Molecular Phylogenetics and Evolution 196 (2024) 108086

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Investigation of heterotrophs reveals new insights in dinoflagellate evolution

Elizabeth C. Cooney [a,b,*], Corey C. Holt [a,b], Elisabeth Hehenberger [c], Jayd A. Adams [a], Brian S. Leander [a,d], Patrick J. Keeling [a]

[a] Department of Botany, University of British Columbia, 3156-6270 University Blvd., Vancouver, BC V6T 1Z4, Canada
[b] Hakai Institute, 1747 Hyacinthe Bay Rd., Heriot Bay, BC V0P 1H0, Canada
[c] Institute of Parasitology, Biology Centre Czech Academy of Sciences, České Budějovice, Czech Republic
[d] Department of Zoology, University of British Columbia, 4200 - 6270, University Blvd., Vancouver, BC V6T 1Z4, Canada

ARTICLE INFO

Keywords:
Dinoflagellates
Single cell transcriptomics
Phylogenomics
Character evolution

ABSTRACT

Dinoflagellates are diverse and ecologically important protists characterized by many morphological and molecular traits that set them apart from other eukaryotes. These features include, but are not limited to, massive genomes organized using bacterially-derived histone-like proteins (HLPs) and dinoflagellate viral nucleoproteins (DVNP) rather than histones, and a complex history of photobiology with many independent losses of photosynthesis, numerous cases of serial secondary and tertiary plastid gains, and the presence of horizontally acquired bacterial rhodopsins and type II RuBisCo. Elucidating how this all evolved depends on knowing the phylogenetic relationships between dinoflagellate lineages. Half of these species are heterotrophic, but existing molecular data is strongly biased toward the photosynthetic dinoflagellates due to their amenability to cultivation and prevalence in culture collections. These biases make it impossible to interpret the evolution of photosynthesis, but may also affect phylogenetic inferences that impact our understanding of character evolution. Here, we address this problem by isolating individual cells from the Salish Sea and using single cell, culture-free transcriptomics to expand molecular data for dinoflagellates to include 27 more heterotrophic taxa, resulting in a roughly balanced representation. Using these data, we performed a comprehensive search for proteins involved in chromatin packaging, plastid function, and photoactivity across all dinoflagellates. These searches reveal that 1) photosynthesis was lost at least 21 times, 2) two known types of HLP were horizontally acquired around the same time rather than sequentially as previously thought; 3) multiple rhodopsins are present across the dinoflagellates, acquired multiple times from different donors; 4) kleptoplastic species have nucleus-encoded genes for proteins targeted to their temporary plastids and they are derived from multiple lineages, and 5) warnowiids are the only heterotrophs that retain a whole photosystem, although some photosynthesis-related electron transport genes are widely retained in heterotrophs, likely as part of the iron-sulfur cluster pathway that persists in non-photosynthetic plastids.

1. Introduction

Dinoflagellates are an abundant and ecologically important group of protists that inhabit most aquatic environments (Taylor et al., 2008). This group is characterized by a huge range of diversity in morphology, nutritional strategies, and life histories, and possesses many unique traits that have expanded our understanding of eukaryotic biology (Gómez, 2020; Keeling and del Campo, 2017). However, reconstructing the evolution of these traits has been challenging because of limitations in the availability of genome-wide molecular data from dinoflagellates. Their nuclear genomes are notoriously problematic: within their structurally divergent nucleus, the dinoflagellate genome can be over 80 times the size of the human genome (Veldhuis et al., 1997), containing many highly conserved copies of some genes (Nand et al., 2021;

Abbreviations: FC, Snauq (False Creek); GA, Gulf of Alaska; HR, heliorhodopsin; JP, Jericho beach Pier; OP, Ogden Point; PR1, proteorhodopsin 1; PR2, proteorhodopsin 2; QI, Quadra Island; RSD, Ross Sea Dinoflagellate; SR, sensory rhodopsin.

* Corresponding author at: Department of Botany, University of British Columbia, 3156-6270 University Blvd., Vancouver, BC V6T 1Z4, Canada.
E-mail addresses: lizcooney22@gmail.com (E.C. Cooney), corey.holt@ubc.ca (C.C. Holt), elisabeth.hehenberger@paru.cas.cz (E. Hehenberger), jaydadams99@gmail.com (J.A. Adams), bleander@mail.ubc.ca (B.S. Leander), pkeeling@mail.ubc.ca (P.J. Keeling).

https://doi.org/10.1016/j.ympev.2024.108086
Received 10 December 2023; Received in revised form 12 April 2024; Accepted 24 April 2024
Available online 26 April 2024

Stephens et al., 2020). The size and repetitive nature of dinoflagellate genomes has rendered genome sequencing an inefficient tool for the study of this lineage as a whole, and transcriptomic sequencing has been favored to explore the genetic underpinnings of dinoflagellate biology, providing nearly all genomic diversity to date (Janouškovec et al., 2017; Keeling et al., 2014). However, even transcriptomes are heavily biased, as they are almost exclusively sequenced from cultured taxa, the vast majority of which are photosynthetic (Keeling et al., 2014). About half of known dinoflagellate species are heterotrophic (Gómez, 2012), but these lineages are generally more sparse in the environment and difficult to culture, resulting in a persistent lack of genomic data. Correcting this imbalance is necessary for a well-informed understanding of how features of dinoflagellate diversity evolved.

Insights provided by Janouškovec et al. (2017) on the appearance of inflated genome size and condensed chromosomes demonstrate the importance of robust phylogenomic analysis in determining the evolution of complex traits. These features correspond in phylogenies with the appearance of horizontally-acquired dinoflagellate viral nucleoproteins (DVNP) and bacterial histone-like proteins (HLP), which are together used instead of histones for chromatin packaging (Gornik et al., 2012; Wong et al., 2003). Interestingly, there are two distinct types of HLP that appear to have been acquired at different times, never coexisting in the same species, at least based on their distribution in this early phylogeny with limited taxa (Janouškovec et al., 2017).

Plastid diversity is another major area of interest in dinoflagellates, but has yet to be comprehensively explored due to the sparseness of genomic data. "Normal" dinoflagellate plastids are very unusual, with a bacterial type II ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCo), only three membranes, and a unique pigment, peridinin. But more unusually, the dinoflagellates contain the only undisputed cases of serial secondary plastid endosymbiosis (Matsumoto et al., 2011; Sarai et al., 2020) as well as plastids derived from endosymbiosis with another secondary plastid containing algae, e.g. tertiary plastids (Burki et al., 2014; Hehenberger et al., 2014; Keeling, 2013; Tengs et al., 2000). The group has also seen a high number of independent losses of photosynthesis giving rise to the many non-photosynthetic taxa (Saldarriaga et al., 2001). But exactly how many times this happened, how the resulting lineages adapted, and how the non-photosynthetic taxa relate to ones with tertiary or serial secondary plastids are often obscure, as key heterotrophic lineages are badly under-sampled. Some heterotrophs have been shown to retain ancestral plastid genes for heme, isoprenoid, and iron-sulfur cluster synthesis (Cooney et al., 2022, 2020; Hehenberger et al., 2014; Janouškovec et al., 2017), but recently one such lineage has been shown to also retain photosystem genes that may have been repurposed for a sensory role (Cooney et al., 2023): this possibility has not been comprehensively explored in other lineages, nor has outright plastid loss. Bacterial rhodopsin is another mechanism of photoactivity that has gained some attention in dinoflagellates after it was found to be highly expressed in the lineage, *Oxyrrhis marina* (Slamovits et al., 2011). It has since been hypothesized to function as a sensory mechanism in dinoflagellates, although its function in the group is still unknown and its distribution barely examined (Gavelis et al., 2017).

To provide a more balanced sampling of dinoflagellate diversity, we generated transcriptomes from 85 single cells targeting under-sampled, mostly heterotrophic taxa. These data significantly expand the representation of heterotrophs in the dinoflagellate tree, resulting in a different overall structure of the tree and allowing for a broader and more accurate view of molecular character evolution. The phylogeny reveals that previous sampling biases missed large blocks of diversity and misled interpretations of character evolution.

2. Materials and methods

2.1. Single cell collection and imaging

Cells were collected mostly from around the Salish Sea in British Columbia, Canada. Sample sites were at Snauq (also known as False Creek; FC) and Jericho Beach pier (JP) in Vancouver, the Hakai Ecological Observatory on Quadra Island (QI), and the Ogden Point sea wall (OP) in Victoria. Three cells were collected from the surface water in the Gulf of Alaska (GA) by Dr. Suzanne Strom during a 2018 Long Term Environmental Research cruise. Samples from the Salish Sea were collected from shore via net tow (20 μm mesh), with the exception of the Ogden Point sample, which was collected by scooping the top 2 cm of intertidal sand into a container along with ambient sea water. Dates and sources of each sample are listed in Table 1.

As soon as possible after collection, environmental samples were examined for unfamiliar or under-sampled dinoflagellates on a Leica DM IL microscope. For the OP sample, cells were extracted from the sand using the melting seawater ice method through a 100 μm mesh filter into a petri dish before examination (Uhlig, 1964). Cells of interest were isolated with a microcapillary pipet and washed with filtered sea water (0.2 μm) from the corresponding sample. Cells were imaged using a Sony A7r III before being placed into lysis buffer (Picelli et al., 2014). Videos of each cell can be found at https://doi.org/10.5281/zenodo.10553634. Samples were immediately frozen and stored at −70 °C until cDNA extraction could take place.

2.2. Sequencing and transcriptome assembly

For each sample, cDNA was extracted according to the Smart-seq2 protocol (Kolisko et al., 2014; Picelli et al., 2014). Nextera Flex or XT library preparations were performed by the Sequencing and Bioinformatics Consortium, University of British Columbia, and sequenced on Nextseq or Miseq platforms (Table 1, Table S1). Resulting forward and reverse raw reads were trimmed in Cutadapt (Martin, 2011) before assembly with rnaSPAdes v3.15.1 (Bankevich et al., 2012). Nucleotide assemblies were converted to amino acid sequences using TransDecoder v5.5.0 for open reading frame identification and annotation (Haas et al., 2013) and BLASTP (Poux et al., 2017) to search against the Uniprot database with an e-value threshold of $\leq 1e10^{-5}$ (assemblies available at https://doi.org/10.5061/dryad.x95x69prq). Final transcriptome quality was assessed using BUSCO v5.0.0 against alveolata_odb10 (Manni et al., 2021).

2.3. Identification and phylogenomic analysis

While some cells could be identified from their morphology, classification was mostly accomplished using small subunit ribosomal RNA gene sequences (SSU rDNA; Genbank OR921467 − OR921549). The most complete SSU sequences were extracted from the nucleotide assembly of each cell using a query sequence and BLASTN (e-value threshold of $\leq 1e10^{-25}$). Resulting hits were submitted as queries in megaBLAST searches against Genbank. Many cells could be classified to the genus or species level based on the SSU rDNA sequence similarity to known taxa, but the sequences of many other cells were not highly similar to any in the database. All SSU rDNA sequences compiled in this study were added to a curated collection of sequences from dinoflagellates and their close relatives. All sequences were aligned in MAFFT v.7.481 (Katoh and Standley, 2013) and trimmed with a 30 % gap threshold in trimAl v.3 (Capella-Gutiérrez et al., 2009). A tree was generated from the trimmed alignment with IQ-TREE v1.6.12, using ModelFinder to select the best fit model, GTR + F + I + G4 (Kalyaanamoorthy et al., 2017; Nguyen et al., 2015). The SSU rDNA sequences of *Oxyrrhis marina* and *Erythropsidinium* sp. were omitted from this analysis as they formed long branches with unstable and therefore uninformative placement in the tree.

Table 1
Sample information for single cell transcriptomes. Classifications are determined from maximum likelihood and small subunit ribosomal RNA gene sequence (SSU rDNA) phylogenies and percent identity with existing SSU rDNA gene sequences in Genbank. Cells that could not be classified to the genus level were given descriptive names inspired by cell appearance. Shorthand IDs are used to annotate and differentiate cells in phylogenies. Dates reflect the day cells were picked from environmental samples. Locations (Loc) reflect the areas that samples were collected. Photosynthetic (Phot) and benthic (Ben) cells are denoted with "Y" ("Y*" for photosynthesis in kleptoplastid cells).

