Mobile group I introns at nuclear rDNA position L2066 harbor sense and antisense homing endonuclease genes intervened by spliceosomal introns

Kjersti Lian, Betty M. N. Furulund, Anders A. Tveita, Peik Haugen and Steinar D. Johansen

Abstract

Background: Mobile group I introns encode homing endonucleases that confer intron mobility initiated by a double-strand break in the intron-lacking allele at the site of insertion. Nuclear ribosomal DNA of some fungi and protists contain mobile group I introns harboring His-Cys homing endonuclease genes (HEGs). An intriguing question is how protein-coding genes embedded in nuclear ribosomal DNA become expressed. To address this gap of knowledge we analyzed nuclear L2066 group I introns from myxomycetes and ascomycetes.

Results: A total of 34 introns were investigated, including two identified mobile-type introns in myxomycetes with HEGs oriented in sense or antisense directions. Intriguingly, both HEGs are interrupted by spliceosomal introns. The intron in Didymium squamulosum, which harbors an antisense oriented HEG, was investigated in more detail. The group I intron RNA self-splices in vitro, thus generating ligated exons and full-length intron circles. The intron HEG is expressed in vivo in Didymium cells, which involves removal of a 47-nt spliceosomal intron (I-47) and 3′ polyadenylation of the mRNA. The D. squamulosum HEG (lacking the I-47 intron) was over-expressed in E. coli, and the corresponding protein was purified and shown to confer endonuclease activity. The homing endonuclease was shown to cleave an intron-lacking DNA and to produce a pentanucleotide 3′ overhang at the intron insertion site.

Conclusions: The L2066 family of nuclear group I introns all belong to the group IE subclass. The D. squamulosum L2066 intron contains major hallmarks of a true mobile group I intron by encoding a His-Cys homing endonuclease that generates a double-strand break at the DNA insertion site. We propose a potential model to explain how an antisense HEG becomes expressed from a nuclear ribosomal DNA locus.

Keywords: Antisense, Diderma, Didymium, Homing endonuclease, Intron evolution, Mobile DNA, Myxomycete, Ribozyme
intron insertion sites are currently known, equally distributed between the SSU and LSU rRNA genes [3–5]. A common nomenclature of rDNA group I introns, based on the *E. coli* rRNA gene numbering system, has been established [6].

Group I introns encode ribozymes responsible for RNA processing reactions, and occasionally contain endonuclease genes involved in intron homing mobility. A group I ribozyme possesses a characteristic structural fold at the secondary and tertiary levels [7], organized into three structural domains (catalytic, substrate and scaffold) and at least nine paired segments (P1 to P9) [8]. Two subgroups (group IC1 and group IE) have been noted among the nuclear group I introns [2]. The main structural difference between the subgroups is found in the scaffold domain where group IC1 introns have a larger and more complex P5 RNA structure than group IE introns. Detailed structural information is available from the *Tetrahymana* group IC1 intron (Tth.L1925) ribozyme based on RNA crystallization and cryo-EM [9–11], and a computer-based model reports three-dimensional structural features of the *Didymium* group IE intron (Dir.S956–1) ribozyme [12].

Some nuclear group I introns are capable of self-splicing in vitro as naked RNA following a two-step trans splicing reaction [2, 13]. The self-splicing reaction, which is catalysed by the group I ribozyme and dependent on an exogenous guanosine cofactor (exoG), leads to perfectly ligated exon sequences and excised linear intron RNA. In an alternative and competing reaction the intron RNA may self-process independently of exoG, leading to full-length intron circles and fragmented RNA exons [2, 14].

About 5–10% of all known nuclear group I introns carry homing endonuclease genes (HEGs), and thus appear as potential mobile genetic elements [2]. Experimental support for homing mobility has been reported in two nuclear group I intron systems, Ppo.L1925 in *Physarum polycephalum* [15] and Dir.S956–1 in *Didymium iridis* [16]. Here, intron homing occurs in sexual crosses between intron-containing and intron-lacking strains. Homing is initiated by a double-strand break close to, or at, the intron insertion site, performed by the intron encoded homing endonuclease. The event is completed by homology-dependent repair that results in a highly efficient unidirectional transfer of the intron [2, 17]. All nuclear homing endonucleases belong to the His-Cys family [17–19], which act as zinc-coordinating homodimers [20] and generate tetra- or pentanucleotide 3’-overhangs at their cleavage sites [21–24].

An intriguing question is how the endonuclease proteins become expressed in an rDNA context. Current knowledge suggests that the mRNAs are generated from processed RNA pol I transcripts (sense HEGs) or from a separate RNA pol II transcript (antisense HEGs) [25]. Some mRNAs are 5’-capped by a 2’,5’ lariat catalyzed by a separate ribozyme [26–29], and most appear polyadenylated [24, 29–33]. Furthermore, some HEGs also harbor small spliceosomal introns that need to be removed [30, 32, 33] and thus may participate in homing endonuclease expression. One intriguing and unsolved question remains. How are HEGs that are organized on the antisense strand expressed? Expression of such HEGs has the potential to cause serious problems for normal cell growth since their mRNAs may induce antisense interference to precursor rRNA transcripts. The *D. iridis* intron Dir.S956–2 was reported to carry and express an antisense HEG from a potential internal promoter, and its mRNA maturation includes both polyadenylation and spliceosomal intron removal [24, 34].

