Complete Sequence of the Mitochondrial DNA of the Red Alga Porphyra purpurea: Cyanobacterial Introns and Shared Ancestry of Red and Green Algae

Gertraud Burger,a,b,1 Diane Saint-Louis,b Michael W. Gray,a,c and B. Franz Langa,b

a Program in Evolutionary Biology, Canadian Institute for Advanced Research, 180 Dundas Street West, Toronto, Ontario M5G 1Z8, Canada
b Département de Biochimie, Université de Montréal, 2900 Boulevard Édouard-Montpetit, Montréal, Québec H3T 1J4, Canada
c Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

The mitochondrial DNA (mtDNA) of Porphyra purpurea, a circular-mapping genome of 36,753 bp, has been completely sequenced. A total of 57 densely packed genes has been identified, including the basic set typically found in animals and fungi, as well as seven genes characteristic of protist and plant mtDNAs and specifying ribosomal proteins and subunits of succinate:ubiquinone oxidoreductase. The mitochondrial large subunit rRNA gene contains two group II introns that are extraordinarily similar to those found in the cyanobacterium Calothrix sp, suggesting a recent lateral intron transfer between a bacterial and a mitochondrial genome. Notable features of P. purpurea mtDNA include the presence of two 291-bp inverted repeats that likely mediate homologous recombination, resulting in genome rearrangement, and of numerous sequence polymorphisms in the coding and intergenic regions. Comparative analysis of red algal mitochondrial genomes from five different, evolutionarily distant orders reveals that rhodophyte mtDNAs are unusually uniform in size and gene order. Finally, phylogenetic analyses provide strong evidence that red algae share a common ancestry with green algae and plants.

INTRODUCTION

Historically considered to be “red plants,” red algae (rhodophytes) were grouped together with protists only in the middle of this century; however, debate continues as to when this taxonomic group first appeared in the evolutionary history of mitochondria-containing eukaryotes. Because the red algal cell lacks flagellar basal bodies and centrioles, some workers have claimed that rhodophytes are the most early diverging and primitive group among photosynthetic eukaryotes (e.g., Lee, 1989). This view also has been suggested by certain molecular phylogenies (e.g., Lipscomb, 1989; Stiller and Hall, 1997). Some authors even advocate that red algae, in particular Cyanidioschyzon spp and Cyanidium spp, represent the evolutionary bridge between cyanobacteria and eukaryotes (Seckbach, 1994). An opposite interpretation contends that the red algae represent a derived group that has secondarily lost many of the distinctive features of the protistan cytoskeleton (Pueschel, 1990; Scott and Broadwater, 1990). Finally, it has been proposed that red algae may have given rise to the fungi (Demoulin, 1985) or that red algae are affiliated with the plant kingdom via common ancestry with green algae (see, e.g., Bhattacharya et al., 1993; Cavalier-Smith, 1993; McFadden et al., 1994; Schlegel, 1994; Paquin et al., 1997; Lang et al., 1998). This broad spectrum of coexisting, mutually exclusive hypotheses highlights our limited knowledge about the phylogenetic relationship between rhodophytes and other eukaryotic phyla.

Rhodophytes form a morphologically heterogeneous phylum that has more species (~2500 to 6000 in at least 12 orders; Woelkerling, 1990) than all other seaweeds combined. Multicellular as well as unicellular rhodophyte genera exist, and there is also considerable variety in morphology and physiology throughout the phylum. On the basis of the observed divergence in nuclear small subunit (SSU) rRNA sequences, Ragan et al. (1994) concluded that “rhodophytes are more divergent among themselves than are fungi or green algae and green plants together.”

In 1995, the first complete mitochondrial DNA (mtDNA) sequence from a red alga, Chondrus crispus (Gigartinales), was published (Leblanc et al., 1995a). In molecular phylogenies using single nuclear genes, members of the Gigartinales form one of the late-diverging rhodophyte orders (Ragan et al., 1994; Saunders and Kraft, 1997). In many respects, the
genome structure and gene repertoire of C. crispus mtDNA resemble those of plants and protists, but the C. crispus mitochondrial genome displays neither exceptionally primitive characters nor especially derived traits. No "new" mitochondrial genes are present, as occurs in the mtDNA of the heterotrophic flagellate Reclinomonas americana (Lang et al., 1997), nor are otherwise ubiquitous mitochondrial genes missing, as occurs, for example, in Chlamydomonas spp (see Nedelcu, 1998) or animals (see Wolstenholme, 1992). However, the extent to which these unexpectedly "normal" features of the C. crispus mtDNA are characteristic of the red algal assemblage as a whole remains to be explored.

We chose to sequence the mtDNA of Porphyra purpurea, a member of the Bangiales, because this order is considered to have diverged before the Gigartinales and most other orders (all except Compsopogonales and some members of the Porphyridiales; Ragan et al., 1994; Saunders and Kraft, 1997). In addition, we have sequenced several mitochondrial genes of Gracilariapis lemaneiformis (Gracilariales; GenBank accession number AF11819). In the meantime, a 10-kb portion of the mtDNA from Cyanidium caldarium, a member of the Porphyridiales, was published (Viehmann, 1995), and at a late stage in the preparation of this manuscript, the complete mtDNA sequence of Cyanidioschyzon merolae was released (Ohta et al., 1998). With these new data, we are now in a position to undertake comparative gene and genome analysis encompassing several early and late diverging red algal orders. Such phylogenetically broad information should not only provide us with a better understanding of mitochondrial genome diversity within this lineage but also allow us to reinvestigate questions about when red algae emerged in the evolutionary history of eukaryotes and how rapidly this group has been evolving.

RESULTS

Physical Properties, Gene Content, and Overall Organization of P. purpurea mtDNA

The mtDNA of P. purpurea maps as a circular molecule of 36,753 bp. The overall AT content of the mitochondrial genome is 66.5%, with intergenic spacers being significantly richer in A+T (72.0%) than are coding regions (66.0%). Figure 1 shows that genetic information is densely packed in the mtDNA, with 91% of the sequence specifying genes and open reading frames (ORFs) and only 9% without detectable coding content. Genes are encoded on both strands, and their orientation suggests that the genome is transcribed in two units, starting from a bidirectional promoter located between trnQ and rtl (Figure 1, double-headed arrow). In fact, this intergenic region contains a stem-loop structure that might be involved in transcription initiation, as suggested for the chicken mitochondrial promoter (L’Abbé et al., 1991). In Chondrus crispus mtDNA as well, the bipolar transcription initiation site coincides with a palindromic repeat (Richard et al., 1998). However, we were unable to detect in the trnQ to rtl intergenic region of P. purpurea a sequence element resembling the promoter motif proposed for C. crispus mtDNA.

Table 1 lists the genes in P. purpurea mtDNA that code for 26 structural RNAs (i.e., the SSU and large subunit [LSU] rRNAs and 24 transfer RNAs) and 21 proteins, including 17 components of the respiratory chain and ATP synthase and four ribosomal proteins. In addition, two conserved ORFs are found in P. purpurea mtDNA, one of which (ytmf16) has been identified recently (Table 1). Open reading frames contained within the two r1 introns are of the reverse transcriptase type (Michel and Lang, 1985; Xiong and Eickbush, 1990). Also present in P. purpurea mtDNA are four unique ORFs lacking significant similarity to any entry in the public domain sequence databanks, as well as fragmented DNA polymerase (dpo) and reverse transcriptase (rtl) genes.

