ORIGINAL PAPER

Debashish Bhattacharya · Simon Damberger
Barbara Surek · Michael Melkonian

Primary and secondary structure analyses of the rDNA group-I introns of the Zygnematales (Charophyta)

Abstract The Zygnematales (Charophyta) contain a group-I intron (subgroup IC1) within their nuclear-encoded small subunit ribosomal DNA (SSU rDNA) coding region. This intron, which is inserted after position 1506 (relative to the SSU rDNA of *Escherichia coli*), is proposed to have been vertically inherited since the origin of the Zygnematales approximately 350–400 million years ago. Primary and secondary structure analyses were carried out to model group-I intron evolution in the Zygnematales. Secondary structure analyses support genetic data regarding sequence conservation within regions known to be functionally important for in vitro self-splicing of group-I introns. Comparisons of zygnematalean group-I intron secondary structures also provided some new insights into sequences that may have important roles in in vivo RNA splicing. Sequence analyses showed that sequence divergence rates and the nucleotide compositions of introns and coding regions within any one taxon varied widely, suggesting that the "1506" group-I introns and rDNA coding regions in the Zygnematales evolve independently.

Key words Group-I introns · Secondary structure · Small subunit ribosomal DNA · Zygnematales

Introduction

Group-I introns are found in organellar genomes, in eubacteria, in phages, in *Chlorella* spp. viruses, and in eukar-

D. Bhattacharya (✉)
Max-Planck-Institut für biophysikalische Chemie,
Abteilung Biochemie, Am Fassberg 11, D-37077 Göttingen, Germany

S. Damberger
University of Colorado, MCD Biology, Campus Box 347,
Boulder, CO 80309-0347, USA

B. Surek · M. Melkonian
Botanisches Institut, Lehrstuhl I, Universität zu Köln,
Gyrhofstrasse 15, D-50931 Köln, Germany

Communicated by R. W. Lee

yotic nuclear-encoded ribosomal DNA (rDNA, Lambowitz and Belfort 1993; Saldanha et al. 1993; Yamada et al. 1994). The distinct secondary structure (Burke et al. 1987; Cech et al. 1994) and the capacity of some group-I introns to catalyze their own excision from coding regions, via the ribozyme, characterize this family of mobile elements (Burke 1988; Cech 1990). In spite of the sporadic distribution of many nuclear-encoded rDNA group-I introns lacking endonuclease-encoding open reading frames, some of these sequences appear to have a protracted stability in the genome after the lateral transfer event (De Jonckheere and Brown 1994; Vader et al. 1994). The capacity of group-I introns to efficiently catalyze their own excision from rRNAs, resulting in no apparent phenotype, may explain their stable inheritance (Dujon 1989). Non-mobile group-I introns provide the opportunity to study the primary and secondary structure evolution of these sequences. An apparent example of group-I intron stability is found within the Zygnematales (green algae, Charophyta).

The zygnematalean intron, which is inserted after position 1506 of the nuclear-encoded SSU rDNA (position relative to the coding region of *Escherichia coli*), is found in all members of this lineage except *Spirogyra* spp. (supporting a single loss of the intron in this taxon) and is proposed to be 350–400 million years old on the basis of the first appearance of a member of the Zygnematales in the fossil record (Bhattacharya et al. 1994). Relatively high sequence conservation among the zygnematalean group-I introns (1506 introns) allows the unambiguous alignment of most of this sequence and the analysis of both primary and secondary structure evolution. Using the zygnematalean 1506 group-I introns as a model system, the objectives of the present study were (1) to compare the primary and secondary structure conservation of the zygnematalean 1506 introns to existing genetic data regarding sequence constraints in group-I intron evolution, and (2) to determine if sequence evolution is coupled in group-I introns and the SSU rDNAs containing these sequences.

