Supporting Information

Janouškovec et al.: Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics

SI Materials and Methods:

Dinoflagellate culturing, sequencing, and sequence assembly. *Noctiluca scintillans* SPMC136 (MMETSP0253) was grown on *Prorocentrum micans* CCMP691 (primary preferred prey) in filtered (0.2 µm) autoclaved seawater (30 psu) with a dilute trace metal amendment (1). Scaled-up cultures were captured on a 80 µm sieve and maintained on *Dunaliella tertiolecta* (secondary non-preferred prey) for 25 days to ensure a complete removal of *Prorocentrum* (which was visually absent by the day 15 following the transfer). *Noctiluca* cells were then re-captured on the sieve and their total RNA was extracted by using the RNAqueous kit (Ambion). *Togula jolla* CCCM725 (project MMETSP0224), *Protoceratium reticulatum* CCCM535 (project MMETSP0228), and *Polarella glacialis* CCMP1383 (project MMETSP0227) were grown in the natural seawater medium HESNW (2), and their RNA was purified by the TRIzol Plus RNA kit (Thermo Fisher). Sequencing libraries were built and transcriptomic reads generated, processed, and assembled at the National Center for Genome Resources as described previously (3). Assembled contigs and predicted proteomes were downloaded from the MMETSP website (http://data.imicrobe.us/project/view/104). A second independent assembly of *Noctiluca* reads was generated by using Trinity v2.0.6 at default settings; the resulting contigs were found to contain longer 5' regions compared to the MMETSP assembly and were used in the analysis of N-termini of plastid-targeted proteins. This transcriptomic assembly has been deposited in GenBank (TSA) under the accession GELK00000000. Each *Noctiluca* protein used in this study (extracted from either of the assemblies) was first screened by BLASTP against the predicted proteome of *Prorocentrum minimum* CCMP2233 (Table S1), and its affiliation to dinoflagellates was then verified in a Maximum likelihood phylogeny. No sequences of *Prorocentrum* and rare, well-identifiable sequences of *Dunaliella* were detected in the assemblies (all assembled nuclear ribosomal RNA contigs belong to *Noctiluca*). The transcriptome assembly of *Hematodinium* sp. was deposited in GenBank (TSA; GEMP00000000. Data from *Amphidinium carterae* and two *Amoebophrya* isolates were generated as described in (4, 5). Data from other species were obtained as detailed in Table S1.

Multiprotein phylogenies. Dinoflagellate sequences were added into alignments of conserved proteins that were previously used in eukaryotic phylogenies (6), and those with 30% or less of taxa missing were selected. The alignments were re-aligned by the 'localpair' algorithm in MAFFT v7.215 (7), stripped of hypervariable sites (-b 4 -g 0.4 settings) in BMGE v1.1 (8) and the orthology of sequences within was verified by comparing their RAxML v.8 (9) maximum likelihood phylogenies (LG + Gamma 4 + F model) with known relationships based on published phylogenies. Paralogous, highly divergent or contaminant sequences were identified in several species and removed; where ambiguous, all paralogs for a given species were removed, and where multiple ambiguities were identified, the whole gene alignment was discarded. Single protein alignments were concatenated in Scafos v1.25 (10) by using 'o=gclv gamma=yes l=1 m=1' settings. Chimeric sequences were created for species where overlapping fragments or non-overlapping fragments of a congruent phylogenetic position were recovered (Table S1). A total of 12 phylogenetic matrices were concatenated independently: three variants of the outgroup times four variants of species presence among thecate taxa (Table S1; Fig. 1). Maximum Likelihood phylogenies of the concatenated matrices were inferred in IQ-Tree v1.41 (11) by using the LG + I + GAMMA4 + F settings (-m TEST was run first to select this model) with 1000 ultrafast bootstraps, and RAxML by using the LG + GAMMA4 + F with 300 non-parametric bootstraps. Bayesian phylogenies were inferred in PhyloBayes MPI v1.5a (12) on CIPRESS Science Gateway (13) by using GTR + CAT + GAMMA4, -dc, and maxdiff<0.1 settings. Approximately unbiased (AU) and Expected likelihood weights (ELW) test scores for alternative tree topologies were computed in Consel (14) and IQ-Tree (11), respectively (Tables S2 and S3).

Theca evolution and dinosterol. Protein sequences of dCel1 and dCel2 cellulases (accessions in Table S4) were each used to retrieve 250 closest hits from the NCBI nr database, which were complemented by dinoflagellate sequences from our dataset (Table S1; primarily MMETSP and NCBI databases). The dataset was reduced to a smaller number of unique, phylogenetically representative sequences: sequences that were largely incomplete and sequences that were closely related to one another (including all sequences from *Kryptoperidinium foliaceum* CCAP 1116/3, a close relative of *K. foliaceum* CCMP1326) or formed very long branches in preliminary phylogenies were removed. The final phylogenetic matrix of the GH7 dataset (Fig. 2B and Fig. S1) contained 184 sequences and 260 amino acid sites and was prepared by an alignment in MAFFT and removal of hypervariable sites in BMGE (as above), and phylogenies were inferred in IQ-Tree, as described above (see Multiprotein phylogeny). Dinosterol distribution in dinoflagellates was mapped by surveying the available literature, in part by using the reference list at https://doi.pangaea.de/10.1594/PANGAEA.819698.

Plastid metabolism and protein targeting. Sequences of plastid and nuclear protein (Figs. 2-4) were identified by BLASTP searches in datasets listed in Table S1, in addition to transcriptomes from two *Oxyrrhis marina* strains (MMETSP1424-1426 and MMETSP0468-471 projects), and *Pyrodinium bahamense* (NCBI: PRJNA169246). Contaminant sequences were identified in several projects (e.g., *Oxyrrhis* MMETSP projects) and carefully removed based on phylogenetic incongruence. In the phylogenetic reconstruction of dependency on plastids in non-photosynthetic species, (Fig. 3A) each enzyme of pathways was analysed separately. New dinoflagellate sequences were included in single-protein alignments from an earlier study (15) and the protein origin was assessed by RAxML phylogenies (computed as above) or analyzed in newly prepared datasets (Fd, FNR, SufB, SufC, SufD, TPT; Table S5). The final phylogenetic matrix of IspC (Fig. 3B) contained 66 sequences and 331 amino acid sites and was computed in IQ-Tree as described above (see GH7 dataset preparation). Dinoflagellate FASI / PKS polyproteins were reconstructed (Fig. 3A) from mutually overlapping fragments comprising at least two domains (Table S5); domain order and functional specificity of mature FASI / PKS forms remain unknown, but both FAS and PKS are likely to be present (most individual domains exist in multiple sequence contexts). Atypical plastid FabI was identified in *Dinophysis acuminata* that was closely related to homologs in the Kareniaceae (Fig. 3A) but other proteins of the pathway were not. It is unclear whether this protein sequences may be a contaminant (other Kareniaceae-like proteins were found in the *Dinophysis* transcriptome), but it remains unlikely that *Dinophysis* possess the plastid fatty acid biosynthesis pathway. Plastid targeting signals in *Noctiluca*, *Dinophysis*, and *Oxyrrhis* were analysed in plastid proteins carrying N-terminal extensions as compared their bacterial orthologs. The most complete sequence for each protein was selected from the following: MMETSP-predicted proteins, proteins newly predicted from MMETSP-assembled contigs (*Oxyrrhis* and *Dinophysis*), or proteins predicted from newly assembled MMETSP reads (*Noctiluca* Trinity assembly; note that N-termini of some MMETSP-predicted proteins were incomplete). Proteins that screened positively for signal peptides in SignalP 4.1 (D-score cutoff 0.45) were further tested for the presence of transit peptides in ChloroP 1.1 at 0.45 cTP-score cutoff (16) and the strongest candidates for plastid targeting were listed in Table S6 (trans-membrane regions were predicted in TMHMM v2.0). The cleavage site of *Plasmodium falciparum* ferredoxin (Fig. S2) was predicted by PATS (http://gecco.org.chemie.uni-frankfurt.de/pats/pats-index.php), a species-specific tool for prediction of targeting pre-sequences. Partial *sufB* sequences were identified in *Noctiluca (*55% GC, c20274_g1_i1, and c20274_g2_i1 contigs in the Trinity assembly), *Oxyrrhis marina* MMETSP0468 *(*57.7% GC; contig CAMNT_0034061651), and *Dinophysis* (66.7% GC; contig CAMNT_0021013865). Plastid *clpC* fragments were identified in *Noctiluca* (55.4% GC; c23770_g4_i1 and c23770_g5_i1 contigs in the Trinity assembly), *Dinophysis* (69.1% GC; contig CAMNT_0020950785), and *Oxyrrhis marina* MMETSP0468 (60.8% GC; contig

CAMNT_0034034689).

Character evolution. Protein sequences of HLP-I and HLP-II (accessions in Table S4) were each used to retrieve 250 closest hits from the NCBI nr database; top hits among environmental and NCBI EST entries were also included. The final phylogenetic matrix (Fig. 4 and Fig. S3) contained 114 sequences and 99 amino acid sites and was prepared by adding sequences of dinoflagellates, removing closely related sequences, and alignment processing, and phylogenies were inferred in IQ-Tree, all as described above (see GH7 dataset preparation). The data presented in the character map were compiled from the literature and ancestral states were reconstructed by parsimony on the consensus of dinoflagellates relationships as established in this study (Fig. 1), taking known lower-level relationships into account (e.g., (17, 18)) . Transcripts of the three mitochondrion-localized protein-coding genes in *Noctiluca* were identified in the Trinity assembly by homology searches (*cox1* on the contig c33015_g1_i5, *cox3* on the contig c32288_g1_i2, and *cob* on the contig c32214_g1_i1). The *cox3* transcript was found to be complete and contain a canonical UAA stop codon at the expected position (i.e., one that is not generated by oligoadenylation of its 3' terminus as observed in other core dinoflagellates (19)); the only canonical stop codon in core dinoflagellates reported so far is in the *cob* of *Symbiodinium minutum*; (20)). Trascriptomic paired-end Illumina 50bp reads were mapped onto the assembled *cox3* contig by using Bowtie2 (v2.2.9); reads mapped continually across the region where the split occurs in other core dinoflagellates with paired-end reads connecting both sides of the split (no indication of two separate RNA fragments, trans-splicing, or oligoA tailing was observed). PCR corresponding to a near-full length *cox3* spanning both sides of the split was done by using Pfu polymerase, Rnase-treated genomic DNA of *Noctiluca* and specific primers. The reaction yielded a single product of the correct size (no product was observed in 'Dnase-treated template' and 'no template' controls) and the sequence of this product corresponded to *Noctiluca cox3*. Polymorphism at multiple sites was observed in the chromatogram and the DNA consensus differed in several nucleotides from the transcriptomic contig, where we also observed extensive polymorphism by read mapping: thus, the number of and variation among *cox3* copies and whether editing of their transcripts is present remain to be established.

