

The Complete Mitochondrial Genome Sequence of the Haptophyte *Emiliana huxleyi* and its Relation to Heterokonts

M. Virginia SÁNCHEZ PUERTA, Tsvetan R. BACHVAROFF, and Charles F. DELWICHE*

Department of Cell Biology and Molecular Genetics, University of Maryland College Park, College Park, MD 20742-5815, USA

(Received 11 November 2003)

Abstract

The complete nucleotide sequence of the mitochondrial genome of *Emiliana huxleyi* (Haptophyta) was determined. *E. huxleyi* is the most abundant coccolithophorid, key in many marine ecosystems, and plays a vital role in the global carbon cycle. The mitochondrial genome contains genes encoding three subunits of cytochrome *c* oxidase, apocytochrome *b*, seven subunits of the NADH dehydrogenase complex, two ATPase subunits, two ribosomal RNAs, 25 tRNAs and five ribosomal proteins. One potentially functional open reading frame was identified, with no counterpart in any other organism so far studied. The *cox1* gene transcript is apparently spliced from two distant segments in the genome. One of the most interesting features in this mtDNA is the presence of the *dam* gene, which codes for a DNA adenine methyltransferase. This enzyme is common in bacterial genomes, but is not present in any studied mitochondrial genome. Despite the great age of this group (ca. 300 Ma), little is known about the evolution of haptophytes or their relationship to other eukaryotes. This is the first published haptophyte organellar genome, and will improve the understanding of their biology and evolution and allow us to test the monophyly of the chromoalveolate clade.

Key words: *Emiliana huxleyi*; mitochondria; Haptophyta; genome sequencing; chromoalveolates

1. Introduction

Emiliana huxleyi (Lohmann) Hay & Mohler is the most abundant of the coccolithophorids, a key group of marine phytoplankton. It has been the subject of numerous studies, but only a few of these employed genetic approaches. This species has been relatively well studied because of its potential importance in the global carbon cycle.^{1,2} It is capable of forming large blooms in all oceans, particularly at mid latitudes, can reach cell densities of 10⁷ cells/L and cover thousands of square kilometers.^{3–7} *E. huxleyi* blooms emit large amounts of dimethylsulfide (DMS), which upon oxidation in the atmosphere is considered an active component in the nucleation of refractive clouds.^{8,9} Therefore, this species is regarded as a key component of the greenhouse effect, natural acid rain and albedo regulation.

Little is known about the evolution of haptophytes and their phylogenetic relationships to other living organisms.^{10–13} On the basis of pigmentation, it has been suggested that the haptophytes belong to a monophyletic group of organisms with chlorophyll *c* containing plastids,

called chromoalveolates^{14–16} but the evolutionary history of plastids does not necessarily reflect that of the whole cell. Mitochondrial genome analysis has been recognized as a valuable tool for resolving evolutionary relationships among the various eukaryotic lineages.^{17,18} The diversity in mitochondrial genome size, gene content and organization is an important tool to elucidate the mechanisms and to reconstruct the pathway by which this evolutionary diversification has occurred.¹⁷ Because this is the first genome to be sequenced in the phylum Haptophyta, it is a unique opportunity to increase our understanding in the biology and evolution of the members that comprise this group, in particular the widely distributed and environmentally important *E. huxleyi*. The phylogenetic position of haptophytes has been examined with single-gene phylogenies.^{10–12,17} Our approach is to understand protist evolution using comparative analyses of whole mitochondrial genomes, which allow comprehensive phylogenetic analysis. In this work, we report the complete sequence of the mitochondrial DNA (mtDNA) of the coccolithophorid *E. huxleyi*, and describe the main features and phylogenetic hypotheses derived from the newly available data.

Communicated by Masahiro Sugura

* To whom correspondence should be addressed. Tel. +1-301-405-8286, Fax. +1-301-314-9081, E-mail: delwiche@umd.edu

2. Materials and Methods

The complete mtDNA sequence of *Emiliania huxleyi* has been deposited in GenBank (accession number AY342361).

2.1. Culture of *E. huxleyi* and mtDNA isolation

The axenic strain of *E. huxleyi* was obtained from Provasoli-Guillard National Center for Culture of Marine Phytoplankton (CCMP # 373). Cultures were grown in Guillard's f/2 medium¹⁹ at 17°C with a 14h/10h L:D cycle. Approximately 6 L of culture were harvested by centrifugation, flash frozen in liquid nitrogen, and stored at -80°C. For total DNA extraction and organellar DNA purification, the protocol designed by Chesnick and Cattolico²⁰ was followed. Mitochondrial, chloroplast and nuclear DNA were separated through CsCl-bisbenzimidazole isopycnic centrifugation in a TLA-110 Rotor at 70 krpm for 17 hr, by which mtDNA forms the least dense band.

2.2. Cloning and DNA sequencing

Mitochondrial DNA was digested with the restriction endonuclease *Hind*III and the resulting fragments were cloned in pGEM -3Zf(+) (Promega, WI) using *Escherichia coli* XL-10 Gold Ultracompetent Cells (Stratagene, CA) as the host bacterium. Plasmids from individual clones were isolated using the 'miniprep' procedure,²¹ and sequenced using dye terminator chemistry (ABI). The M13-20 primer was used for 5' and T7 primer for 3' sequencing. Primer walking was used to determine the full sequence of longer clones and to obtain double stranded sequencing reads. The polymerase chain reaction was used to order the fragments, fill gaps and obtain double-stranded coverage.