Three classes of genes detected in the P. purpurea mtDNA have counterparts in only a few other protist taxa. (1) Among the four ribosomal protein genes (rps3, rps11, rps12, and rpl16), rps11 has been found only in a dozen species, including members of early diverging plants, green algae, jakobids, and other protists (Gray et al., 1998; G. Burger, B.F. Lang, and M.W. Gray, unpublished results). The remaining three ribosomal protein genes are invariably present in protist mtDNAs when these genomes contain ribosomal protein genes. (2) To date, only four non-rhodophyte species are known in which subunits of the succinate:ubiquinone oxidoreductase (respiratory complex II) are mitochondrially encoded: the liverwort Marchantia polymorpha (Oda et al., 1992; Daignan-Fornier et al., 1994; Burger et al., 1996); the jakobid flagellates R. americana (Burger et al., 1996; Lang et al., 1997) and j akoba libera (G. Burger, B.F. Lang, C. O’Kelly, and M.W. Gray, unpublished results); and the cryptophyte alga Rhodomonas salina (Gray et al., 1998). (3) Mitochondrially encoded DNA polymerase (dpo) genes are also present in only a few eukaryotic species, with the majority being fungi and plants. These sequences are generally located on mitochondrial plasmids (reviewed in Kempken et al., 1992; Weber et al., 1995), whereas the P. purpurea dpo is contained within the mtDNA itself. All known mitochondrial dpo genes, whether plasmid encoded or not, belong to the B-type family, related to DNA polymerase II of Escherichia coli (classification according to Braithwaite and Ito, 1993). (4) Finally, mitochondrial reverse transcriptase (or reverse transcriptase-like) genes (rtl) that are free-standing, that is, not in introns, have been detected in only two other species, M. polymorpha (Oda et al., 1992) and Chlamydomonas reinhardtii (Boer and Gray, 1988).

Mitochondrial Gene Content and Order in P. purpurea and Other Red Algae

The gene content of P. purpurea and Chondrus crispus mtDNAs (Leblanc et al., 1995a) is identical, except that rm5
Porphyra purpurea
Mitochondrial DNA 1677

(specifying 5S rRNA) and rpl20 are missing from P. purpurea mtDNA, whereas trnS(gcu), dpo, and rtl are absent from C. crispus mtDNA (identification of C. crispus rns is described below). Compared with the mtDNA of the former two rhodophytes, Cyanidioschyzon merolae mtDNA (Ohta et al., 1998) encodes six or seven more ribosomal proteins and four components involved in c-type cytochrome biogenesis (yejU, yejV, yejW, and yejR [ccl1]), which are otherwise known from plant, ciliate, and jakobid mtDNAs (Gray et al., 1998). A 5S rRNA gene is also present in Cyanidioschyzon merolae mtDNA, located downstream of rpl6, at the same position at which we have identified a homolog in the Cyanidium caldarium mtDNA sequence (see below).

Comparison of gene order in the three completely sequenced rhodophyte mtDNAs reveals that the highest resemblance is between the P. purpurea and Chondrus
Table 1. Genes Identified in P. purpurea mtDNA

<table>
<thead>
<tr>
<th>rRNAs (2)</th>
<th>Ribosomal proteins (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small subunit (1): rns</td>
<td>Small subunit (3): rps3, rps11, rps12</td>
</tr>
<tr>
<td>Large subunit (1): rml</td>
<td>Large subunit (1): rpl16</td>
</tr>
<tr>
<td>Transfer RNAs (24) (see Figure 1 and Table 2)</td>
<td>Electron transport and oxidative phosphorylation (17)</td>
</tr>
<tr>
<td>Ribosomal proteins (4)</td>
<td>Respiratory chain (14)</td>
</tr>
<tr>
<td>Small subunit (3): rps3, rps11, rps12</td>
<td>NADH dehydrogenase (7): nad1, nad2, nad3, nad4, nad4L, nad5, nad6</td>
</tr>
<tr>
<td>Large subunit (1): rpl16</td>
<td>Succinate:ubiquinone oxidoreductase (3): sdh2, sdh3, sdh4</td>
</tr>
<tr>
<td>Electron transport and oxidative phosphorylation (17)</td>
<td>Ubiquinol:cytochrome c oxidoreductase (1): cob</td>
</tr>
<tr>
<td>Respiratory chain (14)</td>
<td>Cytochrome c oxidase (3): cox1, cox2, cox3</td>
</tr>
<tr>
<td>NADH dehydrogenase (7): nad1, nad2, nad3, nad4, nad4L, nad5, nad6</td>
<td>ATP synthase (3): atp6, atp8, atp9</td>
</tr>
<tr>
<td>Succinate:ubiquinone oxidoreductase (3): sdh2, sdh3, sdh4</td>
<td>Group II intronic ORFs of unknown function (1)</td>
</tr>
<tr>
<td>Ubiquinol:cytochrome c oxidoreductase (1): cob</td>
<td>ymf16b (homolog of E. coli mttB [taIC])</td>
</tr>
<tr>
<td>Cytochrome c oxidase (3): cox1, cox2, cox3</td>
<td>Conserved ORFs of unknown function (1)</td>
</tr>
<tr>
<td>ATP synthase (3): atp6, atp8, atp9</td>
<td>ymf39b</td>
</tr>
<tr>
<td>Group II intronic ORFs of unknown function (1)</td>
<td>ORFs unique to P. purpurea mtDNA (4)</td>
</tr>
<tr>
<td>ymf16b</td>
<td>orf132</td>
</tr>
<tr>
<td>Conserved ORFs of unknown function (1)</td>
<td>orf176</td>
</tr>
<tr>
<td>ymf39b</td>
<td>orf238</td>
</tr>
<tr>
<td>ORFs unique to P. purpurea mtDNA (4)</td>
<td>orf284</td>
</tr>
</tbody>
</table>

| a Genes are classified according to their function. Numbers within parentheses indicate the number of genes in a particular class. b ymf designations for ORFs follow the recommendations of the Commission on Plant Gene Nomenclature (1993). ymf39 has been shown to specify a membrane component of unknown function in angiosperms (Stamper et al., 1987; Prioli et al., 1993). Note that ymf19 (orf8) homologs have recently been identified as atp8 (Gray et al., 1998) and that ymf16 has been recognized as a homolog of E. coli taIC (or mttB), which codes for a protein involved in a Sec-independent protein translocation pathway (Bogsch et al., 1998; Weiner et al., 1998). |

The mitochondrial genome of P. purpurea encodes conventional, eubacteria-like LSU and SSU rRNAs. Their predicted sizes are 2546 and 1407 nucleotides, respectively, shorter than those of their E. coli counterparts (2904 and 1542 nucleotides, respectively). This length difference reflects the somewhat truncated variable regions found in the P. purpurea mitochondrial rRNAs compared with the E. coli ones; otherwise, the secondary structures of the mitochondrial and eubacterial homologs are essentially superimposable. (Secondary structures are available from the WWW rRNA database maintained by R.R. Gutell at the University of Texas; http://pundit.icmb.utexas.edu/RNA/. Direct links to these structures are also provided through GObase, the Organelle Genome Database [Korab-Laskowska et al., 1998; http://megasun.bch.umontreal.ca/gobase/].)

In the region between positions ~240 and 265, the mitochondrial LSU rRNA of P. purpurea lacks a recognizable secondary structure element analogous to that found between positions ~300 and 340 in the E. coli LSU rRNA and in a separate small 3S rRNA in some naturally fragmented chloroplast LSU rRNAs (Turmel et al., 1991). In addition, the P. purpurea mitochondrial LSU rRNA sequence contains a deletion in the region corresponding to E. coli 23S rRNA positions ~485 and 510. Both of these atypical features characterize animal and most fungal mitochondrial LSU rRNAs. Finally, a tertiary base pair interaction corresponding to G570:C866 in E. coli SSU rRNA is an unusual G•G pair in the case of P. purpurea.