SI References:

1. Gifford DJ (1985) Laboratory culture of marine planktonic oligotrichs(Ciliophora, Oligotrichida). Mar Ecol Prog Ser 23(3):257–267.

2. Harrison PJ, Waters RE, Taylor FJR (1980) A Broad Spectrum Artificial Sea Water Medium for Coastal and Open Ocean Phytoplankton1. J Phycol 16(1):28–35.

3. Keeling PJ, et al. (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(6):e1001889.

4. Bachvaroff TR, et al. (2014) Dinoflagellate phylogeny revisited: Using ribosomal proteins to resolve deep branching dinoflagellate clades. Mol Phylogenet Evol 70:314–322.

5. Gornik SG, et al. (2015) Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. Proc Natl Acad Sci 112(18):5767–5772.

6. Burki F, Okamoto N, Pombert J-F, Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc R Soc B Biol Sci 279(February):2246–2254.

7. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780.

8. Criscuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new

software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol 10(1):210.

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

10. Roure B, Rodriguez-Ezpeleta N, Philippe H (2007) SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Evol Biol 7 Suppl 1:S2–S2.

11. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32(1):268–274.

12. Lartillot N, Rodrigue N, Stubbs D, Richer J (2013) PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. Syst Biol 62(4):611–615.

13. Miller MA, Pfeiffer W, Schwartz T (2011) The CIPRES science gateway: a community resource for phylogenetic analyses. Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery (ACM), p 41.

14. Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17(12):1246–1247.

15. Janouškovec J, et al. (2015) Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. Proc Natl Acad Sci U S A 112(33):10200–7.

16. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2(4):953–71.

17. Orr RJS, Murray SA, Stüken A, Rhodes L, Jakobsen KS (2012) When naked became armored: an eight-gene phylogeny reveals monophyletic origin of theca in dinoflagellates. PLoS ONE 7(11):e50004.

18. Saldarriaga J (2004) Molecular data and the evolutionary history of dinoflagellates. Eur J Protistol 40(1):85–111.

19. Jackson CJ, et al. (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. BMC Biol 5:41–41.

20. Shoguchi E, Shinzato C, Hisata K, Satoh N, Mungpakdee S (2015) The Large Mitochondrial Genome of Symbiodinium minutum Reveals Conserved Noncoding Sequences between Dinoflagellates and Apicomplexans. Genome Biol Evol 7(8):2237–2244.

SI Figures and Tables:

Fig. S1: Phylogeny of cellulases of the Glycosyl Hydrolase 7 family (next page). Analysis of Glycosyl hydrolase family 7 (GH7) reveals cellulases in athecate dinoflagellates (bold) and their radiation in the thecates (color-coded). Unrooted best Maximum likelihood tree (IQ-Tree) was inferred from 119 dinoflagellate and 65 non-dinoflagellate eukaryotic protein sequences limited to taxonomically representative, near full-length, and slow-evolving entries. GH7 cellulases from *Pyrocystis lunula* (dCel1) and *Lingulodinium polyedrum* (dCel2) are highlighted. Eight putative paralogs are shown that were likely present in the ancestor of all living thecate orders are shown by vertical arrows; the exact number of ancestral paralogs is difficult to predict due to the low resolution of the tree, apparent incomplete lineage sorting and putative horizontal gene transfer (see, e.g., *Bigelowiella natans*), but was likely even higher - see 13-16 paralogous forms in the Gonyaulacales only.

