Algae
Volume 17(1): 59-68, 2002

The Chloroplast *rpl23* Gene Cluster of *Spirogyra maxima* (Charophyceae) Shares Many Similarities with the Angiosperm *rpl23* Operon

Jungho Lee* and James R. Manhart

Department of Biology, Texas A&M University, College Station, TX, 77843-3258, U.S.A.

A phylogenetic affinity between charophytes and embryophytes (land plants) has been explained by a few chloroplast genomic characters including gene and intron (Manhart and Palmer 1990; Baldauf *et al.* 1990; Lew and Manhart 1993). Here we show that a charophyte, *Spirogyra maxima*, has the largest operon of angiosperm chloroplast genomes, *rpl23* operon (*trnI-rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rps11-rpoA*) containing both embryophyte introns, *rpl16.i* and *rpl2.i*. The *rpl23* gene cluster of *Spirogyra* contains a distinct eubacterial promoter sequence upstream of *rpl23*, which is the first gene of the green algal *rpl23* gene cluster. This sequence is completely absent in angiosperms but is present in non-flowering plants. The results imply that, in the *rpl23* gene cluster, early charophytes had at least two promoters, one upstream of *trnI* and another upstream of *rpl23*, which partially or completely lost its function in land plants. A comparison of gene clusters of prokaryotes, algal chloroplast DNAs and land plant cpDNAs indicated a loss of numerous genes in chlorophyll a+b eukaryotes. A phylogenetic analysis using presence/absence of genes and introns as characters produced trees with a strongly supported clade containing chlorophyll a+b eukaryotes. *Spirogyra* and embryophytes formed a clade characterized by the loss of *rpl5* and *rps9* and the gain of *trnI* (CAU) and introns in *rpl2* and *rpl16*. The analyses support the hypothesis that the *rpl23* gene cluster and the *rpl2* and *rpl16* introns of land plants originated from a common ancestor of *Spirogyra* and land plants.

Key Words: Charophytes, Chloroplast, embryophytes, Introns, operon, promoter, *rpl2*, *rpl16*, *rpl23*, *Spirogyra*

INTRODUCTION

More than 450 million years ago, land plants evolved from a green algal ancestor (Graham 1993; Kenrick and Crane 1997). A group of green algae called charophytes (the Charophyceae) are generally accepted as the closest algae to land plants based on analysis of phenetic and molecular data (Mishler *et al.* 1994). There are nuclear and chloroplast genome characters that support a close relationship of charophytes and land plants. Charophytes and embryophytes have *tufA* (Baldauf *et al.* 1990) in their nuclear genome while non-charophyte green algae have *tufA* in their chloroplast genome only. Some charophytes have chloroplast *trnA* and *trnI* group II introns (Manhart and Palmer 1990), which are absent in non-charophyte green algae. The charophyte taxa included in these studies were *Spirogyra* and *Sirogonium* of the Zygnematales (Mattox and Stewart sensu 1983), *Coleochaete orbicularis* of the Coleochaetales, and *Chara* and *Nitella* of the Charales. Among these taxa, *Spirogyra* and *Sirogonium* of the Zygnematales do not contain chloroplast *tufA* like land plants while the others have the gene in their chloroplast genomes (Baldauf *et al.* 1990). *Spirogyra* chloroplast DNA (cpDNA) also has an unusual structure of *rps12* and *rps7* (Lew and Manhart 1993), that contains two introns including a trans-splicing intron. This unusual structure of *rps12* and *rps7* has been documented only in *Spirogyra* and land plants.

The question remains whether or not charophyte chloroplasts have additional chloroplast genomic characters that support the hypothesized close relationship between charophytes and embryophytes. The gross structure of *Spirogyra* chloroplast DNA, the only characterized charophyte cpDNA, differs from those of land plants (Palmer 1991) by the absence of the inverted repeats (IRs) and the order of the limited number of mapped genes (Manhart *et al.* 1990). Among land plants, *Pinus* does not contain large IRs (Wakasugi *et al.* 1994). Therefore, *Spirogyra* is worth investigating in detail in order to reveal possible similarities or transitional states

*Corresponding author (jlee@mail.bio.tamu.edu or LeeGreenplant@aol.com)

