Complete nucleotide sequence of the chlorarachniophyte nucleomorph: Nature's smallest nucleus

Paul R. Gilson[*,†], Vanessa Su[†,‡], Claudio H. Slamovits[§], Michael E. Reith[¶], Patrick J. Keeling[§], and Geoffrey I. McFadden[‡,‖]

[*]Infection and Immunity Division, The Walter and Eliza Hall Institute of Medical Research, Parkville 3050, Australia; [‡]School of Botany, University of Melbourne, Victoria 3010, Australia; [¶]Institute for Marine Biosciences, National Research Council, Halifax, NS, Canada B3H 3Z1; and [§]Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Edited by Jeffrey D. Palmer, Indiana University, Bloomington, IN, and approved April 19, 2006 (received for review January 26, 2006)

The introduction of plastids into different heterotrophic protists created lineages of algae that diversified explosively, proliferated in marine and freshwater environments, and radically altered the biosphere. The origins of these secondary plastids are usually inferred from the presence of additional plastid membranes. However, two examples provide unique snapshots of secondary-endosymbiosis-in-action, because they retain a vestige of the endosymbiont nucleus known as the nucleomorph. These are chlorarachniophytes and cryptomonads, which acquired their plastids from a green and red alga respectively. To allow comparisons between them, we have sequenced the nucleomorph genome from the chlorarachniophyte *Bigelowiella natans*: at a mere 373,000 bp and with only 331 genes, the smallest nuclear genome known and a model for extreme reduction. The genome is eukaryotic in nature, with three linear chromosomes containing densely packed genes with numerous overlaps. The genome is replete with 852 introns, but these are the smallest introns known, being only 18, 19, 20, or 21 nt in length. These pygmy introns are shown to be miniaturized versions of normal-sized introns present in the endosymbiont at the time of capture. Seventeen nucleomorph genes encode proteins that function in the plastid. The other nucleomorph genes are housekeeping entities, presumably underpinning maintenance and expression of these plastid proteins. Chlorarachniophyte plastids are thus serviced by three different genomes (plastid, nucleomorph, and host nucleus) requiring remarkable coordination and targeting. Although originating by two independent endosymbioses, chlorarachniophyte and cryptomonad nucleomorph genomes have converged upon remarkably similar architectures but differ in many molecular details that reflect two distinct trajectories to hypercompaction and reduction.

plastid | secondary endosymbiosis | intron | endosymbiosis

The origin of plastids through endosymbiosis of a cyanobacterium-like prokaryote transferred photosynthesis into eukaryotes and launched a massive wave of diversification that subsequently generated a tremendous range of algae and plants (1). This initial event is referred to as primary endosymbiosis (Fig. 1) and created a plastid with two membranes such as those of green algae, plants, red algae, and glaucophyte algae (1). Transfer of genes from the endosymbiont to the nuclear genome of the host initially led to dependence of the endosymbiont on the host that was necessary to stabilize the partnership (2). Ongoing transfer has resulted in reduction of the prokaryotic genome, so that plastid DNA now represents probably <10% of its original gene content, and increasingly sophisticated regulation of the endosymbiont by the host has resulted in endosymbiont replication, gene expression, metabolic activity, and even death being managed by the eukaryotic host (3). Indeed, primary plastids seem to retain some autonomy only in the synthesis and deployment of redox proteins involved in photosynthetic electron transfer (4).

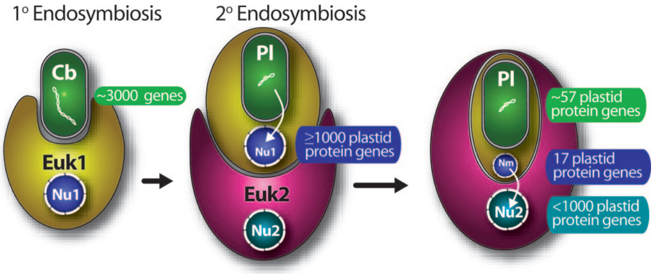

Fig. 1. Evolution of chlorarachniophytes such as *B. natans* by sequential endosymbioses. Enslavement of a photosynthetic, cyanobacterium-like prokaryote (Cb) introduces photosynthesis into a eukaryotic host (Euk 1), whose nucleus (Nu1) acquires at least 1,000 cyanobacterial genes over time. Secondary endosymbiosis involves capture and retention of the primary photosynthetic eukaryote by another eukaryote (Euk 2), producing a plastid with four bounding membranes such as those of cryptomonads, chlorarachniophytes, haptophyte and heterokont algae, and malaria parasites. Essential plastid protein genes are transferred from the endosymbiont nucleus (Nm, nucleomorph) to the nucleus (Nu2) of the second eukaryotic host. Here, we show that only 17 of the original plastid protein genes remain in the nucleomorph of *B. natans*, preventing its loss.

A key element of the host's ability to assert control over its "little green slaves" is its ability to deliver a regulated supply of essential components to the plastid. Proteins, particularly those encoded by endosymbiont genes appropriated by the host, are a vital part of these supplied components. Delivery of these proteins by a sophisticated protein import apparatus embedded in the plastid membranes was a crucial initial requirement for endosymbiotic gene transfer and went on to become a key factor in regulation of plastids by the host (5). It is estimated that the host delivers at least 1,000 different proteins to its plastid, the majority of which are targeted using an N-terminal extension known as a transit peptide, which is necessary and sufficient for translocation across the two plastid membranes (5).

