View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by OceanRep

EC

Journals.ASM.org

Endosymbiotic Gene Transfer in Tertiary Plastid-Containing Dinoflagellates

Fabien Burki,[a] Behzad Imanian,[a] Elisabeth Hehenberger,[a] Yoshihisa Hirakawa,[a]* Shinichiro Maruyama,[b,c]* Patrick J. Keeling[a]

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada[a]; Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada[b]; Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada[c]

Plastid establishment involves the transfer of endosymbiotic genes to the host nucleus, a process known as endosymbiotic gene transfer (EGT). Large amounts of EGT have been shown in several photosynthetic lineages but also in present-day plastid-lacking organisms, supporting the notion that endosymbiotic genes leave a substantial genetic footprint in the host nucleus. Yet the extent of this genetic relocation remains debated, largely because the long period that has passed since most plastids originated has erased many of the clues to how this process unfolded. Among the dinoflagellates, however, the ancestral peridinin-containing plastid has been replaced by tertiary plastids on several more recent occasions, giving us a less ancient window to examine plastid origins. In this study, we evaluated the endosymbiotic contribution to the host genome in two dinoflagellate lineages with tertiary plastids. We generated the first nuclear transcriptome data sets for the "dinotoms," which harbor diatom-derived plastids, and analyzed these data in combination with the available transcriptomes for kareniaceans, which harbor haptophyte-derived plastids. We found low level of detectable EGT in both dinoflagellate lineages, with only 9 genes and 90 genes of possible tertiary endosymbiotic origin in dinotoms and kareniaceans, respectively, suggesting that tertiary endosymbioses did not heavily impact the host dinoflagellate genomes.

The process of endosymbiosis led to some of the most dramatic turns in the evolution of life by giving rise to plastids, the light-gathering organelles of plants and algae (1). Plastid evolution began with the so-called "primary endosymbiosis" between a heterotrophic eukaryote and cyanobacteria, and almost all plastids are derived either directly or indirectly from this pivotal event, most likely in the common ancestor of the Plantae supergroup (2; see reference 3 for a different scenario). Subsequently, other eukaryotic lineages acquired plastids through endosymbioses with primary algae, a process known as secondary endosymbiosis (4). The exact number of secondary endosymbioses is debated, but it is evident that secondary plastids originated more than once, because plastids of green and red algal origin have spread across the eukaryotic tree (5). The evolution of red algal plastids is particularly contentious, and numerous scenarios involving serial transfers between the major "red lineages" rather than ancestral inheritance were recently put forward (for examples, see references 6 to 8). The complex history of plastids is well illustrated in one lineage, the dinoflagellates (1, 9). Not all dinoflagellate species possess a plastid, or at least not a photosynthetically active one, but it has become increasingly clear that their ancestor already had a plastid of red algal origin (10, 11). Most photosynthetic dinoflagellates still harbor this original plastid, which is surrounded by 3 membranes and is characterized by chlorophylls *a* and *c* and the pigment peridinin (12), but others have replaced it with green algal plastids through serial endosymbiosis (13, 14). On several occasions, dinoflagellates have also replaced their original peridinin plastid with plastids derived from other algae, themselves harboring red plastids, in a process referred to as tertiary endosymbiosis (1, 9).

Regardless of the level of endosymbiosis, one common theme of plastid establishment is that plastid genomes are highly reduced compared to their modern-day free-living cyanobacterial relatives, encoding only a small fraction of the proteins required for full functionality. The "missing" genes in plastid genomes are believed to have been either lost or transferred to the host genomes during the course of plastid acquisition, through the process of endosymbiotic gene transfer (EGT) (15). To sustain the plastid needs, some of these endosymbiotic proteins are targeted back to the plastid where they function, together with host-derived proteins and proteins of various bacterial origins other than cyanobacteria, such as chlamydiae (16–21). The impact of EGT in primary photosynthetic eukaryotes has been assessed in an evolutionary framework, using phylogenomics to identify cyanobacterial genes in host nuclear genomes. Various amount of EGT were found in all main Plantae lineages, ranging from 132 cyanobacterial genes in the glaucophyte *Cyanophora paradoxa* (22) to 295 in the red alga *Porphyridium purpureum* (16), 897 in the green alga *Chlamydomonas reinhardtii* (23), and as high as about 4,500 in the land plant *Arabidopsis thaliana* (24).

In contrast to primary endosymbioses, secondary endosymbioses added extra layers of complexity to the EGT inference because host genomes have potentially integrated genes from the

Received 14 November 2013 Accepted 24 November 2013

Published ahead of print 2 December 2013

Address correspondence to Fabien Burki, burkif@mail.ubc.ca, or Patrick J. Keeling, pkeeling@mail.ubc.ca.

* Present address: Yoshihisa Hirakawa, Faculty of Life and Environmental Sciences, University of Tsukuba, Tennodai, Tsukuba, Japan; Shinichiro Maruyama, Division of Environmental Photobiology, National Institute for Basic Biology, Okazaki, Japan.

F.B. and B.I. contributed equally to this article.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/EC.00299-13.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/EC.00299-13

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

secondary plastid but also from the nuclear genome of the algal endosymbiont, which was itself the recipient of endosymbiotic genes during primary endosymbiosis (4). Despite these difficulties, genome-scale sequencing has begun to reveal the footprint of endosymbiosis in taxa with secondary plastids. In the rhizarian *Bigelowiella natans*, which possesses a green algal plastid, 353 genes were deemed to have an algal origin. As expected, the majority was derived from the green lineage (207), but 45 genes displayed a red algal origin (18), consistent with earlier reports based on transcribed genes (25). Large numbers of endosymbiotic genes were also inferred in red algal plastid-containing species, such as the diatom *Phaeodactylum tricornutum* (171 red genes) (26), the phaeophycean *Ectocarpus siliculosus* (>600 red genes) (27), the cryptophyte *Guillardia theta* (100 red genes) (18), and the chromerid *Chromera velia* (263 red genes) (28). Similar to the case with *B. natans*, however, the nuclear genome of these lineages also contained genes with a green algal ancestry, and this green signal often dominated the red signal (18, 27, 28). The most striking example relates to the diatoms, in which over 1,700 genes, representing 16% of all nuclear genes, were recently reported to be of green algal origin (29). A signal of such magnitude was interpreted as deriving from a cryptic green algal endosymbiont predating the acquisition of the current red algal plastid (29). This genomic footprint of endosymbiosis has also been claimed to remain long after plastid loss. Oomycetes and ciliates are two plastid-lacking lineages that are possibly derived from photosynthetic ancestors (for examples, see reference 30), so the finding of 855 and 16 genes reported to have a red algal origin, respectively, was taken as supporting a photosynthetic past for these taxa (31, 32).

However, confidently detecting EGT is a complicated task because most endosymbioses represent very ancient events, likely taking place more than 1,000 million years ago (33, 34). Moreover, the current scarce sampling of genomic data for primary algae often prevents unambiguous identification of the origin of the transferred genes, challenging large-scale assessments of EGT. One way to increase the confidence in evaluating the genomic impact of EGT is to look at more recent endosymbiotic events, for which the origins of individual genes should be easier to infer. The independent primary endosymbiosis in the rhizarian *Paulinella chromatophora*, which happened about 60 million years ago (35), is one such example and has allowed parallels to be drawn with the ancestor of Plantae. In this case, the endosymbiont genome contains only 867 protein-coding genes (about 3-fold fewer than the closest free-living cyanobacterial genome), and this reduction was accompanied by more than 30 EGT events so far identified (36–38). Tertiary endosymbioses in dinoflagellates represent another class of relatively recent events that offer a good opportunity to look at EGT, but this time in the context of eukaryote-to-eukaryote gene transfer. Dinoflagellates have been shown to possess tertiary plastids of haptophyte (39–41) and diatom (42, 43) origins in the so-called Kareniaceae and "dinotom" groups, respectively. The genus *Dinophysis* also possesses tertiary endosymbionts, which are derived from cryptophytes, but it is uncertain whether the host cells permanently maintain them or they represent only transient association (44, 45).