Classification	ID	Date	Loc	Phot	Ben
Protoperidinium sp. 1	Ps1-FC1	07-Sep-20	False Creek	N	N
Protoperidinium sp. 1	Ps1-FC2	27-Sep-20	False Creek	N	N
Protoperidinium sp. 1	Ps1-QI3	03-Nov-21	Quadra Island	N	N
Protoperidinium sp. 1	Ps1-QI4	27-Jul-22	Quadra Island	N	N
Protoperidinium pellucidum	Ppe-QI2	27-Jul-22	Quadra Island	N	N
Protoperidinium conicum	Pco-FC1	22-Nov-20	False Creek	N	N
Protoperidinium conicum	Pco-QI2	04-Nov-21	Quadra Island	N	N
Protoperidinium sp. 2	Ps2-QI	07-Aug-21	Quadra Island	N	N
Protoperidinium depressum	Pde-GA	23-Apr-18	Gulf of Alaska	N	N
Protoceratium reticulatum	Pre-FC	10-Sep-20	False Creek	Y	N
Dinophysis sp. 1	Ds1-JP	04-Dec-20	Jericho Pier	N	N
Dinophysis sp. 2	Ds2-QI	14-Sep-21	Quadra Island	N	N
Dinophysis infundibulus	Din-FC1	30-Jun-21	False Creek	Y*	N
Dinophysis infundibulus	Din-FC2	30-Jun-21	False Creek	Y*	N
Dinophysis infundibulus	Din-FC3	30-Jun-21	False Creek	Y*	N
Phalacroma oxytoxoides	Pox-FC1	29-May-20	False Creek	N	N
Phalacroma oxytoxoides	Pox-FC2	29-May-20	False Creek	N	N
Phalacroma oxytoxoides	Pox-JP3	14-Jul-20	Jericho Pier	N	N
Phalacroma oxytoxoides	Pox-FC4	07-Sep-20	False Creek	N	N
Phalacroma oxytoxoides	Pox-FC5	10-Sep-20	False Creek	N	N
"Capsule cell"	Cce-QI	03-Nov-21	Quadra Island	N	N
Oxytoxum sp.	Osp-FC	01-Dec-20	False Creek	N	N
Cochlodinium pulchellum	Cpu-JP1	10-Dec-20	Jericho Pier	N	N
Cochlodinium pulchellum	Cpu-QI2	19-Oct-22	Quadra Island	N	N
Cochlodinium pulchellum	Cpu-QI3	19-Oct-22	Quadra Island	N	N
Cochlodinium pulchellum	Cpu-QI4	20-Oct-22	Quadra Island	N	N
"False warnowiid"	Fwa-FC1	02-Oct-20	False Creek	N	N
"False warnowiid"	Fwa-FC2	22-Nov-20	False Creek	N	N
"False warnowiid"	Fwa-FC3	22-Nov-20	False Creek	N	N
"False warnowiid"	Fwa-QI4	06-Aug-21	Quadra Island	N	N
"False warnowiid"	Fwa-QI5	14-Sep-21	Quadra Island	N	N
Akashiwo sanguinea	Aks-FC1	06-Oct-20	False Creek	Y	N
Akashiwo sanguinea	Aks-QI2	27-Jul-22	Quadra Island	Y	N
Balechina pachydermata	Bpa-FC1	26-Feb-21	False Creek	N	N
Balechina pachydermata	Bpa-QI2	15-Sep-21	Quadra Island	N	N
Balechina pachydermata	Bpa-QI3	16-Sep-21	Quadra Island	N	N
Balechina pachydermata	Bpa-QI4	16-Sep-21	Quadra Island	N	N
"Acorn cell"	Acc-JP1	12-May-20	Jericho Pier	N	N
"Acorn cell"	Acc-QI2	06-Aug-21	Quadra Island	N	N
Gymnodiniales sp. 1	Gsp-QI	20-Oct-22	Quadra Island	N	N
Gymnodiniales sp. 2	Gsp-OP	02-Jun-21	Ogden Point	N	Y
Gymnodiniales sp. 3	Gsp-FC	24-May-20	False Creek	Y	N
Togula jolla	Tjo-FC	03-Feb-21	False Creek	Y	Y
"Blob cell"	Bce-FC1	07-Nov-20	False Creek	N	N
"Blob cell"	Bce-FC2	07-Nov-20	False Creek	N	N
"Blob cell"	Bce-FC3	01-Dec-20	False Creek	N	N
"Blob cell"	Bce-JP4	04-Dec-20	Jericho Pier	N	N
"Blob cell"	Bce-JP5	04-Dec-20	Jericho Pier	N	N
"Blob cell"	Bce-JP6	04-Dec-20	Jericho Pier	N	N
Ankistrodinium semilunatum	Ans-OP1	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP2	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP3	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP4	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP5	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP6	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP7	02-Jun-21	Ogden Point	N	Y
Ankistrodinium semilunatum	Ans-OP8	02-Jun-21	Ogden Point	N	Y
Apicoporus sp. 1	As1-OP	02-Jun-21	Ogden Point	N	Y
Gyrodinium rubrum	Gru-JP1	10-Dec-20	Jericho Pier	N	N
Gyrodinium rubrum	Gru-JP2	10-Dec-20	Jericho Pier	N	N
Gyrodinium rubrum	Gru-FC3	07-Nov-20	False Creek	N	N
Gyrodinium rubrum	Gru-QI4	15-Sep-21	Quadra Island	N	N
Gyrodinium sp. 1	Gs1-QI	12-Sep-21	Quadra Island	N	N
Gyrodinium dominans	Gdo-FC1	02-Oct-20	False Creek	N	N
Gyrodinium dominans	Gdo-FC2	06-Oct-20	False Creek	N	N
Gyrodinium viridescens	Gvi-OP	08-Sep-21	Ogden Point	N	Y
Gyrodinium sp. 2	Gs2-FC1	09-Oct-20	False Creek	N	N
Gyrodinium sp. 2	Gs2-QI2	12-Sep-21	Quadra Island	N	N
Gyrodinium sp. 2	Gs2-QI3	14-Sep-21	Quadra Island	N	N
Gyrodinium sp. 2	Gs2-QI4	12-Sep-21	Quadra Island	N	N

(*continued on next page*)

Table 1 (*continued*)

Classification	ID	Date	Loc	Phot	Ben
Gyrodinium sp. 2	Gs2-QI5	19-Oct-22	Quadra Island	N	N
Gyrodinium spirale	Gsp-GA1	23-Apr-18	Gulf of Alaska	N	N
Gyrodinium spirale	Gsp-FC2	07-Nov-20	False Creek	N	N
Gyrodinium spirale	Gsp-FC3	09-Oct-20	False Creek	N	N
Gyrodinium spirale	Gsp-FC4	07-Nov-20	False Creek	N	N
Gyrodinium spirale	Gsp-FC5	09-Oct-20	False Creek	N	N
Gyrodinium spirale	Gsp-FC6	07-Nov-20	False Creek	N	N
Noctiluca scintillans	Nsc-QI	15-Sep-21	Quadra Island	N	N
Lebouridinium glaucum	Lgl-QI1	07-Aug-21	Quadra Island	N	N
Lebouridinium glaucum	Lgl-QI2	07-Aug-21	Quadra Island	N	N
Lebouridinium glaucum	Lgl-QI3	12-Sep-21	Quadra Island	N	N
Torodinium robustum	Tro-QI1	06-Aug-21	Quadra Island	Y	N
Torodinium robustum	Tro-QI2	13-Sep-21	Quadra Island	Y	N
Torodinium robustum	Tro-QI3	13-Sep-21	Quadra Island	Y	N
Torodinium robustum	Tro-QI4	19-Oct-22	Quadra Island	Y	N

Curated alignments of 263 conserved genes were used as queries in BLASTP searches to extract homologs from all single cell transcriptomes (Burki et al., 2008). Genes were searched using an e-value threshold of $1e^{-20}$, and the output was parsed for hits with $\geq$ 50 % query coverage. The sequences retrieved as BLAST hits were aligned with their corresponding queries using MAFFT-LINSI and trimmed with an 80 % gap threshold in trimAl. Trimmed alignments were then used to generate trees in IQ-TREE (model: LG + G). The resulting single gene trees were inspected manually for contaminants, paralogs, and isoforms, which were removed. Cleaned alignments contained at most one sequence from each transcriptome. To combine alignments for a multi-gene analysis, relevant taxa were first selected based on gene coverage using SCaFoS v4.55 (Roure et al., 2007). Only the best transcriptome for each species was included in the analysis, plus all other transcriptomes containing more than 50 % of the 263 genes. Genes present in < 60 % of these selected taxa were then eliminated. The final concatenated alignment composed of 199 genes (available at https://doi.org/10.5061/dryad.x95x69prq) was used to infer a maximum likelihood (ML) phylogeny in IQ-TREE with −bb 1000 (Hoang et al., 2018) and the empirical profile mixture model LG + C60 + F + G4 (Quang et al., 2008). To compare node support, a tree was constructed from the same alignment with LG + C60 + F + G4 PMSF and −b 100 (Wang et al., 2018). Fastest evolving sites were also removed from the alignment by 5 % increments up to 50 % removal, and phylogenies were generated at each step to assess changes in node support. Approximately unbiased (Shimodaira, 2002) and other topology tests were performed on the alignment to assess the integrity of major nodes in the spine of the tree (Table S2).

2.4. Protein search

To search for relevant protein transcripts, a database was curated that included all single cell transcriptomes sequenced in this study as well as dinoflagellate transcriptomes available from the Sequence Read Archive (Leinonen et al., 2011) and all taxa in the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; Keeling et al., 2014). Query files of relevant nuclear proteins, rhodopsins, and proteins involved in photosynthesis and plastid-associated synthetic pathways were compiled and used in BLASTP searches against the curated database using the same pipeline described above for the 263 gene search. For rhodopsin and chromatin packaging proteins, BLASTP searches were also performed against an annotated prokaryotic database of Genbank reference sequences. No prokaryotic hits were returned for DVNP, but recovered sequences from HLP and rhodopsins were concatenated with their corresponding eukaryotic sequence collections. The sequences collected for all relevant proteins were aligned in MAFFT-LINSI and trimmed in trimAl with a 70 % gap threshold (30 % for HLP and DVNP). Trees were constructed in FastTree for each protein (Price et al., 2010) and inspected for contaminant sequences, which were removed from original alignments. After decontamination followed by realignment and trimming, final trees were constructed in IQ-TREE using the model LG + F + I + G4 (trees and corresponding gene collections available at https://doi.org/10.5061/dryad.x95x69prq). For the chromerid lineage, which served as the outgroup in our phylogenomic analyses, genes of interest were reported based on BLASTP searches of transcriptomes and nuclear genomes, as well as annotated plastid genomes (Janouškovec et al., 2010).

For the genes involved in isoprenoid biosynthesis, transcripts from heterotrophic dinoflagellates were inspected for N-terminal extensions with characteristics of dinoflagellate plastid-targeting sequences as reported (Hehenberger et al., 2014; Patron et al., 2005). For Class I extensions, this typically consists of a signal sequence with a cleavage site demarcated by an “FVAP” motif, and a transit peptide enriched in serine and threonine followed by a hydrophobic transmembrane domain, with an arginine-rich region immediately upstream of the mature protein (Patron et al., 2005). Class II extensions similarly contain the signal sequence, “FVAP” motif, and transit peptide, but lack the transmembrane domain and arginine-rich region. While “FVAP” motifs were originally defined as four-letter motifs that 1) begin with phenylalanine, 2) contain valine, 3) lack acidic and basic residues, and 4) consist of no more than one polar uncharged residue (Patron et al., 2005), motifs that do not meet each of these criteria and sequences that seem to lack an “FVAP” motif altogether have been found in dinoflagellates (Hehenberger et al., 2014; Patron et al., 2005). For this reason, we assigned “FVAP” motifs based on whether they aligned closely with a predicted cleavage site and met at least three of the aforementioned criteria.

