The Complete Mitochondrial DNA Sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: Two Radically Different Evolutionary Patterns within Green Algae

Monique Turmel,[a,1] Claude Lemieux,[a] Gertraud Burger,[b] B. Franz Lang,[b] Christian Otis,[a] Isabelle Plante,[b] and Michael W. Gray[c]

[a] Program in Evolutionary Biology, Canadian Institute for Advanced Research, and Département de Biochimie, Université Laval, Québec, Québec G1K 7P4, Canada
[b] Program in Evolutionary Biology, Canadian Institute for Advanced Research, and Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada
[c] Program in Evolutionary Biology, Canadian Institute for Advanced Research, and Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Green plants appear to comprise two sister lineages, Chlorophyta (classes Chlorophyceae, Ulvophyceae, Trebouxiophyceae, and Prasinophyceae) and Streptophyta (Charophyceae and Embryophyta, or land plants). To gain insight into the nature of the ancestral green plant mitochondrial genome, we have sequenced the mitochondrial DNAs (mtDNAs) of *Nephroselmis olivacea* and *Pedinomonas minor*. These two green algae are presumptive members of the Prasinophyceae. This class is thought to include descendants of the earliest diverging green algae. We find that Nephroselmis and Pedinomonas mtDNAs differ markedly in size, gene content, and gene organization. Of the green algal mtDNAs sequenced so far, that of Nephroselmis (45,223 bp) is the most ancestral (minimally diverged) and occupies the phylogenetically most basal position within the Chlorophyta. Its repertoire of 69 genes closely resembles that in the mtDNA of *Prototheca wickerhamii*, a later diverging trebouxiophycean green alga. Three of the Nephroselmis genes (*nad10*, *rpl14*, and *rnpB*) have not been identified in previously sequenced mtDNAs of green algae and land plants. In contrast, the 25,137-bp Pedinomonas mtDNA contains only 22 genes and retains few recognizably ancestral features. In several respects, including gene content and rate of sequence divergence, Pedinomonas mtDNA resembles the reduced mtDNAs of chlamydomonad algae, with which it is robustly affiliated in phylogenetic analyses. Our results confirm the existence of two radically different patterns of mitochondrial genome evolution within the green algae.

INTRODUCTION

Of the three main kingdoms of evolutionarily advanced eukaryotes (animals, fungi, and land plants), plants have the most clearly defined antecedents. Combined molecular, biochemical, and ultrastructural data demonstrate that the unicellular progenitors of land plants lie within the green algae, which together with land plants form a monophyletic lineage that is characterized by the presence of chloroplasts surrounded by two membranes and containing stacked thylakoids and chlorophylls *a* and *b* (Bhattacharya and Medlin, 1998; Chapman et al., 1998). The evolutionary coherence of land plants and green algae is apparent at the levels of both the nuclear and chloroplast genomes; however, at the level of the mitochondrial genome, this affiliation breaks down (see below; Gray et al., 1989).

Phycologists generally recognize five classes of green algae: the Charophyceae, which include the closest relatives of land plants; the Chlorophyceae; the Trebouxiophyceae; the Ulvophyceae; and a nonmonophyletic group known as the Prasinophyceae (reviewed in Friedl, 1997; Bhattacharya and Medlin, 1998; Chapman et al., 1998). The latter class is extremely diverse and is thought to include descendants of the earliest diverging green algae (Melkonian, 1990a). Comparison of nuclear small subunit (SSU) rRNA gene sequences suggests that extant green eukaryotes ("green plants") divide into two major evolutionary lineages: the Streptophyta, containing the charophytes, land plants, and possibly some prasinophytes; and the Chlorophyta, containing the rest of the green algae (Friedl, 1997; Bhattacharya and Medlin, 1998; Chapman et al., 1998). The precise order of divergence of the four classes of algae in the latter lineage has not been resolved. The earliest offshoot of the chlorophyte lineage is considered to be occupied by prasinophytes; from this group probably emerged the Ulvophyceae,

whereas the Trebouxiophyceae and Chlorophyceae appeared later.

Complete mitochondrial DNA (mtDNA) sequences have been determined for four green algae and two land plants, and these data indicate that the mitochondrial genome has followed distinctly different evolutionary pathways in the two major lineages of green plants. In the streptophyte lineage, complete sequence is available for the mtDNA of two representatives of land plants, the liverwort *Marchantia polymorpha* (Oda et al., 1992) and the angiosperm Arabidopsis (Unseld et al., 1997). In the chlorophyte lineage, mitochondrial genome sequence is known for the trebouxiophycean green alga *Prototheca wickerhamii* (Wolff et al., 1994) and for three chlorophycean green algae belonging to the derived family Chlamydomonadales (*Chlamydomonas reinhardtii*, Boer and Gray, 1991; Vahrenholz et al., 1993; *C. eugametos*, Denovan-Wright et al., 1998; and *Chlorogonium elongatum*, Kroymann and Zetsche, 1998). Although the Marchantia and Arabidopsis mtDNAs are three- to sixfold larger (at 186,609 and 366,924 bp, respectively) than Prototheca mtDNA (55,328 bp), the extra size does not reflect an increased coding capacity.

Indeed, only a few additional genes have been identified in the Marchantia and Arabidopsis mitochondrial genomes compared with that of Prototheca (Unseld et al., 1997; Gray et al., 1998). Open reading frames (ORFs) accounting for a small fraction of these two land plant mtDNAs have also been found; however, many of them are unique to one or the other genome and thus may be nonfunctional sequences. Even though Arabidopsis mtDNA is twofold larger than Marchantia mtDNA, it contains fewer genes. The missing ones encode seven tRNAs, seven ribosomal proteins, and two subunits of the succinate:ubiquinone oxidoreductase complex (respiratory complex II). This observation is consistent with the well-documented notion that mitochondrial genes are continuously transferred to the nucleus over evolutionary time (Schuster and Brennicke, 1994) and that land plant mitochondrial genomes are prone to expansion by enlargement of intergenic spacers (Ward et al., 1981; Palmer, 1990). Novel mtDNA sequences can be generated through various processes, for example, duplications as well as integration of both chloroplast and nuclear sequences (Unseld et al., 1997).

In contrast, the mtDNAs of green algae from the chlorophyte lineage appear to be constrained to remain tightly packed with genes. Remarkably, a subgroup of these green algae bears mtDNAs that are severely reduced both in size and gene content and feature an unusual rRNA gene organization. The *C. reinhardtii* (15,758 bp) and *C. eugametos* (22,897 bp) mitochondrial genomes are twofold smaller than their Prototheca homolog and contain only 12 genes, that is, 49 genes fewer than in the latter green alga. Considering that *C. reinhardtii* and *C. eugametos* represent the two main clades that were identified in the large and nonmonophyletic genus *Chlamydomonas* (Buchheim et al., 1996), their shared unusual mtDNA features undoubtedly represent char-

acteristics that appeared before the emergence of this group.

In both *C. reinhardtii* and *C. eugametos*, the mitochondrial large subunit (LSU) and SSU rRNA genes are fragmented into subgenic modules that are scrambled and interspersed with other mitochondrial genes instead of being continuous, as in almost all other mtDNAs examined to date. This exceptional gene organization has also been observed in the chlamydomonad alga *Chlorogonium elongatum* (Kroymann and Zetsche, 1998), a member of the clade represented by *Chlamydomonas eugametos* (Buchheim et al., 1996). Another distinctive feature of chlamydomonad mtDNAs is the accelerated rate of sequence divergence of their genes, manifested as very long branches in phylogenetic trees. Large differences can lead to a "long branches attract" artifact (Felsenstein, 1988), which most probably explains why chlamydomonad mitochondrial gene sequences consistently fail to form a monophyletic group with those of Prototheca and land plants in phylogenetic analyses (Gray, 1995; Gray and Spencer, 1996).

Thus, available data indicate the existence of distinct patterns of mitochondrial genome evolution not only between Streptophyta and Chlorophyta but within Chlorophyta as well, raising a number of questions. What accounts for the radical differences observed to date in the evolutionary pathways followed by the mitochondrial genome within the chlorophyte algae? How do we explain the markedly distinct patterns of mtDNA organization between the chlorophyte and streptophyte lineages? How widespread are these diverse patterns, and do other types exist? What evolutionary mechanisms underlie these differences? What did the mitochondrial genome of the common ancestor of all green algae and land plants look like? Answers to these questions can only be obtained by examining the mitochondrial genome in a broader range of green algae.

In this study, we sought to gain insight into the nature of the ancestral green algal mitochondrial genome by undertaking the complete sequencing of the mtDNAs of two presumed prasinophytes, *Nephroselmis olivacea* and *Pedinomonas minor*. Our results demonstrate that these two mitochondrial genomes differ radically in gene content and organization, falling into one of the two chlorophyte patterns previously described. These data, together with an exhaustive phylogenetic analysis of mitochondrial protein-coding sequences, further illuminate the evolution of green algal mtDNA and the interrelationships of the relevant green algae themselves.

