Published in final edited form as:

J Eukaryot Microbiol. 2016 ; 63(1): 123–137. doi:10.1111/jeu.12261.

Complex Ancestries of Isoprenoid Synthesis in Dinoflagellates

Bastian Bentlage^{#a}, Travis S. Rogers^{#a}, Tsvetan R. Bachvaroff^b, Charles F. Delwiche^{a,c}

^aCMNS-Cell Biology and Molecular Genetics, University of Maryland, 2107 Bioscience Research Building, College Park, Maryland 20742

^bInstitute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, 701 E Pratt St., Baltimore, Maryland 21202

^cMaryland Agricultural Experiment Station, AGNR, University of Maryland, College Park, Maryland 20742

These authors contributed equally to this work.

Abstract

Isoprenoid metabolism occupies a central position in the anabolic metabolism of all living cells. In plastid-bearing organisms, two pathways may be present for de novo isoprenoid synthesis, the cytosolic mevalonate pathway (MVA) and nuclear-encoded, plastid-targeted nonmevalonate pathway (DOXP). Using transcriptomic data we find that dinoflagellates apparently make exclusive use of the DOXP pathway. Using phylogenetic analyses of all DOXP genes we inferred the evolutionary origins of DOXP genes in dinoflagellates. Plastid replacements led to a DOXP pathway of multiple evolutionary origins. Dinoflagellates commonly referred to as dinotoms due to their relatively recent acquisition of a diatom plastid, express two completely redundant DOXP pathways. Dinoflagellates with a tertiary plastid of haptophyte origin, by contrast, express a hybrid pathway of dual evolutionary origin. Here, changes in the targeting motif of signal/transit peptide likely allow for targeting the new plastid by the proteins of core isoprenoid metabolism proteins. Parasitic dinoflagellates of the *Amoebophyra* species complex appear to have lost the DOXP pathway, suggesting that they may rely on their host for sterol synthesis.

Keywords

Dinoflagellata; DOXP; endosymbiont gene transfer; Haptophyta; MEP; mevalonate; MVA

ISOPRENOIDS (also referred to as terpenoids) are at the core of an organism's metabolism, performing an array of functions vital to all forms of life. On a cellular level, iso-prenoids are responsible for protein prenylation, sterol, and quinone synthesis, and the formation of pigments (carotenoids) involved in photosynthesis. Isoprenoids are derived from two branched five-carbon (C₅) precursor isomers, dimethylallyl diphosphate (DMAPP) and

Correspondence B. Bentlage, Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, MRC-163, PO Box 37012, Washington, DC 20013, USA Telephone number: +1 (202) 633-1295; FAX number: +1 (202) 633-0182; bastian.bentlage@gmail.com.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

isopentenyl diphosphate (IPP). These isomers are the common products of two distinct pathways, the mevalonate (MVA) and nonmevalonate (DOXP) pathways (Eisenreich et al. 2004; Hunter 2007; Rohmer 1999). The MVA pathway was believed to be the universal metabolic route for the synthesis of DMAPP and IPP. The discovery and subsequent characterization of the prokaryotic DOXP pathway (Fig. 1 top left panel) pathway in the 1990s (Rohmer et al. 1993) led to phylogenetic studies on how the two pathways are distributed across the tree of life (e.g., Lohr et al. 2012; Lombard and Moreira 2011).

The current assumption is that eukaryotes containing a plastid gained the genes for the DOXP pathway from Cyanobacteria (Gould et al. 2008; Lombard and Moreira 2011). In the absence of a plastid, the cytosolic MVA pathway is utilized exclusively (Lombard and Moreira 2011). However, the distribution of the MVA pathway throughout the tree of life is more complex. For example, the MVA pathway is present alongside the plastid-derived DOXP pathway in streptophytes, the lineage including charophyte green algae and land plants (Lohr et al. 2012). By contrast, chlorophyte green algae rely exclusively on the DOXP pathway (Lichtenthaler 1999). As the example of the streptophyte lineage shows, the acquisition of a plastid and subsequent transfer of DOXP pathway genes to the host nucleus does not necessitate the loss of the host MVA pathway. In the vascular plant *Arabidopsis thaliana* the two pathways can work in concert with one another, with the MVA pathway synthesizing sterols inside the cytosol and the plastid-localized DOXP pathway playing a central role in quinone and carotenoid biosynthesis (Laule et al. 2003).

Knowledge of isoprenoid biosynthesis in protists is more limited, although there is considerable interest in the topic because these pathways have the potential to provide natural products for applications ranging from animal feed, human nutrition, cosmetics, and biofuels to pharmaceutical drugs (e.g., Bohlmann and Keeling 2008; Kirby and Keasling 2009; Lohr et al. 2012). Dinoflagellates are important members of both marine and freshwater environments, but knowledge of their isoprenoid metabolism is fragmentary. Along with their sister-lineage, the oyster parasite *Perkinsus marinus* (Bachvaroff et al. 2011, 2014; Reece et al. 1997; Saldarriaga et al. 2003), dinoflagellates are closely related to apicomplexan parasites (Bachvaroff et al. 2014). In apicomplexans, only the DOXP pathway is present (e.g., Lim and McFadden 2010), which has, for example, allowed for the development of drugs interrupting the DOXP pathway to treat multi-drug resistant malaria caused by the apicomplexan parasite *Plasmodium* (Jomaa et al. 1999). Among dinoflagellates the presence of the DOXP pathway was demonstrated by the discovery of select genes of the pathway in expressed sequence tags from the nonphotosynthetic dinoflagellates *Oxyrrhis marina* (Slamovits and Keeling 2007) and *Cryptothecodinium cohnii* (Sanchez-Puerta et al. 2007), implying gene transfer to the nucleus from an ancestral plastid, now lost or cryptic. The full complement of DOXP pathway genes has been described in *P. marinus*, which has been used as evidence to infer the presence of a cryptic plastid in *P. marinus* (Matsuzaki et al. 2007). More recently, the DOXP pathway was characterized for a select number of photosynthetic dinoflagellates (Hehenberger et al. 2014).

In dinoflagellates, the ancestral plastid is thought to have been of red-algal origin (e.g., Bodyl et al. 2009; Howe et al. 2008; Sanchez-Puerta and Delwiche 2008). More than half of the roughly 2,300 described dinoflagellate species are photosynthetic, containing

chlorophylls *a* and *c* as their major light-harvesting pigments and the distinctive carotenoid peridinin (Dodge and Lee 2000; Hackett et al. 2004). This peridinin-type plastid is thought to be the ancestral plastid, but in several taxa the peridinin plastid was lost or replaced with other plastid types. In addition, the earliest branching dinoflagellates are heterotrophic (free-living or parasitic; e.g., Bachvaroff et al. 2014), lacking the ability to photosynthesize, and consequently possess reduced cryptic plastids or possibly lack a plastid (e.g., the loss of the plastid in the crustacean parasite *Hematodinium* sp. Gornik et al. 2015).

