Functional Relationship between a Dinoflagellate Host and Its Diatom Endosymbiont

Elisabeth Hehenberger,*[,1] Fabien Burki,[‡,1] Martin Kolisko,[1] and Patrick J. Keeling[1]

[1]Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, BC, Canada

[‡]Present address: Science for Life Laboratory, Program in Systematic Biology, Department of Organismal Biology, Uppsala University, Uppsala, Sweden

***Corresponding author:** E-mail: helisabe@mail.ubc.ca.

Associate editor: John True

Abstract

While we know much about the evolutionary patterns of endosymbiotic organelle origins, we know less about how the actual process unfolded within each system. This is partly due to the massive changes endosymbiosis appears to trigger, and partly because most organelles evolved in the distant past. The dinotoms are dinoflagellates with diatom endosymbionts, and they represent a relatively recent but nevertheless obligate endosymbiotic association. We have carried out deep sequencing of both the host and endosymbiont transcriptomes from two dinotoms, *Durinskia baltica* and *Glenodinium foliaceum*, to examine how the nucleocytosolic compartments have functionally integrated. This analysis showed little or no functional reduction in either the endosymbiont or host, and no evidence for genetic integration. Rather, host and endosymbiont seem to be bound to each other via metabolites, such as photosynthate exported from the endosymbiont to the host as indicated by the presence of plastidic phosphate translocators in the host transcriptome. The host is able to synthesize starch, using plant-specific starch synthases, as a way to store imported photosynthate.

Key words: **endosymbiosis, metabolism, integration, transporter, dinotom.**

Introduction

That eukaryotic organelles, mitochondria and plastids, originated from endosymbiosis is widely accepted, but the process whereby a transient endosymbiont turned into a permanent organelle remains debated. The depth of this problem is reflected in our lack even of clear definitions for what really distinguishes organelles and endosymbionts. Most commonly organelles are defined (and distinguished from endosymbionts) by their genomic integration, that is the translocation of genes from the endosymbiont to the host genome, and by the presence of a system targeting proteins—of endosymbiont origin or other—back to the endosymbiont. Several discoveries in recent years have begun to blur those definitions, however, such as the description of gene-product targeting to endosymbionts not regarded as classic organelles (Nakabachi et al. 2014) or the characterization of endosymbionts stably integrated into the host without gene transfer and targeting events (Burki et al. 2014).

Part of the challenge to studying the transformation from a transient endosymbiotic association to a permanent one is that these events happened so long ago, and so much change has taken place since the establishment of plastids and especially mitochondria that it is hard to tease apart the order of events. One approach is to examine more recent but nevertheless stable endosymbiotic associations, which we may or may not call organelles, thereby allowing the detection of possible general principles of endosymbiosis. The stable association of a diatom endosymbiont within a dinoflagellate host as represented by the so-called 'dinotoms' (Tomas and Cox 1973; Kite and Dodge 1985; Chesnick et al. 1996, 1997) is one such promising model system. The dinoflagellate has undergone what has been described as tertiary endosymbiosis, by replacing its ancient secondary plastid of red algal origin with another alga itself already containing a secondary plastid also of red algal origin. Compared with the primary (uptake of a cyanobacterium by a eukaryote) or secondary (uptake of a primary plastid-containing alga) endosymbiotic events, the tertiary endosymbiosis has happened more recently. Dinotoms may therefore retain clues to endosymbiotic evolution that have been lost over the long time span that has passed since earlier endosymbiotic events. In addition, the lineages of host and endosymbiont are both phylogenetically well characterized (Horiguchi and Takano 2006; Imanian and Keeling 2007; Pienaar et al. 2007; Hehenberger et al. 2014), allowing a much more reliable identification of which lineage any given gene is from, and by extension potential endosymbiotic gene transfer (EGT) events that would underpin the kind of genomic integration usually looked for in an organelle. Furthermore, in contrast to all primary and secondary plastids as well as the only other described tertiary plastid of the dinoflagellate group Kareniaceae (Tengs et al. 2000), the dinotom endosymbiont has preserved not only its plastid but also a number of other cellular features, such as a large, multilobed nucleus, mitochondria, the endoplasmatic reticulum and ribosomes that are all separated from the host cytosol by a single membrane (Dodge 1971; Tomas and Cox 1973; Tomas et al. 1973; Horiguchi and Pienaar 1988). The endosymbiont is present in all stages of the host cell cycle and division of host and endosymbiont occur simultaneously

(Tippit and Pickettheaps 1976; Figueroa et al. 2009), indicating a stable endosymbiotic relationship. Whether the level of endosymbiont reduction is indicative of a more recent endosymbiotic event, or whether these organisms have developed a particular form of association is currently unknown. It is also unknown whether they are genetically integrated or how the two partners are dependent on one another. Prior comparative analyses have shown that host and endosymbiont mitochondria as well as the endosymbiont plastid retained nearly all characteristics of the corresponding organelles in the respective free-living relatives (Imanian and Keeling 2007; Imanian et al. 2010, 2012) and single metabolic pathways have been found to be present in host and endosymbiont in parallel (Hehenberger et al. 2014; Imanian and Keeling 2014). Some of these pathways seem to be targeted to a remnant of the ancestral peridinin-plastid (Hehenberger et al. 2014), which may be represented by an eyespot structure present in the host cytosol (Dodge 1984). Fully functional, photosynthetically active plastids are however only found in the diatom endosymbiont and the dinoflagellate host likely profits from photosynthate provided by the endosymbiont plastid. During primary plastid endosymbiosis, carbon produced by photosynthesis in the cyanobacterium was exported and integrated into the host metabolism, probably via the storage polysaccharide starch, thereby linking host and endosymbiont metabolism (Ball et al. 2011). A similar scenario can be imagined for the establishment of a metabolic connection between dinotom host and endosymbiont, as the host is able to synthesize starch indicated by the presence of starch granules in the host cytoplasm (Tomas and Cox 1973).

Here, we investigate the endosymbiont–host relationship of the dinotom *Durinskia baltica* on a global scale by transcriptomic analysis. Previous studies have examined data from the plastid or mitochondrial genomes, or focused on genes only from the host nucleus (Imanian and Keeling 2007; Imanian et al. 2010; Burki et al. 2014), revealing no clear functional basis for the partnership and leaving open the possibility that cytosolic metabolism was somehow functionally partitioned. Here, we examined the complete transcriptome of the dinotom, representing a much more comprehensive dataset that includes cytosolic metabolism. From a large database of 7,010 genes, we sorted genes into host- and endosymbiont-encoded transcripts based on their phylogenetic association, revealing a very similar number of genes in the host and endosymbiont. We did not detect a significant number of EGT events, suggesting no genomic integration. Moreover, comparing metabolic pathways between host, endosymbiont, and a free-living diatom revealed seemingly no impact of the endosymbiosis on the endosymbiont metabolism, while metabolic pathways affected in the host are almost exclusively plastid related. In an effort to detect a metabolic link between the two, we investigated the host-encoded starch metabolism and discovered that dinotoms (and dinoflagellates in general) unexpectedly synthesize cytosolic starch using starch synthases that were so far only found in plants and green algae. Host starch synthesis may in turn be fueled by photosynthate from the endosymbiont via several plastidic phosphate transporters that we identified in the host transcriptome.

Results and Discussion

Identifying Host and Endosymbiont Transcripts in the Dinotom Transcriptome

To investigate the degree of endosymbiotic integration in a dinotom, we generated transcriptomes based on two biological replicates from cultures of the closely related dinotoms *D. baltica* and *Glenodinium foliaceum*, which have likely diverged only after uptake and integration of a diatom endosymbiont. This resulted in a total of 100,211,872 and 97,658,774 reads, respectively, which were assembled into 332,209 and 301,986 unique contigs, respectively. This is not surprising given that the host and the endosymbiont in those two dinotoms have large genomes, and dinoflagellates at least are known to have many duplicates of many or most genes.

To distinguish transcripts derived from the host from transcripts derived from the endosymbiont, the phylogenetic position of each gene would need to be determined. However, the total number of unique transcripts we identified was far too large to allow for any thorough phylogenetic reconstruction to be feasible. Since we are primarily interested in determining the functional interactions between the partners, we therefore focused exclusively on genes with an already annotated function to produce a dataset with all known functional information but at the same time of a scale that could feasibly be analyzed. The KEGG Automatic Annotation Server (KAAS) allows functional annotation of genes by BLAST comparison against curated databases, thereby generating a list of genes orthologous to the KEGG GENES database as well as pathways representing the biological system (Kanehisa and Goto 2000; Kanehisa et al. 2014). We chose the *D. baltica* dataset for the sorting process as an initial clustering yielded a higher number of unique clusters than the *G. foliaceum* dataset. By first clustering and then annotating the transcriptome of *D. baltica* using the KAAS tool (see Material and Methods), we reduced the number of transcripts generated by open reading frame (ORF) prediction from >150,000 to 7,010 transcripts homologous to a gene in the KEGG GENES database.

Those 7,010 transcripts were used to carry out phylogenetic analysis employing a custom database comprising both dinotom transcriptomes, all other dinoflagellate and all diatom data available, as well as selected outgroups (supplementary table 1, Supplementary Material online). By manual investigation of the reconstructed trees, >98% of the transcripts could be assigned to be either dinoflagellate- or diatom-associated (3,951 vs. 2,970 transcripts, fig. 1). The sorted transcripts correspond to 2,415 dinoflagellate- and 2,211 diatom-associated unique KEGG identifiers, respectively, as several genes are represented by more than one transcript in both datasets (fig. 1). The number of individual genes that we were able to annotate in endosymbiont is therefore almost as high as the number of individual genes in the host, suggesting a high level of functionality in the endosymbiont. Eighty-nine transcripts (<2% of the annotated transcripts) did not have a sufficient number of homologs in the datasets used to be assigned to a group, or were

determined to be most likely bacteria associated (data not shown).