Materials and methods

Total nucleic acids were isolated from an exponentially growing uni-algal culture of *Gonatozygon aculeatum* (Mix 1980) as in Surek et al. (1994), and the SSU rDNA and group-I intron sequences were determined as in Bhattacharya et al. (1994). The rDNA sequence of *G. aculeatum* will appear in the EBI/Genbank database under the following accession number, X91346. Group-I introns and coding regions from ten Zygnematales were manually aligned; these alignments are available from the authors. Secondary structures were constructed for all zygnematalean group-I introns and a consensus secondary structure was prepared showing intron positions that were identical or conserved in 90% and 80% of the sequences. Only conserved, alignable positions (232 nucleotides, nt) were used in the construction of the consensus secondary structure. The individual zygnematalean 1506 group-I intron secondary structures, drawn according to the conventions of Cech et al. (1994), are available in PICT format from D.B. and in PostScript format from the World Wide Web (URL:http://pundit.colorado.edu:8080/RNA/GRPI/introns.html). To compare sequence-divergence rates in zygnematalean SSU rDNAs and the 1506 group-I introns, the relative-rate test (Sarich and Wilson 1973) was applied to pairwise evolutionary distances estimated according to Jukes and Cantor (1969). The relative-rate test facilitates the comparison of sequence divergence within two taxa since they last shared a common ancestor (point 0 in Table 1) by using as reference a third taxon that has branched prior to the test taxa (i.e. serves as an outgroup, Li and Graur 1991). The method of Kimura and Ohta (1972) was used to calculate the variance for each pairwise distance estimate and these values were used to calculate the test statistic for each rate comparison according to Li and Tanimura (1987) with *Chlorella mirabilis* as the outgroup in all relative-rate measurements. A total of 1723 aligned nt could be used in the evolutionary analyses of the coding regions whereas a total of 232 nt (the same alignment used for the secondary structure analysis) were used in the group-I intron analyses. The base compositions of the 1506 group-I introns and the SSU rDNAs were also compared to further characterize these sequences.

Results and discussion

Secondary structure evolution

The pattern of sequence conservation within the summary secondary structure of the 1506 group-I introns from the Zygnematales shown in Fig. 1 A agrees with existing models. The P, Q, R and S regions (i.e. P4, P7, J8/7, respectively) that form the catalytic core of group-I introns (Cech 1988; Michel and Westhof 1990) are highly conserved in the zygnematalean introns as is the G-C pair in stem P7 that is involved in guanosine (G)-binding (Michel et al. 1989). The internal guide sequence (IGS, Davies et al. 1982; Hall et al. 1987; Burke 1988) at the 5′ terminal region of the zygnematalean group-I introns (i.e. P1) is also well supported. The positioning of the 5′ splice site is conserved, this being either the fourth or fifth paired nucleotide distal from the base of P1. There is also good support for the existence of element P10 (Davies et al. 1982) which is hypothesized to pair a portion of the intron near the 5′ end with the 3′ splice site and so promote the splicing reaction.

Secondary structural elements that are peripheral to the core regions are less well conserved in the group-I introns from the Zygnematales in agreement with genetic studies (i.e. processive deletions) of other group-I intron sequences (reviewed in Burke 1988). The distal region of the P6 element, for example, is the site of the largest insertion in the zygnematalean 1506 group-I introns, a unique 147-nt extension that is found after the first loop of P6 in *Mougeotia scalaris*. High sequence conservation of the proximal stem of P2 (position ① in Fig. 1 A) in the 1506 introns supports a function for this region in group-I intron splicing. The sequence, $^{5'}$C-G,U-A,A,G-C,U·G,G$^{3'}$, that forms the stems and linker region of P2–P2.1 (J2/2.1), is identical in all zygnematalean 1506 introns and is a conserved feature of many other subgroup-IC1 introns (see alignment in Bhattacharya et al. 1994).

Comparisons of secondary structures of the zygnematalean group-I introns also show high conservation of domain P5abc (region ② in Fig. 1 A). This extension of P5 is characteristic of many group-I introns (Collins 1988; Michel and Westhof 1990). Conserved sequence motifs in group-I introns containing an extension of P5 include an A-rich bulge in P5a and a GAAA tetraloop in L5b (Michel and Cummings 1985; Collins 1988). The A-rich bulge has a crucial role in subdomain structure in the *T. thermophila* LSU intron with the second invariant adenosine (A186 in *T. thermophila*) being the most important (Murphy and Cech 1994). Murphy and Cech (1994) showed a long-range interaction between the fourth position (A153) within the GAAA tetraloop and the last base-paired nucleotides preceding J6a/6b (C223–G250) in the *T. thermophila* intron. This interaction results in a compact structure of the *T. thermophila* group-I intron that could be directly visualized with the electron microscope (Wang et al. 1994). Analyses of the zygnematalean 1506 group-I introns provide support for the interaction of the GAAA tetraloop and the C-G base pair preceding the J6a/6b loop. The group-I introns from *Mougeotia caldariorum*, *M. scalaris*, *Zygnemopsis circumcarinata*, *Genicularia aculeatum* and *Genicularia spirotaenia* (see Bhattacharya et al. 1994) all contain these sequence motifs (as do the *Chlorella* spp. 1506 group-I introns; see Fig. 1 B). The later-diverging members of the Desmidiaceae, *Cosmarium botrytis*, *Cosmocladium saxonicum*, *Sphaerozosma granulatum* and *Staurastrum* spp., however, all lack the GAAA tetraloop and have highly divergent L5b sequences, these taxa contain a C-G base pair in the region throught to interact with L5b (region ③ in Fig. 1 A).

