Molecular Microbiology (2011) 80(6), 1439–1449 ■
doi:10.1111/j.1365-2958.2011.07643.x
First published online 6 April 2011

Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga

Yoshihisa Hirakawa, Fabien Burki and Patrick J. Keeling*
Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, V6T 1Z4, Canada.

Summary

The plastid of chlorarachniophytes is distinguished by the retention of a relict nucleus (nucleomorph) derived from a green algal endosymbiont, which is located in the periplastidal compartment (PPC). The nucleomorph genome of a chlorarachniophyte, *Bigelowiella natans*, encodes several plastid-targeted proteins and hundreds of housekeeping proteins, but it lacks many fundamental genes to maintain itself. Here we report the first two host nucleus-encoded genes for proteins targeted to the nucleomorph, histone H2A and H2B. We identified 20 histone genes from the host nuclear genome, and based on phylogenetic analyses predicted that most of these are derived from the host, but that two histone genes are symbiont-derived. The genes both encode N-terminal extensions resembling PPC targeting signals, further suggesting they function in the nucleomorph. Using green fluorescent protein (GFP) fusion proteins expressed in transformed cells, we confirmed that the putative symbiont H2A and H2B were targeted into the nucleomorph, whereas putative host proteins were localized to the host nucleus. Furthermore, we have developed a method to temporarily synchronize *B. natans* cells, and confirmed that both host and symbiont histone expression is controlled during the cell cycle. Our findings provide the first evidence of how the nucleomorph may be regulated by host-encoded gene products.

Introduction

Plants and diverse lineages of algae possess plastids that have been acquired by multiple endosymbiotic events between a non-photosynthetic and a photosynthetic organism. The plastids of land plants, green algae, red algae and glaucophytes are surrounded by two envelope membranes, and are derived from a single primary endosymbiosis with a cyanobacterium (Moreira *et al.*, 2000; Rodríguez-Ezpeleta *et al.*, 2005). Many other algal groups, heterokonts, haptophytes, dinoflagellates, apicomplexans, cryptophytes, euglenophytes and chlorarachniophytes, acquired plastids by multiple secondary endosymbioses with green or red algae (Keeling, 2004; 2010); their plastids are surrounded by three or four envelope membranes and are referred to as secondary, or complex plastids (Cavalier-Smith, 2000). Following these events, many of the genes encoded by the endosymbiont genomes are lost or transferred to the host nucleus (Martin *et al.*, 1998; Bock and Timmis, 2008). The majority of these transferred genes commonly encode N-terminal pre-sequences, which mediate their targeting into the plastid. In organisms with primary plastids, precursors of nucleus-encoded plastid proteins have a plastid targeting sequence called a transit peptide (TP), and are post-translationally transported into the plastid stroma across two envelope membranes (Bruce, 2001; Steiner and Löffelhardt, 2002). In the case of secondary plastids, nucleus-encoded plastid pre-proteins have a bipartite targeting sequence consisting of a signal peptide (SP) followed by a transit peptide-like sequence (TPL) at their N-termini (Patron and Waller, 2007). The SP is involved in co-translational targeting to the endoplasmic reticulum (ER), and the TPL is necessary for delivering proteins from the ER into the plastid stroma across three or four plastid membranes (Bolte *et al.*, 2009).

In most secondary plastids, the nucleus of the endosymbiont has been lost. However two algal groups, chlorarachniophytes and cryptophytes, still maintain a relic nucleus of their green and red algal endosymbionts respectively (Archibald, 2007). The relic nucleus, called a nucleomorph, is located between the inner and outer pair of plastid membranes in the periplastidal compartment (PPC), which is the remnant cytoplasm of the endosymbionts. The genome size of these nucleomorphs is in all cases highly reduced, estimated to be 330–610 kb in chlorarachniophytes (Silver *et al.*, 2007) and 485–845 kb in cryptophytes (Tanifuji *et al.*, 2010). Nucleomorph genomes have been sequenced for a chlorarachniophyte, *Bigelowiella natans*

Accepted 17 March, 2011. *For correspondence. E-mail pkeeling@mail.ubc.ca; Tel. (+1) 604 822 2845; Fax (+1) 604 822 6089.

(373 kb: Gilson *et al.*, 2006), and three cryptophytes, *Guillardia theta* (551 kb: Douglas *et al.*, 2001), *Hemiselmis andersenii* (572 kb: Lane *et al.*, 2007) and *Cryptomonas paramecium* (486 kb: Tanifuji *et al.*, 2011). These nucleomorph genomes encode predominantly housekeeping genes and several putative plastid-targeted protein genes. This implies that nucleomorphs are an essential organelle for expressing these plastid-targeted proteins and maintaining plastids. However, all sequenced nucleomorph genomes lack many fundamental genes for maintenance and self-replication (notably, for example, DNA polymerases). It is thought that such genes were most likely transferred to the host nuclear genome and their products should be transported into the PPC.

In a previous study, we described a nucleus-encoded gene for a protein targeted to the PPC from several species of chlorarachniophytes, translation elongation factor-like (EFL) (Gile and Keeling, 2008). EFL pre-proteins have an N-terminal bipartite targeting sequence, consisting of a SP and a TPL, similar to many plastid-targeted pre-proteins. Both plastid- and PPC-targeted pre-proteins appear to be transported by the same pathway, and are sorted in the PPC according to their ability or inability to cross the inner two membranes; TPLs of plastid-targeted pre-proteins possess several positively charged amino acids, resulting in a pronounced net positive charge that is essential for passing through the inner pair of plastid membranes, whereas PPC-targeted TPLs have a net negative or near-neutral charge (Hirakawa *et al.*, 2009; 2010). Although the nucleomorph genome of *B. natans* lacks many fundamental protein genes, to date EFL is the only PPC-targeted protein that has been detected. Identifying additional nucleus-encoded PPC-targeted proteins is essential for a better understanding of the evolutionary history of nucleomorph, how these complex targeting systems function, as well as whether and how the host exerts control over the endosymbiont in matters such as timing of cell cycle transitions.

