Diurnal Transcriptional Regulation of Endosymbiotically Derived Genes in the Chlorarachniophyte *Bigelowiella natans*

Shigekatsu Suzuki[1,†], Ken-Ichiro Ishida[2], and Yoshihisa Hirakawa[2,*,†]

[1]Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan

[2]Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan

*Corresponding author: E-mail: hirakawa.yoshi.fp@u.tsukuba.ac.jp.

[†]These authors contributed equally to this work.

Abstract

Chlorarachniophyte algae possess complex plastids acquired by the secondary endosymbiosis of a green alga, and the plastids harbor a relict nucleus of the endosymbiont, the so-called nucleomorph. Due to massive gene transfer from the endosymbiont to the host, many proteins involved in plastid and nucleomorph are encoded by the nuclear genome. Genome sequences have provided a blueprint for the fate of endosymbiotically derived genes; however, transcriptional regulation of these genes remains poorly understood. To gain insight into the evolution of endosymbiotic genes, we performed genome-wide transcript profiling along the cell cycle of the chlorarachniophyte *Bigelowiella natans*, synchronized by light and dark cycles. Our comparative analyses demonstrated that transcript levels of 7,751 nuclear genes (35.7% of 21,706 genes) significantly oscillated along the diurnal/cell cycles, and those included 780 and 147 genes for putative plastid and nucleomorph-targeted proteins, respectively. Clustering analysis of those genes revealed the existence of transcriptional networks related to specific biological processes such as photosynthesis, carbon metabolism, translation, and DNA replication. Interestingly, transcripts of many plastid-targeted proteins in *B. natans* were induced before dawn, unlike other photosynthetic organisms. In contrast to nuclear genes, 99% nucleomorph genes were found to be constitutively expressed during the cycles. We also found that the nucleomorph DNA replication would be controlled by a nucleus-encoded viral-like DNA polymerase. The results of this study suggest that nucleomorph genes have lost transcriptional regulation along the diurnal cycles, and nuclear genes exert control over the complex plastid including the nucleomorph.

Key words: algae, cell cycle, endosymbiosis, nucleomorph, secondary plastid, RNA-seq.

Introduction

Plastids evolved by several endosymbiotic events between a photosynthetic organism and a phagotrophic eukaryote. Streptophytes and three algal groups (chlorophytes, rhodophytes, and glaucophytes) acquired plastids by a single primary endosymbiosis with a cyanobacterium (Rodríguez-Ezpeleta et al. 2005; Price et al. 2012). Many other algae evolved complex plastids secondarily by engulfing plastid-bearing primary algae, and these events, the so-called secondary endosymbioses, occurred in multiple distinct lineages (Gould et al. 2008; Keeling 2013; Burki et al. 2016). Through endosymbioses, many endosymbiotic genes have been lost or transferred to the host nuclear genomes, the

bulk of which encode proteins that are targeted back to plastids where they function (Timmis et al. 2004; Archibald 2015). Therefore, plastid function is mainly sustained by a large number of nucleus-encoded proteins.

In photosynthetic organisms, many cellular processes universally respond to alternating light and dark conditions, and diurnal regulation occurs in photosynthesis, carbon metabolism, and fatty acid synthesis. In particular, in unicellular algae with a limited number of plastids, cell and plastid proliferations are usually synchronized by diurnal cycles (Lien and Knutsen 1979; Farinas et al. 2006; Imoto et al. 2011; Miyagishima et al. 2012). Transcriptome studies on photosynthetic organisms have confirmed that abundant nuclear genes show periodic

expression patterns corresponding to diurnal cycles, which include several genes of plastid-associated proteins (Eberhard et al. 2008; Noordally and Millar 2015). This suggests that nuclear-transferred endosymbiotic genes have obtained transcriptional regulation mechanisms, which are very important for the control of modern-day plastids. However, diurnal transcriptome studies are mostly limited to some model algae with complete genome sequences and synchronous culture methods, such as the green algae *Chlamydomonas reinhardtii* (Zones et al. 2015), *Ostreococcus tauri* (Monnier et al. 2010), and the red alga *Cyanidioschyzon merolae* (Fujiwara et al. 2009; Kanesaki et al. 2012). To gain insight into the evolution of transcriptional regulation for endosymbiotic genes, further studies are required on diverse photosynthetic organisms.

Chlorarachniophytes are marine unicellular algae that have complex secondary plastids acquired by the uptake of a green algal endosymbiont (Rogers et al. 2007; Tanifuji et al. 2014; Suzuki et al. 2016). Interestingly, their plastids still harbor a relict nucleus of the endosymbiont, which has disappeared in most cases of secondary endosymbioses. The relict nucleus is called the nucleomorph, located in the periplastidal compartment (PPC) between the inner and outer pair of plastid membranes, which is the remnant cytoplasm of the endosymbiont (Hibberd and Norris 1984). Due to the retention of the nucleomorph, chlorarachniophytes offer an interesting opportunity to study the evolution of endosymbiotically derived genes. Genomic sequencing of the chlorarachniophyte *Bigelowiella natans* has revealed that the nucleomorph contains an extremely reduced and compacted eukaryotic genome that encodes approximately 300 proteins, including 17 plastid-associated proteins and hundreds of housekeeping proteins (Gilson et al. 2006). However, many fundamental genes for the maintenance of the plastid and nucleomorph are missing in the nucleomorph genome, and endosymbiotically derived nuclear genes are supposed to take over their function (Curtis et al. 2012). Indeed, several nucleus-encoded proteins were demonstrated to be targeted to the plastid and PPC (Gile and Keeling 2008; Hirakawa et al. 2009, 2010, 2011). Although the complete genome sequence of *B. natans* has revealed the fate of endosymbiotic genes, little is known about their transcriptional regulation.

Our previous studies have shown that a couple of nuclear genes for plastid- and nucleomorph-targeted proteins of *B. natans* were transcriptionally regulated under a diurnal cycle (Hirakawa et al. 2011, Hirakawa and Ishida 2015). To obtain a comprehensive overview, we performed genome-wide transcript profiling using RNA-seq data along the *B. natans* cell cycles temporally synchronized by diurnal light and dark cycles. Our comparative analyses demonstrated that mRNA levels of 7,751 nuclear genes (35.7% of total 21,706 genes) significantly oscillated during the diurnal/cell division cycles, and those included over a thousand genes for proteins that were predicted to be targeted to the plastid, nucleomorph,

and PPC. Clustering analysis of gene expression patterns revealed the existence of transcriptional networks related to specific biological processes such as photosynthesis, carbon metabolism, translation, and DNA replication. In contrast to the nuclear genes, 99% nucleomorph genes (283 of 285 genes) were found to be constitutively expressed during the cycles. Overall, our data suggest that nucleomorph-retained genes have lost diurnal/cell cycle-regulated transcription through their reductive evolution, and nuclear genes exert control over the complex plastid including a nucleomorph in the chlorarachniophyte.

