Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate

Sebastian G. Gornik,a,b, Febrimarsa,a, Andrew M. Cassin,c, James I. MacRae,d,e, Abhinay Ramaprasad,f Zineb Rchiad,f, Malcolm J. McConville,g, Antony Baicic,a, Geoffrey I. McFadden,a, Arnab Painf,g, and Ross F. Waller,a,h,1

aSchool of Botany, University of Melbourne, Parkville, Victoria 3010, Australia; bSchool of Natural Sciences, National University of Ireland, Galway, Ireland; cAustralian Research Council Centre of Excellence in Plant Cell Walls, School of Botany, University of Melbourne, Victoria 3010, Australia; dThe Francis Crick Institute, London, NW7 1AA, United Kingdom; eBio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia; fBiological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia; gGlobal Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE), Hokkaido University, N20 W10 Kita-ku, Sapporo, 001-0020 Japan; and hDepartment of Biochemistry, University of Cambridge, Cambridge CB2 1QW, United Kingdom

Edited by John M. Archibald, Dalhousie University, Halifax, NS, Canada, and accepted by the Editorial Board March 20, 2015 (received for review December 8, 2014)

Organelle gain through endosymbiosis has been integral to the origin and diversification of eukaryotes, and, once gained, plastids and mitochondria seem seldom lost. Indeed, discovery of nonphotosynthetic plastids in many eukaryotes—notably, the apicoplast in apicomplexan parasites such as the malaria pathogen Plasmodium—highlights the essential metabolic functions performed by plastids beyond photosynthesis. Once a cell becomes reliant on these ancillary functions, organelle dependence is apparently difficult to overcome. Previous examples of endosymbiotic organelle loss (either mitochondria or plastids), which have been invoked to explain the origin of eukaryotic diversity, have subsequently been recognized as organellar reduction to cryptic forms, such as mitosomes and apicoplasts. Integration of these ancient symbionts with their hosts has been too well developed to reverse. Here, we provide evidence that the dinoflagellate Hematodinium sp., a marine parasite of crustaceans, represents a rare case of endosymbiotic organelle loss by the elimination of the plastid. Extensive RNA and genomic sequencing data provide no evidence for a plastid organelle, but, rather, reveal a metabolic decoupling from known plastid functions that typically impede organelle loss. This independence has been achieved through retention of ancestral anabolic pathways, enzyme relocation from the plastid to the cytosol, and metabolic scavenging from the parasite’s host. Hematodinium sp. thus represents a further dimension of endosymbiosis—life after the organelle.

Dinoflagellates belong to a plastid-bearing algal lineage that offers wide scope for studying plastid evolution. Photosynthetic dinoflagellates help power ocean carbon fixation and food webs, and symbiotic dinoflagellates drive tropical reef formation through their coral hosts. However, approximately half of dinoflagellates are heterotrophic and adapted to micropredation through their coral hosts. Nevertheless, all heterotrophic dinoflagellates that have been closely scrutinized apparently retain reduced forms of a plastid organelle, likely because they are metabolically dependent on plastid-based de novo synthesis of compounds such as isoprenoids (Perkinsus, Oxyrrhis, Cryptecodiun), tetrapyrroles (Oxyrrhis), and Fe–S clusters (Perkinsus) (9–12). Genes for these pathways are nucleus-encoded, and targeting signals direct the enzymes into plastid organelles. Retention of the same anabolic pathways makes the apicoplast indispensable throughout almost all of Apicomplexa (13–15).

Metabolic functions that drive the retention of plastids, even when photosynthesis is lost, are typically those found within the cytosol of eukaryotes before the gain of a plastid. Endosymbiosis, therefore, leads to pathway duplication. The presence of the plastid often leads to loss(es) of these original host-cell–based pathways, making the plastid indispensable. Loss of the plastid can only occur if the cell avoids or overcomes these dependencies, by: (i) maintaining cytosolic pathways; (ii) scavenging an exogenous supply of metabolites; (iii) eliminating the requirement for the metabolite; (iv) relocating a plastid pathway; or (v) a combination of these. Only one apicomplexan group, Cryptosporidium, has solved this complex puzzle by maintaining cytosolic fatty acid synthesis and scavenging host isopre precursor and the tetrapyrrole

Significance

Endosymbiotic organelles are a defining feature of eukaryotes—the last common ancestor and all extant eukaryotes possess at least a mitochondrial derivative. Although mitochondria and plastids are identified with aerobic ATP synthesis and photosynthesis, respectively, their retention by their host cells requires the merging and integration of many, often redundant, metabolic pathways. As a result, complex metabolic interdependencies arise between these formerly independent cells. Complete loss of endosymbiotic organelles, even where aerobic respiration or photosynthesis is lost, is exceedingly difficult, as demonstrated by persistence of organelles throughout secondary anaerobes and parasites. Here, we identify a rare but clear case of plastid loss in a parasitic alga and detail the metabolic disentanglement that was required to achieve this exceptional evolutionary event.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.M.A. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KP739886–KP739907).

1To whom correspondence should be addressed. Email: rfw26@cam.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1423400112.

www.pnas.org/cgi/doi/10.1073/pnas.1423400112

PNAS May 5, 2015 vol. 112 no. 18 5767–5772
We have investigated the basal, nonphotosynthetic dinoflagellate *Hematodinium* sp. for evidence of a plastid. *Hematodinium* species are members of the Syndiniales, an exclusively parasitic basal lineage of dinoflagellates (2). *Hematodinium* species parasitize decapod crustaceans, have a broad host range, and cause significant impact on commercial fisheries and wild stocks (20). We generated extensive transcriptomic RNA-sequencing (RNA-seq) and genomic data and analyzed these for evidence of a plastid and plastid-associated metabolic pathways. Our data indicate that through some previously unidentified metabolic pathway contingents, *Hematodinium*—like *Cryptosporidium*—has also solved the plastid dependency puzzle and eliminated this organelle.

**Results and Discussion**

No plastid or candidate plastid organelle has been reported in any ultrastructural studies of *Hematodinium* species (21), but molecular evidence is increasingly a more sensitive and informative approach to testing for cryptic endosymbiotic organelles. We therefore generated extensive molecular data from *Hematodinium* sp. and interrogated these data for evidence of either an extant or past plastid.

