

Molecular evolution of catalytic RNA introns in nuclear ribosomal DNA from eukaryotic microorganisms

Betty Martine Normann Furulund

FACULTY OF BIOSCIENCES AND AQUACULTURE

Molecular evolution of catalytic RNA introns in nuclear ribosomal DNA from eukaryotic microorganisms

Betty Martine Normann Furulund

A thesis for the degree of
Philosophiae Doctor (PhD)

PhD in Biosciences no. 62 (2024)
Faculty of Biosciences and Aquaculture

PhD in Biosciences no. 62 (2024)

Betty Martine Normann Furulund

Molecular evolution of catalytic RNA introns in nuclear ribosomal DNA from eukaryotic microorganisms

© Betty Martine Normann Furulund

ISBN: 978-82-93165-61-3

Print: Trykkeriet NORD

Nord University

N-8049 Bodø

Tel: +47 75 51 72 00

www.nord.no

All rights reserved.

No part of this book may be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying or otherwise, without the prior written permission from Nord University.

Preface

This doctorate thesis is submitted in fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) at the Faculty of Bioscience and Aquaculture (FBA), Nord University, Bodø, Norway. The studies included in this dissertation represent original research as a part of the Stipendiat program.

The PhD project team consists of the following members:

Betty Martine Normann Furulund, MSc: PhD candidate

Steinar Daae Johansen, Professor, FBA, Nord University: primary supervisor

Bård Ove Karlsen, Researcher, Nordlandssykehuset: co-supervisor

Igor Szczepan Babiak, Professor, FBA, Nord University: co-supervisor

Acknowledgements

This PhD at Nord University has been an experience I will never forget. I can compare it with an alpine climb where you go through several mental stages while struggling through climbing pitch after climbing pitch, which ends up with “cruxes” you never thought you could manage. Then end up on the top of the mountain with a lot of joy and happiness. Like the climbing trips I have done, I have never been alone, even though I sometimes felt alone. Several people have been involved during my PhD period, and they have all earned my gratitude, but mentioning all here would take a whole new thesis to write.

First, I would like to express my deep gratitude to my main supervisor, Steinar Daae Johansen. Thank you for giving me this chance to become part of the group I intron world, for your inspiring guidance, patience, valuable suggestions, and support.

Also, a like to thank my co-supervisors, Bård Ove Karlsen and Igor Szczepan Babiak, for your support. Especially, Bård, thank you for your tricking me into this “Climb”, for your inspiration, encouragement, and our scientific conversations.

I would also like special thanks to Joost Raeymaekers for including me to your scientific meetings, such as the LGG meetings and writing retreats. But in general, help me with all types of questions.

I am thankful to all my colleagues from FBA. Tor Erik, Martina, and Jeanett thank you for your help and guidance answering all my questions. Next, Lars-Martin and Suzanne, thank you for the interesting conversations, while struggling up the mountain with our skis on and going down in the white fluffy snow.

I want to thank to my co-authors, Peik Haugen, Kristin Lian, and Andres. Thank you all for your invaluable contributions—especially Peik, for your time, and valuable conversations.

Then I also want to thank Åse Emblem for answering all my questions and helping when needed it.

Then, a big thanks to all my amazing friends and fellow PhDs! Especially Amalia, Hirono, Thijse, and Leona, for the writing club, and for all the wonderful time we had during the PhD period. Especially a big thanks to Amalia, for all the amazing times together.

I want to thank my friends and colleagues at Nordlandssykehuset. I want to thank Torild, Margaret, May-Lill, Lisbet, Kristin, Alice, Guro, Ingvild, Vigdis, for all their support.

Next, my other friends the “Supreme Beta givers”. Especially Emmi, for all your talks, support, trips, hikes, and climbs we have done. Then to Espen, Ditte, Elisabeth, and Trine for running, climbing or walking with me, and just in general being there and bringing laughter and smiles into sometimes gray and cloudy days. And for all the other amazing friends and people in Bodø.

Then, my family, especially mom, dad, and my brother. You have always been my solid rope always to grab on when I need it. Then, Lill-Tove and Sture. Thank you for the support and fantastic dinners.

Last, I would like to thank the best “Beta” giver and Belayer, Runar Helskog. For all the conversations about slime mold and ribozymes, when you most wanted to talk about medicine, anesthetics, ultrasound, and fighter jets. Then our Paris trip ended up with a whole day at the Zoo of Paris to gaze upon the mighty slime mold *Physarum Polycephalum* “Le Blòb”. I have now transformed you into a proper slime mold- and Ribozyme- enthusiast, so you can give me critical feedback on my work, project ideas, and discussions on RNA and slime mold research.

Table of contents

Preface	I
Acknowledgements	III
Table of contents	V
List of figures	VII
List of abbreviations	IX
List of papers	XI
Summary	XIII
Sammendrag på norsk	XVII
1 Introduction	1
1.1 Exploring the RNA	1
1.2 Catalytic RNA.....	3
1.3 Group I intron ribozymes	3
1.3.1 A glimpse of the self-splicing activity of group I introns	5
1.4 RNA structure elements of nuclear group I introns.....	7
1.4.1 RNA structure of IE intron subgroup	8
1.4.2 RNA structure of the IC1 intron subgroup	9
1.5 The peripheral RNA segments of nuclear group I introns	12
1.5.1 A brief view of repetitive sequences	12
1.5.2 The homing endonuclease genes - a hallmark of intron mobility	14
1.6 The evolutionary history of nuclear group I introns.....	17
1.6.1 Evolutionary aspects of group I intron mobility by the homing	18
1.7 Molecular phylogenetics to understand the evolutionary relationship of nuclear group I introns.....	20
1.8 Myxomycetes: a host model system in studying the evolutionary history of nuclear group I introns.....	23
1.8.1 Myxomycetes model species for nuclear group I introns	26
2 Aims of study	27
3 Summary of papers	29
4 General discussion	33
4.1 A cyclic evolution model based on nuclear group I introns	33
4.2 Small spliceosomal introns in nuclear group I intron HEGs.....	38

4.3	Complex direct repeat arrays in nuclear group I introns.....	42
4.4	RNA sequence variations in the G-site	48
4.4.1	An exception from the invariant G-site.....	49
5	Future perspectives.....	53
6	Reference list.....	55

List of figures

Figure 1: Major types of RNA.....	2
Figure 2: Nuclear group I intron insertion sites in rDNA..	4
Figure 3: Processing pathways of nuclear group I introns	6
Figure 4: The nuclear group I intron RNA secondary structure	8
Figure 5: Schematic representation of structural motifs of IC1.....	10
Figure 6: Guanosine binding site in the P7.....	11
Figure 7: Repeat motif organization.....	13
Figure 8: Homing.....	15
Figure 9: The conserved His-Cys box present in mobile-type nuclear group I intro.....	16
Figure 10: The cyclic model..	19
Figure 11: Nuclear group I intron introns within myxomycete rDNA	25
Figure 12: Revision of the Goddard-Burt cyclic evolutionary model	37
Figure 13: Small spliceosomal introns.....	38
Figure 14: Homing endonuclease gene..	40
Figure 15: Direct repeat features.	44
Figure 16: Direct repeat features in <i>Tricia varia</i> S516 introns..	46
Figure 17: Direct repeat features in the <i>Licea parasitica</i> S516 intron.	47
Figure 18: The G-site and the variants	49

List of abbreviations

3D	Three-dimensional
BI	Bayesian inference
Cryo-EM	Cryo-electron microscopy
DR	Direct repeats
exoG	Exogenous guanosine
FLC	Full-length circularization
G-site	Guanosine binding site
HE	Homing endonuclease
HEG	Homing endonuclease gene
IGS	Internal guide sequence
Inteins	In-frame protein sequence intron
ITS	Internal transcribed spacers
LCrz	Lariat-capping ribozyme
lncRNA	Long-ncRNA
LSU rDNA	Large subunit ribosomal DNA
m7G	7-methyl guanosine
ML	Maximum likelihood
MP	Maximum parsimony
mRNA	Messenger RNA
miRNA	Micro-RNA
ncRNA	Non-coding RNA
NJ	Neighbor-joining
piRNA	Piwi-RNA (piRNA)
polyA	Polyadenylation
rDNA	Ribosomal DNA
RNP	RNA-protein complexes.
RNAP	RNA polymerase
rRNA	Ribosomal RNA
siRNA	Small-interfering (siRNA)

snoRNA

Small nucleolar RNA

SS

Splice site

SSU rDNA

Small subunit ribosomal DNA

tRNA

Transfer RNA

Ψ

pseudo

ωG

Omega guanosine

List of papers

Paper I

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

Paper II

Furulund BMN, Karlsen BO, Babiak I, Haugen P, Johansen SD. (2022) **Structural organization of S516 group I introns in myxomycetes.** Genes 13: 944. <https://doi.org/10.3390/genes13060944>

Paper III

Lian K, Furulund BMN, Tveita AA, Haugen P, Johansen SD. (2022) **Mobile group I introns at nuclear rDNA position L2066 harbor sense and antisense homing endonuclease genes intervened by spliceosomal introns.** Mobile DNA 13: 23. <https://doi.org/10.1186/s13100-022-00280-4>

Summary

Group I introns, with their unique and complex RNA structures, encode self-splicing ribozymes and are abundant in myxomycetes (plasmodial slime molds). Myxomycetes are eukaryotic microorganisms with more than a billion years of separate evolution and have maintained genetic features that exhibit remarkable evolutionary conservation, which indicating a stable preservation of group I introns. Their distinct life strategy, characterized by possessing millions of nuclei within a single and mobile plasmodium, positions myxomycetes as a great source for nuclear rDNA group I intron sequences.

The aim of this project was to analyze the evolutionary patterns of catalytic RNA introns, focusing on RNA sequence conservation and structural variability. This was conducted through a studying three nuclear groups I intron family datasets (S516, S1389, and L2066), mainly from myxomycetes. The specific objectives included: (1) characterization and investigation of RNA sequence- and structural- variations within the host-dependent S1389 family of group I introns in myxomycetes by molecular phylogenetic assessments; (2) RNA structural characterization of the self-splicing S516 family from myxomycetes by a comparative and phylogenetic approach; (3) investigate the RNA structural and mobility features of the mobile group I self-splicing L2066 intron family by comparative- and biochemical assessment methods; and (4) revise the evolutionary history model of nuclear group I self-splicing introns in myxomycetes.

This PhD thesis overall results showed conservation within the intron families ribozyme cores and evolutionary inheritance, but some unique RNA sequences variations, and structural features were also detected. Our observations resulted in the revision of the Goddard-Burt cyclic evolutionary model, which presents a cycle of intron invasion, periodic functionality of homing endonuclease genes (HEG), HEG interrupted by small spliceosomal introns, HEG degradation, eventual loss of HEG, and then intron loss. Certain RNA sequence- and structural features observed were further discussed.

The first feature observed was the identification of HEG and small spliceosomal introns. 7% of the studied myxomycete group I introns harbored HEG located in peripheral segments. Interestingly, at least 75% of these group I intron HEGs were observed to be interrupted by GT-AG class of spliceosomal introns. The evolutionary origin of these introns is hypothesized based on the possibility for spliceosomal introns to integrate into HEG from their local host nucleus. In general, the spliceosomal introns are known to facilitate the nuclear-cytoplasmic transport of mRNA. Their biological role was hypothesized to facilitate HEG expression from nuclear rDNA by recruiting spliceosomal components and, subsequently, exon junction complexes.

The second feature within the intron families was the repetitive sequences in some of their peripheral segments. 8% of the myxomycete group I introns studied were observed to contain direct repeat sequences, representing the S1389- and S516 families. A phylogenetic analysis supported a sporadic inheritance pattern for the direct repeat motifs in the S1389 intron isolates. However, sequence heterogeneity, variable length, and suggestion of a taxa-specific distribution among the direct repeats motifs from the S516 introns were observed. From our results, interesting direct repeat features were observed in both families: (I) putative internal guide sequences were reported in three S1389 intron isolates. These motifs were suggested to have a regulatory role; (II) a link between truncated HEG and gain of direct repeat motifs in *Trichia varia* S516 intron isolate. This suggested that the HEG deletion has given rise to two direct repeat motifs; and (III) a complex direct repeat motif within the *Licea parasitica* S516 intron folds into structural hairpin motifs.

The third feature was the identification of RNA sequence variations in the normally highly conserved guanosine binding site (G-site) in the S1389 family. Two types of P7 structural variations were observed. The first type was located in the base pair adjacent to the essential G264-C311 pair and was the well-conserved A:U (corresponds to *Tetrahymena* positions A265:U310), and presented the majority of S1389 introns (68%). The second type of P7 structural variants observed is variation at the bulge nucleotide.

Whereas the great majority of introns had the highly conserved A (93%), five introns have either a bulge-U (four isolates) or a -G (one isolate), in this position. It was previously reported that the S1389 family was unable to self-splice *in vitro* at standard reaction conditions. Perhaps the observed G-site variants represent an additional step in escaping the Goddard-Burt cycle of intron evolution by becoming fixed and more host-dependent.

Our results revealed overall conservation within the nuclear group I intron families in structural organization. However, some interesting exceptions were observed, which resulted in a revision of their evolutionary cyclic model. This model revealed some of the complex evolutionary patterns that catalytic RNA introns may undergo. This gives us a broader insight to enhance our understanding of the unique properties and dynamics of nuclear group I introns in myxomycetes, thereby improving the field of ncRNA and the study of intron evolution.

Sammendrag på norsk

Gruppe I-introner er katalytiske RNA-sekvenser med unike og komplekse RNA-strukturer, som koder for selvpleisende ribozymer. Myxomyceter (plasmodiale slimsopper) er eukaryote mikroorganismer med mer enn en milliard år av separat evolusjon. De er en god kilde til nukleær rDNA gruppe I-introner på grunn av deres unike livsstrategi.

Målet med dette PhD-prosjektet var å analysere de evolusjonære mønstrene til katalytiske RNA-introner, med fokus på RNA-sekvensbevaring og strukturell variabilitet. Dette ble utført gjennom en fylogenetisk vurdering av tre datasett, hvert bestående av en samling av nukleære gruppe I-intronfamilier (S516, S1389, og L2066), hovedsakelig fra myxomyceter. De spesifikke målene inkluderte: (1) karakterisering og undersøkelse av RNA-sekvens- og strukturvariasjoner innen den vertsavhengige S1389-familien av gruppe I-introner i myxomyceter ved molekylærfylogenetiske vurderinger; (2) RNA-strukturell karakterisering av den selv spleisende S516-familien fra myxomyceter ved en komparativ og fylogenetisk tilnærming; (3) undersøkelse av RNA-strukturelle og mobilitetsfunksjoner i den mobile gruppe I selvpleisende L2066 intronfamilien ved komparative og biokjemiske tilnærminger; og (4) revidering av evolusjonsmodellen for nukleære gruppe I selv-spleisende introner i myxomyceter.

Våre resultater viste en generell konservering innenfor intronfamilienes ribozymkjerner og deres evolusjonære nedarvingsmønster, men noen unike RNA-sekvensvariasjoner og strukturelle egenskaper ble også observert blant disse familiene. Våre observasjoner førte til en revisjon av Goddard-Burts sykliske evolusjonsmodell, som presenterer en syklus av introninvasjon, periodisk funksjonalitet av homing endonukleasegener (HEG), integrering av små spleisosomale introner inn i HEG, nedbrytning av HEG, eventuelt tap av HEG og deretter introntap. Denne PhD-avhandlingen diskuterer spesifikke temaer basert på våre observasjoner og revisjon av evolusjonshistorien til selvpleisende nukleære gruppe I-introner i myxomyceter.

En av observasjonene var identifikasjonen av HEG og små spleisosomale introner. 7% av myxomycete gruppe I-introner viste seg å inneholde HEG i sine perifere segmenter. Overraskende nok inneholdt 75% av disse HEG-ene i tillegg et lite spleisosomalt intron, som er medlemmer av den store GT-AG-klassen av spleisosomale introner. Deres evolusjonære opprinnelse ble diskutert basert på hypotesen at HEG-spleisosomale intronene kan ha en lokal opprinnelse fra vertskjernen. Videre er spleisosomale introner generelt kjent for å være involvert i prosesseringen av mRNA og gjøre det klart for transport fra kjerne til cytoplasma. En mulig biologisk rolle for disse HEG-intronene er å rekruttere spleisosomale komponenter og deretter ekson-krysskomplekser, og dermed bidra til HEG-uttrykket fra nukleær rDNA.

En annen observasjon innenfor intronfamiliene var de lange RNA sekvensene i de perifere segmentene. Disse sekvensene var direkte repeterte sekvenser og ble identifisert i 8% av våre intron familier. Direkte repeterte sekvenser ble observert i 1389- og S516-familiene, men ikke i L2066-familien. En fylogenetisk analyse støtter et sporadisk arvemønster for de direkte repetisjonsmotivene i S1389-intronisolatene. I tillegg observerte vi sekvensheterogenitet med variable lengder, mulig taksa-spesifikk fordeling blant noen direkte repetisjonsmotiver i S516-intronene. Andre interessante direkte repetisjonsfunksjoner ble observert gjennom de tre studiene: (I) nye interne guide-sekvens motiv ble observert i tre S1389-intron isolater. Vi foreslår en mulig regulerende rolle for disse motivene; (II) en mulig sammenheng mellom et nedbrutt HEG og et direkte repetisjonsmotiv i ett av S516-intronene fra *Trichia varia* isolatene. Dette antyder at HEG-delesjonen har gitt opphav til direkte repetisjonsmotiver; og (III) direkte repetisjonsmotiver i S516-intronet fra *Licea parasitica* viser komplekse direkte repetisjonsmotiver, som foldes til strukturelle hårnålmotiver.

Den siste observasjonen som er diskutert i avhandlingen var identifikasjonen av RNA-sekvensvariasjoner i det vanligvis høyt konserverte guanosisinbindingsstedet (G-stedet) i S1389-familien. To typer P7-strukturelle variasjoner ble observert. Den første typen var lokalisert i baseparet ved siden av det essensielle G264-C311-paret og var en A:U i

flertallet av S1389-intronene (68 %) (tilsvarer *Tetrahymena*-posisjoner A265:U310). Den andre typen var variasjon ved bulgenukleotidet. Flertallet av intronene hadde det høyt konserverte A (93 %), men fem introner hadde enten en bulge-U (fire isolater) eller en -G (ett isolat). Det ble tidligere rapportert at S1389-familien var ute av stand til å selv-spleise *in vitro* under standard reaksjonsforhold. Disse variantene kan representere en ekstra mulighet i å unnsnippe Goddard-Burt-syklusen ved å bli mer vertsavhengige.

Våre resultater viser en generell bevaring innen de nukleære gruppe I-intronfamiliene i strukturell organisasjon, men med noen interessante unntak. I tillegg ser vi at de ulike observasjonene også viser hvor kompleks intronevolusjonshistorien kan være. Dette gir oss en bredere innsikt for å forbedre vår forståelse av de unike egenskapene og dynamikken til nukleære gruppe I-introner i myxomyceter, og dermed inkludere ny informasjon til feltet for katalytisk RNA og intronevolusjon.

1 Introduction

1.1 Exploring the RNA

RNA molecules are rich in form and function. Their ability to fold into three-dimensional (3D) structures and to undergo conformational transitions between multiple states enable RNA to carry out a wide range of cellular functions (Cech and Steitz 2014; Ganser et al. 2019). RNAs have critical functions in cellular processes such as DNA replication, protein synthesis (Chen et al. 2012), intron splicing (Will and Lührmann 2011) and ligand recognition through riboswitches (Garst et al. 2011). RNA is commonly classified into two main types: the messenger RNA (mRNA), and non-coding RNA (ncRNA) (see Figure 1).

mRNA carries genetic information from genes to proteins (Brown 2017). Eukaryotic mRNA genes are transcribed by RNA polymerase (RNAP)-II located in the nucleoplasm. Maturation of mRNAs in the nucleus involves processes like 5' capping, 3' polyadenylation, and removal of spliceosomal introns, all of which are associated with RNAP-II (Schier and Taatjes 2020). These crucial steps ensure the correct processing of eukaryotic mRNA and facilitate transportation to the cytoplasm, where the mRNA becomes translated by the ribosome (Rodriguez et al. 2004). The ribosomal RNA (rRNA) genes are transcribed by RNAP-I, located in the nucleolus (Pitts and Laiho 2022). Transfer RNA (tRNA) and a range of other small ncRNAs genes are transcribed by the RNAP-III in the nucleoplasm (Turowski and Tollervey 2016).

ncRNAs are composed of expressed housekeeping RNA such as tRNAs and rRNAs, and regulatory RNAs (Figure 1). The latter group is categorized broadly according to their size, into small ncRNA (<200 nt; such as micro-RNA, small-interfering RNA, small nucleolar RNA, and piwi-RNA), and long ncRNA (lncRNA, >200 nt) (Zhang et al. 2019). Both the small- and the large ncRNAs have been reported to take part in important cellular processes including gene expression, cell cycle regulation, modification of

chromatin structure, and disease development (Pauli et al. 2011; Prensner and Chinnaiyan 2011; López-Jiménez and Andrés-León 2021; Bhatti et al. 2021).

A notable class of ncRNA molecules, the catalytic RNA, are regulatory ncRNAs. They have a wide range of sizes and catalyze chemical reactions without the essential help of proteins (Tanner 1999; Fedor and Williamson 2005).

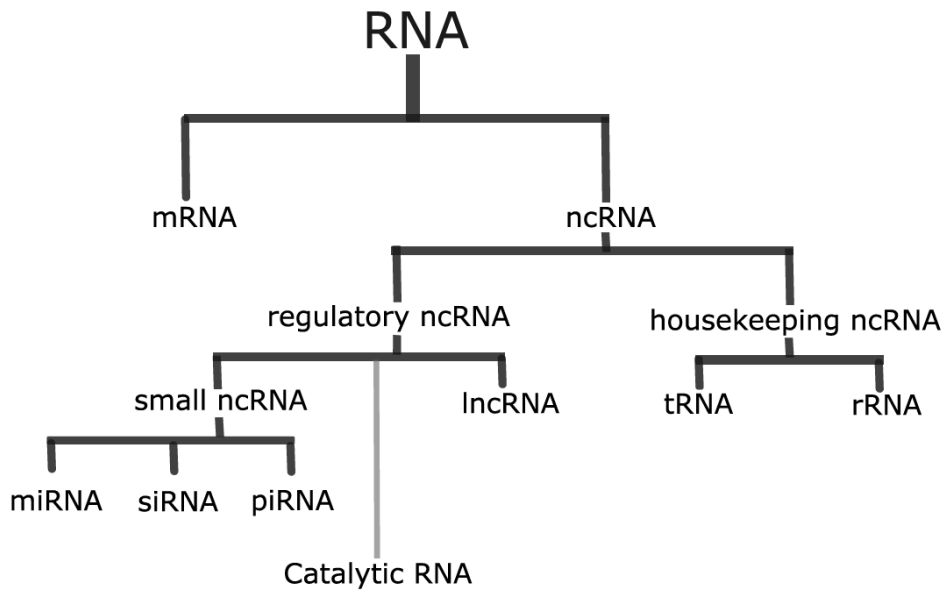


Figure 1: Major types of RNA. RNA can be divided into protein-coding messenger RNA (mRNA) and non-coding RNA (ncRNA). ncRNA is further divided into housekeeping ncRNA (e.g., transfer RNA (tRNA) and ribosomal RNA (rRNA)), and regulatory small ncRNA (e.g., micro-RNA (miRNA), small-interfering (siRNA), small nucleolar RNA (snoRNA), and piwi-RNA (piRNA)), and the long ncRNA (lncRNA). The catalytic RNA could be classified among both the small and large RNAs.

1.2 Catalytic RNA

The catalytic RNAs, also known as ribozymes, are RNA enzymes with the ability to catalyze chemical reactions, such as transesterification (larger majority), phosphoryl transfer, and hydrolysis reactions (reviewed in Lilley and Eckstein 2008). The catalysis is dependent on RNA structures, such as intricate secondary and 3D structures, which may create active sites and stabilize folded domains. Ribozymes usually interact with external factors like metal ions (e.g., magnesium) or cofactors (e.g., guanosine) at specific binding sites, which serve to stabilize them by neutralizing negative charges or are essential for the catalytic reaction of ribozymes (Tanner 1999; Fedor and Williamson 2005). Ribozymes are abundant and widespread in nature and identified both in prokaryotic (e.g., eubacteria, archaea), eukaryotic genomes (e.g., nuclei, mitochondria, plastids) (Haugen et al. 2005b; Deng et al. 2022), viruses (Nishida et al. 1998) and phages (Edgell et al. 2000; Sandegren and Sjöberg 2004).

Ribozymes have various roles in molecular biology, including RNA replication, RNA modification, intron splicing, and protein synthesis (Lilley 2011; Assmann et al. 2023).

1.3 Group I intron ribozymes

The group I intron self-splicing ribozyme belongs to the large natural ribozymes class. The group I intron positioned at the 1925 site in the large subunit (LSU) rRNA gene in the ciliate *Tetrahymena thermophila* was one of the first ribozyme type discovered (Kruger et al. 1982). Today, group I introns are known to be widespread, but scattered, in nature. Ribozyme coding genes have been identified within the nuclear, mitochondrial, and chloroplast genomes of various protist phyla (Bhattacharya 1998; Bhattacharya et al. 2001), in genomes of chloroplast and mitochondria of higher plants (Cho et al. 1998; Cho and Palmer 1999), in nuclei and mitochondria of fungi (Shinohara et al. 1996; Nikoh and Fukatsu 2001; Haugen et al. 2004) in mitochondria of sponges and hexacoral animals (Schuster et al. 2017; Johansen and Emblem 2020) in eubacteria (Kuhse et al. 1990), in archaea (Nawrocki et al. 2018), in bacteriophages (Quirk et al.

1989; Lavigne and Vandersteegen 2013), and in eukaryotic viruses of green algae and soil amoeba (Yamada et al. 1994).

Nuclear rRNA genes in a number of eukaryotic microorganisms (protists and fungi) have been reported to host nuclear group I introns (Cannone et al. 2002; Haugen et al. 2005b; Nielsen and Johansen 2009; Hedberg and Johansen 2013). Among these, we find the myxomycetes (plasmodial slime molds), whose nuclear rDNA is loaded with group I introns (Lundblad et al. 2004; Wikmark et al. 2006; Nandipati et al. 2012). The sites where nuclear group I introns are inserted represent phylogenetically conserved and functionally important positions in the small subunit (SSU) and the LSU rRNA genes (Figure 2). Interestingly, a defined insertion site is commonly occupied by related group I introns (defined as an “intron family”), even in distantly related species. Group I intron insertion sites in rDNA are named according to a standard nomenclature based on the *E. coli* rRNA gene numbering system (Johansen and Haugen 2001).

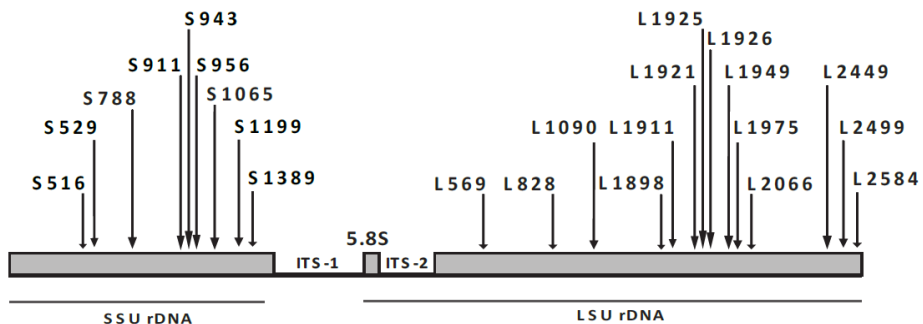


Figure 2: Nuclear group I intron insertion sites in rDNA. Schematic representation of nuclear group I intron sites in rDNA of myxomycetes. The 23 intron sites in the SSU and LSU rDNA are shown with arrows and numbers according to the accepted nomenclature (Johansen and Haugen 2001). The SSU region is interrupted by nine intron families. The LSU region is interrupted by 14 intron families.

Group I introns are not randomly inserted within rRNA genes but are mostly clustered at conserved sites that are essential for ribosome function (Jackson et al. 2002). Several nuclear group I introns interrupt crucial sequences in the peptidyl transferase (LSU

rRNA) or decoding (SSU rRNA) centers of ribosomes (Decatur and Fournier 2002; Jackson et al. 2002). Examples are introns of the S516 family which interrupts the tRNA binding and decoding center (Noller 2006), another is the S1389 family which interrupts the head-and-neck flexibility region of the SSU rRNA (Mohan et al. 2014), and the L2066 family that interfere with the peptidyl transferase center of LSU rRNA (Tirumalai et al. 2021).

1.3.1 A glimpse of the self-splicing activity of group I introns

Group I introns are mobile-type genetic elements, capable of self-splice *in vivo*. They have unique RNA structures, self-splicing properties, and inheritance patterns (Nielsen and Johansen 2009; Hedberg and Johansen 2013). As ribozymes at the RNA level, they catalyze their own excision from a primary RNA transcript by two transesterification reactions (intron splicing) (Cech 1990; Nielsen and Johansen 2009; Hedberg and Johansen 2013). In addition, many group I ribozymes catalyze hydrolysis-initiated intron RNA circularization (Cech 1990; Nielsen et al. 2003; Andersen et al. 2016). The *Tetrahymena* group IC1 intron (Tth.L1925) serves as one of the key model systems in group I intron self-splicing. Tth.L1925 performs self-splicing *in vitro*, resulting in ligated exons and usually truncated intron RNA circles (Cech 1990; Hedberg and Johansen 2013). Another model system in group I intron self-splicing is the myxomycete *Didymium* group IE intron (Dir.S956-1), which performs two parallel self-catalytic reactions resulting in ligated exons and full-length intron circles, respectively (see Figure 3).

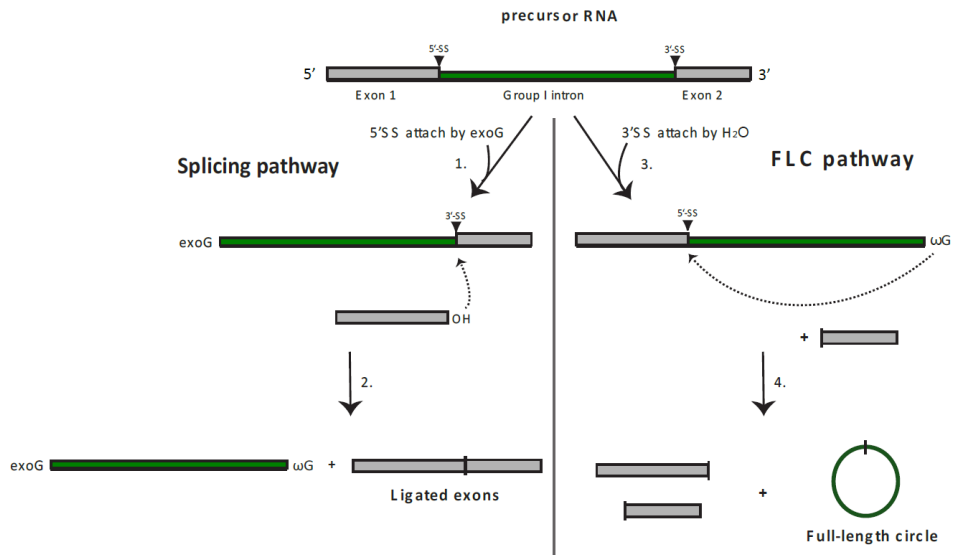


Figure 3: Processing pathways of nuclear group I introns; The possessing pathways are exemplified by the *Didymium* group IE intron Dir.S956-1. The splicing pathway (left) is the self-splicing reaction conducted by two transesterification reactions. The first reaction is initiated by an exogenous guanosine (*exoG*) cofactor binding to the G-site at the catalytic core of the ribozyme. This guanosine acts as a nucleophile through its 3'-hydroxyl group (3'-OH) and attacks the 5'splice site (SS). The outcome entails the ligation of *exoG* to the 5' end of the intron, simultaneously releasing the 5'exon and generating a terminal-free 3'-OH. Before initiating the second step, the ribozyme undergoes a conformational change where the *exoG* exits the active site and is replaced by the conserved guanosine (ωG) at the 3'terminus of the intron (Guo et al. 2004). Consequently, the 3'splice site is strategically positioned for nucleophilic attack by the 3'-OH, thereby promoting the ligation of the exons and the release of a linear intron. The other processing pathway is known as the full-length circularization (FLC) pathway (right). The FLC pathway is initiated by water-induced hydrolysis, which results in cleavage at the 3' SS. This reaction is followed by a transesterification reaction where the hydroxyl group at ωG attacks the 5' splice site. The result is a full-length intron circle and non-ligated exons.

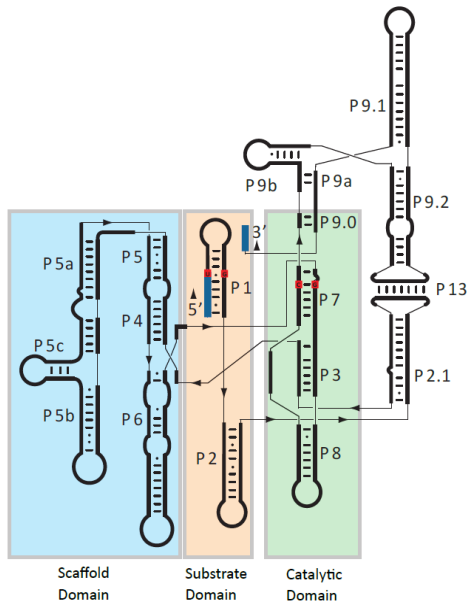
1.4 RNA structure elements of nuclear group I introns

Group I introns have been categorized into at least ten distinct subgroups, which fall under five principal families labeled IA to IE. This classification is based on structural and sequence features, which are supported by phylogenetic analysis (Michel and Westhof 1990; Lehnert et al. 1996; Jaeger et al. 1997; Suh et al. 1999; Li and Zhang 2005; Zhou et al. 2008). Nuclear group I introns are represented by the IC1 and IE subgroups (see Figure 4).

The functional RNA part of the intron is described as the group I ribozyme core. The group I ribozyme core folds into about nine paired segments (P1-P9), which further are organized into the scaffold (P4, P5, P6), the substrate (P1, P2), and the catalytic (P3, P7, P8, P9) helical stack domains (Figure 4). This organization exhibits an intricate folding that is stabilized through both canonical and non-canonical base pair interactions (Michel and Westhof 1990; Vicens and Cech 2006). The IC1 and IE subgroups exhibit distinct characteristics through divergences in their guanosine binding site (G-site) (within the P7 segment), and the scaffold domain, characterized by the supplementary P5abc segment (Figure 4). Recent 3D structural models of Tth.L1925 and Dir.S956-1 catalytic cores present these characteristics in more detail (Andersen et al. 2016; Su et al. 2021).

Group IC1 intron

(*Tetrahymena* intron; *Tth.L1925*)



Group IE intron

(*Didymium* intron; *Dir.S956-1, Glrz*)

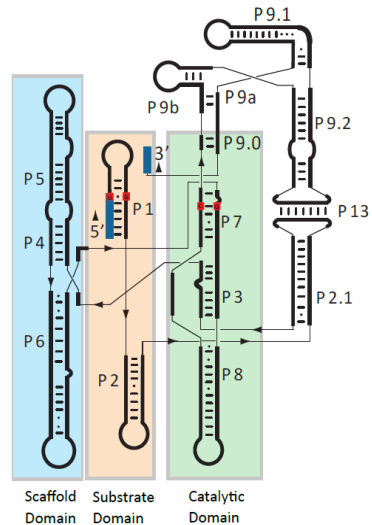


Figure 4: The nuclear group I intron RNA secondary structure. Schematic structure of IC1 intron from *Tth.L1925* (left) and IE intron from *Dir.S956-1* (right). The paired segments (P1-P9) are divided into principal domains and highlighted in boxes: Scaffold (blue), substrate (orange), and catalytic (green) domains. The conserved U:G at the 5' splice site within P1, and the conserved G-C pair in P7 are highlighted in red. *Glrz*, group I intron ribozyme. Adopted from (Hedberg et al. 2013).

1.4.1 RNA structure of IE intron subgroup

The intron subgroup IE is represented by the *Didymium* *Dir.S956-1* intron. The overall structure of the IE subgroup mirrors many of the features of IC1 introns (see section 1.4.2) but appears to diverge from this group by a less complex P5 segment (Haugen et al. 2003; Li and Chang 2005; Zhou et al. 2008) and a more GC-rich catalytic P7 segment. A high-resolution 3D model of the subgroup IE catalytic core presents these features in more detail (Andersen et al. 2016).

1.4.2 RNA structure of the IC1 intron subgroup

The IC1 intron subgroup is represented by the *Tetrahymena* Tth.L1925. Tth.L1925 has a complex structured scaffold domain with a more complex P5abc segment compared to the Dir.S956-1. The Tth.L1925 is a well-studied model system, and high-resolution 3D structural features have been derived from advanced techniques such as RNA crystallization and cryo-electron microscopy (Guo et al. 2004; Su et al. 2021; Bonilla et al. 2022). The crystal structure and 3D models of Tth.L1925 ribozyme core unveiled a highly compact RNA architecture and provide insights into long-range interactions. In particular, the peripheral interactions were suggested to contribute to stabilizing the core structure (Vicens and Cech 2006).

1.4.2.1 Navigating the core: critical RNA structure motifs within Tetrahymena group IC1 intron

The Tth.L1925 intron exhibits a strict and hierarchical folding pattern essential for maintaining its structure and activity. This pattern is defined by several conserved nucleotides within the group I intron, forming the critical core structure (Michel and Westhof 1990; Zhou et al. 2008). The emergence of advanced technologies in RNA structure research, such as cryo-electron microscopy (cryo-EM), has improved our understanding of the intricate connections between secondary and tertiary structures of group I intron (Su et al. 2021; Bonilla et al. 2022; Li et al. 2022; Zhang et al. 2023; Li et al. 2023a).

Key primary and secondary structural motifs of group IC1 introns have been identified based on extensive biochemical and phylogenetic analyses (Figure 5). The first of these motifs are related to substrate specificity, which is governed by the 4-6 nucleotide base pairing between the internal guide sequence (IGS) and 5`exon, forming the 5`splice site (SS) in the P1 segment (Knitt et al. 1994). Earlier studies have emphasized the importance of P1 and P10, which serve as substrates for cleavage and ligation reactions (Doudna et al. 1989). The second motif resides within the P7 segment houses a conserved G-site with a universally conserved G:C pair (G264:C311 in the Tth.L1925

intron) (Michel et al. 1989; Couture et al. 1990; Michel and Westhof 1990; Been and Perrotta 1991). Other conserved sequence features include an A-rich bulge in P5a, the A-rich P4/P5 region (Ikawa et al. 2002), the joining of P4 to the catalytic core, and the terminal ω G at the 3' end (Michel and Westhof 1990).

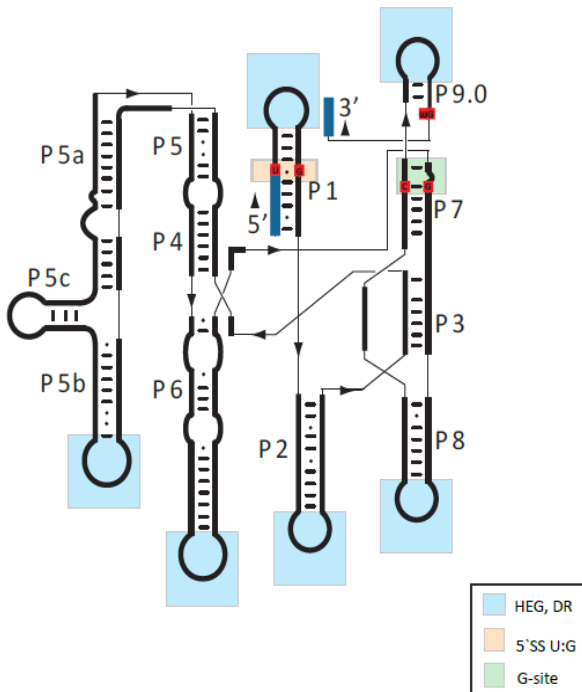


Figure 5: Schematic representation of structural motifs of IC1. Structural features within the group IC1 intron fold. First, the conserved U:G pair within 5' splice site (SS) is indicated in the orange box. Second, homing endonuclease genes (HEG) or direct repeats (DR) are located in the peripheral segments (blue box). Third, the guanosine binding site (G-site) within P7 is indicated in the green box. Lastly, the terminal guanosine (ω G) at the 3' end (red box). (-) Watson–Crick base pair; (•) non-Watson–Crick G:U base pair. The 5' and 3' exon sequences are shown as blue bold lines. This figure is adapted from (Hedberg and Johansen 2013).

1.4.2.2 G-site organization in the P7 segment

The G-site within the P7 segment has the ability to bind guanosine (either exoG or ω G) and is formed by the G264:C311 base pair in Tth.L1925 (Michel et al. 1989). This universally conserved G:C is adjacent to a bulge nucleotide A or C (A263 in Tth.1925). The conserved G264 binds the guanosine through Hoogsteen pairing (Michel et al. 1989). Conserved base triplet interactions within the G-site in the P7 segment have earlier been described (Figure 6) (Guo et al. 2004; Vicens and Cech 2006) and the organization of the G-site is dependent on these base triplets. A study from Tth.L1925 reveals a coplanar base triple surrounded by other base triplets, describes as a base-trippel sandwich (Adams et al. 2004; Guo et al. 2004; Golden et al. 2005). These triplets are formed by stacking base triples at the active site, forming a sandwich structure crucial for stabilizing the active site. In addition, magnesium ions have been reported to coordinate the ribozyme and are essential for the catalytic activity (Guo et al. 2004).

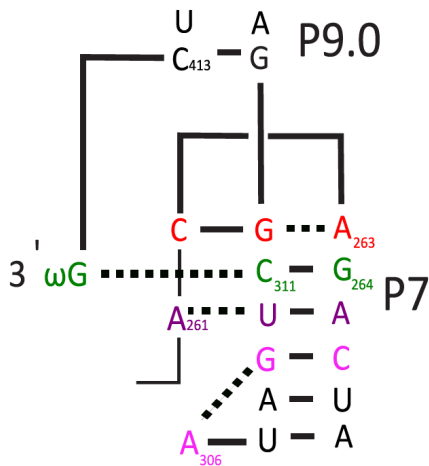


Figure 6: Guanosine binding site in the P7. The RNA structure of the guanosine-binding site (G-site), including the 3' terminal ω G residue, has been elucidated from the crystallographic data of the *Tetrahymena* ribozyme, with base triple interactions indicated by dashed lines. The base triples are shown as red letters (C262-G312:A263), green letters (exoG/ ω G:C311-G264), purple letters (A261:U310-A265), and pink letters (A306:G309-C266). The figure is adapted from (Guo et al. 2004).

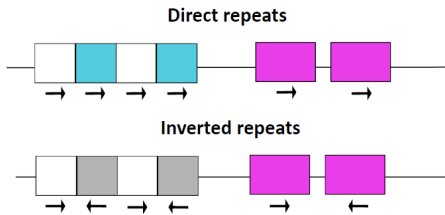
1.5 The peripheral RNA segments of nuclear group I introns

Several nuclear group I introns exhibit distinct characteristics from one another. At the RNA level, additional nucleotides compose peripheral loop segments, extending from the conserved core structure (Figure 5) (Haugen et al. 2004; Hedberg and Johansen 2013). Peripheral components orchestrate an array of tertiary interactions that together significantly contribute to the organizing of the P4-P6 and P3-P9 domains into a compact and active configuration (Michel and Westhof 1990; Golden et al. 1998). These interactions could be between two hairpin loops (kissing loops) or between a hairpin loop and an internal loop (Michel and Westhof 1990; Jaeger et al. 1994; Lehnert et al. 1996). In addition, specific classes of structural and functional motifs have been identified within the peripheral loop segments. These include direct repeat sequence motifs (Vader et al. 1994), open reading frames (i.e., coding for homing endonuclease proteins) (Johansen and Vogt 1994), and insertion elements (i.e., additional ribozymes, such as the lariat capping ribozyme (LCrZ)) (Nielsen et al. 2005).

1.5.1 A brief view of repetitive sequences

Repetitive sequence motifs, widely known as repeats, constitute a prominent component of eukaryotic genomes, comprising at least 50 % of the human genome (Lander et al. 2001). The formation of repeats is attributed to duplications and divergence events of ancestral sequences. They often serve as hotspots of genome rearrangements and evolutionary variations, explained by frequent genetic rearrangements and evolutionary changes (Dunn and Anderson 2019). Repeats have been associated with various biological processes, encompassing evolution and structural stability, and are necessary for genomic structures, such as centromeres and telomeres. Repetitive sequences can be classified based on the characteristics of their repeated motifs, including features like size, orientation, function, and genomic distribution pattern (reviewed in Sperling and Li 2013). Repeated sequences can be broadly grouped into two primary categories: "dispersed repeats" and "tandem repeats" (Figure 7). Both categories are further split into smaller subfamilies.

A Tandem repeats



B Disperse repeats



Figure 7: Repeat motif organization. (A) Tandem repeats could be direct repeats (head-to-tail) or inverted repeats (head-head). The repeat directions are illustrated with arrows. (B) Dispersed repeats are copies of repetitive sequences dispersed throughout the genome. Forward-slash describes the long distance between the repeats.

Dispersed repeats refer to copies of repetitive sequences spread throughout the genome, including transposons, tRNA genes, and gene paralogues (Figure 7b). Tandem repeats are described as sequences repeated in clusters at a particular chromosomal locus, either as direct repeats (head-to-tail pattern) or inverted repeats (head-head pattern) (Figure 7a). Direct repeats are frequently observed in eukaryotic genomes and are exemplified by rDNA repeat arrays and satellite DNA. Satellite DNAs are further classified into satellites, minisatellites, and microsatellites based on the size of repeat motifs (see Richard et al. 2008). New findings indicate that tandem repeats intersecting with regulatory and/or intronic regions might act as functional DNA components, having the potential to control or adjust the process of gene transcription (Fotsing et al. 2019). Identification of repetitive DNA sequences can be done either by manual inspection or by specific bioinformatic tools. Several bioinformatic tools, including Repeat Finder tool in Geneious prime® 2019.2.3 (Geneious 2019), facilitate the identification of repetitive sequences in large datasets.

1.5.1.1 Nuclear group I introns contain direct repeats

A subset of nuclear group I introns has been reported to carry direct repeat motifs, and almost all examples so far have been found in myxomycete rDNA introns. Here, direct repeat arrays were noted in peripheral loop segments (see Figure 5), and most of these introns were not self-splicing as naked RNA *in vitro*. Examples are P1 and P9 arrays in various of L1949 and L2449 introns (Vader et al. 1994; Lundblad et al. 2004; Wikmark et al. 2007a; Nandipati et al. 2012). The current thesis significantly expands our knowledge of group I intron direct repeats into additional intron families.

1.5.2 The homing endonuclease genes - a hallmark of intron mobility

Homing endonucleases (HEs) are enzymes encoded by homing endonuclease genes (HEGs). These endonucleases make a double-stranded DNA cleavage at specific sites, similar to Type 2 restriction enzymes (reviewed in Chevalier and Stoddard 2001; Stoddard 2005; Hafez and Hausner 2012). In addition to code for proteins, HEGs function as mobile genetic elements by integrating into homologous intron-less sites in DNA. This mobility is facilitated by introducing a DNA cleavage by the HE at a specific target site within a cognate allele that lacks a group I introns. The DNA break is then repaired by a gene conversion repair mechanism using a HEG-containing group I intron as a template. This process, termed homing, is schematically presented in Figure 8. In addition to group I introns, HEGs have been identified within in-frame protein sequence introns (inteins) and as free-standing gene elements (reviewed in Stoddard 2005).

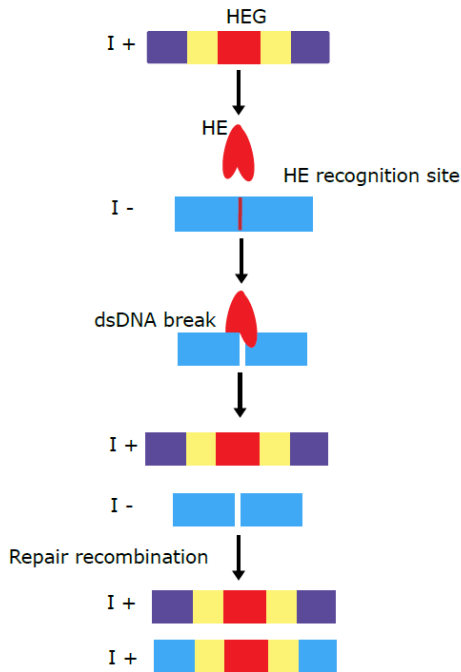


Figure 8: Homing. Intron mobility by a homing endonuclease gene (HEG). The process of mobility at the DNA level requires a double-stranded DNA (dsDNA) cleavage facilitated by the homing endonuclease (HE), at the HE recognition site. This cleavage is followed by a gene conversion recombination and repair process. The mobility dependent on HE is unidirectional and usually highly effective. Intron sequences are indicated by yellow/red (yellow for ribozyme coding and red for HE coding). I+, intron-containing allele; I-, intron-lacking allele; dsDNA, double-stranded DNA. Adapted from (Hedberg and Johansen 2013).

Four major HE families have been described based on conserved amino acid sequence motifs in their active sites (Belfort and Roberts 1997); LAGLIDADG, H-N-H, His-Cys box, and GIN-YING families (reviewed in Galburt and Stoddard 2002; Stoddard 2005). However, all currently known nuclear group I introns HE belong to the His-Cys box family, characterized by a highly conserved series of histidine and cysteine (Figure 9) (Johansen et al. 1993). These residues participate in two zinc-coordinating structures (Flick et al. 1998) (Figure 9). The His-Cys box endonucleases recognize large DNA target sequences (14-40 bp) at or near the insertion site in an intron-less allele (Galburt and Stoddard 2002).

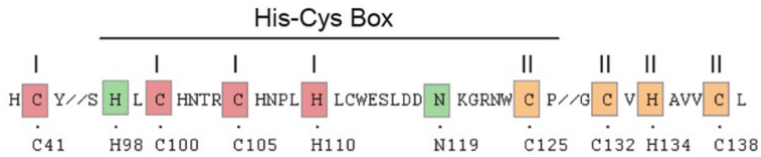


Figure 9: The conserved His-Cys box present in mobile-type nuclear group I introns. Illustration of the conserved Histidine-Cysteine Box (His-Cys box) and the residues participate in the two zinc-coordinating structures. The most important amino acids within the *I-PpoI* His-Cys box are marked with boxes. The residues involved in zinc-binding motifs I (red box) and II (orange box); The green boxes are associated with the active site of the homing endonuclease.

Interestingly, the reading frames of some of the nuclear HEGs have been found to be interrupted by small spliceosomal introns (about 50 nt in size). These spliceosomal introns appear to be members of the major GT-AG class with well-conserved 5' and 3' splice sites, as well as branch sites (Vader et al. 1999; Haugen et al. 2005a; Johansen et al. 2006). Structural, functional, and evolutionary aspects of the HEG-specific spliceosomal introns have not been studied in detail, but the current thesis expands our knowledge about these GT-AG introns.

1.5.2.1 Expression of HEG from rDNA loci in eukaryotes

The fact that all known HEGs from eukaryotic nuclei are located and expressed from rDNA is challenging for the cell. Eukaryotic rDNAs are typically associated with the nucleoli and transcribed by RNAP-I (Pitts and Laiho 2022). Thus, the expression of HEGs must overcome this problem by either introducing RNAP-II to rDNA (such as proposed in anti-sense oriented HEGs) or adding additional features that facilitate nucleolar protein gene expression (reviewed in Johansen et al. 2007). The most common added feature is a polyadenylation signal (AAUAAA) at the 3' region of HE mRNA, which facilitates 3' end cleavage and polyadenylation, and thus mRNA stability. HE mRNA polyadenylation (polyA) tails have been detected *in vivo* from *Didymium* myxomycetes (Vader et al. 1999; Haugen et al. 2005a; Johansen et al. 2006).

The second feature added is the presence of small spliceosomal introns. These introns have been noted and reported removed *in vivo* in *Didymium* species (Vader et al. 1999;

Johansen et al. 2006). Spliceosomal introns may facilitate the nuclear-cytoplasmic transport of the mRNA. The last feature, noted for a subset of mobile nuclear group I introns, is the addition of 5' cap to the HE mRNA, probably to further stabilize the mRNA. This modification is performed by a distinct ribozyme, the LCrz (recently reviewed in Nielsen et al. 2021), known as lariat capping. Lariat capping has been detected in *Didymium* (Nielsen et al. 2005) and in the amoeba-flagellates *Naegleria* (Tang et al. 2011) and *Allovahlkampfia* (Tang et al. 2014).

1.6 The evolutionary history of nuclear group I introns

Group I introns likely originated early in the evolution and probably before the divergence of the three domains of life (reviewed in Haugen et al., 2005). The capacity of group I introns to function as mobile genetic elements is widely believed to have played an essential role in their long-term persistence within genomes of some organisms over evolutionary time (reviewed in Edgell et al. 2011). Moreover, certain group I introns exhibit additional catalytic activities beyond that of self-splicing, including the LCrz (Johansen and Vogt, 1994; Nielsen and Johansen 2009; Nielsen et al. 2021).

