5242–5250 Nucleic Acids Research, 1993, Vol. 21, No. 22

Group I introns interrupt the chloroplast *psa*B and *psb*C and the mitochondrial *rrn*L gene in *Chlamydomonas*

Monique Turmel*, Jean-Patrick Mercier and Marie-José Côté+
Département de Biochimie, Faculté des Sciences et de Génie, Université Laval, Québec G1K 7P4, Canada

Received June 2, 1993; Revised and Accepted September 8, 1993 GenBank accession nos M90639–M90641 (incl.)

ABSTRACT

The polymerase chain reaction was used to identify novel IAI subgroup introns in cpDNA-enriched preparations from the interfertile green algae *Chlamydomonas eugametos* and *Chlamydomonas moewusii*. These experiments along with sequence analysis disclosed the presence, in both green algae, of a single IA1 intron in the *psa*B gene and of two group I introns (IA2 and IA1) in the *psb*C gene. In addition, two group I introns (IA1 and IB4) were found in the peptidyltransferase region of the mitochondrial large subunit rRNA gene at the same positions as previously reported *Chlamydomonas* chloroplast introns. The 188 bp segment preceding the first mitochondrial intron revealed extensive sequence similarity to the distantly spaced rRNA-coding modules L7 and L8 in the *Chlamydomonas reinhardtii* mitochondrial DNA, indicating that these two modules have undergone rearrangements in *Chlamydomonas*. The IAI introns in *psa*B and *psb*C were found to be related in sequence to the first intron in the *C.moewusii* chloroplast *psb*A gene. The similarity between the former introns extends to the immediate 5′ flanking exon sequence, suggesting that group I intron transposition occurred from one of the two genes to the other through reverse splicing.

INTRODUCTION

Group I introns are widespread, occurring in the organelles and nuclei of eukaryotes, in bacteria and in bacteriophages. They are characterized by the short conserved sequences P, Q, R and S that allow the formation of the typical core structure necessary for splicing (1). More than 100 group I introns have been reported in a variety of genes (2–3) and classified into ten subgroups (IA1–IA3, IB1–IB4 and IC1–IC3) on the basis of comparative sequence analysis (2). Although group I introns are present in a wide diversity of organisms and genomes, they appear to be of recent origin (4–5), with the exception of the introns present in the tRNA-Leu genes (UAA anticodon) of most chloroplast and all cyanobacterial lineages (6–7) examined and in the tRNA-Ile and tRNA-Arg genes of other bacterial phyla (8). The restricted and idiosyncratic distribution of group I introns from organelles together with the proven mobility of these genetic elements support the view that group I introns spread into protein-coding genes and rRNA genes after the origin of mitochondria and chloroplasts (4). Similarly, the restricted distribution of bacteriophage and nuclear group I introns, some of which are mobile and share similarity to mitochondrial introns, suggests that they are of recent vintage (5). In the case of the *Tetrahymena* nuclear introns, it has been shown that they result from several independent insertions (9).

Several circumstantial lines of evidence support that most group I introns arrived relatively recently at their present locations through transposition events and/or lateral transfers. These two types of intron mobility (10) are thought to be initiated by double-strand DNA breaks caused by intron-encoded endonucleases (10) and/or by reversal of the self-splicing reaction followed by reverse transcription and homologous recombination (11). In principle, the latter RNA-based mechanism would be expected to yield a larger number of novel intron insertion sites for the two following reasons. First, the reverse self-splicing reaction requires only a limited target site, i.e. a very short RNA sequence (4–7 nt) that base-pairs to the 5′ intron sequence called the internal guide sequence (11); and second, the integration of intron-containing cDNAs into genomic DNA is greatly facilitated by long regions of perfect sequence identity between the cDNA sequences flanking the intron and the corresponding genomic locus (5). In the DNA-based mechanism initiated by intron-encoded endonucleases, the creation of novel intron insertion sites would be limited by the high specificity of these enzymes (12). In support of the reverse self-splicing mechanism as the preferential mode of intron transposition, the rRNA regions corresponding to most of the known group I intron insertion positions in small and large subunit rRNA genes have been found to coincide with functional sites (13). Assuming that intact ribosomal subunits are the substrate for the reverse self-splicing reaction, such regions would be the only ones accessible for the insertion of intron RNA sequences.

* To whom correspondence should be addressed

+ Present address: Agriculture Canada, Central Plant Health Laboratory, 3851 Fallowfield Road, Nepean, Ontario K2H 8P9, Canada

A detailed characterization of the abundant group I introns in the chloroplast genomes of unicellular green algae from the highly diversified *Chlamydomonas* genus could provide further insights into the mode of insertion and proliferation of these genetic elements. Eleven of the 17 known intron insertion sites in the *rrn*L gene have been identified specifically in *Chlamydomonas* chloroplast DNAs (cpDNAs) during a phylogenetic study involving 17 taxa (13). Five of the 12 *Chlamydomonas* chloroplast *rrn*L intron insertion sites were observed in the two major *Chlamydomonas* evolutionary lineages, represented by *Chlamydomonas reinhardtii* and *Chlamydomonas moewusii* (13–14); whereas the 7 remaining ones were found to be restricted to one or the other of these lineages, a result suggesting that the origin of these sites is more recent. Besides rrn*L*, two *Chlamydomonas* chloroplast genes, *psb*A (15–16) and *rrn*S (17) have been shown to contain group I introns. The *psb*A genes of *C.reinhardtii* (15) and *C.moewusii* (16) display a total of six group I introns inserted at five positions. A single group I intron interrupts the *C.moewusii* chloroplast *rrn*S gene (17), whereas the corresponding *Chlamydomonas eugametos* (17) and *C.reinhardtii* (18) genes are unsplit. The sequences of the *C.moewusii psb*A (16), *rrn*S (17) and *rrn*L (19) introns as well as those of the *C.eugametos* (20), *C.reinhardtii* (21) and *Chlamydomonas humicola rrn*L (22) introns have been reported; they fall within the IAI, IA3 and IB4 subgroups (20). Three of these 15 chloroplast introns, CeLSU·5 (23); CrLSU·1 (24) and ChuLSU·1 (22), have been shown to encode distinct endonucleases that cleave specifically the exon junction sequence in the corresponding intronless genes.

In the present study, we have used the polymerase chain reaction (PCR) to identify novel IAI subgroup introns in cpDNA-enriched preparations of the interfertile green algae *C.eugametos* and *C.moewusii*. Comparative sequence analysis of the newly characterized introns with those already reported in *Chlamydomonas* revealed close relationships between introns inserted at different sites, with one pair of closely related introns having possibly arisen by transposition through reverse self-splicing.