2.3. Data analysis

Sequences were edited using the program Sequencher (GeneCodes Corp., Ann Arbor, MI). Vector and low quality bases were removed, and manual editing was performed. Sequence reads were assembled using the contig assembly function of Sequencher.

Putative open reading frames (ORFs) were identified by performing BLAST searches of the GenBank databases at the National Center for Biotechnology Information (NCBI). Similarity searches to detect tRNAs were performed with tRNAscan SE Search Server (Washington University, St. Louis) (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE/>). General codon usage analyses were performed using GCUA.²² Correspondence analysis of codon usage by genes and the indices, frequency of optimal codons (Fop) and codon adaptation index (CAI), were calculated using CodonW (University of Nottingham) (<http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html>).

2.4. Phylogenetic trees

The *cob*, *cox1*, *cox2* and *cox3* genes were first aligned using individual protein alignments with ClustalW (www.cmbi.kun.nl/bioinf/tools/clustalw.shtml) output as a starting point, then manually edited with MacClade 4.0.²³ Unalignable regions were excluded and the four protein-coding genes were concatenated as a nucleotide alignment. Trees were constructed both with and without the third codon position. For maximum likelihood (ML) analysis using PAUP* 4b10 parameters were estimated from a Fitch-Margoliash tree using LogDet distances. In the likelihood analysis, the General Time Reversible model with Invariant site and gamma correction was used (GTR + I + Γ) with four rate categories. Bootstrap analysis was performed using three random additions with nearest neighbor interchange. Bayesian analysis was performed using the MrBayes program²⁴ with the same likelihood model as the PAUP* searches, i.e., GTR + I + Γ 4. Four Markov chains were run with one heated for 2×10^6 generations, sampled every hundredth generation with a burn-in period of 1000 generations.

3. Results and Discussion

3.1. Overall organization of *E. huxleyi* mtDNA

The mitochondrial genome of *E. huxleyi* is a circular molecule of 29,013 bp. Figure 1 depicts the physical and gene map of the mtDNA. The overall A + T content is 71.7%, with protein-coding regions being 72% A + T and intergenic spacers 76% A + T. This base composition is comparable to those of *Cyanidioschyzon merolae* (72.8%),²⁵ *Chondrus crispus* (72.1%)²⁶ and *Reclinomonas americana* (73.9%),²⁷ but higher than that of *Marchantia polymorpha* (57.6%)²⁸ and *Arabidopsis thaliana* (55.2%).²⁹ The genetic information is densely packed, with 78% of sequence specifying genes, ORFs and structural RNAs, and only 22% without detectable coding content. All the genes are encoded on the same strand suggesting that the genome is transcribed in one unit, like the mitochondrial genomes of *Monosiga brevicollis*, *Acanthamoeba castellanii*, *Dictyostelium discoideum*, *Chlamydomonas eugametos* and *Pedinomonas minor*.¹⁷ Table 1 lists all the genes and ORFs in the mtDNA of *E. huxleyi*. No overlapping genes were detected.

A comparison of gene order between *E. huxleyi* and *Pavlova lutheri* mtDNAs shows that no gene clusters are conserved between these two organellar genomes, although the gene content is not strikingly different (<http://megasun.bch.umontreal.ca/ogmp/projects/pluth/gen.html>). There are several features that separate the members of the class Pavlovophyceae (to which *P. lutheri* belongs) from members of the Pymnesiophyceae including *E. huxleyi*. Chloroplast genes³⁰ and the 18S ribosomal DNA gene³¹ based phylogenies showed that

Table 1. Genes identified in *E. huxleyi* mtDNA.^a

rRNA	Small subunit (1): <i>rns</i>
	Large subunit (1): <i>rnl</i>
Transfer RNAs (25) (See figure 1 and table 2)	
Ribosomal protein	Small subunit (4): <i>rps3, rps8, rps12, rps14</i>
	Large subunit (1): <i>rpl16</i>
Electron transport and oxidative phosphorylation	
Respiratory chain	NADH dehydrogenase (7): <i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6</i>
	Ubiquinol: cytochrome <i>c</i> oxidoreductase (1): <i>cob</i>
	Cytochrome <i>c</i> oxidase (3): <i>cox1, cox2, cox3</i>
	ATP synthase (3): <i>atp4, atp6, atp9</i>
Other proteins	DNA adenine methyltransferase (1): <i>dam</i>
ORFs unique to <i>E. huxleyi</i> mtDNA (1)	ORF104

^a Genes are classified according to their function. Numbers in parentheses indicate the number of genes in a particular class. The number indexing the ORF refers to the number of amino acid residues in the deduced polypeptide.

coded on the same strand of the genome, interspersed with other genes, and are separated in the genome by almost 10 kbp. Because only *cox1a* has a translational initiation codon and other genes are interspersed between the two *cox1* coding regions, splicing events would be required to produce the mature message. The *nad1* gene is cis and trans-spliced in plant mitochondrial genomes where the gene is split into different exons.^{35–37} An alternative hypothesis would be that the *cox1* sequence in the mitochondrion is no longer functional, but the high degree of sequence conservation means that it must have been under selection until very recently.