Results and Discussion

Differentially Expressed Genes of Nuclear Genomes during Diurnal/Cell Cycles

To identify genes that exhibit diurnal and/or cell cycle-regulated expression patterns in the chlorarachniophyte *B. natans*, cell division was temporally synchronized under 12h/12h light/dark (L/D) cycles. Cell division occurred during the early dark periods (fig. 1A). RNA were sampled every 4 h during two L/D cycles at 13 time points from 0 to 48 h (dark to light transition at 0/24/48 h and light to dark transition at 12/36 h), and processed for high throughput sequencing. An average of 11,119,018 of RNA-seq reads were mapped on the genomic sequence in each time point, and read counts were obtained for 21,706 and 285 predicted protein-coding genes of the nucleus and nucleomorph, respectively (supplementary table S1 Supplementary Material online). In the nuclear genome, 287 predicted genes (1.3% of the total genes) were not covered with any of the read counts, suggesting that these would be pseudogenized or wrongly annotated genes. To detect differentially expressed genes (DEGs) during the diurnal/cell cycles, normalized read counts of each gene were compared among time points by baySeq with a false discovery rate of ≤ 0.001, and the data from two L/D cycles were treated as biological replicates. A total of 7,751 nuclear and two nucleomorph genes (35.7% of the total genes) were identified as DEGs, which were grouped into five major clusters (cluster A1, A2, B, C1, and C2), including 17 subclusters based on the similarity of their expression patterns analyzed by a hierarchical clustering algorithm (fig. 1B; supplementary table S1 Supplementary Material online). Cluster A1 and A2 comprised 1,775 and 1,650 genes, respectively, that were mainly expressed in the mid and late dark periods, and transcripts of cluster A1 increased earlier than those of cluster A2 (fig. 1B). Cluster B genes of 874 were expressed in the light periods. Genes in cluster C1 and C2 (2,154 and 824, respectively) were highly expressed in the late light and early dark periods across the light to dark transition (fig. 1B). Unexpectedly, transcripts of abundant DEGs were induced in the dark periods despite being a photosynthetic organism. To investigate transcriptional networks of DEGs related to
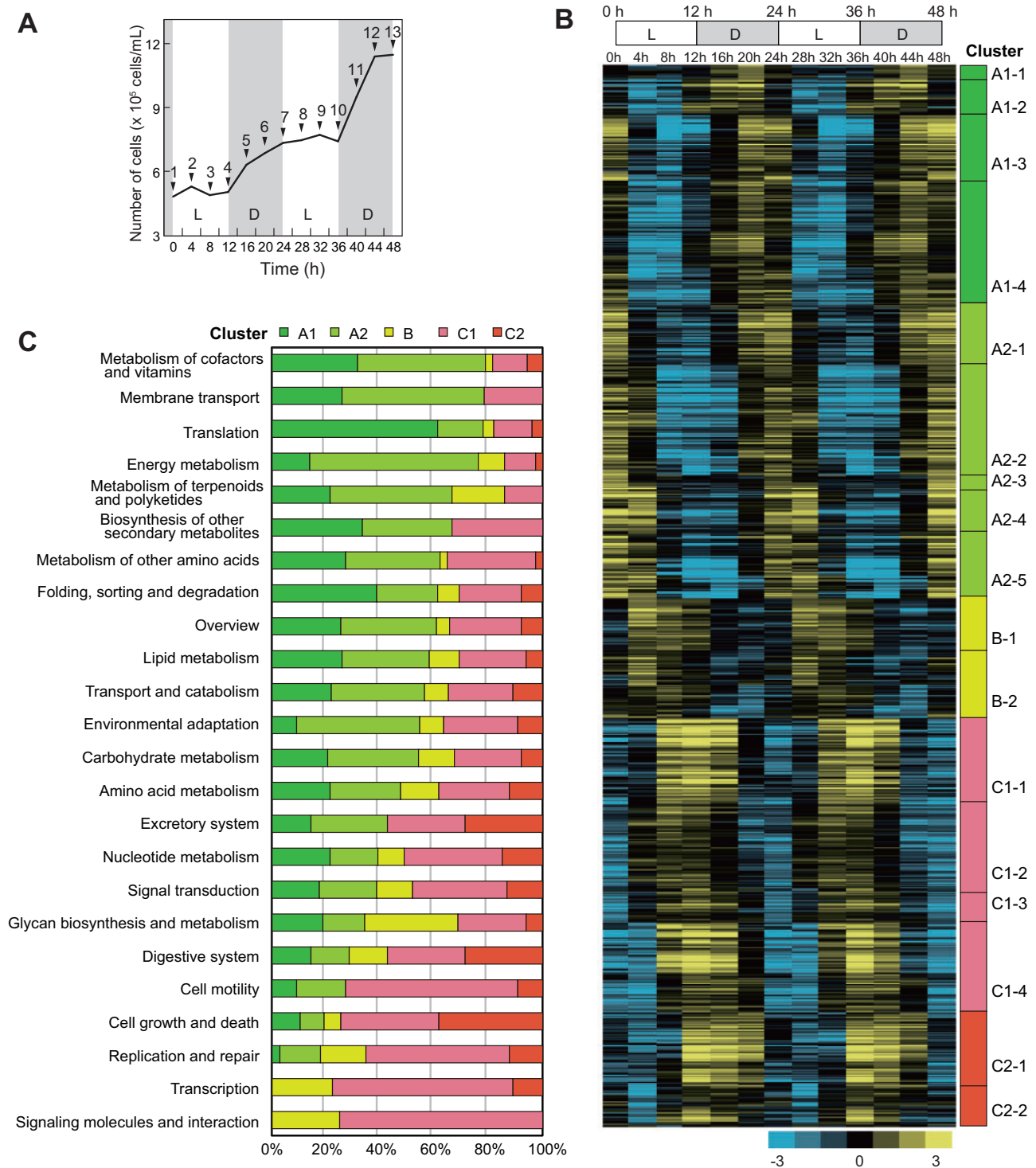
FIG. 1.—Differentially expressed genes during the diurnal cycles. (A) Cell number in temporally synchronized culture of *Bigelowiella natans* along the diurnal cycles. Total RNA were extracted every 4 h at 13 time points indicated by arrowheads. Dark periods are indicated by gray shading. (B) Expression heat map of 7,753 differentially expressed genes (DEGs) that are grouped into five major clusters (A1, A2, B, C1, C2) and 17 subclusters (A1-1 to C2-2) based on their coexpression patterns. Normalized relative transcript levels (log$_2$ values) are shown by color gradient (higher transcription levels in yellow and lower levels in blue). (C) Relationship between expression cluster and functional category of DEGs. Approximately 1,600 DEGs are categorized into 24 classifications by the KEGG annotation.

specific biological processes, DEGs were functionally annotated by the KEGG Automatic Annotation Server (Moriya et al. 2007). Approximately 1,600 DEGs were categorized into 24 classifications, including various metabolic, genetic, and cellular processes (fig. 1C; supplementary table S1 Supplementary Material online). Of these, categories related to translation, metabolism of cofactors and vitamins, and energy metabolism were mainly comprised of DEGs in cluster A, whereas DEGs in cluster C occupied categories of transcription, DNA replication and repair, cell motility, and cell growth and death (fig. 1C). A relatively large number of DEGs for cluster B were found in the category of glycan biosynthesis and metabolism (fig. 1C). Overall, the bulk of DEGs in cluster A were involved in plastid function, and those in cluster C were related to cell cycle regulation.

