Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics

Jan Janouškovec, Gregory S. Gavelis, Fabien Burki, Tsvetan R. Bachvaroff, Sebastian G. Gornik, Kelvin J. Bright, Behzad Imanian, Suzanne L. Strom, Charles F. Delwiche, Ross F. Waller, Robert A. Fensome, Brian S. Leander, Forest L. Rohwer, and Juan F. Saldarriaga

Dinoflagellates are key species in marine environments, but they remain poorly understood in part because of their large, complex genomes, unique molecular biology, and unresolved inter-group relationships. We created a taxonomically representative dataset of dinoflagellate transcriptomes and used this to infer a strongly supported phylogeny to map major morphological and molecular transitions in dinoflagellate evolution. Our results show an early branching position of Noctiluca, monophyly of thecate (plate-bearing) dinoflagellates, and paraphyly of thecate ones. This represents an ambiguous phylogenetic evidence for a single origin of the group’s cellulosic theca, which we show coincided with a radiation of cellulases implicated in cell division. By integrating dinoflagellate molecular, fossil, and biogeochemical evidence, we propose a revised model for the evolution of thecal tabulations and suggest that the late acquisition of dinosterol in the group is inconsistent with dinoflagellates being the source of this biomarker in pre-Mesozoic strata. Three distantly related, fundamentally nonphotosynthetic dinoflagellates, Noctiluca, Oxyrrhis, and Dinophysis, contain cryptic plastidial metabolisms and lack alternative cytosolic pathways, suggesting that all free-living dinoflagellates are metabolically dependent on plastids. This finding led us to propose general mechanisms of dependency on plastid organelles in eukaryotes that have lost photosynthesis; it also suggests that the evolutionary origin of bioluminescence in nonphotosynthetic dinoflagellates may be linked to plasticid tetrapyrrole biosynthesis. Finally, we use our phylogenetic framework to show that dinoflagellate nuclei have recruited DNA-binding proteins in three distinct evolutionary waves, which included two independent acquisitions of bacterial histone-like proteins.

dinoflagellates | phylogeny | theca | plastids | dinosterol

Dinoflagellates comprise approximately 2,400 named extant species, of which approximately half are photosynthetic (1). However, this represents a fraction of their estimated diversity: in surface marine waters, dinoflagellates are some of the most abundant and diverse eukaryotes known (2). Dinoflagellates’ ecological significance befits their abundance: photosynthetic species are dominant marine primary producers, and phagotrophic species play an important role in the microbial loop through predation and nutrient recycling. Approximately 75–80% of the toxic eukaryotic phytoplankton species are dinoflagellates, and they cause shellfish poisoning and harmful algal blooms of global importance. Symbiotic genera like Symbiodinium participate in interactions with metazoans and are essential for the formation of reef ecosystems, and parasitic forms play a central role in the collapse of harmful algal blooms, including those caused by dinoflagellates themselves (3). Dinoflagellates synthesize important secondary metabolites including steroids, polyketides, toxins, and dimethylsulfide, and several of them have evolved bioluminescence. They have a nonnucleosomal system of nuclear DNA packaging, widespread trans-splicing in mRNAs, and highly unusual plastid and mitochondrial genomes with complex transcript modifications (4–8). Their photosynthesis relies on unique light-harvesting complexes, and its frequent loss in the group makes dinoflagellates a model for understanding the basis of evolutionary reliance on nonphotosynthetic plastid organelles.

Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics


www.pnas.org/cgi/doi/10.1073/pnas.1614842114 PNAS

Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics


This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the iMi-robe database (project code CAM_P_0001000) and GenBank Transcriptome Shotgun Assembly (TSA) Sequence Database (accession nos. GELK00000000 and GEMP00000000).

Significance

We created a dataset of dinoflagellate transcriptomes to resolve internal phylogenetic relationships of the group. We show that the dinoflagellate theca originated once, through a process that likely involved changes in the metabolism of cellulose, and suggest that a late origin of dinosterol in the group is at odds with dinoflagellates being the source of this important biomarker before the Mesozoic. We also show that nonphotosynthetic dinoflagellates have retained nonphotosynthetic plastids with vital metabolic functions, and propose that one of these may be the evolutionary source of dinoflagellate bioluminescence. Finally, we reconstruct major molecular and morphological transitions in dinoflagellates and highlight the role of horizontal gene transfer in the origin of their unique nuclear architecture.


This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the iMi-robe database (project code CAM_P_0001000) and GenBank Transcriptome Shotgun Assembly (TSA) Sequence Database (accession nos. GELK00000000 and GEMP00000000).
based on their tabulation, the arrangement of vesicles in the cell cortex that may or may not contain cellulose thecal plates (collectively the theca). Whether the dinoflagellate theca originated once or multiple times has been controversial. Dinoflagellates have left a fossil record that is one of the richest among protists, and many preserve a detailed record of tabulation through reflection of thecal plates that provide insights into the history of some modern taxa, as well as extinct groups. They have also left an extensive biogeochemical record (i.e., sterols), but reconciling this evidence with poorly resolved gene phylogenies has been difficult (15, 16).

We circumvented the difficulties inherent to the sequencing of large dinoflagellate genomes by compiling a phylogenetically representative transcriptomic dataset to illuminate dinoflagellate biology and evolution. We infer a strongly resolved phylogeny for dinoflagellates and provide phylogenetic evidence for a single origin of the theca, which coincides with major predicted changes in cellulose metabolism. We propose a model for the evolution of tabulation, and show that pre-Mesozoic markers that have often been associated with the group are unlikely to have come from dinoflagellate sources. Three distantly related, nonphotosynthetic dinoflagellates were found to be dependent on plastid metabolism, and we propose that this dependency is likely to apply to all freeliving (i.e., nonparasitic) dinoflagellates and that plastidial metabolites are likely to represent the evolutionary origin of dinoflagellate bioluminescence. Finally, we reconstruct character evolution in dinoflagellates and show that their modern-day biology was shaped by stepwise molecular, metabolic, and morphological innovations, including nuclear DNA-binding proteins of a bacterial origin.

