

Nucleomorph genomes: structure, function, origin and evolution

John M. Archibald

Summary

The cryptomonads and chlorarachniophytes are two unicellular algal lineages with complex cellular structures and fascinating evolutionary histories. Both groups acquired their photosynthetic abilities through the assimilation of eukaryotic endosymbionts. As a result, they possess two distinct cytosolic compartments and four genomes—two nuclear genomes, an endosymbiont-derived plastid genome and a mitochondrial genome derived from the host cell. Like mitochondrial and plastid genomes, the genome of the endosymbiont nucleus, or 'nucleomorph', of cryptomonad and chlorarachniophyte cells has been greatly reduced through the combined effects of gene loss and intracellular gene transfer. This article focuses on the structure, function, origin and evolution of cryptomonad and chlorarachniophyte nucleomorph genomes in light of recent comparisons of genome sequence data from both groups. It is now possible to speculate on the reasons that nucleomorphs persist in cryptomonads and chlorarachniophytes but have been lost in all other algae with plastids of secondary endosymbiotic origin. *BioEssays* 29:392–402, 2007. © 2007 Wiley Periodicals, Inc.

"The optimum size for a nucleomorph genome is zero..."
Thomas Cavalier-Smith (*Ann Bot* 2005 95:147–175)

Eukaryotic genome diversity—the big picture

The nuclear genome is tremendously diverse in size and structure. Taking into account the full breadth of eukaryotic diversity, nuclear genomes vary in size ~200,000 fold⁽¹⁾ and can be composed of a single chromosome ($n=1$ in the Australian ant *Myrmecia*) or well over 100 ($n=630$ in the fern *Ophioglossum*).⁽²⁾ Unlike the large genomes of animals and

plants, which often harbor copious amounts of repetitive and non-coding DNA,^(3,4) the genomes of unicellular eukaryotes can be very small and compact. For example, the genomes of the fungus *Ashbya gossypii* and the picoplanktonic alga *Ostreococcus tauri* are only 9.2 and 12.5 megabase pairs (Mbp), respectively,^(5–7) and the genome of the microsporidian parasite *Encephalitozoon cuniculi* is a mere 2.9 Mbp,⁽⁸⁾ the smallest known genome of a self-replicating eukaryote. For comparison, the entire *E. cuniculi* genome is about the same size as the human dystrophin gene, <1% of which is coding sequence.⁽⁹⁾

How small can nuclear genomes get? At the very bottom of the size spectrum are the 'nucleomorph' genomes of cryptomonad and chlorarachniophyte algae. Nucleomorphs are the remnant nuclei of algal endosymbionts that took up residence inside non-photosynthetic host cells^(10–12) and, at well under 1 Mbp in size, nucleomorph genomes are the smallest known. In this article, I provide an overview of the discovery of nucleomorphs, summarize the last decade of research on the origins, structure and function of these unusual organelles, and focus on exciting new insights that have come from comparisons of complete nucleomorph genome sequences. As we shall see, the evolutionary forces that have influenced the size and coding capacity of nucleomorph genomes are similar to those that have shaped the evolution of mitochondria and plastids. This is not a coincidence, as the origin and evolution of nucleomorphs is intricately linked to the birth and spread of oxygenic photosynthesis in eukaryotes.

A brief history of endosymbiosis and plastid origins

Plastids (or chloroplasts) are the solar-powered, energy-generating, oxygen-producing organelles of plants and algae. They are widely recognized as the descendants of once-free living cyanobacteria that became permanent fixtures within the cytoplasm of a non-photosynthetic eukaryote via endosymbiosis (Fig. 1A). The 'primary' endosymbiotic origin of plastids is believed to have occurred only once, ~0.7–1.5 billion years ago^(13,14) and, although clearly cyanobacterial in origin, plastid genomes are highly derived entities, encoding at most ~200 of the thousand or more proteins necessary for organelle function.⁽¹⁵⁾ The bulk of the genes present in the genome of the cyanobacterial progenitor of plastids were either lost or transferred to the nuclear genome of the eukaryotic host,

The Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Biochemistry and Molecular Biology, Dalhousie University, Sir Charles Tupper Medical Building, 5850 College Street, Halifax, Nova Scotia, B3H 1X5, Canada.

E-mail: john.archibald@dal.ca

Funding agency: JMA is a Scholar of the Canadian Institute for Advanced Research, Program in Evolutionary Biology. Research in the Archibald Laboratory is supported by Genome Atlantic, operating grants from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes for Health Research, as well as a New Investigator Award from the Dalhousie Medical Research Foundation.

DOI 10.1002/bies.20551

Published online in Wiley InterScience (www.interscience.wiley.com).

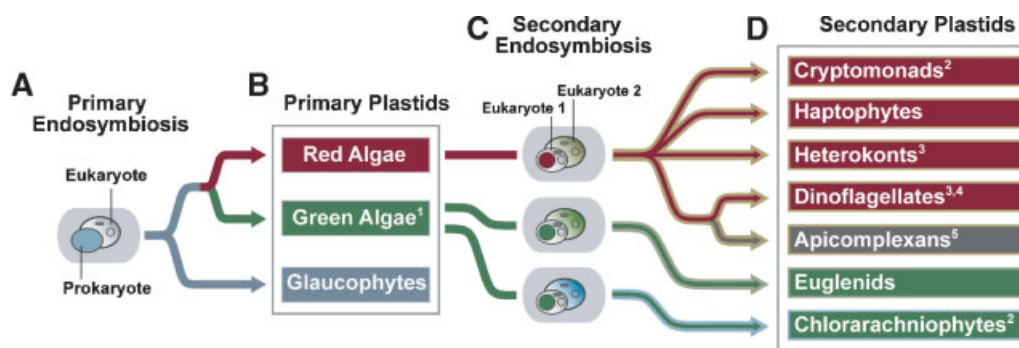


Figure 1. Origin and spread of plastids in eukaryotic cells. **A:** Primary endosymbiosis between a heterotrophic eukaryote and a photosynthetic prokaryote. **B:** Three modern-day eukaryotic lineages, red and green algae and glaucophytes, possess 'primary' plastids whose ancestry traces directly back to the primary endosymbiosis. On balance, molecular phylogenetic data suggest that red and green algae are each other's closest relatives (see text). **C:** At least three independent secondary (i.e. eukaryote–eukaryote) endosymbioses between unrelated host cells and red and green algal endosymbionts have probably occurred (see text and references therein). **D:** Diversity of eukaryotes harboring secondary plastids. Cryptomonads, haptophytes, heterokonts, many dinoflagellates and (probably) apicomplexans harbor plastids derived from a red algal endosymbiont, while euglenids and chlorarachniophytes acquired their plastids from green algae, probably independent of one another. The precise number of secondary endosymbioses that have occurred during eukaryotic evolution is controversial but it is likely at least three, as depicted here (see text).⁽¹⁾ The green algae include multicellular lineages such as land plants.⁽²⁾ The cryptomonads and chlorarachniophytes are the only known secondary plastid-containing algae to still possess the nucleus and nuclear genome of their algal endosymbionts.⁽³⁾ Heterokonts and dinoflagellates both possess significant numbers of non-photosynthetic (and, in some cases, aplastidic) species.⁽⁴⁾ Some dinoflagellate lineages have replaced their ancestral red algal secondary plastid with a 'tertiary' plastid through the uptake of cryptomonad, heterokont or haptophyte algae.⁽⁵⁾ The apicomplexans are an entirely non-photosynthetic lineage of parasites, some of which have completely lost their plastid.

where they are now expressed, translated in the host cytosol, and their protein products targeted back to organelle in which they were originally encoded by a dedicated plastid protein import system.^(16,17)