We set out to characterize the isoprenoid synthesis pathway in a broad sample of dinoflagellates using transcriptomic data. One of our particular aims was to investigate the possible mosaic of isoprenoid synthesis in taxa that have undergone plastid replacement, including species with endosymbiotic diatoms (“dinotoms”) and species in the Kareniaceae with tertiary plastids of haptophyte origin (e.g., Archibald 2009). In the Kareniaceae (genera *Karlodinium* and *Karenia*), the peridinin plastid was replaced by endosymbiosis of a haptophyte alga, as evidenced by the presence of 19’hexanoyloxy-fucoxanthin (fucoxanthin hereafter) instead of peridinin as a light-harvesting pigment, as well as differences in plastid morphology (Tengs et al. 2000). In dinotoms reduction of the newly acquired diatom endosymbiont is incomplete, likely representing an early stage of the process of establishing endosymbiosis (Chesnick et al. 1997). Plastid replacements have led to a mosaic pattern of evolutionary origins of several nuclear-encoded plastid-targeted genes in dinoflagellates (Nosenko et al. 2006; Wisecaver et al. 2013; Yoon et al. 2005). One topic we address is whether dinoflagellates in which the peridinin-type plastid was replaced contain isoprenoid synthesis genes from their ancestral endosymbiont, the peridinin plastid, or from the more recently acquired tertiary plastids of haptophyte and diatom origin.

MATERIALS AND METHODS

Ingroup dinoflagellates

Culture and RNA extraction—Dinoflagellates were obtained from several culture collections and grown in either 15ppt or 32ppt sterilized L1 medium (minus Na₂SiO₃; Guillard and Hargraves 1993) at 18 °C (4 °C for *Polarella glacialis*) on a 16:8 h light:dark cycle (*Gyrodinium instriatum* was grown under a range of conditions and the sequence data used for assembly were a combination of six samples). Cells of the thecate (“armored”) dinoflagellates *Prorocentrum hoffmannianum* (NCBI Biosample SAMN03075519), *Prorocentrum minimum* (NCBI TSA GBFP01000000, BioSample SAMN02768558; see Cooper et al. 2014), *Prorocentrum micans* (NCBI Biosample SAMN03075521), and *Prorocentrum* sp. (NCBI Biosample SAMN03075526) were disrupted using a 1 min cycle on a Mini-Beadbeater-1 (Biospec Products, Bartlesville, OK) while the athecate dinoflagellates *Amphidinium carterae* (NCBI Biosample SAMN03015296), and *P. glacialis* (NCBI Biosample SAMN03075516) were not disrupted using a beadbeater prior to RNA extraction. For *G. instriatum* (NCBI Biosample SAMN03075546), cells were pelleted by centrifugation and frozen in liquid nitrogen for lysis; frozen cells were resuspended in RNeasy lysis buffer (Qiagen, Crawley, UK) and subjected to 1 min of bead beating as described above. RNA was extracted using either the RNeasy spin kit (Qiagen) or the Nucleospin Plant Total RNA kit (Machery-Nagel, Düren, Germany).

The dinoflagellate parasite *Amoebophyra* sp.—Two strains of the dinoflagellate parasite *Amoebophyra* sp. from the Chesapeake Bay were grown infecting their specific hosts *Akashiwo sanguinea* and *Karlodinium veneficum* respectively. Stock cultures were maintained and near synchronous infection of host cultures was established following the methods described in Coats and Park (2002) and Bachvaroff et al. (2009). Infective *Amoebophyra* sp. dinospores were collected within 6 h of formation by gravity filtration through 5 µm Nucleopore filters (Whatman, Piscataway, NJ) and were then used to infect cultures of *K. veneficum* and *A. sanguinea*. As described above under RNA extraction, RNA was extracted from dinospores for *Amoebophyra* sp. infecting *A. sanguinea* (*Amoebophyra* sp. ex. *A. sanguinea* dinospores; NCBI Biosample SAMN03075505). In addition, *Amoebophyra* sp. RNA was extracted from combined host-parasite cultures 48 h post infection for *A. sanguinea* (*Amoebophyra* sp. ex. *A. sanguinea* 48 h; NCBI Biosample SAMN03075505). Dinospores of *Amoebophyra* sp. ex. *K. veneficum* (NCBI Biosample SAMN03075506) were enriched roughly 60 h post infection using gravity filtration without fully removing host cells.

Transcriptome sequencing and assembly—Libraries for paired-end 100 bp Illumina HiSeq sequencing were prepared and sequenced at Macrogen (Seoul, South Korea) or the Institute for Bioscience & Biotechnology Research sequencing core (University of Maryland, College Park, MD) after poly-A selection of the total RNA. Additional 50 bp paired-end Illumina sequence data for *Ceratium fusus*, *C. cohnii*, *Dinophysis acuminata*, *Heterocapsa triquetra*, *Karenia brevis*, *Karlodinium micrum* (= *K. veneficum*; Bergholtz et al. 2006), *Kryptoperidinium foliaceum* and *Glenodinium foliaceum* (note that *G. foliaceum* is considered a synonym of *K. foliaceum*; Guiry and Guiry 2015), *Lingulodinium polyedrum*, *Noctiluca scintillans*, *O. marina*, and *Scrippsiella* sp. were obtained from the Marine Microbial Eukaryote Transcriptome Sequencing Project (Keeling et al. 2014). All Illumina datasets were assembled de novo using Trinity release 2013-02-16 (Grabherr et al. 2011) with default settings. Additional transcriptome assemblies were obtained for *Alexandrium tamarense* (NCBI TSA accession GAFP00000000.1) and *Symbiodinium* sp. clades A (http://medinalab.org/zoox/kb8_assembly.fasta.bz2; Bayer et al. 2012) and D (NCBI TSA accession GAFP01000001). A full list of transcriptomes used in this study can be found in Supporting information.

Outgroup taxon sampling

Outgroup taxa included a broad sampling of Eubacteria, Cyanobacteria, Rhodophyta, Stramenopiles (heterokonts), Cryptophyta, Haptophyta, and the “green lineage” comprising Chlorophyta and Streptophyta (Supporting information). The dataset published in Matsuzaki et al. (2007) provided the backbone of the outgroup taxon sampling (excluding haptophytes and cryptophytes). Additional sequence datasets were obtained from NCBI’s RefSeq database (<http://www.ncbi.nlm.nih.gov/refseq>) and the Joint Genome Institute’s Genome Portal (<http://genome.jgi-psf.org>). 50 bp paired-end Illumina data for several haptophytes and cryptophytes were obtained from the Marine Microbial Eukaryote Transcriptome Sequencing Project (Keeling et al. 2014) and assembled using Trinity as described above under dinoflagellate taxon sampling. Additional chlorophyte and streptophyte sequences

were obtained from unpublished 100 bp paired-end Illumina sequencing datasets that were assembled using Trinity as described above.

We included several apicomplexan sequences in initial analyses, but these sequences were difficult to align to the remainder of the taxa due to large divergences and resulted in long branches in the phylogenies; we excluded these sequences from further analyses following the example of Matsuzaki et al. (2007). Note that our alignments include a large taxonomic selection of groups with putatively red algal-derived plastids despite the exclusion of apicomplexans.

Identification of isoprenoid genes and phylogenetic analysis

Homologs of DOXP and MVA pathway genes were identified in the sequence datasets of both in- and outgroup taxa using *A. thaliana* protein queries (Supporting information) using tblastn searches (Altschul et al. 1990) with an *e*-value cut-off of 10^{-30} . All hits above the *e*-value cut-off were retained for the dinoflagellate datasets; only the top hits were considered for outgroup taxa. In several cases protein-coding sequences in the de novo assembled transcriptomes were fragmented into several contigs belonging to the same graph component, suggesting that they are connected to each other (cf. Grabherr et al. 2011). We assembled these contigs using the overlap-layout-consensus assembly algorithm implemented in Geneious (version 5.3.0; Biomatters Ltd, Auckland, New Zealand). In cases where the secondary assembly of contigs failed to yield one contiguous sequence, all partial protein-coding sequences were included in downstream analyses. In cases in which the Trinity assembly contained multiple potential isoforms (cf., Grabherr et al. 2011) we only retained the one with the best BLAST score. Open reading frames (ORFs) were predicted using NCBI's OrfFinder (<https://www.ncbi.nlm.nih.gov/gorf/gorf.html>) with the standard genetic code. The parasitic chlorophytes *Cephaleuros parasitucs* and *Trentepohlia annulata* presented the exception since they use an alternate genetic code (e.g., Cocquyt et al. 2010) and were translated using the ciliate genetic code.

