Plastid-Derived Genes in the Nonphotosynthetic Alveolate *Oxyrrhis marina*

Claudio H. Slamovits and Patrick J. Keeling

Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

Reconstructing the history of plastid acquisition and loss in the alveolate protists is a difficult problem because our knowledge of the distribution of plastids in extant lineages is incomplete due to the possible presence of cryptic, nonphotosynthetic plastids in several lineages. The discovery of the apicoplast in apicomplexan parasites has drawn attention to this problem and, more specifically, to the question of whether many other nonphotosynthetic lineages also contain cryptic plastids or are derived from plastid-containing ancestors. *Oxyrrhis marina* is one such organism: It is a heterotrophic, early-branching member of the dinoflagellate lineage for which there is no evidence of a plastid. To investigate the possibility that *O. marina* is derived from a photosynthetic ancestor, we have generated and analyzed a large-scale EST database and searched for evidence of plastid-derived genes. Here, we describe 8 genes whose phylogeny shows them to be derived from plastid-targeted homologues. These genes encode proteins from several pathways known to be localized in the plastids of other algae, including synthesis of tetrapyrroles, isoprenoids, and amino acids, as well as carbon metabolism and oxygen detoxification. The 5′ end of 5 cDNAs were also characterized using cap-dependent or spliced leader–mediated reverse transcriptase–polymerase chain reaction, revealing that at least 4 of these genes have retained leaders that are similar in nature to the plastid-targeting signals of other secondary plastids, suggesting that these proteins may be targeted to a cryptic organelle. At least 2 genes do not encode such leaders, and their products may presently function in the cytosol. Altogether, the presence of plastid-derived genes in *O. marina* shows that its ancestors contained a plastid, and the pathways represented by the genes and presence of targeting signals on at least some of the genes further suggests that a relict organelle may still exist to fulfill plastid metabolic functions.

Introduction

The acquisition of photosynthesis by endosymbiosis triggered the radiation of many of the most successful and diverse eukaryotic lineages and had a tremendous impact on the evolution of life in general. The origin of photosynthesis in eukaryotes has been traced back to a single primary endosymbiotic event, but plastids have also subsequently spread between eukaryotic groups by secondary and tertiary endosymbioses (Palmer 2003; Keeling 2004). Details of this complex evolution are still debated, and one major problem in resolving several of these debates lays in the accurate determination of the distribution of plastids. Documenting the distribution of photosynthesis is relatively straightforward because the presence of photosynthesis tends to be very obvious, but documenting the presence of a plastid in the absence of photosynthesis can be more difficult. Plastids in many lineages have lost photosynthesis, and in some cases, the organelles are so degenerate that they are difficult to recognize and these are considered "cryptic" plastids (Williams and Hirt 2004). In addition, we know very little about how and why plastids are lost outright, how likely such an event might be, or what kinds of genes might be retained in cases where the organelle is completely eliminated (Huang et al. 2004; Tyler et al. 2006).

The difficulty in distinguishing between absence of a plastid and absence of photosynthesis is most strikingly illustrated by the Apicomplexa. This group of obligate intracellular parasites is well studied due to their medical significance (e.g., the malaria parasite, *Plasmodium*), and yet their cryptic, nonphotosynthetic plastid, called the apicoplast, was only identified in the 1990s and was only recognizable as a plastid based on the sequences of genes encoded in and derived from its genome (Williamson et al. 1994; Wilson et al. 1996). The apicoplast has now been investigated in some detail and is known to be involved in biosynthesis of fatty acids, isoprenoids, and heme (Roos et al. 2002; Foth and McFadden 2003). All these functions are serviced by proteins encoded in the nuclear genome and the proteins posttranslationally targeted to the organelle, but for the most part, they retain a phylogenetic signal of their cyanobacterial/plastid ancestry (Waller et al. 1998).

The discovery of apicoplasts has renewed interest in plastid evolution in its close relatives too and specifically in the possibility that other nonphotosynthetic relatives evolved from photosynthetic ancestors. The apicomplexans are sisters to the dinoflagellates, about half of which possess a photosynthetic plastid. Interestingly, many of the nonphotosynthetic dinoflagellates seem to branch at or near the base of the clade, raising questions about when the dinoflagellate lineage acquired its plastid and whether plastids have been lost many times or only during photosynthesis. Recently, molecular evidence of a relict plastid has been discovered in the nonphotosynthetic dinoflagellate *Crypthecodinium cohnii* (Sanchez-Puerta et al. 2007). The position of *C. cohnii* in the tree of dinoflagellates already suggested that it evolved from plastid-containing dinoflagellates (Saldarriaga et al. 2001), but these data suggest that several genes were retained and the organelle likely still functions in the cell. More surprisingly, plastid-derived genes have also been described in the oyster parasite *Perkinsus marinus* (Stelter et al. 2007), which is a deep-branching member of the dinoflagellate lineage. At the same time, a 4-membrane–bounded organelle has been also reported in *Perkinsus atlanticus* (Teles-Grilo et al. 2007). The presence of plastids or molecular remnants of plastids in such deep-branching dinoflagellates suggests that plastids may be pervasive in the several nonphotosynthetic lineages that diverged early from the dinoflagellates, but most of these ancient lineages remain unexamined and the distribution of relict plastids or molecular remnants from them is unknown. Even more importantly, with the exception of the apicoplast, little is known about the function of cryptic plastids in these lineages, so it is impossible to contrast the functional degeneration of basal dinoflagellate plastids with that of the apicoplast.

Key words: plastid evolution, Chromalveolates, apicoplast, dinoflagellates, *Oxyrrhis*.

E-mail: pkeeling@interchange.ubc.ca.

Mol. Biol. Evol. 25(7):1297–1306. 2008
doi:10.1093/molbev/msn075
Advance Access publication April 2, 2008

Here, we have investigated the possibility of a plastid-bearing ancestry in the marine flagellate *Oxyrrhis marina*, a free-living, cosmopolitan and ecologically important phagotrophic alveolate that feeds on algae and bacteria. The phylogenetic position of *O. marina* has been studied in some detail and is known to be a deep-branching member of the dinoflagellate lineage, but one that diverged sometime after *Perkinsus* (Saldarriaga et al. 2003; Leander and Keeling 2004; Slamovits et al. 2007). We sequenced 18,000 expressed sequence tags (ESTs) from *O. marina* and found a number of plastid-derived genes. These genes encode proteins involved in many metabolic pathways typically found in plastids, including tetrapyrrole, isoprenoid and amino acid biosynthesis, carbon metabolism, and detoxification, suggesting a greater metabolic contribution than has been retained in the apicoplast. Analysis of N-terminal regions revealed leader sequences with characteristics of plastid-targeting signals on some but not all proteins, suggests that a cryptic plastid may still exist in *O. marina*, but that some plastid-derived proteins might now function in the cytosol.

Materials and Methods

Strains, Cultivation, and EST Library Construction

Oxyrrhis marina strain CCMP 1788 was cultivated axenically in Droop's Ox-7 medium. Twenty liters of culture was harvested in a continuous flow centrifuge and stored in Trizol (Invitrogen, Carlsbad, CA). Total RNA was prepared in 20-ml batches according to the manufacturer's directions, resulting in 2 ug of RNA. A directional cDNA library was constructed in pBluescript II SK using *Eco*R1 and *Xho*I sites (Amplicon Express, Pullman, WA) and shown to contain 5.3×10^5 colony-forming units. In all, 23,702 clones were picked and 5′-end sequenced, resulting in 18,012 EST sequences that assembled into 9,876 unique clusters using TBestDB (O'Brien et al. 2007).