The P. purpurea mitochondrial rRNA secondary structures are very similar to those encoded by the mtDNA of Chondrus crispus (Leblanc et al., 1995b), with the same variable regions closely approximating one another in size and most
of the proven or predicted tertiary interactions found in the homologous red algal rRNA sequences. Overall, however, the structures of the P. purpurea mitochondrial rRNAs appear more derived than those of C. crispus in that atypical features noted above (absence of several secondary structure elements and of tertiary base pair interactions) are not seen in the C. crispus rRNAs. Nevertheless, in phylogenetic analyses of mitochondrial SSU rRNA sequences, P. purpurea and C. crispus form a tight clade without significant differences in branch lengths. In such analyses, the relationship of the monophyletic red algal clade with other phyla is uncertain (D.F. Spencer and M.W. Gray, unpublished results).

Two group II introns are present in the P. purpurea mitochondrial rnl gene; both are inserted within the highly conserved peptidyl transferase center in the 3' half of the LSU rRNA between nucleotides corresponding to 2508 to 2509 and 2610 to 2611 of the E. coli sequence. Intron 1 is located at exactly the same position as intron 4 in the mitochondrial rnl of the brown alga Pyaiella littoralis (Fontaine et al., 1995). The potential RNA structures of these introns and characteristics of the ORFs they contain are described in a later section.

Mitochondrial 5S rRNA Genes in Red Algae

The distribution of the mitochondrial 5S rRNA gene (rrn5) is quite sporadic within various eukaryotic phyla. Originally, this gene was described only in the mtDNA of plants, some green algae, and jakobid flagellates. Subsequently, we detected rrn5 in the mtDNA of Chondrus crispus, situated between nad3 and rps11 (Gray et al., 1998), after the original claim—that a 5S rRNA gene is located between cox2 and cox3 (Leblanc et al., 1995a)—had been discounted (Lang et al., 1996). Moreover, we have determined that Cyanidium caldarium contains a mitochondrially encoded rrn5 in the intergenic region between rpl6 and trnM(cau); at this same genomic location, Ohta et al. (1998) identified a counterpart rrn5 in Cyanidioschyzon merolae mtDNA. However, despite an exhaustive search, we have failed to unearth such a gene in P. purpurea mtDNA.

Figures 2A to 2C depict the mitochondrial 5S rRNAs from Chondrus crispus, Cyanidium caldarium, and Cyanidioschyzon merolae, which share evident sequence identities and secondary structure similarities with 5S rRNA sequences of eubacteria and plant and protist mitochondria (Lang et al., 1996). However, there are two deviations from otherwise universally conserved features of the mitochondrial consensus structure. (1) The Cyanidium caldarium 5S rRNA has a U32:A58 pairing instead of the otherwise conserved A:U pair at the corresponding positions in helical domain III, and CCU (positions 40 to 42) instead of CCA in loop C (the latter feature shared with the Cyanidioschyzon merolae homolog). (2) Compared with their counterparts in other eukaryotes, including Chondrus crispus, helices II and V appear

---

**Figure 2.** Potential Secondary Structures of Red Algal Mitochondrial 5S rRNAs.

(A) Chondrus crispus.

(B) Cyanidium caldarium.

(C) Cyanidioschyzon merolae.

The nucleotide sequences were taken from GenBank accession numbers Z47547 (Leblanc et al., 1995a), Z48930 (Wiehmann, 1995), and D89861 (Ohta et al., 1998), respectively. Structures were modeled after that of wheat mitochondrial 5S rRNA (Spencer et al., 1981; De Wachter et al., 1982). Invariant residues in mitochondrial 5S rRNAs are circled, and broken lines enclose variable regions. Nonstandard pairings are highlighted by small filled (purine-pyrimidine), small open (purine-purine), and large filled (pyrimidine-pyrimidine) circles. Helices (I to V) and loops (A to E) are denoted as in Moore (1995).
thermodynamically much less stable in the mitochondrial SS rRNA secondary structures of Cyanidium caldarium and Cyanidoschyzon merolae.

**Mitochondrial Transfer RNA Genes and Codon Usage in P. purpurea mtDNA**

P. purpurea mtDNA encodes 24 tRNA genes that are scattered over the entire genome, either singly or in groups of two or three. All of these tRNA sequences can assume standard cloverleaf secondary structures, with very few departures from the conventional structure. Most notable is an extra U between positions 12 and 13 in the D stem of tRNA<sup>Asn(GUU)</sup>. Additional atypical features (using the standard tRNA numbering system) include the following: G48 (last nucleotide of the variable loop) versus Y in tRNA<sup>Cys(GCA)</sup>; C9 (between acceptor and D stems) versus R in tRNA<sup>Glu(UUC)</sup>; G8 (also between acceptor and D stems) versus Y, and G14 (in D loop) versus A in tRNA<sup>His(GUG)</sup>; and A32 (first nucleotide of anticodon loop) versus Y in elongator tRNA<sup>Met(CAU)</sup>. The latter feature is found frequently in the mitochondrial elongator tRNA<sup>Met(CAU)</sup> of protists (M.W. Gray, unpublished observation), including that of Chondrus crispus and Cyanidoschyzon merolae.

The single polymorphic site (see below) affecting a tRNA gene falls in the anticodon arm of tRNA<sup>Ala(UGC)</sup>. This polymorphism (C«T) would have a minimal effect on the overall secondary structure stability because both C42 and U42 would pair with G28. All of the unusual features of the mitochondrial tRNAs in P. purpurea are also found in the counterpart genes of C. crispus, with the exception of the extra U nucleotide in the D stem of tRNA<sup>Asn(GUU)</sup>, which occurs in an otherwise strongly conserved stretch of sequence.

Table 2 shows that the 24 P. purpurea mitochondrial tRNAs are sufficient to decode 57 of 62 sense codons that occur in mtDNA-encoded protein genes of this organism (TAA/TAG being used as stop codons). tRNA genes not found in this genome are those recognizing isoleucine (ATA) and threonine (ACN) codons; in these cases, tRNA import from the cytosol presumably makes up the deficit. Interestingly, a tmt gene is also absent from the mtDNAs of golden algae, stramenopiles, a jakobid flagellate, and land plants (see Gray et al., 1998), suggesting a number of independent losses of this particular gene from mtDNA in the course of evolution.

P. purpurea mtDNA encodes two glycine tRNAs, tRNA<sup>Gly(GCC)</sup> and tRNA<sup>Gly(UCC)</sup>, which would be functionally redundant if the latter species were able to recognize all four glycine codons. However, in E. coli, a C to U substitution at...
position 32 (first position in anticodon loop) restricts the decoding capacity of a tRNA^Gly(UCC) to the codons GGA and GGG (Lustig et al., 1993). To recognize the GGY codons, the second tRNA^Gly with a GCC anticodon is required. In fact, P. purpurea mitochondrial tRNA^Gly(UCC) does have a U rather than a C nucleotide at position 32. The two trmG genes show significant sequence similarity (64% identical residues), suggesting that they arose via a recent gene duplication and divergence from a common, ancestral trnG gene. Also, tRNA^Ser(GCU) and tRNA^Ser(UGA) have presumably arisen by gene duplication because they are 71% identical in sequence.

The tRNA gene set encoded in the P. purpurea and Chondrus crispus mtDNAs is identical, except that the latter lacks trnS(gcc). P. purpurea and Cyanidioschyzon merolae also have the same mitochondrial tRNA gene set, except that the latter encodes an additional trnL(caa) and its trnW has a CCA rather than a UCA anticodon.