Results and Discussion

Dinoflagellate Phylogeny.

Representative, strongly resolved phylogeny for dinoflagellates. An inability to resolve dinoflagellate relationships has hindered evolution-driven predictions of their biology and a full integration of the group’s rich fossil record with molecular-based schemes of evolution. Our aim was to overcome these limitations by erecting a framework for character mapping rooted in a representative phylogeny of all major dinoflagellate lineages. We generated trascriptomes from key species lacking deep-coverage sequence data—Noctiluca scintillans, Togula jolla, Protoceratium reticulatum, Polarkea glacialis, Hematodinium spp., Amphidinium carterae, and two isolates of Amoebophrya sp. parasites together with their hosts, Karlodinium veneficum and Akashio sanguinea—and complemented these with data from recent sequencing projects (9, 17–19) (SI Appendix, Table S1). Sequences were added into alignments of conserved proteins previously used in eukaryotic phylogenies (20), and their orthology was verified in individual protein trees (Materials and Methods); 101 orthologous alignments with the fewest missing data were selected and concatenated into three phylogenetic matrices that differ by the root (Fig. 1 A and SI Appendix, Table S1). The matrices include six dinoflagellate lineages previously absent in multiprotein phylogenies: Noctilucales, Gymnodiniales s.s., Togula, Akashiwo, Prorocentrales, and Dinophysiales, representing a broadly sampled large dinoflagellate datasets. Maximum-likelihood and Bayesian inferences on all three matrices gave consistent and well-supported topologies (Fig. 1 A and B). Relationships between the outgroups and the early-branching Oxyrhis, Hematodinium, Amoebophrya, Babesia, and Plasmodium, as well as the modern clades Oxyrrhis, Gonyaulax, Gymnodinium, and Noctiluca, are well characterized in the phylogenies. These three clades plus the outgroups form together an extensive biogeochemical record (i.e., sterols) that provides insights into the history of some modern taxa, as well as extinct groups. They have also left an extensive biogeochemical record (i.e., sterols), but reconciling this evidence with poorly resolved gene phylogenies has been difficult (15, 16).

We circumvented the difficulties inherent to the sequencing of large dinoflagellate genomes by compiling a phylogenetically representative transcriptomic dataset to illuminate dinoflagellate biology and evolution. We infer a strongly resolved phylogeny for dinoflagellates and provide phylogenetic evidence for a single origin of the theca, which coincides with major predicted changes in cellulose metabolism. We propose a model for the evolution of tabulation, and show that pre-Mesozoic markers that have often been associated with the group are unlikely to have come from dinoflagellate sources. Three distantly related, nonphotosynthetic dinoflagellates were found to be dependent on plastid metabolism, and we propose that this dependency is likely to apply to all freeliving (i.e., nonparasitic) dinoflagellates and that plastidial metabolites are likely to represent the evolutionary origin of dinoflagellate bioluminescence. Finally, we reconstruct character evolution in dinoflagellates and show that their modern-day biology was shaped by stepwise molecular, metabolic, and morphological innovations, including nuclear DNA-binding proteins of a bacterial origin.

Results and Discussion

Dinoflagellate Phylogeny.

Representative, strongly resolved phylogeny for dinoflagellates. An inability to resolve dinoflagellate relationships has hindered evolution-driven predictions of their biology and a full integration of the group’s rich fossil record with molecular-based schemes of evolution. Our aim was to overcome these limitations by erecting a framework for character mapping rooted in a representative phylogeny of all major dinoflagellate lineages. We generated trascriptomes from key species lacking deep-coverage sequence data—Noctiluca scintillans, Togula jolla, Protoceratium reticulatum, Polarkea glacialis, Hematodinium spp., Amphidinium carterae, and two isolates of Amoebophrya sp. parasites together with their hosts, Karlodinium veneficum and Akashio sanguinea—and complemented these with data from recent sequencing projects (9, 17–19) (SI Appendix, Table S1). Sequences were added into alignments of conserved proteins previously used in eukaryotic phylogenies (20), and their orthology was verified in individual protein trees (Materials and Methods); 101 orthologous alignments with the fewest missing data were selected and concatenated into three phylogenetic matrices that differ by the root (Fig. 1 A and SI Appendix, Table S1). The matrices include six dinoflagellate lineages previously absent in multiprotein phylogenies: Noctilucales, Gymnodiniales s.s., Togula, Akashiwo, Prorocentrales, and Dinophysiales, representing a broadly sampled large dinoflagellate datasets. Maximum-likelihood and Bayesian inferences on all three matrices gave consistent and well-supported topologies (Fig. 1 A and B). Relationships between the outgroups and the early-branching Oxyrhis, Hematodinium, Amoebophrya, Babesia, and Plasmodium, as well as the modern clades Oxyrrhis, Gonyaulax, Gymnodinium, and Noctiluca, are well characterized in the phylogenies. These three clades plus the outgroups form together an extensive biogeochemical record (i.e., sterols) that provides insights into the history of some modern taxa, as well as extinct groups. They have also left an extensive biogeochemical record (i.e., sterols), but reconciling this evidence with poorly resolved gene phylogenies has been difficult (15, 16).

We circumvented the difficulties inherent to the sequencing of large dinoflagellate genomes by compiling a phylogenetically representative transcriptomic dataset to illuminate dinoflagellate biology and evolution. We infer a strongly resolved phylogeny for dinoflagellates and provide phylogenetic evidence for a single origin of the theca, which coincides with major predicted changes in cellulose metabolism. We propose a model for the evolution of tabulation, and show that pre-Mesozoic markers that have often been associated with the group are unlikely to have come from dinoflagellate sources. Three distantly related, nonphotosynthetic dinoflagellates were found to be dependent on plastid metabolism, and we propose that this dependency is likely to apply to all freeliving (i.e., nonparasitic) dinoflagellates and that plastidial metabolites are likely to represent the evolutionary origin of dinoflagellate bioluminescence. Finally, we reconstruct character evolution in dinoflagellates and show that their modern-day biology was shaped by stepwise molecular, metabolic, and morphological innovations, including nuclear DNA-binding proteins of a bacterial origin.

