ORIGINAL ARTICLE

Putative group I introns in the eukaryote nuclear internal transcribed spacers

Daniele Corsaro[1] · Danielle Venditti[1]

Abstract
Group I introns are mobile genetic elements that interrupt genes encoding proteins and RNAs. In the rRNA operon, introns can insert in the small subunit (SSU) and large subunit (LSU) of a wide variety of protists and various prokaryotes, but they were never found in the ITS region. In this study, unusually long ITS regions of fungi and closely related unicellular organisms (*Polychytrium aggregatum*, *Mitosporidium daphniae*, *Amoeboaphelidium occidentale* and *Nuclearia simplex*) were analysed. While the insertion of repeats is responsible for long ITS in other eukaryotes, the increased size of the sequences analysed herein seems rather due to the presence of introns in ITS-1 or ITS-2. The identified insertions can be folded in secondary structures according to group I intron models, and they cluster within introns in conserved core-based phylogeny. In addition, for *Mitosporidium*, *Amoeboaphelidium* and *Nuclearia*, more conventional ITS-2 structures can be deduced once spacer introns are removed. Sequences of five shark species were also analysed for their structure and included in phylogeny because of unpublished work reporting introns in their ITS, obtaining congruent results. Overall, the data presented herein indicate that spacer regions may contain introns.

Keywords Group I introns · Internal transcribed spacer · ITS-2 · Spacer introns · Spintrons

Introduction

In most eukaryotes, the nuclear ribosomal RNA genes are present in multiple copies as tandem repeats separated by intergenic spacers (IGS). The rRNA genes are usually arranged as two distinct units, the 5S and the main rDNA unit, transcribed by the RNA polymerase III and I, respectively (Orioli et al. 2012; Goodfellow and Zomerdijk 2013). The rDNA unit contains the 18S, 5.8S and 28S rDNA organized in a single operon, where the genes are separated by two internal transcribed spacers (ITS-1 and ITS-2) and regulatory elements for the transcription by RNA polymerase I in a non-transcribed spacer (NTS) (Goodfellow and Zomerdijk 2013; Henras et al. 2015). The rDNA unit is transcribed into a long 45S pre-rRNA that undergoes a complex maturation process to release the three rRNAs, in which conformational folding interactions with ITS-1 and ITS-2 play a role in the intervention of various nucleolytic cleavage factors (Fernández-Pevida et al. 2015; Henras et al. 2015; de la Cruz et al. 2018).

Microsporidia are a notable exception to such an organization because their genes are considerably reduced in size and structure (e.g. 16S-like SSU), and the 5.8S gene is fused at the 5′ of the LSU, giving an overall prokaryote-like resemblance (Torres-Machorro et al. 2010). This is certainly related to the evolution of microsporidians as extreme intracellular parasites, and it does not reflect their ancestral nature, as originally thought. They are indeed characterized by highly compact genomes and unusual morphological features (Corradi 2015; Han and Weiss 2017). The 5.8S/LSU fusion occurred with the loss of ITS-2. The reduction in size and structure can be traced from the mitochondriate ancestors of Microsporidia (these have DNA-free mitosomes) within the Rozellomycota (Corsaro et al. 2014, 2016), a basal lineage of zoosporic fungi (chytrids). Some of these organisms (i.e. *Mitosporidium*, *Nucleophaga*, *Paramicrosporidium*) have indeed microsporidia-like morphologies while preserving fungal genomic traits (Haag et al. 2014; Quandt et al. 2017). Although their SSU rRNAs are typically

✉ Daniele Corsaro
corsaro@gmx.fr

[1] CHLAREAS, 12 rue du Maconnais,
Vandoeuvre-lès-Nancy 54500, France

eukaryotic in size (~ 1800 nt) and structure, the length of their ITS-2 is reduced by almost half (~ 150 nt) compared to that of other chytrids (~ 300 nt). A greater reduction in size and structure of ITS-2 is present in the most basal members of Microsporidia, the Chytridiopsida and Metchnikovellida, which, however, have SSU of eukaryotic structure, but of reduced size (Corsaro et al. 2014, 2019a). Microsporidia and Rozellomycota form a basal branch of the holomycotan lineage, which includes the kingdom Fungi and their closest protist relatives (Aphelidea, Nucleariidae) (Liu et al. 2009; Corsaro et al. 2014).

Surprisingly, the recent release in Genbank of the entire rDNA sequence of *Mitosporidium daphniae* (Genbank ID MF278562) shows that a much longer ITS-2 (722 nt) is present in this organism. Nevertheless, our preliminary analysis proved to be incompatible with the information reported in the database, suggesting instead the possible presence of a group I intron within the ITS.

Group I introns are mobile genetic elements that insert into the genes encoding for RNAs and proteins in viral, prokaryotic and nuclear and organellar genomes (Haugen et al. 2005; Hedberg and Johansen 2013). They are self-splicing RNAs, folding into complex structures that generally consist of ten paired regions (P1–P10). On the basis of structure and sequence variations, the introns can be divided into five main classes: IA, IB, IC, ID and IE (Michel and Westhof 1990; Suh et al. 1999). Some introns can further encode for different types of endonucleases that increase their spread (Chevalier and Stoddard 2001), although reverse splicing also plays an important role in their mobility (Bhattacharya et al. 2005). Introns have been widely studied, especially those occurring into the rRNA genes (i.e. SSU and LSU rDNA), but they have never been found in the ITS. The only note on this subject comes from communications to annual meetings by the team of Rogers, who reported the possible presence of introns in nuclear ITS of chondrichthyan fishes. The authors named them spintrons (for spacer introns) and hypothesized that they invaded a common ancestor of the existing sharks (Shivji et al. 1997; Walker et al. 1999). We therefore included along holomycotan rDNA, some of these shark sequences in our analysis. Our results indicate the possible presence of spacer introns (spintrons) in various organisms.

Materials and methods

Retrieval of ITS sequences

Along with the sequence of *Mitosporidium*, we retrieved from Genbank a set of sequences containing the complete portions of ITS-1, ITS-2 and 5.8S, for other members of Rozellomycota. Since the Rozellomycota form a basal group

near the root of the Fungi (Corsaro et al. 2014), we have included in the comparison the sequences of additional chytrids (Chytridiomycota) and other two basal holomycotan groups, the Aphelidea and Nucleariida (Table 1). We selected ITS-1 and ITS-2 sequences from five shark species (Chondrichthyes, Elasmobranchii, Selachimorpha) that have been shown forming distinct lineages (Cooper 2018). Species are: *Hexanchus griseus* (JN003428, JN003446) and *Notorynchus cepedianus* (JN003429, JN003447) (Squalimorphii, Hexanchiformes); *Carcharhinus leucas* (JN003436, JN003437) (Galeomorphii, Carcharhiniformes); *Isurus oxyrinchus* (JN003432, JN003441) and *Lamna nasus* (JN003431, JN003440) (Galeomorphii, Lamniformes).