Differentially Expressed Genes of Nucleomorph Genomes

Among the 285 predicted nucleomorph genes, we detected only two DEGs: one encoding the plastidal chaperonin Cpn60 and the other a species-specific hypothetical protein (Bnatchr2110). Surprisingly, none of the other genes exhibit any obvious diurnal oscillation in transcript levels (fig. 2A). All nucleomorph genes showed unexpected decreases in transcript levels at 20 h during the first L/D cycle, whereas in the second cycle (biological replicate) there were no such drops (fig. 2A). To confirm the reliability of RNA-seq analysis, we performed real-time quantitative PCR (RT-qPCR) using another batch of synchronized cultures as the third biological replicate. Relative transcript levels were determined for four nuclear and four nucleomorph genes (supplementary material and methods, supplementary fig. S1 Supplementary Material online). Expression patterns of the nuclear genes were essentially identical for the two different assays, RNA-seq and RT-qPCR. For the nucleomorph genes, no decreases of transcript levels were detected at 20 h in the RT-qPCR assay (supplementary fig. S1, Supplementary Material online). Although the reason for the decreases in nucleomorph transcripts during the first diurnal cycle of the RNA-seq analysis remains unclear, we concluded that most nucleomorph genes are constitutively expressed during the diurnal/cell cycles, and, interestingly, these encodes some proteins generally associated with cell cycles, for example DNA replication (histone H3, H4, PCNA, MCM2, and MCM4) (fig. 2A; supplementary table S1, Supplementary Material online). This might be related to the highly compact structure of the nucleomorph genome with very short intergenic regions (Gilson et al. 2006). In reductive evolution of the nucleomorph genome, the majority of the nucleomorph genes would have lost transcriptional promoters regulated by diurnal/cell cycles. In contrast, two nucleomorph genes (*cpn60* in chromosome #1 and *bnatchr2110* in chromosome #2) were found as DEGs in the same subcluster, A1–4, suggesting that they retain transcriptional regulation. These two genes have a conserved motif of TGCATGCAAAT(T/A)A(T/
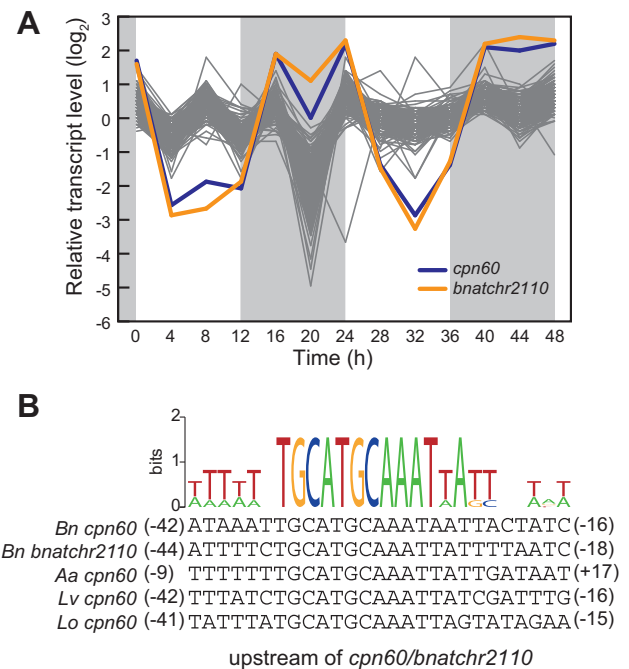

Fig. 2.—Expression profiles of nucleomorph genes during the diurnal cycles. (A) Plots of relative transcript levels for 285 nucleomorph-encoded proteins. Most of the genes show no obvious oscillations during the cycles (grey lines). The two differentially expressed genes (DEGs), *cpn60* and *bnatchr2110*, are shown with the blue and orange line, respectively. Dark periods are indicated by gray shading. (B) Consensus motif in upstream regions of *bnatchr2110* and four *cpn60* genes in nucleomorph genomes (DQ158856, AB996604, AB996599, CP006629). Bn, *Bigelowiella natans*; Aa, *Amorphochlora amoebiformis*; Lv, *Lotharella vacuolata*; Lo, *Lotharella oceanica*.

G)(T/C) upstream of their start codons, and the same motif was also found upstream of nucleomorph *cpn60* genes from three other chlorarachniophytes (fig. 2B), whereas this sequence was not found in any other genes. These findings suggest that this motif would be involved in the transcriptional regulation of the nucleomorph *cpn60* and *bnatchr2110* genes. Although there is no sequence homology between Cpn60 and Bnatchr2110, the hypothetical protein Bnatchr2110 might possibly function in a similar biological process with Cpn60, taking into consideration their coexpression patterns.

Predicted Subcellular Localization of Nucleus-Encoded Proteins

In a previous study by Curtis et al. (2012), 694 of the plastid- and 1,002 of the PPC/nucleomorph-targeted proteins were predicted in the nuclear genome of *B. natans*, based on their N-terminal targeting signals. Our study provided massive RNA-seq data through diurnal/cell cycles (total 80.6 Gbp reads) that covered almost all transcripts even for genes

with low or periodic expression, and these data improved some previous gene models. Total 3,862 of previous gene models had modification with changing their predicted start codons. Total 306 pairs of adjacent coding sequences were combined by filling the gaps. Modified gene models would represent potential full-length proteins. We performed subcellular localization prediction using reconstructed gene models based on their N-terminal features and sequence homology with functionally annotated proteins (see "Methods" section). In all, 780 plastid-targeted, 147 nucleomorph-targeted, 134 PPC-targeted, 2,331 possibly plastid/PPC-targeted, 408 mitochondrion-targeted, and 170 ER/Golgi-targeted proteins were predicted in the *B. natans* nuclear genome. Of these, over 75% genes for the plastid-targeted proteins were identified as DEGs, whereas rates of DEGs for the nucleomorph, PPC, mitochondria, and ER/Golgi were estimated to be 34.0–52.2% (fig. 3A). The majority of DEGs for plastid-targeted proteins were categorized into cluster A1 and A2, and

nucleomorph-targeted proteins comprised of a relatively high rate of DEGs in cluster C1 (fig. 3A). Plastid function appears to be strictly controlled by the transcription of plastid-targeted protein genes during diurnal/cell cycles in the chlorarachniophyte, similar to other photosynthetic organisms (Ma et al. 2001; Monnier et al. 2010; Zones et al. 2015). These results suggest that nuclear-transferred genes encoding many plastid-associated proteins have obtained transcriptional regulation mechanisms along the diurnal/cell cycles through the secondary endosymbiosis, whereas nucleomorph-retained genes no longer show diurnal regulation.