Results and Discussion

Dinoflagellate Phylogeny.

Representative, strongly resolved phylogeny for dinoflagellates. An inability to resolve dinoflagellate relationships has hindered evolution-driven predictions of their biology and a full integration of the group’s rich fossil record with molecular-based schemes of evolution. Our aim was to overcome these limitations by erecting a framework for character mapping rooted in a representative phylogeny of all major dinoflagellate lineages. We generated trascriptomes from key species lacking deep-coverage sequence data—Noctiluca scintillans, Togula jolla, Protoceratium reticulatum, Polarkea glacialis, Hematodinium spp., Amphidinium carterae, and two isolates of Amoebophrya sp. parasites together with their hosts, Karlodinium veneficum and Akashio sanguinea—and complemented these with data from recent sequencing projects (9, 17–19) (SI Appendix, Table S1). Sequences were added into alignments of conserved proteins previously used in eukaryotic phylogenies (20), and their orthology was verified in individual protein trees (Materials and Methods); 101 orthologous alignments with the fewest missing data were selected and concatenated into three phylogenetic matrices that differ by the root (Fig. 1 A and SI Appendix, Table S1). The matrices include six dinoflagellate lineages previously absent in multiprotein phylogenies: Noctilucales, Gymnodiniales s.s., Togula, Akashiwo, Prorocentrales, and Dinophysiales, representing a broadly sampled large dinoflagellate datasets. Maximum-likelihood and Bayesian inferences on all three matrices gave consistent and well-supported topologies (Fig. 1 A and B). Relationships between the outgroups and the early-branching Oxyrhis, Hematodinium, Amoebophrya, Babesia, and Plasmodium, as well as the modern clades Oxyrrhis, Gonyaulax, Gymnodinium, and Noctiluca, are well characterized in the phylogenies. These three clades plus the outgroups form together an extensive biogeochemical record (i.e., sterols) that provides insights into the history of some modern taxa, as well as extinct groups. They have also left an extensive biogeochemical record (i.e., sterols), but reconciling this evidence with poorly resolved gene phylogenies has been difficult (15, 16).

We circumvented the difficulties inherent to the sequencing of large dinoflagellate genomes by compiling a phylogenetically representative transcriptomic dataset to illuminate dinoflagellate biology and evolution. We infer a strongly resolved phylogeny for dinoflagellates and provide phylogenetic evidence for a single origin of the theca, which coincides with major predicted changes in cellulose metabolism. We propose a model for the evolution of tabulation, and show that pre-Mesozoic markers that have often been associated with the group are unlikely to have come from dinoflagellate sources. Three distantly related, nonphotosynthetic dinoflagellates were found to be dependent on plastid metabolism, and we propose that this dependency is likely to apply to all freeliving (i.e., nonparasitic) dinoflagellates and that plastidial metabolites are likely to represent the evolutionary origin of dinoflagellate bioluminescence. Finally, we reconstruct character evolution in dinoflagellates and show that their modern-day biology was shaped by stepwise molecular, metabolic, and morphological innovations, including nuclear DNA-binding proteins of a bacterial origin.
and *Amoebozyga* spp. are fully resolved and congruent with earlier studies (11). Core dinoflagellates are monophyletic, and several longstanding issues about their relationships can be resolved (Fig. 1).

**Early position of Noctilucales and athecate paralogy.** Athecate dinoflagellates have long confounded dinoflagellate molecular phylogenies as a result of their intermixing with thecate taxa, for example within the so-called Gymnodiniales–Peridiniales–Prorocentrales (GPP) complex (21), or as a result of the unstable position of certain outliers like the Noctilucales, which have at times been placed as basal or nested deeply inside the group (12, 16, 22). Our analyses resolve these issues and help reconcile dinoflagellate morphological and molecular data in several important ways. First, we find that athecate dinoflagellates represent a paraphyletic assemblage with respect to the thecates (Fig. L.4), suggesting that earlier mixed groupings like the GPP complex are artifacts caused by limited phylogenetic resolution. Second, *N. scintillans* and *A. carterae* are the earliest-branching core dinoflagellates, with *Noctiluca* positioned at the base in most analyses, except for Bayesian inferences on Root 2 matrix, in which it is also basal but together with *Amphidinium* (Fig. 1B). Statistical evaluation of alternative tree topologies by approximately unbiased test and expected-by limited phylogenetic resolution. Second, we find that athecate dinoflagellates represent a paraphyletic assemblage as basal or nested deeply inside the group (12, 16, 22). Our certain outliers like the Noctilucales, which have at times been (GPP) complex (21), or as a result of the unstable position of nies as a result of their intermixing with thecate taxa, for example gellates have long confounded dinoflagellate molecular phyloge-
groups, recover thecate dinoflagellates as monophyletic, always with monophyletic in specific datasets and with low support (12, 14).

Early position of Noctilucales and athecate paraphyly. (Fig. 1).

eral longstanding issues about their relationships can be resolved earlier studies (11). Core dinoflagellates are monophyletic, and sev-
and

and tree topology tests (Fig. 1 –

suggested, for ex-

and their sister group, the Borghiellaceae (27), are also derived from a common ancestor. Four independent lines of evidence support this: monophyly of the modern thecates in multiprotein phylogenies (Fig. 1), rapid emergence of fossils reflecting the poss-

sion of the theca during the early Mesozoic (30), similarities in tabulation patterns between different thecate lineages (15, 26), and the presence of theca-associated cellulases of a common evolu-
tionary origin in modern thecates (Fig. 2).

