The Chloroplast Genomes of the Green Algae *Pedinomonas minor*, *Parachlorella kessleri*, and *Oocystis solitaria* Reveal a Shared Ancestry between the Pedinomonadales and Chlorellales

Monique Turmel, Christian Otis, and Claude Lemieux

Département de Biochimie et de Microbiologie, Université Laval, Québec (Québec) Canada

The green algae belonging to the Chlorophyta—the lineage sister to that comprising the land plants and their charophycean green algal relatives (Streptophyta)—have been subdivided into four classes (Prasinophyceae, Ulvophyceae, Trebouxiophyceae, and Chlorophyceae). Yet the Pedinomonadales, an assemblage consisting of tiny, naked uniflagellates with a second basal body, has no clear affiliation with these classes and the branching order of the crown chlorophytes remains unknown. To gain an insight into the phylogenetic position of the Pedinomonadales and the relationships among the recognized chlorophyte classes, we have sequenced the chloroplast genomes of *Pedinomonas minor* (Pedinomonadales) and of two trebouxiophyceans belonging to the Chlorellales, *Parachlorella kessleri* (Chlorellaceae) and *Oocystis solitaria* (Oocystaceae), and compared these genomes with those of previously examined streptophytes and chlorophytes, including *Chlorella vulgaris* (Chlorococcales). Unlike their *Chlorella* homolog, the three newly investigated chloroplast DNAs (cpDNAs) carry a large rRNA-encoding inverted repeat (IR) that divides the genome into large and small single-copy regions. In contrast to the situation found for ulvophycean and chlorophycean cpDNAs, the gene contents of the IR and single-copy regions are strikingly similar to that inferred for the common ancestor of chlorophytes and streptophytes. The intronless 98,340-bp *Pedinomonas* genome is among the chlorophyte cpDNAs featuring the smallest size and most ancestral gene organization. All 105 conserved genes encoded by this genome are included in the gene repertoires of *Oocystis* (111 genes) and *Chlorella* (113 genes), with just *trnR* (ccg) missing from *Parachlorella* cpDNA. Trees inferred from 71 cpDNA-encoded genes/proteins of 16 chlorophytes and nine streptophytes showed that *Pedinomonas* is nested in the Chlorellales, a group of algae lacking flagella. This phylogenetic conclusion is independently supported by uniquely shared gene linkages. We hypothesize that chlorellalean and pedinomonadalean green algae are reduced forms of a distant biflagellate ancestor that might have also given rise to the other known trebouxiophycean lineages. Our structural cpDNA data suggest that the Chlorellales and Pedinomonadales represent a deep branch of core chlorophytes, strengthening the notion that the Trebouxiophyceae emerged before the Ulvophyceae and Chlorophyceae. Our results further emphasize the importance of secondary reduction at both the cellular and genome levels during chlorophyte evolution.

Introduction

The Chlorophyta constitutes a morphologically and ecologically diverse assemblage of green algae. This major lineage, together with that including all land plants and their charophycean green algal relatives (Streptophyta), forms the Viridiplantae (Lewis and McCourt 2004; Pröschold and Leliaert 2007). Most chlorophytes have been divided into four classes (Prasinophyceae, Ulvophyceae, Trebouxiophyceae, and Chlorophyceae) based on the ultrastructure of the flagellar apparatus and cytokinesis during mitosis (Mattox and Stewart 1984). The afflux of molecular data, in particular nuclear-encoded 18S rRNA gene sequences, has improved considerably our understanding of the relationships among chlorophytes and prompted taxonomic revision. Trees inferred from 18S rDNA data have uncovered a number of monophyletic groups within each class but have been unable to unravel the branching orders of these lineages as well as the divergence order of the classes (Lewis and McCourt 2004; Pröschold and Leliaert 2007). Nonetheless, they have robustly resolved the Prasinophyceae as the earliest divergences of the Chlorophyta (Steinkötter et al. 1994; Nakayama et al. 1998; Fawley et al. 2000; Guillou et al. 2004), thus strongly supporting the long-held hypothesis that scaly flagellates are the ancestors of the core chlorophytes (Chlorodendrales, Prasinophyceae + Ulvophyceae + Trebouxiophyceae + Chlorophyceae) (Mattox and Stewart 1984).

Phylogenetic approaches, such as the comparative analysis of whole chloroplast genomes, have proven valuable for inferring relationships among photosynthetic eukaryotes at deep levels (Martin et al. 1998; Qiu et al. 2006; Jansen et al. 2007; Lemieux et al. 2007; Rogers et al. 2007; Turmel et al. 2008, 2009). In addition to providing a large suite of genes for phylogenetic analyses, chloroplast genomes offer genomic structural features for validation of tree topologies (Turmel et al. 2006, 2008; Lemieux et al. 2007). To date, the complete chloroplast genome sequences of 14 chlorophytes and 6 streptophyte green algae have been reported (see references in the Materials and Methods) and their comparisons have revealed that chlorophyte chloroplast DNAs (cpDNAs) display much more variability at the levels of size, gene content, intron content, and gene order than their streptophyte counterparts. The genomes of the prasinophycean green algae *Nephrosemis olivacea* (Turmel, Otis, and Lemieux 1999) and *Pyramimonas parkeae* (Turmel et al. 2009) are those most closely resembling their streptophyte counterparts. Trees inferred from the concatenated sequences of 70 chloroplast protein-coding genes and their predicted proteins have recently uncovered the existence of a sister relationships between the prasinophycean orders Pyramimonadales and Mamiellales and have also shown that the euglenids, a group of photosynthetic eukaryotes belonging to the Excavata, acquired their chloroplasts secondarily from a member of the Pyramimonadales (Turmel et al. 2006, 2008; Lemieux et al. 2007). To date, the complete chloroplast genome sequence of *Oocystis solitaria* is among the smallest and contains the fewest genes of any green algal cpDNA, with just 105 conserved genes encoded in the gene repertoires of *Pedinomonas* (113 genes), *Chlorella* (113 genes), and *Oocystis* (111 genes), with just *trnR* (ccg) missing from *Parachlorella* cpDNA. Trees inferred from 71 cpDNA-encoded genes/proteins of 16 chlorophytes and nine streptophytes showed that *Pedinomonas* is nested in the Chlorellales, a group of algae lacking flagella. This phylogenetic conclusion is independently supported by uniquely shared gene linkages. We hypothesize that chlorellalean and pedinomonadalean green algae are reduced forms of a distant biflagellate ancestor that might have also given rise to the other known trebouxiophycean lineages. Our structural cpDNA data suggest that the Chlorellales and Pedinomonadales represent a deep branch of core chlorophytes, strengthening the notion that the Trebouxiophyceae emerged before the Ulvophyceae and Chlorophyceae. Our results further emphasize the importance of secondary reduction at both the cellular and genome levels during chlorophyte evolution.
2318 Turmel et al.

Earlier chloroplast phylogenetic analyses in which representatives of the five main lineages recognized in the Chlorophyceae were sampled had revealed the dichotomy of this class: One major clade comprises the Chlamydomonadales and Sphaeropleales, whereas the other contains the Oedogoniales, Chaetopeltidales, and Chaetophorales (Turmel et al. 2008). In both investigations, phylogenetic conclusions received strong support from comparisons of chloroplast genomic features.

