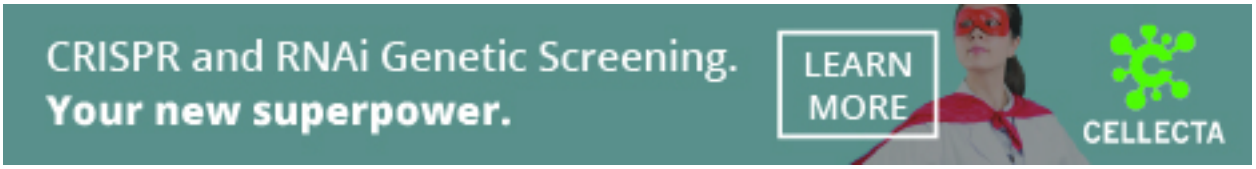
The Complete Mitochondrial DNA Sequence of *Scenedesmus obliquus* Reflects an Intermediate Stage in the Evolution of the Green Algal Mitochondrial Genome

Aurora M. Nedelcu, Robert W. Lee, Claude Lemieux, et al.

Genome Res. 2000 10: 819-831
Access the most recent version at doi:10.1101/gr.10.6.819



Letter

The Complete Mitochondrial DNA Sequence of *Scenedesmus obliquus* Reflects an Intermediate Stage in the Evolution of the Green Algal Mitochondrial Genome

Aurora M. Nedelcu,[1,5,6] Robert W. Lee,[1] Claude Lemieux,[2] Michael W. Gray,[3] and Gertraud Burger[4]

[1]*Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada;* [2] *Département de biochimie, Université Laval, Québec, Québec G1K 7P4, Canada;* [3] *Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada;* [4] *Département de biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada*

Two distinct mitochondrial genome types have been described among the green algal lineages investigated to date: a reduced–derived, *Chlamydomonas*-like type and an ancestral, *Prototheca*-like type. To determine if this unexpected dichotomy is real or is due to insufficient or biased sampling and to define trends in the evolution of the green algal mitochondrial genome, we sequenced and analyzed the mitochondrial DNA (mtDNA) of *Scenedesmus obliquus*. This genome is 42,919 bp in size and encodes 42 conserved genes (i.e., large and small subunit rRNA genes, 27 tRNA and 13 respiratory protein-coding genes), four additional free-standing open reading frames with no known homologs, and an intronic reading frame with endonuclease/maturase similarity. No 5S rRNA or ribosomal protein-coding genes have been identified in *Scenedesmus* mtDNA. The standard protein-coding genes feature a deviant genetic code characterized by the use of UAG (normally a stop codon) to specify leucine, and the unprecedented use of UCA (normally a serine codon) as a signal for termination of translation. The mitochondrial genome of *Scenedesmus* combines features of both green algal mitochondrial genome types: the presence of a more complex set of protein-coding and tRNA genes is shared with the ancestral type, whereas the lack of 5S rRNA and ribosomal protein-coding genes as well as the presence of fragmented and scrambled rRNA genes are shared with the reduced–derived type of mitochondrial genome organization. Furthermore, the gene content and the fragmentation pattern of the rRNA genes suggest that this genome represents an intermediate stage in the evolutionary process of mitochondrial genome streamlining in green algae.

[The sequence data described in this paper have been submitted to the GenBank data library under accession no. AF204057.]

Two distinct mitochondrial genome types have been described among the green algal lineages investigated to date. The *Chlamydomonas*-like type displays a reduced-derived organizational pattern characterized by small genome size (16–25 kb), limited gene content (no ribosomal protein or 5S rRNA genes and only a few respiratory protein and tRNA genes), and the presence of fragmented and scrambled rRNA coding regions. The *Prototheca*-like type represents an ancestral form of green algal mitochondrial genome that features a larger size (45–55 kb), a more complex set of protein-coding genes (including ones for ribosomal proteins), a complete or almost complete set of tRNA genes, and 5S rRNA as well as conventional continuous rRNA genes (Nedelcu 1998; Gray et al. 1998; Turmel et al. 1999). To date, six green algal mitochondrial genomes have been completely sequenced. Of these, four belong to the reduced–derived type (i.e., *Chlamydomonas reinhardtii* [Michaelis et al. 1990; Boer and Gray 1991; Vahrenholz et al. 1993], *Chlamydomonas eugametos* [Denovan-Wright et al. 1998], *Chlorogonium elongatum* [Kroymann and Zetsche 1998], and *Pedinomonas minor* [Turmel et al. 1999]), and two are members of the ancestral type (i.e., *Prototheca wickerhamii* [Wolff et al. 1994] and *Nephroselmis olivacea* [Turmel et al. 1999]). In phylogenetic analyses using mitochondrial rDNA sequences (Denovan-Wright et al. 1996) the above two types of mitochondrial genome fail to affiliate with each other. Furthermore, whereas in mitochondrial protein trees the ancestral mitochondrial sequences directly affiliate, as expected, with their land plant counterparts, the reduced–derived (fast-evolving) sequences form a clade separate from both ancestral green algal and land plant homologs (Turmel et al. 1999).

[5]*Present address: University of Arizona, Department of Ecology and Evolutionary Biology, Tucson, Arizona 85721.*
[6]Corresponding author.
E-MAIL nedelcua@u.arizona.edu.

The causes, factors, and mechanisms responsible for the extensive changes sustained by the mitochondrial genome in different green algal lineages are not known yet, although some suggestions have been made (Nedelcu 1998; Nedelcu and Lee 1998a,b). To decipher the processes involved in the evolution of the green algal mitochondrial genome in particular, and to understand the mechanisms involved in the evolution of the mitochondrial genome in general, we need more knowledge about the extent of mitochondrial genome diversity and the specific evolutionary trends in mitochondrial genome organization within each lineage.

The phylogeny of green algae continues to undergo revision. Green algal lineages are distributed between two phyla, Chlorophyta and Streptophyta (see Bremer 1985) (Figure 1). The reduced-derived green algal mitochondrial genomes sequenced to date belong to lineages placed in the Chlorophyceae (i.e., *Chlamydomonas* spp. and *Chlorogonium*) and Pedinophyceae (i.e., *Pedinomonas*), whereas the ancestral-like mitochondrial genomes belong to the trebouxiophyte (i.e., *Prototheca*) and prasinophyte (i.e., *Nephroselmis*) lineages (Fig. 1). Comparisons among the gene contents of ancestral and reduced–derived mitochondrial genomes show that most of the derived features shared by all of the *Chlamydomonas*-like mitochondrial genomes (such as lack of ribosomal protein and 5S rRNA genes and the presence of fragmented and scrambled rRNA coding regions) are also found in the primitive-like taxon, *Pedinomonas*; however, the rRNA genes are less fragmented and a slightly larger number of genes

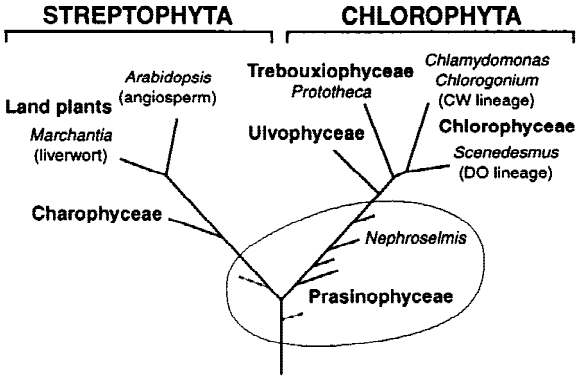


Figure 1 Evolutionary relationships among green plants (green algae and land plants). The two main evolutionary branches depict the two sister phyla (Chlorophyta and Streptophyta) into which all green plants fall. Chlorophyta comprises three classes (Ulvophyceae, Trebouxiophyceae, and Chlorophyceae). The two deeply diverging chlorophycean lineages defined by flagellar apparatus configuration (CW, clockwise; DO, directly opposed) are shown. The Prasinophyceae is a non-monophyletic assemblage of primitive green algae, some of which are of uncertain phylogenic affiliation (*dashed lines*). All the green plants for which complete mtDNA sequences have been determined are indicated, with the exception of *Pedinomonas minor*, a primitive unicellular green flagellate whose phylogenetic position has not yet been established.

are encoded in the mitochondrial genome of this taxon, relative to other *Chlamydomonas*-like counterparts. Do these observations mean that the evolutionary processes leading to the very peculiar type of mitochondrial genome organization in *Chlamydomonas* were initiated long before the divergence of the chlorophycean group, namely in a *Pedinomonas*-like green flagellate ancestor, and continued since? Alternatively, is the similarity in organization between the pedinophycean and chlorophycean mitochondrial genomes an example of convergent evolution?