Here, we describe two PPC-targeted pre-proteins found in the nuclear genome of *B. natans*, histone H2A and H2B, the first known proteins to be targeted to a nucleomorph, and also the first such proteins whose regulation by the host has implications for control over endosymbiont functions. Histone proteins are essential components of eukaryotic chromosomes and are divided into five classes, namely histone H2A, H2B, H3, H4 and H1. Two copies of each core histone, H2A, H2B, H3 and H4, assemble in an octamer, which forms a nucleosome with DNA (Luger *et al.*, 1997), and nucleosomes further assemble into a chromatin structure with linker histones H1 (Thomas, 1999). While the nucleomorph genome of *B. natans* encodes H3 and H4 protein genes, other histone genes are absent (Gilson *et al.*, 2006). We found 20 histone genes of every class from the nuclear genome sequence of *B. natans*. While H1, H3 and H4 genes all appear to encode host histone proteins, two phylogenetically distinct types of each of histone H2A and H2B were identified, one of which was predicted to be derived from the endosymbiont (no putative symbiont H1 could be identified in either host or nucleomorph genome). Green fluorescent protein (GFP) fusion proteins for host and symbiont H2A and H2B genes were expressed in the related chlorarachniophyte *Amorphochlora amoebiformis* (*Lotharella amoebiformis* was placed into a new genus *Amorphochlora* by Ishida *et al.*, 2010), confirming that the putative symbiont-specific histones were targeted into the nucleomorph. To examine whether the host-encoded genes were co-ordinated with the cell cycle, we also developed a method to synchronize *B. natans* cultures and confirmed that the expression of both host- and symbiont-specific histone genes was controlled during the cell cycle using real-time quantitative PCR (RT-qPCR). These findings suggest that the nucleomorph's nucleosome consists of core histones encoded by the nuclear and nucleomorph genomes, and that the composition of the symbiont nucleosome is regulated by the host.

Results

Phylogeny of B. natans *histone H2A and H2B proteins*

The *B. natans* nucleomorph genome encodes histone H3 and H4 genes, but lacks histone H2A, H2B and H1 genes. To detect these missing genes, we used the final assembly genome data of *B. natans* sequenced by JGI (DOE Joint Genome Institute). Twenty histone genes were identified in the nuclear genome; these genes encoded five homologues of H2A (JGI Protein ID: #26990, #46712, #48254, #56563 and #92772), four of H2B (#42386, #50018, #68607 and #136283), six of H3 (#35667, #39940, #49481, #50205, #57231 and #14157), three of H4 (#38173, #46709 and #56340) and two H1 (34697 and 87300). A copy of both H2A (#56563) and H2B (#50018) encode an N-terminal extension that is found in no other histone protein. The phylogenetic relationships of *B. natans* H2A and H2B genes were analysed by phylogenetic reconstruction (Fig. 1). In both cases, *B. natans* histone sequences were separated into two distantly related clades. Multiple slow-evolving paralogues were found to be related to EST sequences from another chlorarachniophyte species, *Gymnochlora stellata*, and in turn weakly related to heterokonts (Fig. 1). However, for both histones the single sequence with the N-terminal extension was found to be a fast-evolving gene, and to branch elsewhere in the tree: the H2A was found to be the sister group to haptophytes, whereas the H2B branched with the streptophytes/green algae clade (Fig. 1). Nucleomorph-derived genes are expected to

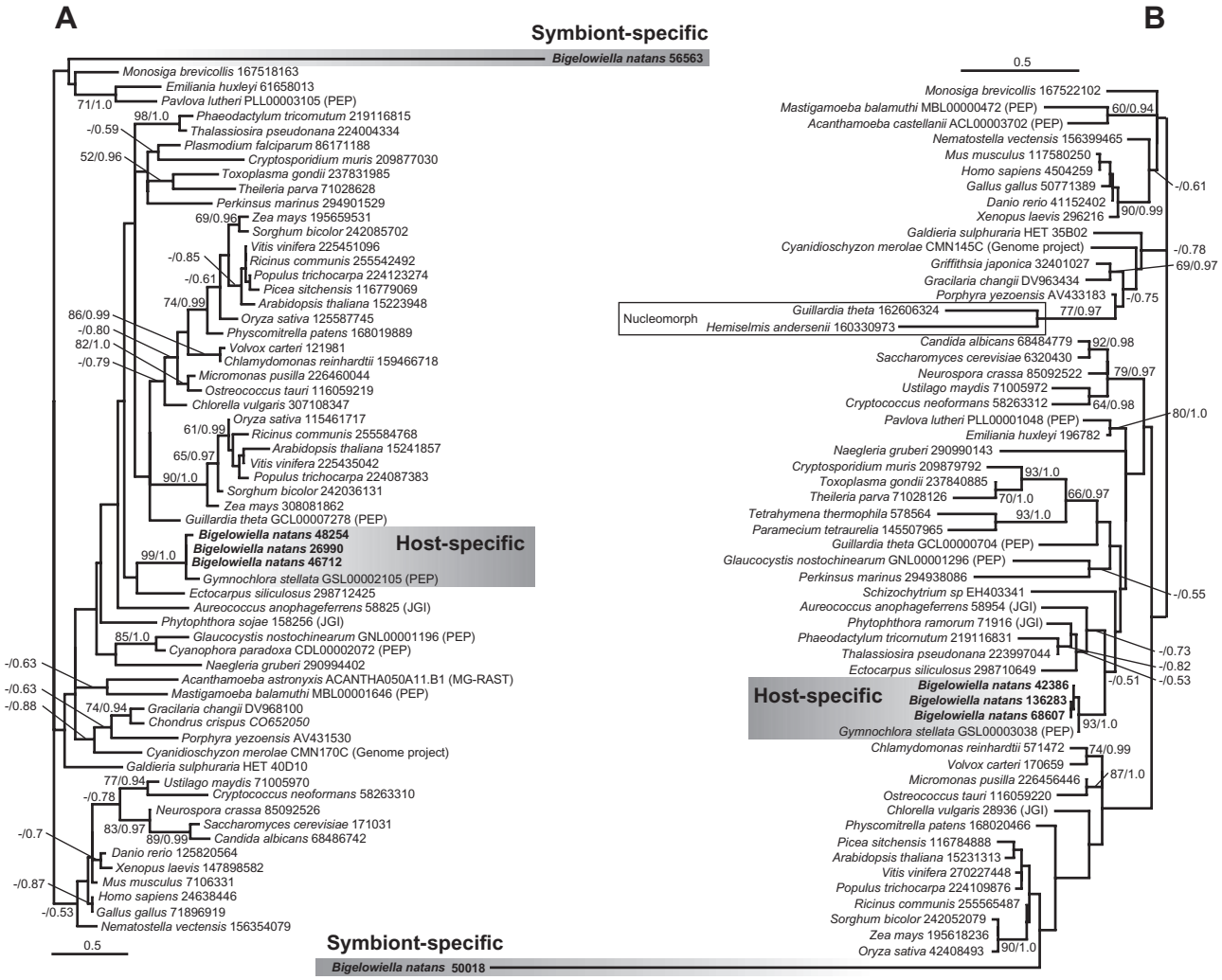

Fig. 1. Unrooted phylogenetic trees representing the eukaryotic diversity of histone H2A (A) and H2B (B) proteins. *Bigelowiella natans* sequences are shown in bold, and chlorarachniophytes are highlighted with grey boxes. The topologies correspond to the best-scoring ML tree as obtained with RAxML, and the values at nodes the bootstrap and posterior probability supports when higher than 50% and 0.5 respectively. The scale bar represents the estimated number of amino acid substitutions per site.

be related to green algal homologues. However, the phylogenetic position of neither gene is statistically supported in our trees, likely due to the high rate of sequence substitution (as indicated by their long branches). Thus, neither the green algal relationship of H2B nor the haptophyte relationship of H2A should be accepted at face value. These phylogenetic data indicate that the nuclear genome of *B. natans* encodes two types of histone H2A/H2B genes, possibly with different evolutionary origins. Because the chlorarachniophyte endosymbiont is a green alga and the nucleomorph genome lacks histone H2A and H2B genes, the most obvious prediction is that the divergent, leader-encoding H2A/H2B genes are symbiont-specific histones, whereas the more conserved genes that lack leaders are host-specific histones. Interestingly, we could find no putatively symbiont-specific H1 gene in the genome sequence, despite the nucleomorph genome lacking H1.