RESULTS

The Mitochondrial Genome of Nephroselmis

Figure 1 shows the structure and gene map of Nephroselmis mtDNA, the sequence of which assembles as a circle of

45,223 bp, with an overall A+T content of 67.2%. More than 78% of the sequence has a recognized coding role, including 63 genes of known function, two conserved ORFs (*ymf16* and *ymf39*), and four intron ORFs, all present in single copy (see Tables 1 and 2). Also present is an ORF (*orf220*) encoding a protein of 220 amino acids that is not obviously homologous to any known protein. In Tables 1 and 2, the gene content of Nephroselmis mtDNA is compared with those of mtDNAs from land plants and other green algae. It can be seen that the gene repertoire of Nephroselmis mtDNA closely matches that of Prototheca mtDNA. If tRNA genes, intron-encoded genes, and unique ORFs are not considered, then there are only four additional genes in Nephroselmis mtDNA relative to its Prototheca homolog: two ribosomal protein genes (*rps8* and *rpl14*), a respiratory chain protein gene (*nad10*), and a gene encoding the RNA component of RNase P (*rnpB*).

Despite their similarity in gene content, the mtDNAs of Nephroselmis and Prototheca differ markedly in gene ar-
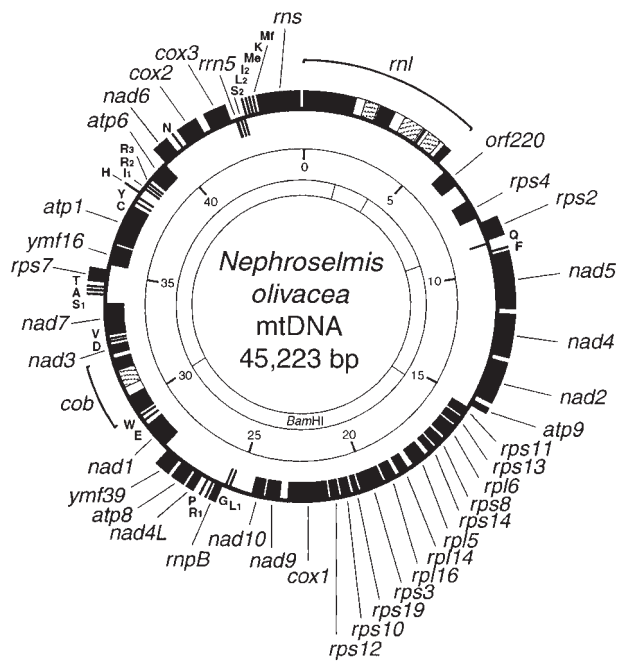

Figure 1. Gene Map of Nephroselmis mtDNA.

Genes (filled rectangles) shown on the outside of the circle are transcribed in a clockwise direction, whereas those on the inside of the circle are transcribed counterclockwise. The innermost circles show a size scale (in kb) and the BamHI restriction map. Transfer RNA genes are indicated by the one-letter amino acid code (Me, elongator methionine; Mf, initiator methionine), with subscripts denoting different genes specific for the same amino acid. The anticodons of the tRNA genes indicated by subscripts are as follows: I_1, gau; I_2, cau; L_1, uaa; L_2, uag; R_1, ucu; R_2, ucg; R_3, acg; S_1, gcu; and S_2, uga. Four group I introns (open rectangles), each containing an ORF (hatched rectangle), were identified.

rangement and in the distribution of genes between the two strands. In Prototheca mtDNA, all contiguous genes in almost half of the genome are encoded by the same strand, whereas those in the rest of the genome are present on the complementary strand. This distribution suggests that there could be as few as two transcription units in Prototheca mtDNA. In contrast, at least 14 potential transcriptional units are predicted from the distribution of genes in the Nephroselmis mitochondrial genome, assuming completely asymmetric transcription of the genome.

Like its homologs in Prototheca, the Nephroselmis *rns* and *rnl* genes have a conventional structure. The products of these two genes, the SSU and LSU rRNAs, respectively, as well as the 5S RNA and the RNase P RNA, fold into potential secondary structures that to a high degree resemble those of their counterparts in eubacteria. Figure 2 shows the secondary structure model of the Nephroselmis RNase P RNA (352 nucleotides long). This model displays most of the conserved primary and secondary structural motifs of the consensus eubacterial RNase P RNA model (Siegel et al., 1996). In this respect, Nephroselmis mitochondrial RNase P RNA closely resembles the 311-nucleotide-long mitochondrial RNase P RNA of the heterotrophic jakobid flagellate *Reclinomonas americana* (Lang et al., 1997).

The 26 identified tRNA genes encode products with conventional cloverleaf secondary structures, and this complement is sufficient to decode the entire set of codons found in Nephroselmis mtDNA. This is similar to the situation in Prototheca mitochondria but differs from what is observed in *Chlamydomonas* spp mtDNAs. Although the mitochondrial tRNA complement of Prototheca also comprises 26 species, two tRNAs (tRNA[Gly] [gcc] and tRNA[Trp] [ugu]) are not encoded by Nephroselmis mtDNA (Table 2). The two Nephroselmis mitochondrial tRNAs having no equivalents in Prototheca are tRNA[Arg] (ucg) and tRNA[Trp] (ggu).

Each of the four introns in Nephroselmis mtDNA belongs to the group I family and contains an ORF potentially encoding a specific DNA endonuclease bearing the LAGLIDADG motif. Three introns reside within the *rnl* gene at sites corresponding to residues 1931/1932, 2500/2501, and 2593/2594 of the *Escherichia coli* LSU rRNA sequence. Interestingly, ORF-containing introns similar to the Nephroselmis site 1931 and site 2593 introns are located at exactly the same positions in the mitochondrial *rnl* gene of the nonphotosynthetic protist *Acanthamoeba castellanii* (Lonergan and Gray, 1994) as well in the chloroplast DNAs of various green algae (Rochaix et al., 1985; Turmel et al., 1995; Wakasugi et al., 1997; M. Turmel, C. Otis, and C. Lemieux, unpublished results). These observations raise the possibilities that these introns were vertically inherited from a green algal/Acanthamoeba common ancestor and that they were transferred between the chloroplast and mitochondrial compartments (see Turmel et al., 1995). Similarly, striking sequence similarities are apparent between the Nephroselmis site 2500 intron and the *Chlamydomonas humicola* chloroplast *rnl* intron (Côté et al., 1993) inserted at the identical position. For each

Table 1. Comparison of Gene Content in Completely Sequenced Land Plant and Green Algal Mitochondrial Genomes

Gene/Genetic Element	Ath[a]	Mpo	Pwi	Nol	Pmi	Cre	Ceu	Gene/Genetic Element	Ath[a]	Mpo	Pwi	Nol	Pmi	Cre	Ceu
Complex I (*nad*)								SSU ribosomal proteins (*rps*)							
1	+	+	+	+	+	+	+	1	−	+	−	−	−	−	−
2	+	+	+	+	+	+	+	2	−	+	+	+	−	−	−
3	+	+	+	+	+	−	−	3	+	+	+	+	−	−	−
4	+	+	+	+	+	+	+	4	+	+	+	+	−	−	−
4L	+	+	+	+	+	−	−	7	+	+	+	+	−	−	−
5	+	+	+	+	+	+	+	8	−	+	−	+	−	−	−
6	+	+	+	+	+	+	+	10	−	+	+	+	−	−	−
7	+	Ψ[b]	+	+	−	−	−	11	−	+	+	+	−	−	−
9	+	+	+	+	−	−	−	12	+	+	+	+	−	−	−
10[c]	−	−	−	+	−	−	−	13	−	+	+	+	−	−	−
Complex II (*sdh*)								14	Ψ[b]	+	+	+	−	−	−
3	−	+	−	−	−	−	−	19	Ψ[b]	+	+	+	−	−	−
4	−	+	−	−	−	−	−	LSU ribosomal proteins (*rpl*)							
Complex III (*cob*)	+	+	+	+	+	+	+	2	+	+	−	−	−	−	−
Complex IV (*cox*)								5	+	+	+	+	−	−	−
1	+	+	+	+	+	+	+	6	−	+	+	+	−	−	−
2	+	+	+	+	−	−	−	14	−	−	−	+	−	−	−
3	+	+	+	+	−	−	−	16	+	+	+	+	−	−	−
Complex V (*atp*)								Unidentified conserved ORFs							
1	+	+	+	+	−	−	−	*ymf16*[g]	+	+	+	+	−	−	−
6	+	+	+	+	+	−	−	*ymf39*[h]	+	+	+	+	−	−	−
8	+	+	+	+	+	−	−	Ribosomal RNAs							
9	+	+	+	+	−	−	−	*rns*	+	+	+	+	+	+[i]	+[i]
Cytochrome *c* biogenesis (*yej*)								*rnl*	+	+	+	+	+[i]	+[i]	+[i]
R[d]	+	+	−	−	−	−	−	*rrn5*	+	+	+	+	−	−	−
U[e]	+	+	−	−	−	−	−	RNase P RNA (*rnpB*)	−	−	−	+	−	−	−
V[f]	+	+	−	−	−	−	−	Transfer RNAs (*trn*)[j]	17	27	26	26	8	3	3
								Introns (group I/group II)	0/23	7/25	5/0	4/0	0/1	0/0	9/0
								Intronic ORFs (group I/group II)	0/1	2/8	2/0	4/0	0/0	0/0	7/0

[a] Ath, *A. thaliana*; Mpo, *M. polymorpha*; Pwi, *P. wickerhamii*; Nol, *N. olivacea*; Pmi, *P. minor*; Cre, *C. reinhardtii*; Ceu, *C. eugametos*. Presence of a gene is denoted by a (+) and its absence by a (−).