The goals of our work were to determine whether the observed split in mitochondrial genome organization and sequence affiliation in green algae is real or is due to insufficient or biased sampling, and to define trends in mitochondrial genome evolution within the green algal group. The chlorophycean group consists of two very distinct evolutionary lineages that diverged early (Wilcox et al. 1992; Steinkötter et al. 1994) (Fig. 1). Because all of the chlorophycean mitochondrial genomes sequenced to date belong to only one lineage and feature reduced–derived organizational types, we decided to investigate mitochondrial genome organization within the other lineage. The specific questions addressed by this study follow:

1. Do the mitochondrial genomes in the two (early diverged) chlorophycean lineages resemble each other and are they of the reduced–derived mitochondrial genome type?
2. Alternatively, are the mitochondrial genomes very different between the two evolutionarily distinct chlorophycean lineages?
3. Will the acquisition of information about mitochondrial genome organization from both chlorophycean lineages contribute to suggesting evolutionary mechanisms and pathways in the streamlining process leading toward the derived chlamydomonadalean mitochondrial genome?
4. Could mitochondrial genome traits help decipher phylogenetic relationships among green algal lineages?

In light of available data on phylogenetic affiliation (Wilcox et al. 1992; Steinkötter et al. 1994), mitochondrial genome size (Kück 1989), and rRNA gene organization (Nedelcu et al. 1996; Nedelcu 1997) for *Scenedesmus obliquus*, we decided to determine the entire mitochondrial genome sequence of this green alga.

RESULTS

Genome Organization

Genome Size, Base Composition, and Map

The complete mtDNA sequence of *S. obliquus* (GenBank no. AF204057) suggests a 42,919-bp genome with

Table 1. Comparison of Mitochondrial Genome Traits in Green Algae[a]

Trait	Cre	Ceu	Cel	Sob	Pmi	Pwi	Nol
Size (bp)	15,758	22,897	22,704	42,919	25,137	55,328	45,223
A+T%							
overall	54.8	65.4	62.2	63.7	77.8	74.2	67.2
coding	54.9	65.7	63.1	60.6	76.3	69.9	65.7
non-coding	54.5	63.6	54.2	69.3	80.2	84.6	72.7
Map	linear	circular	circular	circular	circular	circular	circular
Coding (%)	83.1	84.6	89.1	60.6	60.9	70.6	78.4
Gene content (excluding duplicates)							
overall number	13	19	18	47	21	65	69
protein							
respiratory	7	7	7	13	11	17	18
ribosomal	0	0	0	0	0	13	15
rRNA	2	2	2	2	2	3	3
tRNA	3	3	3	27	8	26	26
ORFs							
free	1	0	0	4	0	4	3
intronic (I/II)	0/0	7/0	6/0	1/0	0/0	2/0	4/0
Introns							
group I	0	9	6	2	0	5	4
group II	0	0	0	2	1	0	0

[a](Cre) *Chlamydomonas reinhardtii;* (Ceu) *Chlamydomonas eugametos;* (Cel) *Chlorogonium elongatum;* (Pmi) *Pedinomonas minor;* (Sob) *Scenedesmus obliquus;* (Pwi) *Prototheca wickerhamii;* (Nol) *Nephroselmis olivacea.*

an overall A +T of 63.7%. Table 1 summarizes features of this mitochondrial genome and Figure 2 depicts the circular genome map deduced from the *Scenedesmus* mtDNA sequence. A circular restriction map for the mitochondrial genome of another strain of *S. obliquus,* KS3/2, has previously been reported (Kück 1989).

Gene Content

Scenedesmus mtDNA features 42 conserved genes and five additional open reading frames (ORFs), all present as single copies. Tables 1 and 2 provide a summary of gene content. Conserved genes code for fragmented large subunit (LSU) and small subunit (SSU) rRNAs, 27 tRNAs, and 13 respiratory proteins of mitochondrial complexes I, III, IV, and V. Notable among the 27 mitochondrial tRNA genes (*trn*) is a *trnL(cua)* coding for a tRNA that would recognize UAG (normally a stop codon) as leucine. Two $tRNA^{Met}$ coding regions have been identified in *Scenedesmus* mtDNA. Comparisons of the two inferred $tRNA^{Met}$ secondary structures with other green algal mitochondrially encoded as well as eubacterial initiator and elongator $tRNA^{Met}$ sequences suggest that one (encoded by Mf) is the initiator $tRNA^{Met}$ whereas the other (encoded by Me) is the elongator $tRNA^{Met}$ (Figure 2). The four free-standing ORFs (*orf130*, *orf148*, *orf345*, and *orf390*) have no known homologs whereas the intronic *orf215* might code for an endonuclease/maturase-like protein. No 5S rRNA (*rrn5*) or ribosomal protein-coding genes have been identified in *Scenedesmus* mtDNA, and transcription and biosynthesis genes are also absent.

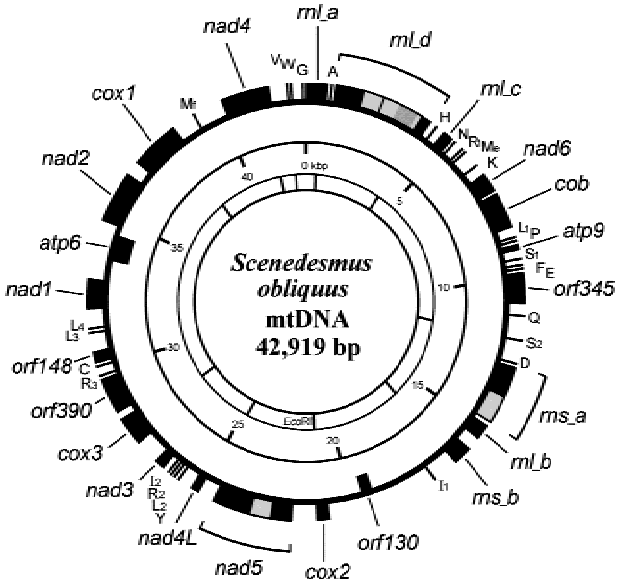


Figure 2 Physical and genome map of the *S. obliquus* mtDNA. Genes, exons, and non-intronic ORFs are depicted as *black blocks*, introns are indicated as *light-gray blocks*, and intronic ORFs are shown as *dark-gray blocks*. Gene abbreviations are listed in Table 2. Gene blocks outside and inside the circle are transcribed in clockwise and counterclockwise directions, respectively. Transfer RNA genes are shown as *thin black bars*, with letters indicating amino acid specificity (see Table 2), and numbers denoting different genes specific for the same amino acid. The anticodons (*lowercase letters enclosed in parentheses*) of the numbered tRNA genes are: I1, (uau); I2, (gau); L1, (cag); L2, (aag); L3, (caa); L4, (cua); R1, (acg); R2, (ccu); R3, (ucu); S1, (gcu); S2, (gga). Me and Mf denote the elongator and initiator *trnM(cau)*, respectively.

Table 2. Comparison of Gene Content in Green Algal Mitochondrial Genomes[a]