In the present cpDNA study, we sampled members of the Pedinomonadaceae (also designated as Pedinomonadaceae) and of the Chlorellales, a major lineage of the Trebouxiophyceae. Pedinomonadacean green algae have no clear affiliation with other groups of chlorophytes (Pickett-Heaps and Ott 1974; Melkonian 1990; Moestrup 1991). Representing only three genera (Pedinomonas, Resultor, and Marsupionomonas), they consist of small, naked flagellates lacking scales on their unique flagellum. The presence of two basal bodies, each associated with two microtubular roots, suggests that they are not primarily uniflagellates (Melkonian 1990). Although the persistent telophase spindle during mitosis and the eye-spot located opposite the flagellar insertion have been considered to be ancient features shared with some prasinophyceans, an affinity with the Ulvophyceae has been proposed based on the configuration of the flagellar apparatus and the structure of the stellate pattern of the flagellar transition region (Melkonian 1990; Moestrup 1991).

Members of the Trebouxiophyceae exhibit a high degree of morphological heterogeneity (Lewis and McCourt 2004) and form at least six distinct monophyletic groups in 18S rDNA trees (Krienitz et al. 2003, 2004; Henley et al. 2004; Karsten et al. 2005; Eliaš et al. 2008; Sluiman et al. 2008). Just three trebouxiophycean chloroplast genomes, all lacking a large inverted repeat (IR) encoding the rRNA operon, have been fully sequenced thus far: the miniature genome of the parasitic Resultor (Michel et al. 1989) and Michel and Westhof (1990), respectively. For each genome, the regions contain- ing nonoverlapping repeated elements were mapped with RepeatMasker (http://www.repeatmasker.org/) running under the WU-Blast 2.0 search engine (http://blast.wustl.edu/), using the repeats > 30 bp identified with REPuter 4.8 (Gene Codes Corporation, Ann Arbor, MI).

Cloning and Sequencing of Chloroplast Genomes

For the procedures used to sequence the Pedinomonas chloroplast genome, the reader should consult the report describing the mitochondrial genome of this chlorophyte (Turmel, Lemieux, et al. 1999). Because these organelle genomes copurified in CsCl–bisbenzimide gradients, they were decrypted in the course of the same sequencing project. The Oocystis and Parachlorella chloroplast genome sequences were determined essentially as reported by Turmel, Lemieux, et al. (1999), except for the following modifications concerning the sizes of the fragments selected for cloning, the choice of cloning vector, the preparation of DNA templates, and the software for sequence assembly. Fragments of 2000–4000 bp were cloned into pSMART-HCKan (Lucigen Corporation, Middleton, WI). After hybridization of the resulting clones with the original DNA used for cloning, plasmids from positive clones were amplified with the Illustra TempliPhi Amplification Kit (GE Healthcare, Baie d’Urfé, Canada) to produce templates for sequencing reactions. The DNA sequences generated in these reactions were assembled using SEQUENCER 4.8 (Gene Codes Corporation, Ann Arbor, MI).

Chloroplast Genome Analyses

Genes and open reading frames (ORFs) were identified as described previously (Turmel et al. 2006). Secondary structures of group I and group II introns were modeled according to Michel et al. (1989) and Michel and Westhof (1990), respectively. For each genome, the regions containing nonoverlapping repeated elements were mapped with RepeatMasker (http://www.repeatmasker.org/) running under the WU-Blast 2.0 search engine (http://blast.wustl.edu/), using the repeats > 30 bp identified with REPuter 2.74 (Kurtz et al. 2001) as input sequences. Conserved clusters of genes with identical polarities in selected chloroplast genomes were identified using a custom-built program.
Phylogenetic Inferences

An amino acid data set and the corresponding nucleotide data set with first and second codon positions were derived from the chloroplast genomes of 16 chlorophytes and nine streptophytes. Species names and accession numbers for the chlorophytes are as follows. Prasinophyceae: *N. olivacea* (Pseudoucouphfidaeae, N.000927 (Turmel, Otis, and Lemieux 1999); *Ostreococcus tauri* (Mamiellaeales), N.008289 (Robbens et al. 2007); *Monosmptis* sp. (Mamiellales), N.012101 (Turmel et al. 2009); *Pycnococcus provasoli* (Pseudocouphfidaeae, Pycnococcaceae), N.012097 (Turmel et al. 2009); *P. parkeae* (Pyramimonadales), N.012099 (Turmel et al. 2009); Ulvophyceae: *Oltmannsiellopsis viridis* (Oltmannsiellopsidales), N.008099 (Pombert et al. 2006); *Pseudendoclonium akinetum* (Ulotrichales), N.008114 (Pombert et al. 2005); Trebouxiophyceae: *C. vulgaris* (Chlorellales), N.001865 (Wakasugi et al. 1997); *L. terrestris* (incerta sedis), N.009681 (de Cambiaire et al. 2007); *O. solitaria* (Chlorellales), FJ968739 (this study), *P. kessleri* (Chlorellales), FJ968741 (this study); Pedinomonadales: *P. minor* FJ968740 (this study); Chlorophyceae: *Chlamydomonas reinhardtii* (Chlamydomonadales), N.005353 (Maul et al. 2002); *Schedesmus obliquus* (Sphaeropleales), N.008101 (de Cambiaire et al. 2006); *Stigeclonium helveticum* (Chaetophorales), N.008372 (Belanger et al. 2006); *Oedogonium cardiaum* (Oedogoniales), N.011031 (Brouard et al. 2008). Note that the non-photosynthetic parasite *Heliocystisporidium* sp. (Chlorellales, Trebouxiophyceae) was not selected because of its highly reduced chloroplast genome repertoire (26 protein-coding genes) (de Koning and Keeling 2006). Representing the Streptophyta were the charophycean green algae *Mesostigmu viride* (Mesostigmatales), N.002186 (Lemieux et al. 2000); *Chlorokybus aphthios* (Chlorokybales), N.008822 (Lemieux et al. 2007); *Staurospore punctulatum* (Zygnematales), N.008116 (Turmel et al. 2005); *Zygonea circumcarinatum* (Zygnematales), N.008117 (Turmel et al. 2005); *Chaetosphaeridium globosum* (Coleochaetales), N.004115 (Turmel et al. 2002); and *Chara vulgaris* (Charales), N.008097 (Turmel et al. 2006) as well as the embrophytes *Anthoceros formosae* (homwort), N.004543 (Kugita et al. 2003); *Marchantia polymorpha* (liverwort), N.001319 (Ohyama et al. 1986); and *Physcomitrella patens* (moss), N.005087 (Sugiuji et al. 2003).

The data sets were allowed to contain missing data; however, limitations were imposed to the proportion of missing data by selecting for analysis the protein-coding genes (eight categories) (minimal number of sequences for a flank position) set to 15, and the protein alignments were concatenated. To obtain the nucleotide data set, the multiple sequence alignment of each protein was converted into a codon alignment, the poorly aligned and divergent regions in each codon alignment were excluded using GBLOCKS 0.91b with the –b2 option (options –b2 = 15 and –t = c (the latter specifying that selected sequences are complete codons), the individual codon alignments were concatenated, and finally third codon positions were excluded with PAUP*4.0b10 (Swoford 2003). Missing characters represented 5.2% and 5.1% of the amino acid and nucleotide data sets, respectively.

