Curr Genet (1989) 15:277–282

Current Genetics

An optional group I intron between the chloroplast small subunit rRNA genes of *Chlamydomonas moewusii* and *C. eugametos*

Vincent Durocher, Antonin Gauthier, Guy Bellemare, and Claude Lemieux
Département de Biochimie, Faculté des Sciences et de Génie, Université Laval, Québec, Québec G1K 7P4 Canada

Summary. We report the presence of a 402 bp group I intron in the chloroplast small subunit (SSU) rRNA gene of *Chlamydomonas moewusii.* The intron is inserted within the highly conserved '530 loop', at a site corresponding to positions 531–532 of the *E. coli* 16rRNA. Residues surrounding the insertion site almost certainly play an important role in ribosomal proofreading function as they proved to be protected by tRNAs in *E. coli* 16S rRNA (Moazed and Noller 1986; Stern et al. 1986). The *C. moewusii* intron revealed a secondary structure model which differs substantially from those of the typical subgroup IA and IB introns. This model, however, shows striking similarities with the structures of the *C. reinhardtii* chloroplast 23S rRNA gene intron (Rochaix et al. 1985), the *S. cerevisiae* mitochondrial COB3 intron (Holl et al. 1985) and the three introns of phage T4 in the *nrdB, td* and *sunY* genes (Shub et al. 1988). The SSU rRNA gene intron is absent from *C. eugametos,* an alga that is interfertile with *C. moewusii.* The presence/absence of the intron account for a 390 bp restriction fragment length polymorphism between the two algal SSU rRNA genes, a polymorphic locus that is strictly co-inherited with a tightly linked streptomycin resistance mutation (*sr-2*) in interspecific hybrids between the two algae.

Key words: Secondary structure model – '530 loop' – Streptomycin resistance mutation – *Chlamydomonas* evolution

Offprint requests to: C. Lemieux

Introduction

Optional introns in the mitochondrial genomes of closely related yeast and fungal species represent a major source of restriction fragment length polymorphisms (RFLPs) among these genomes (Dujon 1981). For example, the gene coding for cytochrome oxidase subunit 1 in various strains of *Neurospora crassa* contains 0–4 introns (Burger et al. 1982; Collins and Lambowitz 1983), whereas the same gene in *Saccharomyces cerevisiae* has 4–9 introns (Hensgens et al. 1983). Mitochondrial introns have been classified into two families, called group I and group II, on the basis of distinctive primary sequence stretches and a number of potential RNA secondary structures (Michel et al. 1982). This classification also applies to chloroplast introns and nuclear introns that do not conform to the GU..AG rule characteristic of eucaryotic pre-mRNAs (Michel and Dujon 1983).

A large number of RFLPs were also detected in the chloroplast genomes of the two pairs of interfertile algae *Chlamydomonas eugametos/C. moewusii* (Turmel et al. 1987) and *Chlamydomonas reinhardtii/C. smithii* (Palmer et al. 1985). Although little is known about their nature, one has been shown to result from the presence of an optional intron. The third of the four group I introns in the *C. reinhardtii* gene coding for the D1 polypeptide of photosystem II (*psbA*) is absent from the interfertile alga *C. smithii* (Erickson et al. 1984; Palmer et al. 1985).

We report here the presence of an optional group I intron between the chloroplast small subunit (SSU) rRNA genes of *C. moewusii* and *C. eugametos.* To our knowledge, this is the first intron that has been detected in a SSU rRNA gene. In *C. moewusii* it is inserted at a position corresponding to a highly conserved region of primary sequence and secondary structure among SSU rRNA genes. Its secondary structure model is very similar to that proposed for the unique intron in the