Localization of nucleus-encoded H2A and H2B proteins

The N-terminal leader sequences encoded by the putative symbiont-specific H2A (#56563) and H2B (#50018) proteins share many characteristics with the pre-sequence of PPC-targeted proteins (Fig. 2A and B). Each N-terminal sequence possessed a hydrophobic region that was predicted to be a SP by the NN method of SP prediction server SignalP (Bendtsen *et al.*, 2004). The sequence following SP was predicted to be a TPL sequence by the chloroplast TP prediction server ChloroP (Emanuelsson *et al.*, 1999). Although there were no conserved sequence motifs among predicted TPL sequences for PPC targeted

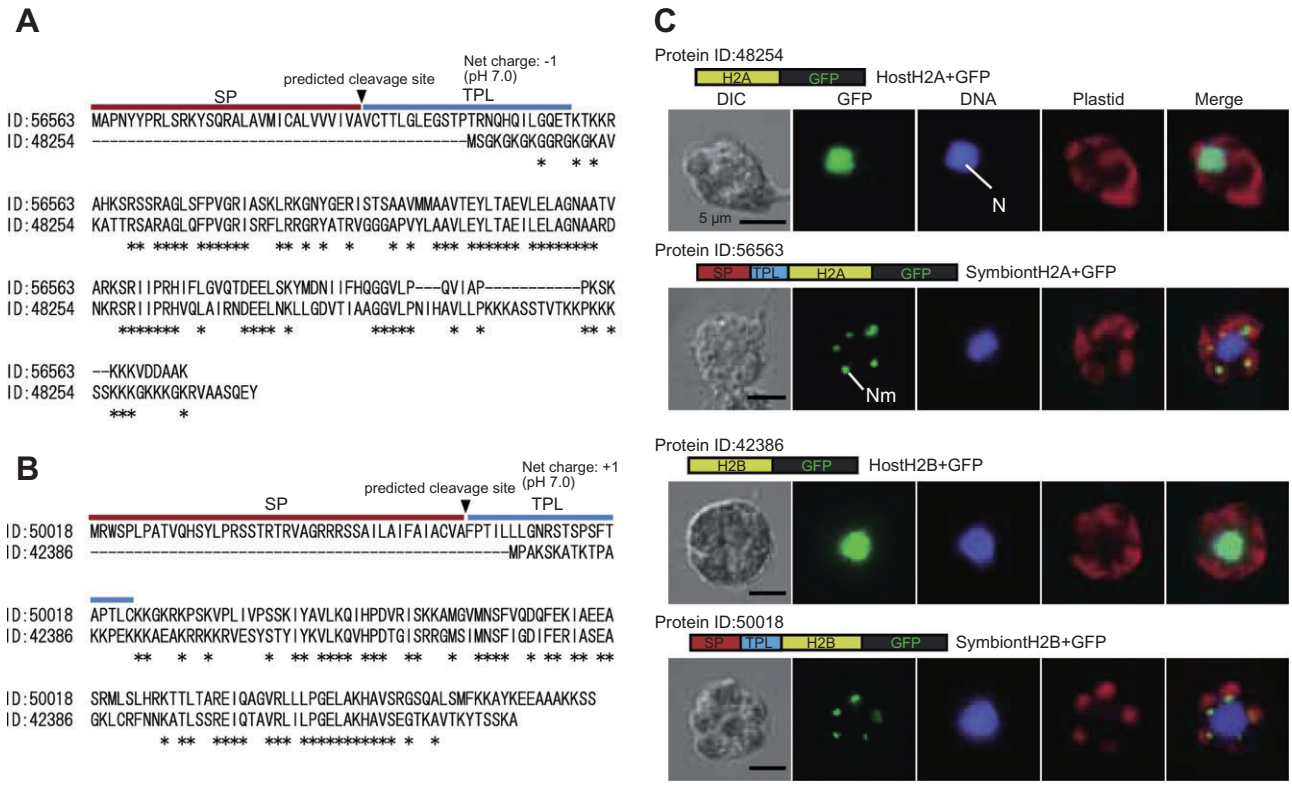

Fig. 2. *In vivo* targeting of GFP fused with each histone H2A and H2B protein.
A and B. Sequences of host- and symbiont-specific H2A and H2B proteins. Red and blue bars above the sequences show predicted signal peptide (SP) and transit peptide-like (TPL) regions respectively. Asterisks indicate conserved amino acids between host- and symbiont-specific histones.
C. GFP localization of a transformed cell expressing HostH2A+GFP, SymbiontH2A+GFP, HostH2B+GFP or SymbiontH2B+GFP. The images labelled 'GFP', 'DNA' and 'Plastid' show GFP localization (green), nuclear localization (blue) and chlorophyll-autofluorescence (red) respectively. 'Merge' indicates the combined image of these three images, and 'DIC' is a differential interference contract image. Scale bars: 5 µm. N, nucleus; Nm, nucleomorph.

pre-proteins, the predicted TPL sequences of histones were calculated to possess a net negative charge (–1) or a weakly net positive (+1) at pH 7.0, which is a feature of PPC targeting signals; the TPLs of plastid targeting signals have a pronounced net positive charge (+5.6 ± 2.6) (Hirakawa *et al*., 2010). In order to confirm the localizations of these H2A and H2B proteins, we used the GFP and the genetic transformation system of a chlorarachniophyte, *A. amoebiformis*. We constructed plasmids that express GFP fused with full-length histone proteins of four types: hostH2A+GFP, hostH2B+GFP, symbiontH2A+GFP and symbiontH2B+GFP. Each of these constructs was used to transform *A. amoebiformis* cells. Both hostH2A+GFP and hostH2B+GFP resulted in green fluorescence colocalizing with Bisbenzimide H-stained host nuclei (Fig. 2C). In contrast, fluorescence of both symbiontH2A+GFP and symbiontH2B+GFP was restricted in small spots that were adjacent to chlorophyll autofluorescence of plastids, and distinct from the Bisbenzimide H-stained nuclei (Fig. 2C). Although we could not detect Bisbenzimide H-stained nucleomorph because of a low amount of nucleomorph DNA, these GFP localizations were similar to the localization of PPC containing a nucleomorph; the localization pattern of PPC was analysed by GFP fluorescence and immunogold localizations in our previous studies (Hirakawa *et al*., 2009; 2010). These results suggest that the host-specific H2A and H2B proteins function in the host nucleus, and that the symbiont-specific H2A and H2B are targeted into the PPC by their N-terminal bipartite targeting signals, and should therefore function in the nucleomorph.