Here we characterized 34 nuclear L2066 group I introns from myxomycete and ascomycete rDNA and identified two differently organized mobile-type introns (Dal.L2066 and Dsq.L2066) in *Diderma alpinum* and *D. squamulosum* with HEGs in sense and antisense orientation, respectively. A common feature to both HEGs was consensus polyadenylation signals and the presence of spliceosomal introns. The antisense HEG intron Dsq.L2066 was investigated in more detail including in vitro self-splicing activity, in vivo HEG expression, as well as homing endonuclease activity and cleavage site mapping.

**Results**

**Structural features of group I introns at position L2066 in the nuclear LSU rRNA gene**

Introns at site L2066 site in LSU rRNA (*E. coli* LSU rRNA numbering [6]) disrupt helix 74 in domain V, which is in proximity to the peptidyl transferase center. L2066 group I introns have so far been noted in two orders of myxomycete protists and three orders of ascomycete fungi (Table 1). Consensus structure diagrams of myxomycete (16 taxa) and ascomycete (18 taxa) L2066 group I intron RNA are shown in Fig. 1 and Fig. S1, which highlight conserved core structures and more variable peripheral regions. The introns fold into a typical group IE ribozyme organization including a less complex P4-P6 scaffold domain, a core-stabilizing P13 helix, and strong exon-binding segments (P1 and P10).

Structural conservation and variation among introns were further assessed by a phylogenetic analysis based on 186 aligned core sequence positions (Fig. S2) representing all the 34 taxa. A Neighbor Joining analysis is presented in Fig. 2 and shows that all ascomycete L2066 introns cluster together with high bootstrap value, but relationships among myxomycete introns appeared more scattered and not strictly organized according to current taxonomy.
Reconstructing group I intron in long-term phylogeny appears challenging due to a limited number of aligned positions, and to biological factors such as horizontal transfers and intron gain-and-loss. The L2066 introns in ascomycetes were uniform in size varying only from 287 bp to 400 bp (Table 1). Size variation was more pronounced among the myxomycete L2066 introns. All introns vary in length between 400 bp and 500 bp, except two introns that are significantly longer, i.e., 1480 bp in D. alpinum It-K56 and 1197 bp in D. squamulosum CR10 (Table 1).
**D. alpinum** and **D. squamulosum** L2066 group I introns harbor HEGs in sense or antisense orientations that are interrupted by small spliceosomal introns

Intron phylogeny (Fig. 2) and secondary structure diagram (Fig. 3a) show that the *D. alpinum* isolates It-K56 and Uk-K78 possess identical catalytic RNA core sequences. Interestingly, they only differ due to an insertion of 1013 nt into P9 of the Italian isolate (i.e., It-K56). A closer inspection of the P9 insertion identified a HEG, putatively encoding a homing endonuclease protein consisting of 235 amino acids (named I-Dali) and belonging to the His-Cys family (see Fig. 4c). In addition, the HEG is associated with a consensus polyadenylation signal (AAUAAA). Finally, the *D. alpinum* HEG was found to be interrupted by a 49-nt spliceosomal intron (I-49) in proximal distance to the 5′ end of the reading frame (Figs. 3b and 4b).

The L2066 intron in *D. squamulosum* (named Dsq. L2066) has a more unusual structural organization...
Specifically, the catalytic RNA core corresponds to a canonical group IE type structure, but the peripheral paired segment P9.2 contains a large HEG-encoding insertion. Interestingly, the HEG is positioned in an anti-sense orientation and is apparently transcribed from the opposite rDNA strand compared to that of the rRNA genes and intron ribozyme. Moreover, the HEG has recognizable start and stop codons, two consensus polyadenylation signals in proximity to the stop codon, and a 47-bp spliceosomal intron (I-47) closer to the 5' end (Fig. 4b). If the I-47 is removed, then the HEG encodes for a putative homing endonuclease (named I-DsqI) consisting of 204 amino acids and belonging to the His-Cys family (Fig. 4c).