Gene			Cre	Ceu	Cel	Sob	Pmi	Pwi	Nol
Complex I									
*nad*1			+	+	+	+	+	+	+
*nad*2			+	+	+	+	+	+	+
*nad*3			–	–	–	+	+	+	+
*nad*4			+	+	+	+	+	+	+
*nad*4L			–	–	–	+	+	+	+
*nad*5			+	+	+	+	+	+	+
*nad*6			+	+	+	+	+	+	+
*nad*7			–	–	–	–	–	+	+
*nad*8			–	–	–	–	–	–	+
*nad*9			–	–	–	–	–	+	+
Complex III (*cob*)									
			+	+	+	+	+	+	+
Complex IV									
*cox*1			+	+	+	+	+	+	+
*cox*2			–	–	–	+	–	+	+
*cox*3			–	–	–	+	–	+	+
Complex V									
*atp*1			–	–	–	–	–	+	+
*atp*6			–	–	–	+	+	+	+
*atp*8			–	–	–	–	+	+	+
*atp*9			–	–	–	+	–	+	+
Ribosomal RNAs									
rnl			+ (8)[b]	+ (6)	+ (6)	+ (4)	+ (2)	+ (1)	+ (1)
rns			+ (4)	+ (3)	+ (3)	+ (2)	+ (1)	+ (1)	+ (1)
rrn5			–	–	–	–	–	+ (1)	+ (1)
Transfer RNAs									
AA	*Anticodon*	*Codon*							
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	(gau)	AUY	–	–	–	+	–	+	+
I[c]	**(uau)**[c]	**AUA**[c]	–	–	–	+	–	–	–
I	(cau)	AUA	–	–	–	–	–	+	+
K	(uuu)	AAR	–	–	–	+	–	+	+
L	**(cua)***	**UAG**	–	–	–	+	–	–	–
L	(uaa)	UUR	–	–	–	–	–	+	+
L	(caa)	UUG	–	–	–	+	+	–	–
L	(uag)	CUN	–	–	–	–	–	+	+
L	**(aag)***	**CUN**	–	–	–	+	–	–	–
L	**(cag)***	**CUG**	–	–	–	+	–	–	–
Me	(cau)	AUG	+	+[d]	+	+	–	+	+
Mf	(cau)	AUG	–	–	–	+	–	+	+
N	(guu)	AAY	–	–	–	+	–	+	+
P	(ugg)	CCN	–	–	–	+	–	+	+
Q	(uug)	CAR	+	+	+	+	+	+	+
R	(acg)	CGN	–	–	–	+	–	+	+
R	(ucg)	CGR	–	–	–	–	–	–	+
R	(ucu)	AGR	–	–	–	+	–	+	+
R	**(ccu)***	**AGG**	–	–	–	+	–	–	–
S	(gcu)	AGY	–	–	–	+	–	+	+
S	(uga)	UCN	–	–	–	–	–	+	+
S	**(gga)***	**UCY**	–	–	–	+	–	–	–
T	(ggu)	ACY	–	–	–	–	–	–	+
T	(ugu)	ACN	–	–	–	–	–	+	–
V	(uac)	GUN	–	–	–	+	–	+	+
W	(cca)	UGG	+	+	+	+	–	+	+
W	(uca)	UGR	–	–	–	–	+	–	–
Y	(gua)	UAY	–	–	–	+	+[d]	+	+

[a]See Table 1 for abbreviations of species names.
[b]Numbers in parentheses indicate the number of coding modules for that particular rRNA species.
[c]Mitochondrially-encoded tRNAs unique to *Scenedesmus* are in bold and marked with an asterisk; the codons they read and the amino acids they specify are also set in bold.
[d]Transfer RNA coded for by a duplicated gene.

Genetic Code and Codon Usage

From the sequence data, we infer that the *Scenedesmus* mitochondrion uses a deviant genetic code. Multiple protein alignments, as well as the presence of a mitochondrially encoded $tRNA^{Leu}$ with anticodon CUA suggest that UAG is read as leucine. In addition, these alignments revealed that if TGA and TAA were assumed to be the sole termination codons in the *Scenedesmus* mitochondrion, protein reading frames would be considerably longer at the carboxyl termini than their homologs from other species. Possible explanations are: (1) a standard sense codon is used as a stop codon; (2) a regular termination codon is generated through RNA editing; or (3) a UAA termination codon is generated through site-specific cleavage or polyadenylation.

Our preliminary data support the first possibility. We observed that in all standard mitochondrial protein-coding genes in *Scenedesmus*, the expected end of the reading frame fell precisely at or close to a serine residue that is always specified by a TCA codon. Moreover, additional TCA codons are not present upstream of this position in any of these protein-coding genes. Also, although there are 27 mitochondrially encoded tRNAs in *Scenedesmus*, the mitochondrial genome lacks a *trnS(uga)* specifying a tRNA that would decode UCA. Consistent with the lack of a *trnS(uga)* is (1) the additional absence in all protein-coding genes of the TCG codon, which is the other serine codon that could be decoded by a $tRNA^{Ser}$(uga); and (2) the presence of a $tRNA^{Ser}$(gga) with the capacity to decode UCU and UCC but not UCG or UCA. To test whether RNA editing or cleavage/polyadenylation might be introducing a termination codon in the primary transcript, we performed a series of reverse transcriptase-PCR (RT-PCR) experiments. Preliminary data (not shown) did not reveal any changes in the mRNA sequence relative to the corresponding DNA coding sequence.

Table 3 presents the codon usage in the 13 standard mitochondrial protein-coding genes in *Scenedesmus*. A pronounced codon bias is indicated by the fact that up to 87% of the codons in four-codon families end in A or T. However, in these cases, there does not seem to be any preference for third-position A or T. Seven codons in total (TTA, Leu; ATA, Ile; TCG, Ser; CGG and AGA, Arg; TAA and TGA, termination) are not found in *Scenedesmus* mtDNA. The set of 27 tRNAs coded for by *Scenedesmus* mtDNA is sufficient to recognize all of the remaining codons except ACN(Thr), assuming that a single tRNA is able to read all codons in the four-codon families GTN(Val), CCN(Pro), CGN(Ala) and GGN(Gly). Gene sequence predicts that in the corresponding tRNAs, the wobble position of the anticodon would be occupied by a U residue; if unmodified, this wobble U would allow a single tRNA species to decode all of the synonymous codons in each four-codon family. In two other tRNAs, $tRNA^{Leu}$(aag) and $tRNA^{Arg}$(acg), the A residue in the wobble position is presumably modified posttranscriptionally to inosine, potentially allowing the tRNAs to read all four synonymous codons (Pfitzinger et al. 1990) (note, however, that an additional $tRNA^{Arg}$(ucg) is available to read CGA codons). Two tRNAs, $tRNA^{Ile}$(uau) and $tRNA^{Arg}$(ucu), appear to be redundant

Table 3. Codon Usage (%) in the 13 Standard Mitochondrial Protein-coding Genes of *Scenedesmus*

Codon	%	AC[a]	AA[a]	Codon	%	AC	AA	Codon	%	AC	AA	Codon	%	AC	AA
TTT	75	(gaa)	F	TCT	34	(gga)	S	TAT	81	(gua)	Y	TGT	83	(gca)	C
TTC	25	(")[b]	F	TCC	6	(")	S	TAC	19	(")	Y	TGC	17	(")	C
TTA	0	(–)	L	***TCA***[c]	100	(–)	*[d]	TAA	0		*	TGA	0		*
TTG	55	(caa)	L	TCG	0	(–)	S	***TAG***	17	(cua)	L	TGG	100	(cca)	W
CTT	21	(**a**ag)[e]	L	CCT	56	(ugg)	P	CAT	51	(gug)	H	CGT	35	(**a**cg)	R
CTC	4	(")	L	CCC	7	(")	P	CAC	49	(")	H	CGC	1	(")	R
CTA	2	(")	L	CCA	33	(")	P	CAA	92	(uug)	Q	CGA	13	(ucg)	R
CTG	1	(cag)	L	CCG	4	(")	P	CAG	8	(")	Q	CGG	0	(–)	R
ATT	93	(gau)	I	ACT	37	(–)	T	AAT	67	(guu)	N	AGT	51	(gcu)	S
ATC	7	(")	I	ACC	11	(–)	T	AAC	33	(")	N	AGC	9	(")	S
ATA	0	(uau)	I	ACA	50	(–)	T	AAA	92	(uuu)	K	AGA	0	(ucu)	R
ATG	100	(cau)	M	ACG	2	(–)	T	AAG	8	(")	K	AGG	51	(ccu)	R
GTT	44	(uac)	V	GCT	43	(ugc)	A	GAT	69	(guc)	D	GGT	33	(ucc)	G
GTC	3	(")	V	GCC	5	(")	A	GAC	31	(")	D	GGC	5	(")	G
GTA	42	(")	V	GCA	45	(")	A	GAA	83	(uuc)	E	GGA	55	(")	G
GTG	11	(")	V	GCG	6	(")	A	GAG	17	(")	E	GGG	6	(")	G

[a]AC and AA denote the tRNA anticodon and the amino acid specified, respectively.
[b](") indicates identity with the anticodon in the line above, and (–) indicates the lack of mitochondrially-encoded tRNAs with anticodons that can read the codons indicated at the left.
[c]Deviations from the universal genetic code are set in bold italics.
[d]Asterisks indicate termination codons.
[e]Adenosines in the wobble position that are expected to be deaminated to inosine are represented in bold.

because the corresponding codons (ATA and AGA, respectively) are not used in *Scenedesmus* mtDNA (Table 3).

