampis, K. H. Lamour, M.-K. Lee, W. H. McDonald, M. Medina, H. J. G. Meijer, E. K. Nordberg, D. J. Maclean, M. D. Ospina-Giraldo, P. F. Morris, V. Phuntumart, N. H. Putnam, S. Rash, J. K. C. Rose, Y. Sakihama, A. A. Salamov, A. Savidor, C. F. Scheuring, B. M. Smith, B. W. S. Sobral, A. Terry, T. A. Torto-Alalibo, J. Win, Z. Xu, H. Zhang, I. V. Grigoriev, D. S. Rokhsar, and J. L. Boore. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science*, 313(5791):1261–1266, Sept. 2006.

90. R. F. Waller, P. J. Keeling, R. G. Donald, B. Striepen, E. Handman, N. Lang-Unnasch, A. F. Cowman, G. S. Besra, D. S. Roos, and G. I. McFadden. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. U. S. A.*, 95(21):12352–12357, Oct. 1998.

91. R. F. Waller and G. I. McFadden. The apicoplast: A review of the derived plastid of apicomplexan parasites. *Curr. Issues Mol. Biol.*, 7(1):57–79, Jan. 2005.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

92. S. Wicke, K. F. Müller, C. W. dePamphilis, D. Quandt, S. Bellot, and G. M. Schneeweiss. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences*, 113(32):9045–9050, Aug. 2016.

93. K. H. Wolfe, C. W. Mordent, S. C. E, and J. D. Palmer. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: Loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *J. Mol. Evol.*, 35(4):304–317, Oct. 1992.

94. S. K. Wyman, R. K. Jansen, and J. L. Boore. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, 20(17):3252–3255, Nov. 2004.

95. D. Yan, Y. Wang, T. Murakami, Y. Shen, J. Gong, H. Jiang, D. R. Smith, J. F. Pombert, J. Dai, and Q. Wu. *Auxenochlorella protothecoides* and *Prototheca wickerhamii* plastid genome sequences give insight into the origins of non-photosynthetic algae. *Sci. Rep.*, 5(1):14465, Nov. 2015.

96. N. D. Young and C. W. DePamphilis. Rate variation in parasitic plants: Correlated and uncorrelated patterns among plastid genes of different function. *BMC Evol. Biol.*, 5(1):16, Feb. 2005.

97. M. Yu, M. P. Ashworth, N. H. Hajrah, M. A. Khiyami, M. J. Sabir, A. M. Alhebshi, A. L. Al-Malki, J. S. M. Sabir, E. C. Theriot, and R. K. Jansen. Evolution of the plastid genomes in diatoms. *Adv. Bot. Res.*, 2018.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information

Table S1. Primers used to amplify and sequence *cob* and *nad1* fragments for study taxa.

Primer name	Primer Sequence 5'– 3'
CobnF	GGAGTTTYGGYTCTTTAGCWGG
CobnR	GGHARAAAATACCACTCWGGVAC
Nad1nF[a]	ATGGGAGCAATYCAAAGRCG
Nad1nR[a]	CATTTCCAACCTAAATACATTAATTG
Nad1F-mod[b]	AAAGACGACGAGGWCCWAATGTKATAGGTT
Nad1R-mod[b]	CATTAATTGGTCATAYCGGTAWCGWGG

[a]Forward and reverse primers for initial amplification reaction.
[b]Forward and reverse primers for second amplification reaction when nested PCR was performed.

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2. Overlapping and adjacent genes in the plastid genomes of *Nitzschia* sp. nitz4 and *Nitzschia* sp. NIES-3581. Groups of overlapping or adjacent genes are separated by empty rows with the coordinates showing the degree of overlap. All but two groups are shared between the two genomes.