Three distinct eukaryotic lineages, the green algae, red algae and glaucophytes, harbor plastids whose ancestry can be traced directly back to the original primary endosymbiosis (Fig. 1B). The green algae, which include the plants and trees that inhabit dry land, are perhaps the best-known primary plastid-containing lineage, while the red algae are generally less well known, despite their abundance and economic significance. The least understood of the three groups are the glaucophytes (or glaucocystophytes), an exclusively unicellular lineage of particular interest to cell evolutionists due to the fact that their plastids are distinctly cyanobacterial in general morphology and ultrastructure.^(18–21) The latest multi-gene phylogenetic studies indicate that red, green and glaucophyte algae are each other's closest relatives,^(22,23) although the question of whether the glaucophytes are the deepest diverging of the three groups, as depicted in Fig. 1, remains to be proven.⁽²³⁾

In addition to the primary endosymbiotic origin of red, green and glaucophyte plastids, subsequent endosymbioses involving the uptake of red and green algae by unrelated eukaryotes have occurred. This process is referred to as secondary endosymbiosis (Fig. 1C) and has given rise to a large number of ecologically significant algal groups, including the eugle-

nids, haptophytes, heterokonts (e.g. diatoms and kelp) and dinoflagellates, as well as secondarily non-photosynthetic lineages such as the apicomplexan parasites (e.g. the malaria parasite *Plasmodium*; Fig. 1D). Unlike primary plastids, which are surrounded by two membranes (and in glaucophytes, a layer of peptidoglycan), secondary plastids are characterized by the presence of either three or four plastid membranes, a consequence of the phagotrophic mechanism by which they were ingested.⁽²⁴⁾ The exact number of secondary endosymbioses that have occurred during eukaryotic evolution is hotly debated, although a minimum of two (and probably three) events are required to account for the known spectrum of secondary plastid-containing organisms (Fig. 1C,D). Finally, tertiary endosymbioses have also occurred, in which dinoflagellates have taken up cryptomonad, diatom, haptophyte, and green algal endosymbionts (see Ref. 25 and references therein). Interested readers can refer to any number of recent reviews for more a comprehensive discussion of the origin and evolution of plastids.^(10–12,24,26–30)

Nucleomorphs—the 'smoking guns' of secondary endosymbiosis

In terms of contributing to our knowledge of the spread of photosynthesis in eukaryotes, arguably the most-significant secondary plastid-containing organisms are the chlorarachniophytes and cryptomonads (Fig. 2). These organisms harbor plastids derived from the ingestion of green and red

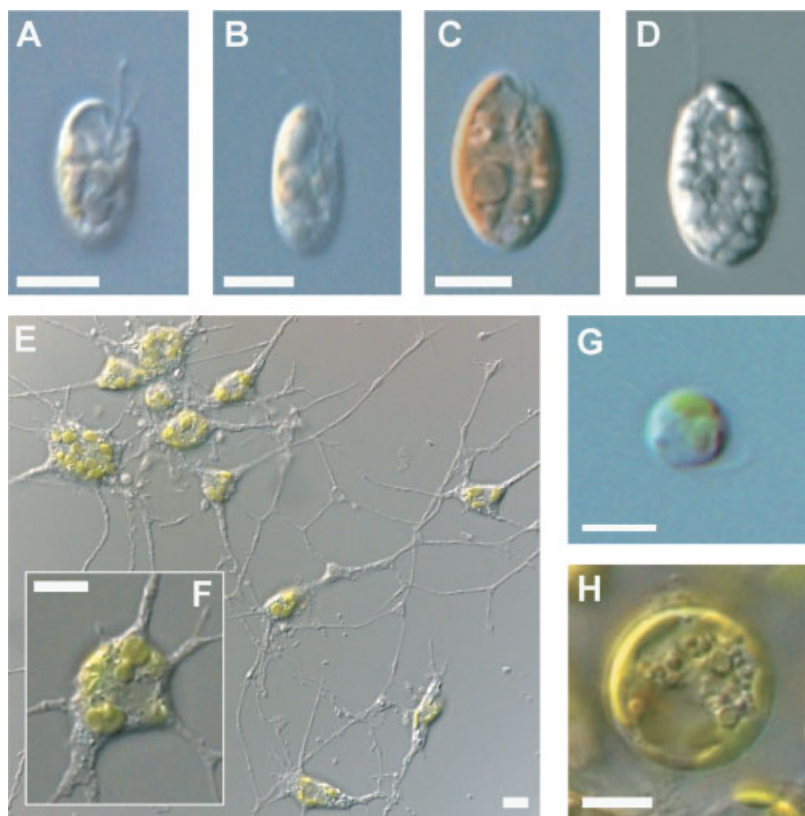
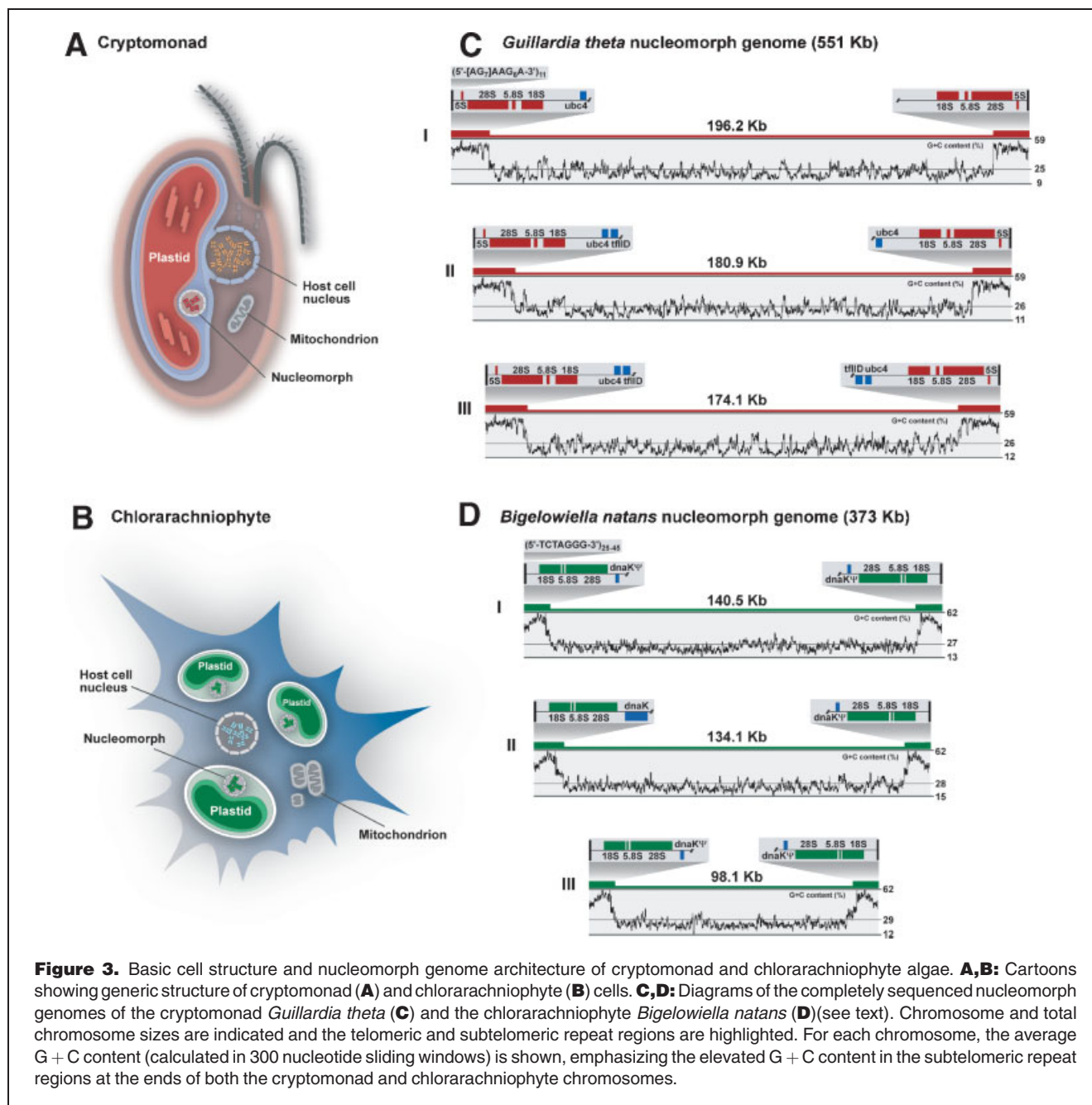