Gene Organization and Structure

The mitochondrial SSU and LSU rRNA genes (*rns* and *rnl*) are fragmented and scrambled in *Scenedesmus*. The *rnl* and *rns* coding modules are found clustered in two regions of the genome and are flanked exclusively by each other or by tRNA genes (Fig. 2). The LSU and SSU rRNA genes are fragmented into four (i.e., *rnl_a, rnl_b, rnl_c, rnl_d*) and two (i.e., *rns_a*, *rns_b*) coding modules, respectively. These modules are transcribed and the size of the transcripts (Figure 3) (Nedelcu 1997) is consistent with that suggested by their corresponding gene sequence. A number of insertions (relative to the *Escherichia coli* homologs) ranging from 79 bp to 150 bp are present in *rnl_a, rnl_c, rnl_d*, and *rns_a*. Northern blot analyses indicate that these extra sequences are present in the mature transcripts (Nedelcu 1997).

The mitochondrial protein-coding regions were identified based on sequence similarity and alignments with homologs in other lineages. The mitochondrial protein sequences were inferred from DNA sequences and amino acid alignments; the genetic code used was a deviant one, as discussed above. The predicted proteins are generally similar in organization to their homologs in other green algae. However, a few exceptions have been noted. First, Nad2 appears 40–51 amino acids longer at its amino-terminus than the *Prototheca* and *Nephroselmis* counterparts; second, Atp6 contains a 27-amino acid insertion; and third, Cox3 has extensions of approximately 12 and 32 residues at its amino- and carboxyl termini, respectively. A more significant deviation is the deduced sequence of Cox2, which is approximately 100 amino acids shorter at the carboxyl terminus than other Cox2 proteins. Because this region represents almost half of the protein sequence in other Cox2 counterparts and contains conserved amino acid motifs, it is conceivable that *cox2* is, in fact, a pseudogene.

Three mitochondrial genes are populated by a total of four introns (two group I and two group II) in *Scenedesmus*. One of the group I introns (1273 bp) is inserted in *rnl_d* (*E. coli* coordinate 2505), belongs to the IA1 subfamily, and contains the only intronic ORF, *orf215*, present in the *Scenedesmus* mitochondrial genome. This ORF is free-standing and potentially codes for a polypeptide that is similar to other group I intronic endonucleases/maturases, especially from fungal mitochondria. The second group I intron (611 bp) is located in *nad5*. A group II intron (608 bp) is inserted in *rnl_d* (*E. coli* coordinate 2455), 49 nt upstream of the *rnl_d* group I intron; this group II intron belongs to the 2B subfamily and represents the smallest reported self-splicing group II intron from an organelle gene (Kück et al., 1990). The second group II intron (848 bp) is located in *rns_a* (*E. coli* coordinate 740).

Genome Structure

Gene density in the mitochondrial genome of *Scenedesmus* is low, with identified genes accounting for only 60.6% of the sequence. Coding regions are rather evenly dispersed throughout the genome. Protein-coding genes are highly interspersed with tRNA and rRNA coding regions, and all but two genes, namely, *atp6* and *orf130*, are encoded on the same DNA strand. Intergenic spacers occupy a total of 16.9 kb (~ 39% of the genome) and vary in length from 0 (i.e., the *trnA(ugc)/rnl_d* and *trnG(ucc)/rnl_c* spacers) to almost 2 kb (i.e., the *trnI(uau)/orf130* spacer); most spacers are between 2 and 600 bp in length, with only one being larger than 1 kb. The A +T content of intergenic spacers is higher than that of coding regions (Table 1).

Numerous repetitive sequences are distributed throughout the genome, and are especially prominent in the intergenic regions. Of the 50 intergenic spacers, 34 are populated by repeated sequences; the intergenic regions that lack such repeats are smaller than 60 bp, and all but one are flanked on at least one end by a tRNA gene. The repeat motifs are between 16 and 118 bp long. In some cases, short repeats (identical or with one mismatch) are organized as tandem arrays (e.g., multiple copies of a 27-bp sequence in the *rns_b/trnI(aua)* spacer and of a 40-bp repeat in the *trnI(aua)/orf130* spacer); the tandem copies are either directly adjacent or separated by sequences that appear to be degenerate or truncated copies of the repeat unit. Most of the longer repeat regions are composed of several different repeat units that are also found isolated or in combination with other repeated sequences throughout the genome. Although most copies of the repetitive sequences are located in intergenic spacers, some are also found in rRNA coding regions (e.g., a 37-bp sequence present in both *rns_c* and the *trnN(guu)/trnR(acg)* intergenic spacer) and introns (e.g., a 43-bp repeat present in the *rns_a* group II intron as well as the *trnS(gga)/trnD(guc)* spacer). In addition, repetitive sequences are found in some ORFs (e.g., *orf345, orf130, and orf148*) and protein-coding genes (e.g., a 47-bp sequence present in the 3′ region of *nad2* and the *nad2/cox1* spacer, and a 28-bp sequence present in the *orf345/trnQ(uug)* spacer as well as overlapping the last 12 bp at the 3′ end of *cob*).

Phylogenetic Analyses

Phylogenetic analyses were conducted using a data set comprising 1949 amino acid positions from the concatenated protein sequences of seven genes (*cob*, *cox1*, *nad1*, *nad2*, *nad4*, *nad5*, and *nad6*) that are common to

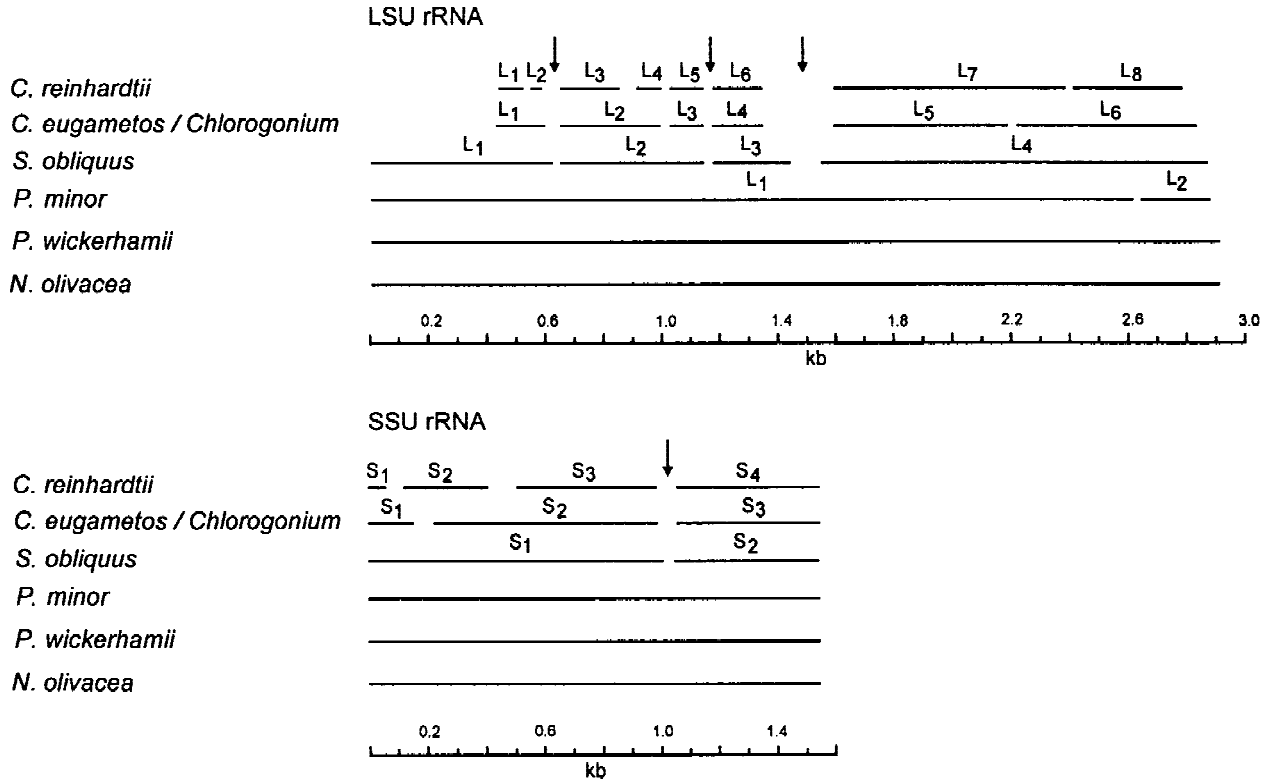


Figure 3 Mitochondrial rRNA structure in green algae. Ribosomal RNA fragmentation patterns are depicted for chlorophyte algae whose mitochondrial genomes have been completely sequenced; the rRNAs are represented to the scale of the covalently continuous *E. coli* homologs. *Arrows* indicate the positions of breakpoints that occur in corresponding variable regions in the *C. reinhardtii*, *C. eugametos, Chlorogonium*, and *S. obliquus* rRNAs; S_1–S_4 and L_1–L_8 denote SSU and LSU rRNA fragments, respectively.

the mitochondrial genomes of *Scenedesmus*, six other green algae (*C. eugametos*, *C. reinhardtii*, *Chlorogonium*, *Nephroselmis*, *Pedinomonas*, *Prototheca*), two land plants (*Arabidopsis thaliana* and *Marchantia polymorpha*) and two red algae (*Chondrus crispus* and *Porphyra purpurea*). The mitochondrial genome of the chytridiomycete fungus, *Allomyces macrogynus*, was used as an outgroup. Figure 4 depicts the best tree inferred with PROTML, which shows that *Scenedesmus*, *Chlorogonium*, and *Chlamydomonas* mitochondrial sequences form a strongly supported clade. However, the *Scenedesmus* and chlamydomonadalean sequences, together with those of *Pedinomonas*, form a cluster that does not directly affiliate with the rest of the green algae and land plants. The remarkably long branches displayed by taxa in this cluster undoubtedly reflect a highly accelerated rate of sequence divergence relative to other protist and land plant mtDNA sequences.