To identify these characteristics, each sequence was entered into SignalP v.3 (Bendtsen et al., 2004) and TMHMM v.2 to look for predicted signal peptides, cleavage sites, and hydrophobic transmembrane regions. We used version 3 of SignalP because signal peptides and the associated cleavage sites in dinoflagellates seem to be better recognized by prediction models employed by this version than the models implemented in more recent versions. In parallel, sequences were viewed in the respective alignments underlying the phylogenetic reconstructions, to differentiate the N-terminal extensions from mature proteins. As some dinoflagellate extensions with unambiguous plastid targeting characteristics can yield weak, ambiguous, or even misleading outputs from predictive software, we visually inspected sequences to make sure predicted features were consistent with alignment characteristics and residue content, and factored this into our interpretation of predictive outputs. For sequences that did not have a motif that met our “FVAP” criteria, we differentiated the signal peptide from the transit peptide based on predicted cleavage site. Some sequences were considered to have plastid targeting signals despite being truncated, because enough of the N-terminus was present to detect at least two of the contiguous regions expected to be adjacent to the mature protein.

To determine if photosynthetic or other plastid related genes in kleptoplastidic *Dinophysis acuminata* and *D. infundibulum* are encoded in the host nucleus, we searched for transcripts with dinoflagellate spliced leaders. During transcription, dinoflagellates transsplice their transcripts with highly conserved spliced leaders unique to this lineage (Zhang et al., 2007). For this reason, the presence of a dinoflagellate spliced leader on a transcript is compelling evidence that the

corresponding gene is encoded in the dinoflagellate genome, differentiating it from genes retained in a cryptic endosymbiont nucleus. Here, we isolated transcripts with dinoflagellate spliced leaders from assembled nucleotide transcriptomes by searching for the ten spliced leader residues closest to the adjoining transcript (TTGGCTCAAG), taking into account that spliced leaders may be N–terminally truncated due to the inherent incompleteness of (single-cell) transcriptomic data. Extracted transcripts were then translated to peptides and searched for relevant genes using BLASTP, as described above. All plastid-related sequences found were added to their corresponding alignments, which were trimmed and used to generate trees via FastTree so phylogenetic affinities could be confirmed.

3. Results and discussion

3.1. A more representative dinoflagellate phylogeny

Previously, the most comprehensive exploration of dinoflagellate character evolution was performed using a phylogeny of 101 orthologous genes and 27 core dinoflagellate taxa, only five of which were heterotrophic (Janouškovec et al., 2017). Here, we expanded the number of genes and species sampled, but most dramatically increased the proportion of heterotrophs, generating a phylogeny of 199 genes with 89 distinct core dinoflagellate taxa including 39 heterotrophs (44 % vs. the previous 18 %: Fig. 1, Table 1, Fig. S1). Using SSU rDNA sequence identity comparisons from Genbank and a comprehensive SSU phylogenetic analysis, collected cells were assigned to 22 different genera (Table 1; Fig. S2) and in total, 34 distinct species-level taxa were categorized.

As expected, transcriptome assemblies varied in quality between cells (Table S1) and based on coverage, only 52 of the 85 cells sequenced were included in the phylogenomic analysis. Transcriptomes for MGD, TGD, *Abedinium*, Noctilucales, Amphidiniales, and Gymnodiniales were also included (Cooney et al., 2023, 2022, 2020; Sarai et al., 2020). All tree-building models yielded the same topology with varying node support, even with progressive fastest-evolving site removal (Fig. 1). This topology had the highest statistical support in AU and related topology tests, although three other topologies could not be eliminated (Table S2).

In addition to increasing the representation of heterotrophs over six-fold, our samples combined with transcriptomes sequenced in recent studies also greatly expand the representation of athecate taxa. The so-called armored lineages of dinoflagellates, Peridiniales, Symbiodiniales, Gonyaulacales, and *Dinophysis* and relatives, form a derived, monophyletic clade (Janouškovec et al., 2017). Most newly added taxa (all except *Protoperidinium*, *Dinophysis* spp., *Phalacroma*, “Capsule cell”, and *Oxytoxum*) are athecate and fill in areas of the tree with previously little representation (Fig. 1).

Our phylogeny places MGD as the earliest-diverging core dinoflagellate, followed by *Torodinium*, *Lebouridinium*, TGD, and then *Abedinium*, all forming independent branches. Our analysis is also the first to include diverse members of the *Gyrodinium* genus, showing that *Gyrodinium* may be sister to the Noctilucales and Amphidiniales (Fig. 1). We also reveal that two non-photosynthetic lineages, *Ankistrodinium semilunatum/Apicoporus* sp. and the unidentified “Blob cell”, are sister to the photosynthetic lineages *Karenia* and *Karlodinium* (which contain haptophyte-derived tertiary plastids), and a Gymnodiniales clade, respectively. Our analysis confirms previous SSU rDNA sequence-based analyses showing the order Gymnodiniales is non-monophyletic, with Gymnodiniales *sensu stricto* represented in our study by the warnowiids, *Gymnodinium catenatum*, *Lepidodinium chlorophorum*, the polykrikoids, and *Paragymnodinium shiwhaense* (Fig. 1; Gómez, 2020). Several taxa collected in the present study cluster with moderate support immediately sister to the thecate dinoflagellate clade. This group consists of four heterotrophic lineages and the genus *Akashiwo*.

3.2. Early acquisition of two histone-like proteins

The dinoflagellate nucleus is structurally unlike the nuclei of other eukaryotes. Its condensed, fibrillar form is the result of non-nucleosomal chromatin packaging facilitated by the histone alternatives DVNP and HLP (Gornik et al., 2012; Wong et al., 2003). The horizontal acquisition of DVNP in the common ancestor of dinoflagellates corresponds with the loss of nucleosomal chromatin packaging and likely helped facilitate the depletion of histones (Gornik et al., 2012; Irwin et al., 2018). A major phylogenomic character mapping study revealed that DVNP is ubiquitous among dinoflagellates (although a more recent study found it may be absent in MALV-I and psammosids; Holt et al., 2023), and also concluded that genes for HLP were gained twice from different bacterial donors: one acquired early in the common ancestor of core dinoflagellates, and a second type acquired later, replacing the early version (Chan and Wong, 2007; Janouškovec et al., 2017). In no case were both HLP types found coexisting in the same species (Janouškovec et al., 2017).

Our expanded phylogeny confirms the presence of DVNP in all dinoflagellates, but reveals a new HLP acquisition timeline (Fig. 2a). Most notably, we find that the HLP type found in thecate taxa (referred to here as *Crypthecodinium*-type, or HLPc), which was proposed to be the type acquired later, is in fact present in the earliest-branching member of the core dinoflagellates, MGD. The second type (*Noctiluca*-type, or HLPn) was absent from MGD, indicating that it may actually have been acquired after HLPc. Despite this, HLPn dominates in most athecate taxa. Moreover, we find both HLP types in several taxa (Fig. 2a), which also contrasts with original findings that they never coexist in the same species, and reveals there was not a simple replacement of one with the other. This raises the possibility that both genes may also be present in some of the single-HLP taxa we sampled, but that expression levels of one type have fallen short of the sampling threshold. It is also possible that one dominates over the other in these instances through expression levels, perhaps with variation in which type is expressed in different contexts. That a single HLP type was found in most taxa does suggests that the coexistence of both may more often be transient, perhaps due to functional redundancy. The overwhelming absence of HLPn transcripts across thecate taxa suggests this type was indeed lost in an early ancestor of the Peridiniales, Prorocentrales, and Symbiodiniales. True losses of one HLP type in the common ancestors of *Ankistrodinium/Apicoporus/Karenia/Karlodinium* and *Gyrodinium* are similarly probable. Ultimately, the absence of either HLP type in any lineage cannot be definitively proven without further investigations beyond transcriptomics.

Out of all core dinoflagellates, there are five lineages from which no HLP transcripts were recovered: *Abedinium*, *Lebouridinium glaucum*, *Apicoporus* sp., *Gymnodinium catenatum*, and the unclassified species, “False warnowiid”. Undersampling is the likely explanation for HLP absence in *Apicoporus* sp., but *Abedinium* was represented by three high-coverage transcriptomes from different species (Cooney et al., 2022, 2020), and *L. glaucum* and “False warnowiid” were represented by three and five single cell transcriptomes of the same species, respectively. *G. catenatum* was represented by a single MMETSP transcriptome with otherwise deep coverage. The lack of HLP transcripts in these taxa is striking considering that all other lineages, including ones with comparatively low coverage, yielded at least one HLP type and the gene seems to be expressed at high levels. Overall, the evidence that HLPs have been lost in some taxa is mounting, but further investigations are essential for determining this.

3.3. Punctate distribution of diverse rhodopsins

Microbial rhodopsins are light activated retinal-binding proteins that have proliferated via horizontal gene transfer (HGT) into all domains of life (Spudich et al., 2000). In bacteria, ion pumping rhodopsins have been implicated in growth, survival during starvation, and carbon fixation (Gómez-Consarnau et al., 2010, 2007; Palovaara et al., 2014).

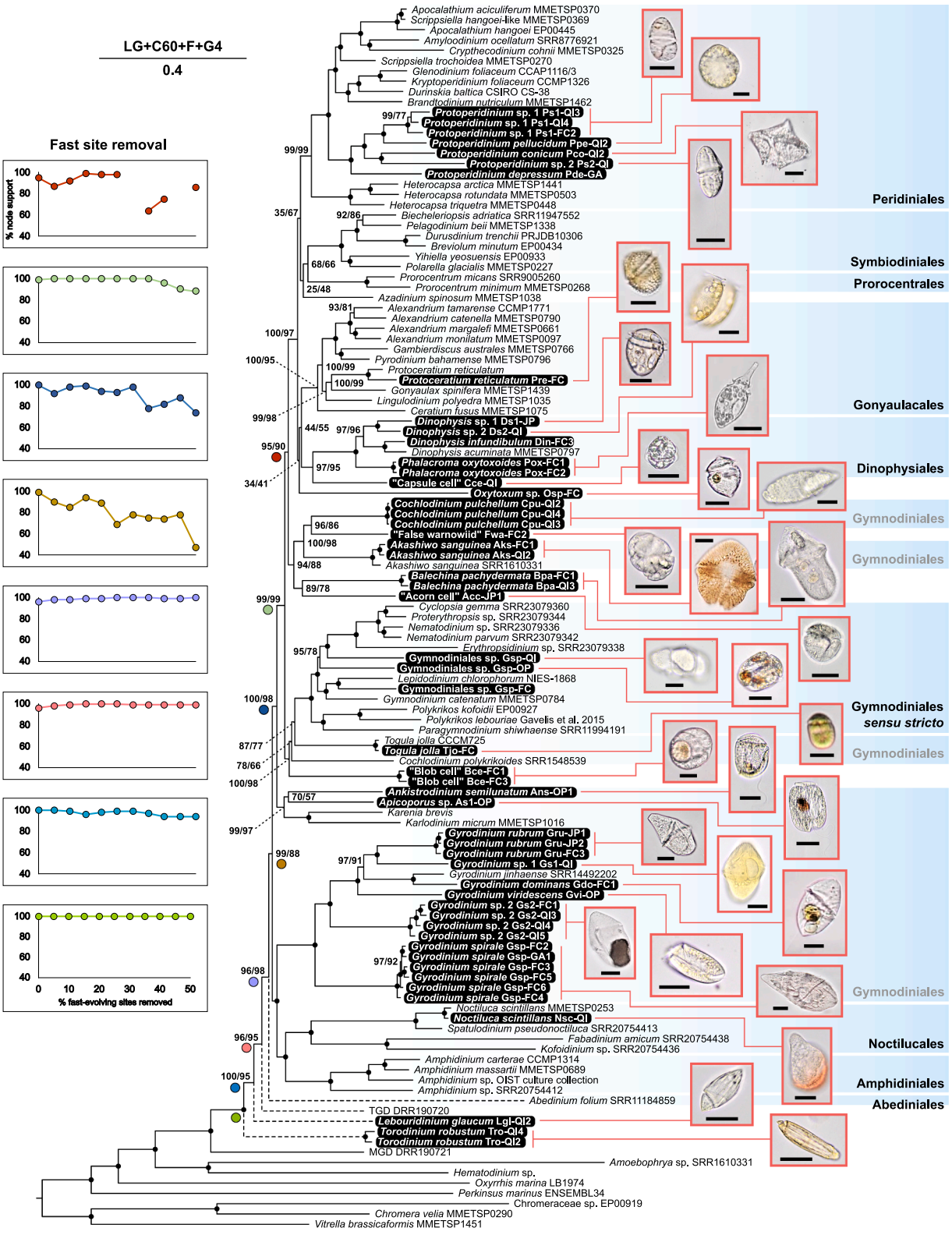

Fig. 1. Maximum likelihood phylogeny of dinoflagellates and close relatives inferred from 199 orthologous genes. Taxa sampled in this study are highlighted with black. Images of select taxa are shown; scale bars in all images are 25 μm. Numbers at nodes are bootstrap values ([LG + C60 + F + G4 −bb 1000]/[LG + C60 + F + G4 PMSF −b 100]) with black dots representing 100 for both analyses. Dotted branches are associated with alternate topologies that could not be ruled out in an approximately unbiased test (Table S2). The scale bar provides reference for the estimated number of amino acid substitutions per site and shows the primary model for the tree used in this figure. Graphs to the left of the tree show percentage node support (y-axis) with incremental fast-evolving site removal (x-axis). Marker color in each graph matches the colored circle on the corresponding tree node. Blue highlighted clades show higher taxonomic designations appearing in past major phylogenetic and phylogenomic analyses (Cooney et al., 2023, 2022, 2020; Gómez, 2020; Janouškovec et al., 2017).