**RNA-Seq Provides Extensive Coverage of *Hematodinium* Genes.** An RNA-seq analysis of different growth stages (trophonts and dinospores) of *Hematodinium* sp. was undertaken and the data interrogated for evidence of either an extant or past plastid. We assembled 222,704 unique transcripts of >200 bp (Table S1). To assess what portion of the complete gene content of *Hematodinium* sp. these data represent, we compared them to known databases and recovered 15,510 Swiss-Prot (UniProtKB) matches (<2.0) and 9,207 Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations. Using the Core Eukaryotic Genes Mapping Approach (CEGMA) (22), we recovered 85.9% of the expected core eukaryotic genes, more than the *Symbiodinium* draft genome or earlier dinoflagellate transcriptomes (Table S2). The large proportion of unique transcripts (222,704) is likely an overestimation of the true coding capacity due to: (i) observed coding sequence redundancy where UTRs showed either divergent or unrelated sequence (the reasons for which are under investigation); and (ii) high representation of very short ORFs. Clustering of identical ORFs results in 118,749 unique predicted proteins, and exclusion of ORFs <180 amino acid length still retains 29,510 unique ORFs of mean and median length of 443 and 330 amino acids, respectively, which is comparable to known gene length means and medians (472 and 361, respectively) from a wide range of eukaryotes (23, 24). Together, these analyses suggest that we have captured a substantial fraction of *Hematodinium* sp. genes.

**Genes Encoding Plastid Proteins Are Absent in *Hematodinium***

A hallmark of endosymbiosis is that the vast majority of organelle proteins are coded for in the cell nucleus and posttranslationally targeted back to the organelle. Proteins sorted to secondary plastids, including those of dinoflagellates and apicomplexans, are first directed into the endomembrane system via a canonical N-terminal signal peptide (SP) (25, 26). Plastid proteins are then directed into the plastid lumen via a chloroplast-type transit peptide (cTP) downstream of the SP, which shares conserved, recognizable features for almost all plastids (27). Most of the biochemistry of the nonphotosynthetic plastids of apicomplexans is defined by nucleus-encoded plastid proteins bearing this distinctive bipartite (SP + cTP) targeting leader (28). To search for evidence of nucleus-encoded plastid genes in *Hematodinium* sp., we used a curated set of 157 predicted plastid proteins from *Plasmodium vivax* (proteins with bipartite leaders, plastid-type functions, and apicoplast-located orthologs) (29). By using RNA-seq data, permisive TBLASTN searches (E value < e^-3) were performed and scrutinized by reciprocal BLASTP searches against the National Center for Biotechnology Information (NCBI) RefSeq database (E value < e^-3), and 81 proteins were recovered (Fig. 1A and Dataset S1). Many of the *Plasmodium* protein set represent generic cell functions (e.g., ribosomal function, tRNA ligases, transporters), and so these *Hematodinium* sp. matches could also represent mitochondrion or cytosolic homologs. To validate the matches as potential plastid proteins, we then analyzed them for the presence of a bipartite leader. Transcripts were first checked for the presence of the dinoflagellate 5′-splice leader (SL) and/or upstream in-frame termination codon as evidence that they are 5′-complete and that the protein N terminus is represented. Of the 81 matches, 77 contained complete ORFs. When these full-length ORFs were tested for a SP (SignalP3.0 and PredSL) followed by a cTP (TargetP1.1 and PredSL), none satisfied these criteria (Fig. 1A and Dataset S1). Of the four incomplete *Hematodinium* sp. sequences that could not be tested for plastid-targeting signals, two were transporters and two were unknown proteins with conserved domains, so all have credible nonplastid roles. Thus, from a test set of 157 nucleus-encoded plastid proteins, no credible transcripts were found for plastid-targeted proteins in *Hematodinium* sp. In contrast, the presence of recognizable bipartite targeting peptides provides evidence of reduced plastids in *Perkinsus*, *Oxyrrhis*, and *Cryptophyceae* (9–12).

To assess the expected likelihood of recovering nucleus-encoded organelle-targeted proteins from our dataset, an equivalent search was performed for mitochondrial proteins, with the exception of scoring for mitochondrion targeting peptides (mTP). By using a set of 109 mitochondrial proteins from *P. falciparum* (30), 94 putative mitochondrial proteins were identified in the *Hematodinium* sp. dataset. Of these proteins, 49 (TargetP1.1) and 51 (PredSL) contained credible mitochondrial targeting information—an overall recovery rate of 45 and 47%, respectively (Fig. 1B and Dataset S2). Targeting peptide predictors by their nature are not perfect and will produce some false predictions. A previous assessment of cTP prediction (ChloroP) in dinoflagellates indicated 45% success (27), similar to that shown for mTP prediction here (56–58%). Given the detection of mitochondria-targeted proteins in this system, and that of plastid proteins in other low branching dinoflagellates, we believe the most plausible explanation is that plastid-targeted proteins are no longer present in
Hematodinium sp. It is conceivable, however, that plastid-targeted proteins do exist, but are of very low abundance, too divergent, or exceptional in lacking identifiable bipartite leaders.

Plastid genes can also be encoded directly in the organelle, although dinoflagellates are unusual in having reduced the number of plastid-encoded genes to only 14 (3). We searched for all known protein genes found in alveolates plastids [82 genes in total, found in dinoflagellates, apicomplexans, or chromerids (3)] and found no matches in our RNA-seq data (Fig. 1C and Dataset S3). In contrast, all five known alveolate mtDNA genes were readily detected (31), and we note that 11 of 14 dinoflagellate plastid-coding genes are represented in equivalent Lingulodinium poly-A-derived RNA-Seq data (32). Furthermore, we have now generated a draft genomic sequence for Hematodinium sp. The estimated genome size is ~4,800 Mbp (33), consistent with the very large genomes of other dinoflagellates (ref. 34 and references therein). We currently have 4,769 Mbp of data assembled (30× coverage, N50 = 17,235 bp, 118,385 scaffolds >10 kbp, 653 scaffolds >100 kbp). Preliminary analysis of these data revealed that genes are densely populated with very large introns and are, therefore, spread over tens of kilobase pairs. Thus, many nucleus-encoded gene sequences are currently incomplete, even on relatively large scaffolds, and sequencing and assembly of this genome is ongoing. Organelle genomes, however, are typically gene-dense, contain few or no introns, and are well represented in genomic sequence data. We searched these genomic data for the 82 alveolate plastid-encoded genes and, similar to the RNA-seq data, found none. Thus, there is a conspicuous absence of any genes for plastid proteins in Hematodinium sp., consistent with loss of this organelle.