**Thecal Evolution and Dinoflagellate Paleohistory.**

Phylogeny-driven model for theca origin, evolution, and loss. Most the-
cate dinoflagellates (both living and fossil) belong to the Gonyaulacales and Peridiniales, two orders with tabulations in- volving five to six latitudinal series of thecal plates. The details of these tabulations are consistently distinct and longstanding in the fossil record, a pattern consistent with the fact that, in molecular phylogenies, the two orders are not closely related within the thecates (Fig. 1). These patterns suggest that dinoflagellates with gonyaulacoid–peridinoid tabulations originated comparatively early: the extinct rhaetogonyaulacoids (Fig. 2) in the Middle to Late Triassic (31) and true, modern-looking gonyaulacoids and peridinoids in the later Early Jurassic. Even if the phylogenetic position of the Dinophysiales and Prorocentrales in molecular trees remains unresolved, their tabulation patterns are mor-

phologically divergent and unlikely to represent ancestral or transitional states: the fossil *Nannoceratopsis* suggests, for ex-

ample, that the dinophysiid tabulation type is evolutionarily derived (Fig. 2A). As explained earlier, we suggest that the suessioloid and gymnodinoid tabulations of the Symbiodiniaceae and their sister group, the Borghiellaceae (27), are also derived secondarily from gonyaulacoid–peridinoid ancestors and origi-
nated by a secondary increase in plate number (Fig. 2A); they do not represent early intermediates in theca evolution, as con-
sidered by some earlier models (15, 32). In contrast, the Late Triassic suessioloid fossils such as *Suessia* could represent an intermediate stage between gymnodinoid and gonyaulacoid–peridinoid tabulation types or an independent example of de-
crease in primary plate number from gymnodinoid ancestors (Fig. 2A). All in all, paleontological and molecular phylogenetic data suggest that all living thecate dinoflagellates originated from ancestors with a gonyaulacoid–peridinoid tabulation and
argue for the derived position of the Symbiodiniaceae. The model is limited by the incompleteness of the fossil record and will be further developed by understanding the tabulations and phylogenies of little known or morphologically divergent incertae sedis thecate types like *Heterodinium*, *Thecadinium*, or *Cladopix* (26). No simple scenario [plate decrease, increase, and fragmentation models (32)] can account for the evolution of thecal tabulation from a phylogeny-driven perspective (Fig. 1): secondary increase in plate number is observed not only in symbiodiniaceans but also in *Pyrophacus* (Gonyaulacales), a genus with a multiply tabulated derivation from ancestors with a gonyaulacoid tabulation, whereas other thecates have gone through a process of plate decrease, e.g., Dinophysiales and Prorocentrales (in the hyposome) and the Late Triassic to Middle Jurassic fossil *Valvaleodinium*. Our model also strongly suggests that the theca can be lost: some species in the Symbiodiniaceae and Borghelliaceae lack visible cellulose in amphiesmal vesicles altogether (28, 33), and their phylogenetic positions suggest that their thecae were lost more than once (Fig. 24). Finally, a broad, negative relationship between the number and relative surface area of amphiesmal vesicles and the amount of cellulose contained in them emerges. The Gymnodiniaceae have numerous, small amphiesmal vesicles that lack cellulose, whereas the Gonyaulacales, Peridiniidae, Prorocentrales, and Dinophysiales have few, large amphiesmal vesicles containing thick thecal plates, the ancestral state for all living thecate dinoflagellates (Fig. 24). Symbiodiniaceans that have moderate plate numbers in 7–10 latitudinal series have only thin cellulosic plates, but those members of the Symbiodiniaceae and Borghelliaceae that reverted to a gymnodinoid tabulation often lack cellulose altogether (Fig. 24) (e.g. refs. 28, 33, but see also ref. 27). Additional data for example from the Borghelliaceae and *Pyrophacus* will make it possible to test these trends, but, as things stand now, it seems that the acquisition of thick cellulosic plates within amphiesmal vesicles is constrained by their surface area and number. Subsequent reductions and losses of cellulose in the Symbiodiniaceae and Borghelliaceae relaxed this constraint, leading to a partial or complete reversal to numerous small-sized amphiesmal vesicles. 

Origin of theca coincides with onset of cellulase radiation. The origin of the dinoflagellate theca is intimately linked to the biosynthesis of cellulose, its building material, but investigations into the details of cellulose production in dinoflagellates have been limited to rare ultrastructural and labeling studies (34). Recently, production of a highly expressed cellulase [dCel1 from Glycosyl hydrolase family 7 (GH7)] was shown to be coupled to the cell cycle progression in *Crypteodinium colombii* and was immunolocalized to the cell wall in several dinoflagellates, suggesting an important role in cellulose processing during division (31). We identified multiple diversified paralogs of GH7 genes in all thecates and one to three closely related paralogs in four athecate dinoflagellates in our dataset (SI Appendix, Table S4). A eukaryote-wide phylogeny of 184 slow-evolving GH7 protein sequences (Fig. 25 and SI Appendix, Fig. S1 and SI Materials and Methods) suggests that the thecate
paralogs are derived by multiple rounds of duplication followed by selective lineage sorting. The branching pattern is poorly resolved, but indicates a common origin for most thecate GH7 proteins together with sequences from the athecate *Kareния brevis* and *A. carterae* and algae *Bigelowiella natans* and *Thalassiosira oceanica* (the latter two are nested within dinoflagellates and were presumably spread horizontally). Some duplications in the thecate GH7 occurred at the level of genera or orders, but at least eight and possibly twice as many paralogs apparently originated earlier (SI Appendix, Fig. S1)—presumably in the common ancestor of all thecates. These observations suggest that the radiation of GH7 genes in thecate dinoflagellates is linked to the evolutionary origin and subsequent evolution of the theca. The GH7 protein identified in *K. brevis* (SI Appendix, Table S4) likely corresponds to the dCellI homolog previously immunolocalized in the cell cortex (31). Interestingly, *A. sanguinea*, the likely sister group of thecate dinoflagellates, is immunopositive for that same protein (31), although the corresponding GH7 sequence remains unknown (our mixed transcriptome of *Akashiwo* cells infected by *Amoebobrya* sp. lacks it). The function of this GH7 enzymes in athecate species has not been studied, but they are likely involved in the metabolism of cellulose or related polysaccharides, which may have been an important precondition for the acquisition of the cellulosic thecal plates. Unlike cellulose breakdown, cellulose biosynthesis in dinoflagellates is not understood at the molecular level (34). We identified three types of algal cellulose synthase (CESA-like) homologs in thecate and athecate dinoflagellates, candidates for elucidating their cellulose biosynthesis (SI Appendix, Table S4).