The group-I introns from the later-diverging Desmidiaceae are also characterized by the loss of a secondary structure element (Fig. 1 C) which is found in the other zygnematalean introns. High sequence conservation allows the unambiguous identification of element "C"; this sequence is also present in the 1506 group-I introns of *Chlorella* spp., the red algae *Porphyra* spp. (Oliveira and Ragan 1994) and *Hildenbrandia rubra* (Ragan et al. 1993), and the fungus *Pneumocystis carinii* (Sogin and Edman 1989). If the phylogenetic framework presented in Bhattacharya et al. (1994) reflects the true evolutionary history of the 1506 introns, then the loss/disruption of element "C" may be a unique event in the Zygnematales lineage that marks the derived Desmidiaceae. Concomitant with the loss of the

Fig. 1 A–C Secondary structure analysis of zygnematalean group-I introns. **A** putative consensus secondary structure of the 1506 group-I introns within the SSU rRNA precursor of ten zygnematalean taxa drawn according to the conventions of Cech et al. (1994). The 5′ and 3′ splice junctions are shown with *arrows*, as are the locations of the pairing segments P1–P10, the internal guide sequence (IGS), the A-rich bulge, the L5b tetraloop and the putative guanosine(G)-binding site. *Circled regions* ① to ③ are as discussed in the text. Positions that are identical in all zygnematalean 1506 group-I introns are inside *open boxes* whereas positions that are conserved in 90% of the introns are in the *dotted boxes* (the nucleotide found in 9/10 introns is also shown) and positions conserved in 80% of the introns are shown as *hatched boxes*. *Broken lines* were used to position secondary structure elements that are believed to interact in close proximity, whereas the *solid lines* indicate non-conserved intron sequences. **B** the tetraloop in L5B that has been lost from the later-diverging Desmidiaceae. **C** the optional helix "C" that has been lost from the later-diverging Desmidiaceae
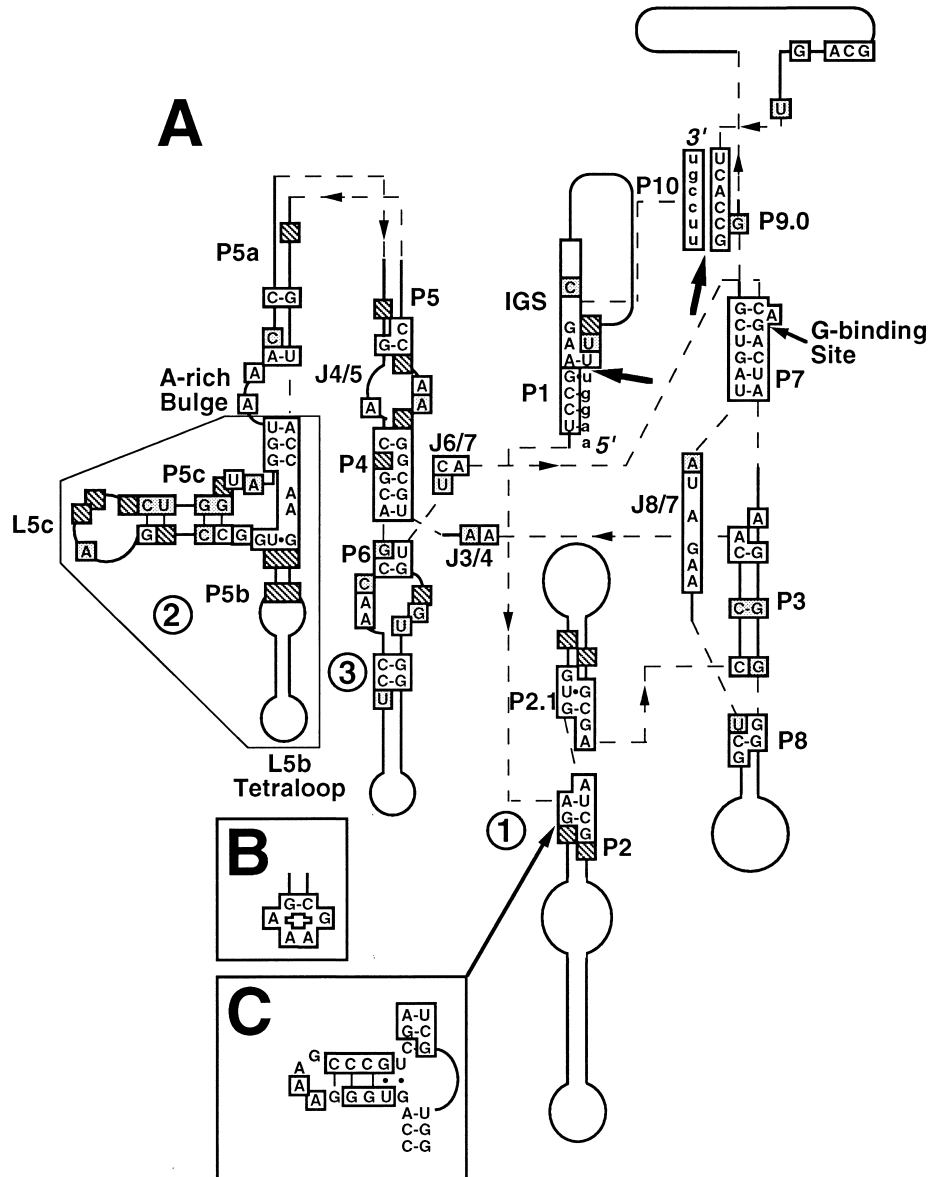