Plastids also occur in a diverse range of eukaryotes apparently unrelated to the direct descendants of the primary plastid endosymbiosis. It is hypothesized that these eukaryotes acquired plastids by a process known as secondary endosymbiosis (Fig. 1) whereby a eukaryotic phagotroph engulfed and retained another plastid-containing eukaryote that was a descendant of the pri-

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ158856, DQ158857, and DQ158858).

See Commentary on page 9379.

[†]P.R.G. and V.S. contributed equally to this work.

[‖]To whom correspondence should be addressed. E-mail: gim@unimelb.edu.au.

mary endosymbiotic event (1). In this way, plastids were spread laterally into disparate branches of the eukaryotic tree (1). These photosynthetic (sometimes mixotrophic) organisms eventually assumed dominance in the world's oceans and are a major component of today's global phytoplankton community that supports ocean primary productivity and carbon recycling (6).

A number of independent secondary endosymbioses have occurred, but exactly how many is not yet resolved (1). Because numerous endosymbiont genes underwent relocation into the host nucleus as part of the primary endosymbiosis, the domesticated cyanobacterium was no longer fully autonomous. Secondary acquisition of a plastid is therefore not as straight forward as simply extracting the plastids and relocating them into a new host cell; numerous genes essential to plastid biogenesis and maintenance located in the primary host nucleus are also necessary for permanent integration of an acquired plastid (3). Thus, secondary endosymbioses probably initially involved engulfing an entire eukaryotic alga followed by a long period of integration. Examples of transient detentions of complete, photosynthetic eukaryotic cells are common in nature (7), and their long-term retention and conversion into a secondary endosymbiont requires coordination of division and segregation of the endosymbiont into daughter host cells (7). Endosymbiont-to-host gene transfer is probably an integral part of this process. Ultimately, this gene transfer may result in the secondary host acquiring all of the necessary genetic resources to maintain the plastid, at which point the primary host nucleus is no longer required and could disappear, leaving a plastid with multiple membranes but no trace of the endosymbiont nucleus or cytoplasm (8). Based on this line of reasoning, all plastids with three or four membranes (which includes those of euglenoids, dinoflagellates, haptophytes, diatoms, brown algae, chrysophytes, and apicomplexan parasites) are thought to have undergone secondary endosymbioses that have resulted in a complete loss of all endosymbiont components other than the membranes and the plastid. In this respect, these organisms can be regarded as examples of secondary endosymbiosis having run to completion, whereby the plastid, perhaps a membrane, and a large collection of genes now residing in the secondary host nucleus are all that remains of the eukaryotic endosymbiont.

To understand this reduction process it is useful to have intermediate stages, and fortunately two examples of secondary endosymbiosis-in-action provide a unique window into the mechanism of endosymbiont diminution. Cryptomonad and chlorarachniophyte algae clearly arose by separate secondary endosymbioses of a red alga and a green alga, respectively (9, 10). The nuclear genomes and the cytoplasms of the endosymbionts clearly persist but are drastically reduced. In cryptomonads the endosymbiont nucleus, known as the nucleomorph, is only 550 kb and encodes a mere 531 genes (10). Thirty of these genes encode plastid proteins, which explains the persistence of the cryptomonad nucleomorph and its expression machinery; until these genes are relocated to the secondary host nucleus, the proteins need to be supplied by the endosymbiont, just as they were when it was a free living alga (10).

In chlorarachniophyte *Bigelowiella natans*, the nucleomorph is even more reduced: with an estimated 380 kb of DNA, it is the smallest known nuclear genome (11). Here, we present the entire nucleotide sequence of this tiny genome and examine its extraordinary molecular structure and coding content.

Results and Discussion

Chromosomes and Genes. The three small chromosomes of the *B. natans* nucleomorph (12) were isolated from pulsed-field gels and were shotgun sequenced. A total of 3,400 sequencing reactions and 1,090 linking PCRs resulted in three contigs (5-fold coverage) equating to the three chromosomes of 140.5 kb, 134.1 kb, and 98.1 kb (Fig. 2). The nucleomorph genome size is 372,870 bp, which is in good agreement with previous estimates based on pulsed-field gel electrophoresis (11, 12). The three chromosomes share perfect inverted repeats of 8.74 kb at their termini comprising 25–45 copies of a typical eukaryotic telomeric repeat (5′-TCTAGGG-3′) and a ribosomal RNA cistron transcribed inwards from the telomere (Fig. 2). No 5S rRNA was identified in the repeat or anywhere else in the genome. No proteins are encoded within the repeat, but a pseudogene comprising the 3′ end of the plastid-targeted DnaK is incorporated in five of the six ends (Fig. 2). The only complete copy of DnaK resides on the left arm of chromosome II, suggesting that the pseudogene fragment has been propagated across the other ends by ectopic recombination that homogenizes rRNA genes and increases rRNA gene dosage (Fig. 2).