Protein sequences were aligned using MUSCLE (Edgar 2004) as implemented in Seaview (version 4.4.2; Gouy et al. 2010). Flanking regions containing missing protein sequence data were identified using the least stringent settings of Gblocks (Talavera and Castresana 2007) as implemented in SeaView (version 4.4.2; Gouy et al. 2010). Instead of utilizing only the conserved regions identified by Gblocks we retained all residues between and including the first and last conserved region to maximize the amount of information for phylogenetic analyses (alignments available through TreeBase: <http://purl.org/phylo/treebase/phylo/phylo/TB2:S18088>). The maximum likelihood (ML) substitution model (including the +I, +G, +F add-ons), fitting the data in these alignments best was inferred using the Bayesian information criterion (BIC) framework implemented in ProtTest (version 2.4; Abascal et al. 2005) with the search restricted to the following models: JTT, LG, DCMut, Dayhoff, WAG, CpRev, Blo-sum62, and VT (Abascal et al. 2005 and references cited in the software's documentation). ML tree searches were performed using PhyML (version 20120412; Guindon and Gascuel 2003) and robustness of the phylogenies evaluated using bootstrapping. We performed 100 nonparametric bootstrap replicates for alignments that included distant eubacterial outgroups with the intention of identifying potential bacterial

contaminants for our sequences of interest. Note that 100 bootstrap replicates can still be considered to produce useful estimates of node support on phylogenies (e.g., Efron et al. 1996; see also Pattengale et al. 2009), making this approach efficient and computationally cheap for the limited scope of identifying artifacts that may affect downstream analyses. By contrast, we performed 500 bootstrap replicates for those alignments containing Cyanobacteria as outgroups on which our subsequent interpretations were based. For these datasets, we further performed a Bayesian analysis using PhyloBayes 3.3 (Lartillot et al. 2009) with the ratecat model to estimate posterior probabilities of nodes. Two chains were sampled at an interval of 100 following a burn-in of 500 cycles, stopping the run when convergence among chains reached 0.1 or lower and effective sample size was greater than 100 (`-nchain 2 100 0.1 100 -ratecat -cat`).

For *Amoebophyra* spp. we performed an additional screen using hidden markov model (HMM) searches to scan the transcriptomes for additional sequences of interest. Following the procedure outlined in Parra et al. (2007), we calculated simulated cutoff scores for each DOXP protein. In short, we aligned each individual protein sequence to an HMM profile generated from the remaining corresponding protein sequences using *hmmsearch* (Eddy 2011). From this we obtained scores that were averaged, yielding a cutoff score for each of the 7 DOXP genes. Lastly, ORFs for the three independent *Amoebophyra* spp. transcriptomes were identified using the ORF caller TransDecoder with default settings (Haas et al. 2013). ORFs were then aligned against HMMs generated from the seven multiple sequence alignments of each DOXP gene, omitting every hit above the cutoff threshold.

Hypothesis testing: plastid replacement and the origin of isoprenoid genes

To test whether genetic mosaicism of nuclear-encoded isoprenoid pathway genes is present in the kareniaacean dinoflagellates *K. brevis* and *K. veneficum* (including the synonymous *K. micrum*) that possess tertiary plastids of haptophyte origin (Nosenko et al. 2006), we performed likelihood-based topology tests using the approximately unbiased (AU) test (Shimodaira 2002) implemented in Consel (Shimodaira and Hasegawa 2001). Similarly, we tested the evolutionary origins of DOXP pathway genes for *K. foliaceum* (and the synonymous *Glenodinium foliaceum*) that contain a tertiary plastid of diatom (Stramenopiles) origin (e.g., Chesnick et al. 1997). For this purpose, we created several minimal topologically constrained phylogenetic trees for each amino acid alignment (see <https://github.com/bastodian/DOXP-Constrained-Analysis>). In short, we constrained the topology to a polytomy containing six monophyletic clades of rhodophytes, Stramenopiles, cryptophytes, haptophytes, dinoflagellates (plus *Perkinsus*), and Viridiplantae; Cyanobacteria served as outgroup. For each protein the H_0 tree corresponded to the placement of sequences of interest into the larger clades observed in the best ML tree found as described in the previous section under phylogenetic analysis. We then tested whether sequences in dinoflagellates with replacement plastids were derived from the peridinin plastid (i.e., fall within the dinoflagellate clade) or the replacement plastid (i.e., fall within a haptophyte or stramenopile clade). Phylogenetic analyses under these topological constraints were performed in PhyML (version 3.1) to calculate per-site likelihoods.

GC content calculation and signal peptide prediction

GC content was calculated for every contig in the transcriptome assemblies using a custom python script. Signal peptide presence and location of cleavage sites for protein sequences were predicted using the SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>; Petersen et al. 2011) with the sensitive *D*-value cut-off settings.

RESULTS

Transcriptome assemblies

Dinoflagellate assemblies generated for this study are available through NCBI Bioprojects PRJNA257290 and PRJNA81753 (Biosample numbers provided in ingroup sampling above and Supporting information). Assemblies generated from short-read datasets described in Keeling et al. (2014) are available through figshare (<http://dx.doi.org/10.6084/m9.figshare.1152727>).

BLAST and HMM search results

While we found the full complement of the seven DOXP pathway genes (Dxs, Dxr, IspD, IspE, IspF, IspG, and IspH; Supporting information) in almost all of the dinoflagellate transcriptomes surveyed, we did not find the full complement of MVA pathway genes in any dinoflagellate. Several assemblies contained transcripts with BLAST hits to the MVA genes Thiolase, HMGS, HMGR, and IPI, possibly as a result of shared functional domains (*K. foliaceum* had a similar hit to MPD as well). In the absence of evidence of a functioning MVA pathway in the dinoflagellate transcriptomes we focused our phylogenetic analyses on the DOXP pathway. The copy of IspF found in *O. marina* was extremely short and we removed it from the analysis to avoid introducing noise into the dataset in the form of excessive missing data.

A notable exception in our dataset was the transcriptome of *Amoebophrya* sp. ex. *A. sanguinea* dinospores, comprised of free-living spores of *Amoebophrya* sp. (see Miller et al. 2012 for a description of the life cycle), which contained no evidence for the presence of DOXP pathway genes, neither using BLAST as a search tool or HMMs. By contrast, *Amoebophrya* sp. ex *A. sanguinea* 48 h post infection and *Amoebophrya* sp. ex. *K. veneficum*, both of which comprised host and parasite cells, contained the full complement DOXP pathway.