3.3. A novel feature in mtDNA: adenine methyltransferase

A unique feature of the mtDNA of *E. huxleyi* is the presence of the *dam* gene, which codes for DNA adenine methyltransferase (A-Mtase). This enzyme catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to the N6 position of a specific adenine in their cognate sequence.³⁸ This gene is not included in the standard set of mitochondrial genes and has not been reported for any mitochondrial genome so far studied. Significant BLAST hits of this gene were exclusively bacteria, viruses and members of Archaea, with the exception of *Anopheles gambiae* (EAA02205). Adenine methylation plays an important role in replication, mismatch repair and segregation of chromosomal DNA in *E. coli*, as well as regulation of gene expression and attenuation of the virulence of a number of pathogens.³⁹ The putative functional role of the *dam* gene in *E. huxleyi* could be related to mitochondrial DNA replication, modulation of gene expression or/and control of virulence of some pathogens, such as viruses. It is known that viruses infecting *E. huxleyi* can control and terminate blooms of

this organism⁷ and some viruses are known to target the mitochondrial genome.⁴⁰

Sequence comparisons among the members of this group of methyltransferases have revealed nine conserved motifs, of which motif I and motif IV are highly conserved.⁴¹ These two motifs are involved in AdoMet binding and methyl group transfer. The two conserved domains of this protein are recognized in *E. huxleyi*, suggesting that it may be functional. Furthermore, the complete sequence does not suggest that it is a pseudogene.

This type of methyltransferase is commonly found in bacteria, Archaea and viruses. N6-Methylated adenine has been found in DNA of eukaryotes, such as protozoa,^{42,43} fungi,⁴⁴ higher plants⁴⁵ and animals;⁴⁶ however, the putative genes responsible for the methylation are found in the nuclear genome and correspond to a different type than the *E. huxleyi* *dam* gene. A nuclear encoded N6 A-Mtase isolated from wheat coleoptiles seems to be responsible for mitochondrial DNA modification that might be involved in the regulation of replication of mitochondria in plants.⁴⁵

Correspondence analysis of the protein-coding genes in mtDNA of *E. huxleyi* (see below) revealed differences in some of the genes, including the *dam* gene. The location of the *dam* gene in the mitochondrial genome is also suggestive, since it occurs adjacent to the tandem repeats. In angiosperm mitochondria, inverted and also direct repeats appear to promote major genome rearrangements.⁴⁷ Three alternative scenarios could account for the presence of the *dam* gene in the mtDNA of *E. huxleyi*: (1) lateral transfer of the *dam* gene from a phage or bacterial DNA; (2) vertical transmission of an ancient gene that was present in the proteobacterial progenitor of the mitochondrion; and (3) lateral transfer of the gene from the nucleus or chloroplast to the

Table 2. Codon usage in the mitochondrial genome of *E. huxleyi*.^a

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	482	1.62	Ser	UCU	147	1.93
	UUC	114	0.38		UCC	10	0.13
Leu	UUA	424	3.44		UCA	133	1.75
	UUG	80	0.65		UCG	42	0.55
Tyr	UAU	140	1.33	Cys	UGU	51	1.48
	UAC	71	0.67		UGC	18	0.52
ter	UAA	19	1.81	Trp	UGA	73	1.73
ter	UAG	2	0.19		UGG	11	0.26
Leu	CUU	116	0.94	Pro	CCU	73	1.73
	CUC	6	0.05		CCC	9	0.21
	CUA	99	0.80		CCA	68	1.61
	CUG	15	0.12		CCG	19	0.45
His	CAU	70	1.43	Arg	CGU	53	2.16
	CAC	28	0.57		CGC	11	0.45
Gln	CAA	109	1.85		CGA	29	1.18
	CAG	9	0.15		CGG	2	0.08
Ile	AUU	336	2.10	Thr	ACU	141	1.87
	AUC	64	0.40		ACC	11	0.15
	AUA	81	0.51		ACA	119	1.58
Met	AUG	139	1.00		ACG	31	0.41
Asn	AAU	200	1.40	Ser	AGU	95	1.25
	AAC	86	0.60		AGC	29	0.38
Lys	AAA	221	1.61	Arg	AGA	46	1.88
	AAG	53	0.39		AGG	6	0.24
Val	GUU	228	2.38	Ala	GCU	146	1.74
	GUC	17	0.18		GCC	33	0.39
	GUA	111	1.16		GCA	126	1.50
	GUG	27	0.28		GCG	31	0.37
Asp	GAU	89	1.47	Gly	GGU	152	1.92
	GAC	32	0.53		GGC	24	0.30
Glu	GAA	111	1.45		GGA	107	1.35
	GAG	42	0.55		GGG	34	0.43

^a Codons are indicated in upper case letters, amino acid residues are in three-letter code; termination codons (UAA and UAG) are indicated by ter. The total number of individual codons (N) and the relative synonymous codon usage (RSCU) are shown.

mitochondria. To distinguish among these alternatives the distribution and phylogeny of this gene need to be examined more completely.