Nitzschia sp. nitz4				*Nitzschia* sp. NIES.3581			
Gene	Coordinates	over[a]	adj[b]	Gene	Coordinates	over	adj
rpl19	complement(3679..3993)			rpl19	complement(3681..3995)		
orf96	complement(3986..4276)	+		urf97	complement(3970..4260)	+	
atpB	complement(4267..5694)	+					
orf122	6041..6409						
rpoB	6406..9513	+					
rpoC1	9517..11580			rpoC1	9410..11479		
rpoC2	11577..14900	+		rpoC2	11476..14820	+	
rps2	14938..15627			rps2	14858..15547		
sufB	15624..17072	+		sufB	15544..16992	+	
sufC	17072..17815		+	sufC	16992..17735		+
atpG	18859..19311			atpG	18777..19229		
atpF	19298..19732	+		atpF	19216..19650	+	
atpA	20300..21805			atpA	20206..21720		
rpl35	21783..21977	+		rpl35	21714..21908	+	
rps12	complement(36097..36471)			rps12	complement(36024..36398)		
rpl31	complement(36471..36680)		+	rpl31	complement(36398..36604)		+
rps9	complement(36687..37091)			rps9	complement(36612..37016)		
rpl13	complement(37088..37537)	+		rpl13	complement(37013..37462)	+	
rpoA	complement(37534..38466)	+		rpoA	complement(37459..38391)	+	
rps11	complement(38487..38891)			rps11	complement(38412..38822)		
rps13	complement(38830..39273)	+		rps13	complement(38755..39198)	+	
secY	complement(39409..40455)			secY	complement(39334..40398)		
rps5	complement(40427..40930)	+		rps5	complement(40370..40873)	+	
rps17	complement(43317..43556)			rps17	complement(43288..43527)		
rpl29	complement(43546..43782)	+		rpl29	complement(43517..43756)	+	
				dnaK	48083..49930		
				tRNA-Arg	49930..50003		+
TOTAL		12	2			10	3

