Check for updates

A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis

Elisabeth Hehenberger[a,1,2], Rebecca J. Gast[b], and Patrick J. Keeling[a,1]

[a]Department of Botany, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; and [b]Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

Plastid endosymbiosis has been a major force in the evolution of eukaryotic cellular complexity, but how endosymbionts are integrated is still poorly understood at a mechanistic level. Dinoflagellates, an ecologically important protist lineage, represent a unique model to study this process because dinoflagellate plastids have repeatedly been reduced, lost, and replaced by new plastids, leading to a spectrum of ages and integration levels. Here we describe deep-transcriptomic analyses of the Antarctic Ross Sea dinoflagellate (RSD), which harbors long-term but temporary kleptoplasts stolen from haptophyte prey, and is closely related to dinoflagellates with fully integrated plastids derived from different haptophytes. In some members of this lineage, called the Kareniaceae, their tertiary haptophyte plastids have crossed a tipping point to stable integration, but RSD has not, and may therefore reveal the order of events leading up to endosymbiotic integration. We show that RSD has retained its ancestral secondary plastid and has partitioned functions between this plastid and the kleptoplast. It has also obtained genes for kleptoplast-targeted proteins via horizontal gene transfer (HGT) that are not derived from the kleptoplast lineage. Importantly, many of these HGTs are also found in the related species with fully integrated plastids, which provides direct evidence that genetic integration preceded organelle fixation. Finally, we find that expression of kleptoplast-targeted genes is unaffected by environmental parameters, unlike prey-encoded homologs, suggesting that kleptoplast-targeted HGTs have adapted to posttranscriptional regulation mechanisms of the host.

plastid endosymbiosis | kleptoplasty | dinoflagellates | plastid integration

Endosymbiosis, or the uptake and retention of one cell within another, is an important process in eukaryotic evolution, resulting in countless cell–cell interactions and increasing cellular complexity, at the extreme leading to the origins of mitochondria and plastid organelles. While we are familiar with the important outcomes of endosymbiosis, we know much less about the process itself at a mechanistic level, partly because the best-studied endosymbiotic events are so ancient that most clues as to how they came to be are now lost. This is particularly true of the mitochondria, which originated once, before the diversification of all known eukaryotes, and have since largely evolved by relatively rare functional change or reduction (1). The evolution of the plastid is less straightforward, but its complexity offers more glimpses into the process. Plastids (defined here as not including the parallel case of *Paulinella chromatophora*) (2), also originated once, from cyanobacteria, in a process described as primary endosymbiosis. Unlike mitochondria, however, plastids have since spread from one lineage to multiple others horizontally. In this process, called secondary endosymbioses, one eukaryote takes up a eukaryotic alga with a primary plastid, resulting in a secondary plastid. Algae with a secondary plastid can also be taken up yet again by another eukaryote, giving rise to tertiary plastids. The resulting complex distribution of secondary and tertiary plastids includes wide spectra of both age and degree of integration with the host. At one end are fully integrated plastids

where all other traces of the endosymbiont except the plastid are erased (3), but there are also plastids that retain varying degrees of their original complexity (4–7).

Secondary and tertiary plastids are diverse, but share one fundamental characteristic: they have all stably integrated with their host and are retained over long periods of evolutionary time. In contrast, many temporary associations have also been observed where an alga is engulfed and its plastid taken up for a period of time but ultimately digested. These are called kleptoplasts, or "stolen" plastids. Kleptoplasty is known in several lineages of protists and animals and may persist anywhere from days to months (8–11). The best-investigated examples are found in the dinoflagellate *Dinophysis* and the animal *Elysia*. *Elysia* is a sea slug that sequesters its plastids directly from its macroalgal food and stores them in cells of the digestive tract and has been the center of a long debate about the degree of integration between the plastid and animal host (12). *Dinophysis*, on the other hand, obtains its kleptoplasts through a series of transfers, by feeding on the ciliate *Mesodinium rubrum*, which in turn acquired the kleptoplast by eating cryptophytes possessing secondary plastids (13).

Significance

Kleptoplasty is the process by which a heterotrophic predator eats an algal prey cell and then steals and temporarily retains the alga's photosynthetic plastid organelle. Kleptoplasty is relatively common in nature, but also represents a key step in the early stages of integration of stable endosymbiotic organelles. We characterized a kleptoplastidic dinoflagellate at the genomic level and compared it with relatives that have fully integrated a closely related plastid, to better understand the tipping point between temporary and fully integrated plastids. We find that genetic integration of the plastid and host and host control over the plastid function occur early, before the plastid is fully fixed in the cell, allowing us to see the order of key events in plastid organelle origins.

This spectrum of endosymbiotic integration should not automatically be viewed as steps in some inevitable process where every system is heading to the same endpoint. It does, however, offer a chance to compare a variety of stages in several parallelly evolving systems. And in some circumstances even more direct links between stages can be made. These are clearest in the dinoflagellates, where plastid evolution is at its most dynamic. The archetypical dinoflagellate plastid is the "peridinin plastid," a secondary plastid derived from a red alga and named for its carotenoid pigment. This plastid is ancestral to the lineage and distinctive functionally, morphologically, and at the genomic level (14). The peridinin plastid genome is the most reduced known, since most plastid genes have moved to the nucleus and their products are targeted back to the plastid via a bipartite leader sequence consisting of a signal peptide (SP) followed by a transit peptide. This transit peptide begins with a 4-residue-long motif devoid of basic or acidic residues and is further distinguished by a transmembrane region (TMR) that mediates the transit of plastid proteins through the Golgi (15, 16). Many dinoflagellates have lost the plastid genome, or even lost the plastid itself, and several have replaced or supplemented the peridinin plastid with new secondary or tertiary plastids, or kleptoplasts.

The Kareniaceae are one such lineage, with some members harboring stable, genetically integrated tertiary plastids derived from haptophyte algae with secondary plastids (17). Although the Kareniaceae hosts are closely related, the plastids of 2 genera, *Karenia* and *Karlodinium*, have been shown to derive from 2 different haptophytes (18). In contrast to peridinin plastid-targeting sequences, basic residues are particularly conserved in the first 2 positions of transit peptides associated with such tertiary haptophyte plastids. Also the transmembrane region typical for the peridinin plastid transit peptide is absent (19). This host lineage also includes a kleptoplastidic member, the psychrophilic Antarctic Ross Sea dinoflagellate (RSD), which is a heterotrophic Kareniaceae that obtains kleptoplasts from its haptophyte prey, *Phaeocystis antarctica* (20). RSD is unable to grow in the dark, even if supplied with prey, and is therefore an obligate kleptoplastidic dinoflagellate (18). But the RSD system is also unique because it is able to survive and maintain its kleptoplasts in the absence of prey for at least 30 mo (20), which is distinctly longer than has been described for any other kleptoplast so far.