Identification of Plastid-Derived Genes and Analysis of N-Terminal Leaders

We used netblast to compare the 9,876 clusters assembled by tbESTdb (O'Brien et al. 2007) against the nonredundant NCBI database. Candidate plastid-derived genes were identified by closest sequence similarity to plant, algal, and/or cyanobacterial genes. Integrity of the 5′ end was assessed by looking for the presence of the spliced leader described by Zhang et al. (2007). None of the 8 selected clusters encoded the leader. We compared the translated protein sequences to known homologues to evaluate the amount of missing sequence. Only the 2 versions of carbonic anhydrase appeared to be complete.

Characterization of N-Terminal Sequences by 5′-Rapid Amplification of cDNA Ends and Spliced Leader–Mediated PCR

Total *O. marina* RNA was used to amplify 5′ ends of genes of potential plastid ancestry using the 5′ cap–mediated RML–rapid amplification of cDNA ends (RACE) procedure, following the manufacturer's direction (Ambion, Austin, TX). In all, 20-bp gene-specific reverse primers were designed to regions between 150 and 250 bp downstream of the start of the ESTs for the following genes: carbonic anhydrase, cysteine synthetase, hydroxymethylbilane synthetase (HMBS), ribulose 5-P isomerase (RPI), ascorbate peroxidase, glutamine synthetase, ketol-acid reductoisomerase (KARI), and 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR). For spliced leader–mediated polymerase chain reaction (PCR), reverse transcriptase–polymerase chain reaction (RT-PCR) was performed using the reverse gene-specific primers in combination with a forward primer based on the sequence of the spliced leader described in (Zhang et al. 2007). New sequences were deposited in GenBank under accession numbers EU551745–EU551753.

Phylogenetic Analyses

Candidate *O. marina* genes with plastid ancestry were subject to phylogenetic analyses with a wide variety of homologues. Phylogenies were initially screened using a rapid Neighbor-Joining (NJ) analysis implemented in PAUP 4.0b10 (Swofford 2002). In cases where this did not exclude the possibility of a plastid origin, phylogenies were inferred by maximum likelihood (ML) using PhyML (Guindon et al. 2005) and Bayesian analysis using MrBayes 3 (Ronquist and Huelsenbeck 2003). For ML, a WAG model of amino acid substitution was used with 8 categories of rate variation under a gamma distribution plus invariable sites. Bootstrap support was tested with 300 replicates under the same conditions. For Bayesian analyses, a gamma correction with 8 categories and proportion of invariable sites was used, and 4 Monte Carlo Markov chains were also used (default temperature = 0.2). In each case, a total of 1,000,000 generations was calculated with trees sampled every 100 generations and with burn-in of 200,000 generations. A majority rule consensus tree was constructed from post–burn-in trees with PAUP* 4.0.

Results and Discussion

Oxyrrhis marina Encodes Nucleus-Encoded Plastid-Derived Genes

About 18,012 *O. marina* ESTs were assembled to 9,876 unique clusters. From these, 20 genes were identified as candidate plastid-derived genes by similarity to plant and algal homologues using BlastX. The phylogeny of 8 genes was consistent with a plastid origin (table 1), whereas the remainder were either unresolved or a plastid origin was ruled out by more detailed analyses (data not shown). Interestingly, the deduced metabolic functions of these 8 genes are also consistent with a plastid origin because the pathways to which they belong are known from many plastid lineages, and in several cases, the gene in question is only known from plastids or bacteria. In some cases, homologues are known from *P. marinus* and various dinoflagellates for which EST sequencing has also been carried out and plastid-targeted genes completely sequenced (e.g., *C. cohnii*, *Karlodinium micrum*, or *Heterocapsa triquetra*) (Patron et al. 2005, 2006; Waller et al. 2006; Sanchez-Puerta et al. 2007).

Table 1
***Oxyrrhis marina* genes derived from plastid-targeted homologues. 5′ End: Indicates whether or no we obtained the complete 5′ end of the gene. SP: Signal peptide prediction.**

GenBank Gene	5′ End	SP	Function
HMBS	Complete[a]	Yes	Tetrapyrrole synthesis
DXR	Incomplete	Yes	Isoprenoid synthesis
KARI	Incomplete	No	Amino acid metabolism
Glutamine synthetase	Complete[b]	No	Amino acid metabolism
Cysteine synthetase	Incomplete	No	Amino acid metabolism
Carbonic anhydrase-1	Complete[a]	Yes	CO_2/H_2CO_3 interconversion
Carbonic anhydrase-2	Complete[a]	Yes	CO_2/H_2CO_3 interconversion
RPI	Incomplete	Yes	Calvin cycle/pentose phosphate pathway
Ascorbate peroxidase	Complete[b]	No	Ascorbate metabolism/ peroxide detoxification

[a] 5′ RACE.
[b] Spliced leader RT-PCR.

Phylogenetic analysis of 6 genes recovered a typical "plastid" clade consisting of proteins known or inferred to be plastid targeted from a variety of plants and algae (figs. 1 and 2). In each case, *O. marina* fell with this clade, although in most instances its position within the plastid clade was not resolved with any support. This pattern is most clearly illustrated by ascorbate peroxidase, where the *O. marina* sequence falls within a strongly supported clade of plastid-targeted homologues from plant and algae (100% support), and within this clade, the *O. marina* gene branches with a subgroup of dinoflagellates with 100% support (fig. 1*A*). We also identified a previously unrecognized but related ascorbate peroxidase from *P. marinus* ESTs, and this gene is a member of the same strongly supported (94%) subclade of algal sequences as *O. marina* (ascorbate peroxidase has not been found in Apicomplexa to date). Support for *O. marina* within the plant/algal clade is also very strong in KARI phylogeny (100%: fig. 1*B*). A similar relationship between *O. marina* and plant and algal plastid homologues was also observed with moderate support in DXR phylogeny (76%: fig. 2*A*). Partial DXR sequences were also found in EST surveys of the dinoflagellates *C. cohnii* and *Alexandrium tamarense*. The *C. cohnii* sequence represents only the extreme 5′ end and its position in the tree is not resolved (data not shown), but its 5′ end is consistent with the presence of a plastid-targeted presequence (Sanchez-Puerta et al. 2007). The *A. tamarense* sequence represents only the extreme 3′ end of the gene, and it branches within the plastid clade with moderate support (data not shown). In RPI phylogeny, a similar plastid clade is also recovered with weak support (61% bootstrap; fig. 2*B*), and the *O. marina* sequence is not resolved within that clade. In HMBS phylogeny (fig. 2*C*), the support for *O. marina* with plants is also weak, but there is strong support (95%) for an extended clade including a number of alpha-proteobacteria, all to the exclusion of eukaryotic cytosolic homologues. In this phylogeny, we did not recover a relationship between plastids and cyanobacteria, as it was reported by Obornik and Green (2005), but the backbone of this tree was not supported so this is not unexpected. Lastly, the *O. marina* glutamine synthetase also weakly associated with plants in ML trees and in Bayesian analyses is sister to a homologue from the diatom *Skeletonema costatum* (fig. 2*D*), which has been shown to function in the plastid (Robertson and Tartar 2006).