Phylogenetic inferences were carried out using the maximum likelihood (ML) and Bayesian methods. Treefinder (version of April 2008) (Jobb et al. 2004) was used to perform the ML analyses of both data sets and to identify the best models fitting the data under the Akaike information criterion. ML trees were inferred from the amino acid data set using the LG model with the observed amino acid frequencies + a gamma distribution of rates across sites with eight categories + invariable sites (LG + F + I). For the analysis of the nucleotide data set, we used the general time reversible model (GTR) + invariable site (GTR + I). Confidence of branch points was estimated by 100 bootstrap replications.

*PhyloBayes* 2.3c (http://www.lirmm.fr/mab/article.php?3d_article=329) analysis of the amino acid data set was performed with the category amino acid site-heterogeneous mixture (CAT) model + a gamma distribution (eight categories) (CAT + I). Two independent runs were carried out with default *PhyloBayes* conditions until a maxdiff value < 0.15 was achieved to ensure chain equilibration (30,000 generations). The first 4,500 points were discarded as burn-in, and the posterior consensus was computed on the remaining trees. The nucleotide data were analyzed with MrBayes 3.1.2 (Ronquist and Huelsenbeek 2003) using the GTR + I (eight categories) + I model. Two independent Markov chain Monte Carlo runs, each consisting of three heated chains in addition to the cold chain, were carried out using the default parameters. The length of each run was 3 million generations after a burn-in phase of 500,000 generations. Trees were sampled every 100 generations. Convergence of the two independent runs was verified according to the output of the “sump” command; this output was also used to determine the burn-in phase. Posterior probability values were estimated from all the sampled trees using the “sump” command.

Results

The *Pedinomenas, Parachlorella*, and *Oocystis* Chloroplast Genomes Feature a Similar Quadruplicate Structure

Circular maps encompassing 98,340 and 123,994 bp were assembled for the cpDNAs of *Pedinomenas* and *Parachlorella*, respectively (figs. 1 and 2). With regards to the *Oocystis* chloroplast genome, the collection of plastid clones and polymerase chain reaction (PCR) fragments...
examined yielded a single linear contig of 96,287 bp featuring virtually all of the genes previously identified in Chlorella cpDNA (fig. 3). A segment of 5 kbp at one of the termini of this contig harbors large ORFs sharing no homology with known DNA sequences. As observed for all chloroplast genomes fully sequenced thus far, the map of the Oocystis cpDNA is almost certainly circular. However, PCR analyses using various sets of primers did not enable us to circularize the map shown in figure 3.

When one glances at the maps of the Pedinomonas and Parachlorella genomes (figs. 1 and 2), the most striking similarities that can be discerned include the large rRNA operon-encoding IR that divides each genome into small and large single-copy (SSC and LSC) regions, the relatively high density of coding sequences (table 1), and the paucity of introns (table 1). The genes encoded by the two genomes are similarly partitioned among the three main genomic regions and the direction of transcription of the rRNA operon is identical, that is, oriented toward the SSC region as in virtually all IR-containing cpDNAs of streptophytes (Palmer 1991; Raubeson and Jansen 2005) and those of the prasinophyceans Nephroselmis and Pyramimonas (Turmel, Otis, and Lemieux 1999; Turmel et al. 2009). The gene partitioning pattern matches that inferred for the common ancestor of chlorophytes and streptophytes (Turmel, Otis, and Lemieux 1999), with the exception of a few contiguous genes that were transferred from the LSC region to a location immediately downstream of the rRNA operon (the genes marked with arrows in figs. 1 and 2).

The sequence data for the Oocystis genome also uncovered the presence of an IR, but only the boundaries of this sequence with the LSC region were localized (fig. 3). None of the clones downstream of the rRNA operon revealed divergent sequences indicative of IR–SSC junctions. The SSC region is likely to contain the missing stretch of the Oocystis genome sequence because this
segment lies at the proximity of a suite of 11 genes usually found in the SSC domain in ancestral-type genomes. Even though the IR–SSC boundaries were not identified, the currently available data clearly indicate that the *Oocystis* genome features an ancestral-type quadripartite architecture, is intron-poor, and is more densely packed than its *Parachlorella*, *Chlorella*, and *Pedinomonas* counterparts (table 1). In *Oocystis*, the sole departures from the ancestral quadripartite architecture include the migration of *ycf20*, a gene typically located in the SSC region, to a locus near one of the LSC–IR boundaries and the finding that one copy of the duplicated *trnN*(guu), a gene also characteristic of the SSC region, lies in the LSC region (fig. 3). The second copy of *trnN*(guu) resides immediately downstream of the rRNA operon next to the other genes usually found in the SSC partition. Considering that the relocalized *ycf20* and *trnN*(guu) are far apart in the *Oocystis* genome, there is little evidence that these genes were cotransferred to the LSC region.

Slight modifications from the ancestral quadripartite structure have also been observed for the prasinophyceans *Nephroselmis* and *Pyramimonas* (Turmel, Otis, and Lemieux 1999; Turmel et al. 2009). In contrast, the IR-containing genomes of the ulvophyceans (Pombert et al. 2005, 2006) and chlorophyceans (Maul et al. 2002; de Cambiaire et al. 2006; Brouard et al. 2008) examined thus far as well as the reduced genome of the tiny prasinophycean *Ostreococcus* (Robbens et al. 2007) differ...
substantially from the ancestral pattern, implying that numerous genes were exchanged between opposite sides of the IR and that the rRNA operon itself might have undergone inversion during this process. Gene shuffling, however, was more extensive in chlorophycean lineages than in the lineages that led to *Ostreococcus* and the ulvophyceans *Oltmannsiellopsis* and *Pseudendoclonium*. The moderate changes observed for the latter genomes allowed inferences to be made regarding the direction of gene movements (Pombert et al. 2005, 2006).

The *Pedinomonas* Genome Contains a Subset of the Conserved Genes Found in *Oocystis*, *Parachlorella*, and *Chlorella* cpDNAs

*Pedinomonas* cpDNA encodes 105 conserved genes and is one of the smallest chloroplast genomes reported for photosynthetic chlorophytes. Only the completely sequenced chloroplast genomes of the prasinophyceans *Ostreococcus* (88 genes, 71,666 bp) and *Pycnococcus* (98 genes, 80,211 bp) are known to be smaller than that of *Pedinomonas* (Robbens et al. 2007; Turmel et al. 2009). On the basis of gene content data alone, it is hard to predict which of the core chlorophyte chloroplast genomes examined to date is the most closely related to *Pedinomonas* (Robbens et al. 2007; Turmel et al. 2009). On the basis of gene content data alone, it is hard to predict which of the core chlorophyte chloroplast genomes examined to date is the most closely related to *Pedinomonas* cpDNA. Although the loosely packed genomes of the trebouxiophycean *Leptosira* (195,081 bp, 106 conserved genes) and of the ulvophyceans *Oltmannsiellopsis* (151,933 bp, 104 genes) and *Pseudendoclonium* (195,867 bp, 105 genes) most closely resemble their *Pedinomonas* counterpart in terms of gene number, they lack a number of genes that are present in the *Pedinomonas* chloroplast (e.g., *ycf47*, *cysA*, *cysT*, *trnS*(gga), and *trnT*(ggu) in the case of ulvophyceans) and also differ by the presence of genes that are lacking in the pedinomonadalean organelle (table 2). Interestingly, the chloroplast gene repertoire of *Pedinomonas* overlaps almost entirely with those of
Introns

Oocystis, Parachlorella, and Chlorella (table 2). With 110 conserved genes in common, the gene complements of the three chlorellalean algae are nearly identical. The 113-gene repertoire of Chlorella differs from that of Parachlorella (112) by the presence of trnR(cgg) and from that of Oocystis (minimum of 111 genes) by the presence of trnL(cua) and trnL(gag). All Pedinomonas genes are included in the Chlorella and Oocystis gene complements, with just trnR(cgg) missing from the Parachlorella genome. In contrast, six of the Pedinomonas genes have no homolog in the chloroplast of the trebouxiophycean Leptosira (table 2).