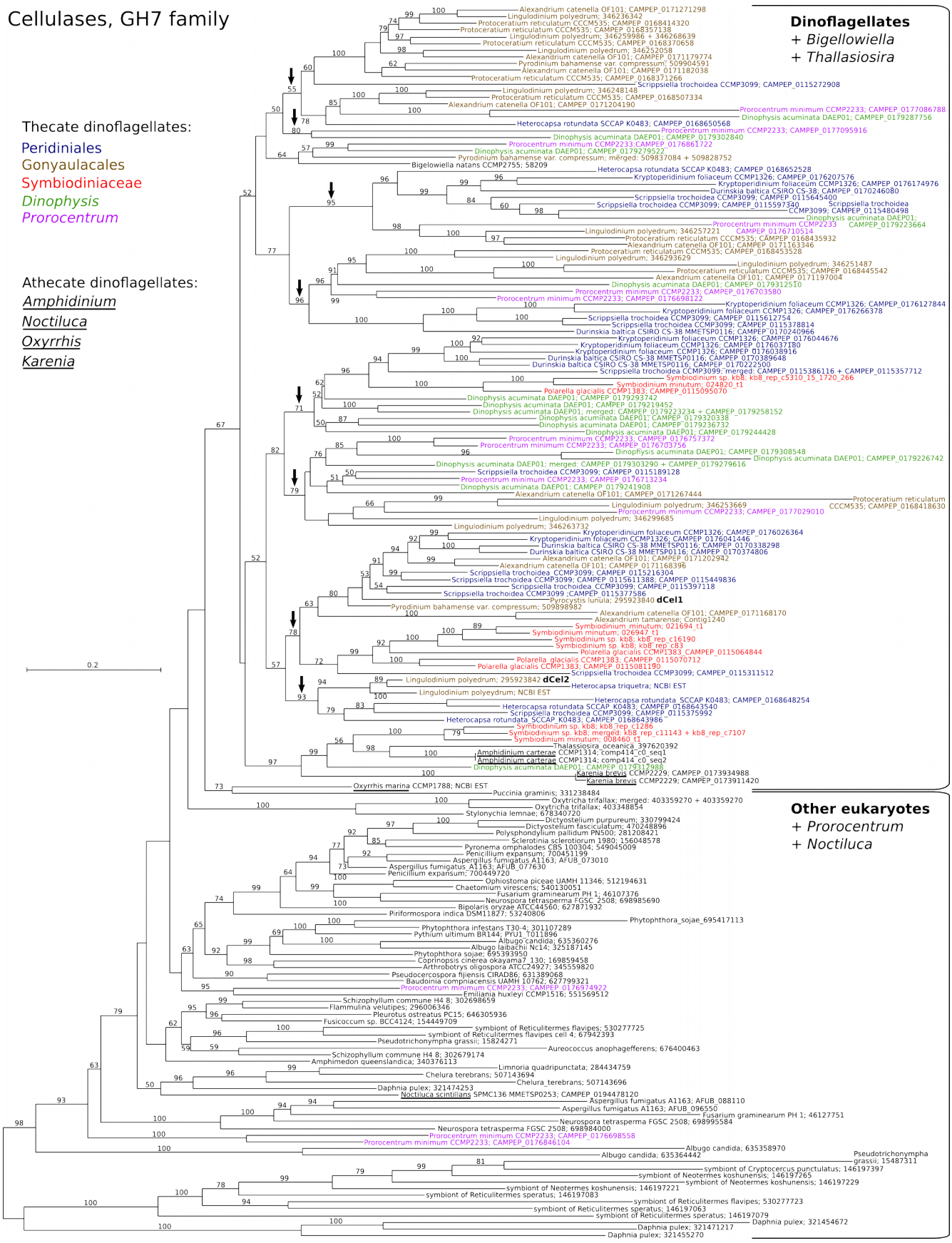
Cellulases, GH7 family
Thecate dinoflagellates:
Peridiniales
Gonyaulacales
Symbiodiniaceae
Dinophysis
Prorocentrum
Athecate dinoflagellates:
Amphidinium
Noctiluca
Oxyrrhis
Karenia
0.2
Dinoflagellates
+ Bigellowiella
+ Thallasiosira
Other eukaryotes
+ Prorocentrum
+ Noctiluca
Alexandrium catenella OF101; CAMPEP_0171271298
Linguladinium polyedrum; 346236342
Protoceratium reticulatum CCCM535; CAMPEP_0168414320
Protoceratium reticulatum CCCM535; CAMPEP_0168357138
Linguladinium polyedrum; 346259986 + 346268639
Protoceratium reticulatum CCCM535; CAMPEP_0168370658
Linguladinium polyedrum; 346252058
Alexandrium catenella OF101; CAMPEP_0171179774
Pyrodinium bahamense var. compressum; 509904591
Alexandrium catenella OF101; CAMPEP_0171182038
Protoceratium reticulatum CCCM535; CAMPEP_0168371266
Scrippsiella trochoidea CCMP3099; CAMPEP_0115272908
Linguladinium polyedrum; 346248148
Protoceratium reticulatum CCCM535; CAMPEP_0168507334
Alexandrium catenella OF101; CAMPEP_0171204190
Procentrum minimum CCMP2233; CAMPEP_0177086788
Dinophysis acuminata DAEP01; CAMPEP_0179287756
Heterocapsa rotundata SCCAP K0483; CAMPEP_0168650568
Procentrum minimum CCMP2233; CAMPEP_0177095916
Dinophysis acuminata DAEP01; CAMPEP_0179302840
Prorocentrum minimum CCMP2233; CAMPEP_0176861722
Dinophysis acuminata DAEP01; CAMPEP_0179279522
Pyrodinium bahamense var. compressum; merged: 509837084 + 509828752
Bigelowiella natans CCMP2755; 58209
Heterocapsa rotundata SCCAP K0483; CAMPEP_0168652528
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176207576
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176174976
Durinskia baltica CSIRO CS-38; CAMPEP_0170246080
Scrippsiella trochoidea CCMP3099; CAMPEP_0115645400
Scrippsiella trochoidea CCMP3099; CAMPEP_0115597340
Scrippsiella trochoidea CCMP3099; CAMPEP_0115480498
Dinophysis acuminata DAEP01;
Procentrum minimum CCMP2233; CAMPEP_0179223664
Linguladinium polyedrum; 346257221 CAMPEP_0176710514
Protoceratium reticulatum CCCM535; CAMPEP_0168435932
Alexandrium catenella OF101; CAMPEP_0171163346
Protoceratium reticulatum CCCM535; CAMPEP_0168453528
Linguladinium polyedrum; 346293629
Linguladinium polyedrum; 346251487
Protoceratium reticulatum CCCM535; CAMPEP_0168445542
Alexandrium catenella OF101; CAMPEP_0171197004
Dinophysis acuminata DAEP01; CAMPEP_0179312510
Procentrum minimum CCMP2233; CAMPEP_0176703580
Prorocentrum minimum CCMP2233; CAMPEP_0176698122
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176127844
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176266378
Scrippsiella trochoidea CCMP3099; CAMPEP_0115612754
Scrippsiella trochoidea CCMP3099; CAMPEP_0115378814
Durinskia baltica CSIRO CS-38 MMETSP0116; CAMPEP_0170240966
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176044676
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176037180
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176038916
Durinskia baltica CSIRO CS-38 MMETSP0116; CAMPEP_0170389648
Durinskia baltica CSIRO CS-38 MMETSP0116; CAMPEP_0170222500
Scrippsiella trochoidea CCMP3099; merged: CAMPEP_0115386116 + CAMPEP_0115357712
Symbiodinium sp. kb8; kb8_rep_c5310_15_1720_266
Symbiodinium minutum; 024820_t1
Polarella glacialis CCMP1383; CAMPEP_0115095070
Dinophysis acuminata DAEP01; CAMPEP_0179293742
Dinophysis acuminata DAEP01; CAMPEP_0179219452
Dinophysis acuminata DAEP01; merged: CAMPEP_0179233234 + CAMPEP_0179258152
Dinophysis acuminata DAEP01; CAMPEP_0179320338
Dinophysis acuminata DAEP01; CAMPEP_0179236732
Dinophysis acuminata DAEP01; CAMPEP_0179244428
Procentrum minimum CCMP2233; CAMPEP_0176757372
Procentrum minimum CCMP2233; CAMPEP_0176703756
Dinophysis acuminata DAEP01; CAMPEP_0179308548
Dinophysis acuminata DAEP01; CAMPEP_0179226742
Dinophysis acuminata DAEP01; merged: CAMPEP_0179303290 + CAMPEP_0179279616
Scrippsiella trochoidea CCMP3099; CAMPEP_0115189128
Prorocentrum minimum CCMP2233; CAMPEP_0176713234
Dinophysis acuminata DAEP01; CAMPEP_0179241908
Alexandrium catenella OF101; CAMPEP_0171267444
Protoceratium reticulatum CCCM535; CAMPEP_0168418630
Linguladinium polyedrum; 346253669
Prorocentrum minimum CCMP2233; CAMPEP_0177029010
Linguladinium polyedrum; 346299685
Linguladinium polyedrum; 346263732
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176026364
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176041446
Durinskia baltica CSIRO CS-38 MMETSP0116; CAMPEP_0170338298
Durinskia baltica CSIRO CS-38 MMETSP0116; CAMPEP_0170374806
Alexandrium catenella OF101; CAMPEP_0171202942
Alexandrium catenella OF101; CAMPEP_0171168396
Scrippsiella trochoidea CCMP3099; CAMPEP_0115216304
Scrippsiella trochoidea CCMP3099; CAMPEP_0115611388; CAMPEP_0115449836
Scrippsiella trochoidea CCMP3099; CAMPEP_0115397118
Scrippsiella trochoidea CCMP3099; CAMPEP_0115377586
Pyrocystis lunula; 295923840 dCel1
Pyrodinium bahamense var. compressum; 509898982
Alexandrium catenella OF101; CAMPEP_0171168170
Alexandrium tamarense; Contig1240
Symbiodinium minutum; 021694_t1
Symbiodinium minutum; 026947_t1
Symbiodinium sp. kb8; kb8_rep_c16190
Symbiodinium sp. kb8; kb8_rep_c83
Polarella glacialis CCMP1383; CAMPEP_0115064844
Polarella glacialis CCMP1383; CAMPEP_0115070712
Polarella glacialis CCMP1383; CAMPEP_0115081190
Scrippsiella trochoidea CCMP3099; CAMPEP_0115311512
Linguladinium polyedrum; 295923842 dCel2
Heterocapsa triquetra; NCBI EST
Linguladinium polyeydrum; NCBI EST
Heterocapsa rotundata SCCAP K0483; CAMPEP_0168648254
Heterocapsa rotundata SCCAP K0483; CAMPEP_0168643540
Scrippsiella trochoidea CCMP3099; CAMPEP_0115375992
Heterocapsa rotundata SCCAP K0483; CAMPEP_0168643986
Symbiodinium sp. kb8; kb8_rep_c1286
Symbiodinium sp. kb8; merged: kb8_rep_c11143 + kb8_rep_c7107
Symbiodinium minutum; 008460_t1
Thalassiosira oceanica 397620392
Amphidinium carterae CCMP1314; comp414_c0_seq1
Amphidinium carterae CCMP1314; comp414_c0_seq2
Dinophysis acuminata DAEP01; CAMPEP_0179312988
Karenia brevis CCMP2229; CAMPEP_0173934988
Karenia brevis CCMP2229; CAMPEP_0173911420
Oxyrrhis marina CCMP1788; NCBI EST
Puccinia graminis; 331238484
Oxytricha trifallax; merged: 403359270 + 403359270
Oxytricha trifallax; 403348854
Stylonychia lemnae; 678340720
Dictyostelium purpureum; 330799424
Dictyostelium fasciculatum; 470248896
Polysphondylium pallidum PN500; 281208421
Sclerotinia sclerotiorum 1980; 156048578
Pyronema omphalodes CBS 100304; 549045009
Penicillium expansum; 700451199
Aspergillus fumigatus A1163; AFUB_073010
Aspergillus fumigatus_A1163; AFUB_077630
Penicillium expansum; 700449720
Ophiostoma piceae UAMH 11346; 512194631
Chaetomium virescens; 540130051
Fusarium graminearum PH 1; 46107376
Neurospora tetrasperma FGSC 2508; 698985690
Bipolaris oryzae ATCC44560; 627871932
Piriformospora indica DSM11827; 53240806
Phytophthora_sojae_695417113
Phytophthora infestans T30-4; 301107289
Pythium ultimum BR144; PYU1_T0118896
Albugo candida; 635360276
Albugo laibachii Nc14; 325187145
Phytophthora sojae; 695393950
Coprinopsis cinerea okayama7 130; 169859458
Arthrobotrys oligospora ATCC24927; 345559820
Pseudocercospora fijiensis CIRAD86; 631389068
Baudoinia compniacensis UAMH 10762; 627799321
Prorocentrum minimum CCMP2233; CAMPEP_0176974922
Emiliania huxleyi CCMP1516; 551569512
Schizophyllum commune H4 8; 302698659
Flammulina velutipes; 296006346
Pleurotus ostreatus PC15; 646305936
Fusicoccum sp. BCC4124; 154449709
symbiont of Reticulitermes flavipes; 530277725
symbiont of Reticulitermes flavipes cell 4; 67942393
Pseudotrichonympha grassii; 15824271
Aureococcus anophagefferens; 676400463
Schizophyllum commune H4 8; 302679174
Amphimedon queenslandica; 340376113
Limnoria quadripunctata; 284434759
Chelura terebrans; 507143694
Chelura_terebrans; 507143696
Daphnia pulex; 321474253
Noctiluca scintillans SPMC136 MMETSP0253; CAMPEP_0194478120
Aspergillus fumigatus A1163; AFUB_088110
Aspergillus fumigatus A1163; AFUB_096550
Fusarium graminearum PH 1; 46127751
Neurospora tetrasperma FGSC 2508; 698985584
Neurospora tetrasperma FGSC 2508; 698984000
Prorocentrum minimum CCMP2233; CAMPEP_0176698558
Prorocentrum minimum CCMP2233; CAMPEP_0176846104
Albugo candida; 635358970
Albugo candida; 635364442
Pseudotrichonympha grassii; 15487311
symbiont of Cryptocercus punctulatus; 146197397
symbiont of Neotermes koshunensis; 146197265
symbiont of Neotermes koshunensis; 146197229
symbiont of Neotermes koshunensis; 146197221
symbiont of Reticulitermes speratus; 146197083
symbiont of Reticulitermes flavipes; 530277723
symbiont of Reticulitermes speratus; 146197063
symbiont of Reticulitermes speratus; 146197079
Daphnia pulex; 321454672
Daphnia pulex; 321471217
Daphnia pulex; 321455270

Fig. S2: *In silico* targeting predictions in plastid ferredoxins (see other plastid-targeted proteins in Table S6). Protein N-termini are shown to scale and show signal peptide D-scores (SignalP) and transit peptide cTP-scores (TargetP) and four conserved cysteines required for Fe-S formation. Spliced leaders or their fragments (SL=length in nucleotides) at the 5' end of dinoflagellate mRNAs indicate N-complete proteins. *Porphyra* ferredoxin is plastid-genome encoded and does not require targeting pre-sequences. The plastid ferredoxin in *Oxyrrhis* is truncated but contains an N-terminal extension carrying (a near-complete or complete) transit peptide region suggesting that the protein is targeted to the plastid like in other dinoflagellates.