Group I introns can facilitate mobility at the DNA level (known as the homing model) or the RNA level (reverse splicing model). These are not mutually exclusive and most likely act in parallel, but at different time frames. However, homing is the most efficient mobility process (as reviewed in Hedberg and Johansen 2013). The first model, involves the function of a HE, encoded by the intron itself and mobility at the DNA level (see Section 1.5). The second model suggests that introns also can integrate at the RNA level through reverse splicing (ribozyme-mediated intron transfer) and may promote intron mobility through an RNA intermediate (review in Hedberg and Johansen 2013). Reverse splicing has been reported *in vitro* and *in vivo* in *E. coli* using the Tth.L1925 (Roman and Woodson 1995; Roman and Woodson 1998; Roman et al. 1999) and in the Dir.S956-1 (Birgisdottir and Johansen 2005). However, the final step, reverse

transcription from RNA to DNA, and a subsequent integration into the genome, has not yet been detected in cells.

1.6.1 Evolutionary aspects of group I intron mobility by the homing

About 30 % of all group I introns, and 5-10 % of those in nuclear rDNA, harbor open reading frames coding for HEs (Jurica and Stoddard 1999; Nielsen and Johansen 2009; Hedberg and Johansen 2013). HEs are involved in the propagation of group I introns and play a crucial role in the transfer of these genetic elements between different host genomes. Nuclear group I introns harbor HEGs positioned within peripheral loop segments (P1, P2, P6, P8, and P9) in both the IC1 and IE intron subgroups (see Figure 5), oriented either in sense- or anti-sense directions compared to that of host rRNA genes (Galburt and Jurica 2005; Hedberg and Johansen 2013).

The interplay between a HEG and its host gene is dynamic. This inheritance pattern of mobile introns ensures rapid fixation in populations. Following fixation, there appears little selective pressure to maintain a functional HE, which may result in an eventual loss of the HEG or even the complete intron. As a result, the host intron insertion site becomes available for reinvasion. The process of rapid invasion, slow degradation, and eventual loss of the intron/ HEG was first characterized in yeast as a cyclic model, which describes the gain and loss of the mitochondrial L2449 ω -intron (Goddard and Burt 1999; Posey et al. 2004; Stoddard 2005). The cyclic model proposed in 1999 (Goddard and Burt 1999) has been crucial in the understanding of the potential evolutionary pathways of genetic elements, such as group I introns, may follow. This model outlines an intron-HEG cycle in which the HEG element integrates into a host gene and undergoes cyclic loss and reinvasion during the time of evolution (Figure 10). Over time, HEGs may accumulate mutations, leading to the degradation of functional HEGs and, eventually loss of introns. However, such mutations can also give rise to novel target recognition sites, facilitating the integration of introns into new genomic locations and promoting their presence in host genes. The net result is a dynamic equilibrium and cyclic intron gain and loss pattern over evolutionary time (Mullineux et al. 2011).

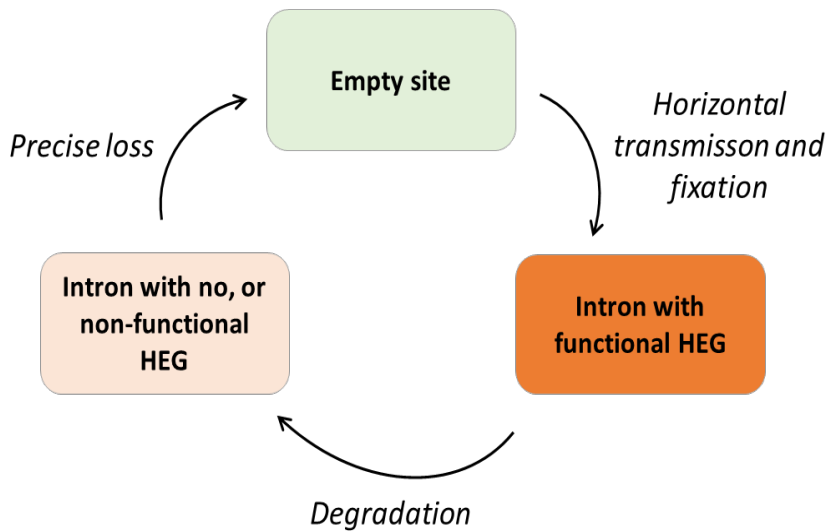


Figure 10: The cyclic model. An illustration of the cyclic model. The cyclic model was introduced by Goddard and Burt in 1999. The cycle initiates with an intron that lacks a homing endonuclease gene (HEG), which then gains a functional HEG through horizontal transmission and fixation. This is followed by degradation and the intron having a non-functional or no HEG due to degradation. The cycle then loops back to the initial state, with the intron precisely losing the HEG. Adapted from (Goddard and Burt 1999).

Group I intron homing has been demonstrated by experimental approaches in different genomic systems in genetic crosses between intron-containing and intron-less strains (Dujon et al. 1989; Lambowitz and Belfort 1993; Sellem et al. 1996). Experimental evidence of homing has so far only been reported in two nuclear group I intron systems, both from myxomycete protists. These are the L1925 intron in *P. polycephalum* (Muscarella and Vogt 1989), and Dir.S956-1 (Johansen et al. 1997). The unidirectional and highly efficient homing has enabled group I introns to integrate into the genomes of diverse organisms. Over time, the group I intron became a more stable component that has gained new roles in RNA splicing, gene regulation, genome organization, and occasionally adapted novel functions (Nielsen and Johansen 2009; Nielsen et al. 2021). A notable and very interesting observation is that group I introns occupying homologous insertion sites exhibit a higher evolutionary relatedness compared to

introns from non-homologous insertion sites, even among evolutionarily distantly related species (reviewed in Haugen et al. 2005b).

1.7 Molecular phylogenetics to understand the evolutionary relationship of nuclear group I introns

The evolutionary history of a specific gene can be assessed by molecular phylogenetics. This approach involves aligning sequences of the genetic element (e.g., a group I intron), comparing it to other homologous sequences, and subjecting data to phylogenetic methods to infer evolutionary relationships at the molecular level. Molecular phylogenetics involves sequence alignment, and construction of phylogenetic trees representing the evolutionary history of a particular genetic element (Vandamme 2009). Assessing horizontal or vertical transmission involves examining whether the phylogenetic history of a genetic element is congruent or not with host gene phylogeny (Farris et al. 1995; Goddard and Burt 1999). In this manner, molecular phylogenetics provides a powerful approach to exploring the evolutionary dynamics of genetic elements and their co-evolutionary relationship with host organisms.

Tracing the evolutionary history of group I introns will increase our understanding of the origin and development of structural variations in catalytic RNAs. Molecular phylogenetics offers a powerful approach to uncovering the molecular evolution of group I introns, using DNA gene sequences, DNA sequences derived from RNA, or amino acid sequences (e.g., derived from HEGs). In general, molecular phylogenetics can be used to investigate gene patterns and reconstruct ancestral genomes (Yang and Rannala 2012).

The first step in phylogenetic tree construction is to generate a reliable sequence alignment. This process involves a comparison of homologous sequences from diverse organisms to identify homologous sequence patterns, which then serve as the foundation to build a phylogenetic tree. The alignment of sequences can be achieved

through manual methods or by employing multiple sequence alignment programs (Higgins and Lemey 2009), such as MAFFT (Kato et al. 2002), T-coffee (Notredame et al. 2000), or Clustal-W (Thompson et al. 1994). These tools all have strengths and weaknesses in their interpretation and performance.

Sequence composition and length conservation are important factors in the alignment process, with the precision of molecular phylogenetics being dependent upon the quality of the derived alignment. As the taxonomic relatedness of sequences declines, determining the ideal alignment becomes increasingly complex. Thus, alignments are typically created using specialized software packages incorporating specific algorithms to address this challenge effectively (Higgins and Lemey 2009).

Genetic sequences with minimal conservation in their primary structure, but with conserved structural motifs (secondary or tertiary), often require manual alignment, utilizing the conserved structure motifs as guides. One such instance is the alignment of group I intron sequences. Group I intron sequence databases have been published for some subtypes (Damberger and Gutell 1994; Zhou et al. 2008; Deng et al. 2022), but the most important sources of group I intron elements are the public databases such as the National Center for Biotechnology Information database (NCBI 1988) and available unpublished collections. Aligning group I introns can be difficult due to their short sequence length (250-500 nt), few conserved nucleotides, and complex folding. Therefore, a manual sequence alignment is typically conducted (see Lundblad et al. 2004; Zhou et al. 2008; Nandipati et al. 2012; Deng et al. 2022).

Computational algorithms are used to generate a phylogenetic tree based on sequence alignment data. Various methodologies are available for tree construction, with the most suitable method determined by factors such as the size and composition of the alignment and the number of sequences to be included. In general, multiple methods are employed to intercept the best tree. Multiple phylogenetic methods have been developed, such as distance- and character-method to determine the best evolutionary tree (Van de Peer 2009; Hall et al. 2011a; Hall et al. 2011e). Among these,

distance-based tree methods convert the aligned sequences into a distance matrix of pairwise differences between sequences. The matrix incorporates factors such as the number of nucleotide base substitutions and is used as the basis from which branching order and branch length are computed. Distance-based methods estimate the actual number of substitutions that have occurred by applying a specific evolutionary model, accounting for assumptions related to the nature of evolutionary changes. Distance-based methods are exemplified by the unweighted pair group methods with arithmetic mean (UPGMA) and neighbor-joining (NJ) (Hall 2011a). UPGMA is the oldest and simplest phylogenetic method; it constructs trees based on a clustering approach, identifying the smallest value from a pairwise distance matrix. UPGMA is built on the assumption that the tree is additive and that all taxa are equidistant from a root, but UPGMA is hardly used today (Hall 2011a). The second distance-based method, the NJ method, provides a fast and efficient approach for constructing phylogenetic trees based on the pairwise distance between sequences (Saitou and Nei 1987). The NJ method does not assume that all taxa are equidistant from a root and offers a unique combination of computational speed and distinctive results. The primary advantage of the NJ method is its remarkable speed, which enables the construction of extensive trees encompassing hundreds of sequences (Van de Peer 2009; Hall 2011a).

Character-based methods are phylogenetic approaches that infer evolutionary relationships by analyzing an individual or a set of characters (e.g., nucleotides, amino acids, or morphological traits) in the data. These methods can find the most parsimonious or probable tree that explains the expected characters, considering specific models of sequence evolution or morphological trait changes. Maximum parsimony (MP), Maximum likelihood (ML), and Bayesian inference (BI) (Van de Peer 2009; Hall 2011a-d) are some of the better-known and widely used character-based methods. MP builds a tree according to the evolutionary scenario with the computationally smallest evolutionary changes, but its drawback is that it can be computationally intensive (Kluge and Farris 1969; Farris 1970; Fitch 1971). ML is also a probabilistic method to find the best-fit tree of the evolutionary scenario that

maximizes the likelihood of observing the data. Similar to MP, ML is a computationally intensive method (Cavalli-Sforza and Edwards 1967; Felsenstein 1981). BI is a probabilistic method that selects trees with the greatest likelihood given the data with a final selection of equal, high-likelihood trees. Even though BI is computationally intense, it provides a probabilistic framework for assessing the uncertainty of the tree (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003).

Each method has its strengths and weaknesses, with its applications relying on the specific research questions, the dataset, and available computational resources. For validating phylogenetic trees, excluding those constructed using BI, a statistical technique called bootstrapping serves as a resampling method for estimating the reliability of a given tree, operating on the principle of resampling with replacement (Felsenstein 1985). Typically, two to four tree-constructing methods are used in analyzing a data set.

1.8 Myxomycetes: a host model system in studying the evolutionary history of nuclear group I introns

The myxomycetes represent a unique and ancient monophyletic protist phylum consisting of unicellular eukaryotes with an evolutionary history spanning over a billion years (Adl et al. 2012, 1019). Myxomycetes hold a distinct position in biology as free-living predators preying upon other protists, fungi, and various prokaryotes. They exhibit a remarkable distribution diversity, having been identified across all studied terrestrial ecosystems and biogeographic regions (Lado and de Basanta 2008; Rojas and Stephenson 2017a; Lado et al. 2018) as well as some aquatic habitats (Lindley et al. 2007).

Over 1,100 valid species have so far been identified (Lado 2005-2022; Lado et al. 2022; Prikhodko et al. 2023). Myxomycetes are represented by 64 genera organized into the dark-spored orders *Physarales*, *Echinosteliales* and *Stemonitidales*, and the bright-spored orders *Trichiales* and *Liceales* (Lado et al. 2022).

The unique and multifaceted life-strategy of myxomycetes comprises various stages, ranging from the haploid microscopic amoeba-flagellates to the diploid macroscopic plasmodium (see Johansen et al. 1997). The plasmodium is formed by synchronous nuclear divisions without cytokinesis that develops from a diploid zygote and consists of thousands of intact nuclei in synchronous mitotic division, enveloped by a single plasma membrane (Collins and Betterley 1982a; Laffler and Tyson 1986; Burland et al. 1993; Fiore-donno et al. 2005; Gorman and Wilkins 2017).

A recent study based on flow cytometry (Li et al. 2023b) reports the haploid genome sizes of myxomycetes to vary from approximately 20 Mb to 500 Mb. Dark-spored species appear to have smaller genomes than bright-spore species. The two myxomycete species *P. polycephalum* and *D. iridis*, both belonging to the order *Physarales*, are separated from each other by probably 4-500 million years (Johansen et al. 1992b). These species have gained significant attention during the last 50-60 years as model systems in biochemistry and cell biology (for reviews see Aldrich and Daniel 1982; Sauer 1982; Dove et al. 1986; Burland et al. 1993; Rojas and Stephenson 2017b). Both species appear to organize their haploid genomes into approximately 40 relatively small-sized chromosomes (Collins and Betterley 1982b; Miller et al. 2022), and sequencing studies estimate their genomes to be approximately 180 Mb and 70 Mb, respectively, in *P. polycephalum* and *D. iridis* (Schaap et al. 2016; S.D. Johansen, pers. com).

An interesting feature among the myxomycetes (and at least among the *Physarales* species) is that rDNA locus is located on a separate mini-chromosome (Figure 12a). These chromosomes are small in size (20-80 kb), linear in structure, and carry regular TTAGGG telomeric repeats at their ends (see Johansen et al. 1992b). Furthermore, they are multicopy nucleosome-containing DNA molecules (ca 150 copies/ haploid genome) that possess non-mendelian inheritance patterns and harbor one or two rRNA transcription units per molecule (see Miller et al. 2022). Of particular interest is that the SSU and LSU rRNA genes are rich in nuclear group I intron insertions (see Figure 2).

For example, *Fuligo septica* (family *Physaraceae*, order *Physarales*) harbors 12 different intron insertions in the same rRNA primary transcript (Lundblad et al. 2004), and *Diderma niveum* (family *Didymiaceae*, order *Physarales*) contains at least 20 group I intron insertions (Nielsen and Johansen 2009; Hedberg and Johansen 2013) (See Figure 11).

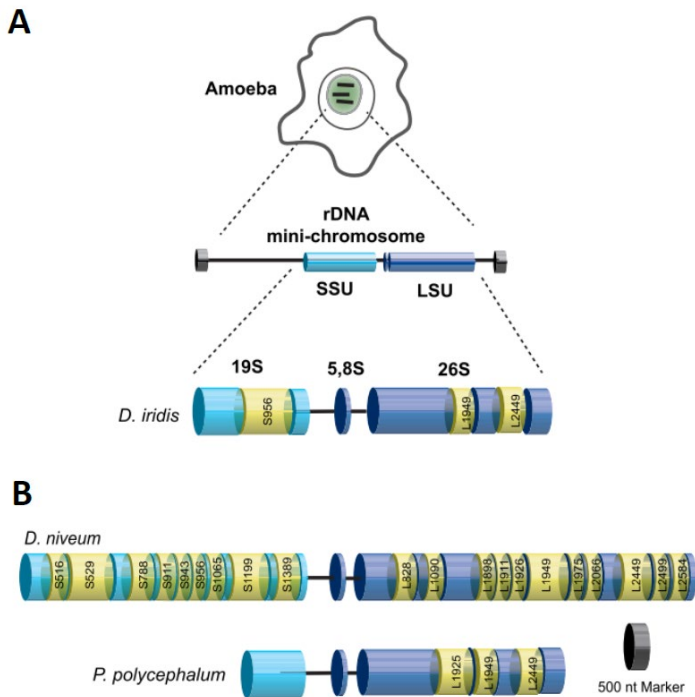


Figure 11: Nuclear group I intron insertions within myxomycete rDNA. (A) The *Didymium iridis* myxoamoeba. A schematic presentation of the rDNA mini-chromosome with SSU and LSU rRNA genes, and regular telomeres (gray boxes). *Didymium iridis* (*D. iridis*) harbors three group I introns (highlighted in yellow); the S956 self-splicing intron in the SSU and the two obligatory introns (L1949 and L2449) in the LSU. (B) group I intron insertions in *Diderma niveum* (Italian isolate) (*D.niveum*) and *Physarum polycephalum* (Carolina isolate) (*P.polycephalum*). All species carry the obligatory introns L1949 and L2449. The *Physarum* LSU rRNA gene also contains a true mobile group I intron at position L1925, which encodes the homing endonuclease *I-Ppol*. Figure from (Hedberg and Johansen 2013).

1.8.1 Myxomycetes model species for nuclear group I introns

Myxomycete group I introns were initially discovered in 1981 in *P. polycephalum* at positions L1949 and L2449 in the LSU rDNA (Nomiyama et al. 1981a; Nomiyama et al. 1981b), and further characterized in the following studies as obligatory introns (Johansen et al. 1992a; Vader et al. 1994; Wikmark et al. 2007a). In addition, an optional LSU rDNA intron, named Ppo.L1925, became the first nuclear mobile group I intron to be discovered (Muscarella and Vogt 1989). Ppo.L1925 carries a HEG that codes for a functional HE named I-*PpoI* (Muscarella et al. 1990; Flick et al. 1998). Furthermore, Ppo.L1925 represents the subgroup IC1 intron and is highly similar in RNA structure organization compared to the well-known Tth.L1925 intron (Ruoff et al. 1992).

The other myxomycete model species, *D. iridis*, contains several group I introns in its rDNA (see Figure 11a). The two LSU introns (L1949 and L2449) are orthologs to introns in *Physarum*. In addition, three optional introns were reported in the SSU rDNA (two at position S956 and one at position S1389) (Johansen and Vogt 1994; Decatur et al. 1995; Haugen et al. 2005; Johansen et al. 2006; Wikmark et al. 2006; Wikmark et al. 2007a; Wikmark et al. 2007b). Both S956 introns are the mobile-type with HEGs (Haugen et al. 2005). While the S956-2 is a group IC1 intron with an anti-sense organized HEG, the S956-1 is a differently organized group IE intron with sense HEG. Furthermore, S956-1 contains a second ribozyme unit, the LCrz, that is responsible for generating a lariat cap structure at the 5' end of the homing endonuclease mRNA (Nielsen et al. 2005; Nielsen et al. 2021).

2 Aims of study

The aim of this study was to investigate and characterize the evolutionary aspects of catalytic RNA introns related to RNA sequence conservation and structural variability by phylogenetic assessments of nuclear group I introns. During this project, identification of structural organizations and evolutionary inheritance patterns of three nuclear group I intron families, S516, S1389, and L2066 were performed following the study objectives:

1. Characterize and investigate RNA sequence- and structural- variations within the host-dependent S1389 family of group I introns in myxomycetes by molecular phylogenetic assessments.
2. RNA structural characterization of the self-splicing S516 family from myxomycetes by a comparative and phylogenetic approach.
3. Investigate the RNA structural and mobility features of the mobile group I self-splicing L2066 intron family by comparative- and biochemical assessment methods.
4. Revise the evolutionary history model of nuclear group I self-splicing introns in myxomycetes.

By approaching these objectives, we expand our knowledge of the molecular evolution of catalytic RNA introns, homing endonuclease mRNA regulation, and how homing endonuclease genes are expressed in an rDNA context.

3 Summary of papers

Paper I: A phylogenetic approach to structural variation in organization of nuclear group I introns and their ribozymes.

The aim of the study was to investigate observed structural variations in myxomycete group I introns belonging to the family S1389 by molecular phylogeny as the main analytic tool. In this study the nuclear rDNA genes of a Norwegian subarctic isolate of the myxomycete *Mucilago crustacea* were sequence analyzed and shown to be interrupted by seven groups I introns; two in the SSU rDNA and five in the LSU rDNA. One of these introns, located at position S1389 in the SSU rDNA, was found unusual in size due to an insertion of a complex direct repeat array in peripheral segment P9. S1389 introns are of particular interest since they interrupt the critical and essential hinge-structure helix H28, which at the RNA level must be spliced out perfectly to obtain a functional SSU rRNA. Earlier studies have shown that group I introns inserted at position S1389 apparently do not self-splice *in vitro* as naked RNA, suggesting essential host-dependence in splicing. To gain comparative information regarding structural variability and molecular evolution of group I introns at position S1389, 72 intron sequences representing four orders, 15 genera, and 37 myxomycete species were collected from databases, as well as from our published sequences, and investigated in more detail. The analyses revealed that the S1389 introns are folded according to a standard IC1 organization with a conserved ribozyme core and a more complex P4-P6 domain, but with structural variations and insertions in peripheral RNA regions. Interestingly, some structural variations appeared taxa-specific. Findings include several unusual sequence variants in the catalytic core (guanosine binding site) in the P7 segment. Next, complex direct repeat arrays were observed in peripheral regions in three of the S1389 introns. Four S1389 introns contain HEGs, indicating mobile intron activity. These HEGs were further interrupted by small spliceosomal introns. Lastly, phylogenetic analyses revealed mixed, scattered, and complex inheritance patterns with candidates representing both vertical and horizontal

inheritance patterns. The distribution of molecular features such as direct repeat arrays, HEGs, and guanosine binding site variants appeared sporadic among S1389 in myxomycete species. We conclude that group I introns belonging to the S1389 family represent an interesting model system in studying molecular organization and evolution of non-self-splicing group I introns more specifically, and more general ncRNA elements in a nuclear environment.

Paper II: Structural organization of S516 group I introns in myxomycetes.

The aim of the study was to perform structural characterizations of S516 group I introns in myxomycetes, based on comparative and phylogenetic approaches, to reconstruct the molecular evolution of self-splicing introns. Nuclear group I introns interrupt highly conserved sites in the rDNA of protists and fungi. One site that is frequently interrupted is at position S516 in the SSU rRNA gene. Position S516 is located in the highly conserved 530-loop of the SSU rRNA, which performs tRNA binding and is involved in mRNA decoding of the ribosome. Thus, group I introns at this position need to be spliced out both efficiently and accurately to retain the ribosome function of the host. In the present study, we compiled and analyzed a dataset consisting of 75 S516 group I introns from myxomycete protists. We investigated previously undescribed structural variations in the molecular organization of introns in an evolutionary context. The 75 intron sequences represented five orders, 21 genera, and 38 myxomycete species, which were collected from public databases as well as our published sequences. The analyses showed that the structural organization of the myxomycete S516 introns represents the IC1 sub-class with a conserved catalytic core region flanked by more variable peripheral regions. Extensions in variable regions consisted of complex and heterogenous direct repeat arrays, as well as HEGs interrupted by small spliceosomal introns. The spliceosomal introns carried well-conserved splice sites and branch sites, which places them in the major GT-AG class of spliceosomal introns. The structural segment P2 most frequently harbored extension sequences in S516 introns. Here, the intron from *Licea parasitica* contained 25 copies of a 23-nt motif, each representing an

RNA hairpin structure with compensatory changes among copies. Several interesting observations were made among the 25 *Trichia varia* intron isolates analyzed. Three isolates contained HEGs with spliceosomal introns in P2, whereas in one isolate the HEG apparently has become truncated and replaced by a direct repeat array. This observation proposes that HEGs and direct repeat insertions may be evolutionarily linked and represents an interesting example of the Goddard-Burt intron gain and loss cycle hypothesis. The phylogenetic analysis identified a complex inheritance pattern that can be explained by repeated intron gain and loss during evolution. Here, a reconstruction of the evolutionary history of S516 intron in myxomycetes was presented from the insertion of a mobile-like intron at unoccupied cognate sites, through HEG degradation and loss, and finally, to the complete loss of introns. The S516 group I introns in myxomycetes thus represent an intron family with surprisingly dynamic structures despite a common function in RNA self-splicing.

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

The aim of this study was to investigate structural features in a comparative setting of the group I intron family L2066, and to perform expression and biochemical analyses of a mobile intron version and its homing endonuclease gene product. Nuclear group I introns are also present in the LSU RNA gene, and the L2066 family of introns interrupt ribosomal RNA in proximity to the peptidyl transferase active site. L2066 introns have so far only been observed within selected orders in myxomycete protists and ascomycete fungi. In this report, 34 L2066 group I introns (16 in myxomycetes and 18 in ascomycetes) were investigated. All L2066 introns could be folded at the RNA level according to the group IE organization. Phylogenetic analysis revealed clustering of the ascomycete and myxomycete L2066 introns into two distinct clades. Whereas the ascomycete introns were uniform in size, large variations were seen among the myxomycete L2066 introns. However, no direct repeat arrays could be observed, but two of the myxomycete introns were mobile-type introns harboring HEGs. Both HEGs

were interrupted by small spliceosomal introns, and interestingly these HEGs were oriented either in sense or in antisense directions. The L2066 intron from *Didymium squamulosum*, which harbors an antisense-oriented HEG, was further investigated in more detail. Firstly, the intron self-spliced *in vitro* as naked RNA, generating perfect ligated exons and full-length intron circles. Secondly, the antisense-oriented HEG was expressed as RNA in *Didymium* cells, resulting in polyadenylated mRNA where the small spliceosomal intron had been removed. Thirdly, the HEG without the spliceosomal intron was over-expressed in *E. coli* and shown to confer endonuclease activity. Finally, the homing endonuclease protein was purified and shown to catalyze a double-stranded break at an intron-lacking allele sequence, which generated a pentanucleotide 3' overhang at the intron insertion site. These results support that the *Didymium* L2066 intron is a true mobile group I intron that could spread by homing between cells through genetic crosses. We also propose a cellular model to explain how an antisense HEG may become expressed from the nuclear rDNA locus.

4 General discussion

This work presents observations based on comparative analyses of three nuclear group I intron family datasets named S516, S1389, and L2066, mainly from myxomycetes. Here, our objectives have been addressed, which resulted in characterizing and identifying RNA sequence- and structural- variations within the host-dependent S1389- and the self-splicing S516 -family in myxomycetes by comparative and phylogenetic assessments (**Paper I** and **Paper II**). We also revealed the mobility features for the *D. squamulosum* L2066 intron, and the RNA structural feature of the mobile group I self-splicing L2066 intron family. These results were conducted by a comparative- and biochemical assessment methods (**Paper III**).

Apart from for the well-conserved RNA structural organization and evolutionary inheritance within the RNA families, we also revealed interesting exceptions that resulted in a revision of the evolutionary history model of nuclear group I self-splicing introns in myxomycetes.

4.1 A cyclic evolution model based on nuclear group I introns

Myxomycetes represent a well-suited model system to study catalytic RNA intron evolution due to their complex sexual life cycle and high abundance of nuclear group I introns in rDNA across diverse orders. **Paper I**, **-II**, and **-III** describe variations in structure and the organization of specific nuclear group I intron families. Phylogenetic analyses reveal an overall congruency within each S1389-, S516-, and L2066- intron families compared to their host genes. These observations suggest that most nuclear group I introns in myxomycetes evolve through a vertical inheritance pattern. However, some introns did not show concurrency toward their host and were apparently horizontally inherited (**Paper I** and **-II**), and some introns present new and surprising sequence variations, such as the G-site variation in the S1389 intron family (**Paper I**), or additional elements, such as, repetitive sequence motifs (**Paper I** and **- II**) and small spliceosomal introns (**Paper I**, **-II**, and **-III**).

The observed intron organizations in the different myxomycete species challenge the current hypothesis of group I intron evolution raised by Goddard and Burt (see Introduction), based on a cyclic model with intron invasion, periodicity of HEG function, and a subsequent intron loss (Goddard and Burt 1999; Goddard et al. 2006). Reconstructing the long-term evolution history of nuclear group I introns faces challenges owing to sequence alignments, horizontal intron transfers, and intron gain and loss events.

Based on our RNA sequence and structural observations (i.e., direct repeats, small spliceosomal introns, and G-site variations) among the S516, S1389, and L2066 groups I introns, we propose a revision of the Goddard-Burt cyclic evolutionary model (Figure 12). The revised model starts with an intron-less host gene. Around 70 % of the myxomycete species with rRNA gene sequences available in the GenBank do not carry group I intron sequences. A short synopsis of the revised model is presented below:

(Step 1): A mobile group I intron containing a functional HEG becomes inserted into the host gene by the intron homing mechanism. Our data show that approximately 7 % of the myxomycete group I introns contain HEGs. In **Paper III** we report that the L2066 intron from *Didymium* represents a mobile type of group I intron where the intron HEG encodes an active homing endonuclease protein, I-*Dsqt*. Other mobile-type introns reported in myxomycetes are from the families S516 (*Lamproderma*, *Echinostelium* and *Trichia*; **Paper II**), S956 (*Didymium*; Johansen and Vogt 1994; Haugen et al. 2005), S1389 (*Brefeldia* and *Trichia*; **Paper I**), and L1925 (*Physarum*; Muscarella and Vogt, 1989).

(Step 2) Group I intron HEGs then gained small spliceosomal introns, probably to facilitate endonuclease expression from a nucleolar environment. At least 75 % of the myxomycete HEGs appear interrupted by spliceosomal introns. This phenomenon was observed and reported in **Papers I, II, and III**, and is discussed in more detail below in Section 4.2.

(Step 3) When no more intron integration sites are available, group I intron HEGs become in excess and subsequently truncated. Such truncated HEGs have been reported in both *Naegleria* (Wikmark et al. 2006) and various red algae (Haugen et al. 1999). A HEG sequence-remnant was observed in one of the S516 *Trichia* introns (**Paper II**). Interestingly, the truncated HEG was found to be associated with a direct repeat, suggesting a structural link between truncated HEGs and direct repeats in some group I introns. Direct repeats in nuclear group I introns are further discussed below in Section 4.3.

(Step 4) A complete loss of HEG results in a 'ribozyme-only' intron in the myxomycete rDNA. Our estimates revealed that 93%, 94%, and 87% of the myxomycete S516, S1389, and L2066 introns, respectively, have a ribozyme-only configuration.

(Step 5) The final step is suggested to have four different outcomes: (a) a complete loss of group I intron and the ability of reinvasion of group I intron by reverse splicing; (b) conversion to a mobile intron by HEG integration; (c) or exit from the Goddard-Burt cycle. Which of these outcomes most probably are currently not known.

Complete loss of a nuclear group I intron (step 5a) may result from homologous recombination between intron-containing and intron-less sequences (e.g., cDNA) of the same allele. Additionally, group I introns may reinvade into rDNA of myxomycete by reverse splicing (Roman and Woodson 1998; Birgisdottir and Johansen 2005). The majority of investigated myxomycete species (approximately 70%) lack a group I intron at a specific site, like S516, S1389, or L2066. An intron-less rRNA gene is then ready to host a new mobile intron element. Alternatively, a ribozyme-only intron can be converted to a mobile intron by HEG integration (Step 5b). HEGs have been identified as free-standing gene elements in some species (reviewed in Stoddard 2005). If peripheral segments contain sequences resembling the homing endonuclease target site, the corresponding HEG could be integrated. There are examples from the literature of putative homing endonuclease target sites in peripheral segments, e.g., in *Didymium iridis* S956 introns (Haugen et al. 2005). Another plausible example of HEG

integration (Step 5b) is the L2066 intron in *Diderma alpinum* (**Paper III**). Here, the ribozyme part was found to be identical between an Italian isolate and a Ukrainian isolate, but only the Italian isolate contains a HEG. The last possible outcome (Step 5c) is that a group I intron can escape from the Goddard-Burt cycle. Such introns are suggested to be more dependent on the host organism for function and subsequently may have gained beneficial roles for the host (see Nielsen et al. 2009; Hedberg and Johansen 2013). Notable examples are the L1949 and L2449 nuclear group I intron (Johansen et al. 1992a; Vader et al. 1994; Wikmark et al. 2007a; Nandipati et al. 2012), which are present in all strains and species of *Physarales* myxomycetes, with an evolutionary time-frame corresponding to approximately 4-500 million years (Johansen et al. 1992b). Whether the G-site mutant variants noted in some S1389 introns (**Paper I**) represent escaping introns will be discussed in more detail below in Section 4.4.

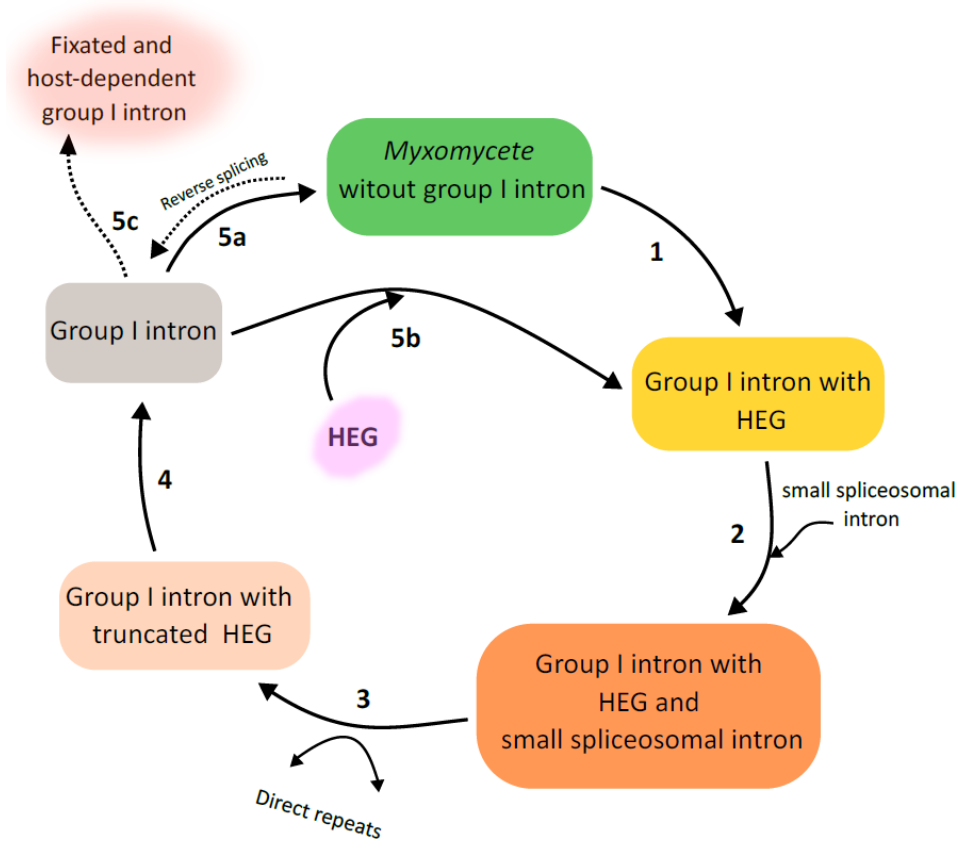


Figure 12: Revision of the Goddard-Burt cyclic evolutionary model. An illustration of a revised Goddard-Burt cycle of intron evolution modified to include observed features of the mobile nuclear group I introns within myxomycetes. This evolutionary model of cyclic intron gain and loss is divided into several steps. (Step 1) Integration of intron with HEG into host gene. (Step 2) Insertion of a small spliceosomal intron within HEG. (Step 3) Degradation and truncation of HEG, can be involved with the gain and loss of direct repeats. (Step 4) Loss of HEG resulting in a stand-alone intron. (Step 5a) Complete loss of group I intron from host gene and the ability of reinvention of group I intron by reverse splicing. (Step 5b) Gain of an independent HEG by the integration directly into group I intron. (Step 5c) Fixation of group I intron in host gene.

4.2 Small spliceosomal introns in nuclear group I intron HEGs

From our S516-, S1389-, and L2066- sequence datasets, we observed several group I introns with large peripheral segments. About 7% of all the S516, S1389, and L2066 introns harbored HEGs, predominantly located in the P1, P2, or P9 segments of the intron structure. A large fraction ($\geq 75\%$) of these HEGs were further interrupted by small spliceosomal introns. Adding spliceosomal introns to HEGs is proposed to occur in Step 2 of the revised Goddard-Burt circle (Figure 12).

The HEG-specific spliceosomal introns are only about 50 bp in size (44-59 bp) and are members of the major GT-AG class of spliceosomal introns (see Introduction). A sequence alignment of ten HEG spliceosomal introns is presented in Figure 13. This alignment shows that sequence motifs at the 5' and 3' splice sites, and branch sites are highly conserved. This level of conservation was observed in all of the identified HEG-specific spliceosomal introns from the S516-, S1389-, and L2066-introns.

Conserved sequences of the small spliceosomal intron

I-58 (S516)	caag	GUAxGxC	GGGCUGAGCUGAACCCGAAAGAUCGGGCAUGA--	RYURAC	CACUCGUGCC	YAG	gacg
I-59 (S516)	aaag	GUAUGAU	CAUUUUGGGAUUGGGAUUAGGGUUAACUGUAU-	ACUGAU	GGAUUUACAUA	CAG	catg
I-50 (S516)	gagu	GUAUGUG	GGCGGUGGCUUUGCGAGGUGCGUGGU-----	ACUAAAC	ACCCCGUC---	CAG	gcug
I-44 (S1389)	caag	GUAGGCC	GUGUAGGCCGCCACCCCU-----	GCUAAU	GAUCAAAA---	CAG	cgca
I-53 (S1389)	aaag	GUAUGUA	UGGGGUUGGGACUUGUUGGGGCUUUUCUG----	ACUGAU	UCCUCUG---	UAG	gaua
I-47 (L2066)	ugug	GUAUGUC	ACCAACUGUUUGCAGGGGCGU-----	ACUGAC	CCUCCUAC---	UAG	guua
I-49 (L2066)	aaag	GUACGGC	AAAUCUUGCAAAAACAGGCUAGUGG-----	ACUAAAC	GGACUAAU---	CAG	cuau
I-56 (S943)	ccau	GUAAGUU	UGCUUCGGUGAACCUUUAGUAGUAUAAGUCUUA	ACUAAAC	UGACUUGC---	UAG	gugu
I-50 (S956)	uaag	GUAGGUC	GGAGUUUGCAGUUAACGCCAUUUAU-----	GUUAAAC	CACCUCUAUA-	UAG	guua
I-51 (S956)	cuag	GUAUGUC	CUUGGAGUGCAAGGGUGGUUAUUUG-----	GCUAAAC	UUAGUUAUU-	CAG	gccc
		**		*		**	

Figure 13: Small spliceosomal introns. Small spliceosomal introns are present in many nuclear group I intron HEGs. The spliceosomal introns from S516, S1389, and L2066 HEGs are aligned and compared with other published spliceosomal introns from the group I intron Emy.S943, Dir.S956-2, and Dir.S956-1 HEGs. All are members of the major GT-AG class of spliceosomal introns. The sequence alignment shows the spliceosomal introns in uppercase letters, and flanking HEG sequences in lowercase letters. Sequence motifs common to the mammalian spliceosomal intron consensus are indicated by green boxes. The invariable positions at the 5' SS (GU), branch site (A), and 3' SS (AG) are marked by * below the alignment.

What is the evolutionary origin of HEG spliceosomal introns?

Two hypotheses may be put forward in explaining the evolutionary origin of these small spliceosomal introns. (1) All HEG spliceosomal introns have evolved from one common ancestral intron and spread by vertical transmission to HEGs in different group I introns in various species. In support of this hypothesis, we observed that 1/3 of the nucleotide positions unambiguously could be aligned between all known HEG spliceosomal introns (Figure 13). However, these intron positions appear to be universally conserved across the spliceosomal introns of all eukaryotes, probably due to functional constraints in the spliceosome. Another important fact that argues against this hypothesis is that the HEG spliceosomal introns are observed in diverse orders of myxomycetes, and in distantly related fungi, with different habitats and geographic distributions. (2) A more plausible hypothesis is that the HEG spliceosomal introns have a local origin from the host nucleus. Whole genome sequence analyses of the myxomycete *P.polycephalum* reported an unusually high load of spliceosomal introns in nuclear genes (>20 introns per gene), and that these introns in general were very small in size (Schaap et al. 2016). If *Physarum* is representative to most myxomycetes, some of these introns may transpose from nuclear genes to HEGs and then become integrated parts of the mobile group I introns, as we observed for S1389 introns in **Paper I**, S516 introns in **Paper II**, and L2066 introns in **Paper III**. According to this hypothesis the high sequence conservations seen between the HEG spliceosomal introns are probably not due to a specific common origin, but rather to the fact of intron size minimization leaving almost only the functional important sequences left.

What is the biological role of HEG spliceosomal introns?

In general, spliceosomal introns may facilitate nuclear-cytoplasmic transport of mRNA (see Introduction). Thus, the HEG spliceosomal introns could facilitate HEG expression from nuclear rDNA by recruiting spliceosomal components and subsequently exon junction complexes (Johansen et al. 2006; Will and Lührmann 2011; Schlautmann and Gehring 2020). However, HEG spliceosomal introns appear not essential for all HEG expression since 1/4 of the mobile group I introns in myxomycetes do not depend on this feature, including the well-characterized intron in *Physarum* rDNA encoding the *I-PpoI* homing endonuclease gene (Muscarella and Vogt 1989; Muscarella et al. 1990).

HEG spliceosomal introns are not the only feature suggested to facilitate HEG expression. A closer inspection of HEG sequences reveals that almost all contain an mRNA polyadenylation signal (polyA signal) “AAUAAA” in proximity to their 3’ stop codon (Figure 14).

Homing endonuclease gene

Small spliceosomal intron

```
ATGTTTCACACGAGAAAGGCCACAACAAGTCCAAACAAAAGCCACCTATCGTTCCTAGACGAGCCGGAGGCACAC
CTTATAGAATATATAAGGGAGCTGAAGGTGTGTGGCCGAGTAATTGGGAGTGTAGTGGGGCGGTGGCTTTGCGAG
GTGGTGTTACTAACAACCCCGTCCAGGCCTCGCCTTGTGGATGGCACATATCATGCGGGCCCTTGCCAAGGCTCGC
CCTGTGAGATGCTGCGGTCCCTGAAAGGCAGGCTGATGACTACTCGGTGCGGGACCACAAGCCAAGCGTGGACA
TCGCTTATGGGTATCAGTTGATAGCCCTGGAGAAGTTTGGGATCGGAATTATGCGGGAGGTTGCGCCGTCTAAGG
ACAGGACTCGATTACAATATCACATTTGTGCGGGACCCTGCTCTGCTGCAGTGCAGATCACCTGGTGTGGAAAC
CCAAATGGCTGAACGACGAGCGGACTTGCTGCCACTTCGTACTGCTTAGTGCCTTTAAGCACGCGGGGTTAGAGG
CAGTGACCGCGTTCTGGAGGCCGCGGGTGCCCGCATACACCTCAGTGTGGCTCCATTGCTGTTAATAAGATAG
TACGCTTCCCACCACATGCTCTGTACCCCGATATAGCCCTACCTCCCCGACGGAGGAGGAATAAAAAATATC
AATACTGA
```

Poly (A) signal

Figure 14: Homing endonuclease gene. The S516 homing endonuclease gene (HEG) of *Tricia varia* (Trichiales, Trichiaceae; GenBank accession numbers KM494993-5). The His-Cys Box HEG and the I-50 spliceosomal intron are indicated. The HEG starts with an initiation codon (ATG) and ends with a termination codon (TGA). The I-50 intron (marked in red) belongs to the major GT-AG class of spliceosomal introns. The polyadenylation signal motif (polyA signal) “AATAAA” (black bold letters) is located close to the termination codon.

The presence of a polyA signal is a strong indication for mRNA polyadenylation. PolyA signals have been noted in almost all group I intron HEGs, indicating that the corresponding homing endonuclease mRNAs are polyadenylated. Some notable examples from this thesis are the *Tricia varia* S1389 introns (**Paper I**, Figure S4), *Tricia varia* S516 introns shown in Figure 14 (**Paper II**, Figure S7), *Lamproderma pseudomaculatum* S516 intron (**Paper II**, Figure S5), *Echinostelium coelocephalum* S516 intron (**Paper II**, Figure S6), *Diderma alpinum* L2066 intron (**Paper III**, Figure 3), and *Didymium squamulosum* L2066 intron (**Paper III**, Figure 4). The mRNA corresponding to the latter intron was reported in **Paper III** to be polyadenylated *in vivo*, resembling that of S956 intron mRNAs from *Didymium iridis* (Vader et al. 1999; Johansen et al. 2006).

Mapping the 5' ends of HE mRNAs were not included in the present study, but we suggest in **Paper III** that the *Didymium squamulosum* L2066 mRNA is transcribed by RNAP-II, and thus may carry a regular m⁷G 5' cap. Interestingly, HE mRNAs from mobile group I introns in *Didymium*, as well as the amoeba-flagellates *Naegleria* and *Allovhalkampia* (Nielsen et al. 2005; Tang et al. 2011; Meyer et al. 2014; Tang et al. 2014; reviewed in Nielsen et al. 2021) have been reported to add a cap structure at the 5' end. These 2',5' lariat cap structures are different from the common type caps (m⁷G caps) and catalyzed by a specific ribozyme (LCrz) coded by the introns. However, the mobile-type introns reported in **Papers I to III** do not encode LCrz.

In summary, processes like spliceosomal intron splicing, polyadenylation, and probably 5' capping HE mRNAs all are important in facilitating HEG expression and HE translation. **Paper III** also presents a model for overcoming some of the challenges related to nucleolar HEG expression by spatial and probably temporal separation of rRNA transcription and antisense mRNA transcription.

4.3 Complex direct repeat arrays in nuclear group I introns

Direct repeat (hereafter referred to as DR) sequences were observed in 13 of the 163 (8%) myxomycete group I introns investigated in this study. While three and eleven DR-introns were noted in the S1389 (**Paper I**) and S516 (**Paper II**) families, respectively, none were detected among the L2066 introns (**Paper III**). DR arrays were previously associated with group I intron lacking *in vitro* self-splicing activity, such as members of the L1949 and L2449 intron families (Vader et al. 1994; Wikmark et al. 2007a; Nandipati et al. 2012). Here, examples of DR arrays were noted in all peripheral segments of the group I introns RNA structure and did probably not interfere with the conserved catalytic core RNA. In **Paper II**, we reported DR features in eleven of the 75 S516 introns (13%), a family of group I introns known to support *in vitro* self-splicing activity (Haugen et al. 2003).

A detailed structure analysis of the S516 DR-introns revealed sequence heterogeneity among the motifs, as well as variations in both length and copy numbers (**Paper II**, Figure S8). DR-motifs ranged in size from only 4 bp to more than 220 bp, and copy numbers from 2 to 25. Interestingly, the different copies in a DR-array appeared non-identical (but highly similar) in sequence, arguing against DNA polymerase slippage mechanisms during replication (Levinson and Gutman 1987; Hancock and Dover 1988) in generating the repeats. Moreover, nine S516 intron isolates, all members of the order Stemonitales, contained DRs and suggested a taxa-specific distribution. In contrast, no such taxonomic association was noted in the S1389 DR-introns. Despite indicating taxa-specific distribution of S516 DR-introns, further phylogenetic analyses support a sporadic occurrence in the S1389 family (**Paper I**, Figure 6) and the S516 family (**Paper II**, Figure S3). Interesting DR features are briefly discussed below.

DR as a putative regulatory element

Structural examination of the large S1389 intron from *Mucilago crustaceum* (Mcr.S1389) revealed an exceptionally long peripheral segment in P9 (**Paper I**). Here, a complex DR array consisting of A-motifs (39 nt) and B-motifs (24 nt) were observed (Figure 15). A closer inspection of motifs A and B showed a significant resemblance to the intron IGS, which base pair to the upstream exon and forming segment P1-like structures. We speculate that Mcr.S1389 may include these copies as putative internal guide sequences (ψ IGS) to regulate 5' SS selection (Figure 15). Interestingly, two other S1389 introns were also found to contain ψ IGS repeats in P9 (**Paper I**, Figure S3). Lar.S1389 and Lca.S1389 harbor DR with a similar ψ IGS feature compared to Mcr.S1389. The DR motifs in Lar.S1389 and Lca.S1389 resemble the ψ IGS from Mcr.S1389 and may support the possibility that DR can function as regulatory elements by their ability to regulate the 5' SS selection and then the splicing activity.

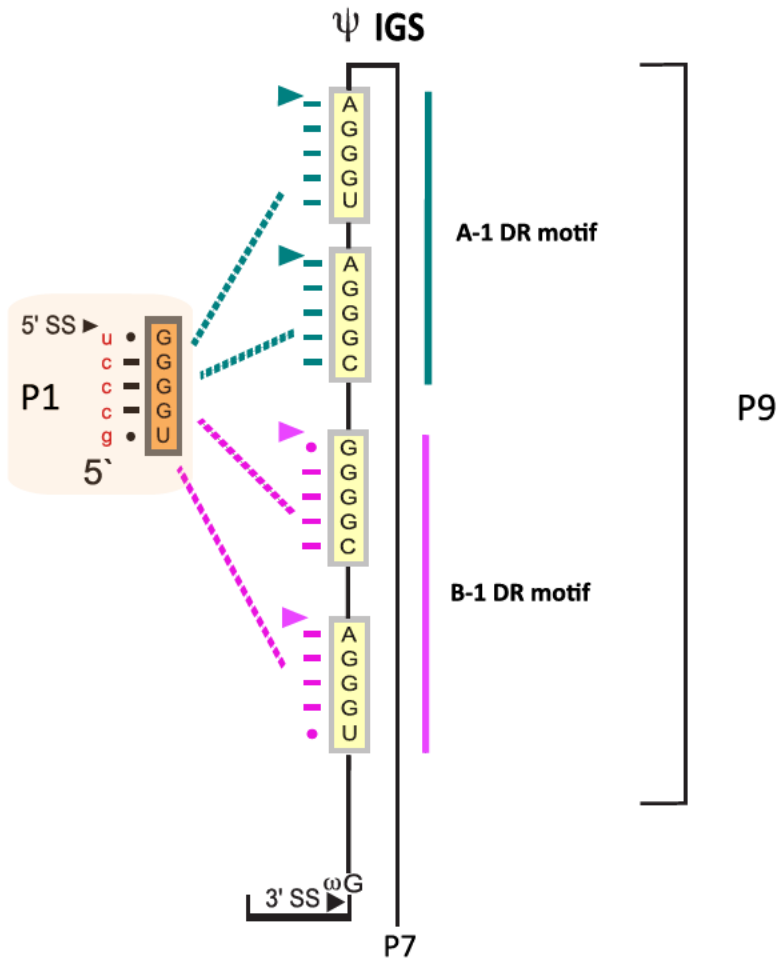
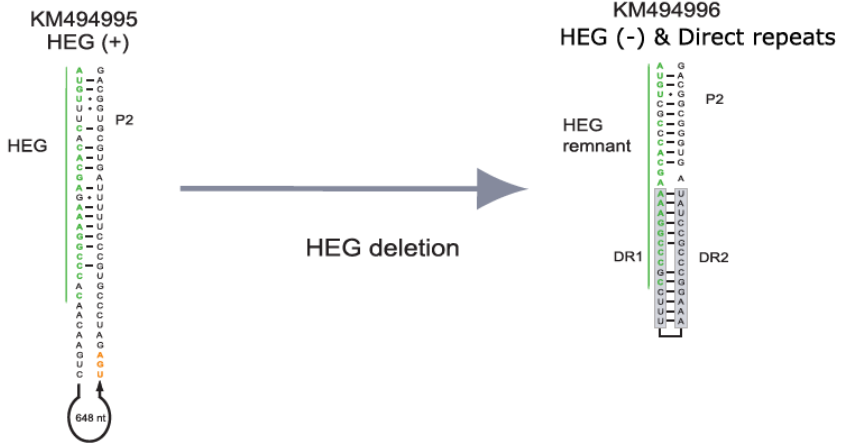


Figure 15: Direct repeat features. An illustration of the direct repeat (DR) features in the P9 segment of Mcr.S1389 intron and how they resemble the internal guide sequence (IGS) (dotted lines). (left) The P1 and the IGS (orange box). (right) DR motifs are observed in P9. Here, putative internal guide sequences (ψ IGS) (yellow box), DR motifs A1 (green), and B1 (pink) are shown. These can interact with the five-nucleotide exon sequence at the 5' splice site (5' SS) and alter the IGS (orange). 5' splice site (5' SS, arrow); 3' splice site (3' SS, arrow); conserved terminal guanosine (ω G); (-) Watson-Crick base pair; (•) non-Watson-Crick base pairing.