This range of plastid integration in close relatives may offer unique insights into several outstanding questions surrounding the mechanism of plastid endosymbiosis. Traditionally, the process is (sometimes vaguely) described as starting with the uptake of the endosymbiont, followed by gene transfer from the endosymbiont to the host and establishment of a protein-targeting system (3, 21), and this genetic integration is sometimes considered the hallmark distinction between an organelle and an endosymbiont. In contrast, recent models (e.g., the "shopping bag hypothesis") propose a period of serial, transient uptakes and the establishment of protein targeting during this phase, followed later by the fixation of one symbiont that ultimately becomes the organelle (3, 21, 22). These models are very different mechanistically and lead to different predictions about the origin of targeted proteins, but are nonetheless difficult to differentiate in extremely old organelles. The plastids of Kareniaceae, however, have a relatively recent origin, well-defined source, and variety of levels of integration, all characteristics that aid in differentiating between broadly defined "targeting-early" and "targeting-late" models for endosymbiotic organelle origins. To take advantage of these characteristics, we characterized deep-transcriptomic datasets from RSD and its prey harvested under various light and temperature conditions, and compared RSD with other Kareniaceae. Overall, we find that RSD is genetically complex, with plastid functions likely partitioned between the kleptoplast and a relict peridinin plastid. We also find

many genes for kleptoplast-targeted proteins in RSD from various sources, and most importantly, find that many of these genes are also in other Kareniaceae with integrated plastids. The shared possession of these genes demonstrates that the ancestor of Kareniaceae had already developed a system to acquire and target plastid genes, and by extension that genetic integration preceded fixation of the actual organelle. This provides the most direct evidence to date in support of the emerging new models for organellogenesis (3, 21, 22).

Results and Discussion

RSD Encodes Genes for Proteins Targeted 2 Functionally Distinct Plastids. The tertiary plastids of Kareniaceae have long been known to rely on numerous nucleus-encoded plastid-targeted proteins (19, 23), but RSD is not closely related to either of the other well-studied Kareniaceae genera (*SI Appendix*, Fig. S1) (24), so whether the RSD kleptoplast is dependent on the nucleus is unknown. To determine this, we searched for genes encoding enzymes from plastid-associated pathways in RSD, *Karenia brevis* and *Karlodinium micrum* (the only Kareniaceae with transcriptomes), and *Dinophysis acuminata* (which has cryptophyte kleptoplasts). Only transcriptome data are available for the lineages in question, and indeed for nearly all dinoflagellates, because dinoflagellate genomes are notoriously large, ranging from 1.5 Gb to over 200 Gb (25). No exact estimates are available for RSD, but the genome size for its closest analyzed relative *K. brevis* was estimated as 56 Gb (26). The pathways and processes we analyzed were chosen because they are directly associated with the photosynthetic function of the plastid (the photosynthetic light reactions, including light-harvesting proteins), or other known plastid metabolic pathways (isoprenoid, heme, iron-sulfur cluster, and fatty acid biosynthesis pathways). For each gene identified, we evaluated 3 questions: 1) The phylogenetic origin of the gene—and particularly whether each gene is specifically related to dinoflagellate or haptophyte homologs; 2) in which genome each gene is encoded—whether specific features of the gene show it is encoded in the genome of the dinoflagellate or the haptophyte; and 3) the probable cellular location of the proteins encoded by those genes—to see if N-terminal sequences suggest targeting to the kleptoplast or any other location within the dinoflagellate. Accordingly, we reconstructed single-gene phylogenies for each gene, analyzed the translated peptides for distinct N-terminal plastid-targeting sequences, and investigated the corresponding transcripts for spliced leader sequences. Spliced leader sequences are transspliced to dinoflagellate transcripts during transcription, but no such system exists in haptophytes, so a spliced leader is strong evidence that a gene is encoded in the dinoflagellate genome (27).

Following this approach, we unexpectedly found that RSD likely retains a relict peridinin plastid. Kareniaceae are thought to have lost the peridinin plastid entirely, but we identified almost complete sets of enzymes for the isoprenoid, heme, and iron-sulfur cluster biosynthesis pathways in RSD, and the proteins are closely related to dinoflagellate homologs (Fig. 1, see also "Metabolic Pathways Phylogenies" in the figshare repository at 10.6084/m9.figshare.7851467 (28)). The single exception is *ispF*, which we did not identify, but is also absent from many dinoflagellate transcriptomes, suggesting a low level of expression. Furthermore, these genes encode N-terminal extensions with typical peridinin plastid-targeting signals (Fig. 1). To unequivocally confirm the presence of a cryptic plastid, direct experimental evidence like protein localization is necessary; however, the presence of 3 nearly complete pathways with proteins consistently predicted to be targeted to the ancestral peridinin plastid strongly suggests the retention of this plastid in RSD. Nonphotosynthetic peridinin plastids have been described from many other dinoflagellates (29), and it has been proposed that the isoprenoid biosynthetic function makes them difficult to lose, since the analogous host cytosolic
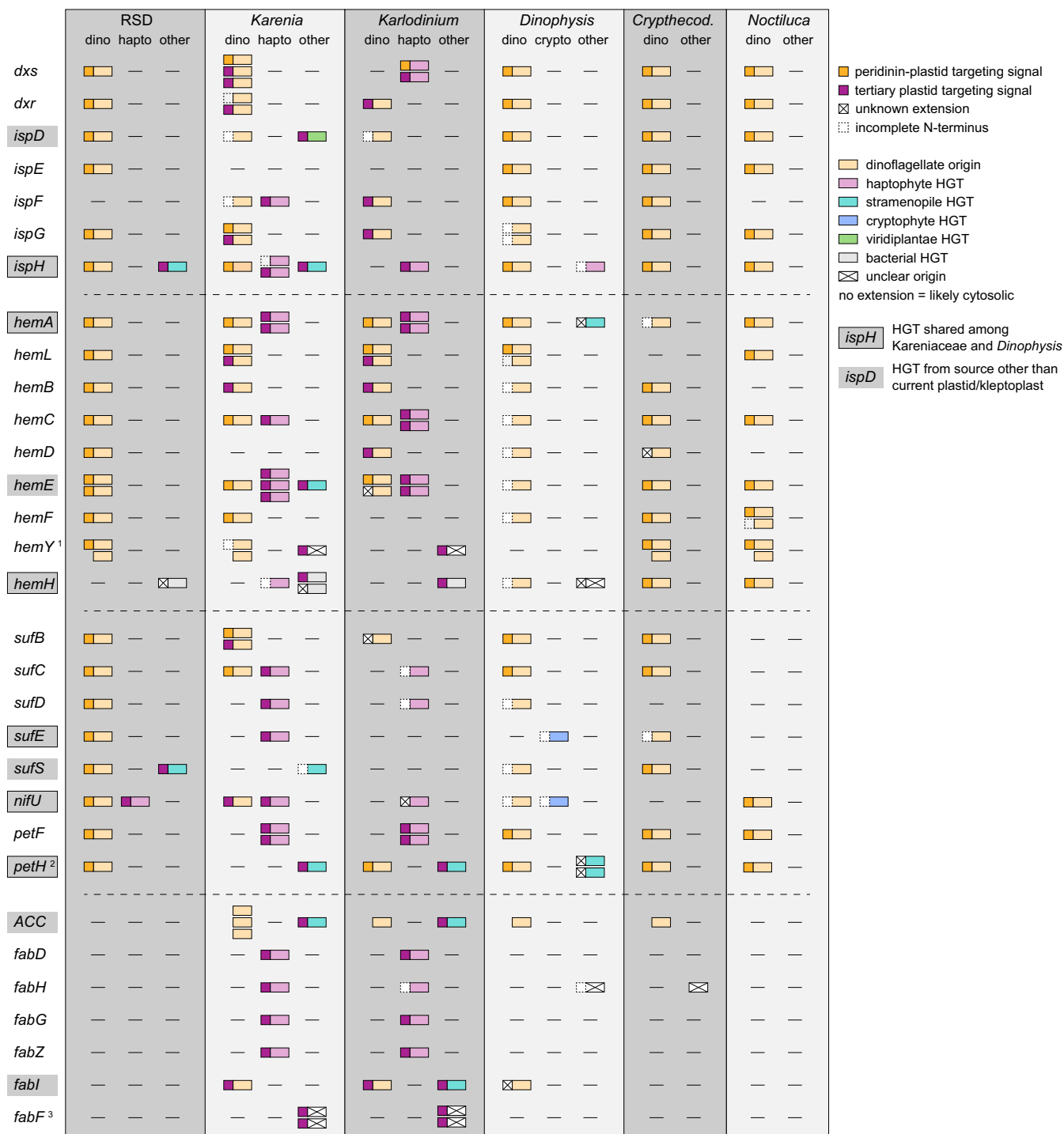
Fig. 1. Schematic depiction of evolutionary origin and N-terminal targeting information of (from top to bottom) plastidial isoprenoid, heme, iron-sulfur cluster, and fatty acid biosynthesis enzymes in RSD, dinoflagellates with tertiary haptophyte plastids (*K. brevis* and *K. micrum*) and another dinoflagellate with kleptoplasts (*D. acuminata*), in comparison with 2 dinoflagellates with nonphotosynthetic plastids (*Noctiluca scintillans* and *Crypthecodinium cohnii*). Abbreviations beneath the dinoflagellate genus name indicate the origin of the gene (dino, dinoflagellate; hapto, haptophyte; crypto, cryptophyte or other). Shaded gene names indicate HGTs from sources other than the current plastid or kleptoplast. Boxed shaded gene names correspond to HGTs shared among Kareniaceae and *Dinophysis*. Mature proteins and N-terminal extensions are depicted with color-coded boxes (see legend). N-terminal extensions marked with a cross indicate either the presence of a SP and/or TMR, but the extensions were not similar to any described targeting signal in dinoflagellates, or no SP and/or TMR was predicted. [1]Dinoflagellates retain the cytosolic as well as the plastidial form of *hemY*. The Kareniaceae group within a clade consisting of haptophytes and stramenopiles, this clade is supported by unique sequence characteristics. [2]In the *petH* phylogeny, at least 5 isoforms were recovered for *Karenia*. [3]The Kareniaceae *fabF* cluster as only dinoflagellate representatives with mitochondrial isoforms in the phylogeny, but carry tertiary plastid-targeting sequences, suggesting retargeting to the plastid. *Crypthecod.*, *Crypthecodinium*.