DISCUSSION

Scenedesmus and Other Green Algal Mitochondrial Genomes: Similarities and Contrasts

Genome Size, Map, and Structure

At 42.9 kb, the mitochondrial genome of *Scenedesmus* is 5% smaller than the ancestral-like green algal mitochondrial genome of *Nephroselmis* (45.2 kb), but considerably larger than the reduced–derived mitochondrial genomes of *Pedinomonas* (25.1 kb) and *Chlamydomonas* spp./*Chlorogonium* (15.8–22.9 kb). Mitochondrial genomes larger than 45 kb (e.g., the 55.3-kb mtDNA of the trebouxiophycean alga *Prototheca*) have been reported in other green algal groups. Like

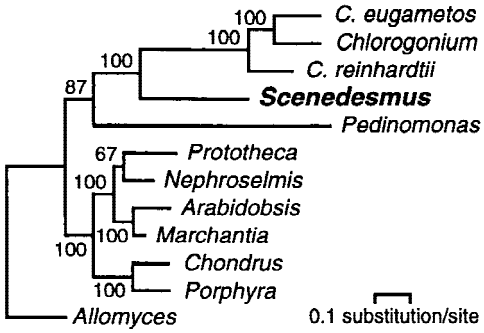


Figure 4 Best tree derived from a phylogenetic analysis of concatenated mitochondrial sequences representing seven respiratory protein-coding genes (1949 amino acid positions) common to all of the green algal mitochondrial genomes sequenced to date. Abbreviations are as in the text.

most, but not all, of the green algal mitochondrial genomes investigated to date, *Scenedesmus* mtDNA is a circular-mapping genome (Table 1).

The gene density in *Scenedesmus* mtDNA (60.6% of the total sequence) is the lowest among the green algal mitochondrial genomes completely sequenced to date (Table 1). A comparably low overall gene density (60.9%) is also seen in *Pedinomonas* mtDNA, but in this case noncoding sequence is mostly localized to an 11-kb region that contains a tRNA gene duplicate (Turmel et al. 1999), rather than being dispersed throughout the mitochondrial genome as in the *Scenedesmus* case. The only gene linkage shared with another green algal mitochondrial genome is the *trnF-trnE* pair that is also present in *Pedinomonas* mtDNA. The *nad4L-nad5-nad4-nad2* gene cluster present in the ancestral mitochondrial genome of the protist, *Reclinomonas americana* (Lang et al. 1997), is completely dispersed in *Scenedesmus*, as it is in *Chlorogonium*. In contrast, *Nephroselmis* mtDNA maintains the *nad5-nad4-nad2* cluster, whereas *Prototheca* and *C. eugametos/Pedinomonas* only retain the *nad5-nad4* and *nad4-nad2* linkage groups, respectively. It is interesting that the arrangement of genes into only one (*Pedinomonas*, *C. eugametos*, and *Chlorogonium*), two (*Prototheca*, *C. reinhardtii*) or three (*Scenedesmus*) transcriptional units seems to be a feature of the more advanced green algae. In contrast, in *Nephroselmis* mtDNA, genes are organized into at least 15 different such units.

The complex organization of the short repetitive sequences found in *Scenedesmus* mtDNA resembles that of the short repeats described in *C. eugametos* mtDNA (Nedelcu and Lee 1998b); however, the repeat units are larger in size and in a higher number of copies when organized in tandem arrays. The *Scenedesmus* short repeated sequences have a less biased base composition when compared with the very GC-rich repeats found in *Chlamydomonas* spp. mtDNA (Boer and Gray 1991; Nedelcu and Lee 1998b) and the very AT-rich repeats described in *Prototheca* mtDNA (Wolff et al. 1994).

Gene Content

Tables 1 and 2 compare gene repertoire in completely sequenced green algal mitochondrial genomes. *Scenedesmus* mtDNA encodes all 12 conserved genes present in its chlamydomonadalean counterpart. Although 29 additional conserved genes are present in the mitochondrial genome of *Scenedesmus* relative to those of *Chlamydomonas* spp./*Chlorogonium* homologs, another 29 such genes are missing in *Scenedesmus* in comparison to the more ancestral *Nephroselmis* mtDNA. In addition, three of the four respiratory protein-coding genes that are present in *Pedinomonas* but not *Chlamydomonas* have also been found in the *Scenedesmus* mitochondrial genome; the only protein-coding gene present in *Pedinomonas* but not in *Scenedesmus* is *atp8*. In fact, *Scenedesmus* is unique in its repertoire of mitochondrially encoded *atp* genes, containing only *atp6* and *atp9*. As in *Pedinomonas* and *Chlamydomonas* spp./*Chlorogonium*, but in contrast to *Prototheca* and *Nephroselmis*, no genes for subunit 1 of the ATPase complex (*atp1*), 5S rRNA or ribosomal proteins have been identified in the *Scenedesmus* mitochondrial genome (Table 1).

The 27 mitochondrially encoded tRNAs in *Scenedesmus* represent a ninefold larger set than that of *Chlamydomonas* spp./*Chlorogonium* mitochondrial genomes and exceed by one the number of tRNAs encoded in the ancestral mitochondrial genome of *Nephroselmis* (Table 2). The *Scenedesmus* mitochondrial genome lacks eight of the tRNA genes that are present in some other green algal mtDNAs; however, six mitochondrial tRNA genes are unique to *Scenedesmus* when compared to other green algal/land plant mitochondrially encoded tRNA sets (Table 2). The mtDNA-encoded tRNA gene repertoire of *Scenedesmus* includes all of the tRNA genes present in either *Chlamydomonas* spp./*Chlorogonium* or *Pedinomonas* mtDNAs except for *trnW(uca)*, which so far is unique to *Pedinomonas* within the green algal/land plant group. A particularly notable feature is the presence of *trnL(caa)* in the mitochondrial genome of *Scenedesmus* and *Pedinomonas* and its absence in the *Prototheca* and *Nephroselmis* counterparts. It is also noteworthy that, whereas *Pedinomonas* lacks any mitochondrially encoded $tRNA^{Met}$ and the *Chlamydomonas* mtDNAs appear to code only for the elongator $tRNA^{Met}$, the *Scenedesmus* mitochondrial genome most likely codes for both initiator and elongator $tRNAs^{Met}$, as do *Nephroselmis* and *Prototheca* mtDNAs.

Genetic Code and Codon Usage

Although most of the green algal mitochondria (such as of *Chlamydomonas* spp., *Chlorogonium*, *Prototheca*, *Nephroselmis*) employ the standard genetic code, deviations have been reported in some green algal groups, including other chlorophycean taxa (direct relatives of *Scenedesmus*) (Hayashi-Ishimaru et al. 1996) and *Pedinomonas* (Turmel et al. 1999). Although the type of variation differs among these green algal lineages, all reported deviations involve the use of a termination codon as a sense codon: UGA as tryptophan in *Pedinomonas* (Turmel et al. 1999) and UAG as alanine or leucine in some lineages within the chlorophycean group (Hayashi-Ishimaru et al. 1996). The *Scenedesmus* mitochondrion does not use UGA to specify tryptophan; however, like some other chlorophycean algae, it does decode UAG as leucine. In addition, *Scenedesmus* is so far unique in using the UCA codon as a signal for termination of translation.

It is interesting that six of the seven codons not

present in *Scenedesmus* mtDNA are also not found in standard protein-coding genes of *C. reinhardtii* mtDNA (Boer and Gray 1988a). The seventh codon is TAA, which is used as a stop codon in all but one of the standard protein-coding genes in *C. reinhardtii* mitochondria but possibly not at all in the *Scenedesmus* counterparts. In *Scenedesmus*, there does not seem to be any preference for third-position A or T among the codons in four-codon families, although such preferences have been noted in other green algal mitochondrial genomes (Denovan-Wright et al. 1998).