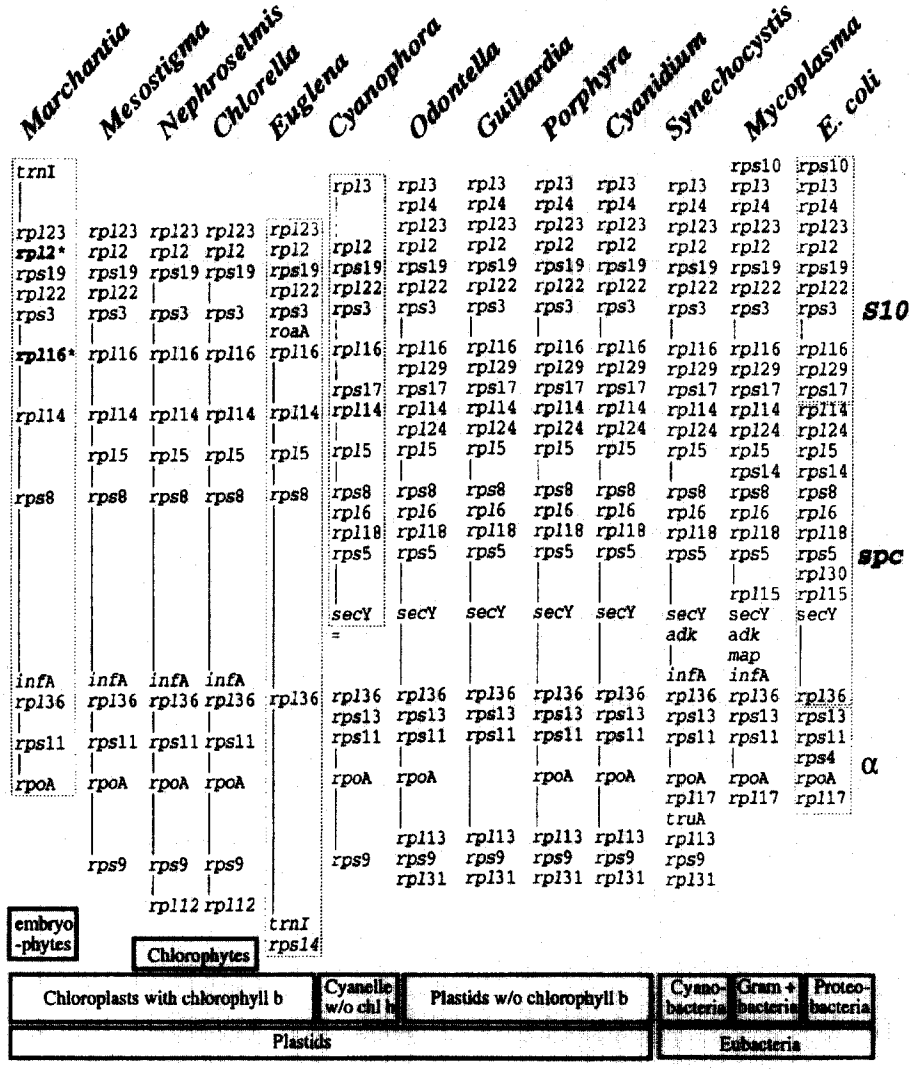

Fig. 1. Alignment of the ribosomal gene cluster (S10, *spc*, α operons) in *E. coli*, *Mycoplasma*, *Synechocystis*, a cyanelle, algal and land plant chloroplasts. [*E. coli* (Blattner *et al.* 1997), *Mycoplasma genitalium* (Fraser *et al.* 1995), *Synechocystis* (Kaneko *et al.* 1996), the cyanelle of *Cyanophora* (Stirewalt *et al.* 1995) and cpDNAs of *Odontella* (Kowallik *et al.* 1995), *Guillardia* (Douglas and Penny 1999), *Porphyra* (Reith *et al.* 1995), *Cyanidium* (Glockner *et al.* 2000), *Euglena* (Christopher and Hallick 1989; Hallick *et al.* 1993), *Chlorella* (Wakasugi *et al.* 1997), *Nephroselmis* (Turmel *et al.* 1999), *Mesostigma* (Lemieux *et al.* 2000) and land plants (see Table 1)] S10, *spc*, and α are three *E. coli* operons.
| : gene absence =: separate

between land plants and non-charophyte green algae.

Land plants have a unique gene cluster, *trnI-rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rps11-rpoA*, that is found in all completely sequenced land plant cpDNAs. In angiosperms, the gene cluster is a transcriptional unit (Kanno and Hirai 1993), the *rpl23* operon (Sugiura 1992). In the bryophyte, *Marchantia*, a distinct eubacterial promoter region is present upstream of *trnI* and there is no other distinct eubacterial promoter region in the *rpl23* gene cluster (Ohyama *et al.* 1988). So it is likely that the *rpl23* gene cluster in *Marchantia* is also a single operon.

The land plant *rpl23* gene cluster shows similarities to three consecutive ribosomal operons (*S10*, *spc*, and *alpha* operons) of *E. coli* (Tanaka *et al.* 1986) with differences in gene and intron content. The genes of the *E. coli* operons are also clustered together in other eubacteria with a few variations in gene contents (Fig. 1). The same gene groupings are found in *Synechocystis*, a cyanobacterium, and its gene contents are more similar than those of *E. coli* to the ribosomal gene cluster of chloroplasts and the cyanelle of *Cyanophora*. The cyanobacterial gene order likely represents the ancestral condition from which the gene clusters found in cyanelle and chloroplast DNAs

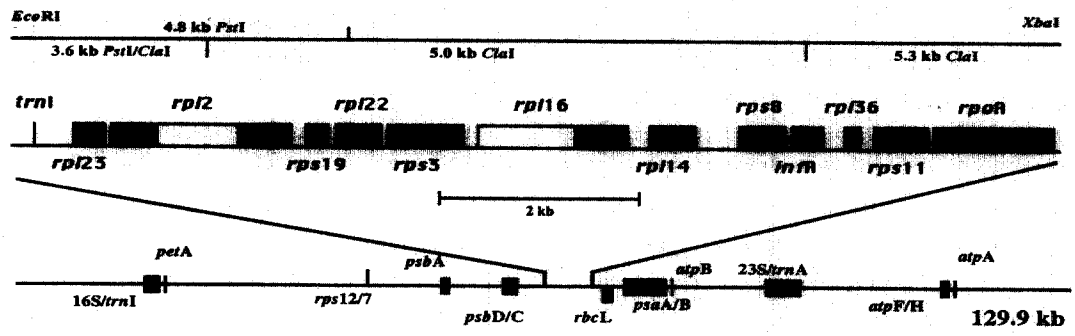

Fig. 2. *rpl23* gene cluster and its physical map in *Spirogyra* cpDNA. Closed boxes indicate genes/exons and open boxes indicate introns.

were derived. Similar gene clusters are found in the cyanelle of the glaucophyte (*Cyanophra paradoxa*) and chloroplast genomes of rhodophytes (*Porphyra purpurea* and *Cyanidium caldarium*), a chromophyte (*Odontella sinensis*), a cryptophyte (*Guillardia theca*). *Porphyra* and *Odontella* cpDNAs contain most of the cyanobacterial genes in the same order (Fig. 1). The cyanelle of *Cyanophra paradoxa* has two separate gene clusters but their combined content is similar to single clusters found in *Porphyra* and *Odontella* cpDNAs. In addition, a chloroplast-like plasmid of *Plasmodium* (McFadden *et al.* 1997) also contains a gene cluster similar to those of *Porphyra* and *Odontella* cpDNAs.