Chloroplast Phylogenomic Analyses Robustly Affiliate Pedinomonas with the Chlorellales

To gain an insight into the positions of the Pedinomonales and Trebouxiophyceae within the Chlorophyta, we generated data sets of 71 concatenated proteins and genes (first and second codon positions) from the completely sequenced chloroplast genomes of 16 chlorophytes and nine streptophytes and analyzed them using the ML and Bayesian methods (fig. 4). The protein trees showed with strong support that Pedinomonas is embedded within the clade

| Table 1 | General Features of cpDNAs from Pedinomonas and the Trebouxiophyceae Oocystis, Parachlorella, Chlorella, and Leptosira |
|---------|---------------------------------------------------------------------------------------------------------------|
| Feature | Pedinomonas | Oocystis | Parachlorella | Chlorella | Leptosira |
| Size (bp) | | | | | |
| Total | 98,340 | >96,287 | 123,994 | 150,613 | 195,081 |
| IR | 10,639 | nd | 10,913 | b | — |
| SSC | 6,664 | nd | 13,871 | b | — |
| LSC | 70,398 | 71,295 | 88,297 | b | — |
| A + T (%) | 65.2 | 71.0 | 70.0 | 68.4 | 72.7 |
| Conserved genes | 105 | 111 | 112 | 113 | 106 |

| Intergenic sequences (%) | | | | | |
| Fraction of genome (%) | 25.6 | 22.2 | 35.2 | 45.4 | 51.5 |
| Average size (bp) | 203 | 184 | 348 | 526 | 913 |

| Short repeats | | | | | |
| Fraction of genome (%) | 0 | 0.7 | 4.1 | 7.8 | 4.9 |

<sup>a</sup> Given the unknown positions of the IR–SSC boundaries, sizes of the IR and SSC could not be determined (nd).
<sup>b</sup> Because Chlorella and Leptosira cpDNAs lack an IR, only the total sizes are given for these genomes.
<sup>c</sup> Conserved genes refer to coding sequences usually found in chloroplast genomes: Unique ORFs and intron ORFs were not taken into account. Genes present in the IR were counted once.
<sup>d</sup> This number differs from that (112) indicated in previous reports for the following reasons: We found that the orf70 and orf504, originally identified by Wakasugi et al. (1997), correspond to ycf67 and tisS, respectively (this study); however, our Blast analyses provided no convincing evidence that the ORF annotated as minE encodes this gene product.
<sup>e</sup> In addition to conserved genes, all ORFs ≥ 100 codons were considered as gene sequences.
<sup>f</sup> Nonoverlapping repeat elements were mapped on each genome as described in the Materials and Methods.

| Table 2 | Differences in Gene Content between the Chloroplast Genomes of Pedinomonas, Trebouxiophyceans, and Ulvophyceans |
|---------|---------------------------------------------------------------------------------------------------------------|
| Gene | Pedinomonas | Oocystis | Parachlorella | Chlorella | Leptosira | Oltmannsiellopsis | Pseudendoclonium |
| ccsA | + | + | + | + | + | + | + |
| chIB | – | – | + | + | + | + | + |
| chlA | + | + | + | + | + | + | + |
| chlL | – | + | + | + | + | + | + |
| chlN | – | + | + | + | + | + | + |
| cysA | – | + | + | + | + | + | + |
| cysT | – | + | + | + | + | + | + |
| infA | – | + | + | + | + | + | + |
| tisS<sup>a</sup> | – | + | + | + | + | – | + |
| ycf12 | + | + | + | + | + | + | + |
| ycf47<sup>a</sup> | + | + | + | + | + | + | + |
| trnL(cua) | + | + | + | + | + | + | + |
| trnV(cua) | + | + | + | + | + | + | + |
| trnR(cgg) | + | + | + | + | + | + | + |
| trnI(ccu) | + | + | + | + | + | + | + |
| trnN(gga) | + | + | + | + | + | + | + |
| trnT(gu) | + | + | + | + | + | + | + |

<sup>a</sup> This gene, previously annotated as ycf712, encodes the tRNA(ile)-lysidine synthetase.

<sup>b</sup> This gene, previously annotated as ycf47, encodes the tRNA(leu)-lysidine synthetase.

<sup>c</sup> This gene, previously annotated as ycf504, encodes the tRNA(leu)-lysidine synthetase.
containing the three members of the Chlorellales (fig. 4A). The representative of the Oocystaceae, Oocystis, represents the first branch of this clade, whereas Pedinomonas is sister to the clade unifying the two representatives of the Chlorellaceae (Chlorella and Parachlorella). Although the gene trees identified the same clade and an identical divergence order for the four lineages, weaker support was observed for the node uniting all four algae as well as that uniting Pedinomonas with Parachlorella and Chlorella. Both the protein and gene trees inferred from the chloroplast data positioned the trebouxiophycean Leptosira as sister to the clade containing the four representatives of the Chlorophyceae (fig. 4), implying that the Trebouxiophyceae are nonmonophyletic. However, given the long branch formed by Leptosira (incertae sedis) and also the fact that this green alga is the sole trebouxiophycean representative outside the Chlorellales, this sister relationship may be the result of phylogenetic reconstruction artifacts. Consistent with the hypothesis that taxon sampling has an influence on the position of Leptosira, a weakly supported clade clustering Leptosira and Chlorella was recovered in recent chloroplast phylogenomic analyses (Turmel et al. 2009).

Analysis of Gene Order Reveals an Ancestral Genome Organization in Pedinomonas and a Close Alliance between the Pedinomonadales and Chlorellales

Because phylogenomic analyses are inherently associated with sparse taxon sampling, they can lead to trees robustly supporting an artifactual clustering of taxa (Brinkmann and Philippe 2008; Heath et al. 2008). However, the use of complete genome sequences in these analyses offers the opportunity to validate topologies by analyzing structural genomic features (Rokas 2006). Considering that gene order analysis has proven to be a promising approach for testing phylogenies inferred from green plant chloroplast genomes (Turmel et al. 2006, 2008; Lemieux et al. 2007), gene orders in the three newly sequenced chlorophyte chloroplast genomes were compared with one another and with those in the previously examined chlorophytes and the streptophytes Mesostigma and Chlorokybus. In addition to providing information on similarity of global gene organization, these comparative analyses allowed us to examine the level of retention of the ancestral gene clusters that predate the divergence of chlorophytes and streptophytes and also to track the gene clusters that arose following the emergence of the Chlorophyta.