Gene Organization and Structure

The *Scenedesmus* mitochondrial genome shares with its *Chlamydomonas* (Boer and Gray 1988b; Denovan-Wright and Lee 1994), *Chlorogonium* (Kroymann and Zetsche 1998), and *Pedinomonas* (Turmel et al. 1999) counterparts the feature of fragmented and scrambled rRNA coding regions. Figure 3 shows that all three *rnl* breakpoints in *Scenedesmus* correspond to variable regions that are interrupted in *Chlamydomonas* spp. and *Chlorogonium* but not *Pedinomonas* LSU rRNA; the breakpoint in the *Pedinomonas rnl* is unique among the organelle LSU rRNA genes investigated to date. The mitochondrial *rns* of *Scenedesmus* is also fragmented as are its homologs in *Chlamydomonas* spp. and *Chlorogonium* but not *Pedinomonas*; the *rns* breakpoint in *Scenedesmus* is within a variable region that is interrupted in all of the chlamydomonadalean SSU rRNA genes described so far.

Extreme variation in intron number has been reported among the green algal mitochondrial genomes completely sequenced to date (Table 1); so far, the *Scenedesmus* mtDNA is the only one that contains both group I and II introns. The number and location of introns have been shown to vary greatly even among quite closely related green algal species (Matagne et al. 1988; Colleaux et al. 1990; Denovan-Wright and Lee 1992; Kroymann and Zetsche 1997); this seems also to be the case among *Scenedesmus* species because a group I intron is present in the *cox1* gene of *Scenedesmus quadricauda* (Watanabe et al. 1998) but not *S. obliquus* (this work). The *Scenedesmus rnl* group II intron is located at a position (i.e., *E. coli* coordinate 2505) that is different from the position (i.e., *E. coli* coordinate 1787) of the *rnl* group II intron in the mtDNAs of both *Pedinomonas* and the brown alga, *Pylaiella littoralis*. No intronic ORF has been identified in either the *Scenedesmus* or *Pedinomonas rnl* group II introns, although an ORF coding for a potential reverse transcriptase is present in the *P. littoralis* counterpart (Fontaine et al. 1995). It is noteworthy that a RT-like gene (*rtl*) is located between two *rnl* coding regions in the mitochondrial genome of *C. reinhardtii* (Boer and Gray 1988c); the sequence and secondary structure of its flanking regions suggest that this *rtl* might be the remnant of an intronic ORF harbored by a group II intron inserted 55 bp upstream of the insertion site of the *Scenedesmus rnl_d* group II intron (Nedelcu and Lee 1998c).

An Intermediate Type of Mitochondrial Genome Organization in Green Algae

The feature that best characterizes the mitochondrial genome of *Scenedesmus* as an evolutionary intermediate between the ancestral and the reduced–derived mtDNA types in green algae is the organization of its rRNA genes. The fragmentation pattern of the rRNA genes in *Scenedesmus* suggests a gradual evolution from continuous rRNA genes, as in *Nephroselmis* and *Prototheca*, to the highly fragmented homologs in *C. reinhardtii* (Figure 3). The four *Scenedesmus rnl* fragments code for an approximately 3-kb discontinuous LSU rRNA that is much larger than its *Chlamydomonas* counterpart (e.g., ca 1.9 kb in *C. eugametos*). The size differential is due in part to the presence in *Scenedesmus* of the 400- and 200-nt 5′- and 3′-terminal regions, respectively, that are completely absent in the *Chlamydomonas* spp./*Chlorogonium* mitochondrial LSU rRNAs but present in all other green algal homologues (Figure 3). The degree of fragmentation of the *Scenedesmus* mitochondrial *rnl* is lower than that observed in the chlamydomonadalean group, but all the breakpoints found in *Scenedesmus rnl* are present in all the chlamydomonadalean counterparts investigated so far. The mitochondrial *rns* of *Scenedesmus* is also less fragmented than its homologs in *Chlamydomonas*, and the breakpoint is shared with its chlamydomonadalean counterparts (Figure 3).

An additional trait that favors an intermediate evolutionary position for the *Scenedesmus* mitochondrial genome is its gene content. Overall, there are 42 identified genes; this value falls between the 12 such genes found in the chlamydomonadalean mitochondrial genome and the 62 genes present in *Nephroselmis* mtDNA. Moreover, for most of the respiratory complexes, the *Scenedesmus* mitochondrial genome encodes an intermediate number of genes (Table 2). For example, seven Nad genes are mitochondrially encoded in *Scenedesmus*, compared with only five in *Chlamydomonas* and 10 in *Nephroselmis*. Furthermore, although the *Scenedesmus* mitochondrial genome codes for all three cytochrome oxidase subunits (as the *Nephroselmis* but not the *Chlamydomonas* mtDNAs), the presence of a truncated *cox2* coding region might reflect an intermediate stage in the evolutionary process leading to the loss of this gene in the *Chlamydomonas*-like mitochondrial genomes.

Mitochondrial Genome Evolution and Green Algal Phylogeny

How does this apparently intermediate type of mito-

chondrial genome organization correlate with the phylogenetic position of *Scenedesmus* within the chlorophycean green algal group? Is the *Scenedesmus* mitochondrial genome a true evolutionary intermediate in the streamlining of the chlorophycean mitochondrial genome toward the very reduced chlamydomonadalean type?

Features that evolved within and therefore are shared by a particular lineage (i.e., synapomorphic characters) are considered to be informative in defining phylogenetic affiliations within a group. In this context, the absence of *atp1*, *rrn5* and ribosomal protein-coding genes as well as the presence of fragmented and scrambled mitochondrial rRNA genes, all of which are shared by the *Chlamydomonas*, *Chlorogonium, Pedinomonas*, and *Scenedesmus* lineages, could be considered indicative of common ancestry for these lineages. Consequently, the absence of *cox2* and *cox3*, a character shared by the *Chlamydomonas*, *Chlorogonium*, and *Pedinomonas* mitochondrial genomes but not the *Scenedesmus* homolog, could argue that the chlamydomonadalean lineage and *Pedinomonas* share a more recent common ancestor with each other than with *Scenedesmus*. Conversely, these mitochondrial traits might well be the consequence of independent evolutionary events. The questions to be addressed then are: (1) How do the derived features of mitochondrial genome organization correlate with the evolutionary history of green algae? (2) To what extent can mitochondrial traits contribute to elucidation of phylogenetic affiliations among green algae?

To address these questions we used concatenated mitochondrial amino acid sequences (Figure 4). The best tree suggests that the chlamydomonadalean group shares a more recent common ancestor with *Scenedesmus* rather than with *Pedinomonas*. However, *Pedinomonas* appears more closely related to the chlorophycean clade than to the prasinophyte *Nephroselmis* or the trebouxiophyte *Prototheca*; this branching order should, nevertheless, be regarded with caution because it could be the result of a long-branch-attraction artifact (Felsenstein 1988).

The observed affiliation of *Scenedesmus* with the chlamydomonadalean lineage is consistent with existing morphologic, ultrastructural, and molecular data (Melkonian 1990; Buchheim and Chapman 1992; Buchheim et al. 1996; Friedl 1997). The derived features of mitochondrial genome organization in both *Chlamydomonas* and *Scenedesmus*, such as the similar pattern of fragmentation of the rRNA coding regions (Figure 3) and the shared absence of specific genes (Table 2), are, therefore, likely to reflect true phylogenetic relationships.

On the other hand, the affiliation of *Pedinomonas* mitochondrial protein-coding sequences with their chlorophycean counterparts (Figure 4) is not consistent with other types of data (Moestrup 1991). Nevertheless, chloroplast *rnl* data are congruent with mitochondrial protein phylogenic analyses in affiliating this taxon with the chlorophycean and not prasinophycean (*Nephroselmis* and *Tetraselmis subcordiiformis*) or trebouxiophycean (*Prototheca*) lineages (Turmel et al. 1999). Does this affiliation reflect a true phylogenic relationship? Could *Pedinomonas* be a descendant of a primitive flagellate chlorophycean ancestor? Interestingly, regardless of whether or not the basal position of *Pedinomonas* relative to the chlorophycean clade reflects a true phylogenic relationship, the absence of *cox2*, *cox3* and a large number of tRNA genes (Table 2) is likely to be an example of evolutionary convergence in the mitochondrial genomes of the chlamydomonadalean and *Pedinomonas* lineages.