Euglena, Chlorella, Nephroselmis and *Mesostigma* cpDNAs have *rpl23* gene clusters (Fig. 1) similar to those in land plants. These algal *rpl23* gene clusters are distinguished from land plants by the absence of *trnI* (CAU) at the 5′ end of the gene clusters and the presence of *rpl5* and *rps9*. The *Euglena rpl23* operon differs from all other taxa in the complete absence of *E. coli alpha* operon genes, the presence of *roaA* (ribosomal operon-associated gene), and the addition of *trnI-rps14* at the 3′ end of the operon (Christopher and Hallick 1990; Jenkins *et al.* 1995). *Chlorella, Nephroselmis* and *Mesostigma* in the Chlorophyceae (Mattox and Stewart 1983) contain all genes of the angiosperm *rpl23* operon except *rpl22* and *trnI* (CAU). *Chlorella, Nephroselmis* and *Mesostigma* are the only chlorophyll a + b taxa that have *rps9* but it is found in *Synechocystis* and non-green plastids (Fig. 1).

Introns have been documented in several of the genes in Fig. 1 but only in land plants and euglenoids. In most land plants, group II introns with the same insertion sites have been documented in *rpl2* and *rpl16*. The *rpl23* operon genes of *Euglena gracilis* contain numerous introns that are unique for euglenoids (Hallick *et al.* 1993). Except for euglenoids and land plant cpDNAs, introns have not been found in the genes of this gene cluster in any chloroplast or cyanelle genome. The introns in land plants and euglenoids were probably introduced independently and relatively recently.

None of genes in the *rpl23* operon has been investigated in charophyte cpDNAs. Therefore, we have mapped, cloned, and sequenced a region corresponding to the *rpl23* gene cluster in *Spirogyra* cpDNA in order to determine whether it resembles that of land plants or other known algal cpDNAs regarding gene content, structure and intron presence/absence. This gene cluster was characterized in *Spirogyra* to improve our understanding of chloroplast genome evolution and phylogenetic relationships in green plants.

MATERIALS AND METHODS

Genomic DNA and cloned cpDNA fragments of *Spirogyra maxima* (UTEX LB2495) in pBluescript II SK+ (Stratagene), as previously described (Manhart *et al.* 1990) were used for this study. The *rpl23* gene cluster was found in 5.0 kb *Cla*I, 5.3 kb *Cla*I, and 4.8 kb *Pst*I fragments by sequencing both ends of the clones (Fig. 2). The three cpDNA fragments of *Spirogyra maxima* containing the members of *rpl23* gene cluster were cut with the restriction endonucleases *Eco*RI, *Pst*I, *Cla*I, and *Xba*I singly and in combination. The cut DNA fragments were cloned into pBluescript II SK+. Ligations were done using ligase and buffer supplied by Boehringer Mannheim, Germany. The plasmids were transformed into *E. coli* strain DH5α and plasmid DNAs were purified using Quiagen Plasmid Midi Kit (Quiagen, Germany). The cloned DNAs were first manually sequenced using the Sequenase kit (United States Biochemical, Cleveland, Ohio) with 7 pM of T3 and T7 primers, 100 ng of template, and S^{35} labeled dATP at 37°C. For further manual sequencing, octomers of 50% GC were designed for sequencing as described by

Hardin *et al.* (1996). Sequencing reactions of octomers were done with annealing temperature of 18°C, 500 ng of template, 7 pM of octomers, and S^{35} labeled dATP. The fragments were sequenced on the other strand using oligomers (16-18mers) and an ABI 377 automatic sequencer in the Gene Technology Lab at Texas A and M University. Automated sequencing reactions were done using AmpliTaq DNA polymerase, FS (Perkin Elmer, Foster City, CA), 250-500 ng of template and 7 pM of oligomers. Connections between sub-cloned DNA fragments were confirmed by sequencing the mother clones. The connection between the 5.0 kb and 5.3 kb *Cla*I fragments (Fig. 2) was confirmed by sequencing a 1.4 kb PCR product produced by the primers, 5'-GGATTTGAGCATACGAC-3' (jl27.oli) and 5'-GGATTTGAGCATACGAC-3' (jl29.oli). 100 ng of *Spirogyra* total DNA, 130 pM of the primers, and Taq Polymerase kit (Boehringer Mannheim, Germany) were used for PCR with 30 cycles of 94°C for 1 min., 40°C for 1 sec and 72°C for 3 min. The DNA sequences were assembled using Sequencher ver. 3.0 (Gene Code Coporation) on a Macintosh Quadra 660. The genes were identified by sequence comparison using the FASTA in the GCG Package (Genetics Computer Group 1991). Stem-loop structures were found using STEMLOOP program in GCG package and the free energy values were calculated using FOLD in the GCG package. *rpl2* and *rpl16* introns were identified by the comparisons of conserved sequences between exons and introns of *Marchantia polymorpha* cpDNA. The presence/absence of aligned genes and introns in conserved ribosomal genes (Fig. 2) were coded as character states in the phylogenetic analysis as shown in Table 2. PAUP ver. 4.0b (Swofford 2000) was used for the analyses with equal weighting of irreversible characters, and *Synechocystis* was used as outgroup. An exhaustive search was used in producing the shortest trees. Decay indices (Bremer 1988; Mishler *et al.* 1991; Donohue *et al.* 1992) were calculated to determine support for the various clades.