pathway was lost in the ancestor of dinoflagellates and apicomplexans (30). The presence of a kleptoplast might make this function appear redundant, but we speculate that its transient nature means selection for the ancestral plastid is maintained.

In contrast to RSD, *Karenia* and *Karlodinium* express multiple isoforms of several proteins in these pathways, particularly so for isoprenoid and heme biosynthesis, sometimes with up to 3 different phylogenetic origins (Fig. 1). Some genes are of dinoflagellate origin and encode the corresponding peridinin plastid-targeting peptides. But others are of haptophyte origin, indicating endosymbiotic gene transfer (EGT) from a haptophyte plastid host, and these encode the distinct targeting peptide associated with tertiary haptophyte plastids. Others still are of dinoflagellate origin, but encode a tertiary haptophyte plastid-targeting peptide, and in 2 cases haptophyte-derived genes encode peptides that would appear to target the protein to the ancestral dinoflagellate plastid (Fig. 1).

These observations suggest the concurrent presence of both plastid types at some point during the evolution of the haptophyte plastids of *Karenia* and *Karlodinium*, as suggested previously (19), but also that the peridinin plastid might still persist. The haptophyte plastid is clearly functionally dominant, with almost all steps in the heme, isoprenoid, and iron-sulfur cluster biosynthesis pathways having at least 1 haptophyte-plastid targeted representative. However, many seemingly peridinin plastid-targeted forms also exist, suggesting either a very recent loss of the peridinin plastid or its persistence. Finally, the phylogenetic origins (dinoflagellate or haptophyte) of individual genes are similar in *Karenia* and *Karlodinium*; however, there are some clear differences (e.g., *dxs* and *hemE*, Fig. 1 and "Metabolic Pathways Phylogenies" in the figshare repository at 10.6084/m9.figshare.7851467 (28)) that possibly correlate with the independent uptake of 2 different haptophyte plastids in those 2 genera, as has already been suggested by plastid SSU rRNA phylogenies and the divergent evolution of the plastid genomes in those 2 lineages (18, 23).

RSD Kleptoplast-Targeted Proteins Are Derived from a Different Haptophyte than the One Providing the Kleptoplast.
One of the predictions of models where organelle-targeting precedes the fixation of the organelle itself is that targeted proteins may be derived from lineages other than the organelle. In the case of RSD, such genes might be retained from the ancestral plastid, or acquired by horizontal gene transfer (HGT) from previous kleptoplastidic prey, or other sources. Since kleptoplasts are so transiently retained, even genes that appear to be related to the current prey lineage cannot unequivocally be attributed to the exact compartment to which they are targeted, so we will refer to these as HGTs rather than EGTs, which specifically implies the gene comes from the endosymbiont.

To identify candidate genes, we followed 3 approaches. First, we reconstructed phylogenies of all spliced leader-encoding (i.e., host) transcripts to identify genes in the host genome, but derived from haptophytes (i.e., food). Second, we specifically investigated the plastid-associated pathways listed above for haptophyte-derived genes. And third, we reconstructed phylogenies for all haptophyte-derived genes already known from *Karenia*/*Karlodinium* to see if any predate the divergence of these hosts. To identify gene transfers with high confidence, we curated a comprehensive database of eukaryotes and prokaryotes and refined all initial phylogenies with recently developed sequence filtering and alignment methods (31–33). All analyses included RSD transcriptomes that were either cleaned of prey sequences using a *P. antarctica* transcriptome ("RSDallclean noPhaeo"), or without cleaning ("RSD Temp01"). The latter version allows us to detect very recent gene transfers from the prey that otherwise might have been removed in the cleaned version due to very high similarity. To distinguish such transfers from a possible contamination, we investigated candidates of interest for

an N-terminal extension indicating targeting to a plastid and/or the presence of a spliced leader.