optional element in the later-diverging Desmidiaceae is an acceleration in the rate of sequence divergence (see Table 1). The relative rates of sequence divergence of the group-I introns from *Staurastrum* spp. and *C. botrytis* are, for example, significantly greater than these rates in taxa that have not lost element C (e.g. *G. spirotaenia*). The connection between the coincidental loss of element C in P2, alteration of the GAAA tetraloop sequence in L5b and sequence-divergence rates is not known, these events may simply coincide with a period of rapid radiation of the group-I introns in the later-diverging Desmidiaceae. Alternatively, it is possible that the loss of element C in P2 and/or the GAAA tetraloop in L5b led to the disruption of the tertiary structure of the later-diverging Desmidiaceae group-I introns that resulted in a high selection pressure for another solution to the tertiary folding requirements of these introns.

Are zygnematalean SSU rDNA coding regions and group-I introns coevolving?

Rates of sequence divergence. Sequence divergence rates (per 100 nt) within SSU rDNAs and group-I introns vary widely in the Zygnematales (Table 1). Taxa such as *G. spirotaenia* and *Z. circumcarinata* contain introns that have diverged at nearly identical rates ($K_{OA}/K_{OB} = 1.1$) since their last common ancestor whereas their coding regions have diverged at a relative rate that is 3.1-times greater in *G. spirotaenia* than in *Z. circumcarinata.* Comparison of the rates of divergence of the rDNA and group-I introns in *G. spirotaenia* and *C. botrytis* show that the coding region of *G. spirotaenia* is evolving at a relative rate that is 2.5-times greater than in *C. botrytis*, whereas the group-I intron in *G. spirotaenia* is evolving at a relative rate that is approximately one-quarter that in *C. botrytis*. The esti-

Table 1 Comparisons of rates of sequence divergence per 100 sites in zygnematalean 1506 group-I introns and SSU rDNA coding regions. The relative-rate test was applied to zygnematalean sequences using *Chlorella mirabilis* (C) as the outgroup (Li and Tanimura 1987). The method of Jukes and Cantor (1969) was used to estimate the evolutionary distances and the variances were computed according to Kimura and Ohta (1972). A total of 1723 nt were compared in the coding regions and a total of 232 nt were compared in the analyses of the group-I intron sequences. Relative-rate differences that are significant at the 5% level are shown with an asterisk (*)