In the case of kareniaceans, initial transcriptomic surveys of *Karenia brevis* and *Karlodinium veneficum* identified several nucleus-encoded plastid-targeted proteins of endosymbiotic origin (46, 47). As expected, the majority of these plastid-targeted proteins were ultimately derived from the haptophyte endosymbiont, either replacing ancestral endosymbiotic copies transferred at the time of the original red algal plastid (48) or constituting novel acquisitions. These studies also pointed to 3 genes in *K. veneficum* retained from the ancient red plastid to service the new haptophyte plastid (47) and 6 genes in *K. brevis* originally derived from green, not red, algae (46). Neither study reported any haptophyte-derived genes that are not currently involved in plastid function, and for the dinotoms, virtually nothing is known of the degree of genetic integration of the diatom plastid. Given the potential of endosymbioses and EGT to remodel nuclear genomes, and the methodological difficulties associated with detecting EGT in ancient primary and secondary endosymbioses, we used in this study a phylogenomic approach to evaluate the more recent endosymbiotic contribution to the host genome in the kareniaceans *K. brevis* and *K. veneficum* and the dinotoms *Durinskia baltica* and *Kryptoperidinium foliaceum*. This approach allowed us to exploit a key advantage tertiary plastids offer in the study of plastid evolution: in both kareniaceans and dinotoms, the host and endosymbiont lineages are extant and relatively well characterized phylogenetically, facilitating the detection of EGT.

MATERIALS AND METHODS

Culturing, cDNA preparation, and sequencing. Cultures of *D. baltica* (*Peridinium balticum*) CSIRO CS-38 and *K. foliaceum* CCMP 1326 were obtained from the CSIRO Microalgae Supply Service (CSIRO Marine and Atmospheric Research Laboratories, Tasmania, Australia) and from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME), respectively. *D. baltica* culture was maintained in GSe medium (49) at 22°C (12-h/12-h light-dark cycle).

Exponentially growing cells were collected and ground as described previously (50). Cells lysis, RNA extractions, precipitations and purifications were performed for both species as described earlier (51). Total RNA for reverse transcription-PCR (RT-PCR) was obtained as described earlier (50). An RNeasy MinElute cleanup kit (Qiagen, Mississauga, Ontario, Canada) was utilized to clean up the total RNA after DNase treatment according to the manufacturer's instructions. An Oligotex mRNA minikit (Qiagen) was used to purify poly(A) RNA from approximately 25 μg of cleaned-up total RNA based on the manufacturer's instructions. Approximately 500 ng of poly(A) RNA from *D. baltica* and *K. foliaceum* was used as the template for constructing first- and second-strand cDNA with a Just cDNA double-stranded cDNA synthesis kit (Agilent Technologies Canada, Mississauga, Ontario, Canada) according to the manufacturer's protocol, with one modification: instead of oligo(dT) and random 9-mer primers, a dinoflagellate-specific splice leader (SL) primer (5′-CCGTAG CCATTTTGGCTCAAG-3′) was used. The resulting double-stranded cDNA samples were amplified through PCR and/or long-range PCR with the SL primer in conjunction with the random 9-mer primer. The amplified cDNA samples were purified using a QIAquick PCR purification kit (Qiagen) and reamplified once more through PCR and/or long-range PCR. The optimized PCR conditions were determined to be 94°C for 2 min, 39 cycles of 94°C for 15 s, 42°C for 30 s, 72°C for 5 min, and 72°C for 6 min, while the long-range PCR conditions were optimized at 92°C for 2 min, 34 cycles of 94°C for 10 s, 45°C for 15 s, 68°C for 20 min, and 68°C for 7 min using buffer 3 from an Expand Long Template PCR System kit (Roche Applied Science, Indianapolis, IN).

The amplified SL cDNAs of *D. baltica* and *K. foliaceum* were sequenced using massively parallel GS-FLX DNA pyrosequencing (Roche 454 Life Sciences, Branford, CT), which was carried out at the Génome Québec Innovation Centre. This pyrosequencing produced totals of 553,695 and 735,618 reads with average lengths of 351 bp and 267 bp for *D. baltica* and *K. foliaceum*, respectively. The reads were assembled *de novo* using gsAssembler 2.5p1 (formerly known as Newbler), edited, and reassembled with CONSED 23 (52) to remove the misaligned reads. The final assembly contained 65% and 67% of all the

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

reads that were assembled into 5,625 and 717 large contigs for *D. baltica* and *K. foliaceum*, respectively.

For *K. brevis* and *K. veneficum*, 65,266 and 17,434 expressed sequence tags (ESTs) were downloaded from GenBank, respectively, and remaining vectors and poly(A) were removed with seqclean (http://compbio.dfci.harvard.edu/tgi/software/). iAssembler (53) was used to assemble the ESTs into clusters, leading to 24,696 unique sequences for *K. brevis* and 11,798 unique sequences for *K. veneficum*.

Phylogenomic pipeline. All nucleotide sequences were translated into amino acids with OrfPredictor using BLASTX against Swissprot to identify coding regions (54). To make a set of unique queries for dinotoms and Kareniaceae, reciprocal blast was performed between *D. baltica* and *K. foliaceum* and between *K. brevis* and *K. veneficum*; when best reciprocal hits were available, only the longest sequence of each copy was retained, which led to 6,210 and 16,207 unique queries for dinotoms and Kareniaceae, respectively (Fig. S1 in the supplemental material describes the full pipeline). These unique queries were used to search with BLASTP a curated in-house protein database (see Table S1 in the supplemental material for the complete list of taxa) that was subjected to CDHIT (55) to remove redundant sequences and close paralogs in order to simplify interpretations of the downstream phylogenetic trees. The BLAST search was followed by a hit-parsing step to retrieve the corresponding sequences, with two initial conditions: (i) to reduce the possibility of retrieving distantly related paralogs and/or sequences with short matches (e.g., protein domain), a stringent E value threshold ($\leq$1e$-$25) and query coverage ($\geq$50%) were mandatory; (ii) to reduce the number of prokaryotic taxa entering the alignments, a maximum of 8 cyanobacteria and 4 taxa in the other prokaryotic groups was allowed, as defined in Table S1. This procedure generated 1,982 and 4,117 fasta files containing at least 5 sequences for dinotoms and Kareniaceae, respectively, which were subjected to a first round of alignment using MAFFT in auto mode (56), poorly aligned site-trimming using TRIMAL (57), and rapid maximum likelihood (ML) tree reconstructions using FastTree and the WAG model of evolution (58) (see Fig. S1). In order to build the final trees with a more rigorous approach, a dereplication strategy was applied to reduce the complexity of the alignments using TreeTrimmer (18, 59). Briefly, this method looks for statistically supported clades in phylogenetic trees and retains only representative taxa in these clades, reducing the overall number of taxa while maintaining the global diversity. The FastTree trees served as the input for this procedure, with a cutoff of 0.9 Shimodaira-Hasegawa support to define a monophyletic assemblage (see Table S6 in the supplemental material for details of the conditions). After dereplication, sequences for the retained taxa were realigned with the FFTNSI algorithm in MAFFT (56), and final trees were reconstructed using RAxML in combination with the LG model of evolution and 100 bootstrap replicates (60). The phylogenetic affinities of the query sequences were determined with a tree sorting PERL script used previously (61), on a prefiltered set of trees containing a minimum of 8 taxa and 3 different taxonomic groups. After this sorting, trees in each affinity bin were further clustered when one or more query sequences overlapped (using the Tree Cluster Tool in PhyloSort [23]).