[b] Pseudogene.

[c] Gene incorrectly listed as *nad8* by Gray et al. (1998).

[d] Gene is split into three separate ORFs in both Mpo (*orf509* or *ymf4*, *orf169* or *ymf3*, and *orf322* or *ymf2*) and Ath (*ccb382*, *ccb203*, and *ccb452*; Unseld et al., 1997). Mpo *orf509* corresponds to Ath *ccb382* and *ccb203*, and Ath *ccb452* corresponds to Mpo *orf169* and *orf322* (Unseld et al., 1997).

[e] Mpo *orf228* or *ymf5* and Ath *ccb256* (Unseld et al., 1997).

[f] Mpo *orf277* or *ymf6* and Ath *ccb206* (Unseld et al., 1997).

[g] Mpo *orf244*, recently identified as a homolog of *E. coli mttB*, a gene encoding a protein involved in a Sec-independent protein translocation pathway (Weiner et al., 1998).

[h] Mpo *orf183* (*orf25* in angiosperms).

[i] Split and rearranged rRNA genes give rise to fragmented rRNAs (Boer and Gray, 1988a; Denovan-Wright and Lee, 1994; this study).

[j] Unique genes (not counting duplicates).

of the three Nephroselmis mitochondrial *rnl* introns, there is evidence that a homologous chloroplast ORF codes for an endonuclease (I-CreI, Dürrenberger and Rochaix, 1991; I-ChuI, Côté et al., 1993; and I-CpaII, Turmel et al., 1995).

The fourth Nephroselmis mitochondrial intron is inserted in the *cob* gene at the same position as the unique intron found in the *Chlamydomonas smithii cob* gene (Colleaux et al., 1990). This *Chlamydomonas* sp intron shares with its Nephroselmis homolog a similar secondary structure (Colleaux et al., 1990) as well as an ORF encoding a specific DNA endonuclease (Ma et al., 1992).

The Mitochondrial Genome of Pedinomonas

The Pedinomonas mitochondrial genome sequence assembles into a circle of only 25,137 bp, with an overall A+T content of 77.8% (Figure 3). This mtDNA is peculiar in both its

Table 2. Comparison of Transfer RNA Gene Content in Completely Sequenced Land Plant and Green Algal Mitochondrial Genomes

Amino Acid	Anticodon	Codons	Ath[a]	Mpo	Pwi	Nol	Pmi	Cre
A	ugc	GCN	−	+	+	+	−	−
C	gca	UGY	+	+	+	+	+	−
D	guc	GAY	+	+	+	+	−	−
E	uuc	GAR	−	+	+	+	+	−
F	gaa	UUY	+	+	+	+	+	−
G	gcc	GGY	+	+	+	−	−	−
G	ucc	GGN	−	+	+	+	−	−
H	gug	CAY	+	+	+	+	+	−
I	cau[b]	AUA	+	+	+	+	−	−
I	gau	AUY	−	−	+	+	−	−
K	uuu	AAR	+[c]	+	+	+	−	−
L	caa	UUG	−	+	−	−	+	−
L	uaa	UUR	−	+	+	+	−	−
L	uag	CUN	−	+	+	+	−	−
Me	cau	AUG	+	+	+	+	−	d
Mf	cau	AUG	+	+	+	+	−	d
N	guu	AAY	+	+	+	+	−	−
P	ugg	CCN	+	+	+	+	−	−
Q	uug	CAR	+	+	+	+	+	+
R	acg[e]	CGN	−	+	+	+	−	−
R	ucg	CGR	−	+	−	+	−	−
R	ucu	AGR	−	+	+	+	−	−
S	gcu	AGY	+[c]	+	+	+	−	−
S	uga	UCN	+[c]	+	+	+	−	−
S	gga	UCY	+	−	−	−	−	−
T	ggu	ACY	−	+	−	+	−	−
T	ugu	ACN	−	−	+	−	−	−
V	uac	GUN	−	+	+	+	−	−
W	cca	UGG	+	+	+	+	−	+
W	uca	UGR	−	−	−	−	+	−
Y	gua	UAY	+[f]	+	+	+	+[c]	−
Unique *trn* genes			17	27	26	26	8	3

[a] See Table 1 for abbreviations of organism names. Presence of a gene is denoted by a (+) and its absence by a (−).
[b] C assumed to be post-transcriptionally modified to lysidine, which pairs with A rather than G.
[c] Duplicated gene.
[d] Genome specifies a single *trnM*(cau).
[e] A assumed to be post-transcriptionally modified to inosine, which pairs with A, G, C, and U.
[f] Two identical and one nonidentical copies.

high proportion of noncoding sequence (almost 40%) and its highly asymmetric distribution of coding and noncoding regions. The 22 genes identified in Pedinomonas mtDNA are tightly packed within a 16-kb segment, a region whose size approximates that of *C. reinhardtii* mtDNA. All 22 genes are encoded by the same DNA strand, as is also the case for *C. eugametos* mtDNA (Figure 3).

As shown in Tables 1 and 2, Pedinomonas mtDNA shares with *C. reinhardtii* and *C. eugametos* mtDNAs five *nad* genes, *cox1*, *cob*, *rns*, *rnl*, and two tRNA genes. In addition, the Pedinomonas mitochondrial genome encodes two *atp*

genes (*atp6* and *atp8*), two extra *nad* genes (*nad3* and *nad4L*), and six more tRNA genes. Pedinomonas mtDNA thus resembles its *Chlamydomonas* spp homologs in harboring (1) a reduced set of genes encoding proteins involved in respiration, electron transport, and oxidative phosphorylation; (2) the absence of any 5S rRNA and ribosomal protein genes; (3) a limited number of tRNA genes; and (4) a fragmented and rearranged *rnl* gene (Tables 1 and 2). Particularly noteworthy is the absence of *cox2* and *cox3* from the two *Chlamydomonas* spp and Pedinomonas mtDNAs. Both *cox2* and *cox3* are encoded in all but three of 23 other completely sequenced protist mtDNAs, whereas, like their Pedinomonas and *Chlamydomonas* spp counterparts, most of these mtDNAs lack the 5S rRNA gene and a number of ribosomal protein genes (Gray et al., 1998).
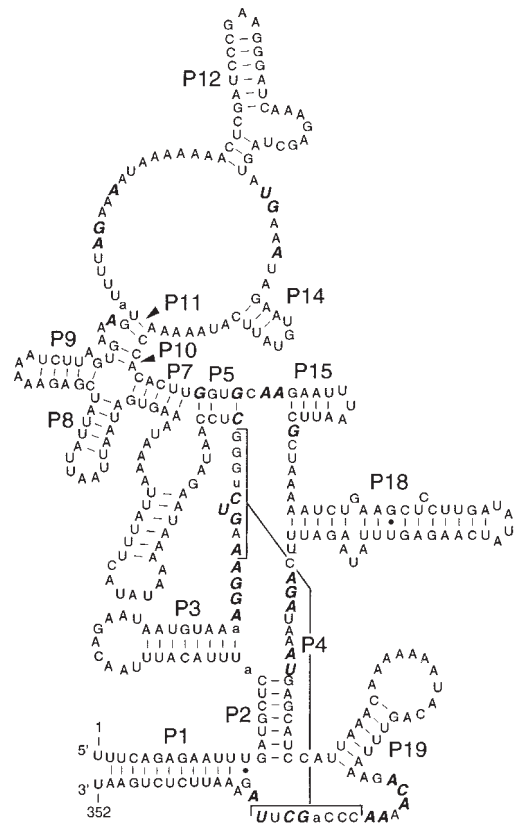

Figure 2. Secondary Structure Model of Nephroselmis RNase P RNA.