### Dinosterol is absent in deep-branching dinoflagellates.

The diversity and abundance of dinoflagellates in Mesozoic and younger sediments correlates with levels of triaromatic dinosterones, derivatives of the fossilizing biomarker 4-methyl sterol, dinosterol (4α, 23, 24R-trimethyl-5α-cholest-22E-en-3β-ol) (15, 35). Dinosteranes also occur in Late Proterozoic and early Paleozoic sediments that are often enriched with acritarchs (microfossils of uncertain origin, some of which have been speculatively attributed to dinoflagellates or their direct ancestors), and this has led to the proposal that dinoflagellates are ancient and acquired dinosterol biosynthesis early in their evolution (35–37). We compared this hypothesis (Fig. 2C, H1) to a Mesozoic origin of the dinoflagellate dinosterol (Fig. 2C, H2) by contrasting our updated phylogeny of modern thecate and athecate dinoflagellates (Fig. 1). Dinosterol and other 4-methyl sterols are absent from all dinoflagellate relatives with known sterol profiles, including ciliates, perkinsids, apicomplexans, *Chromera* (e.g., refs. 41, 43), and *Virella*, but also *Oxyrrhis* (38) and *Amoebobrya*, which likely only acquires 4-methyl sterols from its host (39, 40). In core dinoflagellates, 4-methyl sterols are ubiquitous, but dinosterol itself is absent in three of their earliest branches: *Nocatilua*, *Amphidinium*, and the *Kareния*aeae (e.g., refs. 41–43). *Gyrodinium dominans*, likely another early core dinoflagellate (14), also lacks dinosterol (38). This suggests that dinosterol appeared first in the last common ancestor of Gymnodinaceae s.s., *Akashiwo*, and thecate dinoflagellates (although broader testing for its presence in early-branching dinoflagellates is needed). We suggest that pre-Mesozoic dinosterines are unlikely to originate from dinoflagellates for four reasons. First, dinosteranes from the Late Proterozoic and early Paleozoic greatly predate unambiguous dinoflagellate fossils, and dinosterol presence in modern species is restricted to close relatives of the thecates (Fig. 1), which originated in the early Mesozoic. Second, Paleozoic acritarch microfossils bear no demonstrable morphological similarity to dinoflagellates (26). Third, dinosterane prevalence in Paleozoic and Proterozoic samples is highly variable compared with Mesozoic samples (35). They seem to be entirely absent from the Carboniferous and Permian (35), a discontinuity that contrasts with their almost universal preservation in Mesozoic and younger sediments and species. Finally, small amounts of dinosterol are known from a modern species of diatom (44), and traces of dinosteranes are also present in Archean bitumens, where dinoflagellates could not have possibly existed (45). All this suggests that different organisms in different geological eras evolved dinosterol biosynthesis independently of dinoflagellates and that dinosterol production by certain acritarchs ended with their mid-Paleozoic extinction. We also note that the phylogenetic distance between the origin of dinosterol-producing thecates and the origin of modern thecate dinoflagellates (see Fig. 1) is consistent with the time lapse between the Early Triassic dinosterane increase and the appearance of modern thecate orders in the Early Jurassic sediments (Fig. 2C). We therefore suggest that abundant dinosteranes in some Scythian (Early Triassic) sediments predating the earliest thecate fossils (Middle Triassic) (35) are derived from athecate dinoflagellates alone, which gained the ability to produce dinosterols near the Permian/Triassic boundary and became abundant shortly after it (Fig. 2C, H2).

### Plastid Metabolism and Dependency.

**Plastid metabolism in nonphotosynthetic dinoflagellates.** Approximately half of the described dinoflagellate species are nonphotosynthetic and are traditionally considered to lack plastids. The other half contains a photosynthetic peridinin-pigmented plastid that, in some lineages, has been replaced by other types of plastids. The peridinin plastid was inherited from the plastid in the common photosynthetic ancestor of dinoflagellates and apicomplexans (46, 47), but whether cryptic, nonpigmented plastids have been retained in nonphotosynthetic dinoflagellates remains contentious: *Crypthecodinium* and *Oxyrrhis* appear to contain plastid-derived genes (48, 49), whereas *Hematodinium* lacks all traces of the organelle (50). We investigated whether plastid and cytosolic pathways for isoprenoid, tetrapyrrole, and fatty acid biosynthesis were present in two distinctly related nonphotosynthetic dinoflagellates, *N. scintillans* and *O. marina*, as well as in *Dinophysis acuminata*, a fundamentally nonphotosynthetic species that nevertheless carries kleptoplastids. For each metabolic enzyme in these pathways, we elaborated a single protein phylogeny and classified its origin as plastidic (in a clade with photosynthetic eukaryotes only), cytosolic (in a clade containing heterotrophic eukaryotes), or bacterial (in a clade with bacteria, putative recent horizontal transfer), a methodology informed by published localizations in model eukaryotes (e.g., ref. 51) and by in silico targeting predictions in selected proteins (Fig. 3 and SI Appendix, SI Materials and Methods).