Expression of histone genes during a cell cycle

The expression of eukaryotic histone genes is generally tightly controlled during the cell cycle, and is largely limited to the S-phase when new chromatin production is at its peak (Osley, 1991). To examine the expression pattern of the host- and the symbiont-specific H2A and H2B genes during a cell cycle, we calculated expression levels by RT-qPCR in synchronized cell cultures of

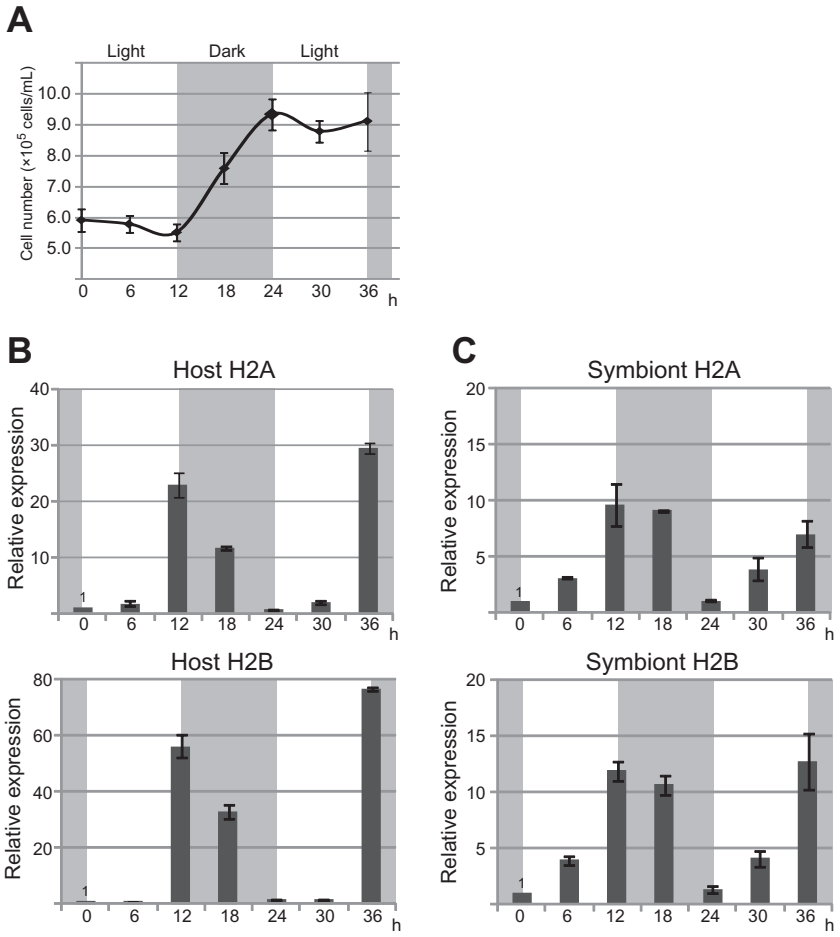

Fig. 3. Expression pattern of host- and symbiont-specific histone genes.
A. Cell densities (cells ml^{-1}) in the synchronized culture of *B. natans* measured at 6 h intervals. Dark periods are indicated by shaded areas.
B. Relative expression of host-specific histone genes detected at 6 h intervals during the second light, dark, and the third light periods. Each expression was normalized against levels of GAPDH, and values set to 1 at time point 0 h.
C. Relative expression of symbiont-specific histone genes.
Error bars represent the SD of triplicate experiments.

B. natans. Since no method to synchronize *B. natans* existed, we first developed a system based on light deprivation. Stationary-phase cells were transferred into a fresh ESM medium and maintained in darkness for 36 h, followed by 12:12 h light : dark cycles. Synchronized cell division appeared in the second dark period of the 12:12 h light : dark cycle (Fig. 3A). Cells in mitosis and cytokinesis were observed during the second dark period (Fig. S1), but these cells did not abound in the culture (approximately 2% at the middle of the dark period). This implies that cell divisions definitely occur during the dark period although the synchrony of cell division is not high. The cells were collected at 6 h intervals during the second light and dark periods, and the third light period, and total RNA was extracted. RNA from seven time points (0, 6, 12, 18, 24, 30 and 36 h) was analysed for expression of the host-specific H2A/H2B and the symbiont-specific H2A/H2B by RT-qPCR.

Relative expressions of histone genes were given based on normalizations to two reference genes, GAPDH (glyceraldehydes-3-phosphate dehydrogenase) or 18S rRNA, and expression patterns were similar in both cases using these two reference genes. We show the relative expressions of each histone gene normalized against GAPDH in Fig. 3 (expressions normalized by 18S rRNA show in Fig. S2). In the case of host-specific histones, both expression patterns of the H2A and the H2B were similar, and two peaks of expressions were found at 12 and 36 h time points: immediately prior to the initiation of a dark cycle (Fig. 3B). These expression levels were 23- to 76-fold higher than the expression level at time point 0 h. This, together with the cell division pattern, suggests that S-phase cells were dominant just before the dark period, that M-phase takes place during the dark period, and that the early light periods would be dominated by G1-phase cells.

Expression of both symbiont-specific H2A and H2B also peaked at 12 and 36 h, similar to the expression of host-specific histones (Fig. 3C). However, there are a few differences between the expression patterns of host- and symbiont-specific histones. The expression increase of symbiont-specific histones appeared at mid-light periods (at 6 and 30 h), which was earlier than the host-specific homologues. Furthermore, the expression decrease of