Species comparisons		Group-I intron			SSU rRNA		
A	B	K_{AB}	K_{AC-BC}	K_{OA}/K_{OB}	K_{AB}	K_{AC-BC}	K_{OA}/K_{OB}
Staurastrum M752	+ *G. spirotaenia*	38.9 ± 5.2	$13.3 \pm 5.2*$	2.0	6.0 ± 0.6	$-2.2 \pm 0.2*$	0.5
Staurastrum M752	+ *M. caldariorum*	36.8 ± 5.0	$19.0 \pm 5.2*$	3.1	3.9 ± 0.5	0.3 ± 0.5	1.2
C. botrytis	+ *G. spirotaenia*	42.0 ± 5.5	$25.8 \pm 5.7*$	4.2	5.7 ± 0.6	$-2.8 \pm 0.6*$	0.4
G. spirotaenia	+ *M. caldariorum*	28.9 ± 4.3	5.7 ± 4.3	1.5	6.6 ± 0.6	$2.5 \pm 0.6*$	2.2
G. spirotaenia	+ *Z. circumcarinata*	33.7 ± 4.7	1.8 ± 4.7	1.1	6.2 ± 0.6	$3.2 \pm 0.6*$	3.1
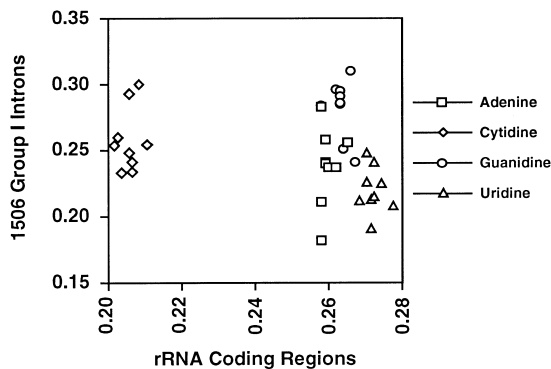
Fig. 2 Analysis of base compositions in zygnematalean SSU rDNAs and 1506 group-I introns

mates presented in Table 1 show that no studied member of the Zygnematales has approximately equal relative rates of divergence within its SSU rDNA and group-I intron; the group-I introns always have a higher rate of sequence divergence than do the rDNA coding regions.

Nucleotide composition. To determine if group-I intron and SSU rDNA coding-region base compositions are similar, the frequences of the four nucleotides were compared pairwise within complete intron and rDNA sequences. This analysis shows that base compositions are widely discordant within group-I introns and SSU rDNAs (Fig. 2). Whereas the zygnematalean SSU rDNAs are relatively poor in cytidine (C)-content and rich in uridine (U)-content, the group-I introns within these coding regions are relatively C-rich and U-poor. Further, the adenine(A)-content varies from 25–27% in the rDNAs while in the group-I introns this value ranges from 18–28%. In general, the group-I introns have much greater variation in their base compositions whereas the coding regions are more constrained.

Conclusions

Analyses of the zygnematalean group-I intron secondary structures support existing views regarding regions vital for RNA-splicing. In addition, the identification of new regions of primary and secondary structure conservation within the zygnematalean 1506 introns (e.g. P5abc, J2/2.1) suggests that these structures may hold clues to the in vivo splicing of these introns that have not yet been addressed in existing in vitro self-splicing assays. Alternatively, these patterns of conservation may reflect 350 million years of evolution within the Zygnematalean cell that is not readily applicable to all group-I introns, in particular regarding the highly divergent later-diverging Desmidiaceae. The naturally variable subset of group-I introns found in the Zygnematales provides a tool to use in the biochemical analyses of structure/function relationships that may provide answers to these questions.

Primary and secondary structure analyses of the zygnematalean 1506 introns also provided insights into the dynamics of group-I intron evolution. The 1506 introns have diverged greatly within the zygnematalean rDNAs since the origin of this evolutionarily distinct lineage (Surek et al. 1994) such that this intron family is now characterized by members with relatively widely varying sequence evolution rates, base compositions and secondary structures. The data presented here suggest that sequence evolution of the zygnematalean 1506 group-I introns is not tightly linked to the evolution of the SSU rDNAs.