This resulted in more than 1,500 single-gene trees, which were manually searched for RSD-encoded transcripts that fall within the haptophytes. We identified 13 candidates that we conclude are encoded in the dinoflagellate genome (based on the presence of a spliced leader and/or the absence of a corresponding transcript in the prey transcriptome) but cluster with haptophyte homologs in phylogenies. These will be termed "haptophyte HGTs" throughout this study (Fig. 2). Importantly, these candidates are, with 1 possible exception, not closely related to *Phaeocystis* (*SI Appendix*, Fig. S2 *A–M*) and are therefore unlikely to have originated from the same lineage as the kleptoplast itself. It is more likely that these genes have been transferred from other haptophytes previously associated with the host (e.g., preyed upon). It is impossible to know if those haptophytes were stable endosymbionts, transient food, or even kleptoplasts. Four of the gene transfer cases play a role in photosynthesis, another
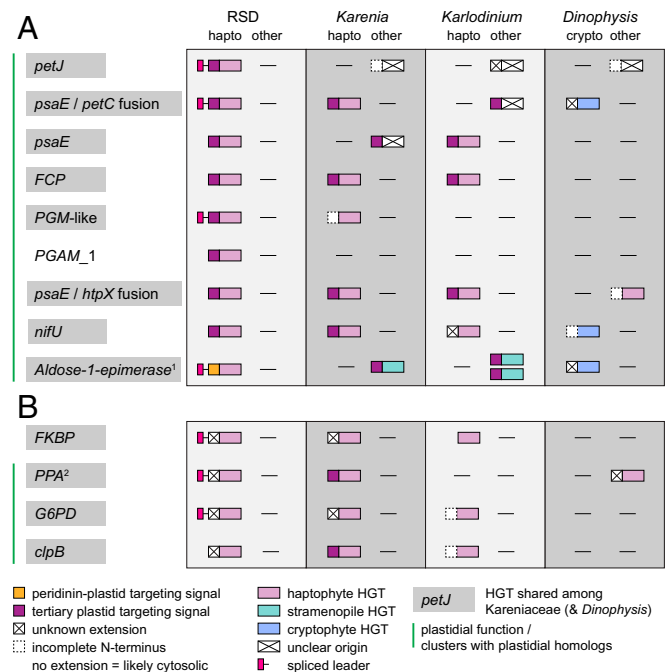
Fig. 2. Schematic depiction of evolutionary origin and N-terminal targeting information of the 13 haptophyte HGTs in RSD that are (*A*) putatively targeted to the kleptoplast or (*B*) carry extensions that could not be identified as indicating targeting to either plastid and the corresponding homologs in dinoflagellates with tertiary haptophyte plastids (*K. brevis* and *K. micrum*) and another dinoflagellate with kleptoplasts (*D. acuminata*). (*A*) HGTs putatively targeted to the kleptoplast (based on the presence of a signal peptide, at least 1 basic amino acid residue in the first 2 positions of the transit peptide and the absence of a transmembrane domain) or the peridinin plastid (based on the presence of 2 TMRs, the second one followed by basic residues). (*B*) HGTs with N-terminal extensions that could not be identified as indicating targeting to either plastid. See also Dataset S1. Shaded gene names correspond to HGTs shared among Kareniaceae and, in some cases, *Dinophysis*. Mature proteins and N-terminal extensions are depicted with color-coded boxes (see legend). N-terminal extensions marked with a cross indicate either the presence of a SP and/or TMR, but the extensions were not similar to any described targeting signal in dinoflagellates, or no SP and/or TMR was predicted. A green vertical line represents HGTs with plastidial function and/or clustering with plastidial homologs in their respective phylogenies. Spliced leader information is only depicted for RSD. [1]The 2 stramenopile HGTs in *Karlodinium* have separate origins in different groups of stramenopiles. [2]The *Karenia*/*Karlodinium* and *Dinophysis* HGTs for *PPA* are in a plastidial clade paralogous to the clade containing the RSD HGT.

EVOLUTION

4 in carbohydrate metabolism, 3 encode chaperones, 1 is an iron-sulfur cluster assembly protein, and the last one is an inorganic pyrophosphatase (Dataset S1). Interestingly, several of haptophyte HGT genes exist as numerous isoforms, and some are distinguished by unusually long N-terminal extensions to the mature protein (relative to haptophyte homologs: *SI Appendix*, Fig. S3). Most of these extensions are similar to the above-described targeting peptides in tertiary haptophyte plastid-targeted proteins in the other Kareniaceae, also showing conservation of basic residues in the beginning of, and lacking the transmembrane region within the transit peptide (Dataset S1). Such extensions are present on 8 of the 13 haptophyte HGT candidates (Fig. 2), suggesting that at least those candidates may be targeted to the kleptoplast. The 2 fusions, *psaE-petC* and *psaE-htpX* (Fig. 2), are specific for RSD and were recovered in both independently assembled transcriptomes (Dataset S1), suggesting that they are not misassembled.

The electron transport protein, cytochrome c6 (*petJ*), represents the only case where the gene may have been transferred from a close relative of the current kleptoplast (*SI Appendix*, Fig. S2A). In the corresponding alignment we recovered an isoform of this gene (RSD_Temp01@32625) that, in contrast to the other RSD-encoded *petJ* isoforms, is identical to homologs from *P. antarctica* over the length of the predicted cytochrome *c* domain, suggesting a very recent HGT. However, the transcript includes a spliced leader, confirming it is encoded in the host, and its N-terminal extension is different from the extension of the *P. antarctica* homolog, suggesting functional changes to targeting, as would be expected (*SI Appendix*, Fig. S3A, isoform 2 and Dataset S1). The remaining isoforms are distinct at the nucleotide level, but also cluster with the *Phaeocystis* clade, and may therefore represent transfers from earlier stages of the current association.

In a few cases, we cannot distinguish recent transfer from prey contamination because the presence of spliced leaders cannot be confirmed. For example, the plastidial Clp protease subunit, *clpB*, clusters with *Phaeocystis* but contains long insertions absent from *Phaeocystis*. However, it lacks a spliced leader, either because the transcript is truncated, because spliced leaders are not universal to all dinoflagellate transcripts (34), or possibly because it is not actually encoded in RSD.

One haptophyte HGT, aldose-1-epimerase, clusters with plastid-targeted homologs in the tree, but appears to encode a peridinin plastid-targeting peptide (*SI Appendix*, Figs. S2I and S3A). This suggests a possible role for this enzyme in the relict peridinin plastid, which is not consistent with the apparent functions of the organelle. The remaining 3 haptophyte HGTs, peptidyl-prolyl cis-trans isomerase chaperone (*FKBP*), soluble inorganic pyrophosphatase (*PPA*), and glucose-6-phosphate 1-dehydrogenase (*G6PD*), also all encode extensions, but not clearly matching the characteristics of a plastid-targeting peptide. However, *G6PD* clusters with plastid-targeted homologs in the respective phylogeny (*SI Appendix*, Fig. S2L), the *PPA* tree contains plastid-targeted paralogs (*SI Appendix*, Fig. S2K), while the *FKBP* extension is predicted to encode a signal peptide (*SI Appendix*, Fig. S3B), suggesting putative plastid function for all but one of the identified HGTs (Fig. 2). Some genes were present as multiple isoforms (*SI Appendix*, Fig. S3B), suggesting that more than one HGT can be integrated and targeted to the kleptoplast, or that a single HGT was followed by duplication. In the case of *FKBP* (*SI Appendix*, Fig. S2J) and *G6PD*, repeated transfer from different haptophytes seems more likely, since RSD isoforms are related to different haptophytes.

Most importantly, homologs for nearly all of the haptophyte HGTs found in RSD are also found in *Karenia* and/or *Karlodinium*, where they are predicted to function in the tertiary haptophyte plastid. The exceptions are *PGAM_1* in both of the other Kareniaceae, *PPA* and *PGM*-like in *Karlodinium*, and the *petC* in *Karlodinium*; and the origin of the *Karlodinium petC* cannot be unambiguously resolved (Fig. 2). The presence of homologous proteins targeted to both the RSD kleptoplast and the fully integrated but independently derived plastids of *Karenia* and *Karlodinium* suggest very strongly that the plastid-targeted proteins and targeting systems must both have predated the fixation of the organelles.

A Complex History of HGT in Tertiary Plastids and Kleptoplasts. In addition to the horizontal gene transfer candidates identified by their similarity to haptophyte homologs described above, we also identified several more through searches for any gene relevant to specific plastid pathways, including heme, isoprenoid, and iron-sulfur cluster biosynthesis, as well as nuclear-encoded photosystem subunit genes (Fig. 1 and *SI Appendix*, Table S2, "Metabolic Pathways Phylogenies" and "Photosynthesis Phylogenies" in the figshare repository at 10.6084/m9.figshare.7851467 (28)). RSD encodes 2 genes from the unrelated stramenopiles (*ispH* and *sufS*), both of which encode N-terminal extensions consistent with targeting to the kleptoplast. *Karenia* encodes a previously reported gene transfer from the unrelated Viridiplantae (35), and several stramenopile genes. *Karlodinium* also encodes several genes from stramenopiles, and *Dinophysis* encodes several genes from stramenopiles and haptophytes. Additionally, genes of bacterial origin were found in RSD, *Karenia*, *Karlodinium*, and/or *Dinophysis*. In the *hemH* phylogeny, a clade consisting of RSD, *Karenia*, and *Dinophysis* branches with Firmicutes, distinct from a bacterial HGT previously identified in *Chromera*/*Vitrella* and Apicomplexans ("Metabolic Pathways Phylogenies" in the figshare repository at 10.6084/m9.figshare.7851467 (28). The RSD and *Dinophysis* genes encode N-terminal extensions (in contrast to their homologs encoded in bacteria) of unknown function, while the *Karenia* sequence is probably incomplete. A second bacterial *hemH*, was identified in *Karenia* and *Karlodinium*, which clusters with the cyanobacterium *Synechococcus* and, unlike the cyanobacterium, encodes bipartite N-terminal extensions putatively targeting them to the haptophyte plastid.

