Curr Genet (2001) 40: 82–90
DOI 10.1007/s002940100227

RESEARCH ARTICLE

Debashish Bhattacharya · Jamie J. Cannone
Robin R. Gutell

Group I intron lateral transfer between red and brown algal ribosomal RNA

Received: 12 February 2001 / Accepted: 26 April 2001 / Published online: 17 July 2001

Abstract How group I introns originate in nuclear ribosomal (r)RNA genes is an important question in evolutionary biology. Central to this issue is the multitude of group I introns present in evolutionarily distantly related plant, fungal, and protist lineages, together with an understanding of their origin and lateral transfer from one exon to another, between cell organelles, and between cells. These introns vary considerably in primary and secondary structure; and their provenance from a few or perhaps many mobile elements that have spread in rRNAs is unknown. Here we show that a novel lineage of group IC1 introns inserted at position 516 (*Escherichia coli* gene numbering) in the small subunit rRNA in bangiophyte red algae and a brown alga (*Aureoumbra lagunensis*) are specifically related, although their host cells are not. These bangiophyte and *Aureoumbra* introns are the only known cases that have a helical insertion in the P5b helix. The highly conserved primary and secondary structure of the extra P5b helix suggests that it is important, although its specific function is unknown. Our study attempts to understand the origin and movement of these IC1 introns.

Keywords Group I intron · Stramenopiles · Lateral transfer · Phylogeny

Communicated by R.W. Lee

D. Bhattacharya (✉)
Department of Biological Sciences, University of Iowa,
239 Biology Building, Iowa City, IA 52242-1324, USA
E-mail: dbhattac@blue.weeg.uiowa.edu

J.J. Cannone · R.R. Gutell
Institute for Cellular and Molecular Biology,
University of Texas, 2500 Speedway, Austin,
TX 78712-1095, USA

Introduction

Group I introns are a distinct class of RNA enzymes that are characterized by a conserved RNA primary and secondary structure essential for splicing and are often capable of "self-splicing" (Kruger et al. 1982; Cech 1985). Clarifying the origin and evolutionary history of group I introns helps us understand how these sequences have persisted for billions of years and, more generally, addresses the possible role(s) of catalytic RNAs in modern cells. The sporadic and wide distribution of group I introns in the nuclear-encoded small (SSU) and large (LSU) subunit ribosomal (r)RNA and organellar genes of green algae, red algae, fungi, ciliates, amoebae, and other organisms (Cannone et al., unpublished data) suggests that these sequences are highly successful at invading and maintaining themselves in eukaryotic genomes (Bhattacharya et al. 1994, 1996a; Turmel et al. 1995; Bhattacharya 1998; Nishida et al. 1998; Watanabe et al. 1998; Suh et al. 1999; Friedl et al. 2000).

In this study, we use secondary structure and phylogenetic analyses to address the lateral transfer of group I introns in nuclear rRNA. To approach this issue, unique structural features of introns are tracked through evolutionary time. We also use traditional phylogenetic methods to determine the evolutionary relationships of these introns and their host cells. Currently, there are approximately 1,214 group I introns available in GenBank. Our analyses of these sequences [summarized at the Comparative RNA web site (CRW), http://www.rna.icmb.utexas.edu/] resulted in their classification into five major structural subgroups [Michel and Westhof 1990 (subgroups A–D); Suh et al. 1999 (subgroup E)]. Within the order Bangiales in the Rhodophyta (red algae), IC1 introns at position 516 revealed a unique insertion in the P5b helix, which is absent from all other IC1 introns (Gutell and Cannone, unpublished data). Subsequently, we found a group I intron in a brown algal stramenopile, *Aureoumbra lagunenesis*, with the same unusual insertion that, coincidentally, also

occurred at position 516 in small subunit rRNA. This collection of "tagged" 516 introns and the sequences for the SSU rRNA in the host cells have presented us with the opportunity to explore the lateral transfer of a group I intron. There are two major scenarios that could explain this situation:

1. Introns early. The 516 intron was present in the SSU rRNA of the common ancestor of red and brown algae and has been maintained only in the Bangiales and *Aureoumbra*. This scenario would be supported by a close evolutionary relationship between red and brown algae, whereas a distant relationship would support independent acquisition of the intron.
2. Introns late. The presence of the 516 intron in the Bangiales and *Aureoumbra* is best explained by independent acquisition of the intron after the separation of red and brown algae. We used phylogenetic analyses of introns and SSU rRNA coding regions to address the origin of these 516 introns.

Materials and methods

Intron and rRNA sequences

The publicly available sequences analyzed here were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/Genbank/index.html), at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). The rRNA and rRNA introns were usually identified by searching for the words "rRNA" and "introns", and for "rRNA" in the feature key field in the Entrez web-searching utility (http://www.ncbi.nlm.nih.gov/Entrez/). However, we have utilized the BLAST utility (http://www.ncbi.nlm.nih.gov/BLAST/) to identify some intron sequences that are similar to the red algal rRNA intron sequences, but are not always annotated properly in their GenBank entry. Although the boundaries for the rRNA exons and introns are usually annotated properly, we have independently determined the intron/exon borders.

Alignment and secondary structure models and diagrams

Secondary structure diagrams for the red and brown algal group I introns were modeled from the current comparative structure models for the IC1 introns (Michel and Westhof 1990; Damberger and Gutell 1994; Gutell and Cannone, unpublished data). These structural models are derived by comparative sequence analysis, a method based on the simple premise that RNAs within the same family (e.g., tRNA, group I intron) that have similar functions will have similar secondary and tertiary structures (reviewed in Gutell 1996). The red and brown algal intron sequences were aligned with the IC1 sequences, because these were the most similar in sequence and structure. All sequences were manually aligned using the SUN Microsystems UNIX-based alignment editor AE2 (T. Macke, Scripps Research Institute, La Jolla, Calif.) to maximize sequence and structure identity, to maintain (when possible) previously proposed base-pairings, and to minimize the number of insertion and deletion events. Once the secondary structure model had been determined, the computer version of the diagram was drawn with the interactive SUN Microsystems UNIX-based, interactive graphics program XRNA (B. Weiser and H. Noller, University of California at Santa Cruz, Calif.).