Collectively, the Pedinomonas and chlorellalean genomes share 16 blocks of colinear sequences encoding 65% of the genes (67/104) common to these genomes (table 3). The genomes having the highest similarity in overall gene order are those of Parachlorella and Chlorella (Chlorellaceae): Seventeen synthetic blocks encoding up to 20 genes, for a total of 94 genes (i.e., 83% of the genes in each genome) were identified in these genomes. Although the Pedinomonas genome was found to more closely resemble the Oocystis genome, with 17 synthetic blocks encoding 79 of its 105 genes (i.e., 75% of the genome), the extent of similarity with the Parachlorella genome was only marginally different (17 synthetic blocks encoding 77 genes, i.e., 73% of the genome). These relative similarities in gene order are in accordance with the genetic distances separating the Pedinomonas chloroplast from its Oocystis and Parachlorella homologs in the phylogenomic trees reported in this study (fig. 4). The pairwise comparison of the Pedinomonas and Chlorella genomes, however, revealed less similarity in gene organization (68%) than expected, with a deficit of six genes (71 genes) in the 17 observed synthetic blocks. As reported for other IR-lacking chloroplast genomes (Palmer et al. 1987; Strauss et al. 1988; Turmel et al. 2005; Bélauger et al. 2006; de Cambiaire et al. 2007), loss of the IR in Chlorella cpDNA was most likely accompanied by the breakdowns and rearrangements of a number of gene blocks (see figs. 5 and 6).
Numerous gene linkages shared by *Pedinomonas* and/or chlorellalean chloroplast genomes arose during chlorophyte evolution (fig. 6). Because of the limited number of chlorophyte lineages investigated so far and the uncertainty regarding the branching pattern of core chloroplasts, it is difficult to infer the timings of emergence of these gene linkages. A subset of the chlorophyte-specific gene linkages found in *Pedinomonas* and/or chlorellaleans have been observed in *Leptosira* and/or other major groups of chlorophytes, indicating that they were inherited from a distant ancestor (often a prasinophycean). The set of 11 gene linkages restricted to *Pedinomonas* and one or more chlorellaleans represents putative synapomorphies that support a shared ancestry of the Chlorellales and Pedinomonadales. On the other hand, the 10 gene linkages unique to *Para-chlorella* and *Chlorella* are consistent with the notions that these green algae belong to the same monophyletic group (Chlorellaceae) and display the most similar chloroplast gene arrangements among the members of the Chlorellales + Pedinomonadales clade examined so far.

Overall, the gene order analyses reported above are in total agreement with our phylogenomic inferences in showing that the *Pedinomonas* chloroplast genome and its three chlorellalean homologs form a coherent group featuring gene linkages not previously identified in other core chlorophytes. Besides the specific gene linkages uniting these genomes, it is worth mentioning that all three IR-containing cpDNAs display the *trnD*(guc) and *trnMe* genes in their IR and that both the *Pedinomonas* and *Oocystis* IRs include as well the *trnG*(gcc)–*trnS*(gca) pair. These tRNA genes appear to be unique to the IRs of these algae and thus are likely to contribute additional synapomorphies supporting the Chlorellales + Pedinomonadales clade.

### Discussion

**Evolutionary Implications of the Inferred Alliance Between the Pedinomonadales and Chlorellales**

Our chloroplast phylogenomic analyses suggest that the small group of uniflagellates making up the Pedinomonadales represents a chlorellalean lineage distinct from the Oocysteaceae and Chlorellaceae (fig. 4). The concept that the Pedinomonadales are allied with the Chlorellales or more generally with the Trebouxiophyceae was anticipated. Pedinomonadales green algae have been regarded to have an affinity to the Prasinophyceae because they share morphological features with some members of this ancient group of algae, namely, a persistent telophase spindle during mitosis and an eye-spot opposite the flagella and in the cleavage plane of the cell during cytokinesis (Melkonian 1990; Moestrup 1991). On the other hand, derived ultrastructural characters of the flagellar apparatus, more specifically the counterclockwise orientation of basal bodies and the cruciate flagellar root system, support the notion that pedinomonadales evolved after the divergence of the prasinophycean lineages but do not belong to the Chlorophyceae. Although the combination of the abovementioned morphological characters led Melkonian (1990) to believe that the Pedinomonadales are related to the Ulvophyceae, he proposed that this group be treated

---

| Table 3 Conserved Gene Clusters in the *Pedinomonas*, *Oocystis*, *Parachlorella*, and *Chlorella* Chloroplast Genomes |
|---------------------------------------------------------------|
| accD–cysA | cysF–ycfI |
| petB–petD | psaK–psaB |
| psaB–psaC | rps2–atpI |
| rplF–ycfA–ycf7 | atpF–atpE |
| petE–petL–petG | rps2–atpH–atpF–atpA |
| psbK–ycf12–psaM | rrs–trnGau–trnAugg–rrl–rrf |
| atpB–atpE–rps4–trnS(gca) | rpl20–rps18–trnW(gca)–trnP(ugg)–psaD–rps12–rps7–atpA–rpl19–ycf4 |
| psbE–psbF–psBL–psB | rpl23–rps19–rps3–rpl16–rpl14–rpl5–rps8–rps6–rps11–rpoA–rpoB |
| rpl2–atpL–atpH–atpF–atpA | rps9–rpl12 |

Underlined are the gene linkages that arose during the evolution of chlorophytes (fig. 6); all other gene linkages predate the divergence of the Chlorophyta and Streptophyta (fig. 5). Genes separated by a dash are encoded on the same DNA strand, whereas those separated by a diagonal bar are encoded on different strands.
as a chlorophyte order of uncertain affinity. It should be noted that the motile cells of trebouxiophyceans also display a counterclockwise orientation of basal bodies and a cruciate flagellar root system; however, in contrast to ulvophyceans, a nonpersistent telophase spindle is generally observed during mitosis (Mattox and Stewart 1984).

Independent lines of evidence indicate that the chloroplast phylogenetic trees reported here reflect the true relationships among the members of the Chlorellales and Pedinomonadales. First, the chloroplast phylogenetic analyses are congruent with published phylogenies inferred from 18S rDNA sequences in recovering a clade that comprises representatives of the Oocystaceae and Chlorellaceae and in resolving these two families of the Chlorellales as separate lineages (Krienitz et al. 2003, 2004; Henley et al. 2004; Sluiman et al. 2008). However, considering that no 18S rRNA gene sequence has been reported for a pedinomonadalean taxon, it remains to be seen whether the inclusion of the Pedinomonadales within the Chlorellales is in agreement with nuclear data. Second, consistent with the notion that the Pedinomonadales and Chlorellales descend from a common ancestor, we have uncovered in the present study structural cpDNA features (gene linkages and gene content of IR) that are uniquely shared by these
green algae. Third, the possibility that the placement reported here for *Pedinomonas* results from systematic errors of phylogenetic reconstructions (Delsuc et al. 2005; Brinkmann and Philippe 2008; Verbruggen and Theriot 2008) is unlikely because this taxon remained nested in the Chlorellales in recent chloroplast phylogenomic analyses with a better sampling of trebouxiophycean lineages (Turmel M, Otis C, Lemieux C, unpublished results). In this context, it should be noted that we used the site-heterogeneous CAT model in our Bayesian analyses of the amino acid data set to overcome some of the risks of systematic errors (Lartillot and Philippe 2004). Models accommodating site-specific biochemical constraints in the amino acid replacement pattern at each position of the alignment have been shown to alleviate long-branch attraction artifacts (Lartillot et al. 2007; Rodriguez-Ezpeleta et al. 2007).

