

# The Mitochondrial Genome of *Chara vulgaris*: Insights into the Mitochondrial DNA Architecture of the Last Common Ancestor of Green Algae and Land Plants<sup>W</sup>

Monique Turmel,<sup>1</sup> Christian Otis, and Claude Lemieux

Département de Biochimie et de Microbiologie, Université Laval, Québec, Québec G1K 7P4, Canada

**Mitochondrial DNA (mtDNA) has undergone radical changes during the evolution of green plants, yet little is known about the dynamics of mtDNA evolution in this phylum. Land plant mtDNAs differ from the few green algal mtDNAs that have been analyzed to date by their expanded size, long spacers, and diversity of introns. We have determined the mtDNA sequence of *Chara vulgaris* (Charophyceae), a green alga belonging to the charophycean order (Charales) that is thought to be the most closely related alga to land plants. This 67,737-bp mtDNA sequence, displaying 68 conserved genes and 27 introns, was compared with those of three angiosperms, the bryophyte *Marchantia polymorpha*, the charophycean alga *Chaetosphaeridium globosum* (Coleochaetales), and the green alga *Mesostigma viride*. Despite important differences in size and intron composition, *Chara* mtDNA strikingly resembles *Marchantia* mtDNA; for instance, all except 9 of 68 conserved genes lie within blocks of colinear sequences. Overall, our genome comparisons and phylogenetic analyses provide unequivocal support for a sister-group relationship between the Charales and the land plants. Only four introns in land plant mtDNAs appear to have been inherited vertically from a charalean algal ancestor. We infer that the common ancestor of green algae and land plants harbored a tightly packed, gene-rich, and relatively intron-poor mitochondrial genome. The group II introns in this ancestral genome appear to have spread to new mtDNA sites during the evolution of bryophytes and charalean green algae, accounting for part of the intron diversity found in *Chara* and land plant mitochondria.**

## INTRODUCTION

It is well established that land plants evolved from green algae belonging to the Charophyceae (Graham et al., 2000). Five orders of freshwater green algae have been recognized in this class: Chlorokybales, Klebsormidiales, Zygnematales, Coleochaetales, and Charales (Mattox and Stewart, 1984). Recently, a phylogenetic analysis of the combined sequences of four genes originating from the nucleus (small subunit rRNA gene), chloroplast (*atpB* and *rbcL*), and mitochondrion (*nad5*) of various charophycean green algae and green plants revealed that the Charales form a highly supported clade with land plants, suggesting that they represent the closest green algal relatives of land plants (Karol et al., 2001). On the basis of morphological characteristics alone, both the Charales and the Coleochaetales had been considered to be closely affiliated with land plants (Bower, 1908; Fritsch, 1935; Pickett-Heaps and Marchant, 1972; Mattox and Stewart, 1984). Given their poor resolution, previously reported phylogenies of charophycean green algae inferred from single gene sequences derived from the nucleus (small subunit rRNA gene) or the chloroplast (*rbcL*) and from the combined chloroplast small subunit and large subunit rRNA gene sequences failed to trace the charophycean lineage that gave rise to land plants (Turmel et al., 2002a). Nevertheless,

these phylogenies are congruent with the four-gene phylogeny mentioned above in placing the charophyceans and land plants in the same monophyletic group. This group, termed Streptophyta, is sister to the Chlorophyta, a major lineage that contains all members belonging to the other classes of green algae (Prasinophyceae, Ulvophyceae, Trebouxiophyceae, and Chlorophyceae), with perhaps the exception of the prasinophyte *Mesostigma viride*.

Green plant mitochondrial DNAs (mtDNAs) display high variability in structure and size (Palmer et al., 2000). The complete mtDNA sequences of *Mesostigma* (Turmel et al., 2002b), of the charophycean alga *Chaetosphaeridium globosum* (Coleochaetales) (Turmel et al., 2002c), of the bryophyte *Marchantia polymorpha* (a liverwort) (Oda et al., 1992a), and of the angiosperms *Arabidopsis* (Unseld et al., 1997), sugar beet (Kubo et al., 2000), and rice (Notsu et al., 2002) indicate that the streptophyte mitochondrial genome underwent major restructuring events either late during the evolution of charophycean green algae (i.e., after the emergence of the Coleochaetales) or concomitant with the emergence of the first land plant lineage (Turmel et al., 2002c), which is thought to be represented by liverworts (Mishler, 1994; Kenrick and Crane, 1997; Lewis et al., 1997; Qiu et al., 1998). As detailed below, these events resulted in a substantial expansion of the mitochondrial genome, the insertion of introns at many sites, and the acquisition of editing at the transcript level.

With a size of 56,574 bp, a coding capacity of 67 genes, and its existence as a unique circular-mapping molecule, *Chaetosphaeridium* mtDNA closely resembles its homologs in *Mesostigma* and *Nephroselmis olivacea*, the earliest diverging chloro-

<sup>1</sup> To whom correspondence should be addressed. E-mail monique.turmel@rsvs.ulaval.ca; fax 418-656-5036.

<sup>W</sup> Online version contains Web-only data.

Article, publication date, and citation information can be found at [www.plantcell.org/cgi/doi/10.1105/tpc.013169](http://www.plantcell.org/cgi/doi/10.1105/tpc.013169).

phyte whose complete mtDNA sequence has been determined (Turmel et al., 1999). It is 3.3 times smaller than the unique circular-mapping mtDNA molecule of *Marchantia*, which contains almost the same number of genes (65), and is at least 6.5 times smaller than its angiosperm homologs, which encode ~50 genes and occur as collections of differently sized circular isomers. The increased size of *Marchantia* mtDNA relative to its *Chaetosphaeridium* homolog is accounted for mainly by the enlargement of intergenic regions (Turmel et al., 2002c), whereas the extra size of angiosperm mtDNAs compared with *Marchantia* mtDNA is essentially attributed to further enlargement of these regions, to duplications, and to uptake of DNA from the chloroplast and the nucleus (Unsel et al., 1997; Kubo et al., 2000; Notsu et al., 2002).

The highly variable intron composition observed among streptophyte mtDNAs suggests that introns have been gained independently in several lineages. Of the 11 introns in *Chaetosphaeridium* mtDNA, only the fourth intron in *cox1* exhibits positional and structural conservation with one of the seven introns in *Mesostigma* mtDNA, and only the second intron in the same gene is conserved positionally in land plant mtDNAs (more specifically, in *Marchantia* mtDNA). Similarly, a single intron in *Mesostigma* mtDNA shows positional and structural homology with a land plant mitochondrial intron (again, a *Marchantia cox1* intron). All of the 32 introns in *Marchantia* mtDNA, with the exception of the *nad2* intron, reside at distinct positions relative to the 20 to 23 introns in angiosperm mtDNAs, suggesting that most, if not all, of the liverwort introns arose independently from their angiosperm counterparts. The distribution patterns of mitochondrial introns among basal land plants (both bryophytes and vascular plants) are consistent with this hypothesis (Malek et al., 1997; Malek and Knoop, 1998; Qiu et al., 1998; Beckert et al., 1999; Pruchner et al., 2001). The alternative hypothesis that all of the introns in the four completely sequenced land plant mtDNAs were present in the common ancestor of all land plants and that subsequently many were lost independently in early-diverging lineages appears unlikely in light of the finding that mitochondrial introns are stable in bryophytes belonging to the same class (Beckert et al., 1999, 2001; Pruchner et al., 2001) and also in both flowering and non-flowering vascular plants (Cho et al., 1998; Qiu et al., 1998; Vangerow et al., 1999). A notable exception is the group-I intron in *cox1* that has been identified in 48 of the 335 genera of angiosperms examined in the course of an extensive DNA gel blot survey (Cho et al., 1998; Cho and Palmer, 1999). This angiosperm intron, inserted at the same site as a group-I intron in *Marchantia* mtDNA, was acquired recently through at least 32 separate events of horizontal transfer from a fungal donor.

RNA-editing events involving the conversions of cytidines to uridines and uridines to cytidines have been observed in the mitochondria of angiosperms and basal land plants (Hiesel et al., 1994; Maier et al., 1996; Malek et al., 1996; Sper-Whitis et al., 1996; Giegé and Brennicke, 1999; Steinhauser et al., 1999) but appear to be absent in the subclass of complex thaloid liverworts (*Marchantiidae*) (Ohyama et al., 1993; Steinhauser et al., 1999) and the few algae examined to date, including *Mesostigma*, *Chaetosphaeridium* (Turmel et al., 2002c), and members of the Charales (Malek et al., 1996; Steinhauser et al.,

1999). This phylogenetic distribution supports the idea that RNA editing was acquired by streptophyte mitochondria coincidentally with the emergence of land plants but was lost secondarily or extremely reduced in incidence in the *Marchantiidae* (Steinhauser et al., 1999).

Did the streptophyte mitochondrial genome begin to expand at a late stage during the evolution of charophycean green algae? Were several of the introns in land plant mtDNAs acquired by the streptophyte mitochondrial genome before the emergence of land plants? Was the gain of RNA editing by streptophyte mitochondria concurrent with the emergence of the first terrestrial plants, as suggested by the sequences of the few charophycean mitochondrial genes investigated to date? To address these questions and to gain insight into the mtDNA architecture of the last charophycean ancestor of land plants, we have undertaken the sequencing of mtDNA from a member of the Charales, *Chara vulgaris*. We report here that *Chara* mtDNA bears strong resemblance to its *Marchantia* homolog despite important differences in size and intron composition. Overall, our genome comparisons and phylogenetic analyses unequivocally support the idea that the Charales are the closest green algal relatives of land plants. Moreover, our results suggest that land plant evolution was accompanied by the expansion of the mitochondrial genome, the insertion of introns at numerous sites, and the appearance of editing at the RNA level.

## RESULTS

### Main Features of *Chara* mtDNA

The *Chara* mtDNA sequence assembles as a circle of 67,737 bp (Figure 1). Table 1 compares the main features of this green algal mtDNA with those of its counterparts in *Mesostigma*, *Chaetosphaeridium*, and land plants. In terms of size, *Chara* mtDNA more closely resembles the mtDNAs of *Mesostigma* and *Chaetosphaeridium* than those of land plants; however, with respect to A+T content, it is related more closely to its land plant homologs. The highest density of coding sequences (90.7%) among all streptophyte mtDNAs examined to date is found in *Chara* mtDNA. Considering that this charophycean mtDNA encodes 68 conserved genes (Table 2), its coding capacity is approximately the same as those of *Mesostigma*, *Chaetosphaeridium*, and *Marchantia* mtDNAs. Note that the term "conserved genes" here designates genes of known or unknown function that have been identified previously in mtDNA; pseudogenes and intron open reading frames (ORFs) are not included in this category. A single pseudogene,  $\Psi trL(caa)$ , is present in *Chara* mtDNA; this pseudo-tRNA gene lacks the sequence encoding the variable D-loop region. Fourteen group-I introns and 13 group-II introns account for 38.5% of the *Chara* mitochondrial genome. Eight of these introns are homologous both positionally and structurally with previously identified introns in *Chaetosphaeridium*, *Mesostigma*, or land plant mtDNAs (see below).