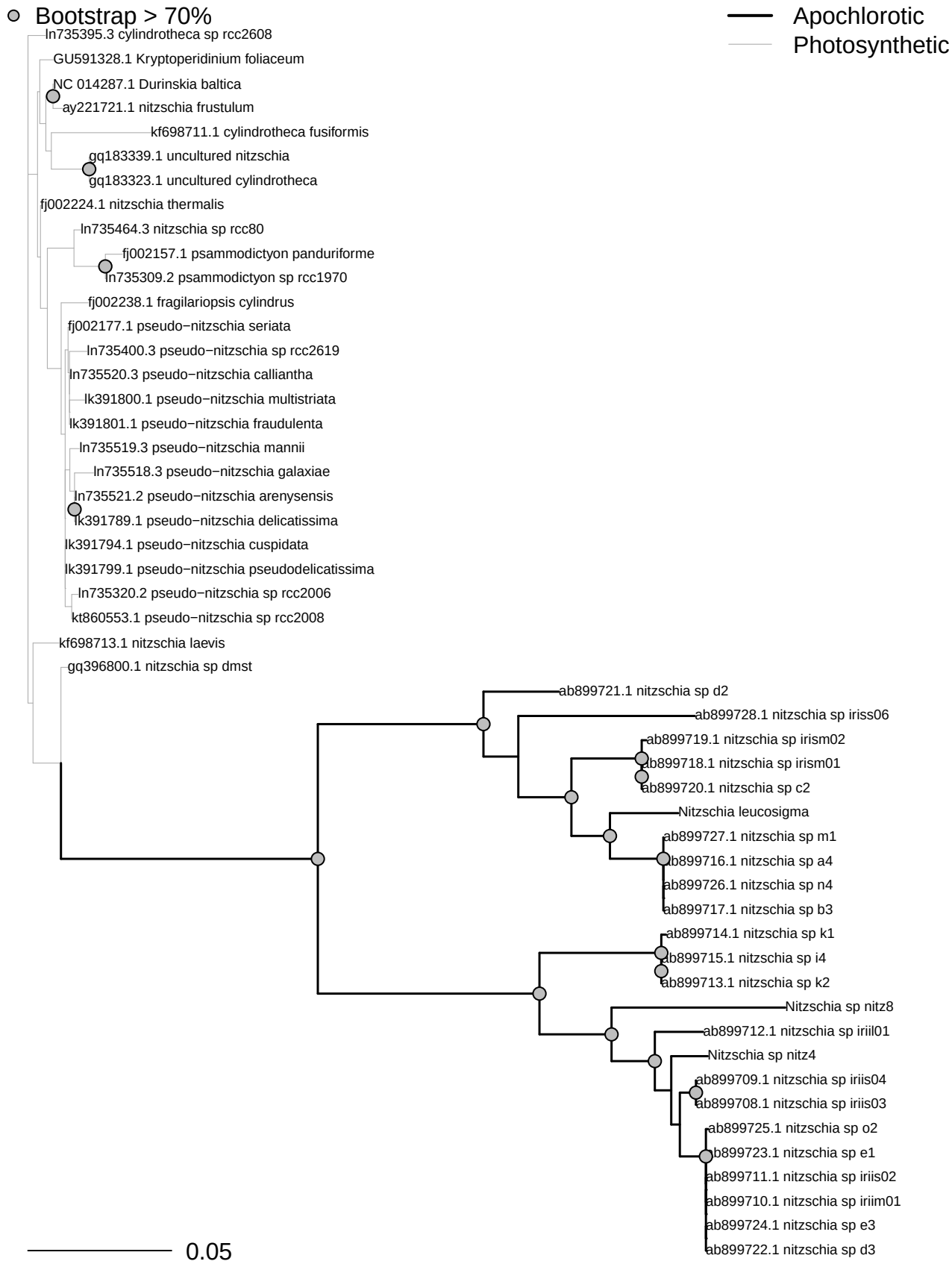
[a]Overlapping
[b]Adjacent

bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S1. Phylogenetic trees for plastid 16S (A), plastid 16S transformed into purine/pyrimidine (R/Y) coding (B), mitochondrial *cob* (C), *nad1* (D), a densely sampled 28S d1–d2 matrix with sequences from this study and GenBank (E), *cob* and *nad1* combined (F), and a large combined nuclear and mitochondrial gene matrix with sequences from this study and GenBank (G). In the 16S tree, *Nitzschia* sp. NIES-3581 is identified by its synonym, iriis04.

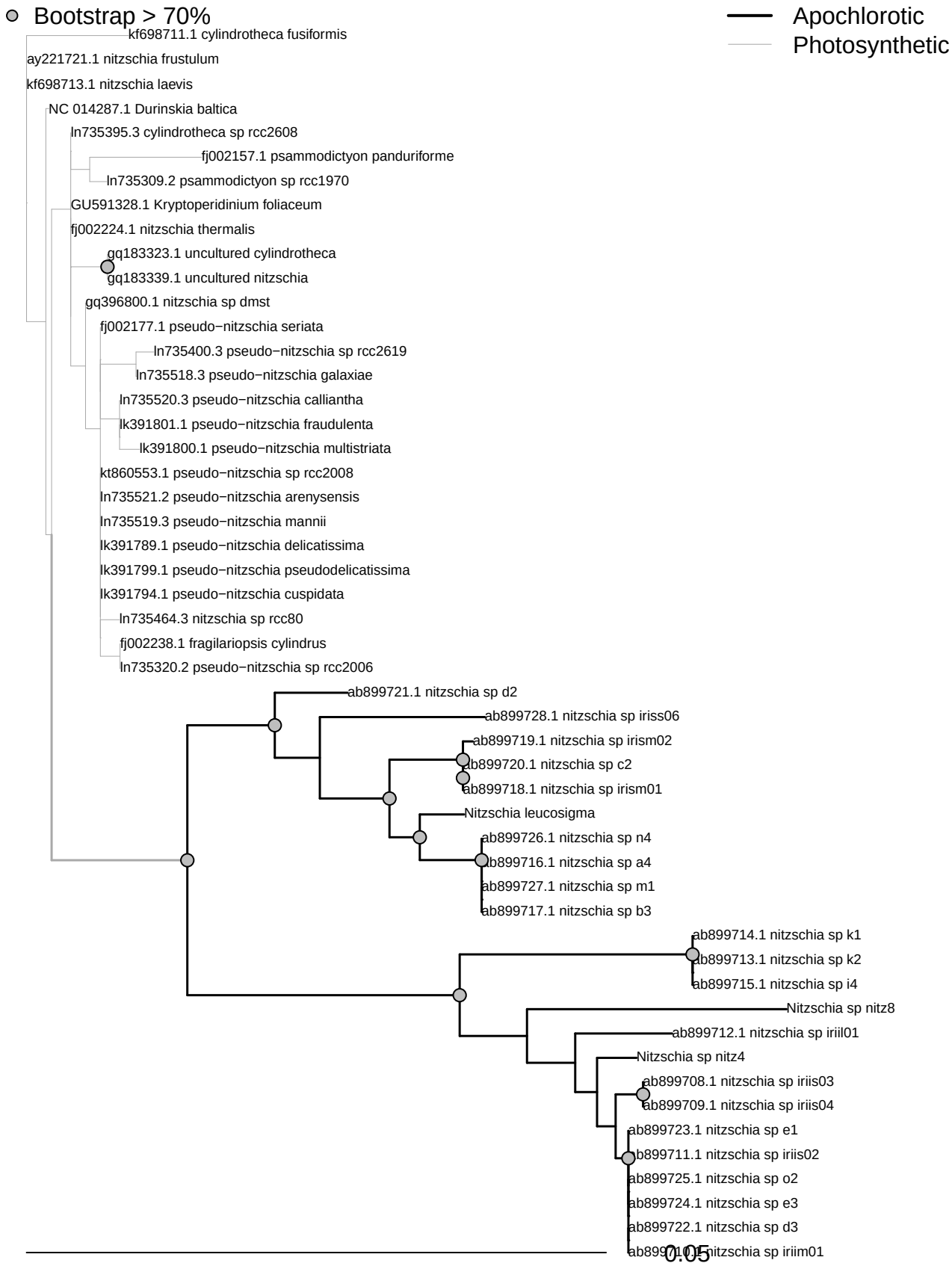
bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A. 16s