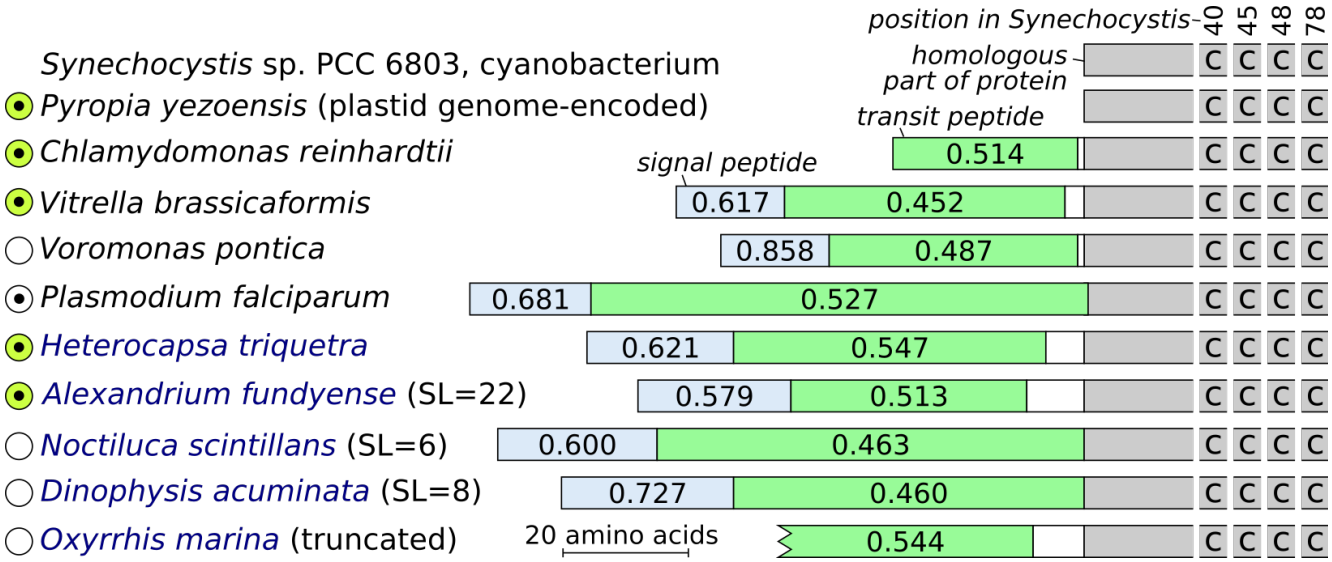

Fig. S3: A novel dinoflagellate histone-like protein (next page). Phylogeny of bacterial (HU-like) and dinoflagellate histone-like proteins (HLP) reveals a novel dinoflagellate type (HLP-II), which has a mutually exclusive distribution with HLP-I. Best Maximum likelihood tree (IQ-Tree) with ultrafast bootstrap supports at branches (only >50 supports are shown). Several environmental sequences (ENV) including those derived from dinoflagellate spliced-leader libraries (ENV dinoSL) are shown. The previously characterized HCc3 in *Crypthecodinium cohnii* (HLP-I) is highlighted in bold. Other histone-like proteins in eukaryotes (e.g., HU-like proteins in the plastid) are not closely related to either of the dinoflagellate HLP forms and were not included in the phylogeny.

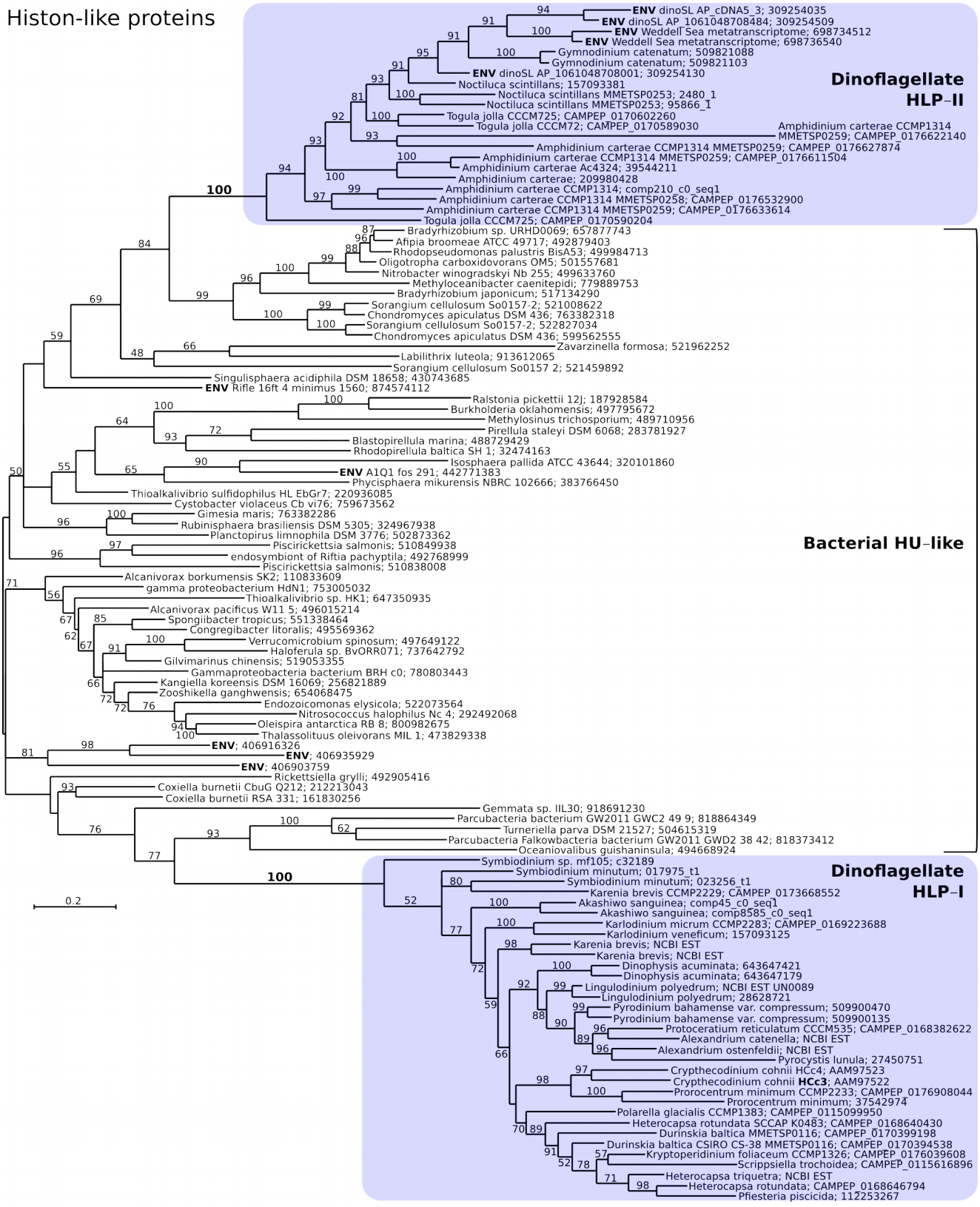
Histon-like proteins
Dinoflagellate HLP–II
ENV dinoSL AP_cDNA5_3; 309254035
ENV dinoSL AP_1061048708484; 309254509
ENV Weddell Sea metatranscriptome; 698734512
ENV Weddell Sea metatranscriptome; 698736540
Gymnodinium catenatum; 509821088
Gymnodinium catenatum; 509821103
ENV dinoSL AP_1061048708001; 309254130
Noctiluca scintillans; 157093381
Noctiluca scintillans MMETSP0253; 2480_1
Noctiluca scintillans MMETSP0253; 95866_1
Togula jolla CCCM725; CAMPEP_0170602260
Togula jolla CCCM72; CAMPEP_0170589030
Amphidinium carterae CCMP1314 MMETSP0259; CAMPEP_0176622140
Amphidinium carterae CCMP1314 MMETSP0259; CAMPEP_0176627874
Amphidinium carterae CCMP1314 MMETSP0259; CAMPEP_0176611504
Amphidinium carterae Ac4324; 39544211
Amphidinium carterae; 209980428
Amphidinium carterae CCMP1314; comp210_c0_seq1
Amphidinium carterae CCMP1314 MMETSP0258; CAMPEP_0176532900
Amphidinium carterae CCMP1314 MMETSP0259; CAMPEP_0176633614
Togula jolla CCCM725; CAMPEP_0170590204
Bacterial HU–like
Bradyrhizobium sp. URHD0069; 657877743
Afipia broomeae ATCC 49717; 492879403
Rhodopseudomonas palustris BisA53; 499984713
Oligotropha carboxidovorans OM5; 501557681
Nitrobacter winogradskyi Nb 255; 499633760
Methyloceanibacter caenitepidi; 779889753
Bradyrhizobium japonicum; 517134290
Sorangium cellulosum So0157-2; 521008622
Chondromyces apiculatus DSM 436; 763382318
Sorangium cellulosum So0157-2; 522827034
Chondromyces apiculatus DSM 436; 599562555
Zavarzinella formosa; 521962252
Labilithrix luteola; 913612065
Sorangium cellulosum So0157 2; 521459892
Singulisphaera acidiphila DSM 18658; 430743685
ENV Rifle 16ft 4 minimus 1560; 874574112
Ralstonia pickettii 12J; 187928584
Burkholderia oklahomensis; 497795672
Methylosinus trichosporium; 489710956
Pirellula staleyi DSM 6068; 283781927
Blastopirellula marina; 488729429
Rhodopirellula baltica SH 1; 32474163
Isosphaera pallida ATCC 43644; 320101860
ENV A1Q1 fos 291; 442771383
Phycisphaera mikurensis NBRC 102666; 383766450
Thioalkalivibrio sulfidophilus HL EbGr7; 220936085
Cystobacter violaceus Cb vi76; 759673562
Gimesia maris; 763382286
Rubinisphaera brasiliensis DSM 5305; 324967938
Planctopirus limnophila DSM 3776; 502873362
Piscirickettsia salmonis; 510849938
endosymbiont of Riftia pachyptila; 492768999
Piscirickettsia salmonis; 510838008
Alcanivorax borkumensis SK2; 110833609
gamma proteobacterium HdN1; 753005032
Thioalkalivibrio sp. HK1; 647350935
Alcanivorax pacificus W11 5; 496015214
Spongiibacter tropicus; 551338464
Congregibacter litoralis; 495569362
Verrucomicrobium spinosum; 497649122
Haloferula sp. BvORR071; 737642792
Gilvimarinus chinensis; 519053355
Gammaproteobacteria bacterium BRH c0; 780803443
Kangiella koreensis DSM 16069; 256821889
Zooshikella ganghwensis; 654068475
Endozoicomonas elysicola; 522073564
Nitrosococcus halophilus Nc 4; 292492068
Oleispira antarctica RB 8; 800982675
Thalassolituus oleivorans MIL 1; 473829338
ENV; 406916326
ENV; 406935929
ENV; 406903759
Rickettsiella grylli; 492905416
Coxiella burnetii CbuG Q212; 212213043
Coxiella burnetii RSA 331; 161830256
Gemmata sp. IIL30; 918691230
Parcubacteria bacterium GW2011 GWC2 49 9; 818864349
Turneriella parva DSM 21527; 504615319
Parcubacteria Falkowbacteria bacterium GW2011 GWD2 38 42; 818373412
Oceaniovalibus guishaninsula; 494668924
Dinoflagellate HLP–I
Symbiodinium sp. mf105; c32189
Symbiodinium minutum; 017975_t1
Symbiodinium minutum; 023256_t1
Karenia brevis CCMP2229; CAMPEP_0173668552
Akashiwo sanguinea; comp45_c0_seq1
Akashiwo sanguinea; comp8585_c0_seq1
Karlodinium micrum CCMP2283; CAMPEP_0169223688
Karlodinium veneficum; 157093125
Karenia brevis; NCBI EST
Karenia brevis; NCBI EST
Dinophysis acuminata; 643647421
Dinophysis acuminata; 643647179
Lingulodinium polyedrum; NCBI EST UN0089
Lingulodinium polyedrum; 28628721
Pyrodinium bahamense var. compressum; 509900470
Pyrodinium bahamense var. compressum; 509900135
Protoceratium reticulatum CCCM535; CAMPEP_0168382622
Alexandrium catenella; NCBI EST
Alexandrium ostenfeldii; NCBI EST
Pyrocystis lunula; 27450751
Crypthecodinium cohnii HCc4; AAM97523
Crypthecodinium cohnii HCc3; AAM97522
Prorocentrum minimum CCMP2233; CAMPEP_0176908044
Prorocentrum minimum; 37542974
Polarella glacialis CCMP1383; CAMPEP_0115099950
Heterocapsa rotundata SCCAP K0483; CAMPEP_0168640430
Durinskia baltica MMETSP0116; CAMPEP_0170399198
Durinskia baltica CSIRO CS-38 MMETSP0116; CAMPEP_0170394538
Kryptoperidinium foliaceum CCMP1326; CAMPEP_0176039608
Scrippsiella trochoidea; CAMPEP_0115616896
Heterocapsa triquetra; NCBI EST
Heterocapsa rotundata; CAMPEP_0168646794
Pfiesteria piscicida; 112253267
0.2