### Intergenic Spacers

Intergenic spacers in *Chara* mtDNA vary from 0 to 1173 bp, with an average size of 91 bp and only three spacers exceeding



**Table 1.** General Features of mtDNAs from Chara, Other Streptophytes, and Mesostigma

Genome Feature	Mesostigma	Chaetosphaeridium	Chara	Marchantia	Arabidopsis	Sugar Beet	Rice
Size (bp)	42,424	56,574	67,737	186,609	366,924	368,799	490,520
A+T content (%)	67.8	65.6	59.1	57.6	55.2	56.1	56.2
Coding sequences (%) <sup>a</sup>	86.6	76.3	90.7	65.0	36.8	33.0	–
Gene content <sup>b</sup>	65	67	68	69	49	48	53
Introns							
Group I	4	9	14	7	0	0	0
Group II							
Cis-spliced	1	2	13	25	18	14	17
Trans-spliced	2	0	0	0	5	6	6

<sup>a</sup> Conserved genes, unique ORFs of >100 codons, introns, and intron ORFs were considered coding sequences. The values for Marchantia and Arabidopsis mtDNAs were taken from Lang et al. (1999) and Marienfeld et al. (1999), respectively. The proportion of coding sequences in rice mtDNA could not be determined because coordinates for unique ORFs of <150 codons and duplicated coding regions are not reported in the genome accession.

<sup>b</sup> Pseudogenes, unique ORFs, and intron ORFs were not considered.

encoded by Chara mtDNA are not sufficient to decode all of the 61 codons identified in this genome; missing is one or more species that can recognize the Thr ACR codons. Of all green plant mtDNAs sequenced to date, only that of the chlorophyte *Prototheca wickerhamii* specifies a tRNA species [tRNA<sup>Thr</sup>(ugu)] that can read these codons (Wolff et al., 1994). The absence of *trnL*(caa), *trnS*(acu), and *trnR*(ucg) from Chara mtDNA does not affect the decoding of the codons UUG, AGU, and CGR, because the products of *trnL*(uaa), *trnS*(gcu), and *trnR*(acg) also recognize them.

Analysis of codon usage in conserved genes and unique ORFs indicates that Chara mtDNA is more like Marchantia mtDNA than Chaetosphaeridium mtDNA. The high similarity of codon usage observed for these mtDNAs also is seen at the level of stop codons. Five of the 11 Chara mitochondrial genes ending with TGA or TAG have counterparts with the same stop codon in Marchantia and/or angiosperm mtDNAs (*atp4*, *atp8*, *rpl5*, *rps12*, and *rps14*); however, none of these genes has an equivalent with the same stop codon in Chaetosphaeridium mtDNA.

At the primary sequence level, conserved genes in Chara mtDNA exhibit more similarity to their Marchantia and/or angiosperm homologs than to their Chaetosphaeridium homologs. This finding is illustrated clearly in Figure 2, which reports the results of overall genome comparisons with MultiPipMaker (Schwartz et al., 2000). It can be seen that for many Chara genes, the greatest scores of sequence similarity are found with Marchantia and/or angiosperm mtDNAs. In addition, for some Chara genes (e.g., *rpl5*, *rpl6*, *rpl16*, and *rps3*), the regions showing similarity to land plant mtDNAs are more extensive than those homologous with Chaetosphaeridium and Mesostigma mtDNAs. In multiple sequence alignments of individual mitochondrial gene or protein sequences, the existence of close affinities between Chara and Marchantia is supported further by the following observations. In the case of four genes varying in the positions of their 5' termini (*cox2*, *nad1*, *rpl2*, and *rps14*), the 5' termini of the Chara genes match precisely those of their Marchantia counterparts. Six Chara genes (*nad2*, *nad3*, *nad4*, *nad5*, *rpl2*, and *yeyR*) feature internal insertions/deletions that have exactly the same lengths as positionally equivalent inser-

tions/deletions in Marchantia genes. Moreover, the Chara and Marchantia mitochondrial *rps3* genes share a deletion of one codon, which is present in all other streptophyte mtDNAs examined here. Unexpectedly, we found that the last 13 codons of the Chara *nad4L* gene are highly divergent from the corresponding sequence in land plant mtDNAs, whereas the 3' termini of the Chaetosphaeridium and land plant *nad4L* genes bear strong similarity. It is not clear why the evolutionary pressure to maintain a high degree of sequence conservation in this region of *nad4L* has been relaxed in the lineage leading to Chara.

Based on the positions of initiation and stop codons in gene sequences, we found no evidence for the editing of Chara mitochondrial transcripts. The positions of these codons in Chara mitochondrial genes show no obvious peculiarities relative to their counterparts in most mitochondrial genes of other eukaryotes. To identify potential editing sites in Chara mtDNA that affect internal codons, we inspected alignments of the deduced amino acid sequences of conserved mitochondrial genes from Chara and other green plants. No sites in Chara sequences were identified that differ from the consensus sites (i.e., the sites exhibited by all other sequences in the alignment, except those edited in Arabidopsis) and that can be converted into the latter sites via conversions of cytidines to uridines or uridines to cytidines. Therefore, these analyses failed to provide any compelling evidence for the presence of potential editing sites in Chara mtDNA.

### Phylogenetic Analysis of Coding Sequences and Pattern of Gene Loss

Various phylogenetic inference methods were used to analyze concatenated data sets of nucleotides (17,370) and amino acids (5470) that were derived from the coding sequences of the 23 protein-coding genes common to the mtDNAs of Chara, Chaetosphaeridium, Mesostigma, Marchantia, Arabidopsis, and sugar beet (see supplemental data online). In all of the analyses, the Mesostigma sequences were used as an outgroup to root the streptophyte trees. For the maximum likelihood and distance-based analyses of the protein sequences, we used the JTT-F model of amino acid replacement, because



**Table 2.** Functions of Conserved Genes in Chara mtDNA

Gene Products <sup>a</sup>	Gene Designations
rRNAs (3)	
Small subunit	<i>rms</i>
Large subunit	<i>rml</i>
5S rRNA	<i>rnr5</i>
tRNAs (26)	<i>trnA(ugc), trnC(gca), trnD(guc), trnE(uuc), trnF(gaa), trnG(gcc), trnG(ucc), trnH(gug), trnI(cau), trnK(uuu), trnL(uaa), trnL(uag), trnM(cau), trnM(cau), trnN(guu), trnP(ugg), trnQ(uug), trnR(acg), trnR(ucg), trnR(ucu), trnS(gcu), trnS(uga), trnT(ggu), trnV(uac), trnW(cca), trnY(gua)</i>
Ribosomal proteins (15)	
Small subunit	<i>rps1, rps2, rps3, rps4, rps7, rps10, rps11, rps12, rps14, rps19</i>
Large subunit	<i>rpl2, rpl5, rpl6, rpl14, rpl16</i>
Electron transport and oxidative phosphorylation (20)	
Respiratory chain	
NADH dehydrogenase	<i>nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad9</i>
Succinate:ubiquinone oxidoreductase	<i>sdh3, sdh4</i>
Ubiquinol:cytochrome c oxidoreductase	<i>cob</i>
Cytochrome c oxidase	<i>cox1, cox2, cox3</i>
ATP synthase	<i>atp1, atp4<sup>b</sup>, atp6, atp8, atp9</i>
Cytochrome c biogenesis (3)	<i>yejR, yejU, yejV<sup>c</sup></i>
Sec-independent protein translocation pathway (1)	<i>mttB<sup>d</sup></i>

<sup>a</sup> Numbers within parentheses indicate the number of genes in a particular class.

<sup>b</sup> Gene identified as *orf183* in Marchantia mtDNA, *orf25* in angiosperm mtDNAs, and *yfm39* in other green plant mtDNAs (Burger et al., 2003).

<sup>c</sup> Different nomenclatures have been used to designate these genes in land plant mtDNAs (for Marchantia and Arabidopsis mtDNAs, see Turmel et al. [1999]; for sugar beet mtDNA, see Kubo et al. [2000]; and for rice mtDNA, see Notsu et al. [2002]).

<sup>d</sup> Gene identified as *orf244* in Marchantia mtDNA, *yfm16* in Mesostigma mtDNA, *tatC* in sugar beet mtDNA, and *orfX* in rice mtDNA.

this model yielded greater likelihood values than the mtREV24-F and WAG-F models in CODEML analyses.

Protein trees constructed with distance, maximum parsimony, and maximum likelihood methods are congruent in showing overwhelming support for the topology in which Chara emerges just before the divergence of the land plants (Figure 3A). The topology of the best maximum likelihood tree that was recovered in PROTML analyses assuming a uniform rate of substitutions across sites (Figure 3A) accounts for 99.99% of the bootstrap samples obtained by resampling of the estimated log likelihood (RELL); the alternative topology recovered in the remaining RELL bootstrap samples identifies Chara and Marchantia in the same clade. The latter topology also is the only alternative topology found in maximum parsimony analyses. On the other hand, analysis of the 105 possible topologies using

CODEML and a  $\Gamma$ -distributed rate of substitutions yields a single topology (Figure 3A), that of the best tree. This topology also is the unique one observed when the 23 individual proteins are assumed to evolve at distinct rates (option Mgene = 0); in these analyses, Cox1 is the slowest evolving protein, with a specific rate of 0.37, and Rps4 is the fastest evolving protein, with a rate of 5.16 (see supplemental data online). As revealed by separate analyses with CODEML (option Mgene = 1), the entire data set is required to obtain unequivocal support for the affinity of Chara with land plants. Thirteen of the 23 proteins investigated, including the slowest and fastest ones, individually support the monophyly of the Charales and land plants, with REML bootstrap values ranging from 23 to 93%, but only three proteins (MttB, Nad4, and Nad5) recover the best topology in >80% of the REML bootstrap samples (see supplemental data online).

Analyses of gene sequences also are congruent and unequivocal in placing Chara immediately before the divergence of the land plants (Figure 3B). Maximum likelihood and maximum likelihood distance trees were constructed under the general time-reversible model of nucleotide evolution (Rodriguez et al., 1990) using a  $\Gamma$  distribution of parameter  $\alpha$  (Yang, 1996) and a proportion of invariable sites, because results from MODELTEST indicated that this model best fits our data set. Maximum likelihood analyses of all three codon positions (Figure 3B) as well as of positions 1 and 2 recovered a single topology: that identified as the best tree in the protein analyses. Maximum parsimony and maximum likelihood-based distance analyses of both sets of codon positions also provide strong and unambiguous support for this same topology.