3.4. Codon usage

Table 2 shows the codon frequency in genes and ORFs of *E. huxleyi* mtDNA. As expected for an extremely A+ T-rich genome, codons ending in A or T vastly outnumber the synonymous codons ending in G or C. The mtDNA of *E. huxleyi* does not use the standard genetic code. The codon UGA, usually serving as a translational termination signal, is used for Tryptophan (Trp). This assignment is based on protein alignments, since the codon UGA occurs where the codon for Trp is conserved in other organisms. This codon (TGA) is used preferentially, accounting for 86.7% of all Trp codons (Table 2). This is the most common deviation from the standard translation code in mitochondria and is also found in other haptophytes,⁴⁸ *C. crispus*,²⁶ *A. castellanii*,⁴⁹ animals, fungi, and ciliates. However, members of the order

Pavloales (phylum Haptophyta) use the universal genetic code.⁵⁰ It is believed that the reassignment of the TGA codon to Trp occurred independently many times in the evolution of the protozoa.⁵⁰

No in-frame ATG start codon is present in the reading frame of *E. huxleyi atp4* or *rps8* genes. Protein alignments suggest that GTG may serve as the codon for translation initiation in these genes. This has been reported for the mitochondrial genome of several different organisms, such as *Porphyra purpurea*,⁵¹ *Paramecium aurelia*,⁵² and *Oenothera berteriana*,⁵³ although not referring to the same genes.

Several analyses were performed to compare the codon usage of the genes present in *E. huxleyi* mtDNA. No significant differences in codon usage, G + C content or G + C content of the silent third position were observed in identified protein-coding genes or the unique ORF. However, by means of correspondence analysis of codon usage, we were able to detect that some genes were considerably different from the others, such as *atp4*, *atp9*

Table 3. Sources of sequences used in phylogenetic analyses.

<i>Acanthamoeba castellani</i>	NC_001637
<i>Allomyces macrogynus</i>	NC_001715
<i>Beta vulgaris</i>	NC_002511
<i>Cafeteria roenbergensis</i>	NC_000946
<i>Chaetosphaeridium globosum</i>	NC_004118
<i>Chondrus crispus</i>	NC_001677
<i>Chrysodidymus synuroideus</i>	NC_002174
<i>Cyanidioschyzon merolae</i>	NC_000887
<i>Cyanidium caldarium</i>	Z48930
<i>Emiliania huxleyi</i>	AY342361
<i>Gracilariopsis lemaneiformis</i>	AF118119
<i>Homo sapiens</i>	NC_001807
<i>Laminaria digitata</i>	NC_004024
<i>Malawimonas jakobiformis</i>	NC_002553
<i>Monosiga brevicollis</i>	NC_004309
<i>Ochromonas danica</i>	NC_002571
<i>Phytophthora infestans</i>	NC_002387
<i>Platymonas subcordiformis</i>	Z47795, Z47797
<i>Podospora anserina</i>	NC_001329
<i>Porphyra purpurea</i>	NC_002007
<i>Prototheca wickerhamii</i>	NC_001613
<i>Pylaiella littoralis</i>	NC_003055
<i>Reclinomonas americana</i>	NC_001823
<i>Rhodomonas salina</i>	NC_002572
<i>Triticum aestivum</i>	AF337547, Y00417, AF336134, X15944
<i>Xenopus laevis</i>	NC_001573

and *dam*. The disparity of the *dam* gene was only detectable in the second axis of ordination. The G content of silent third position in *dam* gene was 20.7%, which is the highest among the protein-coding genes of *E. huxleyi* (average 11.6%). No significant differences were detected for ORF 104 using correspondence analysis.

3.5. tRNAs

The tRNAs are scattered throughout the entire genome, either singly or in groups, and all lack introns. Based on the anticodon sequence of the 25 tRNAs, 24 decode the standard 20 amino acids, while one recognizes UGA, which is used for tryptophan. All the tRNA sequences can assume standard cloverleaf secondary structures, with few departures from the conventional structure, as tested with tRNAscan. This set of mitochondrial-encoded tRNAs is not sufficient to decode the 62 sense codons that occur in protein-coding sequences, even when taking into account wobble and the possible modifications of their anticodons.⁵⁴ The tRNAs, which are not in the mtDNA, may be imported from the cytosol or generated from another tRNA by partial editing or post transcriptional modification, as suggested by Ohta *et al.*²⁵ A minimum of one tRNA gene remains to be identified in order to account for the complete translation of the *E. huxleyi* mitochondrial genetic information; namely *trnL* (CAA) for leu (UUG). In addition,

trnG (GCC) for gly (GGU and GGC) is also missing. The *trnG* (UCC) decodes GGA and GGG anticodons and is possible that it also recognizes GGU and GGC anticodons, as in *C. crispus*.²⁶ Also, the *trnW* (CCA) for Trp (UGG) is not present in *E. huxleyi* mtDNA. However, *trnW* (UCA), which recognizes the UGA codon as tryptophan, may also be able to decode the UGG codon for the same amino acid, as reported for *Tetrahymena pyriformis* mtDNA.⁵⁵ The mitochondrial genome of *E. huxleyi* contains 3 tRNA genes having the methionine anticodon CAU, which include the elongator and initiator methionine-accepting mitochondrial tRNAs.