METHODS

PCR amplifications

PCR amplifications were carried out in a DNA thermal cycler (Perkin Elmer, Norwalk, CT) using the oligodeoxyribonucleotides listed in Table I and cpDNA-enriched preparations (250 ng) as templates (13). Reactions were run in the presence of 0.2 mM of each dNTP, 1.3 μM of each of the primers, 10 mM Tris–HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 0.002% (w/v) gelatin, and 2.5 units of AmpliTaq DNA polymerase (Perkin Elmer, Norwalk, CT) in 100 μl total volume. Thirty cycles of amplification (1 minute at 94°C, 2 minutes at 37°C, 2 minutes plus 5 additional seconds at each consecutive cycle at 72°C) followed by a 10 minute extension at 72°C were done. Aliquots (4 μl) of the PCR reactions were electrophoresed on 1.5% agarose gels.

Southern blot hybridizations

C.eugametos and *C.moewusii* cpDNA-enriched preparations as well as recombinant plasmids whose inserts collectively cover these green algal cpDNAs (25–26) and mitochondrial DNAs (mtDNAs) (27–28) were digested with appropriate restriction enzymes. The resulting fragments were electrophoresed on 0.8% agarose gels, transferred on Hybond-N™ nylon membranes (Amersham, Arlington Heights, IL) as recommended by the manufacturer, and subsequently hybridized with ^{32}P-labelled PCR-amplified fragments under the conditions outlined by Lemieux and Lemieux (29). PCR products were purified by electroelution from 5% acrylamide or 1% agarose gels and labelled with [α-^{32}P]dCTP (3000 Ci/mmol) using the multiprime DNA labelling system (Amersham, Arlington Heights, IL).

DNA sequence determination and analysis

Three recombinant clones isolated from distinct shotgun libraries served as the primary source of DNA for sequencing of the *C.eugametos psb*C, the *C.moewusii psa*B and part of the *C.moewusii* mitochondrial *rrn*L gene. The clone carrying the *C.eugametos* chloroplast *Eco*RI fragment 2′ (25) and that carrying the *C.moewusii* chloroplast *Eco*RI fragment 10′ (26) contain the entire *psb*C and *psa*B genes, respectively, whereas the *C.moewusii* mtDNA clone 13.38 (27) carrying a partially digested *Sau*3AI fragment encodes a segment of the *rrn*L gene. Defined restriction fragments mapping in the regions where PCR-amplified intron IA1 sequences revealed hybridization signals were recovered from agarose gels and cloned into pBluescript KS– or SK– (Stratagene, La Jolla, CA) using *Escherichia coli* strain DH5αF′IQ (Bethesda Research Laboratories, Gaithersburg, MD) as a host. Double stranded DNA templates were sequenced as described previously (13). Sequencing reactions were initiated with T3 and T7 primers as well as with synthetic oligodeoxyribonucleotides (17-mers) complementary to exon or intron sequences. Sequence analysis was performed using the University of Wisconsin GCG software package (30). An alignment of conserved sequences in *Chlamydomonas* IA1 introns was done manually and was based on that published by Michel and Westhof (2). Sequence comparison was restricted to 104 positions constituting the conserved secondary structure elements for which no aligment ambiguities occur. Only one of the two members in pairs of compensatory bases was included in the data set. Evolutionary distance between pairs of introns was estimated by means of K_{nuc} values using DNADIST in the PHYLIP package, version 3.4 (31). The consensus neighbor-joining tree shown in Fig. 4 was constructed by bootstrapping with 1000 iterations using the SEQBOOT, NEIGHBOR (options M and O) and CONSENSE programs of PHYLIP. Maximum parsimony analyses were carried out with the HEURISTIC subroutine of PAUP, version 3.1.1 (32) using the following options: random addition, MULPARS, TBR branch-swapping and 50% majority rule.

RESULTS

PCR amplification and identification of novel IA1 introns in the organelle genomes of *C.eugametos* and *C.moewusii*

Eight IA1 introns representing five insertion sites have been described in the chloroplast *psb*A (16) and *rrn*L (19–20) genes of *C.eugametos* and *C.moewusii*; they consist of CmpsbA·1, CmLSU·4 and three pairs of rDNA introns occupying the same insertion sites: CeLSU·2/CmLSU·1, CeLSU·3/CmLSU·2, CeLSU·4/CmLSU·3. To identify additional IA1 introns in the organelle genomes of these interfertile algae, we amplified by PCR the intron regions comprised between the conserved P and S elements using as primers eight degenerate oligonucleotides representing all the variations found in the P and S sequences of the previously reported *Chlamydomonas* IA1 introns. As

shown in Table I, the two oligos complementary to the P element represent a collection of eight distinct sequences, whereas the six oligos complementary to the S element represent 24 sequences. A+T-rich DNA fractions recovered from equilibrium CsCl density gradients of total cellular DNA were used as templates for the PCR amplifications; in addition to cpDNA, these cpDNA-enriched preparations contain mtDNA (27). For each alga, the PCR amplifications were carried out using all 12 pairwise combinations of P and S oligonucleotide primers, and the resulting PCR products were analyzed by agarose gel electrophoresis. Each PCR amplification typically yielded one to five bands, with fragments ranging from 180 to 3000 bp in size (Table II). As expected, a number of products were found to be derived from the previously identified *psb*A and *rrn*L introns, their sizes being in agreement with those predicted for the sequences delimited by the P and S elements of these introns. To determine the origin of the remaining PCR fragments, we purified and hybridized them to Southern blots containing *Ava*I, *Bst*EII, *Eco*RI and *Hin*dIII fragmens of *C.eugametos* and *C.moewusii* cpDNA-enriched preparations as well as to Southern blots containing cloned *Eco*RI fragments from a collection of plasmids covering the entire chloroplast genome of these algae (25–26). Because the hybridization signals of two PCR products (one from each alga) to the cpDNA-enriched preparations apparently corresponded to mtDNA fragments, these products and three others that revealed no hybridization to cpDNA were also hybridized to Southern blots of *C.eugametos* and *C.moewusii* mtDNA clones that were previously digested with *Hin*dIII and *Sau*3A1, respectively, to release the inserts from the vectors.

Most of the hybridization signals of chloroplast origin were mapped to the *rrn*L gene, while the others were located to two distinct cpDNA regions where *psa*B and *psb*C gene probes were previously found to hybridize (33) (Table II). The 20 PCR products larger than 750 bp, which mapped to the *rrn*L locus, all correspond to regions delimited by the P and S elements of separate IA1 introns in this gene. A *C.eugametos* PCR product of 235 bp and an homologous *C.moewusii* fragment of 240 bp hybridized to a 4.8 kbp cpDNA region containing *psa*B, with the *C.moewusii* cpDNA fragments detected being *Eco*RI 10′, *Ava*I 2′ and *Bst*EII 1. In addition, a *C.eugametos* PCR product of 250 bp and a corresponding *C.moewusii* product of identical size were found to hybridize to a 3.5 kbp cpDNA region encoding *psb*C; i.e., to the *C.eugametos* fragments *Eco*RI 2, *Ava*I 1 and *Bst*EI1 13.