Localization prediction. In dinoflagellates with peridinin plastids that are surrounded by three membranes, nucleus-encoded plastid proteins possess an N-terminal bipartite sequence consisting of a signal peptide followed by a transit peptide (62, 63), and they are consequently directed to plastids via the endomembrane system (64). N-terminal signal peptides have also been predicted in plastid proteins of tertiary plastid-containing dinoflagellates, which contain 4 plastid membranes (47, 65). In the present study, the subcellular localization of 166 haptophyte-derived sequences in kareniaceans and 17 diatom-derived sequences in dinotoms was performed with 4 presequence prediction programs, which all predict whether sequences contain an N-terminal signal peptide or mitochondrial targeting peptide: PredSL (66), TargetP (67), Predotar (68), and iPSORT (69). A "nonplant" setting was applied for all programs. Partial sequences were removed from the prediction analysis based on alignments with homologous proteins provided by BLASTP searches. Both plastid-targeted and endoplasmic reticulum (ER)- or Golgi-targeted proteins were predicted by their N-terminal signal peptides, and endomembrane proteins were manually discriminated from plastid ones based on their putative functions. Mitochondrial proteins have mitochondrial targeting peptides. The results are indicated in Tables S2 and S4 in the supplemental material.

5′ RACE. Total RNA of *D. baltica* was obtained and purified as described above. RNA was processed for rapid amplification of 5′ cDNA ends (5′ RACE), and RT-PCR was performed using the FirstChoice RLM-RACE kit (Ambion, Burlington, Ontario, Canada) according to the manufacturer's instructions. 5′ ends of genes of interest were amplified in a nested PCR employing two nested primers provided by the kit and two nested antisense primers specific to the genes of interest (see Table S7 in the supplemental material). Nested PCR products were sequenced and assembled with the corresponding contigs.

Accession numbers. The transcriptome shotgun assembly projects have been deposited at DDBJ/EMBL/GenBank under accession numbers GAAT00000000 and GABR00000000. The versions described in this paper are the first versions, GAAT01000000 and GABR01000000. The sequences for nested PCR products assembled with the corresponding contigs have been deposited at DDBJ/EMBL/GenBank under accession numbers KC878014 to KC878022.

RESULTS AND DISCUSSION

Phylogenetic inferences from dinotoms and kareniaceans. Dinotom plastids are complex and have preserved several features normally lost after the endosymbiotic event. In addition to the plastid itself, the endosymbiont has retained its nucleus, cytosol, cytosolic ribosomes, endoplasmic reticulum, and even mitochondria, all separated from the host components by a single membrane (70–74). The presence of a large endosymbiont nucleus actively transcribing genes means that the total RNA pool in dinotoms includes a mixture of dinoflagellate and diatom transcripts. To significantly enrich the *D. baltica* and *K. foliaceum* cDNA libraries for nucleus-encoded dinoflagellate transcripts, we took advantage of the spliced leader (SL) sequence that occupies the 5′ end of dinoflagellate mRNA molecules by using an SL primer in the cDNA construction and cDNA amplification steps (see Materials and Methods). This short SL sequence is *trans*-spliced from a small noncoding RNA (SL RNA) to the splice acceptor site in the 5′ untranslated region of pre-mRNAs and has been found in all dinoflagellate species studied to date (75–79). Both dinotom cDNA libraries were sequenced by 454 pyrosequencing. The *D. baltica* 454 sequences were assembled into 5,625 contigs with an average length of 1,082 nucleotides. Constructing and amplifying the SL cDNA library for *K. foliaceum* proved to be more challenging, and we were able to assemble the 454 reads to only 717 contigs with an average length of 506 nucleotides. To generate a set of unique dinotom queries for our phylogenomic pipeline, we further removed the shortest copies of homologous genes occurring in both data sets, as defined by reciprocal best blast hits, which led to 6,210 contigs (see Fig. S1 in the supplemental material). The same general approach was employed for kareniaceans, but in this case the initial data sets were constituted by the 65,266 and 17,434 available EST sequences for *K. brevis* and *K. veneficum*, respectively, which were assembled into 24,696 and 11,798 contigs; the clustering into unique queries led to 16,207 contigs larger than 500 nucleotides (see Fig. S1).

These 6,210 and 16,207 contigs were used as queries in blast searches and several filtering steps to retrieve homologs from a broad sampling of all major eukaryotic and prokaryotic groups,

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

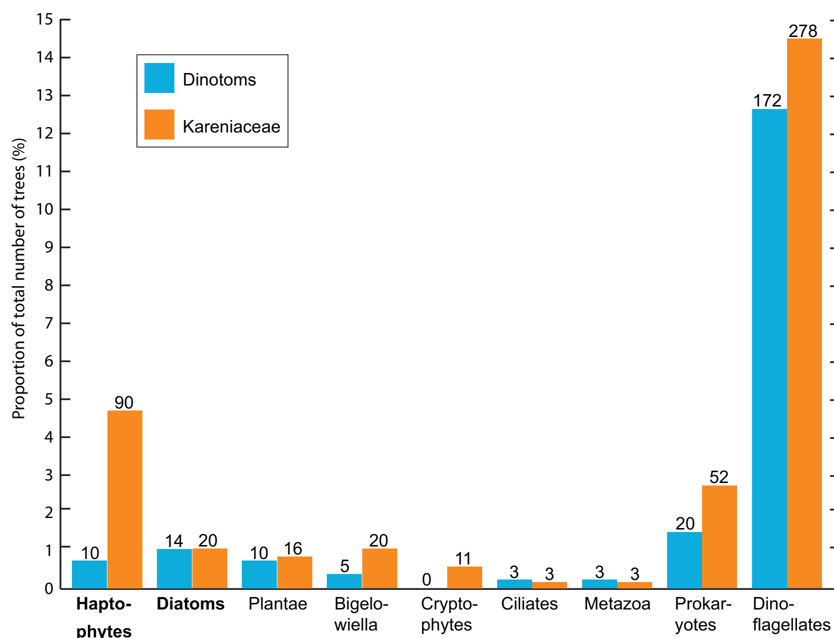

FIG 1 Taxonomic affinities in dinotoms and kareniaceans. A bar chart shows the proportion of genes corresponding to the sorted tree clusters and their inferred origin; bar height is relative to the total number of sorted trees for dinotoms and kareniaceans, and the raw counts of tree clusters are indicated on the bars. Trees not counted here include uninformative trees with respect to the sorting criteria discussed in the text.

including 13 dinoflagellates, 4 haptophytes, and 4 diatoms (see Table S1 in the supplemental material). Maximum likelihood phylogenetic trees were reconstructed for 1,972 and 2,930 sequence alignments corresponding to dinotom and kareniacean queries, respectively, which all included at least 5 homologous sequences. In order to increase the phylogenetic value of these trees, the sorting procedure that determined the phylogenetic affinity of each query sequence was applied only to the trees including at least 8 sequences and 3 taxonomic groups, which corresponded to 1,363 trees for dinotoms and 1,923 trees for kareniaceans (see Fig. S1 and Table S1). Importantly, this taxonomic binning strategy relied on monophyletic groupings between the query sequences and other taxa supported by more than 80% bootstrap support; as such, it does not necessarily constitute an inference of the gene origins in dinotoms and kareniaceans *per se* but rather a simpler measure of the taxonomic affinities of a large part of their transcriptome. The idea is that if the endosymbiotic integration in both systems was accompanied by a massive gene flow to the host nuclear genome, as generally acknowledged (17, 22, 24, 29), then the dominant foreign taxonomic signal is expected to be that of diatoms in dinotoms and haptophytes in kareniaceans.