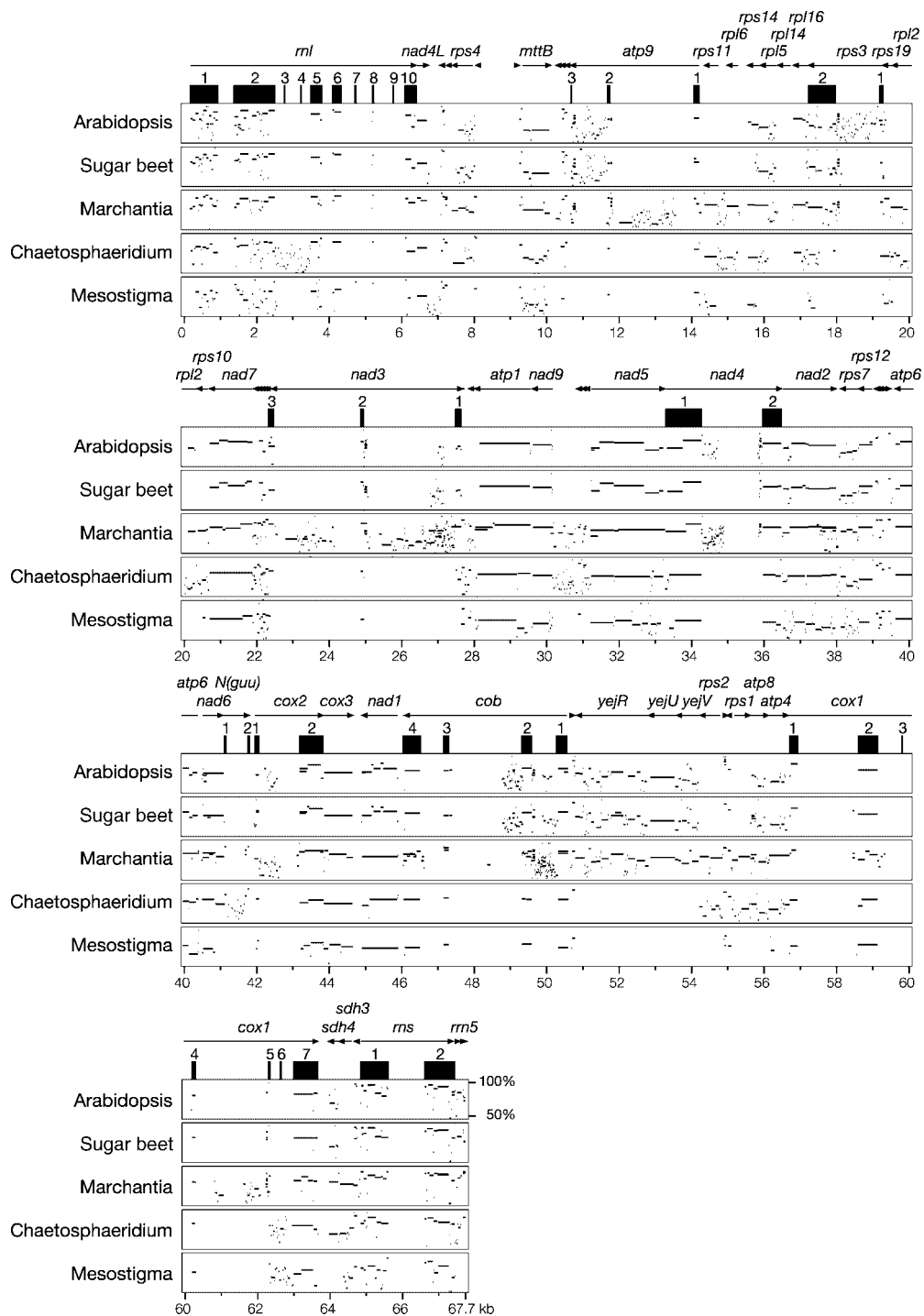
When the distribution of conserved genes in the mtDNAs of the six species examined is mapped on the best tree topology, the pattern of gene losses presented in Figure 4 can be inferred using the principle of maximum parsimony. It can be seen that three genes [*rps8*, *rps13*, and *trnL(caa)*] were lost in the lin-

**Table 3.** Differences between the Mitochondrial Gene Complements of Chara, Other Streptophytes, and Mesostigma

Gene <sup>a</sup>	Mesostigma	Chaetosphaeridium	Chara	Marchantia
<i>nad7</i>	+	+	+	– <sup>b</sup>
<i>rpl2</i>	–	+	+	+
<i>rpl14</i>	+	–	+	–
<i>rps8</i>	–	–	–	+
<i>rps13</i>	+	+	–	+
<i>yejR</i>	–	–	+	+
<i>yejU</i>	–	–	+	+
<i>yejV</i>	–	–	+	+
<i>trnI(cau)</i>	–	+	+	+
<i>trnI(gau)</i>	+	+	+	–
<i>trnL(caa)</i>	+	+	– <sup>b</sup>	+
<i>trnR(ucg)</i>	+	–	–	+
<i>trnS(ucu)</i>	–	+	–	–
<i>trnT(ggu)</i>	–	+	+	+

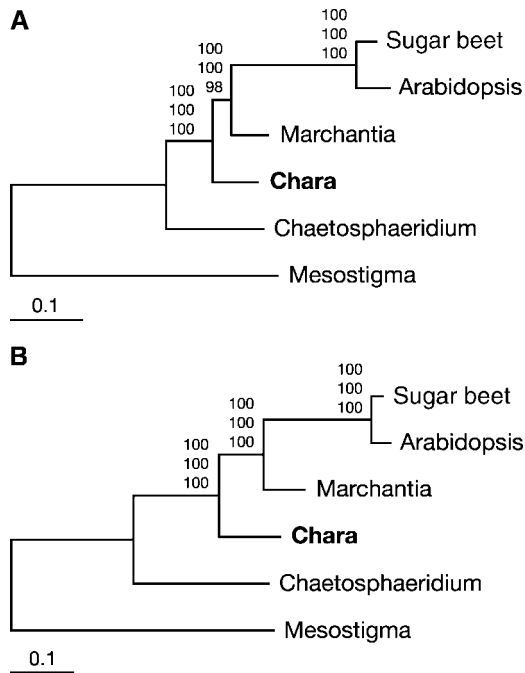
<sup>a</sup> Only the conserved genes that are missing in one or more genomes are indicated. Plus and minus signs denote the presence and absence of genes, respectively.

<sup>b</sup> *trnL(caa)* and *nad7* are pseudogenes in Chara and Marchantia, respectively.



**Figure 2.** MultiPipMaker Analysis of mtDNAs from Chara, Other Streptophytes, and Mesostigma.

The reference genome, Chara mtDNA, was aligned against five other green plant mtDNAs. At the top of the alignment, genes and their polarities are denoted by horizontal arrows, and exons are represented by closed boxes. Positions on the Chara genome are given at the bottom of the alignment. Homologies between aligned regions are shown as average percentage identity between 50 and 100% (the scale is shown at the end of the Arabidopsis plot).



**Figure 3.** Phylogenetic Position of Chara as Inferred from Complete Mitochondrial Genome Sequences of Green Plants.

**(A)** Phylogenetic analysis of a concatenated data set of 23 mitochondrial proteins. The best maximum likelihood tree computed with CODEML using a  $\Gamma$ -distributed rate of substitutions across sites is shown. Bootstrap values obtained in maximum likelihood, distance, and maximum parsimony analyses are indicated above the nodes in the top, middle, and bottom positions, respectively. Bootstrapping in both the maximum likelihood and distance analyses was performed using a  $\Gamma$ -distributed rate of substitutions across sites.

**(B)** Phylogenetic analysis of a concatenated data set of 23 genes. This nucleotide data set corresponds to the protein data set. The best maximum likelihood tree is shown. Bootstrap values obtained in maximum likelihood, distance, and maximum parsimony analyses are indicated above the nodes in the top, middle, and bottom positions, respectively.

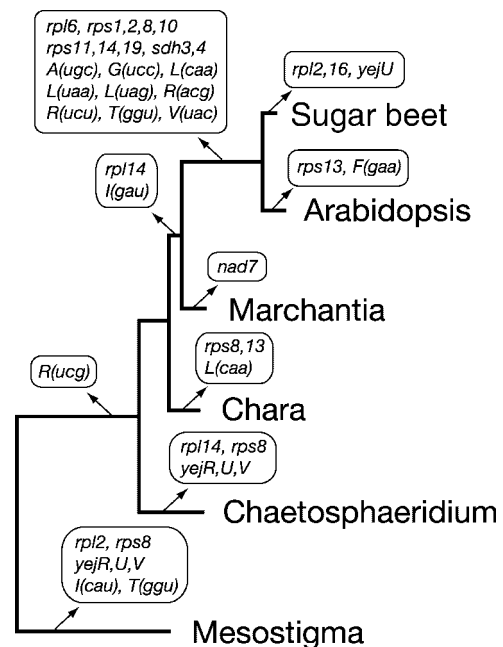
eage leading to Chara but that this charophycean green alga shares no gene loss with its land plant counterparts. These events of gene loss do not represent unique gene losses, because *rps8* has undergone independent losses in two separate green plant lineages and *rps13* and *trnL(caa)* have sustained one independent loss. Note that three tRNA genes with a narrow phylogenetic distribution in green plant mtDNAs [Marchantia *trnR(ucg)*, angiosperm *trnS(gga)*, and Chaetosphaeridium *trnS(acu)*] are not associated with any loss event in the scenario we postulated, because there is reason to believe that they were not inherited ancestrally.

Although *trnR(ucg)* is present in Mesostigma mtDNA, this gene appears to have been lost early during the evolution of the streptophyte mitochondrial genome. It has been proposed that the Marchantia *trnR(ucg)* gene arose from a mutation in the anticodon region of a duplicated copy of *trnR(ucu)* in the mitochondrial genome (Oda et al., 1992b). Considering that one copy of a duplicated segment of Marchantia mtDNA contains *trnR(ucu)* and

*trnY(gua)* and that the second copy contains *trnR(ucg)* and *trnY(gua)*, our finding that the Chara *trnR(ucu)* and *trnY(gua)* genes also lie next to each other on the mitochondrial genome and have the same polarity as their Marchantia homologs offers independent support for the origin of the Marchantia *trnR(ucg)* gene from *trnR(ucu)*. With regard to the *trnS(gga)* gene present in angiosperm mtDNAs, there is evidence that this gene originated from transfer of the homologous chloroplast gene (Joyce and Gray, 1989). Other cases of interorganellar transfer of tRNA genes have been reported in angiosperms (Joyce and Gray, 1989; Unseld et al., 1997; Kubo et al., 2000; Notsu et al., 2002); however, in each of these cases, the transferred chloroplast gene replaced an existing mitochondrial copy. Finally, we consider that the *trnS(acu)* gene of Chaetosphaeridium represents a gain, because this tRNA gene has not been identified in any previously sequenced green plant mtDNAs.