Between the inverted repeats of each chromosome, we found a unique, single-copy sequence with a high average AT content (>65%) that contrasts with the more balanced AT content within the genes of the rRNA cistron (50%). We identified 331 protein-coding genes plus 18 rRNAs, 20 tRNAs (all codons are used, so some tRNAs have either not been identified or are imported from the cytosol), 284 proteins, 5 DnaK pseudogenes, and 4 small nuclear RNAs, resulting in a gene density of 0.88 genes per kb and 22.4% noncoding DNA (Fig. 2). Spacers between genes are as small as 1 bp, and numerous coding regions overlap by as much as 101 bp (Fig. 2). Most genes are flanked by minimal or nonexistent upstream and downstream sequence, and transcription initiation and termination components must be located within neighboring genes (13). The *B. natans* nucleomorph genome is thus extremely compact and the smallest known eukaryotic genome.

Abundant Pygmy Introns. The vast majority of protein-coding genes (240) contain at least one spliceosomal intron, and a total of 852 introns were identified for a density of 2.9 introns per kb or 3.1 introns per gene (Figs. 2 and 3). The introns are all 18, 19, 20, or 21 nt in size (Fig. 3). No larger introns are invoked by our gene model, manual scrutiny, or extensive cDNA analysis (Table 1, which is published as supporting information on the PNAS web site). All introns have canonical GT-AG boundaries, with a single exception (intron 2 in ORF1–329 has GT-CG boundaries, which has been confirmed from cDNA) (Fig. 2). However, no polypyrimidine tract or branch donor site consensus is identifiable (Fig. 3*A*). Moreover, no obvious consensus for exon borders is evident, with the possible exception of a predominant A at −2 (Fig. 3*A*). The most frequent size introns are 19 nt, with decreasing numbers of 20-, 18-, and 21-nt introns (Fig. 3*B*). Because our gene model sought introns on the basis of their excision's increasing the size of any theoretical ORF, the 18- and 21-nt introns are potentially underrepresented because those without internal stop codons cannot easily be identified *in silico*. We tested the authenticity of the nucleomorph introns and our gene model by sequencing cDNAs for 53 nucleomorph genes (Table 1). Of 161 predicted pygmy introns, 160 were removed as predicted (Fig. 2).

Intron Splicing. The nucleomorph genome encodes numerous spliceosome components (proteins and small nuclear RNAs U1, U2, U6, and U5) (Table 2, which is published as supporting information on the PNAS web site, and Fig. 4), which suggests spliceosome-based removal of pygmy introns, but exactly how a spliceosome recognizes pygmy introns with no discernible exon border or internal consensus (Fig. 3*A*) is mysterious. We examined all exon sequences with GT(N)$_{14-17}$AG motifs that are clearly part of coding regions (based on homology with other proteins), which we refer to as pseudointrons. Comparing the core AT content of genuine pygmy introns (88%) with that of pseudointrons from coding regions (73%) reveals a marked difference that may be a basis for discrimination by the spliceosome (Fig. 3 *A* and *B*). We did not find any pseudointrons (even those with very high AT content) to have been excised from our nucleomorph cDNAs, which implies a high level of fidelity on the part of the spliceosome in discriminating real introns from pseudointrons. We propose a calliper model