A third set of homologous *C.eugametos* and *C.moewusii* PCR products of 340 bp and 335 bp, respectively, revealed positive hybridization signals to a common 2.1 kbp mtDNA region where

Table I. Oligodeoxyribonucleotide primers used in this study

Name	Sequence (5′ → 3′)
P (C)	AAAA(T/C)(T/C)GGGTGAATTGC
P (T)	AAAA(T/C)(T/C)GGGTGAATTGT
S (AA)	(A/G)GACTATATC(A/T)TCAACT
S (AC)	(A/G)GACTATATC(A/T)TCACCT
S (AT)	(A/G)GACTATATC(A/T)TCATCT
S (GA)	(A/G)GACTATGTC(A/T)TCAACT
S (GC)	(A/G)GACTATGTC(A/T)TCACCT
S (GT)	(A/G)GACTATGTC(A/T)TCATCT

Table II. Size and identity of the PCR products generated from *C.eugametos* and *C.moewusii* cpDNA-enriched preparations with the P and S sequence-specific oligonucleotides described in Table 1

	C. eugametos products (bp)		*C. moewusii* products (bp)	
	P(C)	P(T)	P(C)	P(T)
S (AA)	800 (LSU•2/LSU•3) 260 (LSU•2)	340 (mtLSU•1) 300 (LSU•3) 250 (psbC•2)	1200 (LSU•3/LSU•4) 800 (LSU•1/LSU•2) 200 (LSU•4)	335 (mtLSU•1) 310 (LSU•2) 250 (psbC•2)
S(AC)	260 (LSU•2) 235 (psaB•1)	340 (mtLSU•1) 250 (psbC•2)	2100 (psbA•1) 240 (psaB•1) 180 (ND)	250 (psbC•2)
S(AT)	3000 (ND 2100 (LSU•2/LSU•4) 210 (LSU•4)	1600 (LSU•3/LSU•4)	2100 (LSU•1/LSU•3) 1600 (LSU•2/LSU•3) 210 (LSU•3)	1600 (LSU•2/LSU•3)
S(GA)	800 (LSU•2/LSU•3) 260 (LSU•2)	340 (mtLSU•1) 300 (LSU•3) 250 (psbC•2)	1200 (LSU•3/LSU•4) 800 (LSU•1/LSU•2) 260 (LSU•1) 220 (ND) 200 (LSU•4)	335 (mtLSU•1) 310 (LSU•2) 250 (psbC•2)
S(GC)	2100 (LSU•2/LSU•4) 800 (LSU•2/LSU•3) 260 (LSU•2) 210 (LSU•4)	1600 (LSU•3/LSU•4) 340 (mtLSU•1) 250 (psbC•2)	210 (LSU•3)	1600 (LSU•2/LSU•3) 335 (mtLSU•1) 250 (psbC•2)
S(GT)	2100 (LSU•2/LSU•4) 210 (LSU•4)	1600 (LSU•3/LSU•4) 340 (mtLSU•1)	2100 (LSU•1/LSU•3) 1600 (LSU•2/LSU•3) 210 (LSU•3)	1600 (LSU•2/LSU•3) 335 (mtLSU•1) 250 (psbC•2)

In parentheses are indicated the introns from which the PCR products were derived. When fragments were amplified from the P sequence of one intron and the S sequence of another intron, the names of the two introns are separated by a/. References for the introns are as follows: CeLSU·2-CeLSU·4 (20), CmLSU·1-CmLSU·4 (19), CmpsbA·1 (16), CepsaB·1 and CmpsaB·1 (this study), CepsbC·2 and CmpsbC·2 (this study), CemtLSU·1 and CmmtLSU·1 (this study). Introns names are abbreviated as in reference 2. ND, not determined.

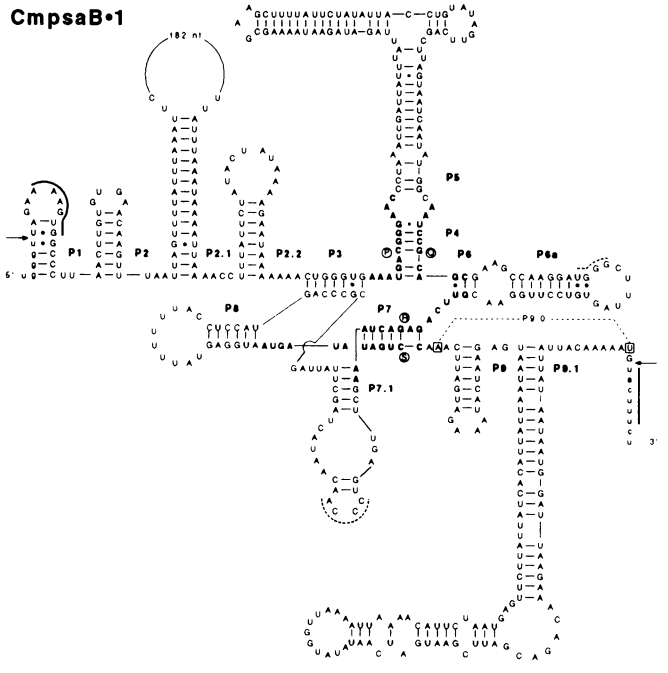

Figure 1. Secondary structure model of the *C.moewusii psa*B intron (CmpsaB·1). Splice sites are denoted by arrows between the exon (lowercase) and intron (uppercase) nucleotides. The nucleotides that are part of the conserved sequence elements P, Q, R and S are indicated in boldface. The intron nucleotides from the internal guide sequence and the exon nucleotides at the 3′ splice site, which can base pair to form P10, are denoted by thick continuous lines. The intron nucleotides in the loops subtending P7.1 and P6a, which can base-pair to form P11, are denoted by broken lines.

*rrn*L and *rrn*S gene probes had been previously mapped (28); the hybridizing *C.moewusii* mtDNA fragments detected consisted of *Ava*I 1, *Bst*EII 2, *Hin*dIII 3, and the *Sau*3AI partial digest fragment 13.38. The three other PCR products (a 3 kbp *C.eugametos* fragment and *C.moewusii* fragments of 220 and 180 bp) that were probed to the cpDNA and mtDNA failed to reveal any positive hybridization signals. It is possible that they were amplified from nuclear group I introns, as these genetic elements appear to be relatively common in the nuclei of green algae (34–35).

To determine the identity of the chloroplast and mitochondrial genes hybridizing with the PCR-amplified fragments, we sequenced the regions of three recombinant plasmids in which these hybridization signals were confined.