Figure 2. Differential interference contrast light micrographs of diverse cryptomonad (**A–D**) and chlorarachniophyte (**E–H**) species. **A:** *Rhodomonas salina* CCMP1319, **B:** *Proteomonas sulcata* CCMP704, **C:** *Rhodomonas lens* CCMP739, **D:** *Cryptomonas paramecium* CCAP977/2A, **E and F:** *Chlorarachnion reptans* CCMP238, **G:** *Bigelowiella natans* CCMP621, **H:** *Lotharella vacuolata* CCMP240. Scale bars are all 5 μm . All images taken by the author.

algae, respectively (Fig. 1C,D), and nested between the inner and outer pairs of their plastid membranes is the proof: the remnant nucleus of the eukaryotic endosymbiont—the *nucleomorph* (Fig. 3A,B). In the 1970s, biochemical and cell biological investigations of algae such as *Euglena* prompted Taylor⁽³¹⁾ and Gibbs⁽³²⁾ to consider the possibility that photosynthesis had spread between unrelated organisms by endosymbioses between two eukaryotes. However, it was not until the plastid-associated, nucleus-like organelles of chlorarachniophytes and cryptomonads were scrutinized using a battery of cytochemical and molecular techniques that secondary endosymbiosis became a widely accepted phenomenon. More than a decade's worth of study on nucleomorphs and their genomes have opened up a fascinating window on the transition from eukaryotic endosymbiont to fully integrated organelle and, more generally, on the process of genome reduction in eukaryotes.

Greenwood^(33,34) coined the term nucleomorph and was the first to speculate on its possible evolutionary significance. As reviewed by McFadden and Gilson,^(35,36) Greenwood's

ultrastructural studies of cryptomonads (Fig. 2A–D) demonstrated that their plastid is surrounded by four membranes and that a tiny (<1 μm in diameter) membrane-bound body, replete with pores and an electron-dense region reminiscent of a nucleolus, resides in the so-called periplastid space between membranes two and three. Independent microscopic studies of the amoeboflagellate alga *Chlorarachnion reptans* (Fig. 2E–F) revealed a similar organelle in close association with its chlorophyll *b*-containing plastid,⁽³⁷⁾ and it was eventually proven that (1) the nucleomorphs of both cryptomonads and chlorarachniophytes contain DNA,^(38–41) (2) the periplastid space contains eukaryotic-sized ribosomes,⁽³⁸⁾ as would be expected of a cytoplasm in which translation is still occurring, and (3) nucleomorphs encode 18S ribosomal DNA (rDNA) genes that are evolutionarily distinct from those present in the host nuclear genome.^(42–45) Having established that nucleomorphs are indeed degenerate nuclei, these studies set the stage for more focused studies on the size, structure and composition of the cryptomonad and chlorarachniophyte nucleomorph genomes



with the goal of elucidating their function and evolutionary history.

Nucleomorph genomes

Eschbach et al.⁽⁴⁶⁾ and Maier et al.⁽⁴⁴⁾ were the first to investigate the karyotypic structure of nucleomorph genomes. Using pulsed-field gel electrophoresis (PFGE), a technique that can resolve intact high molecular-weight chromosomes, these researchers demonstrated that the nucleomorph genome of *Rhodomonas salina* (then *Pyrenomonas salina*)

was a mere 660 Kbp in size—far smaller than any other nuclear genome known at the time—and consisted of three similarly sized chromosomes (~195, 225, 240 Kbp). More-recent and broadly focused PFGE studies have shown that the presence of three nucleomorph chromosomes appears to be a universal feature of cryptomonads and that nucleomorph genome size in this group varies from ~450 Kbp in the secondarily non-photosynthetic *Cryptomonas paramecium* to ~845 Kbp in *Rhodomonas* sp. CCMP1178.^(47–50) The presence of rDNA operons on each chromosome is also a

widespread feature of nucleomorph genomes,^(47,49,50) with the sole exception thus far having been recently described in members of the genus *Hemiselmis*.⁽⁵¹⁾ Where investigated, the rDNA operons are located adjacent to the telomeric repeats on the chromosome ends.^(47,52,53)

Parallel studies of chlorarachniophyte algae by McFadden, Gilson and colleagues revealed intriguing similarities between the structure of their nucleomorph genomes and those of cryptomonads. Three nucleomorph chromosomes exist in all chlorarachniophyte species examined to date and, as in cryptomonads, subtelomeric rDNA operons are present on each chromosome end (although in the reverse orientation and without a 5S rDNA gene^(45,54,55)). The nucleomorph genomes of studied chlorarachniophytes are smaller than those of cryptomonads, ranging from ~380–450 Kbp in size.⁽⁵⁶⁾ Overall, the structural similarities of nucleomorph genomes in both groups are striking when one considers that they are of independent origin (Fig. 1). As noted above, a variety of biochemical and molecular data indicate that the chlorarachniophyte endosymbiont is derived from an ancestor of modern-day green algae while, in cryptomonads, the endosymbiont is of red algal origin.^(10,11,26,57,58) During the process of secondary endosymbiosis, the nucleomorph genomes of both groups converged upon similar sizes and structures and identical karyotypes from much larger genomes with (presumably) dozens of chromosomes. This may or may not be a coincidence. Unlike the high degree of condensation seen in canonical mitotic chromosomes, nucleomorph chromatin may not condense above the level of 30 nm fibers. Douglas, Cavalier-Smith and colleagues^(52,59) therefore suggest that nucleomorph chromosome number is a balance between having chromosomes that are small enough to be properly segregated in the tight confines of the nucleomorph, yet large enough to ensure they are not lost during mitosis.^(52,59) Given that haploid nucleomorph genome size is roughly similar in cryptomonads and chlorarachniophytes, this means partitioning the genome among three chromosomes between ~100 and 200 Kbp.⁽⁵⁵⁾

Bonsai genomics—complete nucleomorph genome sequences

Given their obvious evolutionary importance, the nucleomorphs of the model cryptomonad *Guillardia theta* and the chlorarachniophyte *Bigeloviella natans* quickly became the focus of genome sequencing projects by researchers in Canada, Germany and Australia. Preliminary data from *G. theta* provided an initial snapshot of the fine-scale structure of its nucleomorph chromosome ends,⁽⁵³⁾ and the complete genome was published to much fanfare by Douglas et al. in 2001⁽⁵²⁾: at 551 Kbp, it was the smallest nuclear genome ever sequenced.