Phylogenetic analyses

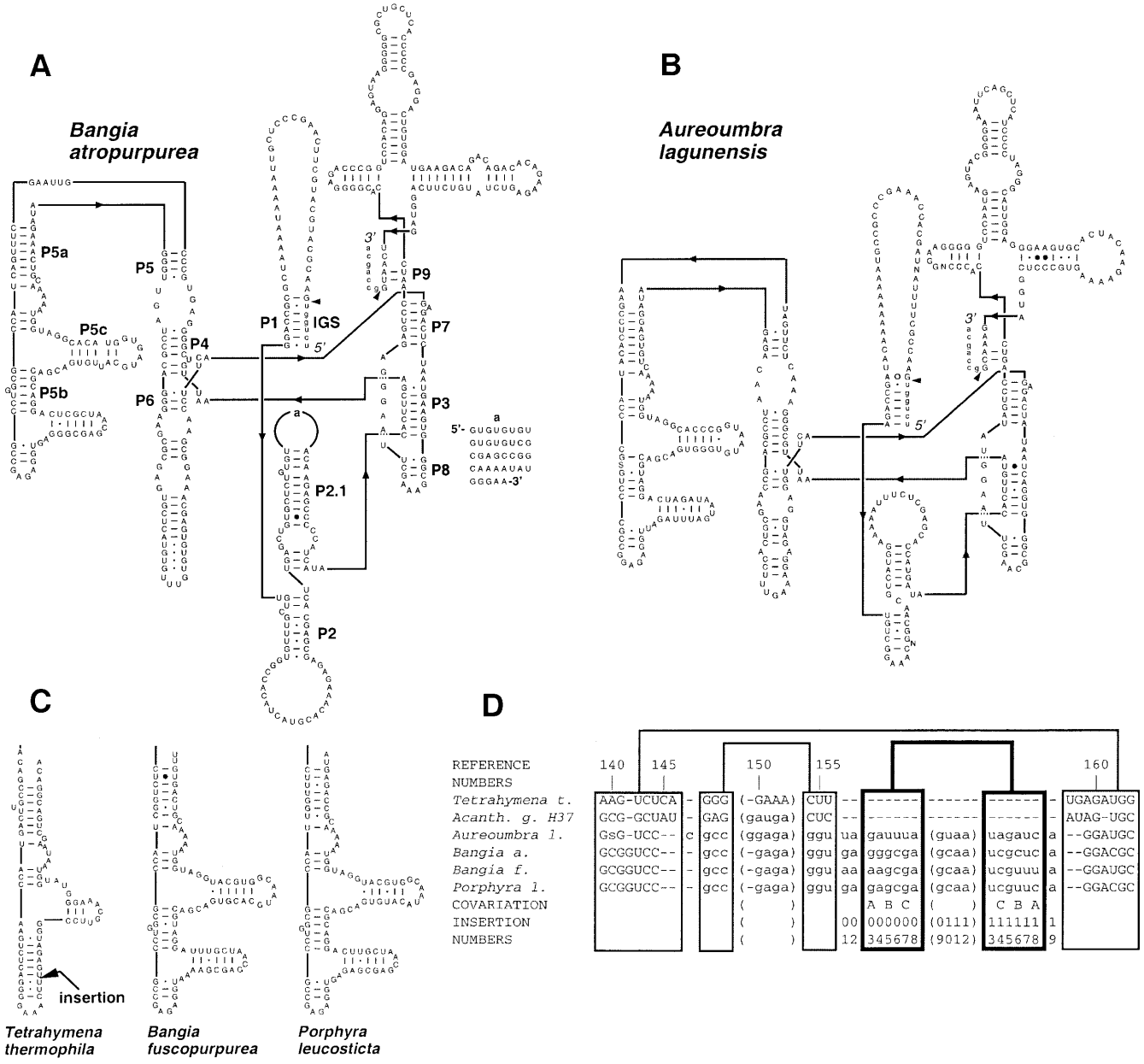
For the group I intron data, a set of 62 group IC1 and IE intron sequences from nuclear (16S and 16S-like) SSU rRNA in green algae, red algae, fungi, *Acanthamoeba griffini*, and *Naegleria* spp. were manually aligned, using primary and secondary structure similarity to juxtapose homologous regions (e.g., Michel and Westhof 1990; Bhattacharya et al. 1994; Damberger and Gutell 1994). An alignment with a total length of 4,523 characters, including insertions in the different introns, was submitted to a pair-wise distance analysis, using the HKY-85 model (Hasegawa et al. 1985) with equal rates across sites and the transition/transversion ratio = 2. Missing data and gaps were excluded from each pair-wise comparison. This distance matrix was used as input for neighbor-joining tree-building. A total of 1,000 bootstrap replicates (Felsenstein 1985) were analyzed with the distance method. We also used the Modeltest program (ver. 3.04; Posada and Crandall 1998) to survey 56 possible models of DNA substitution, to identify the model that best fit the intron data. This analysis showed the general time-reversible model (Rodríguez et al. 1990) with estimations of nucleotide frequencies (A = 0.244, C = 0.251, G = 0.271, T = 0.234), the shape parameter of the gamma distribution ($a = 0.764$) to accommodate rate variations across sites, and the proportion of invariant sites (0.016) that are unable to accept substitutions (GTR + G + I) as best fitting the intron data. The GTR + G + I model was used to calculate bootstrap values (1,000 replicates) for monophyletic groups identified by the HKY-85 distance tree. In this way, we had estimates of support for different groups in the intron tree, using evolutionary models that were either relatively simple (HKY-85), or parameter-rich (GTR + G + I). The IE introns were used to root the intron tree.