All three investigated dinoflagellates contain an isoprenoid pathway of plastid origin (all seven enzymes are present in *Nocatilua* and *Dinophysis*) and lack the cytosolic pathway variant (Fig. 34). This is exemplified by their retention of cyanobacterial *IspC* enzymes (Fig. 3B), which branch among orthologs from photosynthetic dinoflagellates and other algae. Similarly, all three nonphotosynthetic dinoflagellates contain multiple components of the plastid tetrapyrrole pathway (an essentially complete enzyme set is present in *Nocatilua* and *Dinophysis*), but only two to three components of that in mitochondria and the cytosol. Comparing our data to the *Symbiodinium minutum* genome, we propose that a single tetrapyrrole pathway of a predominantly plastid origin that initiates from glutamate (Fig. 3A, GTR and GSA) is present in all core dinoflagellates, a feature typical of eukaryotic plastids [mitochondrial aminolevulinic acid synthase (ALA) synthase is present in the early-branching *Hematodinium*, *Oxyrrhis*, and *Perkinsus* (50)]. None of the three nonphotosynthetic dinoflagellates contain proteins for plastid fatty acid biosynthesis, suggesting that this pathway is dispensable in dinoflagellates in the absence of photosynthesis (Fig. 3f; FabI in *Dinophysis* is unusual; SI Appendix, SI Materials and Methods). Genes for plastid iron–sulfur cluster assembly (SufB, C, D), ferredoxin (Fd) redox system [i.e., Fd NADP+ reductase (FNR)], and triose phosphate membrane translocators (TPTs) are also present in the three species (SI Appendix, Table S5).
Plastid protein targeting and genome loss. We further investigated 56 protein sequences in *Noctiluca*, *Oxyrrhis*, and *Dinophysius* of a plastidic origin (*SI Appendix, Table S5*). Most are incomplete, but seven are complete (they contain a partial spliced leader at the 5′ terminus of the corresponding transcript), and another 28 carry an extension of more than 50 aa at their N terminus. Proteins from the latter two categories were tested for the presence of plastid-targeting peptides in silico, and 17 of them carry bipartite targeting signatures comprising signal and transit peptides (*SI Appendix, Table S6*). Thirteen of these contain a phenylalanine at or near the predicted signal peptide cleavage site, and three *Oxyrrhis* proteins contain a second transmembrane region, all characteristics of targeting to plastids but not to other subcellular compartments in dinoflagellates (52, 53). In silico predictions have limited accuracy, but the consistent presence of N-terminal extensions and signal peptides in proteins is congruent only with a plastidic origin. For example, cyanobacterial Fds in *Noctiluca* and *Dinophysius* with four conserved cysteine residues required for Fe-S formation as to their presence/absence and origin. The data suggest that *Oxyrrhis*, *Noctiluca*, and *Dinophysius* are metabolically dependent on plastids. Metabolic dependency (metabolite uptake) was summarized from the literature. (B) Maximum-likelihood phylogeny (IQ-Tree) reveals IspCs of cyanobacterial origin in nonphotosynthetic dinoflagellates and relatives (bold); ultrafast bootstraps at branches are shown (>50 shown; ≥95 highlighted; filled circles, 100). (C) Three grades in functional organization of core metabolic pathways in non-photosynthetic plastids and non-plastid variants of dinoflagellates (blue) and relatives (*"P"* represents parasites). (D) Model for evolutionary dependency on plastids in dinoflagellates and relatives, and which is applicable to other eukaryotes. Ancestral dependency (marked as "d") on plastid metabolism (loss of cytosolic isoprenoid biosynthesis; later reinforced by the loss of C4 tetrapyrrole biosynthesis in some taxa) led to retention of plastids in all free-living and many parasitic descendants. The dependency can be transferred onto a new plastidial symbiont (Kareniaceae) or host organism (in parasites dependent solely on host-derived metabolites); only the latter leads to an outright loss of the plastid. **Fig. 3.** Plastid metabolism and dependency in nonphotosynthetic dinoflagellates. (A) Phylogeny-driven reconstruction of plastid and nonplastid variants of core metabolism (isoprenoid, tetrapyrrole, and fatty acid biosynthesis) in genomes (marked as "G") or transcriptomes ("T") of dinoflagellates and relatives. Individual enzymes (*SI Appendix, Table S5*) were classified by protein phylogenies and color-coded as to their presence/absence and origin. The data suggest that *Oxyrrhis*, *Noctiluca*, and *Dinophysius* are metabolically dependent on plastids. Metabolite (Met.) uptake was summarized from the literature. (B) Maximum-likelihood phylogeny (IQ-Tree) reveals IspCs of cyanobacterial origin in nonphotosynthetic dinoflagellates and relatives (bold); ultrafast bootstraps at branches are shown (>50 shown; ≥95 highlighted; filled circles, 100). (C) Three grades in functional organization of core metabolic pathways in non-photosynthetic plastids and non-plastid variants of dinoflagellates (blue) and relatives (*"P"* represents parasites). (D) Model for evolutionary dependency on plastids in dinoflagellates and relatives, and which is applicable to other eukaryotes. Ancestral dependency (marked as "d") on plastid metabolism (loss of cytosolic isoprenoid biosynthesis; later reinforced by the loss of C4 tetrapyrrole biosynthesis in some taxa) led to retention of plastids in all free-living and many parasitic descendants. The dependency can be transferred onto a new plastidial symbiont (Kareniaceae) or host organism (in parasites dependent solely on host-derived metabolites); only the latter leads to an outright loss of the plastid.
Cryptic plastids for the biosynthesis of isoprenoid units, and *tiluca* cytosolic pathway variants, and plastid localization of homologs in phototrophs, one of which has genome data available (9). Additionally, distant relatives and also from closely related plastid fatty acid biosynthesis was retained only in apicomplexans and is absent in the plastids of apicomplexans (51). A comparison of amino acids remains insufficiently known in dinoflagellates and plastid tetrapyrroles and the evolution of bioluminescence. Several species of dinoflagellates are bioluminescent (63). In the photosynthetic species *Pyrocystis lunula*, the light-emitting compound luciferin has an open tetrapyrrole structure thought to be synthesized from the structurally similar chlorophyll a (64): the organism incorporates radioactively labeled chlorophyll precursors into chlorophyll and luciferin, suggesting that their biosynthesis is linked (65). However, other bioluminescent dinoflagellates like *Noctiluca*, *Proto- peridinium*, and certain *Pohkrillos* species are nonphotosynthetic (63) and not known to synthesize chlorophyll. The prediction that they acquire chlorophyll from their prey (66) is inconsistent with prey-independent bioluminescence in at least one of them, *Protoperidinium crassipes* (67). Our finding of the plastid tetrapyrrole pathway in *Noctiluca*, which also leads to the precursors of chlorophyll, offers an alternative explanation of luciferin presence: it may be obtained by biosynthesis rather than scavenging, at least in some species. The plastid tetrapyrrole pathway is apparently indispensable as a key requirement for heme synthesis in all core dinoflagellates (Fig. 3A), and could therefore account for luciferin production in any bioluminescent dinoflagellate, irrespective of the presence of photosynthesis. This biosynthesis scenario also opens the possibility that luciferin is not derived via chlorophyll per se, but via an earlier intermediate in its biogenesis, perhaps a chlorophylllike or chlorine-like tetrapyrrole. Although this remains to be tested experimentally, our finding of the plastid tetrapyrrole pathway supports the possibility that bioluminescence in nonphotosynthetic dinoflagellates relies on a biosynthetic machinery repurposed from heme and chlorophyll production.