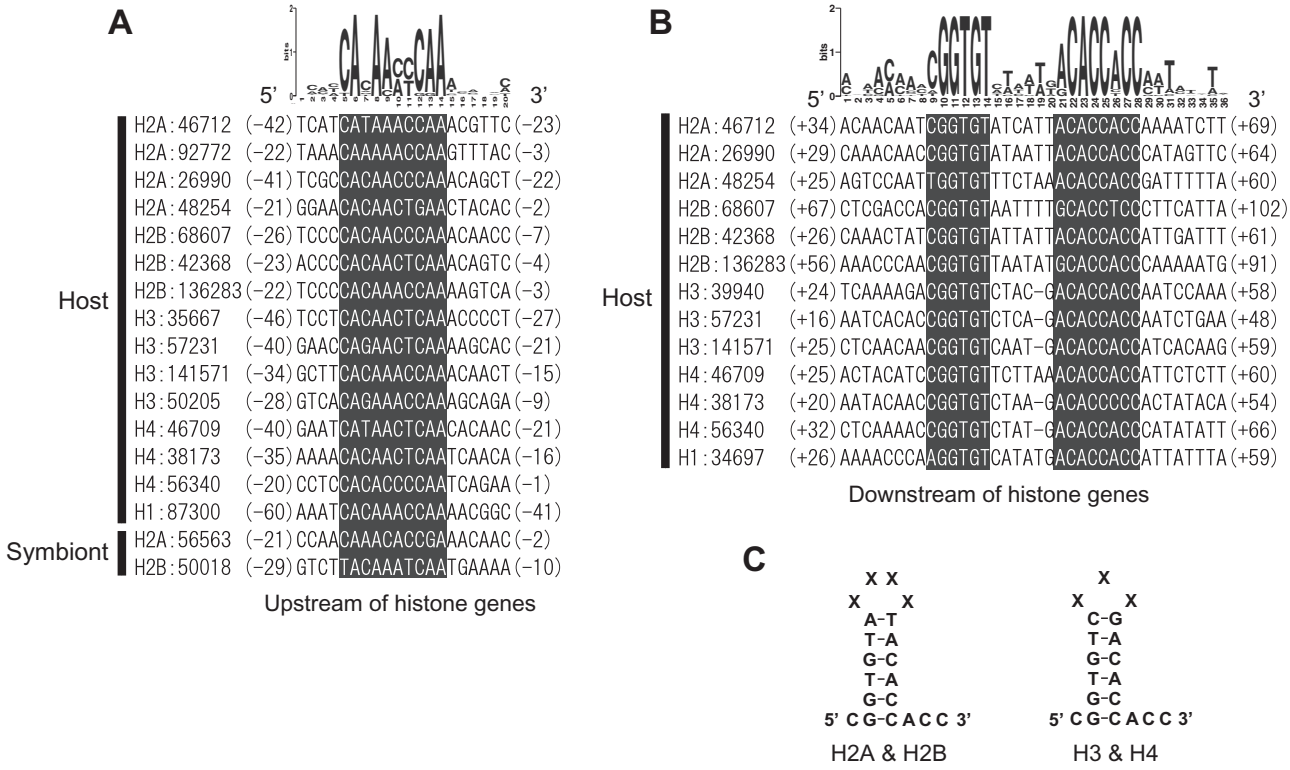

Fig. 4. Alignments of upstream and downstream sequences of *B. natans* histone genes.
A. Conserved motif sequences of upstream region (between positions −57 and −7) among 17 histone genes identified from *B. natans* genome sequence. The schema above the alignment showing the relative frequency of each amino acid at that position was created by WebLogo (Crooks *et al.*, 2004).
B. Conserved motifs of downstream regions (between positions +22 and +94) predicted to form a stem-loop structure.
C. Schema of the predicted stem-loop structures.

symbiont-specific histone did not appear at 18 h in the M-phase, whereas the expression of host-specific histones decreased by half at the same time point.

Conserved motif sequences of promoters and 3′ UTR among histone genes

In diverse eukaryotes, promoter region and 3′ UTR of histone genes generally possess conserved motif sequences that are involved in controlling their expression during a cell cycle (Chowdhary *et al.*, 2005; Mariño-Ramírez *et al.*, 2006; Dávila López and Samuelsson, 2008). To search for such sequences among *B. natans* histone genes, the MEME program (Bailey and Elkan, 1994) was applied to the upstream region (−300 to −1 from start codons) and downstream region (+1 to +300 from stop codons) of each of the 20 histone genes, including five H2A, four H2B, six H3, three H4 and two H1 genes. In the upstream regions, we identified the conserved motif CACAA[C/A][C/T]CAA between positions −57 and −7 (Fig. 4A). This A/C-rich sequence was detected in 17 genes, including both symbiont-specific histone genes. In the downstream regions, the conserved motif CGGTGT[A/C]XXX(X)[T/G]ACACCACC was identified between positions +22 and +94 in 13 genes, all of which were host-specific histones (Fig. 4B). This motif contains a palindrome sequence predicted to form a stem-loop structure (Fig. 4C). Such sequences have been found in replication-dependent histone genes of diverse eukaryotes (Dávila López and Samuelsson, 2008), and an RNA-binding protein called SLBP (stem-loop-binding protein) is bound to it (Jaeger *et al.*, 2005). The SLBP play a pivotal role in processing and stabilization of histone mRNA only during the S-phase. We also identified a SLBP-like gene (#130229) from the *B. natans* genome sequence. Five pairs of host-specific histone genes, such as H2B (#42386) and H3 (#141571), H4 (#38173) and H2A (#48254), H4 (#46709) and H2A (#46712), H4 (#56340) and H2A (#26990), H1 (#34697) and H3 (#57231), were encoded in a neighbouring region of the same scaffold respectively (Fig. S3). Each pair possessed a bidirectional promoter, and an inverted sequence was found in the boundary between the promoters (Fig. S3). In contrast, the symbiont-specific H2A and H2B genes were encoded in distant positions.

Interestingly, the symbiont-specific histone genes all contained introns, whereas most host-specific histones possessed no intron. We looked for introns in homologous histone genes in two green algae, *Chlamydomonas reinhardtii* and *Volvox carteri*, which are comparatively closely related to the green algal endosymbiont of chlorarachniophytes. In their complete genome sequences (Merchant *et al*., 2007; Prochnik *et al*., 2010), we found 38 homologues of H2A and 36 homologues of H2B genes, collectively. While most of these genes contained no introns, three genes, an H2A (JGI ID: #187884) and an H2B (#121902) of *C. reinhardtii*, and an H2A (#109344) of *V. carteri*, possessed introns. However, there is no similarity in the number or position of introns between the symbiont-specific histones in *B. natans* and green algal histones. This implies that the symbiont-specific histones of *B. natans* may have acquired their introns after the gene transfer from the green algal endosymbiont, or if they were in the endosymbiont ancestor that they are not at ancient, conserved positions.