No Endosymbiotic Gene Transfer Could Be Identified

Recent examinations of limited datasets of dinotom transcripts showed that there was no strong support for EGT (Burki et al. 2014; Hehenberger et al. 2014), but did show a distinct GC content for each genome. Burki et al. (2014) had employed that characteristic to identify nine putative cases of EGT events, that is, transcripts that seem to originate from the endosymbiont based on their phylogenetic position but displayed GC contents typical for the host genome.

Here, we analyzed a much larger number of transcripts that were assigned to be diatom-associated, and examined first their GC content. As previously reported, diatom genomes have a GC content of <55%, while most dinoflagellate transcripts have GC contents >55% (Burki et al. 2014). This gives a first indication of whether a transcript is encoded in the host nucleus or the endosymbiont nucleus. Of the 2,971 transcripts deemed to be diatom associated, the majority had GC contents between 32% and 56%, as expected, while only three transcripts had distinctly higher GC contents of 62–65%, equivalent to host-encoded transcripts (fig. 2A).

Those three transcripts correspond to the same protein, a nitrate reductase, and cluster with high support with homologs from all other diatoms (supplementary fig. 1, Supplementary Material online), suggesting that they have been transferred from the endosymbiont to the host nucleus, followed by alteration of their original GC content. However, within the diatom clade their position is not strongly resolved, and they do not branch with the closest relatives of the endosymbiont (e.g., *Fragilariopsis*—Horiguchi and Takano 2006; Imanian et al. 2007; Pienaar et al. 2007; Hehenberger et al. 2014). Moreover, we were not able to detect a dinoflagellate-specific spliced leader (SL) sequence at the 5′ end of any of the three transcripts (see below). The endosymbiont also encodes its own copy of nitrate reductase (GC = 49%), which is phylogenetically distinct from the high-GC copies and clusters with *Nitzschia*, *Pseudo-nitzschia*, *Fragilariopsis*, and *Cylindrotheca* (the position observed throughout the other diatom-associated phylogenies). There is no dinoflagellate-associated homolog from *D. baltica*, even when applying

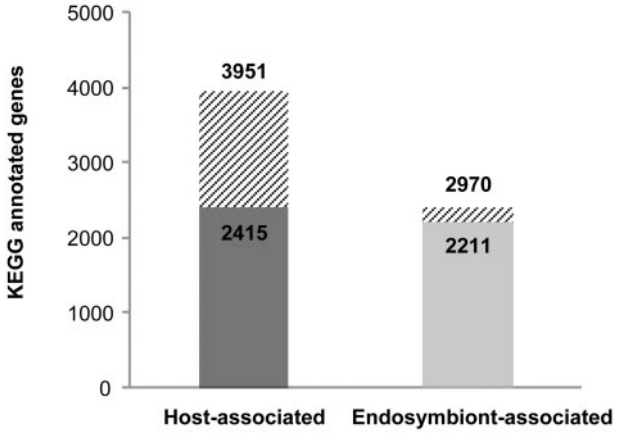

FIG. 1. Host and endosymbiont encode a similar number of annotated genes. A bar chart shows the number of KAAS-annotated host- and endosymbiont-encoded transcripts based on their phylogenetic association. The numbers above the bars indicate the total number of annotated transcripts, while the solid bars (and corresponding numbers) indicate the number of unique KEGG gene identifiers for each group.

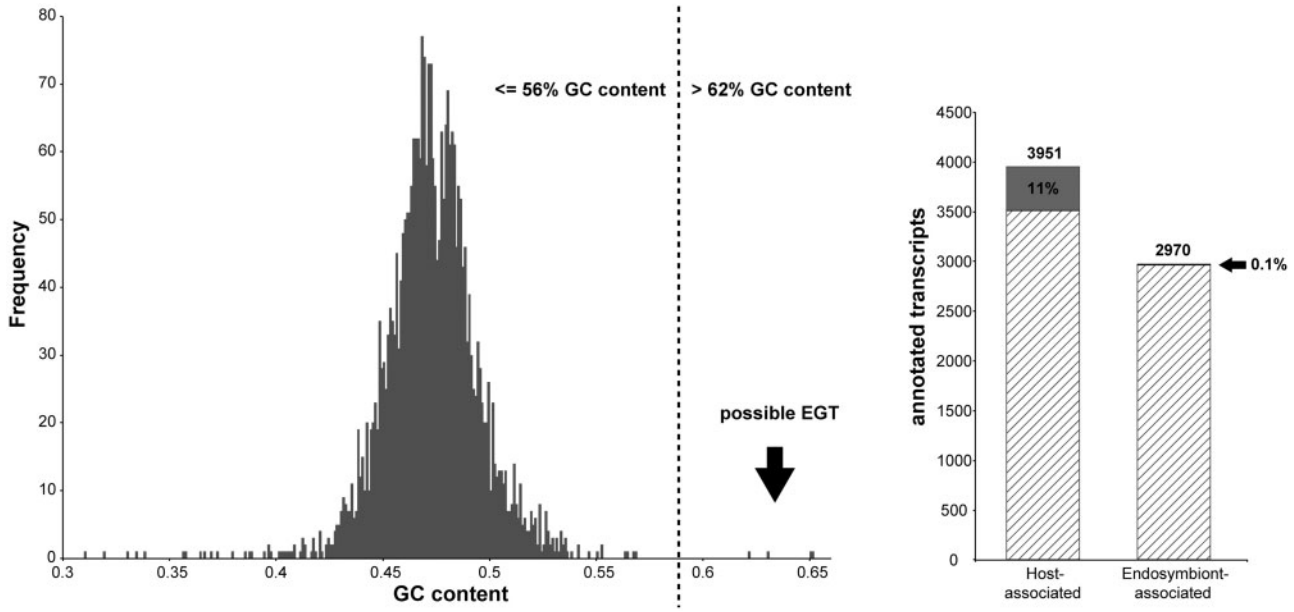

FIG. 2. A lack of EGT in dinotoms. (*A*) A histogram showing the distribution of GC-content for endosymbiont-associated transcripts. An arrow indicates three transcripts with an unusually high GC-content coding for nitrate transporters and that are possibly derived by HGT from a diatom (though phylogenetically maybe not the endosymbiont diatom). (*B*) A bar graph showing the fraction of sorted transcripts that contain at least a 10-nucleotide string of the SL sequence. Solid bars and an arrow indicate transcripts containing at least a 10-nucleotide string of the SL sequence in the host and the endosymbiont, respectively. Numbers above the bars indicate the total number of annotated transcripts.

less stringent filters to the BLAST search output while reconstructing the tree, suggesting the host-encoded nitrate reductase was possibly lost and replaced by a diatom-derived copy. As the high-GC diatom transcripts do not group within the clade that contains the low-GC diatom transcripts, this transfer seemingly is not related to the present endosymbiont. However, this part of the tree is largely unresolved and origin of this gene from the current endosymbiont cannot be ruled out. In that case it would seem that this EGT event is relatively recent—assuming the two dinoflagellate lineages only diverged after having taken up the current endosymbiont—as *G. foliaceum* still encodes a partial host copy of nitrate reductase and we did not find a high-GC homolog within the diatom clade.

Using the much broader taxon sampling described in this article, we were also able reassess the origin of the nine putative cases of EGT reported in Burki et al. (2014). We found that all nine genes actually cluster within the dinoflagellates in our analysis, reducing the number of EGTs to 0.

Another approach to identifying possible cases of EGT in this dataset makes use of the fact that dinoflagellate transcripts have been shown to encode SL sequences at their 5′ ends (Zhang et al. 2007). We searched for those sequences using the last 11 nucleotides of the conserved 22 nucleotide SL sequence, taking into account that many transcripts may not be fully represented at their 5′ end. Applying an additional restriction of a 30-bp-search window (at the 5′ and the 3′ end of each sequence) resulted in 4,232 occurrences of the search string in the original assembly of 332,209 contigs, which corresponded to 4,293 predicted ORFs. We cross-checked those ORFs against the sorted transcripts and found 406 dinoflagellate-associated, and two diatom-associated transcripts with at least the 11 nucleotide sequence, corresponding to 10% and $<$0.1% of the dinoflagellate- and diatom-associated transcripts, respectively. When relaxing the search parameters to a 10-nt search string, these numbers increase to 435 (11%) and three (0.1%) transcripts, respectively (fig. 2B). Investigation of the three diatom-associated transcripts showed that they are relatively large (between 4,600 and 8,600 nt) and code for several ORFs. The presence of contigs coding for more than one ORF is likely one of the reasons we recover more ORFs (4,293) than contigs (4,232) containing the 10-nt search string. While two of the three transcripts represent clearly misassemblies of dinoflagellate and diatom-associated transcripts, the remaining transcript contains the ORF of interest in close proximity to the SL fragments ($\sim$100 nt) and might therefore represent a possible candidate for an EGT event. However, when mapping back the sequence reads to the contig in question, only one read supports the region encompassing the SL sequence, suggesting another, if less obvious, misassembly. Also the GC content of this transcript is typical for diatom-encoded transcripts ($<$ 55%).

The relatively small percentage of SL-containing transcripts, even in the dinoflagellate-associated dataset, may indicate a large fraction of 5′-truncated transcripts in the assembly and/or modified SL sequences that were not identified by our search. However, the fraction of transcripts being trans-spliced in dinoflagellates is contentious and may also differ between dinoflagellate groups (Zhang et al. 2007; Shoguchi et al. 2013; Xiang et al. 2015), making it very difficult to predict how many SL-encoding transcripts to expect. Adding to that the extremely low numbers of identified diatom-associated transcripts carrying the SL, the presence of misassemblies and the contrasting GC contents, we conclude that a substantial number of EGT events in *D. baltica* is not very likely.

Global Analysis of Metabolism in the Host, Endosymbiont, and Their Organelles

Annotated and phylogenetically sorted genes were used to reconstruct the metabolic pathways that make up basic host and endosymbiont metabolism, generating a metabolic pathway map for each with the aim of identifying pathways complete in one but entirely missing in the other. The annotation analysis was based on a stringent bidirectional BLAST against the KEGG GENES database, which contains only one pennate diatom genome and no dinoflagellate genome. The pennate diatom genome available in the database is that of *Phaeodactylum tricornutum*, which is not the closest relative to the endosymbiont. Taking also into account that diatom-specific genes are fast evolving (Bowler et al. 2008), this is a very conservative approach, yielding a high number of false negatives. If a pathway was recovered by KAAS in one but not the other set of sorted transcripts, we manually searched for homologs of the missing transcripts of interest in our dataset to either confirm their absence or recover transcripts that were not detected by KAAS.