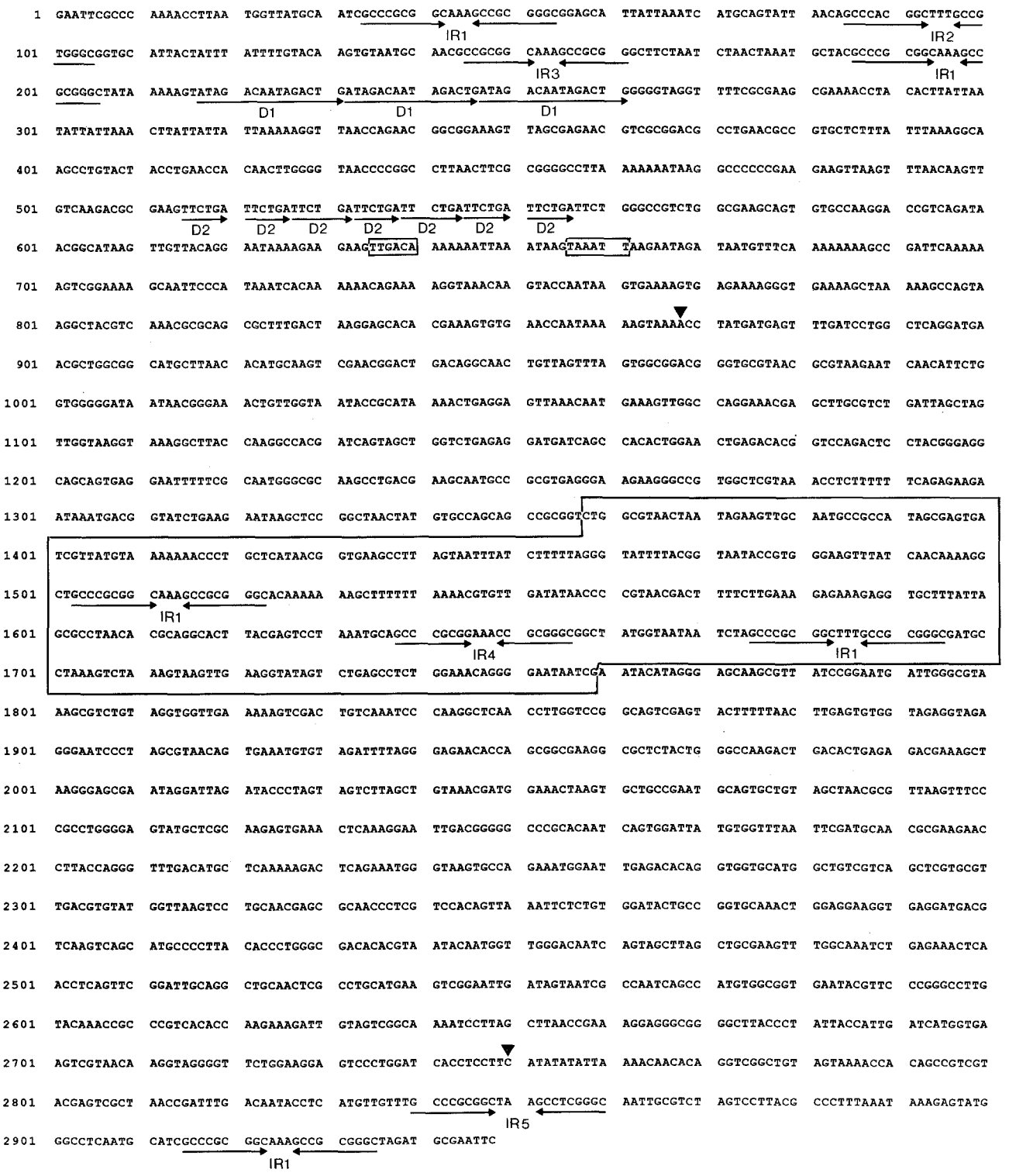