Discussion

The host nuclear genome of the chlorarachniophyte *B. natans* was found to encode two types of histone H2A/H2B proteins, and our phylogenetic analyses indicated that these types have distinct evolutionary origins. One type encodes an N-terminal bipartite targeting sequence similar to PPC-targeting signals, and *in vivo* GFP localization experiments showed these histones are indeed targeted to the PPC, the compartment where the nucleomorph is located. Since the nucleomorph genome lacks H2A and H2B genes, but does encode H3 and H4 genes, the nucleosome of *B. natans* nucleomorph must be formed by two nucleomorph-encoded histones (H3 and H4) and two nucleus-encoded histones (H2A and H2B). Most likely, the H2A and H2B genes were previously encoded in the nucleomorph genome but were transferred to the host nuclear genome, acquired the N-terminal targeting sequences to direct their products into the PPC, and were subsequently lost in the nucleomorph genome. In cryptophytes, a distantly related lineage of algae that have retained a relict nucleomorph of red algal origin, three sequenced nucleomorph genomes of *G. theta*, *H. andersenii* and *C. paramecium* encode histone H2B, H3 and H4 genes, but lack H2A gene (Douglas *et al*., 2001; Lane *et al*., 2007; Tanifuji *et al*., 2011). This implies that the host nuclear genomes of these algae also possess a symbiont-specific H2A gene whose product should be targeted into the PPC, and that their nucleomorph nucleosomes must also be composed on proteins encoded in two different genomes. Indeed, we found a putative nucleomorph-targeted H2A protein (ID: #100176) possessing an N-terminal extension from the preliminary genome sequence data of *G. theta* (JGI).

The assembly of eukaryotic nucleosomes into higher-order structures, chromosomes, is stabilized by a linker histone H1 (Thomas, 1999). In this study, we could find no symbiont-specific H1 gene in either the host nucleus or nucleomorph genome of *B. natans*. Because H1 proteins are more variable in sequence compared with other core histone proteins, it is possible it simply was not detected in our searches of the genome sequence. However, we cannot exclude possibility that the gene has been lost altogether and that some other protein is involved in the compaction of nucleosomes, or that nucleomorph DNA does not assume the typical chromosome structure, especially since no condensed DNA has been observed in the nucleomorph (Moestrup and Sengco, 2001).

Expression of histones is known to be controlled during the cell cycle in many eukaryotes. A massive amount of histone proteins is necessary to compact the newly synthesized genomic DNA during S-phase into chromatin, whereas an excess of histone is generally toxic for cells during G1-, G2- or M-phase (Meeks-Wagner and Hartwell, 1986; Osley, 1991) (although we observed no evidence for toxicity in our transient transformants). Therefore, histone expression is tightly regulated according to the cell cycle. In this study, we demonstrated that the expression of both host-specific H2A/H2B and symbiont-specific H2A/H2B of a chlorarachniophyte was controlled during the cell cycle, using the synchronized cell culture of *B. natans*. The expression of host-specific H2A and H2B was increased in the S-phase just before the start of cell division, and decreased during the M-phase. During G1-phase, the expression was quite low in comparison with the S-phase. In diverse eukaryotes, replication-dependent histone genes share common regulatory elements in their promoter regions and 3′ UTR. The histone promoters of divergent organisms such as yeast, fly and mammals are known to share common potential transcription factor binding sites among heterogeneous histone genes (Chowdhary *et al*., 2005; Mariño-Ramírez *et al*., 2006), and many pairs of histone genes form heterogeneous clusters which share a bidirectional promoter region (Matsumoto and Yanagida, 1985; Kremer and Hennig, 1990). Furthermore, the 3′ UTR of histone genes typically possesses a conserved motif that form a stem-loop structure, which is involved in the processing and stabilization of histone mRNA only during the S-phase (Jaeger *et al*., 2005). Like replication-dependent histone genes of diverse eukaryotes, most of *B. natans* host-specific histone genes (including at least one copy of each H2A, H2B, H3, H4 and H1) have a stem-loop structure in their 3′ UTR, and several pairs of histone genes share a bidirectional promoter. In addition, a conserved motif (A/C-rich sequence) was found in the putative

promoter regions of most host-specific histones. This motif might be involved in controlling the S-phase-specific expressions. All of these common features indicate that the expression of host-specific histone genes might be tightly co-regulated during the cell cycle. In the symbiont-specific H2A and H2B, the maximum expression appeared in the S-phase, like their host-specific counterparts, and their promoters also possess the conserved motif that was found in the promoter region of host-specific histone genes. This might mean that the expression of symbiont- and host-specific histones is co-regulated during the cell cycle, perhaps by the same system. However, the increase of symbiont-specific histone expressions appeared somewhat earlier than that of host-specific histones, and did not decrease immediately during the M-phase. Therefore, the expression control of symbiont-specific histones may occur independently by some host system. Indeed, 3′ UTR of these symbiont-specific histone genes lack a stem-loop structure that is important for the stabilization of histone mRNA only during the S-phase. Interestingly, nucleomorph division has been reported to take place before nuclear mitosis in *B. natans* (Moestrup and Sengco, 2001), which implies that the regulation of symbiont histone expression should be independent of that of the host, and is also consistent with our finding that expression peaks earlier. Moreover, dividing nucleomorph genomes would be expected to require significantly less histone than the host genome simply because the DNA contents of the two genomes are so different, again suggesting the necessity for differences in regulation strategies. In contrast, we would predict that the nucleomorph-encoded histones might share similar regulation profiles to their PPC-targeted counterparts, although the strategies to achieve this may also be different due to the nature of the nucleomorph genome and its gene expression (Williams *et al*., 2005; Gilson *et al*., 2006).

Overall, these findings give new insight that the fundamental components of nucleomorph are encoded in the host nuclear genome and are regulated by the host. The nucleomorph genome lacks many genes for DNA metabolism, especially all DNA polymerases (Gilson *et al*., 2006). These missing genes should also be encoded in the host nuclear genome, and their expression must also be controlled by the host, but to date only two histones and one putative translation factor have been identified. The control exerted by the host over its symbiont must in large part be derived from controlling the expression of such PPC- and nucleomorph-targeted protein genes, so in the future it will be interesting to explore how this control takes place, how independent it is from control of equivalent host systems, and whether control systems are also imported from the symbiont or were co-opted from existing host control systems.

Experimental procedures

Phylogenetic analyses

Homologues of the *B. natans* histone proteins were identified by BLASTP searches against GenBank (or otherwise indicated in Fig. 1: JGI, PEP, MG-RAST and *Cyanidioschyzon merolae* genome project), retrieved and automatically aligned with the L-INS-I method of the MAFFT package (Katoh *et al*., 2005). Gblocks (Castresana, 2000) was used to eliminate poorly aligned positions, with half of the gapped positions allowed, the minimum number of sequences for a conserved and a flank position set to 50% of the number of taxa plus one, the maximum of contiguous non-conserved positions set to 12, and the minimum length of a block set to 4, followed by manual inspection of the alignments using Seaview (Gouy *et al*., 2010). This procedure led to alignments composed of 106 and 108 unambiguously aligned amino acids for H2A and H2B respectively. Maximum likelihood (ML) analyses were performed using RAxML 7.2.8 (Stamatakis, 2006), with the rapid hill-climbing algorithm and the LG + Γ + F model of evolution (-m PROTGAMMALGF, four discrete rate categories). The best-scoring ML tree was determined for both proteins in multiple searches using 20 randomized stepwise addition parsimony starting trees. Statistical support was evaluated with non-parametric bootstrapping using 1000 replicates. Bayesian analyses were performed using MrBayes 3.2 (Huelsenbeck and Ronquist, 2001) and the WAG + Γ + F model (eight gamma categories). The difference in the amino acid replacement matrix between the ML and Bayesian analyses was due to the absence of implementation of LG in MrBayes. The inference has consisted of 1 500 000 generations with sampling every 100 generations, starting from a random starting tree and using 4 Metropolis-coupled Markov Chain Monte Carlo (MCMCMC). Two separate runs were performed to confirm the convergence of the chains with the average standard deviation of split frequencies used to assess the convergence of the two runs. Bayesian posterior probabilities were calculated from the majority rule consensus of the tree sampled after an initial burnin period corresponding to 20% of the samples (300 000 generations) and verified by plotting the maximum likelihood values of the two runs.