Fig. 2 Molecular phylogeny of myxomycete and ascomycete L2066 group I introns. The intron topology was obtained by Neighbor-joining (NJ) analysis of 34 taxa and 186 nt aligned positions. The Didymium iridis intron Dir.S956–1 was included as an outgroup. The NJ and Minimal Evolution (ME) bootstrap replicates (≥ 50%) are given for each node (NJ/ME). Maximum support in NJ and ME (≥ 97%) are shown in bold branches. Black dots at branch points indicate NJ and ME robustness (≥ 90%). The scale-bar indicates the fraction of substitutions per site. GenBank accession numbers for taxa are shown.
Dsq.L2066 self-splices and generates full-length intron RNA circles in vitro

RNA self-splicing and processing of Dsq.L2066 was evaluated from in vitro transcribed and purified RNA, which was subsequently incubated at splicing conditions for 60 min. Ligated exons and intron circles were assessed by eluting the RNA from a polyacrylamide gel, and by subjecting the RNA to RT-PCR using specific primers (Fig. 5a). Sanger sequencing of amplified fragments confirmed self-splicing by identifying the RNA corresponding to ligated exons (Fig. 5b, left panel). Furthermore, an RNA band corresponding to full-length intron RNA circles was identified (Fig. 5b, right panel). Based on these finding we infer that Dsq.L2066 is capable of processing and self-splicing the RNA using the same pathways as previously reported for the *D. iridis* S956–1 group IE intron [14, 35, 36].

I-DSql homing endonuclease is expressed from the rDNA antisense strand in vivo

To examine I-DSql expression in vivo, total RNA from *D. squamulosum* amoeba was isolated and subjected to RT-PCR and Sanger sequencing. We first addressed if the small spliceosomal intron (I-47) was present or not in cellular RNA (Fig. 6a). Two amplicons of expected sizes were observed, corresponding to intron lacking and intron containing RNA (Fig. 6b). Polyadenylation of I-DSql mRNA was assessed using a poly(T) specific primer with a unique sequence tag at the 5’ end in the first-strand synthesis reaction. Polyadenylated mRNA was then
amplified from a primer-set consisting of an upstream primer and a downstream primer, the latter complementary to the poly(T) primer tag (Fig. 6a). The resulting amplicons were separated on an agarose gel, gel purified, and Sanger sequenced. Two polyadenylation sites were detected that correspond to the polyadenylation signal I

Fig. 4 Structural organization of L2066 group I intron in Didymium squamulosum. a Secondary structure diagram of the Costa Rican (CR10) isolate. Antisense homing endonuclease gene (HEG) insertion is indicated. See legend to Fig. 1a for structural annotations. b Schematic presentation of HEG insertion in P9. Start and stop codons are indicated, as well as the 47 nt spliceosomal intron (I-47). Two AAUAAA polyadenylation signals are located close to the termination codon. Below is an alignment of selected spliceosomal introns in group I intron HEGs. I-47, Didymium squamulosum L2066; I-49, Diderma alpinum L2066; I-50, Didymium iridis S956–2; I-51, Didymium iridis S956–1. Yeast spliceosomal consensus sequences at branch and splice sites are indicated above the alignment. (*) indicate hallmark spliceosomal intron positions (GU-A-AG). c Schematic presentation of the I-Dsq I homing endonuclease containing a His-Cys box. Below is an amino acid alignment of His-Cys box features in Didymium squamulosum (I-Dsq), Diderma alpinum (I-Dal), Didymium iridis (I-Dir), Naegleria jamiesoni (I-Nja) and Physarum polycephalum (I-Ppo). I-Nja and I-Ppo represent well-studied His-Cys homing endonucleases. Conserved residues (boxed) corresponding to those presented in the I-Ppo crystal structure [20]. C100, C105, H110 and N119 are involved in zinc binding, and H98 and N119 are associated with the active site
and II located within the 3′ untranslated region of I-DsqI mRNA (Fig. 6c). These experiments support that I-DsqI mRNA formation involves removal of the spliceosomal intron and polyadenylation at two alternative sites.

I-DsqI is an active endonuclease that cleaves the intron-lacking rDNA locus

To heterologously express the I-DsqI homing endonuclease in E. coli, the HEG (lacking I-47) was first PCR amplified from cDNA originating from D. squamulosum cells, then subcloned into the pTH1 expression vector with an N-terminal malE fusion that encodes the maltose binding protein (MBP). MBP-I-DsqI fusion protein was over-expressed after IPTG induction in E. coli. Figure 7a shows an SDS-PAGE gel with protein from samples harvested every 30 min after IPTG induction. A band corresponding to the size of MBP-I-DsqI (approx. 68 kDa) increases in intensity in induced cells. Figure 7b shows total proteins from lysed cells, and from pelleted material after centrifugation. The presence of the expected fusion protein in both fractions indicates that the MBP-I-DsqI fusion is partially soluble under the conditions used. The soluble protein phase was next subjected to affinity purification using 5 ml amylose resin. Figure 7c shows an SDS-PAGE gel with proteins from the collected fractions 1–9 after