Phylogenetic analysis

All phylogenetic analyses were performed under the LG+I+G substitution model that was inferred to fit the data best using the BIC framework implemented in ProtTest. Using Eubacteria as an outgroup we found that the dinoflagellate DOXP sequences identified by us generally do not represent bacterial contaminants (Fig. S8–S14). Note that the sequence from the rhodophyte *Galdieria sulphuraria* falls into the bacterial clade albeit at low support for IspE (Fig. S11); for IspG both rhodophytes in our analysis are nested within Cyanobacteria (Fig. S13). Conversely, Cyanobacteria (*Anabaena variabilis*, *Gloeobacter violaceus*, *Prochlorococcus marinus*, and *Synechococcus* sp.) fall within the clade of eukaryotes for both Dxr (Fig. S9) and IspH (Fig. S14). A *Glenodinium foliaceum* copy of

Dxs (Fig. S8) and IspG (Fig. S12) fell within Bacteria, prompting us to remove it from further analyses after corroborating its bacterial origin with a BLAST search against NCBI's GenBank (Benson et al. 2014). Ingroup phylogenetic analyses were performed after removing bacterial outgroups other than Cyanobacteria.

For the majority of DOXP genes dinoflagellates (or dinoflagellates plus *Perkinsus*) formed moderately to highly supported monophyletic clades (Fig. 1, 2, S1, S2, S5, S6–S8, S9, S12–S14). *Perkinsus marinus* generally groups with dinoflagellates (Fig. 1, 2, S1–S3, S6–S7), but its placement is often weakly supported. While the DOXP sequences in our phylogenetic analyses often form mono-phyletic clades according to their taxonomic origin, the backbone of the phylogeny is at times poorly supported and in several instances topologies are inconsistent with our understanding of the species phylogeny. Cryptophytes, haptophytes, and rhodophytes form monophyletic clades, usually with high support, across all phylogenetic trees (Fig. 1, 2, S1–S14). Stramenopiles, however, are more labile with the diatoms *Thalassiosira pseudonana* and *Phaedactylum tricornutum* often being separate from *Aureococcus anophagefferans* (Pelagophyceae), *Ectocarpus siliculosus* (Phaeophyceae), and *Nannochloropsis gaditana* (Eustigmatophyceae). Viridiplantae is monophyletic in most analyses but is polyphyletic in the analysis of IspF (Fig. 2, S5, S12) and paraphyletic for IspG (Fig. 2, S6 but see Fig. S13 in which it is polyphyletic).

The tertiary plastid-bearing dinoflagellates *Karenia* and *Karlodinium* show a mosaic of placements for their DOXP genes. While *K. brevis* possesses a full complement of DOXP genes that fall among dinoflagellates, *K. brevis* also possesses copies that fall within haptophytes (IspE: Fig. 2, S4, IspF: Fig. 2, S5, IspH: Fig. 2, S7), Stramenopiles (IspH: Fig. 2, S7), and Viridiplantae (IspD: Fig. 1, S3). We did not find evidence for a full dinoflagellate complement of DOXP genes in the dataset from *K. veneficum*, but found copies of all seven DOXP genes for the synonymous *K. micrum*. It appears unlikely that *K. veneficum* is not expressing one of the DOXP genes. Rather, this result may be attributed to differences in sampling depth between the two transcriptome libraries. For the terminal enzymes of the pathway, Dxs and IspH, both *K. veneficum* and *K. micrum* consistently fall within haptophytes (Fig. 1, 2, S1, S7). In contrast with *Karenia* and *Karlodinium*, the dinotom (diatom-containing) *K. foliaceum* (and the synonymous *G. foliaceum*) show a grouping of a full set of DOXP genes with dinoflagellates while a second set groups with Stramenopiles (Fig. 1, 2, S1–S7). It is particularly noteworthy that stramenopile gene copies for *K. foliaceum* and *G. foliaceum* are consistently the sister-lineage to diatoms. Interestingly, a copy of IspH from the predatory and kleptoplastidic *D. acuminata* fell within haptophytes (Fig. 2, S7).

Phylogenetic hypothesis testing

For *K. brevis* we find that we cannot reject the hypothesis of it possessing a copy of every DOXP gene of dinoflagellate origin (Fig. 3). For IspE, IspF, and IspH we further cannot reject that additional copies are of haptophyte origin. While we observed an IspD copy of *K. brevis* within Viridiplantae in our phylogenetic analyses (see above; Fig. 1, S3) we cannot reject the alternative placement of this sequence within dinoflagellates using the AU test. Similarly, we observed a copy of IspH being closely related to the stramenopile *N. gaditana*

(Fig. 2, S7), but alternative placements of that sequence within either dinoflagellates or haptophytes cannot be rejected. In *Karlodinium*, we were unable to reject an alternative placement of the Dxs sequences that were observed as close relatives of haptophytes (Fig. 1, S1) within dinoflagellates (Fig. 3). For IspH, however, we were unable to reject the placement of *Karlodinium* DOXP sequences being exclusively of haptophyte origin (Fig. 3). All other *Karlodinium* sequences seem to be of unequivocal dinoflagellate origin (Fig. 3). For dinotoms (*Kryptoperidinium* and *Glenodinium*), the origin of all DOXP sequences from both dinoflagellates and stramenopiles holds under our hypothesis testing framework (Fig. 3). The unexpected placement of a *D. acuminata* IspH copy among the haptophytes (Fig. 2, S7) could not be rejected. Note, that we found copies for all 7 DOXP genes for *D. acuminata* that fell into the dinoflagellate clade.

GC content

Across dinoflagellates (Fig. 4A), GC content of transcriptome contigs was distributed unimodal with a slight left skew, averaging ~59% GC content with 95% of the data contained in the interval 47.6–72.6%; DOXP transcripts ranged from 47.6% to 72.6% GC content (mean = 59.8%; Fig. 4a). By contrast, GC content in *K. foliaceum* and *G. foliaceum* was bimodally distributed, ranging from 24% to 85.5% (Fig. 4B) and 22.8% to 88.7% (Fig. 4C) GC content. In both cases, the lowest point between the two modes of the distributions fell at ~53% GC content. Here, DOXP transcripts were clearly separated into two groups (Fig. 4B, C).

Karenia brevis (Fig. 4D) and *K. micrum* (Fig. 4E) showed a clear unimodal distribution of transcriptome-wide GC content. While GC content in *K. brevis* ranged from ~30% to ~78% the distribution of *Karlodinium micrum* was somewhat tighter with the majority of the data ranging from ~40% to ~60% GC content. *Karenia brevis* DOXP transcripts ranged in GC content from ~50% to ~60% (Fig. 4D) while *K. micrum* DOXP transcripts displayed a slightly lower GC content (Fig. 4E). Both *K. micrum* and *K. veneficum* (48 h after being infected by *Amoebophyra* sp.) showed overlapping distributions of GC content for all contigs in the transcriptome assemblies and DOXP genes (Fig. 4E). However, *Karlodinium veneficum* had been infected with *Amoebophyra* sp. and showed a bimodal distribution with a second peak falling at ~32% GC content; no DOXP transcript fell into this part of the distribution. The second parasite infection experiment on the other hand did not show a comparable bimodal distribution (Fig. 4F). The *A. sanguinea* culture infected by *Amoebophyra* sp. yielded transcriptome contigs with a GC content similar to *K. brevis* and *K. micrum*. GC content of *Amoebophyra* sp. ex. *A. sanguinea* dinospores occupied the lower end of the distribution without being clearly discernible from the combined *Amoebophyra* sp. ex. *A. sanguinea* distribution.

Signal peptides

57 of the 203 dinoflagellate sequences were predicted by SignalP to contain signal peptide cleavage sites ($D = 0.34$). For dinotoms, we found a moderate degree of per site conservation for the motif A↓FAP (arrow indicates cleavage site; Fig. 5). While the first two amino acids were highly conserved across all sequences the latter two were less conserved. In *K. brevis* and *Karlodinium* spp., the cleavage site and initial residues of the transit peptides deviated

strongly from this pattern (Fig. 5). While the last residues preceding the cleavage site were somewhat ambiguous with G or A being most common, the initial residues of the transit peptide were highly conserved on positions 1, 2, and 4 (RRXQ).