**Character Evolution in Dinoflagellates.**

**Nuclear evolution: Stepwise horizontal gene gain.** Dinoflagellates have unique nuclei that have lost bulk nucleosomal DNA packaging, and instead condense DNA by using two types of basic proteins that are different from histones. Dinoflagellate/viral nucleoproteins (DNVPs) are similar to uncharacterized proteins from phycodnaviruses, are distributed in all dinoflagellates yet examined, and represent a family of basic proteins with high DNA-binding affinity (4). In contrast, dinoflagellate histone-like proteins (HLPs) are distributed in all dinoflagellates yet examined and represent a family of basic proteins with high DNA-binding affinity (4). In contrast, dinoflagellate histone-like proteins (HLPs) are distributed in all dinoflagellates yet examined and represent a family of basic proteins with high DNA-binding affinity (4). In contrast, dinoflagellate histone-like proteins (HLPs) are distributed in all dinoflagellates yet examined and represent a family of basic proteins with high DNA-binding affinity (4).
Evolution of histone-like proteins. Phylogeny of bacterial (HU-like) and dinoflagellate HLPs reveals a dinoflagellate-type histone-like protein, HLP-II, in early-branching core dinoflagellates. HLP-II has a mutually exclusive distribution with HLP-I (e.g., the characterized HC3 in C. cohnii, in bold). Further details are provided in SI Appendix, Fig. S3 and Table S4.

Fig. 4. Evolution of histone-like proteins. Phylogeny of bacterial (HU-like) and dinoflagellate HLPs reveals a dinoflagellate-type histone-like protein, HLP-II, in early-branching core dinoflagellates. HLP-II has a mutually exclusive distribution with HLP-I (e.g., the characterized HC3 in C. cohnii, in bold). Further details are provided in SI Appendix, Fig. S3 and Table S4.

Organellae evolution: Plastid reduction and mitochondrial cox3 split. Evidence of a dependency on plastids in nonphotosynthetic dinoflagellates (Fig. 3) corroborates earlier conclusions that the common ancestor of dinoflagellates and apicomplexans was photosynthetic (46) and dependent on plastid-generated isoprenoids (47). Our phylogeny also supports the prediction that more than a dozen descendant lineages of this dinoflagellate-apicomplexan ancestor have lost photosynthesis (46, 69). At least two parasites, Cryptosporidium and Hapalodinium, have lost the plastid outright, but this is not the case in other parasites and in any free-living lineages that have been investigated with sufficient detail (six independent transitions to heterotrophy). We thus posit that plastid loss in dinoflagellates and apicomplexans is less frequent than their retention after the loss of photosynthesis, and is limited to a few parasites (47). After the split with apicomplexans but at least by the time Amphidinium diverged, the dinoflagellate plastid acquired the photosynthetic carotenoid peridinin, peridinin–chlorophyll binding proteins, and a reduced, minicircular genome (6). Our results suggest that during this transition the plastid sufB and clpC genes (key barriers to plastid genome loss in apicomplexans) were relocalized to the nucleus in dinoflagellates. This made the dinoflagellate plastid genome dispensable in the absence of photosynthesis, likely explaining why all heterotrophic representatives studied to date appear to lack it. In at least four distantly related photosynthetic
dinoflagellates, the expression of plastid genes is accompanied by substitutional editing of corresponding mRNAs (Fig. 5) (7). The origin of plastid editing is, however, uncertain; it appeared some time after the divergence of apicomplexans and chromophorids (70) and possibly became more widespread after the divergence of *Amphidinium* (71), but pinpointing its origin more precisely will require an analysis on deep-branching photosynthetic dinoflagellates such as *Spatodinium pseudonoctiluca* (13).

The mtDNA in at least five lineages of core dinoflagellates including *Amphidinium* contains a unique feature: cox3 is split in the same region into two fragments that are trans-spliced at the RNA level (8, 72). The split is absent in *Hematodinium* and earlier diverging species, but, to our knowledge, its presence in the Noctilucales was not known until now (Fig. 5). We identified a cox3 contig in the *Noctiluca* transcriptome corresponding to a full-length protein (terminated by a canonical stop codon rare in the group; SI Appendix, SI Materials and Methods). Mapping individual RNA read pairs onto the contig demonstrated continuous transcription across the split region and provides no support for the existence of two transcripts and their trans-splicing. PCR amplification by using *Noctiluca* genomic DNA as a template produced a single product spanning both sides of the cox3 split, the identity of which was confirmed by sequencing (SI Appendix, SI Materials and Methods). Because the phylogenetic distribution and the unique character of the cox3 split are indicative of a single evolutionary origin, the uninterrupted cox3 in *Noctiluca* corroborates the early position of the Noctilucales among core dinoflagellates (Fig. 1).