**Fig. 5**  In vitro self-splicing of Dsq.L2066 RNA. **a** The two main processing pathways, splicing and full-length intron circularization (FLC), and approach of RT-PCR/ Sanger sequencing assessments. The self-splicing pathway (left) involves two transesterification reactions. The first reaction (1) is initiated by a nucleophilic attack by an exogenous guanosine cofactor (exoG). The second reaction (2) starts with an attack at the 3′ splice site (3′SS) and results in ligated exons and excised linear intron. The FLC pathway (right) involves hydrolysis at the 3′SS (3) followed by a nucleophilic attack (4) at the 5′SS by the terminal guanosine (ωG). The FLC pathway is independent of exoG and results in full-length intron circles and fragmented exons. **b** Experiments of ligated exons (left) and full-length intron circle (right) junction analyses. RNA sequences corresponding to ligated exon junction and ligated FLC junction are presented below.
addition of maltose. A strong band corresponding to the size of MBP-I-DsqI can be seen in fractions 4 and 5. To test for DNA endonuclease activity, aliquots of fractions 1–9 were incubated at 37 °C with linear target DNA containing the L2066 intron-lacking rDNA allele. After incubation and gel analysis, specific cleavages of target DNA were observed that corresponded to the expected fragment sizes, given that the enzyme cleaves the rDNA near the intron insertion site (Fig. 7d). Optimal temperature for cleavage was assessed over a broad
temperature range of 0°C to 60°C using linear target DNA, 60 min incubation, and 0.1 unit endonuclease (see Materials & Methods). Purified I-DsqI was found active over a broad range of temperatures, from 5°C to 40°C (Fig. 7e), which is similar to that found for the Naegleria I-NjaI homing endonuclease [23].

---

**Fig. 7** (See legend on previous page.)
I-DsqI generates a pentanucleotide 3′-overhang at the L2066 intron insertion site

The precise cleavage site for I-DsqI was determined by primer extension analysis performed at both strands on cleaved target DNA. Sanger sequencing reactions on un-cleaved target DNA using the same primers were run in parallel. Determination of the lower strand cleavage-site (Fig. 8, left panel) indicated cleavage 3′ of a C-residue exactly at the intron insertion site. Similarly, the cleavage site of upper strand was found to be 3′ of an A-residue at position +5 (Fig. 8, right panel). We conclude that I-DsqI generates a five-nucleotide 3′ staggered cut at the L2066 intron insertion site, but with no detectable sequence symmetry (Fig. 8, lower panel).

Discussion

We report analyses of the L2066 family of group I introns based on a dataset of 34 introns representing ascomycete and myxomycete taxa. All introns belong to the group IE subtype and possess highly conserved catalytic RNA cores. Introns from two myxomycete taxa, D. alpinum It-K56 and D. squamulosum CR10, contain large HEG insertions in the RNA segment P9. Both HEGs contain consensus polyadenylation signals in proximity to the termination codons and are interrupted by small spliceosomal introns closer to the initiation codons. Interestingly, the D. squamulosum HEG is located on the opposite (antisense) strand compared to its corresponding self-splicing ribozyme and host rRNA gene and appears to be expressed in myxomycete amoeba. Over-expression in E. coli produces an active I-DsqI homing endonuclease, and primer extension analysis maps a double stranded break with a pentanucleotide 3′-overhang at the intron DNA insertion site.

HEGs in D. alpinum and D. squamulosum were found to contain spliceosomal introns of 49 bp and 47 bp, respectively. Similar spliceosomal introns have been reported in some, but not all, nuclear group I intron HEGs. These include the S943 group I intron in an Ericoid mycorrhizal fungus [37], the S1389 group I introns of Trichiales myxomycetes [32], the S516 introns of Stemonitales and Trichiales myxomycetes [33], and the S956 group I introns in two isolates of the Physarales myxomycete D. iridis (Fig. 4b) [24, 30, 34]. HEG spliceosomal introns may facilitate gene expression, mRNA stability, or mRNA nuclear to cytoplasmic translocation by recruiting spliceosomal components and exon junction complexes [24, 38, 39]. Furthermore, spliceosomal intron sequences may also be involved in additional processes. Recently we suggested that the I-51 spliceosomal intron in D. iridis interacts with the lariat capping ribozyme through strong base-pairings, and thus participate in 5′-end modification of the homing endonuclease mRNA [40].

The nuclear rDNA locus is dedicated to high-level transcription of rRNA genes, but expression of genes coding for proteins embedded in rRNA is not unprecedented. Reverse transcriptases and intron endonucleases are known to be expressed from rDNA harboring non-LTR retrotransposons and group I introns, respectively, in many Arthropoda species and eukaryotic microorganisms [25, 41]. Polyadenylation of mRNA is an indication of mRNA gene expression, and we observed consensus polyadenylation signals in proximity to HEG stop codons in both D. alpinum and D. squamulosum. Interestingly, D. squamulosum contains two polyadenylation signals, which are both used for polyadenylation of the mRNA in vivo. Similar polyadenylation signals, and subsequent polyadenylation, have been noted in several group I intron homing endonuclease mRNA in amoeboid flagellates [29, 31] and myxomycetes [24, 30, 32, 33].