The model is based on the secondary structure of the *E. coli* RNase P RNA, and helical regions (indicated by arrowheads in the case of P10 and P11) are numbered accordingly (Siegel et al., 1996). Bars denote G-C or A-U base pairs, and dots denote G-U pairs. The residues participating in the long-range P4 pairing are indicated by brackets connected with a line. The residues in boldface and italics match the bacterial consensus.
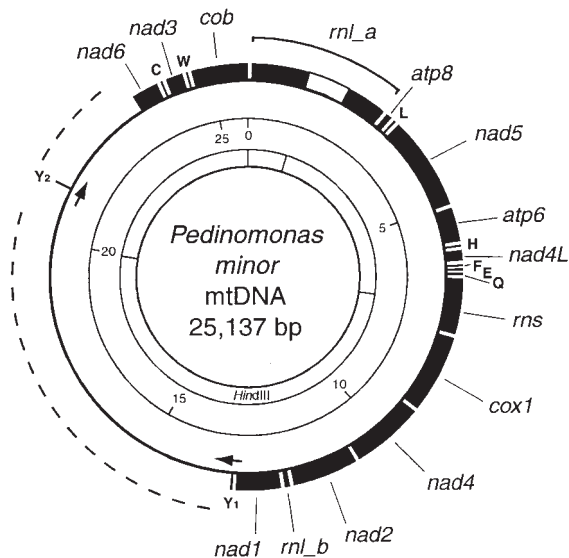
Figure 3. Gene Map of Pedinomonas mtDNA.

Genes (filled rectangles) shown on the outside of the outermost circle are all transcribed in a clockwise direction. The innermost circles show a size scale (in kilobases) and the HindIII restriction map. Transfer RNA genes are indicated by the one-letter amino acid code. The *trnY₁* and *trnY₂* genes are nonidentical versions (differing in sequence at two positions) of the same tRNA gene. The *rnl* gene is split into two modules, *rnl_a* and *rnl_b*. A single group II intron (open rectangle) is located in *rnl_a*. The arrows denote two short direct repeats (ele-13) that contain the *trnY* genes and the 3'-terminal portion of *nad1*. The dashed lines denote two larger repeat regions composed of many repetitive elements.

The eight tRNAs encoded by Pedinomonas mtDNA have conventional structures and are presumably supplemented by imported, nuclear DNA-encoded tRNAs to permit translation of all codons in protein-coding genes. Codon usage in Pedinomonas mtDNA is biased and deviates slightly from the standard genetic code in that UGA is decoded as Trp. This deviation is consistent with the presence of a tRNA^{Trp} having a UCA anticodon, which could recognize both UGG and UGA codons. Except in five species of chlorophycean green algae, in which TAG in mitochondrial protein genes may code for either Ala or Leu (Hayashi-Ishimaru et al., 1996), the standard genetic code is used in all green algal mtDNAs investigated so far. Codon usage in Pedinomonas mtDNA is biased in a different way than in *Chlamydomonas* spp mtDNAs. Five codons (Leu, CTG; Pro, CCG; Arg, CGG; Thr, ACG; and Ala, GCG) are entirely absent from Pedinomonas protein-coding genes, whereas eight sense codons (Leu, TTA; Ser, TCA and TCG; Thr, ACG; Glu, GAA; and Arg, CGG, AGA, and AGG), in addition to TGA (stop), are missing from *C. reinhardtii* mtDNA (Boer and Gray, 1988b). Only two of the codons absent from Pedinomonas mtDNA (CGG and ACG) are also missing from *C. reinhardtii* mtDNA. Two codons that are absent from Pedinomonas mtDNA (Pro, CCG; and Ala, GCG) are used infrequently in *C. reinhardtii* mtDNA, whereas several of the missing *C. reinhardtii* codons (Leu, TTA; Ser, TCA; Glu, GAA; and Arg, AGA) are abundant Pedinomonas codons.

The Pedinomonas LSU rRNA is fragmented at a single site, situated within a variable region separating domains V and VI in the secondary structure. Breaks have not been found previously at this position in other organellar LSU rRNAs, but they have been observed at approximately the same place in trypanosomatid and *Euglena gracilis* nucleocytoplasmic LSU rRNAs (Gray and Schnare, 1996). In Pedinomonas mtDNA, a number of protein-coding and tRNA genes lie between the rRNA coding region for domain VI (*rnl_b*) and the gene specifying the remainder of the LSU rRNA (*rnl_a*) (Figure 3). More extensive fragmentation and rearrangement of rRNA genes are seen in *Chlamydomonas* spp mtDNAs, with SSU and LSU rRNA genes divided into six and at least eight dispersed gene pieces, respectively, in *C. reinhardtii* mtDNA (Boer and Gray, 1988a) and into three and six coding modules, respectively, in *C. eugametos* mtDNA (Denovan-Wright and Lee, 1994). Although their sequences have diverged extensively relative to those of most protist mitochondrial rRNAs, the Pedinomonas mitochondrial SSU and LSU rRNAs are similar to their *Chlamydomonas* spp homologs in having secondary structures that contain the functionally most essential structural elements found in eubacterial rRNAs.

A single intron (group II and lacking an ORF) is inserted within the Pedinomonas *rnl_a* gene at precisely the same site (between positions corresponding to 1787 and 1788 in the *E. coli* secondary structure) as a group II intron in the mitochondrial *rnl* gene of the brown alga *Pylaiella littoralis* (Fontaine et al., 1995). In *Chlamydomonas* spp mtDNAs, the number of introns is extremely variable, and all those reported thus far are members of the group I family (Colleaux et al., 1990; Turmel et al., 1993; Denovan-Wright et al., 1998; Kroymann and Zetsche, 1998).

Pedinomonas mtDNA contains a 9-kb-long region of repeated sequences, located between *nad1* and *nad6*. The primary repeat elements range in size from 6 to 389 bp and comprise 13 distinct families (ele-01 to ele-13) based on sequence relatedness. Although most members within an element family are identical, a few are distinguished by sequence differences of up to 25%. The repeat elements either occur as simple duplications or are integrated into an elaborate superstructure in which primary elements from different families form second-order, and those again third-order, repeat units. An example of a simple duplication is ele-13 (311 bp), which includes a *trnY* gene and the 3'-terminal portion of *nad1* (Figure 3). One copy of ele-13 thus slightly overlaps *nad1*, whereas a second copy sharing 87% sequence identity with the first is found (in the same orientation) 7139 bp downstream of *nad1*. More complex arrangements are present in the region located between the two copies of ele-13 and in a 1926-bp region between the second copy of ele-13 and *nad6*. Nine distinct second-order

elements can be distinguished in these regions; they are characterized by combinations of from three to five primary repeat members that are always arranged in tandem orientation and that contain in their central portion the 6-bp recognition sequence of HpaI. Three of these second-order elements form a third-order unit in which two of the constituent elements are present in reverse orientation. It is unclear what, if any, function these repeat structures have in Pedinomonas mtDNA.

Phylogenetic Analyses of Nephroselmis and Pedinomonas Mitochondrial Protein Sequences

To determine the phylogenetic positions of Nephroselmis and Pedinomonas mitochondria relative to other green algal and land plant mitochondria, we used PROTML (Adachi and Hasegawa, 1996a) to analyze data sets containing concatenated sequences of multiple mitochondrial proteins. All analyses revealed that Pedinomonas and *Chlamydomonas* spp mitochondrial sequences are strongly allied, but their positions are anomalous in showing no connection at all with the green algal/land plant clade. In contrast, Nephroselmis is clearly seen to be a part of the latter clade and to be a member of the chlorophyte lineage.

Figure 4A shows a mitochondrial tree featuring two red algae, *Porphyra purpurea* and *Chondrus crispus*, and two fungi, *Podospora anserina* and *Allomyces macrogynus*, in addition to the green algae and land plants whose complete mtDNA sequences were available at the time of the analysis. The topology depicted is that of the best tree; it is supported with a frequency of 0.79 in RELL bootstrap samples. It can be seen that Nephroselmis clusters robustly with the trebouxiophyte Prototheca, whereas Marchantia and Arabidopsis represent the sister group of this clade. The two red algae form an independent cluster, the sister group of the green algal/land plant clade. Pedinomonas, *C. reinhardtii*, and *C. eugametos* are found entirely outside of the green algal/land plant/red algal clade, forming a strongly supported cluster whose remarkably long branches undoubtedly reflect the highly accelerated rate of sequence divergence of Pedinomonas and *Chlamydomonas* spp mtDNAs relative to other protist and land plant mtDNAs. This accelerated rate of evolution is clearly evident in protein sequence alignments: whereas the Pedinomonas and *Chlamydomonas* spp sequences differ by multiple substitutions at several sites, the other sequences differ mostly by single substitutions (data not shown). Removal of either the two *Chlamydomonas* spp or Pedinomonas from the data set did not alter the position of Pedinomonas or that of *Chlamydomonas* spp relative to other green algae and land plants (data not shown), and neither did the addition of mitochondrial protein sequences from the prasinophyte *Tetraselmis subcordiformis* (Figure 4B), ~40% of whose 42.8-kb mtDNA has been sequenced (Kessler and Zetsche, 1995). In the best tree shown in Figure 4B (supported with a frequency of 0.34 in RELL bootstrap

samples), *T. subcordiformis* mitochondrial sequences cluster with those of Prototheca, a position that is in agreement with a chloroplast phylogeny that includes *T. carteriiformis* (see Figure 4D), a very close relative of *T. subcordiformis*.