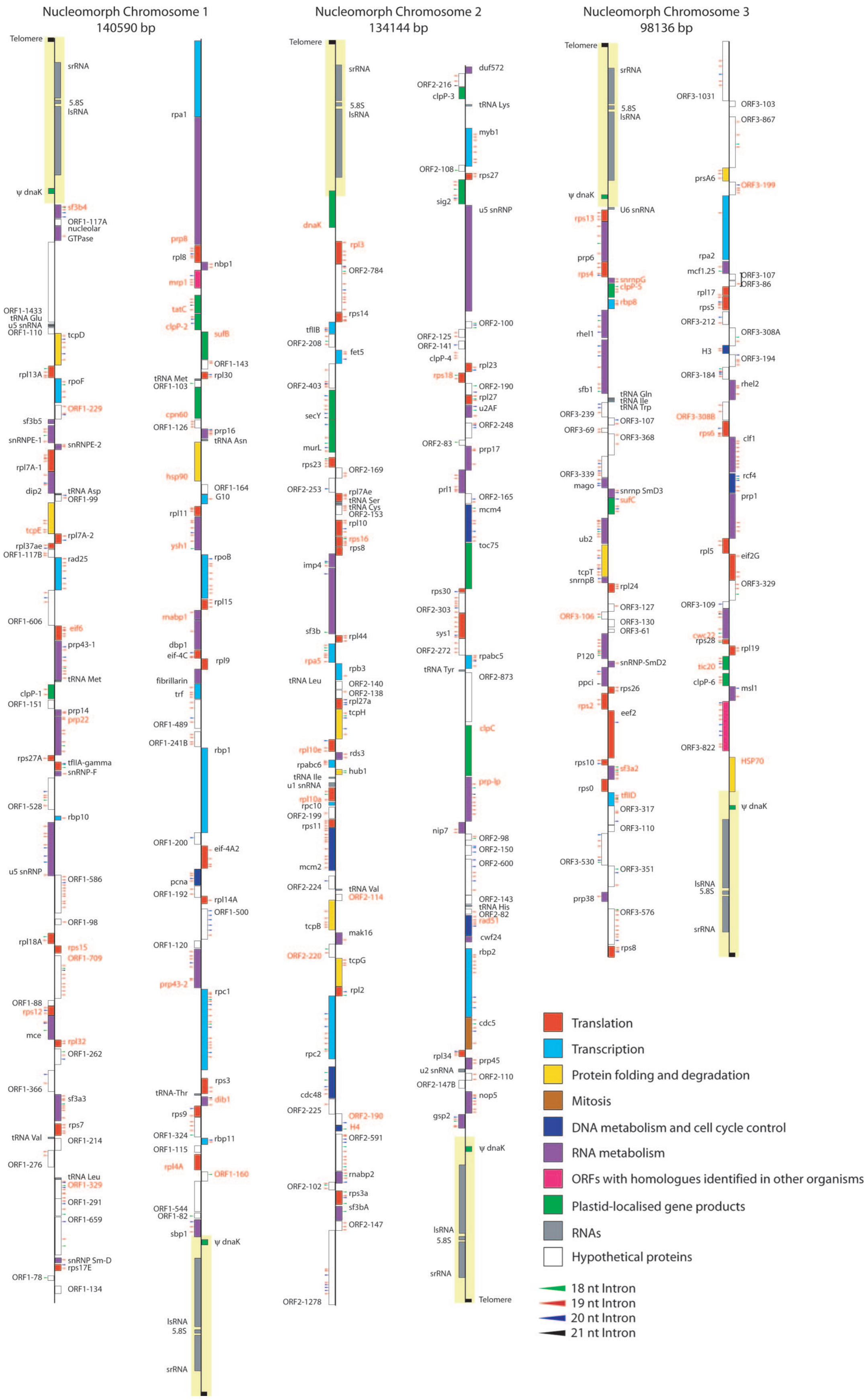

Fig. 2. Gene map of the three *B. natans* nucleomorph chromosomes. Inverted terminal repeats are shaded in yellow. Pygmy introns are depicted as triangles. Genes for which excision of the pygmy introns (where present) has been confirmed by cDNA analysis are named in red type.

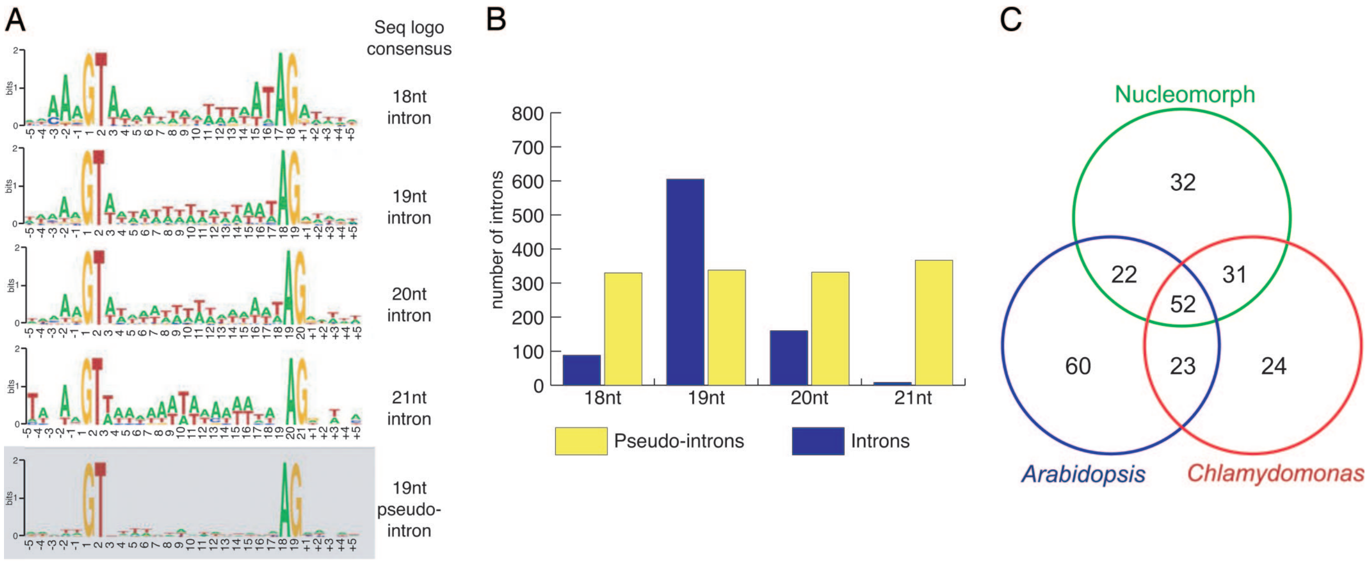

Fig. 3. Nucleomorph genes harbor numerous ultra-small introns. (*A*) Consensus plots of the 18-, 19-, 20- and 21-nt introns showing 5′-GT. . .AG-3′ borders but no exon or intron core consensus other than an A at −2. This A and overall high AT content are the only discernible differences between real introns and 19-nt exonic sequences with 5′-GT. . .AG-3′ borders (pseudointrons) that are not removed by the spliceosome. (*B*) Nineteen nucleotide introns are the most abundant in the nucleomorph, but equivalent numbers of each size category of pseudointrons occur. (*C*) Numbers of introns shared between the nucleomorph, *Arabidopsis*, and *Chlamydomonas*. Most nucleomorph introns occur at the same position as large introns in these closely related genomes, suggesting that the pygmy introns began as normal introns but have been reduced by DNA loss to converge on a minimal spliceable size, ≈19 nt.