First and foremost, the *G. theta* sequence proves beyond all doubt that nucleomorphs are bona fide nuclei of endosym-

biotic origin. The three *G. theta* chromosomes are capped with eukaryote-like telomeres (although with an atypical sequence: 5'-[AG]₇AAG₆A-3') and the genome encodes a variety of protein-coding genes typically associated with basic eukaryotic cellular processes. Somewhat unexpectedly, very few plastid protein-coding genes are present. Early speculation on the raison d'être of nucleomorphs centered on their obvious connection to the plastid and photosynthesis, and it seemed likely that their genomes—small as they may be—would encode a large number of plastid proteins, as did the nuclear genomes of the red and green algal ancestors from which they evolved. Surprisingly, the *G. theta* nucleomorph genome harbors only 30 plastid protein-coding genes (Table 1), even fewer than reside in the *G. theta* plastid genome,⁽⁵⁷⁾ meaning that hundreds of genes for plastid proteins must have moved from the ancestral cryptomonad nucleomorph to the host nuclear genome.^(52,60) This must be the case for genes associated with non-plastid functions as well, since a variety of obviously essential proteins such as DNA polymerases and amino-acyl-tRNA synthetases are not encoded in the *G. theta* nucleomorph.⁽⁵²⁾

The *G. theta* genome is remarkable in its extreme compactness. It packs 513 genes (465 protein coding) into 551 Kbp, and with 1.07 Kbp/gene and a mean intergenic distance of only 70 bp (Box 1), its gene density is on par with that of a typical bacterial genome. 44 gene pairs overlap by as much as 76 bp, a testament to the extreme evolutionary pressures that have compressed the genome since its enslavement. As is observed in other reduced genomes, such as those of mitochondria, plastids, and bacterial endosymbionts,⁽⁶¹⁾ the *G. theta* genome is also very compositionally biased (Fig. 3C). Apart from the rDNA operons and their immediate flanking regions, which have an A + T content of ~55%, the single-copy internal region of each of the three *G. theta* chromosomes is ~75% A + T (Fig. 3C). Preliminary data suggest that this is the case for other cryptomonad nucleomorphs as well.⁽⁴⁷⁾ There are no mobile elements encoded in the genome and as noted above, rDNA operons are adjacent to the telomeric repeats on each chromosome end. These operons are flanked by a region of apparently non-coding DNA, which is itself next to multicopy genes for the ubiquitin-conjugating enzyme *ubc4* and a transcription factor (*tf1ld*) (Fig. 3C). Essentially all other genes are present as single-copy. Taken as a whole, the *G. theta* nucleomorph genome is a model of compaction: its gene set is primarily 'housekeeping' in nature, encoding components of the transcription, translation, cell cycle, and protein folding and degradation machinery, with very few genes for end products of direct use to the cell.⁽⁵²⁾

The recent publication of the nucleomorph genome from the chlorarachniophyte *B. natans* by Gilson, McFadden and co-workers⁽⁶²⁾ has made it possible to compare and contrast two independently derived nucleomorphs and, by extension, better understand the evolutionary forces that have shaped

Table 1. Basic features of the sequenced nucleomorph genomes of the cryptomonad *Guillardia theta* and the chlorarachniophyte *Bigelowiella natans*

Genome characteristics	<i>Guillardia theta</i> ¹	<i>Bigelowiella natans</i> ¹
Evolutionary origin	red algal endosymbiont	green algal endosymbiont
Genome size	551,264 bp	372,870 bp
Chromosome number/size	three (196.2, 180.9 & 174.1 Kbp)	three (140.6, 134.1 & 98.1 Kbp)
Chromosome structure ²	sub-telomeric inverted repeats	sub-telomeric inverted repeats
Telomeric sequence/length	(5'-[AG] ₇ AAG ₆ A-3') ₁₁	(5'-TCTAGGG-3') ₂₅₋₄₅
Genomic A + T content:		
Inverted repeats (inc. rDNA)	~55%	~50%
Single-copy DNA	65–77%	>65%
Number of genes:		
Protein genes	465 ³	293
Non-mRNA	47	42
Pseudogenes	1	5
Total	513	340
Gene density ⁴	1.07 Kb/gene	1.10 Kb/gene
Mean intergenic distance ⁵	70 bp	113 bp
Overlapping genes	44 (max. 76 bp overlap)	not determined (max. 101 bp overlap)
Introns & size range	17 (42–52 bp)	852 (18–21 bp)
Plastid genes	30	17

¹Data taken primarily from Douglas et al.⁽⁵²⁾ and Gilson et al.⁽⁶²⁾ Numbers may vary slightly, depending on updated analyses and method of calculation.

²Refer to Figure 3 for precise structures of inverted repeats.

³Williams et al.⁽⁶⁴⁾ identified an *rpl30* gene not annotated in the original *G. theta* nucleomorph sequence.

⁴Calculated as genome size/total gene number.

⁵Numbers taken from Keeling and Slamovits.⁽⁶³⁾

the structure and composition of their genomes. At ~373 Kbp in size, the *B. natans* genome is even smaller than that of *G. theta*, setting a new standard for nuclear genome reduction. It too harbors protein-coding genes—293 in total—primarily involved in housekeeping functions, with only 17 genes for plastid proteins, even fewer than in *G. theta* (Table 1).⁽⁶²⁾ Overall, the relative proportions of *B. natans* and *G. theta* nucleomorph genes corresponding to discrete functional categories are very similar, although the *B. natans* complement is markedly more reduced. Absent are genes encoding tubulin subunits and proteins involved in proteasomal degradation, as well as genes for 5S rRNA, telomerase RNA and multiple tRNAs.⁽⁶²⁾ As is the case for most of the cells plastid proteins, these gene products are presumably encoded in the host nucleus and imported (see below).

Perhaps the most-striking difference between the *G. theta* and *B. natans* nucleomorph genomes is the abundance and size of their spliceosomal introns. Early work on the *B. natans* genome showed that its protein-coding genes harbored the smallest introns known,⁽⁵⁴⁾ and the recently published complete genome sequence reveals a whopping 852 of them—an average of 3.1 introns per gene and with a size range of only 18–21 nt.⁽⁶²⁾ By contrast, the *G. theta* genome has only 17 introns, 42–52 nt in size.⁽⁵²⁾ These differences probably reflect the intron abundance in the nuclear genomes of the red and green algal ancestors of cryptomonad and chlorarachniophyte nucleomorphs, which were probably intron-poor and intron-rich, respectively.⁽⁶²⁾ In the case of

chlorarachniophytes, while intron loss does not seem to have been a major factor in the compaction of the nucleomorph genome, the introns themselves have shrunk to a surprisingly uniform size that probably represents the absolute minimum size for effective splicing.⁽⁶²⁾ It is perhaps significant that despite its smaller size, the *B. natans* nucleomorph genome retains a variety of genes involved in RNA metabolism and splicing that are absent from the *G. theta* genome.⁽⁶²⁾

Nucleomorph genome reduction—causes and consequences

As discussed by Keeling and Slamovits,⁽⁶³⁾ genomes shrink by eliminating genes or packing the same number of genes into a smaller space. Nucleomorphs have clearly done both. In addition to their limited gene content and ultra-high gene density, nucleomorph genomes possess additional hallmark features of reductive evolution less obviously tied to intracellularly, including elevated A + T content,^(52,62) compositionally biased proteins^(52,64) and accelerated substitution rates.^(64–68) What is known about the mechanistic and evolutionary processes underlying the phenomenon of genome reduction in nucleomorphs?