**Hematodinium Encodes Biochemical Alternatives to Plastid Functions.**

In apicomplexans, ablation of plastid biochemical pathways by gene knockouts and chemical inhibition identified several ana- bolic pathways that are essential for the completion of the life cycle—including fatty acid biosynthesis, tetrapyrrole synthesis, and isopentenyl pyrophosphate (IPP) (isoprene precursor) synthesis (13, 15, 35–37). The indispensability of these pathways is likely the major explanation for plastid retention. In the absence of either molecular or morphological evidence of a plastid in Hematodinium, we have analyzed our data for evidence of non-plastid alternative pathways that could alleviate the need for these typically essential plastid functions (Fig. 2).

**Fatty acids are generated by using cytosolic type I, and not plastid type II, fatty acid synthase.** Fatty acids are assembled de novo in cells by iterative additions of two carbons, donated from malonyl-CoA, by fatty acid synthase (FAS) complexes. Prokaryotic FAS sys- tems typically consists of seven separate monofunctional proteins, and this multiprotein complex, called type II FAS, occurs in plastids. Evolution of FAS in the eukaryotic cytosol has systems typically consists of seven separate monofunctional enzymes currently unidentified in transcriptomes. Enzyme color represents typical location and origin as follows: green, plastid; yellow, cytosol; red, mitochondrion. Hatched (green/white) DAP pathway indicates uncertain origin of this typically plastid-located pathway in Hematodinium. MVA, mevalonate IPP pathway; DOXP, 1-deoxy-D-xylulose-5-phosphate IPP path- way; C15/20, isoprene chains 15 and 20 carbons long derived from IPP (an external source of IPP/isoprenoids for Hematodinium sp. is predicted); SUF, plastid-type iron-sulfur cluster pathway; DAP, diaminopimelate lysine path- way; C4/C5 pathways for tetrapyrrole (TP) synthesis differ only by the reactions to δ-aminolevulinic acid (ALA) and their location.

Fig. 2. Reconstructed metabolic pathways in the common ancestor of apicomplexans and dinoflagellates at the time of plastid gain (A) and in Hematodinium sp. (B), from transcriptomic data. Cytosolic MVA pathway is not present in any apicomplexan or dinoflagellate, but is present in ciliates and is inferred to be present at the time of plastid gain. Dashed lines indicate apicomplexan hybrid tetrapyrrole pathways; dashed circles indicate enzymes currently unidentified in transcriptomes. Enzyme color represents typical location and origin as follows: green, plastid; yellow, cytosol; red, mitochondrion. Hatched (green/white) DAP pathway indicates uncertain origin of this typically plastid-located pathway in Hematodinium. MVA, mevalonate IPP pathway; DOXP, 1-deoxy-D-xylulose-5-phosphate IPP path- way; C15/20, isoprene chains 15 and 20 carbons long derived from IPP (an external source of IPP/isoprenoids for Hematodinium sp. is predicted); SUF, plastid-type iron-sulfur cluster pathway; DAP, diaminopimelate lysine path- way; C4/C5 pathways for tetrapyrrole (TP) synthesis differ only by the reactions to δ-aminolevulinic acid (ALA) and their location.

large, multienzyme complexes, consistent with the presence of a fused type II FAS and the absence of a type II FAS complex (Fig. 2B). Our RNA-seq data recovered three large FAS-enzyme-containing transcripts (GenBank accession nos. KP739886–KP739888), whose full-length ORFs all lack a bipartite leader and which are predicted to be cytosolic, as well as two further incomplete transcripts with multiple FAS-like enzyme domains (GenBank accession nos. KP739889–KP739890; Fig. S1A). Polyketide synthases (PKS) are a related family of proteins that are considered de-generate FASs in which one or more modules have been lost, and PKS are known to occur in dinoflagellates (40). Although two of the transcripts contain all modules expected of complete type I FAS proteins, it is difficult to reliably discern FAS from PKS without biochemical validation. Therefore, we tested for de novo fatty acid synthesis by measuring incorporation of [13C]-U-glucose into Hematodinium sp. fatty acids. [13C]-Glucose is catabolized to pyruvate and acetyl-CoA, providing the precursors for the FAS complexes. Label was uniformly incorporated into saturated C14:0 and C16:0 fatty acids in [13C]-glucose-fed Hematodinium sp., generating a series of fatty-acid isopomers that increased by +2 atomic mass units, consistent with de novo synthesis of fatty acids from a malonyl-CoA donor (Fig. S1 B and C). These analyses suggest that Hematodinium sp. express a large type I FAS complex that would allow loss of a plastid-located type II complex. Tetrapyrrole biosynthesis occurs in the cytosol. Tetrapyrrole bio- synthesis is required for the synthesis of heme and chlorophylls, and it occurs via two related pathways (19) (Fig. 2B). The C4 pathway, which occurs in heterotrophic eucaryotes, commences with the synthesis of δ-aminolevulinic acid (ALA) from glycine and succinyl-CoA by ALA synthase (ALAS) in the mitochondrion. ALA is exported to the cytosol, where the enzymes porphobilinogen synthase (HemB), porphobilinogen deaminase (HemC), uroporphyrinogen III synthase (HemD), and uropor- phyrinogen III decarboxylase (HemE) catalyze the production of coproporphyrinogen, which is then transported back into the mitochondrion to be converted into heme by coproporphyrinogen oxidase (HemF), protoporphyrinogen oxidase (HemG/Y), and ferrochelatase (HemH). In contrast, most photosynthetic
eukaryotes produce tetrapyrroles through the plastid-located C5 pathway. Here, ALA is synthesized from glutamyl–rRNA by glutamyl–rRNA reductase (GTR) followed by glutamate l-semialdehyde aminotransferase (GSA-AT). ALA is then converted into tetrapyrroles via the same enzymes found in the C4 pathway, but these enzymes are mostly derived from the cyanobacterium that gave rise to plastids. In most photosynthetic eukaryotes, the plastid C5 pathway has been retained in place of the ancestral C4 pathway, and the bulk of tetrapyrroles are incorporated into chlorophylls.