DISCUSSION

Among alveolates, ciliates have retained the MVA pathway and lack evidence for any plastid genes or plastid (Eisen et al. 2006), and consequently the DOXP pathway. By contrast, apicomplexan parasites possess nonphoto-synthetic, relic plastids (apicoplasts) and utilize the DOXP pathway exclusively for isoprenoid synthesis (Lim and McFadden 2010). Similarly, we find that dinoflagellates ranging from heterotrophic (*O. marina* and *C. cohnii*) to photosynthetic and mixotrophic (*D. acuminata*) life styles transcribe the full complement of the DOXP pathway while they do not appear to possess or express the alternate cytosolic MVA pathway. Interestingly, the parasitic *Amoebophrya* spp. appear to lack the DOXP pathway (see below).

The generally clear separation of chromist (cryptophytes, haptophytes, and Stramenopiles) DOXP genes from the dinoflagellate DOXP genes, with exception of genes derived from more recent tertiary replacement plastids (see below), is intriguing considering that plastids in alveolates and chromists are all of red algal origin (e.g., Bodyt et al. 2009; Keeling 2009, 2010; Sanchez-Puerta and Delwiche 2008). Apicomplexan DOXP genes were highly diverged from the remainder of the taxa with plastids of red algal origin and were excluded from our analysis, as their long branches may affect phylogenetic reconstructions (cf., Matsuzaki et al. 2007) and consequently our inferences. The rate of sequence evolution in both api-complexans and dinoflagellates may have been accelerated, a pattern expected in parasites (e.g., Paterson et al. 2010; note that some of the earliest branching dinoflagellates are parasites: Bachvaroff et al. 2014), possibly eroding the signal of dinoflagellate DOXP gene origins from chromists. Indeed, phylogenetic relationships of dinoflagellate plastids have been difficult to resolve due to the high relative substitution rates of plastid-associated genes in this clade (e.g., Bachvaroff et al. 2006; Shalchian-Tabrizi et al. 2006).

Plastid replacement and the origins of isoprene synthesis

Isoprene synthesis in dinoflagellates provides insight into gene transfer and replacement during plastid replacement. Most of the dinoflagellate DOXP genes fall within dinoflagellate clades. However, several DOXP genes found in the Kareniaceae, *Karenia* and *Karlodinium*, and dinotoms (*Kryptoperidinium* and *Glenodinium*) grouped with their more recent symbionts. In addition to dinoflagellates with tertiary replacement plastids, our analyses contained *D. acuminata*, a mixotrophic dinoflagellate that is known to temporarily acquire a plastid from the cryptophyte *Gemingera cryophila* (e.g., Park et al. 2008; Wisecaver and Hackett 2010). Surprisingly, one copy of the *D. acuminata* IspH gene fell unambiguously within haptophytes, and placements with cryptophyte or dinoflagellate genes were rejected by the AU tests, suggesting that the additional IspH copy is either a contaminant or was acquired via horizontal gene transfer, as has been documented for a number of other genes in the *D. acuminata* species complex (Wisecaver and Hackett 2010). In both Kareniaceae and

dinotoms, patterns of gene duplications are consistent with their endosymbiont origin rather than other forms of horizontal gene transfer.

For *K. brevis* and *K. veneficum* (including the synonymous *K. micrum*) we find an inconsistent pattern of DOXP gene replacement. *K. brevis* expresses a full complement of dinoflagellate DOXP genes, but we also found three genes that were of likely haptophyte origin (IspE, IspF, and IspH). The unexpected placements of IspD and IspH copies within Viridiplantae and Stramenopiles were poorly supported. This explains why we cannot reject alternative placements of these gene copies (Fig. 3) that are more consistent with our understanding of tertiary endosymbiosis of a haptophyte alga in fucoxanthin-containing dinoflagellates of the Kareniaceae (e.g., Tengs et al. 2000; Yoon et al. 2002), highlighting the need for careful interpretation of the phylogenetic origin of genes from gene trees when evaluating the contributions of horizontal and endosymbiont gene transfer to protist genomes (cf., Burki et al. 2012; Deschamps and Moreira 2012).

In *K. veneficum*, we find potential replacements of dinoflagellate DOXP genes by haptophyte-derived genes. While not unequivocal for Dxs (Fig. 3) we find stronger evidence for a loss of a dinoflagellate derived IspH gene and its replacement by haptophyte derived genes. This result is consistent across two independent datasets for *K. veneficum* (*K. veneficum* and *K. micrum*), suggesting a replacement of vertically transferred DOXP genes with ones of haptophyte origin, leading to a truly hybrid pathway. However, the Dxs and IspH copies of haptophyte origin in *K. veneficum* (and *K. micrum*) were duplicated, leading to some doubt on their exclusive haptophyte origin. For Dxs, we are indeed unable to reject an alternative placement of these sequences within dinoflagellates, leaving the possibility that the placements of both copies within haptophytes could be an artifact of the phylogenetic analysis.

Compared to the Kareniaceae, dinotoms (*Kryptoperidinium* and *Glenodinium*) possess the full complement of DOXP genes of both dinoflagellate and stramenopile origin, a finding consistent with the analysis of Hehenberger et al. (2014). It is well-established that the tertiary endosymbiont in dinotoms is a pennate diatom (Chesnick et al. 1997) and we found that stramenopile DOXP gene copies consistently grouped with diatoms (*T. pseudonana* and *P. tricornutum*) in our analyses. In addition, the two DOXP pathways present in dinotoms can be clearly separated by their nucleotide composition (Fig. 4B, C). The tertiary plastid of diatom origin still possesses a nucleus and its own mitochondrion, and the available evidence suggests that the plastid retains functionality redundant with its host (Imanian and Keeling 2014; Imanian et al. 2012).

Our analyses support the recent finding that the nucleus of the dinotom plastid retained the key metabolic DOXP pathway (Hehenberger et al. 2014), but does not seem to transcribe a complete MVA pathway that is present in diatoms (Lohr et al. 2012). An interesting question that remains to be answered is: what is the target of the dinoflagellate DOXP proteins? Hehenberger et al. (2014) suggested that the host-encoded DOXP proteins may be targeted to the dinotom eyespot which is thought to be a relic plastid and consequently may retain plastid functionality. By contrast, the haptophyte-derived genes in Kareniaceae are likely encoded in the dinoflagellate host genome and targeted to the haptophyte-derived plastid.

The majority of transcripts in our dataset had truncated 5' prime regions. Nonetheless, across 14 dinotom sequences we observed a moderate amount of conservation of the A↓FAP motif (where ↓ represents the cleavage site) (Fig. 5). This is consistent with the transit peptide FAP described for diatoms (Kilian and Kroth 2005; Patron and Waller 2007; see Hehenberger et al. 2014 for a discussion of plastid targeting motifs in dinotoms). For Kareniaceae (21 sequences), our results are in agreement with previous descriptions of a modified transit peptide corresponding to the motif RRXQ (Fig. 5; Patron and Waller 2007). This motif strongly deviates from the peridinin dinoflagellate transit peptide FVAP (Patron and Waller 2007), and likely facilitates DOXP proteins to be targeted to the haptophyte-derived plastid.