Figures 4C and 4D show chloroplast phylogenies that were inferred from *rnl* gene sequences by using a maximum likelihood approach. In contrast to the mitochondrial trees presented above, these chloroplast phylogenies reveal that Pedinomonas and *Chlamydomonas* spp cluster with Prototheca, Tetraselmis, and Nephroselmis. This clade of chlorophytes, the sister group of the land plant clade (formed by Marchantia and tobacco), is strongly supported, with bootstrap values of 94 and 95%.

DISCUSSION

Two Distinct Patterns of Mitochondrial Genome Evolution within the Chlorophyte Algae

Our analyses of the gene content and organization of the Nephroselmis and Pedinomonas mtDNAs as well as our phylogenetic analyses strongly reinforce the notion that the mitochondrial genomes from basal and derived lineages of the chlorophyte phylum followed radically different patterns of evolution. We designate these patterns "ancestral" (Prototheca-like) and "reduced derived" (Chlamydomonas-like), respectively. As detailed below, the mitochondrial genomes from basal lineages, such as those occupied by Nephroselmis and Prototheca, have retained many features inherited from their prokaryotic ancestors, such as the presence of a 5S rRNA gene and other "extra" genes compared with *Chlamydomonas* spp mtDNAs, an almost complete set of tRNAs, a few introns, and eubacteria-like gene clusters. Moreover, their sequences appear to evolve at a relatively slow rate, allowing them to associate with their land plant counterparts in phylogenetic analyses. In contrast, the genomes from the more highly derived lineages containing chlamydomonads and Pedinomonas have retained only a few recognizable ancestral features. Compared with their ancestral counterparts, they are characterized by extensive gene loss, radical departure from conventional rDNA organization and/or rRNA structure, and by the introduction of nonstandard codons (e.g., TGA coding for Trp in Pedinomonas). Their sequences evolve at such an accelerated rate that they cannot be connected with their green algal and land plant homologs in phylogenetic studies.

Nephroselmis mtDNA Represents the Most Ancestral Form of the Mitochondrial Genome within the Green Algae

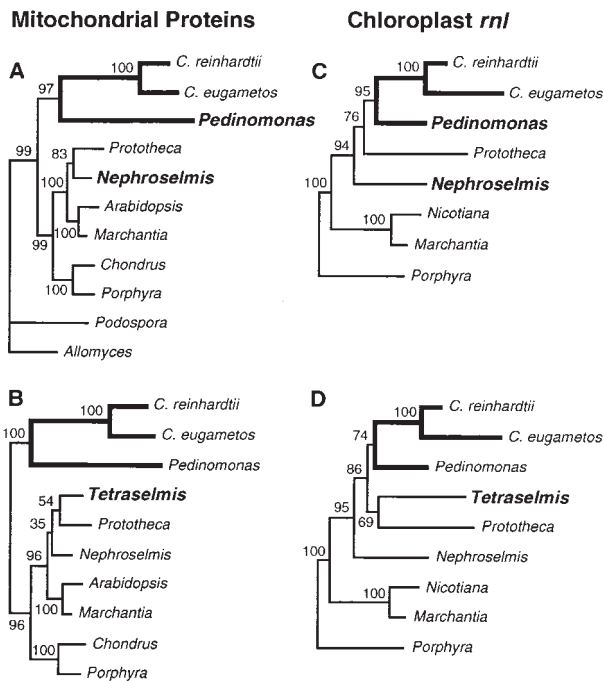Of the green algal mitochondrial genomes that have been analyzed so far, Nephroselmis mtDNA displays the largest

Mitochondrial Proteins   ## Chloroplast *rnl*

Figure 4. Phylogenetic Positions of Nephroselmis and Pedinomonas Deduced from Comparative Analyses of Protein and DNA Sequences.

Concatenated mitochondrial protein sequences from various organisms were analyzed using PROTML (Adachi and Hasegawa, 1996a), whereas chloroplast *rnl* sequences were analyzed using maximum likelihood and the Hasegawa-Kishino-Yano (Hasegawa et al., 1985) model of evolution with rate heterogeneity. Bootstrap values are indicated at the nodes of each tree.
(A) Best tree derived from a phylogenetic analysis of concatenated mitochondrial protein sequences (Cob, Cox1, Nad1, Nad2, Nad4, Nad5, and Nad6) from Nephroselmis, Pedinomonas, Prototheca, *C. reinhardtii*, *C. eugametos*, Marchantia, Arabidopsis, two red algae (Chondrus and Porphyra), and two fungi (Podospora and Allomyces). The data set consisted of 1790 amino acid positions.
(B) Best tree derived from a phylogenetic analysis of concatenated mitochondrial protein sequences (Cob, Cox1, and Nad5) from all organisms examined in **(A)**, except that *T. subcordiformis* (Tetraselmis) was included and the two fungi were omitted. The data set consisted of 1162 amino acid positions.
(C) Majority-rule consensus tree derived from a phylogenetic analysis of chloroplast *rnl* sequences from all five green algae examined in **(A)** and from Marchantia, tobacco (*Nicotiana tabacum*), and Porphyra. The data set consisted of 2025 nucleotide positions.
(D) Majority-rule consensus tree derived from a phylogenetic analysis of chloroplast *rnl* sequences from all organisms examined in **(C)**, except that *T. carteriiformis* (Tetraselmis) was included. As for **(C)**, the data set consisted of 2025 nucleotide positions.
The thick lines in each tree highlight the Pedinomonas/ *Chlamydomonas* spp clade. GenBank accession numbers for mtDNA sequences are as follows: Nephroselmis (AF110138), Pedinomonas (AF116775), *C. reinhardtii* (U03843), *C. eugametos* (AF008237), Prototheca (U02970), *T. subcordiformis* (Z47795, Z47796, Z47797), Marchantia (M68929), Arabidopsis (Y08501, Y08502), Porphyra

number of ancestral features, including the highest coding capacity. Moreover, it occupies the most basal position in the chlorophyte lineage, as revealed by phylogenetic analyses of chloroplast *rnl* sequences (Figures 4C and 4D).

Three of the 69 genes encoded by Nephroselmis mtDNA (*nad10*, *rpl14*, and *rnpB*) had not been identified in the mitochondrial genomes of the other green algae and land plants examined to date. The finding of *nad10* and *rnpB* in Nephroselmis mtDNA was particularly unexpected, because these genes are not widely distributed among mitochondrial genomes. The *nad10* gene is known to be encoded by the mtDNAs of the cryptomonad alga *Rhodomonas salina*, the ciliates *Tetrahymena pyriformis* and *Paramecium aurelia*, and the heterotrophic flagellate Reclinomonas (Gray et al., 1998). The *rnpB* gene has been found in the mitochondrial genomes of the latter flagellate and a number of fungi (Martin and Lang, 1997). *rnpB* is highly variable in sequence and thus difficult to detect in sequence similarity searches; therefore, it may be more widely distributed among protist mtDNAs than is currently appreciated.

Although the gene content of Nephroselmis mtDNA barely exceeds that of Prototheca mtDNA, its organizational pattern is clearly more ancestral. Nephroselmis mtDNA shares five gene clusters with Reclinomonas mtDNA, a minimally derived mitochondrial genome encoding 97 genes and representing the most ancestral form of mtDNA among protists (Lang et al., 1997). These shared clusters are as follows: a group of 12 contiguous ribosomal protein genes, from *rps12* to *rps11*; *nad5-nad4-nad2*; *nad10-nad9*; *rns-rnl*; and *trnR*(ucu)-*rnpB*/*trnG*(ucc), in which *trnR* and *rnpB* are present on the same strand and *trnG* is on the opposite strand. Moreover, Nephroselmis mtDNA shares the *cox2-cox3* linkage with a number of mtDNAs having a pattern more derived than Reclinomonas mtDNA. Prototheca mtDNA exhibits portions of two conserved Nephroselmis gene clusters, the *nad5* cluster and the ribosomal protein cluster (see Figure 5). Because the *ymf39-atp8* pair is the only linkage group specifically shared between the Nephroselmis and Prototheca mtDNAs, it most probably traces back to the most recent common ancestor of these two green algae.

Ribosomal protein genes in Nephroselmis mtDNA are arrayed in a fashion mirroring the order of the homologous genes in the contiguous *str*, S10, *spc*, and α operons of *E. coli* (Figure 5). The same order is seen not only in Reclino-
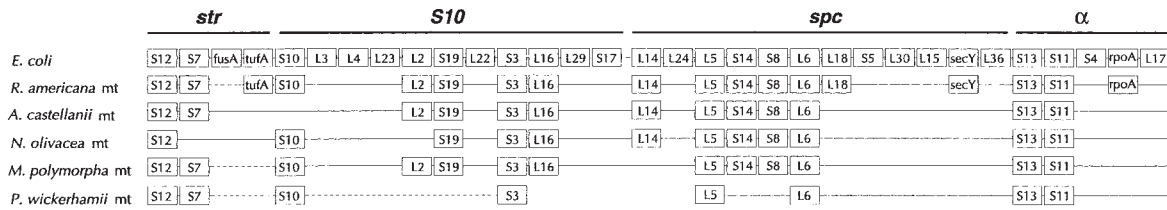
Figure 5. *E. coli*–like Arrangement of Ribosomal Protein Genes in Nephroselmis mtDNA and Other mtDNAs.