Tetrapyrrole synthesis in apicomplexans and their photosynthetic relative Chromera velia is unusual in that elements of both the C4 and C5 pathways have been used in a pathway of hybrid origin (19, 42) (Fig. 2A, dashed lines). Classic C4 synthesis of ALA occurs in the mitochondrion (with plastid enzymes GTR and GSA-AT being lost) and is transported into the plastid. Here, either all seven of the remaining reactions of the C5 pathway either occur in the plastid (Chromera), or only three or four reactions occur in the plastid (other apicomplexans) before protoporphyrinogen or coproporphyrinogen is exported back to the mitochondrion for final heme synthesis. In either case, a linear pathway is envisaged, and both taxonomic groups rely on their plastid for parts of this anabolic process.

Photosynthetic dinoflagellates appear to have retained a C5 plastid pathway, including GTR and GSA-AT (10, 42, 43). Although later-branching dinoflagellates lack ALAS of the C4 pathway, we have previously reported this mitochondrial C4 gene in Hematodinium sp. (44), and it is also in basal lineages Perkinsus and Oxyrrhis. Hematodinium sp. ALAS lacks an obvious mTP; however, when we expressed it as a GFP fusion in Toxoplasma, mitochondrial localization was seen, suggesting that it has a cryptic mitochondrial-targeting signal (Fig. S2). Neither plastid-type GTR nor GSA-AT are found in Hematodinium sp., so although the two alternative pathways for ALA synthesis, C4 and C5, coexisted during the early radiation of dinoflagellates, this redundancy was differentially lost in early and later-branching lineages.

If Hematodinium contains a mosaic pathway, such as in apicomplexans (19, 42), HemB–E could still reside in the plastid. We indeed find Hematodinium sp. does contain genes for these enzymes (GenBank accession nos. KP739891–KP739896). However, they all lack plastid-targeting leaders and are predicted to be cytosolic. Further, HemB, HemC, and HemE enzymes are phylogenetically placed with the cytosolic enzymes found in other heterotrophic eukaryotes (Fig. S3). These data provide further evidence that after the gain of a plastid, duplicated elements of this pathway persisted. Hematodinium sp. HemD is the exception, in that it groups with plastids in phylogenies, and specifically with photosynthetic dinoflagellates (Fig. 3). A plastid-derived HemD provides direct evidence that a plastid was present in the ancestor. Relocation of plastidial HemD to the cytosol has seemingly replaced the former cytosolic enzyme (Fig. 2B). The remainder of the tetrapyrrole pathway apparently takes place in the mitochondrion, because Hematodinium sp. HemY contains a predicted mTP (Fig. 2B). This location is consistent with the mitochondrion being the major site of utilization of tetrapyrroles in the absence of chlorophyll synthesis in these parasites. Only HemF and HemH were not found in the Hematodinium sp. transcriptome; however, the ferrochelatase (HemH) has not been identified in any dinoflagellate, suggesting that a divergent enzyme is present. In summary, although the re-placement of cytosolic HemD with the plastid homolog is previously unidentified, tetrapyrrole synthesis in Hematodinium sp. is otherwise typical of the classic heterotrophic cytosol/mitochondrial C4 pathway and is consistent with nonreliance on a plastid.

Lysine biosynthesis takes place in the cytosol. Prokaryotes and plants synthesize lysine via the intermediate diaminopimelate (DAP) (45). Fungi use a different pathway with an alternative intermediate (α-aminoacidic acid), and animals are unable to synthesize lysine, relying on dietary intake (45). The canonical prokaryotic DAP pathway uses nine enzymes to convert aspartate into lysine: LysC, Asd, DapA, DapB, DapD, DapC, DapE, DapF, and LysA; only DapF was not found; GenBank accession nos. KP739897–KP739900 and KP739902). All plant, green algae, and red algae DapL proteins contain an N-terminal plastid-targeting peptide for plastid targeting, and heterokont proteins (e.g., diatom Phaeodactylum and brown algae Ectocarpus) bear a bipartite leader (SP plus transit peptide), showing plastid retention of this pathway through secondary endosymbiosis. All of the Hematodinium sp. DAP enzymes lack obvious plastid-targeting presequences, and this pathway is therefore predicted to be cytosolic in Hematodinium sp. One explanation for these data is that a plastid-derived DAP pathway was relocated to the cytosol, similar to relocation of HemD. Support for this hypothesis, however, is equivocal. In DapL phylogenies, the position of Hematodinium sp., and other basal dinoflagellates such as Chromera velia and brown algae Perkinus marinus and Oxyrrhis marina, was not clearly resolved, with members of the bacterial phyla Chlamydiae, Spirochaetes, and Plantomycetes interrupting the eukaryotic/plastid clade (Fig. S4). Furthermore, DapL is not found in Apicomplexa [where host lysine is used (51)], and we did not find it in photosynthetic dinoflagellate data, so a precedent and source of a plastid-located DAP pathway in this lineage is not apparent. An alternative hypothesis would be external gain of this pathway by lateral gene transfers. The current phylogenies, however, are not able to provide strong support for this scenario either (Fig. S4).

The origin of the Hematodinium DAP pathway, therefore, is presently unclear. Nevertheless, location of this lysine biosynthetic process is unusual in that elements of both the C4 and C5 pathways have been used in a pathway of hybrid origin (19, 42) (Fig. 2A, dashed lines). Classic C4 synthesis of ALA occurs in the mitochondrion (with plastid enzymes GTR and GSA-AT being lost) and is transported into the plastid. Here, either all seven of the remaining reactions of the C5 pathway either occur in the plastid (Chromera), or only three or four reactions occur in the plastid (other apicomplexans) before protoporphyrinogen or coproporphyrinogen is exported back to the mitochondrion for final heme synthesis. In either case, a linear pathway is envisaged, and both taxonomic groups rely on their plastid for parts of this anabolic process.