DR motifs from truncated HEGs

Structural examination of three larger Tvar.S516 (GenBank accession numbers: KM494994/95/93) intron sequences revealed HEG insertions in their P2 segments. A further sequence comparison with the other smaller Tvar.S516 intron sequences (22 isolates, ca. 443 bp) were performed. Here, the Tvar.S516 (KM494996) intron sequence was found to be a slightly larger in sequence (482 bp) and harbor two DR-motifs in the P2 segment. A more detailed structure analysis was performed. These DR-motifs resemble truncated HEG sequences (Figure 16b), and structural analysis suggests that the HEG deletion has given rise to two DR-motifs (Figure 16a).

A



B

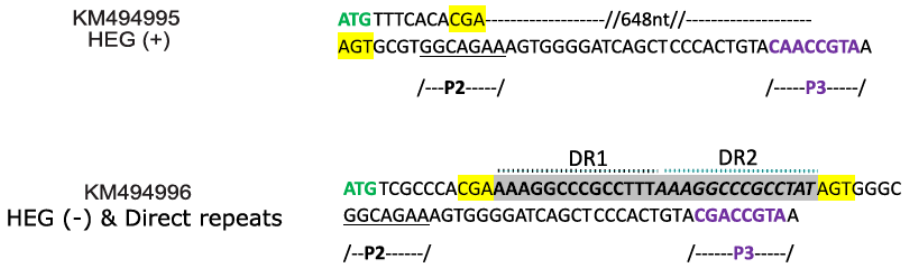


Figure 16: Direct repeat features in *Tricia varia* S516 introns. An illustration of the direct repeat (DR) features in *Tricia varia* S516 introns, GenBank accession numbers KM494995-6. (A) The P2 segment contains HEG sequence in the KM494995 (HEG +) isolate. An illustration of the HEG deletion, which results in HEG remnant (KM494996 (HEG -)) with DR 1 and 2 (DR1/DR2) is shown. (B) A sequence overview of the HEG of KM494995 (HEG +) isolate and KM494996 (HEG and direct repeat) isolate is illustrated. ATG (green letters), start codon. DRI and DR 2, direct repeat motifs 1 and 2; P2 (underlined) and P3 (purple) are indicated.

DR fold into RNA hairpin motifs

In the *Licea parasitica* isolate (Lpar.S516), a complex and interesting DR array with 25 copies of a 23 nt long motif was observed within the P2 segment (Figure 17). A more detailed structure analysis of the Lpar.S516 DR motifs revealed seven motif variants (named I to VII), which are repeated from one time and up to ten times. These motifs fold into structural hairpin motifs containing a tetraloop and a stem region with compensatory mutations. While their functional roles remain undetermined, these features strongly suggest an RNA structural role and a potential regulatory function.

Direct repeats

```

ATGTTGGGAAGCGAAACAAACCGGCTGGAGGGATACCTTTTCCGCGGCAGGAGGGATACCTTTTCCGCGG
GCAGGAGGGATACCTTTTCCGCGGCAGGAGGGATACCTTTTCCGCGGCAGGAGGGACACCTTTTCCGCGG
GCAGGAGGGACACCTTTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTTTTCCGCGG
GCAGGAGGGACACCTTTTCCGCGGCAGGAGGGATACCTTTTCCGCGGCAGGAGGGGACACCTCTTCCGCGG
GCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTTTTCCGCGGCAGGAGGGACACCTTTTCCGCGG
GCAGGAGGGATACCTTTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGG
GCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGG
GCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGG
GCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGG
GCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGG
GCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGGCAGGAGGGACACCTCTTCCGCGG
CCTGGCAGAAAGCCGGGATAAGCTCCGGCT
    
```

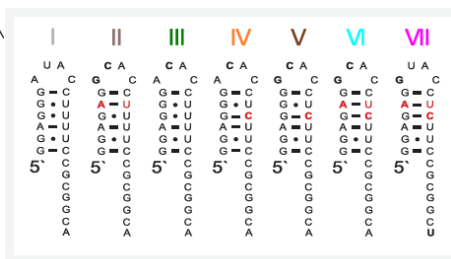


Figure 17: Direct repeat features in the *Licea parasitica* S516 intron. The sequences of the direct repeat (DR) motifs in *Licea parasitica* S516 intron are shown. The 25 DR motifs are marked, and each of the seven repeat motifs (I-VII) is marked with a specific color. An illustration of RNA hairpin structures are shown from each of the seven motifs (I-VII); Red and black bold letters represent mutations.

4.4 RNA sequence variations in the G-site

One RNA structural component (P7) of the group I intron core appears particularly important in self-splicing catalysis. P7 harbors the G-site and has been shown to be highly conserved among group I introns. Pioneer studies published more than 30 years ago and based on comparative sequence analyses of available intron sequences (Michel et al. 1989; Michel and Westhof 1990), identified a strict co-variations of nucleotides in the G-site. The main observations from the pioneer studies were:

- The second base pair in P7 is an invariant G:C pair. These positions correspond to G264 and C311 in the *Tetrahymena* intron (Figure 18, left).
- Adjacent to the G:C pair is an unpaired nucleotide (bulge), which can be either an A or a C. The bulge nucleotide corresponds to A263 in the *Tetrahymena* intron.
- If 'bulge A', then the first base pair in P7 is either a C:G or G:C.
- If 'bulge C', then the first base pair in P7 is either a A:U or U:A.

These simple rules appear still valid for thousands of group I introns noted in nature, supporting an essential role in RNA structure and catalysis. Three-dimensional structures of the G-site in the *Tetrahymena* intron, and a few other group I introns, further highlight the functional implications of P7 (Su et al. 2021; Bonilla et al. 2022; Zhang et al. 2023). In accordance with these rules the G-sites observed among the 75 S516 introns (**Paper II**, Figure S1) and 34 L2066 introns (**Paper III**, Figure S2) respectively, were practically invariant. Here, the S516 introns have a bulge A and G:C as the first P7 base pair, and L2066 introns have a bulge C and an A:U base pair.

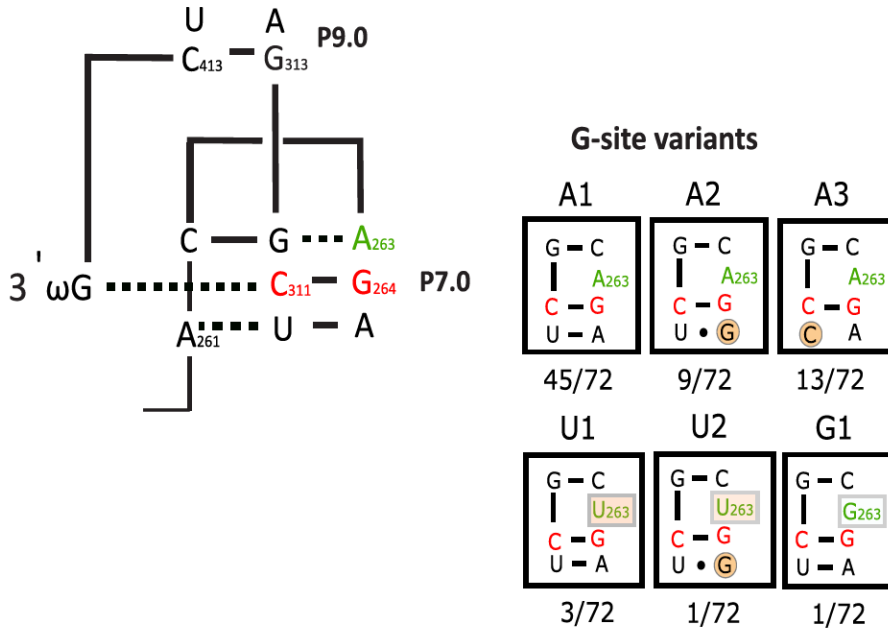


Figure 18: The G-site and the variants. An illustration of the guanosine-binding site (G-site) features within the P7 segment (left). The base triple interactions (marked as dash lines) and the ω G residue at the 3' site are indicated. The illustration is based on *Tetrahymena* intron from Guo et al. 2004. The G-site variations (right) identified in the S1389 intron family are shown in the black boxes (A1, A2, A3, U1, U2, and G1). The “bulge A” variants are boxed. The other nucleotides that differ from the conserved G-site are circled.

4.4.1 An exception from the invariant G-site

Identifying G-site variations within the S1389 intron family (**Paper I**) became a structural surprise compared to the common invariant nucleotide position. Two types of P7 structural variations were seen (Figure 18, right). The first type is located in the base pair adjacent to the essential G264-C311 pair, and is an A:U in the majority (68%) of S1389 introns (corresponds to *Tetrahymena* positions A265:U310). About 14% of the introns have a compensatory change in this base pair due to conversion from A:U to G:U. This variant still closes the P7 helical structure and has probably a minor, if any, effect on the G-site function. Interestingly, 18% of the S1389 introns have an A:C variant, indicating a lack of regular base pair interaction at this position. How an A:C

pair may affect the G-site structure and function in the S1389 introns is unknown, but this A:C pair appears highly conserved and invariant among S1506 nuclear rDNA introns in lichen fungi (Martín et al. 2003), and thus supports its functionality in splicing.

The second type of P7 structural variants observed in the S1389 introns is at the bulge nucleotide. Whereas the great majority (93%) of introns have an A in this position, five introns have other variants (Figure 18, right). Bulge U is represented in four S1389 introns and bulge G was observed only in one isolate. A mutagenesis study in the *Tetrahymena* intron from 1990 assessed different variants of the bulge nucleotide (Couture et al. 1990). They found that when the A-bulge became replaced by a G or a U, the self-splicing activity was reduced to 6.2% and 61%, respectively. Later, an *in vitro* selection experiment was performed on the *Tetrahymena* intron with a focus on P7 (Oe et al. 2000). Like the mutagenesis study by Couture and co-authors, splicing activity was observed on variants that carry both bulge U and bulge G. However, the *in vitro* selection study is hard to compare and interpret due to multiple changes in the P7 segment.

The observed S1389 intron with bulge G in *Diderma testaceum* (Figure 6, **Paper I**) may be due to a spontaneous A to G mutation in this isolate. Considering the mutagenesis study discussed above, a bulge G variant may self-splice very poorly under natural conditions. This intron could be a candidate for new host interactions that facilitate splicing *in vivo*. The different S1389 introns with bulge U were observed in four isolates representing four distantly genera (*Didymium*, *Comatricha*, *Diachea*, and *Lamproderma*) representing two myxomycete orders (Table S1, **Paper I**). The phylogenetic analysis strongly favors an A to U mutation as a sporadic feature, which may be preserved in the natural intron since the splicing activity is only 1/3 reduced (according to the study of Couture et al. 1990).

Bulge U in P7 is unusual among group I introns, but conserved variants are noted in L2449 introns of *Physarum* myxomycetes belonging to the distinct phylogenetic clade III (Nandipati et al. 2012). Different from that of S1389, the *Physarum* introns harbor a

C to U mutation (not A to U) in P7. However, the myxomycete S1389 and L2449 families have an important feature in common. Both intron families appear not able to self-splice *in vitro* at standard conditions, suggesting host dependencies in intron RNA folding and/or catalysis. Perhaps represent the observed G-site variants an additional step in escaping the Goddard-Burt cycle of intron evolution in becoming fixated and more host-dependent (Step 5c, Figure 12).

5 Future perspectives

This thesis provides additional knowledge to the structure organization and molecular evolution of catalytic RNA introns from eukaryotic microorganisms, especially the myxomycetes. We discovered some new and undescribed structural features among the nuclear group I introns. Notably, most assessed introns followed the pioneer rules of structure organization, first reported 30-40 years ago. However, the structural, functional, and evolutionary aspects of the three nuclear group I intron families presented in this thesis are remained to be revealed in more detail. This opens new perspectives for future studies. First, no functional studies on the G-site variation observations were conducted during this study. Therefore, it would be interesting to investigate the effect of the G-site variation by testing the self-splicing activities of introns with different G-site variations, then inspect for signs of intron circularization and lastly gain more structural information by 3D structure analysis (e.g., by Cryo-EM) (Su et al. 2021; Bonilla and Kieft 2022; Li et al. 2023a; Zhang et al. 2023). Second, a deeper investigation of the DR-motifs is needed. Structural and functional activities of introns based on specific DR motifs would be interesting to investigate in greater details. Here, the DR array in Lpar.S516 and the DR motifs in Mcr.S1389 (ψ IGS) are of particular interest. Activities can be assessed on designing intron variants with varying DR-motifs. Third, to expand the analysis to include several other datasets of other intron families. As stated in the introduction section (see Figure 2), at least 23 different group I intron families have been identified in myxomycetes rDNA, and only three of those have been investigated here. This will increase the robustness of the HEG spliceosomal intron hypothesis and the molecular evolution. From another perspective, the ongoing compilation of large-scale datasets and genomes worldwide would increase the probability of identifying novel group I introns and expand the analysis further, to include novel intron families. This may significantly advance our knowledge in molecular evolution and RNA structural characterization of catalytic RNA introns.

6 Reference list

- Adams PL, Stahley MR, Gill ML, Kosek AB, Wang J and Strobel SA. (2004) Crystal structure of a group I intron splicing intermediate. *RNA* 10:1867-1887. <https://doi.org/10.1038/nature02642>
- Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, et al. (2019) Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology* 66:4-119. <https://doi.org/10.1111/jeu.12691>
- Adl SM, Simpson AGB, Lane CE, Lukeš J, Bass D, Bowser SS, et al. (2012) The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology* 59:429-514. <https://doi.org/10.1111/j.1550-7408.2012.00644.x>
- Aldrich HC and Daniel JW. (1982) Cell biology of *Physarum* and *Didymium* (1th edn.). Academic Press.
- Andersen KL, Beckert B, Masquida B, Johansen SD and Nielsen H. (2016) Accumulation of stable full-length circular group I intron RNAs during heat-shock. *Molecules* 21. <http://doi.org/10.3390/molecules21111451>
- Assmann SM, Chou H-L and Bevilacqua PC. (2023) Rock, scissors, paper: How RNA structure informs function. *Plant Cell* 35 (6):1671-1707 <https://doi.org/10.1093/plcell/koad026>
- Been MD and Perrotta AT. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site. *Science* 252:434-437. <https://doi.org/10.1126/science.2017681>
- Belfort M and Roberts RJ. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res* 25: 3379-3388. <https://doi.org/10.1093/nar/25.17.3379>
- Bhattacharya D, Cannone JJ and 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 (1998). The origin and evolution of protist group I introns. *Protist* 149:113-122. [https://doi.org/10.1016/S1434-4610\(98\)70015-X](https://doi.org/10.1016/S1434-4610(98)70015-X)
- Bhatti GK, Khullar N, Sidhu IS, Navik US, Reddy AP, Reddy PH and Bhatti JS (2021). Emerging role of non-coding RNA in health and disease. *Metab Brain Dis* 36:1119-1134. <https://doi.org/10.1007/s11011-021-00739-y>
- Birgisdottir AB and Johansen SD. (2005) Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. *Nucleic Acids Res* 33:2042-2051. <https://doi.org/10.1093/nar/gki341>
- Bonilla SL and Kieft JS. (2022) The promise of cryo-EM to explore RNA structural dynamics. *J Mol Biol* 434:167802. <https://doi.org/10.1016/j.jmb.2022.167802>
- Bonilla SL, Vicens Q and Kieft JS. (2022) Cryo-EM reveals an entangled kinetic trap in the folding of a catalytic RNA. *Sci Adv* 8:eabq4144. <https://doi.org/10.1126/sciadv.abq4144>
- Brown TA. (2017) *Genomes* 4 (4th edn.). Garland Science pp. 12.

- Burland TG, Solnica-Krezel L, Bailey J, Cunningham DB and Dove WF. (1993) Patterns of inheritance, development and the mitotic cycle in the protist *Physarum polycephalum*. *Adv Microb Physiol* 35:1-69.
[https://doi.org/10.1016/s0065-2911\(08\)60096-x](https://doi.org/10.1016/s0065-2911(08)60096-x)
- Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'souza LM, Du Y, et al. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3:2. <https://doi.org/10.1186/1471-2105-3-2>
- Cavalli-Sforza LL and Edwards AW. (1967) Phylogenetic analysis. models and estimation procedures. *Am J Hum Genet* 19:233-257.
- Cech TR. (1990) Self-splicing of group I introns. *Annu Rev Biochem* 59:543-568.
<https://doi.org/10.1146/annurev.bi.59.070190.002551>
- Cech TR and Steitz JA. (2014) The Noncoding RNA Revolution—trashing old rules to forge new ones. *Cell* 157:77-94. <https://doi.org/10.1016/j.cell.2014.03.008>
- Chen J, Tsai A, O'leary SE, Petrov A and Puglisi JD. (2012) Unraveling the dynamics of ribosome translocation. *Curr Opin Struct Biol* 22:804-814.
<https://doi.org/10.1016/j.sbi.2012.09.004>
- Chevalier BS and 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>
- Cho Y and Palmer JD. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial *cox1* gene during evolution of the Araceae family. *Mol Biol Evol* 16:1155-1165.
<https://doi.org/10.1093/oxfordjournals.molbev.a026206>
- Cho Y, Qiu YL, Kuhlman P and Palmer JD. (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc Natl Acad Sci U.S.A* 95:14244-14249.
<https://doi.org/10.1073/pnas.95.24.14244>
- Collins ONR and Betterley DA. (1982a) *Didymium iridis* in past and future research: reproductive cycles. In: Henry C. Aldrich and Daniel J. W. (eds) *Cell biology of Physarum and Didymium: organisms, nucleus, and cell cycle* (1th edn.). Academic Press pp. 28-34.
- Collins ONR and Betterley DA. (1982b) *Didymium iridis* in past and future research: chromosomal numbers and nuclear DNA contents. In: Henry C. Aldrich and Daniel J. W. (eds) *Cell biology of Physarum and Didymium: organisms, nucleus, and cell cycle* (1th edn.). Academic Press pp. 35-37.
- Couture S, Ellington AD, Gerber AS, Cherry JM, Doudna JA, Green R, et al. (1990) Mutational analysis of conserved nucleotides in a self-splicing group I intron. *J Mol Biol* 215: 345-358. [https://doi.org/10.1016/s0022-2836\(05\)80356-0](https://doi.org/10.1016/s0022-2836(05)80356-0)
- Damberger SH and Gutell RR (1994). A comparative database of group I intron structures. *Nucleic Acids Res* 22:3508-3510.
<https://doi.org/10.1093/nar/22.17.3508>
- Decatur WA, Einvik C, Johansen SD and Vogt VM. (1995) Two group I ribozymes with different functions in a nuclear rDNA intron. *EMBO J* 14:4558-4568.
<https://doi.org/10.1002/j.1460-2075.1995.tb00135.x>

- Decatur WA and Fournier MJ. (2002) rRNA modifications and ribosome function. *Trends Biochem Sci* 27: 344-351.
[https://doi.org/10.1016/s0968-0004\(02\)02109-6](https://doi.org/10.1016/s0968-0004(02)02109-6)
- Deng J, Shi Y, Peng X, He Y, Chen X, Li M, et al. (2022) Ribocentre: a database of ribozymes. *Nucleic Acids Res* 51(D1):D262–D268.
<https://doi.org/10.1093/nar/gkac840>
- Doudna JA, Cormack BP and Szostak JW. (1989) RNA structure, not sequence, determines the 5' splice-site specificity of a group I intron. *Proc Natl Acad Sci* 86:7402-7406. <https://doi.org/10.1073/pnas.86.19.7402>
- Dove WF, Dee J, Hatano S, Haugli FB, Wohlfarth-Bottermann K-E. (1986) The molecular biology of *Physarum polycephalum* (1th edn.). Springer NY
- Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, Perlman PS and Vogt VM. (1989) Mobile introns: definition of terms and recommended nomenclature. *Gene* 82:115-118. [https://doi.org/10.1016/0378-1119\(89\)90035-8](https://doi.org/10.1016/0378-1119(89)90035-8)
- Dunn MJ and Anderson MZ. (2019) To repeat or not to repeat: repetitive sequences regulate genome stability in *Candida albicans*. *Genes* 10:866.
<https://doi.org/10.3390/genes10110866>
- Edgell DR, Belfort M and Shub DA. (2000) Barriers to intron promiscuity in bacteria. *J Bacteriol* 182:5281-5289. <https://doi.org/10.1128/jb.182.19.5281-5289.2000>
- Edgell DR, Chalamcharla VR and Belfort M. (2011) Learning to live together: mutualism between self-splicing introns and their hosts. *BMC Biol* 9:22.
<https://doi.org/10.1186/1741-7007-9-22>
- Farris JS. (1970) Methods for computing wagner trees. *Syst Biol* 19:83-92.
<https://doi.org/10.1093/sysbio/19.1.83>
- Farris JS, Kallersjo M, Kluge AG and Bult C. (1995) Constructing a significance test for incongruence. *Syst Biol* 44:570-572. <https://doi.org/10.2307/2413663>
- Fedor MJ and Williamson JR. (2005) The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol* 6:399-412. <https://doi.org/10.1038/nrm1647>
- Felsenstein J (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17:368-376. <https://doi.org/10.1007/BF01734359>
- Felsenstein J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791.
<https://doi.org/10.1111/j.1558-5646.1985.tb00420.x>
- Fiore - Donno AM, Berney C, Pawlowski J and Baldauf SL. (2005) Higher - order phylogeny of plasmodial slime molds (myxogastria) based on elongation factor 1- A and small subunit rRNA gene sequences. *J Eukaryot Microbiol* 52 (3):201-210
<https://doi.org/10.1111/j.1550-7408.2005.00032.x>
- Fitch WM. (1971) Toward defining the dourse of evolution: minimum change for a specific tree topology. *Syst Biol* 20:406-416.
<https://doi.org/10.1093/sysbio/20.4.406>

- Flick KE, Jurica MS, Monnat RJ and Stoddard BL. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* 394:96-101. <https://doi.org/10.1038/27952>
- Fotsing SF, Margoliash J, Wang C, Saini S, Yanicky R, Shleizer-Burko S, et al. (2019) The impact of short tandem repeat variation on gene expression. *Nat Genet* 51:1652-1659. <https://doi.org/10.1038/s41588-019-0521-9>
- Galburt EA and Jurica MS. (2005) His-Cys box homing endonucleases. In: Belfort M., Wood D. W., Stoddard B. L. and Derbyshire V. (eds) *Homing endonucleases and inteins*. Berlin, Heidelberg: Springer Berlin Heidelberg pp. 85-102.
- Galburt EA and Stoddard BL. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry* 41:13851-13860. <https://doi.org/10.1021/bi020467h>
- Ganser LR, Kelly ML, Herschlag D and Al-Hashimi HM. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* 20:474-489. <https://doi.org/10.1038/s41580-019-0136-0>
- Garst AD, Edwards AL and Batey RT. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol* 3(6):a003533 <https://doi.org/10.1101/cshperspect.a003533>
- Geneious. (2019) Geneious Prime 2019.2.3. <https://www.geneious.com>, Accessed 22 Nov 2023
- Goddard MR and Burt A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U.S.A* 96:13880-13885. <https://doi.org/10.1073/pnas.96.24.13880>
- Goddard MR, Leigh J, Roger AJ and 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>
- Golden BL, Gooding AR, Podell ER and Cech TR. (1998) A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* 282:259-264. <https://doi.org/10.1126/science.282.5387.259>
- Golden BL, Kim H and Chase E. (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12:82-89. <https://doi.org/10.1038/nsmb868>
- Gorman JA and Wilkins AS. (2017) Developmental phases in the life cycle of *Physarum* and related myxomycetes. In: Dove W. F. and Rusch H. P. (eds) *Growth and differentiation in Physarum polycephalum* (Volume 5029). Princeton University Press pp. 157-202.
- Guo F, Gooding AR and Cech TR. (2004) Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16:351-362. <https://doi.org/10.1016/j.molcel.2004.10.003>
- Hafez M and Hausen G. (2012) Homing endonucleases: DNA scissors on a mission. *Genome* 55:553-569. <https://doi.org/10.1139/g2012-049>
- Hall BG. (2011a) Major methods for estimating phylogenetic trees. *Phylogenetic trees made easy: a how-to manual* (4th edn.). Sinauer Associates is an imprint of Oxford University Press pp. 64-65.

- Hall BG. (2011b) Parsimony. *Phylogenetic trees made easy: a how-to manual* (4th edn.). Sinauer Associates is an imprint of Oxford University Press pp. 111-113.
- Hall BG. (2011c) Maximum likelihood. *Phylogenetic trees made easy: a how-to manual* (4th edn.). Sinauer Associates is an imprint of Oxford University Press, pp.123-125.
- Hall BG. (2011d) Bayesian inference of trees using MrBayes. *Phylogenetic trees made easy: a how-to manual* (4th edn.). Sinauer Associates is an imprint of Oxford University Press pp.141-142.
- Hall BG. (2011e) Major methods for estimating phylogenetic trees. *Phylogenetic trees made easy: a how-to manual* (4th edn.). Sinauer Associates is an imprint of Oxford University Press pp. 68.
- Hancock JM and Dover GA. (1988) Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol Biol Evol* 5:377-391. <https://doi.org/10.1093/oxfordjournals.molbev.a040505>
- Haugen P, Coucheron DH, Rønning SB, Haugli K and Johansen SD. (2003) The molecular evolution and structural organization of self-splicing group I introns at position 516 in nuclear SSU rDNA of myxomycetes. *J Eukaryot Microbiol* 50:283-292. <https://doi.org/10.1111/j.1550-7408.2003.tb00135.x>
- Haugen P, Huss VA, Nielsen H and Johansen SD. (1999) Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing-endonuclease pseudogenes in the Bangiophyceae. *Curr Genet* 36:345-353. <https://doi.org/10.1007/s002940050509>
- Haugen P, Runge HJ and Bhattacharya D. (2004) Long-term evolution of the S788 fungal nuclear small subunit rRNA group I introns. *RNA* 10:1084-1096. <https://doi.org/10.1261/rna.5202704>
- Haugen P, Wikmark O-G, Vader A, Coucheron DH, Sjøttem E and Johansen SD. (2005a) The recent transfer of a homing endonuclease gene. *Nucleic Acids Research*, 33: 2734-2741. <https://doi.org/10.1093/nar/gki564>
- Haugen P, Simon DM and Bhattacharya D. (2005b) The natural history of group I introns. *Trends Genet* 21:111-119. <https://doi.org/10.1016/j.tig.2004.12.007>
- Hedberg A and 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>
- Higgins D and Lemey P. (2009) Multiple sequence alignment. In: Lemey P., Salemi M. and Vandamme A. M. (eds) *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing* (2th edn.): Cambridge University Press pp. 326-327.
- Huelsenbeck JP and Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Ikawa Y, Yoshimura T, Hara H, Shiraishi H and Inoue T. (2002) Two conserved structural components, A-rich bulge and P4 XJ6/7 base-triples, in activating the group I ribozymes. *Genes Cells* 7:1205-1215. <https://doi.org/10.1046/j.1365-2443.2002.00601.x>

- Jackson S, Cannone J, Lee J, Gutell R and Woodson S. (2002) Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J Mol Biol* 323:35-52. [https://doi.org/10.1016/s0022-2836\(02\)00895-1](https://doi.org/10.1016/s0022-2836(02)00895-1)
- Jaeger L, Michel F and Westhof E. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J Mol Biol* 236:1271-1276. [https://doi.org/10.1016/0022-2836\(94\)90055-8](https://doi.org/10.1016/0022-2836(94)90055-8)
- Jaeger L, Michel F and Westhof E. (1997) The structure of group I ribozymes. In: Eckstein F. and Lilley D. M. J. (eds) *Catalytic RNA*. Berlin, Heidelberg: Springer Berlin Heidelberg pp. 33–51.
- Johansen S, Elde M, Vader A, Haugen P, Haugli K and Haugli F. (1997) *In vivo* mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. *Mol Microbiol* 24:737-745. <https://doi.org/10.1046/j.1365-2958.1997.3921743.x>
- Johansen SD and Emblem Å. (2020) Mitochondrial group I introns in hexacorals are regulatory genetic elements. In: Luis A. S. (ed) *Advances in the studies of the benthic zone*. IntechOpen ch. 7. <https://doi.org/10.5772/intechopen.91465>
- Johansen SD, Embley TM and Willassen NP. (1993) A family of nuclear homing endonucleases. *Nucleic acids research* 21:4405. <https://doi.org/10.1093%2Fnar%2F21.18.4405>
- Johansen SD and Haugen P. (2001) A new nomenclature of group I introns in ribosomal DNA. *RNA* 7:935-936. <https://doi.org/10.1017/s1355838201010500>
- Johansen SD, Haugen P and Nielsen H. (2007) Expression of protein-coding genes embedded in ribosomal DNA. *Biol Chem* 388:679-686. <https://doi.org/10.1515/BC.2007.089>
- Johansen SD, Johansen T and Haugli F. (1992a) Structure and evolution of myxomycete nuclear group I introns: a model for horizontal transfer by intron homing. *Curr Genet* 22:297-304. <https://doi.org/10.1007/BF00317925>
- Johansen SD, Johansen T and Haugli F. (1992b) Extrachromosomal ribosomal DNA of *Didymium iridis*: sequence analysis of the large subunit ribosomal RNA gene and sub-telomeric region. *Curr Genet* 22:305-312. <https://doi.org/10.1007/BF00317926>
- Johansen SD, Vader A, Sjøttem E and Nielsen H. (2006) *In vivo* Expression of a group I intron HEG from the antisense strand of *Didymium* ribosomal DNA. *RNA Biol* 3:157-162. <https://doi.org/10.4161/rna.3.4.3958>
- Johansen SD and Vogt VM. (1994) An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* 76:725-734. [https://doi.org/10.1016/0092-8674\(94\)90511-8](https://doi.org/10.1016/0092-8674(94)90511-8)
- Jurica MS and Stoddard BL. (1999) Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci* 55:1304-1326. <https://doi.org/10.1007/s000180050372>
- Katoh K, Misawa K, Kuma K-I and Miyata T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059-3066. <https://doi.org/10.1093/nar/gkf436>

- Kluge AG and Farris JS. (1969) Quantitative phyletics and the evolution of anurans. *Syst Biol* 18: 1-32. <https://doi.org/10.1093/sysbio/18.1.1>
- Knitt DS, Narlikar GJ and Herschlag D. (1994) Dissection of the role of the conserved G:U pair in group I RNA self-splicing. *Biochemistry* 33: 13864-13879. <https://doi.org/10.1021/bi00250a041>
- Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE and Cech TR. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31:147-157. [https://doi.org/10.1016/0092-8674\(82\)90414-7](https://doi.org/10.1016/0092-8674(82)90414-7)
- Kuhse MG, Strickland R and Palmer JD. (1990) An ancient group I intron shared by eubacteria and chloroplasts. *Science* 250:1570-1573. <https://doi.org/doi:10.1126/science.2125748>
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921. <https://doi.org/10.1038/35057062>
- Lando C. (2005-2022) An on line nomenclatural information system of Eumycetozoa. Real Jardín Botánico, CSIC. Madrid. Available from: <http://www.eumycetozoa.com>. Accessed 10 Nov 2023
- Lado C and De Basanta DW. (2008) A review of neotropical myxomycetes (1828-2008). *An Jard Bot Madr* 65(2):211-254
- Lado C, Estrada-Torres A and Alvarado CR. (2018) New records of genera and species of myxomycetes (Amoebozoa) from the Neotropics. *Check List* 14 (2):509-518 <https://doi.org/10.15560/14.3.509>
- Lado C, Treviño-Zevallos I, García-Martín JM and Wrigley De Basanta D. (2022) *Diachea mitchellii*: A new myxomycete species from high elevation forests in the tropical Andes of Peru. *Mycologia* 114:798-811. <https://doi.org/10.1080/00275514.2022.2072140>
- Laffler TG and Tyson JJ. (1986) The *Physarum* cell cycle. In: Dove W. F., Dee J., Hatano S., Haugli F. B. and Wohlfarth-Bottermann K.-E. (eds) *The molecular biology of Physarum polycephalum* (volume 106). Springer NY pp. 79-109.
- Lambowitz AM and Belfort M. (1993) Introns as mobile genetic elements. *Annu Rev Biochem* 62:587-622. <https://doi.org/10.1146/annurev.bi.62.070193.003103>
- Lavigne R and Vandersteegen K. (2013) Group I introns in *Staphylococcus* bacteriophages. *Future Virology* 8:997-1005. <https://doi.org/10.2217/fvl.13.84>
- Lehnert V, Jaeger L, Michel F and Westhof E. (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem Biol* 3:993-1009. [https://doi.org/10.1016/s1074-5521\(96\)90166-0](https://doi.org/10.1016/s1074-5521(96)90166-0)
- Levinson G and Gutman GA (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203-221. <https://doi.org/10.1093/oxfordjournals.molbev.a040442>

- Li S, Palo MZ, Zhang X, Pintilie G and Zhang K. (2023a) Snapshots of the second-step self-splicing of *Tetrahymena* ribozyme revealed by cryo-EM. *Nat Commun* 14: 1294. <https://doi.org/10.1038/s41467-023-36724-5>
- Li S, Palo MZ, Pintilie G, et al. (2022) Topological crossing in the misfolded *Tetrahymena* ribozyme resolved by cryo-EM. *Proc Natl Acad Sci* 119:e2209146119. <https://doi.org/10.1073/pnas.2209146119>
- Li S, Qi B, Peng X, Wang W, Wang W, Liu P, et al. (2023b) Genome size and GC content of myxomycetes. *Eur J Protistol* 90:125991. <https://doi.org/10.1016/j.ejop.2023.125991>
- Li Z and 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>
- Lilley DMJ. (2011) Mechanisms of RNA catalysis. *Philos Trans R Soc Lond B Biol Sci* 366:2910-2917. <https://doi.org/10.1098/rstb.2011.0132>
- Lilley DM and Eckstein F. (2008) Ribozymes and RNA catalysis: Introduction and primer. In: Lilley D. M. and Eckstein F. (eds) *Ribozymes and RNA catalysis*. Royal Society of Chemistry pp. 1-3.
- Lindley LA, Stephenson SL and Spiegel FW. (2007) Protostelids and myxomycetes isolated from aquatic habitats. *Mycologia* 99:504-509. <https://doi.org/10.3852/mycologia.99.4.504>
- López-Jiménez E and Andrés-León E. (2021) The implications of ncRNAs in the development of human diseases. *Noncoding RNA* 7(1):17. <https://doi.org/10.3390/ncrna7010017>
- Lundblad EW, Einvik C, Rønning S, Haugli K and Johansen SD. (2004) Twelve group I introns in the same pre-rRNA transcript of the myxomycete *Fuligo septica*: RNA processing and evolution. *Mol Biol Evol* 21:1283-1293. <https://doi.org/10.1093/molbev/msh126>
- Martín MP, Coucheron DH and Johansen SD. (2003) Structural features and evolutionary considerations of group IB introns in SSU rDNA of the lichen fungus *Teloschistes*. *Fungal Genet Biol* 40:252-260. <https://doi.org/10.1016/j.fgb.2003.07.001>
- Meyer M, Nielsen H, Oliéric V, Roblin P, Johansen SD, Westhof E and Masquida B. (2014) Speciation of a group I intron into a lariat capping ribozyme. *Proc Natl Acad Sci U S A* 111:7659-7664. <https://doi.org/10.1073/pnas.1322248111>
- Michel F, Hanna M, Green R, Bartel DP and Szostak JW. (1989) The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* 342:391-395. <https://doi.org/10.1038/342391a0>
- Michel F and 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](https://doi.org/10.1016/0022-2836(90)90386-Z)
- Miller D, Padmanabhan R and Sarcar SN. (2022) Chapter 5 - Genomics and gene expression in myxomycetes. In: Rojas C. and Stephenson S. L. (eds) *Myxomycetes* (2th edn.). Academic Press pp. 153-193.

- Mohan S, Donohue JP and Noller HF. (2014) Molecular mechanics of 30S subunit head rotation. *Proc Natl Acad Sci U S A* 111:13325-13330.
<https://doi.org/10.1073/pnas.1413731111>
- Mullineux S-T, Willows K and Hausner G. (2011) Evolutionary dynamics of the mS952 intron: a novel mitochondrial group II intron encoding a LAGLIDADG homing endonuclease gene. *J Mol Evol* 72:433-449.
<https://doi.org/10.1007/s00239-011-9442-7>
- Muscarella DE, Ellison EL, Ruoff BM and Vogt VM. (1990) Characterization of *I-Ppo*, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Mol Cell Biol* 10:3386-3396.
<https://doi.org/10.1128/mcb.10.7.3386-3396.1990>
- Muscarella DE and Vogt VM. (1989) A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* 56:443-454.
[https://doi.org/10.1016/0092-8674\(89\)90247-x](https://doi.org/10.1016/0092-8674(89)90247-x)
- Nandipati SCR, Haugli K, Coucheron DH, Haskins EF and Johansen SD. (2012) Polyphyletic origin of the genus *Physarum* (Physarales, Myxomycetes) revealed by nuclear rDNA mini-chromosome analysis and group I intron synapomorphy. *BMC Evol Biol* 12:166. <https://doi.org/10.1186/1471-2148-12-166>
- Nawrocki EP, Jones TA and Eddy SR. (2018) Group I introns are widespread in archaea. *Nucleic Acids Res* 46:7970-7976. <https://doi.org/10.1093/nar/gky414>
- NCBI. (1988) National Center for Biotechnology Information (NCBI).
<https://www.ncbi.nlm.nih.gov/>, Accessed 22 Nov 2023
- Nielsen H, Einvik C, Lentz TE, Hedegaard MM and Johansen SD. (2009) A conformational switch in the DiGIR1 ribozyme involved in release and folding of the downstream *I-Dirl* mRNA. *RNA* 15:958-967.
<https://doi.org/10.1261/rna.669209>
- Nielsen H, Fiskaa T, Birgisdottir AB, Haugen P, Einvik C and Johansen S. (2003) The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* 9:1464-1475. <https://doi.org/10.1261/rna.5290903>
- Nielsen H and Johansen SD. (2009) Group I introns: Moving in new directions. *RNA Biol* 6:375-383. <https://doi.org/10.4161/rna.6.4.9334>
- Nielsen H, Krogh N, Masquida B and Johansen SD. (2021) The lariat capping ribozyme. pp. 117-142. <https://doi.org/10.1002/9783527814527.ch5>
- Nielsen H, Westhof E and Johansen SD. (2005) An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. *Science* 309:1584-1587.
<https://doi.org/10.1126/science.1113645>
- Nikoh N and Fukatsu T. (2001) Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. *Mol Biol Evol* 18:1631-1642.
<https://doi.org/10.1093/oxfordjournals.molbev.a003952>
- Nishida K, Suzuki S, Kimura Y, Nomura N, Fujie M and Yamada T. (1998) Group I introns found in *Chlorella* viruses: Biological implications. *Virology* 242:319-326.
<https://doi.org/10.1006/viro.1998.9030>

- Noller HF. (2006) Biochemical characterization of the ribosomal decoding site. *Biochimie* 88:935-941. <https://doi.org/10.1016/j.biochi.2006.04.006>
- Nomiyama H, Kuhara S, Kukita T, Otsuka T and Sakaki Y. (1981a) Nucleotide sequence of the ribosomal RNA gene of *Physarum polycephalum*: intron 2 and its flanking regions of the 26S rRNA gene. *Nucleic Acids Res* 9:5507-5520. <https://doi.org/10.1093/nar/9.21.5507>
- Nomiyama H, Sakaki Y and Takagi Y. (1981b) Nucleotide sequence of a ribosomal RNA gene intron from slime mold *Physarum polycephalum*. *Proc Natl Acad Sci U S A* 78:1376-1380. <https://doi.org/10.1073/pnas.78.3.1376>
- Notredame C, Higgins DG and Heringa J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment 11 Edited by J. Thornton. *J Mol Biol* 302:205-217. <https://doi.org/10.1006/jmbi.2000.4042>
- Oe Y, Ikawa Y, Shiraishi H and Inoue T. (2000) Analysis of the P7 region within the catalytic core of the *Tetrahymena* ribozyme by employing *in vitro* selection. *Nucleic Acids Symp Ser* 44(1):197-198. <https://doi.org/10.1093/nass/44.1.197>
- Pauli A, Rinn JL and Schier AF. (2011) Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet* 12:136-149. <https://doi.org/10.1038/nrg2904>
- Pitts S and Laiho M. (2022) Regulation of RNA polymerase I stability and function. *Cancers* 14(23):5776. <https://doi.org/10.3390/cancers14235776>
- Posey KL, Koufopanou V, Burt A and Gimble FS. (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic Acids Res* 32:3947-3956. <https://doi.org/10.1093/nar/gkh734>
- Prensner JR and Chinnaiyan AM. (2011) The emergence of lncRNAs in cancer biology. *Cancer Discov* 1:391-407. <https://doi.org/10.1158/2159-8290.CD-11-0209>
- Prikhodko IS, Shchepin ON, Bortnikova NA, Novozhilov YK, Gmoshinskiy VI, Moreno G, et al. (2023) A three-gene phylogeny supports taxonomic rearrangements in the family *Didymiaceae* (Myxomycetes). *Mycol Prog* 22:11. <https://doi.org/10.1007/s11557-022-01858-1>
- Quirk SM, Bell-Pedersen D and Belfort M. (1989) Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames. *Cell* 56:455-465. [https://doi.org/10.1016/0092-8674\(89\)90248-1](https://doi.org/10.1016/0092-8674(89)90248-1)
- Richard G-F, Kerrest A and Dujon B. (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev* 72:686-727. <https://doi.org/10.1128/MMBR.00011-08>
- Rodriguez MS, Dargemont C and Stutz F. (2004) Nuclear export of RNA. *Biol Cell* 96:639-655. <https://doi.org/10.1016/j.biolcel.2004.04.014>
- Rojas C and Stephenson SL. (2017a) Introduction. *Myxomycetes: Biology, Systematics, Biogeography and Ecology* (1th edn.). Academic Press pp. xvii.
- Rojas C and Stephenson SL. (2017b) *Myxomycetes: biology, systematics, biogeography and ecology* (1th edn.). Academic Press
- Roman J, Rubin MN and Woodson SA. (1999) Sequence specificity of *in vivo* reverse splicing of the *Tetrahymena* group I intron. *RNA* 5:1-13. <https://doi.org/10.1017/s1355838299981244>

- Roman J and Woodson SA. (1995) Reverse splicing of the *Tetrahymena* IVS: evidence for multiple reaction sites in the 23S rRNA. *RNA* 1:478-490.
- Roman J and Woodson SA. (1998) Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing *in vivo*. *Proc Natl Acad Sci U S A* 95:2134-2139. <https://doi.org/10.1073/pnas.95.5.2134>
- Ronquist F and Huelsenbeck JP. (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:572-1574. <https://doi.org/10.1093/bioinformatics/btg180>
- Ruoff B, Johansen SD and Vogt VM. (1992) Characterization of the self-splicing products of a mobile intron from the nuclear rDNA of *Physarum polycephalum*. *Nucleic Acids Res* 20:5899-5906. <https://doi.org/10.1093/nar/20.22.5899>
- Saitou N and Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sandegren L and Sjöberg B-M. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J Biol Chem* 279:22218-22227.
- Sauer HW. (1982) *Developmental biology of Physarum*. Cambridge University Press.
- Schier AC and Taatjes DJ. (2020) Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev* 34:465-488. <https://doi.org/10.1101/gad.335679.119>
- Schlautmann LP and Gehring NH. (2020) A day in the life of the exon junction complex. *Biomolecules* 10:866. <https://doi.org/10.3390/biom10060866>
- Schuster A, Lopez JV, Becking LE, Kelly M, Pomponi SA, Wörheide G, et al. (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>
- Schaap P, Barrantes I, Minx P, Sasaki N, Anderson RW, Bénard M, et al. (2016) The *Physarum polycephalum* genome reveals extensive use of prokaryotic two-component and metazoan-type tyrosine kinase signaling. *Genome Biol Evol* 8:109-125. <https://doi.org/10.1093/gbe/evv237>
- Sellem CH, D'aubenton-Carafa Y, Rossignol M and Belcour L. (1996) Mitochondrial intronic open reading frames in *Podospora*: mobility and consecutive exonic sequence variations. *Genetics* 143:777-788. <https://doi.org/10.1093/genetics/143.2.777>
- Shinohara ML, Lobuglio KF and Rogers SO. (1996) Group-I intron family in the nuclear ribosomal RNA small subunit genes of *Cenococcum geophilum* isolates. *Curr Genet* 29:377-387. <https://doi.org/10.1007/bf02208619>
- Sperling AK and Li RW. (2013) Repetitive sequences. In: Maloy S. and Hughes K. (eds) *Brenner's encyclopedia of genetics* (Second Edition). San Diego: Academic Press pp. 150-154.
- Stoddard BL. (2005) Homing endonuclease structure and function. *Q Rev Biophys* 38:49-95. <https://doi.org/10.1017/S0033583505004063>

- Su Z, Zhang K, Kappel K, Li S, Palo MZ, Pintilie GD, et al. (2021) Cryo-EM structures of full-length *Tetrahymena* ribozyme at 3.1 Å resolution. *Nature* 596:603-607. <https://doi.org/10.1038/s41586-021-03803-w>
- Suh SO, Jones KG and Blackwell M. (1999) A group I intron in the nuclear small subunit rRNA gene of *Cryptosporidium parvum*, 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>
- Tang Y, Nielsen H, Birgisdottir AB and Johansen SD. (2011) A natural fast-cleaving branching ribozyme from the amoeboflagellate *Naegleria gruberi*. *RNA Biol* 8:997-1004. <https://doi.org/10.4161/rna.8.6.16027>
- Tang Y, Nielsen H, Masquida B, Gardner PP and Johansen SD. (2014) Molecular characterization of a new member of the lariat capping twin-ribozyme introns. *Mob DNA* 5:25. <https://doi.org/10.1186/1759-8753-5-25>
- Tanner NK. (1999) Ribozymes: the characteristics and properties of catalytic RNAs. *FEMS Microbiol Rev* 23:257-275. <https://doi.org/10.1111/j.1574-6976.1999.tb00399.x>
- Thompson JD, Higgins DG and Gibson TJ. (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680. <https://doi.org/10.1093/nar/22.22.4673>
- Tirumalai MR, Rivas M, Tran Q and Fox GE. (2021) The peptidyl transferase center: a window to the past. *Microbiol Mol Biol Rev* 85:e0010421. <https://doi.org/10.1128/MMBR.00104-21>
- Turowski TW and Tollervey D. (2016) Transcription by RNA polymerase III: insights into mechanism and regulation. *Biochem Soc Trans* 44:1367-1375. <https://doi.org/10.1042/BST20160062>
- Vader A, Naess J, Haugli K, Haugli F and Johansen SD. (1994) Nucleolar introns from *Physarum flavicomum* contain insertion elements that may explain how mobile group I introns gained their open reading frames. *Nucleic Acids Res* 22:4553-4559. <https://doi.org/10.1093/nar/22.22.4553>
- Vader A, Nielsen H and Johansen SD. (1999) *In vivo* expression of the nucleolar group I intron-encoded *I-Dirt* homing endonuclease involves the removal of a spliceosomal intron. *EMBO J* 18:1003-1013. <https://doi.org/10.1093/emboj/18.4.1003>
- Van de Peer Y. (2009) Phylogenetic inference based on distance methods. In: Lemey P., Salemi M. and Vandamme A. M. (eds) *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing* (2th edn.). Cambridge University Press pp. 149-154.
- Vandamme AM. (2009) Basic concepts of molecular evolution. In: Lemey P., Salemi M. and Vandamme A. M. (eds) *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing* (2th edn.). Cambridge University Press pp. 16-23.
- Vicens Q and Cech TR. (2006) Atomic level architecture of group I introns revealed. *Trends Biochem Sci* 31:41-51. <https://doi.org/10.1016/j.tibs.2005.11.008>


- Wikmark O-G, Einvik C, De Jonckheere JF and Johansen SD. (2006) Short-term sequence evolution and vertical inheritance of the *Naegleria* twin-ribozyme group I intron. BMC Evol Biol 6:39. <https://doi.org/10.1186/1471-2148-6-39>
- Wikmark O-G, Haugen P, Haugli K and Johansen SD. (2007a) Obligatory group I introns with unusual features at positions 1949 and 2449 in nuclear LSU rDNA of Didymiaceae myxomycetes. Mol Phylogenet Evol 43:596-604. <https://doi.org/10.1016/j.ympev.2006.11.004>
- Wikmark O-G, Haugen P, Lundblad EW, Haugli K and Johansen SD. (2007b) The molecular evolution and structural organization of group I introns at position 1389 in nuclear small subunit rDNA of myxomycetes. J Eukaryot Microbiol 54:49-56. <https://doi.org/10.1111/j.1550-7408.2006.00145.x>
- Will CL and Lührmann R. (2011) Spliceosome structure and function. Cold Spring Harb Perspect Biol 3. <https://doi.org/10.1101/cshperspect.a003707>
- Yamada T, Tamura K, Aimi T and Songsri P. (1994) Self-splicing group I introns in eukaryotic viruses. Nucleic Acids Res 22:2532-2537. <https://doi.org/10.1093/nar/22.13.2532>
- Yang Z and Rannala B. (2012) Molecular phylogenetics: principles and practice. Nat Rev Genet 13:303-314. <https://doi.org/10.1038/nrg3186>
- Zhang P, Wu W, Chen Q and Chen M. (2019) Non-Coding RNAs and their Integrated Networks. J Integr Bioinform 16. <https://doi.org/doi:10.1515/jib-2019-0027>
- Zhang X, Li S, Pintilie G, Palo MZ and Zhang K. (2023) Snapshots of the first-step self-splicing of *Tetrahymena* ribozyme revealed by cryo-EM. Nucleic Acids Res 51:1317-1325. <https://doi.org/10.1093/nar/gkac1268>
- Zhou Y, Lu C, Wu Q-J, Wang Y, Sun Z-T, Deng J-C and Zhang Y. (2008) GISSD: group I intron sequence and structure database. Nucleic Acids Res 36:D31-D37. <https://doi.org/10.1093/nar/gkm766>

Paper I

This is an open-access article, reproduced and distributed under the terms of the
Creative Commons Attribution License (CC BY)

Article

A Phylogenetic Approach to Structural Variation in Organization of Nuclear Group I Introns and Their Ribozymes

Betty M. N. Furulund ¹, Bård O. Karlsen ², Igor Babiak ¹  and Steinar D. Johansen ^{1,*}

¹ Genomic Division, Faculty of Biosciences and Aquaculture, Nord University, 8049 Bodø, Norway; betty.m.furulund@nord.no (B.M.N.F.); igor.s.babiak@nord.no (I.B.)

² Research Laboratory, Department of Microbiology, Nordland Hospital Trust, 8005 Bodø, Norway; bard.ove.karlsen@nordlandssykehuset.no

* Correspondence: steinar.d.johansen@nord.no

Abstract: Nuclear group I introns are restricted to the ribosomal DNA locus where they interrupt genes for small subunit and large subunit ribosomal RNAs at conserved sites in some eukaryotic microorganisms. Here, the myxomycete protists are a frequent source of nuclear group I introns due to their unique life strategy and a billion years of separate evolution. The ribosomal DNA of the myxomycete *Mucilago crustacea* was investigated and found to contain seven group I introns, including a direct repeat-containing intron at insertion site S1389 in the small subunit ribosomal RNA gene. We collected, analyzed, and compared 72 S1389 group IC1 introns representing diverse myxomycete taxa. The consensus secondary structure revealed a conserved ribozyme core, but with surprising sequence variations in the guanosine binding site in segment P7. Some S1389 introns harbored large extension sequences in the peripheral region of segment P9 containing direct repeat arrays. These repeats contained up to 52 copies of a putative internal guide sequence motif. Other S1389 introns harbored homing endonuclease genes in segment P1 encoding His-Cys proteins. Homing endonuclease genes were further interrupted by small spliceosomal introns that have to be removed in order to generate the open reading frames. Phylogenetic analyses of S1389 intron and host gene indicated both vertical and horizontal intron transfer during evolution, and revealed sporadic appearances of direct repeats, homing endonuclease genes, and guanosine binding site variants among the myxomycete taxa.

Keywords: catalytic introns; homing endonuclease; intron biology; intron evolution; intron mobility; myxomycete; ribosomal DNA; ribozyme



Citation: Furulund, B.M.N.; Karlsen, B.O.; Babiak, I.; Johansen, S.D. 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>

Academic Editors: Alexander A. Serganov and André P. Gerber

Received: 15 June 2021

Accepted: 21 July 2021

Published: 22 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Group I introns are intervening sequences that interrupt protein-coding or structural RNA genes in diverse taxa [1]. They appear widespread but sporadic in nature [2], and nuclear group I introns have so far only been found within ribosomal RNA (rRNA) genes at phylogenetically conserved sites in protist or fungal eukaryotic microorganisms [3]. A hallmark of group I introns is the characteristic RNA structure that constitutes the group I ribozyme core responsible for intron removal from the precursor RNA. Here, a series of paired segments, named P1 to P13, are organized into higher order helical stack domains (substrate, scaffold, and catalytic domains) [3–5]. Critical structure elements in catalysis, among others, are the U:G wobble pair at the 5' splice site (SS) in segment P1, and the guanosine binding site (G-site) in segment P7. Two subgroups dominate nuclear group I introns, the group IC1 and group IE introns [3,6,7]. These subgroups are easily distinguished due to notable differences in the scaffold domain and the G-site. Crystal structures and three-dimensional models of the *Tetrahymena thermophila* group IC1 ribozyme and the *Didymium iridis* group IE ribozyme, respectively, have been published [8,9].