Conservation of ribosomal protein gene organization in several mitochondrial (mt) genomes is compared with the contiguous *str*, S10, *spc*, and α operons of *E. coli*. Solid lines connect adjacent genes, whereas dashed lines indicate the presence of additional genes that are not shown. L, *rpl*; S, *rps*.

monas mtDNA but also in mtDNAs of Marchantia and Acanthamoeba, testifying to the eubacterial character of the ancestral protomitochondrial genome. The fact that a number of the same gene deletions (relative to the *E. coli* pattern) are shared among these several mtDNAs (Figure 5) strongly suggests that these mitochondrial genomes have all descended from a common ancestor in which these specific deletions had already taken place. Such data reinforce the notion of a single (monophyletic) origin of mitochondria.

Like the ribosomal protein gene cluster, the Nephroselmis mitochondrial gene clusters *nad5-nad4-nad2*, *nad10-nad9*, and *cox2-cox3* have equivalents in a number of mtDNAs as well as in bacterial genomes; thus, they undoubtedly represent vestiges of prokaryotic operon organization. On the other hand, the *trnR*(ucu)-*rnpB*/*trnG*(ucc) and *rns-rnl* clusters have been described so far only in Nephroselmis and Reclinomonas mtDNAs; consequently, they may well represent derived characters of the protomitochondrial genome that have been conserved in these two mtDNAs but lost in most mtDNAs analyzed to date. Although the *rns-rnl* cluster could be fortuitous, this is unlikely in the case of the three-gene *trnR*(ucu)-*rnpB*/*trnG*(ucc) cluster. We therefore believe that the conservation of the genes flanking the eubacteria-like *rnpB* gene in both Nephroselmis and Reclinomonas mtDNAs points to a shared evolutionary history of the green algal and jakobid mitochondrial genomes.

Our conclusion that Nephroselmis mtDNA more closely resembles the protomitochondrial genome than any of the other green algal mtDNAs investigated to date still holds when the partial sequence information available for the 42.8-kb mtDNA of the prasinophyte *T. subcordiformis* (Kessler and Zetsche, 1995) is taken into consideration. Although the mtDNA of this advanced prasinophyte is similar in size to those of Nephroselmis and Prototheca and also falls within the green algal/land plant clade in phylogenetic analyses (Kessler and Zetsche, 1995; see Figure 4B), the arrangement of its 23 known genes bears little similarity to the organization of the corresponding genes in the latter two green algae. Only two ribosomal protein genes, *rps19* and *rpl16*, have so far been identified in *T. subcordiformis* mtDNA; these are unlinked and thus lack an ancestral organization.

Significance of the Similar Gene Contents of Pedinomonas and *Chlamydomonas* spp mtDNAs

Although the number of genes encoded by Pedinomonas mtDNA is almost twice that found in *Chlamydomonas* spp mtDNAs, these severely reduced green algal mtDNAs share essentially the same types of genes (Tables 1 and 2). They have lost all of the 15 ribosomal protein genes and most of the 26 tRNA genes present in Nephroselmis and Prototheca mtDNAs as well as nine other protein-coding genes (*nad7*, *nad9*, *nad10*, *cox2*, *cox3*, *atp1*, *atp9*, *ymf16*, and *ymf19*) plus the 5S rRNA gene. All of these mitochondrial genes might have been lost from a common ancestor before the divergence of the lineages that gave rise to Pedinomonas and chlamydomonads. This scenario implies that approximately half of the genes found in Pedinomonas mtDNA were subsequently lost in the lineage that led to chlamydomonads. Because the unique discontinuity in the Pedinomonas mitochondrial *rnl* gene is not equivalent in position to any of the discontinuous sites in the corresponding gene of *Chlamydomonas* spp, we also infer that the Pedinomonas and chlamydomonad sites of *rnl* discontinuity arose independently. Alternatively, independent events of gene loss might have occurred alongside independent events of *rnl* fragmentation in mtDNAs from the lineages leading to Pedinomonas and chlamydomonads, thereby yielding similar gene contents. Either of these scenarios is compatible with our phylogenetic results. As discussed below, complete mtDNA sequences from other chlorophycean green algae are required to establish unequivocally whether shared or convergent events of gene loss account for the similar gene contents of Pedinomonas and *Chlamydomonas* spp mtDNAs.

The Phylogenetic Position of Pedinomonas: Implications for the Evolutionary History of the Mitochondrial Genome and for Taxonomic Classification

The results of our phylogenetic analyses support the idea that Pedinomonas and chlamydomonad mitochondria specifically shared a common ancestor. The strongly supported clade

formed by Pedinomonas and *Chlamydomonas* spp mitochondria could be attributed to a long-branch attraction artifact due to rapid sequence divergence (Felsenstein, 1988); however, this clade may well reflect the true relationship between Pedinomonas and chlamydomonad algae, considering that the chloroplast phylogenies reconstructed from *rnl* gene sequences (Figures 4C and 4D) are entirely congruent with the mitochondrial phylogenies inferred from protein (Figures 4A and 4B) and rRNA (D.F. Spencer and M.W. Gray, unpublished results) gene sequences in showing that Pedinomonas and chlamydomonads form a strongly supported clade. The positions of Nephroselmis, Prototheca, *T. carteriiformis*, and land plants in these chloroplast phylogenies are in agreement with those found in phylogenies inferred from complete nuclear SSU rRNA gene sequences (Steinkötter et al., 1994; Friedl, 1997). In this context, it is worth noting that there is currently no published phylogeny based on nuclear SSU rRNA gene sequences that includes Pedinomonas.

Nevertheless, Pedinomonas and chlamydomonads cannot be regarded as very close relatives because they show substantial differences at the ultrastructural level (Melkonian, 1990b). In addition, comparative analyses of chloroplast *rnl* sequences suggest that chlamydomonads are more closely related to basal chlorophycean green algae considered to form a separate, major monophyletic lineage within the Chlorophyceae (a lineage defined by green algae with directly opposed basal bodies and comprising members of the Chaetophorales and the polyphyletic Chlorococcales orders [Friedl, 1997]) than to Pedinomonas, a green alga sharing with all members of the Ulvophyceae a counterclockwise absolute orientation of basal bodies (M. Turmel, C. Otis, and C. Lemieux, unpublished results). Consequently, if several mitochondrial genes were lost before the emergence of the Pedinomonas and chlamydomonad lineages, then those species spanning the phylogenetic breadth between Pedinomonas and chlamydomonads (i.e., all or most of the green algae belonging to the major lineages found in the Chlorophyceae) would be expected to harbor reduced mtDNAs lacking all or the great majority of the genes that have been eliminated from Pedinomonas mtDNA.

Our finding that Pedinomonas is not allied with Nephroselmis but rather belongs to more derived lineages related to chlamydomonads supports the views of Moestrup (1982), Sluiman (1985), and Melkonian (1990b), who suggested that this green alga should not be placed in the Prasinophyceae. Moestrup (1982) placed Pedinomonas in the Loxophyceae *sensu* (Christensen, 1962); Sluiman (1985) proposed that it is a specialized member of the Volvocales (Chlorophyceae); and Melkonian (1990b) presented evidence that it is most closely related to the Ulvophyceae. Although phylogenetic data based on chloroplast *rnl* sequences suggest that Pedinomonas is at the base of the Chlorophyceae (M. Turmel, C. Otis, and C. Lemieux, unpublished results), the precise position of this green alga relative to the Chlorophyceae, the Trebouxiophyceae, and the Ulvophyceae remains unknown. Molecular phylogenies featuring representatives of several additional chlorophycean green algal orders as well as trebouxiophytes and ulvophytes will be needed to clarify the phylogenetic position of Pedinomonas.

Mitochondrial Genome Evolution in Green Algae and Land Plants: Current Perspective

From the complete sequences of the green algal and land plant mtDNAs examined to date, it appears that the mitochondrial genome of the common ancestor of the chlorophytes and streptophytes encoded a minimum of 75 genes (i.e., all of the genes listed in Tables 1 and 2, with the exception of the intron ORFs) and that gene losses and rearrangements were common events during the evolution of this genome. Although mitochondrial genes were lost in both the chlorophyte and streptophyte lineages, these events seem to have been far more important in the chlorophyte lineage, in which they literally reshaped the architectures of Pedinomonas and chlamydomonad mtDNAs. Although the evolutionary timing of the transition to the derived "Chlamydomonas-like" mtDNA pattern is not certain, our data argue that this pattern emerged relatively late, after the split between Pedinomonas and *Tetraselmis* spp. It will be necessary to examine mtDNA in green algae spanning the phylogenetic breadth between Pedinomonas and chlamydomonads as well as in other species occupying basal positions relative to Pedinomonas, not only to provide unequivocal evidence for or against the hypothesis that several common events of gene loss account for the reduced gene contents of Pedinomonas and *Chlamydomonas* spp mtDNAs but also to gain insight into the pattern and tempo of the extensive mitochondrial gene losses observed within the Chlorophyceae.