The phylogeny of the remaining 2 genes never recovered a unique clade of plastid-targeted proteins, but in both cases, the position of the *O. marina* protein strongly suggests a plastid origin. In both carbonic anhydrase and

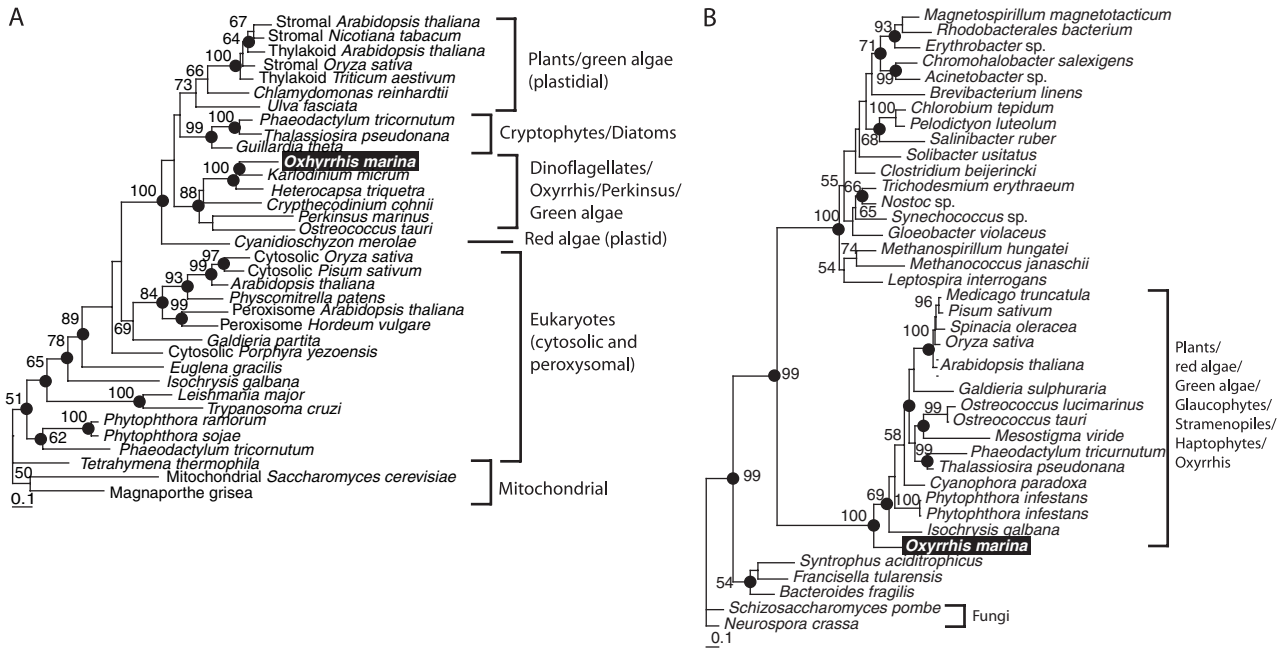

Fig. 1.—Maximum likelihood trees inferred from *Oxyrrhis marina* proteins with a strongly supported affinity to plant and green algal plastid versions. (*A*) Ascorbate peroxidase. (*B*) KARI. Numbers at the nodes represent percentage of bootstrap support over 300 replicas. Only values >50% are shown. Black circles indicate nodes with Bayesian posterior probability of 0.9 or higher.

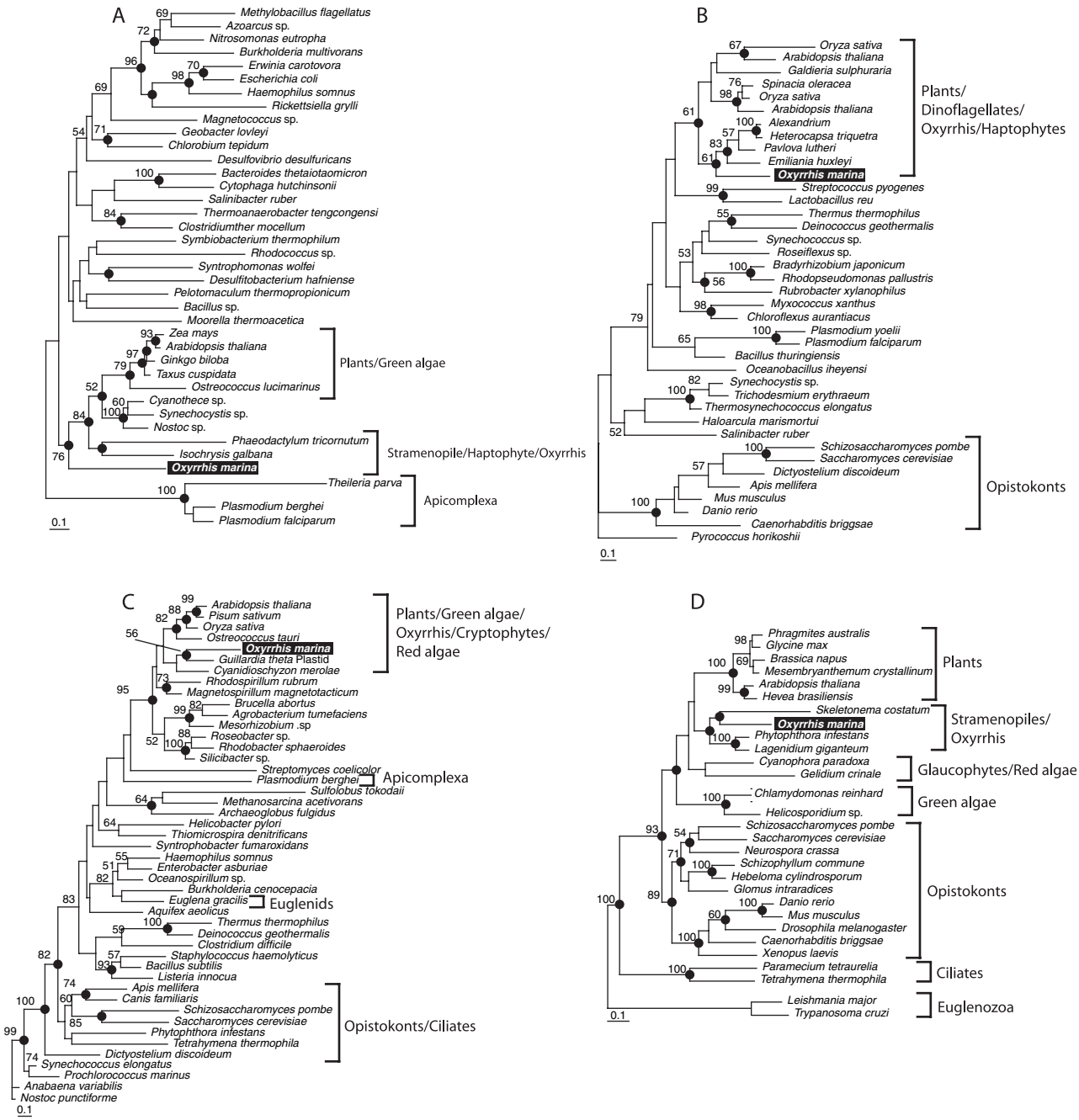

Fig. 2.—Maximum likelihood trees inferred from *Oxyrrhis marina* proteins exhibiting moderate support to a plant-like plastid affinity. (*A*) DXR. (*B*) RPI. (*C*) HMBS. (*D*) Glutamine synthetase. Numbers at the nodes represent percentage of bootstrap support over 300 replicas. Only values >50% are shown. Black circles indicate nodes with Bayesian posterior probability of 0.9 or higher.