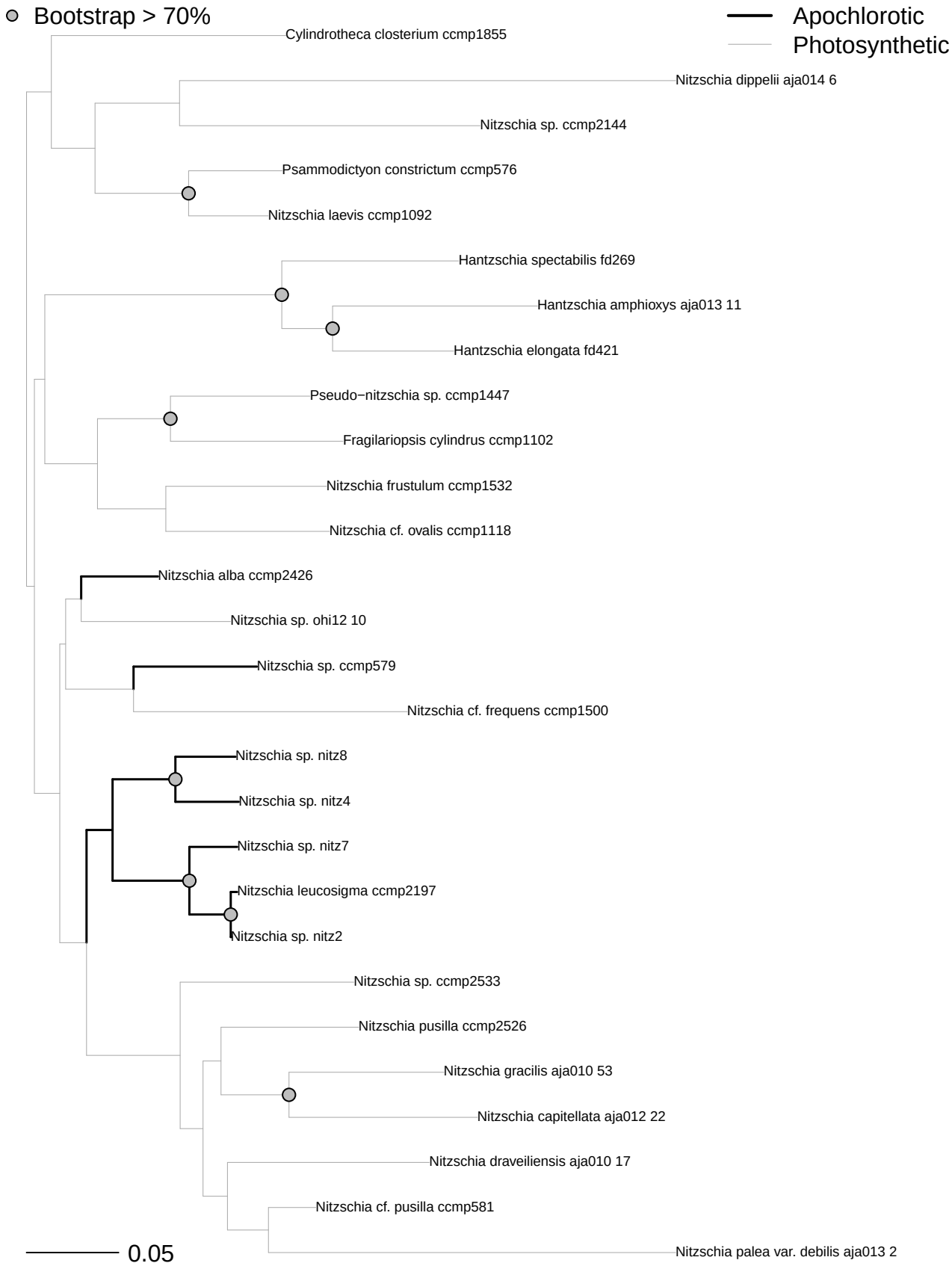
bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B: 16s R/Y coding