Transcriptional Networks Associated with Plastid Biological Processes

Previous transcriptome studies have shown that nuclear genes for plastid-associated proteins are transcriptionally regulated along diurnal cycles in *Arabidopsis* (Ma et al. 2001; Schaffer et al. 2001), the green algae, *Chlamydomonas* (Kucho et al.
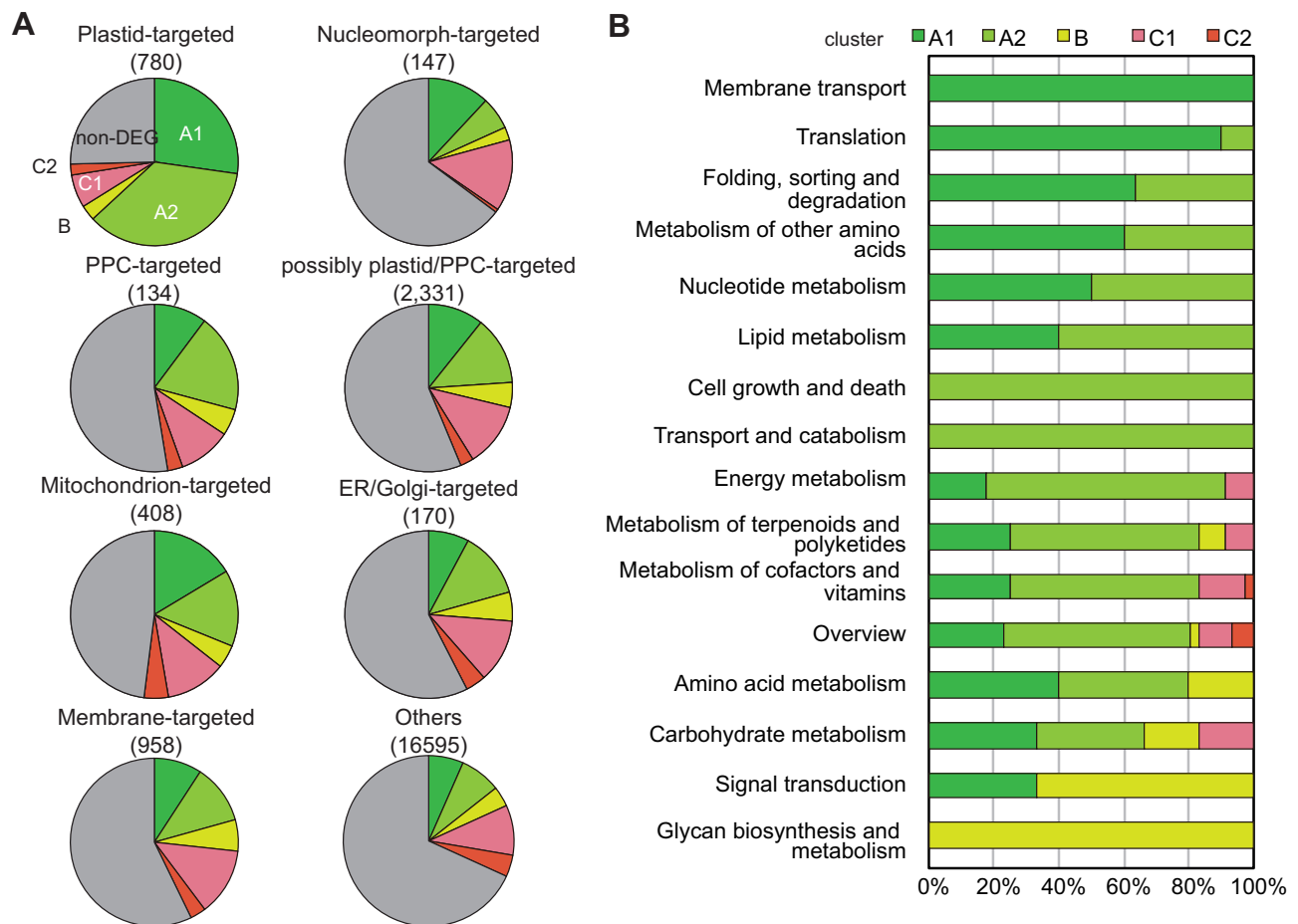

Fig. 3.—Expression profiles of nuclear genes encoding organelle-targeted proteins. (*A*) Proportion of expression clusters in *B. natans* nuclear genes encoding organelle-targeted proteins (plastid, mitochondrion, nucleomorph, PPC, possibly plastid/PPC, ER/Golgi), membrane proteins, and cytoplasmic proteins (Others). The estimated number of proteins is indicated above each pie chart. (*B*) Relationship between expression cluster and functional category in plastid-targeted DEGs (ptDEGs).

2005; Zones et al. 2015) and *Ostreococcus* (Monnier et al. 2010), the diatom, *Seminavis* (Gillard et al. 2008), and the haptophyte, *Chrysochromulina* (Hovde et al. 2015). In our analyses, many genes for plastid-targeted proteins of *B. natans* were identified as DEGs mainly grouped into cluster A1 and A2 that were expressed in the mid and late dark periods, respectively (fig. 3A and B). According to the KEGG classification, some of the functionally categorized plastid-targeted DEGs (ptDEGs) appeared to be coexpressed (fig. 3B). The ptDEGs associated with translation (e.g., ribosomal proteins and aminoacyl-tRNA synthetases) were mostly included in subcluster A1-3 that transcript levels were highest at time point 20/44 h in the dark periods (fig. 4A; supplementary fig. S2, Supplementary Material online). The ptDEGs related to the

porphyrin and chlorophyll metabolism (e.g., protoporphyrinogen oxidase and chlorophyllide a oxygenase) were included in subcluster A2-1 that reached peak levels at the end of the dark period at time point 24/48 h (fig. 4B; supplementary fig. S3, Supplementary Material online). Many ptDEGs involved in photosynthesis (e.g., LHCB and photosystem proteins) were highly coexpressed at the end of the dark and extending to the early light period (24 and 28 h) as in subcluster A2-2 and A2-5 (fig. 4C; supplementary fig. S4, Supplementary Material online). For the DEGs involved in carbon metabolism, the bulk of ptDEGs were mainly included in cluster A2 like photosynthesis-related ptDEGs (fig. 4D; supplementary fig. S5, Supplementary Material online), whereas mitochondrion-targeted DEGs (mtDEGs) were mainly found in cluster C1,
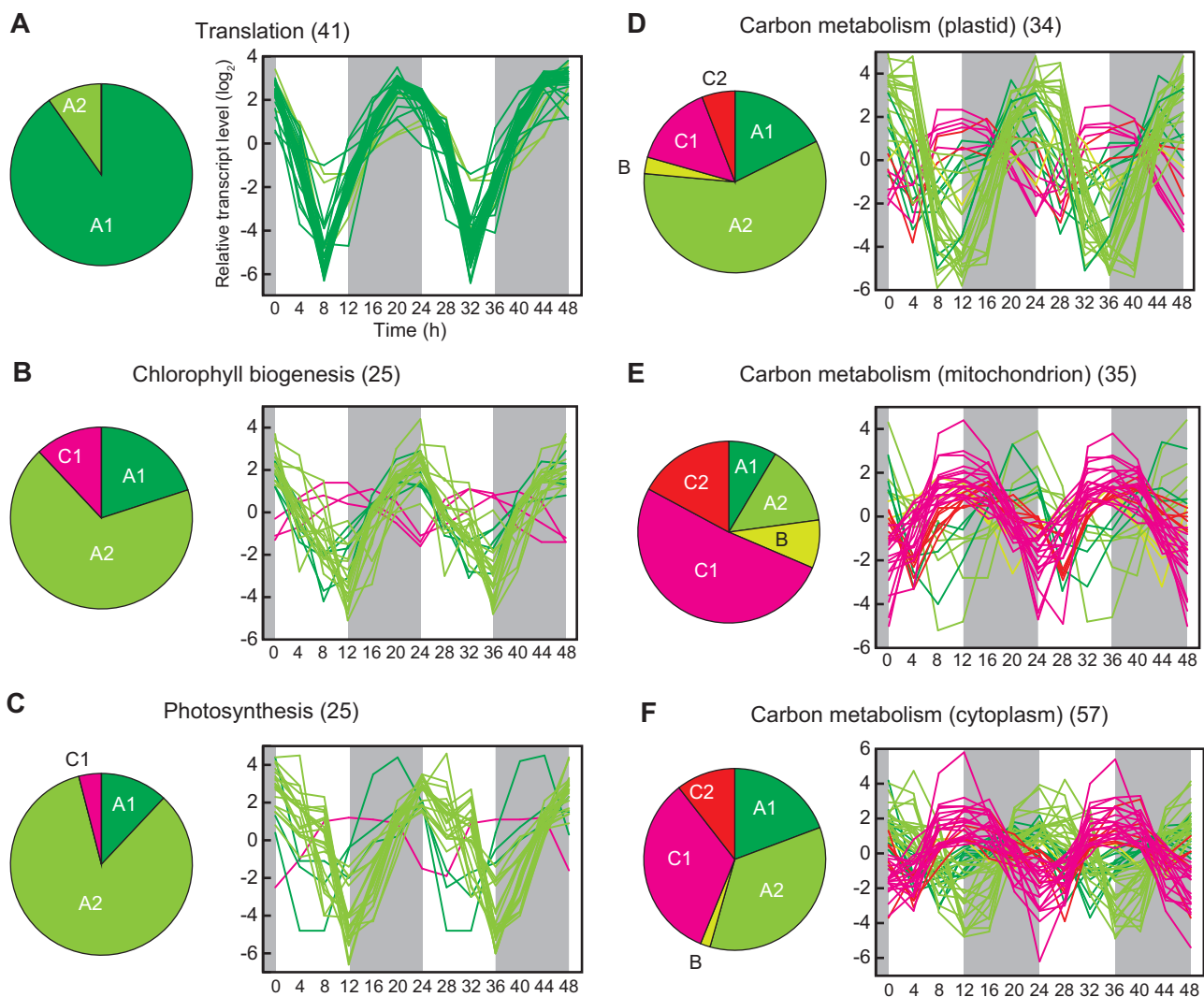