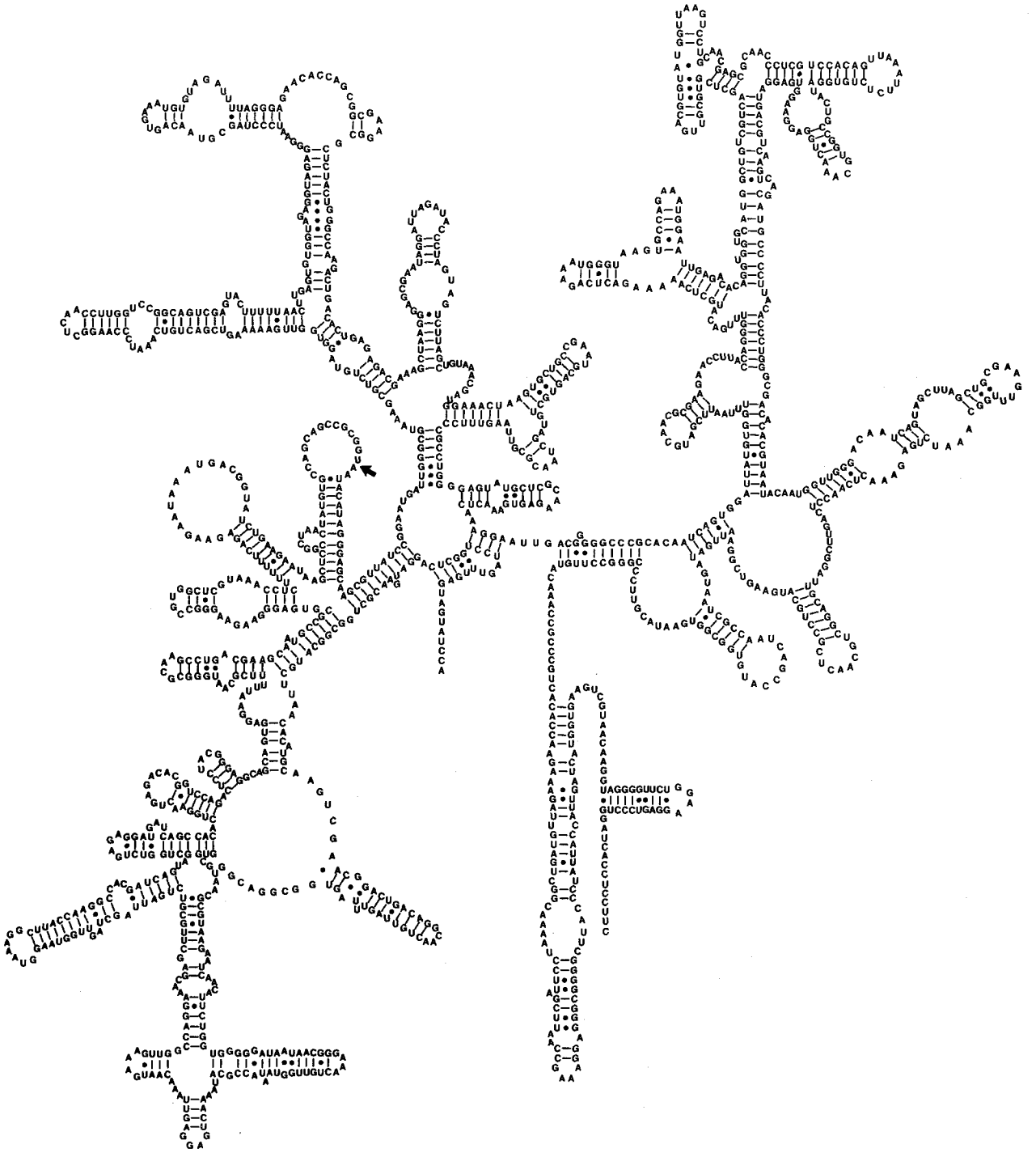