Globally, we recovered a similar number of annotated metabolic pathway genes in the host- and endosymbiont-associated datasets (495 vs. 554); however, those datasets are only partially overlapping (fig. 3A). Closer investigation showed that many of those differences are due to redundancy; for many steps in some metabolic pathways there exists more than one enzyme able to catalyze that reaction. Many of the enzymes unique to one partner fell into this category and do not necessarily suggest a unique function. However, there were also a number of pathways that were not represented at all in one partner, or only partially by enzymes that also function in other pathways (fig. 3B and C, supplementary table 2, Supplementary Material online), and these are summarized under several headings below.

Plastid and Plastid-Related Cytosolic Pathways: Photosynthesis, Fatty Acids, and Terpenoids

Not surprisingly, the host is missing almost all proteins associated with photosynthesis, including the proteins for photosystem I and II, the cytochrome b6/f complex, and the ATP synthase complex. But interestingly several plastid genes were identified, including a plastidial version of the electron transfer protein ferredoxin and the associated ferredoxin-NADP+ reductase (fig. 3C, supplementary table 2, Supplementary Material online). While the transcript for the ferredoxin-NADP+ reductase is truncated at its 5′ end, the ferredoxin transcript is full length and is predicted to encode a signal peptide followed by the phenylalanine-

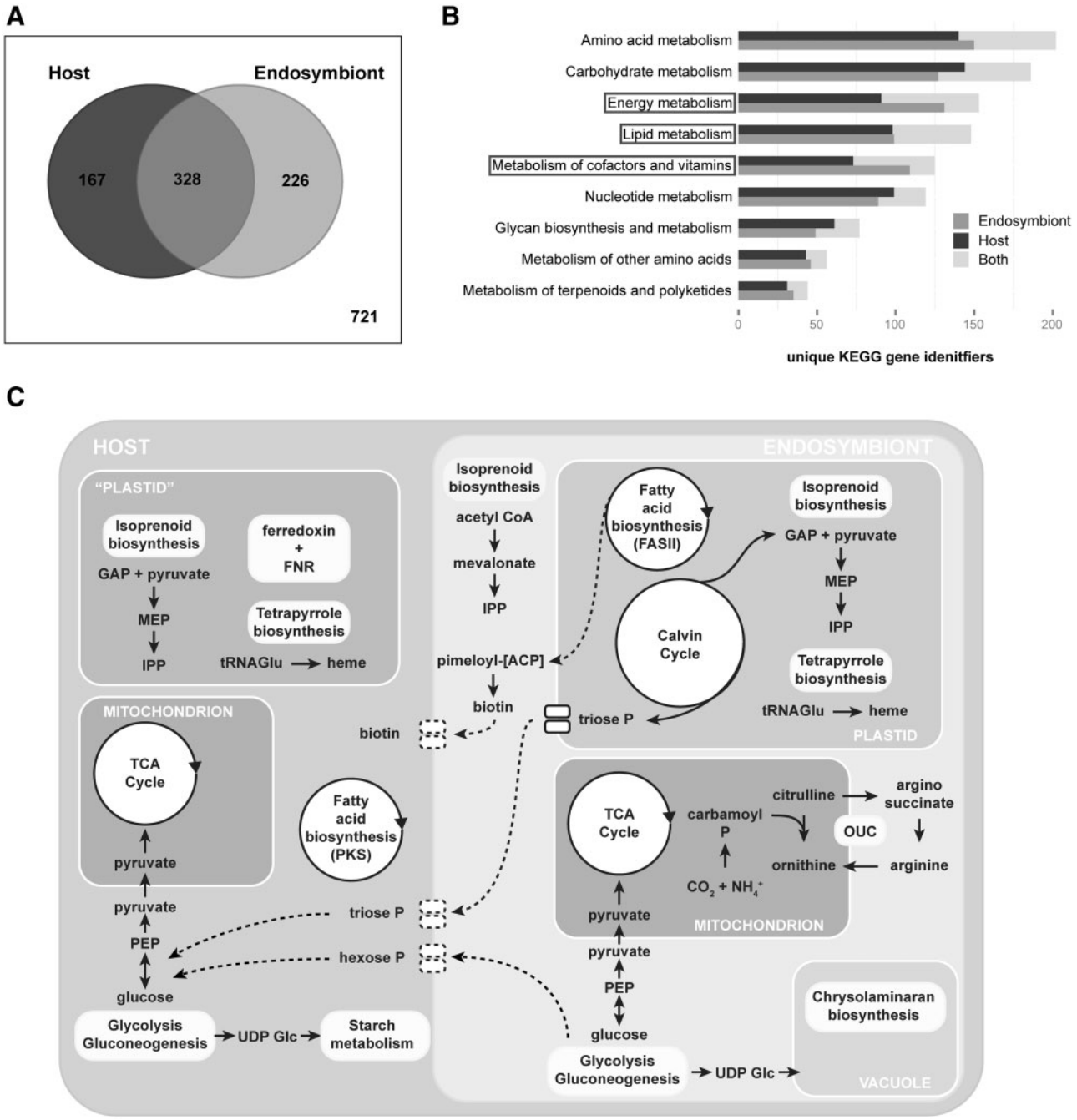

FIG. 3. Pathway loss and retention in host and endosymbiont. (*A*) Annotated transcripts mapping to KEGG metabolic pathways are only partially overlapping in host and endosymbiont. (*B*) Comparison of host- and endosymbiont-associated transcripts mapping to different kinds of metabolism in the KEGG pathway database. The bars represent the numbers of unique KEGG identifiers in the single sorted datasets as well as in both datasets additively combined, indicating that those datasets are only partially overlapping. Metabolic pathways indicated by boxes are those most strongly differentiated in host and endosymbiont. (*C*) Schematic overview of key shared or differentiated energy generating or nutrient metabolizing metabolic pathways. Dashed lines indicate putative metabolite transfer from endosymbiont to host, and their presence implies transport between host and endosymbiont of the relevant metabolites.

containing motif typical for dinoflagellate plastid-targeting transit peptides (Patron et al. 2005; Patron and Waller 2007). This observation is consistent with the suggested presence of a relic plastid and expands its potential metabolic functions (Hehenberger et al. 2014). Ferredoxin and ferredoxin-NADP+ reductase form a redox couple also found in other nonphotosynthetic plastids, and it was proposed that these two proteins may act together in a reductive metabolic pathway other than photosynthesis (Vollmer et al. 2001; Pandini et al. 2002). The host also encodes a number of enzymes in the Calvin cycle that are normally plastid-targeted, however in each case these enzymes can function in another pathway as well and, more importantly, the key enzymes RuBisCo, sedoheptulose-1,7-bisphosphatase, phosphoribulokinase, and plastidial versions of fructose-bisphosphate aldolase class I and II were not found (fig. 3C, supplementary table

2, Supplementary Material online). These observations contrast with the only other analyzed dinoflagellates with a tertiary plastid, the Kareniaceae, where all Calvin cycle proteins in the host nucleus are targeted to the their tertiary plastid, even though some of those proteins are derived from the ancient peridinin plastid (Patron et al. 2006) (supplementary table 2, Supplementary Material online).

Fatty acids are synthesized in plastids via the multi-enzyme FASII complex or in the cytosol typically using a large multi-domain FASI protein. The diatom endosymbiont produces fatty acids in the plastid like its free-living relatives (Armbrust et al. 2004), supported by the presence of all FASII subunits, complete with targeting presequences (supplementary table 2, Supplementary Material online). Dinoflagellates seem to be able to synthesize fatty acids in their peridinin-plastid in the same way, but also in the cytosol. However, the cytosolic enzyme is not the expected FASI, but instead a likely multifunctional polyketide synthase closely related to FASI (Dolah et al. 2013; Janouškovec et al. 2015). We identified a large multi-domain polyketide synthase predicted to be localized in the cytosol of the host, but we were not able to identify any enzymes belonging to the plastid FASII pathway in the host (fig. 3C, supplementary table 2, Supplementary Material online). Interestingly, we found the Kareniaceae express all FASII pathway proteins and they do not cluster with other dinoflagellates in our trees (with one exception, see supplementary table 2, Supplementary Material online), suggesting they are derived from their tertiary endosymbiont. As the Kareniaceae also express a FASI-like polyketide synthase, this suggests a need for the plastidial pathway besides the polyketide synthase-based fatty acid synthesis in the cytosol, and raises the question of whether the dinotom host receives specific fatty acids from the endosymbiont that it cannot synthesize via the cytosolic pathway.

Like fatty acids, terpenoid (or isoprenoid) precursors can also be synthesized via two different pathways: the mevalonate (MVA) pathway localized in the cytosol, and the plastidial nonMVA or MEP/DOXP pathway. Both pathways have been described to co-exist in diatoms, synthesizing the precursors for different kinds of terpenoids in two different compartments of the cell (Massé et al. 2004; Lohr et al. 2012). We reported in an earlier study that in dinotom host and endosymbiont both encode transcripts for genes in the plastidial MEP/DOXP pathway (Hehenberger et al. 2014). Here, we also find diatom-associated homologs for the complete cytosolic MVA pathway, suggesting the endosymbiont, like the free-living diatoms, retains both the cytosolic as well as the plastidial pathway for terpenoid synthesis. The host and the most closely related dinoflagellate for which data are available, *Scrippsiella trochoidea*, encode several but not all proteins from the MVA pathway too (supplementary table 2, Supplementary Material online), but most of these enzymes are known not to be unique to this pathway and most likely function in other pathways in the host as well. However, one of the three enzymes specific for the MVA pathway, phosphomevalonate kinase (PMVK), was identified in the host and several other dinoflagellate lineages, but we were not able to identify the other two MVA-specific enzymes, mevalonate kinase (MVK) or diphosphomevalonate decarboxylase in any of the dinoflagellates. Intriguingly the host-encoded PMVK is nonhomologous to its endosymbiont-encoded counterpart: while the endosymbiont expresses a PMVK found in plants, fungi and bacteria, the host PMVK belongs to the animal-type family of the enzyme (Houten and Waterham 2001). Since the other MVA pathway-specific enzymes are apparently missing in the host, it appears the host and dinoflagellates in general are not synthesizing terpenoids via this pathway, and that the animal-type PMVK may have another function in the dinoflagellates yet to be identified. Taken together, the dinotoms are able to synthesize terpenoid precursors via at least three different pathways, two of them encoded by the endosymbiont and one of them encoded by the host (fig. 3C).