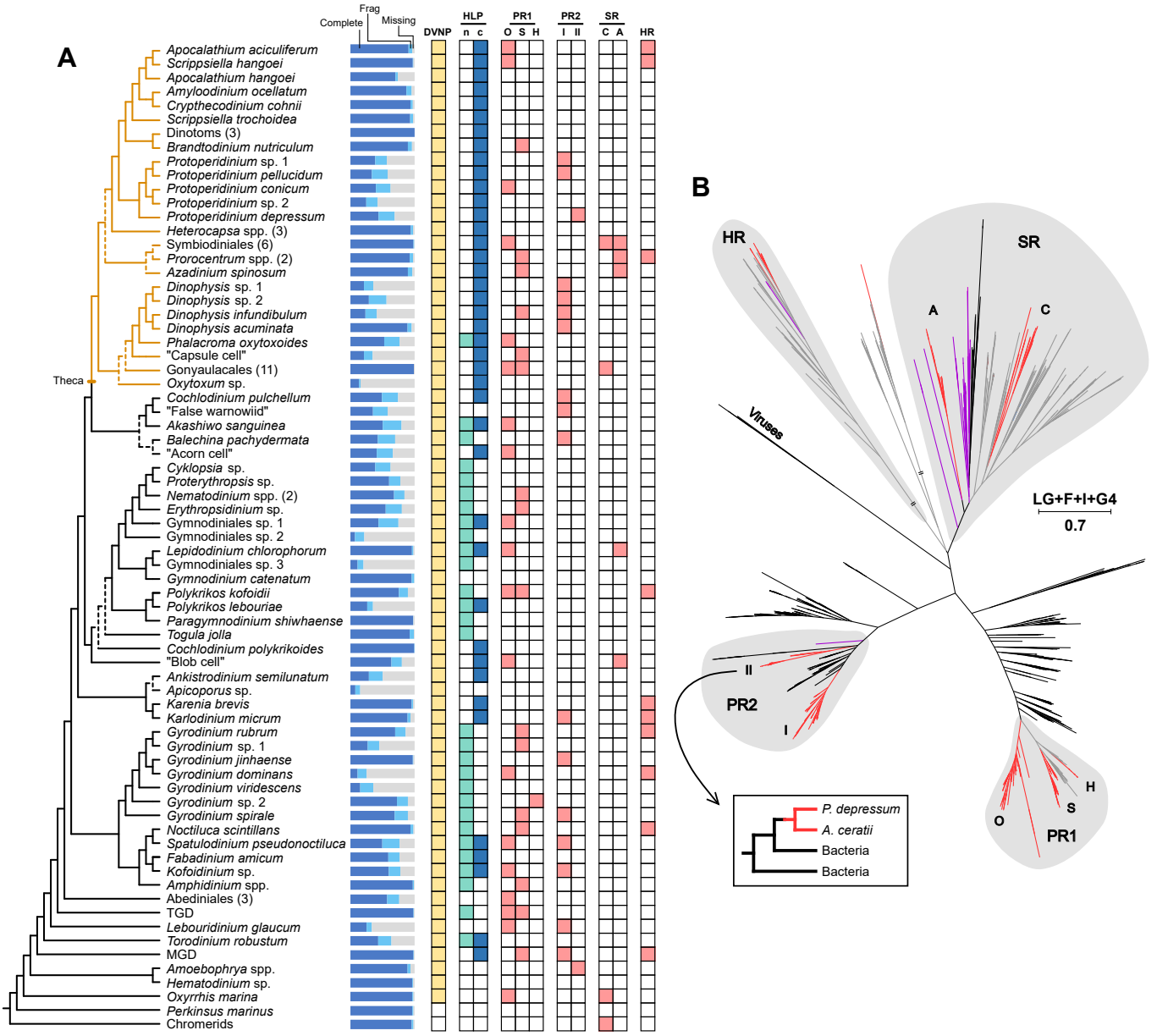

Fig. 2. Presence of chromatin packaging proteins and rhodopsin phylogeny. A) The topology of the tree at the left margin reflects that of the multi-gene phylogeny in Fig. 1. Dotted lines denote nodes with < 95 % bootstrap support and the orange clade corresponds to thecate dinoflagellates. Numbers in parentheses after some branch labels indicate the number of distinct taxa that were included in the search for gene transcripts. Bars to the right of each taxon show the BUSCO completeness scores of the combined transcriptomes that represent each branch (Table S1). Gene presence is indicated by a color-filled box. Frag = fragmented. DVNP = dinoflagellate viral nucleoprotein. HLP = histone-like protein; n = *Noctiluca*-type; c = *Crypthecodinium*-type. PR1 = proteorhodopsin 1; O = *Oxyrrhis* clade; S = stramenopile clade; H = haptophyte clade. PR2 = proteorhodopsin 2; I = clade 1; II = clade 2. SR = sensory rhodopsin; C = cryptophyte clade; A = archaea clade. HR = heliorhodopsin. B) Unrooted maximum likelihood phylogeny of rhodopsins. Bacteria and viruses (labeled) are black, archaea are purple, dinoflagellates are red, and all other eukaryotes are grey. Light grey highlighted sections and the labeling within them correspond to the four major rhodopsin groupings in part A and their subgroups. The branches containing a break have been shortened to half their original length. Scale bar depicts the estimated number of amino acid substitutions per site and shows the model used to generate this tree. Inset shows a simplified depiction of phylogenetic relationships within the PR2 II rhodopsin clade.

Evidence of a more facultative role has been described in diatoms, which seem to use them as an alternate ATP synthesis pathway when iron is limited (Marchetti et al., 2015, 2012). *Oxyrrhis marina*, the first eukaryote in which rhodopsins were described, possesses at least two distinct functional types (Slamovits et al., 2011): one of these is proteorhodopsin (PR), which is a proteobacterial proton pump that is prolific in the marine environment (Béjà et al., 2001; Gómez-Consarnau et al., 2017), and interestingly is the most highly expressed nuclear gene in *Oxyrrhis*. The function of PR is still unknown in dinoflagellates, although it may supplement nutritional requirements under photosynthesis-limiting light conditions (Shi et al., 2015) or when prey is scarce (Guo et al., 2014). *Oxyrrhis* also possesses sensory rhodopsins closely related to those of cryptophytes, which may facilitate phototaxis (Slamovits et al., 2011). Reports of rhodopsins in phylogenetically distant dinoflagellate taxa suggest that they are widespread and functionally important (Sharma et al., 2006; Slamovits et al., 2011; Xiang et al., 2015).

Searching the new data set shows that rhodopsins are indeed even more common, widespread, and diverse in dinoflagellates than previously thought. In total, eight distinct rhodopsins were recovered

(Fig. 2a). These were categorized into four broad groups based on phylogenetic affinities: proteorhodopsin 1 (*PR1*), proteorhodopsin 2 (*PR2*), heliorhodopsin (*HR*), and sensory rhodopsin (*SR*) (Fig. 2b). Dinoflagellates have gained *PR1* three separate times, from stramenopiles (*PR1s*), haptophytes (*PR1h*), and proteobacteria, as described in *Oxyrrhis marina* (*PR1o*; Slamovits et al., 2011). Similarly, *PR2* was gained twice from different proteobacteria (*PR2i* and *PR2ii*). Sensory rhodopsins appear to have been acquired both from cryptophytes (*SRc*) and archaea (*SRa*).

The distribution of most rhodopsin subtypes across the dinoflagellates are patchy, indicative of a complex evolutionary history. In several cases, rhodopsins of the same subtype appear in multiple distantly related subgroups scattered over the tree, rather than clustering in lineages that share a close common ancestor (Fig. 2a). There are multiple scenarios that might explain this distribution. On one hand, it is possible rhodopsin genes are present but were not expressed at a sufficiently high frequency to be detected in many lineages in the present study. On the other hand, rhodopsin genes acquired by an early dinoflagellate ancestor may have been lost independently many times. But the rhodopsin distribution may also be the result of HGT between divergent dinoflagellate taxa, either directly or via other eukaryote or viral vectors (Irwin et al., 2022). Generally, HGT of bacterial rhodopsin