cysteine synthetase phylogenies, *O. marina* falls within an algal clade, but far from plants (fig. 3). *Oxyrrhis marina* has 2 copies of carbonic anhydrase that branch together with strong support (99%)with 2 homologues from the dinoflagellate *H. triquetra*, also with strong support (99%). These are in turn part of a larger clade including other secondary algae, euglenids, and haptophytes. Despite their distant relationship to other plastid-targeted proteins, the *H. triquetra* homologues have been shown to encode typical plastid-targeting leaders (Patron et al. 2005), supporting their location in the plastid. Indeed, beta-carbonic anhydrase is known from many bacteria, but within the eukaryotes is only known in plastid-bearing lineages, where it is generally targeted to the plastid (Smith et al. 1999) (but see Discussion). In the case of cysteine synthase, the *O. marina* gene branches within a weakly supported clade of algal proteins, at least some of which encode plastid-targeting leader sequences, but which are not significantly separated from an apparently cytosolic homologue from *K. micrum* (Patron et al. 2006). When a substantially truncated sequence for the *A. tamarense* plastid-targeted cysteine synthase is included in the analysis, the support for this separation

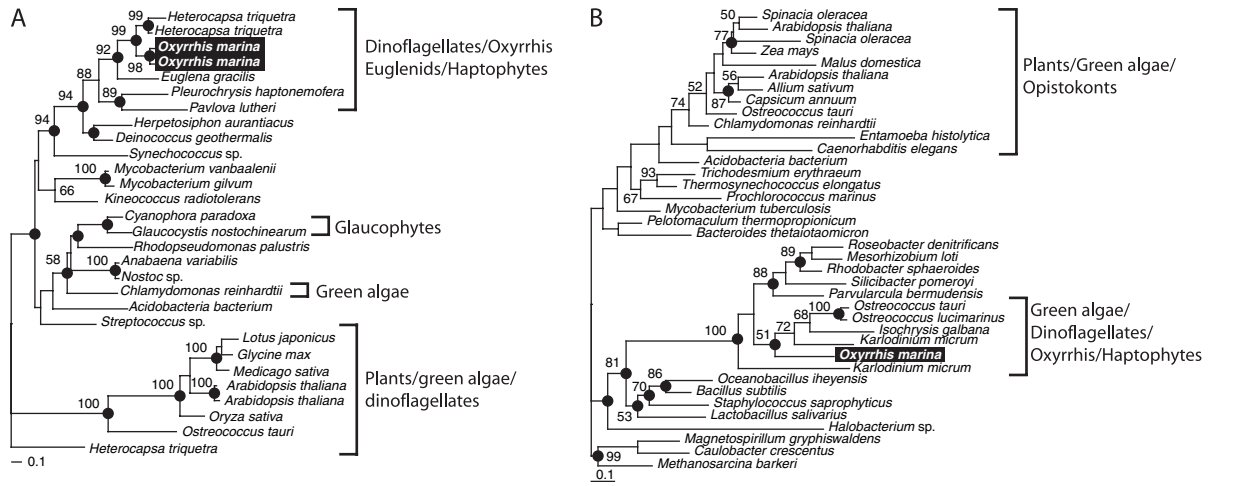

Fig. 3.—Maximum likelihood trees inferred from *Oxyrrhis marina* proteins where no single plastid-targeted clade of proteins is evident, but the *O. marina* protein appears to be derived from a subclass of plastid-targeted proteins. (*A*) Carbonic anhydrase. (*B*) Cysteine synthase. Numbers at the nodes represent percentage of bootstrap support over 300 replicas. Only values >50% are shown. Black circles indicate nodes with Bayesian posterior probability of 0.9 or higher.

increases to 69% (not shown), suggesting that the *O. marina* gene is indeed most closely related to plastid-targeted homologues. Overall, the phylogeny of these 8 proteins is most consistent with these having originated from plastid-targeted homologues, and in 2 cases, the sampling even suggests that the *O. marina* protein is specifically related to plastid-targeted homologues from the dinoflagellates. It is formally possible that these genes are derived by lateral transfer (which has affected plastid-targeted genes in other lineages, including dinoflagellates; Archibald et al. 2003; Bachvaroff et al. 2004; Hackett et al. 2004; Waller et al. 2006), but without direct evidence for this, and given the relationship between *O. marina* and dinoflagellates, the simplest explanation for these genes is that *O. marina* evolved from a plastid-bearing ancestor.

Evidence for Plastid Targeting of Plastid-Derived Proteins

The phylogenetic evidence that *O. marina* retains genes inherited from a photosynthetic ancestor raises the question of whether it also retains the organelle. This can only be definitively determined by directly assessing the cellular location of the putatively plastid-targeted proteins, but the sequences can provide a clue. The majority of proteins targeted to plastids have an N-terminal extension with predictable features. Proteins targeted to secondary plastids have a bipartite leader with a signal peptide followed by a cleavage signal and a chloroplast transit peptide (van Dooren et al. 2001; Archibald and Keeling 2002). The characteristics of transit peptides can vary a great deal and without a sizable number of reliable examples from the organism in question they can be difficult to predict (standard tools for predicting them do not generally work for a phylogenetically broad spectrum of algae). Nevertheless, the characteristics of both apicomplexan and dinoflagellate transit peptides have been examined empirically (van Dooren et al. 2001; Nassoury et al. 2003; Harb et al. 2004; Patron et al. 2005; Tonkin et al. 2006), and the dinoflagellate peptides in particular have a number of distinguishing characteristics (Nassoury et al. 2003; Patron et al. 2005). The characteristics of signal peptides, on the other hand, tend to be highly conserved, and they are relatively easy to predict. More specifically, the signal peptide moiety of plastid-targeted proteins in dinoflagellates and apicomplexans are readily identifiable (van Dooren et al. 2001; Nassoury et al. 2003; Patron et al. 2005).

To examine the possibility that these 8 genes encode targeting information, the 5′ end of all genes was sought using 2 methods. It has recently been shown that *O. marina* and true dinoflagellates both splice a short, identical leader spliced onto the 5′ end of at least a large number of their mRNAs (Zhang et al. 2007). We therefore used both standard cap-dependent 5′ RACE and spliced leader–mediated PCR to capture 5′ ends of as many transcripts as possible. The 5′ end products were sequenced for glutamine synthase, ascorbate peroxidase, HMBS, and both copies of carbonic anhydrase. For all 5 genes, a short 5′ UTR was found to be followed by an in-frame AUG codon and the presence or absence of a potential leader could be inferred. Unfortunately, neither method recovered 5′-end sequence for the remaining 4 genes: based on alignments, RPI and DXR appear to be truncated by only a handful of codons (see below), whereas KARI is truncated by between 30 and 60 codons. The cysteine synthetase cDNA is about 60 codons longer than bacterial genes, but about 40 codons shorter than a *Ostreococcus tauri* plastid-targeted homologue. It includes an ATG codon at a position corresponding to codon number 8 of the plastid-targeted gene from *K. micrum*, so altogether the evidence suggests that this cDNA encodes the complete reading frame, however, without confirming evidence from RACE or spliced leader–mediated RT-PCR it remains uncertain.