Potential Features of Mitochondrial Genome Organization in the Chlorophycean Ancestor

Generally, derived features that are shared by two related lineages are also thought to have been present in their most recent common ancestor. Having data on mitochondrial genome organization from both chlorophycean lineages allows us to make inferences about the nature of the mitochondrial genome in the ancestor of the chlorophycean lineage. Because *Scenedesmus* features a more gene-rich mitochondrial genome than does *Chlamydomonas*, it is probable that the ancestral chlorophycean mitochondrial genome had a larger gene complement than its chlamydomonadalean homolog. Genes that were lost prior to the chlorophycean split most likely include those missing in both chlorophycean lineages, such as *atp1*, *rrn5*, and all ribosomal protein-coding genes. Nevertheless, these genes have also been lost in the *Pedinomonas* lineage, but the shared or independent nature of these losses in the chlorophycean and *Pedinomonas* lineages cannot be deduced with certainty from available data. Therefore, the possibility remains that some of these genes were still present in the chlorophycean ancestor and independently lost in the lineages leading to *Chlamydomonas* and *Scenedesmus*.

A perhaps less contentious feature of mitochondrial genome organization in the chlorophycean ancestor is the presence of fragmented and scrambled rRNA coding regions. All of the rRNA variable regions that are interrupted in *Scenedesmus* are also interrupted in the other chlorophycean taxa investigated to date, suggesting that the rRNA genes in the most common recent ancestor of the chlorophycean group were already fragmented, most likely in a manner similar to that described in *Scenedesmus*. Admittedly, independent events could be responsible for the disruption of rRNA genes in corresponding variable regions in both chlorophycean lineages. However, the fact that the breakpoint in the *Pedinomonas* mitochondrial *rnl* is in a

completely different region of the gene relative to the other green algae investigated to date suggests that it is more likely that independent fragmentation events occurred at different sites rather than in corresponding variable regions in different lineages.

More information is needed to infer the nature of the mitochondrial genome in the chlorophycean common ancestor. If it is determined that *Pedinomonas* is not a true relative of the chlorophycean group, this would indicate that convergent evolutionary events have been rather frequent and extensive in green algal mitochondrial genomes. For this reason, inferences based on shared derived mitochondrial features, especially gene content, will have to be regarded with caution. Based on the available data, we can posit that the ancestral chlorophycean mitochondrial genome could have had a gene repertoire as large as that of its *Nephroselmis* counterpart but not smaller than that of *Scenedesmus*; that it probably featured fragmented and scrambled rRNA coding regions; and that its genes were populated by group I (and possibly group II) introns.

Concluding Remarks

This work was triggered by the observed unexpected dichotomy in mitochondrial genome organization and sequence affiliation among the green algae investigated to date. Although the present study provides information important to understanding the patterns of mitochondrial genome evolution in green algae, data on mitochondrial genomes from additional green algal groups are still needed to decipher the causes and mechanisms responsible for diversity of these genomes within this group. Comparing patterns and evolutionary trends among mitochondrial genomes from distant eukaryotic groups (i.e., animals, land plants, fungi, protists) will ultimately reveal the general mechanisms that govern the evolution of the mitochondrial genome; it is in this light that we think information on mitochondrial genome organization and evolution in green algae will prove useful.

METHODS

Strain, Culture Conditions, and Isolation of mtDNA

Scenedesmus obliquus (Turp.) Kutz strain 78 was obtained from the University of Texas Algal Culture Collection (UTEX) and grown in Basal Medium (Oh-hama and Hase 1980). Cultures were supplied with 1% CO_2 in air and cool-white fluorescent light on a 12 hours light/12 hours dark cycle. Cells were harvested during the exponential phase and broken using a French pressure cell or a mixture of glass beads. Total DNA was prepared using either phenol or guanidinium hydrochloride extraction procedures (Sambrook et al. 1989). Mitochondrial DNA was separated from the nuclear and chloroplast DNA following CsCl-gradient centrifugations. Mitochondrial DNA was initially identified by comparing its *Eco*RI restriction pattern with that previously reported by Kück (1989) for *Scenedesmus* strain KS3/2.

Cloning of mtDNA

Mitochondrial DNA was physically fragmented by nebulization (Okpodu et al. 1994). After fractionation by electrophoresis in an agarose gel, the random, size-selected fragments of mtDNA (500–1000 bp and 1000–3000 bp) were incubated with *E. coli* DNA polymerase I (Klenow fragment) and T7 DNA polymerase to generate blunt ends and then cloned into the *Sma*I site of pBluescriptII KS + (Stratagene). Recombinant plasmids containing mtDNA inserts were identified by colony hybridization using intact mtDNA as a probe. Clones contained in this random library encompassed the entire *Scenedesmus* mitochondrial genome.

Sequencing strategies

As templates, we used single-stranded DNA, obtained by superinfection of recombinant clones with helper phage K07 (Vieira and Messing 1987). Manual DNA sequencing was performed by the dideoxy chain termination method (Sanger et al. 1977). High-resolution polyacrylamide gels, dried onto the glass plate (Lang and Burger 1990), were autoradiographed and sequences were entered manually into computer files. Most readings were obtained by automated sequencing on a Li-Cor 4000L apparatus, using end-labeled primer and a cycle-sequencing protocol (Amersham Pharmacia Biotech, Inc., Baie d'Urfé, Québec, Canada). Both strands were sequenced, and in regions where clone coverage was inadequate, specific primers were synthesized for use in primer walking.

Data Analysis

Sequences were read manually and assembled using the XGAP package V3.6 1996 (Bonfield et al. 1995). Sequence analyses were performed on SUN workstations using software developed by one of the authors (Lang and Burger 1986), as well as with tools included in the Gene Runner (Hastings Software, Inc., Hastings, NY) and GCG (Version 8; Genetics Computer Group, Madison, WI) sequence analysis packages. The FASTA program (Pearson 1990) was used for searches in local databases; sequence similarity searches were also performed at the National Center for Biotechnology Information (NCBI), using the BLAST network service (Altschul et al. 1990). The CLUSTAL V program (Higgins et al. 1992) was used for multiple protein alignments. Both programs were managed in the GDE (Genetic Data Environment) package (Smith et al. 1994). BLAST searches were conducted with the batch utility BBLAST and large-scale output was screened with TBOB (Littlejohn and Rioux 1994). A number of other programs, including multiple sequence file manipulation, preprocessing and conversion utilities for XGAP, FASTA, and GDE have been developed in the Sequencing Unit of the Organelle Genome Megasequencing Program (OGMP). These utilities are available through the OGMP website (http://megasun.bch.umontreal.ca/ogmp/ogmpid.html).

The complete sequence of *Scenedesmus* mtDNA is deposited in GenBank (no. AF204057). Sequences encompassing parts of rRNA genes and flanking regions of another *Scenedesmus* strain, KS3/2, have been published by others (GenBank no. X17375; Kück et al. 1990), but were completely redetermined in the present study. This partial sequence, as well as the complete *Scenedesmus* mtDNA sequence, are also available through the Organelle Genome Database Project

(GOBASE) (http://megasun.bch.umontreal.ca/gobase) (Korab-Laskowska et al. 1998).

Phylogenetic Analyses

Inferred amino acid sequences of seven protein-coding genes (*cob*, *cox1*, *nad1*, *nad2*, *nad4*, *nad5*, and *nad6*) were retrieved from the GOBASE database and aligned using Clustal W 1.74 (Thompson et al. 1994). These genes are found in the seven green algal, two land plant, and two red algal mitochondrial genomes included in these phylogenetic analyses, as well as in the fungal outgroup. Alignments were concatenated, excluding ambiguously aligned regions containing gaps. Phylogenetic analyses were performed using PROTML (Adachi and Hasegawa 1996) and the mtREV-F model of sequence evolution. The RELL bootstrap method was used to assess the statistical significance of tree topologies (Adachi and Hasegawa 1996).