Cytosolic and Mitochondrial Pathways: Biotin, Urea Cycle

One clear difference between host and endosymbiont that could suggest functional loss in the host is that no transcripts for any protein from the biotin (vitamin B_7) synthesis pathway were found in the host. This is a cytosolic pathway that is present in the endosymbiont as well as in the transcriptome of *S. trochoidea* (supplementary table 2, Supplementary Material online). A few dinoflagellates have been reported to be biotin auxotrophs, but *D. baltica* does not require exogenous biotin (Croft et al. 2006), suggesting the host obtains its biotin from the endosymbiont.

In other cases, pathways were found in the endosymbiont but not the host, however these pathways were unlikely to be present in dinoflagellates to begin with, so do not represent functional change resulting from endosymbiosis. An example is the almost complete lack of enzymes involved in the ornithine-urea cycle (OUC) in the host, prompting us to examine this pathway more closely. By integrating core carbon and nitrogen metabolism, the OUC plays a central role in the response of diatoms to nitrogen availability (Allen et al. 2011). The first committed step of this pathway is represented by the mitochondrial carbamoyl phosphate synthase (CPSIII) that utilizes ammonium as a substrate. In diatoms a second type of carbamoyl phosphate synthase (CPSII) is also found, which uses glutamine for the synthesis of pyrimidine in the cytosol. The endosymbiont encodes both types of carbamoyl phosphate synthase, while the host and all dinoflagellates contain only two paralogs of the cytosolic non-OUC form (supplementary table 2, Supplementary Material online). The next step in the pathway also takes place in the mitochondria, and both the endosymbiont and host encode an ornithine transcarbamylase catalyzing this step. However, the host gene (like all other dinoflagellate sequences in the tree) does not encode a mitochondrial targeting transit peptide and these genes are very divergent compared to the diatom sequences. Similarly, the enzymes catalyzing the next two steps in the cycle are present, but share low similarity between dinoflagellates and diatoms (including the host and endosymbiont) (supplementary fig. 2A–E, Supplementary Material online). The last enzyme, arginase, is present in endosymbiont and dinoflagellates but was not found in the dinotom host (sup

plementary table 2, Supplementary Material online). Altogether those results indicate that dinoflagellates contain several enzymes in this pathway, but they were not recognized by KAAS due to their divergent sequences. However, the deviant localization of the host enzymes and particularly the absence of the OUC-committed enzyme CPSIII suggest a different usage than the urea cycle for them (e.g., amino acid synthesis), altogether suggesting that the OUC pathway is probably not present in dinoflagellates in general. This stands in contrast to the findings of a recent genomic survey of the dinoflagellate *Symbiodinium kawagutii* which predicts a full mitochondrial urea cycle in this species (Lin et al. 2015).

Retention and Loss of Diatom-Specific Pathways

As described above, the endosymbiont does not seem to lack any metabolic pathways compared with the host. However, host and endosymbiont are only distantly related, and comparing endosymbiont metabolism to closely related free-living diatoms could also reveal loss of diatom-specific functions. Comparing the diatom-associated genes with the thoroughly analyzed genome of the pennate diatom *P. tricornutum* showed that all general metabolic pathways and genetic information processing activities, such as transcription, translation, and replication, are represented in the endosymbiont very much as in the free-living diatom (fig. 4).

The retention of many general cellular functions is not surprising, but there are also a number of processes that may be expected to be lost or massively altered in an endosymbiont, such as cell wall formation and cell division. Searching for all genes involved in silica cell wall formation revealed several full-length transcripts for silicon transporters and frustulins, as well as a large number of spermine/spermidine synthase transcripts (supplementary table 3, Supplementary Material online) in the diatom-associated genes. Compared with the complement of silicon transporters in the most closely related genera *Nitzschia*, *Pseudo-nitzschia*, *Fragilariopsis*, and *Cylindrotheca* recovered by our phylogenetic reconstruction, the overall number of silicon transporter transcripts in the endosymbiont seems to be reduced (supplementary fig. 3, Supplementary Material online). Such a reduction might be a consequence of the conditions within the host cytosol, for example, there is likely no silica available within the host, as we could not identify host-encoded silicon transporters. The absence of the cell wall might therefore be the result of the absence of silica more than the reduction in the ability to build a wall. Frustulins constitute a major organic component of the diatom silica cell wall (Kröger et al. 1994), and phylogeny of frustulins from *D. baltica* does not suggest any significant losses in this family from the endosymbiont (supplementary fig. 4A–C, Supplementary Material online). Spermine/spermidine synthases are involved in the synthesis of long-chain polyamines, which in turn play a central role in silica morphogenesis (Sumper and Kröger 2004). Spermine/spermidine synthase genes also belong to one of several gene families shown to be over-represented in diatoms (Bowler et al. 2008) and phylogenetic analysis shows that this family is even more expanded in the endosymbiont than in the model diatom *P. tricornutum*. The expansion of spermine/spermidine synthase genes in the endosymbiont has occurred to a similar extent as in the closely related diatom genera *Nitzschia*, *Pseudo-nitzschia*, *Fragilariopsis* and *Cylindrotheca* (supplementary fig. 5, Supplementary Material online).

Another gene family specifically expanded in diatoms is the cyclin family, which plays a major role in the regulation of the cell cycle of eukaryotes. Cell division occurs simultaneously in dinotom host and endosymbiont (Figueroa et al. 2009), suggesting the host may exert cell cycle control over the endosymbiont, which could predict a reduced number of endosymbiont-encoded cyclins. However, searching for cyclin

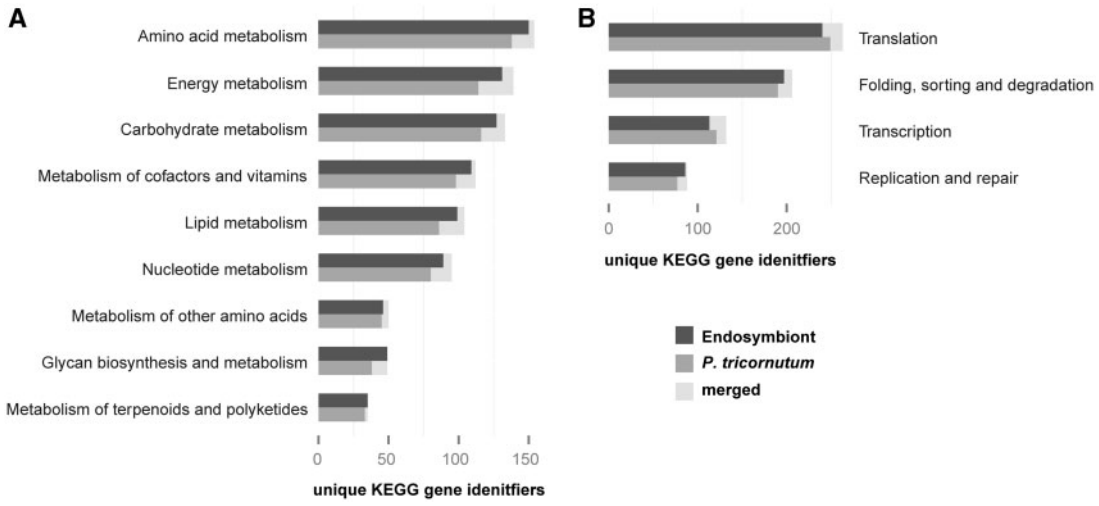

FIG. 4. Comparing the endosymbiont genome to a free-living relative. (*A*) Bar graph showing the number of unique KEGG identifiers mapped to metabolic pathways in endosymbiont and its free-living relative *P. tricornutum*, as well as the unique KEGG identifiers in both datasets additively combined, indicating that those datasets are only partially overlapping. (*B*) Bar graph showing the number of unique KEGG identifiers mapped to genetic information processes in endosymbiont and its free-living relative *P. tricornutum* as well as the unique KEGG identifiers in both datasets combined.

domains and phylogenetic reconstruction showed the endosymbiont cyclin complement actually exceeds that of *P. tricornutum*, and is very similar to that of the more closely related diatom genera *Nitzschia*, *Pseudo-nitzschia*, *Fragilariopsis*, and *Cylindrotheca* (supplementary fig. 6 and table 4, Supplementary Material online). Similarly, other putative cell division components identified in *P. tricornutum* (De Martino et al. 2009) could all be identified in the endosymbiont as well (supplementary table 5, Supplementary Material online). As formation of mitotic spindles has not been observed in the dividing endosymbiont nucleus, it has been proposed that the dinotom endosymbiont does not undergo a 'proper' mitosis (Tippit and Pickettheaps 1976; Figueroa et al. 2009). However, a recent investigation of *P. tricornutum* revealed a nuclear division highly reminiscent of that described for the endosymbiont: among other very similar characteristics the authors were not able to observe a mitotic apparatus or microtubules at any time during the cell cycle (Tanaka et al. 2015), suggesting that the endosymbiont undergoes normal cell division unaffected by the endosymbiosis. Also the high similarity of the endosymbiont transcriptome to its free-living relatives suggests that mitosis is still under strict control mechanisms in the endosymbiont and inheritance of a complete set of chromosomes to the daughter cells still takes place. This does not answer the question how synchronization of host and endosymbiont cell cycle is achieved, but it likely does not depend on genetic transfer from endosymbiont to the host and subsequent complete loss of control of the endosymbiont over its own cell cycle. Rather it may depend on "environmental" cues as suggested by the discovery of diatom-specific cyclins responding to nutritional signals (Huysman et al. 2010).