The nature of the flagellate ancestor of the Chlorellales has remained ambiguous because all members of this group lack flagella (Lewis and McCourt 2004). The placement of the uniflagellate *Pedinomonas* within the Chlorellales indicates that the most recent common ancestor of the Pedinomonadales and Chlorellales was most likely a uniflagellate whose cell architecture was simplified through complete loss of the flagellar apparatus in the lineages leading to the Oocystaceae and Chlorellaceae. In the lineage leading to *Pedinomonas*, reduction of cell size was probably accompanied by loss of the cell wall. Given the relative positions of the chlorellalean and pedinomonadalean lineages in the chloroplast trees, we infer that the flagellar apparatus was independently lost in the Oocystaceae and Chlorellaceae. A multi-layered cell wall, with the crystalline cellulose fibrils in each layer being perpendicular to the orientation found in adjoining layers, distinguishes the oocystaceans (Hepperle et al. 2000) from the chlorellaceans, suggesting that this trait arose after the divergence of these families. In contrast, as in other trebouxiophycean lineages, chlorellaceans feature a trilaminar cell wall with an outer electron-dense layer indicative of sporopollenin (Krienitz et al. 1999; Henley et al. 2004). Interestingly, it was hypothesized that pedinomonadaleans are not primarily uniflagellates based on the presence of a second nonfunctional basal body in some members of the group (Melkonian 1990), implying that the uniflagellate ancestor of the Pedinomonadales and Chlorellales might have originated from a biflagellate related to the trebouxiophyceans that have retained the flagellar apparatus.

**Patterns of Chloroplast Genome Evolution in the Chlorellales + Pedinomonadales Clade**

The IR-containing cpDNAs of *Pedinomonas*, *Oocystis*, and *Parachlorella* resemble one another in overall structure, gene content, and intron content, indicating that the chloroplast genome evolved rather conservatively in the Chlorellales + Pedinomonadales clade. However, relative to their homologs in previously investigated core chlorophytes (ulvophyceans and chlorophyceans), they stand apart by their ancestral gene organization. Their quadripartite structures show striking similarities to that inferred for the common ancestor of chlorophytes and streptophytes. Moreover, their gene repertoires exhibit a few genes (*ycf47, cysA, cysT, trnS(gga), trnT(ggu)*) that are missing from both their ulvophycean and chlorophycean homologs but are present in the ancestral-type genome of the prasinophycean *Nephroselmis*. The higher degree of ancestral character observed for the gene partitioning pattern and gene

---

**Fig. 6.—Conserved gene pairs in core chlorophyte cpDNAs that originated after the emergence of the Chlorophyta.** In this category were classified the gene pairs absent in both *Mesostigma* and *Chlorokybus* and present in at least two core chlorophytes belonging to the same lineage (Trebouxiophyceae, Ulvophyceae, and Chlorophyceae). For each gene pair, adjoining termini of the genes are indicated. Filled boxes indicate the presence of gene pairs, whereas gray or open boxes indicate the absence of gene pairs. A gray box indicates that the two genes associated with a gene pair are found in the genome but are unlinked. An open box indicates that one or both genes associated with a gene pair are absent from the genome. Gene pairs linked by brackets are contiguous on the genome.
content of the Pedinomonas, Oocystis, and Parachlorella genomes relative to those of other core chlorophyte genomes might be indicative of an early origin of the Chloro-ellales + Pedinomonadales clade, which would be consistent with the view of Moestrup (1991) regarding the Pedinophyceae and with the hypothesis that the Ulvophyceae are sister to the Chlorophyceae. Previously reported analyses of chloroplast gene order and gene content together with mitochondrial phylogenomic analyses with poor taxon sampling also favored the view that the Chloro-ellales represents an early-diverging lineage among the core chlorophytes (Pombert et al. 2004). However, the chloro- plast phylogenomic analyses presented in this study and other reports (Lemieux et al. 2007; Turmel et al. 2008, 2009) are inconclusive with regards to the relative positions of the Trebouxiophyceae, Ulvophyceae, and Chlorophyceae. Phylogenies incorporating taxa representing the main lineages recognized in the Trebouxiophyceae and additional chlorophyte lineages will be required to resolve the issue regarding the monophyletic nature of the Treboux-iphycyeae and to unravel the position of this algal group in the radiation of core chlorophytes.

The slight deviations from the ancestral quadripartite structure that we observed in the Pedinomonas, Oocystis, and Parachlorella cpDNAs are associated with distinct sets of genes, indicating that independent events led to the trans- fer of these genes from one side of the rRNA operon to the other. Considering that all relocalized genes, with the exception of the duplicated Oocystis trnN(guu), are present within or near the IR and that the contiguous trnG(gcc) and trnS(uga) genes downstream of the rRNA operon in the Pedinomonas genome are found immediately upstream of the same operon in the Oocystis genome, rearrangements of the IR itself, via expansion–contraction events and inver-sions of two overlapping internal segments containing the entire rRNA operon, might account for these gene transfers. A model involving complex rearrangements of this type has been recently proposed to explain the considerable differences observed for the quadripartite structure of the 218-kb Pelargonium chloroplast genome relative to other land plant cpDNAs (Chumley et al. 2006). Like the cpDNA of the ulvophycean Pseudendoclonium (Pombert et al. 2005), this unusually large angiosperm genome exhibits an IR in which the rRNA operon is transcribed toward the LSC region; to account for this change in orientation, a single inversion was inferred for the region containing the rRNA operon (Chumley et al. 2006).

The compactness of the Pedinomonas chloroplast and mitochondrial genomes lends credit to the notion that tiny unicellular organisms tend to exhibit small genomes. As previously observed for the lineages leading to the prasinophy-ccean coccoids Ostreococcus and Pyrococcus (Robbens et al. 2007; Turmel et al. 2009), the chloroplast genome was reduced in gene content and constrained to maintain a compact gene organization in the Pedinomonas lineage; however, a smaller proportion of conserved genes are missing from Pedinomonas cpDNA compared with its prasinophycean homologs. With 105 conserved genes, the 98-kb cpDNA of Pedinomonas lacks eight of the 113 genes predicted to have been present in the chloroplast of the common ancestor of the Pedinomonadales and Chlorellales (table 2), whereas the 72-kb IR-containing cpDNA of Ostreococcus and the 80-kb IR-less cpDNA of Pyrococcus encode 88 and 98 conserved genes, respectively. Because the nuclear genome of a pedinomonadalean green alga has not been decrypted, the chloroplast genes missing in Pedinomono- nas cannot be unequivocally interpreted in terms of gene transfers from the chloroplast to the nucleus. The infA, tilS, and trnL(cau) genes, whose products play a role in protein synthesis, probably sustained such transfers; however, some might have been definitely extinguished because their presence is dispensable (e.g., trnL(caa) and trnL(gag)) or their requirement is restricted to certain growth and physiological conditions (e.g., the chlB, chlL, and chlN genes associated with chlorophyll synthesis in the dark). Finally, another feature distinguishing the Pedinomonas chloroplast genome from the Ostreococcus and Pyrococcus cpDNAs is the high level of retention of ancestral gene clusters (fig. 5). Of all the chloroplast genomes of core chlorophytes examined thus far, Pedinomonas cpDNA exhibits the most ancestral gene order.