In sharp contrast to the organizational patterns seen in the chlorophyte lineage, the evolutionary pattern of land plant mitochondrial genomes is characterized by a marked expansion in genome size without a corresponding increase in genetic complexity. This expanded pattern has been accompanied by an increase in the size of intergenic spacers and in the number of introns and intron ORFs. Within the angiosperms, there has been a progressive loss of ancestral character, with the breakup of eubacteria-like gene clusters, fragmentation of genes (concomitant with the appearance of *trans*-splicing), transfer of genes (mostly ribosomal protein genes) to the nucleus, incorporation of foreign (e.g., chloroplast) DNA, and loss of tRNA genes. Analyses of mtDNAs from various charophytes may well provide insight into when and how the ancestral pattern of organization shifted to an expanded pattern.

METHODS

Strains and Culture Conditions

Axenic strains of *Nephroselmis olivacea* and *Pedinomonas minor* were grown at 18°C under alternating 12-hr-light/12-hr-dark periods in 8-liter carboys containing modified *Volvox* medium (McCracken et

al., 1980) bubbled with air enriched with 1% CO_2. The axenic strain of Nephroselmis (NIES-484) originated from the National Institute for Environmental Forum (Japan), whereas the Pedinomonas strain (UTEX LB 1350) was obtained from the University of Texas culture collection and was rendered axenic by isolating algal cells away from contaminating bacteria on an agar plate with the aid of a tungsten needle.

Sequencing of Nephroselmis mtDNA

Lambda clones originating from a library of total cellular DNA were the source of DNA templates for sequencing the Nephroselmis mtDNA. The genomic DNA and the library were prepared as follows. Cells (6×10^{11}) were harvested by centrifugation, resuspended in 5 mL of buffer A (10 mM Tris-HCl, pH 8.0, 10 mM EDTA, and 10 mM NaCl), pulverized in liquid nitrogen, and transferred to 250 mL of buffer A containing 1.25% SDS, 100 μg mL^{-1} proteinase K, and 12.5 mM EDTA. After a 2-hr incubation at 50°C, the lysate was extracted successively with phenol, phenol/chloroform, and chloroform. Nucleic acids were precipitated at 4°C with 10% polyethylene glycol 8000 in the presence of 1 M NaCl. After dissolving the resulting pellet in 60 mL of TE (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA), CsCl-bisbenzimide gradients were prepared (1.67 g mL^{-1} CsCl and 200 μg mL^{-1} bisbenzimide) and centrifuged in a 50.2 Ti rotor at 44,000 rpm for 24 hr. The single DNA band that was observed in such gradients was removed, bisbenzimide was extracted with isopropyl alcohol saturated with 3 M NaCl, and DNA was ethanol-precipitated and dissolved in TE. The resulting DNA preparation was used to construct a library in the λGEM-11 vector (Promega). The cloning strategy involved ligating partially digested and partially filled-in (with dGTP and dATP) MboI fragments of 9 to 23 kb to the vector that was digested with XhoI and filled in with dTTP and dCTP. Phages were propagated in the *Escherichia coli* KW251.

In the course of screening the library, an mtDNA clone carrying the *rnl* gene was identified by hybridization. This clone served as the starting point for a genome walk undertaken to assemble a collection of clones encompassing the Nephroselmis mtDNA, using as probes polymerase chain reaction–amplified fragments complementary to the termini of selected inserts or long polymerase chain reaction–amplified fragments covering gaps between contigs. Several clones from each round of hybridization were selected for DNA isolation. Bacteriophage particles were prepared using the standard procedure described by Sambrook et al. (1989), and DNA was extracted in the presence of EDTA and formamide (Lech, 1990).

All nucleotide sequences were determined with the PRISM dye terminator cycle sequencing kit (1 μg of phage DNA per reaction; Applied Biosystems, Foster City, CA) on a DNA sequencer (model 373; Applied Biosystems). Both ends of each insert were sequenced using the T7 and SP6 primers, whereas internal regions of inserts were sequenced by the primer walking method using specific 22-mer oligonucleotides. Genomic regions not represented in the clones analyzed were sequenced from polymerase chain reaction–amplified fragments. Sequences were assembled using the AUTOASSEM-BLER program (version 1.4.0; Applied Biosystems) and were analyzed using the Genetics Computer Group (Madison, WI) software (version 9.1) package.

Sequencing of Pedinomonas mtDNA

DNA templates used for the sequencing of Pedinomonas mtDNA were derived from phagemids originating from a random clone library of purified mtDNA. This mtDNA was isolated from total cellular DNA

as an A+T-rich satellite band resolved by CsCl-bisbenzimide isopycnic centrifugation. The random clone library was prepared from mtDNA sheared by nebulization, and sequences were determined by a combination of manual and automated methods. Manual methods are described elsewhere (Burger et al., 1995). For automated sequencing with a DNA sequencer (model 4000L; Li-Cor, Lincoln, NE), cycle sequencing reactions were performed. Sequences were assembled with the GAP package (Bonfield et al., 1995) and analyzed using custom-made and third-party software, as described earlier (Gray et al., 1998), and documented in the Organelle Genome Megasequencing Program's web page at http://megasun.bch.umontreal.ca/ogmp/ogmpid.html.

Phylogenetic Analysis

Sequences of individual proteins were retrieved from the GOBASE database (http://alice.bch.umontreal.ca/genera/gobase) and aligned with CLUSTALW, version 1.60 (Thompson et al., 1994). N- and C-terminal regions of each protein alignment as well as internal regions containing gaps were excluded from analysis. Protein alignments were concatenated using the SeqLab editor of the Genetics Computer Group package, and phylogenetic analysis was performed using PROTML (Adachi and Hasegawa, 1996a) and the mtREV-F model (Adachi and Hasegawa, 1996b). The RELL bootstrap method was used to assess the statistical significance of tree topologies (Hasegawa and Kishino, 1994).

The nucleotide sequences of chloroplast *rnl* genes were manually aligned on the basis of secondary structure models by using the sequence editor of the Genetic Development Environment program (Smith et al., 1994). Regions not clearly alignable were excluded. Phylogenetic analyses were conducted using version 4b1 of PAUP* (Sinauer Associates, Sunderland, MA) with the maximum likelihood algorithm and the Hasegawa-Kishino-Yano (Hasegawa et al., 1985) model of evolution with rate heterogeneity. Rates (eight categories) were assumed to follow a gamma distribution with shape parameter estimated via maximum likelihood. Nucleotide frequencies and transition/transversion ratio were estimated via maximum likelihood. Bootstrap values from 100 resamplings were calculated for each data set.

REFERENCES

Adachi, J., and Hasegawa, M. (1996a). MOLPHY Version 2.3B3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. Computer Science Monographs, No. 28. (Tokyo: Institute of Statistical Mathematics).

Adachi, J., and Hasegawa, M. (1996b). Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42,** 459–468.

Bhattacharya, D., and Medlin, L. (1998). Algal phylogeny and the origin of land plants. Plant Physiol. **116,** 9–15.

Boer, P.H., and Gray, M.W. (1988a). Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. Cell **55,** 399–411.

Boer, P.H., and Gray, M.W. (1988b). Transfer RNA genes and the genetic code in *Chlamydomonas reinhardtii* mitochondria. Curr. Genet. **14,** 583–590.

Boer, P.H., and Gray, M.W. (1991). Short dispersed repeats localized in spacer regions in *Chlamydomonas reinhardtii* mitochondrial DNA. Curr. Genet. **19,** 309–312.

Bonfield, J.K., Smith, K.F., and Staden, R. (1995). A new DNA sequence assembly program. Nucleic Acids Res. **24,** 4992–4999.

Buchheim, M.A., Lemieux, C., Otis, C., Gutell, R.R., Chapman, R.L., and Turmel, M. (1996). Phylogeny of the Chlamydomonadales (Chlorophyceae): A comparison of ribosomal RNA gene sequences from the nucleus and the chloroplast. Mol. Phylogenet. Evol. **5,** 391–402.

Burger, G., Plante, I., Lonergan, K.M., and Gray, M.W. (1995). The mitochondrial genome of the amoeboid protozoon, *Acanthamoeba castellanii*: Complete sequence, gene content, and genome organization. J. Mol. Biol. **245,** 522–537.

Chapman, R.L., Buchheim, M.A., Delwiche, C.F., Friedl, T., Huss, V.A., Karol, K.G., Lewis, L.A., Manhart, J., McCourt, R.M., Olsen, J.L., and Waters, D.A. (1998). Molecular systematics of the green algae. In Molecular Systematics of Plants, P.S. Soltis, D.E. Soltis, and J.J. Doyle, eds (New York: Chapman and Hall), pp. 508–540.

Christensen, T. (1962). Alger. In Botanik, Vol. 2, Systematisk Botanik, T.W. Böcher, M. Lange, and T. Sørensen, eds (Copenhagen: Munksgaard), pp. 1–178.