Expression of the D. squamulosum HEG can potentially create antisense interference between the HEG mRNA and the rRNA in cells. The essential role of rRNAs in ribosomes (and thus in cell growth) can
therefore represent a serious hazard for growing cells. Based on observations presented here and support by a previous study in *D. iridis* [24] we propose a potential model for spatial and probably temporal separation of transcripts that involves different promoters and polymerases (Fig. 9). 1) The rDNA in *Didymium* and other myxomycetes appear as multicopy extrachromosomal mini-chromosomes [16, 42–44]. Thus, copies of the extrachromosomal rDNA in *D. squamulosum* may be located both in the nucleolus and nucleoplasm. 2) The nucleolar rDNA will express rRNA genes from a regular RNA pol I promoter, and subsequently the L2066 intron become excised during ribozyme-catalyzed RNA processing. 3) The nucleoplasm rDNA will express the I-DsqI HEG from a corresponding RNA pol II promoter located on the opposite strand, generating an I-47 lacking and polyadenylated mRNA. 4) Since the RNA pol I and RNA pol II transcriptions are mainly restricted to the nucleolus and nucleoplasm, respectively, in eukaryotic cells, antisense interference between the pre-rRNA and I-DsqI mRNA becomes minimized or avoided.

Homing of a nuclear rDNA group I intron has been reported in experimental settings in two myxomycete species, *P. polycephalum* and *D. iridis* [15, 16]. Homing is dependent on sexual mating and is initiated by a DNA double strand break at the intron insertion site in an intron-lacking allele, catalyzed by the intron-encoded endonuclease. I-DsqI contains strong hallmarks of a His-Cys homing endonucleases, including conserved amino acid residues in zinc coordination and catalysis, and the ability to cleave an intron-lacking allele DNA at the insertion site. I-DsqI was found to generate a pentanucleotide 3’-overhang at the cleavage site. This was similar to that of *Naegleria* I-Njal, I-NitI, and I-NanI homing endonucleases [22, 23] and *Didymium* I-DirI [24], but different from the tetranucleotide 3’-overhang made by I-Ppol from Physarum [20, 21].

Fig. 9 Potential model for antisense HEG expression in *Didymium* cells. Extrachromosomal rDNA copies are located both in nucleolus and nucleoplasm. The nucleolar rDNA (left) generates an RNA pol I transcript covering the SSU and LSU rRNA genes, and group I intron Dsq.L2066. Intron RNA self-splicing and processing results in ligated exons, excised intron, and full-length intron circles. The nucleoplasm rDNA (right) generates an RNA pol II transcript covering the HEG. Maturation of the corresponding I-DsqI mRNA involves removal of spliceosomal intron (I-47) and polyadenylation. These two processes are separated in space and probably time, avoiding antisense interference between pre-rRNA transcripts and mRNA.
Conclusions
Thirty-four nuclear group IE introns at position L2066 from ascomycete and myxomycete rDNAs were analyzed, and mobile-type introns were found in two of the myxomycete taxa; D. alpinum It-K56 and D. squamulosum CR10. Both introns harbor His-Cys HEGs interrupted by spliceosomal introns. In vivo expression of the antisense-organized HEG in D. squamulosum was supported by mRNA polyadenylation and the removal of a 47-nt spliceosomal intron. The corresponding endonuclease protein, I-Dsq1, was over-expressed in E. coli and showed DNA cleavage activity corresponding to a pentanucleotide 3' overhang at the intron-lacking allele. These features are consistent with group I intron mobility. Finally, we propose a potential model to explain how an antisense HEG from a mobile group I intron become expressed from a nuclear rDNA locus due to spatial and probably temporal separation of rRNA and mRNA transcriptions.

Materials and methods
Myxomycete strain, nucleic acid isolation, amplifications, and rDNA sequencing
D. squamulosum was generously provided by Dr. Jim Clark (University of Kentucky). Approximately 10^8 amoeba cells were harvested from DS/2 agar plates, and total DNA extraction and rDNA sequencing were performed as described previously [45]. Extraction of total cellular RNA, PCR, RT-PCR and Sanger sequencing were performed as described previously [24, 29, 46]. Primer information: OP1019, 5'-TCA GAG CGT TGG CTG GTA GC-3'; OP1020, 5'-AAT CTT GAG TTC GTG GTA C-3'; OP1021, 5'-ACG TCT ACT TTG CGA GC-3'; OP313, 5'-AAG CG AGCG ATG CACGCA TT-3'; OP41, 5'-CGA GCAG ATG CAC GCA TTT TTT TTT TTT TTT-3'.

Phylogenetic analysis
The tree-building methods of Neighbor joining (NJ) and Minimum Evolution (ME) interpreted in Geneious prime® 2019.2.3. and MEGA X [47] with default settings. The evolutionary history of intron dataset 1 was reconstructed with NJ and the Jukes-Cantor (JC) model using MEGA X with default settings. The robustness of the nodes of the tree was tested with NJ-JC (500 replicates) and ME-JC (500 replicates).