Group I intron self-splicing is catalyzed by the group I ribozyme and follows a two-step reaction pathway initiated by a nucleophilic attack of exogenous guanosine (exoG)

bound to the G-site [reviewed in 1]. Here, *exoG* attacks the 5' SS and becomes covalently ligated to the 5' end of intron RNA. In the second reaction the terminal intron nucleotide (ω G) replaces *exoG* at the G-site and becomes attacked by the free 3' hydroxyl group of the upstream exon. The outcome of splicing is exon ligation and intron excision. In addition, some nuclear group I introns generate circular intron RNA molecules, either as truncated circles [10] or full-length circles [11].

Homing at the DNA-level is an efficient intron mobility process linked to sexual mating. Group I intron homing is initiated by a double-strand break at the intron-lacking allele generated by an intron-encoded homing endonuclease and completed by gene conversion using the intron-containing allele as a template [reviewed in 3]. There are several distinct families of homing endonucleases, and mobile nuclear group I introns encode homing endonucleases of the His-Cys family [12,13]. Homing by nuclear group I introns has only been shown experimentally in the myxomycetes *Physarum polycephalum* and *D. iridis* [14,15]. Reverse splicing at the RNA-level may also contribute to intron mobility into entopic (homing) or ectopic (transposition) sites [3,16]. Reverse splicing has been reported in some nuclear group I introns [17,18], but experimental support of genomic integration is still lacking.

Myxomycetes (plasmodial slime molds) are unicellular eukaryotes classified as a distinct and ancient protist phylum [19]. The myxomycete ribosomal DNA (rDNA) minichromosome is a frequent host of nuclear group I introns, and 23 genic sites in the small subunit (SSU) and large subunit (LSU) rRNA genes are known to harbor intron insertions [3,20]. *Mucilago* is a genus in the order Physarales that contains only one species, *M. crustacea* [21]. We have isolated and characterized the nuclear rRNA genes in a specimen of *M. crustacea* collected in sub-Arctic Norway. DNA sequencing revealed seven group I introns; two in the SSU rRNA gene (insertion sites S788 and S1389) and five in the LSU rRNA gene (insertion sites L828, L1926, L1949, L2066, and L2449). The intron at insertion site S1389 contained a complex structural organization and was studied in more detail. In an earlier report we showed that the myxomycete S1389 introns constituted a monophyletic clade (the S1389 family) among nuclear group I introns [22]. Here we extended our analyses to involve 72 myxomycete S1389 group I introns identified in a variety of myxomycete taxa representing four taxonomic orders.

2. Results

2.1. The Nuclear rDNA of *Mucilago Crustacea* Harbors Multiple Group I Intron Insertions

Sequencing analysis of the sub-Arctic isolate No-K94 of *M. crustacea* rDNA revealed seven group I introns at conserved insertion sites in the SSU and LSU rRNA genes (Figure 1). Two sites (S788 and S1389) contained introns in the SSU rRNA gene, and five sites (L828, L1926, L1949, L2066, and L2449) in the LSU rRNA gene. The *Mucilago* introns are named according to [23] based on the *E. coli* SSU and LSU rRNAs numbering system. All introns fold into characteristic group I ribozyme core structures (Figure 2a, Supplementary Figure S1a–f), where four belong to the IC1 subgroup (Mcr.S788, Mcr.S1389, Mcr.L828, Mcr.L2449), two to the IE subgroup (Mcr.L1926, Mcr.L2066), and one is unclassified due to a cryptic ribozyme core structure lacking segment P8 (Mcr.L1949). Three *Mucilago* introns (Mcr.L1949, Mcr.S788, and Mcr.S1389) contained direct repeat sequence motifs in peripheral regions of paired segments P1, P2.1, and P9, respectively.

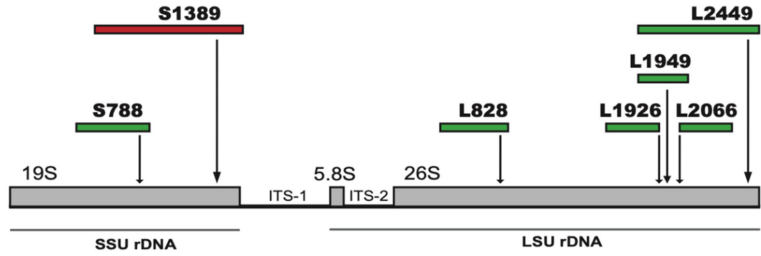


Figure 1. Schematic representation of nuclear rDNA of *Mucilago crustacea*. The SSU rDNA (19S rRNA gene region) is interrupted by group I introns at positions S788 and S1389. The LSU rDNA (5.8S + 26S rRNA gene regions) is interrupted by five group I introns at positions L828, L1926, L1949, L2066, and L2449. See GenBank accession number MZ313554 for details.

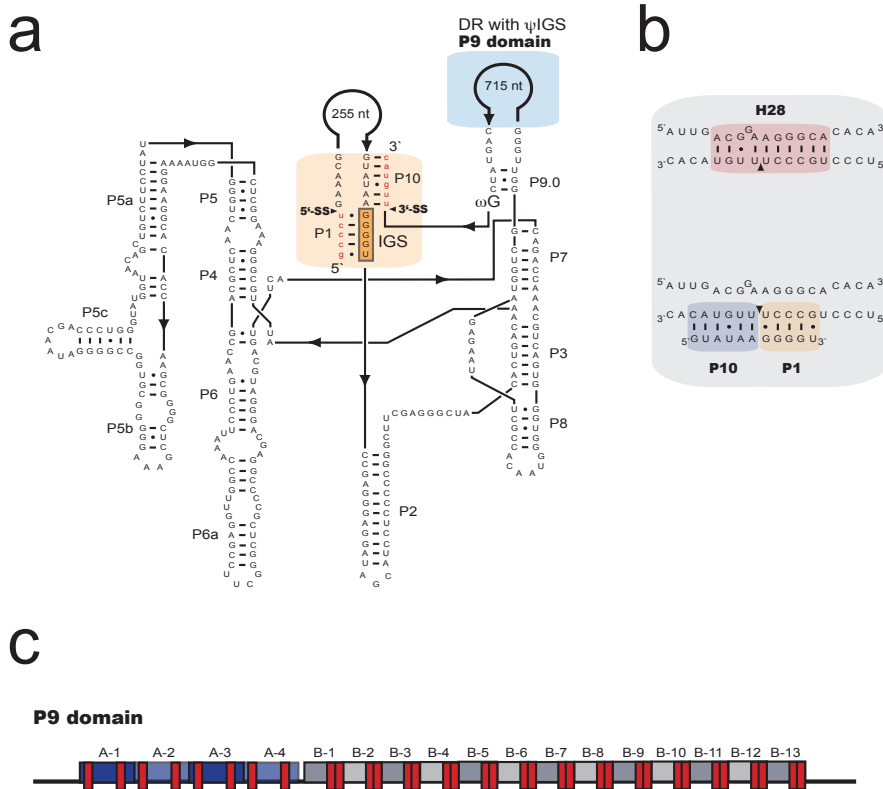


Figure 2. Cont.

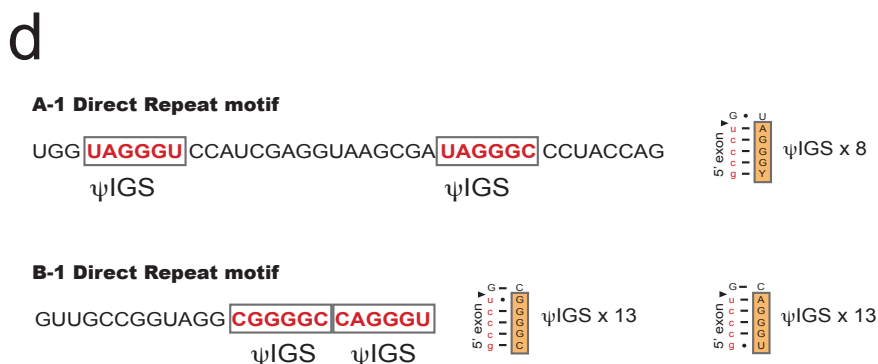


Figure 2. Structural features of the *Mucilago crustacea* S1389 group I intron. (a) Secondary structure diagram of the 1274-nt *M. crustacea* S1389 group IC1 intron S1389, presented according to [4]. The intron has a strong P1/P10 exon binding region of 11 base pairs (red box) and a large extension in P9 that contains direct repeat motifs (blue box). P1-P10, paired RNA segments; 5' SS and 3' SS, exon–intron splice sites; IGS, internal guide sequence. (b) Helix 28 (H28) in SSU rRNA becomes completely disrupted when the S1389 intron is present. (c) The P9 insertion contains two repeat motifs (A and B) with ΨIGS sequences (red boxes). (d) Sequential features of the repeats and putative exon base pair interactions.

2.2. A Complex Group I Intron in *M. Crustacea* SSU rRNA Helix 28 at Insertion Site S1389

Intron Mcr.S1389 has a complex structural organization at the RNA level (Figure 2a). The 1274 nt intron interrupts the critical neck helix (H28) in the SSU rRNA [24] that connects the body and head of the ribosomal small subunit, and presence of the S1389 intron sequences completely disrupts H28 (Figure 2b). The large size of Mcr.S1389 is mainly due to extension sequences in P1 (255 nt) and P9 (715 nt), but no recognizable open reading frames are present. P9, however, contains a complex array of direct repeats (Figure 2c) consisting of three copies of A-motif (39 nt) and 13 copies of B-motif (24 nt). A closer inspection of motifs A and B identified 34 putative internal guide sequences (ψIGS) (Figure 2d).

2.3. Structural Variation among the S1389 Introns

Group I introns at position S1389 are almost exclusively restricted to myxomycete protists, and introns from 72 taxa representing four orders (Physarales, Stemonitales, Liceales and Trichiales), 15 genera, and 37 species were identified. The majority of the SSU rRNA gene sequences were retrieved from the NCBI database, but six Didymiceae sequences (Supplementary Table S1) were generated in this study. To assess structural variations among the S1389 introns, 166 nucleotide positions in the catalytic core region common to all 72 introns were strictly aligned according to secondary structure features (Supplementary Figure S2). The consensus secondary structure diagram (Figure 3) shows a typical group IC1 intron fold with a conserved catalytic domain (P3-P7-P8) and P4-P5a. Interestingly, six sequence variants of the G-site in segment P7 (A1, A2, A3, U1, U2, and G1) were observed, and named according to the bulged P7 nucleotide residue (corresponding to A263 in the *Tetrahymena* intron) (Figure 3, right panel). Bulged U and G have not been reported previously in naturally occurring nuclear group I introns.

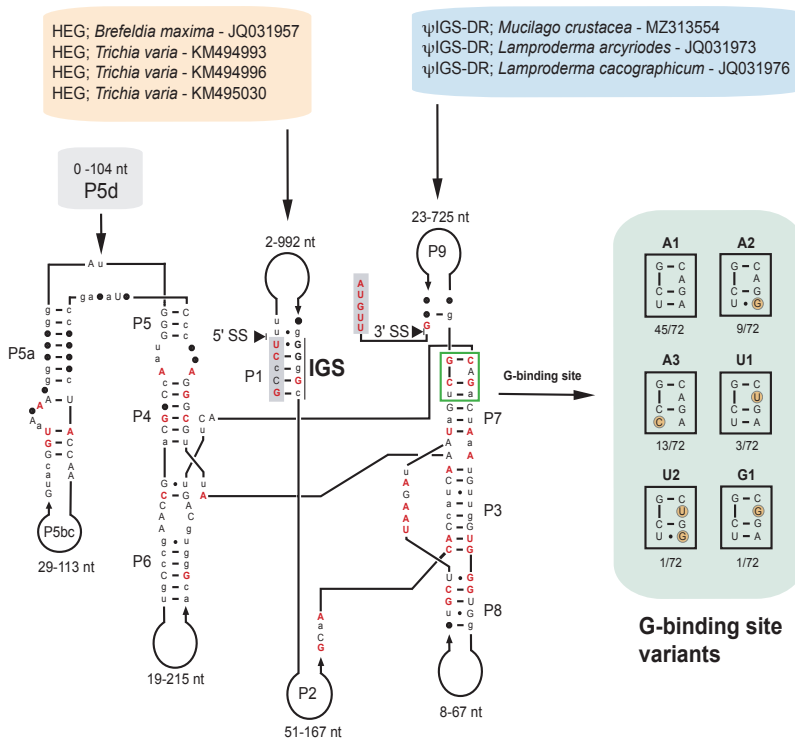


Figure 3. Consensus secondary structure diagram of S1389 group I intron in myxomycetes. The consensus structure is based on 166 nucleotide positions in the catalytic core common among introns from 72 taxa (see Supplementary Table S1). Extensive sequence size variations are noted in most peripheral regions, including the taxa-specific P5d extension. While all homing endonuclease genes (HEGs) are found as P1 extensions, all direct repeat (DR) extensions are located in P9. The latter contains multiple pseudo internal guide sequence (ψIGS) features. Guanosine binding structure variants in P7 are indicated (boxed at right; G-site). P1-P10, paired RNA segments; 5' SS and 3' SS, exon-intron splice sites. Invariant nucleotide positions among the 72 introns are shown as red uppercase letters. Black uppercase letter, >90% conservation; lowercase letters, ≥50% conservation; filled circles, <50% conservation.

Most sequence variations were found in peripheral regions of segments P1, P2, P5, P6, P8, and P9 (Figure 3). The optional helical segment P5d appears taxa-specific and was common in Stemonitales introns but absent in Trichiales and most Physarales introns (Supplementary Table S1). The P9 extension in three S1389 introns host extensive direct repeat motifs, all with multiple ψIGS. Similar to the *M. crustacea* intron (order Physarales), *Lamproderma arcyriodes* and *Lamproderma cacographicum* (both of order Stemonitales) contain 52 copies and 17 copies, respectively, of ψIGS (Supplementary Figure S3). Finally, four S1389 introns contain homing endonuclease gene (HEG) insertions in P1, represented by three *Trichia varia* (order Trichiales) isolates and one in *Brefeldia maxima* (order Stemonitales) (Figure 3; Supplementary Table S1).

2.4. Spliceosomal Introns Interrupt S1389 Group I Intron HEGs

The S1389 HEGs were organized in a sense orientation in P1 compared to that of the intron ribozyme and host SSU rRNA and encode His-Cys homing endonuclease proteins of 244 amino acids and 247 amino acids in *T. varia*, and 188 amino acids in *B. maxima* (Figure 4a). Interestingly, the *Trichia* HEGs were found interrupted by small spliceosomal introns of 44 bp and 53 bp (Figure 4a; Supplementary Figure S4). An amino acid sequence

alignment of the S1389-derived His-Cys boxes to the well-characterized homing endonucleases in *Naegleria* (*I-NjaI*) [25] and *Physarum* (*I-PpoI*) [26] (Figure 4b) confirmed that the intron proteins were members of the His-Cys homing endonuclease family.

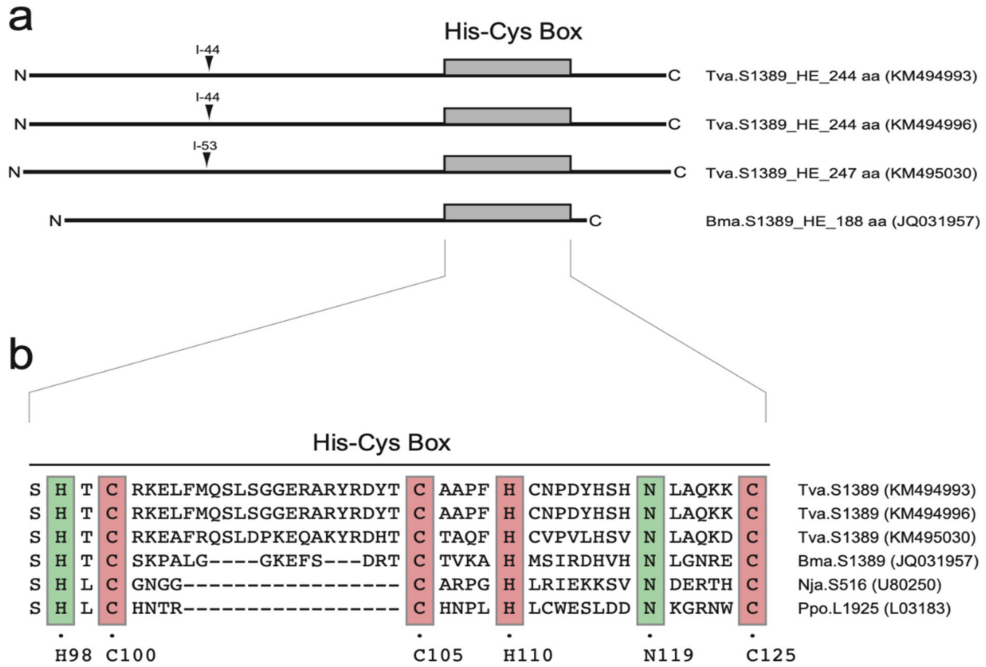


Figure 4. His-Cys homing endonucleases encoded by S1389 group I introns. (a) Schematic organization of S1389 intron homing endonucleases including N and C terminal ends, and the characteristic His-Cys box. *Trichia varia* homing endonuclease genes (HEGs) are interrupted by small spliceosomal introns (I-44 and I-53). (b) Amino acid alignment of His-Cys box features. The *Naegleria* (*I-NjaI*; Nja.S516) and *Physarum* (*I-PpoI*; Ppo.L1925) represent well-studied His-Cys homing endonucleases. Conserved residues (boxed) corresponding to those presented in the *I-PpoI* crystal structure [26]. C100, C105, H110, and C125 are involved in zinc binding, and H98 and N119 are associated with the active site.

2.5. S1389 Group I Introns Appear Vertically and Horizontally Inherited

S1389 introns were identified in distantly related taxa representing four myxomycete orders (Supplementary Table S1). To gain information about the evolutionary history of S1389 introns, phylogenetic analyses based on group I intron core sequences and on the corresponding host SSU rRNA gene sequences were performed. A phylogenetic tree obtained from maximum likelihood (ML) analysis based on 1575 aligned SSU rRNA positions from 72 taxa is shown in Supplementary Figure S5. The tree topology correlates well to the current taxonomy of myxomycetes and more recently reported rDNA-based phylogeny [27–30]. However, we noted that the *Lepidoderma tigrinum* sequence (GenBank accession number DQ903678) appeared embedded within the *Diderma* cluster. Phylogenetic analysis of the corresponding S1389 intron from the same 72 taxa was challenging due to a large dataset and only 166 nt aligned positions. A neighbor-joining tree of the S1389 introns is presented in Supplementary Figure S6. Statistical support was obtained in topologies within several branch clusters, but not well between clusters. Anyway, we observed an agreement between the SSU rDNA and S1389 intron trees, but with some notable exceptions (see below).

To investigate the distribution of intron structural variants and possible horizontal transfer at higher resolution we reduced the selected taxa to 23 in SSU rDNA and corresponding intron phylogenetic analyses. Several interesting and apparently relevant observations were noted when comparing branch patterns of the SSU rDNA-based tree (Figure 5) and the intron-based tree (Figure 6). (1) An overall similarity was found between S1389 introns and their gene hosts (SSU rDNA), suggesting that most S1389 introns possess a vertical inheritance pattern during myxomycete evolution. (2) Three S1389 introns, representing *D. iridis* (AJ938151), *Lamproderma columbinum* (HQ687200), and *L. arcyrioides* (JQ031973) appear horizontally inherited. (3) *L. tigrinum* (DQ903678) appears misannotated in GenBank and should most likely be reclassified as a *Diderma* sp. (4) The Ψ IGS-containing direct repeats in introns from *M. crustacea*, *L. arcyrioides*, and *L. cacographicum* have most likely been independently gained during the evolution. (5) HEGs noted in *B. maxima* and *T. varia* were also independently gained, but we were not able to conclude about evolutionary history of the two *T. varia* HEG variants. (6) The G-site variants (bulge A to U, and bulge A to G mutations) in P7 appear sporadically in closely related S1389 introns. An example is the U1 variant in *D. squamulosum* that appeared generated from the A1 variant by one point mutation (Figures 3 and 6). Similarly, the U2 variant in *L. columbinum* and the G1 variant in *Diderma testaceum* were only one point mutation apart from A2 and A1 variants, respectively (Figures 3 and 6).

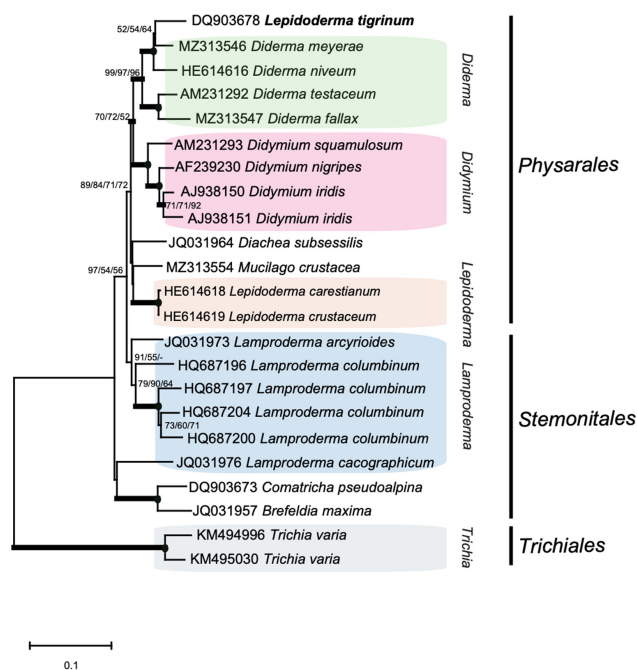


Figure 5. Molecular phylogeny of myxomycete taxa based SSU rDNA sequences. The SSU topology is obtained by neighbor-joining (NJ) analysis of 23 selected taxa and 1575 nt aligned positions (SSU dataset 2; Supplementary Table S1). The tree is rooted with the *T. varia* SSU rDNA sequence. The NJ, maximum likelihood (ML) and maximum parsimony (MP) bootstrap replicates ($\geq 50\%$) are given at each node. Bayesian posterior (BI) probabilities (≥ 0.95) are shown in bold branches. Black dots at branch points; maximum support in NJ, ML, and MP ($\geq 97\%$). Misannotated SSU rDNA in GenBank (*L. tigrinum* DQ903678) is indicated in bold letters. The scale bar indicates the fraction of substitutions per site.

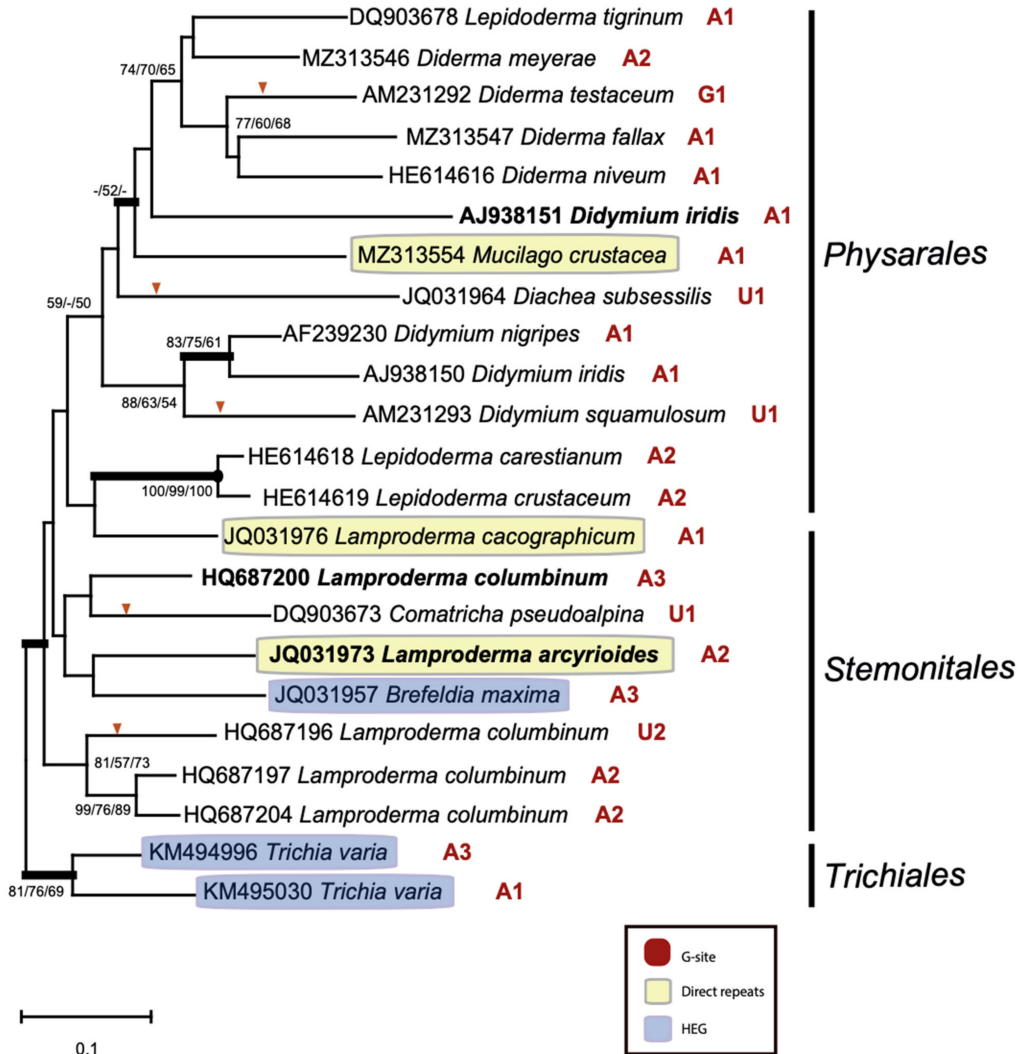


Figure 6. Molecular phylogeny of myxomycete S1389 group I introns. The intron topology is obtained by neighbor-joining (NJ) analysis of 23 taxa and 166 nt aligned positions (intron dataset 2; Supplementary Table S1). The tree is rooted with the *T. varia* S1389 intron sequence. The NJ-, maximum likelihood (ML) and maximum parsimony (MP) bootstrap replicates ($\geq 50\%$) are given for each node. Bayesian posterior (BI) probabilities (≥ 0.95) are shown in bold branches. Black dots at branch points; maximum support in NJ, ML, and MP ($\geq 97\%$). Introns that harbor homing endonuclease genes (HEGs) and direct repeat arrays are indicated by yellow and blue boxes, respectively. G-site variation is indicated at each intron and bulge-nucleotide mutation (A to U; A to G) is indicated on branch by red arrow. Horizontally inherited intron candidates (*D. iridis* AJ938151; *L. columbinum* HQ687200; *L. arcyrrioides* JQ031973) are indicated in bold letters. The scale bar indicates the fraction of substitutions per site.

3. Discussion

We report the nuclear rDNA sequence of the myxomycete *M. crustacea* to contain seven group I introns, which include a complex intron at position S1389 with interesting structural features. A dataset of 72 S1389 introns representing different myxomycete taxa was collected and assessed for structural variation in intron organization and ribozyme core. We identified taxa-specific RNA segments, G-site variants, complex direct repeat arrays, and homing endonuclease genes. The latter contained small spliceosomal introns that have to be removed in order to generate the open reading frames. Phylogenetic reconstruction of intron and host gene suggested an evolutionary history of the S1389 introns that includes both vertical and horizontal intron transfers.

Group I introns are common in myxomycete rDNA, but intron numbers vary between species that have been SSU and LSU rRNA gene sequenced. *P. polycephalum* and *D. iridis*, which are the best studied myxomycete species, carry only 2–3 introns [31–33], but *Diderma niveum* rDNA contains more than 20 group I introns [3]. The seven group I introns reported in *M. crustacea* is an intermediate number, but still apparently challenges the pre-rRNA processing. A study of RNA processing and self-splicing in 12 *Fuligo septica* group I introns (4 in the SSU rRNA gene and 8 in the LSU rRNA gene) revealed that each insertion site represents a distinct intron family and that introns were removed in a non-restricted order [34]. The exception, however, was two compulsory introns in the LSU rRNA gene (L1949 and L2449) that spliced more slowly than the others and were the only *F. septica* introns without in vitro self-splicing activities. The L1949 and L2449 introns are compulsory in Physarales myxomycetes [29,35], and present in *M. crustacea*. The truncated catalytic core lacking segment P8 in the *M. crustacea* L1949 intron (Supplementary Figure S1d) was similar to that of *Physarum* [31,36] and *Fuligo* [34] and may in part explain why these introns do not self-splice as naked RNA at standard reaction conditions in vitro.

S1389 introns in myxomycetes constitute a phylogenetically distinct family of nuclear group I introns with a common origin [22], a notion supported by this work. Interestingly, the S1389 introns appeared dependent on essential host splicing factors since they do not self-splice in vitro [22]. However, the ribozyme core structure folding was in agreement with other well-studied group IC1 introns, such as the in vitro self-splicing *Tetrahymena* LSU rRNA intron [37]. This included the organization of the 5' SS and P1 [38], the A-rich bulges in P5a and P4/P5 [39], and the joining of P4 to the catalytic core [40]. The most surprising structural core variation was at the G-site, which is usually highly conserved among related introns at the same insertion site. Here, a strong covariation has been reported between the P7 bulge nucleotide (corresponding to *Tetrahymena* A263) and the closing base pair (*Tetrahymena* C262:G312) [41,42], forming a conserved base triple [5]. Among the myxomycete S1389 introns we noted six major variants, which included a shift where the A-bulge nucleotide was replaced at U. How this G-site substitution affects catalysis is not known, but a similar sequence variant was obtained by in vitro selection experiments of the *Tetrahymena* intron and found active in self-splicing [43].

The *M. crustacea* S1389 intron harbors a complex direct repeat pattern in peripheral segment P9, and direct repeats are also present in two additional *Mucilago* introns (S788 and L1949). Direct repeat insertions have been reported in several nuclear group I introns in myxomycetes, especially in introns lacking in vitro self-splicing activity. The *F. septica*, *P. flavicomum*, and *D. niveum* L1949 introns contained repeated copies of 45-bp, 101-bp, and 20-bp motifs, respectively, in segment P1 [34–36]. Furthermore, the L2449 introns carry direct repeat arrays in several species including *Didymium dubium* that harbored two copies of two different motifs (71 bp and 303 bp) in segment P1 [35], and *Physarum didermoides* with direct repeats in segments P2 (5 copies of 60 bp) and P9 (5 copies of 100 bp) [29]. Functional roles linked to such direct repeat arrays are currently not known. The presence of multiple copies of putative IGS features in the *M. crustacea* S1389 intron and in two additional myxomycete S1389 introns (*L. arcyriodes* and *L. cacographicum*) may suggest a functional role. On the other hand, large extension sequences in segments P1, P2, or P9 do not interfere with group I ribozyme core structures and may be accepted within introns

as long as the introns are efficiently removed, probably aided by host-encoded factors. Since direct repeat arrays have sporadic occurrences and probably are evolutionary neutral, they could be considered as non-functional junk RNA similar to what is proposed for the majority of vertebrate long non-coding RNAs [44].

Two myxomycete species, *B. maxima* and *T. varia*, harbor S1389 introns with HEGs in segment P1. These HEGs encode His-Cys-type homing endonucleases typical for nuclear group I introns [3,6] and indicate that S1389 introns may perform intron homing at appropriate biological conditions. Interestingly, the *T. varia* HEGs were interrupted by small spliceosomal introns of 44 bp and 53 bp. Related spliceosomal introns have been reported to splice out and restore functional HEGs in two *D. iridis* group I introns [2,45,46]. A plausible role of the spliceosomal introns may be to facilitate the expression of HEGs located within a rDNA locus by recruiting spliceosomes and probably exon junction complexes [47,48].

4. Materials and Methods

4.1. Culturing, DNA Isolation, and DNA Sequencing

The myxomycete *M. crustacea* (isolate No-K94) was collected as spores in sub-Arctic Norway (68°N 45' N, 19°N 42' E). Germinated amoebae from *M. crustacea* and some other Physarales myxomycetes (Supplementary Table S1) were grown on DS/2 media [15] at 26 °C, feeding on heat-inactivated *Escherichia coli*. Total DNA was isolated as previously described [34,49], and rDNA was amplified by PCR using various primers designed from known myxomycete rDNA sequences. Sanger sequencing on both strands was performed as previously described [34].

4.2. Sequence Alignments

Six of the SSU rDNA sequences, including group I introns, were generated in this work (Supplementary Table S1), and 67 SSU rDNA sequences (including group I introns) were downloaded from the NCBI database (<https://www.ncbi.nlm.nih.gov>), accessed 1 May 2021). SSU rDNA sequences (with introns removed) were assembled and aligned in the software program Geneious prime[®] 2019.2.3 (<https://www.geneious.com>, accessed 1 April 2021). Multiple sequence alignments were generated using MAFFT version 7.450 [50,51] with default setting. The SSU dataset 1 (1575 bp) included sequences from all 72 taxa and was based on published information about myxomycete SSU rRNA [30,52] and evaluated and manually trimmed by excluding parts of variable regions that were not confidently aligned. SSU dataset 2 included 23 taxa and was made similar to dataset 1. Group I intron S1389 core sequences (166 bp) were manually aligned and strictly based on secondary structure features as previously described [22,49] using the Geneious prime[®] software program. The intron datasets 1 and 2 corresponded to the SSU datasets 1 and 2, except that *T. thermophila* L1925 and *M. crustacea* S788 introns were included as outgroups.

4.3. Phylogenetic Analysis

The tree-building methods of neighbor joining (NJ), maximum likelihood (ML), and maximum parsimony (MP) in MEGA X [53] with default settings, as well as Bayesian inference (BI) in MrBayes v.3.2.6 [54], were used in SSU datasets 1 and 2 to construct molecular phylogenetics. All sequence alignments were model tested prior to tree constructions by MEGA X software [53]. The topology of NJ, ML, and MP trees was evaluated by bootstrap analyses. The evolutionary history of SSU dataset 1 was reconstructed with ML using the general time reverse model (GTR) [55] with gamma-distributed rate (G) of heterogeneity among sites approximated by five categories (500 replicates) and MP (subtree-pruning-regrafting algorithm and 500 bootstrap replicates). The phylogeny was also tested by BI conducted with MrBayes v.3.2.6 [54] interpreted in Geneious prime[®] 2019.2.3. BI was run with default settings with the substitution model, GTR + G (4 categories), including the chain length (1,100,000), subsampling frequency (200), heated chains (4), heated chain temperature (0.2), burn-in length (100,000), and random seed (23,060). The evolutionary history

of SSU dataset 2 was reconstructed with NJ with Kimura 2 (K2) evolutionary model [56], ML with GTR + G, MP, and BI. The robustness was tested by bootstrapping (500 replicates).

The evolutionary history of intron dataset 1 was reconstructed with NJ and the Jukes-Cantor (JC) model using MEGA X [53] with default settings. To test the robustness of the nodes, the trees were tested with NJ-JC (500 replicates), MP-SPR (500 replicates), and ML-K2 (500 replicates). BI were conducted with MrBayes v.3.2.7 [57,58] with 10 million generations and sampled every 1000, with a burn-in fraction of 0.25. Intron dataset 2 included 23 taxa. The evolutionary history was reconstructed with NJ and JC model using MEGA X [53] with default settings. The robustness of the nodes of the trees was tested with NJ-JC (500 replicates), MP-SPR (500 replicates), and ML-K2 (500 replicates). BI were conducted with MrBayes v.3.2.7 using the same parameters as in intron dataset 1.

5. Conclusions

Myxomycete protists are a frequent source of nuclear group I introns and *M. crustacea* was found to contain seven rDNA group I introns, two in the SSU rRNA gene and five in the LSU rRNA gene. The *M. crustacea* S1389 intron has a complex structural organization and we collected, analyzed, and compared 72 S1389 intron from a wide spectrum of myxomycetes. The S1389 intron sequences and catalytic core structures were clearly related and supported the notion that introns occupying a specific insertion site represent a distinct group I intron family. However, sequence variation at the highly conserved G-site suggested that group I introns may accept more structural changes at the catalytic site than previously anticipated for natural group I ribozymes. Peripheral segments, on the other hand, may harbor large structural insertions containing homing endonuclease genes or direct repeat arrays, RNA sequences that apparently do not interfere with the splicing reaction and the ribozyme core. An interesting and unusual feature was that a second class of introns, the spliceosomal introns, were present within some of the S1389 group I introns. These small introns may have a biological role, probably linked to homing endonuclease gene expression.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ncrna7030043/s1>, Figure S1: Secondary structure diagrams of the *Mucilago crustacea* rDNA group I introns, Figure S2: Sequence alignment of core structure nucleotides of myxomycete S1389 group I intron, Figure S3: Secondary structure diagrams of *Lamproderma arcyrioides* and *Lamproderma arcyrioides* S1389 group I intron, Figure S4: Sequence features in *Trichia varia* S1389 group I introns, Figure S5: Molecular phylogeny of myxomycete taxa based SSU rDNA sequences, Figure S6: Molecular phylogeny of myxomycete S1389 group I introns, Table S1: Key features of 72 myxomycete S1389 group I introns included in this work.

Author Contributions: B.M.N.F., B.O.K., I.B., and S.D.J. contributed to conceptualization and designed the research; S.D.J. performed the sequence-structure analysis and B.M.N.F. performed phylogenetic analysis. B.M.N.F. and S.D.J. analyzed data and wrote the manuscript together with B.O.K. and I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding except general grants from Nord University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Sequencing data are available in GenBank under the accession numbers MZ313546; MZ313547; MZ313548; MZ313549; MZ313552, and MZ313554.

Acknowledgments: We thank Kari Haugli for technical support in collecting myxomycetes, culturing, DNA isolation and Sanger sequencing, and Peik Haugen for advising on intron phylogeny analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cech, T.R. Self-splicing of group I introns. *Ann. Rev. Biochem.* **1990**, *59*, 543–568. [[CrossRef](#)]
2. Haugen, P.; Simon, D.M.; Bhattacharya, D. The natural history of group I introns. *Trends Genet.* **2005**, *21*, 111–119. [[CrossRef](#)]
3. Hedberg, A.; Johansen, S.D. Nuclear group I introns in self-splicing and beyond. *Mob. DNA* **2013**, *4*, 17. [[CrossRef](#)]
4. Cech, T.R.; Damberger, S.H.; Gutell, R.R. Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol.* **1994**, *1*, 273–280. [[CrossRef](#)]
5. Vicens, Q.; Cech, T.R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **2006**, *31*, 41–51. [[CrossRef](#)]
6. 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*, 129–140. [[CrossRef](#)]
7. Zhou, Y.; Lu, C.; Wu, Q.-J.; Wang, Y.; Sun, Z.-T.; Deng, J.-C.; Zhang, Y. GISSD: Group I intron sequence and structure database. *Nucleic Acids Res.* **2008**, *36*, D31–D37. [[CrossRef](#)]
8. Guo, F.; Gooding, A.R.; Cech, T.R. Structure of the *Tetrahymena* ribozyme: Base triple sandwich and metal ion at the active site. *Mol. Cell* **2004**, *16*, 351–362. [[CrossRef](#)]
9. Andersen, K.L.; Beckert, B.; Masquida, B.; Johansen, S.D.; Nielsen, H. Accumulation of stable full-length circular group I intron RNAs during heat-shock. *Molecules* **2016**, *21*, 1451. [[CrossRef](#)]
10. Been, M.D.; Cech, T.R. Selection of circularization sites in a group I IVS RNA requires multiple alignments of an internal template-like sequence. *Cell* **1987**, *50*, 951–961. [[CrossRef](#)]
11. Nielsen, H.; Fiskaa, T.; Birgisdottir, Á.B.; Haugen, P.; Einvik, C.; Johansen, S. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* **2003**, *9*, 1464–1475. [[CrossRef](#)]
12. Johansen, S.; Embley, T.M.; Willassen, N.P. A family of nuclear homing endonucleases. *Nucleic Acids Res.* **1993**, *21*, 4405. [[CrossRef](#)] [[PubMed](#)]
13. Stoddard, B.L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **2005**, *38*, 49–95. [[CrossRef](#)]
14. Muscarella, D.E.; Vogt, V.M. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* **1989**, *56*, 443–454. [[CrossRef](#)] [[PubMed](#)]
15. Johansen, S.; Elde, M.; Vader, A.; Haugen, P.; Haugli, K.; Haugli, F. In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. *Mol. Microbiol.* **1997**, *24*, 737–745. [[CrossRef](#)]
16. Edgell, D.R.; Chalancharla, V.R.; Belfort, M. Learning to live together: Mutualism between self-splicing introns and their hosts. *BMC Biol.* **2011**, *9*, 22. [[CrossRef](#)]
17. Roman, J.; Rubin, M.N.; Woodson, S.A. Sequence specificity of *in vivo* reverse splicing of the *Tetrahymena* group I intron. *RNA* **1999**, *5*, 1–13. [[CrossRef](#)] [[PubMed](#)]
18. Birgisdottir, Á.B.; Johansen, S. Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. *Nucleic Acids Res.* **2005**, *33*, 2042–2051. [[CrossRef](#)]
19. Margulis, L.; Schwartz, K. *Five Kingdoms: An Illustrated Guide to the Phyla of Life in Earth*, 3rd ed.; Freeman: New York, NY, USA, 1997; ISBN 9780716730279.
20. Nielsen, H.; Johansen, S.D. Group I introns: Moving in new directions. *RNA Biol.* **2009**, *6*, 375–383. [[CrossRef](#)]
21. Neubert, H.; Nowotny, W.; Baumann, K. *Die Myxomyceten*; Karlheinz Baumann Verlag: Gomaringen, Germany, 1995; Volume 2, ISBN 3-929822-01-6.
22. Wikmark, O.G.; Haugen, P.; Lundblad, E.W.; Haugli, K.; Johansen, S.D. The molecular evolution and structural organization of group I introns at position 1389 in nuclear small subunit rDNA of myxomycetes. *J. Euk. Microbiol.* **2007**, *54*, 49–56. [[CrossRef](#)] [[PubMed](#)]
23. Johansen, S.; Haugen, P. A new nomenclature of group I introns in ribosomal DNA. *RNA* **2001**, *7*, 935–936. [[CrossRef](#)]
24. Mohan, S.; Donohue, J.P.; Noller, H.F. Molecular mechanics of 30S subunit head rotation. *Proc. Natl Acad. Sci. USA* **2014**, *111*, 13325–13330. [[CrossRef](#)]
25. Elde, M.; Haugen, P.; Willassen, N.P.; Johansen, S. I-Njal, a nuclear intron-encoded homing endonuclease from *Naegleria*, generates a pentanucleotide 3' cleavage-overhang within a 19 base-pair partially symmetric DNA recognition site. *Eur. J. Biochem.* **1999**, *259*, 281–288. [[CrossRef](#)]
26. Flick, K.E.; Jurica, M.S.; Monnat, R.J.J.; Stoddard, B.L. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **1998**, *394*, 96–101. [[CrossRef](#)]
27. Fiore-Donno, A.M.; Mayer, M.; Baldauf, S.L.; Pawlowski, J. Evolution of dark-spored myxomycetes (slime-molds): Molecules versus morphology. *Mol. Phylogenet. Evol.* **2008**, *46*, 878–889. [[CrossRef](#)]
28. Fiore-Donno, A.M.; Kamono, A.; Meyer, M.; Schnittler, M.; Fukui, M.; Cavalier-Smith, T. 18S rDNA phylogeny of lamproderma and allied genera (Stemonirales, Myxomycetes, Amoebozoa). *PLoS ONE* **2012**, *7*, e35359. [[CrossRef](#)]
29. Nandipati, S.C.; Haugli, K.; Coucheron, D.H.; Haskins, E.F.; Johansen, S.D. Polyphyletic origin of the genus *Physarum* (Physarales, Myxomycetes) revealed by nuclear rDNA mini-chromosome analysis and group I intron synapomorphy. *BMC Evol. Biol.* **2012**, *12*, 166. [[CrossRef](#)]
30. Cainelli, R.; de Haan, M.; Meyer, M.; Bonkowski, M.; Fiore-Donno, A.M. Phylogeny of Physarida (Amoebozoa, Myxogastria) based on the small-subunit ribosomal RNA gene, redefinition of *Physarum pusillum* s. str. and reinstatement of *P. gravidum* Morgan. *J. Eukaryot. Microbiol.* **2020**, *67*, 327–336. [[CrossRef](#)]

31. Johansen, S.; Johansen, T.; Haugli, F. Structure and evolution of myxomycete nuclear group I introns: A model for horizontal transfer by intron homing. *Curr. Genet.* **1992**, *22*, 297–304. [[CrossRef](#)]
32. Ruoff, B.; Johansen, S.; Vogt, V.M. Characterization of the self-splicing products of a mobile intron from the nuclear rDNA of *Physarum polycephalum*. *Nucleic Acids Res.* **1992**, *20*, 5899–5906. [[CrossRef](#)]
33. Johansen, S.; Vogt, V.M. An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell* **1994**, *76*, 725–734. [[CrossRef](#)]
34. Lundblad, E.W.; 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*, 1283–1293. [[CrossRef](#)]
35. Wikmark, O.G.; Haugen, P.; Haugli, K.; Johansen, S.D. Obligatory group I introns with unusual features at positions 1949 and 2449 in nuclear LSU rDNA of Didymiaceae myxomycetes. *Mol. Phylogenet. Evol.* **2007**, *43*, 596–604. [[CrossRef](#)]
36. Vader, A.; Naess, J.; Haugli, K.; Haugli, F.; Johansen, S. Nucleolar introns from *Physarum flavicomum* contain insertion elements that may explain how mobile group I introns gained their open reading frames. *Nucleic Acids Res.* **1994**, *22*, 4553–4559. [[CrossRef](#)] [[PubMed](#)]
37. Golden, B.L. Group I introns: Biochemical and crystallographic characterization of the active site structure. In *Ribozymes and RNA Catalysis*; Lilley, D.M.J., Eckstein, F., Eds.; RSC Press: Cambridge, UK, 2008; pp. 178–200. ISBN 978-0-85404-253-1.
38. Knitt, D.S.; Narlikar, G.J.; Herschlag, D. Dissection of the role of the conserved G.U pair in group I RNA self-splicing. *Biochemistry* **1994**, *33*, 13864–13879. [[CrossRef](#)]
39. Ikawa, Y.; Yoshimura, T.; Hara, H.; Shiraishi, H.; Inoue, T. Two conserved structural components, A-rich bulge and P4X]6/7 base-triples, in activating the group I ribozymes. *Genes Cells* **2002**, *7*, 1205–1215. [[CrossRef](#)] [[PubMed](#)]
40. Tanner, M.A.; Cech, T.R. Joining the two domains of a group I ribozyme to form the catalytic core. *Science* **1997**, *275*, 847–849. [[CrossRef](#)]
41. Michel, F.; Westhof, E. Modelling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **1990**, *216*, 585–610. [[CrossRef](#)]
42. Couture, S.; Ellington, A.D.; Gerber, A.S.; Cherry, J.M.; Doudna, J.A.; Green, R.; Hanna, M.; Pace, U.; Rajagopal, J.; Szostak, J.W. Mutational analysis of conserved nucleotides in self-splicing group I intron. *J. Mol. Biol.* **1990**, *215*, 345–358. [[CrossRef](#)]
43. Oe, Y.; Ikawa, Y.; Shiraishi, H.; Inoue, T. Analysis of the P7 region within the catalytic core of the *Tetrahymena* ribozyme by employing in vitro selection. *Nucleic Acids Symp. Ser.* **2000**, *44*, 197–198. [[CrossRef](#)]
44. Palazzo, A.F.; Koonin, E.V. Functional long non-coding RNAs evolve from junk transcripts. *Cell* **2020**, *183*, 1151–1161. [[CrossRef](#)]
45. Vader, A.; Nielsen, H.; Johansen, S. In vivo expression of the nucleolar group I intron-encoded *I-Dir1* homing endonuclease involves the removal of a spliceosomal intron. *EMBO J.* **1999**, *18*, 1003–1013. [[CrossRef](#)]
46. Johansen, S.D.; Vader, A.; Sjøttem, E.; Nielsen, H. In vivo expression of a group I intron HEG from the antisense strand of *Didymium* ribosomal DNA. *RNA Biol.* **2006**, *3*, 157–162. [[CrossRef](#)]
47. Johansen, S.D.; Haugen, P.; Nielsen, H. Expression of protein-coding genes embedded in ribosomal DNA. *Biol. Chem.* **2007**, *388*, 679–686. [[CrossRef](#)]
48. Schlautmann, L.P.; Gehring, N.H. A day in the life of the exon junction complex. *Biomolecules* **2020**, *10*, 866. [[CrossRef](#)]
49. Haugen, P.; Coucheron, D.H.; Rønning, S.B.; Haugli, K.; Johansen, S. The molecular evolution and structural organization of self-splicing group I introns at position 516 in nuclear SSU rDNA of myxomycetes. *J. Eukaryot. Microbiol.* **2003**, *50*, 283–292. [[CrossRef](#)]
50. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
51. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvement in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
52. Johansen, T.; Johansen, S.; Haugli, F.B. Nucleotide sequence of the *Physarum polycephalum* small subunit ribosomal RNA as inferred from the gene sequence: Secondary structure and evolutionary implications. *Curr. Genet.* **1988**, *14*, 265–273. [[CrossRef](#)]
53. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
54. Huelsenbeck, J.P.; Ronquist, F.; Nielsen, R.; Bollback, J.P. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **2001**, *294*, 2310–2314. [[CrossRef](#)] [[PubMed](#)]
55. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press: New York, NY, USA, 2000; ISBN 9780195135855.
56. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [[CrossRef](#)] [[PubMed](#)]
57. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [[CrossRef](#)]
58. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [[CrossRef](#)]

Supplementary materials

Table S1. Key features of 72 myxomycete S1389 group I introns included in this work.