3.6. Phylogenetic analyses of the mitochondrion of *E. huxleyi*

The 1236 aa concatenated alignment was composed of *cob* (358 aa), *cox1* (470 aa), *cox2* (169 aa) and *cox3* (239 aa). Table 3 lists the accession numbers in GenBank of the species included in the phylogenetic analyses. The placement of *E. huxleyi* in phylogenetic trees varies according to the method of analysis and taxa present. Figure 2 shows the placement of *E. huxleyi* as sister to *Malawimonas jakobiformis* when the third codon position is excluded and ML analysis is performed with PAUP*. However, the Fitch-Margoliash tree using LogDet distances places *E. huxleyi* sister to the Opisthokont clade, whereas in the Bayesian analyses *E. huxleyi* forms a clade

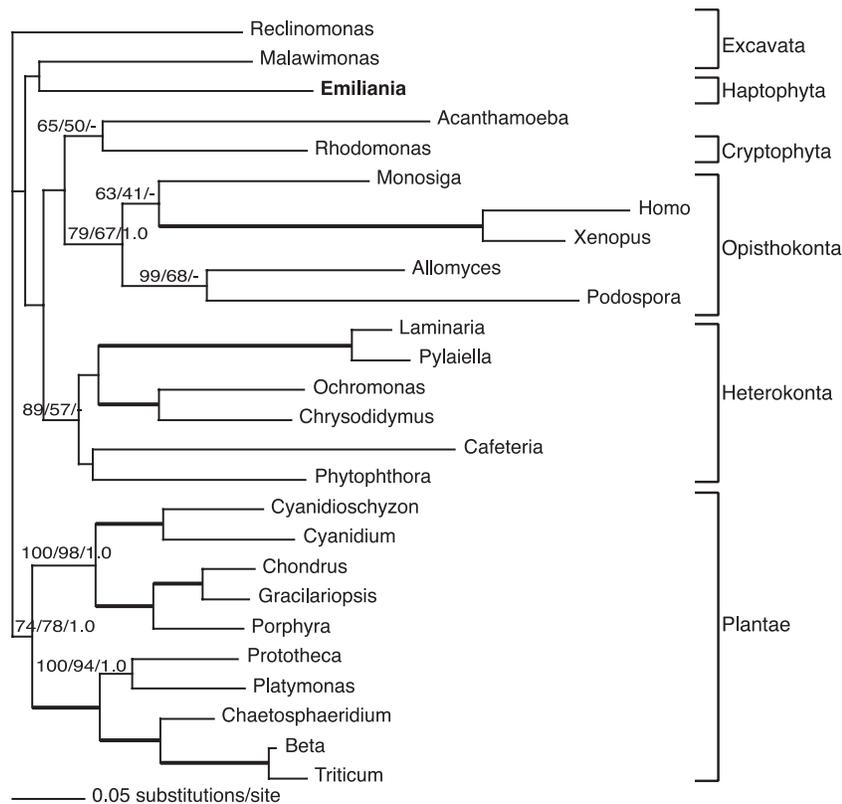


Figure 2. A maximum likelihood tree inferred from the first two codon positions of a concatenated *cob*, *cox1*, *cox2*, and *cox3* alignment using PAUP* with a GTR + I + Γ 4 model of sequence evolution. The numbers above the branches indicate the bootstrap proportion using the first two codon positions, the bootstrap proportion when all positions are used, and the Bayesian posterior probability of the branch when all positions are used with the GTR + I + Γ 4 model, where the same branches were recovered. The thicker branches indicate 100% bootstrap support with both datasets, and a posterior probability of 1.0.

with the Opisthokonts, the *Acanthamoeba castellanii*-*Cafeteria roenbergensis* clade, and *Rhodomonas salina* (see supplemental information). The placement of *E. huxleyi* is affected by the inclusion of the alveolates; with the third position excluded, *E. huxleyi* is sister to the *C. roenbergensis* clade (see supplemental information). The alveolates themselves form a monophyletic group with extremely long branches that has good support with Bayesian methods but little support using PAUP with or without the third codon position. In all trees, regardless the analytical method used, three clades are consistently recovered; a Plantae clade including red and green algae, an opisthokont clade including the choanoflagellate *Monosiga brevicollis* as well as fungi and animals, and a moderately to weakly supported heterokont clade. The placement of *C. roenbergensis* within the heterokont clade was not always supported. These trees do not identify a specific relationship of the haptophytes to heterokonts, cryptophytes or alveolates.

Haptophytes are a monophyletic group of protists that were formerly placed with the heterokonts in the class Chrysophyceae,^{56,57} and some modern authors still emphasize a close relationship among these taxa.⁵⁸ Ultrastructural and molecular evidence indicated that the

haptophytes are a monophyletic group with the primary synapomorphy being a characteristic appendage, the haptonema.^{59–61} Studies using the nuclear genes actin and small subunit ribosomal DNA support the distinctiveness of the haptophytes but have not shown their affinity to any other group.^{10–13} Plastid-encoded or plastid-derived genes do show a relationship to the other chlorophyll *c* containing algae^{11,13} indicating that the plastids of haptophytes, heterokonts, dinoflagellates and cryptophytes form a monophyletic group.^{62,63} However, these data are not informative for relationships among cytosolic genomes.