Photosynthetic dinoflagellates appear to have retained a C5 plastid pathway, including GTR and GSA-AT (10, 42, 43). Although later-branching dinoflagellates lack ALAS of the C4 pathway, we have previously reported this mitochondrial C4 gene in Hematodinium sp. (44), and it is also in basal lineages Perkinsus and Oxyrrhis. Hematodinium sp. ALAS lacks an obvious mTP; however, when we expressed it as a GFP fusion in Toxoplasma, mitochondrial localization was seen, suggesting that it has a cryptic mitochondrial-targeting signal (Fig. S2). Neither plastid-type GTR nor GSA-AT are found in Hematodinium sp., so although the two alternative pathways for ALA synthesis, C4 and C5, coexisted during the early radiation of dinoflagellates, this redundancy was differentially lost in early and later-branching lineages.

If Hematodinium contains a mosaic pathway, such as in apicomplexans (19, 42), HemB–E could still reside in the plastid. We indeed find Hematodinium sp. does contain genes for these enzymes (GenBank accession nos. KP739891–KP739896). However, they all lack plastid-targeting leaders and are predicted to be cytosolic. Further, HemB, HemC, and HemE enzymes are phylogenetically placed with the cytosolic enzymes found in other heterotrophic eukaryotes (Fig. S3). These data provide further evidence that after the gain of a plastid, duplicated elements of this pathway persisted. Hematodinium sp. HemD is the exception, in that it groups with plastids in phylogenies, and specifically with photosynthetic dinoflagellates (Fig. 3). A plastid-derived HemD provides direct evidence that a plastid was present in the ancestor. Relocation of plastidial HemD to the cytosol has seemingly replaced the former cytosolic enzyme (Fig. 2B). The remainder of the tetrapyrrole pathway apparently takes place in the mitochondrion, because Hematodinium sp. HemY contains a predicted mTP (Fig. 2B). This location is consistent with the mitochondrion being the major site of utilization of tetrapyrroles in the absence of chlorophyll synthesis in these parasites. Only HemF and HemH were not found in the Hematodinium sp. transcriptome; however, the ferrochelatase (HemH) has not been identified in any dinoflagellate, suggesting that a divergent enzyme is present. In summary, although the re-placement of cytosolic HemD with the plastid homolog is previously unidentified, tetrapyrrole synthesis in Hematodinium sp. is otherwise typical of the classic heterotrophic cytosol/mitochondrial C4 pathway and is consistent with nonreliance on a plastid.

Lysine biosynthesis takes place in the cytosol. Prokaryotes and plants synthesize lysine via the intermediate diaminopimelate (DAP) (45). Fungi use a different pathway with an alternative intermediate (α-aminoacidic acid), and animals are unable to synthesize lysine, relying on dietary intake (45). The canonical prokaryotic DAP pathway uses nine enzymes to convert aspartate into lysine: LysC, Asd, DapA, DapB, DapD, DapC, DapE, DapF, and LysA; only DapF was not found; GenBank accession nos. KP739897–KP739900 and KP739902). All plant, green algae, and red algae DapL proteins contain an N-terminal plastid-targeting peptide for plastid targeting, and heterokont proteins (e.g., diatom Phaeodactylum and brown algae Ectocarpus) bear a bipartite leader (SP plus transit peptide), showing plastid retention of this pathway through secondary endosymbiosis. All of the Hematodinium sp. DAP enzymes lack obvious plastid-targeting presequences, and this pathway is therefore predicted to be cytosolic in Hematodinium sp. One explanation for these data is that a plastid-derived DAP pathway was relocated to the cytosol, similar to relocation of HemD. Support for this hypothesis, however, is equivocal. In DapL phylogenies, the position of Hematodinium sp., and other basal dinoflagellates such as Chromera velia and brown algae Perkinus marinus and Oxyrrhis marina, was not clearly resolved, with members of the bacterial phyla Chlamydiae, Spirochaetes, and Plantomycetes interrupting the eukaryotic/plastid clade (Fig. S4). Furthermore, DapL is not found in Apicomplexa [where host lysine is used (51)], and we did not find it in photosynthetic dinoflagellate data, so a precedent and source of a plastid-located DAP pathway in this lineage is not apparent. An alternative hypothesis would be external gain of this pathway by lateral gene transfers. The current phylogenies, however, are not able to provide strong support for this scenario either (Fig. S4).
pathway in the cytosol of a eukaryote is both previously unknown and provides no barrier to loss of a plastid in *Hematodinium* sp. A de novo isoprenoid biosynthesis pathway is absent. Isoprenoids are essential lipid molecules that mainly act as precursors for the synthesis of sterols, chlorophyll, and quinones. They are used as lipid anchors for some proteins and as prosthetic groups in tRNAs and proteins. Isoprenoids are built from the five-carbon precursors IPP and its isomer, dimethylallyl pyrophosphate (DMAPP). Two different pathways exist for IPP synthesis in eukaryotes (52). The canonical mevalonate (MVA) pathway consists of six enzymes and occurs in the cytosol of most heterotrophic eukaryotes, as well as some autotrophs, such as plants. A second, nonmevalonate (or DOXP) pathway, occurs in plastids, and uses seven enzymes derived from cyanobacteria. These pathways use distinct enzymes and metabolic intermediates. In particular, the MVA pathway uses acetyl-CoA and acetoacetyl-CoA and has MVA as a key intermediate, whereas the DOXP pathway uses pyruvate and glyceraldehyde-3-phosphate, resulting in production of the distinct intermediate, 1-deoxy-d-xylulose-5-phosphate (DOXP). These pathways converge with the interconversion of IPP and DMAPP by the enzyme IPP isomerase.