RESULTS AND DISCUSSION

Spirogyra rpl23 gene cluster

The gene cluster of *Spirogyra* is composed of *trnI-rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rpl11-rpoA* (Fig. 2, Table 1). This gene cluster in *Spirogyra* is identical to the angiosperm *rpl23* operon in terms of gene contents, gene orientation, and the presence of introns in *rpl2* and *rpl16*. The long spacer between *rpl14* and *rps8* in *Spirogyra* might be a remnant of *rpl5* that is present in all the taxa in Fig. 1 with the exception of land plants (Table 1).

Most land plants have the same genes, gene order and introns as those found in *Spirogyra*. The losses of genes and gene function have been documented in angiosperm *rpl23* operons. *rpl22* is absent in all members of the legume family (Gantt *et al.* 1991). *Nicotiana* has a pseudogene of *infA* (Wolfe *et al.* 1992) and a few angiosperm taxa contain *rpl23* as a pseudogene (Thomas *et al.* 1988; Wolfe *et al.* 1992). These ribosomal genes were presumably transferred to the nucleus before their losses, as suggested for *rpl22* (Gantt *et al.* 1991) and *tufA* (Baldauf *et al.* 1990). Two grasses (Hiratsuka *et al.* 1989; Maier *et al.* 1995), *Oryza* and *Zea*, have an additional *trnH* gene on the other strand (Table 1), which does not affect gene products of the operon. *rpl22* and *rps3* overlap in some land plants (Table 1) but not in *Spirogyra*. Group II introns are present in *rpl2* and *rpl16* in all completely sequenced cpDNAs of land plants, including *Marchantia, Pinus, Nicotiana, Epifagus, Oryza* and *Zea*, but the *rpl2* intron was lost independently in at least five lineages of dicotyledons (Downie *et al.* 1991). Although there are examples of gene loss or loss of gene function, gene gain, and overlapping genes, it is clear that early land plants had the *rpl23* gene cluster and two introns. The presence of the gene cluster and introns in *Spirogyra* implies that the *rpl23* gene cluster found in embryophytes is not unique for land plants but possibly unique for the streptophyte lineage (charophytes and embryophytes).

Chloroplast genomes have eubacterial promoter regions (Ohyama *et al.* 1988). Eubacterial promoters (Fassler and Gussin 1996) have two consensus regions: the -35 motif (TTGACA), usually 30 to 35 nucleotides preceding the transcription start site, and the -10 motif (TATAAT), usually -12 to -7 bases from the transcription initiation sequence. The optimum spacer size between the two motifs is 17 bases. Several eubacterial promoters contain TG at -15 or -14, which allows stronger binding between RNA polymerase and promoter.

Angiosperms produce *trnI-rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rpl11-rpoA* transcripts (Kanno and Hirai 1993), indicating that they use a promoter upstream of trnI. *Marchantia* also contains a putative promoter region upstream of *trnI* (Ohyama *et al.* 1988). In *Spirogyra*, we could not sequence up to the promoter region upstream of *trnI*, which would be present more than 138 bases upstream of *trnI*. *Spirogyra* has a distinct eubacterial promoter region that is also present in

Table 1. Comparison of the sizes of genes, introns, and intergenic spacers in *rpl23* gene cluster of *Spirogyra* and land plants [*Marchantia* (Ohyama *et al.* 1986), *Pinus* (Wakasugi *et al.* 1994), *Nicotiana* (Shinozaki *et al.* 1986), *Oryza* (Hiratsuka *et al.* 1989), and *Zea* (Maier *et al.* 1995)].

	Spirogyra	*Marchantia*	*Pinus*	*Nicotiana*	*Oryza*	*Zea*
trnI	74	74	74	74	74	74
Spacer	[a]235	159	[b]443	166	175	187
rpl23	267	276	276	282	282	282
Spacer	38	37	21	19	19	19
rpl2	834	834	831	825	822	822
Exon-1	397	397	400	391	391	391
Intron	677	545	672	667	664	664
Exon-2	437	437	431	434	431	431
Spacer	73	37	49	61	[c]262	[c]261
rps19	279	279	279	279	282	282
Spacer	31	18	31	54	69	93
rpl22	327	360	428	468	450	447
Spacer	93	50	[d]–16	[d]–16	56	62
rps3	657	654	654	657	720	675
Spacer	111	58	96	147	142	142
rpl16	432	432	405	405	411	411
Exon-1	9	9	9	9	9	9
Intron	786	536	829	1021	1060	1043
Exon-2	423	423	396	396	402	402
Spacer	171	98	121	125	110	110
rpl14	369	369	369	372	372	372
Spacer	346	82	151	169	140	140
rps8	399	399	399	405	411	411
Spacer	118	87	138	-	137	80
infA	255	237	237	-	324	324
Spacer	184	37	103	[e]438	175	188
rpl36	114	114	114	114	114	114
Spacer	152	51	82	103	175	188
rps11	393	393	393	417	432	432
Spacer	11	33	60	66	64	63
rpoA	1,035	1,023	1,008	1,014	1,014	1,020
Size (bp)	8,450	7,256	8,134	8,319	8,883	8,761

a: Stem-loop structure (22 bases in stem), b: Stem-loop structure (24 bases in stem) and trace of inverted repeats, c: an additional *trn*H gene on the other strand, d: Overlapping genes e: *inf*A pseudogene.

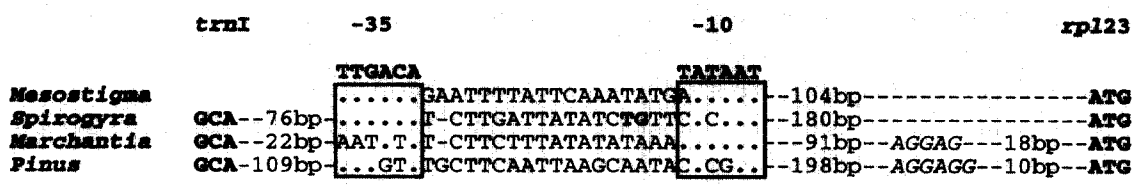

Fig. 3. Putative promoter sequences upstream of *Spirogyra rpl23* and the similar sequences of *Mesostigma*, *Marchantia* and *Pinus*. Boxes indicate eubacterial promoter motifs, -35 and -10 motifs. -15 sequence (TG) is also marked by boldface. Putative ribosome binding sequences (AGGAGG) are italicized.