Interestingly, *Dinophysis* was also found to share a handful of haptophyte-derived genes with Kareniaceae (Figs. 1 and 2), some with N-terminal extensions, which, together with their predicted function, suggest a kleptoplast localization. What this means is unclear, since *Dinophysis* and Kareniaceae are not considered to be closely related.

The taxonomic diversity of sources for HGT suggests that Kareniaceae (and *Dinophysis*) have previously interacted with numerous transient endosymbionts and/or food, some of which left longer-term footprints in the form of HGT. Indeed, *Dinophysis* isolates have been reported to contain plastids from not just cryptophytes but also stramenopiles and chlorophytes (36). *Karenia*, *Karlodinium*, and *Takayama* currently possess fully integrated plastids acquired from a haptophyte, but are nevertheless known to ingest other food: *K. brevis* can feed on cyanobacteria (37), *Karlodinium armiger* is an omnivorous feeder taking up on many different types of prey (38, 39), *K. micrum* can feed on cryptophytes (40), and *Takayama helix* even feeds on a wide range of other dinoflagellates (41). The possibility that they harbored a variety of kleptoplasts during their evolutionary history is likely and consistent with the presence of the diverse shared HGTs. As with the haptophyte HGTs, the presence of additional nonhaptophyte HGTs provides further evidence most compatible with the protein-targeting system predating the current organelle.

Overall, the data are increasingly consistent with models for plastid origins where the first stage has the host taking up and retaining prey in a series of transient relationships (3, 10, 21). Over time, this is hypothesized to lead to the establishment of a protein-targeting system, which allows for the accumulation of genes for plastid-targeted proteins that are retained and retargeted to subsequent transient kleptoplasts, and ultimately targeted to the endosymbiont that is finally fixed to become the

organelle. This series of events could conceivably explain the origin of other endosymbiotic organelles, but there is no reason to assume that all organelles must have originated by exactly the same process as one another, so each case should be considered independently. In the case of mitochondria, for example, whether the host had already evolved phagocytosis is currently debated (42), and this is a requirement for this model. Interestingly, however, similar models are emerging from observations of diverse endosymbiotic associations not generally classified as organelles, like those between bacteria and insects (43), suggesting this process might be relevant beyond secondary or tertiary plastids and may be one plausible explanation for endosymbiotic integration in other systems more generally.

RSD Encodes Genes Mediating Cyclic Electron Transport in the Kleptoplast. The suite of genes for kleptoplast-targeted proteins raises the question: Does the function of the organelle change when it is taken over by its new host? The RSD nucleus encodes kleptoplast-targeted subunits of photosystem I (PSI; i.e., *psaE*) and the cytochrome b_6/f complex (i.e., *petC*), as well as the electron transport protein *petJ* (cytochrome c6). All of these are also nucleus encoded in the prey, *P. antarctica*, and intriguingly *psaE* and *petC* are the only subunits of their respective complexes that are nucleus encoded in *Phaeocystis* (*SI Appendix*, Table S2). Moreover, *petJ* provides a link between the 2 complexes by mediating electron transfer from cytochrome b_6/f to PSI. In

contrast, however, we did not find a single subunit of photosystem II (PSII) in the RSD nucleus, whereas *P. antarctica* encodes several (*SI Appendix*, Table S2). This correlates with measurements showing a severely diminished PSII activity in the kleptoplast compared to *P. antarctica* plastids (44).

The lack of PSII activity and protein-targeting together suggest an absence of canonical linear flow through the photosynthetic electron transport chain, and that the kleptoplast performs cyclic electron transport instead (Fig. 3 *A* and *B*). Cyclic electron transport requires only PSI and the cytochrome b_6/f complex to drive ATP synthesis without the production of NADPH. Cyclic electron flow is less efficient, so why would the kleptoplast lose PSII function? One interesting possibility is simply because PSII function requires more HGT. The host can sustain PSI function after only 2 HGT events, but to support PSII function requires both of these plus another 7 independent HGT events. There is a clear tradeoff between the relative likelihood of acquiring the few genes necessary to establish cyclic electron flow and the greater benefit but lower likelihood of acquiring many more genes to maintain linear electron flow.

However, for cyclic electron flow to work, it requires 1 additional step. In angiosperms it can take place through either of 2 pathways: 1 dependent on Proton Gradient Regulation 5 (PGR5) and PGR5-like protein 1 (PGRL1), and the other on a type I NADH dehydrogenase complex (NDH-1) (45), which may have been substituted by a plastidial type II NADH single-subunit
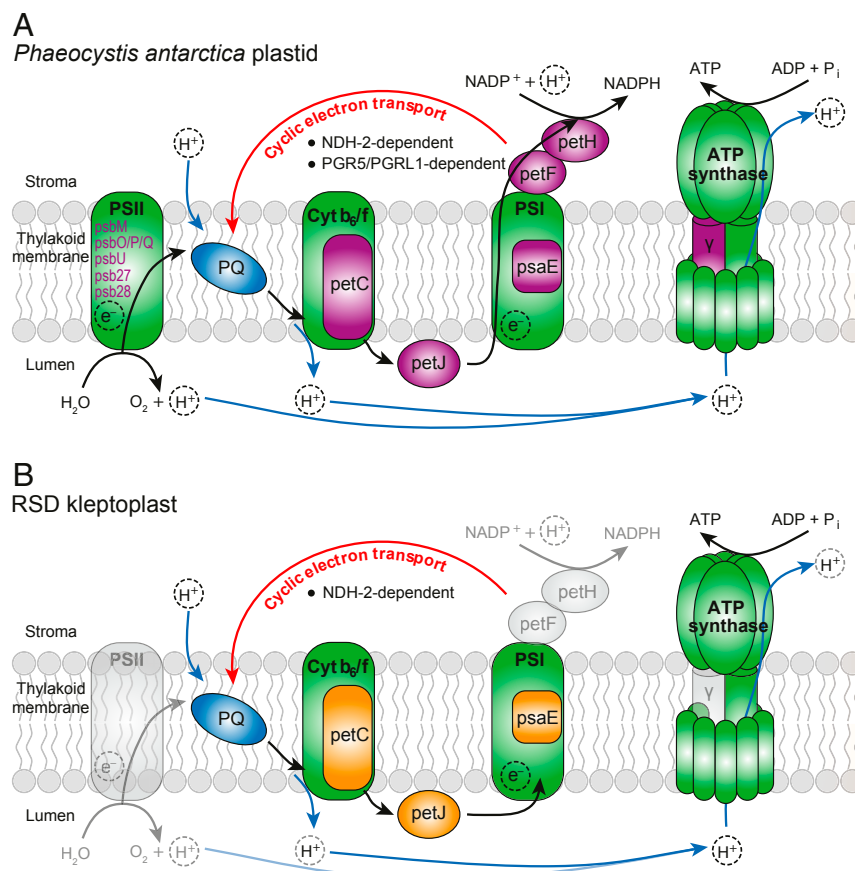
Fig. 3. Scheme of the photosynthetic electron transport chain in (*A*) the plastid of *P. antarctica* and (*B*) the kleptoplast of RSD, adapted from ref. 51. Green fill indicates plastid-encoded proteins; purple fill/writing, haptophyte-nucleus encoded proteins; orange fill, dinoflagellate-nucleus encoded proteins; and blue fill, plastoquinone. Black arrows indicate electron flow; the red arrow shows cyclic electron flow; and the blue arrows indicate proton movement. Transparent features indicate absence of the corresponding proteins/reactions in RSD. γ, ATP synthase subunit gamma; Cyt b_6/f, cytochrome b_6/f complex; petF, ferredoxin; petH, ferredoxin-NADP reductase; petC, cytochrome b_6/f complex iron-sulfur subunit; petJ, cytochrome c6; PQ, plastoquinone; PSI/II, photosystem I/II; psaE, photosystem I subunit IV.