Host species	Acc No	Isolate	Size¹	Insert²	P5d³	P7⁴
Order Physarales						
<i>Diderma fallax</i> ⁵	MZ313547	It-K52	532 bp	-	-	A1
<i>D. globosum</i>	DQ903677	AMFD110	722 bp	-	-	A1
<i>D. meyerae</i>	HE614614	It-K61	735 bp	-	-	A2
<i>D. meyerae</i> ⁵	MZ313546	It-K68	921 bp	-	-	A2
<i>D. niveum</i> ⁵	MZ313548	Fr-M26	749 bp	-	-	A1
<i>D. niveum</i>	AM231291	Fr-K10	682 bp	-	-	A1
<i>D. niveum</i>	HE614616	It-K66	683 bp	-	-	A1
<i>D. niveum</i>	HE614617	Uk-K79	683 bp	-	-	A1
<i>D. saundersii</i> ⁵	MZ313549	Mx-K30	532 bp	-	-	A1
<i>D. testeceum</i>	AM231292	Pr3-1	529 bp	-	-	G1
<i>Didymium iridis</i>	AJ938151	CR19-1	523 bp	-	-	A1
<i>D. iridis</i>	AJ938150	CUR1-4	637 bp	-	-	A1
<i>D. nigripes</i>	AF239230	-	654 bp	-	-	A1
<i>D. squamulosum</i>	AM231293	CR10	792 bp	-	-	U1
<i>Fuligo leviderma</i>	DQ903676	AMFD130	517 bp	-	-	A1
<i>Lepidoderma carestianum</i>	HE614618	It-K71	393 bp	-	-	A2
<i>L. crustaceum</i>	HE614619	It-K62	396 bp	-	-	A2
<i>L. peyerimhoffii</i> ⁵	MZ313552	It-K63	388 bp	-	+	A2
<i>L. tigrinum</i>	DQ903678	AMFD192	785 bp	-	-	A1
<i>Mucilago crustacea</i> ⁵	MZ313554	No-K94	1274 bp	DR-P9	-	A1
Order Stemonitales						
<i>Brefeldia maxima</i>	JQ031957	MM24519	1304 bp	HEG-P1	+	A3
<i>Colloderma oculatum</i>	JQ031959	HS2885	421 bp	-	+	A3
<i>C. robustum</i>	JQ031960	AMFD270	384 bp	-	+	A1
<i>Comatricha nigra</i>	DQ903683	AMFD155	788 bp	-	+	A1
<i>C. pseudoalpina</i>	DQ903673	MM23892	815 bp	-	+	U1
<i>Diacheopsis pauxilla</i>	JQ031966	MM29883	628 bp	-	-	A1
<i>Diachea subsessilis</i>	JQ031964	MM24463	919 bp	-	-	U1
<i>Lamproderma aeneum</i>	JQ031969	MM36255	545 bp	-	-	A3
<i>L. arcyrioides</i>	JQ031973	MM37005	1006 bp	DR-P9	+	A2
<i>L. cacographicum</i>	JQ031976	AMFD310	957 bp	DR-P9	-	A1
<i>L. columbinum</i>	HQ687204	F2	503 bp	-	-	A2
<i>L. columbinum</i>	HQ687196	106	563 bp	-	-	U2
<i>L. columbinum</i>	HQ687197	63b	474 bp	-	-	A2
<i>L. columbinum</i>	HQ687200	132	480 bp	-	+	A3
<i>L. disseminatum</i>	JQ031978	AMFD38	498 bp	-	-	A1
<i>L. echinosporum</i>	JQ031980	AMFD136	486 bp	-	+	A3
<i>L. echinosporum</i>	JQ031979	AK06016	481 bp	-	+	A3
<i>L. pseudomaculatum</i>	JQ031985	MM37354	411 bp	-	-	A3
<i>L. puncticulatum</i>	HQ687194	172	514 bp	-	+	A2
<i>L. puncticulatum</i>	HQ687202	162	511 bp	-	+	A1
<i>L. puncticulatum</i>	HQ687195	3	517 bp	-	+	A1
<i>L. sauteri</i>	DQ903674	AMFD208	543 bp	-	+	A1
<i>L. zonatum</i>	DQ903672	MM21644	434 bp	-	-	A1
<i>Meriderma carestiae</i>	JQ031999	MM35985	566 bp	-	-	A1
<i>M. carestiae</i>	DQ903671	AMFD173	566 bp	-	-	A1
<i>M. cribrarioides</i>	JQ032000	MM37106	538 bp	-	-	A1
Order Liceales						
<i>Licea marginata</i>	JX481296	DWM7368	455 bp	-	+	A3
Order Trichiales						
<i>Calomyxa metallica</i>	JX481284	AMFD483	532 bp	-	-	A3
<i>Trichia varia</i>	KM494993	sc22370	1531 bp	HEG-P1	-	A3
<i>T. varia</i>	KM494994	LE259268	586 bp	-	-	A3
<i>T. varia</i>	KM494995	LE259461	586 bp	-	-	A3
<i>T. varia</i>	KM494996	JVR848	1531 bp	HEG-P1	-	A3
<i>T. varia</i>	KM494997	sc22386	559 bp	-	-	A1
<i>T. varia</i>	KM494998	sc22408	559 bp	-	-	A1
<i>T. varia</i>	KM494999	sc22409	559 bp	-	-	A1
<i>T. varia</i>	KM495003	sc22442	559 bp	-	-	A1
<i>T. varia</i>	KM495005	sc22517	559 bp	-	-	A1
<i>T. varia</i>	KM495006	sc27697	559 bp	-	-	A1

<i>T. varia</i>	KM495009	sc27742	559 bp	-	-	A1
<i>T. varia</i>	KM495010	sc27839	559 bp	-	-	A1
<i>T. varia</i>	KM495018	KRAM M-1585	559 bp	-	-	A1
<i>T. varia</i>	KM495019	sc27686	559 bp	-	-	A1
<i>T. varia</i>	KM495020	sc27737	559 bp	-	-	A1
<i>T. varia</i>	KM495021	sc27850c1	559 bp	-	-	A1
<i>T. varia</i>	KM495022	sc27860c4	559 bp	-	-	A1
<i>T. varia</i>	KM495023	sc27507	559 bp	-	-	A1
<i>T. varia</i>	KM495024	sc27648c1	559 bp	-	-	A1
<i>T. varia</i>	KM495026	sc27667c1	559 bp	-	-	A1
<i>T. varia</i>	KM495027	sc27667c2	559 bp	-	-	A1
<i>T. varia</i>	KM495028	sc27772c2	559 bp	-	-	A1
<i>T. varia</i>	KM495029	sc27850c2	559 bp	-	-	A1
<i>T. varia</i>	KM495030	LE254838	1544 bp	HEG-P1	-	A1

Host species	Acc No	Isolate	Size	Insert	P5d	P7
--------------	--------	---------	------	--------	-----	----

Notes:

¹Size in base pairs of the S1389 group I intron insertion.

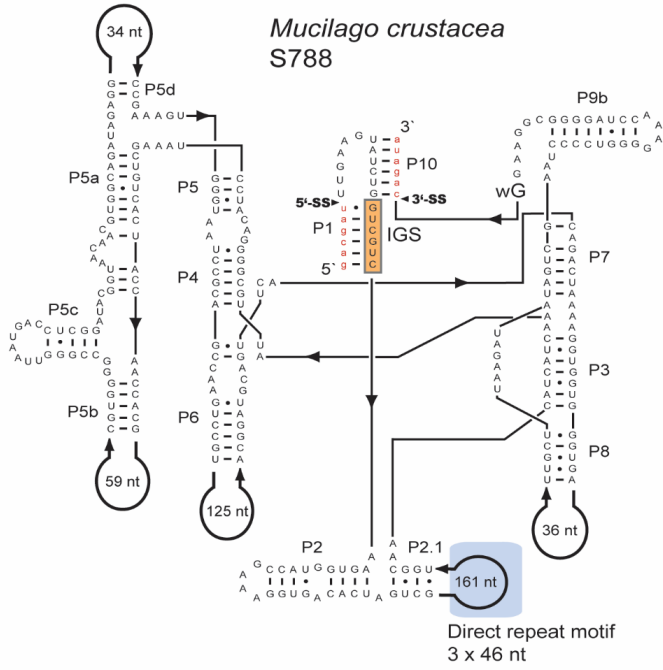
²Type of large insertions in group I ribozyme paired segment P1 or P9. HEG, homing endonuclease gene; DR, direct repeat motifs.

³Presence or absence of the optional paired segment P5d.

⁴G-binding sequence variants in segment P7, as indicated in Figure 3.

⁵This work.

a



b

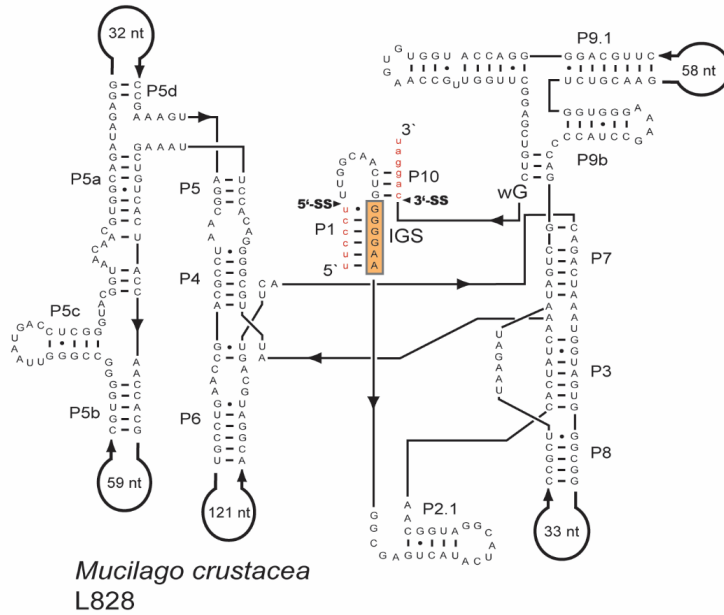
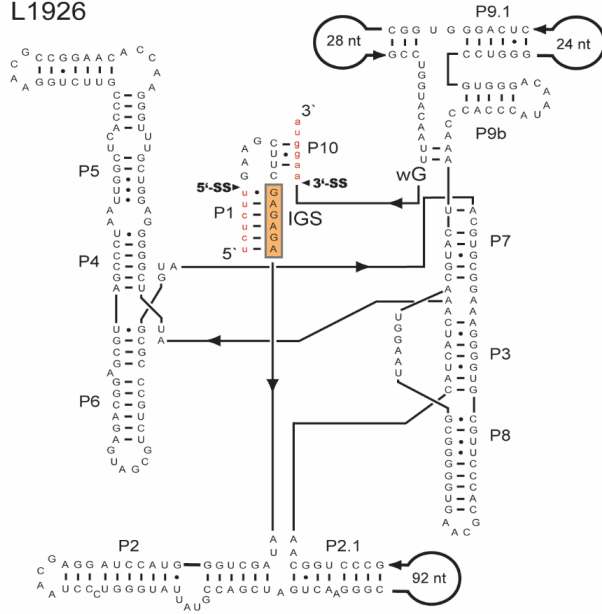


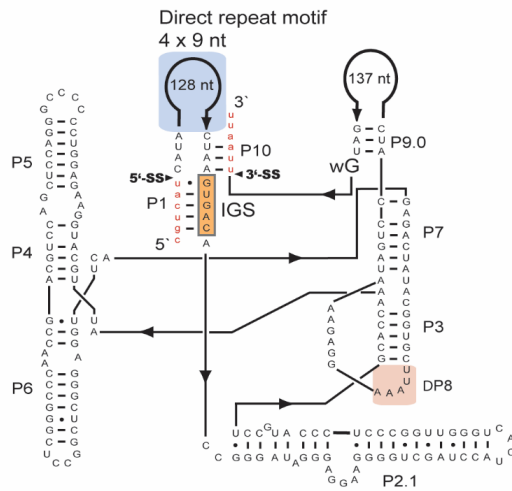
Figure S1 (a&b)

C

Mucilago crustacea
L1926



d

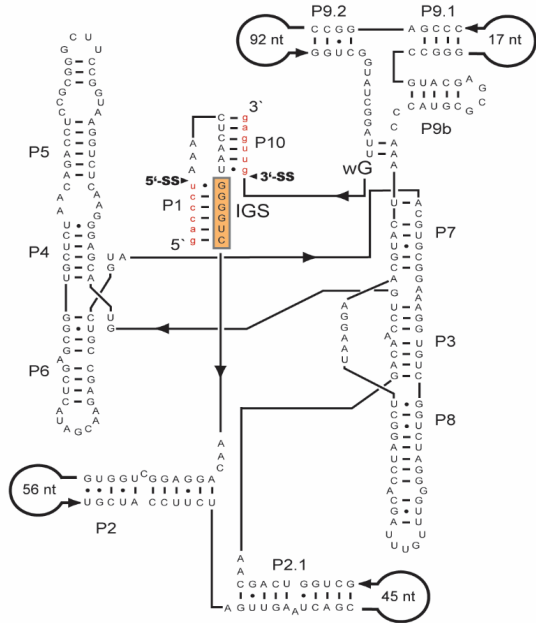


Mucilago crustacea
L1949

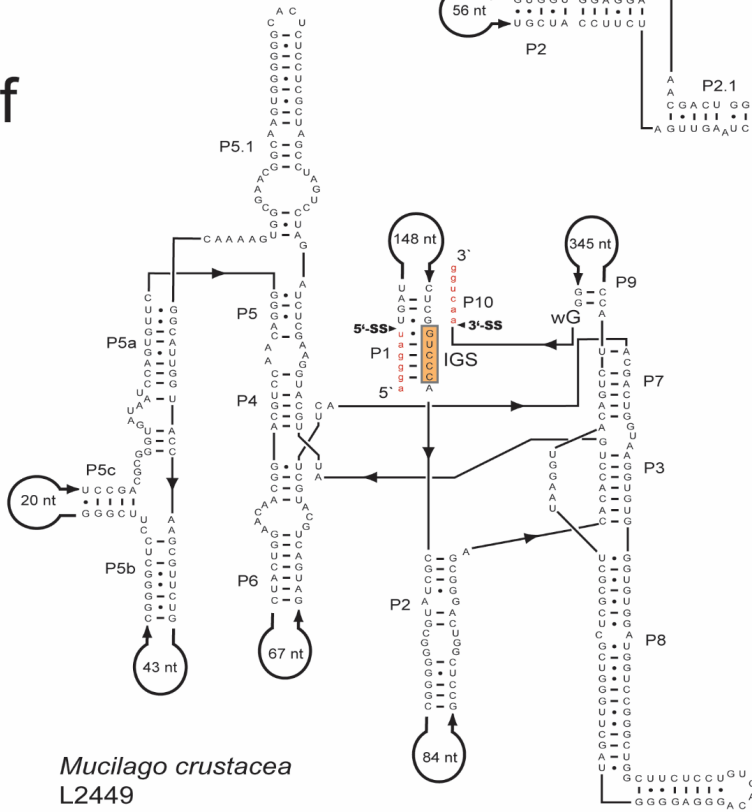
Figure S1 (c&d)

e

Mucilago crustacea
L2066



f



Mucilago crustacea
L2449

Figure S1 (e&f)

Figure S1: Secondary structure diagrams of the *Mucilago crustacea* rDNA group I introns. (a) S788 is a group IC1 intron in the SSU rRNA gene. Note that segment P2.1 (blue box) contains a 46-nt direct repeat motif. (b) L828 is a group IC1 intron in

the LSU rRNA gene. (c) L1926 is a group IE intron in the LSU rRNA gene. (d) L1949 is an unclassified group I intron in the LSU rRNA gene. Note that the P8 segment of the catalytic core is missing (pink box), and that segment P1 (blue box) contains a 9-nt direct repeat motif. (e) L2066 is a group IE intron in the LSU rRNA gene. (f) L2449 is a group IC1 intron in the LSU rRNA gene. Note that P1 and P9 contain extension sequences, but with no reading frame or direct repeat features. Common note to all intron diagrams: P1-P10, paired RNA segments; 5' SS and 3' SS, exon-intron splice sites; IGS, internal guide sequence; ωG, last nucleotide residue of intron; red lowercase letters, exon sequences.


```

<<P5> <<<P4>><<<P6>>>><<<P6>>>>> <<<P7>> <<P3>>><<P8>><P8>> <<P7>> *
M2313554 GGUU-AAACU-TCGACAGCAAGUCUUCAGGGAUGAGUUCACAGACCAAACGUCAGUGGGUGGCGCCUUAAGAGAUGUGCGGUGGuuua
M2313547 GGUACGACC-CGGCCGCAAGCCCGUCAGGGGUCAGCCACAGACCAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCUUUGGuuua
M2313549 GGUACGACC-UGCGGCAGAGCCCGUCAGGGGUCAGCCACAGACCAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCUUUGGuuua
M2313548 GGGC-AAAC-UGCGGCAGAGCCCGUCAGGGGUCAGCUCACAGACCAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGCACGuuua
Mz313546 GGGCCAAAC-UGCAGCAGAGCCCGUUAAGGGGUCAGCUCACAGGCCAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
D0903677 GGGC-AAAC-UGCGGCAGAGCCCGUCAGGGGUCAGCUCACAGACCAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
HE614614 GGGCCAAAC-UGCAGCAGAGCCCGUCAGGGGUCAGCUCACAGGCCAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
AM231291 GGGUUAACC-UGCGGCAGAGCCCGUCAGGGGUCAGCUCACAGAACAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCUUUGGuuua
HE614616 GGGUUAACC-UGCGGCAGAGCCCGUCAGGGGUCAGCUCACAGAACAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCUUUGGuuua
HE614617 GGGUUAACC-UGCGGCAGAGCCCGUCAGGGGUCAGCUCACAGAACAAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCUUUGGuuua
AM231292 GGUU-AAAC-UGCGGCAGAGCCCGUCAGGGGUCAGCCACAGACCAAACUUAAGUGGGUGGUCGCUUAAGAUAUGUGGCGCCUUGGuuua
A1938151 GGUUAUAU-UGCCUCCAAACCCUUGUUGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCUUUGGuuua
A1938150 GGGC-AAAC-UGCAGCAAGCCCGUUAAGGGGUCAGUUCACAGACCUAACGCCAGUGGGUGGUCGCUUAAGAUAUGUGGCGACGuuua
AF239230 GGGC-AAAC-UGCAGCAAGCCCGUUAAGGGGUCAGUUCACAGACCUAACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
AM231293 GGUU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
HE614618 GGUU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
HE614619 GGUU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
D0903678 GGUU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
DQ903676 GGGC-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
JQ031957 GGGG-AAAC-UGCAGCAGAGCCCGUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031959 GGGC-AAAC-UGCAGCAGAGCCCGUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031960 GGGU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAGAUAUGUGGCGAACCCGuuua
DQ903683 GAGU-AAUC-UGCGGCAGAAUUCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
DQ903673 GGGG-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031966 GGGU-AAUC-UGCAGCAGAACCCCGUCAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031964 AGGU-GAAC-UGUAGCAGAGUUCUUAAGGGGUCAG-CAACUUCGCAAAAGCCAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031969 GGGU-GAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031973 GGGG-AAUC-UGCAGCAGAGCCCGUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031976 GGGU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687204 GGGU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687196 GGGU-AAUC-UGCAGCAGAACCCCGUCAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687197 GGGU-AAUC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687200 GGGU-AAUC-UGCAGCAGAACCCCGUCAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031978 GGGU-AAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031980 GGAAC-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031979 GGGG-AAAC-UGCAGCAGAGCCCGUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031985 GGGU-AAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUACGCUUAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687194 UGGC-AAAC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687202 UGGC-AAAC-UGCAGCAGAACCCUUAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
H0687195 GGGC-AAAC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGACCUAACGCGAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
DQ903674 GGGU-AAUC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
DQ903672 GGGU-AAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ031999 GGGU-AAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
D0903671 GGGU-AAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JQ032000 GGGU-AAUC-UGCAGCAGAACCCCGCAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JX481296 GGGC-AAUC-UGCAGCAGAACCCCGGGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494993 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494994 GGGU-GAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494995 GGGU-GAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494996 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494997 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494998 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM494999 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495003 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495005 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495006 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495009 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495010 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495018 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495019 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495020 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495021 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495022 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495023 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495024 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495026 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495027 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495028 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495029 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
KM495030 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
JX481284 GGGU-AAAC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
V01416 UCCU-AAACCGCAGCAGUCCUUAAGGGGUCAGUUCACAGAUCAUAAUGUUAUAGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua
M2313554* GGGU-AAUC-UGCAGCAGAACCCCGUUAAGGGGUCAGUUCACAGAUCAUAAAGUGGGUGGUCGCUUAAAUAUAGCCCGUUAUGuua

```

Figure S2: Sequence alignment of core structured nucleotides of myxomycete S1389 group I intron. Dashes correspond to deleted positions. Secondary structure paired segments (P1-P8) are shown above the alignment. Intron sequences are indicated by GenBank accession numbers. V01416 and MZ313554* correspond to the out-group *Tetrahymena* intron Th.L1925 and *Mucilago* intron Mcr.S788, respectively.

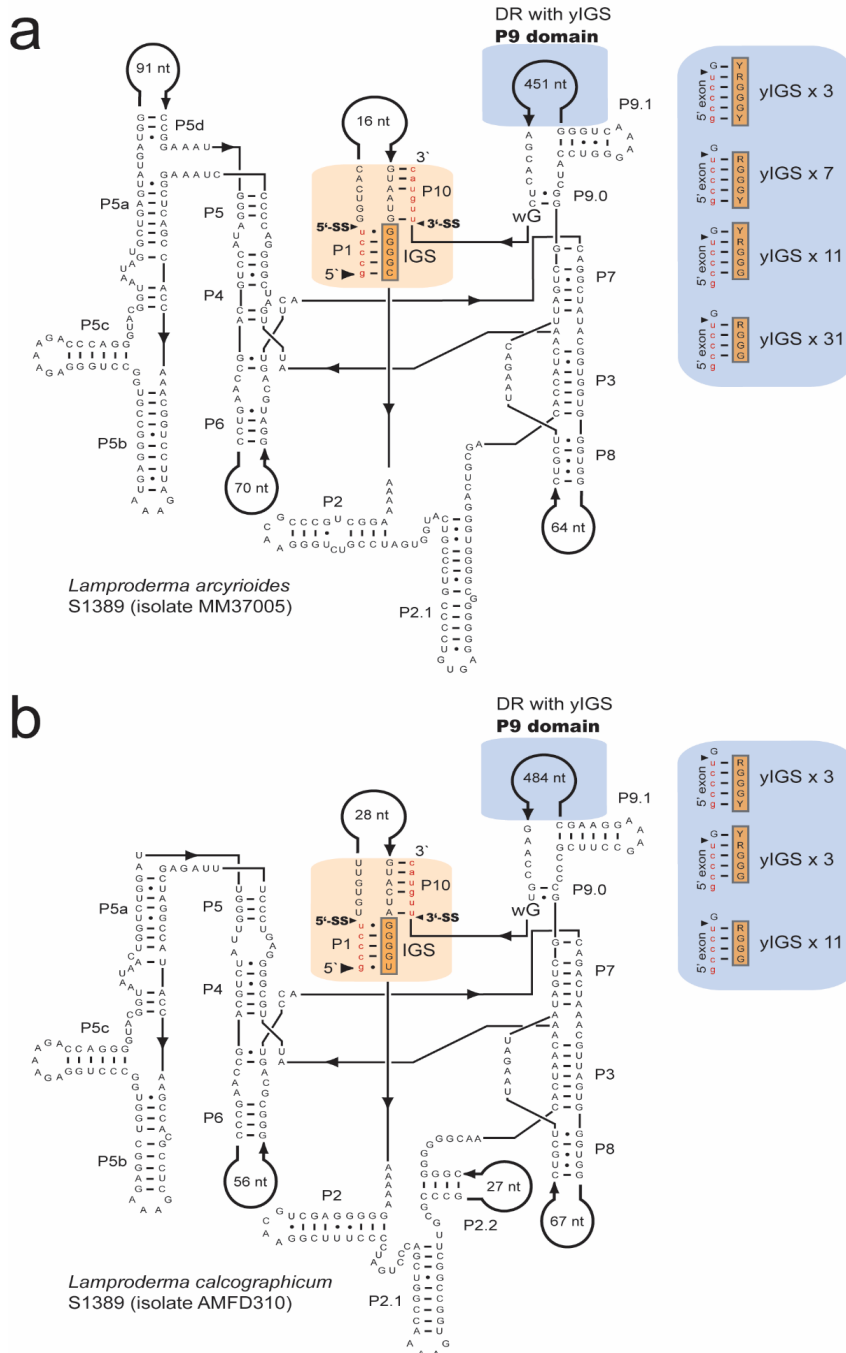


Figure S3

Figure S3: Secondary structure diagrams of (a) *Lamproderma arcyrioides* and (b) *Lamproderma arcyrioides* S1389 group I intron. P1-P10, paired RNA segments; 5' SS and 3' SS, exon-intron splice sites; IGS, internal guide sequence. The introns contain large extensions in P9 that contains direct repeat motifs (blue box) with ΨIGS sequences motifs.

diagram of *T. varia* group I introns (KM494996 and KM495030). Homing endonuclease gene (HEG) insertions are located in segment P1. P1-P10, paired RNA segments; 5' SS and 3'SS, exon-intron splice sites; IGS, internal guide sequence; ωG, last nucleotide residue of intron; red lowercase letters, exon sequences; black lowercase letters, nucleotide positions only present in one of the introns. R represents A or G (purine); Y represents C or U (pyrimidine); K represents G or U; M represents A or C; S represents C or G; W represents A or U.

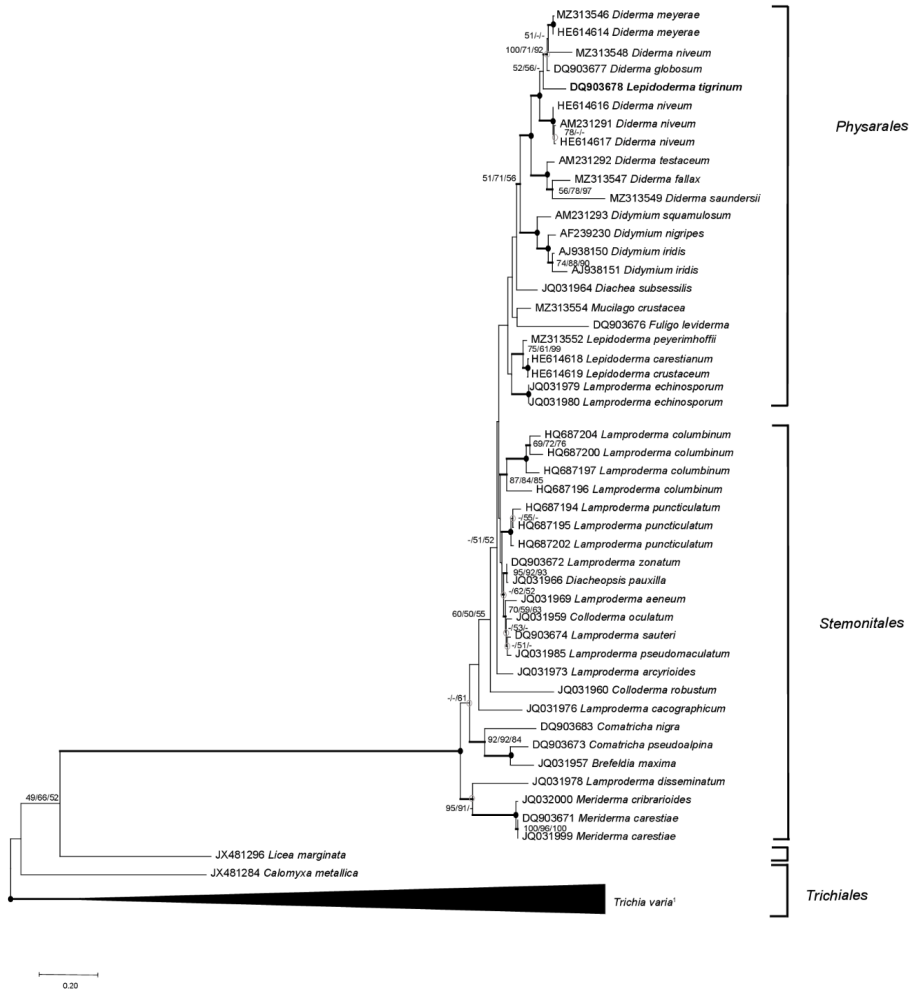


Figure S5: Molecular phylogeny of myxomycete taxa based SSU rDNA sequences. The SSU topology is obtained by neighbor-joining (NJ) analysis of 72 taxa and 1,575 nt aligned positions (SSU dataset 1; Table S1). The tree is rooted with the *T. varia* SSU rDNA sequence. The NJ, maximum likelihood (ML), and maximum parsimony (MP) bootstrap replicates ($\geq 50\%$) are given at each node. Bayesian posterior (BI) probabilities (≥ 0.95) are shown in bold branches. Black dots at branch points; maximum support in NJ, ML, and MP ($\geq 97\%$). The scale bar indicates the fraction of substitutions per site.

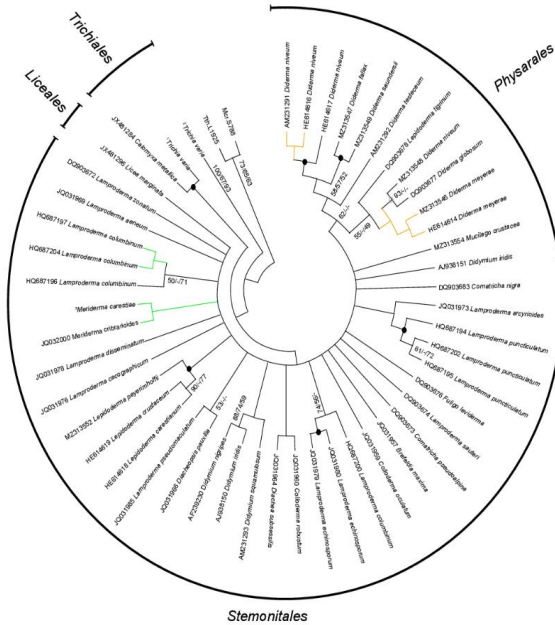



Figure S6: Molecular phylogeny of myxomycete S1389 group I introns. The intron topology is obtained by neighbor-joining (NJ) analysis of 74 taxa (including group introns from *Tetrahymena* (Tth.1925) and *Mucilago* (Mcr.S788), and 166 nt aligned positions (intron dataset 2; Table S1). The NJ-, maximum likelihood (ML), and maximum parsimony (MP) bootstrap replicates ($\geq 50\%$) are given for each node. Bayesian posterior (BI) probabilities (≥ 0.95) are shown in bold branches. Black dots at branch site indicate maximum support in NJ, ML, and MP ($\geq 98\%$). Green line indicates maximum support in NJ, ML and MP ($\geq 90\%$). Orange line indicates maximum support in NJ, ML, and MP ($\geq 70\%$). Gray circles indicate alternative topologies in NJ, ML, and MP. ¹ *Tricia varia* clade is composed of four *T. varia* isolates (KM494993-6). ² *Tricia varia* clade is composed of twenty *T. varia* isolates (KM494997-9, KM495003, KM495005, KM495006, KM495009, KM495010, KM495018-30). ³ *Meriderma carestiae* clade is composed of two *M. carestiae* isolates (JQ031999 and DQ903671). The scale bar indicates the fraction of substitutions per site.

Paper II

This is an open-access article, reproduced and distributed under the terms of the
Creative Commons Attribution License (CC BY)

Article

Structural Organization of S516 Group I Introns in Myxomycetes

Betty M. N. Furulund ¹, Bård O. Karlsen ², Igor Babiak ¹ , Peik Haugen ³ and Steinar D. Johansen ^{1,*}

¹ Genomic Division, Faculty of Biosciences and Aquaculture, Nord University, 8049 Bodø, Norway; betty.martine.furulund@nordlandssykehuset.no (B.M.N.F.); igor.s.babiak@nord.no (I.B.)

² Research Laboratory and Department of Microbiology, Nordland Hospital Trust, 8005 Bodø, Norway; bard.ove.karlsen@nordlandssykehuset.no

³ Department of Chemistry and Center for Bioinformatics, Faculty of Science and Technology, UIT—The Arctic University of Norway, 9037 Tromsø, Norway; peik.haugen@uit.no

* Correspondence: steinar.d.johansen@nord.no

Abstract: Group I introns are mobile genetic elements encoding self-splicing ribozymes. Group I introns in nuclear genes are restricted to ribosomal DNA of eukaryotic microorganisms. For example, the myxomycetes, which represent a distinct protist phylum with a unique life strategy, are rich in nucleolar group I introns. We analyzed and compared 75 group I introns at position 516 in the small subunit ribosomal DNA from diverse and distantly related myxomycete taxa. A consensus secondary structure revealed a conserved group IC1 ribozyme core, but with a surprising RNA sequence complexity in the peripheral regions. Five S516 group I introns possess a twintron organization, where a His-Cys homing endonuclease gene insertion was interrupted by a small spliceosomal intron. Eleven S516 introns contained direct repeat arrays with varying lengths of the repeated motif, a varying copy number, and different structural organizations. Phylogenetic analyses of S516 introns and the corresponding host genes revealed a complex inheritance pattern, with both vertical and horizontal transfers. Finally, we reconstructed the evolutionary history of S516 nucleolar group I introns from insertion of mobile-type introns at unoccupied cognate sites, through homing endonuclease gene degradation and loss, and finally to the complete loss of introns. We conclude that myxomycete S516 introns represent a family of genetic elements with surprisingly dynamic structures despite a common function in RNA self-splicing.

Keywords: catalytic introns; homing endonuclease; intron evolution; ribozyme; twintron



Citation: Furulund, B.M.N.; Karlsen, B.O.; Babiak, I.; Haugen, P.; Johansen, S.D. Structural Organization of S516 Group I Introns in Myxomycetes.

Genes **2022**, *13*, 944. <https://doi.org/10.3390/genes13060944>

Academic Editors: Lina Ma, Tao Huang and Xianyang Fang

Received: 6 May 2022

Accepted: 23 May 2022

Published: 25 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Group I introns are self-splicing and mobile elements sporadically distributed within genomes of bacteria, bacteriophages, chloroplasts, mitochondria, eukaryotic viruses, and the nuclei of eukaryotes [1,2]. Nuclear group I introns are restricted to ribosomal DNA (rDNA) of fungi and protists and appear common in myxomycetes (plasmodial slime molds) [3]. Most nucleolar group I introns interrupt functionally important sequences in the small subunit (SSU) and large subunit (LSU) rRNAs corresponding to the peptidyl transferase center and decoding center [4,5]. Intriguingly, a given insertion site harbors related intron sequences with a common evolutionary history even in distantly related taxa [6–11]. An rDNA group I intron nomenclature has been established based on species name abbreviation, insertion site, and the *E. coli* rRNA gene numbering system [12].

Group I introns catalyze their own splicing due to the intron-encoded ribozyme. The splicing mechanism is well described and involves two succeeding transesterification reactions initiated by an attack of an exogenous guanosine (exoG) cofactor [3,13]. The resulting products are (i) a perfectly ligated exon RNA originating from transcribed host gene and (ii) the excised intron RNA sequence. Most nuclear group I introns also catalyze an exoG-independent reaction pathway leading to full-length intron circles and fragmented exons [14]. The catalytic RNA activities are dependent on a highly organized catalytic core

structure composed of a series of paired segments (P1 to P9), which are further organized into three structural domains [3,15]. A crucial structural segment in catalysis is the P7 element, which is located at the center of the catalytic core [16]. Nuclear group I introns are restricted to either one of two subgroups, i.e., group IC1 and group IE, which possess unique and characteristic structural features [3,17–19].

Nuclear group I introns may also carry additional genes and sequences. Most common are homing endonuclease genes (HEGs), which are found in approximately 5–10% of known nuclear group I introns and code for site-specific endonucleases involved in intron mobility [3,20–22]. Nuclear HEGs are reported to be inserted into different peripheral intron segments in P1, P2, P6, P8, or P9, and the corresponding proteins belong to the His-Cys family of homing endonucleases [23,24]. Several structural features appear to facilitate the cellular expression of nuclear HEGs embedded in ribosomal DNA [25]. Some homing endonuclease mRNAs have a unique lariat cap at their 5' ends catalyzed by a separate intron-encoded ribozyme [26–28], and most appear polyadenylated at their 3' ends [11,29,30]. Interestingly, several HEGs also carry small spliceosomal introns of about 50 nt that have to be removed in order to generate a functional open reading frame [11,29–31] and thus may facilitate the expression.

Several examples of direct repeat (DR) arrays within nuclear group I introns have been reported [7,9,11,32,33]. These arrays appear to have a similar location in the intron RNA structure as HEGs, but their biological role remains unknown. Variations in sequence motifs, number of copies, and size of arrays are generally observed, but a common feature appears that arrays do not interfere with ribozyme functions due to their locations in peripheral segments. Recently, we reported a complex pattern of DR array within the S1389 group I intron in the myxomycete *Mucilago crustacea* that contained 34 copies of the intron internal guide sequence [11], which may suggest a functional role in splicing.

In a previous study, we reported self-splicing and structural organization of nuclear S516 group I introns [6]. These introns interrupt a tRNA binding and decoding domain of the SSU rRNA [34] and have to be precisely removed by splicing in order to restore a functional rRNA. Introns at position S516 were found in ascomycete and basidiomycete fungi, and in various protists phyla such as the red, green, and brown alga, in amoeba and amoeba-flagellates, and in myxomycetes. Some of the introns belonged to the group IE subclass (e.g., ascomycetes and green alga) whereas others to the group IC1 subclass (e.g., red alga, amoeba, amoeba-flagellates, and myxomycetes). The group IC1 introns showed the most variability in structural organization, and HEGs and pseudo-HEGs were identified in red alga [35,36], in amoeba [6], and in amoeba-flagellates [37]. The latter group involved twin-ribozyme introns in several *Naegleria* species and in *Allovahlkamfia* and harbored complex insertions in the P6 segment corresponding to HEGs and lariat-capping ribozymes [28,38,39].

In the original S516 intron study [6], only three myxomycete introns were available. These introns were from related taxa, had a similar RNA structural fold, and contained no HEG insertions. In the present study, we extended the analysis to include 75 S516 group IC1 introns in myxomycetes representing five taxonomic orders. The intron RNA possessed a conserved catalytic core region, but highly variable peripheral regions due to DR arrays and HEGs interrupted with small spliceosomal introns.

2. Materials and Methods

2.1. Small Subunit Ribosomal DNA Sequences

All rDNA sequences included in this work, except that of *Didymium alpinum* isolate Fr-K12, were retrieved from the NCBI database (<https://www.ncbi.nlm.nih.gov>, accessed on 15 March 2022; Table S1). *D. alpinum* was collected as spores in the French Alps by Ms. Kari Haugli (University of Tromsø, Tromsø, Norway). Spores were germinated and approximately 10⁸ amoeba cells were harvested from DS/2 agar plates, total DNA extracted, and rDNA sequenced as described previously [8].

2.2. Sequence Alignments

SSU rDNA sequences (with introns removed) were assembled and aligned in the software program Geneious prime[®] 2022.0.1 (<https://www.geneious.com>, accessed on 15 February 2022). Multiple sequence alignments were generated using MAFFT version 7.450 [40,41] with default settings. The SSU rDNA dataset 1 (1524 nt) and dataset 2 (1784 nt) alignments were manually adjusted according to published information about myxomycete SSU rDNA genes [42,43], excluding parts of variable regions that were not confidently aligned. Group I intron core sequences (143 bp) were manually aligned and strictly based on secondary structure features as described previously [6,8,11] using the Geneious prime[®] software program.

2.3. Phylogenetic Analysis

The tree-building methods of neighbor joining (NJ), maximum likelihood (ML), maximum parsimony (MP), and minimal evolution (ME), interpreted in MEGA X [44], were used in the SSU rDNA and S516 intron datasets with default settings as reported previously [11]. All sequence alignments were model tested prior to tree constructions by the MEGA X software [44]. The topology of NJ, ML, and ME trees was evaluated by bootstrap analyses. The evolutionary relationship generated by SSU rDNA dataset 1 was reconstructed with NJ and ML using the Kimura 2 (K2) evolutionary model [45]. The robustness was tested by bootstrapping (500 replicates). SSU rDNA dataset 2 was reconstructed with NJ, ML, and ME using the K2 evolutionary model [45]. The robustness was tested by bootstrapping (500 replicates). The evolutionary relationships of S516 intron core sequences (intron dataset 1) were reconstructed with the NJ method and the Jukes–Cantor (JC) model using MEGA X [44] with default settings. To test the robustness of the nodes, the trees were tested with NJ–JC (500 replicates) and ME–JC (500 replicates). Intron dataset 2 was reconstructed with NJ, ML, and ME using the JC evolutionary model [45]. The robustness was tested by bootstrapping (500 replicates).

3. Results

3.1. *Diderma alpinum* Fr-K12 Contains a Group I Intron at Position 516 in the Nuclear SSU rRNA Gene

Ribosomal DNA sequence analysis of the myxomycete *D. alpinum* (Fr-K12) identified a 534-bp group I intron (Dalp.S516) at position S516 in the nuclear SSU rRNA gene. An RNA secondary structure diagram is presented in Figure 1. Dalp.S516 represents a typical group IC1 intron fold, similar to that of the *Tetrahymena* intron [19,46,47], with a well-conserved core organization based on paired RNA segments (P1–P10). Several general structure features common to group IC1 introns are present (Figure 1): (i) A conserved U:G pair at the 5'SS in segment P1, which partly constitutes the internal guide sequence and is critical in the first reaction step of intron self-splicing. (ii) A conserved guanosine binding site in segment P7 (yellow box) is essential in the catalytic site of the ribozyme. (iii) A tertiary segment (P13) that likely contributes to the overall stability of the ribozyme core (green box). (iv) Three GNRA tetra-loops (within the P5b, P8, and P9b extensions) were probably also involved in RNA–RNA interactions and stability. The GAAA-loop in P5b is expected to interact with a predicted receptor in P6. (v) A conserved terminal guanosine (ω G) at the 3'SS is involved in the second reaction step of catalysis. (vi) A proposed and putative 10-bp interaction between P2 and P6 extensions, with a possible role in ribozyme core stability (green box). It is important to note that this interaction has not been tested experimentally.

3.2. Structure Variation among Myxomycetes S516 Group I Introns

Group I introns at position S516 are widely but sporadically distributed among fungi and protists, including myxomycetes [6]. We collected 75 myxomycete S516 introns representing five distantly related orders (Physarales, Stemonitales, Liceales, Echinosteliales, and Trichiales), 21 genera, and 38 distinct species (Table S1). The majority of the SSU rRNA gene sequences were retrieved from the NCBI database, and one sequence (*D. alpinum*

Fr-K12) was provided in this study. To assess structural variations among the S516 introns, 143 nucleotide positions in the ribozyme core region common to all 75 introns were strictly aligned according to secondary structure features (Figure S1). The consensus secondary structure diagram (Figure 2) shows a typical group IC1 intron fold with a conserved substrate (P1–P2), scaffold (P4–P5–P6), and catalytic (P3–P7–P8) domains [see 19,46]. They all show a conserved guanosine binding site in P7, an extended P5d hinge region, and a P13 segment. The latter represents a common but only weakly conserved structural feature among the S516 introns. Most sequence variations are found in their peripheral extension regions, especially the P2 extension. Here, five S516 introns contain homing endonuclease gene (HEG) insertions, and eleven S516 introns were found to contain direct repeat (DR) arrays. HEGs were exclusively identified in P2, whereas DRs were found in all extension regions (Figure 2).

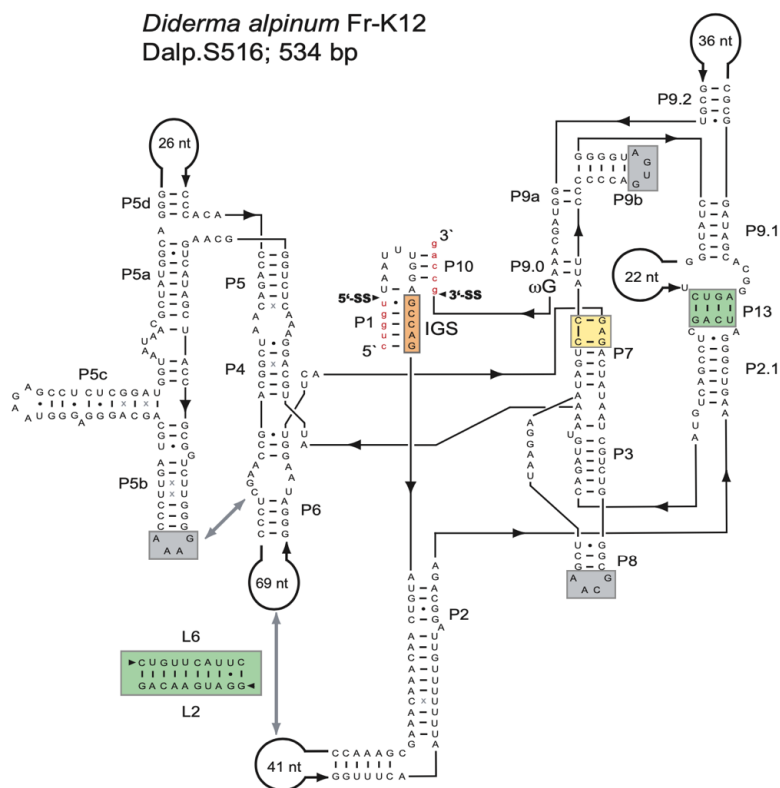


Figure 1. Secondary structure of the *Diderma alpinum* S516 group I intron RNA. The diagram of the representative 534-nt group IC1 intron RNA is presented according to [15]. The intron has a well-conserved core organization based on paired RNA segments (P1–P10). The guanosine binding site in P7 is indicated (yellow box), as well as the universal conserved terminal guanosine (ω G). Long-distance RNA–RNA interactions are commonly found in group I intron RNAs, involving receptors and GNRA-loops (gray boxes), or regular base pairings (green boxes). 5'SS and 3'SS, exon-intron splice sites; IGS, internal guide sequence: (–) Watson–Crick base pair; (•) non-Watson–Crick GU base pairing; (x) additional non-Watson–Crick base pairings with comparative support. Exon sequences are shown as red lowercase letters.

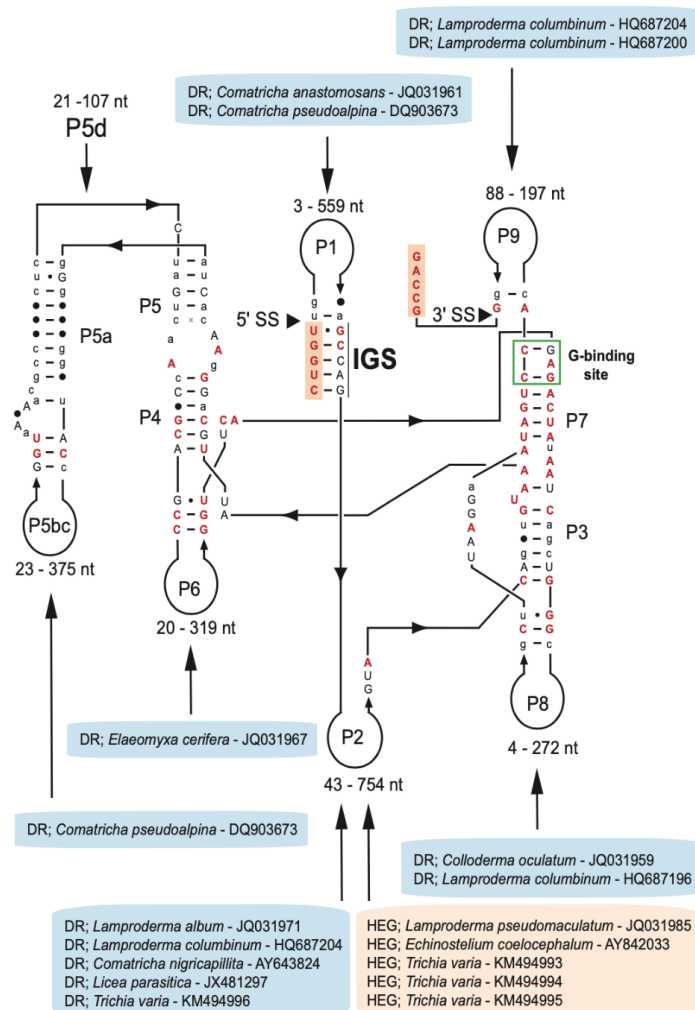


Figure 2. Consensus secondary structure diagram of S516 group I intron RNA in myxomycetes. The consensus structure is based on 143 nucleotide positions in the catalytic core common among introns from 75 taxa (see Table S1). Extensive sequence size variations are noted in peripheral regions. While all homing endonuclease genes (HEGs) are found as P2 extensions, direct repeat (DR) extensions are located in most peripheral regions. The guanosine (G) binding site in P7 is indicated. P1–P9, paired RNA segments; 5' SS and 3' SS, exon-intron splice sites. Invariant nucleotide positions among the 75 introns are shown as red uppercase letters. Black uppercase letter, $\geq 90\%$ conservation; lowercase letters, $\geq 50\%$ conservation; filled circles, $< 50\%$ conservation.

3.3. Complex Inheritance Pattern of Myxomycete S516 Group I Introns

Evolutionary relationships among the myxomycete S516 introns were assessed by phylogenetic analyses. First, a host gene phylogeny was established based on an alignment of SSU rDNA sequences covering 1524 positions from 66 taxa. The topology of a representative neighbor-joining (NJ) tree (Figure S2) is in general agreement with previously reported studies using a similar selection of myxomycete taxa [11,33,48]. Next, we compiled an

intron dataset containing 143 sequence positions from the catalytic core region from a total of 75 myxomycete S516 group I introns (Figure S1) and inferred the evolutionary relationships using the NJ method (Figure S3). Whereas some of the clades are strongly supported by bootstrap analyses, the positioning of the same clades in the overall tree topology is not. As noted earlier for S1389 group I introns [11], intron-family phylogenies are typically based on few aligned positions and a large number of taxa, thus resulting in trees with limited robustness of nodes. We did, however, observe an overall congruency between the SSU rDNA and intron-based phylogenies, particularly between clades with closely related introns (e.g., among *Diderma* taxa and among *Trichia* taxa; compare Figures S2 and S3), suggesting that at least some S516 group I introns possessed a vertical inheritance pattern during the evolution of myxomycetes.

Most introns, however, showed a relationship that is not consistent with stable vertical inheritance. To better resolve the evolutionary relationship between S516 introns, we analyzed a smaller dataset consisting of 14 selected taxa (Figure 3), and from the resulting NJ tree, we made several interesting observations. (i) While the six *L. columbinum* isolates cluster together in the SSU rDNA analysis to a well-defined clade (blue box), the corresponding introns are found scattered on the tree, e.g., the *L. columbinum* (HQ687198) and *L. columbinum* (HQ687197) introns are distantly related to the other *L. columbinum* introns. (ii) SSU rDNA analysis strongly support that *M. aggregatum* and *M. cribrarioides* are two closely related taxa from the same genus (green box), whereas intron analysis of the core region indicates that the S516 introns are distantly related. This is further supported by structural comparisons (Figure S4), where P2 and P6 peripheral extensions were found to be clearly different. (iii) Two closely related *C. nigricapillitia* isolates cluster together in the SSU rDNA analysis (orange box), but their corresponding S516 introns suggest a distant relationship. In summary, our phylogenetic analyses suggest that myxomycete S516 introns share a complex evolutionary history, with several clear examples of both vertical and horizontal transfers.

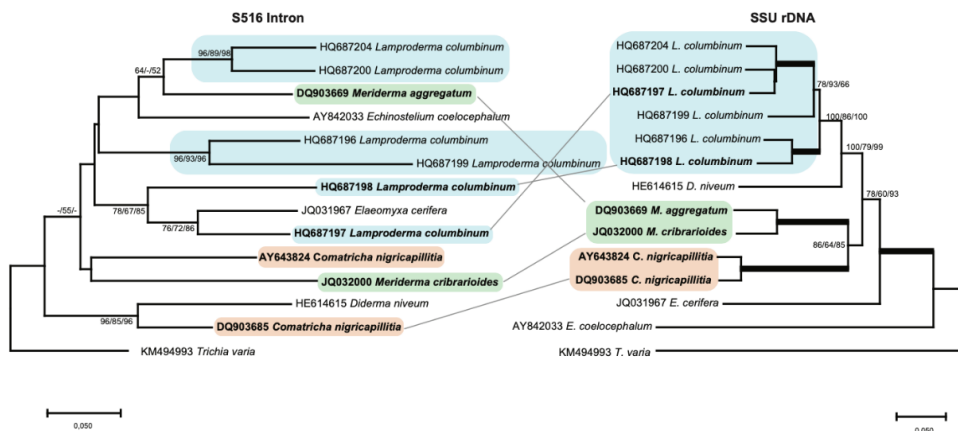


Figure 3. Molecular phylogeny of S516 group I introns as candidates to horizontal transfer. The intron and SSU topology is obtained by neighbor-joining (NJ) analysis of 14 taxa and 143-nt aligned positions (intron dataset 2; Table S1) and 1784-nt aligned positions (SSU rDNA dataset 2; Table S1). The NJ, maximum likelihood (ML), and minimal evolution (ME) bootstrap replicates ($\geq 50\%$) are given for each node. Isolates observed as horizontal inheritance candidates are shown in bold letters. Bold branches in the SSU rDNA tree indicate maximum support in NJ ($\geq 98\%$), ML, and ME ($\geq 98\%$). The scale bar indicates the fraction of substitutions per site.

3.4. Spliceosomal Introns Interrupt Homing Endonuclease Genes of S516 Group I Intron

The S516 HEGs are organized in sense orientation compared to the intron ribozymes and host SSU rDNA. Moreover, the HEGs are located only in the intron P2 segment (Figure 4A) and encode His-Cys homing endonucleases of 187 amino acids in *L. pseudomaculatum* (JQ031985) (Figure S5), 190 amino acids in *E. coelocephalum* (AY842033) (Figure S6), and 210 amino acids in *T. varia* (KM494993/4/5) (Figure S7). Furthermore, the S516 HEGs appear sporadically distributed among myxomycetes except for *T. varia*. Here, the three HEG-containing isolates cluster together on the host phylogeny with one *T. varia* isolate (KM494996) lacking a HEG insertion (Figure S2). Interestingly, all S516 HEGs are interrupted by small spliceosomal introns of either 58 bp, 59 bp, or 50 bp (Figure 4A). A nucleotide sequence alignment of the spliceosomal introns from four different SSU rDNA group I intron sites (S516, S943, S956, and S1389) show that the introns contain well-conserved splice sites and branch sites, which places them in the GT-AG class of spliceosomal introns (Figure 4B).

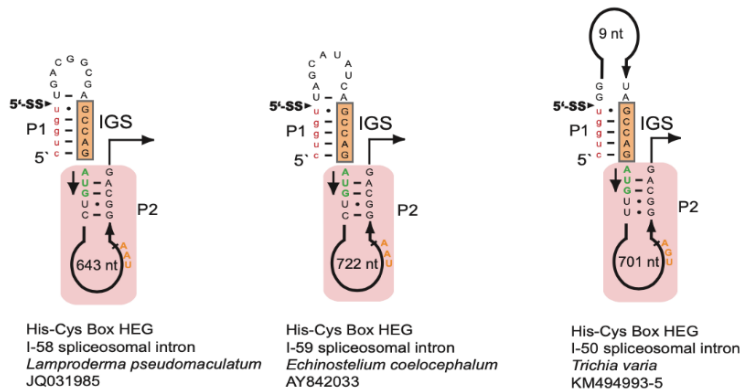
A hallmark of nuclear homing endonucleases is a His-Cys box motif consisting of two zinc coordinating structures (Zn-I and Zn-II) and two catalytic amino acid residues. Figure 4C shows that the His-Cys box motif in myxomycete S516 introns is highly conserved when compared to S516 homing endonucleases in *Porphyra* (red algae) [49], *Allovahlkampfia* (amoeba-flagellate) [39], and *Naegleria* (amoeba-flagellate) [50], as well as the structurally characterized I-Ppo I from the myxomycete *Physarum polycephalum* [51,52].