Endosymbiotic signal in dinotoms and kareniaceans. Several factors can influence a phylogenetic reconstruction, such as poor taxon sampling, lineage sorting, lack of genuine signal, compositional biases, or extreme rate variation among species potentially resulting in the long branch attraction artifact (80–83). At the scale of our phylogenomic approach, these drawbacks will impact a significant proportion of all reconstructed trees and produce artifactual relationships. Accordingly, to define a phylogenetic "noise" baseline above which the endosymbiotic signal should rise if induced by large amounts of transferred genes, the query genes were also assigned to various other taxonomic bins (i.e., Plantae, Bigelowiella, Cryptophytes, Ciliates, and Metazoa). As expected, genes from both dinotoms and kareniaceans showed affinities to these groups, but they represented low proportions, generally below 1% of the sorted trees (Fig. 1). Thus, even if some cases might represent real gene transfers, we favor the interpretation that most are phylogenetic artifacts and as such define the expected level of background noise in the data. In addition, a "dinoflagellate" category and a "prokaryote" category were evaluated as further controls. Here, the expectations were that most genes should show a dinoflagellate affinity—since the host is a dinoflagellate—and a relatively large proportion should have a prokaryote affinity due to the accepted notion that protist genomes have acquired numerous horizontally transferred genes from bacteria (84). Not surprisingly, the vast majority of the sorted trees were indeed consistent with a common origin between the query genes and other dinoflagellates, with 12.6% and 14.5% in dinotoms and kareniaceans, respectively (Fig. 1). Similarly, the prokaryotic component in both groups appeared above the baseline, although not by large margins, with 1.5% in dinotoms and 2.7% in kareniaceans, which corresponded to the figures obtained in a recent investigation of another dinoflagellate taxon (*Alexandrium tamarense*) (85).

More importantly, the number of genes with haptophyte affinity in kareniaceans was clearly above the baseline, with 163 trees recovering a monophyletic grouping between either *K. brevis* or *K. veneficum* and haptophytes (see Tables S2 and S3 in the supplemental material). A detailed analysis revealed that several trees in fact corresponded to the same genes; for example, 48 trees corresponded to an isoform of the chloroplast light-harvesting complex. A high copy number for many genes is a common feature in dinoflagellates (86–89), so the overlapping trees recovered in this study most likely constituted lineage-specific gene duplication and thus were grouped into 90 distinct clusters to prevent the recursive counting of highly similar copies. Still, these 90 tree clus-

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

ters (4.7% of the 1,923 sorted trees) represented an excess of genes with shared ancestry between haptophytes and kareniaceans (Fig. 1). As expected, this pool included proteins likely targeted to the plastids but also included proteins tentatively assigned to function in the mitochondria, the endomembrane system, or without robust subcellular prediction, some of which might be cytoplasmic (see Table S2 and Materials and Methods). EGT has been shown to be involved in remodeling functions beyond those of the plastids in primary and secondary plastid-containing lineages (18, 22, 24), and our results show that they also likely impacted tertiary endosymbiosis, revisiting a recent assessment in *K. veneficum* where no clear evidence of haptophyte-derived mitochondrial genes was found (90).

In the dinotoms, the diatom contribution to the host genome was more modest: only 17 trees, which were grouped into 14 tree clusters (1.02% of the 1,363 sorted trees), showed a diatom affinity (Fig. 1; see also Tables S4 and S5 in the supplemental material). Even though we used the dinoflagellate-specific SL primer during cDNA preparation, a fraction of transcripts encoded in the diatom genome is expected to be found in our data sets due to the presence of very large amounts of transcriptionally active endosymbiotic DNA. The possibility that these 14 clusters in fact represent genes encoded in the diatom genome was investigated by 5′ RACE experiments to assess the presence or absence of the dinoflagellate SL signature. Despite several attempts, 5 reactions failed to produce a complete fragment, but of the 9 complete products that were obtained (see Table S4), none contained the SL sequence. At face value, this result suggests that all 14 clusters most likely corresponded to genes residing in the diatom DNA and thus were not instances of endosymbiotic gene transfers. However, the intriguing possibility exists that SL capping in dinoflagellates might not be a universal feature, an observation also reported for other dinoflagellate species (79, 91).

To clarify the situation, we analyzed the GC contents of the 14 clusters and found that they fall into two distinct categories: five clusters were in the range of 45% to 53% GC, whereas the other 9 clusters were all above 58% GC (Fig. 2A; see also Table S4 in the supplemental material). The GC content at the third codon positions (GC3) showed an even more pronounced difference, with values remaining between 45% and 53% for the five clusters with lower GC percentages but ranging between 71% and 88% for the nine clusters with higher GC content (Fig. 2A). Because most amino acids allow silent substitutions in the third codon position (with the exception of methionine and tryptophan), the mutational pressure strongly influences the GC content at these positions, and it has been shown that the overall GC content of a genome is highly correlated to the GC3 of its genes (92, 93). Moreover, comparing the GC contents of all transcripts from three diatom genomes (*Phaeodactylum tricornutum*, *Fragilariopsis cylindrus*, and *Thalassiosira pseudonana*) with the whole transcriptome of *D. baltica* was consistent with our observation: most diatom transcripts were below 55% GC, whereas the *D. baltica* transcripts were almost entirely above this mark (Fig. 2B). Taken together, these results suggest that there is a significant difference in GC content between the host and endosymbiont genomes in dinotoms, which confirms a previous statement based on only three genes (94), and that the 9 high-GC clusters could be encoded in the dinoflagellate nuclear genome, in spite of their apparent lack of SL sequence.

Impact of EGT in dinotoms and kareniaceans. Our phylogenomic analysis of the nuclear transcriptome of tertiary plastid-containing dinoflagellates identified 9 and 90 cases of possible EGT in dinotoms and kareniaceans, respectively. To understand the global impact of gene transfer on eukaryotic genomes associated with plastid endosymbioses, these figures should be put in perspective to other recently investigated instances of EGT. Comparing the situation in different lineages is a complex task, because the methods of detection and the data sets (e.g., full genomes versus partial transcriptomes) are often different. However, most studies have one characteristic in common: they relied on shared ancestry between putative donor and recipient taxa as a proxy to infer EGT (18, 23, 29, 95). The same approach was used in this study, but in the case of tertiary endosymbioses, the relatively younger age of the transferred genes increased the phylogenetic power for detecting genes with haptophyte affinity in kareniaceans and diatom affinity in dinotoms.

In this context, the 90 haptophyte genes inferred in kareniaceans, even more so the 9 diatom genes in dinotoms, may seem strangely small in comparison to the much larger algal signals generally observed in other eukaryotes (16, 23, 24, 27, 29, 31). However, these figures nicely fit within a series of recent reevaluations of the impact of EGT on nuclear genomes, which have systematically lowered the endosymbiotic signal that can be reliably detected with the current methods. For example, the widespread green algal signal in diatoms constituted by more than 1,700 genes (29) was drastically reduced, to only 28 genes, after using rigorous criteria to define EGT and a broader taxon sampling, in particular more red algae (96). Interestingly, with 126 genes deemed of red algal origin, the same study simultaneously increased the relative proportion of genes derived from the current red plastid in diatoms (96). Similarly, a better algal representation and manual curation of the phylogenetic trees led to a 10× reduction of the algal signal in *C. velia*: from 513 in the original study (28) to only 51 cases of EGT (95). Along the same line, the reanalysis of two *Phytophthora* genomes showed no evidence for an unusually high red algal contribution in oomycetes (97), contrasting with the original finding of 855 red alga-derived genes in this plastid-lacking lineage (31). Finally, in the recent investigation of the nuclear genomes of *B. natans* and *G. theta*, less than 6% and 7%, respectively, of all genes for which phylogenetic trees could be generated were concluded to be algal in nature (18), roughly corresponding to the 4.7% of haptophyte genes in kareniaceans (Fig. 1). Taken together, these results indicate that although EGT has undoubtedly occurred during the course of plastid integration, current evidence does not support the massive flux of algal genes that has been suggested in the past. Rather more modest figures, in the range of a few hundred genes or less, seem more likely.