Examining the N-terminal regions of all full-length or near-full-length genes showed that carbonic anhydrase, HMBS, RPI, and DXR all encode extensions when compared with nonplastid or bacterial homologues. The N-terminal extremity of each extension was predicted to

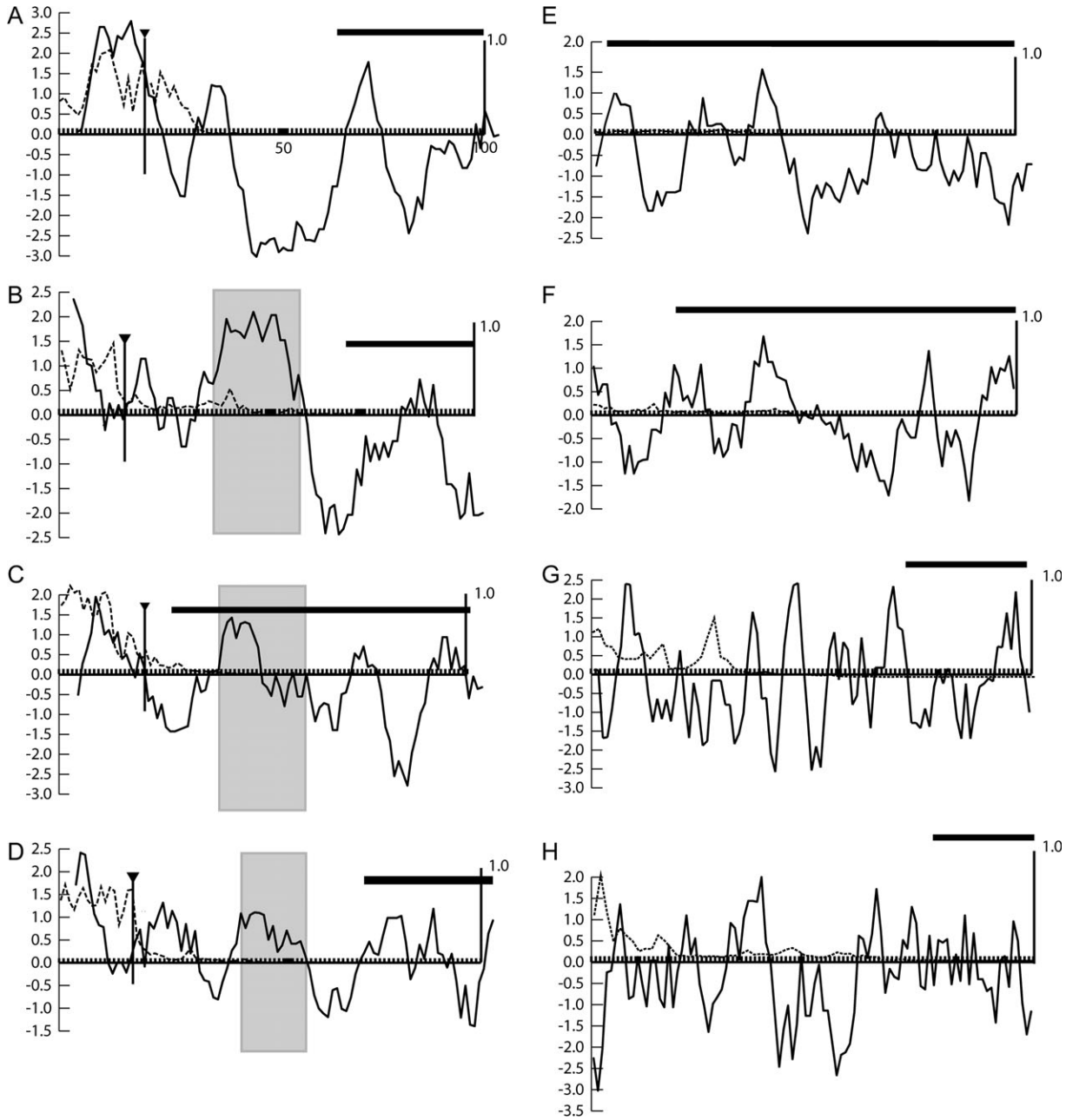

Fig. 4.—Characterization of the amino-termini of the 8 *Oxyrrhis marina* putative plastid-derived proteins. Each graph shows Kyle–Doolittle hydropathy plot (continuous line, left axis) and the SignalP "S" score (dotted line, right axis). Horizontal axis represents position in amino acids starting from position 1 in the available sequence. Most probable cleavage site as calculated by SignalP is indicated (vertical line). When present, transmembrane domains are indicated by a gray box and the region of similarity to amino-termini of known bacterial sequences is marked with a horizontal black rectangle above the graph. (*A*) Carbonic anhydrase; (*B*) HMBS; (*C*) RPI; (*D*) DXR; (*E*) Ascorbate peroxidase; (*F*) Glutamine synthetase; (*G*) KARI; and (*H*) Cysteine synthetase.

encode a signal peptide with a canonical cleavage site (fig. 4*A–D*). Downstream of the signal cleavage site was relatively hydrophobic region ranging from 10 to 20 amino acids, consistent with the transit peptides of dinoflagellates. In addition, HMBS, DXR, and perhaps RPI (see below) were also predicted to encode an extra transmembrane domain downstream of the hydrophobic region (fig. 4*B–D*), a feature that has also been found to characterize a subset of dinoflagellate plastid-targeting leaders (Patron et al. 2005). The RPI and DXR cDNAs are both truncated within extensions relative to the 5′ end of nonplastid homologues, so these extensions may encode truncated leaders. The RPI leader is relatively short, and its putative transmembrane domain actually falls within the region of similarity to prokaryotic homologues, so it is possible this protein lacks a leader. Likewise, the *O. marina* cysteine synthetase sequence is substantially longer at the 5′ end than bacterial counterparts and about the same size as the plastid-targeted

protein from *K. micrum*, suggesting the presence of a leader. However, there is no evidence of transit peptide–like features (fig. 4*H*), and without additional evidence for the location of the 5′ end of the cDNA, we cannot conclude whether this protein has a plastid-targeting leader or not.

The N-termini of 4 of these proteins therefore match our expectations for proteins targeted to a secondary plastid in general (i.e., the presence of signal peptides and sequences with characteristics consistent with transit peptides) and perhaps specifically for targeting to dinoflagellate plastids (2 classes of leader distinguished by the presence or absences of a putative transmembrane domain). The presence of N-terminal leaders with the properties of signal (and possibly transit) peptides suggests that these *O. marina* proteins are targeted to an organelle using a system similar to that of other secondary plastids. This is not conclusive evidence for the presence of a plastid, but there are a number of reasons to argue that this is most probable. First, the presence of a signal peptide in the absence of plastid targeting would lead to a protein being secreted or directed to the endomembrane system, and there is no conceivable functional reason for *O. marina* to secrete these proteins into the marine environment nor is there precedent for their activity in the endomembrane system. Secondly, HMBS and DXR are part of the heme and nonmevalonate isoprenoid biosynthesis pathways, respectively, and are strictly associated with the plastid in photosynthetic protists (and Apicomplexa).