The *C.moewusii psa*B contains a single IA1 intron

The nucleotide sequence of a 3086 bp segment from the *C.moewusii* chloroplast *Eco*RI fragment 10′ (GenBank accession no.: M90641) revealed that the chloroplast gene encoding the P700 chlorophyll *a* apoprotein A2 of photosystem I is composed of two exons containing 591 and 144 codons and of a single IA1 intron (CmpsaB·1) comprising 655 nucleotides. The intron/exon boundaries were predicted by comparison with the intronless *C.reinhardtii psa*B gene (36). The *psa*B gene product of this divergent green alga shows 92% identity with its *C.moewusii* counterpart and features one extra amino acid, either the leucine or the glutamine residues corresponding to codons 203 and 204, respectively. In land plant *psa*B genes, an additional residue is also observed at these positions, suggesting that the missing codon in *C.moewusii* is due to a deletion.

As shown in Fig. 1, the CmpsaB·1 intron sequence can be folded into an RNA secondary structure that exhibits the typical group I ribozyme core structure consisting of the base-paired regions P3 to P8 and connecting segments (1). As is the case in other group I introns, the 5′ splice site of CmpsaB·1 lies in P1 after a U residue that pairs with a G and the 3′ splice site occurs after a G. Regarding the proposed P9.0 (37–38) and P10 (39) pairings, which are widespread but not universal in group I introns, they are found in CmpsaB·1 and consist of one and five Watson–Crick bp, respectively. On the basis of a comparative sequence analysis of the conserved cores of CmpsaB·1 and other group I introns, CmpsaB·1 was classified into subgroup IA1. The P11 pairing (2, 40) as well as most other interactions (2) recently reported in this subgroup are present in CmpsaB·1.

The *C.eugametos psb*C is interrupted by two group I introns: CepsbC·1 belongs to the IA2 subgroup and CepsbC·2 is part of the IA1 subgroup

The nucleotide sequence of a 3230 bp segment from the *C.eugametos* chloroplast *Eco*RI fragment 2′ (GenBank accession no.: M90639) revealed that the gene encoding the 43 kDa chlorophyll *a*-binding protein of photosystem II (also called P6) is composed of three exons containing 181, 113 and 167 codons and of two group I introns comprising 1028 (CepsbC·1) and 510 (CepsbC·2) nucleotides. As for the *psa*B gene, the intron/exon boundaries were predicted by comparison with the corresponding intronless *C.reinhardtii* gene (41). The *C.eugametos* and *C.reinhardtii psb*C gene products share 94.4% sequence identity and the same number of amino acid residues, assuming that the initiation codon is GTG in both algae. In land plants, it is not clear if translation of the *psb*C mRNA is initiated at the GUG corresponding to this same position or at the ATG situated 12 codons upstream (42). Because the N-terminus of the 43 kDa protein of spinach starts with N-acetyl-O-phospho-threonine at residue 15 of the amino acid sequence predicted by this extended reading frame (43), the protein must be necessarily synthesized as a precursor that is processed to the mature form by removing 14 or two residues, depending on whether initiation of translation occurs at the AUG or GUG. In *C.eugametos*, the 43 kDa protein can be initiated at the AUG located 11 codons upstream of the putative GUG initiaton codon, but the resulting protein would show no similarity with the corresponding region of the land plant 43 kDa proteins. For this reason and also because a ribosomal binding site is found nine bases upstream of the conserved GTG in the two *Chlamydomonas* species and in land plants, it is likely that the initiation codon of *psb*C is GTG in all these organisms.

The secondary structure models of CepsbC·1 and CepsbC·2 are typical of group I introns (Fig. 2). As expected, CepsbC·2, the intron detected during our PCR amplifications (Table I), falls within subgroup IAI. Three unusual features are observed in the core structure of this intron: four bases, instead of three, are present in the joining segment between P3 and P4; G·U, instead of a Watson-Crick base-pair, is the first base-pair of P4; and finally, the nucleotides U and C following the latter base-pair are unpaired. CepsbC·1 was classified into subgroup IA2, as it exhibits most of the structural and sequence properties characteristic of this subgroup, which is closely related to the IA3 subgroup (2). Like most IA2 introns, but unlike IA3 introns, CepsbC·1 displays six base-pairs in P3, three bases in the joining segment between P3 and P4, four bases in the joining segment between P4 and P5, and the conserved nucleotides GUA in the segment between P7.1 and P7.2 (2). In addition, the first nucleotide 3′ of P7 can base-pair with the C in the conserved AACA sequence flanking the terminal G of the intron. The P6 of CepsbC·1, however, is unusual in displaying A/U and U/A as the first two base-pairs, instead of G/C,U and C/G. Despite the large size of CepsbC·1, no long open reading frame could be identified in this intron. The 579 nucleotide region residing in the proximity of P6b can fold into a complex secondary structure (data not shown) and contains three short inverted repeat elements that belong to one of the repeat families (the c family) previously identified in other *C.eugametos* and *C.moewusii* chloroplast group I introns and also in noncoding regions flanking the chloroplast genes of these green algae (20). One repeat element of this family as well as one repeat of the a family were observed in the P9.2 stem of CepsbC·2. An intron at the same position as CepsbC·1 is likely to be present in the *C.moewusii* cpDNA as revealed by PCR amplification with oligos complementary to the exon sequences immediately flanking the insertion site of CepsbC·1 (data not shown).

Two group I introns are present in the *C.moewusii* mitochondrial *rrn*L coding region corresponding to the distantly spaced *C.reinhardtii* coding modules L7 and L8

The nucleotide sequence of a 927 bp segment from the *C.moewusii* mtDNA clone 13.38 (GenBank accession no.: M90640) revealed the presence of two group I introns in the segment of the *rrn*L gene which encodes the peptidyltransferase region corresponding to positions 2237 to 2500 in the *E.coli* 23S rRNA. The intron detected during the PCR amplifications carried out in the present study (CmmtLSU·1) comprises 482 nucleotides and is inserted between positions 2449 and 2450 in the *E.coli* 23S rRNA. The second intron, CmmtLSU·2, has been partially