Parallels have been drawn between the evolution of nucleomorphs, mitochondria and plastids.^(55,63) Genome reduction far beyond that observed in other nuclear genomes is tolerated in nucleomorphs in large part because, like organelles, they are able to transfer essential genes to the host cell nuclear genome. The presence of numerous

green or red algal-derived genes in both the *B. natans* and *G. theta* nuclear genomes confirms that, as expected, extensive nucleomorph-to-host-nucleus gene transfer has occurred^(69–71) and the process by which the products of these genes are targeted to the plastid or periplastid space is becoming increasingly well understood.^(71,72) However, little is known about the process of nucleomorph-to-host-nucleus gene transfer in cryptomonads and chlorarachniophytes. A recent analysis⁽⁵¹⁾ of closely related members of the cryptomonad genus *Hemiselmis* indicate that nucleomorph genome sizes can change on the order of ~20 Kbp over very short evolutionary time scales, suggesting that nucleomorph-to-host-nucleus transfer can, in principle, occur via the bulk movement of DNA, as is known to occur with mitochondria and plastids.^(73,74) However, transfer of DNA on a gene-by-gene basis via cDNA cannot at the present time be ruled out, and it will be interesting to compare what we find in nucleomorphs with what has been inferred with organellar systems.^(75,76)

The accelerated rates of sequence evolution seen in organelles and bacterial endosymbionts are often attributed to the effect of ‘Muller’s Ratchet’: mildly deleterious mutations are thought to accumulate in the genome by genetic drift, due to small effective population sizes and lack of sexual recombination.^(77–79) This phenomenon has also been suggested to account for the apparently fast evolution of nucleomorph genomes.^(55,80) The effective population size of nucleomorphs is indeed predicted to be miniscule, given that sex appears to be rare in cryptomonads and even when present, opportunities for genetic exchange between nucleomorph genomes would appear to be hindered by the numerous membranes surrounding the endosymbiont compartment (Fig. 3).⁽⁵⁵⁾ However, a robust explanation for the divergent nature of nucleomorph genomes remains elusive, because it is not known whether the mutation rates of nucleomorph genomes are higher than in canonical nuclear genomes and/or whether the functional constraints acting on the proteins functioning in the nucleomorph, periplastid space and plastid have been relaxed. It is possible that a drive towards an A + T biased genome and compositionally biased protein sequences is at least in part due to the loss and/or alteration of enzymes involved in DNA repair and recombination, as has been proposed in bacterial endosymbionts.^(61,81,82) That said, at least some recombination machinery must function in the nucleomorphs of both cryptomonads and chlorarachniophytes, given that gene conversion appears to be actively homogenizing the ends of the *G. theta* and *B. natans* chromosomes (Fig. 3).⁽⁵⁵⁾

An interesting recent study by Patron, Rogers and Keeling⁽⁸³⁾ systematically addressed the question of the rates of nucleomorph genome evolution, and the results suggest that the processes driving the evolution of nucleomorphs in cryptomonads and chlorarachniophytes may be quite different. These authors compared the evolutionary rates

of nucleus-, nucleomorph- and plastid-encoded proteins in *G. theta* and *B. natans* to their closest homologs in plants and algae, and concluded that, in *B. natans*, nucleomorph-encoded genes have accrued substitutions at a faster rate than homologs in other genomes, regardless of the cellular compartment in which their protein products function. In contrast, although most cryptomonad nucleomorph-encoded proteins have also evolved faster than their homologs in other genomes, nucleomorph-encoded plastid-targeted proteins appear to be evolving at about the same rate as plastid-targeted proteins encoded in the host cell nucleus. Furthermore, comparison of nucleus- and nucleomorph-encoded genes from two different cryptomonads, *G. theta* and *Rhodomonas salina* suggest that the nucleomorph genomes of these two species are in fact diverging less rapidly than their nuclear genomes.⁽⁸³⁾ Future experiments aimed at quantifying the mutation rate of nucleomorph genomes will likely shed further light on this issue.

Regardless of the underlying causes of genome reduction in nucleomorphs, the effects are clear. An interesting consequence of the extreme degree of compaction seen in the *B. natans* and *G. theta* nucleomorph genomes occurs at the level of transcription. Preliminary study of the *B. natans* nucleomorph revealed the co-transcription of two protein-coding genes, ClpP and snRNP E⁽⁵⁴⁾ and a more recent expressed sequence-tag-based study by Williams et al.⁽⁸⁴⁾ revealed a surprisingly high number of *B. natans* and *G. theta* mRNAs encoding more than one gene. Given that such genes are not preferentially located on the same strand, it is doubtful that these mRNAs represent operons. Rather, the implication is that during the extreme compaction of the cryptomonad and chlorarachniophyte nucleomorph genomes, transcriptional regulatory elements such as promoters and terminators were forced within or beyond adjacent genes, resulting in the production of mRNAs with un-translated extensions containing additional genes or gene fragments.⁽⁸⁴⁾

Perhaps the most biologically significant, yet poorly understood, consequence of genome reduction is its impact on the proteome. This can be seen at the level of the amino acid composition of proteins, where A + T-rich gene sequences encode proteins with an increased frequency of ‘FYMINK’ amino acids (phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine).⁽⁸⁵⁾ This trend was apparent in early work on genes encoding molecular chaperones in the *G. theta* nucleomorph genome, where asparagine codons were notably abundant,⁽⁶⁴⁾ and ongoing studies of nucleomorph-encoded proteins in *G. theta* and *Hemiselmis rufescens* reveal that, in extreme cases, more than 50% of a protein can comprise FYMINK residues (Archibald Laboratory, unpublished data). The significance of these observations is not clear but it is interesting that, in the case of bacterial endosymbionts in the genus *Buchnera*, where genomic A + T

content can approach 80%, protein integrity appears to be compromised to the point where folding efficiency is reduced⁽⁸⁶⁾ and protein function may even be lost or modified.^(87,88)

Finally, in addition to producing proteins with biased amino acid compositions, evidence suggests that genome reduction leads to a situation in which the genes and proteins themselves are shorter. Analysis of the 2.9 Mbp genome of the microsporidian *Encephalitozoon cuniculi* revealed that more than 85% of a conserved set of 350 proteins were shorter than their homologs in the related fungus *Saccharomyces cerevisiae*.⁽⁸⁾ The reason(s) underlying this discrepancy is unknown, but it is significant that, like the genomes that encode them, nucleomorph genes appear to have been pushed to extremes not observed in other eukaryotes. For example, a heat-shock transcription factor (HSF) encoded in the *G. theta* nucleomorph^(52,64) and recently identified in the *H. rufescens* genome (Archibald Laboratory, unpublished data) is a third shorter than the *E. cuniculi* homolog and less than half the size of HSFs in most plants, animals and fungi. The amino and carboxy termini of nucleomorph-encoded chaperonin genes in *G. theta* also appear slightly shorter than those in other eukaryotes.⁽⁶⁴⁾ More nucleomorph genomic data from diverse cryptomonad and chlorarachniophyte species will make it possible to investigate further the relationship between genome reduction and protein shortening and, perhaps, determine whether a deletion bias is at the root of nucleomorph gene and genome shrinkage.

Why do nucleomorphs persist and will they eventually disappear?

The nucleomorph genomes of cryptomonads and chlorarachniophytes are clearly the product of intense reductive evolution. Given that endosymbiont nuclei are absent in all other secondary plastid-containing organisms, the burning question is why have they been retained independently in these two lineages? As absurd as it seems, the persistence of at least one nucleomorph gene encoding an essential plastid protein demands that all the cellular machinery necessary for its proper expression and function be maintained, either imported from the host cell or encoded in the nucleomorph genome. In the words of Cavalier-Smith quoted above, a nucleomorph genome size of zero would be “optimal” in the sense that nucleomorph loss would release the cell from the genetic and metabolic load associated with DNA replication and the transcription and translation of hundreds of genes.⁽⁸⁹⁾ Again, the puzzle of nucleomorph retention is reminiscent of the debate over why mitochondria and plastids still retain genomes, and many of the same questions and possible answers are relevant.^(73,90,91) If organelle-to-nucleus gene transfer *can* happen, why hasn't it gone to completion? What are the possible barriers to gene relocation?