**Character map: Framework for evolutionary and functional predictions.** By using parsimony, we reconstructed ancestral character states of major conserved morphological and molecular traits at different points of the dinoflagellate phylogeny (Fig. 5 and SI Appendix, SI Materials and Methods). Newly mapped transitions include the gain of 4-methyl sterols, dinosterol, nuclear HLP-I and II, and the mitochondrial cox3 split, the theca, and the gain of multiple paralogs of GH7 cellulases. Two additional transitions map at the common ancestor of *Amphidinium* and later-diverging taxa: the gain of condensed liquid crystalline chromosomes throughout the life cycle and the gain of a proteinaceous striated rod in the transverse flagellum, which produces a strongly pronounced flagellar wave (Fig. 5). The corresponding characteristics in the Noctilucales are little understood and understood as liquid crystalline chromosomes in one of their life stages, the trophont, are relaxed and the transversal flagellum in their gametes is trailing, wave-less, and contains only a thin filament in place of the striated rod (73). Detailed analysis is required to determine whether these states represent true evolutionarily intermediates or secondary modifications associated with the unusual morphology of this order. The origin of other dinoflagellate characteristics was established previously and is reinforced within our framework: gain of plastids, Rubisco CO form II and oligoU-tailing in plastid mRNAs before the split with apicomplexans (46), and the acquisition of spliced leader trans-splicing of mRNAs in their common ancestor with Perkinsids (Fig. 5). DVNPs, ubiquitous in the species in our dataset, are ancestral to dinoflagellates and associated with changes in protein:DNA ratio and genome size. The ancestor of syndinians and core dinoflagellates had a life stage with a shallow sulcus and cingulum (flagellar grooves), the latter dividing the cell into an upper epi- some and a lower hyposome, a transitional morphology between short flagellar grooves in *Oxyrrhis* and *Psammosa* and deeply en-graved perpendicular flagellar grooves in core dinoflagellates (Fig. 5). Altogether, most transitions map to the branch corresponding to the ancestor of core dinoflagellates, but other characteristics are scattered widely along the evolutionary backbone (Fig. 5). Thus, the ecological success of dinoflagellates has resulted from a series of independent changes to the morphology, metabolism, and molecular biology of their ancestors.

**Conclusions**

We used sequence data to illuminate dinoflagellate biology and evolution. Evidence from our multiprotein phylogenies resolves numerous issues relating to dinoflagellate relationships, provides strong support for the single origin of the theca, and helps reconcile several apparent contradictions in dinoflagellate fossil, biogeochemical, and molecular data (Figs. 1 and 2). The origin of the theca coincides with a radiation of cell wall-localized cellulases involved in cell division (Fig. 2B). Plastid biosynthetic pathways exist in the nonphotosynthetic *Noctiluca*, *Oxyrrhis*, and *Dinophysis*, and cytosolic pathway variants do not (Fig. 3). This suggests that all free-living dinoflagellates are metabolically dependent on plastids that have taken over important cellular functions, apparently early in the evolution of the group; plastidal tetrapyrrole biosynthesis may also explain the existence of bioluminescent luciferin in nonpigmented dinoflagellates. The origin of the liquid crystalline nuclei coincides with the acquisition of bacterial histone-like proteins, which occurred in two distinct evolutionary phases (Fig. 4), suggesting that horizontal gene transfers were the ultimate origin of key dinoflagellate features. By producing a map of the major transitions in the evolutionary history of dinoflagellates (Fig. 5), we provide a predication framework that will facilitate the investigation of many aspects of the group’s cell biology (nuclear organization, plastid evolution), molecular biology, and paleobiology.

**Materials and Methods**

RNA was extracted by RNAqueous kit or TRIzol Plus RNA kit. Paired-end 50-bp or 100-bp illumina sequence reads were generated and assembled in Trinity version 2 or as part of the Marine Microbial Eukaryote Transcriptome Sequencing Project pipeline (19). Phylogenetic matrices were prepared from alignments mapped in MAFFT version 7.215 stripped of hypervariable sites in Block Mapping and Gathering with Entropy version 1.1. Phylogenies were computed in IQ-Tree (1,000 ultrafast bootstrap), RaxML version 8 (300 nonparametric bootstrap), and Phylobayes (where applicable). Plastid targeting signals were analyzed in SignalP 4.1 (D-score cutoff 0.45) and ChloroP 1.1 at 0.45 cTP-score cutoff. Species culturing and sequencing, phylogenetic inferences, and analyses of plastid metabolism and protein targeting are detailed in SI Appendix, SI Materials and Methods.

**ACKNOWLEDGMENTS.** We thank Bill MacMillan for technical support and Patrick Keeling for facilities and support. This work was supported by a University College London Excellence Fellowship (to J.J.), a CIFAR Global Scholar Fellowship (to J.J.), a University of British Columbia Four-Year PhD Fellowship (to J.J.), Gordon and Betty Moore Foundation Grant 2637 to the National Center for Genome Resources (NCGR), National Science and Engineering Research Council of Canada Grant NSERC 2014-05258 (to B.S.L. and G.G.), a Tula Foundation Grant to Patrick Keeling (F.B.), the Centre for Microbial Biodiversity and Evolution (F.B.), Australian Research Council Grant DP1003395 (to S.G.G.), Science Foundation Ireland Grant 13/SGR/2125 (to S.G.G.), NSF Grant EF-0629625 (to C.F.D. and T.R.B.), and Canadian Institute for Health Research Grant MOP-42517 to Patrick Keeling. MMETSP samples were sequenced, assembled, and annotated at NCGR. This is ESS contribution no. 20160099.


Janouškovec et al.