The complete SSU rRNA sequences from a broadly diverse set of red algae and stramenopiles were also aligned, using conserved secondary structures, and 1,581 nt that could be simultaneously aligned in all the sequences were submitted to a distance analysis using the HKY-85 model (as described above). Modeltest also identified the GTR + G + I model as best fitting the rRNA data; and so it was used in the distance bootstrap analysis (1,000 replicates). The parameter value estimates for this data set were: A = 0.251, C = 0.206, G = 0.281, T = 0.262, proportion of invariant sites = 0.359, $a = 0.548$). These data were also analyzed with an unweighted maximum parsimony method. Initial trees were built stepwise with ten random sequence additions; and the trees were rearranged with tree bisection/reconnection, to find the most parsimonious phylogeny. A total of 1,000 bootstrap replicates were analyzed with this method. The stramenopile sequences were used to root the SSU rRNA tree. All phylogenetic analyses were done with PAUP* (ver. 4.0b8; Swofford 2001). The figures presented in this text, along with other secondary structure diagrams, are available at the web site http://www.rna.icmb.utexas.edu/PHYLO/ALGINT and are referred to in this text as the CRW Algal Introns.

Results and Discussion

Secondary structure analysis of the 516 introns

An analysis of the group I introns at the online CRW web site revealed that 59 bangiophytes in the monophyletic red algal order Bangiales (*Bangia* and *Porphyra* spp; Müller et al. 1998, 2001a; Oliveira and Bhattacharya 2000) contained a IC1 intron at SSU rRNA position 516 with the following characteristic: they each had an insertion in the P5b helix (see Fig. 1A,C) which is absent from the other IC1 introns, except for an intron in *A. lagunensis* (Fig. 1B, Table 1). The *Aureombra* group I intron was also in subgroup C1, had the P5b insertion, and mapped to the 516 site in SSU rRNA. Detailed analysis of the P5b helix provided strong evidence for its homology in the red and brown algal 516 introns. The insertion was conserved in size (19 nucleotides) and the



tetraloop hairpin sequence at position 156:9–12 (Fig. 1D) showed high identity (5′-GYAA-3′) to one of the canonical hairpin tetraloop sequences initially identified in 16S rRNA (Woese et al. 1990). All of the base pair positions have some underlying variation, and they all contain G:C, A:U, or G:U base pairs in all of the sequences with this insertion. The three base pairs with the purest covariance are 3:18, 7:14, and 8:13 (see dark boxes in Fig. 1D, Table 2). Taken together, these observations were consistent with a single origin of the P5b insertion and, by extension, of the 516 introns containing the extra helix. The pattern of sequence conservation of the P5b helix also suggested that it is involved in a RNA–RNA tertiary structure interaction within the P5abc subdomain (Cate et al. 1996; Doherty et al. 1999; Doherty and Doudna 2000). Tetraloops often dock into stem-loop structures in other regions of RNA. An example of such an interaction is between domains 5 and 1 of group II introns (Abramovitz and Pyle 1997). The conserved nature of such higher-order interactions also generally requires that the helices capped by tetraloops be of a fixed length (Hedenstierna et al. 2000), as found in the P5b insertions (see Fig. 1D). Regardless of its specific role, upon establishment, the helix has apparently been under strong selective constraint and has not been lost in any of the 31 publicly available introns at position 516 in the SSU rRNA that have been reported from the Bangiales or from *Aureoumbra*.

To determine whether other publicly available group I introns, none of which apparently contain the P5b insertion described here, may share a close evolutionary relationship with the Bangiales and *Aureoumbra*

◀

Fig. 1A–D Analysis of intron secondary structures. **A** Comparative secondary structure model for the group IC1 intron at position 516 in the small subunit (SSU) rRNA of *Bangia atropurpurea* (GenBank D88387), drawn according to the conventions of Cech et al. (1994). The 5′ and 3′ splice junctions are marked with *arrows*, as are the locations of the pairing segments P1–P9 and the internal guide sequence (*IGS*). A portion of P2.1 marked *a* is shown as the primary structure and has not been folded in this model. **B** The secondary structure model of the *Aureoumbra lagunensis* intron (U40258). **C** Comparison of the P5a–c domains in the 516 introns in *B. fuscopurpurea* (AF342745), *Porphyra leucosticta* (AF342746), and *Tetrahymena thermophila* (J01235). **D** Alignment of the P5b region, including the insertion. Base-pairings in helices are identified in the *boxed regions*; and the insertion is identified by the *bold boxes*. The *T. thermophila* large subunit rRNA group I intron is used as the reference for these comparisons. These secondary structure figures, including the *T. thermophila*, *Acanthamoeba* spp, and all the red and brown algal intron diagrams we have determined for sequences that are publicly available, can be accessed at our comparative RNA web site (http://www.rna.icmb.utexas.edu/) and at http://www.rna.icmb.utexas.edu/PHYLO/ALGINT/

516 introns, we used the BLAST program (BLASTN ver. 2.1.2; Altschul et al. 1997) to search for other intron sequences with significant sequence identity with the *Aureoumbra* intron. This analysis identified introns in two strains of the lobose amoeba, *Acanthamoeba griffini* (GenBank S81337, U02540), with a 95% identity over a 48-bp sequence. Intrigued by this result, both *Acanthamoeba* intron sequences (strains ATCC30731, H37) were included in our intron alignment; and secondary structure models were developed. The *Acanthamoeba* introns also interrupt the SSU rRNA at the 516 site, although they each lack the insertion in the P5b helix (CRW Algal Introns). The *Acanthamoeba* introns, however, share many secondary structure features in common with IC1 introns, such as the nine conserved helical domains shown in Fig. 1A, and can conclusively be classified in this group (for details, see Michel and Westhof 1990; Suh et al. 1999).