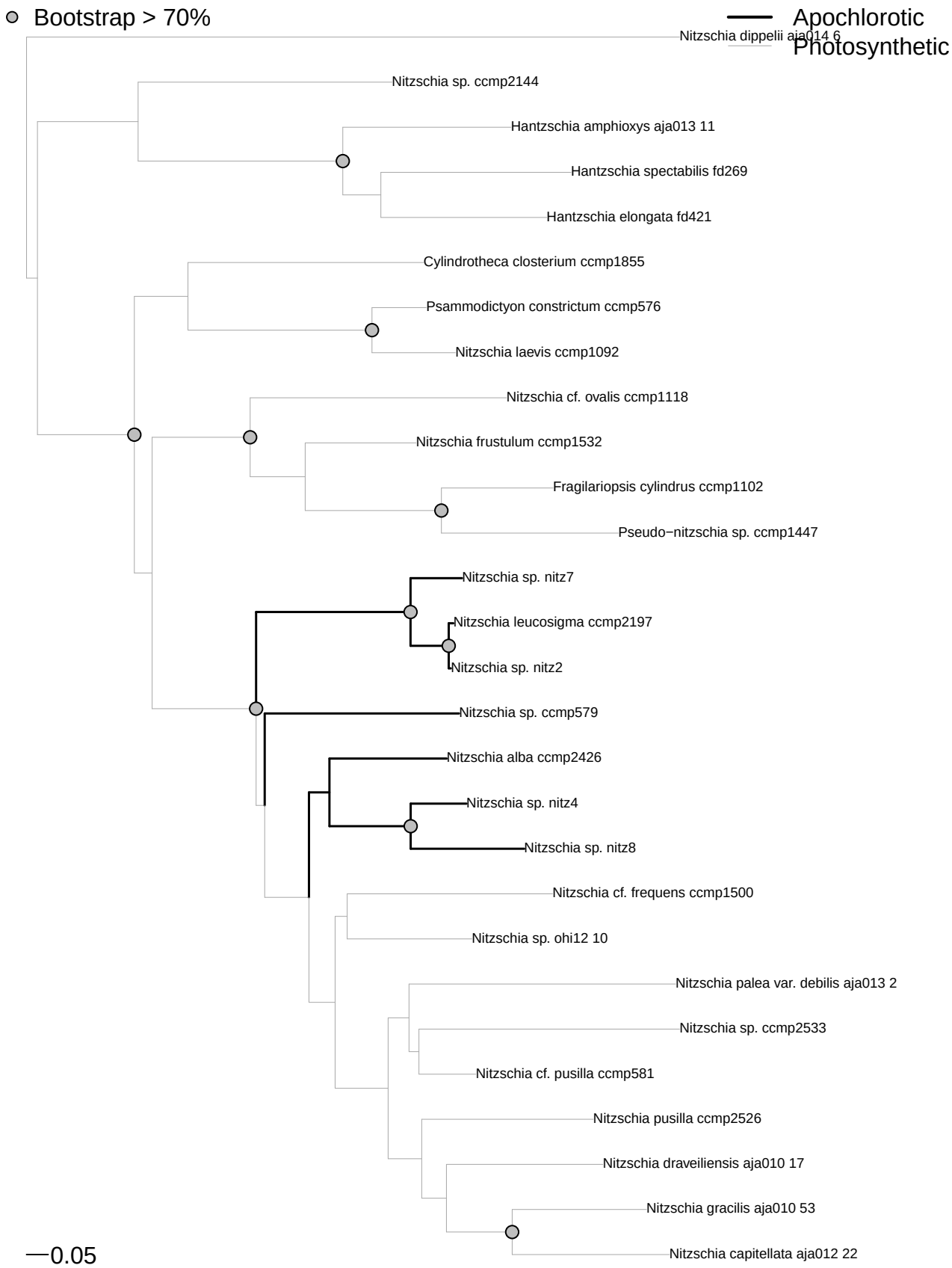
bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