Mesostigma (Fig. 3). The region in *Spirogyra* contains the -35 motif (TTGACA) and the -10 motif (CACAAT) upstream of *rpl23*. The spacer between the two motifs is 17 bases, the optimum size for the eubacterial promoter. In addition, this promoter region has TG at -15, which is a determinant of promoter strength (Fassler and Gussin

1996). This indicates that the *rpl23* gene cluster of *Spirogyra* contains a strong promoter upstream of *rpl23*.

The *Spirogyra rpl23* promoter-like sequence is also present in *Marchantia* and *Pinus* (Fig. 3), but is absent in angiosperms. The sequences of *Marchantia* and *Pinus* show low conservation in one of two motifs (Fig. 3). The promoter regions upstream of *rpl23* in *Marchantia* and *Pinus* might have weak promoter functions. In contrast, strong stem-loop structures, which could halt transcription, are present upstream of the *rpl23* promoter region in *Spirogyra* (ΔG = -23.0, 22 bases in stem) and Pinus (ΔG = -35.6, 24 bases in stem). The *rpl23* promoter region of *Pinus* partially overlaps with its stem-loop structure. In contrast, the large stem-loop structure is absent in *Marchantia*.

Comparison of the ribosomal protein gene clusters among prokaryotes and plastids

The gene cluster corresponding to three *E. coli* operons (S10, *spc* and *alpha*) is conserved throughout eubacteria, although transcriptional units vary (Ohkubo *et al.* 1987; Boylan *et al.* 1989; Lindahl *et al.* 1990; Jahn *et al.* 1991; Sanangelantoni and Tiboni 1993; Sanangelantoni *et al.* 1994; Pfeiffer *et al.* 1995; Suh *et al.* 1996). There are four main differences between the gene clusters of *E. coli* and *Synechocystis*. In *Synechocystis* and other cyanobacteria, *rps10*, the first gene of the *E. coli* S10 operon, is a member of the cyanobacterial *str* operon (*rps12-rps7-fusA-tufA-rps10*). The second is the absence of three genes (*rps14, rpl30*, and *rpl15*) in the *E. coli spc* operon and the presence of two additional genes (*adk* and *infA*) in this region. The third is the absence of *rps4* in the *E. coli alpha* operon. The last is that four additional genes (*truA, rpl13, rps9*, and *rpl31*) occur at the end of the gene cluster in *Synechocystis*.

Gene clusters similar to that of *Synechocystis* are present in the cyanelle *Cyanophora paradoxa* and cpDNAs of the rhodophyte *Porphyra purpurea* and the chromophyte *Odontella sinensis* (Fig. 1). Among these three, *Odontella* and *Porphyra* have most of the cyanobacterial genes. However, they are distinguished from *Synechocystis* by the absence of *adk* and *infA* genes of the *E. coli spc* operon, *rpl17* of the *E. coli* α operon, and *truA* at the 3′ end of the gene cluster. The cyanelle of *Cyanophora* is missing three genes of the *E. coli* S10 operon (*rpl4, rpl23*, and *rpl29*), *rpl24* of the *E. coli spc* operon, and two genes at the 3′ end of the gene cluster (*rpl13* and *rpl31*). In addition, the cyanelle of *Cyanophora paradoxa* contains the ribosomal gene cluster in two separate gene clusters (Fig. 1). A unique character of *Porphyra* and *Odontella* plastids is the presence of cyanobacterial str operon genes (*rps12-rps7-tufA-rps10*) at the end of the gene cluster. A similar organization is also present in a vestigial plastid of apicomplex parasites (McFadden *et al.* 1997).

Chlorophyll a and b eukaryotes (euglenoids, chlorophytes, and charophytes + land plants) show significant similarities with the absence of fifteen genes: *rpl3, rpl4, rpl29, rps17, rpl24, rpl6, rpl18, rps5, secY, adk, rps13, rpl17, truA, rpl13*, and *rpl31* (Fig. 1). Among chlorophyll a + b organisms, *Euglena, Chlorella, Nephroselmis* and *Mesostigma* contain *rpl5*, absent in *Spirogyra* and land plants. *Euglena* cpDNA does not contain any *E. coli* α operon genes but the *Euglena rpl23* gene cluster contains a unique gene, *roaA*. In contrast, *Chlorella, Nephroselmis* and *Mesostigma* and *Spirogyra* + land plants contain most genes of the *E. coli* α operon found in non-green algal cpDNAs with the exception of *rps13*. Unlike *Spirogyra* + land plants, *Chlorella, Nephroselmis* and *Mesostigma* contain *rps9* at the 3′ end of the gene cluster. In addition, *Chlorella* and *Nephroselmis* have additional *rpl12* unlike other green plants. The green plants also contain the *infA* gene, which is not present in non-green algal plastids, but present in cyanobacteria. The inclusion of trnI at the 5′ end of the cluster is unique to land plants + *Spirogyra*.

Phylogenetic analyses

Presence/absence of the genes and introns (Table 2) were used as characters in phylogenetic analyses (Fig. 4). The characters were equally weighted and the cyanobacterium *Synechocystis* was the designated outgroup. A single shortest tree was produced. The tree places the plastids of chlorophyll a organisms at the base of the chloroplast + cyanelle clade, and *Cyanophora* branches off next. The chloroplast + cyanelle clade differs from *Synechocystis* by the loss of three genes (*adk, rpl17*, and *truA*). *Cyanophora* shares the loss of 3 genes (*rpl4, rpl29*, and *rpl24*) with chlorophyll a + b eukaryotes (green plants + *Euglena*).