Table S1: Sequence sources and phylogenetic matrices used. Species and data presence in final concatenated phylogenetic matrices (Root 1-3; Fig. 1) with sequence sources: matrix size (sites #), percentage of missing sites (%MS), genes (%MG), and merged chimeric entries (%CH; Materials and Methods).

Group	Operational taxonomic unit (Fig. 1)	Root 1 sites #	% MS	% MG	% CH	Root 2 sites #	% MS	% MG	% CH	Root 3 sites #	% MS	% MG	% CH	Data source
Dinoflagellates	Akashiwo sanguinea	27237	7	2	4	28820	6	2	4	28779	7	2	4	Bachvaroff et al., 2014 MPE
Dinoflagellates	Alexandrium spp. (A. tamarense, A. catenella OF101, A. minutum, A. ostenfeldii)	28346	4	1	5	29619	4	1	3	29857	4	1	3	NCBI EST; MMETSP0790
Dinoflagellates	Amoebophrya sp. ex Akashiwo	26900	9	6	0	28385	8	6	0	---	---	---	---	Bachvaroff et al., 2014 MPE
Dinoflagellates	Amoebophrya sp. ex Karlodinium	25811	12	3	0	27025	12	3	0	---	---	---	---	Bachvaroff et al., 2014 MPE
Dinoflagellates	Amphidinium carterae CCMP1314	28630	3	0	0	29931	3	0	0	30074	3	0	0	MMETSP0258-MMETSP0259; Bachvaroff et al., 2014 MPE
Dinoflagellates	Dinophysis acuminata DAEP01	25106	15	2	29	26629	13	2	30	26904	13	2	30	MMETSP0797
Dinoflagellates	Durinskia baltica CSIRO CS-38	26195	11	2	3	27245	11	2	2	27338	12	2	2	MMETSP0116-MMETSP0117
Dinoflagellates	Gymnodinium catenatum GC744	28347	4	2	5	29929	3	2	4	30236	2	2	4	MMETSP0784
Dinoflagellates	Hematodinium sp. SG-2012 ex Nephrops	28763	2	2	0	29746	3	2	0	---	---	---	---	NCBI TSA: GEMP00000000
Dinoflagellates	Heterocapsa spp. (H. triquetra CCMP449; H. rotundata SCCAP K-0483)	26353	10	5	21	27614	10	5	21	27799	10	5	21	NCBI EST; MMETSP0503
Dinoflagellates	Karenia brevis (CCMP2229, Wilson)	28973	1	1	1	30373	1	1	1	30567	1	1	1	NCBI EST; MMETSP0027, MMETSP0029-MMETSP0031
Dinoflagellates	Karlodinium veneficum (CCMP2283, CCMP 415, CCMP1974)	29119	1	0	4	30615	1	0	4	30816	1	0	4	NCBI EST; Bachvaroff et al., 2014 MPE; MMETSP1015-MMETSP1017
Dinoflagellates	Kryptoperidinium foliaceum CCAP 1116/3	25472	13	5	3	26649	13	5	2	26986	13	5	3	MMETSP0118-MMETSP0119
Dinoflagellates	Kryptoperidinium foliaceum CCMP 1326	26666	9	3	1	27709	10	3	1	28055	9	3	1	MMETSP0120-MMETSP0121
Dinoflagellates	Lingulodinium polyedrum	27323	7	0	18	28520	7	0	18	28515	8	0	17	NCBI TSA: GABP00000000
Dinoflagellates	Noctiluca scintillans SPMC136	27743	6	2	4	28787	6	2	3	29032	6	2	3	MMETSP0253; NCBI TSA: GELK00000000
Dinoflagellates	Oxyrrhis marina (CCMP1788, 44_PLY01)	11227	62	40	0	11114	64	41	0	---	---	---	---	NCBI EST; Lowe et al., 2012
Dinoflagellates	Polarella glacialis CCMP1383	27590	6	1	12	29182	5	1	12	29304	5	1	12	MMETSP0227
Dinoflagellates	Polykrikos lebourae	10500	64	30	20	10641	65	30	19	10637	66	30	19	Gavelis et al., 2015
Dinoflagellates	Prorocentrum minimum CCMP2233	27625	6	2	8	29208	5	2	8	29518	5	2	8	MMETSP0267-MMETSP0269
Dinoflagellates	Protoceratium reticulatum CCCM535	27388	7	0	4	28527	7	0	4	28762	7	0	4	MMETSP0228
Dinoflagellates	Scrippsiella trochoidea CCMP3099	28580	3	0	6	30275	2	0	6	30340	2	0	5	MMETSP0270-MMETSP0272
Dinoflagellates	Symbiodinium minutum Mf1.05b	21137	28	12	1	21589	30	12	1	21754	30	12	1	Symbiodinium minutum genome database
Dinoflagellates	Symbiodinium sp. CassKB8	25884	12	3	0	26788	13	3	0	26726	14	3	0	NCBI SRA: SRX076696; http://medinalab.org/zoox/
Dinoflagellates	Togula jolla CCCM725	29018	1	0	2	30353	1	0	2	30617	1	0	2	MMETSP0224
Perkinsids	Perkinsus marinus	27188	8	0	0	28253	8	0	0	---	---	---	---	NCBI NR
Apicomplexans	Babesia bovis	26932	8	5	0	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Babesia microti	25736	12	10	0	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Cryptosporidium muris	26442	10	6	0	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Cryptosporidium parvum	25233	14	8	0	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Eimeria tenella	20938	29	16	2	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Plasmodium falciparum	27311	7	4	0	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Theileria annulata	23842	19	11	0	---	---	---	---	---	---	---	---	NCBI NR
Apicomplexans	Toxoplasma gondii	26623	9	5	0	---	---	---	---	---	---	---	---	NCBI NR
Ciliates	Ichthyophthirius multifiliis	25145	14	13	0	---	---	---	---	---	---	---	---	NCBI NR
Ciliates	Paramecium tetraurelia	24211	18	21	0	---	---	---	---	---	---	---	---	NCBI NR
Ciliates	Oxytricha trifallax	25367	14	13	0	---	---	---	---	---	---	---	---	NCBI NR
Ciliates	Tetrahymena thermophila	26960	8	6	0	---	---	---	---	---	---	---	---	NCBI NR
Stramenopiles	Aureococcus anophageferrens	27214	7	4	0	---	---	---	---	---	---	---	---	NCBI NR
Stramenopiles	Ectocarpus siliculosus	28646	3	0	0	---	---	---	---	---	---	---	---	NCBI NR
Stramenopiles	Saprolegnia parasitica	26602	10	8	0	---	---	---	---	---	---	---	---	NCBI NR
Stramenopiles	Schizochytrium aggregatum	27594	6	4	0	---	---	---	---	---	---	---	---	NCBI NR
Stramenopiles	Thalassiosira pseudonana	28137	4	1	0	---	---	---	---	---	---	---	---	NCBI NR
TOTAL		29400	12			30780	12			30988	10			

Table S2: Testing of selected alternative topologies of *Noctiluca* and *Akashiwo*. Probability scores for different topologies as given by Approximately unbiased (AU) and Expected likelihood weights (ELW) tests. Scores of p=0.05 or greater are highlighted in bold; n.a.=not applicable, outgroup missing.