### Gene Order

We reported previously that Chaetosphaeridium and Marchantia mtDNAs share 15 blocks of colinear sequences and that 27



**Figure 4.** Phylogenetic Distribution of Mitochondrial Gene Losses in the Streptophyta.

Events of gene loss inferred from the mitochondrial genome sequences of six green plants were mapped on the tree shown in Figure 3A. A repertoire of 72 genes was predicted for the mitochondrial genome of the green algal ancestor of Mesostigma and the five streptophytes examined. This repertoire includes *rps8*, *rps13*, *trnL(caa)*, and *trnR(ucg)* in addition to the genes found in Chara mtDNA (Table 2). In the lineage leading to angiosperms, the tRNA genes that were lost after the insertion of homologous chloroplast DNA sequences are not indicated. Note also that intron-encoded genes were not considered in our analysis and that there is evidence that the *sdh* and all *rps* genes, except *rps2* and *rps11*, were lost independently in the sugar beet and Arabidopsis lineages (Adams et al., 2002).

inversions (using 34 different end points) would be required to interconvert the gene orders of these genomes (Turmel et al., 2002c). Here, we found that Chara mtDNA more closely resembles Marchantia mtDNA with respect to gene order than does its Chaetosphaeridium homolog. The former mtDNAs share 16 blocks of colinear sequences of up to 10 genes (Figure 1). Only 9 of the 68 conserved genes of Chara are not contained within common blocks. Using the program DERANGE, we inferred that the differences in gene order between Chara and Marchantia mtDNAs are attributable to 20 inversions that share 25 end points. Fifteen of these end points (i.e., 60%) are associated with tRNA genes; in the Chaetosphaeridium/Marchantia mtDNA comparison, the proportion of inversion end points associated with tRNA genes was slightly greater (79%).

With respect to the Chara/Chaetosphaeridium mtDNA comparison, the extent of gene scrambling was found to be comparable to that exhibited by Chaetosphaeridium and Marchantia mtDNAs. Chara and Chaetosphaeridium mtDNAs share 12 blocks of colinear sequences that contain 47 of the conserved genes found in Chara, and a total of 28 inversions (using 34 different end points) account for the rearrangements they display.

### Group-I Introns

The 14 Chara group-I introns in Chara mtDNA can be classified into either subgroup IA or subgroup IB (Michel and Westhof, 1990) (see supplemental data online). Two IA and seven IB introns interrupt the large subunit rRNA gene (*rnl*), four IB introns are found in *cox1*, and the remaining IA intron resides in the small subunit rRNA gene (*rns*). An ORF that encodes a protein of 259 amino acids with two copies of the LAGLIDADG motif is present in the latter intron. All Chara group-I introns, with the exception of *rnl.i4* and *rnl.i7*, are inserted at the same positions as are previously reported group-I introns (Figure 5; see also supplemental data online). Previously, the sites occupied by *cox1.i3*, *rnl.i1*, *rnl.i8*, and *rnl.i9* were known to be restricted to fungal mitochondria, whereas those occupied by *rnl.i2*, *rnl.i5*, and *rns.i1* were observed only in the chloroplasts of chlorophytes. The Chara *cox1.i2*, *cox1.i5*, *cox1.i6*, and *rnl.i3* introns have counterparts in streptophytes; however, only the *cox1.i2* insertion site has been identified in land plants, notably in Marchantia, *Peperomia*, and many other angiosperms, as shown in Figure 6 (Ohta et al., 1993; Cho et al., 1998; Cho and Palmer, 1999). The latter insertion site is shared as well by chlorophyte and nongreen plant introns. All angiosperm *cox1* introns at this site have been found to be related specifically to their fungal counterparts (Cho et al., 1998). The phylogenetic analysis of *cox1* intron sequences presented in Figure 6 corroborates this observation and shows that Chara *cox1.i2* exhibits more affinity with its Marchantia and chlorophyte homologs than with its angiosperm and fungal homologs.

The great majority of Chara mitochondrial group-I introns sharing insertion sites with other organelle introns belong to the same subgroup as their counterparts. Only Chara *rnl.i9* and its *Agroclybe aegerita* homolog represent a notable exception. The structural similarities between Chara *rns.i1* and its *Chlamydomonas pallidostigmatica* chloroplast counterpart extend to their ORFs. These intron ORFs lie at exactly the same position

(L8) relative to the intron secondary structure and display substantial sequence identity. Given that the *Chlamydomonas* chloroplast ORF has been shown to encode a homing endonuclease specific for the insertion site of the corresponding intron (Turmel et al., 1995), the Chara mitochondrial intron probably encodes an endonuclease with a similar sequence specificity.

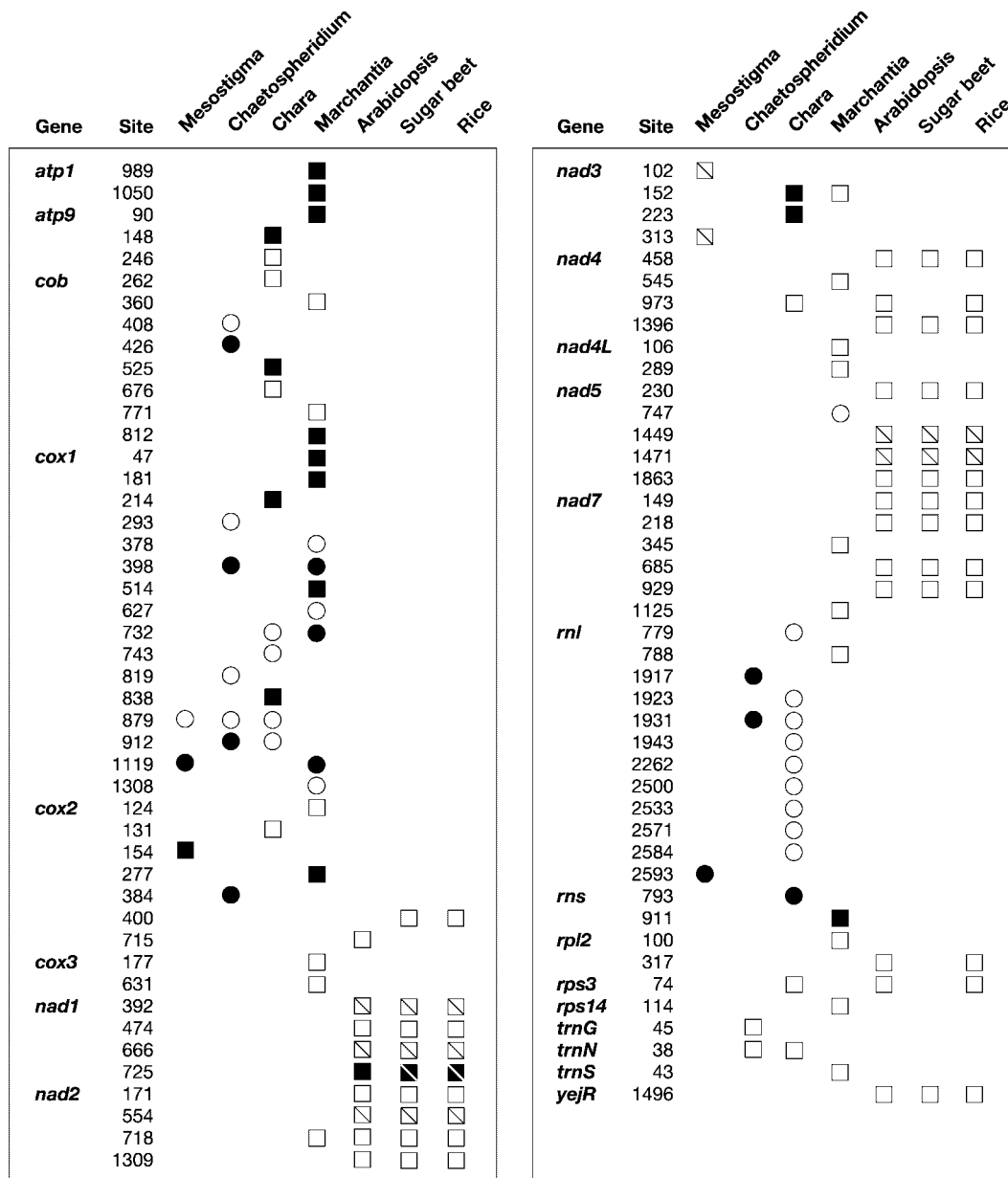
In addition to confirming the close relationships between some Chara introns (*cox1.i2*, *cox1.i3*, *cox1.i6*, and *rnl.i5*) and their counterparts in other eukaryotes, phylogenetic analyses of IB introns revealed that Chara *rnl.i4*, *rnl.i5*, and *rnl.i6* are closely related (data not shown). The latter introns and *C. pallidostigmatica* chloroplast *rnl.i3* were found to form a well-supported clade in distance and maximum likelihood analyses. Chara *rnl.i2* also shows affinity with these four introns but to a lesser degree. For the three IA introns, similar phylogenetic analyses revealed a close relatedness between Chara *rnl.i1* and *rnl.i7*.

### Group-II Introns

The 13 group-II introns in Chara mtDNA are all *cis* splicing and reside in eight genes. The *cox2*, *nad4*, *rps3*, and *trnN(guu)* genes display a single insertion site; *atp9*, *cox1*, and *nad3* exhibit two sites; and *cob* contains three sites. The main features of these introns are reported in the supplemental data online. On the basis of the diagnostic features of subgroups IIA and IIB (Michel et al., 1989), 12 introns were classified in subgroup IIA and one (*cob.i2*) was classified in subgroup IIB. Six introns contain a long ORF coding for a multidomain protein, including one domain related to reverse transcriptase and one related to maturase function (domain X). None of the intron ORFs carries the zinc finger-like/HNH domain, which is required for the endonuclease activity of group-II introns. The catalytic motif of the reverse transcriptase domain in the *cox1.i1*-encoded protein (YEKD) deviates substantially from the consensus sequence (YADD) (Zimmerly et al., 2001), suggesting that this protein lacks reverse transcriptase activity. Only the *nad3.i1* ORF is in frame with the 5' exon, a characteristic of most group-II intron ORFs from fungi and Marchantia (Zimmerly et al., 2001).