To put the finding of the P5b insertion in perspective, of the approximately 1,214 publicly available group I introns (350 of which are in subgroup IC1), approximately 772 of them occur in SSU and LSU rRNA; and nearly 600 of these introns interrupt 35 different SSU rRNA sites. The SSU rRNA sites with the largest number of introns are 1506 (123 introns), 516 (92 introns), and 943 (90 introns). Approximately one-half of the 516 introns are in subgroup IC1, whereas the other half are in the IE subgroup. In the entire group I intron data set, only the Bangiales and *Aureoumbra* IC1 516 introns contain the P5b insertion. Thus, this insertion associates these two sets of introns and suggests that they are specifically related to one another. To resolve the evolutionary relationships of these introns, we re-

Table 1 List of organisms containing the inserted helix in the P5b region of the IC1 516 introns. GenBank accession numbers are listed at http://www.ncbi.nlm.nih.gov/

Organism	Strain/isolate	GenBank accession number
Aureoumbra lagunensis		U40258
Bangia atropurpurea		D88387
B. fuscopurpurea		AF342745
Porphyra carolinensis		AF133792
P. dentata	Isolate Koga Fukuoka	AB013183
P. haitanensis		AB015795
P. katadae	Isolate Kawatana Yamaguchi	AB013184
P. leucosticta	(Baltic)	AF342746
P. lilliputiana		AF136424
P. onoi		AB015794
P. pseudolinearis	Isolate Tohaku Tottori	AB013185
Porphyra sp.	Isolate GRB108	AF136420
Porphyra sp.	Isolate RAK049	AF136425
Porphyra sp.	Isolate Shimonoseki Yamaguchi	AB013182
Porphyra sp.	Strain BRU107	AF136418, AF136419
P. suborbiculata		AB015796
P. suborbiculata	Isolate Kawatana Yamaguchi	AB013180
P. tenera	Isolate Kawaura Kumamoto	AB013176
P. tenera	Isolate Matsukawa Fukushima; strain Matsukawa-3	AB029883
P. tenera	Isolate Matsukawa Fukushima; strain Matsukawa-4	AB029884
P. tenera	Isolate Shinwa Kumamoto	AB013175
P. tenera	Isolate Shinwa Kumamoto; strain Kawaura-4	AB029881
P. tenera	Isolate Shinwa Kumamoto; strain Shinwa-2	AB029879
P. tenera	Isolate Shinwa Kumamoto; strain Shinwa-3	AB029880
P. tenera	Strain T-1	D86237
P. tenera	Strain TU-3	D86236
P. tenera	Strain T8	AB000964
P. tenuipedalis		AB015797
P. yezoensis		D79976
P. yezoensis	Isolate Hakodate Hokkaido	AB013177
P. yezoensis	Isolate Ogatsu Miyagi	AB013178

Table 2 Base-pair frequencies for the inserted helix in P5b found in the red and brown algae. Data are presented for the 31 publicly available sequences containing this inserted helix (30 red algae, 1 brown alga, see Table 1). Base-pairs refer to the "insertion numbers" shown in Fig. 1D. For each base-pair, – represents a combination which is not present at that base-pair. In the *Notes*, A, B, and C refer to the marked base-pairs in Fig. 1D

Base-pair	G:C	C:G	A:U	U:A	G:U	U:G	Notes
03:18	96.8%	–	3.2%	–	–	–	A
04:17	16.1%	–	71.0%	–	12.9%	–	
05:16	38.7%	–	6.5%	3.2%	51.6%	–	B
06:15	–	96.8%	–	–	–	3.2%	
07:14	96.8%	–	–	3.2%	–	–	C
08:13	3.2%	–	96.8%	–	–	–	

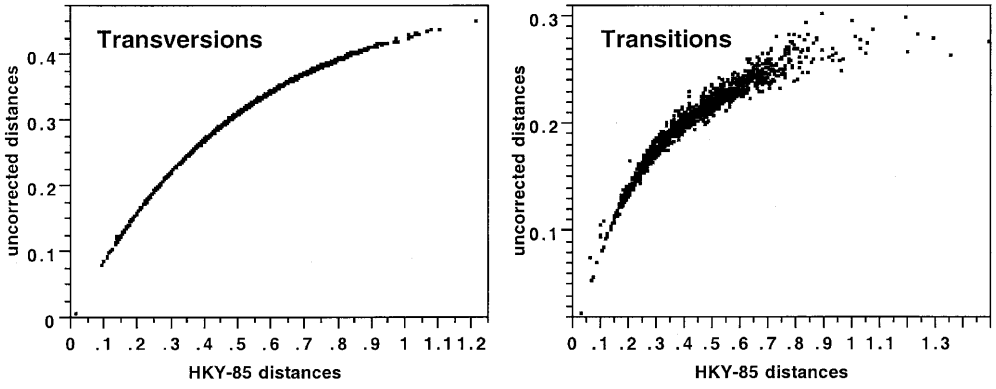


Fig. 2 Plots of uncorrected pair-wise substitutions versus values corrected for multiple hits (HKY-85 model; Hasegawa et al. 1985) for the group I introns included in the phylogeny shown in Fig. 3. Transitions and transversions are studied separately. In this analysis, the bending of the line at higher uncorrected distances indicates mutational saturation (Moritz et al. 1992)

constructed the phylogeny of the 516 introns, other SSU rRNA introns, and the host cells that contain them, using more traditional phylogenetic methods.