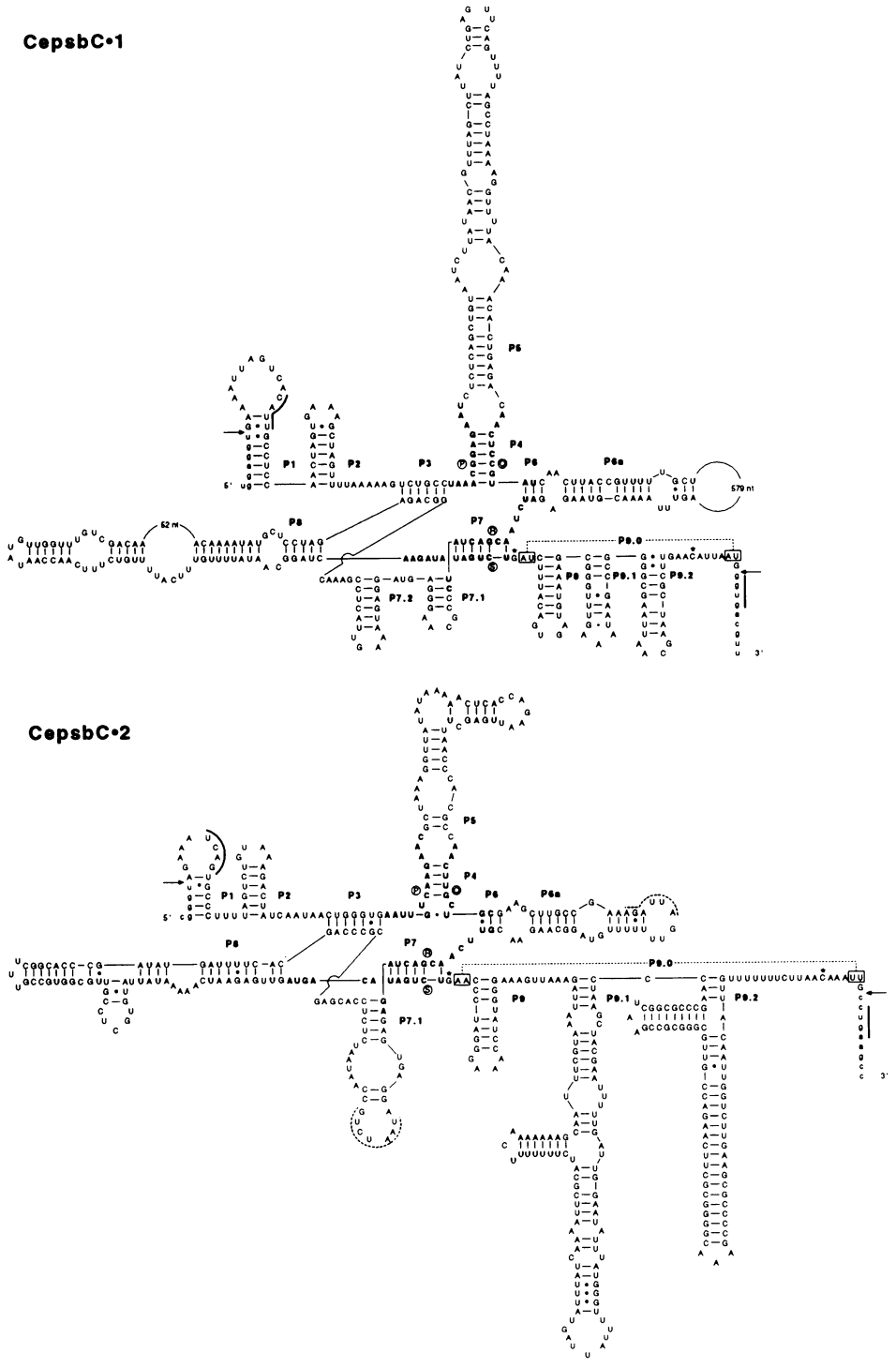

Figure 2. Secondary structure models of the *C.eugametos psb*C introns (CepsbC·1 and CepsbC·2). The intron nucleotides denoted by asterisks can base pair. For more details, see the legend of Fig. 1.

sequenced (206 nucleotides) and found to lie between residues 2500 and 2501 in this bacterial rRNA. The intron/exon boundaries were determined by comparison with the corresponding intronless *C.reinhardtii* mitochondrial *rrn*L sequences (44). In this distantly related green alga, the mitochondrial large subunit rRNA-coding region is fragmented into eight modules that are interspersed with protein-coding genes, transfer RNA genes and small subunit rRNA-coding modules throughout a 6 kbp stretch of the 14.5 kbp genome (44). To our surprise, we found that one of the extremities of the 927 bp *C.moewusii* mtDNA segment sequenced features 188 bp of continuous *rrn*L sequences corresponding to the distantly spaced *C.reinhardtii* modules 7 and 8, or more precisely to the domain V rRNA region spanning positions 1773 and 2020 in the potential

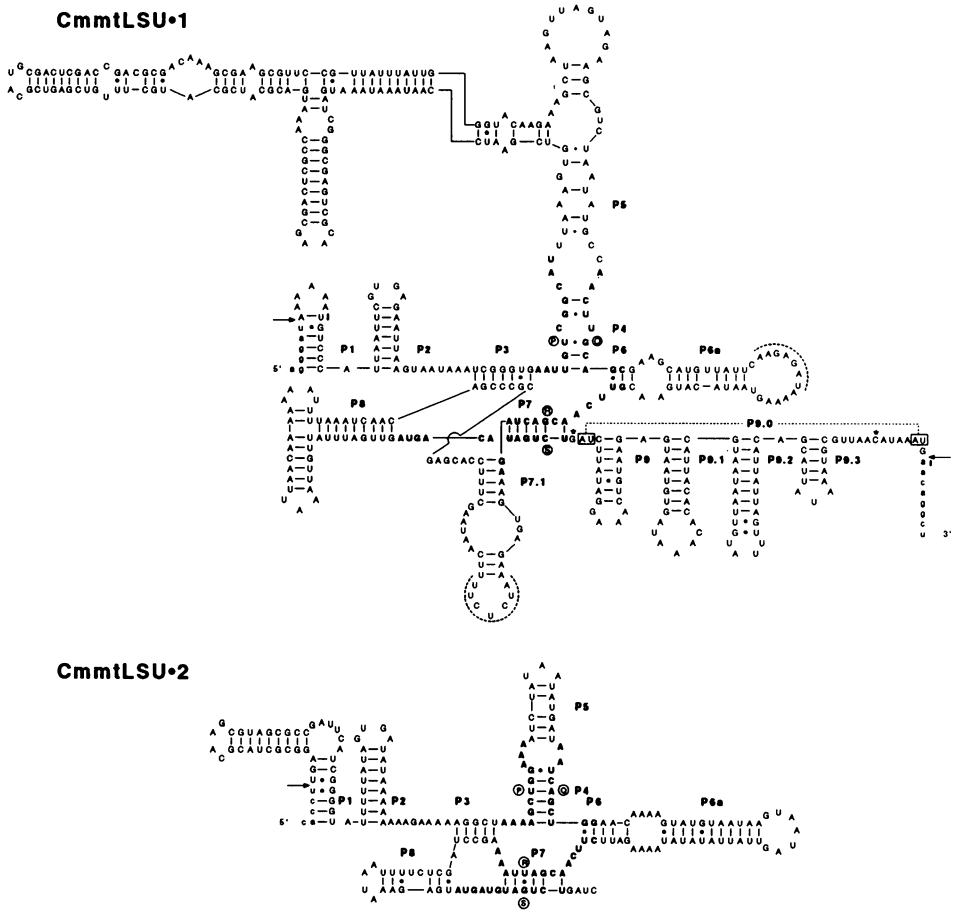

Figure 3. Secondary structure models of the *C.moewusii* mitochondrial *rrn*L introns (CmmtLSU·1 and CmmtLSU·2). The secondary structure of CmmtLSU·2 is partial, the sequence covering P9 and the terminus of this intron being not available. The intron nucleotides denoted by asterisks can base pair. For more details, see the legend of Fig. 1.