In vitro transcription and group I intron splicing
Dsq.L2066 with some flanking exon sequences (approximately 50 nt of the 5' exon and 150 nt of the 3' exon), were PCR amplified from LSU rDNA and inserted into pGEM-T easy vector (Promega) downstream the T7 promoter.

Intron containing RNAs were transcribed from linearized plasmid using T7 RNA polymerase (Stratagene, La Jolla, CA, USA) and RNA species of interest were eluted after PAGE separation and submitted to RT-PCR Sanger sequencing as described previously [29, 48]. Primer information: OP808, 5'-AAT TTA ATA CGA CTC ACT AT A GGG CTT GGC ACA AT T AGC GG-3'; OP809, 5'-GTT AGT TAC AGC TTT ATT AGC GG-3'.

Plasmid construction
The I-Dsq1 HEG, containing or lacking the I-47, was introduced into expression vector pTH1 (Addgene). pTH1 generates N-terminal maltose binding protein (MBP) fusions. Briefly, the I-Dsq1 HEG was PCR amplified using cDNA made from total RNA from D. squamulosum and the first-strand oligonucleotide primer OP41, using the primer pair OP1422 and OP1423. The resulting entry clone pDONR221-I-Dsq1 was used to introduce the HEG into the pTH1 expression vector in an LR reaction. Final construct for N-terminal MBP fusion was named pl-Dsq1-MBP. Target DNA for homing endonuclease cleavage was made by insertion of a 1.59 kb fragment of a L2066-lacking Didymium LSU rDNA (amplified by OP122 and OP747) into the pGEM-T easy vector. The plasmid was linearized with Ndel restriction enzyme (New England Biolabs) prior to activity studies. Primer information: OP122, 5'-CGC GCA TGA ATG GAT TA-3'; OP747, 5'-TCC AAC ACT TAC TGA ATT CT-3'; OP1422, 5'-GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GAA CAA CTA CCA GCA GCC A-3'; OP1423, 5'-GGG GAC CAC TTT GTA CAA AAA AGC AGG TGG GTC CTA CCA GCA TGC TGG GTG GTG TT-3'.

Protein expression
For expression of fusion proteins pl-Dsq1-MBP was introduced into E. coli BL21-CodonPlus (DE3)-RIL cells (Stratagene). Cells were cultured in YT-medium with 100 μg/ml ampicillin and 100 μg/ml chloramphenicol. Cultures were grown to OD_{600 nm} 0.4–0.6, then split into two halves, and expression of fusion proteins was induced in one half by adding Isopropyl-1-thio-β-D-galactopyranoside (IPTG) or 20% L-arabinose, respectively. The one half of the culture with no added arabinose was grown in parallel and used as negative control. Samples of 0.5 ml were harvested every 30 min and subjected to SDS-PAGE to monitor expression of proteins. 90 min after induction the cells were harvested by centrifugation for 20 min at 4000 g (5000 rpm in a Sorwall GSA rotor, Dupont Instruments). The cells were resuspended in 5 ml column buffer and stored over night at −20°C. Next day samples were thawed.
in ice-cold water before they were sonicated in six short pulses of 10 sec with 20-sec pause while kept on ice. After centrifugation for 20 min at 4°C at 14000 g (11,000 rpm in a Sorwall SS-34 rotor, Dupont Instruments) the supernatant (crude extract) was stored on ice. The pellet was resuspended in column buffer (insoluble matter). Samples of supernatant and resuspended pellet were subjected to SDS-PAGE to monitor solubility of fusion proteins.

**Affinity purification of proteins**

I-DSqI-MBP fusion proteins were affinity purified by running crude cell extract through a 16 × 100 mm column packed with 5 ml amylose resin, with the flow rate set at 0.5 ml/min. The column was subsequently washed with 5 column volumes of distilled water and 10 column volumes column buffer. I-DSqI-MBP was eluted by before 5 ml crude extract was added. To release the fusion protein from the matrix column buffer containing 10 mM maltose was added to the column and 1.5 ml fractions were collected. Samples from each fraction were subjected to 10% SDS-PAGE and OD_{280 nm} measurements to monitor protein purification.

**In vitro homing endonuclease cleavage assays and cleavage site mapping**

I-DSqI endonuclease activity was assessed by incubating 1 μg linearized target DNA in a 20 μl reaction containing 10 x Tris Acetate-EDTA (TAE) buffer (1 x TAE buffer with 40 mM Tris-acetate and 1 mM EDTA at pH8.3), 0.1 unit of enzyme and 60 min incubation at 37 °C, and separated on 0.7% agarose gels after incubation. One unit of I-DSqI was defined as the amount of enzyme required to completely digest 1 μg linearized target DNA in 1 h at 37°C in a 20 μl reaction containing 10 x TAE buffer. Cleavage site mapping was performed essentially as described previously [22, 23] based on primer extension analyses on I-DSqI cleaved linearized target DNA using OP122 (located ca 105 nt upstream cleavage site) and OP1451 (located ca 87 nt downstream cleavage site) and α-32P dATP as the label. Primer extension products were separated on a 5% polyacrylamide gel alongside main Sanger sequencing reactions generated from un-cleaved target DNA using the same primers. Primer information: OP1451, 5’-GGT GGT ATT TCA AGG TCG-3’.