Test for mutational saturation of introns

Before embarking on phylogenetic analyses with the group I intron sequences, we wanted to first establish that they encoded a significant phylogenetic signal. To do this, a data set of 62 sequences that spanned the diversity of IC1 and IE intron SSU rRNA insertion sites (see Figs. 2, 3) was used to determine the extent of superimposed substitutions. Uncorrected distances were plotted versus those corrected with the HKY-85 model to detect mutational saturation with regard to transversions and transitions (see Daugbjerg and Andersen 1997; Lopez et al. 1999). In this analysis, the bending of the line at higher, uncorrected distances indicates mutational saturation (Moritz et al. 1992). The plots presented in Fig. 2 suggest that the group I introns encode a significant phylogenetic signal. Transversions show a minor level of mutational saturation, whereas transitions become more rapidly saturated as intron sequence divergence increases. One obviously cannot compare these group I introns to "workhorse" molecular markers such as rRNA coding regions, in terms of utility across a broad phylogenetic spectrum. The introns often have 4–10 times greater rates of sequence divergence (Bhattacharya et al. 1996b). Nonetheless, these sequences contain a phylogenetic signal that can be used to estimate a group I intron evolutionary tree.

Phylogenetic analysis of introns and host cells

To test the monophyly of the Bangiales and *Aureoumbra* 516 introns developed by the comparison of the intron secondary structure features, we first estimated a phylogeny of the IC1 and IE introns from a wide variety of algae and protists. The preliminary distance analysis, which included 322 IC1 and 101 IE introns in nuclear SSU and LSU rRNA demonstrated a specific evolutionary relationship between the *Acanthamoeba*, *Aureoumbra*, and red algal 516 introns within the IC1 cluster (data not shown). This group was distinct from all other IC1 introns and all other publicly available introns at the 516 site, i.e., those in *Naegleria* spp and the IE 516 introns. Knowing this, we reduced the data set to 62 SSU rRNA introns for a more detailed analysis.

The distance tree (HKY-85 model) inferred from the reduced data set of group I introns provides strong support (96% bootstrap value) for the monophyly of the *Acanthamoeba*, Bangiales, and *Aureoumbra* IC1 introns at position 516, the early divergence of the *Acanthamoeba* introns within this clade (86% bootstrap support), and the distant evolutionary relationship of these introns to the 516 introns in the IE group (Fig. 3). The *Naegleria* spp. 516 introns are positioned at the base of the IC1 516 intron clade without bootstrap support. It is, therefore, unclear whether the *Naegleria* introns are monophyletic with the other 516 introns in subgroup IC1. Interestingly, the IC1 subgroup also includes the 1506 SSU rRNA intron in the Bangiales and one member of the derived red algal subclass Florideophycidae (*Hildenbrandia rubra*; Ragan et al. 1993). The Bangiales