Fig. 1. Nucleotide sequence of the *C. moewusii* chloroplast SSU rRNA gene and its flanking regions. The sequence of the 402 bp group I intron is *framed; dark triangles* indicate the most probable positions of the 5′ and 3′ termini of the SSU rRNA. Regions homologous to the −10 and −35 elements of the putative rDNA promoter sequence are *boxed.* Direct (*D*) and inverted (*IR*) repeats are marked with *arrows*

Fig. 2. Secondary structure model of the *C. moewusii* chloroplast SSU rRNA. It is modelled as a strict replica of the secondary structure proposed for *E. coli* 16S rRNA (Noller 1984). The *heavy arrow* indicates the position of the intron within the '530' loop



23S rRNA gene of the *C. reinhardtii* chloroplast (Rochaix et al. 1985).

Materials and methods

The plasmids containing the fragments *Eco*RI 21 of *C. moewusii* and *Eco*RI 19 of *C. eugametos* have been described previously (C. Lemieux et al. 1985). Subfragments of these two plasmids were cloned into M13mp18 and M13mp19 (Norrander et al. 1983). Deletions of the *Eco*RI fragments cloned into M13mp19 were generated in order to produce a series of overlapping sequences (Dale et al. 1985). DNA sequences were determined by the Sanger dideoxynucleotide procedure (Sanger et al. 1977), using α-^{35}SdATP (Biggin et al. 1983). In order to eliminate ambiguities in the sequences, we often employed 7-deaza-dGTP (Boehringer Mannheim) instead of dGTP in the sequencing reactions. A few sequencing reactions were initiated with the two following synthetic oligonucleo-

tide primers: primer 1,5′-TACAGACGCTTTACGCCCAAT-3′ (positions 1811-1791 in Fig. 1); and primer 2,5′-CGCGTTGCATCGAATTAAACC-3′ (positions 2194-2174). They were purified by preparative electrophoresis on denaturing 18% polyacrylamide gels. RNA sequencing across the putative exon 1/exon 2 junction was performed with primer 1 according to Lane et al. (1985). DNA sequence analyses and searches were carried out with the aid of the software package of the University of Wisconsin Genetics Computer Group (Devereux et al. 1984).

Results and discussion

Identification of a group I intron in the chloroplast SSU rRNA gene of C. moewusii

During the construction of the physical map of the *C. moewusii* chloroplast genome, C. Lemieux et al. (1985) localized the gene coding for the chloroplast SSU (16S) rRNA on the *Eco*RI fragment 21. We report here that the nucleotide sequence of this 2,946 bp fragment (Fig. 1) includes the entire 16S rRNA gene and about 850 bp of 5′ flanking sequences.

When compared with eubacterial and other chloroplast SSU rRNA genes, the *C. moewusii* sequence revealed all the domains of a typical 16S rRNA molecule and a colinear order of these sequences. Surprisingly, however, an extra segment of 402 bp was observed between two stretches of structural sequences; one stretch corresponds to positions 1–531 of the eubacterial 16S rRNA and the other to positions 532–1542. This 402 bp extra sequence is inserted within a universally conserved region of primary sequence and potential secondary structure, designated universal region 3 (U3) by Gray et al. (1984). A secondary structure model of the *C. moewusii* SSU rRNA is presented in Fig. 2. This structure is almost an exact replica of the eubacterial 16S rRNA model, with minor variations in length and helical structure being localized to known variable domains of chloroplast SSU rRNA (Gutell et al. 1985).

The additional 402 bp sequence proved to be an intron of the *C. moewusii* SSU rRNA gene. Two pieces of evidence rule out the possibilities that the *C. moewusii* SSU rRNA is 400 bp larger than usual chloroplast 16S rRNAs or that it consists of two discontinuous RNA species whose coding sequences are separated by an internal transcribed spacer excised during the processing of the primary rDNA transcript, as is the case for the SSU rRNAs of *Trypanosoma brucei* (Campbell et al. 1987) and *Crithidia fasciculata* (Spencer et al. 1987). First, hybridization probing of the *Eco*RI fragment 21 to blots of *C. moewusii* total cellular RNA revealed only a single RNA species with a size of 1,500 bases; i.e., the size of typical eubacterial and chloroplast 16S rRNA molecules (Turmel et al. 1988). Second, RNA sequencing across the putative exon 1/exon 2 junction suggested that the additional sequence of 402 bp is absent from the *C. moewusii* SSU rRNA (data not shown). By using total cellular RNA as a template and a synthetic DNA oligonucleotide complementary to the coding region located 32 bp downstream of the junction as a primer, the RNA sequences corresponding to the two putative exons can be shown to be contiguous.