Cell simplification and reduction in cell size had more influence on the mitochondrial genome than on the chloro-plast genome in the Pedinomonas lineage, but the reverse situation has been found for the Ostreococcus lineage (Robbens et al. 2007). Streamlining of the Pedinomonas mitochondrial genome was severe and was accompanied by extensive genome rearrangements, the gain of a nonstan- dard codon (TGA coding for Trp), and acceleration of the rate of sequence evolution (Turmel, Lemieux, et al. 1999). This evolutionary pattern, also characteristic of some chlorophycean mitochondrial DNAs (mtDNAs) (Gray et al. 1998, 2004), contrasts sharply with the ancestral pattern observed for the mitochondrial genomes of Ostreococcus (Robbens et al. 2007), Nephroselmis (Turmel, Lemieux, et al. 1999), and the chlorellalean Prototheca wickerhamii (Wolff et al. 1994). Whereas the latter mtDNAs range from 44 to 55 kb in size and encode 61–65 genes, the 25-kb mtDNA of Pedinomonas specifies only 22 genes, all of which are tightly packed in a 16-kb segment and encoded by the same strand. Unlike the phylogenies reported here, those inferred from mitochondrial genes revealed that Ped- inomonas and chlorophyceans form a highly supported clade that is sister to the ulvophyte Pseudendoclonium (Pombert et al. 2004). Given the long branch displayed by Pedinomonas and chlorophyceans in the mitochondrial trees, the incongruence of these trees with the chloroplast phylogenies is attributed to long-branch attraction artifacts.

In conclusion, the comparative chloroplast genome analyses presented in this study have unveiled the phylo- genetic position of the Pedinomonadales and illuminated the evolutionary history of the core chlorophytes belonging to the Chlorellales. Our results add support to the notion that the Chlorellales + Pedinomonadales clade emerge early during the radiation of core chlorophytes and highlight further the importance of secondary simplification at both the cellular and genome levels during chlorophyte evolution.

Supplementary Material

The data sets used in phylogenetic analyses are available at Molecular Biology and Evolution online
(http://mbe.oxfordjournals.org/). The fully annotated chloroplast genome sequences of *Oocystis, Pedinomonas*, and *Parachlorella* have been deposited in the GenBank database under the accession numbers FJ968739, FJ968740, and FJ968741, respectively.

Acknowledgments

We thank Jean-François Pombert for his help in annotating the *Pedinomonas* chloroplast genome. This study was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (to M.T. and C.L.).

Literature Cited

Andersen RA, Berges JA, Harrison PJ, Watanabe MM. 2005. Appendix A—recipes for freshwater and seawater media. In: Andersen RA, ed. Algal culturing techniques. Berlin (MA): Elsevier Academic Press. p. 429–538.

Bélanger A-S, Brouard J-S, Charlebois P, Otis C, Lemieux C, Turmel M. 2006. Distinctive architecture of the chloroplast genome in the chlorophycean green alga *Stigeoclonium helveticum*. Mol Gen Genomics. 276:464–477.

Brinkmann H, Philippe H. 2008. Animal phylogeny and large-scale sequencing: progress and pitfalls. J Syst Evol. 46:274–286.

Brouard J-S, Otis C, Lemieux C, Turmel M. 2008. Chloroplast DNA sequence of the green alga *Oedogonium cardiacum* (Chlorophyceae): unique genome architecture, derived characters shared with the Chaetophorales and novel genes acquired through horizontal transfer. BMC Genomics. 9:290.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17:540–552.

Chumley T, Palmer J, Mower J, Fourcade H, Calie P, Boore J, Jansen R. 2006. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 23:2175.

de Cambiaire J-C, Otis C, Lemieux C, Turmel M. 2006. The complete chloroplast genome sequence of the chlorophycean green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. BMC Evol Biol. 6:37.

de Cambiaire J-C, Otis C, Lemieux C, Turmel M. 2007. The chloroplast genome sequence of the green alga *Leptosira terrestris*: multiple losses of the inverted repeat and extensive genome rearrangements within the Trebouxiophyceae. BMC Genomics. 8:213.

de Koning AP, Keeling PJ. 2006. The complete plastid genome sequence of the parasitic green alga *Helmcosporidium sp.* is highly reduced and structured. BMC Biol. 4:12.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet. 6:361–375.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Eliás M, Neustupa J, Škaloud P. 2008. *Elliptochloris bilobata* var. *corticola* var. nov. (Trebouxiophyceae, Chlorophyta), a novel subaerial coccoid green alga. Biologia. 63:791–798.

Fawley MW, Yun Y, Qin M. 2000. Phylogenetic analyses of 18S rDNA sequences reveal a new coccoid lineage of the Prasinophyceae (Chlorophyta). J Phycol. 36:387–393.

Gray MW, Lang BF, Burger G. 2004. Mitochondria of protists. Annu Rev Genet. 38:477–524.

Gray MW, Lang BF, Cedergren R, et al. 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. 26:865–878.

Guillou L, Eikrem W, Chrètiennot-Dinet M-J, Le Gall F, Massana R, Romari K, Pedrós-Alói C, Vaulot D. 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. Protist. 155:193–214.

Heath TA, Heddle SM, Hills DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J Syst Evol. 46:239–257.

Henley W, Hironaka J, Guillou L, Buchheim M, Buchheim J, Fawley M, Fawley K. 2004. Phylogenetic analysis of the ‘Nannochloris-like’ algae and diagnoses of *Pichochlorum oklahomensis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). Phycologia. 43:641–652.

Heperde L, Hegewald E, Krienitz L. 2000. Phylogenetic position of the Oocystaceae (Chlorophyta). J Phycol. 36:590–595.

Jansen RK, Cai Z, Raubeson LA, et al. (16 co-authors). 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA. 104:19369–19374.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 4:18.

Karsten U, Friedl T, Schumann R, Hoyer K, Lembcke S. 2005. Mycosporine-like amino acids and phylogenies in green algae: Prasiola and its relatives from the Trebouxiophyceae (Chlorophyta). J Phycol. 41:557–566.

Krienitz L, Hegewald E, Hepperde L, Huis V, Rohr T, Wolf M. 2004. Phylogenetic relationship of *Chlorella* and *Parachlor-ella* gen. nov. (Chlorophyta, Trebouxiophyceae). Phycologia. 43:529–542.

Krienitz L, Hegewald E, Hepperde L, Wolf A. 2003. The systematics of coccoid green algae: 18S rRNA gene sequence data versus morphology. Biologia. 58:437–446.

Krienitz L, Takeda H, Hepperde L. 1999. Ultrastructure, cell wall composition, and phylogenetic position of *Pseudodictyonema jurissii* (Chlorococcales, Chlorophyta) including a comparison with other picoplanktonic green algae. Phycologia. 38:100–107.