secondary structure proposed for this green algal mitochondrial large subunit rRNA (44). The secondary model of this rRNA segment from *C.moewusii* is highly similar to that from *C.reinhardtii*, with the exception of the terminal region of the conserved 9 bp helix which corresponds to the break between the *C.reinhardtii* rRNA fragments encoded by modules 7 and 8 (L7 and L8) (data provided to the reviewers). Because only five single-stranded bases extend this helix in *C.moewusii* as opposed to the stretches of eight and 57 single-stranded bases in the mature *C.reinhardtii* L7 and L8 rRNAs, respectively, it is possible that the *C.moewusii* mitochondrial large subunit rRNA is not fragmented at that site. Obviously, rRNA analysis will be necessary to test this hypothesis. In the interfertile alga *C.eugametos*, an *Hin*dIII fragment containing an insertion of 100 bases relative to *C.moewusii* (28) spans this variable rRNA region as well as the 5′ portion of the adjacent intron, implying that one of these regions is polymorphic.

As expected, the sequence and secondary structure of CmmtLSU·1 are consistent with this intron being part of the IA1 subgroup (Fig. 3). Its most unusual feature is the presence of two unpaired residues (C and U) lying between two wobble bp (G/U) in P4. Given the similarity in size of the *C.eugametos* and *C.moewusii* mitochondrial intron segments amplified with the P and S oligonucleotides at this locus, it is possible that the additional 100 bases found in the aforementioned *C.eugametos* *Hin*dIII fragment represent extra base-paired regions in the segment 3′ to P3.

The partial sequence of CmmtLSU·2 features the conserved regions P3 to P8 found in all group I introns (Fig.3). On the basis of a comparative sequence analysis, this intron was classified into subgroup IB4. It differs from most IB4 introns in displaying A/U as the first bp of P4, instead of U/A, and G as the sixth position of sequence S, instead of A. Because the intron segment analyzed resides at an extreme end of the *Sau*3A1 mtDNA fragment carried by clone 13.38, completion of the secondary structure of CmmtLSU·2 will require a separate mtDNA fragment overlapping this region. Such a fragment could not be cloned during the construction of the *C.moewusii* mtDNA library (27).

DISCUSSION

We identified five group I intron insertion sites in the chloroplast *psa*B and *psb*C and the mitochondrial *rrn*L of *Chlamydomonas*. In *C.eugametos* and *C.moewusii*, each of these genes contains a subgroup-IAI intron; in addition, the *psb*C from both algae is interrupted by a IA2 intron and the *C.moewusii* mitochondrial *rrn*L displays a IB4 intron. Prior to our study, no group I introns had been described in *psa*B and *psb*C, suggesting a relatively recent origin for the three *Chlamydomonas* intron insertion sites

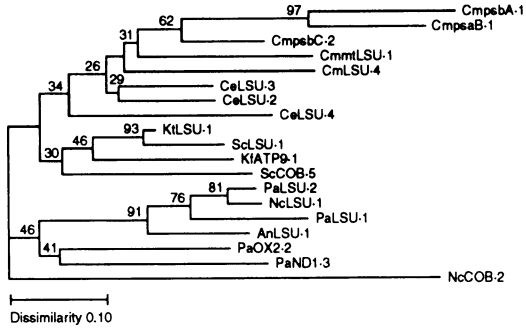

Figure 4. Relationships among organelle IAI introns as determined by neighbor-joining analysis. This tree was derived from a Kimura dissimilarity matrix that was computed from a data set of 104 positions. Branch lengths on the horizontal axis represent the evolutionary distance between the nodes. The numbers at the nodes indicate the percentage of times (≥ 25) that the monophyletic group to the right occurred in 1000 bootstrap samples. Intron sequences were taken from: *Chlamydomonas* chloroplast and mitochondrial introns, see the footnote of Table II; other mitochondrial introns, see the references cited by Michel and Westhof (2).

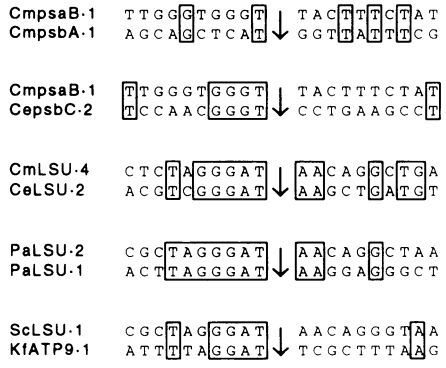

Figure 5. Comparison of exon sequences flanking structurally related IA1 introns inserted at distinct sites. The arrows denote the positions of the introns. Exon sequences were taken from: *Chlamydomonas* chloroplast introns, see the footnote of Table II; PaLSU·1 and PaLSU·2, reference 51; ScLSU·1 and KfATP·9, reference 10.

in these genes. Although group II introns have been identified in the *psa*B and *psb*C of *Euglena gracilis*, their positions differ from those identified here (45–46). Introns occupying the insertion site of the IAI CmmtLSU·1 are present in the mitochondria of *Saccharomyces*, *Neurospora*, *Aspergillus* and *Podospora*, in the chloroplasts of three *Chlamydomonas* taxa belonging to the two major lineages observed in this genus, and in the nucleus of *Physarum* (see 13). The *Physarum* nuclear intron inserted at this position falls within the IC3 subgroup, whereas all of the organelle introns characterized at this same site, including the mobile ScLSU·1 intron encoding the endonuclease I-*Sce*I (12), share with CmmtLSU·1 the IA1 subgroup (2). Similarly, the position of the IB4 CmmtLSU·2 has been found to be occupied by chloroplast introns from four *Chlamydomonas* taxa belonging to the two major lineages (13), and interestingly, the unique intron examined at this position (ChLSU·1) is a member of the IB4 subgroup which encodes a site-specific endonuclease (22).