3.5. Direct Repeat Arrays in Peripheral Extensions of S516 Group I Introns

DR features were found in eleven S516 introns representing six myxomycete genera (*Elaeomyxa*, *Comatricha*, *Colloderma*, *Lamproderma*, *Licea*, and *Trichia*), and all except two (*Licea* and *Trichia*) classified to the myxomycete order Stemonitales (Table S1). According to the consensus diagram (Figure 2), DRs were located within six peripheral extension regions (P1, P2, P5, P6, P8, and P9), all of which do not structurally interfere with the catalytic core. A closer inspection of DR sequences revealed a high degree of sequence heterogeneity in sequence motif, sequence length, and copy number (Table S2). Sequence heterogeneity between individual motifs within the same array was common (Figure S8).

Some isolates showed unique and characteristic DR features (Figure S8): (i) *L. album* (JQ031971) harbors the largest motif (ca. 220 nt), which is located in P2 and repeated twice. (ii) *C. nigricapillitia* (AY643824) harbors the smallest motif (only 4 nt) in P2 and is repeated four times. (iii) *C. pseudoalpina* (DQ903673) harbors DRs in two different extension regions (P1 and P5). While the P1-DR motif is large in size (ca. 110 nt) and repeated four times, the P5-DR is composed of an approximately 60-nt motif and repeated twice. (iv) *L. columbinum* (HQ687204) also harbors DRs in two extension regions, but in P2 and P9. Two DR motifs are located in P2 (ca. 20 nt and ca. 40 nt), in 5 and 3 copies, respectively, and the DR motif in P9 is 28 nt in size and repeated twice. (v) *E. cerifera* (JQ031967) harbors three different DR motifs in P6 of 11 nt, 10 nt, and ca. 80 nt in sizes, repeated 4, 6, and 2 times, respectively. (vi) One each of two closely related *T. varia* introns contains (KM494995; 1145 bp) or lacks (KM494996; 482 bp) a HEG. Intriguingly, the shorter variant contains a DR, including a HEG remnant sequence, at the proposed HEG deletion site (Figure S9). (vii) Finally, *L. parasitica* (JX481297) harbors a complex array in P2. Specifically, the intron folds into a typical group IC1 intron structure (Figure 5A), and the large size (1149 bp) is mainly due to a 577-nt extension in P2. No recognizable open reading frame was found, but P2 harbors a DR array with 25 copies of a 23-nt motif (Figure 5B). Seven motif variants (named I to VII) are present, and repeated either once (III, V, VI, and VII), five times (II), six times (I), or ten times (IV). Interestingly, each motif apparently folds into an RNA hairpin structure where heterogenic positions are either located within the single-stranded loop or present as compensatory base pairs in the stem region (Figure 5C).

A



B

I-58 (S516)	caag	GU AUGGC	GGGCUGAGCUGAACCGAAAGAUCGGGCAUGA--	ACU AAC	CACUCGCGGCC	CAG	gacg
I-59 (S516)	aaag	GU AUGAU	CAUUUUGGGAUUGGGAUUAGGGUUAACUGUAU--	ACU GAU	GGAUUUACAUA	CAG	catg
I-50 (S516)	gagu	GU AGUGG	GGCGGUGGCUUUGCGAGGUGCGUGGU-----	ACU AAC	ACCCCGC--	CAG	gcug
I-44 (S1389)	caag	GU AGGCC	GUGUAGGCCGCCACCCCU-----	GCU AU	GCAUCAAA--	UAG	cgca
I-53 (S1389)	aaag	GU AUGUA	UGGGGUUGGGACUUGUUGGGCUUUUCUG----	ACU GAU	UCCUCCUG--	CAG	gaua
I-56 (S943)	ccau	GU AAGAU	UGCUCGGUGAACCUUUAUGAGUAUAAGUCUUA	ACU AAC	UGACUUGC--	UAG	gugu
I-50 (S956)	uaag	GU AGGUC	GGAGUUGGAGUUAACCCAUUUA-----	GUU AAC	CACCUCUAUA-	CAG	guua
I-51 (S956)	cuag	GU AUGUC	CUUGGAGUGCAAGGGUGGUUAUUU-----	GCU AAC	UUAGUUAUU-	CAG	gccc
		**		*		**	

C

	His-Cys Box																				
	I	I	I	I	II	II	II	II	II	II	II	II									
<i>Lamproderma_S516_HE</i>	R	C	K//S	H	I	C	GTRH	C	LKRT	H	LLLEPKWV	N	DERTH	C	H//A	C	P	H	EPR	C	F
<i>Echinostelium_S516_HE</i>	R	C	R//S	H	V	C	GTRN	C	CRRT	H	LVLEPKHI	N	DERVH	C	H//A	C	P	H	EPR	C	L
<i>Trichia_S516_HE</i>	P	C	E//S	H	L	C	GTHL	C	CSAD	H	LVLEPKWL	N	DERTC	C	H//G	C	P	H	TPQ	C	G
<i>Porphyra_S516_HE</i>	V	C	D//S	H	V	C	GTHH	C	LAAH	H	LMLEPKHV	N	DDRVH	C	H//H	C	P	H	LPR	C	Y
<i>Allovahkampfia_S516_HE</i>	C	C	E//S	H	L	C	GNGG	C	ARPG	H	IIIEPKTV	N	DERVA	C	H//L	C	P	H	KPK	C	F
<i>Naegleria_S516_HE</i>	V	C	K//S	H	L	C	GNGG	C	ARPG	H	LRIEKKS	V	DERTH	C	H//A	C	P	H	TPR	C	F
<i>Physarum_L1925_HE</i>	H	C	Y//S	H	L	C	HNTR	C	HNPL	H	LCWESLDD	N	KGRNW	C	P//G	C	V	H	AVV	C	L
	C41	H98	C100	C105	H110	N119	C125	C132	H134	C138											

Figure 4. His-Cys Box homing endonucleases encoded by S516 group I introns. **(A)** Schematic organization of the S516 intron substrate domain hosting homing endonuclease genes (HEGs) in P2. 5'SS, exon-intron splice site; IGS, internal guide sequence. The HEG start codons (AUG) and stop codons (UAA/UGA) are indicated. HEGs are interrupted by small spliceosomal introns (I-50, I-58, and I-59). **(B)** Sequence alignment of spliceosomal introns (uppercase letters) with some flanking HEG sequences (lowercase letters). Sequence motifs common to the mammalian spliceosomal intron consensus are indicated (green boxes), and invariable positions at the 5'SS (GU), branch site (A), and 3' splice site (AG) are indicated by * below the alignment. **(C)** Amino acid alignment of His-Cys Box features of S516 homing endonucleases (HE). The *Naegleria* and *Physarum* HEs represent well-studied and experimentally verified His-Cys homing endonucleases. Conserved residues (boxed) corresponding to those presented in the *Physarum* L1925 HE crystal structure [52]. C41, C100, C105, and H110 are involved in zinc-binding motif I. C125, C132, H134, and C138 are involved in zinc-binding motif II. H98 and N119 are associated with the active site.

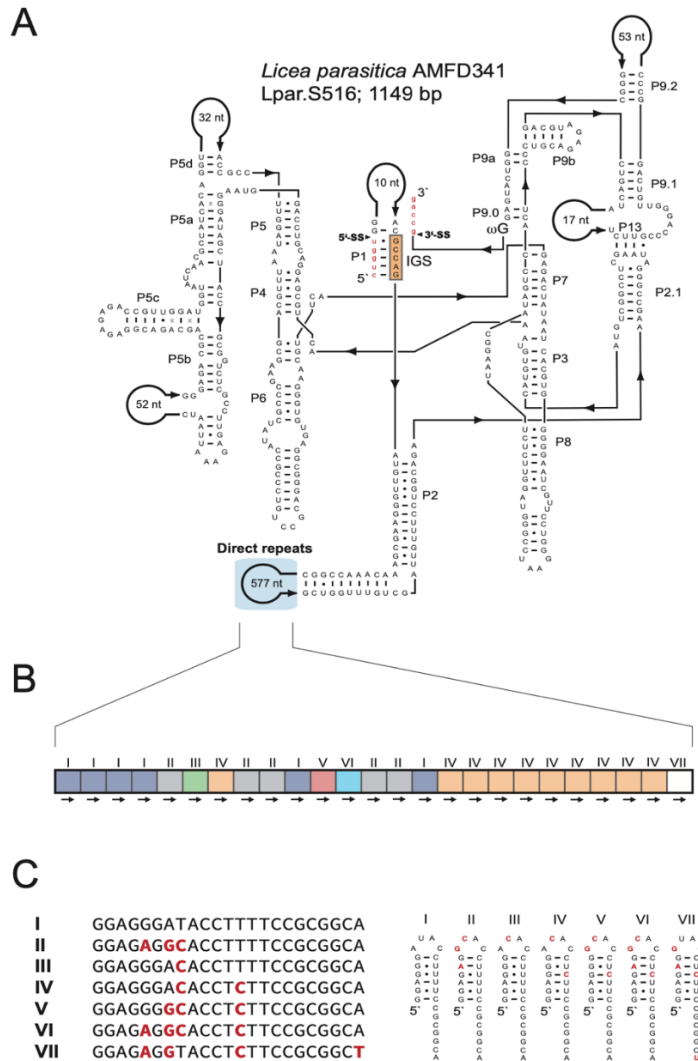


Figure 5. Direct repeat arrays in peripheral regions. **(A)** Secondary structural diagram of the S516 group I intron RNA in *Licea parasitica* containing direct repeats in the P2 peripheral region. For structural annotations, see legend to Figure 1. **(B)** Schematic organization of the direct repeats, which consists of 25 copies and seven motif variants (I to VII). **(C)** Alignment of sequence variants (left) and possible RNA hairpin structure formation (right).

4. Discussion

We report analyses of structural organization and molecular evolution of nucleolar group I intron at position S516 in myxomycetes. All 75 introns belong to the group IC1 subtype and fold at the RNA level into a highly conserved catalytic core, exemplified by the *D. alpinum* Fr-K12 intron. Extension sequences such as HEGs or DR arrays are all located in peripheral segments that do not interfere with the catalytic ribozyme core. HEGs located as insertions into segment P2 in five taxa are interrupted by small spliceosomal introns. Moreover, DR arrays were found in 11 taxa at six peripheral segments, and the *L. parasitica*

intron was found to harbor a highly structured DR motif in 25 copies. Intron evolution analyses indicate a complex inheritance pattern that apparently includes both vertical and horizontal transfers.

All HEGs in myxomycete S516 introns encode His-Cys homing endonucleases with a characteristic set of conserved amino acid residues involved in zinc coordination and catalysis. This His-Cys box was found to be highly conserved among S516 HEGs, also outside the myxomycete phylum. These include the *Porphyra* red algae [49] and the *Allovalhalkampfia* and *Naegleria* amoeba-flagellates [38,39], which support a common origin of S516 HEGs. The *Naegleria* homing endonucleases have been studied in more detail and found to recognize a 19 bp partially symmetrical sequence at the intron-less allele close to the S516 insertion site and to generate a pentanucleotide 3' overhang at the DNA cleavage site [50,53]. The myxomycete S516 HEGs were all found to be interrupted by small spliceosomal introns similar to those found in several other group I intron HEGs at SSU rDNA positions S943, S956, and S1389 [11,29–31]. The presence of spliceosomal introns in HEGs is proposed to facilitate the expression of protein-coding genes embedded in nuclear rDNA [25].

Extension sequences containing DR arrays were noted in six peripheral segments, including P2. The arrays varied from small insertions (4-nt motif in 4 copies) in *C. nigricapillitia* to 23-nt motif in 25 copies in *L. parasitica*. A common feature to most arrays was non-identical sequence repeat motifs, which argues against slippage-like mechanisms during replication [54,55] as a cause of repeat origin and maintenance. Then, what could be the biological role, if any, of the DR arrays in nuclear group I introns? DR arrays have been reported in several nuclear group I introns in myxomycetes, especially in introns that are unable to self-splice in vitro [7,9,11,33]. In *Fuligo septica* L1949 [7] and *M. crustacea* S1389 [11], intron DR arrays resulted in multiple alternative P1 segments (carrying the 5' splice site), suggesting that DRs potentially can interfere with intron splicing. Myxomycete S516 introns, however, appear to self-splice efficiently as naked RNA in vitro [6]. The highly structured hairpins with compensatory nucleotide changes in the *L. parasitica* DR array may suggest a selective structural feature at the RNA level, but no function has currently been assigned.

The S516 group IC1 intron in myxomycetes and other eukaryotic microorganisms constitute a phylogenetic distinct family with a common origin [6,56]. While introns within a short-term evolutionary time frame show strict vertical inheritance patterns [38], most introns indicate a complex evolutionary history not consistent with stable vertical transfers. Based on the idea of the Goddard–Burt cyclic model [57], we propose a scenario for the S516 group I introns that is funded on intron invasion, periodicity of HEG function, and intron loss. The scenario (Figure 6) is described by the four following stages. (1) A HEG-containing intron was gained at site S516 and effectively became spread in a population by homing mobility. Examples of mobile-type introns in myxomycetes are from *Lamproderma*, *Echinostelium*, and *Trichia*, but also from *Naegleria* [38], *Allovalhalkampfia* [39], and *Porphyra* [49]. (2) To inactivate homing endonuclease activity, and subsequent intron homing, frameshift, truncations, and sporadic deletions occurred in the HEG region. Such variants have been noted in the S516 introns of *Naegleria martinezi* [38], *Bangia fuscopurpurea* [58], and an *Acanthamoeba* sp. Isolate [49]. (3) The HEG insertion then became completely lost, resulting in an all-ribozyme organization as seen in *D. alpinum*, as well as 93% of the myxomycete S516 introns. HEG deletion was further supported in one of the *Trichia* S516 introns and suggests a link between deletion of HEG and generation of DR. (4) The complete intron became lost, probably due to homologous recombination with an intron-less allele resulting in SSU rDNA lacking an S516 intron, as noted in the approximately 80% of myxomycetes assessed.

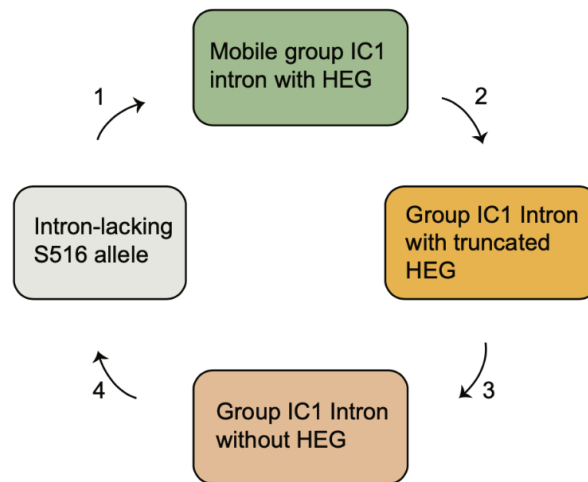


Figure 6. The evolutionary history of nucleolar S516 group I introns based on invasion and extinction.

5. Conclusions

The myxomycete protists represent a unique model system for studying the evolutionary history of self-splicing introns due to an exceptionally high content of nucleolar group I introns. We analyzed 75 group I introns at position 516 in the SSU rDNA of myxomycetes representing five distantly related taxonomic orders. The result reveals a conserved group IC1 ribozyme core, but highly variable peripheral RNA domains containing direct repeat arrays and homing endonuclease genes in 20% of the taxa. All HEGs were interrupted by small spliceosomal introns, which probably facilitate the expression of protein-coding genes embedded in nuclear rDNA. Phylogenetic analysis identified a complex inheritance pattern that can be explained by repeatedly intron gain and loss during the evolution. The current study expands our understanding of the structural organization of group IC1 introns and their evolution.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13060944/s1>, Figure S1: Sequence alignment of 143 core structure nucleotides of myxomycete S516 group I intron. Secondary structure paired segments (P1–P8) are specified with colors and their location is marked above the alignment. Intron sequences are indicated by GenBank accession numbers and family names. Note that DQ903678 is denoted as a *Diderma* sp. (*). Figure S2: Molecular phylogeny of myxomycete taxa-based SSU rDNA sequences. Molecular phylogeny of myxomycete taxa-based SSU rDNA sequences. The SSU-based topology is obtained by neighbor-joining (NJ) analysis of 67 taxa and 1524-nt aligned positions (SSU rDNA dataset 1). Some isolates are left out from the analysis due to partial SSU rDNA sequences. The predicted location of *Reticularia* and *Licea* isolates is shown in gray color letters. The tree is rooted with the *T. varia* SSU rDNA sequence. The NJ, maximum likelihood (ML) and maximum parsimony (MP) bootstrap replicates ($\geq 50\%$) are given at each node. Black dots at branch points; maximum support in ML, NJ, and ME ($\geq 97\%$). ¹ *Tricia varia* clade is composed of twenty *T. varia* isolates (KM495021-27, KM495029, KM495011-13, KM494999, KM495017, KM495005-9, KM495030, and KM495031). Homing endonuclease genes (HEGs) and direct repeats (DRs) are indicated in red and blue letters. The scale bar indicates the fraction of substitutions per site. Figure S3: Molecular phylogeny of myxomycete S516 group I introns. The intron topology is obtained by neighbor-joining (NJ) analysis of 75 taxa and 143-nt aligned positions (intron dataset 2; Figure S1). The NJ and minimal evolution (ME) bootstrap replicates ($\geq 50\%$) are given for each node. Bold lines indicate maximum support in NJ and ME ($\geq 98\%$). ¹ *Tricia varia* clade is composed of twenty *T. varia* isolates (KM495021-27, KM495029, KM495011-13, KM494999, KM495017, KM495005-9, KM495030, and KM495031). Homing endonuclease genes (HEGs) and direct repeats (DRs) are indicated in red and blue letters. The scale bar indicates

the fraction of substitutions per site. Figure S4: Two closely related *Meriderma* species contain distantly related S516 group I introns. (a) Secondary structure diagram of *M. aggregarum* S516 intron RNA. (b) Secondary structure diagram of *M. cribrarioides* S516 intron RNA. Peripheral regions with significant structural differences are boxed. Figure S5: (a) Structural features of homing endonuclease genes (HEG)-containing S516 group I intron in *Lamproderma pseudomaculatum*. (b) Nucleotide sequence of HEG, including the 58-nt spliceosomal intron (red letters). Start (ATG) and stop (TAA) codons are indicated, as well as the polyadenylation signal (AATAAA). (c) Amino acid sequence of the homing endonuclease. Important residues involved in Zn binding (red) and catalysis (blue) are indicated. Figure S6: (a) Structural features of homing endonuclease genes (HEG)-containing S516 group I intron in *Echinostelium coelocephalum*. (b) Nucleotide sequence of HEG, including the 59-nt spliceosomal intron (red letters). Start (ATG) and stop (TAA) codons are indicated, as well as the polyadenylation signal (AATAAA). (c) Amino acid sequence of the homing endonuclease. Important residues involved in Zn binding (red) and catalysis (blue) are indicated. Figure S7: (a) Structural features of homing endonuclease genes (HEG)-containing S516 group I intron in *Trichia varia*. (KM494993-5). (b) Nucleotide sequence of HEG, including the 50-nt spliceosomal intron (red letters). Start (ATG) and stop (TGA) codons are indicated, as well as the polyadenylation signal (AATAAA). (c) Amino acid sequence of the homing endonuclease. Important residues involved in Zn binding (red) and catalysis (blue) are indicated. Figure S8: Sequence alignment of direct repeat motifs in myxomycete S516 group I intron. Direct repeats are located in peripheral regions P1, P2, P5, P6, P8, and P9. Figure S9: RNA structural support of HEG deletion in *Trichia varia* S516 group I introns. HEG deletion in segment P2 (KM494996) leaves a HEG remnant sequence corresponding to the 5' end of the open reading frame (indicated by green colored nucleotides), as well as a direct repeat feature (DR-1 and DR-2). Only paired segments P1 and P2 are shown. Table S1: Key features of 75 nucleolar group I introns at position S516 in myxomycete. Table S2: Schematic organization of direct repeats arrays in S516 group I introns in myxomycetes.

Author Contributions: B.M.N.F., B.O.K., I.B., P.H. and S.D.J. designed the research; S.D.J. performed the sequence structure analysis and B.M.N.F. performed phylogenetic analysis together with P.H. and S.D.J. Additional data analyses were performed by B.M.N.F. and S.D.J. B.M.N.F. wrote the manuscript together with P.H., B.O.K. and I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: New sequencing data are available in GenBank under Accession Number ON155994 (*Didyma alpinum* Fr-K12).

Acknowledgments: We thank Kari Haugli for technical support in collecting and culturing myxomycetes, as well as DNA isolation and Sanger sequencing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haugen, P.; Simon, D.M.; Bhattacharya, D. The natural history of group I introns. *Trends Genet.* **2005**, *21*, 111–119. [[CrossRef](#)] [[PubMed](#)]
2. Michel, F.; Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **1990**, *216*, 585–610. [[CrossRef](#)]
3. Hedberg, A.; Johansen, S.D. Nuclear group I introns in self-splicing and beyond. *Mobile DNA* **2013**, *4*, 17. [[CrossRef](#)] [[PubMed](#)]
4. Jackson, S.A.; Cannone, J.J.; Lee, J.C.; Gutell, R.R.; Woodson, S.A. Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J. Biol. Mol.* **2002**, *323*, 35–52. [[CrossRef](#)]
5. Decatur, W.A.; Fournier, M.J. rRNA modifications and ribosome function. *Trends Biochem. Sci.* **2002**, *27*, 344–351. [[CrossRef](#)]
6. Haugen, P.; Coucheron, D.H.; Rønning, S.B.; Haugli, K.; Johansen, S. The molecular evolution and structural organization of self-splicing group I introns at position 516 in nuclear SSU rDNA of myxomycetes. *J. Eukaryot. Microbiol.* **2003**, *50*, 283–292. [[CrossRef](#)]
7. Lundblad, E.W.; 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*, 1283–1293. [[CrossRef](#)]

8. Wikmark, O.G.; Haugen, P.; Lundblad, E.W.; Haugli, K.; Johansen, S.D. The molecular evolution and structural organization of group I introns at position 1389 in nuclear small subunit rDNA of myxomycetes. *J. Euk. Microbiol.* **2007**, *54*, 49–56. [[CrossRef](#)]
9. Wikmark, O.G.; Haugen, P.; Haugli, K.; Johansen, S.D. Obligatory group I introns with unusual features at positions 1949 and 2449 in nuclear LSU rDNA of Didymiaceae myxomycetes. *Mol. Phylogenet. Evol.* **2007**, *43*, 596–604. [[CrossRef](#)]
10. Reeb, V.; Haugen, P.; Bhattacharya, D.; Lutzoni, F. Evolution of *Pleopsidium* (lichenized Ascomycota) S943 group I introns and the phylogeography of an intron-encoded putative homing endonuclease. *J. Mol. Evol.* **2007**, *64*, 285–298. [[CrossRef](#)]
11. Furulund, B.M.N.; Karlsen, B.O.; Babiak, I.; Johansen, S.D. A phylogenetic approach to structural variation in organization of nuclear group I introns and their ribozymes. *Non-coding RNA* **2021**, *7*, 43. [[CrossRef](#)] [[PubMed](#)]
12. Johansen, S.; Haugen, P. A new nomenclature of group I introns in ribosomal DNA. *RNA* **2001**, *7*, 935–936. [[CrossRef](#)] [[PubMed](#)]
13. Cech, T.R. Self-splicing of group I introns. *Ann. Rev. Biochem.* **1990**, *59*, 543–568. [[CrossRef](#)] [[PubMed](#)]
14. Nielsen, H.; Fiskaa, T.; Birgisdottir, Á.B.; Haugen, P.; Einvik, C.; Johansen, S. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* **2003**, *9*, 1464–1475. [[CrossRef](#)]
15. Cech, T.R.; Damberger, S.H.; Gutell, R.R. Representation of the secondary and tertiary structure of group I introns. *Nature Struct. Biol.* **1994**, *1*, 273–280. [[CrossRef](#)]
16. Vicens, Q.; Cech, T.R. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **2006**, *31*, 41–51. [[CrossRef](#)]
17. Guo, F.; Gooding, A.R.; Cech, T.R. Structure of the *Tetrahymena* ribozyme: Base triple sandwich and metal ion at the active site. *Mol. Cell* **2004**, *16*, 351–362. [[CrossRef](#)]
18. Andersen, K.L.; Beckert, B.; Masquida, B.; Johansen, S.D.; Nielsen, H. Accumulation of stable full-length circular group I intron RNAs during heat-shock. *Molecules* **2016**, *21*, 1451. [[CrossRef](#)]
19. Su, Z.; Zhang, K.; Kappel, K.; Li, S.; Palo, M.Z.; Pintilie, G.D.; Rangan, R.; Luo, B.; Wei, Y.; Das, R.; et al. Cryo-EM structures of full-length *Tetrahymena* ribozyme at 3.1 Å resolution. *Nature* **2021**, *596*, 603–607. [[CrossRef](#)]
20. Muscarella, D.E.; Vogt, V.M. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell* **1989**, *56*, 443–454. [[CrossRef](#)]
21. Johansen, S.; Elde, M.; Vader, A.; Haugen, P.; Haugli, K.; Haugli, F. In vivo mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. *Mol. Microbiol.* **1997**, *24*, 737–745. [[CrossRef](#)] [[PubMed](#)]
22. Belfort, M.; Roberts, R.J. Homing endonucleases: Keeping the house in order. *Nucleic Acids Res.* **1997**, *25*, 3379–3388. [[CrossRef](#)] [[PubMed](#)]
23. Johansen, S.; Embley, T.M.; Willassen, N.P. A family of nuclear homing endonucleases. *Nucleic Acids Res.* **1993**, *21*, 4405. [[CrossRef](#)] [[PubMed](#)]
24. Hafez, M.; Hausner, G. Homing endonucleases: DNA scissors on a mission. *Genome* **2012**, *55*, 553–569. [[CrossRef](#)]
25. Johansen, S.D.; Haugen, P.; Nielsen, H. Expression of protein-coding genes embedded in ribosomal DNA. *Biol. Chem.* **2007**, *388*, 679–686. [[CrossRef](#)]
26. 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*, 1584–1587. [[CrossRef](#)]
27. Meyer, M.; Nielsen, H.; Olieric, V.; Roblin, P.; Johansen, S.D.; Westhof, E.; Masquida, B. Speciation of a group I intron into a lariat capping ribozyme. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7659–7664. [[CrossRef](#)]
28. Nielsen, H.; Krogh, N.; Masquida, B.; Johansen, S.D. The lariat capping ribozyme. In *Ribozymes*; Muller, S., Masquida, B., Winkler, W., Eds.; WILEY-VCH GmbH: Weinheim, Germany, 2021; pp. 118–142; ISBN 978-3-527-34454-3.
29. Vader, A.; Nielsen, H.; Johansen, S. In vivo expression of the nucleolar group I intron-encoded *I-DiI* homing endonuclease involves the removal of a spliceosomal intron. *EMBO J.* **1999**, *18*, 1003–1013. [[CrossRef](#)]
30. Johansen, S.D.; Vader, A.; Sjøttem, E.; Nielsen, H. In vivo expression of a group I intron HEG from the antisense strand of *Didymium* ribosomal DNA. *RNA Biol.* **2006**, *3*, 157–162. [[CrossRef](#)]
31. Haugen, P.; Wikmark, O.G.; Vader, A.; Coucheron, D.; Sjøttem, E.; Johansen, S.D. The recent transfer of a homing endonuclease gene. *Nucleic Acids Res.* **2005**, *33*, 2734–2741. [[CrossRef](#)]
32. Vader, A.; Naess, J.; Haugli, K.; Haugli, F.; Johansen, S. Nucleolar introns from *Physarum flavicomum* contain insertion elements that may explain how mobile group I introns gained their open reading frames. *Nucleic Acids Res.* **1994**, *22*, 4553–4559. [[CrossRef](#)] [[PubMed](#)]
33. Nandipati, S.C.; Haugli, K.; Coucheron, D.H.; Haskins, E.F.; Johansen, S.D. Polyphyletic origin of the genus *Physarum* (Physarales, Myxomycetes) revealed by nuclear rDNA mini-chromosome analysis and group I intron synapomorphy. *BMC Evol. Biol.* **2012**, *12*, 166. [[CrossRef](#)]
34. Noller, H.F. Biochemical characterization of the ribosomal decoding site. *Biochimie* **2006**, *88*, 935–941. [[CrossRef](#)]
35. Oliveira, M.C.; Ragan, M.A. Variant forms of a group I intron in nuclear small-subunit rRNA genes of the marine red alga *Porphyra spiralis* var. *amplifolia*. *Mol. Biol. Evol.* **1994**, *11*, 195–207. [[CrossRef](#)]
36. Haugen, P.; Huss, V.A.; Nielsen, H.; Johansen, S. Complex group-I introns in nuclear SSU rDNA of red and green algae: Evidence of homing-endonuclease pseudogenes in the Bangiophyceae. *Curr. Genet.* **1999**, *36*, 345–353. [[CrossRef](#)] [[PubMed](#)]
37. Einvik, C.; Decatur, W.A.; Embley, T.M.; Vogt, V.M.; Johansen, S. *Naegleria* nucleolar introns contain two group I ribozymes with different functions in RNA splicing and processing. *RNA* **1997**, *2*, 710–720.
38. Wikmark, O.G.; Einvik, C.; De Jonckheere, J.F.; Johansen, S.D. Short-term sequence evolution and vertical inheritance of the *Naegleria* twin-ribozyme group I intron. *BMC Evol. Biol.* **2006**, *6*, 39. [[CrossRef](#)]

39. Tang, Y.; Nielsen, H.; Masquida, B.; Gardner, P.P.; Johansen, S.D. Molecular characterization of a new member of the lariat capping twin-ribozyme introns. *Mobile DNA* **2014**, *5*, 25. [[CrossRef](#)]
40. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
41. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvement in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
42. Johansen, T.; Johansen, S.; Haugli, F.B. Nucleotide sequence of the *Physarum polycephalum* small subunit ribosomal RNA as inferred from the gene sequence: Secondary structure and evolutionary implications. *Curr. Genet.* **1988**, *14*, 265–273. [[CrossRef](#)]
43. Cainelli, R.; de Haan, M.; Meyer, M.; Bonkowski, M.; Fiore-Donno, A.M. Phylogeny of Physarida (Amoebozoa, Myxogastria) based on the small-subunit ribosomal RNA gene, redefinition of *Physarum pusillum* s. str. and reinstatement of *P. gravidum* Morgan. *J. Euk. Microbiol.* **2020**, *67*, 327–336. [[CrossRef](#)] [[PubMed](#)]
44. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
45. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [[CrossRef](#)]
46. Lehnert, V.; Jaeger, L.; Michel, F.; Westhof, E. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: A complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem. Biol.* **1996**, *3*, 993–1009. [[CrossRef](#)]
47. Golden, B.L.; Gooding, A.R.; Podell, E.R.; Cech, T.R. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **1998**, *282*, 259–264. [[CrossRef](#)] [[PubMed](#)]
48. Fiore-Donno, A.M.; Kamono, A.; Meyer, M.; Schnittler, M.; Fukui, M.; Cavalier-Smith, T. 18S rDNA phylogeny of *Lamproderma* and allied genera (Stemonirales, Myxomycetes, Amoebozoa). *PLoS ONE* **2012**, *7*, e35359. [[CrossRef](#)]
49. Haugen, P.; De Jonckheere, J.F.; Johansen, S. Characterization of the self-splicing products of two complex *Naegleria* LSU rDNA group I introns containing homing endonuclease genes. *Eur. J. Biochem.* **2002**, *269*, 1641–1649. [[CrossRef](#)]
50. Elde, M.; Willassen, N.P.; Johansen, S. Functional characterization of isoschizomeric His-Cys box homing endonucleases from *Naegleria*. *Eur. J. Biochem.* **2000**, *267*, 7257–7266. [[CrossRef](#)]
51. Muscarella, D.E.; Vogt, V.M. A mobile group I intron from *Physarum polycephalum* can insert itself and induce point mutations in the nuclear ribosomal DNA of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **1993**, *13*, 1023–1033. [[CrossRef](#)]
52. Flick, K.E.; Jurica, M.S.; Monnat, R.J.J.; Stoddard, B.L. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature* **1998**, *394*, 96–101. [[CrossRef](#)] [[PubMed](#)]
53. Elde, M.; Haugen, P.; Willassen, N.P.; Johansen, S. I-NjaI, a nuclear intron-encoded homing endonuclease from *Naegleria*, generates a pentanucleotide 3' cleavage-overhang within a 19 base-pair partially symmetric DNA recognition site. *Eur. J. Biochem.* **1999**, *259*, 281–288. [[CrossRef](#)] [[PubMed](#)]
54. Levinson, G.; Gutman, G.A. Slipped-strand mispairing: A major mechanism for DNA sequencing evolution. *Mol. Biol. Evol.* **1987**, *4*, 203–221. [[CrossRef](#)] [[PubMed](#)]
55. Hancock, J.M.; Dover, G.A. Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. *Mol. Biol. Evol.* **1988**, *5*, 377–391. [[CrossRef](#)]
56. Corsaro, D.; Kohler, M.; Venditti, D.; Rott, M.B.; Walochnik, J. Recovery of an *Acanthamoeba* strain with two group I introns in the nuclear 18S rRNA gene. *Eur. J. Protistol.* **2019**, *68*, 88–98. [[CrossRef](#)]
57. Goddard, M.R.; Burt, A. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13880–13885. [[CrossRef](#)]
58. Müller, K.M.; Cannone, J.J.; Gutell, R.R.; Sheath, R.G. A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta). *Mol Biol. Evol.* **2001**, *18*, 1654–1667. [[CrossRef](#)]

Supplementary materials

Figure S1. Sequence alignment of core structure nucleotides of myxomycete S516 introns

INTRON

	<IGS>	<<P3>>	<P4>	<P5>	<<P5a>>>		
DaIp. S516_ON155994	CTGGT	TAGAGCCAG	GTACAGATG	TAAATTCGAGG	AACTCTGGAGT	CATAGCTACCGGT	AATAACGGTAT
Dniv. S516-1_HE614615	CTGGT	GGACGCCAG	GTACAGATG	TAAATTCGAGG	GACGCTCTG	AGTCAAGCAATACCGGT	AATAACGCTGT
Dniv. S516-2_HE614616	CTGGT	GGACGCCAG	GTACAGATG	TAAATTCGAGG	GACGCTCTG	AGTCAAGCAATACCGGT	AATAACGCTGT
Dniv. S516-3_HE614617	CTGGT	GGACGCCAG	GTACAGATG	TAAATTCGAGG	GACGCTCTG	AGTCAAGCAATACCGGT	AATAACGCTGT
Dsou. S516_MZ313549	CTGGT	TGTAGCCAG	GTACGACGT	TAAACTG	CAGGAACTCTG	AGCCTCAAGTACCGGT	AACAACGGCTA
Lcru. S516_HE614619	CTGGT	TGAAAGCCAG	GTACAGCCG	TAAAGTTCG	GGGGAAACCGCA	TGGGGGGGATACCGGT	AACAACGGCCC
Dsp. S516*_DQ903678	CTGGT	GGAAAGCCAG	GTACAGCTG	TAAACTG	CAGGAACTCTG	AGTCAAGCAATACCGGT	AATAACGTTAT
Bgra. S516_AJ555451	CTGGT	TATAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGATTGGATACCGGT	AATAACGTTCA
Flev. S516_DQ903676	CTGGT	TACAGCCAG	GTACAGCCG	TAAATTCG	GGGGAAACCGCA	TGAGTGGGTACCGGT	AACAATGCTCA
Fsep. S516-1_AJ584697	CTGGT	AAATGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGAAAGCTACCGGT	AACAATGGCAT
Fsep. S516-2_MF983561	CTGGT	GTATGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGAAAGCTACCGGT	AACAATGGCAT
Fsep. S516-3_AJ555452	CTGGT	AAATGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGAAAGCTACCGGT	AACAATGGCAT
Ecer. S516_JQ031967	CTGGT	TTTAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Pfla. S516_AJ555453/4	CTGGT	TGAAAGCCAG	GTACAGCGT	TAAACTG	CAGGAACTCTG	GGGATTCATGACCGGT	AAAACCATGGA
Pnot. S516_JX035988	CTGGT	GGATGCCAG	GTACATTCTG	TAAATTCG	GGGGAAACCGCA	TGGGAGTGCCACCGGT	AACAGCGCACT
Psp. S516-1_JX035984	CTGGT	GGCAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGACTGGATACCGGT	AACAACGTTCA
Psp. S516-2_MK336180	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Cana. S516_JQ031961	CTGGT	TGAAAGCCAG	GTACAGCTT	TAAACTG	CAGGAAACCGCA	TGGTGAAGCGGATACCGGT	TATAACGGCCC
Cnig. S516-1_DQ903685	CTGGT	GGACGCCAG	GTACAGACG	TAAATTCG	CAGGAACTCTG	TGGTGTGCACTACCGGT	AATAACGTTCCG
Cnig. S516-2_AY643824	CTGGT	GGACGCCAG	GTACACTGT	TAAATTCG	GGGGAAACCGCA	TGAGTGGGTACCGGT	AAAAACCGCCG
Cpse. S516_DQ903673	CTGGT	TGTTGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGAGCGGGATACCGGT	TATAACGTTCC
Cocu. S516_JQ031959	CTGGT	TAAAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATAGACCGGT	AACAGCGTCTA
Crob. S516_JQ031960	CTGGT	TAAAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATAGACCGGT	AACAGCGTCTA
Dpau. S516_JQ031966	CTGGT	TAGAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Laca. S516_JQ031968	CTGGT	TAAAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Laen. S516-1_JQ031969	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Laen. S516-2_JQ031970	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
La1b. S516_JQ031971	CTGGT	TGAGGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lco1. S516-1_HQ687204	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lco1. S516-2_HQ687196	CTGGT	CCGGGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lco1. S516-3_HQ687198	CTGGT	GAAAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lco1. S516-4_HQ687199	CTGGT	ACTTCTGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lco1. S516-5_HQ687197	CTGGT	TATAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lco1. S516-6_HQ687200	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lech. S516-1_JQ031980	CTGGT	GAGGCCAG	GTACAGATG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lech. S516-2_JQ031979	CTGGT	TGAGGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lpse. S516_JQ031985	CTGGT	GTAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lpn. S516_HQ687194	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lsp. S516_JQ031994	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Mobl. S516_DQ903682	CTGGT	AGTGGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Magg. S516_DQ903669	CTGGT	TGAGGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Mcr1. S516_JQ032000	CTGGT	GTAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Sfius. S516-1_MK041079	CTGGT	CAAGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Sfius. S516-2_MK041078	CTGGT	GGACGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Lpar. S516_JX481297	CTGGT	GGACGCCAG	GTACACTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Rlyc. S516_JX481311	CTGGT	TGCAGCCAG	GTACGCTCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Agto. S516_JX481282	CTGGT	TACAGCCAG	GTACAGTGT	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tdec. S516_FJ810499	CTGGT	TATAGCCAG	GTACAATCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-1_KM495030	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-2_KM495011	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-3_KM495025	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-4_KM495027	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-5_KM495026	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-6_KM495022	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-7_KM495010	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-8_KM495009	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-9_KM494999	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-10_KM495017	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-11_KM495008	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-12_KM494996	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-13_KM495029	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-14_KM495023	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-15_KM495012	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-16_KM495024	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-17_KM495021	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-18_KM494993	CTGGT	GGTAGCCAG	GTACAACCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-19_KM495013	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-20_KM495007	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-21_KM495006	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-22_KM495005	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-23_KM495031	CTGGT	TGTAGCCAG	GTACAGCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-24_KM494995	CTGGT	GGTAGCCAG	GTACAACCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Tvar. S516-25_KM494994	CTGGT	GTAGCCAG	GTACAACCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Dniv. S516_JX481289	CTGGT	CACAGCCAG	CTCAACCGG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA
Ecoe. S516_AY842033	CTGGT	TACAGCCAG	GTACAATCTG	TAAATTCG	GGGGAAACCGCA	TGGGATCAGGACATG	AACTGTTGA

INTRON

Dalp. S516 **GGCA****CCAGACA**CAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Dniv. S516-1 **GGCA****CCAGATA**TAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CATCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Dniv. S516-2 **GGCA****CCAGATA**TAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CATCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Dniv. S516-3 **GGCA****CCAGATA**TAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CATCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Dsou. S516 **TGGC****CCAGACA**CAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCCGGGGCGCT**CAAGGAA**TAGTCC**ATT**AGGCCAG**
 Lcru. S516 **TTCA****CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Dsp. S516* **GGCA****CCAGATA**TAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CATCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Bgra. S516 **CTCA****CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Flev. S516 **CTCA****CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Fsep. S516-1 **CCAG****CTGGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Fsep. S516-2 **CCCA****CTGGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Fsep. S516-3 **CCAG****CTGGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Ecer. S516 **TTCC****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGAGTGGGGTCT**TAAGTAT**AGTCC**ATT**AGGCCAG**
 Pfla. S516 **TTCC****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTGTGGGGTCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Pnot. S516 **CTCA****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGAATGGGGCCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Psp. S516-1 **CTCA****CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAACTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Psp. S516-2 **GTC****CTAGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Cana. S516 **TTCA****CCAGIT**TTAT**TTCCG**CCGGTT**CAGAGACTA**CAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Cnig. S516-1 **TTCC****CCAGACA**CAAT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCGCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Cnig. S516-2 **TTCA****CTAGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCGCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Cpse. S516 **TTCC****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**CAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Cocu. S516 **TTCC****CCAGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CATGTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Crob. S516 **GTC****CTAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCGCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Dpau. S516 **TC**TT**CTGATC**GAAT**CCGCA**AGCCGGTT**CAGAGACTA**CAAT**CGACTGGGGCCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Laca. S516 **ACTC****CTGGGCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**CAAT**CGACTGGGGTAT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Laen. S516-1 **CC**TT**CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGTCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Laen. S516-2 **CC**TT**CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGTCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 La1b. S516 **CC**TT**CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGTCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Lcol. S516-1 **CTC****CTGGIT**TAAT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGCCT**TAAGGCA**TAGTCC**ATT**AGGCCAG**
 Lcol. S516-2 **TC**CC**CCAGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**TCGACTGGGGCCT**TAAGGCA**TAGTCC**ATT**AGGCCAG**
 Lcol. S516-3 **CTCC****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**TCGACTGGGGCCT**TAAGGCA**TAGTCC**ATT**AGGCCAG**
 Lcol. S516-4 **TTCC****CCAGGCA**CA**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**TCGGGTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Lcol. S516-5 **CTCC****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAACTGGGGTCT**TAAGGCA**TAGTCC**ATT**AGGCCAG**
 Lcol. S516-6 **TC**CT**CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAACTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Lech. S516-1 **CACA****CTAGATA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CATCTGGGGCCT**TAAGGAA**TAGTCC**AA**TTAGGCCAG**
 Lech. S516-2 **CACA****CTAGACA**CAAT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CATCTGGGGCCT**TAAGGAA**TAGTCC**AA**TTAGGCCAG**
 Lpse. S516 **TC**CC**CTGGTCA**AT**CCGCA**A**CCGGTT****CAGAGACTA**CAAT**CGACTGGGGCCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Lpun. S516 **TC**CT**CCAGCCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CGAGTGGGGTCT**TAAGTAT**AGTCC**ATT**AGGCCAG**
 Lsp. S516 **CC**TT**CTAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAACTGGGGTCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Mob1. S516 **CGGA****CCAGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Magg. S516 **GCTC****CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**CAAT**CGACTGGGGCCT**TAAGGTAT**AGTCC**ATT**AGGCCAG**
 Mcr1. S516 **CC**TC**CAAGGCA**CA**CTGCA**AGCCGGTT**CAGAGACTA**CAAT**CGACTGGGGCCT**TAAGGCA**TAGTCC**AA**TTAGGCCAG**
 Sfus. S516-1 **AT**TC**CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Sfus. S516-2 **TTCC****CTGGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Lpar. S516 **TC**AC**CTGGATA**AT**TTTCCG**CCGGTT**CAGAGACTA**TAAT**TCGACTGGGGTCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Rlyc. S516 **CACA****CCATTT**TAAT**CCGCA**AGCCGGTT**CAGAGACTA**TACA**CCAGCCGGGGCCT**TAAGTAT**AGTCC**ATT**AGGCCAG**
 Ag1o. S516 **TCCA****CTGGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tdec. S516 **GCTC****CTAGCCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-1 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-2 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-3 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-4 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-5 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-6 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-7 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-8 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-9 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-10 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-11 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-12 **TT**TC**CTGGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-13 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-14 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-15 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-16 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-17 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-18 **TT**TC**CTGGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-19 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-20 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-21 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-22 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-23 **ACTC****CTAGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CAGCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-24 **TT**TC**CTGGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Tvar. S516-25 **TT**TC**CTGGTCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGTCTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Dniv. S516 **GGCA****CTAGCCA**AT**CTGCA**AGCCGGTT**CAGAGACTA**TAAT**CGACTGGGGCCT**TAAGGAA**TAGTCC**ATT**AGGCCAG**
 Ecoe. S516 **CT**TC**CTGGTCA**AT**CCGCA**AGCCGGTT**CAGAGACTA**TAAT**CAATGGGGCCT**TAAGTAT**AGTCC**ATT**AGGCCAG**

Molecular phylogeny of myxomycete taxa based on SSU rDNA sequences

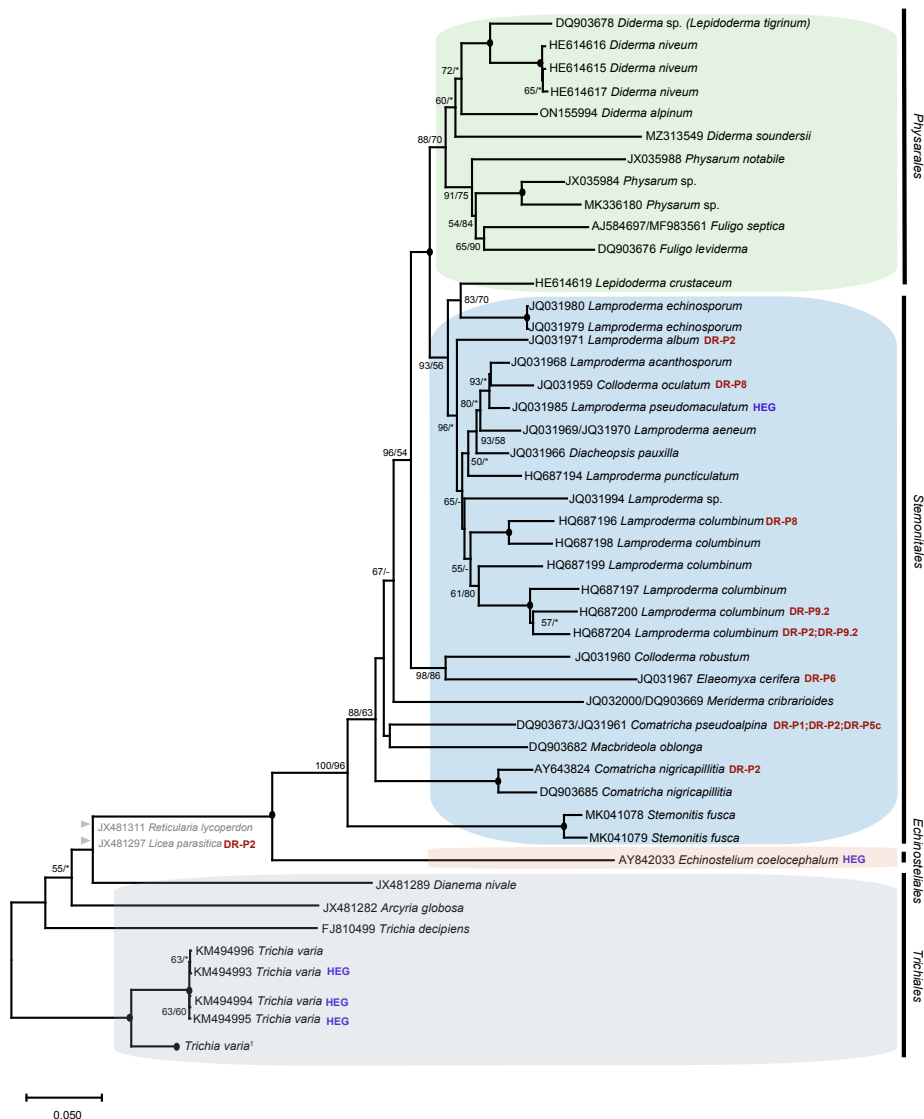


Figure S2

Molecular phylogeny of myxomycete S516 group I introns

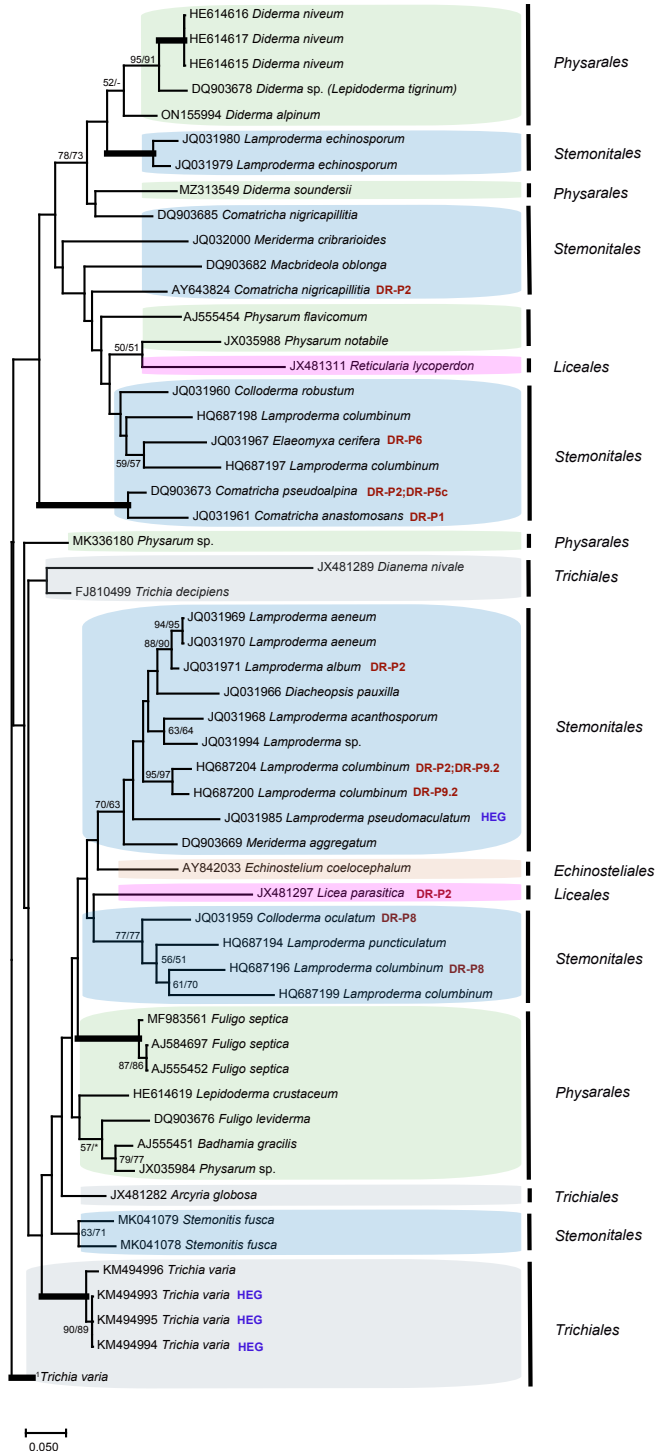


Figure S3

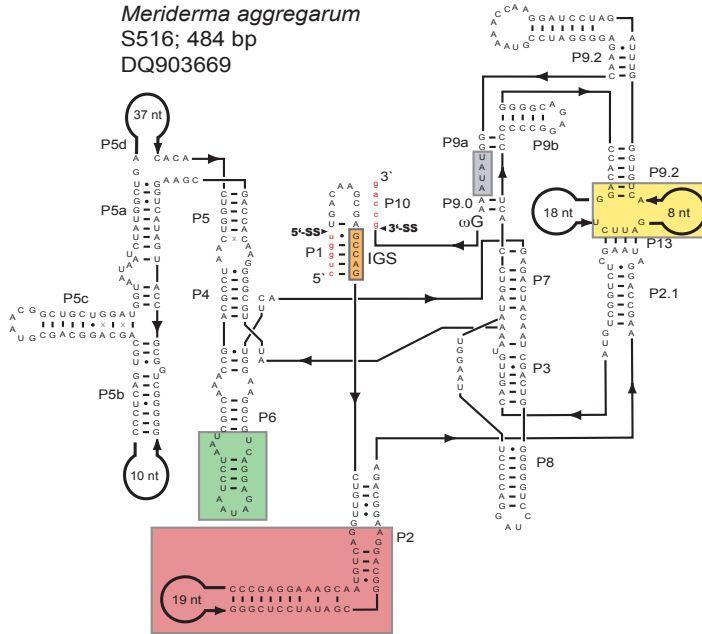
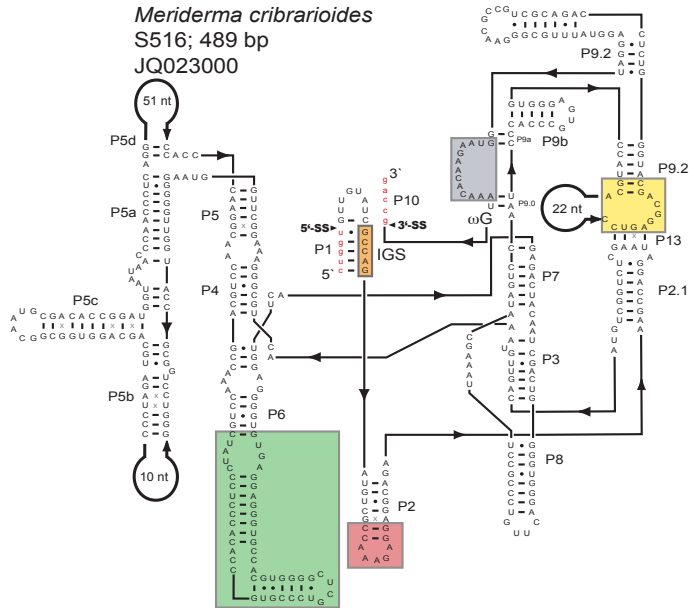
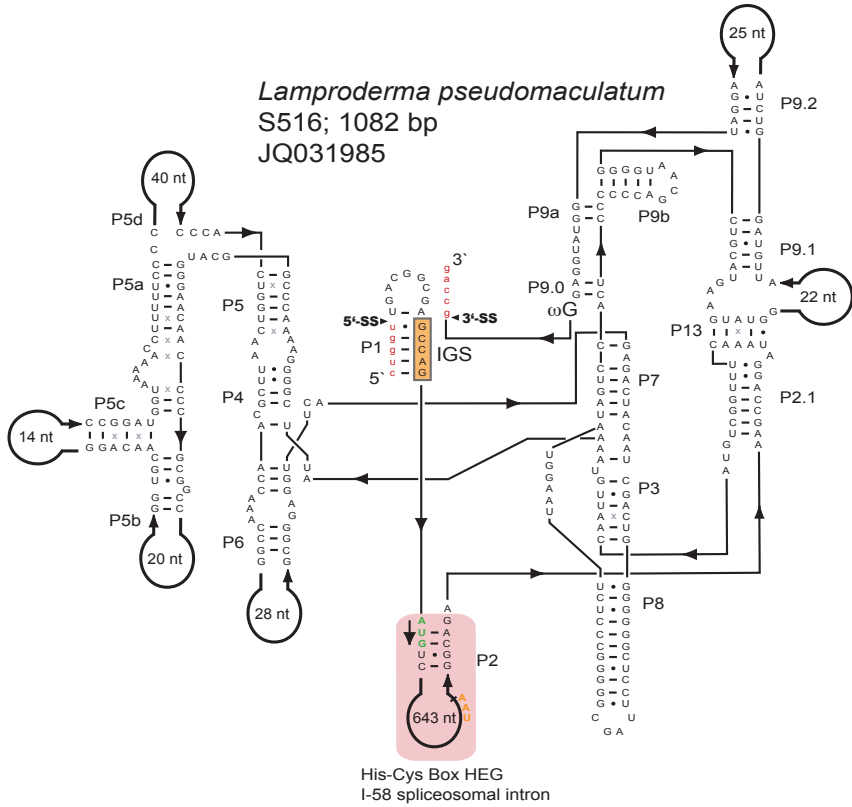
A**B**