Fig. 4.—Expression profiles of nuclear genes encoding plastid-targeted proteins. Plots of relative transcript levels for *B. natans* plastid-targeted proteins related to specific biological processes: translation (A), chlorophyll biogenesis (B), photosynthesis (C), and carbon metabolism in the plastid (D), mitochondrion (E), and cytoplasm (F). The number of proteins for respective biological processes is shown in parentheses. Each pie chart indicates the proportion of expression clusters. Dark periods are indicated by gray shading.

and cytoplasm-targeted DEGs (ctDEGs) were grouped into cluster A2 and C1 at nearly equal rates (fig. 4E and F; supplementary figs. S6 and S7, Supplementary Material online). It is interesting to note that transcript levels of the translation-related ptDEGs increase slightly earlier than those of other ptDEGs involved in photosynthesis and metabolisms (figs. 3A and 4A–F), and a similar feature has been also reported in unicellular green algae (Idoine et al. 2014). Products of the translation-related ptDEGs would be built up in the plastids prior to translation of plastid-encoded proteins, and their preceded expression appears to adjust the timing of protein biogenesis from nuclear and plastid genes involved in photosynthesis and metabolisms. Another interesting feature is that transcripts of B. natans ptDEGs are mostly induced in the dark periods, although the expression of plastid-associated genes generally peaks early in the light periods in other photosynthetic organisms (Monnier et al. 2010; Zones et al. 2015). Expression of plastid-associated genes in anticipation of dawn might be advantageous for efficient use of light energy during daytime. It is likely that transcription of many B. natans ptDEGs is mainly regulated by the cell cycle or circadian rhythm rather than the direct light stress.

Plant-like Circadian Clock Components

In the model plant Arabidopsis, circadian rhythms are driven by complex transcriptional/translational feedback loops constituted by central oscillator genes: Circadian Clock Associated 1 (CCA1), Late Elongated Hypocotoyl (LHY), and Timing of Cab expression 1 (TOC1) (Nagel and Kay 2012). Two paralogs of the Myb-related transcription factors CCA1 and LHY, whose transcript levels peak in the morning, directly repress TOC1 expression during the day. Conversely, transcription of TOC1 increases toward the evening, and its product acts to repress the expression of CCA1 and LHY. The circadian rhythms are entrained by light inputs via photoreceptors (Nefissi et al. 2011; Wenden et al. 2011; Pudasaini and Zoltowski 2013), and several nucleus-encoded genes involved in photosynthesis and carbon metabolism are regulated by circadian output pathways (Schaffer et al. 2001). Our transcript profiling results suggest that many B. natans ptDEGs seem to be controlled by the cell cycle or circadian rhythm, because of their periodic expression in the anticipation of dawn. Moreover, we found two DEGs (JGI#27638 in subcluster A1-4 and JGI#91355 in subcluster A2-1) that encoded proteins containing a conserved motif of Myb-related DNA-binding domain, like plant CCA1/LHY (fig. 5A and B), but we did not find any obvious homologs of TOC1. Three genes for plant-like cryptochromes (JGI#43343, JGI#19879, and JGI#58130) have been previously reported as candidates of blue-light photoreceptor (Fortunato et al. 2014), and all three have shown differential expression patterns that peaked in the late-dark or early-light periods, as in cluster A2 (fig. 5A). Although it remains unknown whether the chlorarachniophyte has a plant-like
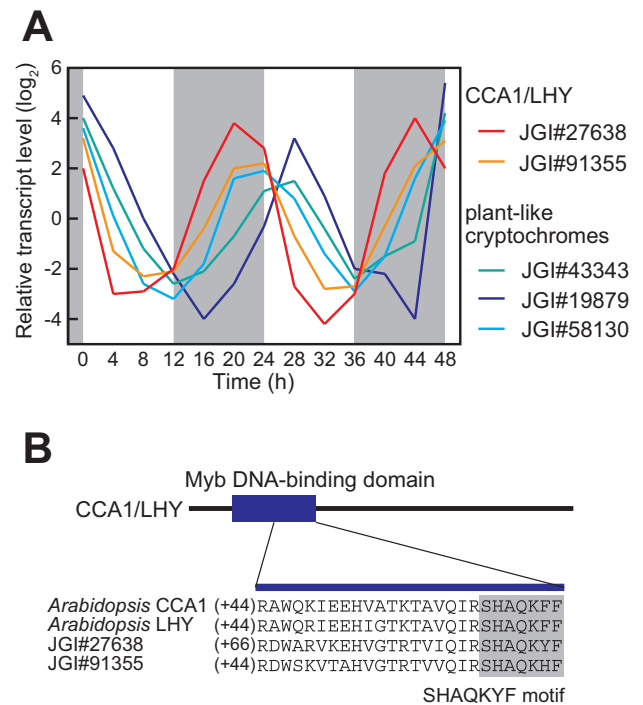

Fig. 5.—Expression profiles of putative plant-like circadian oscillators. (A) Plots of relative transcript levels for CCA1/LHY homologs and plant-like cryptochromes. Dark periods are indicated by gray shading. (B) Alignment of CCA1/LHY homologs of B. natans (JGI#27638 and JGI#91355) with CCA1 and LHY of Arabidopsis thaliana (AT1G01060 and AT2G46830). These sequences contain a conserved motif of Myb DNA-binding domain showing the blue box.

circadian system, some of the above-described candidates might be related to transcriptional regulation of B. natans ptDEGs.