Although phylogenetic analyses clearly support a red algal origin of chromophyte plastids, the number of events that gave rise to them and the phylogenetic relationships among the host cells remain unclear. There are two competing hypotheses regarding the number of endosymbiotic events in the chromophyte algae. In the first of these, a single endosymbiotic event for all the chromophytes would imply that the host cells are also monophyletic,^{14,15,63} and that plastid-loss has occurred in the non-photosynthetic lineages. The clade postulated by this hypothesis is known as the chromoalveolates.^{16,64} An alternative hypothesis would

infer multiple endosymbiotic events in separate chromophyte lineages, would allow host and endosymbiont phylogenies to be incongruent,⁶⁵ and would imply a non-photosynthetic ancestor for each lineage. Several intermediate hypotheses could also be proposed. The latter, polyphyletic hypothesis has fallen into disfavor on the basis of analyses of glyceraldehyde-3-phosphate dehydrogenase GAPDH.^{66,67} However, most authors seem to agree that the question remains unresolved. Our analyses examined the host cell phylogenetic relationships using concatenated mitochondrial genes in order to assess the monophyly of the chromoalveolate clade. Mitochondrial data can be used to resolve relationships among eukaryotes, such as the monophyletic origin of red and green primary plastids and can complement data from the nuclear genome. Like previous analyses of mitochondrial data,⁵¹ our data support a single origin of the primary plastid. The hypothesis of a single origin of the chlorophyll *c* containing (or chromophyte) hosts is not supported by these data, but the bootstrap and posterior probabilities are weak, leaving open the possibility that this hypothesis is correct. *E. huxleyi* is clearly excluded from the heterokonts in these analyses, but its placement as sister taxon to the heterokonts cannot be rejected.

Although the concatenated mitochondrial data presented here do not strongly resolve the relationships among chlorophyll *c* containing algae, they are difficult to reconcile with the chromoalveolate hypothesis and indicate that further study of relationships among these taxa is needed. The identity of the sibling taxon to haptophytes remains an unsolved problem.

Acknowledgements: We are grateful to Kenneth G. Karol, Greg Concepcion, and other members of the Delwiche lab for advice and useful discussion, as well as to Gertraud Burger and Organelle Genome Megasequencing Program for making the *Pavlova lutheri* mitochondrial genome map available on the web. Supported in part by NSF grant MCB-9984284.

References

- Riebesell, U., Zondervan, I., Rost, B., Tortell, P. D., Zeebe, R. E., and Morel, F. M. M. 2000, Reduced calcification of marine plankton in response to increased atmospheric CO₂, *Nature*, **407**, 364–367.
- Falkowski, P., Scholes, R. J., Boyle, E., et al. 2000, The global carbon cycle: a test of our knowledge of earth as a system, *Science*, **290**, 291–296.
- Paasche, E. 2002, A review of the coccolithophorid *Emiliania huxleyi* (Prymnesiophyceae), with particular reference to growth, coccolith formation, and calcification-photosynthesis interactions, *Phycologia*, **40**, 503–529.
- Balch, W. M., Holligan, P. M., and Kilpatrick, K. A. 1992, Calcification, photosynthesis and growth of the bloom-forming coccolithophore, *Emiliania huxleyi*, *Cont. Shelf Res.*, **12**, 1353–1374.
- Winter, A. and Siesser, W. G. 1994, *Coccolithophores*, Cambridge University Press, Cambridge.
- Brown, C. W. and Yoder, J. A. 1994, Coccolithophorid blooms in the global oceans, *J. Geophys. Res.*, **99**, 7467–7482.
- Green, J. C. and Harris, R. P. 1996, EHUX (*Emiliania huxleyi*), *J. Marine Syst.*, **9**, 1–136.
- Malin, G., Turner, S., and Liss, P. 1992, Sulfur: the plankton/climate connection, *J. Phycol.*, **28**, 590–597.
- Malin, G. and Kirst, G. O. 1997, Algal production of dimethyl sulfide and its atmospheric role, *J. Phycol.*, **33**, 889–896.
- Bhattacharya, D., Stickel, S. K., and Sogin, M. L. 1993, Isolation and molecular phylogenetic analysis of actin-coding regions from *Emiliania huxleyi*, a prymnesiophyte alga, by reverse transcriptase and PCR methods, *Mol. Biol. Evol.*, **10**, 689–703.
- Medlin, L. K., Barker, G. L. A., Campbell, L., et al. 1996, Genetic characterisation of *Emiliania huxleyi* (Haptophyta), *J. Marine Syst.*, **9**, 13–31.
- Medlin, L. K., Kooistra, W. H. C. F., Potter, D., Saunders, G. W., and Andersen, R. A. 1997, Phylogenetic relationships of the ‘golden algae’ (haptophytes, heterokonts, chromophytes) and their plastids, *Plant Syst. Evol.*, **11 Supplement**, 187–219.
- Tengs, T., Dahlberg, O. J., Shalchian-Tabrizi, K., et al. 2000, Phylogenetic analyses indicate that the 19′ hexanoxy-fucoanthin-containing dinoflagellates have tertiary plastids of haptophyte origin, *Mol. Biol. Evol.*, **17**, 718–729.
- Calvalier-Smith, T. 1989, In: Green, J., Leadbeater, B. S. C., and Diver, W. (eds) *The Chromophyte Algae: problems and perspectives*, Systematics Association Special Vol. 38. Clarendon Press, Oxford, pp. 381–407.
- Calvalier-Smith, T. 1981, Eukaryote kingdoms: seven or nine?, *BioSystems*, **14**, 461–481.
- Calvalier-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.
- Gray, M. W., Lang, B. F., Cedergren, R., et al. 1998, Genome structure and gene content in protist mitochondrial DNAs, *Nucleic Acids Res.*, **26**, 865–878.
- Gray, M. W., Burger, G., and Lang, B. F. 1999, Mitochondrial evolution, *Science*, **283**, 1476–1481.
- Andersen, R. A., Morton, S. L., and Sexton, J. P. 1997, Provasoli-Guillard National Center for Culture of Marine Phytoplankton 1997 list of strains, *J. Phycol.*, **33**, 1–75.
- Chesnick, J. M. and Cattolico, R. A. 1993, Isolation of DNA from eukaryotic algae, *Methods Enzymol.*, **224**, 168–176.
- Sambrook, J. and Russell, D. W. 2001, *Molecular cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- McInerney, J. O. 1998, GCUA: General Codon Usage Analysis, *Bioinformatics*, **14**, 372–373.
- Maddison, W. and Maddison, P. 2000, *MacClade version 4: analysis of phylogeny and character evolution*, Sinauer Associates, Sunderland, MA.
- Huelsenbeck, J. P. and Ronquist, F. 2001, MRBAYES:

- Bayesian inference of phylogeny, *Bioinformatics*, **17**, 754–755.
25. Ohta, N., Sato, N., and Kuroiwa, T. 1998, Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence, *Nucleic Acids Res.*, **26**, 5190–5198.
 26. Leblanc, C., Boyen, C., Richard, O., Bonnard, G., Grienenberger, J. M., and Kloareg, B. 1995, Complete sequence of the mitochondrial DNA of the rhodophyte *Chondrus crispus* (Gigartinales). Gene content and genome organization, *J. Mol. Biol.*, **250**, 484–495.
 27. Lang, B. F., Burger, G., O'Kelly, C. J., et al. 1997, An ancestral mitochondrial DNA resembling a eubacterial genome in miniature, *Nature*, **387**, 493–496.
 28. Oda, K., Yamato, K., Ohta, E., et al. 1992, Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome, *J. Mol. Biol.*, **223**, 1–7.
 29. Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. 1997, The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides, *Nature Genet.*, **15**, 57–61.
 30. Fujiwara, S., Tsuzuki, M., Kawachi, M., Minaka, N., and Inouye, I. 2001, Molecular phylogeny of the Haptophyta based on the *rbcL* gene and sequence variation in the spacer region of the Rubisco operon, *J. Phycol.*, **37**, 121–129.
 31. Edvardsen, B., Eikrem, W., Green, J. C., Andersen, R. A., Moon-van der Staay, S. Y., and Medlin, L. K. 2000, Phylogenetic reconstruction of the Haptophyta inferred from 18S ribosomal DNA sequences and available morphological data, *Phycologia*, **39**, 19–35.
 32. Young, J., Didymus, J. M., Bown, P. R., Prins, B., and Mann, S. 1992, Crystal assembly and phylogenetic evolution in heterococcoliths., *Nature*, **356**, 516–518.
 33. Richard, O., Bonnard, G., Grienenberger, J. M., Kloareg, B., and Boyen, C. 1998, Transcription initiation and RNA processing in the mitochondria of the red alga *Chondrus crispus*: Convergence in the evolution of transcription mechanisms in mitochondria, *J. Mol. Biol.*, **283**, 549–557.
 34. Burger, G., Lang, B. F., Braun, H. P., and Marx, S. 2003, The enigmatic mitochondrial ORF *yfm39* codes for ATP synthase chain b, *Nucleic Acids Res.*, **31**, 2353–2360.
 35. Conklin, P. L., Wilson, R. K., and Hanson, M. R. 1991, Multiple *trans*-splicing events are required to produce a mature *nad1* transcript in a plant mitochondrion, *Gene Dev.*, **5**, 1407–1415.
 36. Wissinger, B., Schuster, W., and Brennicke, A. 1991, *Trans*-splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon and *Trans*-splicing group II intron sequences, *Cell*, **65**, 473–482.
 37. Chapelaine, Y. and Bonen, L. 1991, The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: A *trans*-splicing model for this gene-in-pieces, *Cell*, **65**, 465–472.
 38. Ahmad, I. and Rao, D. N. 1996, Chemistry and biology of DNA methyltransferases, *Crit. Rev. Biochem. Mol.*, **31**, 361–380.
 39. Heithhoff, D. M., Soinsheimer, R. L., Low, D. A., and Mahan, M. J. 1999, An essential role for DNA adenine methylation in bacterial virulence, *Science*, **284**, 967–970.
 40. Hong, Y. G., Dover, S. L., Cole, T. E., Brasier, C. M., and Buck, K. W. 1999, Multiple mitochondrial viruses in an isolate of the Dutch elm disease fungus *Ophiostoma novo-ulmi*, *Virology*, **258**, 118–127.
 41. Malone, T., Blumenthal, R., and Cheng, X. 1995, Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyl-transferases, and suggests a catalytic mechanism for these enzymes., *J. Mol. Biol.*, **253**, 618–632.
 42. Rae, P. and Spear, B. 1978, Macronuclear DNA of the hypotrichous ciliate *Oxytricha fallax*, *Proc. Natl Acad. Sci. USA*, **75**, 4992–4996.
 43. Capowski, E. E., Wells, J. M., Harrison, G. S., and Karrer, K. M. 1989, Molecular analysis of N6-Methyladenine patterns in *Tetrahymena thermophila* nuclear DNA, *Mol. Cell. Biol.*, **9**, 2598–2605.
 44. Rogers, S., Rogers, M., Saunders, G. W., and Holt, G. 1986, Isolation of mutants sensitive to 2-aminopurine and alkylating agents and evidence for the role of DNA methylation in *Penicillium chrysogenum*, *Curr. Genet.*, **1986**, 557–560.
 45. Fedoreyeva, L. I. and Vanyushin, B. F. 2002, N6-Adenine DNA-methyltransferase in wheat seedlings, *FEBS Letters*, **514**, 305–308.
 46. Kay, P. H., Pereira, E., and Marlow, S. A. 1994, Evidence for adenine methylation within the mouse myogenic gene MYP-D1, *Gene*, **151**, 89–95.
 47. Hanson, M. R. and Folker, T. S. 1992, Structure and function of the higher plant mitochondrial genome, *Int. Rev. Cytol.*, **141**, 129–172.
 48. Hayashi-Ishimaru, Y., Ehara, M., Inagaki, Y., and Ohama, T. 1997, A deviant mitochondrial genetic code in prymnesiophytes (yellow-algae): UGA codon for tryptophan, *Curr. Genet.*, **32**, 296–299.
 49. Burger, G., Plante, I., Lonergan, K. M., and Gray, M. W. 1995, The mitochondrial genome of the amoeboid protozoon, *Acanthamoeba castellanii*. Complete sequence, gene content and genome organization, *J. Mol. Biol.*, **239**, 476–499.
 50. Inagaki, Y., Ehara, M., Watanabe, K. I., Hayashi-Ishimaru, Y., and Ohama, T. 1998, Directionally evolving genetic code: the UGA codon from stop to tryptophan in mitochondria, *J. Mol. Evol.*, **47**, 378–384.
 51. Burger, G., Saint-Louis, D., Gray, M. W., and Lang, B. F. 1999, Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: cyanobacterial introns and shared ancestry of red and green algae, *Plant Cell*, **11**, 1675–1694.
 52. Pritchard, A. E., Seilhamer, J. J., Mahalingam, R., Sable, C. L., Venuti, S. E., and Cummings, D. J. 1990, Nucleotide sequence of the mitochondrial genome of *Paramecium*, *Nucleic Acids Res.*, **18**, 173–180.
 53. Bock, H., Brennicke, A., and Schuster, W. 1994, *Rps3* and *rpl16* genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria?, *Plant Mol. Biol.*, **14**, 811–818.
 54. Crick, F. H. C. 1966, Codon-anticodon pairing: the