Similar to cell division, a lack of a spindle apparatus was described during meiosis in the endosymbiont. However, in-depth microscopic analysis has demonstrated a synchronous sexual cycle in the host and the endosymbiont, reporting plasmogamy and karyogamy of the endosymbionts and their nuclei within fusing host gametes (Chesnick and Cox 1989).

Retention of Ancestral Carbohydrate Storage in Host and Endosymbiont: Starch and Chrysolaminaran

The hypothesis that tertiary endosymbiosis was successful because of the provision of photosynthate from the ingested prey to the heterotrophic predator requires a very early establishment of a metabolic connection between the two partners. Storage polysaccharides such as starch and glycogen are hypothesized to have played a central role in establishing a biochemical connection between host and endosymbiont metabolism during primary endosymbiosis (Ball 2014), so storage polysaccharide synthesis may be a likely process to examine to find evidence for the metabolic link between host and endosymbiont in dinotoms.

All three Archaeplastida lineages that emerged from the primary endosymbiotic event—green algae & plants (Viridiplantae), red algae, and glaucophytes—use starch as a storage polysaccharide. Dinoflagellates as well as some other secondary plastid-bearing lineages also store starch. Unlike plants and green algae, which accumulate starch in their plastids, the glaucophytes, red algae, dinoflagellates and the apicomplexan *Toxoplasma gondii* all synthesize starch in their cytosol or, in the case of cryptophytes, within their periplast. While plastidial starch synthases are related to bacterial enzymes and use ADP-glucose as their substrate, lineages with cytosolic and periplastidial starch were proposed to produce it from UDP-glucose using a host-derived glucan synthase. Indeed, several studies have confirmed that red algae, cryptophytes, *T. gondii* and dinoflagellates use a UDP-glucose-based pathway of starch synthesis (Viola et al. 2001; Coppin et al. 2005; Deschamps et al. 2006; Dauvillée et al. 2009).

The principal energy storage polysaccharide in diatoms, in contrast, is not starch (or glycogen), but the β-1,3-linked glucan chrysolaminaran (Kroth et al. 2008). So the ancestors of dinotoms would have been a cytosolic starch-storing dinoflagellate, and a cytosolic chrysolaminaran-storing diatom.

To investigate storage carbohydrates, we first searched for transcripts involved in starch synthesis. Quite unexpectedly we identified transcripts for ADP-glucose-dependent starch synthases similar to the plastidial enzymes of the green algae and plants, but no eukaryotic UDP-glucose-dependent starch synthases in *D. baltica*. Phylogenetic reconstruction showed that the identified enzymes cluster with the soluble starch synthase group SSI–SSII of green algae and plants (fig. 5) and that this group is present not only in dinotoms but all other dinoflagellate and cryptophytes as well. The SSI–SSII group likely originated from duplication of the granule-bound starch synthase GBSS1 gene that was transferred from the cyanobacterial symbiont to the ancestor of Archaeplastida after primary endosymbiosis (Patron and Keeling 2005; Ball et al. 2013). This duplication event was thought to have taken place only in green algae and plants, but the presence of SSI–SSII-related sequences in dinoflagellates (and also cryptophytes) indicates a diversification of GBSS1 already before the Archaeplastida diverged and subsequent loss in red algae, glaucophytes and *T. gondii*. Alternatively a horizontal gene transfer event from green algae to some secondary red plastid-bearing lineages may have caused the observed scenario, but the topology of our phylogenetic reconstruction does not render such a scenario very likely. Interestingly, while green algae and plants encode two enzymes in this group, dinoflagellates and cryptophytes possess at least four SSI–SSII-group soluble starch synthases. This additional diversification may reflect the absence of soluble starch synthase group SSIII–SSIV in dinoflagellates and cryptophytes, which is essential for starch synthesis in plants (Szydlowski et al. 2009). The SSIII–SSIV group is of bacterial but not cyanobacterial origin and was identified so far only in green algae and plants as well as in the glaucophyte *Cyanophora paradoxa* (Price et al. 2012). An additional glaucophyte sequence was recovered by our phylogenetic analysis (fig. 5). Besides transcripts for soluble starch synthases, dinoflagellates also express the granule-bound starch synthase GBSS1 responsible for the synthesis of amylose within the starch grain, as has been described (Patron and Keeling 2005), and several branching enzymes necessary for the introduction of branches into the linear glucan molecules generated by starch synthases (supplementary table 2, Supplementary Material online).

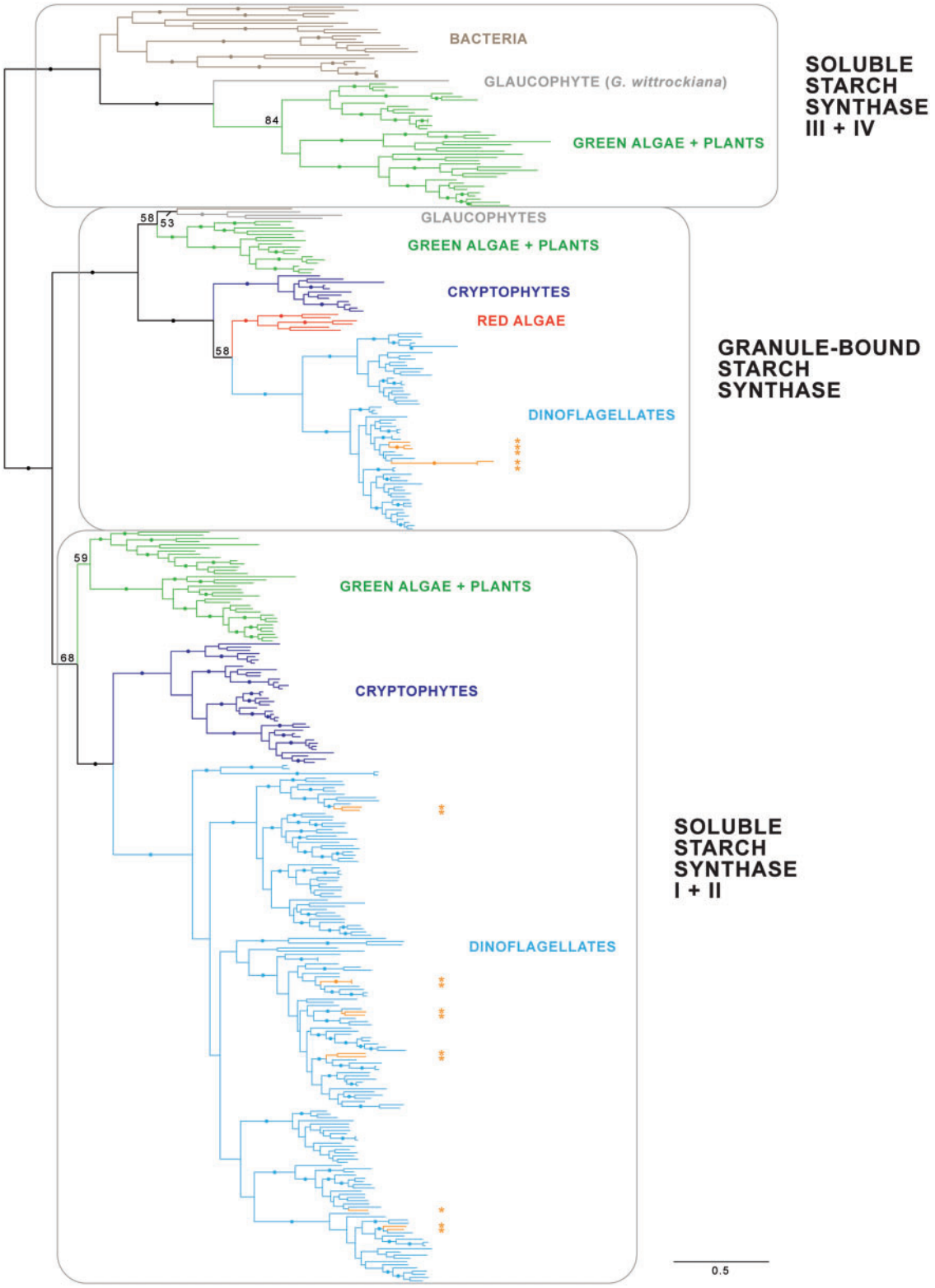

Fig. 5. The host expresses several starch synthases related to the ADP-glucose-specific starch synthases from plants and green algae. Phylogeny of starch synthases as inferred by ML (LG + Γ) and 500 bootstrap replicates, depicting several types of starch synthases in the dinotom host. The granule-bound starch synthase, probably transferred from the cyanobacterial ancestor of the plastid to all three archaeplastidial groups, is still retained in the dinotom host (and all other dinoflagellates) as well as cryptophytes. The SSI–SSII-group soluble starch synthases were thought to have evolved only in green algae and plants from duplication of the granule-bound starch synthase gene, but are also found in form of several paralogs in the dinotom host (and all other dinoflagellates) as well as cryptophytes. No dinoflagellate transcripts were found to cluster with the SSIII–SSIV-group soluble starch synthases. Dots correspond to >95% ML bootstrap support. Otherwise, for increased clarity, only bootstrap support greater than 50% and for deep nodes (= indicating the different algal groups) is shown. Orange tips and asterisks indicate dinotom host-associated transcripts in the tree. The scale bar represents the estimated number of amino acid substitutions per site.