Gene transfer from the peridinin plastid may make up a significant portion of the nuclear genome in dinoflagellates (e.g., Bachvaroff et al. 2004; Boldt et al. 2012; Wisecaver and Hackett 2011). By contrast, the contribution of the replacement plastid to the Kareniaceae and dinotom host nucleus appears much smaller, a likely result of the more recent acquisition of these tertiary plastids and potentially lower degree of host-endosymbiont integration (Burki et al. 2014). The endosymbiont in dinotoms is not as reduced as most secondary and tertiary endosymbionts, maintaining a nucleus and mitochondria in addition to a plastid (Imanian et al. 2012), and hence it is relatively unsurprising that it maintains a redundant DOXP pathway at the core of its anabolic metabolism (cf. Hehenberger et al. 2014). In Kareniaceae, the endosymbiont is much more reduced and the haptophyte-derived DOXP genes appear to have been transferred from the plastid to the host nucleus. While not unequivocal, a truly hybrid DOXP pathway of multiple evolutionary origins may exist in *Karlodinium*. However, in most cases the host pathway seems to prevail, being retargeted to the newly acquired plastid. For example, *K. brevis* transcribes its ancestral DOXP pathway, likely targeting the newly acquired haptophyte plastid.

Absence of DOXP pathway in the parasite *Amoebophyra*

In contrast with the remainder of the dinoflagellates, *Amoebophyra* spp. did not express DOXP (or MVA) pathway genes. While we analyzed only a single dataset derived from essentially pure *Amoebophyra* cells (*Amoebophyra* sp. ex. *A. sanguinea* dinospores), neither combined parasite/host dataset yielded DOXP genes likely to be of *Amoebophyra* spp. origin. Nucleotide composition clearly separated host and parasite in *Amoebophyra* sp. ex. *K. veneficum*, with no DOXP gene copy falling into the part of the distribution attributable to *Amoebophyra* sp. (Fig. 4E). For *Amoebophyra* sp. ex. *A. sanguinea* the nucleotide composition did not allow for clear separation between host and parasite (Fig. 4F). However, we found only a single copy for each DOXP gene, which we interpret as being derived from the host *A. sanguinea*.

In conjunction with other plastid-targeted genes, transcription of DOXP genes in *Amoebophyra* spp.'s oyster-parasitizing relative *P. marinus* was used as evidence to infer possession of a cryptic nonphotosynthetic plastid (Matsuzaki et al. 2007; Robledo et al. 2011). So far, ultra-structural work has not discovered a candidate plastid in *Amoebophyra* sp. (Miller et al. 2012) and we did not find evidence of expression of plastid-targeted DOXP genes here. The complete lack of isoprenoid precursor synthesis genes in our dataset opens

the possibility that *Amoebophyra* spp. are dependent on their host for de novo synthesis of isoprenoid metabolic compounds, for example by importing host sterols (Place et al. 2009). Similarly to *Amoebophyra* spp., the crustacean parasite *Hematodinium* sp. was shown to lack both a plastid and any evidence for a functional DOXP or MVA pathway (Gornik et al. 2015).

Considering that we investigated *Amoebophyra* spp. both prior to and after infection, it seems unlikely that we did not detect the DOXP pathway simply because *Amoebophyra* sp. was not highly transcriptionally active, as might be expected during dispersal via the dinospore, although we cannot exclude the possibility that isoprenoid genes are expressed at some point in the life cycle that we did not sample. Downregulation of DOXP expression does, for example, occur during part of the life cycle of the apicomplexan malaria parasite *Plasmodium* (e.g., Cassera et al. 2007).

Amoebophyra spp. possess a genome that lacks the high copy number of genes typical for many dinoflagellates (Bachvaroff et al. 2009). This has been interpreted as an ancestral feature for dinoflagellates, given that *Amoebophyra* spp. branch near the base of the dinoflagellates (Bachvaroff et al. 2009, 2014). Alternatively, however, genome reduction could be a derived adaptation to the parasitic life-style. Questions to address in light of the data presented here are whether genes downstream of the DOXP pathway are being transcribed that would allow *Amoebophyra* spp. to rely on importing and processing precursor compounds of isoprenoid metabolism from its host. By comparison, apicomplexan parasites possess the full complement of the DOXP pathway (cf. Matsuzaki et al. 2007).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Endymion Cooper for generously sharing unpublished green algal transcriptome assemblies and Gregory Mendez for insightful discussions. We are also indebted for two reviewers who provided critical comments that helped us improve an earlier version of this manuscript. Funding for this project came through NSF grants DEB-1036506, DEB-1046075, and DEB-0629624. TR wishes to acknowledge an HHMI International Research Program Supplement.

LITERATURE CITED

- Abascal F, Zardoya R & Posada D 2005 ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, 21:1–17.
- Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ 1990 Basic local alignment search tool. *J. Mol. Biol.*, 215:403–410. [PubMed: 2231712]
- Archibald JM 2009 The puzzle of plastid evolution. *Curr. Biol.*, 19:R81–R88. [PubMed: 19174147]
- Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM & Delwiche CF 2004 Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist*, 155:65–78. [PubMed: 15144059]
- Bachvaroff TR, Gornik SG, Concepcion GT, Waller RF, Mendez GS, Lippmeier JC & Delwiche CF 2014 Dinoflagellate phylogeny revisited: using ribosomal proteins to resolve deep branching dinoflagellate clades. *Mol. Phylogenet. Evol.*, 70:314–322. [PubMed: 24135237]