Figure 3 shows that the secondary structure model of the *C. moewusii* intron closely resembles those of mitochondrial group I introns with their characteristic loops and helices (Michel and Dujon 1983; Burke et al. 1987). Like other members of the group I family, the intron contains the usual P, Q, R and S elements (Burke et al. 1987) and starts immediately after a U residue and ends with a G residue. Although the *C. moewusii* intron shows many of the typical features of subclass IA within the group I family (Michel et al. 1982), it cannot be unambiguously classified as a member of this subgroup. The SSU rRNA gene intron contains two helices (P7.1 and P7.2) between P3 and P7 instead of the single one usually found in subgroup IA introns, and the P7 pairing includes only four consecutive base pairs instead of five. As proposed by Michel and Cummings (1985), there is an extension of complementarity in P7 between the U residue at position 1732 (see Fig. 1) and the A residue at position 1565 and, like most subgroup IA introns, the bulged nucleotide is a C (Cech 1988). Few examples of introns with two stem loops between P3 and P7 have been reported. These include the *C. reinhardtii* chloroplast 23S rRNA gene intron (Rochaix et al. 1985), the *S. cerevisiae* mitochondrial COB3 intron (Holl et al. 1985) and the three introns of phage T4 in the *nrdB, td* and *sunY* genes (Shub et al. 1988). In contrast to these, the *C. moewusii* intron sequence lacks a long internal open reading frame. Finally, the optional pairing of P10 (Burke et al. 1987) is not present in the *C. moewusii* intron. This pairing allows a precise alignment of the splice junctions, and it links the nucleotides in the first helix of the intron (P1) to the nucleotides in the region downstream of the opposite intron-exon junction (Davies et al. 1982).

The chloroplast SSU rRNA gene of the interfertile species C. eugametos contains no intron

We also sequenced the chloroplast SSU rRNA gene of *C. eugametos,* an alga that is interfertile with *C. moewusii.* No intronic sequence was found in this gene, indicating that the *C. moewusii* intron is optional between the two species. Apart from this intron, *C. eugametos* and *C. moewusii* have an almost identical

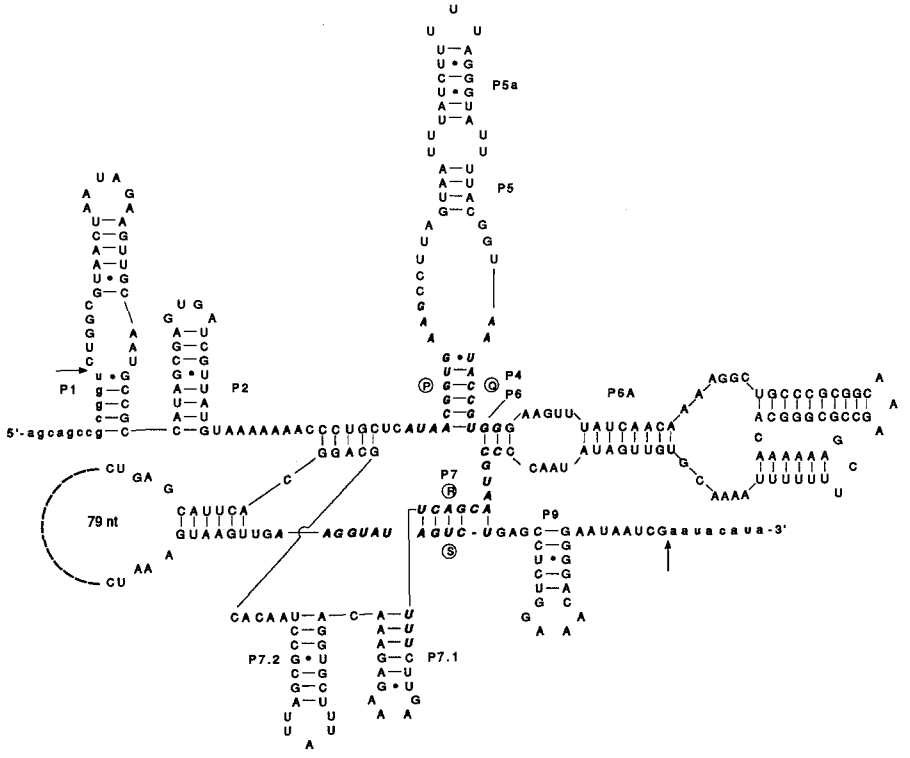


Fig. 3. Secondary structure model of the group I intron in the *C. moewusii* chloroplast SSU rRNA gene. The *arrows* point to the 5' and 3' splice junctions; exon sequences are represented by *lowercase characters* and intron sequences by *uppercase characters*. Helices (*P1* to *P9*) are designated as in Burke et al. (1987), and the conserved sequence elements P, Q, R and S are represented in *italicized characters*

SSU rRNA gene sequence, the only difference being a 'G' in *C. eugametos* instead of an 'A' at position 1767 of *C. moewusii* (position 539 in *E. coli*).