DNA Replication in the Nucleus and Nucleomorph

Eukaryotic nuclear DNA replication is a conserved mechanism comprising multiple proteins showing cell cycle-specific expression. Based on the KEGG classification of B. natans nuclear genes, we identified several DEGs involved in nuclear DNA replication. These DEGs were grouped into cluster C1 and C2 that exhibited high expression before the initiation of cell division in the dark period. In anticipation of dusk, early S phase-specific genes encoding the origin recognition complex (ORC), cell division control protein 6 (CDC6), and minichromosome maintenance proteins (MCM2 to MCM7) showed peak expression at 8/32 h (fig. 6A). Subsequently, transcripts of core components, including DNA polymerases (POLA and POLD), replication factors (RFC and RPA), flap endonuclease and helicase (FEN1 and DNa2), and proliferating cell nuclear antigen (PCNA), peaked at 12/36 h just before dusk. Some DEGs for the condensin I complex (SMC2, SMC4, YCG1, YCG4, and BRN1) involved in mitotic chromosome condensation were
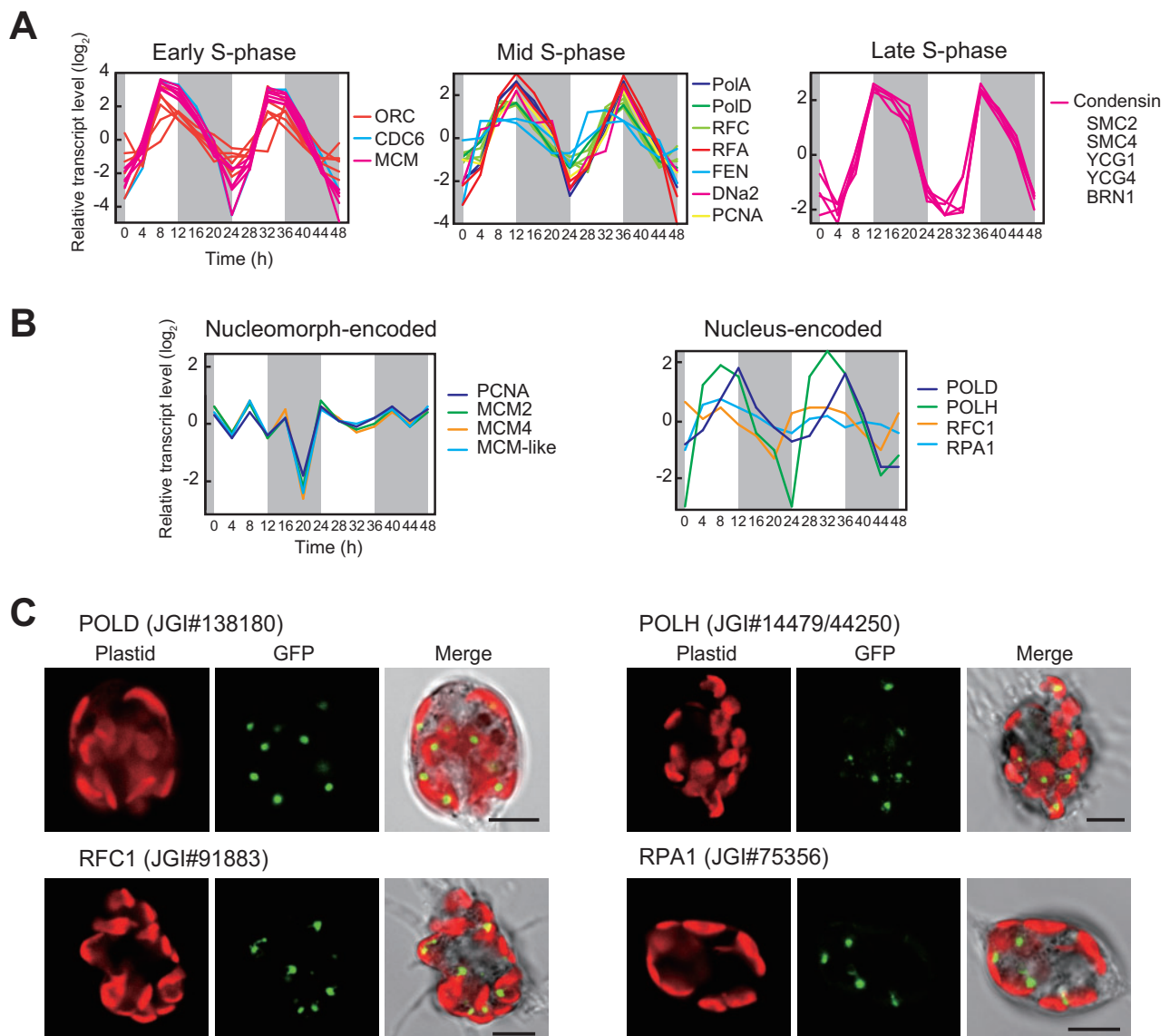
FIG. 6.—Gene expression for nuclear and nucleomorph DNA replication. (*A*) Plots of relative transcript levels of *B. natans* nuclear genes associated with the nuclear DNA replication. The genes are divided into three groups (Early, Mid, and Late S-phase) by their transcription peaks and predicted function. Dark periods are indicated by gray shading. (*B*) Plots of relative transcript levels for nucleus- and nucleomorph-encoded proteins associated with the nucleomorph DNA replication. (*C*) GFP localization of four nucleus-encoded nucleomorph DNA replication proteins (POLD, POLH, RFC1, and RPA1) of *B. natans*. The confocal images show GFP fluorescence heterologously expressed in another chlorarachniophyte, *Amorphochlora amoebiformis*. The red color is the chlorophyll-autofluorescence and the green color represent GFP localization in the nucleomorphs. Scale bars are 5 μm.

grouped into cluster C2, and exhibited high expression early in the dark periods (fig. 6*A*). Overall, the chlorarachniophyte shows typical eukaryotic DNA replication mechanisms in the nucleus. In the cell cycle, the S phase occurs approximately at the time of transition from the light to dark period, and the M phase occurs early in the dark period.

Similar to the nucleus, the nucleomorph should also have a mechanism for DNA replication tightly regulated by the host cell cycles, because *B. natans* commonly has a single plastid with a single nucleomorph per cell (Moestrup and Sengco

2001). The nucleomorph genome carries only a few genes involved in DNA replication (e.g., PCNA, MCM2, and MCM4), the expression of which was not regulated during the cell cycle (fig. 6*B*). This finding suggests that core components of nucleomorph DNA replication are encoded by the nuclear genome. Based on our subcellular localization prediction, 147 nucleus-encoded proteins were predicted to be targeted to the nucleomorph, of which 4 exhibited obvious sequence homology to known DNA replication factors: JGI#138180, JGI#144779/44250 (a fused gene model),