Five Chara group-II introns are inserted at the same positions as previously identified group-II introns. Four have counterparts in streptophytes (Figure 5): *nad3.i1* is inserted at the same position as Marchantia *nad3.i1*; *nad4.i1* shares the insertion sites of Arabidopsis and rice *nad4.i2* as well as those of introns in the mosses *Timmia bavarica* and *Takakia lepidozioides* (Pruchner et al., 2001); *rps3.i1* is inserted at the same site as Arabidopsis and rice *rps3.i1*; and *trnN.i1* resides at the same position as Chaetosphaeridium *trnN.i1*. The remaining Chara group-II intron, *cox1.i4*, has its counterparts in the brown alga *Pylaiella littoralis* and the diatom *Thalassiosira nordenskioldii*. All of these IIA introns are not only positionally homologous, they also are closely related at both the structural and primary sequence levels. As shown in the phylogenetic tree in Figure 7, each of the five Chara introns and its streptophyte counterpart form a highly supported clade, with the Chara and Marchantia *nad3* introns displaying the most relatedness. The core structures of the latter introns are strikingly similar, sharing 86% sequence identity in the alignable regions covering 506 nucleotides (Fig-





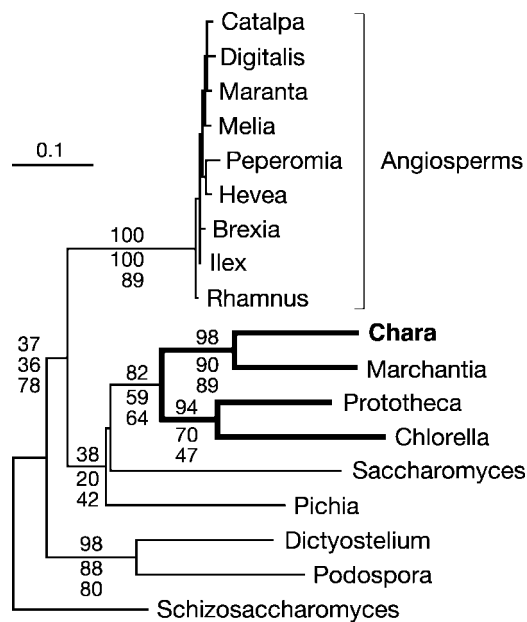
**Figure 5.** Distribution of Introns in Streptophyte and Mesostigma mtDNAs.

Circles denote the presence of a group-I intron, and squares denote the presence of a group-II intron. Divided squares represent *trans*-spliced group-II introns. Open symbols denote the absence of an ORF, and closed symbols denote its presence. For protein-coding and tRNA genes, insertion sites are given relative to the corresponding genes in *Reclinomonas americana*; for rRNA genes, they are given relative to the corresponding genes in *Escherichia coli*. For each site, the position corresponding to the nucleotide immediately preceding the intron is reported. The anticodons of *trnG*, *trnN*, and *trnS* are ucc, guu, and gcu, respectively.

ure 8). In this context, it should be mentioned that unlike its Chaetosphaeridium homolog (Turmel et al., 2002c), which lacks domain VI, Chara *trnN.i1* displays a standard domain-VI structure.

To explore the relationships between Chara IIA introns and the possible affinities of these introns with those of Marchantia, we included in our phylogenetic analysis all Chara IIA introns and 15

Marchantia IIA introns that could be aligned unambiguously with the former introns (Figure 7). The results clearly show that some Chara introns are closely related to one another and to some Marchantia introns. Chara *nad3.i2*, *rps3.i1*, and *atp9.i2* cluster together to form a highly supported clade that is included within a larger clade that also receives strong support and features other Chara introns (the clustered *nad4.i1* and *cob.i1* introns and



**Figure 6.** Phylogenetic Relationships between Chara *cox1.i2* and Homologous Introns Inserted at the Same Site in the *cox1* Gene.

A data set of intron sequences originating from site 732 of the *cox1* gene (177 sites corresponding to unambiguously aligned regions of the intron core) was analyzed using distance, maximum parsimony, and maximum likelihood methods. The distance and maximum likelihood analyses were performed using the Hasegawa-Kishino-Yano model and a uniform rate of substitutions across sites. The unrooted majority-rule distance tree is shown. Bootstrap values obtained in distance, maximum parsimony, and maximum likelihood analyses after 100 replications are indicated above selected nodes in the top, middle, and bottom positions, respectively. The clade containing introns from Chara, Marchantia, and two chlorophytes is highlighted with thick lines.

also *nad3.i1* present in a separate cluster) as well as Marchantia introns (the clustered *nad2.i1*, *cox2.i2*, and *nad4L.i2* introns and *cob.i2* and *nad3.i1* in separate clusters). The Marchantia introns that fall within strongly supported clades have been identified previously as close relatives (Ohshima et al., 1993).

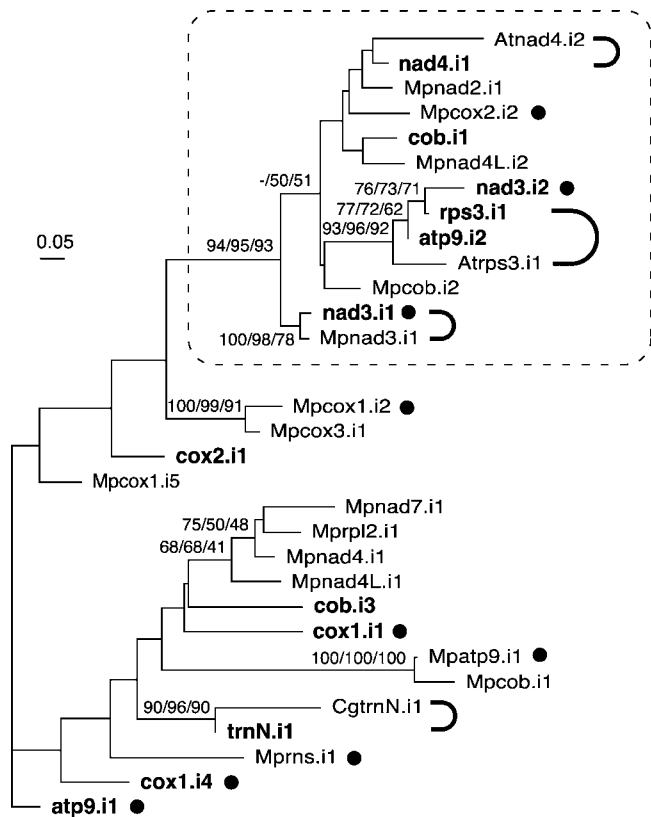
## DISCUSSION

### Extensive Conservation of Genomic Features in Chara and Marchantia mtDNAs

The major finding that emerges from the complete mtDNA sequence of Chara is the remarkable similarity of this genome to its Marchantia counterpart in terms of gene content, A+T composition, gene sequences, codon usage, and gene order. At all of these levels, Chara mtDNA clearly is more like Marchantia mtDNA than its charophycean Chaetosphaeridium counterpart. Several of the genomic features shared between Chara and Marchantia mtDNAs, including the presence of common pairs of overlapping genes (*rps12-rps7* and *atp4-cox1*) and identical gene clusters (Figure 1), represent derived characters, which

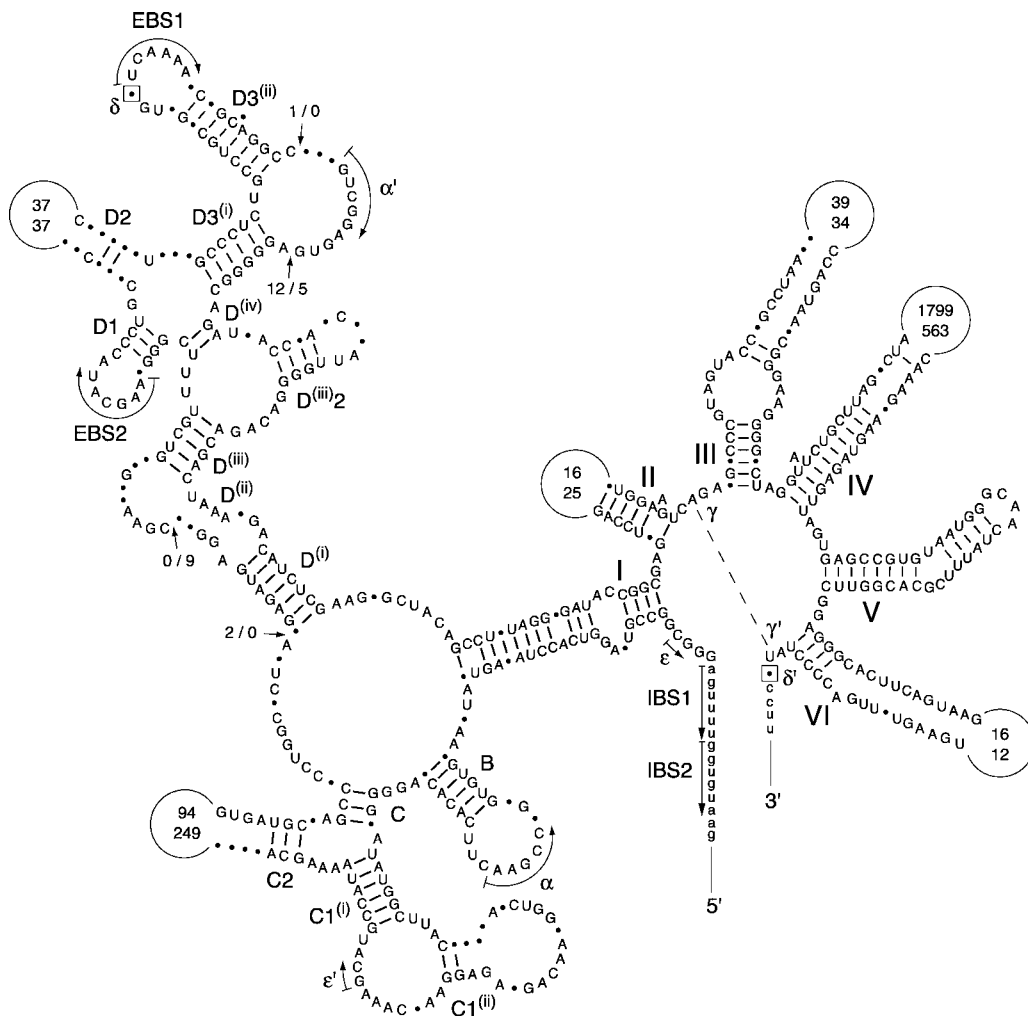
likely were harbored by the common ancestor of the Charales and land plants. Together with the results of our phylogenetic analyses (Figure 3), these observations strongly support the idea that the Charales rather than the Coleochaetales diverged just before the emergence of the land plants.

Chara mtDNA shares 65 genes with Marchantia mtDNA, including the three genes involved in cytochrome *c* biogenesis



**Figure 7.** Phylogenetic Analysis of Group-IIA Mitochondrial Introns from Chara and Marchantia and of Selected Introns from Chaetosphaeridium and Arabidopsis.