Unlike the green algal and plant homologs, dinoflagellate soluble starch synthases seem to act in the cytosol as they do not possess any signal and transit peptides that would target them to the plastid. This is consistent with the observation of starch granules in the dinoflagellate cytosol (Tomas and Cox 1973). We were unable to find transcripts for ADP-glucose phosphorylase, synthesizing ADP-glucose, in any available dinoflagellate transcriptome, but did find a host-encoded bifunctional triosephosphate isomerase/UDP-glucose pyrophosphorylase (supplementary table 2, Supplementary Material online), supporting an UDP-glucose-dependent starch synthesis in dinoflagellates. The absence of enzymes synthesizing ADP-glucose in dinoflagellates while at the same time they express soluble starch synthases that are related to the ADP-glucose dependent forms in Viridiplantae suggests a change in substrate specificity for these enzymes. Such a change has likely taken place in GBSS1, which is specific for ADP-glucose in extant cyanobacteria but displays dual specificity for both nucleotide sugars in Archaeplastida and cryptophytes (Deschamps et al. 2006; Ball et al. 2013), suggesting a similar evolutionary scenario for soluble starch synthases in dinoflagellates.

Taken together, dinotoms (and dinoflagellates in general) synthesize starch in an unusual way that combines elements of green algal and red algal starch synthesis: they generate starch using enzymes of cyanobacterial origin so far only found in plastids of green algae and plants, but in dinoflagellates those enzymes act in the cytosol—like the host-derived enzymes in red algae. This suggests an interesting hypothesis about starch synthesis in species with secondary plastids: only groups that inherited their secondary plastid from a red alga are able to synthesize starch—implying that starch synthesis, once it has moved to the plastid in the lineage leading to green algae and plants, cannot move back to the cytosol where it is necessary to establish a metabolic connection between host and endosymbiont.

Finally, we investigated the storage polysaccharide metabolism of the endosymbiont, as it was observed that parasitic and symbiotic bacteria tend to lose their glycogen storage ability as a consequence of their intracellular lifestyle (Henrissat et al. 2002; Nowack et al. 2008). The principle storage polysaccharide in diatoms under nutrient replete conditions is the β-1,3-linked glucan chrysolaminaran, which accumulates in the vacuole (Kroth et al. 2008). Neither synthesis nor degradation of this molecule is known in detail, but UDP-glucose likely serves as substrate for chrysolaminaran synthesis, and recent genetic evidence confirmed a rate-limiting function for the UDP-glucose producing enzyme UDP-glucose pyrophosphorylase in the biosynthesis of this storage glucan (Roessler 1987; Zhu et al. 2015). Furthermore, increased endo- and exo-1,3-β-glucanase activity was shown to coincide with chrysolaminaran degradation (Vårum et al. 1986; Chauton et al. 2013). For all enzymes predicted to be involved in either chrysolaminaran biosynthesis or degradation in *P. tricornutum* (Kroth et al. 2008), we could find a respective homolog in the endosymbiont transcriptome. The homologous sequences carry signal peptides and C-terminal transmembrane domains as described for their *P. tricornutum* homologs, supporting the same localization for these enzymes as well (supplementary table 6, Supplementary Material online). Based on these results, we conclude that storage polysaccharide metabolism in the endosymbiont is still functional.

Plastidic Phosphate Translocators Are Present in the Host

One current hypothesis about the origin of plastids is that photosynthate was exported from the nascent primary plastid in form of ADP-glucose using a host-endomembrane-derived nucleotide sugar translocator, which was recruited to the endosymbiont inner membrane (Ball 2014). In extant plants, carbon is exported from the plastid by the triose phosphate translocator (TPT), a member of the plastidic phosphate translocator (pPT) family. This family of transporters, found in red and green algae as well as in the lineages with secondary plastids, was shown to be monophyletic and is thought to have originated from the ancient host-endomembrane translocator involved in establishing the metabolic link (Weber et al. 2006). Members of the pPT family are therefore good candidates for export of photosynthate from the dinotom endosymbiont to the host. It presently appears that no endosymbiont targeting system exists in dinotoms, as suggested by lack of either genomic integration or functional reduction of the endosymbiont. But experiments in yeast have shown that a plant TPT was able to reach internal membranes even when deprived of its transit peptide (Loddenkötter et al. 1993), implying that the ancestral ADP-glucose translocator can have reached the inner membrane of the primary endosymbiont by accident. Similarly, at least one outer membrane protein of extant plant chloroplasts seems to be able to reach the chloroplast surface and be integrated into the membrane without requiring a N-terminal targeting peptide or proteins within the chloroplast (Hofmann and Theg 2005; Bodyl et al. 2009).

Similarity searches for members of the pPT family in *D. baltica* using annotated plant proteins resulted in the identification of several dinoflagellate-associated sequences. Phylogenetic reconstruction with those sequences recovered a tree containing all major groups of plant plastid translocators (triose phosphate translocator—TPT, glucose-6-phosphate translocator—GPT, xylulose-5-phosphate translocator—XPT, and phosphoenolpyruvate translocator—PPT), as well as several clades of red algal, dinoflagellate and diatom phosphate translocators (fig. 6). Neither the transcript sequences nor the tree topology provide sufficient information to identify distinct translocators among the dinoflagellate pPTs, but representatives from all dinoflagellate clades feature the same domain architecture as plant homologs: either an EamA-like transporter family followed by a TPT family domain or a TPT domain only (data not shown).

The three most taxonomically widespread dinoflagellate clades in figure 6 (the dinoflagellate clades labeled 1, 5, and 6) also contain dinotom sequences and transcripts in those clades encode N-terminal extensions with characteristics of dinoflagellate plastid targeting transit-peptides. The N-terminus of the dinoflagellate transporters identified in this study

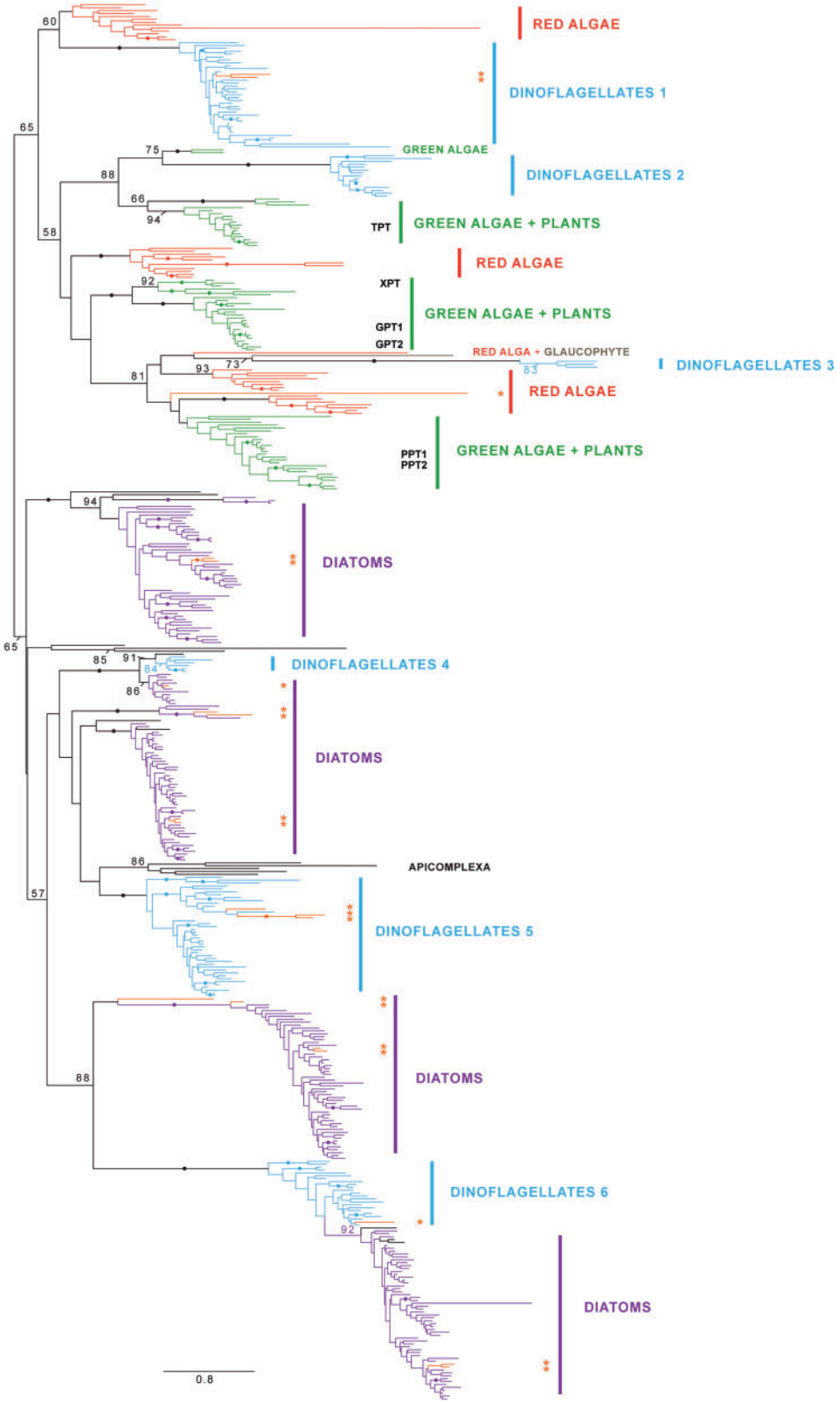

Fig. 6. Host and endosymbiont express several members of the pPTs family. Phylogeny of pPTs as inferred by ML (LG + Γ) and 500 bootstrap replicates, depicting host- and endosymbiont-associated homologs clustering within several dinoflagellate and diatom clades, respectively. Dots correspond to >95% ML bootstrap support. Otherwise, for increased clarity, only bootstrap support greater than 50% and for deep nodes (= indicating the different algal groups) is shown. The position of described *Arabidopsis thaliana* pPTs is indicated by the following abbreviations: TPT = triose phosphate translocator, XPT = xylulose-5-phosphate translocator, GPT ½ = glucose-6-phosphate translocator ½, and PPT1/2 = phosphoenolpyruvate translocator ½. Orange tips and asterisks indicate dinotom host-associated transcripts in the tree. Black tips represent nondiatom stramenopile taxa, unless otherwise indicated. The scale bar represents the estimated number of amino acid substitutions per site.

consists of a TMR that may serve as a signal peptide before the start of the transit peptide which is indicated by the "FVAP" motif that was shown to be the most conspicuous feature of dinoflagellate transit peptides (Patron et al. 2005) (supplementary fig. 7, Supplementary Material online). In dinoflagellates still harboring a photosynthetic peridinin-plastid, those transcripts are likely directed to the plastid, which is also supported by the position of the dinoflagellate clades in the tree with plastid-targeted homologues from other algae (such as red algae and diatoms). Currently we cannot exclude the possibility that the pPT transcripts identified in dinotoms are also targeted to a remnant plastid (possibly represented by their eyespot), similar to the two TPTs in *Plasmodium falciparum* that were hypothesized to provide fuel for apicoplast-located pathways (Mullin et al. 2006), however in that case they would have to function as plastid importers as opposed to exporters. We can therefore only speculate about the function and localization of those transcripts in dinotoms, but a spontaneous, targeting system-independent recruitment of those translocators to another membrane than the peridinin-plastid membranes offers the intriguing possibility that such translocators in dinotoms are connecting endosymbiont and host metabolism.