JGI#91883, and JGI#75356, which are homologs of DNA polymerase delta (POLD), DNA polymerase eta (POLH), RFC1, and RPA1, respectively. Moreover, we confirmed that these four proteins were targeted to the nucleomorph using heterologously expressed GFP fusion proteins in *Amorphochlora amoebiformis*, which is the only transformable chlorarachniophyte species (fig. 6C). GFP fluorescence was observed as several small dots near the plastid autofluorescence, which is the typical pattern of nucleomorph localization (Hirakawa et al. 2009). Transcript levels of the nucleomorph-targeted POLD (JGI#138180) peaked at 12/36 h, similar to that observed for nuclear DNA replication factors, whereas the transcripts of the nucleomorph-targeted RFC1 and RPA1 did not exhibit clear oscillation (fig. 6B). The nucleomorph-targeted POLH, the expression of which was restricted in the light period, might be involved in the nucleomorph DNA repair. Collectively, these findings suggest that nucleomorph DNA replication appears to be regulated by the core DNA polymerase POLD encoded by the nuclear genome through the cell cycle. Interestingly, the gene for nucleomorph-targeted POLD was found to be phylogenetically related to the sequences of giant viruses found in green algae (Blanc et al. 2015) (supplementary fig. S8, Supplementary Material online), and a homologous gene of nucleomorph-targeted POLD was identified in another chlorarachniophyte, *Lotharella globosa* (MMETSP0041# 25121). This suggests that a central part of nucleomorph DNA replication machinery has been replaced by a viral-like DNA polymerase in chlorarachniophytes. It is possible that the ancestral host cell of chlorarachniophytes engulfed a green algal endosymbiont infected by a giant virus during the secondary endosymbiosis. We previously reported that nucleomorph-encoded proteins are evolving considerably faster than their nuclear counterparts in chlorarachniophytes (Hirakawa et al. 2014). The high evolutionary rate of nucleomorph genes might be partially attributed to replication mediated by the non-canonical DNA polymerase. The replacement of bacterial RNA and/or DNA polymerases by viral sequences has been also reported in endosymbiotic organelles, such as mitochondria and plastids (Filée and Forterre 2005); for example, mitochondrial DNA polymerase gamma was shown to be phylogenetically related to DNA polymerases of T3/T7 bacteriophages (Filée et al. 2002). This observation implies the existence of similar selective pressures that have resulted in replacement of the original organelle DNA polymerases by certain viral counterparts during their endosymbiotic evolution.

Materials and Methods

Synchronous Culture and RNA Extraction

Bigelowiella natans CCMP621 cells were maintained at 20 °C under white light condition with 12:12-h light:dark cycles in 250 mL polystyrene flask (Iwaki) containing 125 mL ESM medium (Kasai et al. 2009). Cell division was temporally synchronized as previously described (Hirakawa et al. 2011). Cells were sampled every 4 h during the second and third diurnal cycles after light deprivation for 36 h, and total RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's protocol; 7.7–18.7 µg total RNA was purified from each time point (supplementary table S2, Supplementary Material online). The cell number was monitored using a hemocytometer (fig. 1A).

RNA Sequencing and Genome Mapping Analyses

For mRNA-seq by Illumina HiSeq 2000, library construction and sequencing were performed by Eurofins Genomics. The pair-end libraries were constructed with 200 bp cDNA inserts, and sequenced using TruSeq SBS Kit v3 (200-cycle). At least 22 million paired-end sequences were obtained from each sample, resulting in a total of 80.6 G bases of 798 million paired-end sequences from the 13 samples (supplementary table S2 Supplementary Material online). The raw reads were deposited in the DDBJ Sequence Read Archive under accession numbers DRA004608 to DRA004620. Both ends of RNA reads were trimmed with a quality value less than Q30, and short reads ($\leq$50 bases) were discarded using Trimmomatic 0.32 (Bolger et al. 2014). The resulting reads of each sample were mapped on the nuclear (Curtis et al. 2012) and nucleomorph genome (Gilson et al. 2006) of *B. natans* by using TopHat 2.0.13 (Trapnell et al. 2009) without multi hits (-g 1); the genome sequences were downloaded from the JGI website (http://genome.jgi.doe.gov/Bigna1/Bigna1.home.html, last accessed Aug 9, 2016) and GenBank (accession numbers DQ158856 to DQ158858). Mapped reads on the gene-coding regions were counted by HTSeq 0.6.1 (Anders et al. 2014). Data of the coding regions of the nuclear genome were obtained from the Ensembl website (http://protists.ensembl.org/Bigelowiella_natans/Info/Index, last accessed Aug 9, 2016). For the nucleomorph genome, either copy of duplicated genes (*dnaK*, *rpl23*, *rps8*, and *rps15*) was eliminated in the mapping analysis. The read counts for the 13 samples examined were normalized based on their library sizes using an iDEGES/DESeq2 method (the $DESeq2\text{-}(DESeq2\text{-}DESeq2)_n$ pipeline with $n = 3$, using DESeq2 1.6.3) (Love et al. 2014; Sun et al. 2013). Datasets from the two diurnal cycles were treated as biological replicates for the normalization. To identify DEGs during the cycles, the normalized counts of each gene were compared between different time points by baySeq 2.0.50 (Hardcastle and Kelly 2010) with 100,000 resamplings, and genes showing false discovery rates less than 0.001 were identified as DEGs.

Clustering Analyses of DEGs

To investigate the transcriptional networks of DEGs, we performed a cluster analysis of 7,753 DEGs based on their expression patterns using Cluster 3.0 (de Hoon et al. 2004).

Normalized read counts of each gene were transformed to log$_2$ and centered by median values. A hierarchical clustering with the complete-linkage method was applied. Clustered expression patterns were visualized using Java TreeView 3.0 (Saldanha 2004).

Reconstruction of Gene Models and Prediction of Protein Localization

To identify more accurate gene models, we reconstructed nuclear gene models of *B. natans* using the vast amounts of comprehensive mRNA sequences obtained in the present study. Model reconstruction was performed using the Program to Assemble Spliced Alignments (PASA) reconstruction transcript pipeline (Haas et al. 2003) with RNA contigs assembled using Trinity 2.0.4 (Grabherr et al. 2011). We manually discarded the open reading frames that lacked start codons and/or coded less than 30 amino acids, thus resulting in 21,043 protein-coding genes. The gene models were visualized and checked on the Integrated Genome Browser (Helt et al. 2009). Data of the refined gene models generated by PASA were deposited at the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.6n332). We predicted subcellular localization (plastid, PPC, nucleomorph, ER/Golgi, membrane, and mitochondrion) of nucleus-encoded proteins based on their N-terminal signal regions and sequence homologies with functionally annotated proteins in other organisms. The detailed procedure is described in supplementary figure S9, Supplementary Material online.

Localization Analyses of GFP Fusion Proteins

Subcellular localization of nucleus-encoded nucleomorph-targeted DNA polymerase delta (POLD, JGI#138180), DNA polymerase eta (POLH, JGI#144779/44250), replication factor C (RFC1, JGI#91883), and replication protein A (RPA1, JGI#75356) of *B. natans* was confirmed using GFP fusion proteins. Partial cDNA fragments encoding N-terminal regions of these proteins were amplified by PCR and inserted into the GFP expression vector pLaRGfp + mc; 263 amino acids (aa) of POLD, 214 aa of POLH, 187 aa of RFC1, and 143 aa of RPA were used. To analyze subcellular localization of GFP fusion proteins, *Amorphochlora amoebiformis* cells, which are the only transformable species in chlorarachniophytes, were transiently transformed using a Biolistic PDS-1000/He particle delivery system (Bio-Rad), as described previously (Hirakawa et al. 2009). Confocal microscopy images of GFP expressing cells were obtained by an inverted Zeiss LSM 510 laser scanning microscope.

Supplementary Material

Supplementary materials and methods, figures S1–S9, and tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Literature Cited

Anders S, Pyl PT, Huber W. 2014. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Archibald JM. 2015. Endosymbiosis and eukaryotic cell evolution. Curr Biol. 25:R911–R921.

Blanc G, Gallot-Lavallée L, Maumus F. 2015. Provirophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. Proc Natl Acad Sci U S A. 112:E5318–E5326.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Burki F, et al. 2016. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc Biol Sci. 283:20152802.

Curtis BA, et al. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492:59–65.

de Hoon MJL, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. Bioinformatics 20:1453–1454.