Within alveolates, ciliates use the MVA pathway, whereas apicomplexans exclusively use the plastid-located DOXP pathway (8, 36). Dinoflagellates also only appear to contain a plastid DOXP pathway, irrespective of whether they are photosynthetic or secondarily nonphotosynthetic (9, 10, 53, 54). Indeed, in the basal lineage *P. marinus*, the most substantial evidence for a plastid is present of all seven DOXP enzyme genes (six published, plus GenBank accession no. AB445015) with bipartite plastid-targeting leaders (12). Thus, the MVA pathway appears to have been entirely abandoned in favor of the plastid pathway in the common ancestor of dinoflagellates and apicomplexans since they diverged from ciliates (Fig. 2A).

Unexpectedly, we found no evidence for the presence of either the DOXP or the MVA pathway in *Hematodinium* sp. Nevertheless, we readily identified pathways for assembly and utilization of IPP/DMAPP-containing molecules in *Hematodinium* sp., including a farnesyl–pyrophosphate synthase (GenBank accession no. KP739903), which produces farnesyl–pyrophosphate through a sequential condensation reaction of DMAPP with two units of IPP and the alpha- and beta-subunits of both farnesyl–pyrophosphate synthase (GenBank accession nos. KP739904–KP739907), which add either a C15 or C20 prenylgroup to prenylate proteins. These data strongly suggest that *Hematodinium* sp. can scavenge IPP and/or DMAPP or other isoprenoids (Fig. 2B). A similar salvage mechanism operates in other parasites lacking a de novo pathway, such as *Cryptosporidium* (17). In this context, it is noteworthy that crustaceans produce and circulate the methylated linear C15-isoprenoid molecule, methyl farnesolate, as an important crustacean hormone. Methyl farnesolate is secreted from the mandibular organs into the hemolymph, where *Hematodinium* grows, and in crustaceans it reaches concentrations up to 0.6 μM and plays a role in morphogenesis, molt cycles, and other metabolic responses in these animals (55). Such a molecule might contribute either directly to *Hematodinium* isoprenoids or conceivably be metabolized back to IPP and DMAPP. Regardless of the source of scavenged molecules, the absence of a DOXP pathway is again consistent with loss of a plastid from *Hematodinium* sp.

Conclusion

Compilation of genomic evidence for plastid-type metabolic pathways that occur in apicomplexans, dinoflagellates, and ciliates provides insights into the pathways that were present in the common ancestor at the time of plastid gain (Fig. 2A). These analyses suggest that there have been many different lineage-specific losses and re- tense of host and/or plastid pathways. For example: (i) loss of MVA-derived IPP from the ancestor of apicomplexans and dinoflagellates; (ii) loss of type I FAS from *Plasmodium* and type II FAS from piroplasms (14, 38); and (iii) multiple losses and retentions of C4 and C5 enzymes for tetrapyrrole synthesis and complex rerouting of these pathways in both dinoflagellates and apicomplexans (19, 42). Although the origin of the DAP pathway for lysine biosynthesis in *Hematodinium* sp. is uncertain, this pathway is present in red algal plastids and most red-derived secondary plastids, so it is likely that this pathway was originally present in the apicomplexan/dinoflagellate ancestor also but was lost once or perhaps multiple times. The metabolic roles of plastids, once gained, therefore, appear to have been highly dynamic in the evolution of these lineages, and it is thus possible that dependency upon plastid functions beyond photosynthesis occurred relatively quickly. The chaotic distribution of pathway gains and losses might also suggest rapid radiation after organism acquisition. The rarity of plastid loss among the eukaryotic lineages that have lost photosynthesis suggests the likelihood of gaining metabolic freedom from plastids once they have been integrated is extremely low. *Hematodinium* sp. appears to have achieved plastid loss by retention of cytosolic pathways for synthesis of fatty acids and tetrapyrroles and supplementing loss of the DOXP pathway by scavenging intermediates for isoprenoid synthesis, likely from its crustacean hosts. The lack of plastid-type II FAS and DOXP pathways would also have removed the need for plastid Fe-S clusters that function as prosthetic groups in both pathways, thereby allowing further loss of this pathway, whereas both mitochondrial and cytosolic Fe-S cluster assembly pathways remain present. However, enzyme relocation has also been necessary, with one plastid enzyme for tetrapyrrole (HemD) relocated from the plastid to the cytosol. In comparison with *Hematodinium*, plastid loss from *Cryptosporidium*, the only other substantiated case of endosymbiotic organelle loss, appears to have been achieved primarily by extreme parasite reduction and a more complete reliance on host animal metabolites. Together, these taxa illustrate the complex machinations required to undo endosymbiosis and the reasons why it is so rarely achieved.

Materials and Methods

Transcriptome Sequencing, Assembly, and Analysis. *Hematodinium* sp. was cultured as described (33). *Hematodinium* sp. dinosporas were harvested upon sporulation from infected *Nephrops norvegicus* lobster (courtesy of N. Beevers, University of Glasgow). Total RNA was extracted, polyA-enriched, sequenced, and assembled by using the de novo assembler “Trinity” as described in SI Materials and Methods. Following assembly, ORFs were predicted using EMBoss (Version 6.5.7) getorf (56). The CEGMA test for percentage of highly conserved core eukaryotic genes was performed as described (22).

Plastid Gene Searches, Targeting Sequence Prediction, and Phylogenies. Similarity searches were performed by using NCBI BLAST. SignalP (Version 3.1) and TargetP (Version 1.1) predictions were made at standard settings as described (2007) (57). PredSL was run as described (58). For phylogenies, nearest homologs were identified from public databases including from the MMETSP (50). Alignments were generated by using MAFFT (59), manually corrected, and ambiguous sites were removed. Maximum-likelihood phylogenies were performed by using RAxML (60) using the best-fit model (LG+I+G+8) inferred by Prottest3 (61). Bayesian analyses were performed with MrBayes (Version 3.2.1) (62) by using the WAG+I+G+4 model. Markov Chain Monte Carlo runs of 1,100,000 generations were calculated with trees sampled every 200 generations and with a prior burn-in of 100,000 generations.

Genomic sequencing and metabolomics methods are outlined in SI Materials and Methods and Dataset S4.