In dinotoms, the endosymbiotic contribution to the host nuclear compartment could be even more modest, with only a few genes with diatom affinity. This makes sense in light of the cell biology of the two lineages. In contrast to kareniaceans, in which the tertiary plastids are reduced to an extent similar to that of most secondary plastids (39), the dinotom endosymbiont is much more complex and has preserved several features normally lost in other secondary or tertiary endosymbionts. In addition to the plastid itself, the endosymbiont has retained its nucleus, cytosol, cytosolic ribosomes, endoplasmic reticulum, and even mitochondria, all separated from the host components by a single membrane (70–74). Despite this low degree of cellular reduction, the endosymbi-

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

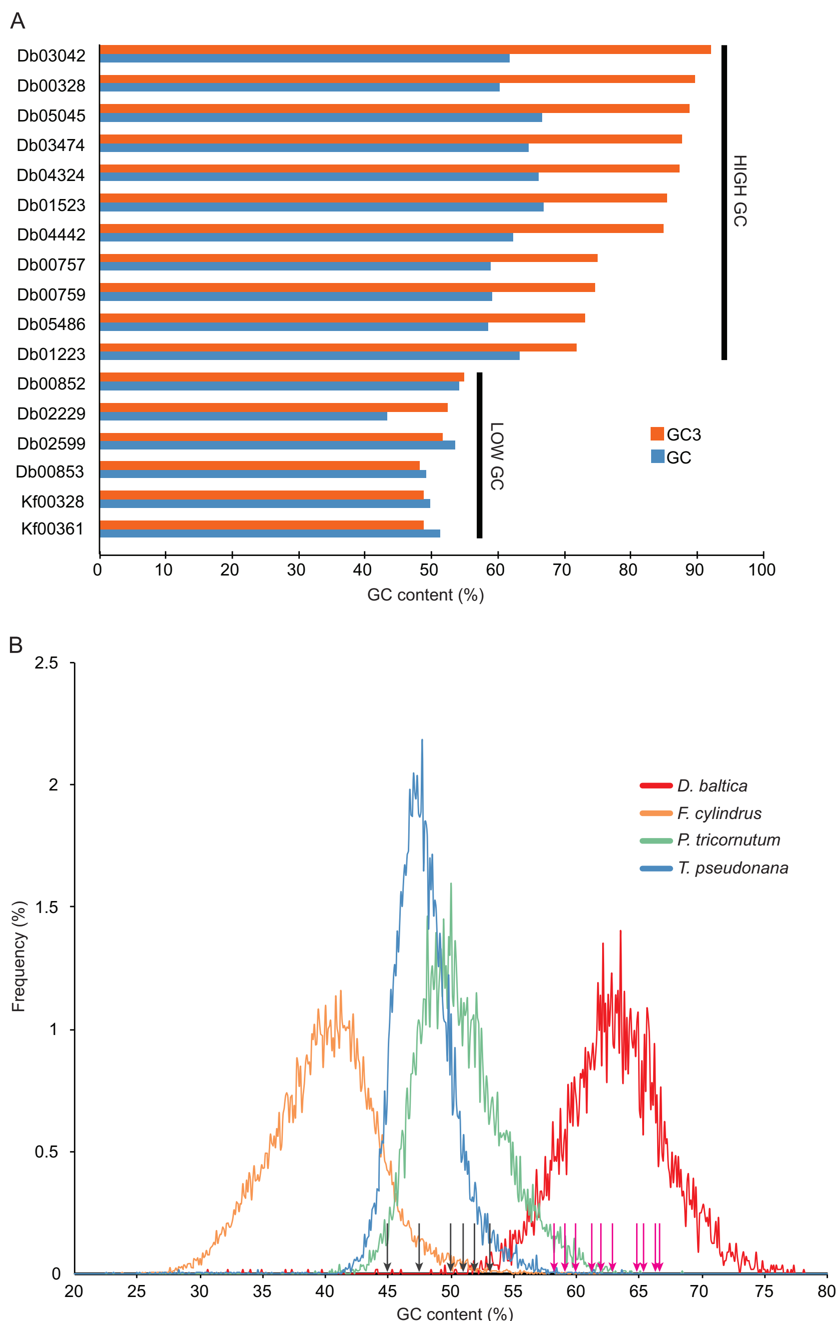

FIG 2 (A) Total GC content and GC content at third codon positions (GC3) for the 17 genes with diatom affinity in dinotoms. The sequence names are abbreviated as follows: Db, *D. baltica*; Kf, *K. foliaceum*. (B) Frequency distribution of the GC content for all transcripts from *F. cylindrus*, *P. tricornutum*, and *T. pseudonana* and all contigs from the *D. baltica* SL library. Gray arrows denote the genes with diatom affinity with low GC content; pink arrows denote the genes with diatom affinity with high GC content.

ont appears to be well integrated within the host system, as not only is the host-endosymbiont relationship mutually essential but also the endosymbiont is present throughout all stages of the cell cycle and its division is associated with the host division (74, 98–101). These unusual cell characteristics, in turn, mean that it may not have been necessary to massively transfer genes to the dinoflagellate host nucleus to maintain a functional plastid, given that the diatom nucleus still contains a large genome that may fully sustain the plastid (51). Indeed, it is possible that this plastid is fixed in the cell but not genetically integrated within its host in the traditional way.

This scenario leads to the more general question of what differentiates an organelle, here the plastid, from an endosymbiont. Plastids are often defined as genetically integrated organelles that depend on their host to provide key proteins through a dedicated import system (102). This explanation fits the general view of plastid reduction, where organelles are thought to severely reduce or lose their genome through gene loss and the transfer of impor-

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

tant genes to the host, which has been shown even in taxa that still possess highly reduced versions of the endosymbiont nuclei, the nucleomorphs (18, 103), and in the more recent primary endosymbiosis in *P. chromatophora* (35). Dinoflagellates are no exception to this process: nuclear genes whose products are targeted to the peridinin plastid have been reported for several lineages (85, 104). Our data now show that this endosymbiont contribution represents the dominant foreign signal in kareniaceans and likely extends beyond strictly plastid-associated genes. On the other hand, the low level of EGT observed in dinotoms puts their endosymbiont on the periphery of the organelle definition, since it is fully integrated within the cell in many ways but not deeply integrated on a genetic level. Thus, in addition to challenging our view of how we define endosymbionts and organelles, this also reinforces the idea that dinotoms may represent an unusual intermediate in the continuum of symbiotic interactions observed in nature (51, 105, 106).

ACKNOWLEDGMENTS

This work was funded by a grant from the Natural Sciences and Engineering Research Council of Canada (227301) and a grant to the Center for Microbial Diversity and evolution from the Tula Foundation. E.H. was supported by a prospective researcher postdoctoral fellowship from the Swiss National Science Foundation. B.I. was supported by a doctoral scholarship from NSERC. P.J.K. is a Fellow of the Canadian Institute for Advanced Research.

We thank Beverly Green for providing the plastid genome of *F. cylindrus*.