**Abbreviations**

exoG: Exogenous guanosine cofactor; FLC: Full length circularization; HEG: Homing endonuclease gene; LSU rRNA: Large subunit ribosomal RNA; rDNA: Ribosomal DNA; SS: Splice site; SSU rRNA: Small subunit ribosomal RNA.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s13100-022-00280-4.

**Additional file 1: Figure S1.** Consensus secondary structure diagrams of L2066 group I introns in myxomycetes and ascomycetes. a) Consensus structure in myxomycetes based on ca 250 nucleotide positions in the catalytic core common among introns from 16 taxa (see Table 1). Sequence size variations are noted in most peripheral regions, and homing endonuclease genes (HEGs) are found as P9 extensions. P1-P10 and P13, paired RNA segments; 5’ SS and 3’ SS, exon-intron splice sites. Invariant nucleotide positions are shown as red uppercase letters. Black uppercase letter, > 90% conservation; lowercase letters, > 50% conservation; filled circles, < 50% conservation. b) Consensus structure in ascomycetes based on ca 260 nucleotide positions in the catalytic core common among introns from 18 taxa (see Table 1).

**Additional file 2: Figure S2.** Sequence alignment of 186 core structure nucleotides of myxomycetes and ascomycetes L2066 group I introns. Secondary structure paired segments (P1-P8) are shown above the alignment, and key segments are colour coded. Intron taxa sequences are indicated by species name abbreviations and GeneBank accession numbers.

**Acknowledgements**

We thank Kari Haugli for technical support in collecting specimens, culturing D. squamulosum, DNA isolation and Sanger sequencing. We also thank Ingrid Skjæveland for intron experiments in the initial stage of this project.

**Authors’ contributions**

K.L., B.M.N.F., PH, and SDJ contributed to the conceptualization and designed the research; S.D.J. wrote the manuscript in collaboration with all authors; K.L. and A.A.T performed plasmid cloning and homing endonuclease expression studies; K.L. performed Sanger sequencing, protein purification and cleavage mapping; A.A.T. performed intron splicing analysis; S.D.J. performed sequence-structure analysis with contributions from B.M.N.F. and PH; Phylogenetic analysis was performed by B.M.N.F. All authors have read and agreed to the published version of the manuscript.

**Funding**

The research received no external funding except general grants from Nord University and UiT-The Arctic University of Norway.

**Availability of data and materials**

Sequencing data are available in GenBank under the accession numbers ON155995 (Lepidoderma alpestroides), ON155996 (Comatricha laxa), and ON155997 (Fusarium sp.).

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1 Nofima AS, Muninbakken 9-13, Breivika, 9291 Tromsø, Norway. 2 Genomics division, Faculty of Biosciences and Aquaculture, Nord University, N-8049 Bodø, Norway. 3 Medical Department, Barun Hospital; Vestre Viken Hospital Trust, Drammen, Norway. 4 Department of Chemistry and Center for Bioinformatics, Faculty of Science and Technology, UiT-The Arctic University of Norway, N-9037 Tromsø, Norway.

Received: 8 April 2022 Accepted: 28 September 2022

Published online: 08 October 2022
References

1. Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. Trends Genet. 2005;21:111–9. https://doi.org/10.1016/j.tig.2004.12.007.

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

3. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D’Souza LM, Du Y, et al. The comparative RNA web (CWRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinform. 2002;3:2. https://doi.org/10.1186/1471-2105-3-2.

4. Jackson SA, Cannone JJ, Lee JC, Gutell RR, Woodson SA. Distribution of rRNA introns in the three-dimensional structure of the ribosome. J Biol Mol. 2002;233:35–52. https://doi.org/10.1016/S0222-2836(02)00895-1.

5. Nielsen H, Johansen SD. Group I introns: moving in new directions. RNA Biol. 2009;6:375–83. https://doi.org/10.4161/rna.6.4.9534.

6. Johansen S, Haugen P. A new nomenclature of group I introns in ribosomal DNA. RNA. 2001;7:935–6. https://doi.org/10.1017/S1355838201015001.

7. Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. J Mol Biol. 2011;414:2105–32.

8. Johansen S, Haugen P, Reeb V, Lutzoni F, Bhattacharya D. The evolution of homing intron HEG from the antisense strand of Didymium ribosomal DNA. RNA Biol. 2003;1:157–62. https://doi.org/10.4161/rna.1.3.43958.

9. Johansen SD, Haugen P, Nielsen H. Expression of protein-coding genes embedded in ribosomal DNA. Biol Chem. 2007;388:679–86. https://doi.org/10.1515/BC.2007.089.