The presence of ultra-small introns in the *B. natans* nucleomorph genome led Gilson et al.⁽⁵⁴⁾ to speculate that

their tiny size might prevent the establishment of such genes in the chlorarachniophyte host nuclear genome. Having not co-evolved with the nucleomorph-encoded introns, the spliceosomal machinery functioning in the host cytosol might be incapable of recognizing them and catalyzing their removal. This may be the case, but the complete *B. natans* nucleomorph genome sequence reveals that six of its 17 plastid genes lack introns,⁽⁶²⁾ suggesting that these genes are, at least in principle, able to be transferred to the host nuclear genome. Scrutinizing the gene complements of the *G. theta* and *B. natans* genomes, Gilson et al.⁽⁶²⁾ could find no obvious barrier to the eventual complete loss of the nucleomorph genome in cryptomonads and chlorarachniophytes. Particularly striking is the almost complete lack of overlap in the suite of plastid proteins remaining in the two genomes—only two of the 17 *B. natans* plastid proteins are part of the set of 30 encoded in the *G. theta* nucleomorph. The author's conclude that “...no single plastid gene may be responsible for the retention of nucleomorphs, and they may yet disappear”.⁽⁶²⁾ It will be interesting to determine whether the suite of nucleomorph-encoded plastid proteins is conserved *within* diverse members of cryptomonads and chlorarachniophytes. If not, this will lend further support to the idea that nucleomorph genomes will ultimately be lost.

One additional point worth considering relates to the availability of nucleomorph DNA for transfer to the host nucleus of cryptomonads and chlorarachniophytes. In the case of plastids, considerable variation in the frequency of organelle-to-nucleus transfers have been documented in laboratory experiments with tobacco and the green alga *Chlamydomonas*,^(73,74,92) the differences being attributed to variation in the number of plastids per cell, and thus the amount of DNA available for nuclear integration through organelle lysis. The amount of time that nucleomorph genomes have been reducing in cryptomonads and chlorarachniophytes is debatable (see Ref. 93 for discussion), but one possible explanation for why the chlorarachniophyte nucleomorph genomes are noticeably smaller than those of cryptomonads is that chlorarachniophyte cells typically harbor multiple plastids and nucleomorphs per cell,^(37,94) and are, in theory, more likely to survive organelle lysis, a possible first step in DNA transfer. In contrast, cryptomonads usually have a single plastid and nucleomorph per cell,⁽⁹⁵⁾ the outer membrane of which is continuous with the nuclear envelope. A notable exception is in *Cryptomonas paramecium* (formerly *Chilomonas*) and its closest relatives, in which two plastids (and two nucleomorphs) are present.^(95–97) It is perhaps significant that *C. paramecium* harbors the smallest cryptomonad nucleomorph genome known, at ~450 Kbp.⁽⁴⁹⁾

Conclusions and future directions

From the above discussion, it is clear that, while complete nucleomorph genome sequences from *G. theta* and *B. natans*

have opened a window onto the origin and evolution of nucleomorphs, they also raise as many questions as they answer. One important area of future research will be to reveal the underlying source of nucleomorph genome size variation within cryptomonads and chlorarachniophytes. The size range observed in cryptomonads is particularly striking—between ~450 and ~845 Kbp^(47,49)—and it will be important to determine whether this is primarily the result of differences in the total number of genes, the amount of non-coding DNA (e.g. introns, intergenic spacers and repetitive DNA), or a combination of the two.⁽⁴⁷⁾ Comparison of the already complete 551 Kbp nucleomorph genome of *G. theta*⁽⁵²⁾ with the *Hemiselmis rufescens* genome (~580 Kbp⁽⁵¹⁾) and the genome of *Cryptomonas paramecium* (~450 Kbp), both of which are currently being sequenced (Archibald Laboratory, unpublished data), will be an important first step. These three organisms are quite distantly related to one another,^(47,51,95,98) and it will be interesting to see the extent of overlap in their complement of nucleomorph genes and what differences exist in terms of gene density. Comparison of the *C. paramecium*, *H. rufescens* and *G. theta* genomes will also provide a first glimpse at the extent of intrachromosomal and interchromosomal rearrangements occurring in nucleomorph genomes. A recent study by Slamovits et al.⁽⁹⁹⁾ revealed a significant reduction in the rate of genomic rearrangements in microsporidian genomes compared to those of other fungi, likely due to the smaller size of their intergenic regions. While the repeat structures on the ends of chlorarachniophyte and cryptomonad nucleomorph chromosomes are clearly subject to very active recombination (Fig. 3C,D)^(51,52,55,62), it is not known whether the structure of the single-copy regions of nucleomorph genomes is constrained by a paucity of non-coding DNA, as in microsporidians.⁽⁹⁹⁾ In the context of a robust phylogeny of cryptomonads and chlorarachniophytes, and with enough sequence data from diverse nucleomorph genomes in both groups, it should be possible to determine whether nucleomorph-to-host-nucleus gene transfer is essentially random or whether there is a core set of genes that appear recalcitrant to relocation. It may be that the reason(s) for nucleomorph retention in cryptomonads and chlorarachniophytes are different.

Although additional nucleomorph genome sequences from diverse cryptomonad and chlorarachniophyte species will be important, they will only tell us so much. During their tenure as endosymbionts, nucleomorphs have surrendered most of their genetic material to the host cell and, for this reason, the nuclear genome is the key to a more-complete understanding of these enigmatic organisms. To that end, *G. theta* and *B. natans* have become the focus of complete nuclear genome sequencing projects by the Joint Genome Institute's Community Sequencing Program (<http://www.jgi.doe.gov/sequencing/cspseqplans2007.html>) which will hopefully provide the missing pieces to many puzzles. For example,

a comprehensive set of nucleus-encoded, endosymbiont-targeted proteins from both organisms will make it possible to reveal the true complexity of the cellular processes still taking place in the nucleomorph and periplastid compartments of cryptomonads and chlorarachniophytes, and to determine whether they are in fact simplified compared to other eukaryotes.⁽⁵²⁾ Complete nuclear and nucleomorph genome sequences should also help to elucidate the process(es) by which nucleomorph-to-host-nucleus gene transfers occur. More generally, complete cryptomonad and chlorarachniophyte nuclear genomes will make it possible to assess the extent to which the nuclear genomes of the eukaryotic endosymbionts of both groups have contributed to their respective host cells. The presence of actin genes of red algal provenance in the nuclear genomes of different cryptomonads^(50,100) suggests that this is a distinct possibility. Conversely, it will be of interest to know whether host-derived genes of cryptomonads and chlorarachniophytes (as well as those obtained by lateral gene transfer^(60,69)) have acquired functions in their respective endosymbiont compartments, similar to the way in which the plastid proteomes of plants comprise significant numbers of proteins of eukaryotic ancestry.^(101,102)

Finally, as fascinating as they are, cryptomonads and chlorarachniophytes are currently experimentally intractable, and complete nucleus and nucleomorph genome sequences from both organisms should jump-start the development and use of genetic systems towards the goal of better understanding their basic biology. A clearer picture of the biochemical activities taking place in the endosymbiont compartments of cryptomonads and chlorarachniophytes will shed light on the question of whether their nucleomorphs have shrunk to a point beyond which further reduction is impossible or whether they are ultimately headed for extinction.

Acknowledgments

I sincerely thank Susan Douglas, Thomas Cavalier-Smith, Uwe Maier and Patrick Keeling for introducing me to nucleomorphs. Geoff McFadden, Paul Gilson, Nicola Patron, Hervé Philippe and the members of my laboratory are thanked for helpful discussion and advice, and Ford Doolittle, Andrew Roger and Chris Lane are thanked for comments on the manuscript. I also acknowledge two anonymous reviewers for providing helpful comments.