enzyme (NDH-2) in some plants and microalgae (46). We found both *PGR5* and *PGRL1* transcripts as well as a homolog for *NDH-2* in the *P. antarctica* prey, indicating it can potentially perform both pathways, in addition to linear electron flow (Fig. 3*A*). Dinoflagellates, on the other hand, contain homologs for *PGR5* and *PGRL1*, including a *PGRL1–PGR5* fusion in *Karenia*, *Karlodinium*, and *Dinophysis*, but neither was identified in RSD transcriptomes (*SI Appendix*, Fig. S4 *A* and *B*), either because they were not sampled or are not present. NDH-2 is less well investigated, but we reconstructed an NDH-2 phylogeny and recovered 2 dinoflagellate plastid clades (*SI Appendix*, Fig. S4*C*). Neither homologs of RSD nor of the other Kareniaceae are represented in either clade, but interestingly genes from both form a separate *NDH-2* clade, which branches within a larger group of plastid-targeted proteins, including the experimentally localized protein from *Chlamydomonas*, Nda2 (*SI Appendix*, Fig. S4*C*) (47). The RSD homologs encode an N-terminal extension, but not one that is clearly a targeting peptide. Many other algal proteins in this extended clade, such as the peridinin-plastid dinoflagellates, however, do encode plastid-targeting leaders. This would suggest another putative HGT event of functional significance, although the plastid localization clearly requires more direct evidence. If this plastid-targeted NDH-2 is kleptoplast targeted as well, a preliminary model for photosynthetic electron flow in the RSD kleptoplast (Fig. 4*B*) would use NDH-2 to recycle electrons from PSI to plastoquinone with either NADH or NADPH as electron donors, like their *Chlamydomonas* counterpart (47). We also failed to identify a transcript for the kleptoplastidic ATP synthase subunit gamma (*atpC*) in RSD, which is the only nucleus-encoded ATP synthase subunit in *P. antarctica* (*SI Appendix*, Table S2). It is therefore unclear whether the ATP synthase complex in the kleptoplast is fully functional, but given photosynthesis is functional, some form of the complex must be operating.

Kleptoplast-Targeted Genes Have Lost Transcriptional Regulation. The presence of kleptoplast-targeted proteins encoded in the RSD genome means the host may be able to regulate its kleptoplasts. As a first step to investigate this, we analyzed expression levels of all kleptoplast-targeted genes with haptophyte origin in RSD

under different light (light vs. deep chlorophyll maximum [DCM] vs. dark) and temperature (0° vs. 5°) conditions (*SI Appendix, SI Materials and Methods*). The presence of about 1% prey *P. antarctica* in the RSD cultures during RNA extraction for transcriptome sequencing prevented us from testing the response of the genes encoded within the kleptoplast genome itself. Of the 13 transcripts for kleptoplast-targeted proteins, only 1, *G6PD*, showed increased expression in light vs. dark conditions under our parameters (minimum 2-fold change, false discovery rate [FDR] < 0.05). However, the N-terminal extension, and therefore the location of this protein, could not be unambiguously identified. All other transcripts were unaffected.

We compared this response with the same genes in *P. antarctica* cultures exposed to the same environmental conditions. In contrast to RSD, 7 out of 13 of the *P. antarctica* transcripts displayed at least a 2-fold change in expression under different light and temperature conditions (Fig. 4). Transcripts significantly up-regulated in the light (and DCM) vs. dark were mostly involved in photosynthesis (*petJ*, *petC*, *FCP*, plus *PPA*) whereas transcripts for carbohydrate metabolism genes *PGAM_1* and *G6PD* increased in the dark. Transcripts for *petJ*, *petC*, and *PPA* were also up-regulated at 0° vs. 5°. Two isoforms of the chaperone *clpB* were up-regulated at the higher temperature and 1 of them was also up-regulated in the dark. *petJ* represents a case where different isoforms of the same gene are expressed under different conditions, suggesting putatively different functions for those isoforms correlating with environmental parameters.

We also analyzed all other transcripts discussed in this study (i.e., transcripts encoding the intermediates of the heme, isoprenoid, and iron-sulfur cluster biosynthesis pathways, and which are likely functioning in a cryptic peridinin plastid) for differential expression in RSD, but observed no significant difference in any. It has been reported that dinoflagellates display only small changes in mRNA expression even under significantly different conditions, suggesting that posttranscriptional or posttranslational mechanisms play a dominant role in the regulation of dinoflagellate expression (48–50). Horizontal transfer of genes from a haptophyte to the RSD nuclear genome subjected them to the dinoflagellate-specific mechanisms of gene regulation,
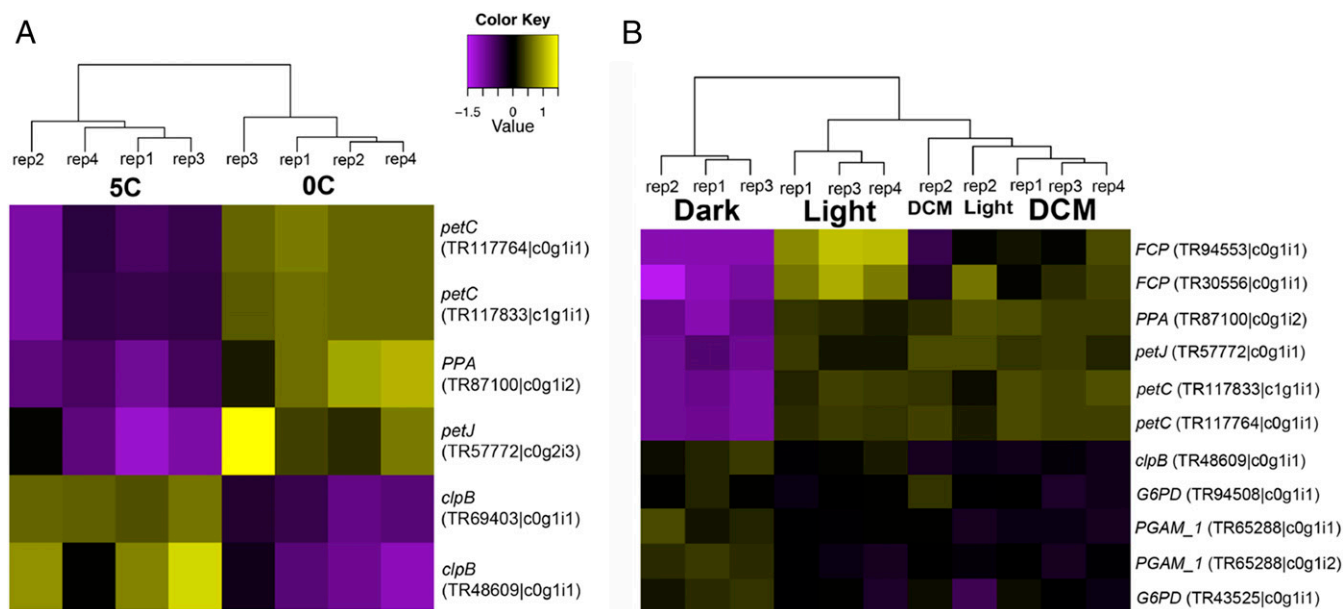

Fig. 4. Heatmaps depicting differential expression of HGT transcripts in *Phaeocystis* in response to (*A*) temperature and (*B*) light at a level of 2-fold. Transcript identifiers correspond to the "*Phaeocystis antarctica* Temp01" predicted peptides present in the HGT phylogenies shown in *SI Appendix*, Fig. S2 *A–M* as listed in *SI Appendix*, Table S3.

which at least for the plastid-targeted genes investigated here appears to have resulted in a loss of transcriptional regulation.