REFERENCES

1. **Keeling PJ.** 2010. The endosymbiotic origin, diversification and fate of plastids. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365:**729–748. http://dx.doi.org/10.1098/rstb.2009.0103.
2. **Palmer JD, Soltis D, Chase M.** 2004. The plant tree of life: an overview and some points of view. Am. J. Bot. **91:**1437–1445. http://dx.doi.org/10.3732/ajb.91.10.1437.
3. **Stiller JW, Reel DC, Johnson JC.** 2003. A single origin of plastids revisited: convergent evolution in organellar genome content. J. Phycol. **39:**95–105. http://dx.doi.org/10.1046/j.1529-8817.2003.02070.x.
4. **Keeling PJ.** 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu. Rev. Plant Biol. **64:**27.1–27.25. http://dx.doi.org/10.1146/annurev-arplant-050312-120144.
5. **Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ.** 2007. The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Mol. Biol. Evol. **24:**54–62. http://dx.doi.org/10.1093/molbev/msl129.
6. **Bodył A, Stiller JW, Mackiewicz P.** 2009. Chromalveolate plastids: direct descent or multiple endosymbioses? Trends Ecol. Evol. **24:**119–121. http://dx.doi.org/10.1016/j.tree.2008.11.003.
7. **Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H.** 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. Mol. Biol. Evol. **27:**1698–1709. http://dx.doi.org/10.1093/molbev/msq059.
8. **Sanchez Puerta MV, Delwiche CF.** 2008. A hypothesis for plastid evolution in chromalveolates. J. Phycol. **44:**1097–1107. http://dx.doi.org/10.1111/j.1529-8817.2008.00559.x.
9. **Wisecaver JH, Hackett JD.** 2011. Dinoflagellate genome evolution. Annu. Rev. Microbiol. **65:**369–387. http://dx.doi.org/10.1146/annurev-micro-090110-102841.
10. **Janouškovec JJ, Horák A, Oborník M, Lukeš J, Keeling PJ.** 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci. U. S. A. **107:**10949–10954. http://dx.doi.org/10.1073/pnas.1003335107.
11. **Shalchian-Tabrizi K, Minge MA, Cavalier-Smith T, Nedreklepp JM, Klaveness D, Jakobsen KS.** 2006. Combined heat shock protein 90 and ribosomal RNA sequence phylogeny supports ultiple replacements of dinoflagellate plastids. J. Eukaryot. Microbiol. **53:**217–224. http://dx.doi.org/10.1111/j.1550-7408.2006.00098.x.
12. **Hackett JD, Anderson D, Erdner D, Bhattacharya D.** 2004. Dinoflagellates: a remarkable evolutionary experiment. Am. J. Bot. **91:**1523–1534. http://dx.doi.org/10.3732/ajb.91.10.1523.
13. **Watanabe MM, Suda S, Inouya I, Sawaguchi T, Chihara M.** 1990. *Lepidodinium viride* gen. et sp. nov. (Gymnodinaiales, Dinophyta), a green dinoflagellate with a chlorophyll a- and b-containing endosymbiont. J. Phycol. **26:**741–751.
14. **Matsumoto T, Shinozaki F, Chikuni T, Yabuki A, Takishita K, Kawachi M, Nakayama T, Inouye I, Hashimoto T, Inagaki Y.** 2011. Green-colored plastids in the dinoflagellate genus *Lepidodinium* are of core chlorophyte origin. Protist **162:**268–276. http://dx.doi.org/10.1016/j.protis.2010.07.001.
15. **Timmis JN, Ayliffe MA, Huang CY, Martin W.** 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. **5:**123–135. http://dx.doi.org/10.1038/nrg1271.
16. **Qiu H, Yoon HS, Bhattacharya D.** 2013. Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. Front. Plant Sci. **4:**1–8. http://dx.doi.org/10.3389/fpls.2013.00366.
17. **Lane CE, Archibald JM.** 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends Ecol. Evol. **23:**268–275. http://dx.doi.org/10.1016/j.tree.2008.02.004.
18. **Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJM, Herman EK, Klute MJ, Nakayama T, Oborník M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, Höppner MP, Ishida K, Kim EE, Koreny L, Kroth PG, Liu Y, Malik S-B, Maier UG, McRose D, Mock T, Neilson JAD, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM.** 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature **492:**59–65. http://dx.doi.org/10.1038/nature11681.
19. **Suzuki K, Miyagishima S.** 2010. Eukaryotic and eubacterial contributions to the establishment of plastid proteome estimated by large-scale phylogenetic analyses. Mol. Biol. Evol. **27:**581–590. http://dx.doi.org/10.1093/molbev/msp273.
20. **Huang J, Gogarten JP.** 2007. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol. **8:**R99. http://dx.doi.org/10.1186/gb-2007-8-6-r99.
21. **Moustafa A, Reyes-Prieto A, Bhattacharya D.** 2008. Chlamydiae has contributed at least 55 genes to Plantae with predominantly plastid functions. PLoS One **3:**e2205. http://dx.doi.org/10.1371/journal.pone.0002205.
22. **Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D.** 2006. Cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr. Biol. **16:**2320–2325. http://dx.doi.org/10.1016/j.cub.2006.09.063.
23. **Moustafa A, Bhattacharya D.** 2008. PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*. BMC Evol. Biol. **8:**6. http://dx.doi.org/10.1186/1471-2148-8-6.
24. **Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D.** 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. U. S. A. **99:**12246–12251. http://dx.doi.org/10.1073/pnas.182432999.
25. **Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ.** 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proc. Natl. Acad. Sci. U. S. A. **100:**7678–7683. http://dx.doi.org/10.1073/pnas.1230951100.
26. **Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret J-P, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A,

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kröger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin Jézéquel Lopez VPJ, Lucas S, Mangogna M, Mcginnis K, Medlin LK, Montsant A, Oudot-Le Secq M-P, Napoli C, Obornik M, Parker MS, Petit J-L, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, Dassow Von P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van de Peer Y, Grigoriev IV. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature 456:239–244. http://dx.doi.org/10.1038/nature07410.

27. Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury J-M, Badger JH, Beszteri B, Billiau K, Bonnet E, Bothwell JH, Bowler C, Boyen C, Brownlee C, Carrano CJ, Charrier B, Cho GY, Coelho SM, Collén J, Corre E, Da Silva C, Delage L, Delaroque N, Dittami SM, Doulbeau S, Elias M, Farnham G, Gachon CMM, Gschloessl B, Heesch S, Jabbari K, Jubin C, Kawai H, Kimura K, Kloareg B, Küpper FC, Lang D, Le Bail A, Leblanc C, Lerouge P, Lohr M, Lopez PJ, Martens C, Maumus F, Michel G, Miranda-Saavedra D, Morales J, Moreau H, Motomura T, Nagasato C, Napoli CA, Nelson DR, Nyvall-Collén P, Peters AF, Pommier C, Potin P, Poulain J, Quesneville H, Read B, Rensing SA, Ritter A, Rousvoal S, Samanta M, Samson G, Schroeder DC, Ségurens B, Strittmatter M, Tonon T, Tregear JW, Valentin K, Dassow Von P, Yamagishi T, Van de Peer Y, Wincker P. 2010. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature 465:617–621. http://dx.doi.org/10.1038/nature09016.
28. Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and Apicomplexa-related *Chromera velia*. Genome Biol. Evol. 3:1220–1230. http://dx.doi.org/10.1093/gbe/evr100.
29. Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. 2009. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. Science 324:1724–1726. http://dx.doi.org/10.1126/science.1172983.
30. Archibald JM. 2009. The puzzle of plastid evolution. Curr. Biol. 19:R81–R88. http://dx.doi.org/10.1016/j.cub.2008.11.067.
31. Tyler BM. 2006. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science 313:1261–1266. http://dx.doi.org/10.1126/science.1128796.
32. Reyes-Prieto A, Moustafa A, Bhattacharya D. 2008. Multiple genes of apparent algal origin suggest ciliates may once have been photosynthetic. Curr. Biol. 18:956–962. http://dx.doi.org/10.1016/j.cub.2008.05.042.
33. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21:809–818. http://dx.doi.org/10.1093/molbev/msh075.
34. Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. U. S. A. 108:13624–13629. http://dx.doi.org/10.1073/pnas.1110633108.
35. Nowack ECM, Melkonian M, Glöckner G. 2008. Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. Curr. Biol. 18:410–418.http://dx.doi.org/10.1016/j.cub.2008.02.051
36. Nakayama T, Ishida K. 2009. Another acquisition of a primary photosynthetic organelle is underway in *Paulinella chromatophora*. Curr. Biol. 19:R284–R285. http://dx.doi.org/10.1016/j.cub.2009.02.043.
37. Reyes-Prieto A, Yoon HS, Moustafa A, Yang EC, Andersen RA, Boo SM, Nakayama T, Ishida K, Bhattacharya D. 2010. Differential gene retention in plastids of common recent origin. Mol. Biol. Evol. 27:1530–1537. http://dx.doi.org/10.1093/molbev/msq032.
38. Nowack ECM, Vogel H, Groth M, Grossman AR, Melkonian M, Glöckner G. 2011. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. Mol. Biol. Evol. 28:407–422. http://dx.doi.org/10.1093/molbev/msq209.
39. 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. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026350.
40. Kite GC, Dodge JD. 1985. Structural organization of plastid DNA in two anomalously pigmented dinoflagellates. J. Phycol. 21:50–56.
41. Kite GC, Dodge JD. 1988. Cell and chloroplast ultrastructure in *Gyrodinium aureolum* and *Gymnodinium galatheanum*. Two marine dinoflagellates containing an unusual carotenoid. Sarsia 73:131–138.
42. Dodge JD, Crawford RM. 1969. Observations on the fine structure of the eyespot and associated organelles in the dinoflagellate *Glenodinium foliaceum*. J. Cell Sci. 5:479–493.
43. 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. http://dx.doi.org/10.1111/j.1550-7408.1997.tb05672.x.
44. Wisecaver JH, Hackett JD. 2010. Transcriptome analysis reveals nuclear-encoded proteins for the maintenance of temporary plastids in the dinoflagellate *Dinophysis acuminata*. BMC Genomics 11:366. http://dx.doi.org/10.1186/1471-2164-11-366.
45. Nishitani G, Nagai S, Hayakawa S, Kosaka Y, Sakurada K, Kamiyama T, Gojobori T. 2012. Multiple plastids collected by the dinoflagellate *Dinophysis mitra* through kleptoplastidy. Appl. Environ. Microbiol. 78:813–821. http://dx.doi.org/10.1128/AEM.06544-11.
46. Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng J-F, Bhattacharya DD. 2006. Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*. Mol. Biol. Evol. 23:2026–2038. http://dx.doi.org/10.1093/molbev/msl074.
47. Patron NJ, Waller RF, Keeling PJ. 2006. A tertiary plastid uses genes from two endosymbionts. J. Mol. Biol. 357:1373–1382. http://dx.doi.org/10.1016/j.jmb.2006.01.084.
48. Ishida K, Green BR. 2002. Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. Proc. Natl. Acad. Sci. U. S. A. 99:9294–9299. http://dx.doi.org/10.1073/pnas.142091799.
49. Blackburn SI, Bolch CJS, Haskard KA, Hallegraeff GM. 2001. Reproductive compatibility among four global populations of the toxic dinoflagellate *Gymnodinium catenatum* (Dinophyceae). Phycologia 40:78–87. http://dx.doi.org/10.2216/i0031-8884-40-1-78.1.
50. 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. http://dx.doi.org/10.1186/1471-2148-7-172.
51. 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. http://dx.doi.org/10.1371/journal.pone.0010711.
52. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202. http://dx.doi.org/10.1101/gr.8.3.195.
53. Zheng Y, Zhao L, Gao J, Fei Z. 2011. iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. BMC Bioinformatics 12:453. http://dx.doi.org/10.1186/1471-2105-12-453.
54. Min XJ, Butler G, Storms R, Tsang A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33:W677–W680. http://dx.doi.org/10.1093/nar/gki394.
55. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. http://dx.doi.org/10.1093/bioinformatics/btl158.
56. Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33:511–518. http://dx.doi.org/10.1093/nar/gki198.
57. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973. http://dx.doi.org/10.1093/bioinformatics/btp348.
58. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. http://dx.doi.org/10.1371/journal.pone.0009490.
59. Maruyama S, Eveleigh RJM, Archibald JM. 2013. Treetrimmer: a method for phylogenetic dataset size reduction. BMC Res. Notes 6:145. http://dx.doi.org/10.1186/1756-0500-6-145.
60. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.
61. Chan CX, Yang EC, Banerjee T, Yoon HS, Martone PT, Estevez JM, Bhattacharya D. 2011. Red and green algal monophyly and extensive