References

1. Gregory TR. 2001. Coincidence, coevolution or causation? DNA content, cell size, and the C-value enigma. *Biol Rev* 76:65–101.
2. Filipki A, Kumar S. 2005. Comparative genomics in eukaryotes. In: Gregory TR, editor. *The evolution of the genome*. London, UK: Elsevier Academic Press. p. 521–583.
3. Gregory TR. 2005. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann Bot* 95:133–146.
4. Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115:49–63.

5. Archibald JM. 2006. Genome complexity in a lean, mean photosynthetic machine. *Proc Natl Acad Sci USA* 103:11433–11434.
6. Derelle E, Ferraz C, Rombauts S, Rouzé P, Worden AZ, et al. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci USA* 103:11647–11652.
7. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, et al. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304:304–307.
8. Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, et al. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* 414:450–453.
9. Hoffman EP, Brown RH Jr, Kunkel LM. 1987. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. *Cell* 51:919–928.
10. Bhattacharya D, Yoon HS, Hackett JD. 2003. Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. *Bioessays* 26:50–60.
11. Keeling PJ. 2004. Diversity and evolutionary history of plastids and their hosts. *Am J Bot* 91:1481–1493.
12. Palmer JD. 2003. The symbiotic birth and spread of plastids: how many times and whodunit? *J Phycol* 39:4–11.
13. Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* 21:809–818.
14. Cavalier-Smith T. 2006. Cell evolution and Earth history: stasis and revolution. *Phil Trans Royal Soc B* 361:969–1006.
15. Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5:123–135.
16. McFadden GI. 1999. Plastids and protein targeting. *J Eukaryot Microbiol* 46:339–346.
17. Soll J, Schleiff E. 2004. Protein import into chloroplasts. *Nat Rev Mol Cell Biol* 5:198–208.
18. Bhattacharya D, Schmidt HA. Division Glaucocystophyta. 1997. In: Bhattacharya D, editor. *Origin of algae and their plastids*. Wein: Springer-Verlag. p. 139–148.
19. Hall WT, Claus G. 1963. Ultrastructural studies on the blue-green algal symbiont in *Cyanophora paradoxa* Korschikoff. *J Cell Biol* 19:551–563.
20. Pflanzagl B, Allmaier G, Schmid ER, de Pedro MA, Löffelhardt W. 1996. N-acetylputrescine as a characteristic constituent of cyanelle peptidoglycan in glaucocystophyte algae. *J Bacteriol* 178:6994–6997.
21. Pflanzagl B, Zenker A, Pittenauer E, Allmaier G, Martinez-Torrecuadrada J, et al. 1996. Primary structure of cyanelle peptidoglycan of *Cyanophora paradoxa*: a prokaryotic cell wall as part of an organelle envelope. *J Bacteriol* 178:332–339.
22. Moreira D, Le Guyader H, Philippe H. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* 405:69–72.
23. Rodríguez-Espeleta N, Brinkmann H, Burey SC, Roure B, Burger G, et al. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol* 15:1325–1330.
24. Archibald JM, Keeling PJ. 2005. On the origin and evolution of plastids. In: Saap J, editor. *Microbial phylogeny and evolution*. New York: Oxford University Press. p. 238–260.
25. Hackett JD, Anderson DM, Erdner DL, Bhattacharya D. Dinoflagellates: 2004. A remarkable evolutionary experiment. *Am J Bot* 91:1523–1534.
26. Archibald JM, Keeling PJ. 2002. Recycled plastids: a green movement in eukaryotic evolution. *Trends Genet* 18:577–584.
27. Bodyl A. 2005. Do plastid-related characters support the chromalveolate hypothesis? *J Phycol* 41:712–719.
28. Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J Eukaryot Microbiol* 46:347–366.
29. Delwiche CF. 1999. Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* 199:154:S164–S177.
30. McFadden GI. 2001. Primary and secondary endosymbiosis and the origin of plastids. *J Phycol* 37:951–959.
31. Taylor FJR. 1974. Implications and extensions of the serial endosymbiosis theory of the origin of eukaryotes. *Taxon* 23:229–258.
32. Gibbs SP. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Can J Bot* 56:2883–2889.
33. Greenwood AD. 1974. The Cryptophyta in relation to phylogeny and photosynthesis. In: Sanders JV, Goodchild DJ, editors. *Electron microscopy 1974*. Canberra: Australian Academy of Sciences. p. 566–567.
34. Greenwood AD, Griffiths HB, Santore UJ. 1977. Chloroplasts and cell compartments in Cryptophyceae. *Brit Phycol J* 12:119.
35. McFadden GI. 1993. Second-hand chloroplasts: Evolution of cryptomonad algae. In: Callow JA, editor. *Advances in Botanical Research*. Volume 19: London, UK: Academic Press Limited. p. 189–230.
36. McFadden GI, Gilson P. 1995. Something borrowed, something green: lateral transfer of chloroplasts by secondary endosymbiosis. *Trends Ecol Evol* 10:12–17.
37. Hibberd DJ, Norris RE. 1984. Cytology and ultrastructure of *Chlorarachnion reptans* (Chlorarachniophyta division nova, Chlorarachniophyceae classis nova). *J Phycol* 20:310–330.
38. Hansmann P. 1988. Ultrastructural localization of RNA in cryptomonads. *Protoplasma* 146:81–88.
39. Hansmann P, Falk H, Sitte P. 1985. DNA in the nucleomorph of *Cryptomonas* demonstrated by DAPI fluorescence. *Z Naturforsch* 40c: 933–935.
40. Ludwig M, Gibbs SP. 1989. Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyta) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. *J Phycol* 25:385–394.
41. Ludwig M, Gibbs SP. 1987. Are the nucleomorphs of cryptomonads and *Chlorarachnion* the vestigial nuclei of eukaryotic endosymbionts? *Ann New York Acad Sci* 501:198–211.
42. Douglas SE, Murphy CA, Spencer DF, Gray MW. 1991. Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* 350:148–151.
43. Hansmann P, Eschbach S. 1991. Isolation and preliminary characterization of the nucleus and the nucleomorph of a cryptomonad, *Pyrenomonas salina*. *Europ J Cell Biol* 52:373–378.
44. Maier UG, Hofmann CJ, Eschbach S, Wolters J, Igloi GL. 1991. Demonstration of nucleomorph-encoded eukaryotic small subunit ribosomal RNA in cryptomonads. *Mol Gen Genet* 230:155–160.
45. McFadden GI, Gilson PR, Hofmann CJ, Adcock GJ, Maier UG. 1994. Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc Natl Acad Sci USA* 91:3690–3694.
46. Eschbach S, Hofmann CJ, Maier UG, Sitte P, Hansmann P. 1991. A eukaryotic genome of 660 kb: electrophoretic karyotype of nucleomorph and cell nucleus of the cryptomonad alga, *Pyrenomonas salina*. *Nucleic Acids Res* 19:1779–1781.
47. Lane CE, Khan H, MacKinnon M, Fong A, Theophilou S, et al. 2006. Insight into the diversity and evolution of the cryptomonad nucleomorph genome. *Mol Biol Evol* 23:856–865.
48. McFadden GI, Gilson P, Douglas SE. 1994. The photosynthetic endosymbiont in cryptomonad cells produces both chloroplast and cytoplasmic-type ribosomes. *J Cell Sci* 107:649–657.
49. Rensing SA, Goddemeier M, Hofmann CJ, Maier UG. 1994. The presence of a nucleomorph hsp70 gene is a common feature of Cryptophyta and Chlorarachniophyta. *Curr Genet* 26:451–455.
50. Tanifuji G, Erata M, Ishida K-I, Onodera N, Hara Y. 2006. Diversity of secondary endosymbiont-derived actin-coding genes in cryptomonads and their evolutionary implications. *J Plant Res* 119:205–215.
51. Lane CE, Archibald JM. 2006. Novel nucleomorph genome architecture in the cryptomonad genus *Hemiselmis*. *J Eukaryot Microbiol* 53:515–521.
52. Douglas SE, Zauner S, Fraunholz M, Beaton M, Penny S, et al. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* 410:1091–1096.
53. Zauner S, Fraunholz M, Wastl J, Penny S, Beaton M, et al. 2000. Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph. *Proc Natl Acad Sci USA* 97:200–205.
54. Gilson PR, McFadden GI. 1996. The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns. *Proc Natl Acad Sci USA* 93:7737–7742.
55. Gilson PR, McFadden GI. 2002. Jam packed genomes—a preliminary, comparative analysis of nucleomorphs. *Genetica* 115:13–28.