Targeting sequence prediction

To predict the SP and the TPL of histone H2A and H2B, we used the NN method of SP prediction server SignalP (Bendtsen *et al*., 2004) and the chloroplast TP prediction server ChloroP (Emanuelsson *et al*., 1999) respectively. The cleavage site of each TPL was also estimated using an alignment of mature regions of homologous histones. The overall charge of each TPL was calculated using the Peptide Property Calculator (Innovagen, Lund, Sweden). Assumptions and the equation used can be viewed at http://www.innovagen.se/custom-peptide-synthesis/peptide-property-calculator/peptide-property-calculator.asp.

Plasmid constructs

In order to construct plasmid vectors, phostH2B+GFP, phostH2B+GFP, psymbiontH2A and psymbiontH2B, cDNA

fragments of four histone gene (JGI Protein ID: #48254, #42386, #56563, #50018) were amplified by PCR (primer sets are listed in Table S1), and each fragment was inserted between HindIII and NcoI sites of the pLaRGfp+mc vector (Hirakawa *et al*., 2009). Since every fragment contained a HindIII or a NcoI site in the coding region, we remove these site using a PCR-based site-directed mutagenesis technique (Higuchi *et al*., 1988). All plasmids were cloned in the DH5α strain of *Escherichia coli* and purified using a QIAprep Spin Miniprep Kit (QIAGEN). These constructs were subsequently sequenced to ensure correct construction.

Transient transformation of A. amoebiformis *cells*

To prepare for transformation, *A. amoebiformis* (CCMP2058) (Ishida *et al*., 2000; Ishida *et al*., 2011) was grown at 20°C under white illumination (55–60 μmol photons m^{-2} s^{-1}) on a 12:12 h light : dark cycle in 500 ml Erlenmeyer flasks containing 250 ml of ESM medium (Kasai *et al*., 2009). We used a Biolistic PDS-1000/He Particle Delivery System (Bio-Rad) for the transformation of *A. amoebiformis* cells. The cells were then bombarded by each plasmid in the best condition as described previously (Hirakawa *et al*., 2008). After the bombardment, the cells were immediately transferred into new plastic plates with 10 ml of fresh ESM medium and incubated under the conditions described above. Twenty-four hours after the bombardment, the cells were fixed in 2% paraformaldehyde/2% NaCl/10 μg ml^{-1} Bisbenzimide H 33258 (Sigma) in PHEM buffer (60 mM PIPES/25 mM HEPES/10 mM EGTA/2 mM $MgCl_2$; pH 7.4), and then GFP-expressing cells were observed under an Axioplan 2 fluorescent microscope (Carl Zeiss AG) with a 3CCD HD video camera XL H1S (Canon). GFP fluorescence, plastid autofluorescence and Bisbenzimide fluorescence were detected with the filter 17 (excitation BP 485/20 nm, emission BP 515–565), the filter 15 (excitation BP 546/12 nm, emission LP 590 nm) and the filter 1 (excitation BP 365/12 nm, emission LP 397 nm) respectively.

Synchronized culture of B. natans

Bigelowiella natans (CCMP621) cells were maintained at 20°C under white illumination (55–60 μmol photons m^{-2} s^{-1}) on a 12:12 h light : dark cycle in 250 ml polystyrene flasks (BD Falcon) containing 125 ml of ESM medium. For the synchronization, stationary-phase culture (cell density was approximately 4×10^6 cells ml^{-1}) was prepared by culturing for 2 weeks. The cells were transferred into fresh ESM medium with the cell density of 5×10^5 cells ml^{-1}, and then maintained in darkness for 36 h, followed by 12:12 h light- : dark cycles. Synchronized cell division occurred during the second dark period in the light : dark cycles. Cell number was monitored at intervals of 6 h with an Improved Neubauer haemocytometer.

RT-qPCR assay

Total RNA was isolated from the synchronized culture of *B. natans* every 6 h during the second light and dark, and the third light period, using TRIzol Reagent (Invitrogen) according to the manufacturer's protocol. Each total RNA sample was quantified by NanoDrop-1000 (Thermo Scientific), and 2 μg was used for cDNA synthesis. The reverse transcription was performed in total volume 20 μl with random hexamers and SuperScript II Reverse Transcriptase (Invitrogen). Primers for quantification/notarization of each histone gene (host-specific H2A, #46712; host-specific H2B, #42386; symbiont-specific H2A, #56563; symbiont-specific H2B, #50018) and two reference genes (GAPDH, #56635 and 18S rRNA, AF054832) were designed using the Primer3Plus online software (primer sequences are listed in Table S2). RT-qPCR was carried out on the iCycler iQ5 Real-Time PCR System (Bio-Rad) under following conditions: 5 μl of diluted cDNA, 1 μl of each primer (0.4 μM), 12.5 μl of iQ SYBR Green Supermix (Bio-Rad) and 5.5 μl of DNase/RNase-free water. The cycling conditions comprised 3 min of denaturation at 95°C followed by 40 cycles of 10 s at 95°C, 30 s at 58°C, and a melting curve programme (55–95°C with a heating rate of 0.5°C per 10 s). Relative quantification was calculated by the ΔΔCT method (Pfaffl, 2001), and the qPCR efficiencies were calculated from the slopes of the standard curves of serial dilutions in four steps, and found to be nearly identical for each primer set (between 1.97 and 2.05).

Acknowledgements

We thank Joyce Wang and the Kronstad lab for granting access to the Particle Delivery System, and Richard White and the Rieseberg lab for access to the Real-Time qPCR System. This work was funded by a grant from the Natural Sciences and Engineering Research Council of Canada (227301), and by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. We thank the US Department of Energy Joint Genome Institute for sequencing the nuclear genome of *B. natans* and J.M. Archibald, M.W. Gray, G.I. McFadden and C.E. Lane for project contributions. F.B. is supported by a prospective researcher postdoctoral fellowship from Swiss National Science Foundation. P.J.K. is a Fellow of the Canadian Institute for Advanced Research and a Senior Scholar of the Michael Smith Foundation for Health Research.