C: cob



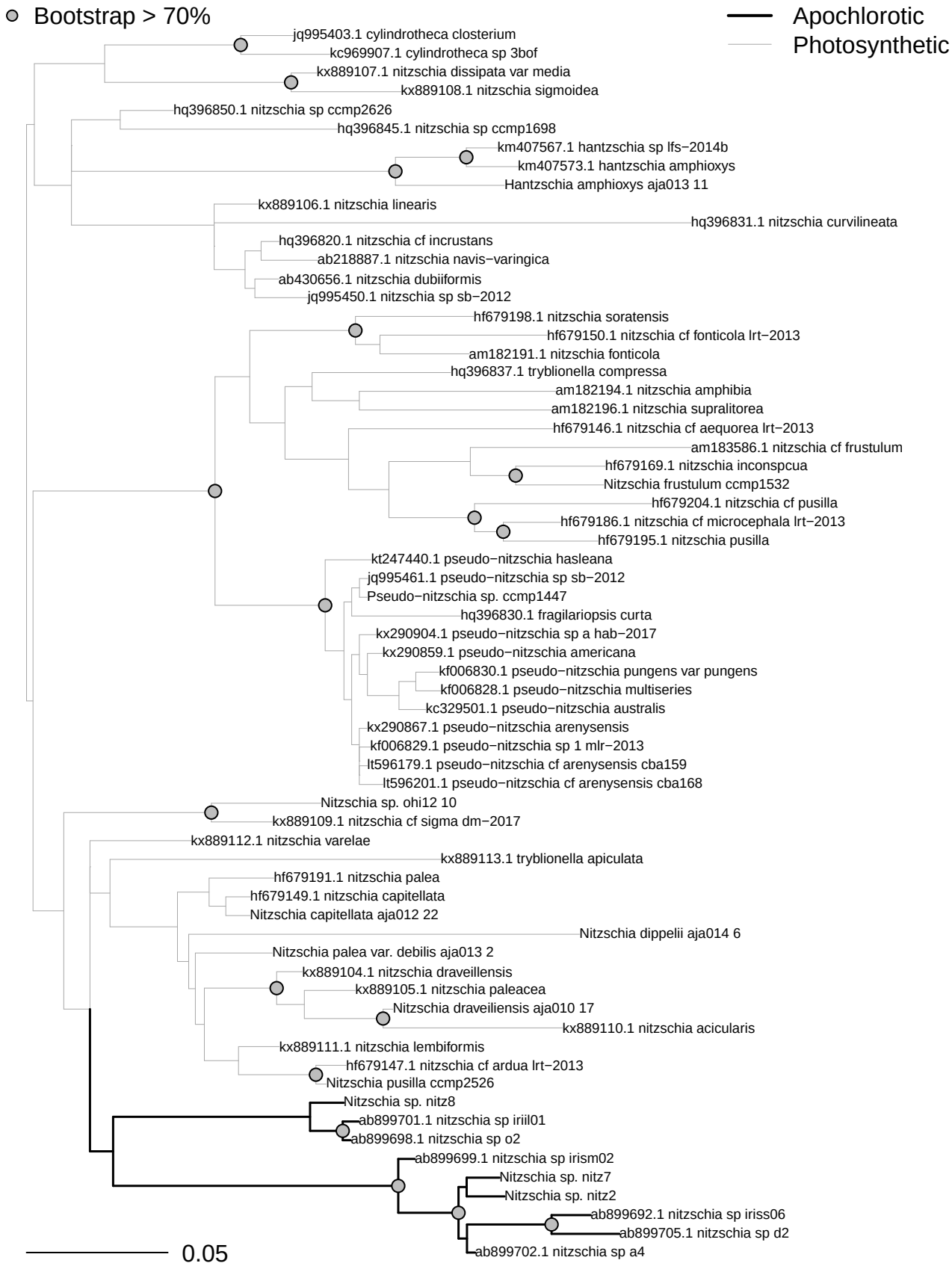
bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D: nad1