Table 2 also lists the codon distribution in P. purpurea protein-coding genes, which, as in most mtDNAs, is biased toward codons ending in an A or T. No significant difference in codon usage was observed in identified protein-coding genes, conserved ORFs, unique ORFs, and intronic ORFs. Moreover, there is a GTG codon in the P. purpurea cox1 gene at exactly the same position as an ATG initiator codon in the corresponding Chondrus crispus gene, strongly suggesting that GTG serves as a start codon in mitochondrial translation in P. purpurea; use of GTG initiation codons has been inferred previously in several plant and protist mitochondrial as well as bacterial systems (e.g., Netzker et al., 1982; Kozak, 1983; Lang et al., 1997). Finally, TGA is translated in P. purpurea mitochondria as tryptophan (inferred from the presence of a tRNA with anticodon UCA and multiple protein sequence alignments). This is the most common deviation from the standard translation code in mitochondria and is also found in C. crispus mtDNA; in contrast, mitochondria of Cyanidium caldarium and Cyanidioschyzon merolae use the universal genetic code.

Unassigned Reading Frames and Pseudogenes in P. purpurea mtDNA

Four unique ORFs (orf132, orf176, orf238, and orf284) are present in the mtDNA of P. purpurea. Whereas the deduced Orf132 and Orf176 proteins exhibit an intermediate degree of polarity, the amino acid composition of Orf238 is characteristic of hydrophobic membrane components (hydrophobicity 59.5; Kyte and Doolittle, 1982). Orf284, in contrast, contains a high proportion of charged and polar residues (hydrophobicity −10.6), comparable with that of ribosomal proteins.

As mentioned earlier, the coding regions of rtl and dpo are discontinuous, that is, broken up into two and four adjacent gene fragments, respectively, whose order and orientation are maintained. Fragmentation of both genes is due to inter-frame TAA-termination codons or frameshifts (apparently generated by single-nucleotide deletions). As discussed below, rtl seems to be a vestige of a group II intron ORF, and therefore we assume that it is not expressed, although its codon usage does not differ significantly from that of conserved mitochondrial genes.

A partial dpo sequence has been obtained from another, uncharacterized Porphyra sp (GenBank accession number X65264). Figure 3 shows that the sequence translates, in contiguous fashion, into that portion of the dpo gene that corresponds to P. purpurea fragments 1 and 2, plus the 5′-terminal part of fragment 3. Unexpectedly, the Dpo proteins of the two Porphyra spp share only 39.5% identity. Such a high fluidity in gene sequence and structure within the same genus indicates that dpo is evolving extremely rapidly and is probably a pseudogene. Split and scrambled dpo sequences also occur in the liverwort M. polymorpha, whose mitochondrial genome contains three ORFs that display significant similarity to the N-terminal, middle, and C-terminal portion of P. purpurea dpo (Figure 3). In contrast, a continuous dpo is found in maize mitochondria, where it is encoded by a plasmid (Paillard et al., 1985), and in the mitochondria of the golden-brown alga Ochromonas danica (G. Burger, A. Coleman, and B.F. Lang, unpublished results), where it is contained in the mtDNA itself. It has been suggested that dpo was not originally contained in the proto-mitochondrial genome but was introduced via mobile plasmids at an advanced point in the evolutionary history of mitochondria (Weber et al., 1995).

Introns and Free-Standing Intron Fragments in P. purpurea mtDNA

Two group II introns, 2483 and 2478 nucleotides long, interrupt the LSU rRNA coding region in P. purpurea mtDNA. According to their RNA secondary structures (Figure 4), these introns belong to subgroup B1 (Michel et al., 1989). Both introns contain ORFs, inserted within domain IV, that code for proteins 544 and 546 residues long. These ORFs possess all of the conserved motifs characteristic of reverse transcriptases of non-LTR retrotransposons (Michel and Lang, 1985; Xiong and Eckbush, 1990) as well as the particular protein domains (reverse transcriptase, Z, zinc finger, and the C-terminal HNH motif) that are believed to be involved in endonucleolytic cleavage and intron homing (Mohr et al., 1993; Shub et al., 1994).

Figures 4A and 4B show that introns 1 and 2 in P. purpurea mtDNA resemble one another to a high degree. The intron RNA secondary structures are virtually identical, the deduced protein sequences of the two intron ORFs (Orf544 and Orf546) share 70% identical residues and 20% conserved changes, and the nucleotide sequence outside the reading frames can be aligned readily (68% identity). Such a pronounced resemblance suggests that intron transposition has occurred in cis into the same gene (Ferat et al., 1994).
Indeed, the sequences surrounding the two intron insertion sites in rnl share nearly 50% identical residues (nine out of 20 nucleotides).

Database searches with the Orf544 and Orf546 proteins reveal striking similarity to intronic reverse transcriptase ORFs of the cyanobacterium Calothrix sp (Ferat and Michel, 1993; BLAST score 344, probability 3.0 \times 10^{-171}). The second highest score was found with mitochondrial intron ORF proteins from the brown alga Pyliaea littoralis; these algal ORFs had earlier been reported to share common features with cyanobacterial counterparts (Fontaine et al., 1995). Overall protein similarity is considerably lower between the Pyliaea littoralis and Calothrix sp (32%) or Pyliaea littoralis and P. purpurea (23%) ORFs than between the P. purpurea and Calothrix sp homologs (44%).

In addition, the introns from P. purpurea mtDNA and Calothrix sp in which these ORFs reside are extraordinarily alike. Figure 4B shows that the few obvious differences in their potential RNA secondary structures are three additional stem–loop elements inserted in domain I of the P. purpurea introns; otherwise, there are only minor variations in helix length and loop size in the six domains (for domain assignments, see Michel et al., 1989). Finally, the P. purpurea and Calothrix sp introns share significant similarities at the nucleotide sequence level, primarily in domain V and in the basal regions of domains I and III, which are generally poorly conserved. Such a striking degree of structural resemblance is not seen between the ml introns from P. purpurea and Pyliaea littoralis or between the Calothrix sp and Pyliaea littoralis introns.

In addition to the two rnl introns described above, we detected relics of a group II intron in P. purpurea mtDNA. A typical domain V consensus structure is located downstream of and partially overlaps with the C-terminal region of rtl; however, a complete intron core structure could not be identified. Presumably, the rtl gene was initially contained in a group II intron, and fragmentation of its reading frame occurred concomitantly with the loss of its conserved RNA secondary structure. Because the intron no longer resided in a vital gene, there would have been no selective pressure to conserve the RNA secondary structure that is required for proper splicing.

Phylogenetic Analysis of Group II Intron ORFs

To investigate the evolutionary distances of the P. purpurea intron ORFs relative to other reverse transcriptase proteins,
Figure 4. Similarities between Group II Introns from P. purpurea Mitochondria and Calothrix sp.

(A) Intron 1 in ml of P. purpurea mtDNA. Nucleotide (nt) positions identical with those in intron 2 of the same genome are shown in blue. Residues identical in the two P. purpurea introns and in the group II intron of the cyanobacterium Calothrix sp (Ferat and Michel, 1993; GenBank accession number X71404) are shown in red.

(B) Intron 2 in ml of P. purpurea mtDNA. Secondary structure elements not present in the Calothrix sp intron are marked by green shading. Tertiary base pairings are indicated by EBS1:IBS1, EBS2:IBS2, and αα′ to γγ′; the involved nucleotides are boxed or marked by arrows (Costa et al., 1997).
phylogenetic analyses were performed on a collection of readily alignable reverse transcriptase proteins of mitochondrial, chloroplast, and bacterial origin. Bootstrap support (Felsenstein, 1985) and likelihood estimations (Kishino et al., 1990) were high with all tree construction methods applied. Figure 5 depicts the resulting phylogenetic reconstruction, which reveals two large clusters of reverse transcriptase proteins, one including the classic mitochondrial intron ORFs of plants and fungi that are all inserted in group IIA introns, the other cluster including members from protist mitochondria, chloroplasts, and bacteria, which reside in group IIB introns. A similar bipartite tree structure of group IIA and IIB intron ORFs has been reported previously (Fontaine et al., 1995). The most closely related reverse transcriptase proteins are those from P. purpurea mitochondria and Calothrix sp., reinforcing the view of a very recent lateral intron transfer between cyanobacteria and P. purpurea mitochondria.