In addition, the clade containing green plants + *Euglena* is distinguished by 10 gene losses (*rpl3, rps17, rps8, rpl6, rpl18, rps5, secY, rps13, rpl13* and *rpl31*), which is the best-supported clade on the tree, though the position of *Chlorella* and *Euglena* varies. In this clade, the clade containing *Spirogyra* and land plants is separated from other chlorophyll a + b containing organisms by *rpl15* loss, *trnI* gain, and the gain of *rpl2* and *rpl16* introns. The grasses are further derived by the gain of the *trnH* gene on the other strand, and *Nicotiana* is fur-

Table 2. Characters, character states, and data matrix used for cladistic analysis (Figure 4).

Characters

1: *trn*I	2: *rpl*3	3: *rpl*4	4: *rpl*23	5: *rpl*2	6: *rps*19
7: *rpl*22	8: *rps*3	9: *roa*A	10: *rpl*16	11: *rpl*29	12: *rps*17
13: *rpl*14	14: *rpl*24	15: *rpl*5	16: *rps*8	17: *rpl*6	18: *rpl*18
19: *rps*5	20: *sec*Y	21: *adk*	22: *inf*A	23: *rpl*36	24: *rps*13
25: *rps*11	26: *rpo*A	27: *rpl*17	28: *tru*A	29: *rpl*13	30: *rps*9
31: *rpl*31	32: *rpl*12	33: *trn*H*			
34: *rpl*2 land plant intron		35: *rpl*16 land plant intron			
36: *Euglena* interons		37: Disjunction in front of *rpl*36			

Character state

-: absence +: presence

			10		20		30		37
Synechocystis	−++++	+++−+	+++++	+++++	+++++	+++++	+−−−−	−−
Cyanophora cyanelle	−+−−+	+++−+	−++−+	+++++	−−+++	+−−++	+−−−−	−+
Odontella/Porphyra/Cyanidium chloroplast	−++++	+++−+	+++++	+++++	−−+++	+−−++	+−−−−	−−
Guillardia chloroplast	−++++	+++−+	+++++	+++++	−−+++	−−−++	+−−−−	−−
Euglena chloroplast	−−−++	+++++	−−+−+	−−−−−	−−+−−	−−−−−	−−−−−	−+
Chlorella/Nephroselmis chloroplast	−−−++	+−+−+	−−+−+	−−−−−	−++−+	+−−−+	−+−−−	−−
Mesostigma chloroplast	−−−++	+++−+	−−+−+	−−−−−	−++−+	+−−−+	−−−−−	−−
Spirogyra & land plant chloroplast	+−−++	+++−+	−−+−−	−−−−−	−++−+	+−−−−	−−−++	−−
Nicotiana chloroplast	+−−++	+++−+	−−+−−	−−−−−	−−+−+	+−−−−	−−−++	−−
Grass chloroplasts	+−−++	+++−+	−−+−−	−−−−−	−++−+	+−−−−	−−+++	−−

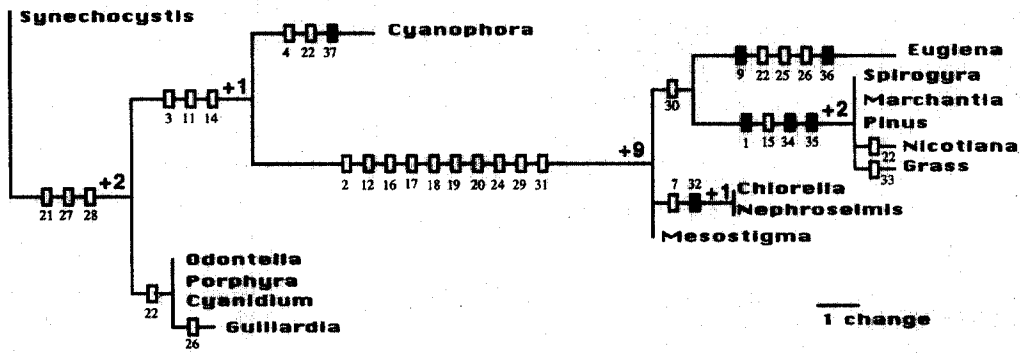

Fig. 4. The shortest tree with equal weighting of irreversible characters and *Synechocystis* the designated outgroup. Open boxes indicate losses and closed boxes mark gains. Decay indices are preceded by + and indicate branches found in all trees with index number +35 steps [+1 found 9 trees; +2, 24; +3, 57; +4, 150 and +5, 391].

ther derived by the loss of *infA*.

The phylogenetic position of the cyanelle of *Cyanophora* is uncertain (Bhattacharya and Medlin 1995; Bhattacharya *et al.* 1995; Helmchen *et al.* 1995) and it was removed from a subsequent analysis to determine if that had any effect on the topology and character changes. The removal of Cyanophra did not affect the topology and character changes. *infA* has apparently been lost in *Cyanophora, Odontella, Porphyra* and *Euglena*, and retained in the green plant lineage, a reverse of the general trend where the former taxa have retained many more genes than in green plants, assuming a common ancestry for all plastids and cyanelles.