References

Archibald, J.M. (2007) Nucleomorph genomes: structure, function, origin and evolution. *BioEssays* **29:** 392–402.

Bailey, T.L., and Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28–36.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: signalP 3.0. *J Mol Biol* **16:** 783–795.

Bock, R., and Timmis, J.N. (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. *BioEssays* **30:** 556–566.

Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S., and Maier, U.G. (2009) Protein targeting into secondary plastids. *J Eukaryot Microbiol* **56:** 9–15.

Bruce, B.D. (2001) The paradox of plastid transit peptides: conservation of function despite divergence in primary structure. *Biochim Biophys Acta* **1541:** 2–21.

Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540–552.

Cavalier-Smith, T. (2000) Membrane heredity and early chloroplast evolution. *Trends Plant Sci* **5:** 174–182.

Chowdhary, R., Ali, R.A., Albig, W., Doenecke, D., and Bajic, V.B. (2005) Promoter modeling: the case study of mammalian histone promoters. *Bioinformatics* **21:** 2623–2628.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14:** 1188–1190.

Dávila López, M., and Samuelsson, T. (2008) Early evolution of histone mRNA 3′ end processing. *RNA* **14:** 1–10.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.T., *et al.* (2001) The highly reduced genome of an enslaved algal nucleus. *Nature* **410:** 1091–1096.

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* **8:** 978–984.

Gile, G.H., and Keeling, P.J. (2008) Nucleus-encoded periplastid-targeted EFL in chlorarachniophytes. *Mol Biol Evol* **25:** 1967–1977.

Gilson, P.R., Su, V., Slamovits, C.H., Reith, M.E., Keeling, P.J., and McFadden, G.I. (2006) Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci USA* **103:** 9566–9571.

Gouy, M., Guindon, S., and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27:** 221–224.

Higuchi, R., Krummel, B., and Saiki, R.K. (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interaction. *Nucleic Acids Res* **16:** 7351–7367.

Hirakawa, Y., Kofuji, R., and Ishida, K. (2008) Transient transformation of a chlorarachniophyte alga, *Lotharella amoebiformis* (chlorarachniophyceae), with *uid*A and *egfp* reporter genes. *J Phycol* **44:** 814–820.

Hirakawa, Y., Nagamune, K., and Ishida, K. (2009) Protein targeting into secondary plastids of chlorarachniophytes. *Proc Natl Acad Sci USA* **106:** 12820–12825.

Hirakawa, Y., Gile, G.H., Ota, S., Keeling, P.J., and Ishida, K. (2010) Characterization of periplastidal compartment-targeting signals in chlorarachniophytes. *Mol Biol Evol* **27:** 1538–1545.

Huelsenbeck, J.P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Ishida, K., Ishida, N., and Hara, Y. (2000) *Lotharella amoeboformis* sp. nov.: a new species of chlorarachniophytes from Japan. *Phycol Res* **48:** 221–229.

Ishida, K., Yabuki, A., and Ota, S. (2011) *Amorphochlora amoebiformis* gen. et comb. nov. (Chlorarachniophyceae). *Phycol Res* **59:** 52–53.

Jaeger, S., Barends, S., Giegé, R., Eriani, G., and Martin, F. (2005) Expression of metazoan replication-dependent histone genes. *Biochimie* **87:** 827–834.

Kasai, F., Kawachi, M., Erata, M., Yumoto, K., Sato, M., and Ishimoto, M. (2009) NIES-Collection List of Strains, 8th edition. *Jpn J Phycol (Sôrui)* **57:** 220.

Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33:** 511–518.

Keeling, P.J. (2004) Diversity and evolutionary history of plastids and their hosts. *Am J Bot* **91:** 1481–1493.

Keeling, P.J. (2010) The endosymbiotic origin, diversification and fate of plastids. *Philos Trans R Soc Lond B Biol Sci* **365:** 729–748.

Kremer, H., and Hennig, W. (1990) Isolation and characterization of a *Drosophila hydei* histone DNA repeat unit. *Nucleic Acids Res* **18:** 1573–1580.

Lane, C.E., van den Heuvel, K., Kozera, C., Curtis, B.A., Parsons, B.J., Bowman, S., and Archibald, J.M. (2007) Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci USA* **104:** 19908–19913.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:** 251–260.

Mariño-Ramírez, L., Jordan, I.K., and Landsman, D. (2006) Multiple independent evolutionary solutions to core histone gene regulation. *Genome Biol* **7:** R122.

Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., and Kowallik, K.V. (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393:** 162–165.

Matsumoto, S., and Yanagida, M. (1985) Histone gene organization of fission yeast: a common upstream sequence. *EMBO J* **4:** 3531–3538.

Meeks-Wagner, D., and Hartwell, L.H. (1986) Normal stoichiometry of histone dimer sets is necessary for high fidelity of mitotic chromosome transmission. *Cell* **44:** 43–52.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., *et al.* (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318:** 245–250.

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

Moreira, D., Le Guyader, H., and Philippe, H. (2000) The origin of red algae and the evolution of chloroplasts. *Nature* **405:** 69–72.

Osley, M.A. (1991) The regulation of histone synthesis in the cell cycle. *Annu Rev Biochem* **60:** 827–861.

Patron, N.J., and Waller, R.F. (2007) Transit peptide diversity and divergence: a global analysis of plastid targeting signals. *BioEssays* **29:** 1048–1058.

Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29:** e45.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., *et al.* (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* **329:** 223–226.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., *et al.* (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* **15:** 1325–1330.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-

based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.
Steiner, J.M., and Löffelhardt, W. (2002) Protein import into cyanelles. *Trends Plant Sci* **7:** 72–77.
Silver, T.D., Koike, S., Yabuki, A., Archibald, J.M., and Ishida, K. (2007) Phylogeny and nucleomorph karyotype diversity of chlorarachniophyte algae. *J Eukaryot Microbiol* **54:** 403–410.
Tanifuji, G., Onodera, N.T., and Hara, Y. (2010) Nucleomorph genome diversity and its phylogenetic implications in cryptomonad algae. *Phycol Res* **58:** 230–237.
Tanifuji, G., Onodera, N.T., Wheeler, T.J., Dlutek, M., Donaher, N., and Archibald, J.M. (2011) Complete nucleomorph genome sequence of the non-photosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. *Genome Biol Evol* **3:** 44–54.
Thomas, J.O. (1999) Histone H1: location and role. *Curr Opin Cell Biol* **11:** 312–317.
Williams, B.A., Slamovits, C.H., Patron, N.J., Fast, N.M., and Keeling, P.J. (2005) A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci USA* **102:** 10936–10941.

Supporting information

Additional supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.