

# Endosymbiosis undone: stepwise elimination of the plastid in the dinoflagellate *Hematodinium* sp.

Sebastian G Gornik<sup>1,2</sup>, Febrimarsa<sup>1</sup>, Andrew Cassin<sup>3</sup>, James I MacRae<sup>4,5</sup>, Abhinay Ramaprasad<sup>6</sup>, Zineb Rchiad<sup>6</sup>, Malcolm J McConville<sup>5</sup>, Antony Bacic<sup>1,3</sup>, Geoffrey I McFadden<sup>1</sup>, Arnab Pain<sup>6</sup> and Ross F Waller<sup>1,7,\*</sup>

<sup>1</sup> School of Botany, University of Melbourne, 3010 Victoria, Australia

<sup>2</sup> School of Natural Sciences, NUI Galway, Galway, Ireland

<sup>3</sup> ARC Centre of Excellence in Plant Cell Walls, School of Botany, University of Melbourne, Victoria 3010, Australia

<sup>4</sup> The Francis Crick Institute, London, NW7 1AA, United Kingdom

<sup>5</sup> Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, Victoria 3010, Australia

<sup>6</sup> BESE Division, King Abdullah University of Science and Technology, Thuwal-Jeddah, Saudi Arabia

<sup>7</sup> Department of Biochemistry, University of Cambridge, CB2 1QW, United Kingdom

\* Corresponding author: Ross F. Waller

Short Title: Elimination of an endosymbiotic organelle

Keywords: organelle loss, plastid, endosymbiosis, metabolism, diaminopimelate aminotransferase (*DapL*)

## Abstract (250 words max)

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 non-photosynthetic 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 organelle 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 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 new dimension of endosymbiosis—life after the organelle.

### **Significance Statement**

Endosymbiotic organelles are a defining feature of eukaryotes—the last common ancestor and all extant eukaryotes possess at least a mitochondrial derivative. While 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.

\body

### **Introduction**

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. But approximately half of dinoflagellates are heterotrophic and adapted to micropredation and parasitism (1). Even the deepest branching members of the dinoflagellate lineage, which are predominately parasites and predators (2), are secondarily non-photosynthetic as there is compelling evidence that the common ancestor of dinoflagellates and their sister phylum, Apicomplexa, possessed a photosynthetic plastid (3). All apicomplexans, which include the malaria parasite, have parasitic lifestyles and contain a relic non-photosynthetic plastid called the apicoplast (except *Cryptosporidium*, see below). Basal members of the apicomplexan lineage *Chromera* and *Vitrella*, however, maintain photosynthesis (4, 5) and their plastids share several distinctive synapomorphies with dinoflagellates (3). Whereas apicomplexan

parasites are thought to stem from a single loss of photosynthesis, photosynthesis has been lost on multiple occasions independently in dinoflagellates (6). Ciliates and colponemids represent the basal members of the infrakingdom Alveolata, but they appear to lack plastid organelles (7, 8). Despite the parallel losses of photosynthesis, 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*, *Crypthecodinium*), 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 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, either by: 1) maintaining cytosolic pathways; 2) scavenging an exogenous supply of metabolites; 3) eliminating the requirement for the metabolite; 4) relocating a plastid pathway; or 5) a combination these. Only one apicomplexan group, *Cryptosporidium*, has solved this complex puzzle by maintaining cytosolic fatty acid synthesis, and scavenging host isoprene precursors and the tetrapyrrole heme (16-18) allowing it to lose the apicoplast. Drastic reduction of heme requirements (*Cryptosporidium* has only one known cytochrome) (19), and limitation to a single host to complete its lifecycle, might also have contributed to overcoming plastid dependency. To date, this is the only substantiated case of plastid loss in any eukaryote, highlighting the difficulty of this evolutionary step.