Evolution of the *rpl23* gene cluster and their phylogenetic implications

With the exception of roaA in *Euglena* chloroplast DNA, the phylogenetic analysis indicates that these cyanelle and chloroplast gene clusters evolved mostly by gene loss. The cyanelle of *Cyanophora* and chloroplasts of *Porphyra* and *Odontella* have retained more of the genes than chlorophyll a and b organisms. Among these plas-

tids, *Porphyra* and *Odontella* are the least derived taxa relative to cyanobacteria. The loss of fewer genes from this gene cluster is expected in *Porphyra* because it has the largest chloroplast genome that has been completely sequenced. *Cyanophora* is more derived than *Porphyra* and *Odontella* by more gene losses and split gene clusters. *Cyanophora* is intermediate between *Odontella* + *Porphyra* and chlorophyll a + b organisms but closer to *Odontella* + *Porphyra*.

The *Euglena* + green plant clade is strongly supported, indicating that the *Euglena* chloroplast was derived from a green alga as suggested by ultracellular evidence (Gibbs 1978). Lastly, *Spirogyra* and land plants are in a clade separated from other organisms containing chlorophyll a+b and share the unique gene cluster, *trnI-rpl23-rpl2-rps19-rpl22-rps3-rpl16-rpl14-rps8-infA-rpl36-rpl11-rpoA*, in addition to *rpl2* and *rpl16* group II introns. This indicates that this gene cluster probably evolved early in the evolution of the charophyte-land plant lineage but that can only be verified by sampling other major lineages of green algae.

The distribution of promoter sequences upstream of *rpl23* indicates that the promoter was present in the common ancestor of *Spirogyra* and land plants but that it has been gradually and completely replaced by the *trnI* promoter in angiosperms. The determination of the presence/absence of *trnI*, *rpl5*, and promoter-like sequence upstream of *rpl23* in other major green algal lineages might provide a better understanding of phylogenetic relationships of land plants, charophytes, and non-charophyte green algae.

ACKNOWLEDGEMENTS

This project was supported in part by NSF Grant 9306462 to JRM and Texas A&M graduate-grant to JL (1996).

REFERENCES

Baldauf S.L., Manhart J.R., and Palmer J.D. 1990. Differential fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *Proc. Natl. Acad. Sci. U.S.A.* **87:** 5317-5321.

Bhattacharya D., Helmchen T., Bibeau C. and Melkonian M. 1995. Comparisons of nuclear-encoded small-subunit ribosomal RNAs reveal the evolutionary position of the Glaucocystophyta. *Mol. Biol. Evol.* **12:** 415-420.

Bhattacharya D. and Medlin L. 1995. The phylogeny of plastids: A review based on comparisons of small-subunit ribosomal RNA coding regions. *J. Phycol.* **31:** 489-498.

Blattner F.R., Plunkett III G., Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vodes J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B. and Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453-1474.

Boylan S.A., Suh J.-W., Thomas S.M. and Price C.W. 1989. Gene encoding the alpha core subunit of Bacillus subtilis RNA polymerase is cotranscribed with the genes for innitiation factor 1 and ribosomal proteins B, S13, S11, and S17. *J. Bact.* **171:** 2553-2562.

Bremer K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42:** 795-803.

Christopher D.A. and Hallick R.B. 1990. Complex RNA mutation pathway for a chloroplast ribosomal protein operon with an internal tRNA cistron. *The Plant Cell* **2:** 659-671.

Donoghue M.J., Olmstead R.G., Smith J.F. and Palmer J.D. 1992. Phylogenetic relationships of Dipsacales based on *rbcL* sequences. *Ann. Missouri Bot. Gard.* **79:** 333-345.

Douglas S.E. and Penny S.L. 1999. The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae. *J. Mol. Evol.* **48:** 236-244.

Downie S.R., Olmstead R.G., Zurawski G., Soltis D.E., Soltis P.S., Watson J.C. and Palmer J.D. 1991. Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: Molecular and phylogenetic implications. *Evolution* **45:** 1245-1259.

Fassler J.S. and Gussin J.S. 1996. Promoters elements and RNA polymerase components. *Method in Enz.* **273:** 3-42.

Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. and Venter J.C. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270:** 397-403.

Gantt J.S., Baldauf S.L., Calie P.J., Weeden N.F. and Palmer J.D. 1991. Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *EMBO. J.* **10:** 3073-3078.

Genetic Computer Group 1991. *Program Manual for the GCG Package*, Version 7. Madison, Wisconsin.

Gibbs S.P. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.* **56:** 2883-2889.

Glockner G., Rosenthal A. and Valentin K. 2000. The structure and gene repertoire of an ancient red algal plastid genome. *J. Mol. Evol.* **51:** 382-390.

Graham L.E. 1993. *Origin of land plants*. John Wiley and Sons, Madison pp. 1-14.

Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A., Orsat B., Spielmann A. and Stutz E. 1993. Complete

sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* **21:** 3537-3544.

Hardin S.H., Jones L.B., Homayouni R. and McCollum J.C. 1996. Octomer-primed cycle sequencing: Design of an optimized primer library. *Genome Research* **6:** 545-550.

Helmchen T.A., Bhattacharya D. and Melkonian M. 1995. Analyses ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J. Mol. Evol.* **45:** 203-210.

Hiratsuka J., Shimada H., Whittier R., Ishibashi T., Sakamoto M., Mori M., Kondo C., Honji Y., Sun C.-R., Meng B.-Y., Li Y.-Q., Kanno A., Nishizawa Y., Hirai A., Shinozaki K. and Sugiura M. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recommbination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217:** 185-194.

Jahn D., Hartmann R.K. and Erdmann V.A. 1991. Analysis of the spc ribosomal protein operon of *Thermus aquaticus*. *Eur. J. Biochem.* **197:** 733-740.

Jenkins K.P., Hong L. and Hallick R.B. 1995. Alternative splicing of the *Euglena gracilis* chloroplast *roaA* transcript. *RNA* **1:** 624-633.

Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y., Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T., Hosouchi T., Matsuno A., Muraki A., nakazaki N., Naruo K., Okunura S., Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M. and Tabata S. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research* **3:** 109-136.