The optional intron thus accounts for the RFLP of 390 bp between the *C. moewusii* and *C. eugametos* chloroplast genomes in the region of the SSU rRNA gene (C. Lemieux et al. 1985; Turmel et al. 1987). This polymorphic region, also known as locus 'A' (Lemieux and Lee 1987), is strictly co-inherited with the locus of the streptomycin resistance mutation *sr-2* in interspecific hybrids between the two algae (Lemieux et al. 1984). The *sr-2* mutation has been recently mapped in the *C. eugametos* SSU rRNA gene at a site corresponding to position 523 in the *E. coli* SSU rRNA (Gauthier et al. 1988). As the mutation is located only eight residues from the insertion site of the SSU rRNA gene intron, it is not surprising that recombination seldom occurs between these genetic and physical markers. This explains why they are strictly co-inherited in interspecific hybrids. Schneider et al. (1985) have suggested that recombinational processes in the *C. reinhardtii* chloroplast could involve numerous short direct and inverted repeats scattered throughout the rDNA region. It would be interesting to see if the repeat elements in the flanking regions of the *C. moewusii* SSU rRNA gene (see Fig. 1) play a role in recombination events.

High level of sequence divergence among the chloroplast SSU rRNAs of Chlamydomonas

The chloroplasts SSU rRNAs of *C. eugametos* and *C. moewusii* display only 86% sequence identity with that of *C. reinhardtii* (Dron et al. 1982). This value is much lower than the 95-99% identity observed between corresponding land plant sequences. One would expect a higher degree of sequence identity between organisms belonging to the same genus. However, since *C. reinhardtii* and the two interfertile algae show important differences in their biological properties, as well as major differences in the order of their chloroplast genes (Lemieux and Lemieux 1985; B. Lemieux et al. 1985), we believe that the algal lineages diverged well before the emergence of land plants. A similar conclusion was recently reached by Jupe et al. (1988) on the basis of their comparative sequence analyses of *Chlamydomonas* nuclear SSU rRNA genes.

Conclusion

The *C. moewusii* intron described in this study is the first group I intron reported in an SSU rRNA gene. As the intron is absent from the interfertile species *C. eugametos,* it is tempting to speculate that *Chlamydomonas* chloroplast genes resemble the mitochondrial genes of yeast and fungi in displaying major variations

in intron composition. Like the introns in the *nrdB, td* and *sunY* genes of phage T4 (Shub et al. 1988), the *C. reinhardtii* chloroplast 23S rRNA gene (Rochaix et al. 1985) and the *S. cerevisiae* mitochondrial COB3 gene (Holl et al. 1985), the *C. moewusii* intron cannot be unambiguously classified in one of the two recognized subclasses of group I introns (i.e., IA or IB, see Michel et al. 1982). These introns may well form part of a new subclass, which remains to be defined.

Clarke et al. (1984) observed that all of the introns that have been reported so far in large subunit (LSU) rRNA genes are inserted within, or very close to, highly conserved regions of LSU rRNA which are implicated in tRNA interactions. The same conclusion can be drawn for the *C. moewusii* SSU rRNA gene intron, which is inserted within the '530' loop. This is a region of the SSU rRNA that is thought to play an important role in the proofreading function, because the bases surrounding the insertion site of the intron have been found to interact with tRNAs in *E. coli* (Moazed and Noller 1986; Stern et al. 1986). It will be important to determine why rDNA introns tend to be located in such regions and whether there are any selective advantages for their presence.

Acknowledgements. We thank Drs. R. Cedergreen, M. W. Gray and M. Turmel for helpful comments, and Dr. B. B. Sears for critical reading of this manuscript. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada. C. L. is a Scholar of the Canadian Institute for Advanced Research.