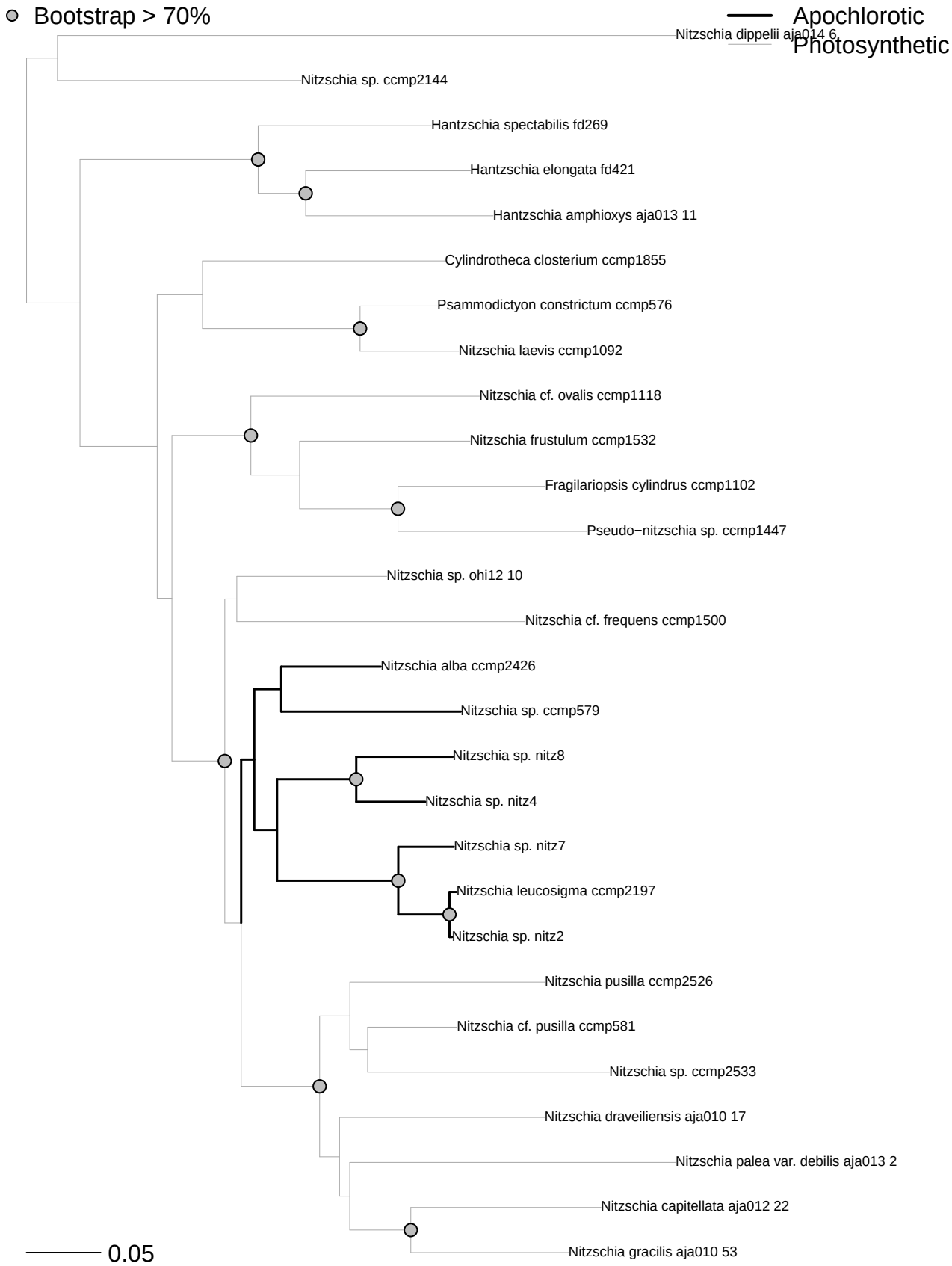
bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