Conclusion

Based on our analysis of more than 7,000 curated transcripts, the stable endosymbiotic relationship between the dinotom *D. baltica* and its diatom endosymbiont is unlikely to be genetically integrated in the same way as in all other plastid-containing organisms. Rather, the integration of host and endosymbiont metabolism seems to be the defining characteristic of this partnership. The effect of endosymbiosis is markedly stronger in the host, where whole pathways appear to have been lost, while the changes in the endosymbiont are more subtle and harder to identify. Obvious effects on the endosymbiont, such as the loss of the silica cell wall, are barely reflected in the transcriptome—by the putative loss of some but not all paralogs (e.g., silica transporters), but not by the loss of complete gene families or pathways.

Although the dinotom system represents a stable relationship accompanied by some morphological reduction of the endosymbiont, the dinotom endosymbiont defies the classical definition of an organelle in that we were not able to clearly identify a single case of EGT. Further testing is needed to evaluate the possibility that the dinotom endosymbiosis is a relatively recent event that has yet to integrate at the genomic level. For now, this possibility is supported by fossil evidence that places the oldest known diatom fossil (Barron 1993) at a younger age than the oldest known coccolithophorid haptophytes (Bown 1998), which are the closest relatives to the haptophyte plastid in Kareniaceae in recent phylogenies (Ševčíková et al. 2015). Another equally likely possibility is that the development of the dinotom endosymbiosis is simply following a different path, which would further extend upon recent discoveries (Nowack and Grossman 2012; Nakabachi et al. 2014) that blur the distinction between organelle and endosymbiont.

Materials and Methods

Cultures, Media, Growth, and Harvest Conditions

Cultures of *D. baltica* (*Peridinium balticum*) CSIRO CS-38, and *G. foliaceum* CCAP 1116/3 were obtained from the Australian National Algae Culture Collection (CSIRO Marine and Atmospheric Research, Hobart, Australia) and the Culture Collection of Algae and Protozoa (CCAP SAMS Research Services Ltd., Scottish Marine Institute, OBAN, Scotland, UK), respectively. The *D. baltica* culture was maintained in GSe medium at 22 °C in 12:12 light:dark cycles and harvested during the light phase, whereas the *G. foliaceum* culture was maintained in and harvested from f/2-Si medium under the same conditions.

Nucleic Acid Extraction and Purification, Library Construction, Sequencing, and Assembly

Exponentially growing cells (200 ml) were collected by centrifugation (1,000 × g, 10 min, 4 °C), cell pellets were homogenized in TRIzol Reagent (Ambion, lot # 47414) using a glass homogenizer, followed by chloroform extraction. Ethanol-precipitated samples were loaded onto a column of the PureLink RNA Mini Kit (Ambion, lot # 1395746), and washed and eluted according to the kit protocol. Sequencing and assembly of the reads is described in Supplementary Material online.

Annotation and Phylogenetic Pipeline

The assembled and translated *D. baltica* peptides were annotated with the web version of the KAAS (Kanehisa and Goto 2000; Kanehisa et al. 2014) using the bi-directional best-hit method and selecting all protist datasets available in the KEGG GENES database as of November 2014 plus *Homo sapiens* and *Arabidopsis thaliana* as template datasets for orthology assignment. The BLAST-based homology search detected 7,010 genes in our dataset orthologous to genes in the KEGG GENES database. The recovered sequences were used in a phylogenetic pipeline described in the Supplementary Material online to reconstruct single-gene trees that were investigated for their phylogenetic association.

General Identification of Genes and Pathways Investigated

All genes, with exception of the animal-type PMVK, listed in supplementary table 2, Supplementary Material online were identified in the *D. baltica* dataset by KAAS. If a gene was only annotated in either the host or the endosymbiont, the respective tree was searched for possible homologs in either partner of the symbiosis and the identity of the putative homolog was confirmed in a BLASTP search against the nr protein database. In the same way homologs of the dinoflagellates *S. trochoidea* and *Karenia brevis* and the diatom *P. tricornutum* were identified. We identified the animal-type PMVK in *D. baltica* by subjecting the *S. trochoidea* transcriptome to KAAS and using the *S. trochoidea* homolog to query the *D. baltica* dataset. Single-gene trees for the genes in supplementary table 2, Supplementary Material online, with exception of the trees depicted as main figures or

supplementary figures, have been deposited to the Dryad data depository (http://datadryad.org) accession http://dx.doi.org/10.5061/dryad.n654g. Reconstruction of the trees shown in figures 5 and 6 as well as in supplementary figures 1, 2A–E, 3, 4A–C, 5, and 6 is described in the Supplementary Material online.

Localization Prediction

Prediction of signal peptides as part of N-terminal bipartite leader sequences in diatom- as well as in dinoflagellate-derived transcripts was performed with the Hidden Markov Model of SignalP3.0 (Nielsen and Krogh 1998; Bendtsen et al. 2004) using the default truncation setting of 70 residues. Mitochondrial targeting signal were predicted with TargetP 1.1 (Emanuelsson et al. 2000). TMHMM v.2.0 transmembrane helix prediction (Sonnhammer et al. 1998) was used to predict N-terminal transmembrane domains in diatom- and dinoflagellate-derived sequences.

Supplementary Material

Supplementary figures S1–S7 and tables S1–S7 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Compute/Calcul Canada for computing resources and assistance, in particular Calcul Quebec's Mammouth Parallèle 2 and Colosse facilities. We also thank Anna Karnkowska for valuable discussion and help during the revision process. We also want to thank the editor and the reviewers of this manuscript for their constructive comments. This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (grant number 227301), and by a grant from the Tula Foundation to the Centre for Microbial Diversity and Evolution.

References

Allen AE, Dupont CL, Oborník M, Horák A, Nunes-Nesi A, McCrow JP, Zheng H, Johnson DA, Hu H, Fernie AR, et al. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473:203–207.

Armbrust EV, Berges JA, Bowler C, Green BA, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306:79–86.

Ball SG. 2014. Evolution of storage polysaccharide metabolism in Archaeplastida opens an unexpected window on the molecular mechanisms that drove plastid endosymbiosis. In: Löffelhardt W, editor. Endosymbiosis. Vienna: Springer. p. 111–134.

Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C. 2011. The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. *J Exp Bot*. 62:1775–1801.

Ball SG, Subtil A, Bhattacharya D, Moustafa A, Weber APM, Gehre L, Colleoni C, Arias M, Cenci U, Dauvillée D. 2013. Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? *Plant Cell* 25:7–21.

Barron JA. 1993. Diatom. In: Lipps JH, editor. Fossil prokaryotes and protists. Cambridge (MA): Blackwell Scientific. p. 155–167.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 340:783–795.

Bodyl A, Mackiewicz P, Stiller JW. 2009. Early steps in plastid evolution: current ideas and controversies. *BioEssays* 31:1219–1232.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–244.

Bown PR. 1998. Calcareous nannofossil biostratigraphy. London: Chapman & Hall.

Burki F, Imanian B, Hehenberger E, Hirakawa Y, Maruyama S, Keeling PJ. 2014. Endosymbiotic gene transfer in tertiary plastid-containing dinoflagellates. *Eukaryot Cell* 13:246–255.

Chauton MS, Winge P, Brembu T, Vadstein O, Bones AM. 2013. Gene regulation of carbon fixation, storage, and utilization in the diatom *Phaeodactylum tricornutum* acclimated to light/dark cycles. *Plant Physiol*. 161:1034–1048.

Chesnick JM, Cox ER. 1989. Fertilization and zygote development in the binucleate dinoflagellate *Peridinium balticum* (Pyrrhophyta). *Am J Bot*. 76:1060–1072.

Chesnick JM, Kooistra WHCF, Wellbrock U, Medlin LK. 1997. Ribosomal RNA analysis indicates a benthic pennate diatom ancestry for the endosymbionts of the dinoflagellates *Peridinium foliaceum* and *Peridinium balticum* (Pyrrhophyta). *J Eukaryot Microbiol*. 44:314–320.

Chesnick JM, Morden CW, Schmieg AM. 1996. Identity of the endosymbiont of *Peridinium foliaceum* (Pyrrophyta): analysis of the *rbcLS* operon. *J Phycol*. 32:850–857.

Coppin A, Varré J-S, Lienard L, Dauvillée D, Guérardel Y, Soyer-Gobillard M-O, Buleon A, Ball S, Tomavo S. 2005. Evolution of plant-like crystalline storage polysaccharide in the protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. *J Mol Evol*. 60:257–267.

Croft MT, Warren MJ, Smith AG. 2006. Algae need their vitamins. *Eukaryot Cell* 5:1175–1183.

Dauvillée D, Deschamps P, Ral J-P, Plancke C, Putaux J-L, Devassine J, Durand-Terrasson A, Devin A, Ball SG. 2009. Genetic dissection of floridean starch synthesis in the cytosol of the model dinoflagellate *Crypthecodinium cohnii*. *Proc Natl Acad Sci U S A*. 106:21126–21130.

De Martino A, Amato A, Bowler C. 2009. Mitosis in diatoms: rediscovering an old model for cell division. *BioEssays* 31:874–884.