We have investigated the basal, non-photosynthetic 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-Seq) and genomic data and analyzed these for evidence of a plastid and plastid-associated metabolic pathways. Our data indicate that through some novel metabolic pathway contortions, *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 this 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 >200 bp (Table S1). To assess what portion of the complete gene content of *Hematodinium* sp. these data represent, we compared it to known databases and recovered 15,510 Swiss-Prot (UniProtKB) matches ( $<e^{-5}$ ) and 9,207 KEGG (Kyoto Encyclopedia of Genes and Genomes) 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 number of unique transcripts (222,704) is likely an overestimation of the true coding capacity due to: 1) observed coding sequence redundancy where UTRs (untranslated regions) showed either divergent or unrelated sequence (the reasons for which are under investigation); and 2) high-representation of very short ORFs (open reading frames). 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). Taken 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 post-translationally 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, that shares conserved, recognizable features for almost all plastids (27). Most of the biochemistry of the non-photosynthetic plastids of apicomplexans is defined by nucleus-encoded plastid proteins bearing this distinctive bi-partite (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 bi-partite leaders, plastid-type functions, and apicoplast-located orthologues) (29). Using RNA-Seq data permissive TBLASTN searches (E value  $<e^{-3}$ ) were performed, and scrutinized by reciprocal BLASTP searches against the NCBI RefSeq database (E value  $<e^{-5}$ ), and 81 proteins were recovered (Fig. 1A, 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 mitochondrial or cytosolic homologues. To validate the matches as potential plastid proteins we then analyzed them for the presence of a bi-partite 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, 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 non-plastid 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 bi-partite targeting peptides provides evidence of reduced plastids in *Perkinsus*, *Oxyrrhis*, and *Cryptothecodinium* (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 mitochondrial targeting peptides (mTP). 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, 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 either of very low abundance, too divergent, or exceptional in lacking identifiable bi-partite leaders.

Plastid genes can also be encoded directly in the organelle, although dinoflagellates are unusual in having reduced this to only 14 genes (3). We searched for all known protein genes found in alveolates plastids (82 genes in total, found in either dinoflagellates, apicomplexans or chromerids (3)) and found no matches in our RNA-Seq data (Fig. 1C, 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 ((34) and references therein). We have currently 4,769 Mbp of data assembled (30x coverage, N50 = 17,235 bp, 118,385 scaffolds >10 kbp, 653 scaffolds >100 kbp). Preliminary analysis of these data reveals that genes are densely populated with very large introns and are, therefore, spread over 10s to 100s of kbps. 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 anabolic pathways that are essential for the completion of the life cycle—including fatty acid biosynthesis, tetrapyrrole synthesis, and isopentenyl pyrophosphate (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 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 systems typically consists of seven separate mono-functional proteins, and this multiprotein complex, called type II FAS, occurs in plastids. Evolution of FAS in the eukaryotic cytosol has entailed fusion of most or all of the necessary enzyme modules to form the multi-enzyme type I FAS complex (Fig. 2). Plastid-containing organisms can, therefore, contain both type I and II FAS capabilities, or they might lose one or the other. In apicomplexans, *Toxoplasma* and *Eimeria* have both, whereas *Plasmodium* retains only the plastid type II form, and *Cryptosporidium* only cytosolic type I (14, 38). Thus, the common ancestor of apicomplexans and dinoflagellates must have had both forms. Although FAS in dinoflagellates is less well studied, these organisms can synthesize fatty acids *de novo* (39, 40). Gene disruption of type II FAS in *Plasmodium* shows that this plastid pathway is essential during the liver stage and, in a human strain, the insect stage of the lifecycle, implying that FAS might be a function mandating plastid retention (13, 35, 41).

We searched the *Hematodinium* sp. transcript data and found coding sequence for all necessary FAS enzymes, but never as separate genes. Instead they were always present on transcripts of large multi-enzyme-complexes, consistent with the presence of a fused type I FAS and the absence of a type II FAS complex (Fig. 2B). Our RNA-Seq data recovered three large FAS-enzyme containing transcripts (GenBank# KP739886-KP739888), whose full-length ORFs all lack a bipartite leader and are predicted to be cytosolic, as well as two further incomplete transcripts with multiple FAS-like enzyme domains (GenBank# KP739889-KP739890; Fig. S1A). Polyketide syntases (PKS) are a related family of proteins that are considered degenerate FASs where one or more modules have been lost, and PKS are known to occur in dinoflagellates (40). While 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 [<sup>13</sup>C]-U-glucose into *Hematodinium* sp. fatty acids. <sup>13</sup>C-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 <sup>13</sup>C-glucose fed *Hematodinium* sp. generating a series of fatty acid isotopomers that increased by +2 atomic mass units (amu), consistent with *de novo* synthesis of fatty acids from a malonyl-CoA donor

(Fig. S1B and C). These analyses suggest that *Hematodinium* sp. expresses a large type I FAS complex that would allow loss of a plastid-located type II complex.