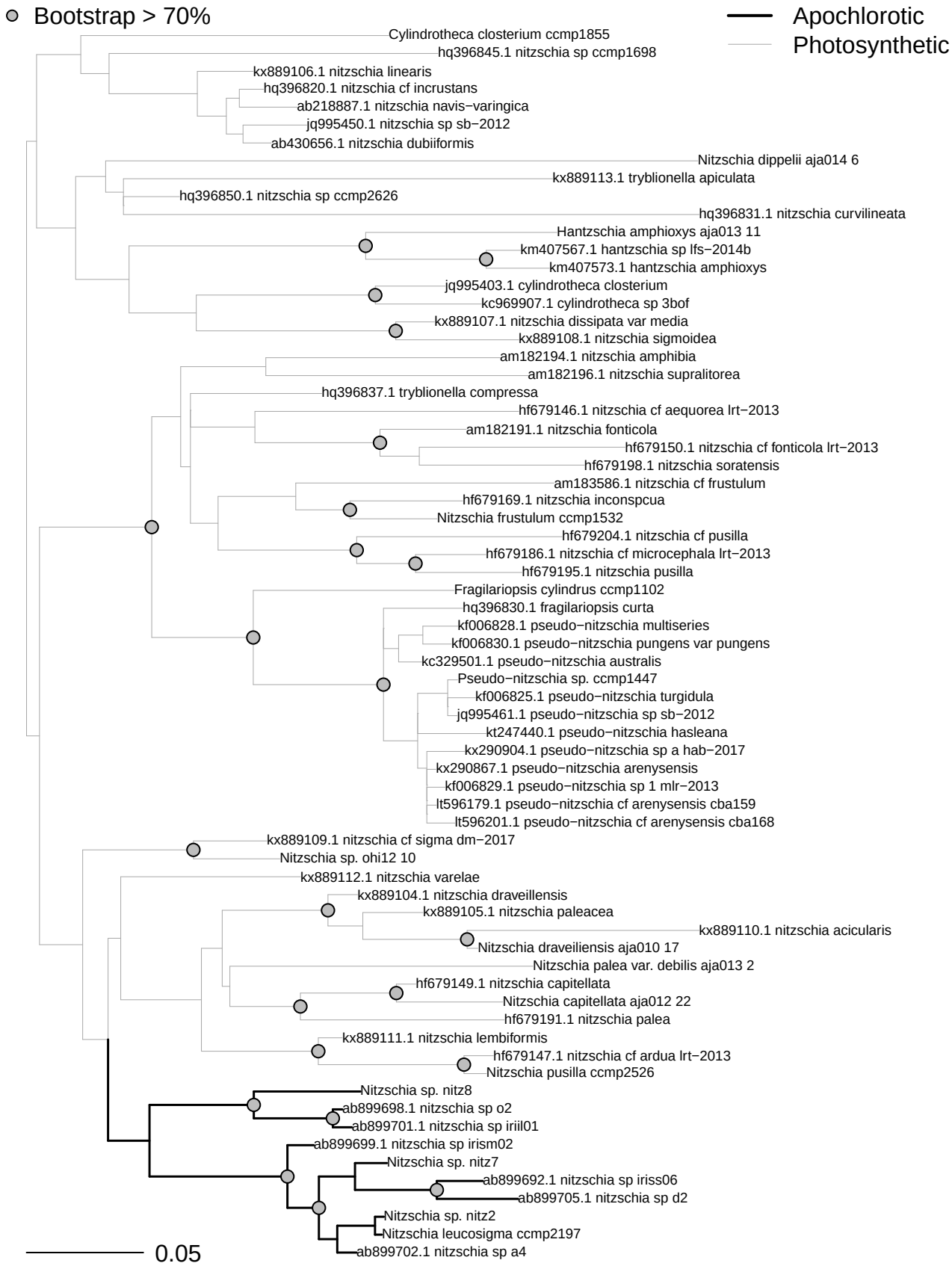
E. lsu (new + ncbi)



bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

F: mito (cob + nad1)



bioRxiv preprint doi: https://doi.org/10.1101/298810; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

G: mito (cob + nad1) + lsu (new + ncbi)