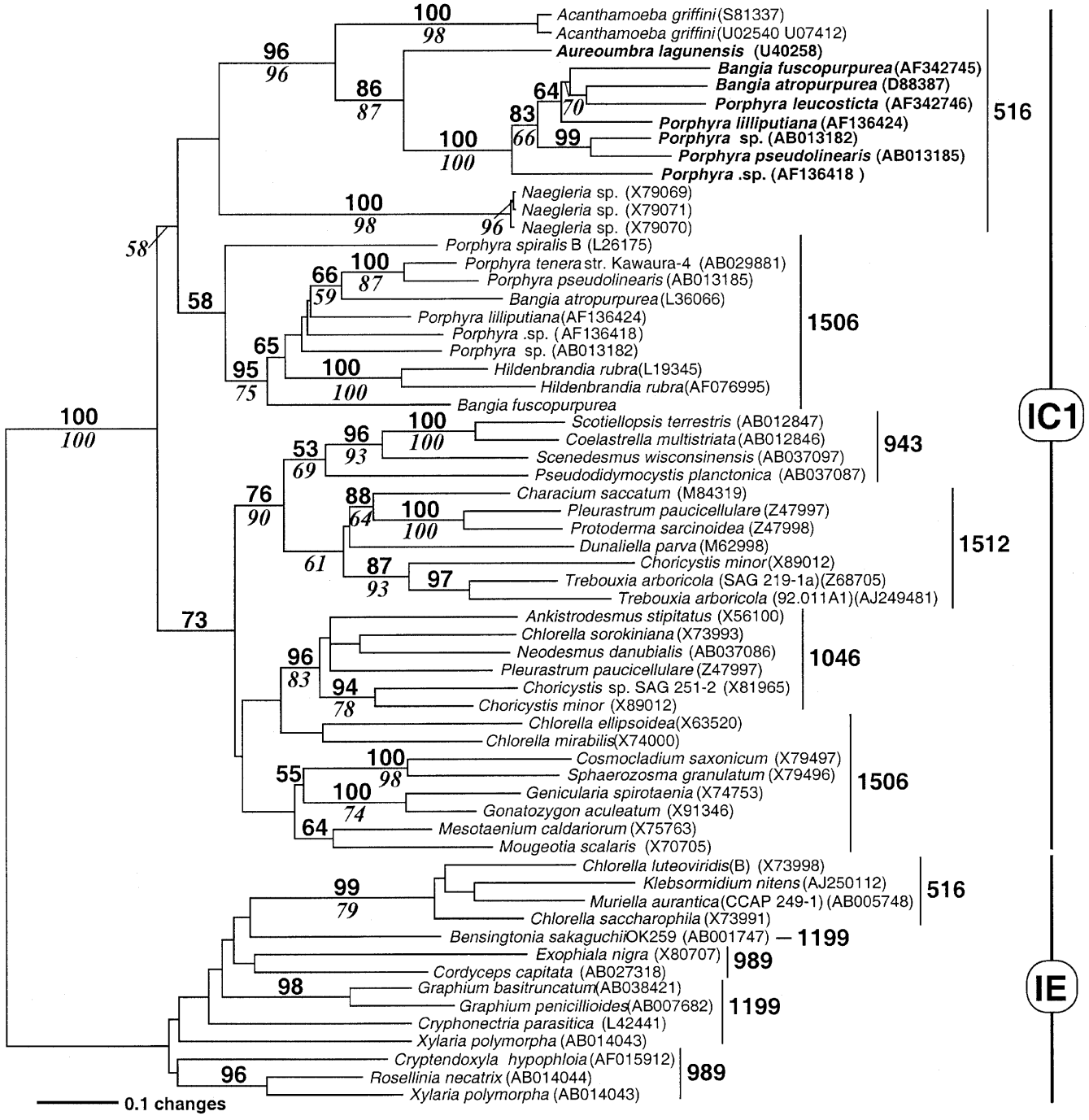


Fig. 3 Phylogeny of IC1 and IE introns interrupting the nuclear SSU rRNA of different eukaryotes. This tree was built with neighbor-joining, using a matrix of corrected distances (HKY-85 model). Bootstrap values (1,000 replications) are shown at the branches. The *values shown below the branches* result from a distance bootstrap analysis (1,000 replications), using the GTR + G + I model. Only values ≥50% are shown. The rRNA intron insertion sites are indicated. The tree has been rooted on the branch leading to the IE introns. GenBank accession numbers are shown (*in parentheses*) for each intron sequence. The group IC1 introns at position 516 that contain the P5b insertion are shown in *bold*

are postulated to be the sister group of the Florideophycidae (Oliveira and Bhattacharya 2000; Müller et al. 2001b), a hypothesis that is supported by the presence of the 1506 intron in only these red algae. We presume here that the common ancestor of the Bangiales/Florideophycidae likely contained the 1506 intron. Bootstrap-distance analysis of the 516 introns with the GTR + G + I model provided similar results to the HKY-85 model. The monophyly of the IC1 introns at position 516 of *Acanthamoeba*, Bangiales, and *Aureoumbra* was strongly supported (96%), although there

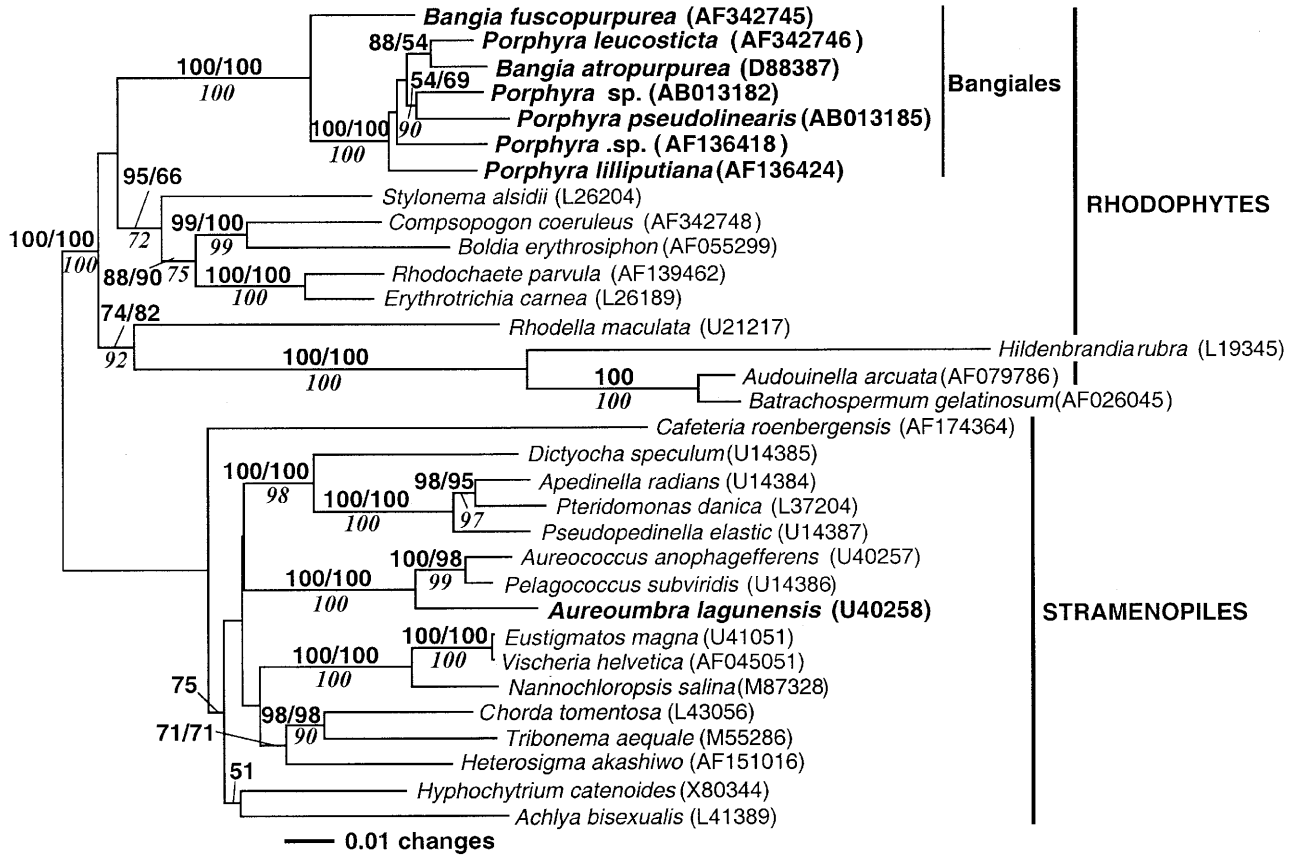


Fig. 4 Phylogeny of selected red algae (Rhodophytes) and stramenopiles, based on SSU rRNA sequence comparisons. This tree was built with neighbor-joining, using a matrix of corrected distances (HKY-85 model). Bootstrap values (1,000 replications) are shown *above the branches on the left of the slash-marks*. The values shown *on the right of the slash-marks* result from a distance bootstrap analysis (1,000 replications) using the GTR + G + I model. Bootstrap values shown *below the branches* are derived from an unweighted maximum parsimony analysis (1,000 replications). Only values ≥50% are shown. The Bangiales and *Aureoumbra* are shown in *bold*. GenBank accession numbers are indicated (*in parentheses*) for each SSU rRNA sequence