#### *Tetrapyrrole biosynthesis occurs in the cytosol*

Tetrapyrrole biosynthesis is required for the synthesis of heme and chlorophylls, and occurs via two related pathways (19) (Fig. 2A). The C4 pathway, which occurs in heterotrophic eukaryotes, commences with the synthesis of  $\delta$ -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 uroporphyrinogen 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-tRNA by glutamyl-tRNA reductase (GTR) followed by glutamate 1-semialdehyde aminotransferase (GSA-AT). ALA is then converted into tetrapyrroles via the same enzymes found in the C4 pathway, but these 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 pathway have been employed 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). While 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

mitochondrial targeting peptide, 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 GTR nor GSA-AT are found in *Hematodinium* sp., so while the two alternative pathways for ALA synthesis, C4 and C5, co-existed 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# 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). This provides 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). This provides direct evidence that a plastid was present in the *Hematodinium* 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, as *Hematodinium* sp. HemY contains a predicted mitochondrial targeting peptide (mTP) (Fig. 2B). This 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 a divergent enzyme is present. In summary, although replacement of cytosolic HemD with the plastid homologue is novel, tetrapyrrole synthesis in *Hematodinium* sp. is otherwise typical of the classic heterotrophic cytosol/mitochondrial C4 pathway, and consistent with non-reliance 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 ( $\alpha$ -amino adipic acid), and animals are unable to synthesize lysine, relying on dietary intake (45). The canonical prokaryotic DAP pathway utilizes nine enzymes to convert aspartate into lysine: LysC, Asd, DapA, DapB, DapD, DapC, DapE, DapF and LysA. Chloroplasts are the site of lysine synthesis from aspartate in plants (46), a function assumed to originate from their cyanobacterial forebears. However, plants lack identifiable *dapD*, *dapC*, and *dapE* genes, suggesting the

presence of an alternative pathway. A novel enzyme, LL-diaminopimelate aminotransferase (DapL), was recently discovered in *Arabidopsis*, which replaces DapC-E, and defines a new variant of the DAP pathway (47, 48). DapL is also reported from green algae, cyanobacteria, and amongst other Bacteria and Archaea (48, 49). In broad genome and transcriptome surveys of diverse eukaryotes (MMETSP) (50) we identify DapL also in red algae and amongst secondary plastid-containing algae (heterokonts, haptophytes, cryptophytes and basal dinoflagellates) (Fig. S4). In eukaryotes DapL is restricted to plastid-containing lineages.

We found *dapL* in *Hematodinium* sp. (GenBank# KP739901), along with the genes for the other enzymes of the DAP pathway (*lysC*, *asd*, *dapA*, *dapB* and *lysA*: only *dapF* was not found; GenBank# KP739897-KP739900, KP739902). All plant, green algal and red algal DapL proteins contain an N-terminal plastid-targeting transit peptide for plastid targeting, and heterokont proteins (e.g. diatom *Phaedactylum* and brown algae *Ectocarpus*) bear a bipartite leader (signal peptide plus transit peptide) showing plastid retention of this pathway through secondary endosymbiosis. All of the *Hematodinium* sp. DAP enzymes lack obvious plastid-targeting pre-sequences, and this pathway is therefore predicted to be cytosolic in *Hematodinium* sp. One explanation for this 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 dinozoan taxa *Perkinsus marinus* and *Oxyrrhis marina*, was not clearly resolved, with members of the bacterial phyla Chlamydiae, Spirochaetes, and Planctomycetes interrupting the eukaryotic/plastid clade (Fig. S4). Furthermore, DapL is not found in Apicomplexa (where host lysine is used (51)), nor did we 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 pathway in the cytosol of a eukaryote is both novel 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 isopentenyl pyrophosphate (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, non-mevalonate (or DOXP) pathway, occurs in plastids, and uses seven enzymes derived from cyanobacteria. These pathways utilize distinct enzymes and metabolic intermediates. In particular, the MVA pathway utilizes acetyl-CoA and acetoacetyl-CoA and has mevalonate as a key intermediate, while the DOXP pathway utilizes pyruvate and glyceraldehyde-3-phosphate resulting in production of the distinct intermediate, 1-deoxy-D-xylulose-5-phosphate (DOXP). These pathways converge with the inter-conversion of IPP and DMAPP by the enzyme IPP isomerase.