All Chara and Marchantia introns that were identified as belonging to group IIA were analyzed, whereas only those of Chaetosphaeridium and Arabidopsis that share insertion sites with their Chara counterparts were examined. Introns originating from Chara are indicated in boldface; Marchantia, Chaetosphaeridium, and Arabidopsis introns are distinguished by the prefixes Mp, Cg, and At, respectively. The introns whose names are followed by closed circles contain ORFs. The introns that are connected by thick, curved lines share common insertion sites. The nucleotide data set (122 sites corresponding to unambiguously aligned regions of the intron core) was analyzed using distance, maximum parsimony, and maximum likelihood methods. The distance and maximum likelihood analyses were performed using the Hasegawa-Kishino-Yano model and a uniform rate of substitutions across sites. The unrooted maximum likelihood tree is shown. Bootstrap values obtained in distance, maximum parsimony, and maximum likelihood analyses after 100 replications are indicated above the nodes in the top, middle, and bottom positions, respectively; only values of >50% are shown. A highly supported clade containing 13 introns is framed.



**Figure 8.** Compared Secondary Structure Models of the *Chara* and *Marchantia nad3* Introns.

Exon sequences are shown in lowercase letters. Roman numerals specify the major structural domains of the introns, and uppercase letters followed by numbers denote the helices in domain I. Dashed lines, curved arrows, and/or Greek letters represent tertiary interactions. EBS and IBS indicate exon binding and intron binding sites, respectively. Nucleotides that potentially participate in the  $\delta$ - $\delta'$  interaction are boxed. Positions exhibiting different nucleotides are denoted by dots. Positions showing deletions/additions are denoted by arrows labeled with two numbers separated by a dash; the first number indicates the number of nucleotides inserted in the *Chara* sequence. Conserved base pairings are represented by dashes. The numbers inside the variable loops indicate the sizes of these loops in the compared introns, with the top number denoting the size of the *Chara* sequence.

(*yejR*, *yejU*, and *yejV*) that are absent from *Mesostigma*, *Chaetosphaeridium*, and all chlorophyte mtDNAs. Losses of tRNA genes and ribosomal protein genes in the lineages leading to *Chara*, *Marchantia*, or the common ancestor of all land plants largely account for the differences in gene composition between *Chara* and *Marchantia* mtDNAs (Figure 4). Interestingly, losses of genes belonging to these two classes also explain most of the gene compositional differences between *Chara*, *Chaetosphaeridium*, and *Mesostigma* mtDNAs. Moreover, it is worth noting that ribosomal protein genes have sustained frequent losses from the mitochondrion during angiosperm and primate evolution (Gray et al., 1998; Lang et al., 1999; Adams et al., 2002). Considering that the *yej* genes are missing in most of the green algal lineages examined to date and that they also are absent in other groups of algae as well

as in most protist mtDNAs (Gray et al., 1998; Lang et al., 1999), one wonders why their presence is so stable in land plants. This intriguing distribution pattern may reflect two opposing trends: the high volatility of the *yej* genes in most mtDNAs, and the strong selection pressure to keep them in land plant mitochondria and their most closely related green algal ancestors. On the other hand, we cannot entirely exclude the possibility that the highly biased distribution of the *yej* genes in green plant lineages reflects a late event of lateral transfer in the Streptophyta. A survey of various streptophyte mtDNAs for the presence of *yejR*, *yejU*, and *yejV* would elucidate the evolutionary dynamics of these genes in the Streptophyta.

Our finding that no more than 20 inversion events account for the differences in gene orders between *Chara* and *Marchan-*

tia mtDNAs agrees with our recent comparison of Chaetosphaeridium and Marchantia mtDNAs and supports the idea that the mitochondrial genome is relatively stable in late-diverging lineages of charophycean green algae (Turmel et al., 2002c). This observation is in contrast to the highly scrambled gene order displayed by the other algal mtDNA sequences available to date. The reduced number of repeated sequences in Chara, Chaetosphaeridium, and Marchantia mtDNAs relative to that in their angiosperm counterparts likely accounts for the substantial conservation of gene order in the former genomes. The presence of repeat sequence families in angiosperm mtDNAs contributes to the fluid structure of these genomes by facilitating recombinational exchanges (Palmer et al., 2000).

The alignments of the deduced amino acid sequences of conserved mitochondrial genes from Chara and other green plants revealed no potential editing sites in this charophycean green alga. This observation suggests that the mitochondrial transcripts of Chara are not edited. If proven by cDNA analysis, this hypothesis will suggest that editing of transcripts in land plant mitochondria most likely emerged during the evolution of early land plants.

#### A Sister-Group Relationship between the Charales and Land Plants

Our phylogenetic analyses of concatenated nucleotide and amino acid data sets provide overwhelming support for the hypothesis that the Charales rather than the Coleochaetales diverged just before the emergence of the land plants (Figure 3). Maximum likelihood analyses of both data sets assuming a  $\Gamma$ -distributed rate of substitutions across sites recovered a single tree, the topology of which was observed in all bootstrap samples examined in distance and maximum parsimony analyses. Likewise, the existence of a sister-group relationship between the Charales and land plants is supported strongly by the four-gene analysis of Karol et al. (2001), which included representatives from all streptophyte lineages. Our separate maximum likelihood analysis clearly shows that several mitochondrial proteins are required to identify unequivocally the relationships between the Charales, Coleochaetales, and land plants. Eight of the 23 proteins investigated (Atp6, Cox1, MttB, Nad2, Nad4, Nad5, Nad6, and Rps4) individually support the monophyly of Charales and land plants (see supplemental data online) but cannot reject alternative hypotheses. Of these proteins, the one encoded by the single mitochondrial gene analyzed by Karol et al. (2001), Nad5, provides the highest bootstrap support (93%) for the sister-group relationships between the Charales and land plants.

#### A Distinctive Intron Composition in Chara mtDNA

Although Chara mtDNA resembles its land plant counterparts in harboring numerous introns, only 4 of its 27 introns are positionally and structurally homologous with land plant mitochondrial introns (Figure 5). On the other hand, 4 of the 11 introns in Chaetosphaeridium mtDNA have homologs in Chara mtDNA, but none is related to any land plant mitochondrial introns. Again, this distribution pattern supports a close affiliation of

Chara with land plants and agrees with the hypothesis that the Charales are the closest green algal ancestors of land plants.

The finding of closely related introns at common sites in Chara and land plant mtDNAs strongly suggests that two mitochondrial introns (homologs of the Chara group-I intron *cox1.i2* and the group-II intron *nad3.i1*) in the common ancestor of the Charales and land plants were transmitted vertically to Marchantia and that two others (homologs of the Chara group-II introns *nad4.i1* and *rps3.i1*) gave rise to their angiosperm homologs through vertical descent. Note that homologs of Chara *nad4.i1* also have been found in mosses (Pruchner et al., 2001). Differential losses of mitochondrial introns in the Marchantia and angiosperm lineages must be invoked to explain why none of the four introns of charophycean ancestry mentioned above is shared by these land plant lineages. In this context, it is possible that the single intron shared by Marchantia and angiosperm mtDNAs (the group-II intron at site 718 in *nad2*) arose within the Charales and was lost subsequently in the Chara lineage. Other possible scenarios of vertical inheritance followed by intron loss can be envisioned from Figure 5; for example, the *cox1* intron at site 398 in Marchantia mtDNA, which occupies the same position as Chaetosphaeridium *cox1.i2*, might have been inherited by vertical descent from a common ancestor of the Coleochaetales and Charales and then disappeared in the Chara lineage. Several cases of intron loss have been documented for land plant species (Qiu et al., 1998; Pruchner et al., 2002). Examination of other members of the Charales and of other bryophytes for their mitochondrial intron compositions would provide more information on the charophycean progenitors of land plant introns and on the origin of the Chara introns.

Considering that 12 of the 14 group-I intron insertion sites in Chara mtDNA have been reported previously in other organelle genomes and that only three (*cox1.i5*, *cox1.i6*, and *ml.i2*) are predicted to have been inherited vertically, horizontal transfers of foreign introns might have been responsible for the origins of many Chara mitochondrial group-I introns. The introns that participated in such transfers might have originated from the mitochondria of fungi and chlorophytes and from the chloroplasts of several chlorophytes (see supplemental data online). Horizontal transfers of group-I introns, and especially those that encode a homing endonuclease, are thought to have been pervasive events during the evolution of the mitochondrial genome, accounting notably for some of the Marchantia *cox1* introns (Ohta et al., 1993) and for the *cox1* intron that occupies site 732 in many angiosperm mtDNAs (Cho et al., 1998; Cho and Palmer, 1999). An organelle group-I intron encoding a homing endonuclease is predicted to undergo efficient horizontal transfer after invading the organelle of a different cell, because homing allows it to readily insert into its cognate site in all organelle DNA molecules and be maintained at this site (Lambowitz and Belfort, 1993). Given the absence of ORFs in all of the Chara group-I introns that are likely to have been acquired by horizontal transfer, with the exception of *rns.i1*, it is possible that homing has played a role in the origins of these introns and that the homing endonuclease gene disappeared soon after their mtDNA insertion. Supporting this hypothesis is the presence of an endonuclease gene or ORF in many introns positioned identically in other organelle genomes.

With regard to the two Chara group-I introns representing novel insertion sites (*rnl.i4* and *rnl.i7*), their close relationships to other introns in Chara mtDNA (*rnl.i5* and *rnl.i1*, respectively) suggests that they arose from intragenomic transposition of the latter introns. Group-I intron transposition is thought to occur mainly by reverse self-splicing of intron RNA into a different mRNA (a reaction requiring only a very short RNA sequence that base pairs to the 5' intron sequence to form P1), followed by reverse transcription and homologous DNA recombination (Woodson and Cech, 1989). Consistent with this mechanism is our finding that sequence similarity between the exons that flank *rnl.i4* and *rnl.i5* and between those that flank *rnl.i7* and *rnl.i1* is restricted mainly to the 5' region included in P1 (data not shown).

Most of the group-II introns in Chara mtDNA appear to have originated from intragenomic transposition events. Apart from the three introns that we postulate to have been inherited from a common ancestor of the Charales and land plants, only two (*cox1.i4* and *trnN.i1*) are structurally and positionally homologous with previously described mitochondrial introns. The *cox1.i4* intron can be attributed to horizontal transfer of a mobile group-II intron from a brown alga or a diatom, whereas *trnN.i1* probably was acquired by vertical descent from a common ancestor of the Coleochaetales and Charales (see supplemental data online). The hypothesis that intragenomic transposition events gave birth to most Chara group-II introns is in agreement with our finding that the core sequences of several introns from this green alga are closely related, forming strongly supported clades in phylogenetic analyses (Figure 7). The origins of several Marchantia mitochondrial group-II introns also have been attributed to intragenomic transposition (Ohyama et al., 1993). Interestingly, a number of Chara and Marchantia introns cluster within the same clade (Figure 7), suggesting that some of the Marchantia introns arose from mitochondrial introns that were present in the last common ancestor of Chara and Marchantia but disappeared subsequently from the Chara or Marchantia lineage after transposition to novel mtDNA sites.