Kugita M, Kaneko A, Yamamoto Y, Takeya Y, Matsumoto T, Yoshinaga K. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res. 31:716–721.

Kurtz S, Choudhuri JV, Ohlebusch E, Schierlremacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of a genomic library. Nucleic Acids Res. 29:4633–4642.

Lartillot N, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl):S4.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol. 21:1095–1109.

Lemieux C, Otis C, Turmel M. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. Nature. 403:649–652.

Lemieux C, Otis C, Turmel M. 2007. A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biol. 5:2.

Lewis LA, McCourt RM. 2004. Green algae and the origin of protists. Annu Rev Genet. 38:477–524.
Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature. 393:162–165.

Mattox KR, Stewart KD. 1984. Classification of the green algae: a concept based on comparative cytology. In: Irvine DEG, John DM, editors. The systematics of the green algae. London: Academic Press. p. 29–72.

Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB. 2002. The Chlamydomonas reinhardtii plastid chromosome: islands of genes in a sea of repeats. Plant Cell. 14:2659–2679.

McCracken DA, Nadakavukaren MJ, Cain JR. 1980. A biochemical and ultrastructural evaluation of the taxonomic position of Glaucotheca vacuolata Korsch. New Phytol. 86:39–44.

Melkonian M. 1990. Chlorophyte orders of uncertain affinities. Order Pedinomonadales. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ, editors. Handbook of protocista. The structure, cultivation, habitats and life histories of the Eukaryotic microorganisms and their descendants exclusive of animals, plants and fungi. Boston (MA): Jones and Bartlett Publishers. p. 649–651.

Michel F, Umeno K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns – a review. Gene. 82:5–30.

Michel F. Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol. 216:585–610.

Moestrup O. 1991. Further studies of presumably primitive green algae including the description of Pedinophyceae class. nov. and Resultor gen. nov. J Phycol. 27:119–133.

Nakayama T, Marin B, Kranz HD, Surek B, Huss VAR, Inouye I, Melkonian M. 1998. The basal position of scaly flagellates among the green algae (Chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences. Protist. 149:367–380.

Oh Yamah K, Fukuzawa H, Kohchi T, et al. (13 co-authors). 1986. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. Nature. 322:572–574.

Palmer JD. 1991. Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil K, editors. The molecular biology of plastids. San Diego (CA): Academic Press. p. 5–53.

Palmer JD, Osorio B, Aldrich J, Thompson WF. 1987. Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr Genet. 11:275–286.

Pickett-Heaps JD, Ott DW. 1974. Ultrastructural morphology and cell division in Pedinomonas minor. Cytobios. 11:41–58.

Pombert J-F, Lemieux C, Turmel M. 2006. The complete chloroplast DNA sequence of the green alga Olmammisellopsis viridis reveals a distinctive quadruplicate architecture in the chloroplast genome of early diverging ulvophytes. BMC Biol. 4:3.

Pombert J-F, Otis C, Lemieux C, Turmel M. 2004. The complete mitochondrial DNA sequence of the green alga Pseudendoclonium akinetum (Ulvophyceae) highlights distinctive evolutionary trends in the Chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. Mol Biol Evol. 21:922–935.

Pombert J-F, Otis C, Lemieux C, Turmel M. 2005. The chloroplast genome sequence of the green alga Pseudendoclonium akinetum (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. Mol Biol Evol. 22:1903–1918.

Pröschold T, Leilaïet F. 2007. Systematics of the green algae: conflict of classic and modern approaches. In: Brodie J, Lewis J, editors. Unravelling the algae: the past, present, and future of algal systematics. Boca Ratón (FL): CRC Press, Taylor & Francis. p. 123–153.

Qiu YL, Li LB, Wang B, et al. (21 co-authors). 2006. The deepest divergences in land plants inferred from phyllogenomic evidence. Proc Natl Acad Sci USA. 103:15511–15516.

Raubeson LA, Jansen RK. 2005. Chloroplast genomes of plants. In: Henry RJ, editor. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. Wallingford (United Kingdom): CABI Publishing. p. 45–68.

Robbens S, Derelle E, Ferraz C, Wuyts J, Moreau H, Van de Peer Y. 2007. The complete chloroplast and mitochondrial DNA sequence of Ostreococcus tauri: organelle genomes of the smallest eukaryote are examples of compaction. Mol Biol Evol. 24:956–968.

Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. Syst Biol. 56:389–399.

Rogers MB, Gilson PR, Su V, McFadden GL, Keeling PJ. 2007. The complete chloroplast genome of the chlorarachniophyte Bigelowiella natans: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbions. Mol Biol Evol. 24:54–62.

Rokas A. 2006. Genomics and the tree of life. Science. 313:1897–1899.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Sluiman H, Guihal C, Mudimu O. 2008. Assessing phylogenetic affinities and species delimitations in Klebsormidiales (streptophyta): nuclear-encoded rDNA phylogenies and its secondary structure models in Klebsormidium, Hormidiella, and Entransia. J Phycol. 44:183–195.

Steinkötter J, Bhattacharya D, Smelmeroth I, Bibeau C, Melkonian M. 1994. Prasinophytes form independent lineages within the Chlorophyta: evidence from ribosomal RNA sequence comparisons. J Phycol. 30:340–345.

Strauss SH, Palmer JD, Howe GT, Doerkens AH. 1988. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. Proc Natl Acad Sci USA. 85:3898–3902.

Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M. 2003. Complete chloroplast DNA sequence of the moss Physcomitrella patens: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. Nucleic Acids Res. 31:5324–5331.

Swafford DL. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Turmel M, Brouard JS, Gagnon C, Otis C, Lemieux C. 2008. Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses. J Phycol. 44:739–750.

Turmel M, Gagnon MC, O’Kelly CJ, Otis C, Lemieux C. 2009. The complete chloroplast DNA sequence of the green alga Nephroselmis olivacea: insights
into the architecture of ancestral chloroplast genomes. Proc Natl Acad Sci USA. 96:10248–10253.

Turmel M, Otis C, Lemieux C. 2002. The chloroplast and mitochondrial genome sequences of the charophyte Chaetosphaeridium globosum: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc Natl Acad Sci USA. 99:11275–11280.

Turmel M, Otis C, Lemieux C. 2005. The complete chloroplast DNA sequences of the charophycean green algae Staurastrum and Zygnema reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. BMC Biol. 3:22.

Turmel M, Otis C, Lemieux C. 2006. The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. Mol Biol Evol. 23:1324–1338.

Verbruggen H, Theriot E. 2008. Building trees of algae: some advances in phylogenetic and evolutionary analysis. Eur J Phycol. 43:229–252.

Wakasugi T, Nagai T, Kapoor M, et al. (15 co-authors). 1997. Complete nucleotide sequence of the chloroplast genome from the green alga Chlorella vulgaris: the existence of genes possibly involved in chloroplast division. Proc Natl Acad Sci USA. 94:5967–5972.

Wolff G, Plante I, Lang BF, Kück U, Burger G. 1994. Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca wickerhamii. J Mol Biol. 237:75–86.

Martin Embley, Associate Editor

Accepted June 29, 2009