Sequence analyses

Boundaries between the different rRNA genes were identified by locating in the alignments a set of highly conserved primers targeting known positions near to the terminal parts of 18S (euk1520F) and 28S (euk28-1R). Positions were further checked by conformational studies (Corsaro et al. 2019a) and research in the ITS-1 (http://www.itsonedb.cloud.ba.infn.it/) and ITS-2 (http://www.its2.bioapps.biozentrum.uni-wuerzburg.de/) databases (Ankenbrand et al. 2015; Santamaria et al. 2018). The selected sequences were screened for the presence of repeats using Tandem Repeat Finder (Benson 1999), and they were analysed performing separate multiple alignments with MAFFT including different groups of ITS-rDNA or introns. Blast search for possible similarities was also carried out for each sequence, both over its entire length and on partial fragments.

Intron characterization and molecular phylogeny

The reference group I intron sequence alignments (Michel and Westhof 1990; Suh et al. 1999; Li and Zhang 2005) were used as guide to identify the conserved core paired helices P3–P7. Thus, secondary core structures were inferred manually, and the remaining portions of the molecules were determined using Mfold. Final structures were drawn using RnaViz (De Rijk et al. 2003). The conserved core pairs (P3–P8) of spacer introns were manually aligned with those of representatives of each intron class (IA–IE), including mostly introns located in the SSU or LSU rDNA of bacteria (b), nucleus or organelles (c, chloroplast; m, mitochondria) of different eukaryotes. The molecular phylogenetic tree was then inferred as described previously (Corsaro et al. 2017, 2019a, b), by using maximum likelihood (ML, GTR + Γ model; 1000 replicates), neighbour joining (NJ, Jukes-Cantor model; 2000 replicates) and maximum parsimony (MP, 1000 replicates) with TBR (tree-bisection-reconnection, search level 1 with 10 random additions), and using IE introns as outgroup.

Table 1 Summary of holomycotan ITS sequences analysed

Taxon	Genbank ID	Length (nt)		Note[a]
		ITS-1	ITS-2	
Chytridiomycota				
Gromochytrium mamkaevae	KF586842	281	449	
Polychytrium aggregatum	NR_119549	**820**	406	Spintron 1
Spizellomyces punctatus	AY546684, AY997092, AY546692	295	233	
Rozellomycota				
Mitosporidium daphniae	MF278562	**414**	**722**	Spintrons 1 and 2
Paramicrosporidium saccamoebae	JQ796369	108	129	
Rozella allomycis	NG_017174, NR_119552	218	256	
unc. clone GL03020_035_G2840	KY687565	211	288	
unc. clone WS_CM1	FJ687268	183	235	
Aphelidea				
Amoeboaphelidium occidentale	JX967274	365	**826**	Spintron 2; IC1 S943[b]
Amoeboaphelidium protococcarum	JX507298	196	168	
Aphelidium desmodesmi	KY249641	218	219	
Nucleariidae				
Nuclearia moebiusi	AF484686	304	375	
Nuclearia simplex	AF484687	236	**589**	Spintron 2

[a]Spintron 1 or 2 refers to the putative presence of intron in ITS1 or ITS2 (in bold), respectively

[b]Intron of IC1 type presents in the 18S rRNA at the 943 site

ITS-2

For *M. daphniae*, *Amoeboaphelidium occidentale* and *Nuclearia simplex*, the identified spacer introns were removed and secondary structures of ITS-2 were inferred manually as previously described (Corsaro et al. 2014, 2019a).

Results

ITS length

Some of the selected sequences of holomycotans revealed high differences in the size of ITS. *Mitosporidium daphniae* and *Polychytrium aggregatum* have ITS-1 two and four times longer, respectively. For ITS-2, sizes exceeding two to five times the lengths of the other members were found in *Mitosporidium*, *A. occidentale* and *N. simplex* (Table 1). These unusually long sequences were analysed further. The ITS-1 of *Polychytrium* and ITS-2 of *Nuclearia* have no repeats, whereas these are present in both ITS of *Mitosporidium*. ITS-1 contains two repeats, long (11 bases: GATGGT GGATC) and short (2 bases: AC), for total lengths of 38 and 64 bp, respectively. ITS-2 has a unique short repeat (4 bases: ATCA; total length of 34 bp). The ITS-2 of *A. occidentale* proves to be particular because it contains a long portion of 409 nt repeated twice, separated by 22 nt corresponding to the first bases of 28S. The true beginning of the 28S is,

however, very likely localized after this double repetition where we can locate, as expected, the primer 28-1R site preceded by this same terminal portion. Another abnormal situation occurs in the ITS-2 of *Aphelidium desmodesmi*, where the 5′-end of 28S consisting of the first 20 nt followed by the 28-1R primer site is repeated twice. We consider these sequences as true, but a possible error in their amplification/assembly cannot be ruled out.

Repeats were not found in the shark sequences, except for the ITS-2 sequences of *C. leucas*, *Hexanchus griseus* and *N. cepedianus* (consensus patterns of 20–26 bases for total lengths of 41–63 bp).