- wobble hypothesis, *J. Mol. Biol.*, **19**, 548–555.
55. Burger, G., Zhu, Y., Littlejohn, T. G., et al. 2000, Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA, *J. Mol. Biol.*, **297**, 365–380.
 56. Pascher, A. 1910, Chrysoomonaden aus dem Hirschberger Grossteiche. Untersuchungen über die Flora des Hirschberger Grossteiches. I. Teil., *Monogr. Abh. Int. Rev. Gesamten Hydrobiol. Hydrogr.*, **1**, 1–66.
 57. Bourrelly, P. 1957, Recherches sur les Chrysophycées: morphologie, phylogénie, systématique, *Rev. Algol. Mém. Hors-Séeri*, **1**, 1–412.
 58. Andersen, R. A. 1991, The cytoskeleton of chromophyte algae, *Protoplasma*, **164**, 143–159.
 59. Christensen, T. 1962, In: Bocher, T., Lange, M., and Sorensen, T. (eds) Botanik, Munksgaard, Copenhagen., pp. 1–178.
 60. Green, J. C. and Leadbeater, B. S. C. 1994, The haptophyte algae, Clarendon Press, Oxford.
 61. Christensen, T. 1989, In: Green, J. C., Leadbeater, B. S. C., and Diver, W. (eds) The Chromophyte Algae: Problems and Perspectives, Systematics Association Special Vol. 38. Clarendon Press, Oxford, pp. 1–12.
 62. Harper, J.T. and Keeling, P.J. 2003, Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids, *Mol. Biol. Evol.*, **20**, 1730–1735.
 63. Yoon, H. S., Hackett, J. D., Pinto, G., and Bhattacharya, D. 2002, The single, ancient origin of chromist plastids, *Proc. Natl. Acad. Sci. USA*, **99**, 15507–15512.
 64. Cavalier-Smith, T. 2002, Genomic reduction and evolution of novel membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae), *Phil. Trans. R. Soc. Lond.*, **358**, 109–134.
 65. Cavalier-Smith, T., Allsopp, M., and Chao, E. 1994, Chimeric conundra: are nucleomorphs and chromists monophyletic or polyphyletic?, *Proc. Natl. Acad. Sci. USA*, **91**, 11368–11372.
 66. Fast, N. M., Kissinger, J. C., Roos, D. S., and Keeling, P. J. 2001, Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids, *Mol. Biol. Evol.*, **18**, 418–426.
 67. Palmer, J. D. 2003, The symbiotic birth and spread of plastids; how many times and whodunit?, *J. Phycol.*, **39**, 4–11.