**Inverted Repeats in P. purpurea mtDNA**

In P. purpurea mtDNA, a sequence stretch of 291 bp occurs twice, with the two copies located roughly opposite one another on the circular mtDNA map and in inverted orientation (Figure 1, magenta arrows). These repeated elements, designated INV1 and INV2, extend into the 5’ portion of orf238 and orf132, whose protein sequences are consequently identical between amino acid positions 1 and 94. In angiosperm mitochondria, inverted (and also direct) repeats appear to promote major genome rearrangements (Hanson and Folkerts, 1992). Hence, we in-

![Figure 5. Phylogenetic Relationships among Group IIB Intron ORF Proteins from Mitochondria and Bacteria.](image_url)

A combination of PROTDIST and FITCH with a variation coefficient of 0.3 was used, as described in Methods. The same tree topology was obtained when maximum likelihood approaches (PROTML and PUZZLE) were used. Bootstrap support, calculated from 1000 replicates, is shown at each branch (in percentages; rounded numbers). Ovals highlight short and poorly supported branches (bootstrap support <80%). mt, cp, and bc indicate mitochondrial, chloroplast, and bacterial sequences, respectively. Designations (GenBank accession numbers within parentheses) are as follows: Calothrix, orf584 bc, “Orf2,” contained in an intron of an unknown gene of Calothrix sp, see legend to Figure 4; Clostridium, orf609 bc, intronic ORF of C. difficile transposon (X98606; Mullant et al., 1996); E. coli, orf416 bc, intronic ORF in E. coli DNA (S50828 [protein]; Ferat et al., 1994); Marchantia, cob-3 mt, ORF in cob intron 3; Marchantia, cox1-2 mt, ORF in cox1 intron 2 of M. polymorpha mtDNA (M68929; Oda et al., 1992); Porphyra, m1-1 mt, Orf544 in m1 intron 1; Porphyra, m1-2 mt, Orf546 in m1 intron 2 of P. purpurea mtDNA (this work); Pylaiella, m1-1 mt, ORF in m1 intron 1; Pylaiella, m1-2 mt, ORF in m1 intron 2 of P. littoralis mtDNA (S58503, S58504 [protein]; Fontaine et al., 1995); Saccharomyces, cox1-1 mt, ORF in cox1 intron 1; Saccharomyces, cox1-2 mt, ORF in cox1 intron 2 of S. cerevisiae mtDNA (V00694; Bonitz et al., 1980); Scenedesmus, petD-1 cp, ORF in petD intron 1 of S. obliquus chloroplast DNA (P19593 [protein]; Kück, 1989); Schizosaccharomyces, cob-1 mt, ORF in cob intron 1 of S. pombe mtDNA (X02819; Lang et al., 1985); and Streptococcus, orf425 bc, maturase-related protein in S. pneumoniae DNA (AF030367; Coffey et al., 1998).
investigated whether this is also the case in P. purpurea mtDNA. Polymerase chain reaction (PCR) amplification was performed with primers that anneal in the single-copy regions, 90 bp upstream of INV1 and 95 bp downstream of INV2, respectively (see Figure 1, genome map). If, in P. purpurea mitochondria, molecules are present that had undergone intramolecular recombination between the inverted repeats, a PCR product of 476 bp should be found, whereas the experimental conditions would not amplify the 14,653-bp fragment that corresponds to the configuration (shown in Figure 1) inferred from sequencing the random clone library. In fact, we obtained a PCR product of ~480 bp and confirmed its identity by DNA sequencing. Although potential PCR artifacts cannot be excluded readily, we infer that the 476-bp fragment reflects the presence of a low proportion of mtDNA molecules in an alternative genomic conformation that results from intramolecular recombination involving the inverted repeats.

**Sequence Polymorphisms in P. purpurea mtDNA**

In P. purpurea mtDNA, we observed 64 polymorphisms of three different types (Table 3). (1) Single-nucleotide substitutions in coding as well as intergenic regions are the most abundant type. These substitutions involve only two of the four possible nucleotides and are predominantly transitions. (2) Deletions/insertions (indels) of from 1 to 43 bp are located exclusively in intergenic regions. (3) A macro substitution of 53 or 494 bp is found after nucleotide 11,409 (Figure 1). The longer substitution, representing the version of the longer substitution, represents the version of the shorter version.

Intergenic regions contain approximately twice as many polymorphic sites per kilobase than coding regions. Among coding regions, rtl, dpo, and ORFs (except orf284, see below) harbor approximately sixfold more polymorphic sites per kilobase and exhibit an approximately fourfold higher ratio of nonsilent to silent single-nucleotide substitutions compared with conserved genes. ORFs and fragmented genes in P. purpurea mtDNA are apparently subject to minimal selective pressure and therefore can evolve at a faster pace than do genes involved in vital functions such as electron transfer, ATP synthesis, and translation. Notably, orf284 contains significantly fewer polymorphic sites (1.1 per kilobase) than do other unique ORFs (Table 3). Considering that the encoded protein has a high number of charged and polar residues, orf284 might code for a poorly conserved, and therefore unrecognizable, ribosomal protein.

We further investigated whether the states of adjacent polymorphic sites correlate at the level of individual clones. In most adjacent pairs examined (43 of 49), we detected covariance, suggesting that the mtDNA used to construct the clone library contained two distinct types of molecules. Remarkably, most noncovariant adjacent pairs of sites are localized to the inverted repeat copies. This finding confirms the conclusion of the previous section that these repeats are involved in intramolecular recombination.

The polymorphic sites within the inverted repeats of P. purpurea mtDNA account for a 1.4% sequence variation within these repeats. This high variability stands in marked contrast to the situation in the chloroplast genome of many green algae and land plants, in which the two inverted repeat copies encompassing rRNA genes; however, unlike in green algal and plant chloroplast DNAs, these repeats are arranged in direct orientation, differ from one another in sequence, and do not appear to support intramolecular recombination (Reith and Munholland, 1993b).

Table 3. Sequence Polymorphisms in P. purpurea mtDNA

<table>
<thead>
<tr>
<th>Gene/Regiona</th>
<th>Substitution</th>
<th>Single ntb</th>
<th>Total</th>
<th>Indel</th>
<th>Macro</th>
<th>Total Number</th>
<th>Number per kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intergenic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regions (55)</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>11</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intron cores (2)</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNAs (26)</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>proteins (23)</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rare proteins (2)</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>1*</td>
<td>9*</td>
<td></td>
<td>3.6</td>
</tr>
<tr>
<td>Intronic ORFs (2)</td>
<td>14</td>
<td>10</td>
<td>0</td>
<td>14</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique ORFs (4)</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>10</td>
<td>4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total mtDNA</td>
<td>59</td>
<td>4</td>
<td>1</td>
<td>64</td>
<td>1.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Classes of genes and genome regions in P. purpurea mtDNA in which polymorphic sites occur. Numbers within parentheses denote the total number of genes or regions within a class.
b nt, nucleotide.
c A dash means not applicable.
d Asterisks indicate that a single macro substitution affects both dpo and rtl.
approximately half of the mitochondrial chromosome relative to the other half. These findings raise questions about the homogeneity of the DNA material that was used for constructing the clone library. For mtDNA extraction, gametophytes (folious thalli) of *P. purpurea* were collected in the wild, offshore Nova Scotia, Canada, and propagated in the laboratory (Reith and Munholland, 1993a). The observed variance in the sequenced DNA material suggests a considerable heterogeneity of the mitochondrial genome within cross-fertilizing populations of *P. purpurea*.