Spacer introns identification and deduced ITS-2 in the holomycotans

The selected sequences do not show significant similarities with any other available sequence. On the other hand, for each of them it was possible to identify short regions (< 10 nt) corresponding to conserved pairs of introns of group I. The 2D reconstructions of the molecules according to the pairs identified proved to be in accordance with the group I intron model (Michel and Westhof 1990; Cech et al. 1994). The complete set of P1–P10 pairs was found (Figs. 1a, b, 2a, b, 3). All have a wobble *u:G* at the 5′ splice site, most with a complementary pair at the *G*-binding site in the P7, normally formed by the highly conserved *C:G* or the rarer *U:G* or *U:A*. While the ITS-1 intron of *Polychytrium*
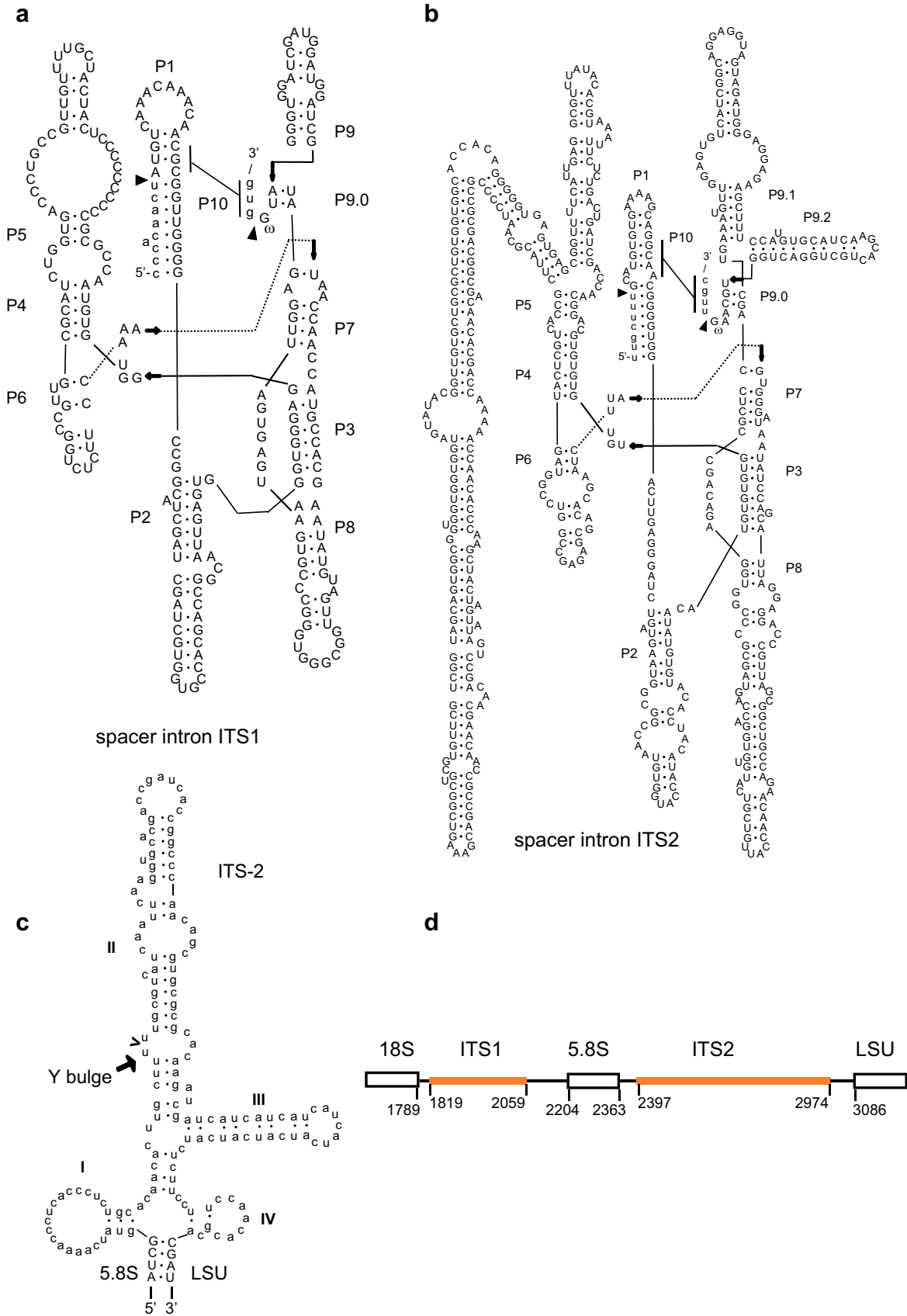
a

spacer intron ITS1

b

spacer intron ITS2

c

d

◄Fig. 1 Predicted secondary structures of the spacer introns present in the ITS-1 (**a**) and ITS-2 (**b**) of *Mitosporidium daphniae*. Sequences of the introns are in uppercase, and portions of the ITS serving as flanking exons are in lowercase. Paired elements (P1–P10) and the 5′ and 3′ splice sites (black arrowheads) are indicated. Predicted secondary structure of the ITS-2 once the spacer intron has been removed (**c**). Ends of 5.8S and LSU rDNA are in uppercase, and the sequence of the ITS-2 is in lowercase. Roman numbers label ITS helices, and arrowhead indicates the insertion site of the spacer intron. Schematic draw of the rDNA operon of *Mitosporidium* (**d**). The rRNA genes are indicated by white boxes, and the ITS by thin lines. Spacer introns are represented by thick lines. Numbers refer to the position of the *M. daphniae* rDNA sequence (Genbank ID MF278562)

possesses as terminal residue at the 3′-end the rare *U*, all the others have the most common *G* ($G\omega$). Furthermore, the ITS-2 intron of *A. occidentale* contains an open reading frame in the P9 loop (Fig. 2a), and its translation in frame-2 gives a theoretical product of 44 amino acids containing an His-Cys box (Fig. 4). The presence of such a highly degenerated gene would be congruent with the intronic nature of these elements because the nuclear introns of several protists, targeting the rRNA genes, can indeed contain homing endonucleases of the His-Cys box family (Haugen et al. 2004; Corsaro and Venditti 2018). For the ITS-2 intron of *Mitosporidium*, *A. occidentale* and *N. simplex*, the putative insertion site of the introns is located in helix 2 of ITS-2 deduced after intron removal, near the pyrimidine bulge. Indeed, once the introns were removed from the original sequences of the three organisms, it was possible to reconstruct 2D models of ITS-2 that appear more consistent with those expected (Figs. 1c, 2c, d). All have a multi-helix structure typical of ITS-2 containing a helix 2 with a characteristic pyrimidine bulge. The ITS-2 of *Mitosporidium* has only 148 nt, a length totally compatible with that envisaged for this group of microsporidia-like organisms (Corsaro et al. 2014, 2019a).

Spacer introns in sharks

As indicated before, the suggestion that the shark ITS could contain introns was advanced by Rogers and colleagues. These works have, to our knowledge, remained unpublished, and they are only available as conference papers or master thesis. In the present study, we were able to effectively reconstruct secondary structures corresponding to the group I intron models from ITS-1 and ITS-2 of five shark species (Fig. 5), which would confirm the spintron hypothesis.

Intron phylogeny

To trace the phylogenetic relationships of the spacer introns, a set of rDNA introns, occurring in prokaryotes and eukaryote nuclear and organellar genomes and representing the five

intron classes IA–IE, was selected. Introns present in the DNA viruses infecting protists were also included. Phylogenetic analysis was performed on the basis of conserved regions P3–P8.

The spacer introns seem to form a unique group within the IC class showing some relationships with introns from viruses of amoebae (*Tupanvirus*) and nuclear introns of *Acanthamoeba lenticulata* and *Physarum polycephalum* (Fig. 6). However, the statistical supports are weak, although all methods recovered the virus intron as the closest relative, as well as the introns of *Acanthamoeba* and *Physarum* as the next relative but in different positions. The spacer introns are divided into subgroups almost in accordance with their location in ITS-1 or ITS-2. The ITS-1 and ITS-2 introns of *Mitosporidium* cluster with the ITS-1 intron of *Polychytrium* and the shark ITS-2 introns, respectively. The ITS-2 intron of *A. occidentale*, which has no relation with the S943 intron located in the 18S rDNA of the same organism we previously shown belonging to IC (Corsaro et al. 2019b), appears rather to form a lineage with the ITS-2 intron of *Nuclearia*, while shark ITS-1 introns form a paraphyletic basal group (holophyletic in NJ and MP). This could be due to the fact that for sharks we only included five of these introns. Indeed, according to Rogers and colleagues, shark introns should form two sister groups that, inherited vertically, have diversified over the evolutionary history of the hosts. In addition, it is possible that the internal topology is affected by the artificial position of the L2449 intron of *Physarum* and that, therefore, ITS-1 and ITS-2 introns actually form sister groups. Overall, although the present analysis probably does not reflect an accurate phylogeny, it would nonetheless confirm that these elements belong to introns.