Deschamps P, Haferkamp I, Dauvillée D, Haebel S, Steup M, Buleon A, Putaux JL, Colleoni C, d'Hulst C, Plancke C, et al. 2006. Nature of the periplastidial pathway of starch synthesis in the cryptophyte *Guillardia theta*. *Eukaryot Cell* 5:954–963.

Dodge JD. 1971. Dinoflagellate with both a mesocaryotic and a eucaryotic nucleus.1. Fine structure of nuclei. *Protoplasma* 73:145–157.

Dodge JD. 1984. The functional and phylogenetic significance of dinoflagellate eyespots. *Biosystems* 16:259–267.

Dolah FM, Zippay ML, Pezzolesi L, Rein KS, Johnson JG, Morey JS, Wang Z, Pistocchi R. 2013. Subcellular localization of dinoflagellate polyketide synthases and fatty acid synthase activity. *J Phycol*. 49:1118–1127.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 300:1005–1016.

Figueroa RI, Bravo I, Fraga S, Garcés E, Llaveria G. 2009. The life history and cell cycle of *Kryptoperidinium foliaceum*, a dinoflagellate with two eukaryotic nuclei. *Protist* 160:285–300.

Hehenberger E, Imanian B, Burki F, Keeling PJ. 2014. Evidence for the retention of two evolutionary distinct plastids in dinoflagellates with diatom endosymbionts. *Genome Biol Evol*. 6:2321–2334.

Henrissat B, Deleury E, Coutinho PM. 2002. Glycogen metabolism loss: a common marker of parasitic behaviour in bacteria? *Trends Genet*. 18:437–440.

Hofmann NR, Theg SM. 2005. Chloroplast outer membrane protein targeting and insertion. *Trends Plant Sci*. 10:450–457.

Horiguchi T, Pienaar RN. 1988. Ultrastructure of a new sand-dwelling dinoflagellate, *Scrippsiella arenicola* sp. nov. *J Phycol*. 24:426–438.

Horiguchi T, Takano Y. 2006. Serial replacement of a diatom endosymbiont in the marine dinoflagellate *Peridinium quinquecorne* (Peridiniales, Dinophyceae). *Phycol Res*. 54:193–200.
Houten SM, Waterham HR. 2001. Nonorthologous gene displacement of phosphomevalonate kinase. *Mol Genet Metab*. 72:273–276.
Huysman MJJ, Martens C, Vandepoele K, Gillard J, Rayko E, Heijde M, Bowler C, Inzé D, Van de Peer Y, De Veylder L, et al. 2010. Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biol*. 11:R17.
Imanian B, Carpenter KJ, Keeling PJ. 2007. Mitochondrial genome of a tertiary endosymbiont retains genes for electron transport proteins. *J Eukaryot Microbiol*. 54:146–153.
Imanian B, Keeling PJ. 2007. The dinoflagellates *Durinskia baltica* and *Kryptoperidinium foliaceum* retain functionally overlapping mitochondria from two evolutionarily distinct lineages. *BMC Evol Biol*. 7:172.
Imanian B, Keeling PJ. 2014. Horizontal gene transfer and redundancy of tryptophan biosynthetic enzymes in dinotoms. *Genome Biol Evol*. 6:333–343.
Imanian B, Pombert J-F, Dorrell RG, Burki F, Keeling PJ. 2012. Tertiary endosymbiosis in two dinotoms has generated little change in the mitochondrial genomes of their dinoflagellate hosts and diatom endosymbionts. *PLoS One* 7:e43763.
Imanian B, Pombert J-F, Keeling PJ. 2010. The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS One* 5:e10711.
Janouškovec J, Tikhonenkov DV, Burki F, Howe AT, Kolísko M, Mylnikov AP, Keeling PJ. 2015. 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.
Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 42:D199–D205.
Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28:27–30.
Kite GC, Dodge JD. 1985. Structural organization of plastid DNA in two anomalously pigmented dinoflagellates. *J Phycol*. 21:50–56.
Kroth PG, Chiovitti A, Gruber A, Martin Jézéquel M, Mock T, Schnitzler Parker M, Stanley MS, Kaplan A, Caron L, Weber T, et al. 2008. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS One* 3:e1426.
Kröger N, Bergsdorf C, Sumper M. 1994. A new calcium binding glycoprotein family constitutes a major diatom cell wall component. *Embo J*. 13:4676–4683.
Lin S, Cheng S, Song B, Zhong X, Lin X, Li W, Li L, Zhang Y, Zhang H, Ji Z, et al. 2015. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* 350:691–694.
Loddenkötter B, Kammerer B, Fischer K, Flugge UI. 1993. Expression of the functional mature chloroplast triose phosphate translocator in yeast internal membranes and purification of the histidine-tagged protein by a single metal-affinity chromatography step. *Proc Natl Acad Sci U S A*. 90:2155–2159.
Lohr M, Schwender J, Polle JEW. 2012. Isoprenoid biosynthesis in eukaryotic phototrophs: a spotlight on algae. *Plant Sci*. 185–186:9–22.
Massé G, Belt ST, Rowland SJ, Rohmer M. 2004. Isoprenoid biosynthesis in the diatoms *Rhizosolenia setigera* (Brightwell) and *Haslea ostrearia* (Simonsen). *Proc Natl Acad Sci U S A*. 101:4413–4418.
Mullin KA, Lim L, Ralph SA, Spurck TP, Handman E, McFadden GI. 2006. Membrane transporters in the relict plastid of malaria parasites. *Proc Natl Acad Sci U S A*. 103:9572–9577.
Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S-Y. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr Biol*. 24:R640–R641.
Nielsen H, Krogh A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol*. 6:122–130.
Nowack ECM, Grossman AR. 2012. Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci U S A*. 109:5340–5345.
Nowack ECM, Melkonian M, Gloeckner G. 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr Biol*. 18:410–418.
Pandini V, Caprini G, Thomsen N, Aliverti A, Seeber F, Zanetti G. 2002. Ferredoxin-NADP+ reductase and ferredoxin of the protozoan parasite *Toxoplasma gondii* interact productively *in vitro* and *in vivo*. *J Biol Chem*. 277:48463–48471.
Patron NJ, Keeling PJ. 2005. Common evolutionary origin of starch biosynthetic enzymes in green and red algae. *J Phycol*. 41:1131–1141.
Patron NJ, Waller RF, Archibald JM, Keeling PJ. 2005. Complex protein targeting to dinoflagellate plastids. *J Mol Biol*. 348:1015–1024.
Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. *J Mol Biol*. 357:1373–1382.
Patron NJ, Waller RF. 2007. Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *BioEssays* 29:1048–1058.
Pienaar RN, Sakai H, Horiguchi T. 2007. Description of a new dinoflagellate with a diatom endosymbiont, *Durinskia capensis* sp. nov. (Peridiniales, Dinophyceae) from South Africa. *J Plant Res*. 120:247–258.
Price DC, Chan CX, Yoon HS, Yang EC, Qiu H, Weber APM, Schwacke R, Gross J, Blouin NA, Lane C, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 335:843–847.
Roessler PG. 1987. UDPglucose pyrophosphorylase activity in the diatom *Cyclotella cryptica*. Pathway of chrysolaminarin biosynthesis. *J Phycol*. 23:494–498.
Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, et al. 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Curr Biol*. 23:1399–1408.
Sonnhammer EL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol*. 6:175–182.
Sumper M, Kröger N. 2004. Silica formation in diatoms: the function of long-chain polyamines and silaffins. *J Mater Chem*. 14:2059–2065.
Szydlowski N, Ragel P, Raynaud S, Lucas MM, Lucas I, Montero M, Muñoz FJ, Ovecka M, Bahaji A, Planchot V, et al. 2009. Starch granule initiation in *Arabidopsis* requires the presence of either class IV or class III starch synthases. *Plant Cell* 21:2443–2457.
Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S, Jenkins J, Schmutz S, Přibyl P, Fousek J, et al. 2015. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep*. 5:10134.
Tanaka A, De Martino A, Amato A, Montsant A, Mathieu B, Rostaing P, Tirichine L, Bowler C. 2015. Ultrastructure and membrane traffic during cell division in the marine pennate diatom *Phaeodactylum tricornutum*. *Protist* 166:506–521.
Tengs T, Dahlberg OJ, Shalchian-Tabrizi K, Klaveness D, Rudi K, Delwiche CF, Jakobsen KS. 2000. Phylogenetic analyses indicate that the 19'Hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol Biol Evol*. 17:718–729.
Tippit DH, Pickettheaps JD. 1976. Apparent amitosis in the binucleate dinoflagellate *Peridinium balticum*. *J Cell Sci*. 21:273–289.
Tomas RN, Cox ER. 1973. Observations on the symbiosis of *Peridinium balticum* and its intracellular alga. I. Ultrastructure. *J Phycol*. 9:304–323.
Tomas RN, Cox ER, Steidinger KA. 1973. *Peridinium balticum* (Levander) Lemmermann, an unusual dinoflagellate with a mesocaryotic and an eucaryotic nucleus. *J Phycol*. 9:91–98.
Vårum KM, Østgaard K, Grimsrud K. 1986. Diurnal rhythms in carbohydrate metabolism of the marine diatom *Skeletonema costatum* (Grev.) Cleve. *J Exp Mar Biol Ecol*. 102:249–256.
Viola R, Nyvall P, Pedersén M. 2001. The unique features of starch metabolism in red algae. *Proc Biol Sci*. 268:1417–1422.

Vollmer M, Thomsen N, Wiek S, Seeber F. 2001. Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast-localized, plant-type ferredoxin-NADP(+) reductase and ferredoxin. *J Biol Chem*. 276:5483–5490.

Weber APM, Linka M, Bhattacharya D. 2006. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. *Eukaryot Cell* 5:609–612.

Xiang T, Nelson W, Rodriguez J, Tolleter D, Grossman AR. 2015. *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.

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

Zhu B-H, Shi H-P, Yang G-P, Lv N-N, Yang M, Pan K-H. 2015. Silencing UDP-glucose pyrophosphorylase gene in *Phaeodactylum tricornutum* affects carbon allocation. *New Biotechnol*. 33:237–244.