Kanno A. and Hirai A. 1993. A transcription map of the chloroplast genome from rice (*Oryza sativa*). *Curr. Genet.* **23:** 166-174.

Kenrick P. and Crane P.R. 1997. The origin and early evolution of plants on land. *Science* **389:** 33-39.

Kowallik K.V., Stoebe B., Schaffran I., Kroth-Panic P. and Frieier U. 1995. The chloroplast genome of a chlorophyll a+c-containing alga, *Odontella sinensis*. *Plant Mol. Biol. Rep.* **13:** 336-342.

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.

Lew K.A. and Manhart J.R. 1993. The *rps12* gene in *Spirogyra maxima* (Chlorophyta) and its evolutionary significance. *J. Phycol.* **29:** 500-505.

Lindahl L., Sor F., Archer R.H. and Nomura M. 1990. Transcriptional organization of the S10, *spc* and *alpha* operons of *Escherichia coli*. *Bioch. Biophys. Acta.* **1050:** 337-342.

Maier R.M., Neckermann K., Igloi G.L. and Kossel H. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* **251:** 614-628.

Manhart J.R. and Palmer J.D. 1990. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. *Nature* **345:** 268-270.

Manhart J.R., Hoshaw R.W. and Palmer J.D. 1990. Unique chloroplast genome in *Spirogyra maxima* (Chlorophyta) revealed by physical and gene mapping. *J. Phycol.* **26:** 490-494.

Mattox K.R. and Stewart K.D. 1983. Classification of the green algae: A concept based on comparative cytology. In Irvine D.E.G. and John D.M. (eds) *Systematics of green algae.* Academic press, London, pp. 29-72

McFadden G.I., Waller R.F., Reith M.E. and Lang-Unnasch N. 1997. Platids in apicomplexan parasites. *Pl. Syst. Evol. (Suppl.)* **11:** 261-287.

Mishler B.D., Donoghue M.J. and Albert V.A. 1991. *The decay index as a measure of relative robustness within a cladogram.* Willi Hennig Society Meeting. Toronto, Ontario.

Mishler B.D., Lewis L.A., Buchheim M.A., Renzagolia K.S., Garbary D.J., Delwiche C.F., Zechman F.W., Kantz T.S. and Chapman R.L. 1994. Phylogenetic relationships of the Green algae and Bryophytes. *Ann. Missouri Bot. Gard.* **81:** 451-483.

Ohkubo S., Muto A., Kawauchi Y., Yamao F. and Osawa S. 1987. The ribosomal protein gene cluster of *Mycoplasma capricolm. Mol. Gen. Genet.* **210:** 314-322.

Ohyama K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S., Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S.-I., Inokuchi H. and Ozeki H. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* **322:** 572-574.

Ohyama K., Fukuzawa H., Kohchi T., Sano T., Sano S. and Shrai H. (1988) Structure and organization of *Marchantia polymorpha* chloroplast genome: I. Cloning and gene identification. *J. Mol. Biol.* **203:** 281-298.

Palmer JD 1991. Plastid chromosomes: Structure and evolution. In Bogorad L. and Vasil I.K. (eds) *The molecular biology of plastids.* Academic Press, Boston pp. 5-53.

Pfeiffer T., Jorcke D., Feltens R., and Hartmann R.K. 1995. Direct linkage *str-*, S10- and *spc-* related gene cluster in *Thermus thermophilus* HB8, and sequences of ribosomal proteins L4 and S10. *Gene* **167:** 141-145.

Reith M. and Munholland J. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Rep.* **13:** 333-335.

Sanangelantoni A.M. and Tiboni O. 1993. The chromosomal location of genes for elongation factor Tu and ribosomal protein S10 in cyanobacterium *Spirulina platensis* provides clues to the ancestral organization of str and S10 in prokaryotes. *J. Gen. Micro.* **139:** 2579-2584.

Sanangelantoni A.M., Bocchetta M., Cammarano P., and Tiboni O. 1994. Phylogenetic depth of *S10* and *spc* operons: Cloning and sequencing of a ribosomal protein gene cluster from the extremely thermophilic bacterium *Thermotoga maritima. J. Bact.* **176:** 7703-7710.

Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayasida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,

Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H. and Sugiura M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* **5:** 2043-2049.

Stirewalt V.L., Michalowski C.B., Loffelhardt W., Bohnert H.J. and Bryant D.A. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol. Biol. Rep.* **13:** 327-332.

Sugiura M. 1992. The chloroplast genome. *Plant Mol. Biol.* **19:** 149-168.

Suh J.-W., Boylan S.A., Oh S.-H., and Price C.W. 1996. Genetic and transcriptional organization of the *Bacililus subtilis spc-alpha* region. *Gene* **169:** 17-23.

Swofford D.L. 2000. *Phylogenetic analysis using parsimony, version 4.0 beta.: User's manual*. Illinois Natural History Survey, Champaign, IL

Tanaka M., Wakasugi T., Sugita M. and Shinozaki K. 1986. Genes for the eight ribosomal proteins are clustered on the chloroplast genome of tobacco (*Nicotiana tabacum*): Similarity to the *S10* and *spc* operons of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA.* **83:** 6030-6034.

Thomas F., Massenet O., Dorne A.M., Briat J.F. and Mache R. 1988. Expression of the *rpl23*, *rpl2*, and *rps19* genes in spinach chloroplasts. *Nucleic Acids Res.* **16:** 2461-2472.

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. U.S.A.* **96:** 10248-10253.

Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T. and Sugiura M. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 9794-9798.

Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S., Tsudzki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T., Inamura A., Yoshinaga K. and Sugiura M. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of the genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 5967-5972.

Wolfe K.H., Morden C.W., Ems S.C. and Palmer J.D. 1992. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: Loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *J. Mol. Evol.* **35:** 304-317.

Accepted 19 January 2002