Discussion

Exceptionally long ITS have already been found in very distant organisms such as some oomycetes or arthropods, where the increase in size of the sequences is caused by the inclusion of tandem repeated elements (e.g. Hlinka et al. 2002; Thines 2007; Paredes-Esquivel and Townson 2014). These elements generally consist of about ≥ 100 nt, and their repetition gives ITS often exceeding 1.5 kb. Such a tandem repetition of long fragments does not seem to occur in the sequences analysed herein. Indeed, repeats are absent in most ITS of sharks as well as of *Polychytrium* and *Nuclearia*, and those identified have only short fragments for total lengths of about 50 nt. The only exception is the sequence of *Amoeboaphelidium*. On the other hand, portions of these sequences can be folded, giving congruent configurations with the 2D models of the group I introns (Figs. 1a, b, 2a, b, 3), and the phylogenetic analysis conducted on the conserved regions corroborates their membership to group I introns
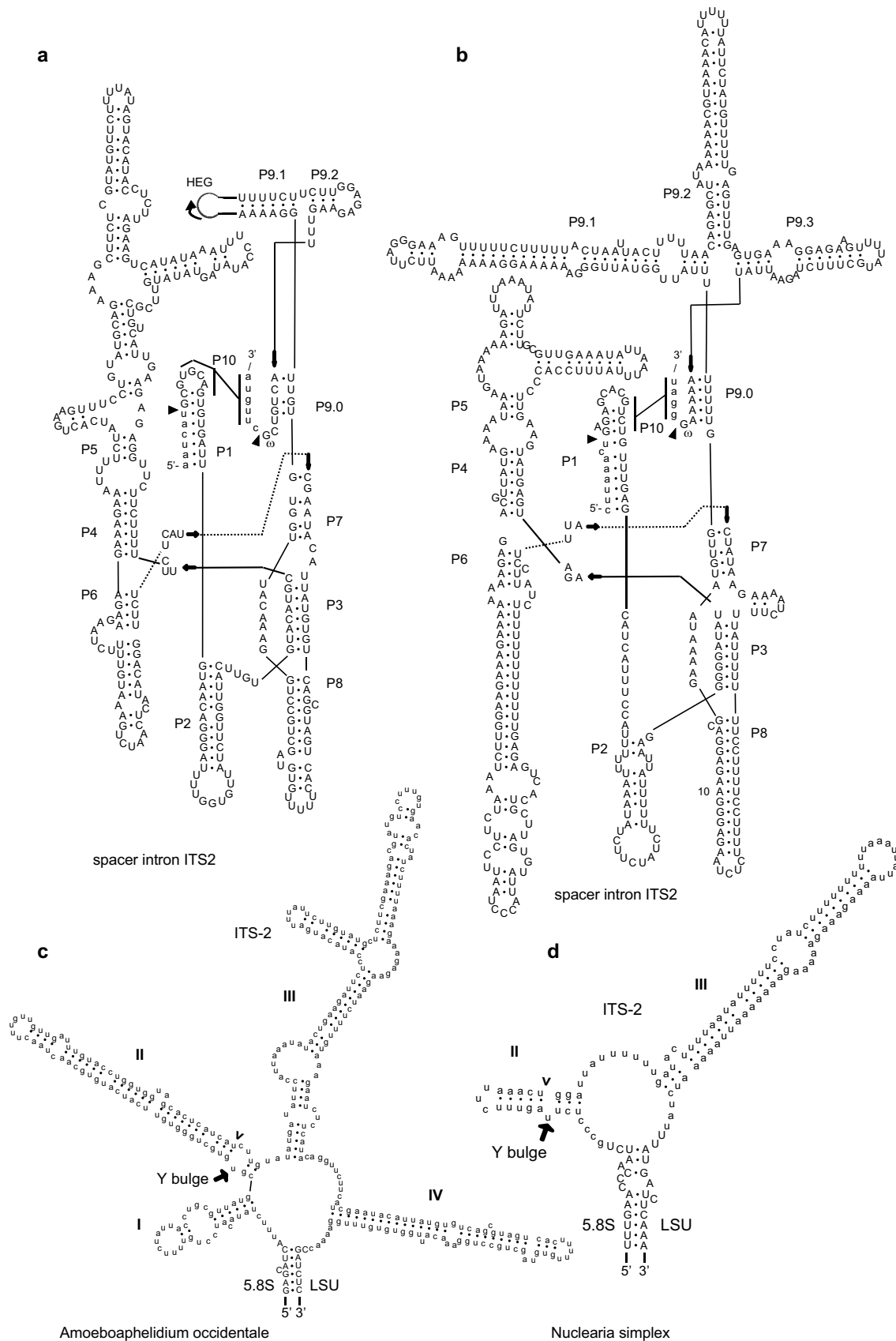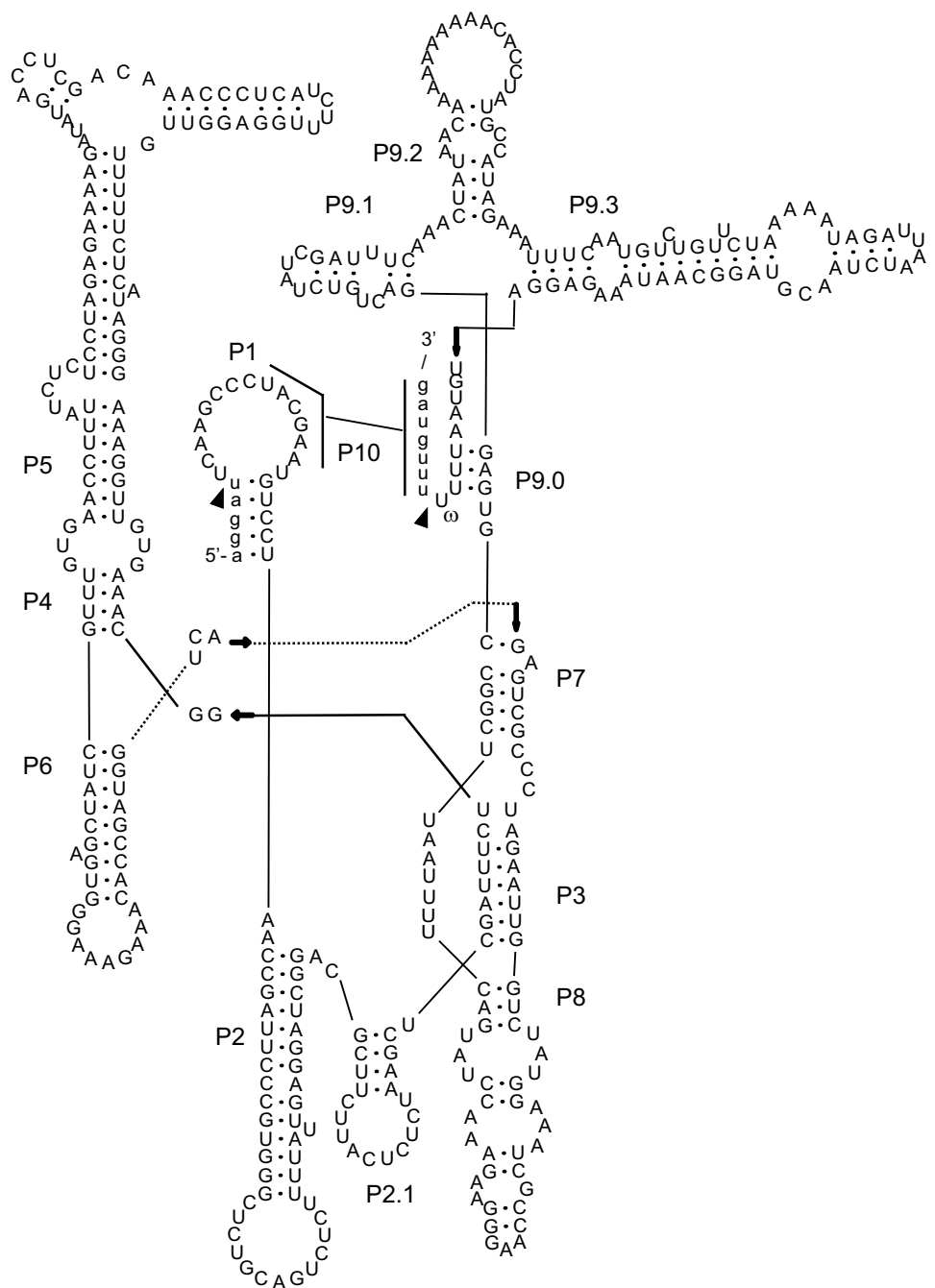
Fig. 2 Predicted secondary structures of (top) the spacer introns present in the ITS-2 and of (bottom) the ITS-2 after intron removal, of *Amoeboaphelidium occidentale* (**a**, **c**) and *Nuclearia simplex* (**b**, **d**). Labelling as in Fig. 1