for the nucleomorph spliceosome, whereby it scans mRNAs for $GT(N)_{14-17}AG$ motifs with high AT content and then splices out these elements. Apparent absence of U4 small nuclear RNA, which normally has a role in bringing together the two remote splice sites, is consistent with this model, but other recognition elements must also be at play to prevent the accidental splicing of the most AT-rich pseudointrons.

Intron Origin. The abundance of introns (although minute ones) in a highly compacted genome seems paradoxical. Two explanations for the presence of so many introns are apparent: either the pygmy introns are selfish elements that propagated throughout the genome recently or, alternatively, they are stripped-down relicts of longer spliceosomal introns present in the endosymbiont genome from the outset. We compared intron positions from 44 highly conserved nucleomorph genes with homologues from *Chlamydomonas reinhardtii* and *Arabidopsis thaliana*, two organisms in the same part of the evolutionary tree as the endosymbiont (9). Of the 137 pygmy introns in these nucleomorph genes, 38% are shared by all three genomes, and 77% are conserved in either *Chlamydomonas* and/or *Arabidopsis* (Fig. 3*C*), which strongly suggests that the pygmy introns are derived from canonical spliceosomal introns present in the endosymbiont before capture. The *Chlamydomonas* and *Arabidopsis* introns are much longer than their nucleomorph homologues (14). We hypothesize that, as nucleomorph introns underwent compaction, they converged on a minimal spliceable size, which might relate to the minimal physical distance between the components of the spliceosome that recognize intron boundaries (14).

Nucleomorph intron density is similar to intron density of *Chlamydomonas* and *Arabidopsis* (14–16), suggesting little intron loss in nucleomorphs despite widespread DNA attrition. Intron loss by DNA deletion must be base-perfect to avoid corrupting the gene and is presumably a rare event. Intron loss by recombination with cDNA circumvents this problem (17). Perhaps intron loss was prevented in the nucleomorph by loss of either

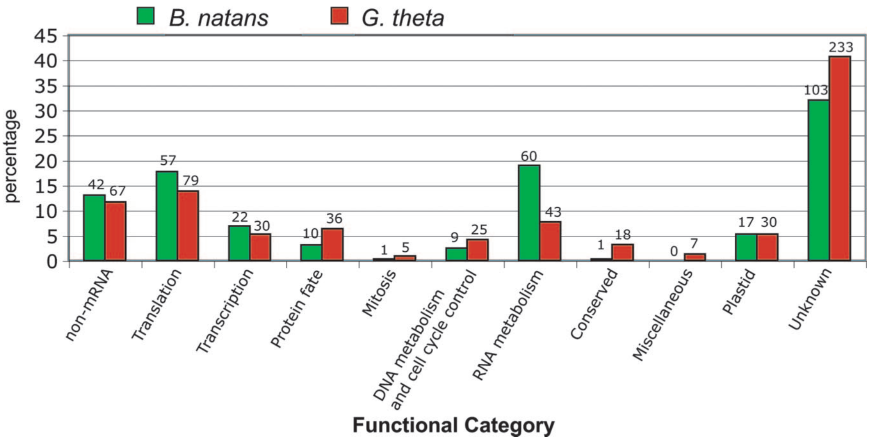

Fig. 4. Comparison of gene functional categories from the nucleomorph genomes of *B. natans* and *G. theta*. Percentages of genes belonging to different functional categories are indicated by the scale bar on the left, and the total numbers of genes belonging to these categories are shown above the class bars.

reverse transcriptase or recombination activities. Interestingly, the cryptomonad nucleomorph contrasts with *B. natans* in that it contains only 17 introns, all of which are normal size (10). So far as we know, red algae are depauperate in introns (18), so the cryptomonad endosymbiont nucleus likely began with a modest complement of introns that did not change much in size or density. Intriguingly, of the 17 intron-containing genes in the *Guillardia theta* nucleomorph, 9 genes are also found in *B. natans*, and 4 of the cryptomonad introns are shared with *B. natans* and are generally found in other green and red algae (Fig. 5, which is published as supporting information on the PNAS web site). In cryptomonads, these introns remain relatively large in comparison with the pygmy introns of *B. natans*, suggesting that DNA loss has been less severe in this element of the cryptomonad genome or that the cryptomonad spliceosome has not adapted to remove pygmy introns.