Test	Dataset	Root1		Root2		Root3		Topology
		p(AU)	p(ELW)	p(AU)	p(ELW)	p(AU)	p(ELW)	
Noctiluca an early branch among core dinoflagellates, a sister lineage to Amphidinium, or branching as a second lineage after Amphidinium	All	**1.000**	**1.000**	**1.000**	**1.000**	n.a.	n.a.	(OUT,(Noc,(Amp,OCDs)));
		2E-036	0	1E-005	0	n.a.	n.a.	(OUT,((Noc,Amp),OCDs));
		1E-060	0	2E-004	0	n.a.	n.a.	(OUT,(Amp,(Noc,OCDs)));
	All-Din	**1.000**	**1.000**	**1.000**	**1.000**	n.a.	n.a.	(OUT,(Noc,(Amp,OCDs)));
		3E-004	0	1E-005	0	n.a.	n.a.	(OUT,((Noc,Amp),OCDs));
		2E-045	0	1E-004	0	n.a.	n.a.	(OUT,(Amp,(Noc,OCDs)));
	All-Pro	**1.000**	**1.000**	**1.000**	**1.000**	n.a.	n.a.	(OUT,(Noc,(Amp,OCDs)));
		2E-008	0	3E-005	0	n.a.	n.a.	(OUT,((Noc,Amp),OCDs));
		2E-008	0	1E-007	0	n.a.	n.a.	(OUT,(Amp,(Noc,OCDs)));
	All-Din&Pro	**1.000**	**1.000**	**1.000**	**1.000**	n.a.	n.a.	(OUT,(Noc,(Amp,OCDs)));
		5E-006	0	3E-004	0	n.a.	n.a.	(OUT,((Noc,Amp),OCDs));
		1E-089	0	3E-004	0	n.a.	n.a.	(OUT,(Amp,(Noc,OCDs)));
Akashiwo a sister lineage of thecate dinoflagellates, a sister lineage of Togula+Gymnodiniaceae s.s., or branching ealier than Togula+Gymnodiniaceae s.s.	All	**0.718**	**0.692**	**0.889**	**0.865**	**0.928**	**0.912**	(OUT,(Gym+Tog,(Aka,The)));
		0.315	**0.305**	**0.133**	**0.129**	**0.082**	**0.088**	(OUT,((Aka,Gym+Tog),The));
		0.008	0.003	0.009	0.005	0.001	0.001	(OUT,(Aka,(Gym+Tog,The)));
	All-Din	**0.748**	**0.725**	**0.896**	**0.876**	**0.933**	**0.907**	(OUT,(Gym+Tog,(Aka,The)));
		0.290	**0.266**	**0.139**	**0.120**	**0.083**	**0.091**	(OUT,((Aka,Gym+Tog),The));
		0.017	0.010	0.019	0.004	0.008	0.002	(OUT,(Aka,(Gym+Tog,The)));
	All-Pro	**0.655**	**0.619**	**0.871**	**0.832**	**0.877**	**0.866**	(OUT,(Gym+Tog,(Aka,The)));
		0.397	**0.373**	**0.165**	**0.162**	**0.143**	**0.134**	(OUT,((Aka,Gym+Tog),The));
		0.021	0.009	0.014	0.006	0.003	0.000	(OUT,(Aka,(Gym+Tog,The)));
	All-Din&Pro	**0.596**	**0.565**	**0.868**	**0.840**	**0.904**	**0.891**	(OUT,(Gym+Tog,(Aka,The)));
		0.468	**0.422**	**0.179**	**0.152**	**0.117**	**0.105**	(OUT,((Aka,Gym+Tog),The));
		0.032	0.013	0.026	0.008	0.006	0.004	(OUT,(Aka,(Gym+Tog,The)));

Table S3: Testing of all alternative topologies among thecate dinoflagellate orders. Probability scores for different topologies as given by Approximately unbiased (AU) and Expected likelihood weights (ELW) tests. All topologies in which at least on of the tests gave p=0.01 or greater are shown; p=0.05 or greater are highlighted in bold.

Dataset	Root1		Root2		Root3		Topology (OUT=outgroup)
	p(AU)	p(ELW)	p(AU)	p(ELW)	p(AU)	p(ELW)	
All	**0.580**	**0.279**	**0.429**	**0.156**	**0.366**	**0.092**	(OUT,((Din,Pro),(Gon,(Per,Sym))));
	0.570	**0.245**	**0.742**	**0.351**	**0.773**	**0.376**	(OUT,((Din,Gon),(Per,(Pro,Sym))));
	0.524	**0.208**	**0.458**	**0.108**	**0.366**	**0.059**	(OUT,(Din,(Gon,(Per,(Pro,Sym)))));
	0.515	**0.227**	**0.565**	**0.267**	**0.611**	**0.331**	(OUT,((Gon,(Din,Pro)),(Per,Sym)));
	0.110	0.017	**0.178**	0.033	**0.255**	**0.053**	(OUT,(Din,(Per,(Gon,(Pro,Sym)))));
	0.058	0.011	**0.202**	0.045	**0.150**	0.014	(OUT,(Din,((Per,Gon),(Pro,Sym))));
	0.044	0.003	**0.071**	0.005	**0.067**	0.007	(OUT,(Din,((Gon,Pro),(Per,Sym))));
	0.038	0.001	**0.125**	0.026	**0.180**	0.046	(OUT,(Per,((Din,Gon),(Pro,Sym))));
	0.032	0.002	0.041	0.004	0.007	0	(OUT,((Din,Gon),(Pro,(Per,Sym))));
	0.021	0.002	0.031	0	0.005	0	(OUT,(Din,(Gon,(Pro,(Per,Sym)))));
	0.017	0	0.001	0	0.020	0	(OUT,((Din,(Pro,Gon)),(Per,Sym)));
	0.013	0	<0.001	0	0.008	0	(OUT,(Gon,(Din,(Per,(Pro,Sym)))));
	0.012	0	<0.001	0	<0.001	0	(OUT,(Gon,((Pro,(Din,Per)),Sym)));

	0.009	0.004	0.016	0.002	**0.053**	0.005	(OUT,((Din,Pro),(Per,(Gon,Sym))));
	0.007	0	0.028	0.001	0.048	0.005	(OUT,((Pro,(Din,Gon)),(Per,Sym)));
	0.005	0	0.019	0	0.006	0	(OUT,(Din,(Pro,(Gon,(Per,Sym)))));
	<0.001	0	<0.001	0	0.017	0	(OUT,((Pro,Gon),(Din,(Per,Sym))));
	<0.001	0	0.021	0	<0.001	0	(OUT,(Pro,(Din,(Per,(Gon,Sym)))));
	<0.001	0	0.016	0	<0.001	0	(OUT,(Per,(Pro,(Din,(Gon,Sym)))));
	<0.001	0	0.048	0.002	**0.059**	0.011	(OUT,((Per,(Din,Gon)),(Pro,Sym)));
	<0.001	0	0.023	0	<0.001	0	(OUT,(Pro,(Din,(Gon,(Per,Sym)))));
	<0.001	0	0.008	0	0.010	0.001	(OUT,(Pro,((Din,Gon),(Per,Sym))));
	<0.001	0	<0.001	0	0.012	0	(OUT,((Pro,(Din,Per)),(Gon,Sym)));
	<0.001	0	<0.001	0	0.018	0	(OUT,(Pro,(Gon,((Din,Per),Sym))));
	<0.001	0	<0.001	0	0.036	0	(OUT,(Pro,(Gon,(Din,(Per,Sym)))));
All-Din	**0.874**	**0.681**	**0.659**	**0.344**	**0.638**	**0.345**	(OUT,(Gon,(Per,(Pro,Sym))));
	0.336	**0.159**	**0.544**	**0.290**	**0.613**	**0.312**	(OUT,(Per,(Gon,(Pro,Sym))));
	0.200	**0.111**	**0.303**	**0.138**	**0.289**	**0.158**	(OUT,((Pro,Gon),(Per,Sym)));
	0.126	0.035	**0.449**	**0.211**	**0.369**	**0.159**	(OUT,((Gon,Per),(Pro,Sym)));
	0.037	0.012	**0.090**	0.015	**0.086**	0.026	(OUT,(Pro,(Gon,(Per,Sym))));
	0.012	0.003	0.008	0	<0.001	0	(OUT,(Gon,(Pro,(Per,Sym))));
	<0.001	0	0.011	0	0.009	0	(OUT,((Pro,Per),(Gon,Sym)));
	<0.001	0	<0.001	0	0.014	0	(OUT,(Pro,(Per,(Gon,Sym))));
All-Pro	**0.603**	**0.569**	**0.761**	**0.713**	**0.824**	**0.771**	(OUT,((Din,Gon),(Per,Sym)));
	0.466	**0.429**	**0.329**	**0.274**	**0.278**	**0.212**	(OUT,(Din,(Gon,(Per,Sym))));
	0.008	0.002	0.012	0.011	0.033	0.016	(OUT,(Din,(Per,(Gon,Sym))));
	0	0	<0.001	0	0.010	0	(OUT,(Per,(Gon,(Din,Sym))));
All-Din&Pro	**0.991**	**0.988**	**0.984**	**0.966**	**0.960**	**0.945**	(OUT,(Gon,(Per,Sym)));
	0.016	0.009	0.028	0.025	**0.067**	**0.050**	(OUT,(Per,(Gon,Sym)));
	0.007	0.003	0.020	0.009	0.013	0.006	(OUT,(Sym,(Per,Gon)));

Table S4: Reference accessions for dinoflagellate cellulases, CESA-like and histon-like proteins. NCBI, CAMPEP=MMETSP (see Table S1), and *S. minutum* genome Db protein accession are shown.