was again no support found for the inclusion of the *Naegleria* spp. 516 introns within this clade.

Given this phylogenetic framework for the group I introns at SSU rRNA site 516, we studied the phylogeny of the host cells containing these introns. Analysis of the SSU rRNA in red algae and stramenopiles showed that taxa containing the 516 intron are evolutionarily distantly related (Fig. 4). Recent phylogenetic analyses of more extensive data sets also suggested that the red algae share a most recent common ancestry with the green algae and land plants (Burger et al. 1999; Moreira et al. 2000). None of the green algae and land plants, however, contain an IC1 intron at position 516. The stramenopiles are often allied with alveolate protists, *Acanthamoeba* spp. are members of the Mycetozoa (a sister group of fungi/animals), and *Naegleria* (Heterolobosea) is a sister group of the Euglenozoa (e.g., Baldauf et al. 2000; Van de Peer et al. 2000). There is, therefore, no direct evidence to support a close evolutionary relationship between red algae, brown algae, acanthamebids, and *Naegleria* spp. The most parsimonious explanation for the close relationship of the IC1 516 introns, whereas the host cells containing them are not closely related, is that the intron was independently inserted into the homologous SSU rRNA site in the Bangiales, *Aureoumbra*, *Acanthamoeba*, and *Naegleria* (i.e., introns late). An observation that is consistent with this hypothesis is that, of 373 stramenopile SSU rRNA coding regions that have been determined (organism names and GenBank accession numbers available at CRW-algal introns), only the relatively late-diverging *Aureoumbra* (Pelagophyceae) contains the IC1 516 intron (see Fig. 4). To explain this distribution in terms of introns early, one would need to invoke the non-parsimonious hypothesis of widespread intron loss in all stramenopiles, except for the relatively late-diverging *Aureoumbra* (see Fig. 4). Furthermore, the 516 intron is limited to and sporadically distributed in the Bangiales within the Rhodophyta; and it is only found in three closely related acanthamebids of about 61 nuclear SSU rRNA sequences of *Acanthamoeba* strains/species that are currently available in GenBank (organism names and GenBank accession numbers available at CRW Algal Introns).

Another fascinating possibility for the origin of the *Aureoumbra* IC1 516 intron is through symbiotic gene

transfer. It is now generally believed that the plastid of photosynthetic stramenopiles has arisen through a single secondary (eukaryotic) symbiosis involving a red alga (Douglas et al. 1991; Bhattacharya and Medlin 1995, 1998; Daugbjerg and Andersen 1997; Oliveira and Bhattacharya 2000). It is, therefore, conceivable that the *Aureoumbra* intron may trace its origin to an intron in the rRNA gene of the red algal secondary symbiont. The intron could have been transferred from the symbiont nucleus to that of the stramenopile prior to the elimination of the red algal nuclear genome. However, an observation that argues against this scenario is that, as stated above, only *Aureoumbra* (of the many nuclear SSU rRNA sequences that have been determined from photosynthetic stramenopiles) contains an IC1 intron at position 516 in the SSU rRNA. If the intron was present in the photosynthetic ancestor of all stramenopiles, then it must have been lost in all of these taxa, except for the relatively late-diverging *Aureoumbra*.

In conclusion, our analyses provide clear insights into the origin of the IC1 introns at the 516 site of SSU rRNA. The host-cell phylogenetic data suggest that it is unlikely that this intron was vertically inherited from the common ancestor of the red algae, brown algae, acanthamebids, and *Naegleria* spp., because these lineages do not share a specific evolutionary relationship. This hypothesis is supported by the restricted distribution of the intron in each of the host-cell lineages. The intron phylogeny shows that the IC1 516 intron is evolutionarily distantly related to the IE 516 intron, supporting independent intron insertions into the homologous SSU rRNA site. The intron tree also provides strong support for the monophyly of the IC1 516 intron containing the insertion in helix P5b within the Bangiales and *Aureoumbra*. This group also includes the *Acanthamoeba* 516 intron that lacks the insertion. The *Naegleria* spp 516 introns are only distantly related to this clade.

Taken together, these data are consistent with the second scenario that the Bangiales and *Aureoumbra* IC1 516 introns were acquired after the separation of the red and brown algal lineages. We suggest this intron has been inserted independently into the SSU rRNA of the Bangiales, into a single stramenopile brown alga, *Aureoumbra*, and into derived members of the genera *Acanthamoeba* and *Naegleria*. Finally, our data suggest not only that lateral transfer may establish new intron lineages in different organisms but also that, over time, these sequences may evolve into distinct secondary structure variants (i.e., containing the P5b helix insertion), perhaps reflecting selection pressure to adapt to the novel splicing environment. The IC1 introns at position 516 in the SSU rRNA may, therefore, be a valuable example of how novel intron groups are created through the dual process of lateral transfer followed by rapid secondary structure evolution. In this regard, the P5b insertion has unambiguously marked the intron in the Bangiales and *Aureoumbra* and has allowed us to identify these homologous introns in evolutionarily divergent protists. Detailed analysis of group I intron secondary structure promises, therefore, to provide an important source of comparative data to elucidate intron origin and evolution.