Within alveolates, ciliates use the mevalonate pathway, while 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 non-photosynthetic (9, 10, 53, 54). Indeed, in the basal lineage *Perkinsus marinus*, the most substantial evidence for a plastid is presence of all seven DOXP enzyme genes (six published, plus GenBank# 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 DOXP or the mevalonate pathways 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 (FPPS; GenBank# 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 farnesyltransferase and geranylgeranyltransferase (GenBank# KP739904- KP739907), which add either a C15 or C20 prenyl-group to prenylate proteins. This strongly suggests 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 farnesoate, as an important crustacean hormone. Methyl Farnesoate is secreted from the mandibular organs into the haemolymph where *Hematodinium* grows, and in crustaceans reaches concentrations up to 0.6  $\mu\text{M}$  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 retentions of host and/or plastid pathways. For example: 1) loss of MVA-derived IPP from the ancestor of apicomplexans and dinoflagellates; 2) loss of type I FAS from *Plasmodium* and type II FAS from piroplasmids (14, 38); and 3) 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 following organelle acquisition. The rarity of plastid loss amongst 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 this 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, while both mitochondrial and cytosolic Fe-S cluster assembly pathways remain present. Yet enzyme relocation has also been necessary, with one plastid enzyme for tetrapyrrole (HemD) relocated from the plastid to cytosol. In comparison to *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 metabolites. Together these taxa illustrate the complex machinations required to undo endosymbiosis and the reasons why it is so rarely achieved.

## **Material and Methods**

*Transcriptome sequencing, assembly and analysis* *Hematodinium* sp. was cultured as previously described (33). *Hematodinium* sp. dinospores were harvested upon sporulation from infected *Nephrops norvegicus* lobster (courtesy of Dr N. Beevers, University of Glasgow, UK). Total RNA was extracted, polyA-enriched, sequenced and assembled using the *de novo* assembler 'Trinity' as described in the Supplementary Information. Following assembly, open reading frames were predicted using EMBOSS (v6.5.7) *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 using NCBI BLAST. SignalP 3.1 and TargetP 1.1 predictions were made at standard settings as previously described (2007) (57). PredSL was run as described (58). For phylogenies, nearest homologues were identified from public databases including from the MMETSP (50). Alignments were generated using MAFFT (59), manually corrected and ambiguous sites removed. Maximum likelihood phylogenies were performed using RAxML (60) using the best-fit model (LG+I+G+8) inferred by Prottest3 (61). Bayesian analyses were performed with MrBayes 3.2.1 (62) using the WAG+I+G+4 model. Markov Chain Monte Carlo (MCMC) 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 the Supplementary Information.

## **Acknowledgements**

The authors thank Nick Katris for his assistance with *Toxoplasma* transformation, and Ellen Nisbet for critically reading this report. This work was supported by ARC grants DP130100572 and DP1093395, a KAUST- faculty Baseline Research Fund, and VLSCI grant VR0254. SGG acknowledges support from the Science Foundation Ireland (grant 13/SIRG/2125), F from an Australia Award, AC and AB from the ARC Centre of Excellence in

Plant Cell Walls (grant CE110001007), and MJM from the NHMRC as a Principal Research Fellow.

## References

1. Coats DW (1999) Parasitic life styles of marine dinoflagellates. *J Euk Micro* 46:402–409.
2. Bachvaroff TR et al. (2014) Dinoflagellate phylogeny revisited: Using ribosomal proteins to resolve deep branching dinoflagellate clades. *Mol Phylogenet Evol* 70:314–322.
3. Janouškovec J, Horák A, Obornik M, Lukes J, Keeling PJ (2010) A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proc Natl Acad Sci USA* 107:10949–10954.
4. van Dooren GG, Striepen B (2013) The algal past and parasite present of the apicoplast. *Annu Rev Microbiol* 67:271–289.
5. Moore RB et al. (2008) A photosynthetic alveolate closely related to apicomplexan parasites. *Nature* 451:959–963.
6. Saldarriaga J et al (2001) Dinoflagellate nuclear SSU rRNA phylogeny suggests multiple plastid losses and replacements. *J Mol Evol* 53:204–13.
7. Janouškovec J et al. (2013) Colponemids represent multiple ancient alveolate lineages. *Curr Biol* 23:2546–2552.
8. Eisen JA et al. (2006) Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLOS Biology* 4:e286.
9. Slamovits CH, Keeling PJ (2008) Plastid-derived genes in the nonphotosynthetic alveolate *Oxyrrhis marina*. *Mol Biol Evol* 25:1297–1306.
10. Sánchez Puerta MV, Lippmeier JC, Apt KE, Delwiche CF (2007) Plastid genes in a non-photosynthetic dinoflagellate. *Protist* 158:105–117.
11. Robledo JAFN et al. (2011) The search for the missing link: A relic plastid in *Perkinsus*? *Int J Parasitol* 41:1217–1229.