10. Nielsen H, Westhof E, Johansen S. An mRNA is capped by a 2',5' lariat catalyzed by a group I-like ribozyme. Science. 2005;309:1584–7. https://doi.org/10.1126/science.1113645.

11. Tang Y, Nielsen H, Briggsdotter AB, Johansen SD. A natural fast-cleaving branching ribozyme from the amoeboid flagellate Naegleria pringshei. RNA Biol. 2011;8:997–1004. https://doi.org/10.4161/rna.8.6.16027.

12. Meyer M, Nielsen H, Olieric V, Roblin P, Johansen SD, Westhof E, et al. Specialization of a group I intron into a lariat capping ribozyme. Proc Natl Acad Sci U S A. 2014;111:7659–64. https://doi.org/10.1073/pnas.1322248111.

13. Tang Y, Nielsen H, Masqueda B, Gardner PP, Johansen SD. Molecular characterization of a new member of the lariat capping twin-ribozyme introns. Mob DNA. 2014;5:25. https://doi.org/10.1186/1759-8753-5-25.

14. Vanden Berghe SH, Gutell RR. Representation of the secondary and tertiary structure group of I introns. Nature Struct Biol. 2001;8:450–6. https://doi.org/10.1038/8792.

15. Johansen S, Haugen P, Elde M, Willassen NP, Johansen SD. I-DirI, a nuclear intron-encoding I-DirI homing endonuclease involves the removal of a splicingosine complex. EMBO J. 1999;18:1003–13. https://doi.org/10.1093/embob/18.4.1003.

16. Haugen P, De Jongeheere JF, Johansen S. Characterization of the self-splicing products of two complex Naegleria LSU rRNA group I introns containing homing endonuclease genes. Eur J Biochem. 2002;269:1641–9. https://doi.org/10.1046/j.1432-1327.2002.02802.x.

17. Furulund BMN, Karlsen BO, Babjak I, Johansen SD. A phylogenetic approach to structural variation in organization of nuclear group I introns and their ribozymes. Non-coding RNA. 2021;7:43. https://doi.org/10.3390/ncrna7030043.

18. Johansen S, VágØM VM. An intron in the nuclear ribosomal DNA of Didymium irdis codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. Cell. 1994;76:725–34. https://doi.org/10.1016/0092-8674(94)90111-8.

19. Decatur WA, Einvik C, Johansen S, Krøgh M, Babjak I. Two group I ribozymes with different functions in a nuclear rDNA intron. EMBO J. 1995;14:4558–68. https://doi.org/10.1002/j.1460-2075.1995.tb03135.x.

20. Haugen P, Reeb V, Lutzoni F, Bhattacharya D. The evolution of homing endonuclease genes and group I introns in nuclear rDNA. Mol Biol Evol. 2004;21:129–40. https://doi.org/10.1093/molbev/msj005.

21. Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3:a003707. https://doi.org/10.1101/cshperspect.a003707.

22. Schlautmann LP, Gehring NH. A day in the life of the exon junction complex. Biomolecules. 2020;10:866. https://doi.org/10.3390/biom10060866.

23. Nielsen H, Krøgh M, Masqueda B, Johansen SD. The lariat capping ribozyme. In: Muller S, Masquida B, Winkler W, editors. Ribozymes: WILEY-VCH GmbH; 2021. p. pp118–42. ISBN: 978–3–527–34454-3.

24. Johansen SD, VágØM VM. Two group I ribozymes with different functions in a nuclear rDNA intron. EMBO J. 1995;14:4558–68. https://doi.org/10.1002/j.1460-2075.1995.tb03135.x.

25. Hafner M, Auget D. Splicing products of a group I intron in the nuclear DNA of Didymium irdis. EMBO J. 1999;18:1003–13. https://doi.org/10.1093/embob/18.4.1003.

26. Hafner M, Auget D. Splicing products of a group I intron in the nuclear DNA of Didymium irdis. EMBO J. 1999;18:1003–13. https://doi.org/10.1093/embob/18.4.1003.

27. Hafner M, Auget D. Splicing products of a group I intron in the nuclear DNA of Didymium irdis. EMBO J. 1999;18:1003–13. https://doi.org/10.1093/embob/18.4.1003.
46. Wikmark OG, Einvik C, De Jonckheere JF, Johansen SD. Short-term sequence evolution and vertical inheritance of the Naegleria twin-ribozyme group I intron. J BMC Evol Biol. 2006;6:39. https://doi.org/10.1186/1471-2148-6-39.

47. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9. https://doi.org/10.1093/molbev/msy096.

48. Lundblad EW, Einvik C, Rønning S, Haugli K, Johansen S. Twelve group I introns in the same pre-rRNA transcript of the myxomycete Fuligo septica: RNA processing and evolution. Mol Biol Evol. 2004;21:1283–93. https://doi.org/10.1093/molbev/msh126.

**Publisher’s Note**
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.