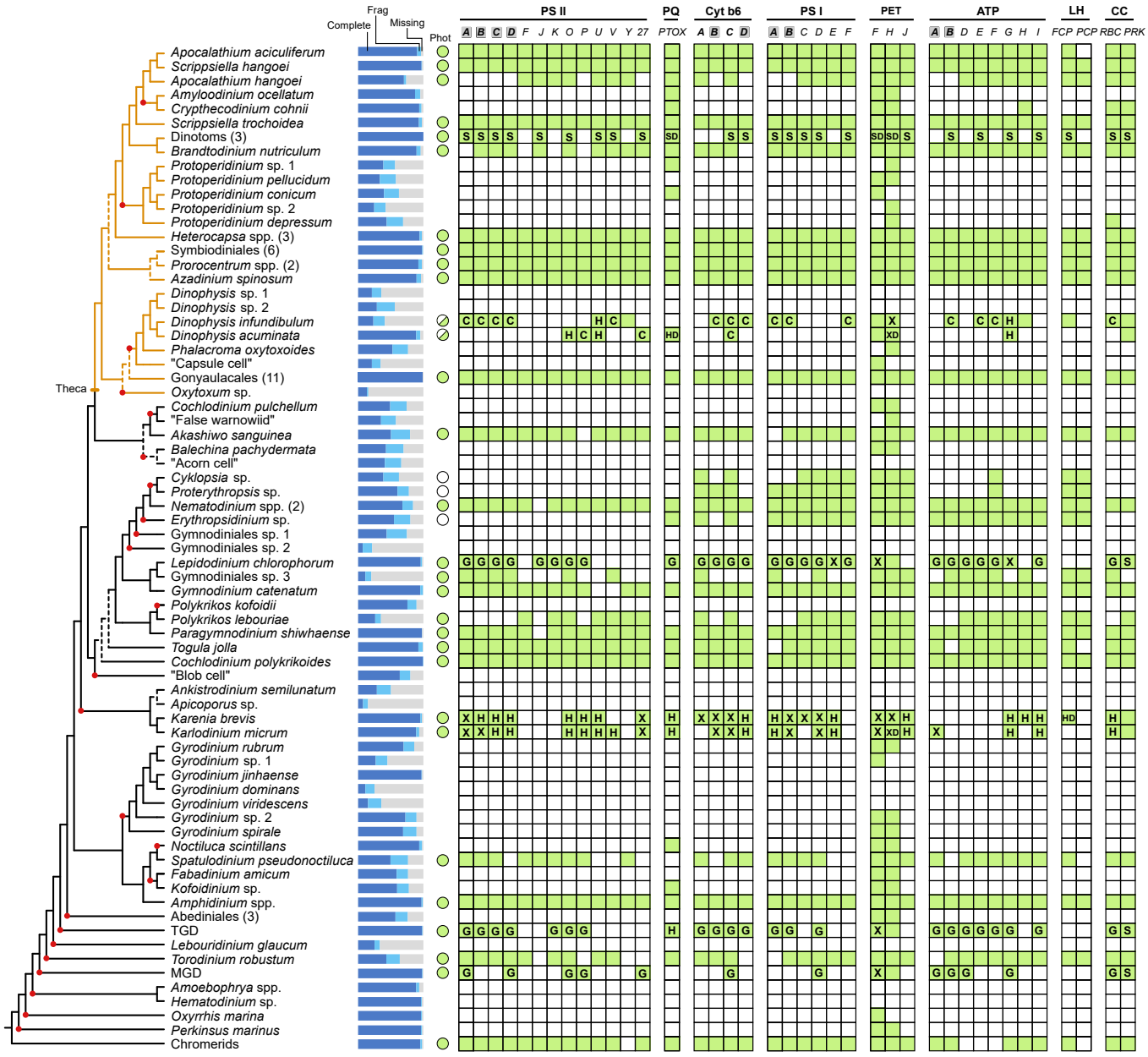

Fig. 3. Presence of photosynthesis-related gene transcripts. See Fig. 2A caption for a description of the tree and bar graph. Red dots indicate losses of photosynthesis. Letters indicate the eukaryotic lineage(s) that recovered sequences originate from. S = stramenopiles, H = haptophytes, G = green algae, C = cryptophytes, D = dinoflagellates, and X = ambiguous origins. Green filled circles denote taxa that are photosynthetic (Phot) while white filled circles denote taxa that possess photosystem genes but are not canonically photosynthetic. Half-filled circles represent kleptoplasty. PS II = photosystem II; corresponding letters are shorthand for gene names beginning with "*psb*". PTOX = plastoquinol terminal oxidase. Cyt b6 = cytochrome b6; letters are shorthand for "*pet*" genes. PS I = photosystem I; letters are shorthand for "*psa*" genes. PET = plastid electron transport; letters are shorthand for "*pet*" genes. ATP = ATP synthase; letters are shorthand for "*atp*" genes. LH = light harvesting; FCP = fucoxanthin-chlorophyll binding protein; PCP = peridinin-chlorophyll binding protein. CC = Calvin cycle; RBC = ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCo); PRK = phosphoribulokinase.

is common (de la Torre et al., 2003; Frigaard et al., 2006), and other horizontally acquired eukaryotic genes have shown similar patchy distributions, possibly explained by eukaryote-to-eukaryote transfer (Keeling and Inagaki, 2004; Rogers et al., 2007).

The complex distribution of different rhodopsins may also be the result of a combination of scenarios occurring in parallel. The patchy distribution of *PR1s* and *PR1o* is consistent with early acquisition of both genes followed by many independent losses of one type or the other due to redundancy. This pattern is reminiscent of the HLPs, although rhodopsins are presumably less vital, permitting loss of both types in some taxa. In contrast, the more restricted presence of *PR2b* in *Amoebophrya* and *Protoperidinium depressum* could be due to HGT between these divergent taxa (Fig. 2b, inset). As *Amoebophrya* parasitizes core dinoflagellates, it is not hard to imagine HGT being facilitated by infection in this case. The evolutionary histories of other rhodopsin subtypes are more ambiguous, some showing small-scale synapomorphic distributions in distant lineages (Fig. 2a). These may also be cases of inter-dinoflagellate HGT followed by proliferation into descendent taxa. However, ancestral horizontal acquisition followed by abundant independent losses in these rhodopsin types cannot be ruled out.

3.4. Photosynthesis gene retention in heterotrophs

Dinoflagellate plastid evolution is characterized by multiple independent losses of photosynthesis and plastid reduction (Janouškovec et al., 2017; Saldarriaga et al., 2001). The discovery of plastidial photosynthesis-related genes in ostensibly heterotrophic warnowiid dinoflagellates has raised the question of whether other heterotrophic taxa have also retained parts of the photosynthetic machinery (Cooney et al., 2023). In dinoflagellates, most plastid genes have moved to the nucleus (Bachvaroff et al., 2004; Hackett et al., 2004; Howe et al., 2008b). Many of these genes encode proteins for non-photosynthetic pathways, and not surprisingly many heterotrophic taxa have been shown to retain these plastid-derived genes even after the plastid genome and photosynthesis have been lost. The proteins are targeted back to the non-photosynthetic organelle to carry out functions such as isoprenoid, heme, and iron-sulfur cluster biosynthesis (Hehenberger et al., 2019, 2014). These pathways have been relatively well-studied, but since photosynthesis is absent in these lineages, a comprehensive exploration of genes pertaining to the photosynthetic mechanism has never been conducted in the few heterotrophs for which data were available.

Our search reveals that at least 21 independent losses of photosynthesis have occurred across the core dinoflagellates (Fig. 3). In general, massive retention of photosynthesis genes in non-photosynthetic lineages, like the whole photosystem complexes retained in warnowiids, is not common to other heterotrophs, but elements of the photosynthetic pathways are retained in some heterotrophic lineages, sometimes recurrently. Transcripts for the electron transport gene ferredoxin (*petF* or *Fd*) were recovered in *Amyloodinium ocellatum*, *Crypthecodinium cohnii*, *Protoperidinium* spp., "Capsule cell", *Cochlodinium pulchellum*, *Balechina pachydermata*, *Gyrodinium* spp., Noctilucales, Abediniales, and *Oxyrrhis*, as well as the non-dinoflagellate oyster parasite, *Perkinsus marinus*. Ferredoxin-NADP$^{\pm}$ reductase (*petH* or *FNR*) is similarly prevalent, appearing in these taxa (except "Capsule cell" and *Oxyrrhis*) as well as *Phalacroma oxytoxoides* and "False warnowiid". Although these proteins are best known for their role in producing NADPH via electron transport from photosystem I, they have been observed in several divergent non-photosynthetic lineages (Donaher et al., 2009; Dorrell et al., 2019; Füssy et al., 2020; Hehenberger et al., 2016; Kim et al., 2020). In a close sister group to the dinoflagellates, the apicomplexans, *petF* and *petH* help facilitate the production of iron-sulfur clusters in the non-photosynthetic apicoplast (Röhrich et al., 2005; Seeber and Soldati-Favre, 2010). As the plastidial pathway for iron-sulfur cluster synthesis has been retained in many non-photosynthetic dinoflagellates as well, this likely explains the prevalence of these genes in both groups (Cooney et al., 2020; Hehenberger et al., 2014; Janouškovec et al., 2017; Mathur et al., 2019). Interestingly, the CO_2 fixation protein, RuBisCo was found in *Protoperidinium depressum*. The presence of this gene is unexpected in heterotrophs, as it is not known to serve a role outside of photosynthesis. However, RuBisCo has also been observed before in the closely related heterotroph, *Crypthecodinium cohnii* (Sanchez-Puerta et al., 2007), and has been retained in non-dinoflagellate organisms where photosynthesis was lost (Sekiguchi et al., 2002; Wolfe and DePamphilis, 1998). Here, we show that *C. cohnii* is also the only heterotroph that retains phosphoribulokinase (PRK), a protein thought to exclusively function in the Calvin cycle.

Unlike typical heterotrophic dinoflagellates, the kleptoplastidic species *Dinophysis acuminata* and its sister, *D. infundibulum*, possess several photosynthesis and other plastid-related genes from a mix of haptophyte, cryptophyte, and peridinin plastid origins. In an effort to identify whether any of these genes are encoded in the dinoflagellate nucleus, we sought to identify dinoflagellate spliced leaders on the transcripts for these genes, since many dinoflagellate genes have this sequence at the 5′ end, while none of the photosynthetic lineages from which kleptoplasts are derived do. Initial searches for global spliced leader transcripts returned fewer than 15 sequences per transcriptome in *D. infundibulum*, none of which were plastid-related. This is likely due to N-terminal degradation of transcripts, which poses a particular challenge in single cell sequencing due to the limited amount of sequenceable starting material. In contrast, over 6000 spliced leader sequences were recovered in *D. acuminata*, which was sequenced from culture (MMETSP; Keeling et al., 2014). From these, five plastid-related genes (*psbU*, *PTOX, petC*, *petH*, *atpG*) originating from non-ancestral plastids were identified, indicating that they were gained through HGT and are now encoded in the host nucleus (Fig. 3; supplementary spreadsheet: https://doi.org/10.5061/dryad.x95x69prq). Furthermore, *psbU*, *PTOX*, and *atpG* share a phylogenetic affinity with haptophytes (as do *psbU* and *atpG* in *D. infundibulum*), and are therefore also very unlikely to be derived from remnants of a cryptophyte nucleus. This is consistent with previous findings of non-ancestral plastid-related genes with spliced leader transcripts (including *psbU*) in *D. acuminata*, (Wisecaver and Hackett, 2010).

Transcripts for other haptophyte and cryptophyte genes have been observed in *D. acuminata* before (Hehenberger et al., 2019). The latter are derived from the cryptophyte kleptoplast that *D. acuminata* attains from its ciliate prey (Kim et al., 2012), while the origin of haptophyte transcripts is unknown. As in the Ross Sea Dinoflagellate (RSD), the presence of genes from mixed origins is consistent with a "shopping bag" scenario (Howe et al., 2008a; Keeling, 2024, 2013; Larkum et al., 2007), where proteins targeted to temporary plastids accumulate over time from symbionts, food, and other sources, facilitating the ability of the host to retain and use kleptoplasts from its prey (Hehenberger et al., 2019; Karnkowska et al., 2023). The present study includes several lineages that have permanently adopted tertiary or serial secondary plastids (Hehenberger et al., 2014; Sarai et al., 2020; Tengs et al., 2000). The photosynthesis-related genes in these taxa originate predominantly from their new plastids, however every one of these lineages possesses at least one gene of stramenopile or haptophyte origin, and some retain ancestral dinoflagellate genes as well (Fig. 3).

Notably, there is only one lineage of ostensibly non-photosynthetic dinoflagellates that carries ancestral photosystem genes. In addition to cytochrome b6/f and electron transport, the heterotrophic warnowiids, *Proterythropsis*, *Cyklopsia*, and *Erythropsidinium* retain genes for plastidial ATP synthase, light harvesting proteins, and photosystem I (Cooney et al., 2023). This unique clade, which includes one photosynthetic genus, *Nematodinium* (*Warnowia* sp. from Cooney et al., 2023 was recently reclassified as *Nematodinium parvum* by García-Portela et al., 2023), is characterized by the ocelloid, a complex camera-type eye (Greuet, 1977, 1965). The retina analog of the ocelloid is a highly modified plastid that has been shown to contain remnants of a plastid genome (Gavelis et al., 2015; Greuet, 1977; Hayakawa et al., 2015). While *Nematodinium* has a typical assemblage of photosynthetic genes,

transcriptomes of heterotrophic taxa do not contain transcripts associated with photosystem II, RuBisCo, or the Calvin cycle, suggesting they have adapted the photosynthetic machinery for a sensory function (Cooney et al., 2023). Our data demonstrate that the warnowiids are an outlier in this respect, as we found no other cases in which ancestral photosystem genes were retained in heterotrophic dinoflagellates (Fig. 3).

Some photosynthetic species with high transcriptomic coverage yielded comparatively low coverage of photosynthesis-related gene transcripts. MGD had a sparser assemblage of photosynthetic genes, possibly because they are derived from its green algal plastid and the photosynthetic mechanism may have reduced to contain only the most critical components for photosynthetic function. Interestingly, many of the missing transcripts are also absent from the other two green alga-containing lineages, *Lepidodinium* and TGD (*psbF*, *psbU*, *psbV*, *psbY*, *petJ*, *atpH*). If these genes are in fact missing, their absence is consistent with loss due to reduction.