Figure S4

A



B

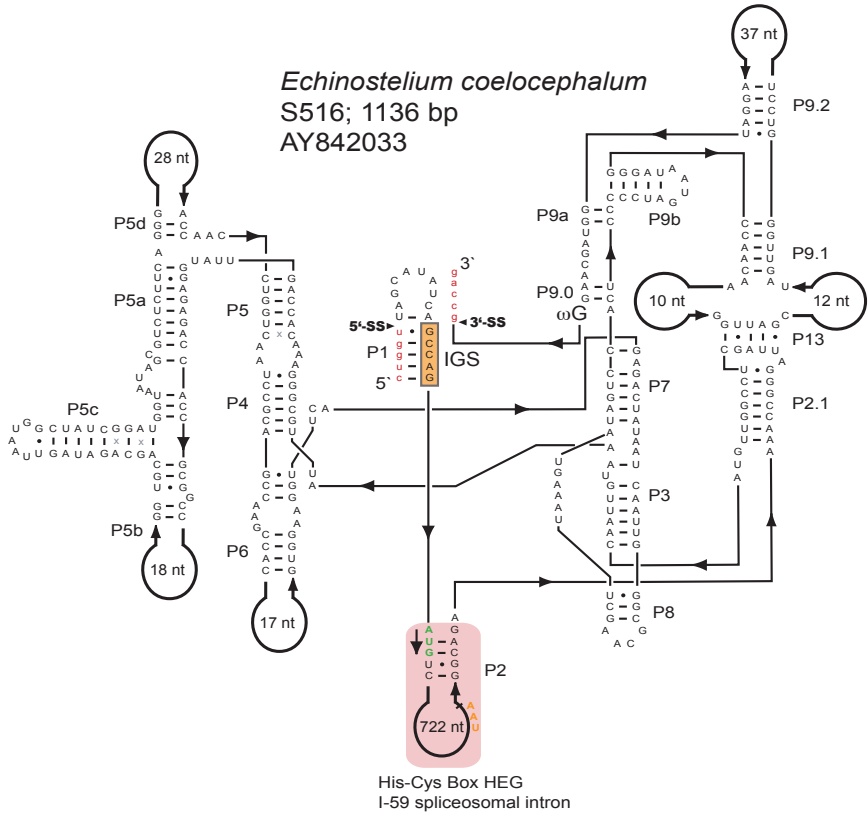
ATGTCGAACAATGAGGGAAACAACCGAACGTCGAAAGCACAAACAGTCTGGGCTGACCACGCTGAGCGTACTGAAG
 TCGAGTGATGAAATGATCCTGGCGCTAGAGCCTCTCCTAAAAAGGAAGGCACGGGCAATAGGCTTCGAAGTAGTC
 AAGCGATCCAGGTGCAAAGTATGGCCGAACAAAAGTCCAAGATCACC AAGGGCAAG**G TATGGCGGGCTGAGCTG**
AACCGAAAGATCGGGCATGAACTAACCACCTCGCTGCCAGGACGTCGGGGAGGCCTATCATGTTAGGGCATTTC
 AGTATTCGGCCGGAACGTTTGGGGCTGGTACCCCGAAC AAGCGCCAAGGCACCTCGTGGTATCGCATATATG
 CGGACCCGTCACCTGCCTGAAGCGTACCCACCTGCTGCTGGAGCCAAAGTGGGTGAATGACGAACGAACCCACTG
 CCACCTTTCGCCACATTCATGGCGGGCAGGGGCAAGGATCCGGCGGAGACTGTGAAGGCTGTTAGAAAGGCGTG
 CCCTCATGAGCCTAGGTGCTTCTCGTACCTGGGCAATTCGAACCTGATGAACGTGTAATTTGTGAGGCCCTGTGCC
 AGTGGT**AATAAAC**CACCATG**TAA**

C

MSMSNNEGTTERRKHKNSGLTTLVSLKSSDEMILALEPLLKRKARAIGFEVVKRSR**CK**VWPNKSKITKGDVGE
 AYHVRAFQVFRERLGLVPPNKRQGDLLVVS**HC**GTRH**CL**KR**H**LLLEPKW**N**DER**TH**CH**F**LPHSMAARGKDP**AET**
 VKAVRK**AC**PH**EP**RC**FS**YLG**N**FE**PD**ERVIVRPV**V**INTM 187

Figure S5

A



B

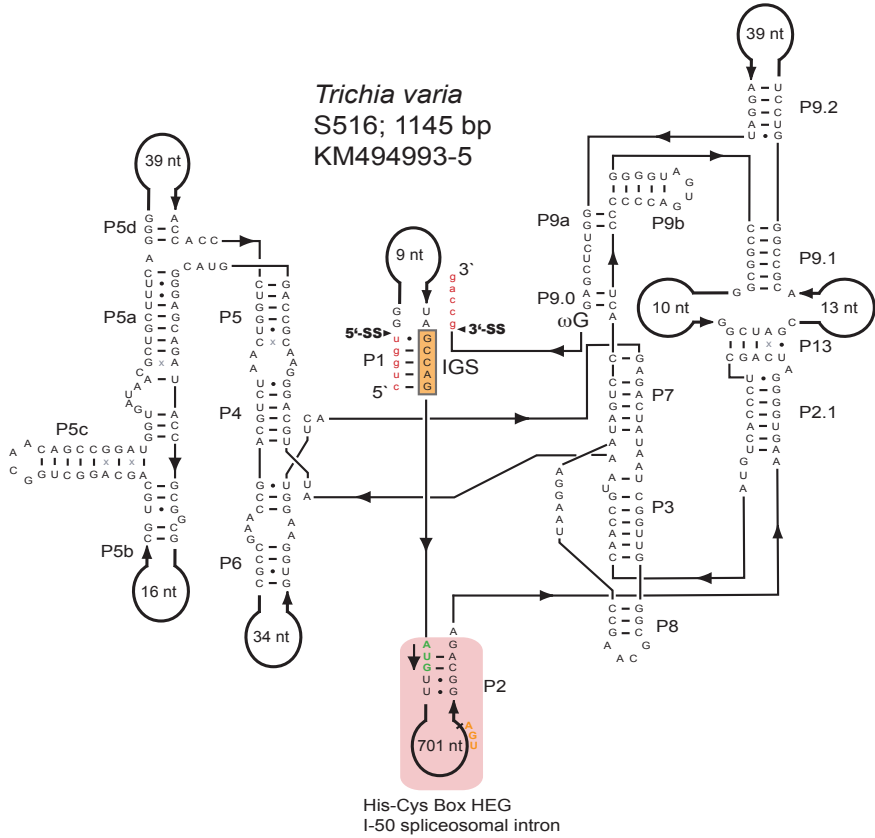


C

MSDRLHNKSGLTGLSFLQEDDVLAEWTKTLNRRVKDQKEARSAPKVI PHSR**C**RLLANHKGTLVMSSKHVCYGYQ
LRALVFPGRDRMAQVPPNKAGDSLIS**H V C G T R N C R R T H L V I E P K H I N**DERVH**C H M V A H W M W K K S K R Y G R M A K**
VRAAVQLA**C P H E P K C L T S V R D L D V K D R F V E S K P S N A I K Y D** 190

Figure S6

A



B

ATGTTTCACACGAGAAAGGCCACAACAAGTCCAACAAAAGCCACCTATCGTTCCTAGACGAGCCGGAGGCACAC
CTTATAGAATATATAAGGGAGCTGAAGGTGTGTGGCCGAGTAATTGGGAGTGTAGTGGGGCGGTGGCTTTGCGAG
GTGCGTGGTACTAACACCCCGTCCAGGCTCGCCTTGTGGATGGCACAATATCATGCGGGCCTTTGCCAAGGCTCGC
CCTGTGAGATGCTGCGGTCCCTGAAAGGCAGGCTGATGACTACTCGGTTGCGGGACCACAAGCCAAAGCGTGGACA
TCGCTTATGGGTATCAGTTGATAGCCCTGGAGAAGTTTGGGATCGGAATTATGCGGGAGGTTGCGCCGTCTAAGG
ACAGGGACTCGATTACAATATCACATTTGTGCGGGACCCTGCTCTGCTGCAGTGCAGGATCACCTGGTGTGGAAAC
CCAAATGGTGAACGACGAGCGGACTTGTGCCACTTCGTACTGCTTAGTGCTTTAAGCACGGCGGGTTAGAGG
CAGTGACCCCGTCTGGAGGCCGGGGTGCCCGCATAACCTCAGTGTGGCTCCATTGCTGTTAATAAGATAG
TACGCTTCCACCACATGCTCTGTACCCGATATAGCCCTACTCCTCCCGACGGAGGAGAGAAATAAAAAATATC
AATACTGA

C

MMFHTRKAHNKSNKSHLSFLDEPEAHLIEYIRELKVCGRVIGSARLVDGTYHAGLCQGSPCEMLRSLKGRMLMTR
LRDHKPSVDIAYGYQLIALEKFGIGIMREVAPSKDRDSITISHLCGTLLCCSADHLVLEPKWLNDECTCHFVLL
SAFKHGGLAETAVLEAGCPHTPQCGSIAVNKIVRFPFHALYPDIALPPPTEENKYYQ 210

Supplementary Figure S8

Direct repeat features in S516 intron segments

P1:

Comatricha anastomosans - JQ031961

```
1: AGGGGCAGATGGGAAGCGAGCGGGG----ATTTTCATGGTCCCCC---GCTTGAAAGCGGCAAGGAAGC-  
2: AGGGGTGGACGGGAAGCGAGCGGGG----ATTTTCATGGTCCCCC---GCTTGAAAGCGGCAAGGAAGCAGGGGAGGAA  
3: --GGGTGGACGGGAAGCGAGCGGGGGGGGATCTCACGGTCCCCCCCCTTGAAAGCGGCAAGGAAACAGGGGGGAA  
4: AGGGGTGGACAAAAAGCGAGCGGGG---ATCTCATGTTCTCCC---GCTTGAAAGCGGCAAGGAAGCAGGGGGGAA
```

```
1: TCTCCCC-GC-TTCCCCACTGCGTTCCAGATGCCCTG  
2: TCTCCCCTGCCTTCCCCACTGCGTTCCAGATGCCCTG  
3: TCTTCCCTGCCTTCC-TACTGCGTTCCAGATGCCCTG  
4: TCTCCCCTGCCTTCCCCACTGCGTTCTCGGACACCCTG
```

Copy no: 4
Motif size: 107 nt – 114 nt

Comatricha pseudoalpina - DQ903673

```
1: AGGGGGCGGATGGGAAGCGGGCGGGG--ATTTTCATGATCCCCC-GCTTGAAAGCGGCAAGGAAGCGTGGGG--TCA  
2: AGGGG--CGGACGGGAAGCGGGCGGGGGATCTCATGTTCTTCCCCTTGAAAGCGGCAAGGAAGCGACGGG-GCCA  
3: GGGGG--TGGACGGGAAGCGGGCGGGG-ATTTTCATGTTCCCCCGCTCGAAAGCGGCAAGGAAGCGAGGGG-CCCA  
4: AGGGG--TGGACGAGAAGCGGGCGGGG--ATTTTCATGTCCCCC-GCTTGAAAGCGGCAAGGAAGCGGGGGGTAA
```

```
1: TTCCCC-----GCTTCCCCACTGCGTTCCAGATGTCCAG  
2: TTCCCC---TAGCTTCCCCACTGCGTTCCAGATG-CCTCT  
3: TTACCC-----TGCTTCCCCACTGCGTTCCCGATAACCCTG  
4: TCTCCCCCCTGCCTTCCCCACTGCGTTCTCGGACGCCCTG
```

Copy no: 4
Motif size: 109 nt – 114 nt

P2:

Comatricha nigricapillitia - AY643824

```
1: ACAA  
2: ACAA  
3: ACAA  
4: ACAA
```

Copy no: 4
Motif size: 4 nt

Lamproderma album - JQ031971

1: ACCGATCTCACGCCAACGAACCAAGCATCCTCGGGGAGATGATGGTTGCCACACGGAACGACGCACCTGGTCA
2: ACCGATCTCGCGCCCAACGAACCAAGCATCCTCGGGGAGATGATGGTGCACATGGAACGACGCACCTGGTCA

1: GGTGCCGAGTTCCAGGCTCGATGC-----GATCTCTG-CCCTCGGGGTGGCAGGGAACGCGCTTGTGGATCGACA
2: GGTGCCGAGTTCAGGCTCGATGCACAGCGATCTTGCCCTCGGGGTGGCAGGGAACCCGCTCGTGGTTCGACG

1: GAAGGCGTGGGGGAGTATGAAGCGGAACGAGCCCTCGGGGCGGCTCTGGAGGCGGACGGGCAAGGTGT-GCAGG
2: GAAGGCGTGGGGAGAACCATAAGCGGAACGAGCCCTCGGGGTGGCTCTGGAAGCGGACGGGCGAGGTGTTGCAGG

Copy no: 2
Motif size: 219 and 226 nt

Lamproderma columbinum - HQ687204

Motif I:
1: GACCAATCGTGATTCTTGTT
2: GACCAATCGTGATTCTTGTT
3: GACCAATCGTGATTCTTGTT
4: GACCA-TCGTGATTCTTGTT
5: GACCAATCGTGATTCTTGTT

Copy no: 5
Motif size: 19 and 20 nt

Motif II:
1: GACTAAAAGGGAGGGGGTAAAT
2: GACTAAA-GGAAGGG--TAAAT
3: GACTAAA-GGGAGGGGGTAA-T

Copy no: 3
Motif size: 19 - 22 nt

Licea parasitica - JX481297

1: GGAGGGATACCTTTTCCGCGGCA I
2: GGAGGGATACCTTTTCCGCGGCA I
3: GGAGGGATACCTTTTCCGCGGCA I
4: GGAGGGATACCTTTTCCGCGGCA I
5: GGAGAGGCACCTTTTCCGCGGCA II
6: GGAGGGACACCTTTTCCGCGGCA III
7: GGAGGGACACCTTTTCCGCGGCA IV
8: GGAGAGGCACCTTTTCCGCGGCA II
9: GGAGAGGCACCTTTTCCGCGGCA II
10: GGAGGGATACCTTTTCCGCGGCA I
11: GGAGGGGACCTCTTCCGCGGCA V
12: GGAGAGGCACCTTTTCCGCGGCA VI
13: GGAGAGGCACCTTTTCCGCGGCA II
14: GGAGAGGCACCTTTTCCGCGGCA II
15: GGAGGGATACCTTTTCCGCGGCA I
16: GGAGGGACACCTCTTCCGCGGCA IV
17: GGAGGGACACCTCTTCCGCGGCA IV
18: GGAGGGACACCTCTTCCGCGGCA IV
19: GGAGGGACACCTCTTCCGCGGCA IV
20: GGAGGGACACCTCTTCCGCGGCA IV
21: GGAGGGACACCTCTTCCGCGGCA IV
22: GGAGGGACACCTCTTCCGCGGCA IV
23: GGAGGGACACCTCTTCCGCGGCA IV
24: GGAGGGACACCTCTTCCGCGGCA IV
25: GGAGAGGTACCTCTTCCGCGGCT VII

Copy no: 25
Motif size: 23 nt

Trichia varia – KM494996

1: AAAGGCCCGCCTTT
2: AAAGGCCCGCCTAT

Copy no: 2
Motif size: 14 nt

P5:

Comatricha pseudoalpina - DQ903673

1: GGGTGAGATCGATAAAGGTTTGGGCTATTAACGCCTGCGTGCCTGCTTCTTTCAA-GCCA
2: GGGCGAGATCGATAAAGGTTTGGGCTATTAACGCCTGCGTACCTGCTTCTTTCAAAGCCA

Copy no: 2
Motif size: 59 and 60 nt

P6:

Elaeomyxa cerifera - JQ031967

Motif I:

1: TAGGGTTACGC
2: TAGAGTTACGT
3: TAGGGTTACGC
4: CAGGTTTACGC

Copy no: 4
Motif size: 11 nt

Motif II:

1: TAACCCTAGA
2: TAACCCTAGA
3: TAACCCTAGG
4: AAACCTTAGC
5: TGACCCTAGG
6: TAACCCTAGG

Copy no: 6
Motif size: 10 nt

Motif III:

1: CAACCCTATGCAAAATCTTATGTTTCGGTGCCTCTAGGCCCCCGCGGTGAAAACCATGGTGGTTTGGTAAAAATCGAAGC
2: CAACC-TATGCAAAACCTTATGTTTCGGTGCCTCTGGGCCACAGTGGTAAAACCTATAGTGGTTTGGTAAAAATCGAAGC

Copy no: 2
Motif size: 79 and 80 nt

P8:

Colloderma oculatum - JQ031959

1: CCCCGCCGTCGACCCCTTGGCCCCCCTCGTCATGAGGGGGGGTCTCGGCGAGGGTAGGGTCTGCCCGTCG
2: TCCTGCCGTCGACCCCTTGGCCTGCCCTCGTCATGAGGGGGGGTCTCGGCGAGGGT----GTCTGCCCTGTCG

1: ACCCTTGGCCCGCCCTCGTCATGAGGGGGTGTCTCGGCGAGGG-TGGGG
2: ACCCTTGGCCTGCCCTCGTCGTGAGGGGGGTCTCGGCGAGGGGTGGG

Copy no: 2
Motif size: 120 and 123 nt

Lamproderma columbinum HQ687196

Motif I:
1: CTTT-AAAGGAACTG
2: CTTT-AAGGGA-CTC
3: CTTT-AAGGGA-CTG
4: CTTT-AAGGGA--TC
5: CTTTGAAGAGT-CAG

Copy no: 5
Motif size: 11 - 14 nt

Motif II:
GGCTTTAGTC
GGTTTTAGTC

Copy no: 2
Motif size: 10 nt

P9:

Lamproderma columbinum - HQ687204

1: AAAGCAGTCAAACCAAGTTAAAAGCAGTC
2: AATCCAGTCAAACCAAGTTAAAAGCAGGC

Copy no: 2
Motif size: 28 nt

Lamproderma columbinum - HQ687200

1: AAACCAAGTT
2: AAACCAAGTC
3: AAACCAAGC

Copy no: 3
Motif size: 9 nt

***Trichia varia* S516 Group I intron with and without HEG**

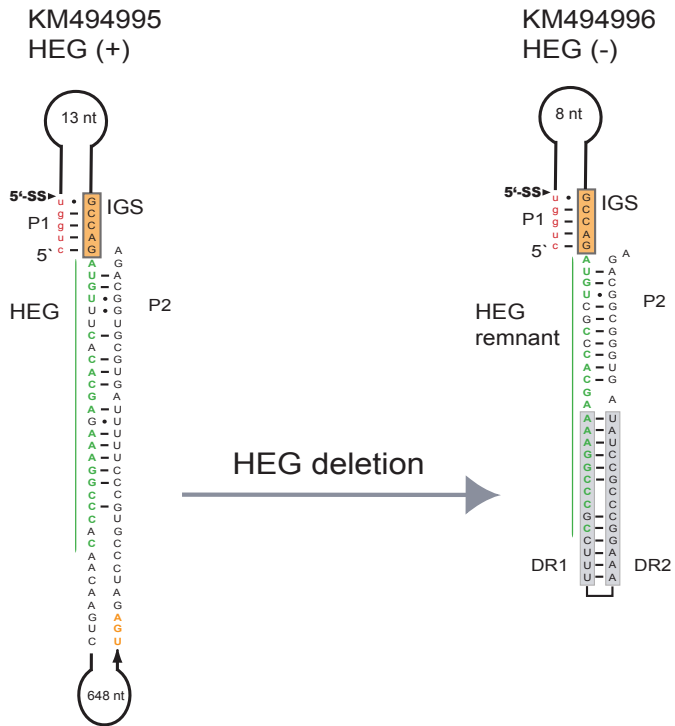


Figure S9

Table S1

Key features of 75 nucleolar group I introns at position S516 in myxomycetes.

Host species	Strain/ isolate	S516	Acc. Number	Insert
Myxomycetes				
Order: Physarales; Family: Didymiceae				
<i>Diderma alpinum</i>	Fr-K12	534 bp	ON155994	-
<i>Diderma niveum</i>	Fr-K10	496 bp	HE614615	-
<i>Diderma niveum</i>	It-K66	496 bp	HE614616	-
<i>Diderma niveum</i>	Uk-K79	496 bp	HE614617	-
<i>Diderma soundersii</i>	Mx-K30	482 bp	MZ313549	-
<i>Lepidoderma crustaceum</i>	It-K62	474 bp	HE614619	-
<i>Lepidoderma tigrinum</i>	AMFD192	501 bp	DQ903678	-
Order: Physarales; Family: Physarceae				
<i>Badhamia gracilis</i>	Az4-1	428 bp	AJ555451	-
<i>Fuligo leviderma</i>	AMFD130	402 bp	DQ903676	-
<i>Fuligo septica</i>	NY-1	534 bp	AJ584697	-
<i>Fuligo septica</i>	HMJAU1009	538 bp	MF983561	-
<i>Fuligo septica</i>	unknown	534bp	AJ555452	-
<i>Physarum flavicomum</i>	UFF6	>493 bp	AJ555453/4	-
<i>Physarum notabile</i>	LE47491	574 bp	JX035988	-
<i>Physarum</i> sp.	LE255721	441 bp	JX035984	-
<i>Physarum</i> sp.	AMFD-2018a	445 bp	MK336180	-
Order: Stemonitales; Family: Stemonitaceae				
<i>Comatricha anastomosans</i>	Now12905	1411 bp	JQ031961	DR-P1
<i>Comatricha nigricapillitia</i>	MM21077	489 bp	DQ903685	-
<i>Comatricha nigricapillitia</i>	Isolate 3	564 bp	AY643824	DR-P2
<i>Comatricha pseudoalpina</i>	MM23892	1471 bp	DQ903673	DR-P1; DR-P5c
<i>Colloderma oculatum</i>	HS2885	719 bp	JQ031959	DR-P8
<i>Colloderma robustum</i>	AMFD270	510 bp	JQ031960	-
<i>Diacheopsis paucilla</i>	MM29883	516 bp	JQ031966	-
<i>Elaeomyxa cerifera</i>	MM24498	725 bp	JQ031967	DR-P6
<i>Lamproderma acanthosporum</i>	MM36058	536 bp	JQ031968	-
<i>Lamproderma aeneum</i>	MM36255	497 bp	JQ031969	-
<i>Lamproderma aeneum</i>	AK06013	>490 bp	JQ031970	-
<i>Lamproderma album</i>	MM37151	993 bp	JQ031971	DR-P2
<i>Lamproderma columbinum</i>	Isolate F2	821 bp	HQ687204	DR-P2; DR-P9.2
<i>Lamproderma columbinum</i>	Isolate 106	885 bp	HQ687196	DR-P8
<i>Lamproderma columbinum</i>	Isolate 90	699 bp	HQ687198	-
<i>Lamproderma columbinum</i>	Isolate 94	650 bp	HQ687199	-
<i>Lamproderma columbinum</i>	Isolate 63b	594 bp	HQ687197	-
<i>Lamproderma columbinum</i>	Isolate 132	650 bp	HQ687200	DR-P9.2
<i>Lamproderma echinosporum</i>	AMFD136	510 bp	JQ031980	-
<i>Lamproderma echinosporum</i>	AK06016	509 bp	JQ031979	-
<i>Lamproderma pseudomaculatum</i>	MM37354	1082 bp	JQ031985	HEG-P2; I-58
<i>Lamproderma puncticulatum</i>	Isolate 172	623 bp	HQ687194	-
<i>Lamproderma</i> sp.	AMFD-2011a	449 bp	JQ031994	-
<i>Macbrideola oblonga</i>	unknown	462 bp	DQ903682	-
<i>Meriderma aggregatum</i>	AMFD135	484 bp	DQ903669	-
<i>Meriderma cribrarioides</i>	MM37106	489 bp	JQ032000	-
<i>Stemonitis fusca</i>	MM39482	460 bp	MK041079	-
<i>Stemonitis fusca</i>	AMFD519	482 bp	MK041078	-
Order: Liceales; Family: Liceaceae				
<i>Licea parasitica</i>	AMFD341	1149 bp	JX481297	DR-P2
Order: Liceales; Family: Reticulariaceae				
<i>Reticularia lycoperdon</i>	AMFD262	494 bp	JX481311	-

Order: Trichiales; Family: Trichiaceae

<i>Arcyria globosa</i>	AMFD252	459 bp	JX481282	-
<i>Trichia decipiens</i>	unknown	457 bp	FJ810499	-
<i>Trichia varia</i>	LE254838	443 bp	KM495030	-
<i>Trichia varia</i>	MdH-FR1210005	443 bp	KM495011	-
<i>Trichia varia</i>	sc22520	443 bp	KM495025	-
<i>Trichia varia</i>	sc27667c2	443 bp	KM495027	-
<i>Trichia varia</i>	sc27556c1	443 bp	KM495026	-
<i>Trichia varia</i>	sc27860c4	443 bp	KM495022	-
<i>Trichia varia</i>	sc27839	443 bp	KM495010	-
<i>Trichia varia</i>	sc27742	443 bp	KM495009	-
<i>Trichia varia</i>	sc22409	443 bp	KM494999	-
<i>Trichia varia</i>	AMDF451	443 bp	KM495017	-
<i>Trichia varia</i>	sc27664c1	443 bp	KM495008	-
<i>Trichia varia</i>	JVR848	482 bp	KM494996	DR-P2
<i>Trichia varia</i>	sc27850c2	443 bp	KM495029	-
<i>Trichia varia</i>	sc27507	443 bp	KM495023	-
<i>Trichia varia</i>	sc27745	443 bp	KM495012	-
<i>Trichia varia</i>	sc27648c1	443 bp	KM495024	-
<i>Trichia varia</i>	sc27850c1	443 bp	KM495021	-
<i>Trichia varia</i>	sc22370	1145 bp	KM494993	HEG-P2; I-50
<i>Trichia varia</i>	sc27772c3	443 bp	KM495013	-
<i>Trichia varia</i>	sc22556	443 bp	KM495007	-
<i>Trichia varia</i>	sc27697	443 bp	KM495006	-
<i>Trichia varia</i>	sc22517	443 bp	KM495005	-
<i>Trichia varia</i>	LE256579	443 bp	KM495031	-
<i>Trichia varia</i>	LE259461	1145 bp	KM494995	HEG-P2; I-50
<i>Trichia varia</i>	LE259268	1145 bp	KM494994	HEG-P2; I-50

Order: Trichiales; Family: Dianemataceae

<i>Dianema nivale</i>	MM29888	538 bp	JX481289	-
-----------------------	---------	--------	----------	---

Order: Echinosteliales; Family: Echinosteliaceae

<i>Echinostelium coelocephalum</i>	ATCC MYA-2984	1136 bp	AY842033	HEG-P2; I-59
------------------------------------	---------------	---------	----------	--------------

Table S2

Key features of direct repeat arrays in myxomycete S516 introns.

DR region	S516 intron ¹	Motif ²	Copy no	Heterogeneity ²	
P1	<i>Comatracha anastomosans</i> - JQ031961	ca 110 nt	4	+	
	<i>Comatracha pseudoalpina</i> - DQ903673	ca 110 nt	4	+	
P2	<i>Comatracha nigricapillitia</i> - AY643824	4 nt	4	-	
	<i>Lamproderma album</i> - JQ031971	ca 220 nt	2	+	
	<i>Lamproderma columbinum</i> - HQ687204	ca 20 nt	5	+	
		ca 40 nt	3	+	
	<i>Licea parasitica</i> - JX481297	23 nt	25	+	
<i>Trichia varia</i> - KM494996	14 nt	2	+		
P5	<i>Comatracha pseudoalpina</i> - DQ903673	ca 60 nt	2	+	
P6	<i>Elaeomyxa cerifera</i> - JQ031967	11 nt	4	+	
		10 nt	6	+	
		Ca 80 nt	2	+	
P8	<i>Colloderma oculatum</i> - JQ031959	ca 120 nt	2	+	
		<i>Lamproderma columbinum</i> - HQ687196	ca 12 nt	5	+
			10 nt	2	+
P9	<i>Lamproderma columbinum</i> - HQ687204	28 nt	2	+	
	<i>Lamproderma columbinum</i> - HQ687200	9	3	+	

Notes:

¹⁾ More detailed information about the S516 introns is provided in Tabel S1.²⁾ Motif sequences and heterogeneity are presented in Figure S8.

Paper III

This is an open-access article, reproduced and distributed under the terms of the
Creative Commons Attribution License (CC BY)

RESEARCH

Open Access



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

Kjersti Lian¹, Betty M. N. Furulund², Anders A. Tveita³, Peik Haugen⁴ and Steinar D. Johansen^{2*}

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

Background

Nuclear group I introns are intervening sequences so far exclusively reported in ribosomal DNA (rDNA) that interrupt highly conserved sites in the small subunit (SSU) and large subunit (LSU) rRNA genes [1, 2]. Group I introns appear restricted to some eukaryotic lineages, and among these are the ascomycete fungi and various groups of protists such as the ciliates, amoeba-flagellates, green algae and myxomycetes. About 100 rDNA

*Correspondence: steinar.d.johansen@nord.no

² Genomics division, Faculty of Biosciences and Aquaculture, Nord University, N-8049 Bodø, Norway
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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 complexed P5 RNA structure than group IE introns. Detailed structural information is available from the *Tetrahymena* 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 transesterification reaction [2, 13]. The self-splicing reaction, which is catalyzed 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

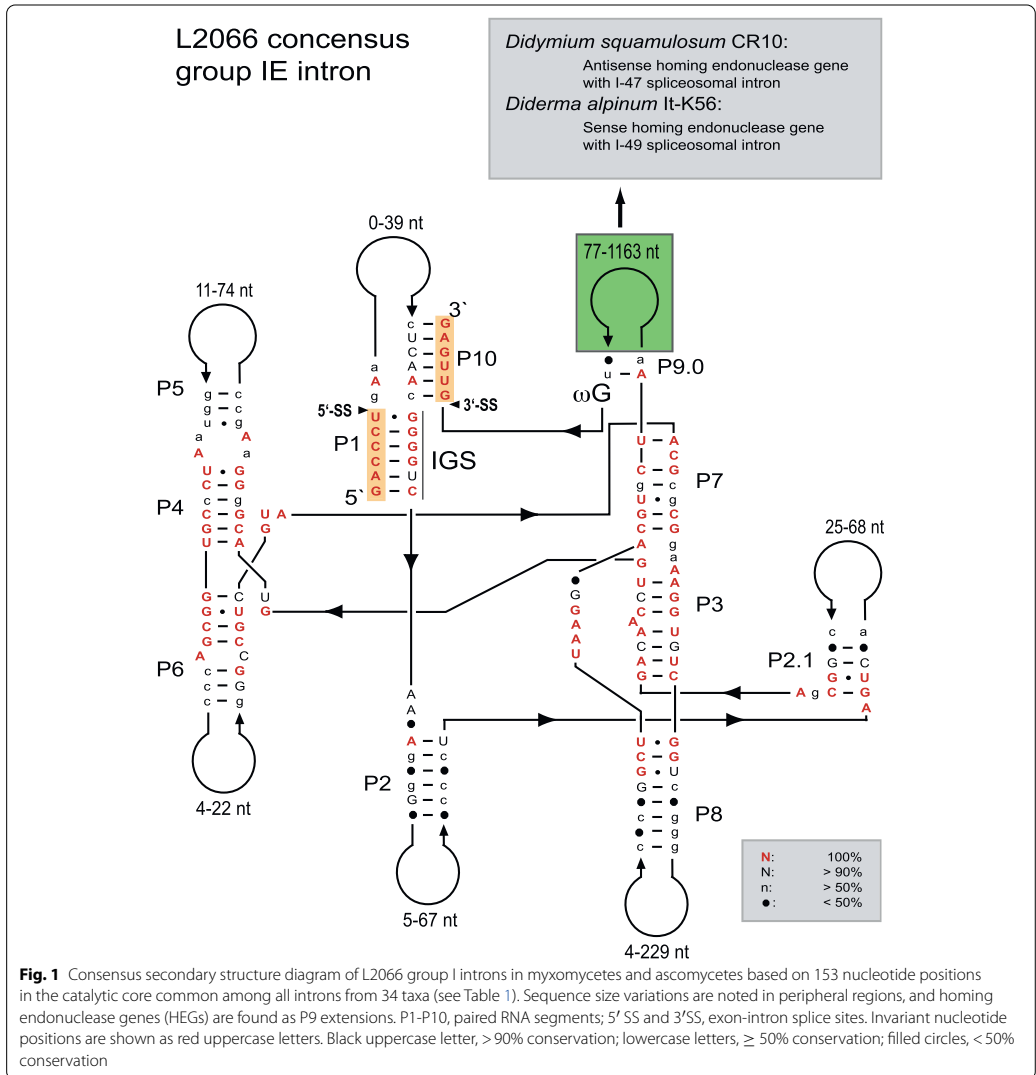
Table 1 Key features of L2066 group I introns in myxomycetes and ascomycetes

Host species	Strain/ isolate	L2066	HEG ^a	Acc. Number
MYXOMYCOTA				
Order: Physariales				
<i>Craterium minutum</i>	It-IG38 (Italy)	536 bp	–	HE655081
<i>Diderma alpinum</i>	It-K56 (Italy)	1480 bp	+ (S; I-49)	HE655057
<i>Diderma alpinum</i>	Uk-K78 (Ukraine)	467 bp	–	HE655058
<i>Diderma meyeræ</i>	It-K61 (Italy)	414 bp	–	HE655059
<i>Diderma niveum</i>	Fr-K10 (France)	438 bp	–	AM407429
<i>Diderma niveum</i>	It-K66 (Italy)	434 bp	–	HE655060
<i>Diderma niveum</i>	Uk-K79 (Ukraine)	434 bp	–	HE655061
<i>Diderma saundersii</i>	Mx-K30 (Mexico)	595 bp	–	AM407428
<i>Didymium squamulosum</i>	CR10 (Costa Rica)	1197 bp	+ (AS; I-47)	AM407427
<i>Fuligo septica</i>	IW1 (Iowa-USA)	389 bp	–	HE655082
<i>Lepidoderma aggregatum</i>	Uk-K86 (Ukraine)	449 bp	–	HE655062
<i>Lepidoderma alpestroides</i>	Fr-K17 (France)	435 bp	–	ON155995
<i>Lepidoderma caestianum</i>	Fr-K18 (France)	449 bp	–	AM407430
<i>Lepidoderma crustaceae</i>	It-K62 (Italy)	476 bp	–	HE655064
<i>Mucilago crustacea</i>	No-K94 (Norway)	468 bp	–	HE655067
Order: Stemonitales				
<i>Comatricha laxa</i>	EdHa (USA)	493 bp	–	ON155996
ASCOMYCOTA				
Order: Capnodiales				
<i>Sphaerulina quercicola</i>	CBS 663.94 (Netherlands)	367 bp	–	GU214496
Order: Hypocreales				
<i>Beauveria bassiana</i>	DAOM216540	358 bp	–	EU334679
<i>Beauveria bassiana</i>	ARSEF654	358 bp	–	KJ701419
<i>Beauveria bassiana</i>	ARSEF502	358 bp	–	KJ701420
<i>Beauveria bassiana</i>	ARSEF2991	359 bp	–	EU334676
<i>Beauveria bassiana</i>	V55714 (Germany)	358 bp	–	MG654725
<i>Beauveria bassiana</i>	VSP11 (Germany)	358 bp	–	MG654726
<i>Beauveria bassiana</i>	STB	358 bp	–	JF429894
<i>Cordyceps kanzashiana</i>	(Japan)	380 bp	–	AB044639
<i>Cordyceps militaris</i>	ATCC 34164	360 bp	–	CP023322
<i>Cordyceps proflicca</i>	97,009 (Japan)	379 bp	–	AB044641
<i>Fusarium</i> sp.	Uk-K90 (Ukraine)	287 bp	–	ON155007
<i>Levanicillium</i> sp.	CEP 419 (Argentina)	364 bp	–	MH013330
<i>Ophiocordyceps sinensis</i>	AMI1080 (China)	341 bp	–	FJ461354
<i>Ophiocordyceps sinensis</i>	AMI1081 (China)	341 bp	–	FJ461355
<i>Paecilomyces tenuipes</i>	(Japan)	400 bp	–	AB044642
Order: Saccharomycetales				
<i>Myxozyma monticola</i>	NRRL Y-17726	376 bp	–	DQ518989
<i>Myxozyma nipponensis</i>	NRRL Y-27625	342 bp	–	DQ518993

^a (+) denotes presence of homing endonuclease gene (HEG). (S) denotes sense orientation and (AS) antisense orientation. (I-49) and (I-47) denote the presence of spliceosomal introns of 49 bp and 47 bp, respectively

(Table 1). 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).

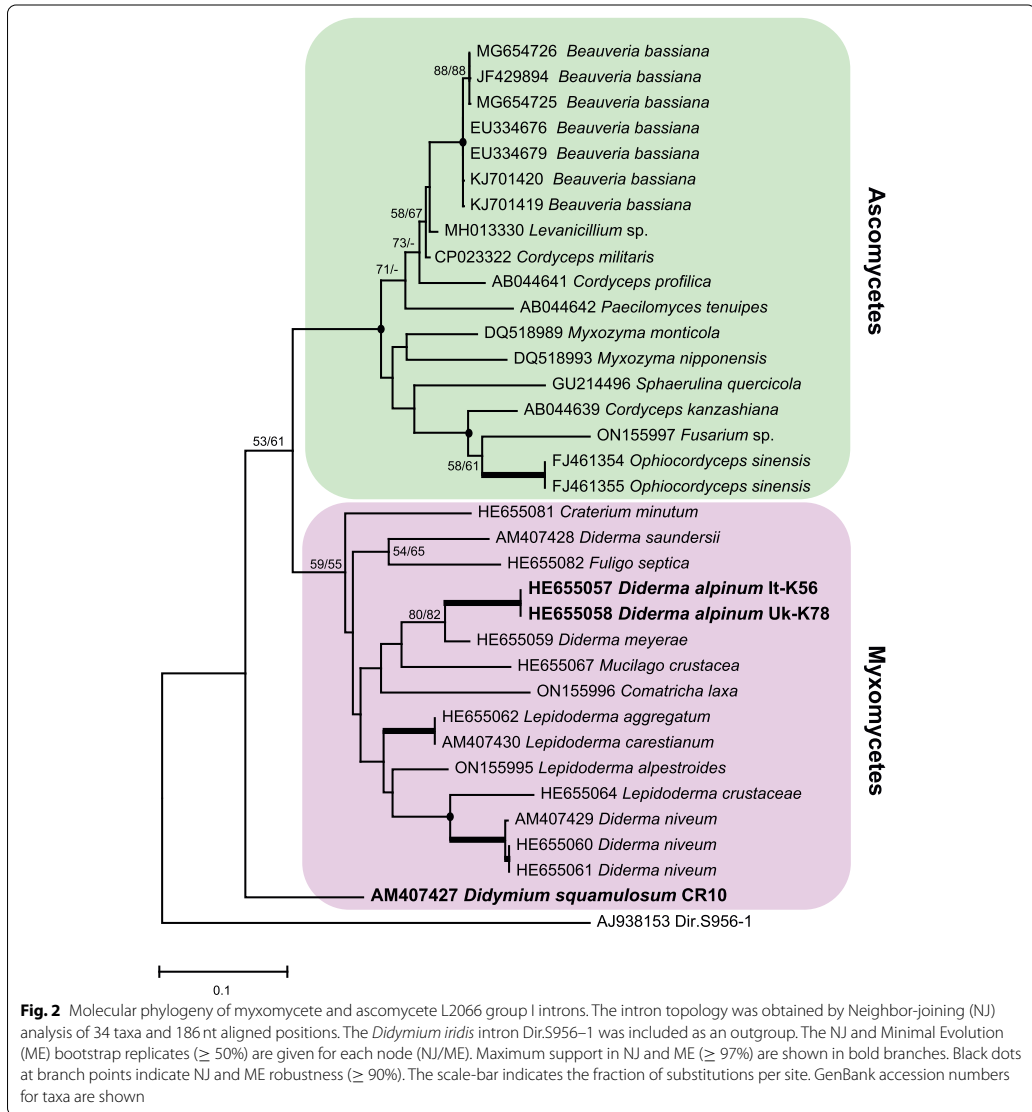


***D. alpinum* and *D. squamulosum* L2066 group I introns harbor HEGs in sense or antisense orientations that are interrupted by small spliceosomal introns**

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

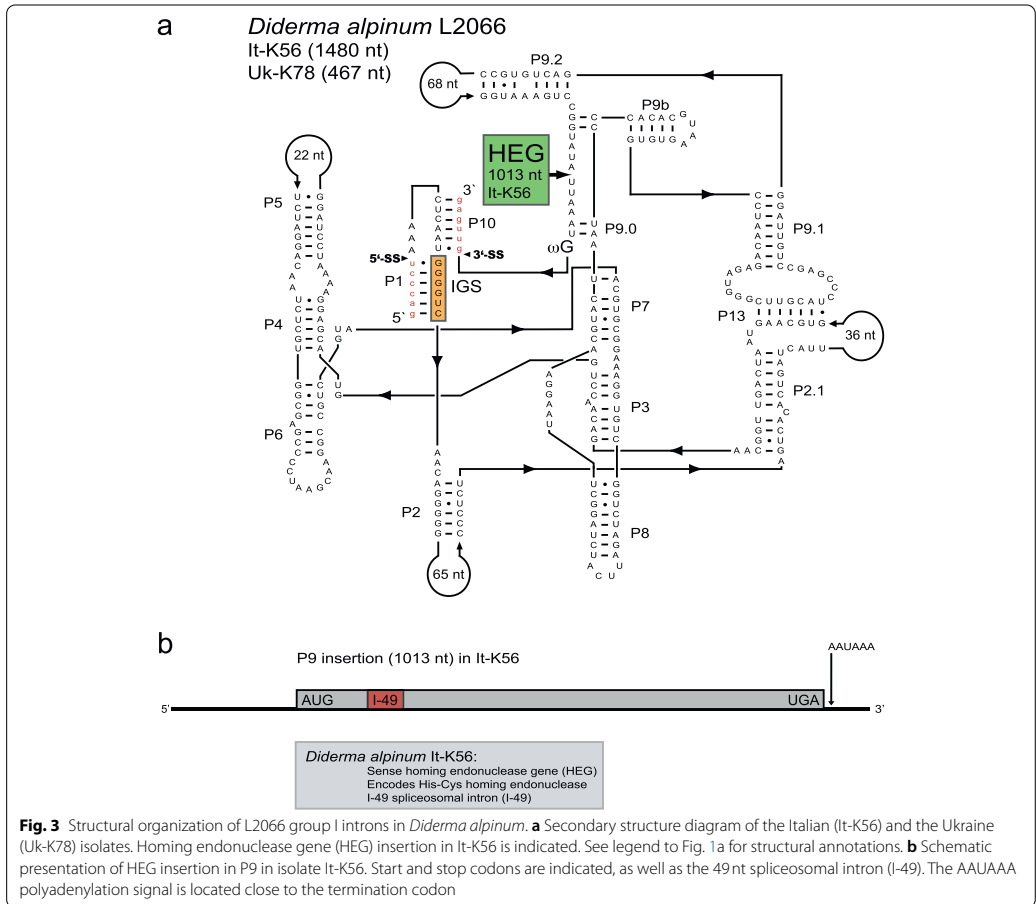
encoding a homing endonuclease protein consisting of 235 amino acids (named *I-Dall*) and belonging to the His-Cys family (see Fig. 4c). In addition, the HEG is associated with a consensus polyadenylation signal (AAUAAA). Finally, the *D. alpinum* HEG was found to be interrupted by a 49-nt spliceosomal 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



(Fig. 4a). 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-Dsq1*) consisting of 204 amino acids and belonging to the His-Cys family (Fig. 4c).