Plastid Protein Genes. In addition to compaction, genome reduction can obviously involve the loss of genes. The *B. natans* nucleomorph began as a fully autonomous eukaryotic genome but is now reduced to only 331 genes. The nucleomorph was indispensable at the outset of endosymbiosis because it encoded a large cohort of proteins (probably ≈1,000) essential to biogenesis and function of the plastid. We identified only 17 genes encoding plastid proteins in the *B. natans* nucleomorph (Figs. 1 and 4 and Table 2). These proteins participate in various plastid functions such as division (MurL), transcription (Sig2), protein transport and folding [Toc75, Tic20, TatC, SecY, ClpC (Hsp93), DnaK, Hsp60 (GroEL)], iron sulfur cluster formation (SufB, SufC), and protein degradation (ClpP). The remaining plastid protein genes have apparently relocated to the secondary host nucleus, and a number have already been identified (19, 20). The small residue of plastid protein genes in the nucleomorph obliges *B. natans* to retain the nucleomorph and attendant expression machinery for synthesis and delivery of these 17 proteins to the plastid. The nucleomorph-encoded plastid proteins bear N-terminal transit peptides that likely direct the proteins through a translocation apparatus (which would include the nucleomorph-encoded Toc75 and Tic20) spanning the inner two membranes of the plastid. Until now, no Toc component had been identified in a secondary endosymbiotic alga (21), so nucleomorph Toc75 suggests that nucleus-encoded plastid proteins (19, 20) could also travel through the Toc75/Tic20-containing apparatus after crossing the two outer membranes surrounding the plastid via the secretory pathway (22).

Cryptomonad and Chlorarachniophyte Nucleomorphs. Comparison of the cryptomonad and chlorarachniophyte nucleomorphs identifies a strikingly similar breakdown in proportions of components (Fig. 4 and Table 2). Both genomes are dominated by housekeeping components that exist to express a minor proportion (4–5%) of plastid protein genes (Fig. 4 and Table 2). The chlorarachniophyte nucleomorph retains more spliceosome components but is otherwise more depleted (Fig. 4 and Table 2). Both nucleomorphs lack DNA-modifying enzymes and DNA polymerases, which, along with numerous other essential housekeeping proteins, must be imported from the host. Intriguingly, both nucleomorphs retain only one aminoacyl tRNA synthase, and it is the same one (serine). The chlorarachniophyte nucleomorph lacks fundamental components retained by cryptomonad nucleomorphs such as α, β, and γ tubulin (23), the proteasome, 5S rRNA, telomerase RNA, U4 small nuclear RNA, and numerous tRNAs (Table 2). Both nucleomorphs encode centromere-binding proteins, suggesting a role for centromeres in nucleomorph mitosis, so *B. natans* potentially imports tubulins for a spindle. Centromere DNA has not been identified in either nucleomorph (10), but a few noncoding regions on the nucleomorph chromosomes are candidates (Fig. 2).

Nucleomorphs have undergone extreme compaction during their tenure as endosymbiotic nuclear genomes, rivaled only by microsporidian intracellular parasites (24). Why compaction has been so extreme is not yet clear. It is commonly argued that selection for smaller genomes drives this process (see, e.g., refs. 24–26), and such selection would have to be very strong to account for the unusual characteristics of nucleomorph genomes. It is also possible that a DNA loss ratchet is in operation, whereby nonlethal losses of DNA accumulate faster than the addition of new DNA through recombination or the activity of transposons that thrive in sexually reproducing populations (no reverse transcriptase or transposons are present in the genome) (27, 28). Distinguishing between these possibilities will require comparative analysis of related nucleomorph genomes, in particular including those that are larger than *G. theta* or *B. natans*. Whatever the cause of nucleomorph reduction, the gross similarities in the resulting cryptomonad and chlorarachniophyte nucleomorph architecture (three chromosomes of ≈100–200 kb) argue that DNA loss has proceeded to some minimal endpoint whereby the genomes cannot get much smaller. An attractive idea is that the chromosomes cannot be much shorter than 100 kb to avoid loss during mitotic segregation (29). However, the chromosomes also need to be small enough when condensed to fit within the nucleomorph volume, forcing the residual 373–550 kb of DNA to be distributed between three chromosomes (10).

Conclusions

Why do nucleomorphs persist? In *B. natans*, a mere 17 plastid protein genes await transfer to the secondary host nucleus to potentially allow the nucleomorph to disappear. Is there a reason they haven't relocated? It was argued that pygmy introns might impede relocation of plastid protein genes to the host nucleus (30), but six of the *B. natans* nucleomorph plastid protein genes identified here lack introns, and none of the cryptomonad nucleomorph plastid protein genes have introns (10), so this explanation is refuted. Retention of mitochondrial and plastid genomes has been rationalized by a requirement to encode redox proteins, which are prone to oxidative damage, within the compartment of function (4), but no redox reactions are evident in the nucleomorph compartment, so this explanation can also be dismissed. Another explanation for organellar genome persistence is refractoriness of the products to intracellular transfer because of protein hydrophobicity (31, 32), but the residual plastid protein genes in the two nucleomorphs are not especially hydrophobic. Failing any mechanistic rationalization of nucleomorph persistence, the remaining explanation is simply that there may not yet have been time for transfer of all essential genes to the host nucleus: nucleomorphs as evolutionary intermediates.