3.5. Other plastid-derived genes are present in heterotrophs

Aside from photosynthesis, four other plastid pathways have been comparatively well-studied due to their presence in non-photosynthetic plastids of apicomplexans: these are biosynthesis of heme, fatty-acids, isoprenoids, and iron-sulfur clusters. In a few heterotrophic dinoflagellates, plastid-derived genes with N-terminal plastid-targeting sequences have been observed, and based on these it is generally thought that most non-photosynthetic dinoflagellates retain reduced plastid organelles that continue to perform these functions (Cooney et al., 2022, 2020; Hehenberger et al., 2019, 2014; Janouškovec et al., 2017). However, this has really only ever been investigated in three lineages containing heterotrophs – *Crypthecodinium*, Noctilucales, and Abediniales (Cooney et al., 2022, 2020; Sanchez-Puerta et al., 2007), and complete plastid loss is known in one dinoflagellate lineage, the Syndiniales (Gornik et al., 2015; Holt et al., 2023). Our expanded phylogeny shows there have been many independent losses of photosynthesis across the core dinoflagellates, so we performed searches for genes of all plastid-derived biosynthetic pathways to see if they are in fact retained across heterotrophic dinoflagellates as a rule.

The starkest difference between photosynthetic and heterotrophic lineages was in the presence of type II fatty acid synthesis pathway (FASII) gene transcripts. Based on the apparent absence of this pathway in the few taxa explored previously, it is assumed to be commonly lost in heterotrophs, and fatty acids are instead thought to be produced in these lineages by a synthase related to FASI (Hehenberger et al., 2016; Janouškovec et al., 2015; Van Dolah et al., 2013). Due to the inherent incompleteness of transcriptomic data, it has thus far been impossible to confirm whether this is the case. Here, we find that most heterotrophs show no sign of FASII genes, but there are some exceptions. *Protoperidinium conicum*, *P. depressum*, and Gymnodiniales sp. 1 possess multiple FASII genes, with *P. depressum* expressing the whole suite. Transcripts for only *fabI* were recovered in *Erythropsidinium* sp., which is thought to have lost photosynthesis despite retaining photosystem I (Cooney et al., 2023). Unfortunately, the N-termini of all these heterotrophic sequences are truncated, making it impossible to determine if they have plastid targeting extensions.

Gene transcripts from isoprenoid, heme, and iron-sulfur cluster synthesis pathways were generally present across heterotrophic lineages, although low transcriptome completeness in some species precluded any conclusion. The total absence of some pathways from low-coverage lineages (*Protoperidinium* sp. 2, *Dinophysis* sp. 1, *D. infundibulum*, Capsule cell, *Oxytoxum* sp., Gymnodiniales sp. 2, Gymnodiniales sp. 3, *Apicoporus* sp., *G. dominans*, *Lebouridinium glaucum*) was impossible to interpret, since under-sampling is as likely an explanation as absence from the genome. Overall, however, the sprinkling of transcripts from these pathways across all lineages suggest they are generally present in heterotrophic dinoflagellates.

Some species possessed multiple different homologs of certain proteins. This was the case for three components of the heme synthesis pathway (*hemE*, *hemH*, *hemY*) and acetyl-CoA carboxylase (*ACC*) from the FASII pathway (Fig. S3A). Phylogenetic analysis revealed three separate homologous clades of *hemE*, all containing at least some sequences with putative plastid targeting extensions (Fig. S3B). One homolog is ancestral and of eukaryotic origin, shared with apicomplexans and perkinsids (c in Fig. S3), while two group with plastidial lineages, suggesting they originate from plastid acquisition (a and b in Fig. S3). In contrast, two homologs are present for both *hemY* and *hemH*, with plastidial and cytosolic forms of *hemY*, and eukaryotic and bacterial forms in *hemH*. Together, these genes support the hypothesis that heme biosynthesis has a long evolutionary history of redundancy, as it has been observed in early-branching dinoflagellate lineages (Holt et al., 2023) and more distant lineages like the archaeplastids (Gawryluk et al., 2019). For *ACC*, all sequences with plastid targeting extensions clustered into one clade, and therefore all sequences in this clade are likely of plastidial origin while those in other clades are presumed to be cytosolic (Fig. S3A).

The presence of plastid targeting sequences on isoprenoid biosynthesis genes in heterotrophs was not always clear due to frequent N-terminal truncation. However, several lineages previously unexamined for evidence of a plastid organelle (*Amyloodinium ocellatum*, *Phalacroma oxytoxoides*, *Polykrikos kofoidii*, "Blob cell", *Gyrodinium* spp., *Fabadinium amicum*) had at least one sequence possessing an N-terminal extension with dinoflagellate plastid targeting characteristics (Fig. S4; supplementary spreadsheet: https://doi.org/10.5061/dryad.x95x69prq). Other lineages had partial N-terminal extensions, some of which were too truncated to identify. As reported in previous explorations in dinoflagellates, most transcripts were preceded by Class I transit peptides, while *ispF* genes possessed the Class II type (Hehenberger et al., 2014; Patron et al., 2005). An interesting exception were two sequences of *ispF* from two different *Erythropsidinium* transcriptomes that had Class II transit peptides. Since little is known about how targeting extensions come to associate with their respective transcripts or how these associations vary and change, it is unclear why this lineage would be different from the rest. While more exploration is required, our data demonstrate the variability in length and composition among dinoflagellate plastid targeting sequences and further support that all heterotrophic core dinoflagellates retain a reduced plastid organelle.

4. Conclusions

By expanding the available transcriptome data for dinoflagellates with an emphasis on heterotrophic, athecate taxa, we were able to explore the molecular character evolution of this group with more resolution and with a much more representative phylogeny. While the phylogeny resulting from our dataset is weakly supported in some areas, many new relationships have been resolved and general patterns have emerged that alter existing narratives about the unique traits of this group. Specifically, we have shown that the two types of HLP found in dinoflagellates were horizontally acquired early in the core dinoflagellates, rather than sequentially (Fig. 2A). One or the other HLP may have been subsequently lost in many lineages, likely due to redundancy, resulting in the close to mutually exclusive distribution we see today. Our exploration of rhodopsins reveals an even more convoluted history of horizontal acquisition and loss (Fig. 2A,B). While these proteins are abundant and diverse in dinoflagellates, their erratic distribution suggests that this prevalence has been impacted by HGT and their likelihood of short-term selective advantage. The distribution of photosynthesis-related genes reveals that electron transport and other components are present in some taxa that have lost photosynthesis, but that warnowiids are the only lineage that retain a photosystem (Fig. 3). Finally, our search for plastid-derived biosynthesis pathway genes reveals that most heterotrophs have lost FASII (with some clear exceptions), and generally retain isoprenoid, heme, and iron-sulfur cluster

pathways (Fig. 4). Biases still remain, as freshwater dinoflagellates have yet to be included in robust phylogenomic analyses and parasitic lineages are also greatly underrepresented, due to the technical challenges of finding and sequencing them. But taken together, these observations clarify and add complexity to the already complex evolutionary story of dinoflagellates.

Author Contributions

The manuscript was conceptualized and written by ECC and PJK, with PJK as project administrator. Investigation was performed by ECC with assistance from CCH and JAA. Field work and collections were done by ECC and PJK. Methodology was designed by ECC with

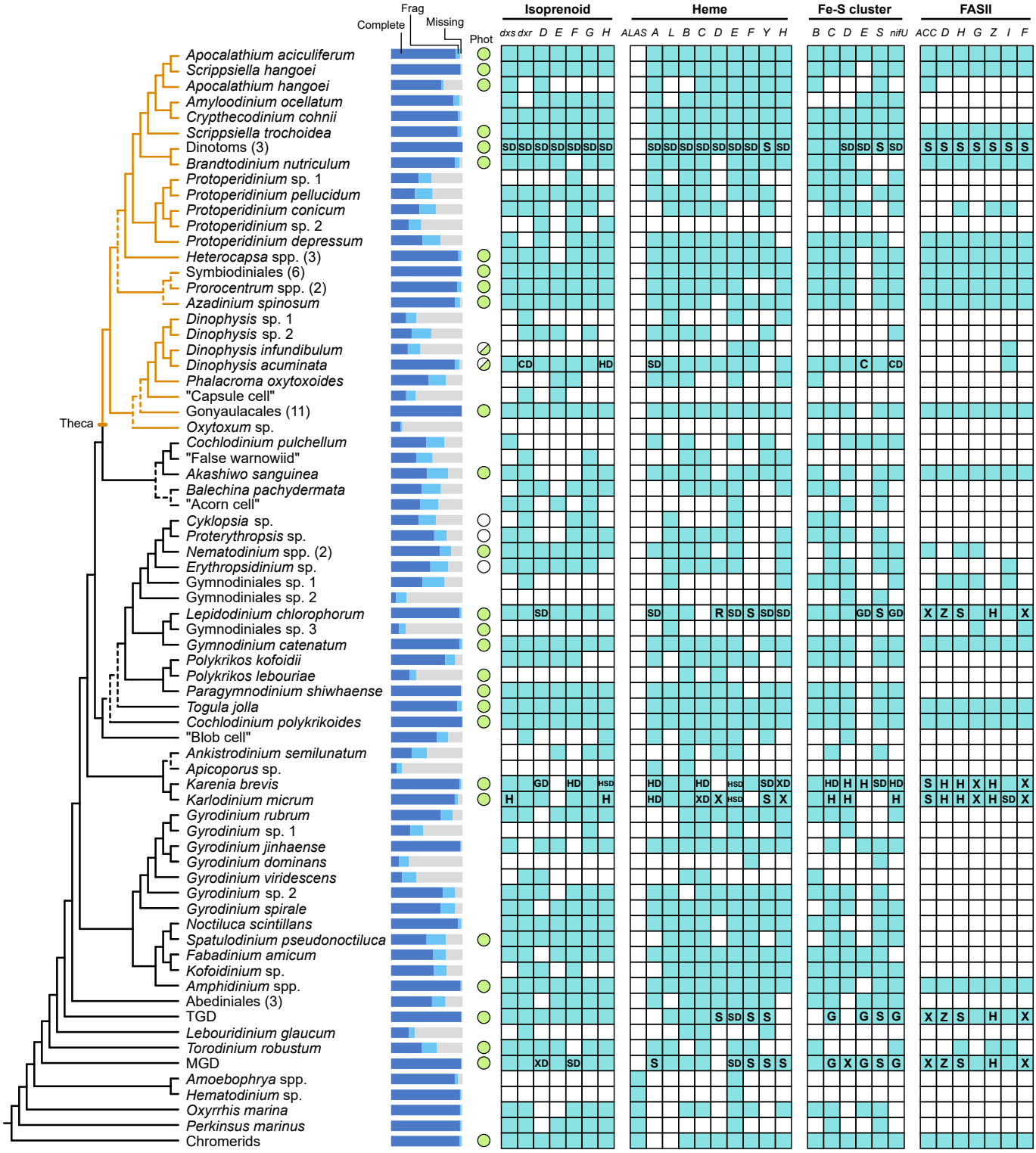

Fig. 4. Presence of plastid-originating synthesis pathway gene transcripts. See Fig. 2A caption for a description of the tree and bar graph. See Fig. 3 caption for an explanation of circles and letters within boxes. Z = rhizarian origin. Gene presence columns headed with single letters are shorthand for gene names beginning with "*iso*" (isoprenoid), "*hem*" (heme), "*suf*" (iron-sulfur or "Fe-S" cluster), and "*fab*" (fatty acid synthesis II or "FASII").

assistance from EH. Funding acquisition and supervision were performed by PJK and BSL. Review and editing of the manuscript were performed by all authors.

CRediT authorship contribution statement

Elizabeth C. Cooney: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Corey C. Holt:** Writing – review & editing, Investigation. **Elisabeth Hehenberger:** Writing – review & editing, Methodology. **Jayd A. Adams:** Writing – review & editing, Investigation. **Brian S. Leander:** Writing – review & editing, Supervision, Funding acquisition. **Patrick J. Keeling:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Assemblies and other data are are available on Dryad at https://doi.org/10.5061/dryad.x95x69prq and https://doi.org/10.5281/zenodo.10553634.

Acknowledgments

We thank Dr. Suzanne Strom for contributing single cell samples to this study.

Funding

This work was supported by Hakai and the Tula Foundation; the Gordon and Betty Moore Foundation [PJK: GBMF9201 (https://doi.org/10.37807/GBMF9201)]; and the Natural Sciences and Engineering Research Council of Canada to BSL [NSERC 2019-03986] and PJK [NSERC 2019-03994].