Conclusions

The kleptoplastidic Ross Sea dinoflagellate is both more complex and highly chimeric than was previously appreciated. In addition to the haptophyte kleptoplast, it also appears to retain a cryptic peridinin plastid and has partitioned normal plastid functions between them; the kleptoplast carries out functions directly related to photosynthesis, whereas the peridinin plastid carries out a range of nonphotosynthetic plastid functions found in other dinoflagellate plastids. But the photosynthetic function of the kleptoplast also changes when it is stolen from its *Phaeocystis* host: most obviously in the loss of linear electron flow, and in coming under the influence of its new host due to host-encoded genes for proteins targeted to the kleptoplast (including proteins required to maintain photosystem I).

This presence of protein targeting to the kleptoplast is particularly significant. The possible existence of such genes has been a major point of debate in other kleptoplast systems, such as *Elysia* (12, 51, 52), but in RSD they illuminate not only the evolution of the kleptoplast, but also the fully integrated tertiary plastids in its close relatives. Specifically, the ancestor of the kleptoplastidic RSD and its relatives with fully integrated plastids already seems to have acquired several genes for plastid/kleptoplast-targeted proteins. This shows that the establishment of a protein-targeting system preceded the stable fixation of a plastid organelle. Some of these shared genes are haptophyte derived, but they are not obviously from either the kleptoplast or plastid lineages, and many are clearly derived from other kinds of algae, which supports the emerging "shopping bag" model of organelle origins (3, 21). This is also consistent with data emerging from various endosymbiotic systems that proteins targeted to endosymbionts or nascent organelles in the very earliest stages of integration are not generally derived from that endosymbiont (3, 21, 43, 53). An alternative explanation is that the endosymbiont had a chimeric genome resulting from multiple serial endosymbiotic events (54, 55), but the number of different donor lineages for HGTs, the fact that some genes have been transferred from multiple donor lineages (e.g., *SI Appendix*, Fig. S1*I*), and the general lack of such a condition in algae most closely related to the ancestors of the endosymbionts all argue against this view. Altogether, this system provides a unique look at close relatives on either side of an important tipping point in the transient-to-fully integrated transition of an organelle.

Materials and Methods

The RSD and *P. antarctica* were isolated from the Ross Sea, Antarctica, and grown under 2 different temperatures and 3 different light conditions for differential expression analysis. Culturing, detailed experimental setup, RNA extraction, sequencing, transcriptome assembly, and expression analysis are described in *SI Appendix, SI Materials and Methods*. The small subunit ribosomal RNA (SSU rRNA) phylogeny was generated from SILVA Ref NR sequences for Kareniaceae and SSU sequences from selected cultured taxa from the Marine Microbial Eukaryote Transcriptome project (MMETSP) and NCBI (*SI Appendix, SI Materials and Methods*). Sequences were aligned, ambiguously aligned sites were removed, and trees were reconstructed as described in *SI Appendix, SI Materials and Methods*. Phylogenies for plastid-associated proteins and described endosymbiotic gene transfer candidates in *Karenia* and *Karlodinium* were reconstructed using RSD homologs for those proteins as BLASTP queries against a custom database (see *SI Appendix, SI Materials and Methods* for details). BLASTP hits were parsed, aligned, ambiguously aligned sites were removed, and initial trees were reconstructed as described in *SI Appendix, SI Materials and Methods*. After removal of contaminations and other problematic sequences, cleaned sequences were prefiltered, aligned, ambiguously aligned sites were removed, and final phylogenies were calculated (see *SI Appendix, SI Materials and Methods* for details). RSD transcripts with spliced leader sequences were identified and filtered as described in *SI Appendix, SI Materials and Methods*, and initial and final phylogenies for proteins encoded by transcripts with spliced leader sequences were reconstructed as described above. Phylogenies for cyclic electron transport protein phylogenies are described in *SI Appendix, SI Materials and Methods*; trees were reconstructed as above. Prediction of signal peptides and N-terminal transmembrane domains and identification of domains and their coordinates for proteins of interest were performed as described in *SI Appendix, SI Materials and Methods*.

1. J. M. Archibald, Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* **25**, R911–R921 (2015).
2. E. C. Nowack, M. Melkonian, G. Glöckner, Chromatophore genome sequence of Paulinella sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* **18**, 410–418 (2008).
3. P. J. Keeling, The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64**, 583–607 (2013).
4. S. E. Douglas, C. A. Murphy, D. F. Spencer, M. W. Gray, Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**, 148–151 (1991).
5. G. I. McFadden, P. R. Gilson, C. J. Hofmann, G. J. Adcock, U. G. Maier, Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3690–3694 (1994).
6. B. Imanian, J.-F. Pombert, R. G. Dorrell, F. Burki, P. J. Keeling, Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. *PLoS One* **7**, e43763 (2012).
7. E. Hehenberger, F. Burki, M. Kolisko, P. J. Keeling, Functional relationship between a dinoflagellate host and its diatom endosymbiont. *Mol. Biol. Evol.* **33**, 2376–2390 (2016).
8. M. G. Park, J. S. Park, M. Kim, W. Yih, Plastid dynamics during survival of Dinophysis caudata without its ciliate prey. *J. Phycol.* **44**, 1154–1163 (2008).
9. M. D. Johnson, The acquisition of phototrophy: Adaptive strategies of hosting endosymbionts and organelles. *Photosynth. Res.* **107**, 117–132 (2011).
10. R. G. Dorrell, C. J. Howe, What makes a chloroplast? Reconstructing the establishment of photosynthetic symbioses. *J. Cell Sci.* **125**, 1865–1875 (2012).
11. R. Onuma, T. Horiguchi, Morphological transition in kleptochloroplasts after ingestion in the dinoflagellates Amphidinium poecilochrum and Gymnodinium aeruginosum (Dinophyceae). *Protist* **164**, 622–642 (2013).
12. C. Rauch et al., Why it is time to look beyond algal genes in photosynthetic slugs. *Genome Biol. Evol.* **7**, 2602–2607 (2015).
13. M. Kim, S. W. Nam, W. Shin, D. W. Coats, M. G. Park, Dinophysis caudata (Dinophyceae) sequesters and retains plastids from the mixotrophic ciliate prey Mesodinium rubrum. *J. Phycol.* **48**, 569–579 (2012).
14. C. J. Howe, R. E. Nisbet, A. C. Barbrook, The remarkable chloroplast genome of dinoflagellates. *J. Exp. Bot.* **59**, 1035–1045 (2008).
15. N. Nassoury, M. Cappadocia, D. Morse, Plastid ultrastructure defines the protein import pathway in dinoflagellates. *J. Cell Sci.* **116**, 2867–2874 (2003).
16. N. J. Patron, R. F. Waller, J. M. Archibald, P. J. Keeling, Complex protein targeting to dinoflagellate plastids. *J. Mol. Biol.* **348**, 1015–1024 (2005).
17. T. Tengs et al., Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol.* **17**, 718–729 (2000).
18. R. J. Gast, D. M. Moran, M. R. Dennett, D. A. Caron, Kleptoplasty in an Antarctic dinoflagellate: Caught in evolutionary transition? *Environ. Microbiol.* **9**, 39–45 (2007).
19. N. J. Patron, R. F. Waller, P. J. Keeling, A tertiary plastid uses genes from two endosymbionts. *J. Mol. Biol.* **357**, 1373–1382 (2006).
20. C. G. Sellers, R. J. Gast, R. W. Sanders, Selective feeding and foreign plastid retention in an Antarctic dinoflagellate. *J. Phycol.* **50**, 1081–1088 (2014).
21. A. W. Larkum, P. J. Lockhart, C. J. Howe, Shopping for plastids. *Trends Plant Sci.* **12**, 189–195 (2007).
22. F. Husnik et al., Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**, 1567–1578 (2013).
23. R. G. Dorrell, G. A. Hinksman, C. J. Howe, Diversity of transcripts and transcript processing forms in plastids of the dinoflagellate alga Karenia mikimotoi. *Plant Mol. Biol.* **90**, 233–247 (2016).
24. R. J. Gast et al., Abundance of a novel dinoflagellate phylotype in the Ross Sea, Antarctica. *J. Phycol.* **42**, 233–242 (2006).
25. Y. Hou, S. Lin, Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: Gene content estimation for dinoflagellate genomes. *PLoS One* **4**, e6978 (2009).
26. T. C. LaJeunesse, G. Lambert, R. A. Andersen, M. Coffroth, D. W. Galbraith, Symbiodinium (Pyrrhophyta) genome sizes (DNA content) are smallest among dinoflagellates. *J. Phycol.* **41**, 880–886 (2005).
27. H. Zhang et al., Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4618–4623 (2007).