Our finding that *Chlamydomonas* mitochondrial and chloroplast *rrn*L introns inserted at common positions belong to the same subgroup suggests that group I introns can move from one organelle to another within the cell and spread themselves at their cognate site. The movement of introns between organelle genomes is a likely possibility given that mitochondria have been found associated with the chloroplast during the vegetative cell-cycle of *Chlamydomonas* (47) and that transfer of cpDNA sequences to the mtDNA appears to be an ongoing evolutionary process in land plants (48–49). Mobile introns encoding endonucleases would be predicted to integrate into organelle genomes much more efficiently than their non-mobile counterparts. In this context, it is not unexpected to find introns encoding site-specific endonucleases (ScLSU·1 and ChLSU·1) at the two insertion sites identified here in *Chlamydomonas* mitochondria. As *Chlamydomonas* mitochondrial and chloroplast *rrn*L introns inserted at common positions do not appear to be very closely related (see Fig. 4), it is likely that the postulated movement of group I introns did not occur within recent times in *Chlamydomonas* lineages and/or that equivalent introns have evolved independently in separate cytoplasmic genomes. Obviously, the possibility cannot be eliminated that mitochondrial rrnL introns from one organism were transferred to the cpDNA of a separate organism or vice-versa. To gain insight into the evolutionary origin of organelle *rrn*L introns, it will be necessary to analyze the distribution of chloroplast and mitochondrial *rrn*L introns among the green algae and to undertake comparative sequence analyses of these introns as well as of their flanking exons.

Two of the three *Chlamydomonas* chloroplast IA1 introns identified in this study were found to be closely related to the previously reported chloroplast CmpsbA·1 intron. As shown in Fig. 4, our neighbor-joining analysis of organelle IAI intron sequences from *Chlamydomonas* and other organisms revealed that two groups of closely related introns featuring distinct insertion sites clustered together more than 60% of the time: 1) CmpsbA·1, CmpsaB·1 and CepsbC·2, and 2) PaLSU·2, NcLSU·1, PaLSU·1 and AnLSU·1. The latter group of fungal mitochondrial *rrn*L introns (featuring two insertion sites, one of which coincides with that occupied by CmLSU·4 and CmmtLSU·1) was expected, as Cummings *et al.* (50) reported extensive sequence and structural similarities between the members of this group. Although Dujon *et al.* (51) observed great similarities between ScLSU·1 and KfATP9·1, these yeast mitochondrial introns were not frequently found together during our bootstrap analysis. Using a method of analysis based on maximum parsimony, we found essentially the same intron relationships as determined by the neighbor-joining method; however, bootstrap analysis indicated a weaker support for the alliance of CepsbC·2 with the CmpsbA·1 and CmpsaB·1 pair. The most similar *Chlamydomonas* IAI introns are CmpsaB·1 and CmpsbA·1; 129 out of 150 positions spanning the regions P3 to P7, P7.1 and P8 share identical residues in these introns. The secondary structure of CmpsaB·1 mainly differs from that of CmpsbA·1 by the lengths of the terminal loops subtending P5, P6a, P8 and P9.1, and by the presence of two extra P2 stems (P2.1 and P2.2).

Like the pairs of related mitochondrial IAI introns PaLSU·2/PaLSU·1 and ScLSU·1/KfATP9·1, the related CmpsaB·1 and CepsbC·2 exhibit significant similarity in their flanking exon sequences, suggesting that one of the members in each of these intron pairs arose from transposition (Fig. 5). Because the sequence similarity between the exons of CmpsaB·1

and CepsbC·2 and between those of ScLSU·1 and KfATP·9 is mainly restricted to the 5′ region that base-pairs with the intron to form P1, we favour the idea that these related introns were generated through reverse self-splicing. PaLSU·2 and PaLSU·1 show more sequence similarity in their 5′ and 3′ flanking exons, notably in the 5′ exon sequence that is not involved in the formation of P1, implying that intron transposition might have occurred via cleavage of the recipient locus by an endonuclease. Such an enzyme could be encoded by the internal open reading frame present in PaLSU·1 (51). Although a close relationship between the chloroplast IA1 introns CmLSU·4 and CeLSU·2 is not supported by our bootstrap analysis (see Fig. 4), it is of interest to mention that both the 5′ and 3′ flanking sequences of these introns are similar (Fig. 5). It is thus possible that these two intron insertion sites arose from a transposition event that occurred before the divergence of the *Chlamydomonas* lineages. With regard to the CmpsaB·1 and CmpsbA·1 introns featuring almost identical core structures but different flanking exons, this similarity can be attributed to nonreciprocal homologous recombination (gene conversion). CmpsbA·1 is located in the inverted repeat, a cpDNA region that is thought to participate actively in gene conversion events and that has been shown to be involved in both intra- and intermolecular recombinations (48). Evidence for gene conversion between a sequence located in the inverted repeat and a cpDNA sequence in a single-copy region has been reported in wheat (52).

CepsbC·1 is the first organelle intron ever assigned to the IA2 subgroup, a category of introns comprising bacteriophage introns (2). Because mobile IA2 introns appear to be widespread among T-even phages (12), CepsbC·1 might have been acquired through horizontal transfer of an intron from such phages. We have previously reported that the 5′ exon sequences flanking CepsbC·1 and the IA3 CeLSU·1 intron are very similar (13). Such resemblance may be due to chance or to a common evolutionary relationship between these two types of introns. As IA2 and IA3 introns share similar P and Q sequences and differ from other group I introns by the presence of two stem-loop structures between P3 and P7 (P7.1 and P7.2), the latter possibility is more likely. Like their IA2 counterparts, IA3 introns appear to be restricted in their distribution, having been identified so far exclusively in the *Chlamydomonas* chloroplasts (17, 19–21). More knowledge about the distributions of IA2 and IA3 introns will be required to understand the evolutionary relationships between these two types of introns.