ACKNOWLEDGMENTS. We thank Nick Katris for assistance with Toxoplasma transformation and Ellen Nisbet for critically reading this report. This work was supported by Australian Research Council (ARC) Grants DP130100572 and DP1093135; a King Abdullah University of Science and Technology Faculty Baseline Research Fund; and Victorian Life Sciences Computation Initia- lity Grant VR0254. S.G.G. was supported by Science Foundation Ireland Grant 13/1/SIRG/1215; F. was supported by an Australia Award; A.M.C. and A.B. were supported by ARC Centre of Excellence in Plant Cell Walls Grant CE110001007; and M.J.M. was supported by the National Health and Medical Research Council as a Principal Research Fellow.

Gornik et al.


Supporting Information

Gornik et al. 10.1073/pnas.1423400112

SI Materials and Methods

**Hematodinium Transcriptome Sequencing and Assembly.** Hematodinium cells were cultured and dinosporas harvested as described in the main body of the manuscript. Total RNA was extracted using TRIzol reagent (Invitrogen). Polyadenylated RNA was extracted from total RNA by using oligo(dT) Dynabeads, and libraries were prepared by using TrueSeq Stranded mRNA LT Sample Prep kits (Illumina). First-strand cDNA was synthesized by using randomly primed oligos. Strand specificity was achieved through dUTPs incorporation during second-strand synthesis. The cDNA was adapter-ligated, and the libraries were amplified. The 100-bp PE sequencing was performed on a HiSeq2000 platform. RNA-Seq reads were cleared of adapter sequences and trimmed to a quality value of Q30 by using fastq-mcf (code.google.com/p/ea-utils/). Assemblies were generated by using Trinity (r2013.08.14) (1). The assembly was clustered by using CD-HIT-EST (Version 4.6) (2) into the most similar cluster that meets the threshold (-g 1) with a sequence identity threshold of 90% (-c 0.9), a word size of n = 8 (-n 8), comparing both strands (+ -r). Further clustering was carried out by using CAP3 (3) with default settings as described in Yang and Smith (4).

**Hematodinium Genome Sequencing and Assembly.** Using a Qiagen “Blood&Tissue” extraction, ~120 μg of Hematodinium high-molecular-weight genomic DNA was purified. From this DNA, three PCR-free 350-bp-insert libraries were prepared, one PCR-free 500-bp-insert library, and mate-pair libraries of 2.9, 3.5, 4, 5.7, 6, and 9 kb by using either Nextera (4-, 6-, and 9-kb libraries) or TrueSeq (2.9-, 3.5-, and 5.7-kb libraries) library preparation kits. For the 350-bp and TrueSeq MP libraries 100-bp, paired-end (PE) sequencing was performed on the Illumina HiSeq2000 platform. For the Nextera MP libraries, 250-bp PE sequencing was performed on the Illumina MySeq platform. More than 1.6 billion reads across all 10 libraries have been generated to date. Preliminary assembly using SOAPdenovo [soap.genomics.org.cn/soapdenovo.html (last accessed 02/02/2015)] yielded a 4,769-Mbp genome with 30x coverage and a GC content of 47.31%. The genome sequence currently is spread across 869,500 scaffolds with an N50 of 17,235 bp, 118,385 scaffolds >10 kbp, 653 scaffolds >100 kbp, and a longest scaffold at 186 kbp. The assembly and annotation of this genome is ongoing.

**Metabolite Extraction.** Hematodinium was cultured as described in the main text. Total RNA was extracted from 2.5 × 10^6 mL^-1 cells, and cDNA was synthesized by using random primers (Invitrogen). Polyadenylated RNA was extracted from total RNA by using oligo(dT) Dynabeads, and libraries were prepared by using TrueSeq Stranded mRNA LT Sample Prep kits (Illumina). First-strand cDNA was synthesized by using randomly primed oligos. Strand specificity was achieved through dUTPs incorporation during second-strand synthesis. The cDNA was adapter-ligated, and the libraries were amplified. The 100-bp PE sequencing was performed on a HiSeq2000 platform. RNA-Seq reads were cleared of adapter sequences and trimmed to a quality value of Q30 by using fastq-mcf (code.google.com/p/ea-utils/). Assemblies were generated by using Trinity (r2013.08.14) (1). The assembly was clustered by using CD-HIT-EST (Version 4.6) (2) into the most similar cluster that meets the threshold (-g 1) with a sequence identity threshold of 90% (-c 0.9), a word size of n = 8 (-n 8), comparing both strands (+ -r). Further clustering was carried out by using CAP3 (3) with default settings as described in Yang and Smith (4).

**Hematodinium Genome Sequencing and Assembly.** Using a Qiagen “Blood&Tissue” extraction, ~120 μg of Hematodinium high-molecular-weight genomic DNA was purified. From this DNA, three PCR-free 350-bp-insert libraries were prepared, one PCR-free 500-bp-insert library, and mate-pair libraries of 2.9, 3.5, 4, 5.7, 6, and 9 kb by using either Nextera (4-, 6-, and 9-kb libraries) or TrueSeq (2.9-, 3.5-, and 5.7-kb libraries) library preparation kits. For the 350-bp and TrueSeq MP libraries 100-bp, paired-end (PE) sequencing was performed on the Illumina HiSeq2000 platform. For the Nextera MP libraries, 250-bp PE sequencing was performed on the Illumina MySeq platform. More than 1.6 billion reads across all 10 libraries have been generated to date. Preliminary assembly using SOAPdenovo [soap.genomics.org.cn/soapdenovo.html (last accessed 02/02/2015)] yielded a 4,769-Mbp genome with 30x coverage and a GC content of 47.31%. The genome sequence currently is spread across 869,500 scaffolds with an N50 of 17,235 bp, 118,385 scaffolds >10 kbp, 653 scaffolds >100 kbp, and a longest scaffold at 186 kbp. The assembly and annotation of this genome is ongoing.