- Bachvaroff TR, Handy SM, Place AR & Delwiche CF 2011 Alveolate phylogeny inferred using concatenated ribosomal proteins. *J. Eukaryot. Microbiol*, 58:223–233. [PubMed: 21518081]
- Bachvaroff TR, Place AR & Coats DW 2009 Expressed sequence tags from *Amoebophrya* sp. infecting *Kariodinium veneficum*: comparing host and parasite sequences. *J. Eukaryot. Microbiol*, 56:531–541. [PubMed: 19883441]
- Bachvaroff TR, Sanchez-Puerta MV & Delwiche CF 2006 Rate variation as a function of gene origin in plastid-derived genes of peridinin-containing dinoflagellates. *J. Mol. Evol*, 62:42–52. [PubMed: 16408243]
- Bayer T, Aranda M, Sunagawa S, Yum LK, DeSalvo MK, Lindquist E, Coffroth MA, Voolstra CR & Medina M 2012 Symbiodinium transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS ONE*, 7: e35269. doi:10.1371/journal.pone.0035269.
- Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J & Sayers EW 2014 GenBank. *Nucleic Acids Res*, 42: D32–D37.
- Bergholtz T, Daugbjerg N, Moestrup O & Fernandez-Tejedor M 2006 On the identity of *Kariodinium veneficum* and description of *Kariodinium armiger* sp. nov. (Dinophyceae), based on light and electron microscopy, nuclear-encoded LSU rDNA, and pigment composition. *J. Phycol*, 42:170–193.
- Bodt A, Stiller JW & Mackiewicz P 2009 Chromalveolate plastids: direct descent or multiple endosymbioses? *Trends Ecol. Evol*, 24:119–121. [PubMed: 19200617]
- Bohlmann J & Keeling CI 2008 Terpenoid biomaterials. *Plant J*, 54:656–669. [PubMed: 18476870]
- Boldt L, Yellowlees D & Leggat W 2012 Hyperdiversity of genes encoding integral light-harvesting proteins in the dinoflagellate *Symbiodinium* sp. *PLoS ONE*, 7:e47456. doi:10.1371/journal.pone.0047456.
- Burki F, Flegontov P, Oborník M, Cihlář J, Pain A, Lukeš J & Keeling PJ 2012 Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. *Genome Biol. Evol*, 4:626–635.
- Burki F, Imanian B, Hehenberger E, Hirakawa Y, Maruyama S & Keeling PJ 2014 Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot. Cell*, 13:246–255. [PubMed: 24297445]
- Cassera MB, Merino EF, Peres VJ, Kimura EA, Wunderlich G & Katzin AM 2007 Effect of fosmidomycin on metabolic and transcript profiles of the methylerythritol phosphate pathway in *Plasmodium falciparum*. *Mem. Inst. Oswaldo Cruz*, 102:377–383.
- Chesnick JM, Wellbrock U & Medlin LK 1997 Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J. Eukaryot. Microbiol*, 44:314–320. [PubMed: 9225445]
- Coats DW & Park MG 2002 Parasitism of photosynthetic dinoflagellates by three strains of *Amoebophrya* (Dinophyta): parasite survival, infectivity, generation. *J. Phycol*, 528:520–528.
- Cocquyt E, Gile GH, Leliaert F, Verbruggen H, Keeling PJ & De Clerck O 2010 Complex phylogenetic distribution of a non-canonical genetic code in green algae. *BMC Evol. Biol*, 10:327. [PubMed: 20977766]
- Cooper ED, Bentlage B, Gibbons TR, Bachvaroff TR & Delwiche CF 2014 Metatranscriptome profiling of a harmful algal bloom. *Harmful Algae*, 37:75–83. [PubMed: 25484636]
- Deschamps P & Moreira D 2012 Reevaluating the green contribution to diatom genomes. *Genome Biol. Evol*, 4:683–688. [PubMed: 22684208]
- Dodge JD & Lee JJ 2000 Phylum Dinoflagellata (Bütschli, 1885) In: Lee JJ, Leedale GF & Bradbury P (ed.), *An Illustrated Guide to the Protozoa: Organisms Traditionally Referred to as Protozoa, or Newly Discovered Groups*. Allen Press, Lawrence, KS.
- Eddy SR 2011 Accelerated profile HMM searches. *PLoS Comput. Biol*, 10:e1002195.
- Edgar RC 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113. doi:10.1186/1471-2105-5-113. [PubMed: 15318951]
- Efron B, Halloran E & Homes S 1996 Bootstrap confidence levels for phylogenetic trees. *PNAS*, 93:7085–7090. [PubMed: 8692949]
- Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M,

- Carlton JM, Smith RK Jr, Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz AP, Asai DJ, Wilkes DE, Wang Y, Cai H, Collins K, Stewart BA, Lee SR, Wilamowska K, Weinberg Z, Ruzzo WL, Wloga D, Gaertig J, Frankel J, Tsao CC, Gorovsky MA, Keeling PJ, Waller RF, Patron NJ, Cherry JM, Stover NA, Krieger CJ, del Toro C, Ryder HF, Williamson SC, Barbeau RA, Hamilton EP & Orias E 2006 Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol*, 4:e286. doi:10.1371/journal.pbio.0040286.
- Eisenreich W, Bacher A, Arigoni D & Rohdich F 2004 Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cell. Mol. Life Sci*, 61:1401–1426. [PubMed: 15197467]
- Gornik SG, Febrimarsa AMC, Cassin AM, MacRae JI, Ramaprasad A, Rchiad Z, McConville MJ, Bacic A, McFadden GI, Pain A & Waller RF 2015 Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *PNAS*, 112:5767–5772. [PubMed: 25902514]
- Gould SB, Waller RF & McFadden GI 2008 Plastid evolution. *Annu. Rev. Plant Biol*, 59:491–517. [PubMed: 18315522]
- Gouy M, Guindon S & Gascuel O 2010 SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol*, 27:221–224. [PubMed: 19854763]
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N & Regev A 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol*, 29:644–652.
- Guillard RRL & Hargraves PE 1993 *Stichochrysis immobilis* is a diatom, not a chrysophyte. *Phycologia*, 32:234–236.
- Guindon S & Gascuel O 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*, 52:696–704. [PubMed: 14530136]
- Guiry MD & Guiry GM 2015 *AlgaeBase*. Natl. Univ. Ireland, Galw. <http://www.algaebase.org> (accessed 29 January 2015).
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N & Regev A 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc*, 8:1494–1512. [PubMed: 23845962]
- Hackett JD, Anderson DM, Erdner DL & Bhattacharya D 2004 Dinoflagellates: a remarkable evolutionary experiment. *Am. J. Bot*, 91:1523–1534. [PubMed: 21652307]
- Hehenberger E, Imanian B, Burki F & Keeling PJ 2014 Evidence for the retention of two evolutionary distinct plastids in dinoflagellates with diatom endosymbionts. *Genome Biol. Evol*, 6:2321–2334. [PubMed: 25172904]
- Howe CJ, Barbrook C, Nisbet RER, Lockhart PJ & Larkum WD 2008 The origin of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci*, 363:2675–2685. [PubMed: 18468982]
- Hunter WN 2007 The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J. Biol. Chem*, 282:21573–21577. [PubMed: 17442674]
- Imanian B & Keeling PJ 2014 Horizontal gene transfer and redundancy of tryptophan biosynthetic enzymes in dinotoms. *Genome Biol. Evol*, 6:333–343. [PubMed: 24448981]
- Imanian B, Pombert JF, Dorrell RG, Burki F & Keeling PJ 2012 Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. *PLoS ONE*, 7:e43763. doi:10.1371/journal.pone.0043763.
- Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Tu I, Eberl M, Zeidler J, Lichtenthaler HK, Soldati D & Beck E 1999 Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science*, 285:1573–1576. [PubMed: 10477522]
- Keeling PJ 2009 Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J. Eukaryot. Microbiol*, 56:1–8. [PubMed: 19335769]
- Keeling PJ 2010 The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci*, 365:729–748. [PubMed: 20124341]

- Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Rynearson T, Schilling KB, Schroeder DC, Simpson AG, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaultot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A & Worden AZ 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:e1001889. doi:10.1371/journal.pbio.1001889.
- Kilian O & Kroth PG 2005 Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. *Plant J*, 41:175–183. [PubMed: 15634195]
- Kirby J & Keasling JD 2009 Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu. Rev. Plant Biol*, 60:335–355. [PubMed: 19575586]
- Lartillot N, Lepage T & Blanquart S 2009 PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, 25:2286–2288. [PubMed: 19535536]
- Laule O, Fürholz A & Chang H 2003 Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *PNAS*, 100:6866–6871. [PubMed: 12748386]
- Lichtenthaler HK 1999 The 1-deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol*, 50:47–65. [PubMed: 15012203]
- Lim L & McFadden GI 2010 The evolution, metabolism and functions of the apicoplast. *Philos. Trans. R. Soc. Lond. B Biol. Sci*, 365:749–763. [PubMed: 20124342]
- Lohr M, Schwender J & Polle JEW 2012 Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci*, 185–186:9–22.
- Lombard J & Moreira D 2011 Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. *Mol. Biol. Evol*, 28:87–99. [PubMed: 20651049]
- Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K & Nozaki H 2007 A cryptic algal group unveiled: a plastid biosynthesis pathway in the oyster parasite *Perkinsus marinus*. *Mol. Biol. Evol*, 25:1167–1179.
- Miller JJ, Delwiche CF & Coats DW 2012 Ultrastructure of *Amoebophrya* sp. and its changes during the course of infection. *Protist*, 163:720–745. [PubMed: 22209083]
- Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng J-F & Bhattacharya D 2006 Chimeric plastid proteome in the Florida “red tide” dinoflagellate *Karenia brevis*. *Mol. Biol. Evol*, 23:2026–2038. [PubMed: 16877498]
- Park MG, Park JS, Kim M & Yih W 2008 Plastid dynamics during survival of *Dinophysis caudata* without its ciliate prey. *J. Phycol*, 44:1154–1163. [PubMed: 27041712]
- Parra G, Bradnam K & Korf I 2007 CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23:1061–1067. [PubMed: 17332020]
- Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, Fenton A, Hall N & Brockhurst MA 2010 Antagonistic coevolution accelerates molecular evolution. *Nature*, 464: 275–278. [PubMed: 20182425]
- Patron NJ & Waller RF 2007 Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *BioEssays*, 29:1048–1058. [PubMed: 17876808]
- Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME & Stamatakis A 2009 How many bootstrap replicates are necessary? *J. Comput. Biol*, 3:337–354.
- Petersen TN, Brunak S, von Heijne G & Nielsen H 2011 SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, 8:785–786. [PubMed: 21959131]
- Place AR, Bai X, Kim S, Sengco MR & Coats WD 2009 Dinoflagellate host-parasite sterol profiles dictate karlotoxin sensitivity. *J. Phycol*, 45:375–385. [PubMed: 27033816]