Interestingly, evidence for plastid-targeting leader sequences was not found for all proteins. Both ascorbate peroxidase and glutamine synthase cDNAs appear to be full length as demonstrated by spliced leader PCR (table 1), but neither encodes a leader and the N-termini do not share any of the characteristics expected of a signal peptide (SignalP predicted a significant *C* value for a cleavage site in glutamine synthetase, but did not predict other characteristics of a signal peptide). Unless an as yet unknown targeting mechanism is present, these proteins most likely function in the cytosol. The location of KARI is inconclusive, but it too appears to be cytosolic. Its N-terminus lacks evidence of a signal peptide (fig. 4), and although it encodes a potential initiator AUG, the 6 EST sequences that extend furthest are approximately 70 amino acids shorter than plant sequences on average. Attempts to obtain the complete 5′ end of the transcript by spliced leader PCR failed, and 5′ RACE generated products (12 clones) that were similar to the EST clones up to an AUG codon at position 98 of the ESTs, but upstream of that encoded between 138 to 247 bp of nonhomologous sequence. The 98 bp present in the ESTs but absent from these RACE products share a high level of similarity to plant homologues. In contrast, the 138–247 bp present in the RACE products but absent in the EST shared no detectable similarity to any other sequence and is interrupted by termination codons in all reading frames. The RACE products may be derived from expressed pseudogenes, but the fact that the region of shared sequence similarity begins at an AUG codon also suggests that the RACE products and ESTs could correspond to different versions of the gene, or differential splicing products, with different 5′ ends. Nevertheless, because of these other products, we were never able to characterize the 5′ end of the EST version of these gene, and so it is possible that it represents a plastid-targeted version.

Functional Implications of *O. marina* Plastid-Derived Proteins

The *O. marina* plastid-derived proteins described here do not completely account for any one biochemical pathway. Instead, each protein is a single representative of various potential pathways. Further sampling would be needed to determine if any complete plastid-derived pathways exist today in *O. marina*, but in the meantime, it is interesting to compare which proteins are present with the much better studied nonphotosynthetic apicoplast (Foth and McFadden 2003). Plastid-based fatty acid biosynthesis using the type II pathway is known in apicoplasts, but has yet to be demonstrated in any dinoflagellate, and we did not detect any enzymes from this pathway in *O. marina*. Isoprenoids are synthesized in plastids by condensation of the basic unit isopentenyl phosphate, which is made through the mevalonic acid pathway in the cytosol and the 2-*C*-methyl-D-erythritol 4-phosphate or DOXP pathway in bacteria and plastids (Eisenreich et al. 2004). Plants have maintained both pathways, and many secondarily photosynthetic algae acquired the DOXP pathway with their plastids. Genes for proteins involved in this pathway are now known from *Perkinsus*, *Oxyrrhis*, the colorless dinoflagellate *C. cohnii* (Sanchez-Puerta et al. 2007; Stelter et al. 2007), and *Plasmodium*, where DXR is a strong candidate for a therapeutic strategy (Lichtenthaler 2000). This suggest that this pathway may be ancestral to the apicomplexans and dinoflagellates, but unfortunately the sampling and resolution of the DXR phylogeny do not allow us to discriminate between a single or multiple origins of this pathway.

Tetrapyrroles, precursors of heme and chlorophyll, are synthesized in all eukaryotes, but the pathways and subcellular locations of the steps differ between photosynthetic and nonphotosynthetic species. In animals, heme biosynthesis is partitioned between the cytosol and mitochondria, and the enzymes involved are of either eukaryotic or alphaproteobacterial (mitochondrial) origin (Obornik and Green 2005). In plants and some algae, the whole process occurs inside the chloroplast and most genes are of cyanobacterial origin, although some have different origins (Obornik and Green 2005). In Apicomplexa the story is more complex because the pathway seems to be partitioned between the plastid and the mitochondrion (Sato et al. 2004; Varadharajan et al. 2004; van Dooren et al. 2006). Two genes from this pathway, delta-aminolevulinic acid dehydratase (HemB) and protoporphyrinogen oxidase (HemG), were identified in the nonphotosynthetic dinoflagellate, *C. cohnii* (Sanchez-Puerta et al. 2007). *Oxyrrhis marina* has maintained the plastidial HMBS (HemC), suggesting that it uses at least part of the plastid-derived pathway.

Another function of plastids in plants is the synthesis of several amino acids (Singh and Matthews 1994; Kleffmann et al. 2004). Interestingly, some amino acid biosyntheses appear to have been retained in dinoflagellate plastids (Sanchez-Puerta et al. 2007) but has been lost in Apicomplexa: although there was some debate about the shikimate pathway, these enzymes are not plastid derived

and function in the cytosol (Roberts et al. 1998; Keeling et al. 1999; Richards et al. 2006). The situation in *O. marina* appears to be different again: genes for 3 plastid-derived proteins with roles in biosynthesis of amino acids were identified (glutamine synthetase, cysteine synthetase, and KARI, which participate in the synthesis of glutamine, cysteine, and branched amino acids, respectively), but we found no evidence that at least glutamine synthetase is targeted to the plastid (cysteine synthetase and KARI are incomplete at the 5′ end). In plants, glutamine synthesis occurs in the chloroplast in a light-dependent manner and glutamine synthetase is driven by photophosphorylation-generated adenosine triphosphate (Givan 1975). It is possible, therefore, that glutamine synthesis moved to the cytosol after photosynthesis was lost as it would no longer be light dependent.

Ascorbate peroxidase plays an important role in hydrogen peroxide detoxification in plants, where multiple isoforms are directed to the chloroplast, cytosol, and peroxisome. Although uncertain, it seems likely that the cytosolic and plastid paralogues diverged before the divergence of the red and green plastid lineages. This would mean that a plastid ascorbate peroxidase from the endosymbiont was redirected to the cytosol sometime in the evolution of *O. marina*, but whether or not the original cytosolic paralogue is also present is unclear without an exhaustive genomic survey. Interestingly, a related ascorbate peroxidase gene has been described from both *C. cohnii* and *P. marinus* and extracts of the latter were found to contain ascorbate peroxidase activity, indicating that the enzyme is present. Unfortunately in neither of these can its subcellular localization be inferred (Schott et al. 2003; Sanchez-Puerta et al. 2007).