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

gene sharing found in a rich repertoire of red algal genes. Curr. Biol. **21**:328–333. http://dx.doi.org/10.1016/j.cub.2011.01.037.
62. **Patron NJ, Waller RF, Archibald JM, Keeling PJ.** 2005. Complex protein targeting to dinoflagellate plastids. J. Mol. Biol. **348**:1015–1024. http://dx.doi.org/10.1016/j.jmb.2005.03.030.
63. **Bodył A, Mackiewicz P.** 2007. Analysis of the targeting sequences of an iron-containing superoxide dismutase (SOD) of the dinoflagellate *Lingulodinium polyedrum* suggests function in multiple cellular compartments. Arch. Microbiol. **187**:281–296. http://dx.doi.org/10.1007/s00203-006-0194-5.
64. **Nassoury N.** 2003. Plastid ultrastructure defines the protein import pathway in dinoflagellates. J. Cell Sci. **116**:2867–2874. http://dx.doi.org/10.1242/jcs.00517.
65. **Minge MA, Shalchian-Tabrizi K, Tørresen OK, Takishita K, Probert I, Inagaki Y, klaveness D, Jakobsen KS.** 2010. A phylogenetic mosaic plastid proteome and unusual plastid-targeting signals in the green-colored dinoflagellate *Lepidodinium chlorophorum* BMC Evol. Biol. **10:** 191. http://dx.doi.org/10.1186/1471-2148-10-191.
66. **Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ.** 2006. PredSL: a tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics Proteomics Bioinformatics **4**:48–55. http://dx.doi.org/10.1016/S1672-0229(06)60016-8.
67. **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. http://dx.doi.org/10.1006/jmbi.2000.3903.
68. **Small I, Peeters N, Legeai F, Lurin C.** 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics **4**:1581–1590. http://dx.doi.org/10.1002/pmic.200300776.
69. **Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S.** 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics **18**:298–305. http://dx.doi.org/10.1093/bioinformatics/18.2.298.
70. **Dodge JD.** 1971. A dinoflagellate with both a mesocaryotic and a eucaryotic nucleus. I. Fine structure of the nuclei. Protoplasma **73**:145–157.
71. **Horiguchi T, Pienaar RN.** 1994. Ultrastructure of a new marine sand-dwelling dinoflagellate, *Gymnodinium quadrilobatum* sp. nov. (Dinophyceae) with special reference to its endosymbiotic alga. Eur. J. Phycol. **29**:237–245. http://dx.doi.org/10.1080/09670269400650691.
72. **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. http://dx.doi.org/10.1007/s10265-006-0047-y.
73. **Tamura M, Shimada S, Horiguchi T.** 2005. *Galeidiniium rugatum* gen. et sp. nov. (Dinophyceae), a new coccoid dinoflagellate with a diatom endosymbiont. J. Phycol. **41**:658–671. http://dx.doi.org/10.1111/j.1529-8817.2005.00085.x.
74. **Tomas RN, Cox ER.** 1973. Observations on the symbiosis of *Peridinium balticum* and its intracellular alga. I. Ultrastructure. J. Phycol. **9**:304–323.
75. **Lidie KB, Van Dolah FM.** 2007. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. J. Eukaryot. Microbiol. **54:** 427–435. http://dx.doi.org/10.1111/j.1550-7408.2007.00282.x.
76. **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. http://dx.doi.org/10.1073/pnas.0700258104.
77. **Zhang H, Lin S.** 2009. Retrieval of missing spliced leader in dinoflagellates. PLoS One **4**:e4129. http://dx.doi.org/10.1371/journal.pone.0004129.
78. **Slamovits CH, Keeling PJ.** 2008. Widespread recycling of processed cDNAs in dinoflagellates. Curr. Biol. **18**:R550–R552. http://dx.doi.org/10.1016/j.cub.2008.04.054.
79. **Bachvaroff TR, Place AR.** 2008. From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. PLoS One **3**:e2929. http://dx.doi.org/10.1371/journal.pone.0002929.
80. **Jeffroy O, Brinkmann H, Delsuc F, Philippe H.** 2006. Phylogenomics: the beginning of incongruence? Trends Genet. **22**:225–231. http://dx.doi.org/10.1016/j.tig.2006.02.003.
81. **Stiller JW.** 2011. Experimental design and statistical rigor in phylogenomics of horizontal and endosymbiotic gene transfer. BMC Evol. Biol. **11**:259. http://dx.doi.org/10.1186/1471-2148-11-259.
82. **Philippe H, Laurent J.** 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. **8**:616–623. http://dx.doi.org/10.1016/S0959-437X(98)80028-2.
83. **Felsenstein J.** 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**:401–410. http://dx.doi.org/10.2307/2412923.
84. **Keeling PJ, Palmer JD.** 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. **9**:605–618. http://dx.doi.org/10.1038/nrg2386.
85. **Chan CX, Soares MB, Bonaldo MF, Wisecaver JH, Hackett JD, Anderson DM, Erdner DL, Bhattacharya D.** 2012. Analysis of *Alexamdrium tamarense* (Dinophyceae) genes reveals the complex evolutionary history of a microbial eukaryote. J. Phycol. **48**:1130–1142. http://dx.doi.org/10.1111/j.1529-8817.2012.01194.x.
86. **Okamoto OK, Liu L, Robertson DL, Woodland Hastings JW.** 2001. Members of a dinoflagellate luciferase gene family differ in synonymous substitution rates. Biochemistry **40**:15862–15868. http://dx.doi.org/10.1021/bi011651q.
87. **Reichman JR, Wilcox TP, Vize PD.** 2003. PCP gene family in *Symbiodinium* from *Hippopus hippopus*: low levels of concerted evolution, isoform diversity, and spectral tuning of chromophores. Mol. Biol. Evol. **20**:2143–2154. http://dx.doi.org/10.1093/molbev/msg233.
88. **Kim S, Bachvaroff TR, Handy SM, Delwiche CF.** 2011. Dynamics of actin evolution in dinoflagellates. Mol. Biol. Evol. **28**:1469–1480. http://dx.doi.org/10.1093/molbev/msq332.
89. **Lidie KB, Ryan JC, Barbier M, Van Dolah FM.** 2005. Gene expression in Florida red tide dinoflagellate *Karenia brevis*: analysis of an expressed sequence tag library and development of DNA microarray. Mar. Biotechnol. **7**:481–493. http://dx.doi.org/10.1007/s10126-004-4110-6.
90. **Danne JC, Gornik SG, Waller RF.** 2012. An assessment of vertical inheritance versus endosymbiont transfer of nucleus-encoded genes for mitochondrial proteins following tertiary endosymbiosis in *Karlodinium micrum*. Protist **163**:76–90. http://dx.doi.org/10.1016/j.protis.2011.03.002.
91. **Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, Takeuchi T, Hisata K, Tanaka M, Fujiwara M, Hamada M, Seidi A, Fujie M, Usami T, Goto H, Yamasaki S, Arakaki N, Suzuki Y, Sugano S, Toyoda A, Kuroki Y, Fujiyama A, Medina M, Coffroth M-A, Bhattacharya DD, Satoh N.** 2013. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. Curr. Biol. **23**:1399–1408. http://dx.doi.org/10.1016/j.cub.2013.05.062.
92. **Muto A, Osawa S.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. U. S. A. **84**:166–169. http://dx.doi.org/10.1073/pnas.84.1.166.
93. **Nei M, Kumar S.** 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.
94. **McEwan ML, Keeling PJ.** 2004. HSP90, tubulin and actin are retained in the tertiary endosymbiont genome of *Kryptoperidinium foliaceum*. J. Eukaryot. Microbiol. **51**:651–659. http://dx.doi.org/10.1111/j.1550-7408.2004.tb00604.x.
95. **Burki F, Flegontov P, Obornik M, Cihlar J, Pain A, Lukes J, Keeling PJ.** 2012. Reevaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. Genome Biol. Evol. **4**:626–635. http://dx.doi.org/10.1093/gbe/evs036.
96. **Deschamps P, Moreira D.** 2012. Re-evaluating the green contribution to diatom genomes. Genome Biol. Evol. **4**:795–800.
97. **Stiller JW, Huang J, Ding Q, Tian J, Goodwillie C.** 2009. Are algal genes in nonphotosynthetic protists evidence of historical plastid endosymbioses? BMC Genomics **10**:484. http://dx.doi.org/10.1186/1471-2164-10-484.
98. **Chesnick JM, Cox ER.** 1987. Synchronized sexuality of an algal symbiont and its dinoflagellate host, *Peridinium balticum* (Levander) Lemmermann. Biosystems **21**:69–78. http://dx.doi.org/10.1016/0303-2647(87)90007-4.
99. **Chesnick JM, Cox ER.** 1989. Fertilization and zygote development in the binucleate dinoflagellate *Peridinium balticum* (Pyrrhophyta). Am. J. Bot. **76**:1060–1072. http://dx.doi.org/10.2307/2444528.
100. **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. http://dx.doi.org/10.1016/j.protis.2008.12.003.
101. **Tippit DH, Pickett-Heaps JD.** 1976. Apparent amitosis in the binucleate dinoflagellate *Peridinium balticum*. J. Cell Sci. **21**:273–289.
102. **Cavalier-Smith T, Lee JJ.** 1985. Protozoa as hosts for endosymbioses and the conversion of symbionts into organelles. J. Eukaryot. Microbiol. **32**:376–379.

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE

103. **Felsner G, Sommer MS, Gruenheit N, Hempel F, Moog D, Zauner S, Martin W, Maier UG.** 2011. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. Genome Biol. Evol. **3:**140–150. http://dx.doi.org/10.1093/gbe/evq074.
104. **Bachvaroff TR, Concepcion GT, Rogers CR, Herman EM, Delwiche CF.** 2004. Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. Protist **155:**65–78. http://dx.doi.org/10.1078/1434461000165.
105. **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. http://dx.doi.org/10.1111/j.1550-7408.2007.00245.x.
106. **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. http://dx.doi.org/10.1371/journal.pone.0043763.

Downloaded from http://ec.asm.org/ on October 10, 2019 at INSTITUT FUER MEERESKUNDE