Protein or Enzyme type	Protein name	Dinoflagellate species	Reference accession	Reference source
Cellulose/polysacharide metabolism	Glycosyl hydrolase family 7, dCel1	Pyrocystis lunula	ADG63073	NCBI nr
	Glycosyl hydrolase family 7	Amphidinium carterae	comp414_c0_seq2	NCBI TSA
	Glycosyl hydrolase family 7	Karenia brevis CCMP2229	CAMPEP_0173626610	MMETSP0027, MMETSP0029-MMETSP0031
	Glycosyl hydrolase family 7	Noctiluca scintillans SPMC136	CAMPEP_0194478120	MMETSP0253
	Glycosyl hydrolase family 7	Oxyrrhis marina LB1974	CAMPEP_0190395562	MMETSP1424-MMETSP1426
	Glycosyl transferase CESA-like, type 1	Alexandrium catenella OF101	CAMPEP_0171158290	MMETSP0790
	Glycosyl transferase CESA-like, type 1	Gymnodinium catenatum GC744	CAMPEP_0117466612	MMETSP0784
	Glycosyl transferase CESA-like, type 1	Karenia_brevis CCMP2229	CAMPEP_0173802094	MMETSP0027, MMETSP0029-MMETSP0031
	Glycosyl transferase CESA-like, type 1	Protoceratium reticulatum CCCM535	CAMPEP_0168370212	MMETSP0228
	Glycosyl transferase CESA-like, type 1	Scrippsiella trochoidea CCMP3099	CAMPEP_0115463718	MMETSP0270-MMETSP0272
	Glycosyl transferase	Karlodinium micrum CCMP2283	CAMPEP_0169068018	MMETSP1015-

	CESA-like, type 2			MMETSP1017
	Glycosyl transferase CESA-like, type 2	Prorocentrum minimum CCMP2233	CAMPEP_0177019396	MMETSP0267-MMETSP0269
	Glycosyl transferase CESA-like, type 3	Symbiodinium minutum Mf1.05b	018942.t1	*S. minutum genome Db*
Histon-like proteins, DNA-binding	Histone-like protein, HLP-I	Amphidinium carterae CCMP1314	ACJ04919	NCBI nr
	Histone-like protein, HLP-I	Noctiluca scintillans SPMC136	CAMPEP_0194550744	MMETSP0253
	Histone-like protein, HLP-II, HCC3	Crypthecodinium cohnii	AAM97522	NCBI nr
	Histone-like protein, HLP-II	Symbiodinium minutum Mf1.05b	017975.t1	*S. minutum genome Db*

Table S5: Proteins in non-photosynthetic dinoflagellate plastids. Presence of genes encoding for plastid, and cytosolic and mitochondrial proteins in *Noctiluca*, *Oxyrrhis*, and *Dinophysis*; full protein names, enzyme commission numbers, protein and pathway abbreviations (used in Fig. 3A), and sequence sources (Noctiluca TSA assembly, CAMNT=MMETSP contigs, or CAMPEP=MMETSP proteins; see Table S1) are shown.

Pathway	Localization	Abbreviation	EC no.	Protein name	in Noctiluca?	in Oxyrrhis? [a]	in Dinophysis?
Isoprenoid precursor biosynthesis (Isopentenyl diphosphate= IPP/ Dimethylallyl diphosphate= DMAP)							
	Cytosolic (mevalonate pathway = MEV)						
		HMGCS	2.3.3.10	hydroxymethylglutaryl-CoA synthase			
		HMGCR	1.1.1.34	hydroxymethylglutaryl-CoA reductase			
		MVK	2.7.1.36	mevalonate kinase			
		PMVK	2.7.4.2	phosphomevalonate kinase			
		MVD	4.1.1.33	diphosphomevalonate decarboxylase			
	Plastid (non-mevalonate pathway = MEP/DOXP)						
		DXS	2.2.1.7	1-deoxy-D-xylulose-5-phosphate synthase	c37102_g1_i1	CAMNT_0034174429 [a]	CAMPEP_0179347270
		IspC (DXR)	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase	c8491_g1_i1	CAMPEP_0190399830 [a]	CAMNT_0021046817, CAMNT_0020960785
		IspD	2.7.7.60	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	c34245_g1_i1		CAMNT_0020996201
		IspE	2.7.1.148	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	c34076_g1_i1	CAMNT_0034113123	CAMNT_0021004701
		IspF	4.6.1.12	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	c27998_g3_i1, c27998_g2_i1	CAMPEP_0190331156	CAMNT_0020973523, CAMNT_0021088175
		IspG	1.17.7.1	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	c4614_g1_i1	CAMPEP_0190339536, CAMPEP_0190336160	CAMPEP_0179237040, CAMPEP_0179316044
		IspH (LytB)	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	c37338_g1_i1		CAMNT_0020995807
Tetrapyrrole biosynthesis (heme, chlorophyll, etc.)							
	Mitochondrial (C4 pathway) only						
		ALAS	2.3.1.37	5-aminolevulinate synthase		CAMNT_0034123849	
	Plastid (C5 pathway) only						
		GTR (HemA)	1.2.1.70	glutamyl-tRNA reductase	c16606_g1_i1		CAMNT_0020962887
		GSA (HemL)	5.4.3.8	glutamate-1-semialdehyde 2,1-aminomutase	c6482_g2_i1		CAMNT_0021017431
	Mitochondrial/Cytosolic (C4 pathway) or Plastid C5 pathway)						
		ALAD (HemB)	4.2.1.24	5-aminolevulinate dehydratase/ porphobilinogen synthase	c17529_g1_i1, c17529_g3_i1	CAMNT_0034089891	CAMNT_0020991467
		PBGD (HemC)	2.5.1.61	porphobilinogen deaminase hydroxymethylbilane synthase	c1741_g1_i1_17	CAMNT_0034053199	CAMNT_0020995631
		UROS (HemD)	4.2.1.75	uroporphyrinogen-III synthase			CAMPEP_0179295544
		UROD (HemE)	4.1.1.37	uroporphyrinogen decarboxylase	c18473_g1_i1, c18473_g2_i1, c5162_g1_i1,	CAMNT_0034059593	CAMPEP_0179265208, CAMPEP_0179260602, CAMPEP_0179236930