Colleaux, L., Michel-Wolwertz, M.-R., Matagne, R.F., and Dujon, B. (1990). The apocytochrome *b* gene of *Chlamydomonas smithii* contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. Mol. Gen. Genet. **223,** 288–296.

Côté, V., Mercier, J.P., Lemieux, C., and Turmel, M. (1993). The single group-I intron in the chloroplast *rrnL* gene of *Chlamydomonas humicola* encodes a site-specific DNA endonuclease (I-*Chu*I). Gene **129,** 69–76.

Denovan-Wright, E.M., and Lee, R.W. (1994). Comparative structure and genomic organization of the discontinuous mitochondrial ribosomal RNA genes of *Chlamydomonas eugametos* and *Chlamydomonas reinhardtii*. J. Mol. Biol. **241,** 298–311.

Denovan-Wright, E.M., Nedelcu, A.M., and Lee, R.W. (1998). Complete sequence of the mitochondrial DNA of *Chlamydomonas eugametos*. Plant Mol. Biol. **36,** 285–295.

Dürrenberger, F., and Rochaix, J.-D. (1991). Chloroplast ribosomal intron of *Chlamydomonas reinhardtii*: *In vitro* self-splicing, DNA endonuclease activity and *in vivo* mobility. EMBO J. **10,** 3495–3501.

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. **22,** 521–565.

Fontaine, J.M., Rousvoal, S., Leblanc, C., Kloareg, B., and Loiseaux-de Goër, S. (1995). The mitochondrial LSU rDNA of the brown alga *Pylaiella littoralis* reveals α-proteobacterial features and is split by four group IIB introns with an atypical phylogeny. J. Mol. Biol. **251,** 378–389.

Friedl, T. (1997). The evolution of the green algae. Plant Syst. Evol. **11** (suppl.), 87–101.

Gray, M.W. (1995). Mitochondrial evolution. In The Molecular Biology of Plant Mitochondria, Vol. 2, Advances in Cellular and Molecular Biology of Plants, C.S. Levings III and I.K. Vasil, eds (Dordrecht, The Netherlands: Kluwer Academic Publishers), pp. 635–658.

Gray, M.W., and Schnare, M.N. (1996). Evolution of rRNA gene organization. In Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis, R.A. Zimmermann and A.E. Dahlberg, eds (Boca Raton, FL: CRC Press), pp. 49–69.

Gray, M.W., and Spencer, D.F. (1996). Organellar evolution. In Evolution of Microbial Life, D.M. Roberts, P. Sharp, G. Alderson, and M. Collins, eds (Cambridge, UK: Cambridge University Press), pp. 109–126.

Gray, M.W., Cedergren, R., Abel, Y., and Sankoff, D. (1989). On the evolutionary origin of the plant mitochondrion and its genome. Proc. Natl. Acad. Sci. USA **86,** 2267–2271.

Gray, M.W., Lang, B.F., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T.G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., and Burger, G. (1998). Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. **26,** 865–878.

Hasegawa, M., and Kishino, H. (1994). Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. Mol. Biol. Evol. **11,** 142–145.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating the ape–human split by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22,** 160–174.

Hayashi-Ishimaru, Y., Ohama, T., Kawatsu, Y., Nakamura, K., and Osawa, S. (1996). UAG is a sense codon in several chlorophycean mitochondria. Curr. Genet. **30,** 29–33.

Kessler, U., and Zetsche, K. (1995). Physical map and gene organization of the mitochondrial genome from the unicellular green alga *Platymonas* (*Tetraselmis*) *subcordiformis* (Prasinophyceae). Plant Mol. Biol. **29,** 1081–1086.

Kroymann, J., and Zetsche, K. (1998). The mitochondrial genome of *Clorogonium elongatum* inferred from the complete sequence. J. Mol. Evol. **47,** 431–440.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M.W. (1997). An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature **387,** 493–497.

Lech, K. (1990). Preparing lambda DNA from phage lysates. In Current Protocols in Molecular Biology, Vol. 1, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds (Media, PA: John Wiley and Sons), pp. 1.13.1–1.13.10.

Lonergan, K.M., and Gray, M.W. (1994). The ribosomal RNA gene region in *Acanthamoeba castellanii* mitochondrial DNA: A case of evolutionary transfer of introns between mitochondria and plastids? J. Mol. Biol. **239,** 476–499.

Ma, D.-P., King, Y.-T., Kim, Y., and Luckett, W.S., Jr. (1992). The group I intron of apocytochrome *b* gene from *Chlamydomonas smithii* encodes a site-specific endonuclease. Plant Mol. Biol. **18,** 1001–1004.

Martin, N.C., and Lang, B.F. (1997). Mitochondrial RNase P: The RNA family grows. Nucleic Acids Symp. Ser. **36,** 42–44.

McCracken, D.A., Nadakavukaren, M.J., and Cain, J.R. (1980). A biochemical and ultrastructural evaluation of the taxonomic position of *Glaucosphaera vacuolata* Korsh. New Phytol. **86,** 39–44.

Melkonian, M. (1990a). Phylum Chlorophyta. Class Prasinophyceae. In Handbook of Protoctista: The Structure, Cultivation, Habitats and Life Histories of the Eukaryotic Microorganisms and Their Descendants Exclusive of Animals, Plants and Fungi, L. Margulis, J.O. Corliss, M. Melkonian, and D.J. Chapman, eds (Boston: Jones and Bartlett Publishers), pp. 600–607.

Melkonian, M. (1990b). Chlorophyte orders of uncertain affinities. Order Pedinomonadales. In Handbook of Protoctista: The Structure, Cultivation, Habitats and Life Histories of the Eukaryotic Microorganisms and Their Descendants Exclusive of Animals, Plants and Fungi, L. Margulis, J.O. Corliss, M. Melkonian, and D.J. Chapman, eds (Boston: Jones and Bartlett Publishers), pp. 649–651.

Moestrup, Ø. (1982). Flagellar structure in algae: A review, with new observations particularly on the Chrysophyceae, Phaeophyceae (Fucophyceae), Euglenophyceae, and *Reckertia.* Phycologia **21,** 427–528.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and Ohyama, K. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. J. Mol. Biol. **223,** 1–7.

Palmer, J.D. (1990). Contrasting modes and tempos of genome evolution in land plant organelles. Trends Genet. **6,** 115–120.

Rochaix, J.D., Rahire, M., and Michel, F. (1985). The chloroplast ribosomal intron of *Chlamydomonas reinhardtii* codes for a polypeptide related to mitochondrial maturases. Nucleic Acids Res. **13,** 975–984.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schuster, W., and Brennicke, A. (1994). The plant mitochondrial genome: Physical structure, information content, RNA editing, and gene migration to the nucleus. Annu. Rev. Plant Physiol. Plant Mol. Biol. **45,** 61–78.

Siegel, R.W., Banta, A.B., Haas, E.S., Brown, J.W., and Pace, N.R. (1996). *Mycoplasma fermentans* simplifies our view of the catalytic core of ribonuclease P RNA. RNA **2,** 452–462.

Sluiman, H.J. (1985). A cladistic evaluation of the lower and higher green plants (Viridiplantae). Plant Syst. Evol. **149,** 217–232.

Smith, S.W., Overbeek, R., Woese, C.R., Gilbert, W., and Gillevet, P.M. (1994). The genetic data environment: An expandable GUI for multiple sequence analysis. Comput. Appl. Biosci. **10,** 671–675.

Steinkötter, J., Bhattacharya, D., Semmelroth, I., Bibeau, C., and Melkonian, M. (1994). Prasinophytes form independent lineages within the Chlorophyta: Evidence from ribosomal RNA sequence comparisons. J. Phycol. **30,** 340–345.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22,** 4673–4680.

Turmel, M., Mercier, J.-P., and Côté, M.-J. (1993). Group I introns interrupt the chloroplast *psaB* and *psbC* and the mitochondrial *rrnL* gene in *Chlamydomonas.* Nucleic Acids Res. **21,** 5242–5250.

Turmel, M., Côté, V., Otis, C., Mercier, J.-P., Gray, M.W., Lonergan, K.M., and Lemieux, C. (1995). Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplasts and mitochondrion). Mol. Biol. Evol. **12,** 533–545.

Unseld, M., Marienfeld, J.R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat. Genet. **15,** 57–61.

Vahrenholz, C., Rieman, G., Pratje, E., Dujon, B., and Michaelis, G. (1993). Mitochondrial DNA of *Chlamydomonas reinhardtii*: The structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. Curr. Genet. **24,** 241–247.

Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K., and Sugiura, M. (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.

Ward, B.L., Anderson, R.S., and Bendich, A.J. (1981). The mitochondrial genome is large and variable in a family of plants. Cell **25,** 793–803.

Weiner, J.H., Bilous, P.T., Shaw, G.M., Lubitz, S.P., Frost, L., Thomas, G.H., Cole, J.A., and Turner, R.J. (1998). A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. Cell **93,** 93–101.

Wolff, G., Plante, I., Lang, B.F., Kück, U., and Burger, G. (1994). Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. Gene content and genome organization. J. Mol. Biol. **237,** 75–86.