Fig. 3 Predicted secondary structure of the spacer intron present in the ITS-1 of *Polychytrium aggregatum*. Labelling as in Fig. 1

(Fig. 6). Residing in the ITS, which are much more variable compared to 18S and 28S rRNA, spacer introns could therefore be highly derived lineages representing new subtypes or even new classes. In addition, ITS-2 of plausible structure can be modelled for all three holomycotans after intron exclusion (Figs. 1c, 2c, d). These reconstructions indicate that a vestigial ITS-2 can theoretically form once the spacer intron is removed, which could occur naturally during the processing of the pre-rRNA if these elements are actually self-splicing introns. In the same way, self-splicing should also occur for introns in the ITS-1 by releasing a functional

pre-rRNA intermediate. Therefore, assuming that these elements are true introns, they should behave as such, with little or no effect at the DNA level, whereas at the RNA level their self-splicing should allow the release of functional products from the invaded genes. The protistan nature of these organisms could also explain how the introns may have invaded them. Introns are indeed widespread, although scattered among protists at the level of nuclear rRNA, and they are particularly common in fungi and algae (Bhattacharya et al. 1996; Haugen et al. 2004; Simon et al. 2005). Once the intron is inserted into the host DNA, it propagates by vertical

```
                                                 His-Cys box
Njam 165 -TVSSSKKDPDARTISHLCGNGGCARPGHLRIEKKSVNDERTHCHF---LLRRSQSVAQS--EMIRLACPHTPRC 233
Ngru 127 -THNSFA-QNSGKEHTHICGDTNCCRPSHVIYAPKGTNISRTNCPG---YVRWKDN-NNK--EIFYKVCKHDPPC 193
Aspe 172 -KVSQRKKEKNAYSVSHLCGNGGCARPGHIIIEPKTVNDERVACHR---FLRSCKCKWDA--RVIRRLCPHKPKC 240
Cmoj  86 -RLNAGK---KRLQVSHLCHRGWCFNPEHLVVESQRLNAKRNNCKGnniFITNNGVTINP-------CPHGKNG 148
Pumb  70 -IFNAGRVPVEGEQASHRCHRGACVNPQHLVFETERVNKTRQYCDH---FKDHPDHN----------CLHQPQC 131
Ppol  83 fLLEPDNINGKTCTASHLCHNTRCHNPLHLCWESLDDNKGRNWCPG---PN-------------gGCVHAVVC 138
Acan  46 -QVEASK-DRSSKLLSHICGTRNCCNVDHIVIETKGINDERTHCHF---CMLNVFNASGHSGVARFLGCPHTPRC 118
Aocc   1 aCTQVTKINHKSLLQSH-CTHDECYHQVQINNKVSCTHDTPSTRA---------------------------  44
            o *    *     *         o       *                            * *
```

Fig. 4 The deduced amino acid sequence (44 aa) of the putative homing endonuclease found in the ITS-2 intron of *Amoeboaphelidium occidentale* (Aocc) is aligned with the sequences of His-Cys endonucleases present in nuclear introns of various protists by using COBALT. Selected sequences were from three heterolobosean amoebae, *Naegleria jamiesoni* (Njam), *Naegleria gruberi* (Ngru) and *Allovahlkampfia spelaea* (Aspe), the zoopagomycotan fungus *Coemansia* *mojavensis* (Cmoj), the rhodophyte *Porphyra umbilicalis* (Pumb), the myxogastrid *Physarum polycephalum* (Ppol) and *Acanthamoeba* sp. KA/E5 genotype T4 (Acan) (see Corsaro and Venditti 2018). The sequence carried by the aphelid is very small, but an almost complete His-Cys box can be identified. Asterisks and open circles mark the conserved zinc-binding and active site residue, respectively

inheritance to the entire lineage. The possible presence of endonuclease and reverse splicing are responsible for its mobility at the DNA or RNA level, respectively (Bhattacharya et al. 2005; Hedberg and Johansen 2013). Such mobility includes either the change of the insertion site within the same host or the change of host by horizontal transfer, the latter generally requiring close cell-to-cell interactions that would promote gene transfer even between very distant organisms (Bhattacharya et al. 2001; Haugen et al. 2005).

More intriguing is the situation of ITS shark introns. In animals, group I introns were found only in the non-bilaterian basal lineages, i.e. poriferans, cnidarians and placozoans, and they are confined to the mitochondrial genomes, located in the cytochrome c oxidase subunit I (*cox1*) or the NADH dehydrogenase subunit 5 (*nad5*). These introns are usually of class IB and often the cox1 intron encodes for a LAGLIDADG endonuclease (Goddard et al. 2006; Emblem et al. 2014; Celis et al. 2017; Osigus et al. 2017; Schuster et al. 2017). Studies have shown the occurrence of vertical inheritances and horizontal transfers between species, between sponges and corals, as well as the possible fungal origin of some of these introns. The intron spread could be explained by the absence or weak germline barrier in non-bilaterian animals. In bilaterians, segregation between germline and somatic cells is thought to be a major obstacle to horizontal gene transfers, although those involving functional genes are possible via symbionts or parasites, or during the early developmental stages exposed to foreign DNA (Huang 2013; Boto 2014; but see Ku and Martin 2016). In their analyses, Rogers and colleagues (Shivji et al. 1997; Walker et al. 1999; Cooper 2018) found that spacers introns are widespread among sharks, and they suggested that intron

invasion probably occurred in an ancestor of present-day elasmobranchs (sharks, skates and rays), followed by vertical inheritance. A possible conformational role of the spintrons helping in the pre-rRNA processing could explain their conservation over the millions of years of shark evolution. Also, one could speculate that such early intron invasion occurred possibly via an infectious microorganism such as a virus or protist as a donor. Additional studies are obviously necessary.