References

Bhattacharya D, Surek B, Rüsing M, Damberger S, Melkonian M (1994) Group-I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). Proc Natl Acad Sci USA 91:9916–9920
Burke JM (1988) Molecular genetics of group-I introns: RNA structures and protein factors required for splicing – a review. Gene 73:273–294
Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW, Tabak HF (1987) Structural conventions for group-I introns. Nucleic Acids Res 15:7217–7221
Cech TR (1988) Conserved sequences and structures of group-I introns: building an active site for RNA catalysis – a review. Gene 73:259–271

Cech TR (1990) Self-splicing of group-I introns. Annu Rev Biochem 59:543–568

Cech TR, Damberger SH, Gutell RR (1994) Representation of the secondary and tertiary structure of group-I introns. Nature Struc Biol 1:273–280

Collins RA (1988) Evidence of natural selection to maintain a functional domain outside of the "core" in a large subclass of group-I introns. Nucleic Acids Res 16:2705–2715

Davies RW, Waring RB, Ray JA, Brown TA, Scazzocchio C (1982) Making ends meet: a model for RNA splicing in fungal mitochondria. Nature 300:719–724

De Jonckheere JF, Brown S (1994) Loss of the ORF in the SSU rDNA group-I intron of one *Naegleria* lineage. Nucleic Acids Res 22:3925–3927

Dujon B (1989) Group-I introns as mobile genetic elements: facts and mechanistic speculations – a review. Gene 82:91–114

Hall DH, Povinelli CM, Ehrenman K, Pedersen-Lane J, Chu F, Belfort M (1987) Two domains for splicing in the intron of the phage T4 *td* gene established by non-directed mutagenesis. Cell 48:63–71

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munoro HN (ed) Manual of protein metabolism. Academic Press, New York

Kimura M, Ohta T (1972) On the stochastic model for estimation of mutational distance between homologous proteins. J Mol Evol 2:87–90

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Michel F, Cummings DJ (1985) Analysis of class-I introns in a mitochondrial plasmid associated with senescence of *Podospora anserina* reveals extraordinary resemblance to the *Tetrahymena* ribosomal intron. Curr Genet 10:69–79

Michel F, Westhof E (1990) Modelling the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610

Michel F, Hanna M, Green R, Bartel DP, Szostak JW (1989) The guanosine binding site of the *Tetrahymena* ribozyme. Nature 342:391–395

Mix M (1980) Liste der Sammlung von Conjugaten-Kulturen im Institut für Allgemeine Botanik der Universität Hamburg Mitt Staatsinst. Allg Bot Hamburg 14:135–169

Murphy FL, Cech TR (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group-I intron domain. J Mol Biol 236:49–63

Oliveira MC, Ragan MA (1994) Variant forms of a self-splicing group-I intron in nuclear small-subunit rRNA genes of the marine red alga *Porphyra spiralis* var. *amplifolia*. Mol Biol Evol 11:195–207

Ragan MA, Bird CJ, Rice EL, Singh RK (1993) The nuclear 18 *s* ribosomal RNA gene of the red alga *Hildenbrandia rubra* contains a group-I intron. Nucleic Acids Res 211:3898

Saldanha R, Mohr G, Belfort M, Lambowitz AM (1993) Group-I and group-II introns. FASEB J 7:15–24

Sarich VM, Wilson AC (1973) Generation time and genomic evolution in primates. Science 179:1144–1147

Sogin ML, Edman JC (1989) A self-splicing intron in the small subunit rRNA gene of *Pneumocystis carinii*. Nucleic Acids Res 17:5349–5359

Surek B, Beemelmanns U, Melkonian M, Bhattacharya D (1994) Ribosomal RNA sequence comparisons demonstrate an evolutionary relationship between Zygnematales and Charophytes. Syst Evol 191:171–181

Vader A, Naess J, Haugli K, Haugli F, Johansen S (1994) Nuclear introns from *Physarum polycephalum* contain insertion elements that may explain how mobile group-I introns gained their open reading frames. Nucleic Acids Res 22:4553–4559

Wang Y-H, Murphy FL, Cech TR, Griffith JD (1994) Visualization of a tertiary structural domain of the *Tetrahymena* group-I intron by electron microscopy. J Mol Biol 236:64–71

Yamada T, Tamura K, Aimi T, Songsri P (1994) Self-splicing group-I introns in eukaryotic viruses. Nucleic Acids Res 22: 2532–2537