References

Biggin MD, Gibson TJ, Heng GF (1983) Proc Natl Acad Sci USA 80:3963–3965
Burger G, Scriven C, Machleidt W, Werner S (1982) EMBO J 1:1385–1391
Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW, Tabak HF (1987) Nucleic Acids Res 15:7217–7221
Campbell DA, Kubo K, Clark CG, Boothroyd JC (1987) J Mol Biol 196:113–124
Cech TR (1988) Gene 73:259–271
Clarke CG, Tague BW, Ware VC, Gerbi SA (1984) Nucleic Acids Res 12:6197–6220
Collins RA, Lambowitz AM (1983) Plasmid 9:53–70
Dale RM, McClure BA, Houchins JP (1985) Plasmid 13:31–40
Davies RW, Waring RB, Ray JA, Brown TA, Scazzocchio C (1982) Nature 300:719–724
Devereux J, Haeberli P, Smithies O (1984) Nucleic Acids Res 12:387–395
Dron M, Rahire M, Rochaix JD (1982) Nucleic Acids Res 10:7609–7620
Dujon B (1981) In: Strathern JN, Jones EW, Broach JR (eds) Molecular biology of the yeast *Saccharomyces.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 505–635
Erickson JM, Rahire M, Rochaix JD (1984) EMBO J 3:2753–2762
Gauthier A, Turmel M, Lemieux C (1988) Mol Gen Genet 214:192–197
Gray MW, Sankoff D, Cedergreen RJ (1984) Nucleic Acids Res 12:5837–5852
Gutell RR, Weiser B, Woese CR, Noller HF (1985) Prog Nucleic Acids Res 32:155–216
Hensgens LAM, Arnberg AC, Van der Veen R, Van Ommen GJB, Cerivell LA (1983) J Mol Biol 164:35–58
Holl J, Schmidt C, Schweyen RJ (1985) In: Quagliariello et al. (eds) Achievements and perspectives of mitochondrial research, vol II: Biogenesis. Elsevier Science, Amsterdam, pp 227–236
Jupe ER, Chapman RL, Zimmer EA (1988) BioSystems 21:223–230
Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Proc Natl Acad Sci USA 82:6955–6959
Lemieux B, Lemieux C (1985) Curr Genet 10:213–219
Lemieux B, Turmel M, Lemieux C (1985) BioSystems 18:293–298
Lemieux C, Lee RW (1987) Proc Natl Acad Sci USA 84:4166–4170
Lemieux C, Turmel M, Seligy VL, Lee RW (1984) Proc Natl Acad Sci USA 81:1164–1168
Lemieux C, Turmel M, Lee RW, Bellemare G (1985) Plant Mol Biol 5:77–84
Michel F, Cummings DJ (1985) Curr Genet 10:69–79
Michel F, Dujon B (1983) EMBO J 2:33–38
Michel F, Jacquier A, Dujon B (1982) Biochimie 10:867–881
Moazed D, Noller HF (1986) Cell 47:985–994
Noller HF (1984) Annu Rev Biochem 53:119–162
Norrander F, Kampe T, Messing J (1983) Gene 26:101–106
Palmer JD, Boynton JE, Gilham NW, Harris EH (1985) In: Steinbuck RE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 269–278
Rochaix JD, Rahire M, Michel F (1985) Nucleic Acids Res 13:975–984
Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74:5463–5467
Schneider M, Darlix JL, Erickson J, Rochaix JD (1985) Nucleic Acids Res 23:8531–8541
Shub DA, Gott JM, Xu MQ, Lang BF, Michel F, Tomaschewski J, Pedersen-Lane J, Belfort M (1988) Proc Natl Acad Sci USA 85:1151–1155
Spencer DF, Collings JC, Schnare MN, Gray MW (1987) EMBO J 6:1063–1071
Stern S, Wilson RC, Noller HF (1986) J Mol Biol 192:101–110
Turmel M, Bellemare G, Lemieux C (1987) Curr Genet 11:543–552
Turmel M, Lemieux B, Lemieux C (1988) Mol Gen Genet 214:412–419

Communicated by Barbara B. Sears

Received October 13, 1988