- Reece KS, Siddall ME, Bureson EM & Graves JE 1997 Phylogenetic analysis of Perkinsus based on actin gene sequences. *J. Parasitol*, 83:417–423. [PubMed: 9194820]
- Robledo JF, Caler E, Matsuzaki M, Keeling PJ, Shanmugam D, Roos DS & Vasta GR 2011 The search for the missing link: a relic plastid in Perkinsus? *Int. J. Parasitol*, 41:1217–1229. [PubMed: 21889509]
- Rohmer M 1999 The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep*, 16:565–574. [PubMed: 10584331]
- Rohmer M, Knani M, Simonin P, Sutter B & Sahm H 1993 Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J*, 295:517–524. [PubMed: 8240251]
- Saldarriaga JF, McEwan ML, Fast NM, Taylor FJR & Keeling PJ 2003 Multiple protein phylogenies show that Oxyrrhis marina and Perkinsus marinus are early branches of the dinoflagellate lineage. *Int. J. Syst. Evol. Microbiol*, 53:355–365. [PubMed: 12656195]
- Sanchez-Puerta MV & Delwiche CF 2008 A hypothesis for plastid evolution in chromalveolates. *J. Phycol*, 44:1097–1107. [PubMed: 27041706]
- Sanchez-Puerta MV, Lippmeier JC, Apt KE & Delwiche CF 2007 Plastid genes in a non-photosynthetic dinoflagellate. *Protist*, 158:105–117. [PubMed: 17150410]
- Shalchian-Tabrizi K, Skånseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen A, Tengs T & Jakobsen KS 2006 Heterotachy processes in rhodophyte-derived secondhand plastid genes: implications for addressing the origin and evolution of dinoflagellate plastids. *Mol. Biol. Evol*, 23:1504–1515. [PubMed: 16699169]
- Shimodaira H 2002 An approximately unbiased test of phylogenetic tree selection. *Syst. Biol*, 51:492–508. [PubMed: 12079646]
- Shimodaira H & Hasegawa M 2001 CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, 17:1246–1247. [PubMed: 11751242]
- Slamovits CH & Keeling PJ 2007 Plastid-derived genes in the nonphotosynthetic alveolate Oxyrrhis marina. *Mol. Biol. Evol*, 25:1297–1306.
- Talavera G & Castresana J 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol*, 56:564–577. [PubMed: 17654362]
- Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF & Jakobsen KS 2000 Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol*, 17:718–729. [PubMed: 10779532]
- Wisecaver JH, Brosnahan ML & Hackett JD 2013 Horizontal gene transfer is a significant driver of gene innovation in dinoflagellates. *Genome Biol. Evol*, 5:2368–2381. [PubMed: 24259313]
- Wisecaver JH & Hackett JD 2010 Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate Dinophysis acuminata. *BMC Genom*, 11:366. doi:10.1186/1471-2164-11-366.
- Wisecaver JH & Hackett JD 2011 Dinoflagellate genome evolution. *Annu. Rev. Microbiol*, 65:369–387. [PubMed: 21682644]
- Yoon HS, Hackett JD & Bhattacharya D 2002 A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *PNAS*, 99:11724–11729. [PubMed: 12172008]
- Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL & Bhattacharya D 2005 Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol. Biol. Evol*, 22:1299–1308. [PubMed: 15746017]

Figure 1.

Schematic of the DOXP pathway (top left). In addition, the maximum likelihood phylogenies for Dxs, Dxr, and IspD are shown. Branch shading represents support measured by 500 nonparametric bootstrap replicates. Placement of sequences for Kareniaceae and dinotoms are indicated by circles (*Karenia*), diamonds (*Karlodinium*), and squares (dinotoms).

Figure 2.

Maximum likelihood phylogenies for IspE, IspF, IspG, and IspH. Branch shading represents support measured by 500 nonparametric bootstrap replicates. Placement of sequences for Kareniaceae and dinotoms are indicated by circles (*Karenia*), diamonds (*Karlodinium*), and squares (dinotoms).

Figure 3. Summary of results from the constrained tree-topology tests using the approximately unbiased test. Solid symbols indicate that the placement of a particular gene was observed in the best tree while hollow symbols indicate that the placement was not observed but could not be rejected despite the lower likelihood of the resulting phylogenetic tree. 1: the alternative placement of this gene copy within dinoflagellates cannot be rejected; 2: alternative placements of this gene copy within dinoflagellates or haptophytes cannot be rejected.

Figure 4.

Nucleotide composition of transcriptomes and DOXP genes. **A.** GC content across all contigs in the peridinin dinoflagellate transcriptome assemblies analyzed; crosses indicate DOXP transcript distribution. **B., C.** Bimodal GC content distributions of *Kryptoperidinium foliaceum* and *Glenodiinium foliaceum* transcriptome assemblies with DOXP transcript placements (crosses). The GC content of mRNAs for *Thalassiosira pseudonana* CCMP1335 and *Phaeodactylum tricornutum* CCAP1055/1 was calculated from NCBI Genbank data and overlaid for comparison. **D.** GC content across all contigs in the *Karenia brevis* transcriptome assembly with *K. brevis* DOXP genes indicated by crosses. GC content across all haptophyte assemblies overlaid for comparison. **E.** Bimodal distribution of GC content for *Karlodinium micrum* and *Amoebophyra* sp. ex. *Karlodinium veneficum* 48 h after infection. GC content across all haptophyte assemblies overlaid for comparison. **F.** GC content of *Amoebophyra* sp. ex. *Akashiwo sanguinea* dinospores transcriptome assembly and *Amoebophyra* sp. ex. *A. sanguinea* 48 h after infection. Crosses indicate DOXP genes identified from the *Amoebophyra* sp. ex. *A. sanguinea* datasets that contained *A. sanguinea*

48 h after being infected by *Amobophyra* sp.; no DOXP genes were identified in the *Amobophyra* sp. dinospore transcriptome assembly.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Sequence logos of the N terminal region of the inferred transit peptide (position +1 through +6) including cleavage site and last residue of the signal peptide (position -1). Vertical arrow indicates location of predicted cleavage site. Upper logo (14 sequences): dinotoms (*Kryptoperidinium foliaceum* and *Glenodinium foliaceum*). Lower logo (21 sequences): fucoxanthin dinoflagellates (*Karenia* and *Karlodinium*).