12. Matsuzaki M, Kuroiwa H, Kuroiwa T, Kita K, Nozaki H (2008) A cryptic algal group unveiled: a plastid biosynthesis pathway in the Oyster parasite *Perkinsus marinus*. *Mol Biol Evol* 25:1167–1179.
13. Vaughan AM et al. (2009) Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. *Cell Microbiol* 11:506–520.
14. Seeber F, Soldati-Favre D (2010) Metabolic pathways in the apicoplast of apicomplexa. *Int Rev Cell Mol Biol* 281:161–228.
15. Ramakrishnan S et al. (2012) Apicoplast and endoplasmic reticulum cooperate in fatty acid biosynthesis in apicomplexan parasite *Toxoplasma gondii*. *J Biol Chem* 287:4957–4971.
16. Xu P et al. (2004) The genome of *Cryptosporidium hominis*. *Nature* 431:1107–1112.
17. Bessoff K, Sateriale A, Lee KK, Huston CD (2013) Drug repurposing screen reveals FDA-approved inhibitors of human HMG-CoA reductase and isoprenoid synthesis that block *Cryptosporidium parvum* growth. *Antimicrob Agents Chemother* 57:1804–1814.
18. Zhu G, Marchewka MJ, Keithly JS (2000) *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiol* 146:315–321.
19. van Dooren GG, Kennedy AT, McFadden GI (2012) The use and abuse of heme in apicomplexan parasites. *Antioxid Redox Signal* 17:634–56.
20. Stentiford GD, Shields JD (2005) A review of the parasitic dinoflagellates *Hematodinium* species and *Hematodinium*-like infections in marine crustaceans. *Dis Aquat Org* 66:47–70.
21. Appleton PL (1996) Investigations on the cytology and life-cycle of the parasitic dinoflagellate *Hematodinium* sp associated with mortality of *Nephrops norvegicus*. Dissertation (University of Glasgow).
22. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–1067.
23. Tiessen A, Pérez-Rodríguez P, Delaye-Arredondo LJ (2012) Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial

species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Res Notes* 5:85.

24. Brocchieri L, Karlin S (2005) Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* 33:3390–3400.
25. Nassoury N, Cappadocia M, Morse D (2003) Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J Cell Sci* 116:2867–2874.
26. Waller RF, Reed MB, Cowman AF, McFadden GI (2000) Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J* 19:1794–1802.
27. Patron NJ, Waller RF (2007) Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *BioEssays* 29:1048–1058.
28. Ralph SA et al. (2004) Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat Rev Microbiol* 2:203–216.
29. Carlton JM et al. (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455:757–763.
30. van Dooren GG, Stimmler LM, McFadden GI (2006) Metabolic maps and functions of the *Plasmodium* mitochondrion. *FEMS Microbiol Rev* 30:596–630.
31. Jackson CJ, Gornik SG, Waller RF (2012) The mitochondrial genome and transcriptome of the basal dinoflagellate *Hematodinium* sp.: character evolution within the highly derived mitochondrial genomes of dinoflagellates. *Gen Biol Evol* 4:59–72.
32. Beauchemin M et al. (2012) Dinoflagellate tandem array gene transcripts are highly conserved and not polycistronic. *Proc Natl Acad Sci USA* 109:15793–15798.
33. Gornik SG et al. (2012) Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Curr Biol* 22:2303–2312.
34. LaJeunesse TC, Lambert G, Andersen RA, Coffroth MA, Galbraith DW (2005) *Symbiodinium* (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. *J Phycol* 41:880–886.
35. Yu M et al. (2008) The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. *Cell Host Microbe* 4:567–578.