**Global Phylogenetic Analysis Using Mitochondrial Protein-Coding Genes**

We have analyzed the phylogenetic relationship within rhodophytes and between rhodophytes and chlorophytes by including new mitochondrial sequence data from the red algae *P. purpurea* (reported here), *Cyanidioschyzon merolae* (Ohta et al., 1998), and *Gracilaria lemaneiformis* (Lang and Goff, 1999) as well as of the green alga *Tetraselmis* maculata (this report). For the analysis, we used the concatenated, deduced protein sequences of four genes, cox1, cox2, cox3, and cob, from 33 different taxa. Both distance and likelihood algorithms were applied, with the analyses generating identical topologies and similar bootstrap support and likelihood estimations.

Figure 6 shows within the red algal clade a deep divergence between *Cyanidium caldarium* and *Cyanidioschyzon merolae*, on the one hand, and *Porphyridiales*, *Bangiales*, and *Gracilariaceae* on the other. *P. purpurea* (representing *Bangiales*) emerges after the divergence of the “Cyanidium complex” but before *Porphyridiales* and *Gracilariaceae*. Notably, the *Cyanidium caldarium* and *Cyanidioschyzon merolae* branch lengths are approximately two times as long as those of the other three rhodophyte taxa, reflecting an accelerated rate of evolution at the sequence level in these acidothermophilic species.

The tree as a whole is characterized by coherent clades of rhodophytes, chlorophytes/plants, animals, and fungi; the overall branching pattern implies that green and red algae are sister taxa that evolved from a common ancestor to the exclusion of animals and fungi. The distance between the bacterial and mitochondrial clades is conspicuously large, whereas within the mitochondrial clade, the intervals between the major nodes are small, implying a nearly simultaneous, “explosive” radiation of all eukaryotic taxa.

To resolve further the branching order of green and red algae relative to that of the common ancestor of animals and fungi, we included in the analysis additional protists, such as *R. americana*, *Malawimonas jakobiformis*, and *J. akohia libera* (jakobid flagellates; Lang et al., 1997; G. Burger, B.F. Lang, C. O’Kelly, and M.W. Gray, unpublished results), *Acanthamoeba castellanii* (rhizopod amoeba; Burger et al., 1995), *Phytophthora infestans* (oomycete; B.F. Lang and L. Forget, unpublished results), and *O. danica* (chrysophyte alga; G. Burger, B.F. Lang, and A. Coleman, unpublished results) (Figure 6). Stramenopiles (*O. danica* and *P. infestans*) on the one hand and jakobids on the other form strongly supported clades that, together with *A. castellanii*, do not specifically group with either the green or red algae, or with animals and fungi. However, stramenopiles, jakobids, and rhizopods cannot be positioned on the tree at a significant confidence level. We attribute this topological instability particularly to the absence of a suitable outgroup (reflected by the large distance between bacterial and mitochondrial proteins), as well as to a lack of data from early diverging eukaryotic taxa.

**DISCUSSION**

**Very Recent Intron Transfer between the Genomes of *P. purpurea* Mitochondria and a Cyanobacterium**

ORFs in group II introns fall into two distinct subclasses of reverse transcriptases. Those contained in group IIB introns are small and share distinctive protein signatures that set them apart from those residing in group IIA introns (Ferat et al., 1994; Fontaine et al., 1995). A large majority of the 100 or so group II introns known to date belong to subgroup IIA and are found in mitochondrial and chloroplast genomes. The two mitochondrial introns from *P. purpurea* belong to subgroup B, as do the only three currently known bacterial group II introns.

Our phylogenetic analysis of reverse transcriptase ORFs (Figure 4) as well as primary sequence and RNA secondary structure comparisons show that the *P. purpurea* mitochondrial introns are very closely related to counterparts from the cyanobacterium *Calothrix* sp, a finding of relevance to the long-standing question about intron origins and propagation. Several ORF-containing group I and II introns have been shown to be mobile elements in the sense that in genetic crossovers, they invade intronless copies of the genes in which they are inserted (Dujon et al., 1986; Skelly et al., 1991). This has led to the inference in a number of cases that introns are also able to spread laterally between unrelated species, and between genomes from different cellular compartments (e.g., Lang et al., 1985; Lonergan and Gray, 1994; Turmel et al., 1995; Vaughn et al., 1995; Cho et al., 1998). In the case of *P. purpurea*, cyanobacterial introns might have invaded mitochondria via chloroplasts that inherited their introns directly from a cyanobacterial ancestor. However, the chloroplast DNA of *P. purpurea*, which has been sequenced completely (Reith and Munholland, 1995), does not contain any introns or reverse transcriptase ORFs, making the hypothesis of intron transfer via chloroplast DNA less likely. A direct, recent intron transmission between a cyanobacterium and a red algal mitochondrion remains the most plausible scenario to account for the above observations.
Figure 6. Phylogeny of Red and Green Algal Mitochondrial Genomes.

The phylogenetic tree was inferred from an alignment of the concatenated protein sequences of Cob, Cox1, Cox2, and Cox3. A combination of PROTDIST and FITCH was used, as described in Methods, with a variation coefficient of 0.5. The same tree topology was obtained when maximum likelihood approaches (PROTML and PUZZLE) were used, except for branches with <80% bootstrap support. Bootstrap support (in percentages), calculated from 1000 replicates, is shown for all branches. Short and poorly supported branches (<80% bootstrap support) are indicated by thin lines. Animals, fungi, rhizopods, red algae, green algae/plants, jakobids, stramenopiles, and bacteria are indicated and grouped together by square brackets. Ciliates, chlamydomonas, kinetoplastids, apicomplexans, and slime molds have been excluded from the analysis because these taxa cannot be placed reliably in the tree due to the accelerated evolutionary rate of their mitochondrial genes. Species (with GenBank accession numbers in parentheses) are as follows: Homo, H. sapiens (J01415; Anderson et al., 1981); Mus, M. musculus (J01420; Bibb et al., 1981); Xenopus, X. laevis (M10217; Roe et al., 1985); Strongylodiscus, S. purpuratus (X12631; Jacobs et al., 1988); Drosophila, D. yakuba (fruitfly; X03240; Clary and Wolstenholme, 1985); Metridium, M. senile (AF000023; Beagley et al., 1998); Rhizopus, R. stolonifer (chytridomycete fungus; Laforest et al., 1997; Saccharomyces, S. cerevisiae (ascomycete fungus; X84842, V00694, V00695, J01478; de Zamaroczy and Bernardi, 1986); Pichia, P. canadensis (Hansenula wingei) (ascomycete fungus; D31785; Sekito et al., 1995); Aspergillus, A. (Emericella) nidulans (ascomycete fungus; J01387, X15441, X06960; Netzker et al., 1982; Brown, 1993); Neurospora, N. crassa (ascomycete fungus; K01181, A28755 [protein], X01850, K00825, V00668; Collins, 1993); Podospora, P. anserina (ascomycete fungus; X55026; Cummings et al., 1990); Alomyces, A. macrognus (chytridomycete fungus; U41288; Paquin and Lang, 1996; Paquin et al., 1997); Acanthamoeba, A. castellanii (rhizopod; U12386; Burger et al., 1995); Malawimonas, M. jakobiformis (jakobid); G. Burger, B.F. Lang, C. O’Kelly, and M.W. Gray, unpublished data); Cyanidioschyzon, C. merolae; Cyanidium, C. caldarium; Gracilariosis, G. lemaneiformis, Cox1, Cox2, Cox3 (rhodophyte; AF118119); Condrus, C. crispus (rhodophytes; see legend to Figure 2); Porphyra, P. purpura (rhodophyte; this report); Triticum, T. aestivum (P07747 [protein], Y00417, X01108, P15953 [protein]; Bonen et al., 1984, 1987; Boer et al., 1985; Gualberto et al., 1990); Marchantia, M. polymorpha (liverwort; see legend to Figure 5); Prototheca, P. wickerhamii (green alga; U02970; Wolff et al., 1994); Tetraselmis, T. maculata, Cob (300 amino acids), Cox1 (266 amino acids), Cox2 (complete) (green alga; this report); Tetraselmis s., T. (Platymonas) subcordiformis (green alga; 247797; Kessler and Zetsche, 1995); Reclinomonas, R. americana (jakobid; AF007261; Lang et al., 1997); Jakoba, J. libera; Phytophthora, P. infestans (oomycete; B.F. Lang and L. Forget, unpublished data); Ochromonas, O. danica (chytridomyx; G. Burger, A. Coleman, and B.F. Lang, unpublished data); Rickettsia, R. prowazekii (α-Proteobacterium; A1235270 to A1235273; Andersson et al., 1998); Bradyrhizobium, B. japonicum (α-Proteobacterium; J03176, X68547; Thony-Meyer et al., 1989; Bott et al., 1992); Rhodobacter, R. sphaeroides (α-Proteobacterium; X56157, X62645, M57680, C45164 [protein]; Yun et al., 1990; Cao et al., 1992; Shapleigh and Gennis, 1992); and Paracoccus, P. denitrificans (α-Proteobacterium; X05829, M17522, X05934, X05828; Kurowski and Ludwig, 1987; Raibo et al., 1987). Unpublished protein sequences are available at http://megasun. bch.umontreal.ca/People/lang/FMGP/proteins.html.
Rhodophytes Emerged Simultaneously with Chlorophytes