If transfer of genes from nucleomorph to host nucleus is predominantly random, we would not expect any marked overlap in the plastid proteins still encoded by the two nucleomorphs, and indeed, other than Hsp60 (GroEL) and two isoforms of the catalytic subunit of the Clp protease (ClpP1 and ClpP2), the nucleomorphs retain a different slate of plastid protein genes (Table 2). Moreover, the nucleomorph Hsp60 genes in *B. natans* and *G. theta* are paralogues, deriving from different duplication events of cyanobacterial precursor genes (33), so their common retention is not necessarily significant. Similarly, both nucleomorphs retain multiple isoforms of ClpP (six in the *B. natans* and two in *G. theta*), but extensive ClpP paralogy exists in other plastid-containing organisms (34), and it is not clear whether the nucleomorph genes are orthologous, so this overlap may not be significant either. Thus, taking into consideration that the two different nucleomorph precursors began with ≈1,000 plastid protein genes, the lack of overlap in the current complement suggests that there is no specific impediment to gene relocation. This observation is supported by the fact that, in two groups where endosymbiont nuclei have vanished altogether and for which we have complete genomes, namely diatoms (*Thallasiosira pseudonana*) and apicomplexans (e.g., *Plasmodium* and *Theile-*

ria), Hsp60 and ClpP have been relocated to the secondary host nucleus (35–39). Thus, we conclude that no single plastid gene may be responsible for the retention of nucleomorphs, and they may yet disappear. Acquisition of suitable genes from prey organisms in the case of *B. natans* may even foreshorten the process (19). Nucleomorphs, therefore, seem to have converged on a minimal viable size with a tiny residue of plastid protein genes acting as an anchor to force the retention of a genome and attendant expression machinery. Similarities between the overall form of cryptomonad and chlorarachniophyte nucleomorph genomes can thus be viewed as a remarkable convergence, but our comparison of the complete genomes reveals that they have traveled substantially different trajectories from different starting points, emphasizing the diverse responses to strong reducing selection on the genome.

Materials and Methods

Library Construction and Sequencing. Motile forms of *B. natans* (CCMP 621) were cultured in f/2 media with continuous lighting at 24°C in a flask bubbled with filter-sterilized air. To obtain enough nucleomorph DNA to construct libraries, 8 liters of algal cells were harvested by centrifugation, and the cell pellet was resuspended in 32 ml of ice-cold NET buffer [150 mM NaCl/50 mM EDTA/10 mM Tris, pH 8.8/1 mg/ml Pronase E (CalBiochem)]. The cells were partially lysed in 2% (wt/vol) sodium-deoxycholate followed by 1% (wt/vol) SDS at 4°C for 2 h. Debris was removed by centrifugation (3,000 × *g* at 4°C for 30 min), and the supernatant was extracted with phenol:chloroform:isoamylalcohol (25:24:1) followed by chloroform:isoamylalcohol (24:1). DNA was precipitated after the addition of 1/9 volume of 5 M potassium acetate and 1 vol of isopropanol. After dissolving in TE buffer, nucleomorph chromosomes from the DNA samples were separated by pulsed-field gel electrophoresis (12). The nucleomorph chromosomes were excised from the gel, and the chromosomal DNA was partially digested with MseI to yield fragments of 0.5–3 kb. The DNA was ligated into the NdeI site of pGEM 5Zf(−) (Promega) to create a shotgun library. The plasmid library was grown in ElectroMAX DH10B (GIBCO/BRL), and randomly picked clones were sequenced. Sequence editing and contig assembly were done as described (10, 30). Gaps between contigs that could not be filled by shotgun sequencing were amplified by long-range PCRs by using organelle DNA and eLONGase (Invitrogen) (94°C for 15 s; 58°C for 30 s; 65°C for 15 min for 35 cycles). These PCR products were cloned into pCR-XL-TOPO vector (Invitrogen). Inverse PCR (40) was also used to bridge some gaps. Double-stranded sequencing of large plasmid inserts was achieved by primer walking.

Gene Model and cDNA Analyses. Preliminary analysis identified abundant small introns in nucleomorph genes (30). A gene model based on maximizing the size of ORFs was implemented with a PERL script we call COWPIE that searches both strands of DNA for conserved intron splice junctions spaced 18, 19, 20, or 21 bp apart. The script endeavors to generate the largest possible ORF by removing putative introns. RT-PCR was used to confirm intron excision for selected genes. A cDNA library containing transcripts of nucleomorph genes was constructed as described (30), and EST sequences were generated as per ref. 19. tRNA genes were identified as described in ref. 10. Identified genes were annotated by using ARTEMIS genome browser (41). GenBank accession numbers of nucleomorph chromosomes I, II, and III are DQ158856, DQ158857 and DQ158858, respectively.

G.I.M. is a Howard Hughes Medical Institute International Scholar, an associate of the Canadian Institute for Advanced Research (CIAR), and an Australian Research Council Professorial Fellow. P.J.K. is a fellow of the CIAR and a New Investigator of the Canadian Institutes for Health Research. This work was supported by a Program Grant from the National Health and Medical Research Council (to G.I.M.) and the Natural Sciences and Engineering Research Council (P.J.K.).