It appears that the ancestral group-II intron(s) that gave rise to the Chara and Marchantia introns encoded a multifunctional protein that interacted with intron RNA and flanking exon sequences to promote the insertion of the intron(s) into homologous sites (retrohoming) (Bonen and Vogel, 2001). This hypothesis is supported by the distribution pattern of the intron ORFs shown in Figure 7 and by a comparative analysis of the corresponding protein products, which revealed relationships that are congruent with those inferred from the intron core sequences (data not shown). Earlier evidence that group-II intron core structures coevolve with their intron ORFs came from two studies (Fontaine et al., 1997; Zimmerly et al., 2001); interestingly, the more comprehensive of these studies led to the proposal that most ORF-less introns are derivatives of mobile introns and that the shared evolutionary history of the ribozyme core and the intronic ORF predates the split into subgroups IIA and IIB (Zimmerly et al., 2001).

Group-II introns also occasionally can insert into novel genomic sites through reverse transcriptase-mediated movement (retrotransposition). One of the two possible pathways of ret-

rotransposition involves reverse splicing of the intron RNA (as a ribonucleoprotein) into a cellular mRNA, a step that requires relaxed complementarity between the target site and the exon binding sites of the intron (Bonen and Vogel, 2001). In the second pathway, the group-II intron (as a ribonucleoprotein) is inserted directly into a DNA site that closely resembles the homologous site by a mechanism analogous to retrohoming. Analyses of 5' and 3' sequences flanking Chara and Marchantia mitochondrial introns are compatible with the idea that the spread of group-II introns in green plant organelle genomes occurred by the two retrotransposition pathways mentioned above (Ohyama et al., 1993; this study). In our analysis of Chara introns, we found that *atp9.i1* and *cox1.i4* provide the most compelling evidence for retrotransposition by direct insertion of intron RNA into DNA; the core and ORF sequences of these introns are closely related, and significant portions of their flanking exon sequences are conserved (the 40 nucleotides immediately preceding and following these introns show 47.5 and 50% sequence identity, respectively).

### A Highly Compact Mitochondrial Genome in Chara

With its coding sequences representing 68 conserved genes and 91% of the genome, Chara mtDNA is the most compact of the streptophyte mtDNAs analyzed to date (Table 1). These genome features reinforce the hypothesis that the green algal ancestor of land plants possessed a densely packed mitochondrial genome with a gene content similar to that of its Marchantia homolog (Turmel et al., 2002c). As a corollary, we conclude that the pressure to maintain a compact genome became relaxed during the evolution of early land plants, leading to the large and loosely packed mitochondrial genome characteristic of all land plant mtDNAs investigated to date.

### Conclusion

The completely sequenced and annotated mtDNA of Chara represents an important source of information for our understanding of mitochondrial genome evolution in the Streptophyta lineage. Given the sister-group relationship between the Charales and land plants, this genome sequence offers significant insights into the mtDNA architecture of the green algal ancestor of all land plants and into the events that shaped the mitochondrial genome during the transition from aquatic algae to terrestrial plants.

### METHODS

#### DNA Isolation

Plant material (the green alga *Chara vulgaris*) was collected from a pond located in Quebec City; a voucher (number QFA468020) of this material is held at the Herbarium Louis-Marie of Laval University. A fraction containing chloroplast DNA and mitochondrial DNA (mtDNA) was obtained using the isolation procedure described by Palmer (1986) with minor modifications. Organelles were purified by banding, in a two-step gradient with sucrose concentrations of 30 and 52% (w/w), a cell homogenate that was prepared with a 4:1 (v/w) ratio of isolation buffer (0.35 M sorbitol, 50 mM Tris-HCl, pH 8.0, 5 mM EDTA, 0.1% [w/v] BSA, and 0.1%



[v/v] 2-mercaptoethanol) to wet material. The centrifugation conditions were as reported previously (Palmer, 1986). The organelle fraction at the 30%-52% interface was diluted with seven volumes of a buffer containing 0.35 M sorbitol, 50 mM Tris-HCl, pH 8.0, and 25 mM EDTA, and after centrifugation, the pellet was resuspended in the same buffer. The suspension was incubated in the presence of 1 mg/mL proteinase K and 1% (w/v) sodium sarcosinate at 37°C for 15 min. The lysate then was centrifuged in a CsCl-ethidium bromide gradient (1.55 g/mL and 100 µg/mL ethidium bromide) in a NVT-65.2 rotor (Beckman Instruments, Fullerton, CA) for 16 h at 50,000 rpm, and the major DNA band (the top band) was collected. After extraction with isopropyl alcohol saturated with 3 M NaCl, this fraction was diluted with one volume of TE (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA), DNA was pelleted by sedimentation in a TH-641 rotor (DuPont Company, Wilmington, DE) for 12 h at 35,000 rpm, and the pellet was dissolved in TE.

### DNA Cloning and Sequencing

A random clone library was prepared from the Chara organelle DNA fraction as described previously (Lemieux et al., 2000). DNA was sheared by nebulization, and 1500- to 2000-bp fragments were recovered by electroelution after agarose gel electrophoresis. These fragments were treated with *Escherichia coli* Klenow fragment and T7 DNA polymerase and cloned into the SmaI site of Bluescript II KS+ (Stratagene). After hybridization of the clones with the original DNA used for cloning, DNA templates from positive clones were prepared with the QIAprep 96 Miniprep kit (Qiagen, Mississauga, Canada). Nucleotide sequences were determined with the PRISM BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA), the PRISM dGTP BigDye terminator ready reaction kit (Applied Biosystems), and the DYE-namic ET terminator cycle sequencing kit (Amersham Pharmacia Biotech, Baie d'Urfé, Canada) on ABI model 373 and 377 DNA sequencers (Applied Biosystems) using T3 and T7 primers as well as oligonucleotides complementary to internal regions of plasmid DNA inserts. Genomic regions not represented in the clones analyzed were sequenced from PCR-amplified fragments. Sequences were assembled using SEQUENCHER 4.1.1 (Gene Codes Corp., Ann Arbor, MI).

### DNA Sequence Analysis

Sequences were analyzed using the Genetics Computer Group (Madison, WI) software (version 10.3) package. Protein-coding and rRNA genes were identified by Basic Local Alignment Search Tool (BLAST) searches (Altschul et al., 1990) of the nonredundant database at the National Center for Biotechnology Information, whereas tRNA genes were found using tRNAscan-SE (Lowe and Eddy, 1997). Repeated sequence elements were searched using REPuter (Kurtz and Schleiermacher, 1999). Introns were modeled according to the nomenclatures proposed for group-I (Michel and Westhof, 1990) and group-II (Michel et al., 1989) introns. The program DERANGE2 (M. Blanchette and D. Sankoff, University of Ottawa, Canada) was used to determine the number of gene permutations by inversions. Pair-wise comparisons of genome sequences were performed using MultiPipMaker (Schwartz et al., 2000). Proteins encoded by Chara group-II introns were compared with other group-II intron-encoded proteins by adding these Chara sequences to the alignment in GenBank accession ALIGN\_000044 (Zimmerly et al., 2001) using the profile alignment option of CLUSTAL W 1.81 (Thompson et al., 1994).

### Phylogenetic Analyses

Mitochondrial genome sequences were retrieved from GenBank for Chara (this study), *Chaetosphaeridium globosum* (Turmel et al., 2002c), *Mesostigma viride* (Turmel et al., 2002b), *Marchantia polymorpha* (Oda et

al., 1992a), *Arabidopsis thaliana* (Unsel et al., 1997), sugar beet (*Beta vulgaris*; Kubo et al., 2000), and rice (*Oryza sativa*; Notsu et al., 2002). A data set consisting of concatenated protein sequences derived from these genomes (Bruno et al., 2000) was prepared as follows: the deduced amino acid sequences from individual genes were aligned using T-COFFEE 1.37 (Notredame et al., 2000), the ambiguously aligned regions in each alignment were removed with GBLOCKS 0.91 (Castresana, 2000), and the protein alignments were concatenated. Maximum likelihood analyses of the concatenated sequences were performed using PROTML in MOLPHY2.3b3 (Adachi and Hasegawa, 1996) and CODEML in PAML 3.12 (Yang, 1997). In these analyses, resampling of the estimated log likelihood bootstrap probabilities was performed after 10,000 replications (Adachi and Hasegawa, 1996). In addition, maximum parsimony and distance trees were inferred from the protein sequences using PROTPARS in PHYLIP 3.6a3 (Felsenstein, 1995) and WEIGHBOR 1.2 (Bruno et al., 2000), respectively, and their robustness was assessed by bootstrap percentages after 100 or 1000 replications. For the distance analyses, maximum likelihood distances were computed with PUZZLE-BOOT 1.03 and TREE-PUZZLE 5.0 (Strimmer and von Haeseler, 1996).

A separate data set consisting of concatenated mitochondrial gene sequences was prepared as follows: individual gene alignments in which the positions of codons perfectly match those of amino acid residues in the protein alignments mentioned above were generated, ambiguously aligned codons from each alignment were removed with GBLOCKS, and the gene alignments were concatenated. Maximum likelihood, distance, and maximum parsimony analyses of the gene data set were performed using PAUP\* 4.0b10 (Swofford, 2000). Trees were searched with the full heuristic option, and optimization was performed by branch swapping using tree bisection and reconnection. For maximum likelihood analyses, starting trees were obtained by the stepwise addition of sequences. For distance analyses, the minimum evolution function was applied for tree reconstruction, and starting trees were obtained by the stepwise random addition of sequences with one tree held per addition. The model of DNA substitutions that best fitted the data was selected using MODELTEST 3.06 (Posada and Crandall, 1998) based on hierarchical likelihood ratio tests and the Akaike information criterion. The proportion of invariable sites,  $\Gamma$  distribution shape parameters (four categories), and base frequencies were estimated with PAUP\* 4.0b10. The confidence of branch points was estimated by 1000 bootstrap replications in distance and maximum parsimony analyses and by 100 bootstrap replications in maximum likelihood analyses.

Data sets of intron sequences were analyzed with PAUP\* 4.0b10 using the same methods, except that a uniform rate of substitutions across sites was used and the distance and maximum likelihood analyses were inferred under the Hasegawa-Kishino-Yano model. The confidence of branch points was estimated by 100 bootstrap replications in all analyses.

Upon request, materials integral to the findings presented in this publication will be made available in a timely manner to all investigators on similar terms for noncommercial purposes. To obtain materials, please contact M. Turmel, monique.turmel@rsvs.ulaval.ca.