56. Gilson PR, McFadden GI. 1999. Molecular, morphological and phylogenetic characterization of six chlorarachniophyte strains. *Phycol Res* 47:7–19.
57. Douglas SE, Penny SL. 1999. The plastid genome of the cryptophyte alga, *Guillardia theta*: complete sequence and conserved syntenic groups confirm its common ancestry with red algae. *J Mol Evol* 48:236–244.
58. Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte *Bigeloviella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol Biol Evol* 24:54–62.
59. Cavalier-Smith T. 2002. Nucleomorphs: enslaved algal nuclei. *Curr Opin Microbiol* 5:612–619.
60. Archibald JM. 2005. Jumping genes and shrinking genomes-probing the evolution of eukaryotic photosynthesis using genomics. *IUBMB Life* 57:539–547.
61. Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583–586.
62. Gilson PR, Su V, Slamovits CH, Reith ME, Keeling PJ, et al. 2006. Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc Natl Acad Sci USA* 103:9566–9571.
63. Keeling PJ, Slamovits CH. 2005. Causes and effects of nuclear genome reduction. *Curr Opin Genet Dev* 15:601–608.
64. Archibald JM, Cavalier-Smith T, Maier U, Douglas S. 2001. Molecular chaperones encoded by a reduced nucleus- the cryptomonad nucleomorph. *J Mol Evol* 52:490–501.
65. Cavalier-Smith T, Couch JA, Thorsteinsen KE, Gilson P, Deane JA, et al. 1996. Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. *Eur J Phycol* 31:315–328.
66. Ishida K, Green BR, Cavalier-Smith T. 1999. Diversification of a chimaeric algal group, the chlorarachniophytes: phylogeny of nuclear and nucleomorph small-subunit rRNA genes. *Mol Biol Evol* 16:321–331.
67. Keeling PJ, Deane JA, Hink-Schauer C, Douglas SE, Maier UG, et al. 1999. The secondary endosymbiont of the cryptomonad *Guillardia theta* contains alpha-, beta-, and gamma-tubulin genes. *Mol Biol Evol* 16:1308–1313.
68. Van de Peer Y, Rensing SA, Maier U-G. 1996. Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc Natl Acad Sci USA* 93:7732–7736.
69. Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ. 2003. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci USA* 100:7678–7683.
70. Deane JA, Fraunholz M, Su V, Maier UG, Martin W, et al. 2000. Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. *Protist* 151:239–252.
71. Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ, et al. 2006. Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. *Mol Biol Evol* 23:2413–2422.
72. Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, et al. 2006. Protein targeting into the complex plastid of cryptophytes. *J Mol Evol* 62:674–681.
73. Martin W. 2003. Gene transfer from organelles to the nucleus: frequent and in big chunks. *Proc Natl Acad Sci USA* 100:8612–8614.
74. Stegemann S, Hartmann S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci USA* 100:8828–8833.
75. Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* 29:380–395.
76. Nugent JM, Palmer JD. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66:473–481.
77. Clark MA, Moran NA, Baumann P. 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol Biol Evol* 16:1586–1598.
78. Lynch M. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol Biol Evol* 14:914–925.
79. Moran NA. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 93:2873–2878.
80. Gilson PR, McFadden GI. 2001. Genome sequencing. A grin without a cat. *Nature* 410:1040–1041.
81. Moran NA, Mira A. 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* 2:0054.1–0054.12.
82. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* 407:81–86.
83. Patron NJ, Rogers MB, Keeling PJ. 2006. Comparative rates of evolution in endosymbiotic nuclear genomes. *BMC Evol Biol* 6:46.
84. Williams BA, Slamovits CH, Patron NJ, Fast NM, Keeling PJ. 2005. A high frequency of overlapping gene expression in compacted eukaryotic genomes. *Proc Natl Acad Sci USA* 102:10936–10941.
85. Singer GA, Hickey DA. 2000. Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol* 17:1581–1588.
86. van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, et al. 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci USA* 100:581–586.
87. Risse C, Delmotte F, van Ham RC, Moya A. 2004. Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. *Genome Res* 14:44–53.
88. Shigenobu S, Watanabe H, Sakaki Y, Ishikawa H. 2001. Accumulation of species-specific amino acid replacements that cause loss of particular protein functions in *Buchnera*, an endocellular bacterial symbiont. *J Mol Evol* 53:377–386.
89. Cavalier-Smith T. 2005. Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann Bot* 95:147–175.
90. Martin W, Herrmann RG. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol* 118:9–17.
91. Race HL, Herrmann RG, Martin W. 1999. Why have organelles retained genomes? *Trends Genet* 15:364–370.
92. Lister DL, Bateman JM, Purton S, Howe CJ. 2003. DNA transfer from chloroplast to nucleus is much rarer in *Chlamydomonas* than in tobacco. *Gene* 316:33–38.
93. Cavalier-Smith T. 2006. The tiny enslaved genome of a rhizarian alga. *Proc Natl Acad Sci USA* 103:9379–9380.
94. Ota S, Ueda K, Ishida K-I. 2005. *Lotharella vacuolata* sp. nov., a new species of chlorarachniophyte algae, and time-laps video observations on its unique post-cell division behavior. *Phycol Res* 53:275–286.
95. Deane JA, Strachan IM, Saunders GW, Hill DRA, McFadden GI. 2002. Cryptomonad evolution: nuclear 18S rDNA phylogeny versus cell morphology and pigmentation. *J Phycol* 38:1236–1244.
96. Santore UJ. 1985. A cytological survey of the genus *Cryptomonas* (Cryptophyceae) with comments on its taxonomy. *Arch Protistenkd* 130:1–52.
97. Hoef-Emden K, Melkonian M. 2003. Revision of the genus *Cryptomonas* (Cryptophyceae): a combination of molecular phylogeny and morphology provides insights into a long-hidden dimorphism. *Protist* 154:371–409.
98. Marin B, Klingberg M, Melkonian M. 1998. Phylogenetic relationships among the Cryptophyta: Analyses of nuclear-encoded SSU rRNA sequences support the monophyly of extant plastid-containing lineages. *Protist* 149:265–276.
99. Slamovits CH, Fast NM, Law JS, Keeling PJ. 2004. Genome compaction and stability in microsporidian intracellular parasites. *Curr Biol* 14:891–896.
100. Stibitz TB, Keeling PJ, Bhattacharya D. 2000. Symbiotic origin of a novel actin gene in the cryptophyte *Pyrenomonas helgolandii*. *Mol Biol Evol* 17:1731–1738.
101. Abdallah F, Salamini F, Leister D. 2000. A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci* 5:141–142.
102. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* 99:12246–12251.