For the first time, mitochondrial sequence information is available from five rhodophyte orders, including Bangiales (P. purpurea, complete sequence; described here), Gigartiniales (Chondrus crispus, complete sequence; Leblanc et al., 1995a), Porphyridiales (Cyanidium caldarium, 10-kb fragment; Viehmann, 1995), Gracilariiales (Gracilariales lemaneiformis, 3-kb fragment; Lang and Goff, 1999; GenBank accession number AF118119), and Cyanidioschyzon (C. merolae, complete sequence; Ohta et al., 1998). In addition, we have information about mtDNA size and shape in Gracilariales pulvinata (Goff and Coleman, 1995). Taken together, these data should fully encompass mtDNA diversity within the red algae. It should be noted that some taxonomic hierarchies include the genus Cyanidioschyzon within the Cyanidiales, which comprises several small, unicellular, acidothermophilic red algae (Ott and Seckbach, 1994). In contrast, the National Center for Biotechnology Information taxon database consortium recognizes Cyanidioschyzon as a separate order comprising a single genus and species.

In nuclear and chloroplast phylogenies, the genus Cyanidium forms the deepest branch of the red algal clade (Ragan et al., 1994; Saunders and Kraft, 1996; Palmer and Delwiche, 1998), although Porphyridiales and Gracilariiales are more recently diverging clades (reviewed in Saunders and Kraft, 1997). The red algal topology in the tree that we infer from mitochondrial encoded proteins (Figure 6) is in complete agreement with these nuclear and chloroplast phylogenies.

 Whereas the phylogenetic relationships within the red algae are well established, the association of rhodophytes with other eukaryotic phyla is still highly controversial. In nuclear SSU rRNA trees, the majority of eukaryotic taxa, including rhodophytes, cluster tightly together in a nearly simultaneous radiation (termed the crown group), basal to which emerge slime molds and a number of other groups (Sogin, 1997). On the other hand, analyses of β-tubulin proteins place Chondrus crispus together with slime molds (Liaud et al., 1995). Finally, in a tree based on the large subunit of RNA polymerase II (RNAPII; Stiller and Hall, 1997), green algae group together with fungi and animals, to the exclusion of red algae. Although bootstrap support is high in this study, the green algal/fungal/animal alliance is likely an artifact due to the pronounced divergence of the red algal RNAPII sequences compared with those of green algae; also, insufficient taxonomic sampling (only one red and two green algal species) is a concern. As reviewed in detail by Palmer and Delwiche (1998), nuclear phylogenies that include representatives of red algae together with members of the major eukaryotic kingdoms are largely unresolved, often suffering from long-branch attraction, weak support, or poor overall sampling (e.g., Zhou and Ragan, 1995; Liu et al., 1996; Keeling and Doolittle, 1997).

In contrast to nuclear trees, several recent phylogenetic analyses of mitochondrial data point to a sister-group relationship between red and green algae; however, these analyses also lack either rigorous statistical or bootstrap support (e.g., Bhattacharya and Schmidt, 1997) or include only a limited number of eukaryotic taxa (e.g., Boyen et al., 1994) or of rhodophyte taxa in particular (e.g., Paquin et al., 1997; Lang et al., 1998). To our knowledge, the mitochondrial phylogeny presented here (Figure 6) is the only analysis to date combining high bootstrap support with broad sampling of mitochondrial eukaryotes, showing robust coherence of members of the red algae, green algae/plants, animals, and fungi, and providing strong evidence for the divergence of red and green algae from a shared ancestor. The sister-group relationship of red and green algae is congruent with results obtained from the analysis of chloroplast genes in which green, red, and glaucocystophyte algae cluster together (reviewed in Palmer and Delwiche, 1998). This congruence supports the idea that the origin of plastids is monophyletic and that the primary plastid-containing organisms diverged from a single common ancestor.

Rhodophytes Exhibit a Derived Pattern of mtDNA Organization and Complexity

The mtDNAs of P. purpurea, Chondrus crispus, Cyanidioschyzon merolae, Cyanidium caldarium, and Gracilariales spp are quite similar in size (25 to 36.8 kb), and the four mtDNAs of the first four species are similar in gene complement and gene order as well. Apart from ribosomal protein genes, six more of which are present in Cyanidioschyzon merolae (and probably also Cyanidium caldarium) mtDNA than in P. purpurea mtDNA, the mitochondrial gene repertoire in rhodophytes is relatively small. Having a total of ~60 genes and encoding essentially the animal set of subunits of respiratory complex I, III, IV, and V, rhodophyte mtDNAs have apparently lost a substantial number of genes to the nucleus. These features set rhodophyte mtDNAs apart from mitochondrial genomes present in certain green algae, cryptophyte algae, and many nonphotosynthetic protists, which all display a more ancestral pattern of mtDNA organization.

In fact, rhodophyte mitochondrial genomes are reminiscent of animal (particularly coelenterate) mtDNAs, both with respect to their conservation of gene order and constancy of genome size (reviewed in Wolstenholme and Fauron, 1995). In stark contrast to rhodophytes and animals, the chlorophyte/plant lineage displays extraordinary variability in mtDNA size, gene content, and overall organization. The mtDNAs of Prototheca wickerhamii, Nephroselmis olivacea, and M. polymorpha, which are of the ancestral type, contain numerous extra genes compared with animal and fungal mtDNAs. In contrast, derived mitochondrial genomes such as those of Chlamydomonas spp and Pedinomonas minor
have been subject to considerable gene loss, exceeding even that seen in animal mtDNAs, as well as fragmentation and scrambling of rRNA genes (reviewed in Gray et al., 1998).