36. Yeh E, DeRisi JL (2011) Chemical rescue of malaria parasites lacking an apicoplast defines organelle function in blood-stage *Plasmodium falciparum*. *PLoS Biology* 9:e1001138.
37. Botté C, Dubar F, McFadden G (2012) *Plasmodium falciparum* apicoplast drugs: targets or off-targets? *Chem Rev* 112:1269-83.
38. Zhu G (2004) Current progress in the fatty acid metabolism in *Cryptosporidium parvum*. *J Eukaryot Microbiol* 51:381-388.
39. Sonnenborn U, Kunau W-H (1982) Purification and properties of the fatty acid synthetase complex from the marine dinoflagellate, *Cryptothecodinium cohnii*. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* 712:523-534.
40. Van Dolah FM et al. (2013) Subcellular localization of dinoflagellate polyketide synthases and fatty acid synthase activity. *J Phycol* 49:1118-1127.
41. van Schaijk BCL et al. (2014) Type II fatty acid biosynthesis is essential for *Plasmodium falciparum* sporozoite development in the midgut of *Anopheles* mosquitoes. *Eukaryot Cell* 13:550-559.
42. Koreny L, Sobotka R, Janouškovec J, Keeling PJ, Obornik M (2011) Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. *Plant Cell* 23:3454-3462.
43. Patron NJ, Waller RF, Keeling PJ (2006) A tertiary plastid uses genes from two endosymbionts. *J Mol Biol* 357:1373-1382.
44. Danne JC, Gornik SG, Macrae JI, McConville MJ, Waller RF (2012) Alveolate mitochondrial metabolic evolution: dinoflagellates force reassessment of the role of parasitism as a driver of change in Apicomplexans. *Mol Biol Evol* 30:123-39.
45. Velasco AM, Leguina JI, Lazcano A (2002) Molecular evolution of the lysine biosynthetic pathways. *J Mol Evol* 55:445-459.
46. Mills WR, Wilson KG (1978) Amino acid biosynthesis in isolated pea chloroplasts: metabolism of labeled aspartate and sulfate. *FEBS Lett* 92:129-132.

47. Hudson AO, Singh BK, Leustek T, Gilvarg C (2006) An LL-diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway in plants. *Plant Physiol* 140:292–301.
48. McCoy AJ et al. (2006) L, L-diaminopimelate aminotransferase, a trans-kingdom enzyme shared by *Chlamydia* and plants for synthesis of diaminopimelate/lysine. *Proc Natl Acad Sci USA* 103:17909–17914.
49. Dobson RCJ, Girón I, Hudson AO (2011) L,L-diaminopimelate aminotransferase from *Chlamydomonas reinhardtii*: a target for algaecide development. *PLOS ONE* 6:e20439.
50. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biol* 12:e1001889.
51. Shanmugasundram A, Gonzalez-Galarza FF, Wastling JM, Vasieva O, Jones AR (2013) Library of Apicomplexan Metabolic Pathways: a manually curated database for metabolic pathways of apicomplexan parasites. *Nucleic Acids Res* 41:D706–13.
52. Eisenreich W, Bacher A, Arigoni D, Rohdich F (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cell Mol Life Sci* 61:1401–1426.
53. 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.
54. Bachvaroff TR, Place AR (2008) From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLOS ONE* 3:e2929.
55. Nagaraju GPC (2007) Is methyl farnesoate a crustacean hormone? *Aquaculture* 272:39–54.
56. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277.
57. Emanuelsson O, Brunak S, Heijne von G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2:953–971.

58. Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ (2006) PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. *Genomics Proteomics Bioinformatics* 4:48–55.
59. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780.
60. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
61. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.
62. Ronquist F et al. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biol* 61:539–542.

## Figure legends

Fig. 1: Search strategy for plastid-encoded genes, and nucleus-encoded genes for plastid- (or mitochondrial-) targeted proteins. Gene query sets were from: A) predicted apicoplast-targeted proteins common to *Plasmodium*; and B) predicted *P. falciparum* mitochondrion-targeted proteins; C) plastid-encoded genes common to apicomplexans, dinoflagellates and chromerids. A list of genes and full search results are shown in Datasets S1-3. FL, full length; SP, signal peptide; cTP, chloroplast transit peptide; mTP, mitochondrial targeting peptide.

Fig. 2: Reconstructed metabolic pathways in (A) the common ancestor of apicomplexans and dinoflagellates at the time of plastid gain, and (B) in *Hematodinium* sp., 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: green=plastid, yellow=cytosol and red=mitochondrion. Hatched (green/white) DAP pathway indicates uncertain origin of

this typically plastid-located pathway in *Hematodinium*. FAS I/II, type I/II Fatty Acid Synthase; MVA, mevalonate isopentenyl pyrophosphate (IPP) pathway; DOXP, 1-deoxy-D-xylulose-5-phosphate IPP pathway; 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-sulphur cluster pathway; DAP, diaminopimelate lysine pathway; C4/C5 pathways for tetrapyrrole (TP) synthesis differ only by the reactions to ALA ( $\delta$ -aminolevulinic acid) and their location.

Fig. 3: HemD (uroporphyrinogen III synthase) phylogeny showing a plastid-derived protein in *Hematodinium* sp. Presence/absence of targeting leader sequences for the plastid clade is indicated. Support values (ML bootstraps/Bayesian posterior probabilities) only shown for major clades. Wedges indicate collapsed clades shown in full in Fig. S3.