The Complete Mitochondrial Genome Sequence of the Haptophyte Emiliania 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 Emiliania 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: Emiliania huxleyi; mitochondria; Haptophyta; genome sequencing; chromoalveolates

1. Introduction

Emiliania 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 10^7 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 chromoalveolates14–16 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 coccolithophor E. huxleyi, and describe the main features and phylogenetic hypotheses derived from the newly available data.

Communicated by Masahiro Sugiura
* 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 medium\(^{19}\) 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 organelar DNA purification, the protocol designed by Chesnick and Cattolico\(^{20}\) was followed. Mitochondrial, chloroplast and nuclear DNA were separated through CsCl-bisbenzimide isopycnic centrifugation in a TLA-110 Rotor at 70 krpm for 17 hr, by which mtDNA forms the densest 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 Sequencer (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 Sequencer.

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 program\(^{24}\) with the same likelihood model as the PAUP* searches, i.e., GTR + I + Γ. Four Markov chains were run with one heated for 2 × 10⁸ 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 castellani*, *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 organelar genomes, although the gene content is not strikingly different (http://megasun.bch.umontreal.ca/ogmp/projcts/pluth/gen.html). There are several features that separate the members of the class Pavlovophyceae (to which *P. lutheri* belongs) from members of the Prymnesiophyceae including *E. huxleyi*. Chloroplast genes\(^{30}\) and the18S ribosomal DNA gene\(^{31}\) based phylogenies showed that
members of these two groups form two distinct clades, which presumably diverged between 220 and 300 Ma. With additional complete mtDNA sequences from other members of the phylum Haptophyta and related groups of organisms, the comparative analysis of gene order pattern should be useful to understand the evolutionary changes that these genomes have undergone since the species diverged.

Intergenic regions vary in size from 1 to 2624 bp, with the majority being 1–100 bp long and only two exceeding 250 bp. The longest of these, located downstream of trnI and upstream of dam, encompasses two types of direct repeat motifs. One is a 147-bp repeat that occurs as a tandem array of 5 motifs. The other, which occurs adjacent to the first, is 246 bp long and is also arranged in a tandem array of 5 motifs. The intergenic region contains a 45-nt stem loop structure which was detected downstream of the tandem repeats and upstream of the dam gene (Fig. 1). The G + C content of this region is 26% and this value is approximately equal to the average of the mtDNA of E. huxleyi (28.3%). This region may play a role in transcription initiation or DNA replication as in the red alga C. crispus.

However, no significant sequence similarity between E. huxleyi and C. crispus stem loops is discernable.

3.2. Gene content

The mitochondrial genome of E. huxleyi codes for 21 proteins, including 14 components of the respiratory chain and 5 ribosomal proteins (Table 1). None of the protein-coding genes contain introns, although introns are present in other mitochondrial genomes, including the haptophyte P. lutheri, the red alga C. crispus and the liverwort M. polymorpha. Three of the genes encoding subunits of the ATP synthase complex, atp4 (ymf39), atp6 and atp9, were identified. Recently, ymf39 was identified as atp4. Seven genes encoding NADH dehydrogenase subunits, nad1-6 and nad4L, were detected; these are not grouped together. In addition to these 21 protein-coding genes, a total of 27 RNA genes are present in the E. huxleyi mitochondrial genome, coding for 25 tRNAs and the small and large subunit (rrs, rrl) rRNAs. A 5S rRNA gene was not detected. One unique 104 amino acid (aa) ORF (ORF 104) was present, and it lacks significant similarity to any entry in the public domain sequence databanks. A distinguishing organizational feature in E. huxleyi mtDNA is the presence of two separate coding regions that show similarity to coxl. The coxl segment encodes the N-terminal 88 aa residues, and coxlb specifies the C-terminal 433 residues. Both segments are en-
Other proteins
- DNA adenine methyltransferase (1): *dam*
- ORFs unique to *E. huxleyi* mtDNA (1)
- ORF104

Table 1: Genes identified in *E. huxleyi* mtDNA.