It is likely that the great variability of the ITS does not offer many possibilities for intron insertion, which requires the recognition of specific sequences corresponding to the flanking exons. On the other hand, this is ensured by the conserved portions of the rRNA genes. The existence of spacer introns, if any, should therefore be a rare event, probably due to fortuitous circumstances. However, the data presented herein seem consistent with the interpretation of these elements as spacer introns. In addition, we analysed only a very small number of ITS because our initial interest was to study the sequence of *Mitosporidium* which was abnormally long according to our prediction (Corsaro et al. 2019a). More detailed and extended analyses, including a larger number of sequences, can thus shed light on the real existence of these putative spacer introns and their diffusion. In the event that these elements prove to be really present, and in several other sequences, this would open up a new field of research. In addition, there may also be direct impacts as the ITS region is considered a promising DNA barcode for fungi.
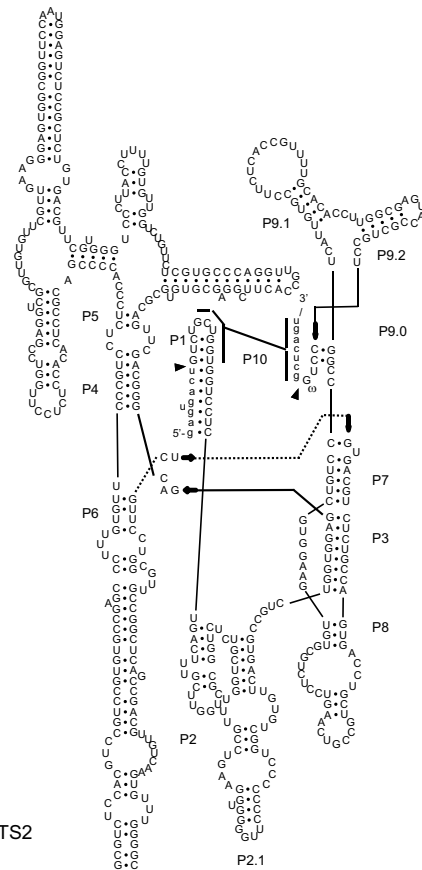
Fig. 5 Examples of predicted secondary structures of spacer introns ▶ present in the ITS-1 and ITS-2 of chondrichthyan fishes, *Carcharhinus leucas* (**a**, **b**) and *Isurus oxyrinchus* (**c**, **d**)
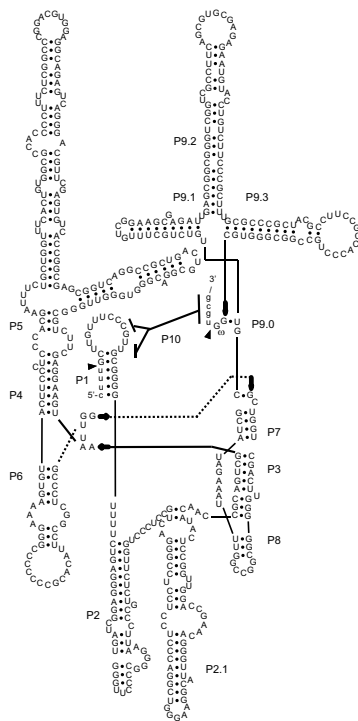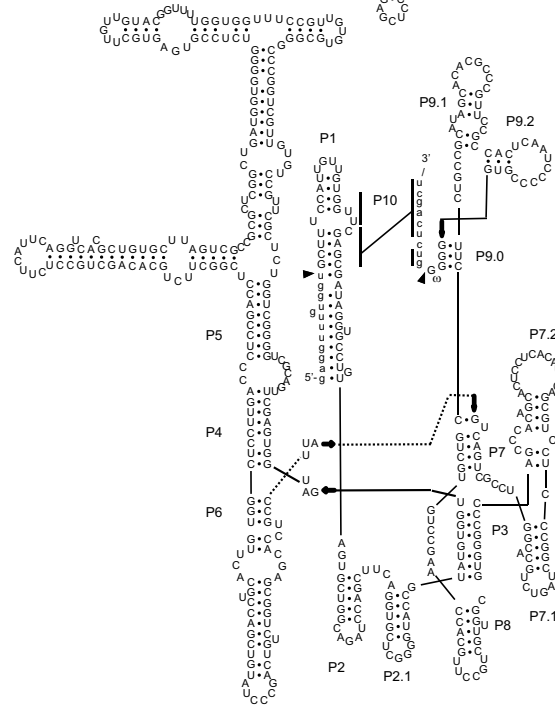
a
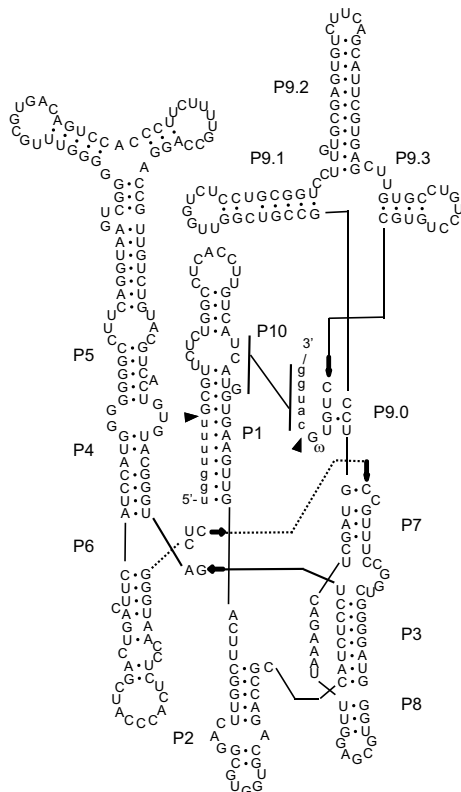
Carcharhinus leucas ITS1