1. Archibald, J. M. (2005) *IUBMB Life* **57,** 539–547.
2. Cavalier-Smith, T. & Lee, J. (1985) *J. Protozool.* **32,** 376–379.
3. McFadden, G. I. (2001) *J. Phycol.* **37,** 1–9.
4. Allen, J. F. (2003) *Philos. Trans. R. Soc. London B* **358,** 19–38.
5. Soll, J. & Schleiff, E. (2004) *Nat. Rev. Mol. Cell Biol.* **5,** 198–208.
6. Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O. & Taylor, F. J. (2004) *Science* **305,** 354–360.
7. Okamoto, N. & Inouye, I. (2005) *Science* **310,** 287.
8. Nisbet, R. E., Kilian, O. & McFadden, G. I. (2004) *Curr. Biol.* **14,** R1048–R1050.
9. Van de Peer, Y., Rensing, S. A., Maier, U.-G. & de Wachter, R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7732–7736.
10. Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L.-T., Wu, X., Reith, M., Cavalier-Smith, T. & Maier, U.-G. (2001) *Nature* **410,** 1091–1096.
11. McFadden, G. I., Gilson, P. R., Hofmann, C. J., Adcock, G. J. & Maier, U.-G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3690–3694.
12. Gilson, P. R. & McFadden, G. I. (1995) *Chromosoma* **103,** 635–641.
13. Williams, B. A., Slamovits, C. H., Patron, N. J., Fast, N. M. & Keeling, P. J. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 10936–10941.
14. Deutsch, M. & Long, M. (1999) *Nucleic Acids Res.* **27,** 3219–3228.
15. Shrager, J., Hauser, C., Chang, C. W., Harris, E. H., Davies, J., McDermott, J., Tamse, R., Zhang, Z. & Grossman, A. R. (2003) *Plant Physiol.* **131,** 401–408.
16. The Arabidopsis Genome Initiative (2000) *Nature* **408,** 796–815.
17. Fink, G. R. (1987) *Cell* **49,** 5–6.
18. Matsuzaki, M., Misumi, O., Shin, I. T., Maruyama, S., Takahara, M., Miyagishima, S. Y., Mori, T., Nishida, K., Yagisawa, F., Yoshida, Y., *et al.* (2004) *Nature* **428,** 653–657.
19. Archibald, J. M., Rogers, M. B., Toop, M., Ishida, K. & Keeling, P. J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 7678–7683.
20. Deane, J., Fraunholz, M., Su, V., Maier, U.-G., Martin, W., Durnford, D. & McFadden, G. (2000) *Protist* **151,** 239–252.
21. McFadden, G. I. & van Dooren, G. G. (2004) *Curr. Biol.* **14,** R514–R516.
22. Nassoury, N. & Morse, D. (2005) *Biochim. Biophys. Acta* **1743,** 5–19.
23. Keeling, P. J., Deane, J. A., Hink-Schauer, C., Douglas, S. E., Maier, U.-G. & McFadden, G. I. (1999) *Mol. Biol. Evol.* **16,** 1308–1313.
24. Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., *et al.* (2001) *Nature* **414,** 450–453.
25. Cavalier-Smith, T. (2005) *Ann. Bot. (London)* **95,** 147–175.
26. Keeling, P. J. & Slamovits, C. H. (2005) *Curr. Opin. Genet. Dev.* **15,** 601–608.
27. Lozovskaya, E. R., Hartl, D. L. & Petrov, D. A. (1995) *Curr. Opin. Genet. Dev.* **5,** 768–773.
28. Gilson, P. R. & McFadden, G. I. (2002) *Genetica* **115,** 13–28.
29. Murray, A. & Szostak, J. (1985) *Annu. Rev. Cell Biol.* **1,** 289–315.
30. Gilson, P. R. & McFadden, G. I. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7737–7742.
31. Claros, M., Perea, J., Shu, Y., Samatey, F., Popot, J.-L. & Jacq, C. (1995) *FEBS Lett.* **228,** 762–771.
32. Daley, D. O., Clifton, R. & Whelan, J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 10510–10515.
33. Zauner, S., Lockhart, P., Stoebe-Maier, B., Gilson, P., McFadden, G. I. & Maier, U. G. (2006) *BMC Evol. Biol.* **6,** 38.
34. Peltier, J. B., Ytterberg, J., Liberles, D. A., Roepstorff, P. & van Wijk, K. J. (2001) *J. Biol. Chem.* **276,** 16318–16327.
35. Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. A., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., *et al.* (2004) *Science* **306,** 79–86.
36. Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M., Foth, B. F., Tonkin, C. J., Roos, D. S. & McFadden, G. I. (2004) *Nat. Rev. Microbiol.* **2,** 203–216.
37. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., *et al.* (2002) *Nature* **419,** 498–511.
38. Gardner, M. J., Bishop, R., Shah, T., de Villiers, E. P., Carlton, J. M., Hall, N., Ren, Q., Paulsen, I. T., Pain, A., Berriman, M., *et al.* (2005) *Science* **309,** 134–137.
39. Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C., *et al.* (2005) *Science* **309,** 131–133.
40. Hartl, D. L. & Ochman, H. (1996) *Methods Mol. Biol.* **58,** 293–301.
41. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. & Barrell, B. (2000) *Bioinformatics* **16,** 944–945.