Eberhard S, Finazzi G, Wollman FA. 2008. The dynamics of photosynthesis. Annu Rev Genet. 42:463–515.

Farinas B, et al. 2006. Natural synchronisation for the study of cell division in the green unicellular alga *Ostreococcus tauri*. Plant Mol Biol. 60:277–292.

Filée J, Forterre P. 2005. Viral proteins functioning in organelles: a cryptic origin? Trends Microbiol. 13:510–513.

Filée J, Forterre P, Sen-Lin T, Laurent J. 2002. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. J Mol Evol. 54:763–773.

Fortunato AE, Annunziata R, Jaubert M, Bouly JP, Falciatore A. 2014. Dealing with light: the widespread and multitasking cryptochrome/photolyase family in photosynthetic organisms. J Plant Physiol. 172:42–54.

Fujiwara T, et al. 2009. Periodic gene expression patterns during the highly synchronized cell nucleus and organelle division cycles in the unicellular red alga *Cyanidioschyzon merolae*. DNA Res. 16:59–72.

Gile GH, Keeling PJ. 2008. Nucleus-encoded periplastid-targeted EFL in chlorarachniophytes. Mol Biol Evol. 25:1967–1977.

Gillard J, et al. 2008. Physiological and transcriptomic evidence for a close coupling between chloroplast ontogeny and cell cycle progression in the pennate diatom *Seminavis robusta*. Plant Physiol. 148:1394–1411.

Gilson PR, et al. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus. Proc Natl Acad Sci U S A. 103:9566–9571.

Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. Annu Rev Plant Biol. 59:491–517.

Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.

Haas BJ, et al. 2003. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31:5654–5666.

Hardcastle TJ, Kelly KA. 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422.

Helt GA, et al. 2009. Genoviz Software Development Kit: Java tool kit for building genomics visualization applications. BMC Bioinformatics 10:266.

Genome Biology and Evolution

SMBE

Hibberd DJ, Norris RE. 1984. Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta divisio nova, Chlorarachniophyceae classis nova). J Phycol 20:310–330.

Hirakawa Y, Nagamune K, Ishida K. 2009. Protein targeting into secondary plastids of chlorarachniophytes. Proc Natl Acad Sci U S A. 106:12820–12825.

Hirakawa Y, Burki F, Keeling PJ. 2011. Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. Mol Microbiol. 80:1439–1449.

Hirakawa Y, Gile GH, Ota S, Keeling PJ, Ishida K. 2010. Characterization of periplastidal compartment-targeting signals in chlorarachniophytes. Mol Biol Evol. 27:1538–1545.

Hirakawa Y, Ishida K. 2015. Prospective function of FtsZ proteins in the secondary plastid of chlorarachniophyte algae. BMC Plant Biol. 15:276.

Hirakawa Y, Suzuki S, Archibald JM, Keeling PJ, Ishida K. 2014. Overexpression of molecular chaperone genes in nucleomorph genomes. Mol Biol Evol. 31:1437–1443.

Hovde BT, et al. 2015. Genome sequence and transcriptome analyses of *Chrysochromulina tobin*: metabolic tools for enhanced algal fitness in the prominent order Prymnesiales (Haptophyceae). PLOS Genet. 11:e1005469.

Idoine AD, Boulouis A, Rupprecht J, Bock R. 2014. The diurnal logic of the expression of the chloroplast genome in *Chlamydomonas reinhardtii*. PLoS One 9:e108760.

Imoto Y, Yoshida Y, Yagisawa F, Kuroiwa H, Kuroiwa T. 2011. The cell cycle, including the mitotic cycle and organelle division cycles, as revealed by cytological observations. Microscopy 60:S117–S136.

Kanesaki Y, Imamura S, Minoda A, Tanaka K. 2012. External light conditions and internal cell cycle phases coordinate accumulation of chloroplast and mitochondrial transcripts in the red alga *Cyanidioschyzon merolae*. DNA Res. 19:289–303.

Kasai F, Kawachi M, Erata M, Yumoto K, Sato M. 2009. NIES-collection list of strains, 8th edition. Jpn J Phycol 57:220.

Keeling PJ. 2013. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu Rev Plant Biol. 64:583–607.

Kucho K, Okamoto K, Tabata S, Fukuzawa H, Ishiura M. 2005. Identification of novel clock-controlled genes by cDNA macroarray analysis in *Chlamydomonas reinhardtii*. Plant Mol Biol. 57:889–906.

Lien T, Knutsen G. 1979. Synchronous growth of *Chlamydomonas reinhardtii* (Chlorophyceae): a review of optimal conditions. J. Phycol 15:191–200.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Ma L, et al. 2001. Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways. Plant Cell 13:2589–2607.

Miyagishima S, Suzuki K, Okazaki K, Kabeya Y. 2012. Expression of the nucleus-encoded chloroplast division genes and proteins regulated by the algal cell cycle. Mol Biol Evol. 29:2957–2970.

Moestrup Ø, Sengco M. 2001. Ultrastructural studies on *Bigelowiella natans*, gen. et sp. nov., A chlorarachniophyte flagellate. J Phycol 37:624–646.

Monnier A, et al. 2010. Orchestrated transcription of biological processes in the marine picoeukaryote *Ostreococcus* exposed to light/dark cycles. BMC Genomics 11:192.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–W185.

Nagel DH, Kay SA. 2012. Complexity in the wiring and regulation of plant circadian networks. Curr Biol. 22:R648–R657.

Nefissi R, et al. 2011. Double loss-of-function mutation in *EARLY FLOWERING 3* and *CRYPTOCHROME 2* genes delays flowering under continuous light but accelerates it under long days and short days: an important role for *Arabidopsis* CRY2 to accelerate flowering time in continuous light. J Exp Bot 62:2731–2744.

Noordally ZB, Millar AJ. 2015. Clocks in Algae. Biochemistry 54:171–183.

Price DC, et al. 2012. *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. Science 335:843–847.

Pudasaini A, Zoltowski BD. 2013. Zeitlupe senses blue-light fluence to mediate circadian timing in *Arabidopsis thaliana*. Biochemistry 52:7150–7158.

Rodríguez-Ezpeleta N, et al. 2005. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. Curr Biol. 15:1325–1330.

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.

Saldanha AJ. 2004. Java Treeview–extensible visualization of microarray data. Bioinformatics 20:3246–3248.

Schaffer R, et al. 2001. Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. Plant Cell 13:113–123.

Sun J, Nishiyama T, Shimizu K, Kadota K. 2013. TCC: an R package for comparing tag count data with robust normalization strategies. BMC Bioinformatics 14:219.

Suzuki S, Hirakawa Y, Kofuji R, Sugita M, Ishida K. 2016. Plastid genome sequences of *Gymnochlora stellata*, *Lotharella vacuolata*, and *Partenskyella glossopodia* reveal remarkable structural conservation among chlorarachniophyte species. J Plant Res. 129:581–590.

Tanifuji G, et al. 2014. Nucleomorph and plastid genome sequences of the chlorarachniophyte *Lotharella oceanica*: convergent reductive evolution and frequent recombination in nucleomorph-bearing algae. BMC Genomics 15:374.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Wenden B, et al. 2011. Light inputs shape the *Arabidopsis* circadian system. Plant J. 66:480–491.

Zones JM, Blaby IK, Merchant SS, Umen JG. 2015. High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. Plant Cell 27:2743–2769.

Associate editor: John Archibald