As inferred from the gene complement in rhodophyte and chlorophyte mtDNAs, the common ancestor of both groups must have possessed a mitochondrial genome set that approximates that of the heterotrophic flagellate R. americana (94 genes; Lang et al., 1997). Approximately half of this set must have been lost in the red algal lineage soon after its divergence from the common progenitor shared with chlorophytes, whereas subsequently, only minor evolutionary changes seem to have occurred.

Are Rhodophytes a Highly Uniform or Highly Diverse Assemblage?

The remarkable similarities evident among red algal mtDNAs, particularly at the level of gene order and mtDNA size, as well as the solid coherence and moderate branch lengths of rhodophyte taxa in the phylogenetic tree shown in Figure 6, exemplify the uniformity of the mitochondrial genome in this phylum. These findings are in stark contrast to the view based on nuclear gene phylogenies, which characterize rhodophytes as an exceptionally diverse phylum (Ragan et al., 1994). Although nuclear and mitochondrial trees are typically congruent in topology, they can differ considerably in phylogenetic distances of particular taxa. The most extreme examples involve the Chlamydomonas-like green algae. Chlamydomonads are strongly affiliated with other green algae and plants in nuclear rRNA phylogenies, whereas in mitochondrial rRNA and protein trees, their branches lengthen considerably, generating unstable and artifactual topologies (reviewed in Gray and Spencer, 1996). In red algae, the opposite is true: red algae exhibit exceptionally high diversity in nuclear SSU rRNA phylogenies but are characterized by compact clustering and moderate distances between one another in mitochondrial protein phylogenies. It appears that the nuclear SSU rRNA gene in rhodophytes has undergone accelerated evolutionary change. Much more nuclear sequence data will be required, including data from nuclear protein-coding genes, to evaluate how accurately the nuclear SSU rRNA sequence reflects the evolutionary pace of the red algal nuclear genome.

**METHODS**

**Isolation, Purification, and Cloning of Mitochondrial DNA**

Porphyra purpurea was collected near Avonport, Nova Scotia, Canada, and then grown in the laboratory in the form of gametophyte cultures (Reith and Munholland, 1993b). An AT-rich fraction of P. purpurea DNA, consisting of a mixture of mitochondrial and chloroplast DNAs, was kindly provided by M. Reith.

Mitochondrial DNA (mtDNA) was isolated by cutting the AT-rich DNA fraction with the restriction enzyme SacI and resolving the resulting fragments by agarose gel electrophoresis. The largest (35 kb) restriction fragment proved to be the linearized mtDNA, whereas the chloroplast DNA was cut into several fragments of ~15 kb (Reith and Munholland, 1993b). The mtDNA was recovered by electroelution and physically sheared by nebulization (Okpodo et al., 1994), and a fraction of 500 to 3000 bp was recovered after agarose gel electrophoresis. The DNA was incubated with a mixture of T7 DNA polymerase and Escherichia coli DNA polymerase I (the Klenow fragment) to generate blunt ends and then cloned into the SmaI site of the phagemid pBluescript II KS+ (Stratagene). Recombinant plasmids containing mtDNA inserts were identified by colony hybridization, using the 35-kb SacI restriction fragment as a probe. Clones contained in this random library encompassed the entire P. purpurea mitochondrial genome. A region of ~500 bp bridging the ScaI site was amplified by the polymerase chain reaction (PCR), cloned, and sequenced, confirming the presence of a single such site in the mtDNA.

Tetraselmis maculata CCMP897 was obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME) and cultured in seawater. An AT-rich DNA fraction, consisting of chloroplast and mitochondrial DNA, was isolated from total cellular DNA by CsCl/Hoechst-33258 dye (Aldrich, Milwaukee, Wl) equilibrium centrifugation. Shearing of DNA and subsequent steps were conducted as described above.

**DNA Sequencing and Data Analysis**

DNA sequencing was performed by the dideoxy chain termination method (Sanger et al., 1977), using single-stranded DNA as template and 35S-dATP as label. Polyacylamide gels, dried onto the glass plate (Lang and Burger, 1990), were autoradiographed, and sequences were entered manually into computer files. In addition, automated sequencing was performed on a Li-Cor (Lincoln, NE) 4000L apparatus, using end-labeled primer and a cycle-sequencing protocol (Amersham).

Data processing and analysis were performed on SUN workstations (Sun Microsystems, Palo Alto, CA). Sequences were assembled using the XBASE package (Dear and Staden, 1991). Multiple sequence alignments were performed with the CLUSTAL W program (Thompson et al., 1994), integrated into the GDE package (Genetic Data Environment; Smith et al., 1994). The FASTA program (Pearson, 1990) was employed for searches in local databases and the BLAST network service (Altschul et al., 1990) for similarity searches in GenBank at the National Center for Biotechnology Information. Custom-made batch utility scripts were developed for submitting queries and browsing the output (BBLAST, TBOB, BFASTA, and FOB). A number of additional programs have been developed by the Organelle Genome Megasequencing Program (OGMP). These utilities are described in more detail and are available through the OGMP website at http://megasun.bch.umontreal.ca/ogmp/ogmpid.html.

The phylogeny programs applied to these sequence data include PROTDIST, FITCH, NEIGHBOUR, PROML, and PUZZLE (Fitch and Margoliash, 1967; Saitou and Nei, 1987; Felsenstein, 1993; Strimmer and Von Haeseler, 1996). The most recent implementation of PROTDIST/FITCH (Phylip, version 3.6; Felsenstein, 1993) was used, which allows a J in/Nei correction for unequal rates of change.
at different amino acid positions. Bootstrap or likelihood estimations were performed according to Felsenstein (1985) and Kishino et al. (1990), respectively.

The complete sequence of P. purpurea mtDNA has GenBank accession number AF114794. At single-nucleotide substitution sites in this published sequence record, we have provided the nucleotide that is the most abundant in the mtDNA population rather than showing the IUB ambiguity code. The long version of the macrosubstitution in the rtl/dpo region (see Results) is given in the record, whereas the short substitution is listed in the "note" field of the corresponding feature.

The DNA sequences of T. maculata cob, cox1, and cox2 have GenBank accession numbers AF116776, AF116777, and AF116778.

ACKNOWLEDGMENTS

We thank Michael Reith (Institute for Marine Biosciences, National Research Council of Canada, Halifax, Nova Scotia) for providing purified organelar DNA and François Michel (Centre de Génétique Moléculaire du CNRS, Gif-sur-Yvette, France) for help in the modeling of intron RNA secondary structures and in reverse transcriptase protein alignments. We thank Charles J. O’Kelly for helpful discussions and Isabelle Plante, Yun Zhu, and Isabelle Robert for excellent technical assistance with library construction and sequencing. Support of the Organelle Genome Megasequencing Program (OGMP) by the Medical Research Council (Grant Nos. SP-34/SP-14226) is also gratefully acknowledged. Generous donations of bacterial protein export system with homologues in plastids and mitochondria. J. Biol. Chem. 273, 18003-18006.


otide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize. EMBO J. 4, 1125–1128.


Paquin, B., and Lang, B.F. (1996). The mitochondrial DNA of Allo-


drial open reading frame orf221 encodes a membrane-


Roe, B.A., Ma, D.P., Wilson, R.K., and Wong, J.F. (1985). The complete nucleotide sequence of the Xenopus laevis mitochondri-


Complete Sequence of the Mitochondrial DNA of the Red Alga *Porphyra purpurea*: Cyanobacterial Introns and Shared Ancestry of Red and Green Algae

Gertraud Burger, Diane Saint-Louis, Michael W. Gray and B. Franz Lang

*Plant Cell* 1999;11:1675-1694

DOI 10.1105/tpc.11.9.1675

This information is current as of February 21, 2021