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108086.

References

Bachvaroff, T.R., Concepcion, G.T., Rogers, C.R., Herman, E.M., Delwiche, C.F., 2004. Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. Protist 155, 65–78. https://doi.org/10.1078/1434461000165.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Béjà, O., Spudich, E.N., Spudich, J.L., Leclerc, M., DeLong, E.F., 2001. Proteorhodopsin phototrophy in the ocean. Nature 411, 786–789. https://doi.org/10.1038/35081051.

Bendtsen, J.D., Nielsen, H., Von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795. https://doi.org/10.1016/j.jmb.2004.05.028.

Burki, F., Shalchian-Tabrizi, K., Pawlowski, J., 2008. Phylogenomics reveals a new "megagroup" including most photosynthetic eukaryotes. Biol. Lett. 4, 366–369. https://doi.org/10.1098/rsbl.2008.0224.

Burki, F., Imanian, B., Hehenberger, E., Hirakawa, Y., Maruyama, S., Keeling, P.J., 2014. Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. Eukaryot. Cell 13, 246–255. https://doi.org/10.1128/EC.00299-13.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

Chan, Y.H., Wong, J.T.Y., 2007. Concentration-dependent organization of DNA by the dinoflagellate histone-like protein HCc3. Nucleic Acids Res. 35, 2573–2583. https://doi.org/10.1093/nar/gkm165.

Cooney, E.C., Okamoto, N., Cho, A., Hehenberger, E., Richards, T.A., Santoro, A.E., Worden, A.Z., Leander, B.S., Keeling, P.J., 2020. Single cell transcriptomics of Abedinium reveals a new early-branching dinoflagellate lineage. Genome Biol. Evol. 12, 2417–2428. https://doi.org/10.1093/gbe/evaa196.

Cooney, E.C., Leander, B.S., Keeling, P.J., 2022. Phylogenomics shows unique traits in Noctilucales are derived rather than ancestral. PNAS Nexus 1, 1–11. https://doi.org/10.1093/pnasnexus/pgac202.

Cooney, E.C., Holt, C.C., Jacko-reynolds, V.K.L., Brian, S., Keeling, P.J., 2023. Photosystems in the eye-like organelles of heterotrophic warnowiid dinoflagellates. Curr. Biol. 1–9 https://doi.org/10.1016/j.cub.2023.08.052.

de la Torre, J.R., Christianson, L.M., Beja, O., Suzuki, M.T., Karl, D.M., Heidelberg, J., DeLong, E.F., 2003. Proteorhodopsin genes are distributed among divergent marine bacterial taxa. Proc. Natl. Acad. Sci. 100, 12830–12835. https://doi.org/10.1073/pnas.2133554100.

Donaher, N., Tanifuji, G., Onodera, N.T., Malfatti, S.A., Chain, P.S.G., Hara, Y., Archibald, J.M., 2009. The complete plastid genome sequence of the secondarily nonphotosynthetic alga Cryptomonas paramecium: Reduction, compaction, and accelerated evolutionary rate. Genome Biol. Evol. 1, 439–448. https://doi.org/10.1093/gbe/evp047.

Dorrell, R.G., Azuma, T., Nomura, M., de Kerdrel, G.A., Paoli, L., Yang, S., Bowler, C., Ishii, K. ichiro, Miyashita, H., Gile, G.H., Kamikawa, R., 2019. Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. Proc. Natl. Acad. Sci. U. S. A. 116, 6914–6923. https://doi.org/10.1073/pnas.1819976116.

Frigaard, N.U., Martinez, A., Mincer, T.J., DeLong, E.F., 2006. Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. Nature 439, 847–850. https://doi.org/10.1038/nature04435.

Füssy, Z., Záhonová, K., Tomcala, A., Krajcovic, J., Yurchenko, V., Oborník, M., Eliáš, M., 2020. The cryptic plastid of euglena longa defines a new type of nonphotosynthetic plastid organelle. mSphere 5, 1–19.

García-Portela, M., Moestrup, Ø., Daugbjerg, N., Altenburger, A., Lundholm, N., 2023. Studies on the complex Warnowiaceae (Dinophyceae) I. Lohmann's Pouchetia parva refound and renamed Nematodinium parvum comb. nov. (= Warnowia parva). Phycologia 00, 1–15. https://doi.org/10.1080/00318884.2023.2244810.

Gavelis, G.S., Hayakawa, S., White, R.A., Gojobori, T., Suttle, C.A., Keeling, P.J., Leander, B.S., 2015. Eye-like ocelloids are built from different endosymbiotically acquired components. Nature 523, 204–207. https://doi.org/10.1038/nature14593.

Gavelis, G.S., Keeling, P.J., Leander, B.S., 2017. How exaptations facilitated photosensory evolution: Seeing the light by accident. BioEssays 39, 1–8. https://doi.org/10.1002/bies.201600266.

Gawryluk, R.M.R., Tikhonenkov, D.V., Hehenberger, E., Husnik, F., Mylnikov, A.P., Keeling, P.J., 2019. Non-photosynthetic predators are sister to red algae. Nature 572, 240–243. https://doi.org/10.1038/s41586-019-1398-6.

Gómez, F., 2012. A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). Syst. Biodivers. 10, 267–275. https://doi.org/10.1080/14772000.2012.721021.

Gómez, F., 2020. Diversity and classification of dinoflagellates, in: Dinoflagellates: Classification Evolution Physiology and Ecological Significance. pp. 1–38. https://doi.org/10.1007/bf02861206.

Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., Pedrós-Alió, C., Pinhassi, J., 2007. Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature 445, 210–213. https://doi.org/10.1038/nature05381.

Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D.L., González, J.M., Pinhassi, J., 2010. Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. PLoS Biol. 8, 2–11. https://doi.org/10.1371/journal.pbio.1000358.

Gómez-Consarnau, L., Levine, N.M., Cutter, L., Wang, D., Seegers, B., Arístegui, J., Fuhrman, J.A., Gasol, J.M., Sañudo-Wilhelmy, S.A., 2017. Marine proteorhodopsins rival chlorophyll in solar energy capture. Nature in review.

Gornik, S.G., Ford, K.L., Mulhern, T.D., Bacic, A., McFadden, G.I., Waller, R.F., 2012. Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. Curr. Biol. 22, 2303–2312. https://doi.org/10.1016/j.cub.2012.10.036.

Gornik, S.G., Cassin, A.M., Macrae, J.I., Ramaprasad, A., Rchiad, Z., McConville, M.J., Bacic, A., McFadden, G.I., Pain, A., Waller, R.F., 2015. Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. Proc. Natl. Acad. Sci. 112, 5767–5772. https://doi.org/10.1073/pnas.1423400112.

Greuet, C., 1965. Structure fine de l'ocelle d'*Erythropsis pavillardi* Hertwig, Peridinien Warnowiidae Lindemann. Comptes Rendus Des Seances L'academie Des Sci. 261, 1904–1907.

Greuet, C., 1977. Evolution structurale et ultrastructurale de l'ocelloide d'Erythropsidinium pavillardi Kofoid et Swezy (Peridinien Warnowiidae Lindemann) au cours des divisions binaire et palintomiques. Protistologica 13, 127–143.

Guo, Z., Zhang, H., Lin, S., 2014. Light-promoted rhodopsin expression and starvation survival in the marine dinoflagellate *Oxyrrhis marina*. PLoS One 9, 1–23. https://doi.org/10.1371/journal.pone.0114941.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference

generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.
Hackett, J.D., Yoon, H.S., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Nosenko, T., Bhattacharya, D., 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr. Biol. 14, 213–218. https://doi.org/10.1016/j.cub.2004.01.032.
Hayakawa, S., Takaku, Y., Hwang, J.S., Horiguchi, T., Suga, H., Gehring, W., Ikeo, K., Gojobori, T., 2015. Function and evolutionary origin of unicellular camera-type eye structure. PLoS One 10, 1–16. https://doi.org/10.1371/journal.pone.0118415.
Hehenberger, E., Imanian, B., Burki, F., Keeling, P.J., 2014. Evidence for the retention of two evolutionary distinct plastids in dinoflagellates with diatom endosymbionts. Genome Biol. Evol. 6, 2321–2334. https://doi.org/10.1093/gbe/evu182.
Hehenberger, E., Burki, F., Kolisko, M., Keeling, P.J., 2016. Functional relationship between a dinoflagellate host and its diatom endosymbiont. Mol. Biol. Evol. 33, 2376–2390. https://doi.org/10.1093/molbev/msw109.
Hehenberger, E., Gast, R.J., Keeling, P.J., 2019. A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis. Proc. Natl. Acad. Sci. 116, 17934–17942. https://doi.org/10.1073/pnas.1910121116.
Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.
Holt, C.C., Hehenberger, E., Tikhonenkov, D.V., Jacko-Reynolds, V.K.L., Okamoto, N., Cooney, E.C., Irwin, N.A.T., Keeling, P.J., 2023. Multiple parallel origins of parasitic Marine Alveolates. Nat. Commun. 1–14 https://doi.org/10.1038/s41467-023-42807-0.
Howe, C.J., Barbrook, A.C., Nisbet, R.E.R., Lockhart, P.J., Larkum, A.W.D., 2008a. The origin of plastids. Philos. Trans. r. Soc. B Biol. Sci. 363, 2675–2685. https://doi.org/10.1098/rstb.2008.0050.
Howe, C.J., Nisbet, R.E.R., Barbrook, A.C., 2008b. The remarkable chloroplast genome of dinoflagellates. J. Exp. Bot. 59, 1035–1045. https://doi.org/10.1093/jxb/erm292.
Irwin, N.A.T., Martin, B.J.E., Young, B.P., Browne, M.J.G., Flaus, A., Loewen, C.J.R., Keeling, P.J., Howe, L.J., 2018. Viral proteins as a potential driver of histone depletion in dinoflagellates. Nat. Commun. 9 https://doi.org/10.1038/s41467-018-03993-4.
Irwin, N.A.T., Pittis, A.A., Richards, T.A., Keeling, P.J., 2022. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. Nat. Microbiol. 7, 327–336. https://doi.org/10.1038/s41564-021-01026-3.
Janouškovec, J., Horak, A., Obornik, M., Lukes, J., Keeling, P.J., 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. 107, 10949–10954. https://doi.org/10.1073/pnas.1003335107.
Janouškovec, J., Tikhonenkov, D.V., Burki, F., Howe, A.T., Kolísko, M., Mylnikov, A.P., Keeling, P.J., 2015. Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1423790112.
Janouškovec, J., Gavelis, G.S., Burki, F., Dinh, D., Bachvaroff, T.R., Gornik, S.G., Bright, K.J., Imanian, B., Strom, S.L., Delwiche, C.F., Waller, R.F., Fensome, R.A., Leander, B.S., Rohwer, F.L., Saldarriaga, J.F., 2017. Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. Proc. Natl. Acad. Sci. 114, E171–E180. https://doi.org/10.1073/pnas.1614842114.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.
Karnkowska, A., Yubuki, N., Maruyama, M., Yamaguchi, A., Kashiyama, Y., Suzaki, T., Keeling, P.J., Hampl, V., Leander, B.S., 2023. Euglenozoan kleptoplasty illuminates the early evolution of photoendosymbiosis. Proc. Natl. Acad. Sci. 120, 1–12. https://doi.org/10.1073/pnas.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Keeling, P.J., 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu. Rev. Plant Biol. 64, 583–607. https://doi.org/10.1146/annurev-arplant-050312-120144.
Keeling, P.J., 2024. Horizontal gene transfer in eukaryotes: aligning theory with data. Nat. Rev. Genet. 1–15 https://doi.org/10.1038/s41576-023-00688-5.
Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A. M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 2014. The marine microbial eukaryote transcriptome sequencing project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol. 12 https://doi.org/10.1371/journal.pbio.1001889.
Keeling, P.J., del Campo, J., 2017. Marine protists are not just big bacteria. Curr. Biol. 27, R541–R549. https://doi.org/10.1016/j.cub.2017.03.075.
Keeling, P.J., Inagaki, Y., 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. Proc. Natl. Acad. Sci. u. s. a. 101, 15380–15385. https://doi.org/10.1073/pnas.0404505101.
Kim, J.I., Jeong, M., Archibald, J.M., Shin, W., 2020. Comparative plastid genomics of non-photosynthetic chrysophytes: Genome reduction and compaction. Front. Plant Sci. 11, 1–14. https://doi.org/10.3389/fpls.2020.572703.
Kim, M., Nam, S.W., Shin, W., Coats, D.W., Park, M.G., 2012. Dinophysis caudata (dinophyceae) sequesters and retains plastids from the mixotrophic ciliate prey Mesodinium rubrum. J. Phycol. 48, 569–579. https://doi.org/10.1111/j.1529-8817.2012.01150.x.
Kolisko, M., Boscaro, V., Burki, F., Lynn, D.H., Keeling, P.J., 2014. Single-cell transcriptomics for microbial eukaryotes. Curr. Biol. 24, R1081–R1082. https://doi.org/10.1016/j.cub.2014.10.026.
Larkum, A.W.D., Lockhart, P.J., Howe, C.J., 2007. Shopping for plastids. Trends Plant Sci. 12, 189–195. https://doi.org/10.1016/j.tplants.2007.03.011.
Leinonen, R., Sugawara, H., Shumway, M., 2011. The Sequence Read Archive. Nucleic Acids Res. 39, 2010–2012. https://doi.org/10.1093/nar/gkq1019.
Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. 38, 4647–4654. https://doi.org/10.1093/molbev/msab199.
Marchetti, A., Schruth, D.M., Durkin, C.A., Parker, M.S., Kodner, R.B., Berthiaume, C.T., Morales, R., Allen, A.E., Armbrust, E.V., 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. Proc. Natl. Acad. Sci. 109, E317–E325. https://doi.org/10.1073/pnas.1118408109.
Marchetti, A., Catlett, D., Hopkinson, B.M., Ellis, K., Cassar, N., 2015. Marine diatom proteorhodopsins and their potential role in coping with low iron availability. ISME J. 9, 1–4. https://doi.org/10.1038/ismej.2015.74.
Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet.journal.
Mathur, V., Kolísko, M., Hehenberger, E., Irwin, N.A.T., Leander, B.S., Kristmundsson, Á., Freeman, M.A., Keeling, P.J., 2019. Multiple independent origins of apicomplexan-Like parasites. Curr. Biol. 29, 2936–2941. https://doi.org/10.1016/j.cub.2019.07.019.
Matsumoto, T., Shinozaki, F., Chikuni, T., Yabuki, A., Takishita, K., Kawachi, M., Nakayama, T., Inouye, I., Hashimoto, T., Inagaki, Y., 2011. Green-colored plastids in the dinoflagellate genus Lepidodinium are of core chlorophyte origin. Protist 162, 268–276. https://doi.org/10.1016/j.protis.2010.07.001.
Nand, A., Zhan, Y., Salazar, O.R., Aranda, M., Voolstra, C.R., Dekker, J., 2021. Genetic and spatial organization of the unusual chromosomes of the dinoflagellate Symbiodinium microadriaticum. Nat. Genet. https://doi.org/10.1038/s41588-021-00841-y.
Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
Palovaara, J., Akram, N., Baltar, F., Bunse, C., Forsberg, J., Pedros-Alio, C., Gonzalez, J. M., Pinhassi, J., 2014. Stimulation of growth by proteorhodopsin phototrophy involves regulation of central metabolic pathways in marine planktonic bacteria. Proc. Natl. Acad. Sci. 111, E3650–E3658. https://doi.org/10.1073/pnas.1402617111.
Patron, N.J., Waller, R.F., Archibald, J.M., Keeling, P.J., 2005. Complex protein targeting to dinoflagellate plastids. J. Mol. Biol. 348, 1015–1024. https://doi.org/10.1016/j.jmb.2005.03.030.
Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., Sandberg, R., 2014. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181. https://doi.org/10.1038/nprot.2014.006.
Poux, S., Arighi, C.N., Magrane, M., Bateman, A., Wei, C.H., Lu, Z., Boutet, E., Bye-A-Jee, H., Famiglietti, M.L., Roechert, B., 2017. On expert curation and scalability: UniProtKB/Swiss-Prot as a case study. Bioinformatics 33, 3454–3460. https://doi.org/10.1093/bioinformatics/btx439.
Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 5, 1–10. https://doi.org/10.1371/journal.pone.0009490.
Quang, L.S., Gascuel, O., Lartillot, N., 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24, 2317–2323. https://doi.org/10.1093/bioinformatics/btn445.
Rogers, M.B., Watkins, R.F., Harper, J.T., Durnford, D.G., Gray, M.W., Keeling, P.J., 2007. A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. BMC Evol. Biol. 7, 1–13. https://doi.org/10.1186/1471-2148-7-89.
Röhrich, R.C., Englert, N., Troschke, K., Reichenberg, A., Hintz, M., Seeber, F., Balconi, E., Aliverti, A., Zanetti, G., Köhler, U., Pfeiffer, M., Beck, E., Jomaa, H., Wiesner, J., 2005. Reconstitution of an apicoplast-localised electron transfer pathway involved in the isoprenoid biosynthesis of *Plasmodium falciparum*. FEBS Lett. 579, 6433–6438. https://doi.org/10.1016/j.febslet.2005.10.037.
Roure, B., Rodriguez-Ezpeleta, N., Philippe, H., 2007. SCaFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol. Biol. 7, 1–12. https://doi.org/10.1186/1471-2148-7-S1-S2.
Saldarriaga, J.F., Taylor, F.J.R., Keeling, P.J., Cavalier-Smith, T., 2001. Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. J. Mol. Evol. 53, 204–213. https://doi.org/10.1007/s002390010210.
Sanchez-Puerta, M.V., Lippmeier, J.C., Apt, K.E., Delwiche, C.F., 2007. Plastid genes in a non-photosynthetic dinoflagellate. Protist 158, 105–117. https://doi.org/10.1016/j.protis.2006.09.004.
Sarai, C., Tanifuji, G., Nakayama, T., Kamikawa, R., Takahashi, K., Yazaki, E., Matsuo, E., Miyashita, H., Ishida, K., Iwataki, M., Inagaki, Y., 2020. Dinoflagellates with relic

endosymbiont nuclei as models for elucidating organellogenesis. Proc. Natl. Acad. Sci. u. s. a. 117, 5364–5375. https://doi.org/10.1073/pnas.1911884117.
Seeber, F., Soldati-Favre, D., 2010. Metabolic pathways in the apicoplast of apicomplexa. Int. Rev. Cell Mol. Biol. 281, 161–228. https://doi.org/10.1016/S1937-6448(10)81005-6.
Sekiguchi, H., Moriya, M., Nakayama, T., Inouye, I., 2002. Vestigial chloroplasts in heterotrophic stramenopiles Pteridomonas danica and Ciliophrys infusionum (Dictyochophyceae). Protist 153, 157–167. https://doi.org/10.1078/1434-4610-00094.
Sharma, A.K., Spudich, J.L., Doolittle, W.F., 2006. Microbial rhodopsins: Functional versatility and genetic mobility. Trends Microbiol. 14, 463–469. https://doi.org/10.1016/j.tim.2006.09.006.
Shi, X., Li, L., Guo, C., Lin, X., Li, M., Lin, S., 2015. Rhodopsin gene expression regulated by the light dark cycle, light spectrum and light intensity in the dinoflagellate *Prorocentrum*. Front. Microbiol. 6, 1–11. https://doi.org/10.3389/fmicb.2015.00555.
Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508. https://doi.org/10.1080/10635150290069913.
Slamovits, C.H., Okamoto, N., Burri, L., James, E.R., Keeling, P.J., 2011. A bacterial proteorhodopsin proton pump in marine eukaryotes. Nat. Commun. 2, 183. https://doi.org/10.1038/ncomms1188.
Spudich, J.L., Yang, C.S., Jung, K.H., Spudich, E.N., 2000. Retinylidene proteins: Structures and functions from archaea to humans. Annu. Rev. Cell Dev. Biol. 16, 365–392. https://doi.org/10.1146/annurev.cellbio.16.1.365.
Stephens, T.G., González-Pech, R.A., Cheng, Y., Mohamed, A.R., Burt, D.W., Bhattacharya, D., Ragan, M.A., Chan, C.X., 2020. Genomes of the dinoflagellate *Polarella glacialis* encode tandemly repeated single-exon genes with adaptive functions. BMC Biol. 18, 1–21. https://doi.org/10.1186/s12915-020-00782-8.
Taylor, F.J.R., Hoppenrath, M., Saldarriaga, J.F., 2008. Dinoflagellate diversity and distribution. Biodivers. Conserv. 17, 407–418. https://doi.org/10.1007/s10531-007-9258-3.
Tengs, T., Dahlberg, O.J., Shalchian-Tabrizi, K., Klaveness, D., Rudi, K., Delwiche, C.F., Jakobsen, K.S., 2000. Phylogenetic analyses indicate flint the 19'hexanoyloxy-fucoxanthin- containing dinoflagellates have tertiary plastids of haptophyte origin. Mol. Biol. Evol. 17, 718–729. https://doi.org/10.1093/oxfordjournals.molbev.a026350.
Uhlig, G., 1964. Eine einfache methode zur extraktion der vagilen, mesopsammalen mikrofauna. Helgoländer Wissenschaftliche Meeresuntersuchungen 11, 178–185. https://doi.org/10.1007/BF01612370.
Van Dolah, F.M., Zippay, M.L., Pezzolesi, L., Rein, K.S., Johnson, J.G., Morey, J.S., Wang, Z., Pistocchi, R., 2013. Subcellular localization of dinoflagellate polyketide synthases and fatty acid synthase activity. J. Phycol. 49, 1118–1127. https://doi.org/10.1111/jpy.12120.
Veldhuis, M.J.W., Cucci, T.L., Sieracki, M.E., 1997. Cellular DNA content of marine phytoplankton using two new fluorochromes: Taxonomic and ecological implications. J. Phycol. 33, 527–541. https://doi.org/10.1111/j.0022-3646.1997.00527.x.
Wang, H.-C., Minh, B.Q., Susko, E., Roger, A.J., 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67, 216–235. https://doi.org/10.1093/sysbio/syx068.
Wisecaver, J.H., Hackett, J.D., 2010. Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate Dinophysis acuminata. BMC Genomics 11. https://doi.org/10.1186/1471-2164-11-366.
Wolfe, A.D., DePamphilis, C.W., 1998. The effect of relaxed functional constraints on the photosynthetic gene *rbcL* in photosynthetic and nonphotosynthetic parasitic plants. Mol. Biol. Evol. 15, 1243–1258. https://doi.org/10.1093/oxfordjournals.molbev.a025853.
Wong, J.T.Y., New, D.C., Wong, J.C.W., Hung, V.K.L., 2003. Histone-like proteins of the dinoflagellate *Crypthecodinium cohnii* have homologies to bacterial DNA-binding proteins. Eukaryot. Cell 2, 646–650. https://doi.org/10.1128/EC.2.3.646.
Xiang, T., Nelson, W., Rodriguez, J., Tolleter, D., Grossman, A.R., 2015. *Symbiodinium* transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. Plant J. 82, 67–80. https://doi.org/10.1111/tpj.12789.
Zhang, H., Hou, Y., Miranda, L., Campbell, D.A., Sturm, N.R., Gaasterland, T., Lin, S., 2007. Spliced leader RNA trans-splicing in dinoflagellates. Proc. Natl. Acad. Sci. 104, 4618–4623. https://doi.org/10.1073/pnas.0700258104.

Glossary

dinoflagellate viral nucleoprotein (DVNP): chromatin packaging proteins of viral origin found only in dinoflagellates
histone-like protein (HLP): chromatin packaging proteins of bacterial origin found only in dinoflagellates
marine alveolate (MALV): an umbrella category for dinoflagellates of Syndiniales and other early-diverging lineages known mostly from environmental sequences, but highly abundant in the ocean
type I fatty acid synthesis pathway (FASI): cytosolic fatty acid biosynthetic pathway
type II fatty acid synthesis pathway (FASII): plastidial fatty acid biosynthetic pathway