b

Carcharhinus leucas ITS2

c

Isurus oxyrinchus ITS1
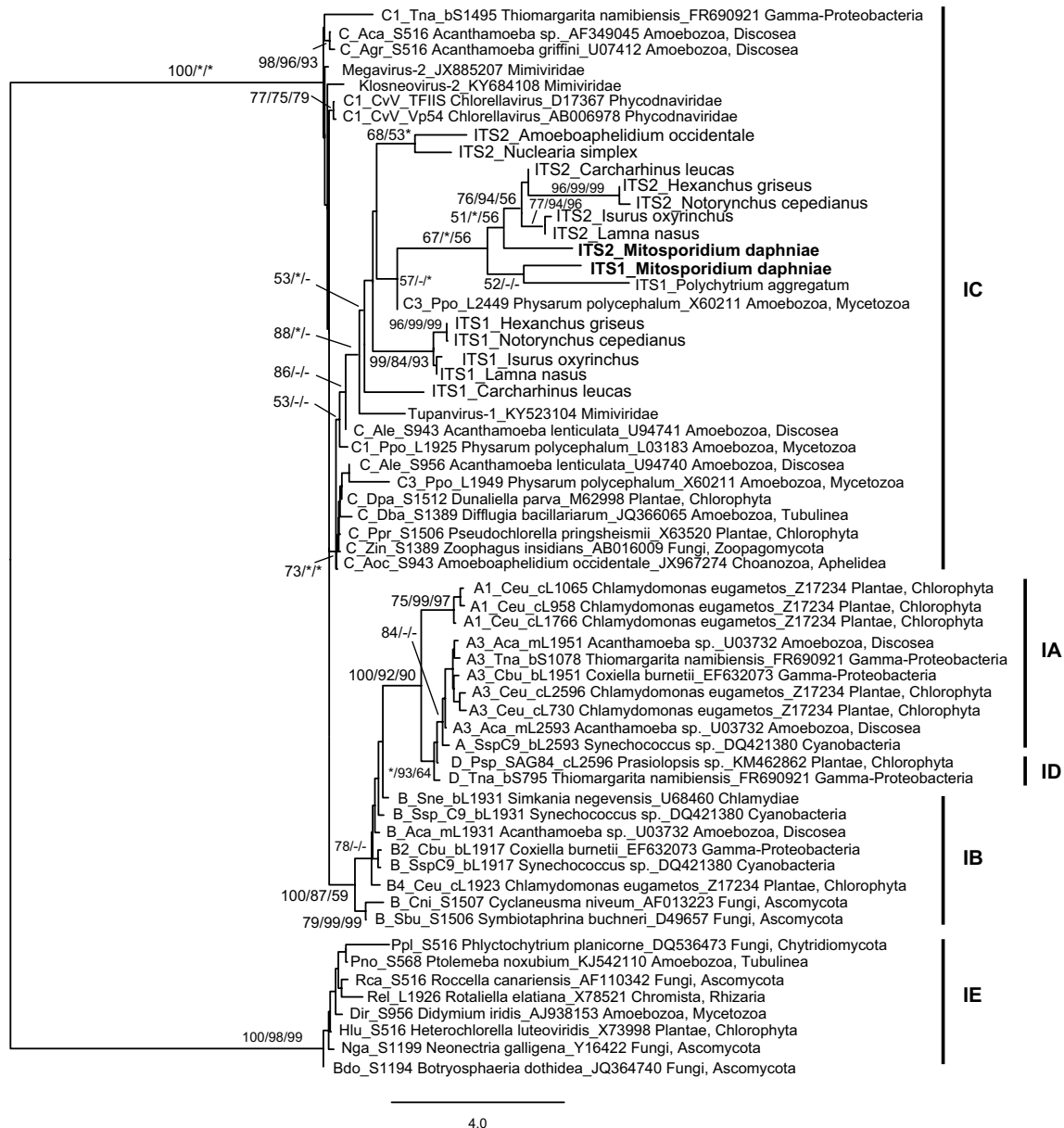
d

Isurus oxyrinchus ITS2

Fig. 6 Phylogenetic relationships of spacer introns within group I introns. Classes IA to IE are indicated. The analysis includes mainly introns located into either bacterial (b), eukaryotic nuclear or organellar (c, chloroplast; m, mitochondria) rDNA, using IE introns as outgroups. The only introns located in protein-encoding genes here considered are those occurring in DNA viruses of different protists (green algae and amoebae). The alignment is based on the conserved core P3–P8. Bootstrap values for ML/NJ/MP are shown at the node (1000 replicates for ML and MP; 2000 for NJ). Asterisk, node < 50%; hyphen, node not supported

Compliance with ethical standards

References

Ankenbrand MJ, Keller A, Wolf M, Schultz J, Förster F (2015) ITS2 Database V: twice as much. Mol Biol Evol 32:3030–3032. https://doi.org/10.1093/molbev/msv174

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Bhattacharya D, Friedl T, Damberger S (1996) Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion

site lineages in the green algae. Mol Biol Evol 13:978–989. https://doi.org/10.1093/oxfordjournals.molbev.a025666

Bhattacharya D, Cannone JJ, Gutell RR (2001) Group I intron lateral transfer between red and brown algal ribosomal RNA. Curr Genet 40:82–90. https://doi.org/10.1007/s002940100227

Bhattacharya D, Reeb V, Simon DM, Lutzoni F (2005) Phylogenetic analyses suggest reverse splicing spread of group I introns in fungal ribosomal DNA. BMC Evol Biol 5:68. https://doi.org/10.1186/1471-2148-5-68

Boto L (2014) Horizontal gene transfer in the acquisition of novel traits by metazoans. Proc R Soc B 281:20132450. https://doi.org/10.1098/rspb.2013.2450

Cech TR, Damberger SH, Gutell RR (1994) Representation of the secondary and tertiary structure of group I introns. Nat Struct Biol 1:273–280. https://doi.org/10.1038/nsb0594-273

Celis JS, Edgell DR, Stelbrink B, Wibberg D, Hauffe T, Blom J, Kalinowski J, Wilke T (2017) Evolutionary and biogeographical implications of degraded LAGLIDADG endonuclease functionality and group I intron occurrence in stony corals (Scleractinia) and mushroom corals (Corallimorpharia). PLoS One 12:e0173734. https://doi.org/10.1371/journal.pone.0173734

Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 29:3757–3774. https://doi.org/10.1093/nar/29.18.3757

Cooper L (2018) Evolutionary investigation of group I introns in nuclear ribosomal internal transcribed spacers in Neoselachii. Master Sci, Graduate College of Bowling Green State University, Ohio

Corradi N (2015) Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. Annu Rev Microbiol 69:167–183. https://doi.org/10.1146/annurev-micro-091014-104136

Corsaro D, Venditti D (2018) Nuclear Group I introns with homing endonuclease genes in *Acanthamoeba* genotype T4. Eur J Protistol 66:26–35. https://doi.org/10.1016/j.ejop.2018.07.002

Corsaro D, Walochnik J, Venditti D, Steinmann J, Müller KD, Michel R (2014) Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. Parasitol Res 113:1909–1918. https://doi.org/10.1007/s00436-014-3838-4

Corsaro D, Michel R, Walochnik J, Venditti D, Müller KD, Hauröder B, Wylezich C (2016) Molecular identification of *Nucleophaga terricolae* sp. nov. (Rozellomycota), and new insights on the origin of the Microsporidia. Parasitol Res 115:3003–3011. https://doi.org/10.1007/s00436-016-5055-9

Corsaro D, Köhsler M, Montalbano Di Filippo M, Venditti D, Monno R, Di Cave D, Berrili F, Walochnik J (2017) Update on *Acanthamoeba jacobsi* genotype T15, including full-length 18S rDNA molecular phylogeny. Parasitol Res 116:1273–1284. https://doi.org/10.1007/s00436-017-5406-1

Corsaro D, Wylezich C, Venditti D, Michel R, Walochnik J, Wegensteiner R (2019a) Filling gaps in the microsporidian tree: rDNA phylogeny of *Chytridiopsis typographi* (Microsporidia: Chytridiopsida). Parasitol Res 118:169–180. https://doi.org/10.1007/s00436-018-6130-1