Lastly, in eukaryotes, carbonic anhydrase is exclusively found in plastid-bearing lineages where it is typically plastid targeted (but see below), and its main function is raising the concentration of CO_2 to increase the rate of carboxylation by rubisco. It is unclear what role this protein might play in absence of rubisco, but some form of acid-base regulation by interconversion between carbon dioxide and carbonic acid is 1 possibility. Intriguingly, a rubisco gene is maintained in certain nonphotosynthetic algae, including the dinoflagellate *C. cohnii* (Sanchez-Puerta et al. 2007), raising the question of whether rubisco might also exist in *O. marina*. Rubisco in *O. marina* would be of particular interest given its phylogenetic place at the base of dinoflagellates as it could provide valuable insights into the origin of the unusual type II rubisco typical of peridinin-containing plastids (Morse et al. 1995). Interestingly, however, extracellular carbonic anhydrase activity has also been reported in several dinoflagellates (Nimer et al. 1999; Rost et al. 2006), so it is also possible that this protein is not directed to the plastid. It could be, for example, that 1 phylogenetic subgroup of algal carbonic anhydrases genes corresponds to extracellular proteins and the other plastid-targeted proteins (note that *H. triquetra* has both: fig. 3*A*). In this case, the signal peptide may direct this protein to the plasma membrane rather than to a plastid. Even in this case, the genes are mostly likely all derived from the plastid because beta-carbonic anhydrase is not known from any aplastidal eukaryote.

Evolutionary Implications of a Photosynthetic Ancestry of *O. marina*

Recent years have demonstrated that the distribution of plastids is much wider than the distribution of photosynthesis, and the distribution of plastid-bearing ancestry is wider still. Cryptic plastids have now been reported in a number of lineages with overtly plastid-bearing relatives. The apicoplast of malaria and *Toxoplasma* are now so well characterized that it is easy to forget how recently they were discovered to be plastids. Plastid genes have also now been reported from oomycete genomes (Tyler et al. 2006) and the apicomplexan *Cryptosporidium* (Huang et al. 2004), both of which appear to lack the actual organelle (Abrahamsen et al. 2004; Tyler et al. 2006). Plastid genes and evidence for targeting have been found in the nonphotosynthetic dinoflagellate *Crypthecodinium* (Sanchez-Puerta et al. 2007), and from *Perkinsus*, both plastid genes and a potential plastid-derived organelle of unknown function are known (Stelter et al. 2007; Teles-Grilo et al. 2007). The data presented here now suggest that *O. marina* also evolved from photosynthetic ancestors, and analysis of putative targeting peptides suggests that it might also retain an organelle, although direct evidence for subcellular location of putatively targeted proteins is needed to conclude this. This widening distribution of photosynthetic ancestry in eukaryotes raises questions about the relationship of these cryptic organelles and their photosynthetic counterparts. Ever since the discovery of the apicoplast, its possible relationship to dinoflagellate plastids has been debated, and this question is intimately tied to the larger hypothesis that all eukaryotes with secondary plastids derived from red algae acquired them in a single event, the so called chromalveolate hypothesis (Cavalier-Smith 1999). A good deal of work to prove or overturn this hypothesis has now been reported (Fast et al. 2002; Harper and Keeling 2003; Patron et al. 2004; Bodyl 2005; Harper et al. 2005; Burki et al. 2007; Hackett et al. 2007), and it remains contested. If the chromalveolate hypothesis turns out to be correct, then it would suggest that the cryptic plastids and plastid-derived genes discovered in the many nonphotosynthetic chromalveolate lineages (e.g., all of those described above) are mostly likely descendent from the same plastid ancestor. Even without the chromalveolate hypothesis, a conclusive demonstration that the apicoplast and dinoflagellate plastids share a common ancestor would at least suggest that the cryptic plastids in their close relatives are also derived from that ancestor. Support for such a conclusion comes from a discovery reported while this paper was under review: *Chromera* is a newly described, deep-branching member of the apicomplexan lineage that is also a free-living, photosynthetic alga (Moore et al. 2008). The very existence of *Chromera* and the phylogenetic relationships of its plastid-encoded genes makes a very strong case for a common origin of apicomplexan and dinoflagellate plastids and by extension suggests that any relict plastid in *Oxyrrhis* or *Perkinsus* is most likely derived from the same source.

Another characteristic of the plastid-derived genes from *O. marina* that is emerging as a topic of general importance is that the evolutionary history of a gene and the cellular location of its product need not correspond. The

location of some pathways seems to be highly conserved, but there is growing evidence that others are shifting either entirely or partly between compartments. Heme and amino acid metabolism in the alveolates are interesting case studies of this process because the retention or cellular distribution of these pathways differs between Apicomplexa and dinoflagellates (Roberts et al. 1998; Keeling et al. 1999; Sato et al. 2004; Varadharajan et al. 2004; Obornik and Green 2005; Richards et al. 2006; van Dooren et al. 2006; Waller et al. 2006; Sanchez-Puerta et al. 2007). The intermediate position of *O. marina* could therefore be of use in reconstructing the evolution of apicoplast and the peridinin plastid metabolism. Indeed, there is already evidence for differences between *O. marina*, Apicomplexa, and dinoflagellates: like dinoflagellates, but in contrast to Apicomplexa, *O. marina* has retained some enzymes for amino acid metabolism, but in at least 1 case, the enzyme is apparently no longer located in the plastid. Taxa such as *Perkinsus* and *Oxyrrhis* can offer great insights into how these pathways evolved and shifted within the cell over time, but for now they only offer snapshots. Only when complete genomes are available and localization of representative enzymes completed will we be able to compare the plastids of these taxa with those of Apicomplexa comprehensively.

Acknowledgments

This work was supported by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. *Oxyrrhis marina* EST sequencing was supported by the Protist EST Program of Genome Canada/Genome Atlantic. We thank Allen Larocque for assistance with cloning and sequencing. P.J.K. is a fellow of the Canadian Institute for Advanced Research and a Senior Scholar of the Michael Smith Foundation for Health Research.

Literature Cited

Abrahamsen MS, Templeton TJ, Enomoto S, et al. (20 co-authors). 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science. 304:441–445.

Archibald JM, Keeling PJ. 2002. Recycled plastids: a green movement in eukaryotic evolution. Trends Genet. 18:577–584.

Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc Natl Acad Sci USA. 100:7678–7683.

Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF. 2004. Dinoflagellate expressed sequence tags data indicate massive transfer of chloroplast genes to the nuclear genome. Protist. 155:65–78.

Bodyl A. 2005. Do plastid-related characters support the chromalveolate hypothesis? J Phycol. 41:712–719.

Burki F, Shalchian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI, Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. PLoS ONE. 2:e790.

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

Eisenreich W, Bacher A, Arigoni D, Rohdich F. 2004. Biosynthesis of isoprenoids via the non-mevalonate pathway. Cell Mol Life Sci. 61:1401–1426.

Fast NM, Xue L, Bingham S, Keeling PJ. 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. J Eukaryot Microbiol. 49:30–37.

Foth BJ, McFadden GI. 2003. The apicoplast: a plastid in Plasmodium falciparum and other apicomplexan parasites. Int Rev Cytol. 224:57–110.

Givan CV. 1975. Light-dependent synthesis of glutamine in pea chloroplast preparations. Planta. 122:281–291.

Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33:W557–W559.

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. Mol Biol Evol. 24:1702–1713.

Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Nosenko T, Bhattacharya D. 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr Biol. 14:213–218.

Harb OS, Chatterjee B, Fraunholz MJ, Crawford MJ, Nishi M, Roos DS. 2004. Multiple functionally redundant signals mediate targeting to the apicoplast in the apicomplexan parasite Toxoplasma gondii. Eukaryot Cell. 3:663–674.

Harper JT, Keeling PJ. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. Mol Biol Evol. 20:1730–1735.