Acknowledgements D.B. was supported by grants from the Carver Foundation. R.R.G. was supported by research grant NIH-GM48207 and start-up funds from the University of Texas, Institute for Cellular and Molecular Biology.

References

Abramovitz DL, Pyle AM (1997) Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop–receptor interaction. J Mol Biol 266:493–506

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Baldauf S, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–977

Bhattacharya D (1998) The origin and evolution of protist group I introns. Protist 149:113–122

Bhattacharya D, Medlin L (1995) The phylogeny of plastids: a review based on comparisons of small subunit ribosomal RNA coding regions. J Phycol 31:489–498

Bhattacharya D, Medlin L (1998) Algal phylogeny and the origin of land plants. Plant Physiol 116:9–15

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

Bhattacharya D, Friedl T, Damberger S (1996a) Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion site lineages in the green algae. Mol Biol Evol 13:978–989

Bhattacharya D, Damberger S, Surek B, Melkonian M (1996b) Primary and secondary structure analyses of the rDNA group I introns of the Zygnematales (Charophyta). Curr Genet 29:282–286

Burger G, Saint-Louis D, Gray MW, Lang BF (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: cyanobacterial introns and shared ancestry of red and green algae. Plant Cell 11:1675–1694

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. Science 273:1678–1685

Cech TR (1985) Self-splicing RNA: implications for evolution. Int Rev Cytol 93:3–22

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

Damberger SH, Gutell RR (1994) A comparative database of group I intron structures. Nucleic Acid Res 22:3508–3510

Daugbjerg N, Andersen RA (1997) Phylogenetic analyses of the *rbcL* sequences from haptophytes and heterokont algae suggest their chloroplasts are unrelated. Mol Biol Evol 14:1242–1251

Doherty EA, Doudna JA (2000) Ribozyme structures and mechanisms. Annu Rev Biochem 69:597–615

Doherty EA, Herschlag D, Doudna JA (1999) Assembly of an exceptionally stable RNA tertiary interface in a group I ribozyme. Biochemistry 38:2982–2990

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Molecular evidence that cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct eukaryotes. Nature 350:148–151

Felsenstein J (1985) Confidence intervals on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Friedl T, Besendahl A, Pfeiffer P, Bhattacharya D (2000) The distribution of group I introns in lichen algae suggests that lichenization facilitates intron lateral transfer. Mol Phylogenet Evol 14:342–352

Gutell RR (1996) Comparative sequence analysis and the structure of 16S and 23S rRNA. In: Zimmermann RA, Dahlberg AE (eds) Ribosomal RNA: structure, evolution, processing, and function in protein biosynthesis. CRC Press, Boca Raton, pp 111–128

Hasegawa M, Kishino H, Yano T (1985) Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hedenstierna KO, Siefert JL, Fox GE, Murgola EJ (2000) Coconservation of rRNA tetraloop sequences and helix length suggests involvement of the tetraloops in higher-order interactions. Biochimie 82:221–227

Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell 31:147–157

Lopez P, Forterre P, Philippe H (1999) The root of the tree of life in the light of the covarion model. J Mol Evol 49:496–508

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

Moreira D, Le Guyader H, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. Nature 405:69–72

Moritz C, Schneider CJ, Wake DB (1992) Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Syst Biol 41:273–291

Müller KM, Sheath RG, Vis ML, Crease J, Cole KM (1998) Biogeography and systematics of *Bangia* (Bangiales, Rhodophyta) based on the rubisco spacer, *rbcL* gene and 18S rRNA gene sequences and morphometric analyses. 1. North America. J Phycol 37:195–207

Müller KM, Cannone JJ, Gutell RR, Sheath RG (2001a) A structural and phylogenetic analysis of the Group IC1 introns in the order Bangiales (Rhodophyta). Mol Biol Evol 18(8) (in press)

Müller KM, Oliveira MC, Sheath RG, Bhattacharya D (2001b) Ribosomal DNA phylogeny of the Bangiophycidae (Rhodophyta) and the origin of secondary plastids. Am J Bot 88(8) (in press)

Nishida K, Suzuki S, Kimura Y, Nomura N, Fujie M, Yamada T (1998) Group I introns found in *Chlorella* viruses: biological implications. Virology 15:319–326

Oliveira M, Bhattacharya D (2000) Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. Am J Bot 87:482–492

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Ragan MA, Bird CJ, Rice EL, Singh RK (1993) The nuclear 18S ribosomal RNA gene of the red alga *Hildenbrandia rubra* contains a group I intron. Nucleic Acids Res 21:3898

Rodríguez F, Oliver JF, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142:485–501

Suh S-O, Jones KG, Blackwell M (1999) A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of group I introns. J Mol Evol 48:493–500

Swofford DL (2001) PAUP*: Phylogenetic analysis using parsimony, ver. 4.0b8. Smithsonian Institution, Washington, D.C.

Turmel M, Côté V, Otis C, Mercier J-P, Gray MW, Lonergan KM, Lemieux C (1995) Evolutionary transfer of ORF-containing group I introns between different subcellular compartments (chloroplast and mitochondrion). Mol Biol Evol 12:533–545

Van de Peer Y, Baldauf SL, Doolittle WF, Meyer A (2000) An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. J Mol Evol 51:565–576

Watanabe KI, Ehara M, Inagaki Y, Ohama T (1998) Distinctive origins of group I introns in the *COXI* genes of three green algae. Gene 213:1–7

Woese CR, Winker S, Gutell RR (1990) Architecture of ribosomal RNA: constraints on the sequence of tetra-loops. Proc Natl Acad Sci USA 87:8467–8471