ACKNOWLEDGEMENTS

We thank Claude Lemieux for helpful discussions. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (GP0003293 to M.T.), and from 'Le Fonds pour la Formation de Chercheurs et l'Aide à la Recherche' (93-ER-0350 to M.T.). M.-J.C. was supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada. M.T. is a Scholar in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

REFERENCES

1. Cech, T.R. (1988) *Gene*, **73**, 259–271.
2. Michel, F., and Westhof, E. (1990) *J. Mol. Biol.*, **216**, 585–610.
3. De Wachter, R., Neefs, J.-M., Goris, A., and Van de Peer, Y. (1992) *Nucleic Acids Res.*, **20**, 1251–1257.
4. Cavalier-Smith, T. (1991) *TIG*, **7**, 145–148.
5. Palmer, J.D., and Logsdon, J.M.Jr. (1991) *Curr. Opin. Genet. Dev.*, **1**, 470–477.
6. Kuhsel, M.G., Strickland, R., and Palmer, J.D. (1990) *Science*, **250**, 1570–1573.
7. Xu, M.Q., Kathe, S.D., Goodrich-Blair, H., Nierzwicki-Bauer, S.A., and Shub, D.A. (1990) *Science*, **250**, 1566–1570.
8. Reinhold-Hurek, B., and Shub, D.A. (1992) *Nature*, **357**, 173–176.
9. Sogin, M.L., Ingold, A., Karlok, M., Nielsen, H., and Engberg, J. (1986) *EMBO J.*, **5**, 3625–3630.
10. Dujon, B. (1989) *Gene*, **82**, 91–113.
11. Woodson, S.A., and Cech, T.R. (1989) *Cell*, **57**, 335–345.
12. Lambowitz, A.M., and Belfort, M. (1992) *Annu. Rev. Biochem.*, **62**, 587–622.
13. Turmel, M., Gutell, R.R., Mercier, J.-P., Otis, C., and Lemieux, C. (1993) *J. Mol. Biol.*, **232**, 446–467.
14. Buchheim, M.A., Turmel, M., Zimmer, E.A., and Chapman, R.L. (1990) *J. Phycol.*, **26**, 689–699.
15. Erickson, J.M., Rahire, M., and Rochaix, J.-D. (1984) *EMBO J.*, **3**, 2753–2762.
16. Turmel, M., Boulanger, J., and Lemieux, C. (1989) *Nucleic Acids Res.*, **17**, 3875–3887.
17. Durocher, V., Gauthier, A., Bellemare, G., and Lemieux, C. (1989) *Curr. Genet.*, **15**, 277–282.
18. Dron, M., Rahire, M., and Rochaix, J.-D. (1982) *Nucleic Acids Res.*, **10**, 7609–7619.
19. Gauthier, A. (1990) *Etude d'introns optionnels entre les rDNAs chloroplastiques de Chlamydomonas moewusii et Chlamydomonas eugametos*. Ph.D. thesis, Université Laval.
20. Turmel, M., Boulanger, J., Schnare, M.N., Gray, M.W., and Lemieux, C. (1991) *J. Mol. Biol.*, **218**, 293–311.
21. Rochaix, J.D., Rahire, M., and Michel, F. (1985) *Nucleic Acids Res*, **13**, 975–984.
22. Côté, V., Mercier, J.-P., Lemieux, C., and Turmel, M. (1993) *Gene*, **129**, 69–76..
23. Gauthier, A., Turmel, M., and Lemieux, C. (1991) *Curr. Genet.*, **19**, 43–47.
24. Dürrenberger, F., and Rochaix, J.-D. (1991) *EMBO J.*, **10**, 3495–3501.
25. Lemieux, C., Turmel, M., Seligy, V.L., and Lee, R.W. (1985) *Curr. Genet.*, **9**, 139–145.
26. Turmel, M., Bellemare, G., and Lemieux, C. (1987) *Curr. Genet.*, **11**, 543–552.
27. Lee, R.W., Dumas, C., Lemieux, C., and Turmel, M. (1991) *Mol. Gen. Genet.*, **231**, 53–58.
28. Denovan-Wright, E.M., and Lee, R.W. (1992) *Curr. Genet.*, **24**, 197–202.
29. Lemieux, B., and Lemieux, C. (1985) *Curr. Genet.*, **10**, 213–219.
30. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
31. Felsenstein, J. (1991) *PHYLIP: Phylogenetic Inference Package*, version 3.41. Computer program, Herbarium, University of California, Berkeley.
32. Swofford, D.L. (1993) *PAUP: Phylogenetic Analysis Using Parsimony*, version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
33. Turmel, M., Lemieux, B., and Lemieux, C. (1988) *Mol. Gen. Genet.*, **214**, 412–419.
34. Dçvila-Aponte, J.A., Huss, V.A.R., Sogin, M.L., and Cech, T.R. (1991) *Nucleic Acids Res.*, **19**, 4429–4436.
35. Wilcox, L.W., Lewis, L.A., Fuerst, P.A., and Floyd, G.L. (1992) *Mol. Biol. Evol.*, **9**, 1103–1118.
36. Kück, U., Choquet, Y., Schneider, M., Dron, M., and Bennoun, P. (1987) *EMBO J.*, **8**, 2185–2195.
37. Michel, F., Netter, P., Xu, M.-Q., and Shub, D.A. (1990) *Genes Dev.*,**4**, 777–788.
38. Burke, J.M., Esherick, J.S., Burfeind, W.R., and King, J.L. (1990) *Nature*, **344**, 80–82.
39. Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A., and Scazzocchio, C. (1982) *Nature*, **300**, 719–724.
40. Jaeger, L, Westhof, E., and Michel, F. (1991) *J. Mol. Biol.*, **221**, 1153–1164.
41. Rochaix, J.-D., Kuchka, M., Mayfield, S., Schirmer-Rahire, M., Girard-Bascou, J., and Bennoun, P. (1989) *EMBO J.*, **8**, 1013–1021.
42. Alt, J., Morris, J., Westhoff, P., and Herrmann, R.G. (1984) *Curr. Genet.*, **8**, 597–606.
43. Michel, H., Hunt, D.F., Shabanowitz, J., and Bennett, J. (1988) *J. Biol. Chem.*, **263**, 1123–1130.
44. Boer, P.H., and Gray, M.W. (1988) *Cell*, **55**, 399–411.

45. Montandon, P.-E., Vasserot, A., and Stutz, E. (1986) *Curr. Genet.*, **11**, 35–39.
46. Cushman, J.C., Hallick, R.B., and Price, C.A. (1988) *Curr. Genet.*, **13**, 159–171.
47. Osafune, T., Mihara, S., Hase, E., and Ohkuro, I. (1972) *Plant Cell Physiol.*, **13**, 981–989.
48. Palmer, J.D. (1985) In MacIntyre, R.J., (ed.) *Monographs in Evolutionary Genetics: Molecular Evolutionary Genetics.* Plenum Press, New York, pp. 131–240.
49. Joyce, P.B.M., and Gray, M.W. (1989) *Nucleic Acids Res.*, **17**, 5461–5476.
50. Cummings, D.J., Domenico, J.M., and Nelson, J. (1989) *J. Mol. Evol.*, **28**, 242–255.
51. Dujon, B., Colleaux, L., Jacquier, A., Michel, F., and Monteilhet, C. (1986) In Wickner, R.B., Hinnebusch, A., Lambowitz, A.M., Gunsalus, I.C., and Hollaender, A., (eds.), *Extrachromosomal Elements in Lower Eukaryotes.* Plenum Press, New York, pp. 5–27.
52. Bowman, C.M., Barker, R.F., and Dyer, T.A. (1988) *Curr. Genet.*, **14**, 127–136.