Harper JT, Waanders E, Keeling PJ. 2005. On the monophyly of the chromalveolates using a six-protein phylogeny of eukaryotes. Int J Sys Evol Microbiol. 55:487–496.

Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC. 2004. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. Genome Biol. 5:R88.

Keeling PJ. 2004. The diversity and evolutionary history of plastids and their hosts. Am J Bot. 91:1481–1493.

Keeling PJ, Palmer JD, Donald RG, Roos DS, Waller RF, McFadden GI. 1999. Shikimate pathway in apicomplexan parasites. Nature. 397:219–220.

Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, Baginsky S. 2004. The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. Curr Biol. 14:354–362.

Leander BS, Keeling PJ. 2004. Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from HSP90 and actin phylogenies. J Phycol. 40:341–350.

Lichtenthaler HK. 2000. Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors. Biochem Soc Trans. 28:785–789.

Moore RB, Obornik M, Janouskovec J, et al. (14 co-authors). 2008. A photosynthetic alveolate closely related to apicomplexan parasites. Nature. 451:959–963.

Morse D, Salois P, Markovic P, Hastings JW. 1995. A nuclear-encoded form II RuBisCO in dinoflagellates. Science. 268: 1622–1624.

Nassoury N, Cappadocia M, Morse D. 2003. Plastid ultrastructure defines the protein import pathway in dinoflagellates. J Cell Sci. 116:2867–2874.

Nimer NA, Brownlee C, Merrett MJ. 1999. Extracellular carbonic anhydrase facilitates carbon dioxide availability for photosynthesis in the marine dinoflagellate Prorocentrum micans. Plant Physiol. 120:105–112.

Obornik M, Green BR. 2005. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. Mol Biol Evol. 22:2343–2353.

O'Brien EA, Koski LB, Zhang Y, Yang L, Wang E, Gray MW, Burger G, Lang BF. 2007. TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). Nucleic Acids Res. 35:D445–D451.

Palmer JD. 2003. The symbiotic birth and spread of plastids: how many times and whodunit? J Phycol. 39:1–9.

Patron NJ, Rogers MB, Keeling PJ. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Eukaryot Cell. 3:1169–1175.

Patron NJ, Waller RF, Archibald JM, Keeling PJ. 2005. Complex protein targeting to dinoflagellate plastids. J Mol Biol. 348:1015–1024.

Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. J Mol Biol. 357:1373–1382.

Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW. 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryot Cell. 5: 1517–1531.

Roberts F, Roberts CW, Johnson JJ, et al. 1998. Evidence for the shikimate pathway in apicomplexan parasites. Nature. 393:801–805.

Robertson DL, Tartar A. 2006. Evolution of glutamine synthetase in heterokonts: evidence for endosymbiotic gene transfer and the early evolution of photosynthesis. Mol Biol Evol. 23:1048–1055.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Roos DS, Crawford MJ, Donald RG, Fraunholz M, Harb OS, He CY, Kissinger JC, Shaw MK, Striepen B. 2002. Mining the Plasmodium genome database to define organellar function: what does the apicoplast do? Philos Trans R Soc Lond B Biol Sci. 357:35–46.

Rost B, Richter KU, Riebesell U, Hansen PJ. 2006. Inorganic carbon acquisition in red tide dinoflagellates. Plant Cell Environ. 29:810–822.

Saldarriaga JF, McEwan ML, Fast NM, Taylor FJR, Keeling PJ. 2003. Multiple protein phylogenies show that *Oxyhrris marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. Int J Sys Evol Microbiol. 53:355–365.

Saldarriaga JF, Taylor FJR, Keeling PJ, Cavalier-Smith T. 2001. Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. J Mol Evol. 53:204–213.

Sanchez-Puerta MV, Lippmeier JC, Apt KE, Delwiche CF. 2007. Plastid genes in a non-photosynthetic dinoflagellate. Protist. 158:105–117.

Sato S, Clough B, Coates L, Wilson RJ. 2004. Enzymes for heme biosynthesis are found in both the mitochondrion and plastid of the malaria parasite *Plasmodium falciparum*. Protist. 155:117–125.

Schott EJ, Pecher WT, Okafor F, Vasta GR. 2003. The protistan parasite Perkinsus marinus is resistant to selected reactive oxygen species. Exp Parasitol. 105:232–240.

Singh BK, Matthews BF. 1994. Molecular regulation of amino acid biosynthesis in plants. Amino Acids. 7:165–174.

Slamovits CH, Saldarriaga JF, Larocque A, Keeling PJ. 2007. The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate Oxyrrhis marina shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. J Mol Biol. 372:356–368.

Smith KS, Jakubzick C, Whittam TS, Ferry JG. 1999. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. Proc Natl Acad Sci USA. 96:15184–15189.

Stelter K, El-Sayed NM, Seeber F. 2007. The expression of a plant-type ferredoxin redox system provides molecular evidence for a plastid in the early dinoflagellate Perkinsus marinus. Protist. 158:119–130.

Swofford DL. 2002. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland (MA).

Teles-Grilo ML, Tato-Costa J, Duarte SM, Maia A, Casal G, Azevedo C. 2007. Is there a plastid in Perkinsus atlanticus (Phylum Perkinsozoa)? Eur J Protistol. 43:163–167.

Tonkin CJ, Roos DS, McFadden GI. 2006. N-terminal positively charged amino acids, but not their exact position, are important for apicoplast transit peptide fidelity in Toxoplasma gondii. Mol Biochem Parasitol. 150:192–200.

Tyler BM, Tripathy S, Zhang X, et al. (52 co-authors). 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. 313:1261–1266.

van Dooren GG, Schwartzbach SD, Osafune T, McFadden GI. 2001. Translocation of proteins across the multiple membranes of complex plastids. Biochim Biophys Acta. 1541:34–53.

van Dooren GG, Stimmler LM, McFadden GI. 2006. Metabolic maps and functions of the Plasmodium mitochondrion. FEMS Microbiol Rev. 30:596–630.

Varadharajan S, Sagar BK, Rangarajan PN, Padmanaban G. 2004. Localization of ferrochelatase in Plasmodium falciparum. Biochem J. 384:429–436.

Waller RF, Keeling PJ, Donald RG, Striepen B, Handman E, Lang-Unnasch N, Cowman AF, Besra GS, Roos DS, McFadden GI. 1998. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc Natl Acad Sci USA. 95:12352–12357.

Waller RF, Patron NJ, Keeling PJ. 2006. Phylogenetic history of plastid-targeted proteins in the peridinin-containing dinoflagellate *Heterocapsa triquetra*. Int J Syst Evol Microbiol. 56:1439–1447.

Williams BA, Hirt RP. 2004. RACE and RAGE cloning in parasitic microbial eukaryotes. Methods Mol Biol. 270: 151–172.

Williamson DH, Gardner MJ, Preiser P, Moore DJ, Rangachari K, Wilson RJ. 1994. The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. Mol Gen Genet. 243:249–252.

Wilson RJMI, Denny PW, Preiser DJ, et al. (11 co-authors). 1996. Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J Mol Biol. 261:155–172.

Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, Gaasterland T, Lin S. 2007. Spliced leader RNA trans-splicing in dinoflagellates. Proc Natl Acad Sci USA. 104:4618–4623.

Andrew Roger, Associate Editor

Accepted March 13, 2008