ACKNOWLEDGMENTS

We thank B. Franz Lang for contributions to mtDNA purification and critical discussion, I. Plante for library construction, and I. Plante, D. Saint-Louis, and Y. Zhu for sequencing. This work was supported by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada to R.W.L (NSERC 9599) and C.L. (NSERC 2830), and the Medical Research Council of Canada to the Organelle Genome Megasequencing Program and M.W.G. (SP-14226 and MT-4124, respectively). A.M.N. was supported by an NSERC grant to R.W.L. and an NSERC Postdoctoral Fellowship. Generous donations of an automated sequencer from Li-Cor (Lincoln, NE) and computer equipment from Sun Microsystems (Palo Alto, CA) greatly assisted this study. G.B. and C.L. are Associates and M.W.G. is a Fellow in the Program in Evolutionary Biology of the Canadian Institute for Advanced Research (CIAR), whom we thank for salary and interaction support.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adachi, J. and M. Hasegawa. 1996. *Computer science monographs, No. 28. MOLPHY version 2.3B3: Programs for molecular phylogenetics based on maximum likelihood*. Institute of Statistical Mathematics, Tokyo.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

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

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

Boer, P.H. and M.W. Gray. 1988c. Genes encoding a subunit of respiratory NADH dehydrogenase (ND1) and a reverse transcriptase-like protein (RTL) are linked to ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *EMBO J.* **7:** 3501–3508.

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

Bonfield, J.K., K.F. Smith, and R. Staden. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* **23:** 4992–4999.

Bremer, K. 1985. Summary of green plant phylogeny and classification. *Cladistics* **1:** 369–385.

Buchheim, M.A. and R.L. Chapman. 1992. Phylogeny of *Carteria* (Chlorophyceae) inferred from molecular and organismal data. *J. Phycol.* **28:** 362–374.

Buchheim, M.A., C. Lemieux, C. Otis, R.R. Gutell, R.L. Chapman, and M. Turmel. 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.

Colleaux, L., M.-R. Michel-Wolwertz, R.F. Matagne, and B. Dujon. 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.

Denovan-Wright, E.M. and R.W. Lee. 1992. Comparative analysis of the mitochondrial genomes of *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. *Curr Genet.* **21:** 197–202.

Denovan-Wright, E.M. and R.W. Lee. 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., D. Sankoff, D.F. Spencer, and R.W. Lee. 1996. Evolution of fragmented mitochondrial ribosomal RNA genes in *Chlamydomonas*. *J. Mol. Evol.* **42:** 382–391.

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

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

Fontaine, J.M., S. Rousvoal, C. Leblanc, B. Kloareg, and S. Loiseaux-de Göer. 1995. The mitochondrial LSU rDNA of the brown alga *Pylaiella littoralis* reveals αproteobacterial 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.* (Suppl.) **11:** 87–101.

Gray M.W., Lemieux, C., Burger, G., Lang, B.F., Otis, C., Plante, I., and Turmel, M. 1998. Mitochondrial genome organization and evolution within the green algae and land plants. In *Plant mitochondria: From gene to function* (eds. I.M. Møller, P. Gardeström, K. Glimelius, and E. Glaser), pp. 1–8. Backhuys Publishers, Leiden.

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

Higgins, D.G., A.J. Bleasby, and R. Fuchs. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* **8:** 189–191.

Korab-Laskowska, M., P. Rioux, N. Brossard, T.G. Littlejohn, M.W. Gray, B.F. Lang, and G. Burger. 1998. The Organelle Genome Database Project (GOBASE). *Nucleic Acids. Res.* **26:** 138–144.

Kroymann, J. and K. Zetsche. 1997. The apocytochrome-b gene in *Chlorogonium elongatum* (*Chlamydomonadaceae*): An intronic GIY-YIG ORF in green algal mitochondria. *Curr. Genet.* **31:** 414–418.

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

Kück, U. 1989. The intron of a plastid gene from a green alga contains an open reading frame for a reverse transcriptase-like enzyme. *Mol. Gen. Genet.* **218:** 257–265.

Kück, U., I. Godenhardt, and U. Schmidt. 1990. A self-splicing group II intron in the mitochondrial large subunit rRNA (LSU rRNA) gene of the eukaryotic alga *Scenedesmus obliquus*. *Nucleic Acids Res.* **18:** 2691–2697.

Lang B.F. and G. Burger. 1986. A collection of programs for nucleic acid and protein analysis, written in FORTRAN 77 for IBM-PC compatible microcomputers. *Nucleic Acids Res.* **14:** 455–465.

Lang, B.F. and G. Burger. 1990. A rapid, high resolution DNA sequencing gel system. *Anal. Biochem.* **188:** 176–180.

Lang, B.F., G. Burger, C.J. O'Kelly, R. Cedergren, G.B. Golding, C. Lemieux, D. Sankoff, M. Turmel, and M.W. Gray. 1997. An

ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387:** 493–497.

Littlejohn, T.G. and P. Rioux, 1994. *TBOB.Bionet.software* Message-IDCpnMwI.LAH@cc.umontreal.ca

Matagne, R.F., D. Rongvaux, and R. Loppes. 1988. Transmission of mitochondrial DNA in crosses involving diploid gametes homozygous or heterozygous for the mating-type locus in *Chlamydomonas*. *Mol. Gen. Genet.* **214:** 257–262.

Melkonian, M. 1990. Phylum Chlorophyta, class Prasinophyceae. In *Handbook of Protoctista* (eds. L. Margulis, J.O. Corliss, M. Melkonian, and D.J. Chapman), pp. 600–607. Jones and Bartlett Publishers, Boston.

Michaelis, G., C. Vahrenholz, and E. Pratje. 1990. Mitochondrial DNA of *Chlamydomonas reinhardtii*: The gene for apocytochrome *b* and the complete functional map of the 15.8 kb DNA. *Mol. Gen. Genet.* **233:** 211–216.

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

Nedelcu, A.M. 1997. Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: A model for their origin and evolution. *Mol. Biol. Evol.* **14:** 506–517.

Nedelcu, A.M. 1998. Contrasting mitochondrial genome organizations and sequence affiliations among green algae: Potential factors, mechanisms, and evolutionary scenarios. *J. Phycol.* **34:** 16–28.

Nedelcu, A.M. and R.W. Lee. 1998a. Modes and tempos of mitochondrial and chloroplast genome evolution in *Chlamydomonas*: A comparative analysis. In *Molecular biology of Chlamydomonas*: Chloroplasts and mitochondria (eds. J.-D. Rochaix, M. Goldschmidt-Clermont, and S. Merchant), pp. 63–91. Kluwer, Dordrecht.

Nedelcu, A.M and R.W. Lee. 1998b. Short repetitive sequences in green algal mitochondrial genomes: Potential roles in mitochondrial genome evolution. *Mol. Biol. Evol.* **15:** 690–701.

Nedelcu, A.M, and R.W. Lee. 1998c. A degenerate group II intron in the intronless mitochondrial genome of *Chlamydomonas reinhardtii*: Evolutionary implications. *Mol. Biol. Evol.* **15:** 918–922.

Nedelcu, A.M., D.F. Spencer, E.M. Denovan-Wright, and R.W. Lee. 1996. Discontinuous mitochondrial and chloroplast large subunit ribosomal RNAs among green algae: Phylogenetic implications. *J. Phycol.* **32:** 103–111.

Oh-hama, T. and E. Hase. 1980. Formation of protochlorophyll(ide) in wild type and cell mutant C-2A' of *Scenedesmus obliquus*. *Plant Cell Physiol.* **21:** 1263–1272.

Okpodu, C.M., D. Robertson, W.F. Boss, R.K. Togasaki, and S.J. Surzycki. 1994. Rapid isolation of nuclei from carrot suspension culture cells using a BioNebulizer. *Biotechniques* **16:** 154–159.

Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183:** 63–98.

Pfitzinger, H., J.H. Weil, D.T.N. Pillay, and P. Guillemaut. 1990. Codon recognition mechanisms in plant chloroplasts. *Plant Mol. Biol.* **14:** 805–814.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

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

Sprinzl, M., C. Horn, M. Brown, A. Ioudovitch, and S. Steinberg. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26:** 148–153.

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

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

Turmel, M., C. Lemieux, G. Burger, B.F. Lang, C. Otis, I. Plante, and M.W. Gray. 1999. The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: Two radically different evolutionary patterns within the green algae. *Plant Cell* **11:** 1717–1730.

Vahrenholz, C., G. Riemen, E. Pratje, B. Dujon, and G. Michaelis. 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.

Vieira, J., and J. Messing. 1987. Production of single-stranded plasmid DNA. *Meth. Enzymol.* **153:** 3–11.

Watanabe, K.I., M. Ehara, Y. Inagaki, and T. Ohama. 1998. Distinctive origins of group I introns found in the COXI genes of three green algae. *Gene* **213:** 1–7.

Wilcox, L.W., L.A. Lewis, P.A. Fuerst, and G.L. Floyd. 1992. Assessing the relationships of autosporic and zoosporic chlorococcalean green algae with 18S rDNA sequence data. *J. Phycol.* **28:** 381–386.

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

Received November 29, 1999; accepted in revised form March 29, 2000.