					c44501_g1_i1		
		CPOX (HemF)	1.3.3.3	coproporphyrinogen III oxidase	c43130_g1_i1, c19480_g3_i1	CAMPEP_0190301822	CAMNT_0021027311
		PPOX (HemY)	1.3.3.4	protoporphyrinogen oxidase	c32625_g3_i1, c7748_g1_i1	CAMNT_0034176619 [a], CAMPEP_0190445206 [a]	CAMNT_0020983029, CAMPEP_0179218520
		FECH (HemH)	4.99.1.1	ferrochelatase	c37180_g1_i1		CAMNT_0021055343
Fatty acid biosynthesis and elongation							
	Cytosolic (polyketide synthase/ fatty acid synthase type I pathway = PKS/FASI)						
		FAAL	2.3.1.86	fatty acid synthase type I, fatty acyl ligase domain	CAMPEP_0194492864, CAMPEP_0194556372, CAMPEP_0194556876, CAMPEP_0194490892	CAMPEP_0190385440, CAMPEP_0179308804	CAMPEP_0179250968
		KS	2.3.1.86	fatty acid synthase type I, ketoacyl synthase domain	CAMPEP_0194492864, CAMPEP_0194518942, CAMPEP_0194531188, CAMPEP_0194533868, CAMPEP_0194544822, CAMPEP_0194547234, CAMPEP_0194547510, CAMPEP_0194549100, CAMPEP_0194553672, CAMPEP_0194555636, CAMPEP_0194556058, CAMPEP_0194556250, CAMPEP_0194556372, CAMPEP_0194556646, CAMPEP_0194556876	CAMPEP_0190386610, CAMPEP_0190386310, CAMPEP_0190385440, CAMPEP_0190321934	CAMPEP_0179225310, CAMPEP_0179230504, CAMPEP_0179231766, CAMPEP_0179233104, CAMPEP_0179240346, CAMPEP_0179274784, CAMPEP_0179301140, CAMPEP_0179302604, CAMPEP_0179309896, CAMPEP_0179340930, CAMPEP_0179347172, CAMPEP_0179358456, CAMPEP_0179372828
		AT	2.3.1.86	fatty acid synthase type I, acyl transferase domain	CAMPEP_0194492864, CAMPEP_0194482594, CAMPEP_0194485702, CAMPEP_0194531888, CAMPEP_0194554940, CAMPEP_0194556372, CAMPEP_0194556876, CAMPEP_0194557686	CAMPEP_0190386610, CAMPEP_0190386310, CAMPEP_0190385440, CAMPEP_0190315816	CAMPEP_0179225352, CAMPEP_0179232228, CAMPEP_0179232754, CAMPEP_0179257472, CAMPEP_0179301464, CAMPEP_0179308134
		DH	2.3.1.86	fatty acid synthase type I, dehydrase domain	CAMPEP_0194551082, CAMPEP_0194552350	CAMPEP_0190386610, CAMPEP_0190323466	CAMPEP_0179257472
		ER	2.3.1.86	fatty acid synthase type I, enoyl reductase domain	CAMPEP_0194551082, CAMPEP_0194552350	CAMPEP_0190323466	CAMPEP_0179257472
		KR	2.3.1.86	fatty acid synthase type I, ketoacyl reductase domain	CAMPEP_0194551082, CAMPEP_0194552350, CAMPEP_0194492864	CAMPEP_0190386310, CAMPEP_0190384048, CAMPEP_0190330588	CAMPEP_0179227628, CAMPEP_0179257472, CAMPEP_0179268324, CAMPEP_0179292270, CAMPEP_0179306616, CAMPEP_0179310080, CAMPEP_0179311052, CAMPEP_0179311360
		ACP	2.3.1.86	fatty acid synthase type I, acyl carrier protein domain	CAMPEP_0194492864, CAMPEP_0194549100, CAMPEP_0194551082, CAMPEP_0194552350, CAMPEP_0194556058, CAMPEP_0194556372, CAMPEP_0194556876	CAMPEP_0190321934, CAMPEP_0190323466, CAMPEP_0190385440	CAMPEP_0179257472
		TRD (SDR)	2.3.1.86	fatty acid synthase type I, terminal reductase domain			
	Endoplasmic reticulum (ER fatty acid elongation pathway)						
		ELO	2.3.1.199	beta-ketoacyl-CoA synthase	CAMPEP_0194480994	CAMPEP_0190376278, CAMPEP_0190332524	CAMPEP_0179225992, CAMPEP_0179261800
		KCR	1.1.1.330	beta-ketoacyl-CoA reductase		CAMPEP_0190306630	
		PHS	4.2.1.134	beta-hydroxyacyl-CoA dehydratase	CAMPEP_0194540052	CAMPEP_0190316596	CAMPEP_0179242246, CAMPEP_0179358262
		TECR	1.3.1.93	trans-2-enoyl-CoA reductase			
	Plastid (fatty acid synthase type II pathway = FASII)						
		FabD	2.3.1.39	malonyl-CoA-acyl carrier protein transacylase			
		FabG	1.1.1.100	beta-ketoacyl-acyl carrier protein reductase			
		FabH	2.3.1.180	beta-ketoacyl-acyl carrier protein synthase III			
		FabZ	4.2.1.59	D-3-hydroxyoctanoyl-acyl carrier protein dehydratas			
		FabI	1.3.1.9	enoyl acyl carrier protein reductase			CAMPEP_0179266880 [b]

		FabB/F	2.3.1.41	beta-ketoacyl-acyl carrier protein synthetase			
		ACP		acyl-carrier protein			
Iron-sulfur (Fe-S) cluster assembly							
	Plastid (Suf pathway)						
		SufA		Iron-sulfur assembly protein SufA			
		SufB		Iron-sulfur assembly protein SufB	c20274_g1_i1, c20274_g2_i1	CAMNT_0034061651	CAMNT_0021013865
		SufC		Iron-sulfur assembly protein SufC		CAMNT_0021025055	CAMNT_0034058317
		SufD		Iron-sulfur assembly protein SufD			CAMNT_0021040041
		SufE		Iron-sulfur assembly protein SufE			
Ferredoxin redox system							
	Plastid (Fd – FNR pathway)						
		Fd (PetF)		[2Fe-2S] ferredoxin	c41939_g1_i1	CAMNT_0034037779	CAMNT_0020927107
		FNR (PetH)	1.18.1.2	Ferredoxin NADP+ reductase	c11604_g1_i1		CAMNT_0020923739
Triosephosphate translocation							
	Plastid (TPT translocon)						
		TPT		Triosephosphate/phosphate transporter	c40379_g1_i1		CAMNT_0021019703, CAMNT_0021010591
Protein folding and processing							
	Plastid (ClpC chaperone/protease)						
		ClpC		Chloroplast molecular chaperone ClpC	c23770_g4_i1 c23770_g5_i1	CAMNT_0034034689	CAMNT_0020950785

[a] sequence was obtained from *Oxyrrhis marina* LB1974 MMETSP1424-1426 combined assembly (all other were obtained from *Oxyrrhis marina* MMETSP0468-471 combined assembly).
[b] closely related to Kareniaceae, a full plastid pathway unlikely to be present - see main text

Table S6: Signal and target peptide predictions. Prediction statistics of proteins predicted to be targeted to the non-photosynthetic plastid in *Noctiluca*, *Oxyrrhis*, and *Dinophysis* are listed (protein abbreviations correspond to Table S5). Protein ID: CAMNT=MMETSP contig at iMicrobe or *Noctiluca scintillans* Trinity assembly contig name in NCBI TSA; predicted cleavage sites of signal (Cmax and Ymax scores) and transit peptides (CS-score) are generally low - this has been observed previously in the dinoflagellate and *Perkinsus* plastid proteins.

No	Dinoflagellate species	Protein and accession		N-terminal region integrity				Signal peptide (SP) prediction in SignalP 4.1							Transit peptide (cTP) prediction in ChloroP 1.1				
		Protein	Protein ID	N-terminal extension?	Met present?	Spliced leader at 5' end?	STOP upstream of 1st Met?	SP Cmax score	SP Ymax score	SP Smax score	SP Smean score	Cleavage site positions	SP D-score	SP Networks-used	Phenylalanine present? (position relative to cleavage site)	cTP score	cTP CS-score	cTP length	domainSecond hydrophobic
1	Noctiluca scintillans	IspD	c34245_g1_i1	yes	yes	6	yes	0.644	0.736	0.961	0.852	30-31	0.798	SignalP-noTM	FAMP (-2)**	0.477	2.767	75	
2	Noctiluca scintillans	IspE	c34076_g1_i1	yes	yes			0.200	0.341	0.803	0.686	24-25	0.479	SignalP-TM	FDLV (+1)	0.560	11.862	70	
3	Noctiluca scintillans	IspG	c4614_g1_i1	yes	yes	12		0.222	0.385	0.814	0.661	21-22	0.534	SignalP-noTM	FVSS (+7)	0.519	4.804	30	
4	Noctiluca scintillans	IspH	c37338_g1_i1	yes	yes	6	yes	0.340	0.447	0.710	0.573	19-20	0.515	SignalP-noTM	FALP (+14)	0.587	13.209	67	
5	Noctiluca scintillans	Fd	c41939_g1_i1	yes	yes	6	yes	0.399	0.581	0.940	0.852	27-28	0.727	SignalP-noTM	FAIA (+7)**	0.460	2.204	55	
6	Noctiluca scintillans	FNR	c11604_g1_i1	yes	yes			0.24	0.47	0.98	0.92	28-29	0.71	SignalP-noTM	FVQM (-3)**	0.55	11.93	51	

7	Dinophysis acuminata	IspC	CAMNT_0021046817	yes	yes			0.520	0.623	0.859	0.735	17-18	0.684	SignalP-noTM	FVPG (+1)	0.554	4.408	33	
8	Dinophysis acuminata	IspF	CAMNT_0020973523	yes				0.323	0.495	0.854	0.749	19-20	0.632	SignalP-noTM	FSHQ(-2)	0.476	1.677	22	
9	Dinophysis acuminata	TPT1	CAMNT_0021019703	yes	yes	6		0.758	0.692	0.765	0.634	23-24	0.661	SignalP-noTM		0.510	2.188	21	
10	Dinophysis acuminata	TPT2	CAMNT_0021010591	yes	yes		yes	0.817	0.740	0.793	0.675	23-24	0.705	SignalP-noTM		0.523	1.992	21	
11	Dinophysis acuminata	SufC	CAMNT_0021025055	yes				0.62	0.73	0.9	0.86	19-20	0.8	SignalP-noTM	FAAA (-2)	0.46	2.91	4 ***	
12	Dinophysis acuminata	FNR	CAMNT_0021011101	yes				0.22	0.42	0.95	0.8	21-22	0.63	SignalP-noTM	FASP (+13)	0.51	7.62	42	
13	Dinophysis acuminata	Fd	CAMNT_0020927107	yes	yes	8*	yes	0.188	0.382	0.922	0.786	25-26	0.600	SignalP-noTM	FVAP (+7)	0.463	4.645	67	
14	Oxyrrhis marina MMETSP0468	IspG	CAMNT_0034079579	yes				0.684	0.591	0.600	0.524	19-20	0.564	SignalP-TM	FSLR (+5)	0.520	3.042	72	yes
15	Oxyrrhis marina MMETSP0468	ALAD	CAMNT_0034089891	yes	yes			0.609	0.612	0.885	0.707	26-27	0.650	SignalP-TM		0.518	0.664	73	yes
16	Oxyrrhis marina MMETSP0468	SufB	CAMNT_0034061651	yes	yes			0.18	0.38	0.89	0.8	15-16	0.55	SignalP-TM		0.45	2.91	64	
17	Oxyrrhis marina MMETSP0468	PBGD	CAMNT_0034053199	yes	yes			0.427	0.486	0.707	0.588	19-20	0.527	SignalP-TM	FLQS (+8)	0.516	1.892	65	yes
	Oxyrrhis marina LB1974	PBGD	CAMNT_0034168717	yes	yes	6		0.159	0.335	0.841	0.710	17-18	0.485	SignalP-TM	FLES (+10)	0.527	1.892	70	yes
18	Oxyrrhis marina MMETSP0468	Fd	CAMNT_0034037779	yes												0.544	5.420	40	

*masked by 3 nucleotides at 5' end

**alternative SP cleavage site at the position +1

***alternative TP cleavage site likely present