### Accession Numbers

The GenBank accession numbers for the sequences mentioned in this article are as follows: *Chara vulgaris*, AY267353; *Chaetosphaeridium globosum*, AF494279; *Mesostigma viride*, AF353999; *Marchantia polymorpha*, M68929; *Arabidopsis thaliana*, Y08501 and Y08502; *Beta vulgaris*, AP000396 and AP000397; and *Oryza sativa*, AB076665 and AB076666.

### ACKNOWLEDGMENTS

We are grateful to Charles F. Delwiche and Kenneth G. Karol for identifying the Chara species analyzed in this study and to Klaus Kowallik for

his comments on the manuscript. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

Received April 22, 2003; accepted June 4, 2003.

## REFERENCES

- Adachi, J., and Hasegawa, M.** (1996). MOLPHY version 2.3: Programs for molecular phylogenetics based on maximum likelihood method. *Comput. Sci. Monogr.* **28**, 1–150.
- Adams, K.L., Qiu, Y.L., Stoutemyer, M., and Palmer, J.D.** (2002). Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer during angiosperm evolution. *Proc. Natl. Acad. Sci. USA* **99**, 9905–9912.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J.** (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
- Beckert, S., Muhle, H., Pruchner, D., and Knoop, V.** (2001). The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: A comparative analysis in mosses. *Mol. Phylogenet. Evol.* **18**, 117–126.
- Beckert, S., Steinhauser, S., Muhle, H., and Knoop, V.** (1999). A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. *Plant Syst. Evol.* **218**, 179–192.
- Bonen, L., and Vogel, J.** (2001). The ins and outs of group II introns. *Trends Genet.* **17**, 322–331.
- Bower, F.O.** (1908). *The Origin of Land Flora: A Theory Based upon the Facts of Alternation*. (London: Macmillan).
- Bruno, W.J., Socci, N.D., and Halpern, A.L.** (2000). Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.* **17**, 189–197.
- Burger, G., Lang, B.F., Braun, H.-P., and Marx, S.** (2003). The enigmatic mitochondrial ORF *ymf39* codes for ATP synthase chain b. *Nucleic Acids Res.* **31**, 2353–2360.
- Castresana, J.** (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–542.
- Cho, Y., and Palmer, J.D.** (1999). Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. *Mol. Biol. Evol.* **16**, 1155–1165.
- Cho, Y., Qiu, Y.L., Kuhlman, P., and Palmer, J.D.** (1998). Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. USA* **95**, 14244–14249.
- Felsenstein, J.** (1995). PHYLIP (Phylogeny Inference Package), version 3.5. (Seattle: Department of Genetics, University of Washington).
- Fontaine, J.M., Goux, D., Kloareg, B., and Loiseaux-de Goër, S.** (1997). The reverse-transcriptase-like proteins encoded by group II introns in the mitochondrial genome of the brown alga *Pylaiella littoralis* belong to two different lineages which apparently coevolved with the group II ribozyme lineages. *J. Mol. Evol.* **44**, 33–42.
- Fritsch, F.E.** (1935). *The Structure and the Reproduction of the Algae*. (London: Cambridge University Press).
- Giegé, P., and Brennicke, A.** (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. USA* **96**, 15324–15329.
- Graham, L.E., Cook, M.E., and Busse, J.S.** (2000). The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proc. Natl. Acad. Sci. USA* **97**, 4535–4540.
- Gray, M.W., et al.** (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* **26**, 865–878.
- Hiesel, R., Combettes, B., and Brennicke, A.** (1994). Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proc. Natl. Acad. Sci. USA* **91**, 629–633.
- Joyce, P.B., and Gray, M.W.** (1989). Chloroplast-like transfer RNA genes expressed in wheat mitochondria. *Nucleic Acids Res.* **17**, 5461–5476.
- Karol, K.G., McCourt, R.M., Cimino, M.T., and Delwiche, C.F.** (2001). The closest living relatives of land plants. *Science* **294**, 2351–2353.
- Kenrick, P., and Crane, P.R.** (1997). *The Origin and Early Diversification of Land Plants: A Cladistic Study*. (Washington, DC: Smithsonian Institution Press).
- Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., and Mikami, T.** (2000). The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic Acids Res.* **28**, 2571–2576.
- Kurtz, S., and Schleiermacher, C.** (1999). REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* **15**, 426–427.
- Lambowitz, A.M., and Belfort, M.** (1993). Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62**, 587–622.
- Lang, B.F., Gray, M.W., and Burger, G.** (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397.
- Lemieux, C., Otis, C., and Turmel, M.** (2000). Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature* **403**, 649–652.
- Lewis, L.A., Mishler, B.D., and Vilgalys, R.** (1997). Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Mol. Phylogenet. Evol.* **7**, 377–393.
- Lowe, T.M., and Eddy, S.R.** (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964.
- Maier, R.M., Zeltz, P., Kossel, H., Bonnard, G., Gualberto, J.M., and Grienerberger, J.M.** (1996). RNA editing in plant mitochondria and chloroplasts. *Plant Mol. Biol.* **32**, 343–365.
- Malek, O., Brennicke, A., and Knoop, V.** (1997). Evolution of *trans*-splicing plant mitochondrial introns in pre-Permian times. *Proc. Natl. Acad. Sci. USA* **94**, 553–558.
- Malek, O., and Knoop, V.** (1998). *Trans*-splicing group II introns in plant mitochondria: The complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort. *RNA* **4**, 1599–1609.
- Malek, O., Lattig, K., Hiesel, R., Brennicke, A., and Knoop, V.** (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.* **15**, 1403–1411.
- Mariefeld, J., Unseld, M., and Brennicke, A.** (1999). The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends Plant Sci.* **4**, 495–502.
- Mattox, K.R., and Stewart, K.D.** (1984). Classification of the green algae: A concept based on comparative cytology. In *The Systematics of the Green Algae*, D.E.G. Irvine and D.M. John, eds (London: Academic Press), pp. 29–72.
- Michel, F., Umesono, K., and Ozeki, H.** (1989). Comparative and functional anatomy of group II catalytic introns: A review. *Gene* **82**, 5–30.
- Michel, F., and Westhof, E.** (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610.
- Mishler, B.D.** (1994). Phylogenetic relationships of the “green algae” and “bryophytes.” *Ann. Mo. Bot. Gard.* **81**, 451–483.
- Notredame, C., Higgins, D.G., and Heringa, J.** (2000). T-COFFEE: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
- Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K.** (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics* **268**, 434–445.

- Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and Ohshima, K. (1992a). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. *J. Mol. Biol.* **223**, 1–7.
- Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., and Ohshima, K. (1992b). Transfer RNA genes in the mitochondrial genome from a liverwort, *Marchantia polymorpha*: The absence of chloroplast-like tRNAs. *Nucleic Acids Res.* **20**, 3773–3777.
- Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohshima, K., and Michel, F. (1993). Group I introns in the liverwort mitochondrial genome: The gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. *Nucleic Acids Res.* **21**, 1297–1305.
- Ohshima, K., Oda, K., Ohta, E., and Takemura, M. (1993). Gene organization and evolution of introns of a liverwort, *Marchantia polymorpha*, mitochondrial genome. In *Plant Mitochondria*, A. Brennicke and U. Kuck, eds (Weinheim, Germany: Verlag Chemie), pp. 115–129.
- Palmer, J.D. (1986). Isolation and structural analysis of chloroplast DNA. *Methods Enzymol.* **118**, 167–186.
- Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L., and Song, K. (2000). Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proc. Natl. Acad. Sci. USA* **97**, 6960–6966.
- Pickett-Heaps, J.D., and Marchant, H.J. (1972). The phylogeny of the green algae: A new proposal. *Cytobios* **6**, 255–264.
- Posada, D., and Crandall, K.A. (1998). MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.
- Pruchner, D., Beckert, S., Muhle, H., and Knoop, V. (2002). Divergent intron conservation in the mitochondrial *nad2* gene: Signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. *J. Mol. Evol.* **55**, 265–271.
- Pruchner, D., Nassal, B., Schindler, M., and Knoop, V. (2001). Mosses share mitochondrial group II introns with flowering plants, not with liverworts. *Mol. Genet. Genomics* **266**, 608–613.
- Qiu, Y.L., Cho, Y., Cox, J.C., and Palmer, J.D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394**, 671–674.
- Rodriguez, F., Oliver, J.F., Marin, A., and Medina, J.R. (1990). The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142**, 485–501.
- Schwartz, S., Zhang, Z., Frazer, K., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. (2000). PipMaker: A Web server for aligning two genomic DNA sequences. *Genome Res.* **10**, 577–586.
- Sper-Whitis, G.L., Moody, J.L., and Vaughn, J.C. (1996). Universality of mitochondrial RNA editing in cytochrome-c oxidase subunit I (*coxI*) among the land plants. *Biochim. Biophys. Acta* **1307**, 301–308.
- Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing. *J. Mol. Evol.* **48**, 303–312.
- Strimmer, K., and von Haeseler, A. (1996). Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969.
- Swofford, D.L. (2000). PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), Version 4. (Sunderland, MA: Sinauer Associates).
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (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., Ehara, M., Otis, C., and Lemieux, C. (2002a). Phylogenetic relationships among streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences. *J. Phycol.* **38**, 364–375.
- Turmel, M., Mercier, J.P., Côté, V., Otis, C., and Lemieux, C. (1995). The site-specific DNA endonuclease encoded by a group I intron in the *Chlamydomonas pallidostigmatica* chloroplast small subunit rRNA gene introduces a single-strand break at low concentrations of Mg<sup>2+</sup>. *Nucleic Acids Res.* **23**, 2519–2525.
- Turmel, M., Otis, C., and Lemieux, C. (1999). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. USA* **96**, 10248–10253.
- Turmel, M., Otis, C., and Lemieux, C. (2002b). The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol. Biol. Evol.* **19**, 24–38.
- Turmel, M., Otis, C., and Lemieux, C. (2002c). The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc. Natl. Acad. Sci. USA* **99**, 11275–11280.
- Unsold, 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.
- Vangerow, S., Teerkorn, T., and Knoop, V. (1999). Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. *Plant Biol.* **1**, 235–243.
- Wolff, G., Plante, I., Lang, B.F., Kück, U., and Burger, G. (1994). Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prothotheca wickerhamii*. *J. Mol. Biol.* **237**, 75–86.
- Woodson, S.A., and Cech, T.R. (1989). Reverse self-splicing of the *Tetrahymena* group I intron: Implication for the directionality of splicing and for intron composition. *Cell* **57**, 335–345.
- Yang, Z. (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* **11**, 367–372.
- Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**, 555–556.
- Zimmerly, S., Hausner, G., and Xu-chu, W. (2001). Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res.* **29**, 1238–1250.