<table>
<thead>
<tr>
<th>Category</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>rRNA</td>
<td>Small subunit (1): <em>rns</em></td>
</tr>
<tr>
<td></td>
<td>Large subunit (1): <em>rln</em></td>
</tr>
<tr>
<td>Transfer RNAs (25)</td>
<td>(See figure 1 and table 2)</td>
</tr>
<tr>
<td>Ribosomal protein</td>
<td>(4): <em>rps3, rps8, rps12, rps14</em></td>
</tr>
<tr>
<td></td>
<td>Large subunit (1): <em>rpl16</em></td>
</tr>
<tr>
<td>Electron transport and oxidative</td>
<td></td>
</tr>
<tr>
<td>phosphorylation</td>
<td></td>
</tr>
<tr>
<td>Respiratory chain</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NADH dehydrogenase (7): <em>nad1</em>,<em>nad2, nad3, nad4, nad4L, nad5, nad6</em></td>
</tr>
</tbody>
</table>
|                                 | Ubiquinol: cytochrome c
|                                 | oxidoreductase (1): *cob*       |
|                                 | Cytochrome c oxidase (3): *cox1, cox2, cox3* |
|                                 | ATP synthase (3): *atp4, atp6, atp9* |

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.

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. 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 gambiace* (EAA02205). Adenine methylation plays an important role in replication, mismatch repair and segregation of chromosomonal DNA in *E. coli*, as well as regulation of gene expression and attenuation of the virulence of a number of pathogens. The putative functional role of the *dam* gene in *E. huxleyi* could be related to mitochondrial DNA replication, modulation of gene expression or control of virulence of some pathogens, such as viruses. It is known that viruses infecting *E. huxleyi* can control and terminate blooms of this organism and some viruses are known to target the mitochondrial genome.

Sequence comparisons among members of this group of methyltransferases have revealed nine conserved motifs, of which motif I and motif IV are highly conserved. 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, fungi, higher plants and animals; 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.

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. 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 plage 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*.\(^a\)

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

\(^a\) 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 Pavlovalles (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*...
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.30 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 \textit{cob}, \textit{cox1}, \textit{cox2}, and \textit{cox3} alignment using PAUP\textsuperscript{*} with a GTR + I + \Gamma_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 + \Gamma_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 \textit{Acanthamoeba castellani-Cafeteria roenbergensis} clade, and \textit{Rhodomonas salina} (see supplemental information). The placement of \textit{E. huxleyi} is affected by the inclusion of the alveolates; with the third position excluded, \textit{E. huxleyi} is sister to the \textit{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 \textit{Monosiga brevicollis} as well as fungi and animals, and a moderately to weakly supported heterokont clade. The placement of \textit{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,\textsuperscript{56,57} and some modern authors still emphasize a close relationship among these taxa.\textsuperscript{58} Ultrastructural and molecular evidence indicated that the haptophytes are a monophyletic group with the primary synapomorphy being a characteristic appendage, the haptonema.\textsuperscript{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.\textsuperscript{10–13} Plastid-encoded or plastid-derived genes do show a relationship to the other chlorophyll \textit{c} containing algae\textsuperscript{11,13} indicating that the plastids of haptophytes, heterokonts, dinoflagellates and cryptophytes form a monophyletic group.\textsuperscript{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,\textsuperscript{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.\textsuperscript{16,64} An alternative hypothesis would
infer multiple endosymbiotic events in separate chlorophyte lineages, would allow host and endosymbiont phylogenies to be incongruent, 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. 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 chloroalveolate 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, our data support a single origin of the primary plastid. The hypothesis of a single origin of the chlorophyll c containing (or chlorophyte) 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 chloroalveolate 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

24. Huelsenbeck, J. P. and Ronquist, F. 2001, MRBAYES:


54. Crick, F. H. C. 1966, Codon-anticodon pairing: the


60. Green, J. C. and Leadbeater, B. S. C. 1994. The hapto-


nase (GAPDH) indicates a single origin for chromalveo-