**Metabolite Extraction.** Hematodinium was cultured as described in the main text. Total RNA was extracted from 2.5 × 10^6 mL^-1 cells, and cDNA was synthesized by using random primers (Invitrogen). Polyadenylated RNA was extracted from total RNA by using oligo(dT) Dynabeads, and libraries were prepared by using TrueSeq Stranded mRNA LT Sample Prep kits (Illumina). First-strand cDNA was synthesized by using randomly primed oligos. Strand specificity was achieved through dUTPs incorporation during second-strand synthesis. The cDNA was adapter-ligated, and the libraries were amplified. The 100-bp PE sequencing was performed on a HiSeq2000 platform. RNA-Seq reads were cleared of adapter sequences and trimmed to a quality value of Q30 by using fastq-mcf (code.google.com/p/ea-utils/). Assemblies were generated by using Trinity (r2013.08.14) (1). The assembly was clustered by using CD-HIT-EST (Version 4.6) (2) into the most similar cluster that meets the threshold (-g 1) with a sequence identity threshold of 90% (-c 0.9), a word size of n = 8 (-n 8), comparing both strands (+ -r). Further clustering was carried out by using CAP3 (3) with default settings as described in Yang and Smith (4).

Fig. S1. (A) Conserved domains and architecture of Hematodinium sp. putative FAS-I and PKS annotated by using the Conserved Domain Database (CDD) from NCBI [Marchler-Bauer, et al. (1)] inferring putative FAS-I or PKS function based on domain occurrences. Protein length is shown for full-length ORFs. (B) Incorporation of stable isotope-labeled $^{13}$C-U-glucose into Hematodinium saturated C14:0 and C16:0 fatty acids. “Molecules labeled” indicates the fraction of all molecules into which $^{13}$C-U-glucose is incorporated; “carbons labeled” reflects the fraction incorporation of carbons that are $^{13}$C in a given labeled molecule. (C) Mass isotopologue fractional abundances of myristic acid (C14:0) and palmitic acid (C16:0) saturated fatty acids (FAs) generated in $^{13}$C-U-glucose–fed Hematodinium cells. “m0, m1, m2, etc.” indicates the monoisotopic mass containing 0, 1, 2, etc. $^{13}$C atoms. Error bars indicate SD (n = 3).


Fig. S2. Plate showing Hematodinium ALAS-GFP overexpression in T. gondii revealing mitochondrial localization using a Tom40 antibody as mitochondrial marker.
HemB

Fig. S3. (Continued)
HemC

Fig. S3. (Continued)
HemD

Fig. S3. (Continued)
Fig. 53. HemB, HemC, and HemE phylogenies showing that Hematodinium enzymes group with the cytosolic proteins of heterotrophic eukaryotes. Support values (ML bootstraps/Bayesian posterior probabilities) shown for major nodes only. Full phylogeny for HemD showing that Hematodinium HemD is of plastid origin. Presence/absence of targeting leader sequences for the plastid clade is indicated. Support values (ML bootstraps/Bayesian posterior probabilities) only shown for major clades.
Phylogeny for DapL showing that the position of Hematodinium DapL and other basal dinozoan DapLs (for *P. marinus* and *O. marina*) is not clearly resolved, with members of the bacterial phyla Chlamydiae, Spirochaetes, and Planctomycetes interrupting the eukaryotic/plastid clade. Presence/absence of targeting leader sequences for the plastid clade is indicated. Support values (ML bootstraps/Bayesian posterior probabilities) are only shown for major clades.

Gornik et al. www.pnas.org/cgi/content/short/1423400112
Table S1. Transcriptome assembly

<table>
<thead>
<tr>
<th>Assembly workflow</th>
<th>Trophont</th>
<th>Dinospore</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of PE reads</td>
<td>85,143,391</td>
<td>74,761,446</td>
<td></td>
</tr>
<tr>
<td>Assembly using Trinity (1)</td>
<td>417,603</td>
<td>407,894</td>
<td></td>
</tr>
<tr>
<td>Mean length in base pairs, bp ±SD (SD)</td>
<td>1,653 (1,430)</td>
<td>1,554 (1,290)</td>
<td></td>
</tr>
<tr>
<td>Clustering</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD-HIT-EST (~0.9) (2)</td>
<td>246,717 (&gt;200 bp)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean length in bp ±SD</td>
<td>1,037 (1,058)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAP3 (final assembly) (3)</td>
<td>222,704 (&gt;200 bp)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean length in bp ±SD</td>
<td>1,106 (1,101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique CDS (EMBOSS ORFs &gt; 50 aa) (4)</td>
<td>118,749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean length, aa ±SD</td>
<td>103 (121)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique CDS (EMBOSS ORFs &gt; 180 aa)</td>
<td>29,510</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean length ±SD</td>
<td>443 (330)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table S2. CEGMA completeness of dinoflagellate transcriptomes and genome

<table>
<thead>
<tr>
<th>Species</th>
<th>No. of ESTs</th>
<th>No. of CEGs</th>
<th>Completeness, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hematodinium sp. (all transcripts)</td>
<td>222,704</td>
<td>213</td>
<td>85.9</td>
</tr>
<tr>
<td>Hematodinium sp. (unique ORFs &gt; 180 aa)</td>
<td>31,020</td>
<td>179</td>
<td>78.2</td>
</tr>
<tr>
<td>Symbiodinium sp. (strains CassKB8 &amp; Mf1.05b) (1)</td>
<td>148,436</td>
<td>176</td>
<td>71.0</td>
</tr>
<tr>
<td>Symbiodinium sp. genome (2)</td>
<td>—</td>
<td>109</td>
<td>44.0</td>
</tr>
<tr>
<td>Lingulodinium polyedrum (3)</td>
<td>74,958</td>
<td>174</td>
<td>70.2</td>
</tr>
<tr>
<td>Karenia brevis*</td>
<td>65,226</td>
<td>17</td>
<td>6.9</td>
</tr>
<tr>
<td>Oxyrrhis marina*</td>
<td>18,024</td>
<td>18</td>
<td>7.3</td>
</tr>
<tr>
<td>Karlodinium veneficum*</td>
<td>11,903</td>
<td>25</td>
<td>10.1</td>
</tr>
</tbody>
</table>

*ESTs from GenBank@NCBI (November 2013).


Other Supporting Information Files

- Dataset S1 (XLSX)
- Dataset S2 (XLSX)
- Dataset S3 (XLSX)
- Dataset S4 (DOC)