Corsaro D, Wylezich C, Venditti D, Rott MR, Walochnik J (2019b) Recovery of an *Acanthamoeba* strain with two group I introns in the nuclear 18S rRNA gene. Eur J Protistol 68:88–98. https://doi.org/10.1016/j.ejop.2019.01.007

de la Cruz J, Gómez-Herreros F, Rodríguez-Galán O, Begley V, de la Cruz Muñoz-Centeno M, Chávez S (2018) Feedback regulation of ribosome assembly. Curr Genet 64:393–404. https://doi.org/10.1007/s00294-017-0764-x

De Rijk P, Wuyts J, De Wachter R (2003) RnaViz 2: an improved representation of RNA secondary structure. Bioinformatics 19:299–300

Emblem Å, Okkenhaug S, Weiss ES, Denver DR, Karlsen BO, Moum T, Johansen SD (2014) Sea anemones possess dynamic mitogenome structures. Mol Phylogenet Evol 75:184–193. https://doi.org/10.1016/j.ympev.2014.02.016

Fernández-Pevida A, Kressler D, de la Cruz J (2015) Processing of pre-ribosomal RNA in *Saccharomyces cerevisiae*. Wiley Interdiscip Rev RNA 6:191–209. https://doi.org/10.1002/wrna.1269

Goddard MR, Leigh J, Roger AJ, Pemberton AJ (2006) Invasion and persistence of a selfish gene in the Cnidaria. PLoS One 1:e3. https://doi.org/10.1371/journal.pone.0000003

Goodfellow SJ, Zomerdijk JC (2013) Basic mechanisms in RNA polymerase I transcription of the ribosomal RNA genes. Subcell Biochem 61:211–236. https://doi.org/10.1007/978-94-007-4525-4_10

Haag KL, James TY, Pombert JF, Larsson R, Schaer TM, Refardt D, Ebert D (2014) Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. Proc Natl Acad Sci USA 111:15480–15485. https://doi.org/10.1073/pnas.1410442111

Han B, Weiss LM (2017) Microsporidia: obligate intracellular pathogens within the fungal kingdom. Microbiol Spectrum. https://doi.org/10.1128/microbiolspec.FUNK-0018-2016

Haugen P, Reeb V, Lutzoni F, Bhattacharya D (2004) The evolution of homing endonuclease genes and group I introns in nuclear rDNA. Mol Biol Evol 21:129–140. https://doi.org/10.1093/molbev/msh005

Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. Trends Genet 21:111–119. https://doi.org/10.1016/j.tig.2004.12.007

Hedberg A, Johansen SD (2013) Nuclear group I introns in self-splicing and beyond. Mob DNA 4:17. https://doi.org/10.1186/1759-8753-4-17

Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE (2015) An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip Rev RNA 6:225–242. https://doi.org/10.1002/wrna.1269

Hlinka O, Murrell A, Barker SC (2002) Evolution of the secondary structure of the rRNA internal transcribed spacer 2 (ITS2) in hard ticks (Ixodidae, Arthropoda). Heredity 88:275–279. https://doi.org/10.1038/sj/hdy/6800040

Huang J (2013) Horizontal gene transfer in eukaryotes: the weak-link model. BioEssays 35:868–875. https://doi.org/10.1002/bies.201300007

Ku C, Martin WF (2016) A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70% rule. BMC Biol 14:89. https://doi.org/10.1186/s12915-016-0315-9

Li Z, Zhang Y (2005) Predicting the secondary structures and tertiary interactions of 211 group I introns in IE subgroup. Nucleic Acids Res 33:2118–2128. https://doi.org/10.1093/nar/gki517

Liu Y, Steenkamp ET, Brinkmann H, Forget L, Philippe H, Lang BF (2009) Phylogenomic analyses predict sistergroup relationship of nucleariids and fungi and paraphyly of zygomycetes with significant support. BMC Evol Biol 9:272. https://doi.org/10.1186/1471-2148-9-272

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610. https://doi.org/10.1016/0022-2836(90)90386-Z

Orioli A, Pascali C, Pagano A, Teichmann M, Dieci G (2012) RNA polymerase III transcription control elements: themes and variations. Gene 493:185–194. https://doi.org/10.1016/j.gene.2011.06.015

Osigus HJ, Eitel M, Schierwater B (2017) Deep RNA sequencing reveals the smallest known mitochondrial micro exon in animals: the placozoan cox1 single base pair exon. PLoS One 12:e0177959. https://doi.org/10.1371/journal.pone.0177959

Paredes-Esquivel CC, Townson H (2014) Functional constraints and evolutionary dynamics of the repeats in the rDNA

internal transcribed spacer 2 of members of the *Anopheles barbirostris* group. Parasit Vectors 7:106. https://doi.org/10.1186/1756-3305-7-106

Quandt CA, Beaudet D, Corsaro D, Walochnik J, Michel R, Corradi N, James TY (2017) The genome of an intranuclear parasite, *Paramicrosporidium saccamoebae*, reveals alternative adaptations to obligate intracellular parasitism. Elife 6:e29594. https://doi.org/10.7554/eLife.29594

Santamaria M, Fosso B, Licciulli F, Balech B, Larini I, Grillo G, De Caro G, Liuni S, Pesole G (2018) ITSoneDB: a comprehensive collection of eukaryotic ribosomal RNA Internal Transcribed Spacer 1 (ITS1) sequences. Nucleic Acids Res 46:D127–D132. https://doi.org/10.1093/nar/gkx855

Schuster A, Lopez JV, Becking LE, Kelly M, Pomponi SA, Wörheide G, Erpenbeck D, Cárdenas P (2017) Evolution of group I introns in Porifera: new evidence for intron mobility and implications for DNA barcoding. BMC Evol Biol 17:82. https://doi.org/10.1186/s12862-017-0928-9

Shivji MS, Stanhope MJ, Rogers SO (1997) Group I introns ("spintrons") are present in shark nuclear ribosomal DNA internal transcribed spacers. Abstract. American Elasmobranch Society, 13th Annual Meeting, 26 June–2 July 1997, Seattle, Washington

Simon D, Moline J, Helms G, Friedl T, Bhattacharya D (2005) Divergent histories of rDNA Group I introns in the lichen family Physciaceae. J Mol Evol 60:434–446. https://doi.org/10.1007/s00239-004-0152-2

Suh SO, 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. https://doi.org/10.1007/PL00006493

Thines M (2007) Characterisation and phylogeny of repeated elements giving rise to exceptional length of ITS2 in several downy mildew genera (Peronosporaceae). Fungal Genet Biol 44:199–207. https://doi.org/10.1016/j.fgb.2006.08.002

Torres-Machorro AL, Hernández R, Cevallo AM, López-Villaseñor I (2010) Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? FEMS Microbiol Rev 34:59–86. https://doi.org/10.1111/j.1574-6976.2009.00196.x

Walker NB, Shivji MS, Stanhope M, Rogers SO (1999) Characterization of Group I Intron-Like insertion elements in shark ribosomal DNA spacers. American Elasmobranch Society, 15th Annual Meeting, June 24–30, State College, Pennsylvania State University