EVOLUTION

28. E. Hehenberger, Metabolic pathways phylogenies/Photosynthesis phylogenies. Fig-share. https://figshare.com/articles/Metabolic_pathways_phylogenies_Photosynthesis_phylogenies/7851467. Deposited March 15, 2019.

29. R. F. Waller, L. Kořený, "Plastid complexity in dinoflagellates: A picture of gains, losses, replacements and revisions" in *Advances in Botanical Research*, Y. Hirakawa, Ed. (Academic Press, Oxford, 2017), **vol. 84**, pp. 105–143.

30. J. Janouškovec *et al.*, Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10200–10207 (2015).

31. K. Katoh, D. M. Standley, A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**, 1933–1942 (2016).

32. S. Whelan, I. Irisarri, F. Burki, PREQUAL: Detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930 (2018).

33. R. H. Ali, M. Bogusz, S. Whelan, Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol. Biol. Evol.* 10.1093/molbev/msz142 (2019).

34. T. Xiang, W. Nelson, J. Rodriguez, D. Tolleter, A. R. Grossman, Symbiodinium transcriptome and global responses of cells to immediate changes in light intensity when grown under autotrophic or mixotrophic conditions. *Plant J.* **82**, 67–80 (2015).

35. B. Bentlage, T. S. Rogers, T. R. Bachvaroff, C. F. Delwiche, Complex ancestries of isoprenoid synthesis in dinoflagellates. *J. Eukaryot. Microbiol.* **63**, 123–137 (2016).

36. M. Kim, S. Kim, W. Yih, M. Park, The marine dinoflagellate genus Dinophysis can retain plastids of multiple algal origins at the same time. *Harmful Algae* **13**, 105–111 (2012).

37. P. Glibert *et al.*, Grazing by Karenia brevis on Synechococcus enhances its growth rate and may help to sustain blooms. *Aquat. Microb. Ecol.* **55**, 17–30 (2009).

38. T. Berge, P. Hansen, Ø. Moestrup, Feeding mechanism, prey specificity and growth in light and dark of the plastidic dinoflagellate Karlodinium armiger. *Aquat. Microb. Ecol.* **50**, 279–288 (2008).

39. T. Berge, P. Hansen, Ø. Moestrup, Prey size spectrum and bioenergetics of the mixotrophic dinoflagellate Karlodinium armiger. *Aquat. Microb. Ecol.* **50**, 289–299 (2008).

40. J. E. Adolf, T. Bachvaroff, A. R. Place, Can cryptophyte abundance trigger toxic Karlodinium veneficum blooms in eutrophic estuaries? *Harmful Algae* **8**, 119–128 (2008).

41. H. J. Jeong *et al.*, Mixotrophy in the phototrophic dinoflagellate Takayama helix (family Kareniaceae): Predator of diverse toxic and harmful dinoflagellates. *Harmful Algae* **60**, 92–106 (2016).

42. M. W. Gray, Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10133–10138 (2015).

43. F. Husnik, J. P. McCutcheon, Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5416–E5424 (2016).

44. K. Stamatakis, D. Vayenos, C. Kotakis, R. J. Gast, G. C. Papageorgiou, The extraordinary longevity of kleptoplasts derived from the Ross Sea haptophyte Phaeocystis Antarctica within dinoflagellate host cells relates to the diminished role of the oxygen-evolving Photosystem II and to supplementary light harvesting by mycosporine-like amino acid/s. *Biochim. Biophys. Acta Bioenerg.* **1858**, 189–195 (2017).

45. W. Yamori, T. Shikanai, Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. *Annu. Rev. Plant Biol.* **67**, 81–106 (2016).

46. G. Peltier, E.-M. Aro, T. Shikanai, NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. *Annu. Rev. Plant Biol.* **67**, 55–80 (2016).

47. C. Desplats *et al.*, Characterization of Nda2, a plastoquinone-reducing type II NAD(P)H dehydrogenase in chlamydomonas chloroplasts. *J. Biol. Chem.* **284**, 4148–4157 (2009).

48. H. Alexander *et al.*, Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5972–E5979 (2015).

49. S. L. Gierz, S. Forêt, W. Leggat, Transcriptomic analysis of thermally stressed Symbiodinium reveals differential expression of stress and metabolism genes. *Front. Plant Sci.* **8**, 271 (2017).

50. X. Shi *et al.*, Transcriptomic and microRNAomic profiling reveals multi-faceted mechanisms to cope with phosphate stress in a dinoflagellate. *ISME J.* **11**, 2209–2218 (2017).

51. S. K. Pierce *et al.*, Transcriptomic evidence for the expression of horizontally transferred algal nuclear genes in the photosynthetic sea slug, Elysia chlorotica. *Mol. Biol. Evol.* **29**, 1545–1556 (2012).

52. D. Bhattacharya, K. N. Pelletreau, D. C. Price, K. E. Sarver, M. E. Rumpho, Genome analysis of Elysia chlorotica Egg DNA provides no evidence for horizontal gene transfer into the germ line of this Kleptoplastic Mollusc. *Mol. Biol. Evol.* **30**, 1843–1852 (2013).

53. A. Nakabachi, K. Ishida, Y. Hongoh, M. Ohkuma, S. Y. Miyagishima, Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr. Biol.* **24**, R640–R641 (2014).

54. J. W. Stiller *et al.*, The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat. Commun.* **5**, 5764 (2014).

55. R. G. Dorrell *et al.*, Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife* **6**, e23717 (2017).