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 findings 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-DsqI homing endonuclease is expressed from the rDNA antisense strand in vivo

To examine *I-DsqI* 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-DsqI* 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

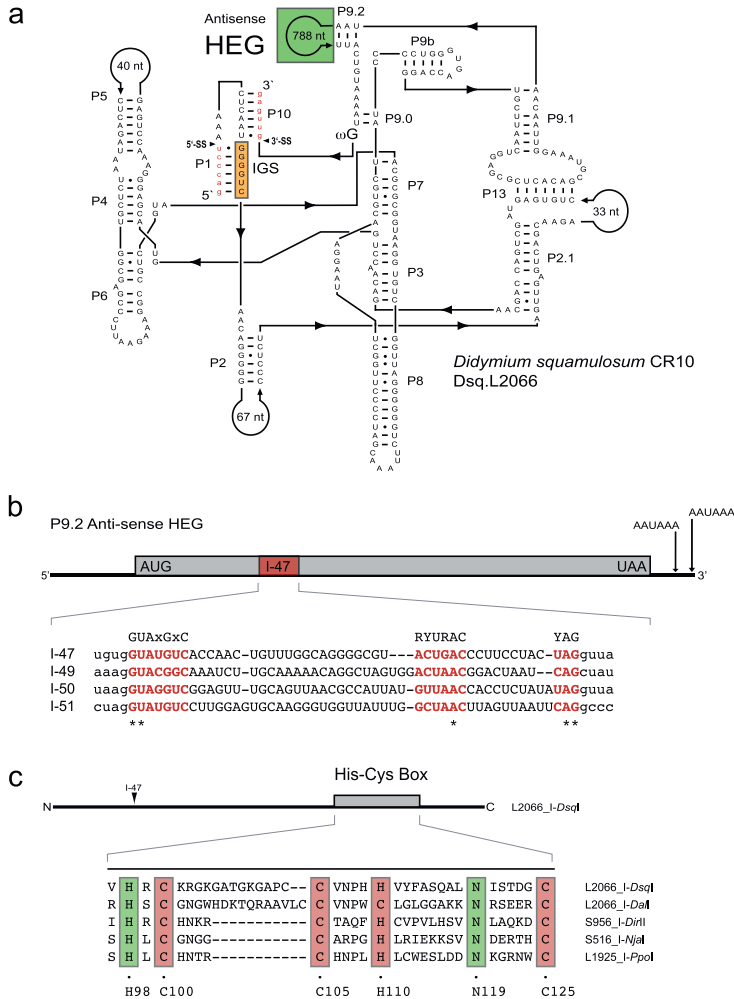
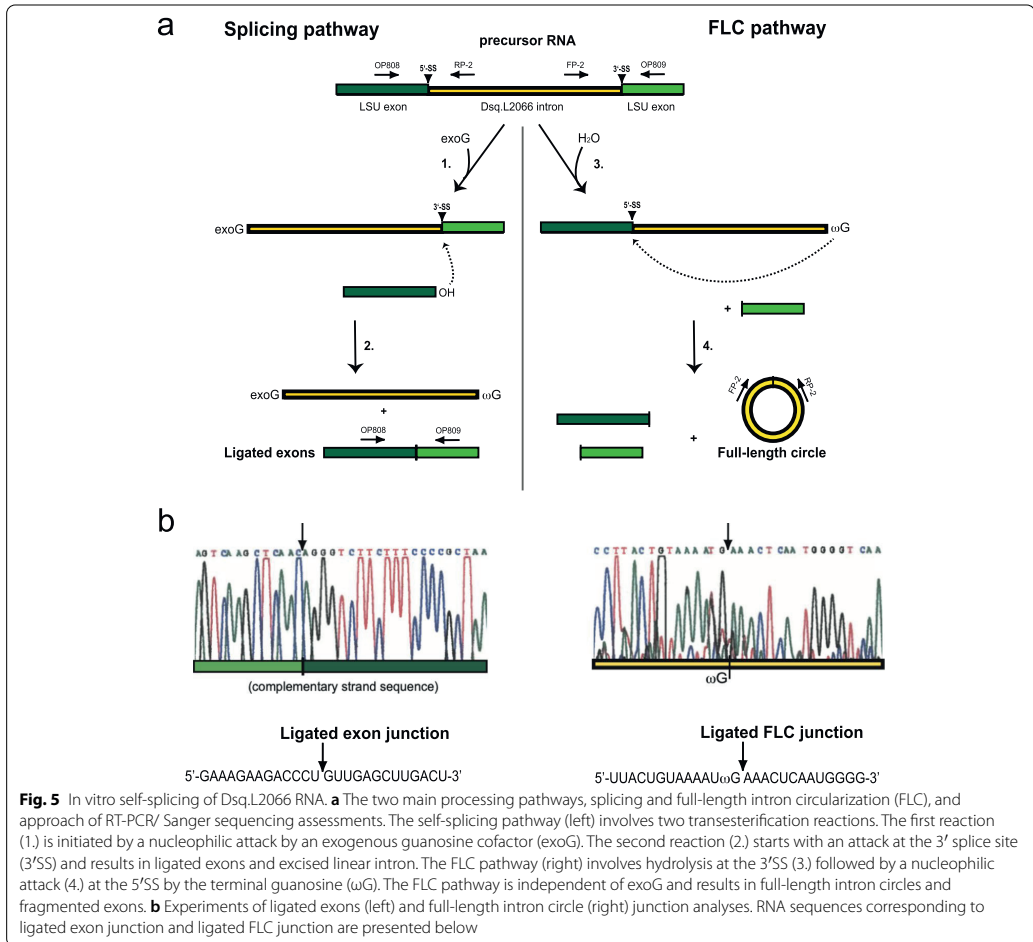


Fig. 4 Structural organization of L2066 group I intron in *Didymium squamulosum*. **a** Secondary structure diagram of the Costa Rican (CR10) isolate. Antisense homing endonuclease gene (HEG) insertion is indicated. See legend to Fig. 1a for structural annotations. **b** Schematic presentation of HEG 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-AAG). **c** Schematic presentation of the I-DsqI homing endonuclease containing a His-Cys box. Below is an amino acid alignment of His-Cys box features in *Didymium squamulosum* (I-DsqI), *Diderma alpinum* (I-DalI), *Didymium iridis* (I-DirI), *Naegleria jamiesoni* (I-Njal) and *Physarum polycephalum* (I-Ppol). I-Njal and I-Ppol represent well-studied His-Cys homing endonucleases. Conserved residues (boxed) corresponding to those presented in the I-Ppol crystal structure [20]. C100, C105, H110 and C125 are involved in zinc binding, and H98 and N119 are associated with the active site

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



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

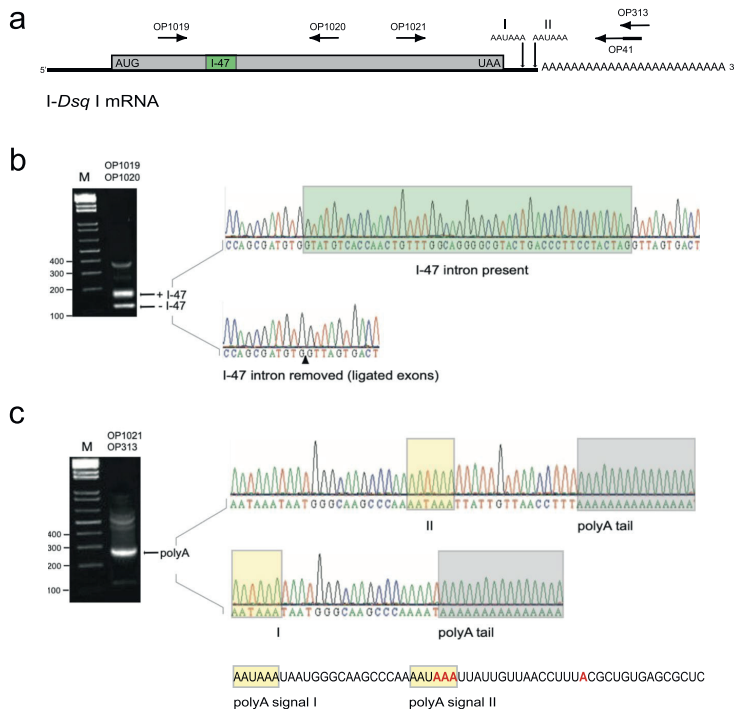


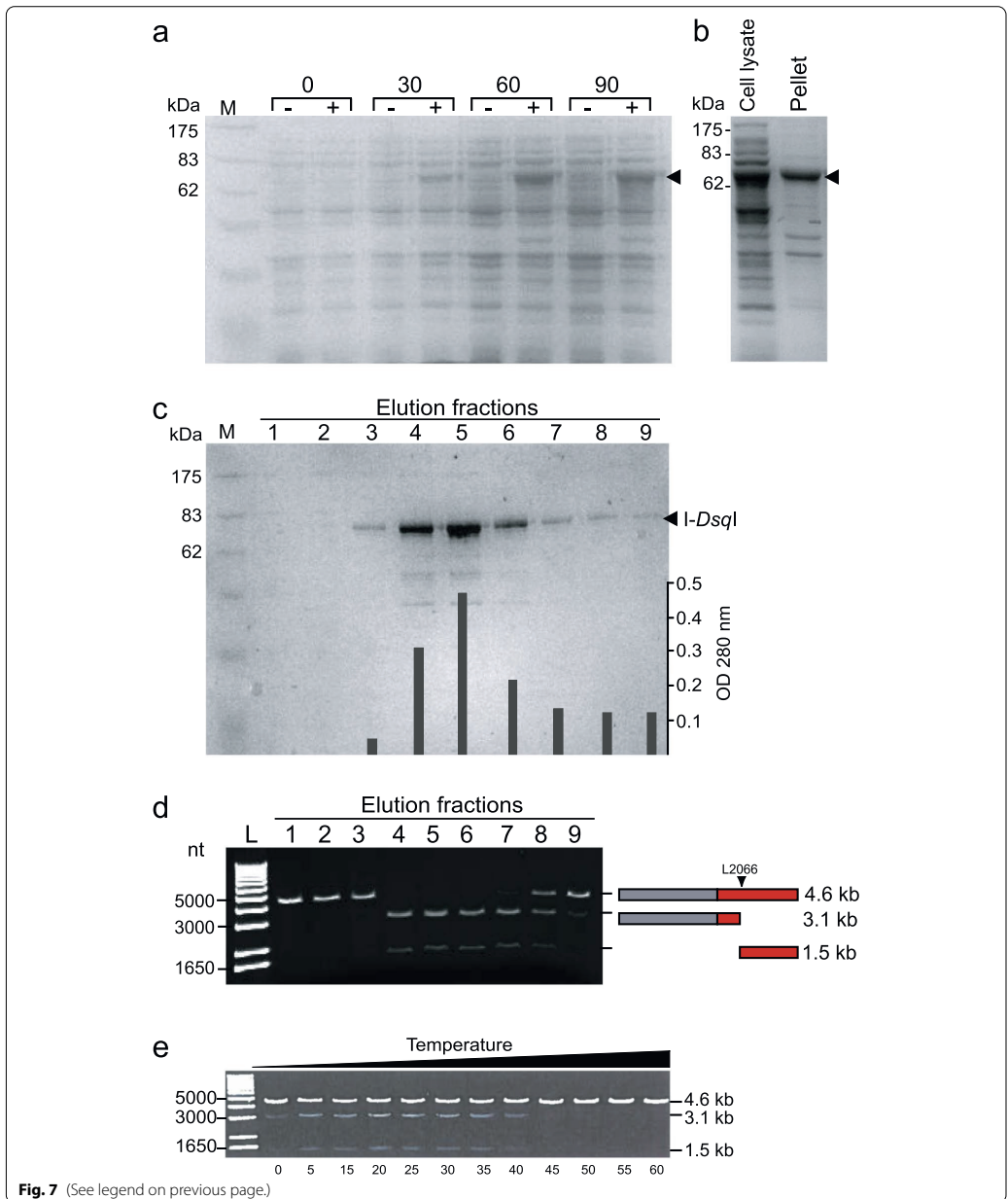
Fig. 6 In vivo maturation of *I-DsqI* mRNA in *Didymium squamulosum*. **a** Schematic presentation of *I-DsqI* mRNA. Start (AUG) and stop (UAA) codons are indicated, as well as the 47 nt spliceosomal intron (I-47). Two AAUAAA polyadenylation signals (I and II) are located close to the termination codon. Primer locations used in assessment of intron removal and polyadenylation are indicated. **b** RNA from amoeba cells was analyzed by an RT-PCR approach using a primer set flanking I-47. Amplicons corresponding to I-47 lacking (–) and I-47 containing (+) mRNAs were detected by gel analysis (left) and confirmed by Sanger sequencing (right). **c** RNA from amoeba cells was analyzed by an RT-PCR approach using a primer set flanking the polyadenylation signals. A poly(T) specific primer was used in the first-strand synthesis reaction that contained a unique sequence tag at the 5' end, corresponding to a tag-complementary downstream primer. Amplicons containing polyA tails were detected by gel analysis (left) and confirmed by Sanger sequencing (right)

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

(See figure on next page.)

Fig. 7 *I-DsqI* homing endonuclease expression, purification and activity. **a** 10% SDS-PAGE gel showing expression of MBP-*I-DsqI* in *E. coli* CodonPlus cells. Samples were harvested from induced (+) and uninduced (–) cells every thirty minutes. M is the molecular mass standard. Incubation times are shown above lanes. Arrowhead indicates the expressed endonuclease. The protein is theoretically 68 kDa. **b** After expression, the cells were lysed and centrifuged. The supernatant (cell lysate) and the pellet material (insoluble matter) were run on a 10% SDS-PAGE to clarify if MBP-*I-DsqI* is soluble or not. Expected size of MBP-*I-DsqI* is indicated with an arrow. **c** SDS-PAGE analysis of affinity purified MBP-*I-DsqI*. Protein purification was executed using amylose resin. Bound MBP-*I-DsqI* was released from the column using maltose. Nine fractions were sampled. M is a broad range prestained protein marker (New England Biolabs). Purified protein is indicated by an arrow. OD 280 nm measurements of each fraction are shown. Fraction 5 contains the highest concentration of MBP-*I-DsqI*. **d** Activity study. The nine fractions from c) were incubated with the linearized target DNA pDan122/747 for 15 min. Two bands at 3.07 kb and 1.53 kb indicate *I-DsqI* activity. **e** Temperature range of activity observed after incubation of 0.1 unit affinity purified *I-DsqI* and linear target DNA for 60 min. Temperatures assessed ranged from 0 °C to 60 °C



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-Dsq 1 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-Nja1 homing endonuclease [23].

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

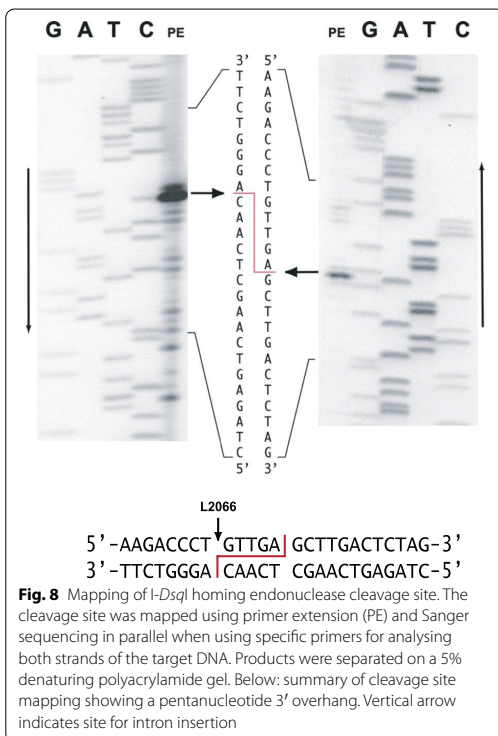
We report analyses of the L2066 family of group I introns based on a dataset of 34 introns representing ascomycete and myxomycete taxa. All introns belong to the group IE subtype and possess highly conserved catalytic RNA cores. Introns from two myxomycete taxa, *D. alpinum*

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 49bp and 47bp, 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 amoeba-flagellates [29, 31] and myxomycetes [24, 30, 32, 33].

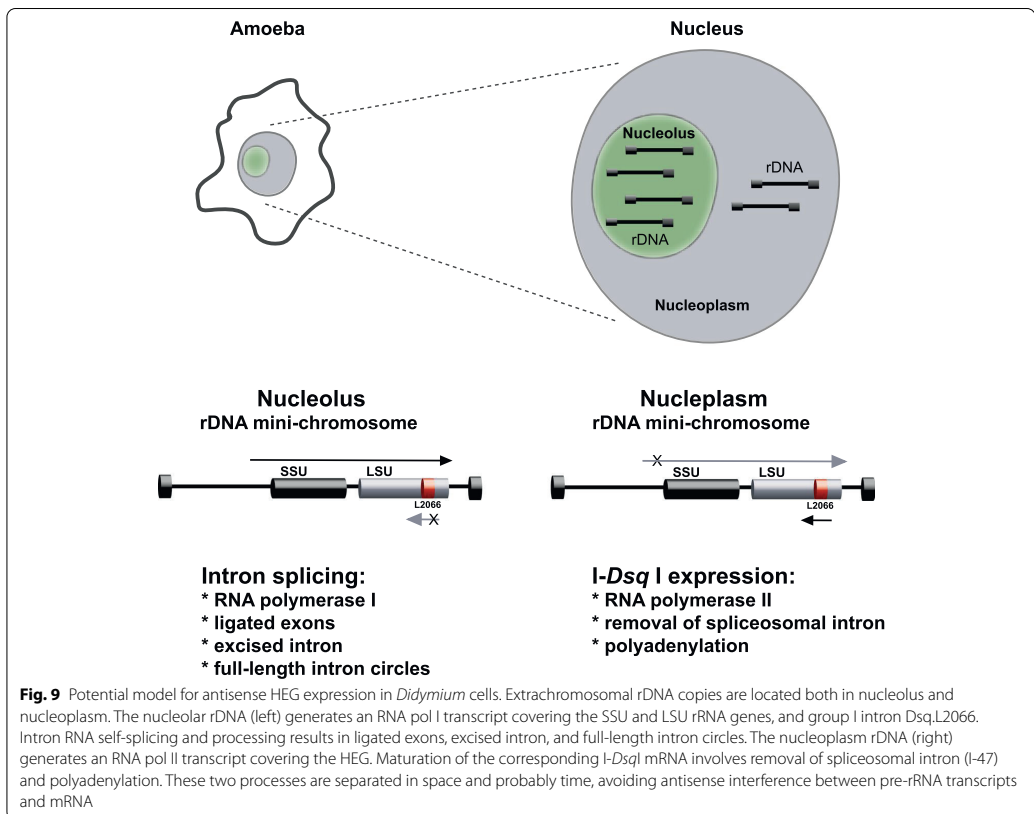
Expression of the *D. squamulosum* HEG can potentially create antisense interference between the HEG mRNA and the rRNA in cells. The essential role of rRNAs in ribosomes (and thus in cell growth) can



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-NjaI*, *I-NitI*, and *I-NanI* homing endonucleases [22, 23] and *Didymium I-DitII* [24], but different from the tetranucleotide 3'-overhang made by *I-PpoI* from *Physarum* [20, 21].



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-*DsqI*, 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 TTC-3'; OP1020, 5'-AAT CTT GAG TTC GTG GTA GC-3'; OP1021, 5'-ACG TCT ACT TTG CGA GC-3'; OP313, 5'-AAG CG ACGC ATG CACGCA TT-3'; OP41, 5'-CGA CGC 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 ATA GGG CTT GGC ACA ATT AGC GG-3'; OP809, 5'-GTT AGT TAC AGC ATT AGC-3'.

Plasmid construction

The I-*DsqI* 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-*DsqI* 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-*DsqI* was used to introduce the HEG into the pTH1 expression vector in an LR reaction. Final construct for N-terminal MBP fusion was named pI-*DsqI*-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 *NdeI* 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 GGC A-3'; OP1423, 5'-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC CTA CCA GCA TGC TGG GGT GTG GTT-3'.

Protein expression

For expression of fusion proteins pI-*DsqI*-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 Sorvall 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 pH 8.3), 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-³³P dATP as the label. Primer extension products were separated on a 5% polyacrylamide gel alongside manual 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 myxomycete and ascomycete 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., P.H. and S.D.J. 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 P.H.; 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 (*Comatracha 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

¹Nofima AS, Muninbakken 9-13, Breivika, 9291 Tromsø, Norway. ²Genomics division, Faculty of Biosciences and Aquaculture, Nord University, N-8049 Bodø, Norway. ³Medical Department, Bærum Hospital, Vestre Viken Hospital Trust, Drammen, Norway. ⁴Department 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

- 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>.
- 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>.
- Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, et al. The comparative RNA web (CRW) 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>.
- 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;323:35–52. [https://doi.org/10.1016/S0022-2836\(02\)00895-1](https://doi.org/10.1016/S0022-2836(02)00895-1).
- Nielsen H, Johansen SD. Group I introns: moving in new directions. *RNA Biol.* 2009;6:375–83. <https://doi.org/10.4161/rna.6.4.9334>.
- Johansen S, Haugen P. A new nomenclature of group I introns in ribosomal DNA. *RNA.* 2001;7:935–6. <https://doi.org/10.1017/S1355838201010500>.
- Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. *Trends Biochem Sci.* 2006;31:41–51. <https://doi.org/10.1016/j.tibs.2005.11.008>.
- Cech TR, Damberger SH, Gutell RR. Representation of the secondary and tertiary structure of group I introns. *Nature Struct Biol.* 1994;1:273–80. <https://doi.org/10.1038/nsb0594-273>.
- Guo F, Gooding AR, Cech TR. Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell.* 2004;16:351–62.
- Su Z, Zhang K, Kappel K, Li S, Palo MZ, Pintilie GD, et al. Cryo-EM structures of full-length *Tetrahymena* ribozyme at 3.1 Å resolution. *Nature.* 2021;596:603–7. <https://doi.org/10.1038/s41586-021-03803-w>.
- Bonilla SL, Vicens Q, Kieft JS. Cryo-EM reveals an entangled kinetic trap in the folding of a catalytic RNA. *Sci Adv.* 2022;8:eabq4144. <https://doi.org/10.1126/sciadv.abq4144>.
- Andersen KL, Beckert B, Masquida B, Johansen SD, Nielsen H. Accumulation of stable full-length circular group I intron RNAs during heat-shock. *Molecules.* 2016;21:1451. <https://doi.org/10.3390/molecules21111451>.
- Cech TR. Self-splicing of group I introns. *Annu Rev Biochem.* 1990;59:543–68. <https://doi.org/10.1146/annurev.bi.59.0701.002551>.
- Nielsen H, Fiskaa T, Birgisdottir AB, Haugen P, Einvik C, Johansen S. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA.* 2003;9:1464–75. <https://doi.org/10.1261/rna.5290903>.
- Muscarella DE, Vogt VM. A mobile group I intron in the nuclear rDNA of *Physarum polycephalum*. *Cell.* 1989;56:443–54. [https://doi.org/10.1016/0092-8674\(89\)90247-x](https://doi.org/10.1016/0092-8674(89)90247-x).
- Johansen S, Elde M, Vader A, Haugen P, Haugli K, Haugli F. *In vivo* mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. *Mol Microbiol.* 1997;24:737–45. <https://doi.org/10.1046/j.1365-2958.1997.3921743.x>.
- Belfort M, Roberts RJ. Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* 1997;25:3379–88. <https://doi.org/10.1093/nar/25.17.3379>.
- Johansen S, Embley TM, Willassen NP. A family of nuclear homing endonucleases. *Nucleic Acids Res.* 1993;21:4405. <https://doi.org/10.1093/nar/21.18.4405>.
- Hafez M, Hausner G. Homing endonucleases: DNA scissors on a mission. *Genome.* 2012;55:553–69. <https://doi.org/10.1139/g2012-049>.
- Flick KE, Jurica MS, Monnat RJ, Stoddard BL. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-Ppo1. *Nature.* 1998;394:96–101. <https://doi.org/10.1038/27952>.
- Muscarella DE, Ellison EL, Ruoff BM, Vogt VM. Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. *Mol Cell Biol.* 1990;10:3386–96. <https://doi.org/10.1128/mcb.10.7.3386-3396.1990>.
- Elde M, Haugen P, Willassen NP, Johansen S. I-Njal, a nuclear intron-encoded homing endonuclease from *Naegleria*, generates a pentanucleotide 3' cleavage-overhang within a 19 base-pair partially symmetric DNA recognition site. *Eur J Biochem.* 1999;259:281–8. <https://doi.org/10.1046/j.1432-1327.1999.00035.x>.
- Elde M, Willassen NP, Johansen S. Functional characterization of isoschizomeric his-Cys box homing endonucleases from *Naegleria*. *Eur J Biochem.* 2000;267:7257–66. <https://doi.org/10.1046/j.1432-1327.2000.01862.x>.
- Johansen SD, Vader A, Sjøttem E, Nielsen H. *In vivo* expression of a group I intron HEG from the antisense strand of *Didymium* ribosomal DNA. *RNA Biol.* 2006;3:157–62. <https://doi.org/10.4161/rna.3.4.3958>.
- 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>.
- 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>.
- Tang Y, Nielsen H, Birgisdottir AB, Johansen SD. A natural fast-cleaving branching ribozyme from the amoeboid flagellate *Naegleria gringisheimi*. *RNA Biol.* 2011;8:997–1004. <https://doi.org/10.4161/rna.8.6.16027>.
- Meyer M, Nielsen H, Olieric V, Roblin P, Johansen SD, Westhof E, et al. Specification 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>.
- Tang Y, Nielsen H, Masquida 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>.
- Vader A, Nielsen H, Johansen S. *In vivo* expression of the nuclear group I intron-encoded I-Diri homing endonuclease involves the removal of a spliceosomal intron. *EMBO J.* 1999;18:1003–13. <https://doi.org/10.1093/emboj/18.4.1003>.
- Haugen P, De Jonckheere JF, Johansen S. Characterization of the self-splicing products of two complex *Naegleria* LSU rDNA 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>.
- Furulund BMN, Karlsen BO, Babiak 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/nrna7030043>.
- Furulund BMN, Karlsen BO, Babiak I, Haugen P, Johansen SD. Structural organization of S516 group I introns in myxomycetes. *Genes.* 2022;13:944. <https://doi.org/10.3390/genes13060944>.
- Haugen P, Wikmark OG, Vader A, Coucheron D, Sjøttem E, Johansen SD. The recent transfer of a homing endonuclease gene. *Nucleic Acids Res.* 2005;33:2734–41. <https://doi.org/10.1093/nar/gki564>.
- Johansen S, Vogt VM. An intron in the nuclear ribosomal DNA of *Didymium iridis* 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\(94\)90511-8](https://doi.org/10.1016/0092-8674(94)90511-8).
- Decatur WA, Einvik C, Johansen S, Vogt VM. Two group I ribozymes with different functions in a nuclear rDNA intron. *EMBO J.* 1995;14:5558–68. <https://doi.org/10.1002/j.1460-2075.1995.tb00135.x>.
- 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/msh005>.
- Will CL, Luhrmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol.* 2011;3:a003707. <https://doi.org/10.1101/cshperspect.a003707>.
- Schlautmann LP, Gehring NH. A day in the life of the exon junction complex. *Biomolecules.* 2020;10:866. <https://doi.org/10.3390/biom10060866>.
- Nielsen H, Krogh N, Masquida B, Johansen SD. The lariat capping ribozyme. In: Muller S, Masquida B, Winkler W, editors. *Ribozymes*: WILEY-VCH GmbH; 2021. p. pp 118–42. ISBN: 978–3–527-34454-3.
- Eickbush TH, Eickbush DG. Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol Spectr.* 2015;3:MDNA3-0011-2014. <https://doi.org/10.1128/microbiolspec.MDNA3-0011-2014>.
- Ferris PJ, Vogt VM, Truitt CL. Inheritance of extrachromosomal rDNA in *Physarum polycephalum*. *Mol Cell Biol.* 1983;3:635–42. <https://doi.org/10.1128/mcb.3.4.635-642.1983>.
- Silliker ME, Collins OR. Non-mendelian inheritance of mitochondrial DNA and ribosomal DNA in the myxomycete, *Didymium iridis*. *Mol Gen Genet.* 1988;213:370–8. <https://doi.org/10.1007/BF00339605>.
- Johansen S, Johansen T, Haugli F. Extrachromosomal ribosomal DNA of *Didymium iridis*: sequence analysis of the large subunit ribosomal RNA gene and sub-telomeric region. *Curr Genet.* 1992;22:305–12. <https://doi.org/10.1007/BF00317926>.
- Wikmark OG, Haugen P, Lundblad EW, Haugli K, Johansen SD. The molecular evolution and structural organization of group I introns at position 1389 in nuclear small subunit rDNA of myxomycetes. *J Euk Microbiol.* 2007;54:49–56. <https://doi.org/10.1111/j.1550-7408.2006.00145.x>.

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, Niyaz 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 *Fulligo 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.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

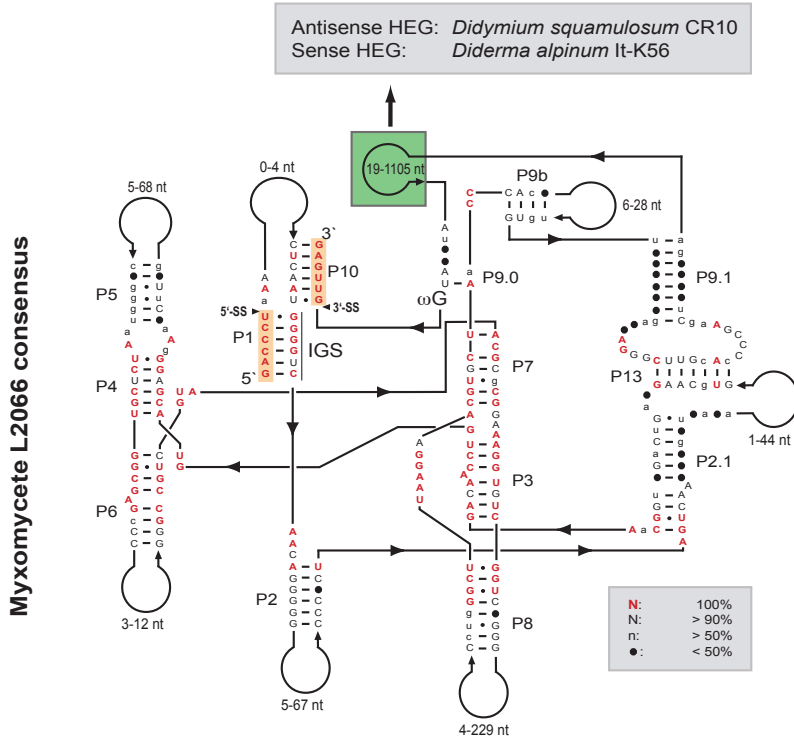
At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Supplementary Information

a



b

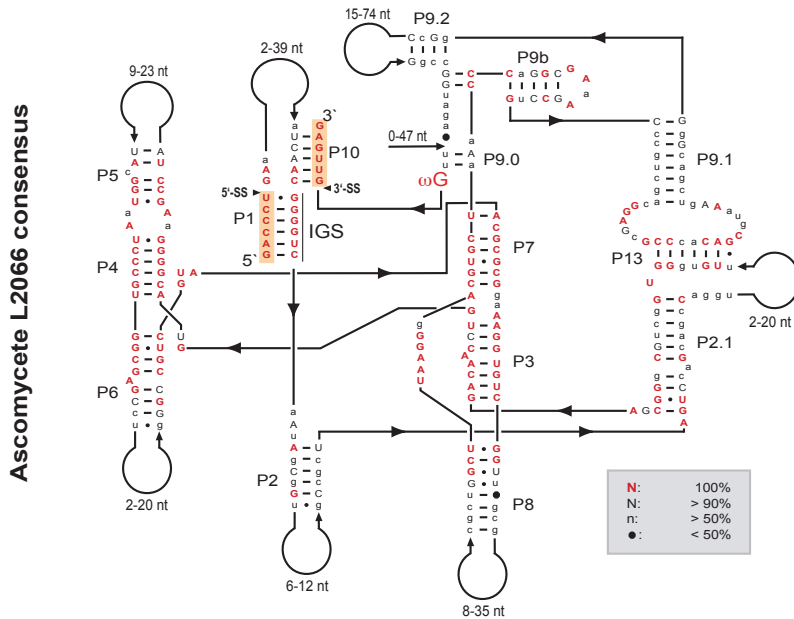


Figure S1

Figure S2: L2066_Consensus alignment (34 introns)

```
<<P1>> <<P10><<P1>> <<<<<<P2>>>>> <P2.1><<P13>><P2.1> <<<P3>>>
Cmi_HE655081 gaccctAAA//TTCGATGGGTC AACAGGGGA//TCCTCTAGTCAA/ GTGCAAG//TTGCCAGAGCAAACTGGT
Dat_HE655057 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCCTCTAGTCAC/ GTGCAAG//TTGCCAGAGCAAACTGGT
Dat_HE655058 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCCTCTAGTCAC/ GTGCAAG//TTGCCAGAGCAAACTGGT
Dme_HE655059 gaccctGAA//CTCAATGGGTC AACAGGGGG//CCCTCTAGTCAA/ GTGCAAG//CTGCCAAGAGCAAACTGGT
Dni_AM407429 gaccctGAA//CTCAATGGGTC AACAGGGGG//CCCTCTAGTCAA/ GTGCAAG//TTGCCAGAGCAAACTGGT
Dni_HE655060 gaccctGAA//CTCAACGGGTC AACAGGGGG//CCCTCTAGTCAA/ GTACAAG//GTGCCAAGAGCAAACTGGT
Dni_HE655061 gaccctGAA//CTCAACGGGTC AACAGGGGG//CCCTCTAGTCAA/ GTACAAG//GTGCCAAGAGCAAACTGGT
Dso_AM407428 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCTCTAGTCTGT/ GTACAAG//GGCCGAGAGCAAACTGGT
Dsq_AM407427 gaccctGAA//CTCAACGGGTC AACAGGGGG//CCCTCTAGTTGA/ CTGTGAG//CGCCGAGAGCAAACTGGT
Fse_HE655082 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCCCCTAGTCTGA/ GTGCAAG//CCGCCGAGAGCAAACTGGT
Lag_HE655062 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCCCCTAGTCAA/ ATGCAAG//CTGCCGAGAGCAAACTGGT
Lat_ON155995 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCCCCTAGTCAA/ GTGCAAG//TTGCCAAGAGCAAACTGGT
Lca_AM407430 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCCCCTAGTCAA/ ATGCAAG//CTGCCGAGAGCAAACTGGT
Lcr_HE655064 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCTCTAGTCAA/ GTACAAG//TGCCGAGAGCAAACTGGT
Mcr_HE655067 gaccctAAA//CTCAATGGGTC AACAGGGGG//CCTCTAGTTGA/ GTGCAAG//TCGCCAAGAGCAAACTGGT
Clf_ON155996 gaccctAAG//TTGCTAGGGTC AAAAAGGGCC//AGCCCTAGTCAA/ GTGCAAG//CTGCCGAGAGCAAACTGGT
Squ_GU214496 gaccctGAA//CTCAACGGGTC AAAAAGTAGT//GCTGCTAGTCTC/ TTGCAAG//CAGCCGAGAGCAAACTGGT
Bba_EU334679 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Bba_KJ701419 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Bba_KJ701420 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Bba_EU334676 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Bba_MG654725 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Bba_MG654726 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Bba_JF429894 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Cka_AB044639 gaccctGAA//CTCAACGGGTC GAAAACGGC//GCCGCTAGTCCA/ TTGCAAG//CGGCCGAGAGCAAACTGGT
Cmi_CP023322 gaccctGAA//ATCAACGGGTC CAATAGCGGT//GCCGCTAGTCCA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Cpr_AB044641 gaccctGAA//ATCAACGGGTC CAGTAGCGGT//GCCGCTAGTCCA/ CTGTGAG//CGGCCGAGAGCAAACTGGT
Fsp_ON155997 gaccctGAG//CTCAACGGGTC CAAACAATAGC//GCTATTAGTCCA/ TTACGGG//CGGCCGAGAGCAAACTGGT
Lsp_MH013330 gaccctGAA//ATCAACGGGTC CAAGAGCGGT//GCCGCTAGTCCA/ CTGCAAG//CGGCCGAGAGCAAACTGGT
Ost1_F3461354 gaccctGAG//CTCAACGGGTC CGAAGACGGC//GCCGCTAGTCCA/ CTGCAAG//CGGCCGAGAGCAAACTGGT
Ost1_F3461355 gaccctGAG//CTCAACGGGTC CGAAGACGGC//GCCGCTAGTCCA/ CTGCAAG//CGGCCGAGAGCAAACTGGT
Pte_AB044642 gaccctGAA//ATGACGGGTC CAAGAGCGGT//AGCCTCAGTCCG/ CTGTGAG//CGGCCGAGAGCAAACTGGA
Mmo_DQ518989 gaccctGAA//CATAACGGGTC AAAAAGCAGT//GCTGACTGCTA/ TTGTGAG//CGGCCGAGAGCAAACTGGT
Mni_DQ518993 gaccctGAA//ATCAACGGGTC CAATAGCAGT//ACTGCTAGTCTA/ TTGTGAG//CAGCCGAGAGCAAACTGGA
<<P1>> <<P10><<P1>> <<<<<<P2>>>>> <P2.1><<P13>><P2.1> <<<P3>>>
```

```
<<P4>> <<<P5>>> <<P4>><<<<<<<<P6>>>>>>> <<<P7>>> <<P3>>><<<<<<<<P8
Cmi_HE655081 ACGAGG GAGGC//TGCTA CTCGTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGGAG//
Dat_HE655057 ACGAGGAAAAT//GACAA CTCGTGGCGAGCCC//AGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Dat_HE655058 ACGAGGAAAAT//GACAA CTCGTGGCGAGCCC//AGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Dme_HE655059 ACGAGGAAAAC//GACAA CTCGTGGCGAGCCC//AGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Dni_AM407429 ACGAGGAAGGC//AGTAA CTCGTGGCGAGTCC//GGGCCGTGCTA ACGCAC GAAAGGTGTGGTTCGAG//
Dni_HE655060 ACGAGGAAGGC//AGTAA CTCGTGGCGAGTCC//GGGCCGTGCTA ACGCAC GAAAGGTGTGGTTCGAG//
Dni_HE655061 ACGAGGAAGGC//AGTAA CTCGTGGCGAGTCC//GGGCCGTGCTA ACGCAC GAAAGGTGTGGTTCGAG//
Dso_AM407428 ACGAGGGAAC//GGTAA CTCGTGGCGAGTCC//TGGCCGTGCTA ACGCAC GAAAGGTGTGGTTCGAG//
Dsq_AM407427 ACGAGGGAAC//GATAA CTCGTGGCGAGCCC//AGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Fse_HE655082 ACGGGGAGCC//GGGAA CTCGTGGCGAGTCA//AGGCCGTGCTA ACGCAC GAAAGGTGTGGTTCGAG//
Lag_HE655062 ACGAGGGAAGC//GCTAA CTCGTGGCGAGCCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Lat_ON155995 ACGAGGGAAGC//GGTAA CTCGTGGCGAGTCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Lca_AM407430 ACGAGGGAAGC//GGTAA CTCGTGGCGAGCCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Lcr_HE655064 ACGAGGAAGGC//AGTAA CTCGTGGCGAGCCC//GGGCCGTGCTA ACGCAC GAAAGGTGTGGTTCGAG//
Mcr_HE655067 ACGAGGGAAC//GACGA CTCGTGGCGAGTCC//GAGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Clf_ON155996 ACGAGGGAAGC//GCCTA CTCGTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Squ_GU214496 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_EU334679 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_KJ701419 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_KJ701420 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_EU334676 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_MG654725 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_MG654726 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Bba_JF429894 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Cka_AB044639 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Cmi_CP023322 ACGGGGAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Cpr_AB044641 ACGGGGAGCC//GGTAA CCGCTGGCGACCT//GAGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Fsp_ON155997 ACGGGGAGCC//GGTAA CCGCTGGCGAGTCC//GAGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Lsp_MH013330 ACGGGGAAGCC//GGTAA CCGCTGGCGAGCCT//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Ost1_F3461354 ACGGGGAGCC//GGTAA CCGCTGGCGAGCCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Ost1_F3461355 ACGGGGAGCC//GGTAA CCGCTGGCGAGTCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Pte_AB044642 ACGGGGAGCC//GGTAA CCGCTGGCGAGTCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Mmo_DQ518989 ACGGGGAGCC//GGTAA CCGCTGGCGAGTCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
Mni_DQ518993 ACGGGGAGCC//GGTAA CCGCTGGCGAGTCC//GGGCCGTGCTA ACGCGG GAAAGGTGTGGTTCGAG//
<<P4>> <<<P5>>> <<P4>><<<<<<<<P6>>>>>>> <<<P7>>> <<P3>>><<<<<<<<P8
```



```

>>>>>>> <<P7>> <<<<P9b>>> <<P13>> <P10>>
Cmi_HE655081 CTCTGGCTTAAGGAA<CGTGC>AA//CCAC//GTGGCTGG//<CTGCAC>//CCGG//TTGgttgag
Da1_HE655057 TCTAGGCTTAAGGAA<CGTAC>AA//CACA//TGTGCCTA//<CTGCAT>//TAGG//ATGgttgag
Da1_HE655058 TCTAGGCTTAAGGAA<CGTAC>AA//CACA//TGTGCCTA//<CTGCAT>//TAGG//ATGgttgag
Dme_HE655059 CCCGGGCTTAAGGAA<CGTAC>AA//CACA//TGTGCCTG//<CTGCAT>//TAGA//ATGgttgag
Dni_AM407429 CTCTGGGCTTAAGGAA<CGTGC>AT//CAGC//TGTGTTCT//<CTGTAC>//AGAA//AAAgttgag
Dni_HE655060 CTCTGGGCTTAAGGAA<CGTGC>AT//CAGC//TGTGTTCT//<CTGTAC>//AGAA//AAAgttgag
Dni_HE655061 CTCTGGGCTTAAGGAA<CGTGC>AT//CAGC//TGTGTTCT//<CTGTAC>//AGAA//AAAgttgag
Dso_AM407428 CCCCGGCTTAAGGAA<CGTGC>AA//CACA//TGTGCCTG//<CTGTAC>//CAGG//CTGgttgag
Dso_AM407427 CCTTGGCTTAAGGAA<CGTGC>AT//CCTG//CAGGTGCT//<CTCACAG>//ACAA//ATGgttgag
Fse_HE655082 AACGGGCTTAAGGAA<CGTGC>AA//TACA//TGTACTGG//<CTGCAC>//CTGG//TTGgttgag
Lag_HE655062 CCTGGGCTTAAGGAA<CGTGC>AA//CACA//TGTGCTG//<CTGCAT>//TGGA//ATGgttgag
La1_ON155995 CCTGGGCTTAAGGAA<CGTGC>AA//CAGC//CGTGTCA//<CTACAC>//TGAG//TTGgttgag
Lca_AM407430 CCTGGGCTTAAGGAA<CGTGC>AA//CACA//TGTGCTG//<CTGCAT>//TGGA//ATGgttgag
Lcr_HE655064 CCTGGGCTTAAGGAA<CGTGC>AT//CACA//TGTGTTCC//<CTGTAC>//GGAA//AAAgttgag
Mcr_HE655067 CCTAGGCTTAAGGAA<CGTAC>AA//CATG//CATGCCGG//<CTGCAC>//CCGA//TTGgttgag
Cl_a_ON155996 CCTAGGCTTAAGGAA<CGTGC>AT//CGTG//CACGCCTG//<CTGCAC>//CCGA//GAGgttgag
Squ_GU214496 GCCTGGCTTAAGGAA<CGTGC>AA//CATG//CATGTTCT//<CTGCAG>//AGGA//TTGgttgag
Bba_EU334679 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Bba_KJ701419 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Bba_KJ701420 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Bba_EU334676 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Bba_MG654725 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Bba_MG654726 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Bba_JF429894 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCTG//<CCACAG>//CGGG//TCGgttgag
Cka_AB044639 TCTCGGCTTAAGGTA<CGTGC>AA//CAGG//CCTGCCCG//<CTGCAG>//GGCG//ATGgttgag
Cmi_CP023322 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCCG//<CCACAG>//CGGG//ATGgttgag
Cpr_AB044641 CGCCGGCTTAAGGAA<CGTGC>AA//CTGG//CCAGCCCG//<CCCCAG>//CGGG//TTGgttgag
Fsp_ON155997 CCGGGGCTTAAGGTA<CGTGC>AA//CAGG//CCTGTGCC//<CCCGTAG>//GGCA//TTGgttgag
Lsp_MH013330 CGCTGGCTTAAGGAA<CGTGC>AA//CAGG//CCTGCCCG//<CCACAG>//CGGG//ATGgttgag
Osi_FJ461354 CCGGGGCTTAAGGTA<CGTGC>GA//CGGG//CCCGGCCCG//<CTGCAG>//GGCG//TTGgttgag
Osi_FJ461355 CCGGGGCTTAAGGTA<CGTGC>GA//CGGG//CCCGGCCCG//<CTGCAG>//GGCG//TTGgttgag
Pte_AB044642 CGCCGGCTTAAGGTA<CGTGC>AA//CAGG//CCTGCCCG//<CCACAG>//CGGG//ATGgttgag
Mme_EU518989 CCCCGGCTTAAGGTA<CGTGC>AA//CTGG//CCTGCCCT//<CCACAG>//AGGG//ATGgttgag
Mni_DQ518993 AATCAGCTTAAGGAA<CGTGC>AT//CAGG//CCTGCTGT//<CCACAG>//ACGG//ATGgttgag
>>>>>>> <<P7>> <<<<P9b>>> <<P13>> <P10>>

```


List of previously published theses for PhD in Aquaculture / PhD in Aquatic Biosciences / PhD in Biosciences, Nord University

No. 1 (2011)

PhD in Aquaculture

Chris André Johnsen

Flesh quality and growth of farmed Atlantic salmon (*Salmo salar* L.) in relation to feed, feeding, smolt type and season

ISBN: 978-82-93165-00-2

No. 2 (2012)

PhD in Aquaculture

Jareeporn Ruangsri

Characterization of antimicrobial peptides in Atlantic cod

ISBN: 978-82-93165-01-9

No. 3 (2012)

PhD in Aquaculture

Muhammad Naveed Yousaf

Characterization of the cardiac pacemaker and pathological responses to cardiac diseases in Atlantic salmon (*Salmo salar* L.)

ISBN: 978-82-93165-02-6

No. 4 (2012)

PhD in Aquaculture

Carlos Frederico Ceccon Lanes

Comparative Studies on the quality of eggs and larvae from broodstocks of farmed and wild Atlantic cod

ISBN: 978-82-93165-03-3

No. 5 (2012)

PhD in Aquaculture

Arvind Sundaram

Understanding the specificity of the innate immune response in teleosts: Characterisation and differential expression of teleost-specific Toll-like receptors and microRNAs

ISBN: 978-82-93165-04-0

No. 6 (2012)

PhD in Aquaculture

Teshome Tilahun Bizuayehu

Characterization of microRNA during early ontogeny and sexual development of Atlantic halibut (*Hippoglossus hippoglossus* L.)

ISBN: 978-82-93165-05-7

No. 7 (2013)

PhD in Aquaculture

Binoy Rajan

Proteomic characterization of Atlantic cod skin mucosa – Emphasis on innate immunity and lectins

ISBN: 978-82-93165-06-04

No. 8 (2013)

PhD in Aquaculture

Anusha Krishanthi Shyamali Dhanasiri

Transport related stress in zebrafish: physiological responses and bioremediation

ISBN: 978-82-93165-07-1

No. 9 (2013)

PhD in Aquaculture

Martin Haugmo Iversen

Stress and its impact on animal welfare during commercial production of Atlantic salmon (*Salmo salar* L.)

ISBN: 978-82-93165-08-8

No. 10 (2013)

PhD in Aquatic Biosciences

Alexander Jüterbock

Climate change impact on the seaweed *Fucus serratus*, a key foundational species on North Atlantic rocky shores

ISBN: 978-82-93165-09-5

No. 11 (2014)

PhD in Aquatic Biosciences

Amod Kulkarni

Responses in the gut of black tiger shrimp *Penaeus monodon* to oral vaccine candidates against white spot disease

ISBN: 978-82-93165-10-1

No. 12 (2014)

PhD in Aquatic Biosciences

Carlo C. Lazado

Molecular basis of daily rhythmicity in fast skeletal muscle of Atlantic cod (*Gadus morhua*)

ISBN: 978-82-93165-11-8

No. 13 (2014)

PhD in Aquaculture

Joanna Babiak

Induced masculinization of Atlantic halibut (*Hippoglossus hippoglossus* L.): towards the goal of all-female production

ISBN: 978-82-93165-12-5

No. 14 (2015)

PhD in Aquaculture

Cecilia Campos Vargas

Production of triploid Atlantic cod: A comparative study of muscle growth dynamics and gut morphology

ISBN: 978-82-93165-13-2

No. 15 (2015)

PhD in Aquatic Biosciences

Irina Smolina

Calanus in the North Atlantic: species identification, stress response, and population genetic structure

ISBN: 978-82-93165-14-9

No. 16 (2016)

PhD in Aquatic Biosciences

Lokesh Jeppinamogeru

Microbiota of Atlantic salmon (*Salmo salar* L.), during their early and adult life

ISBN: 978-82-93165-15-6

No. 17 (2017)

PhD in Aquatic Biosciences

Christopher Edward Presslauer

Comparative and functional analysis of microRNAs during zebrafish gonadal development

ISBN: 978-82-93165-16-3

No. 18 (2017)

PhD in Aquatic Biosciences

Marc Jürgen Silberberger

Spatial scales of benthic ecosystems in the sub-Arctic Lofoten-Vesterålen region

ISBN: 978-82-93165-17-0

No. 19 (2017)

PhD in Aquatic Biosciences

Marvin Choquet

Combining ecological and molecular approaches to redefine the baseline knowledge of the genus *Calanus* in the North Atlantic and the Arctic Oceans

ISBN: 978-82-93165-18-7

No. 20 (2017)

PhD in Aquatic Biosciences

Torvald B. Egeland

Reproduction in Arctic charr – timing and the need for speed

ISBN: 978-82-93165-19-4

No. 21 (2017)

PhD in Aquatic Biosciences

Marina Espinasse

Interannual variability in key zooplankton species in the North-East Atlantic: an analysis based on abundance and phenology

ISBN: 978-82-93165-20-0

No. 22 (2018)

PhD in Aquatic Biosciences

Kanchana Bandara

Diel and seasonal vertical migrations of high-latitude zooplankton: knowledge gaps and a high-resolution bridge

ISBN: 978-82-93165-21-7

No. 23 (2018)

PhD in Aquatic Biosciences

Deepti Manjari Patel

Characterization of skin immune and stress factors of lumpfish, *Cyclopterus lumpus*

ISBN: 978-82-93165-21-7

No. 24 (2018)

PhD in Aquatic Biosciences

Prabhugouda Siriyappagoudar

The intestinal mycobiota of zebrafish – community profiling and exploration of the impact of yeast exposure early in life

ISBN: 978-82-93165-23-1

No. 25 (2018)

PhD in Aquatic Biosciences

Tor Erik Jørgensen

Molecular and evolutionary characterization of the Atlantic cod mitochondrial genome

ISBN: 978-82-93165-24-8

No. 26 (2018)

PhD in Aquatic Biosciences

Yangyang Gong

Microalgae as feed ingredients for Atlantic salmon

ISBN: 978-82-93165-25-5

No. 27 (2018)

PhD in Aquatic Biosciences

Ove Nicolaisen

Approaches to optimize marine larvae production

ISBN: 978-82-93165-26-2

No. 28 (2019)

PhD in Aquatic Biosciences

Qirui Zhang

The effect of embryonic incubation temperature on the immune response of larval and adult zebrafish (*Danio rerio*)

ISBN: 978-82-93165-27-9

No. 29 (2019)

PhD in Aquatic Biosciences

Andrea Bozman

The structuring effects of light on the deep-water scyphozoan *Periphylla periphylla*

ISBN: 978-82-93165-28-6

No. 30 (2019)

PhD in Aquatic Biosciences

Helene Rønquist Knutsen

Growth and development of juvenile spotted wolffish (*Anarhichas minor*) fed microalgae incorporated diets

ISBN: 978-82-93165-29-3

No. 31 (2019)

PhD in Aquatic Biosciences

Shruti Gupta

Feed additives elicit changes in the structure of the intestinal bacterial community of Atlantic salmon

ISBN: 978-82-93165-30-9

No. 32 (2019)

PhD in Aquatic Biosciences

Peter Simon Claus Schulze

Phototrophic microalgal cultivation in cold and light-limited environments

ISBN: 978-82-93165-31-6

No. 33 (2019)

PhD in Aquatic Biosciences

Maja Karoline Viddal Hatlebakk

New insights into *Calanus glacialis* and *C. finmarchicus* distribution, life histories and physiology in high-latitude seas

ISBN: 978-82-93165-32-3

No. 34 (2019)

PhD in Aquatic Biosciences

Arseny Dubin

Exploration of an anglerfish genome

ISBN: 978-82-93165-33-0

No. 35 (2020)

PhD in Aquatic Biosciences

Florence Chandima Perera Willora Arachchilage

The potential of plant ingredients in diets of juvenile lumpfish (*Cyclopterus lumpus*)

ISBN: 978-82-93165-35-4

No. 36 (2020)

PhD in Aquatic Biosciences

Ioannis Konstantinidis

DNA hydroxymethylation and improved growth of Nile tilapia (*Oreochromis niloticus*) during domestication

ISBN: 978-82-93165-36-1

No. 37 (2021)

PhD in Aquatic Biosciences

Youngjin Park

Transcriptomic and cellular studies on the intestine of Atlantic salmon

Discovering intestinal macrophages using omic tools

ISBN: 978-82-93165-34-7

No. 38 (2021)

PhD in Aquatic Biosciences

Purushothaman Kathiresan

Proteomics of early embryonic development of zebrafish (*Danio rerio*)

ISBN: 978-82-93165-37-8

No. 39 (2021)

PhD in Aquatic Biosciences

Valentin Kokarev

Macrobenthic communities of sub-Arctic deep fjords: composition, spatial patterns and community assembly

ISBN: 978-82-93165-38-5

No. 40 (2021)

PhD in Aquatic Biosciences

Aurélien Delaval

Population genomics of a critically endangered data-deficient elasmobranch, the blue skate *Dipturus batis*

ISBN: 978-82-93165-39-2

No. 41 (2021)

PhD in Aquatic Biosciences

Isabel Sofía Abihssira García

Environmental impact of microplastics in relation to Atlantic salmon farming

ISBN: 978-82-93165-40-8

No. 42 (2022)

PhD in Aquatic Biosciences

Yousri Abdelmutalab Ahmed Abdelhafiz

Insights into the bacterial communities of Nile tilapia – core members and intergenerational transfer

ISBN: 978-82-93165-41-5

No. 43 (2022)

PhD in Aquatic Biosciences

Fredrik Ribsskog Staven

Interaction studies on lumpfish exposed to Atlantic salmon: behavioural observations and the underlying physiological and neurobiological mechanisms

ISBN: 978-82-93165-42-2

No. 44 (2022)

PhD in Aquatic Biosciences

Solveig Lysfjord Sørensen

Influence of feed ingredients and additives on mucosal health with focus on the intestine of Atlantic salmon (*Salmo salar*)

ISBN: 978-82-93165-43-9

No. 45 (2022)

PhD in Aquatic Biosciences

Apollo Marco Dalonos Lizano

Examining challenges in species-level taxonomy among *Calanus* copepods in the Northern seas using genome and transcriptome data

ISBN: 978-82-93165-44-6

No. 46 (2022)

PhD in Aquatic Biosciences

Sowmya Ramachandran

Ribosomal RNA, ribose methylation, and box C/D snoRNAs during embryonic development of teleosts zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*)

ISBN: 978-82-93165-45-3

No. 47 (2023)

PhD in Aquatic Biosciences

William John Hatchett

The brown algal genus *Fucus*: A unique insight into reproduction and the evolution of sex-biased genes

ISBN: 978-82-93165-46-0

No. 48 (2023)

PhD in Aquatic Biosciences

Ying Yen

Application of ensiled *Saccharina latissima* and *Alaria esculenta* as feed: ensilibility, digestibility and bioactivity

ISBN: 978-82-93165-47-7

No. 49 (2023)

PhD in Aquatic Biosciences

Deepak Pandey

Marine macroalgae as an alternative, environment-friendly, and bioactive feeding resource for animals

ISBN: 978-82-93165-48-4

No. 50 (2023)

PhD in Aquatic Biosciences

Nimalan Nadanasabesan

Dietary approaches to improve mucosal health of Atlantic salmon (*Salmo salar*)

ISBN: 978-82-93165-49-1

No. 51 (2023)

PhD in Aquatic Biosciences

Likith Reddy Pinninti

Biosystematics and evolutionary genomics of deep-sea fish (lumpsuckers, snailfishes, and sculpins) (Perciformes: Cottoidei)

ISBN: 978-82-93165-50-7

No. 52 (2023)

PhD in Biosciences

Mie Prik Arnberg

Directed endozoochory: a hitchhiker's guide to successful sexual reproduction in clonal ericaceous plants

ISBN: 978-82-93165-51-4

No. 53 (2023)

PhD in Aquatic Biosciences

Adnan Hussain Gora

Insights from a zebrafish model to combat dyslipidemia using microbe-derived bioactive compounds

ISBN: 978-82-93165-52-1

No. 54 (2023)

PhD in Aquatic Biosciences

Saima Rehman

Diet-induced inflammation in zebrafish and its alleviation by functional oligo- and polysaccharides

ISBN: 978-82-93165-53-8

No. 55 (2023)

PhD in Biosciences

Clara Isabel Wagner

Exploring the nuclear genome of the spiny dogfish (*Squalus acanthias*)

ISBN: 978-82-93165-54-5

No. 56 (2023)

PhD in Aquatic Biosciences

Miirõ Ilmari Virtanen

Internal tagging in Atlantic salmon (*Salmo salar L.*): A study of welfare, wounds, and stress

ISBN: 978-82-93165-55-2

No. 57 (2023)

Md Golam Rbbani

PhD in Biosciences

Characterization of Circular RNAs in Nile Tilapia Muscle and Thermal Modulation of their Expression in Relation to Growth

ISBN: 978-82-93165-56-9

No. 58 (2023)

Èric Jordà Molina

PhD in Aquatic Biosciences

Spatio-temporal dynamics of soft-bottom macrobenthic communities in a rapidly changing Arctic: a case study of the Northwestern Barents Sea

ISBN: 978-82-93165-57-6

No. 59 (2024)

Shubham Varshney

PhD in Aquatic Biosciences

From nanoplastics to chemical pollutants: Exploring mixture toxicity in zebrafish

ISBN: 978-82-93165-58-3

No. 60 (2024)

PhD in Aquatic Biosciences

Charles Patrick Lavin

The cumulative impacts of climate change and fishing on marine communities

ISBN: 978-82-93165-59-0

No. 61 (2024)

PhD in Biosciences

Dhurba Adhikari

Evolutionary genomics of pelvic spine reduction in *Gasterosteus aculeatus*

ISBN: 978-82-93165-60-6

Nuclear group I introns, characterized by unique and complex RNA structures that encode self-splicing ribozymes, are abundant in myxomycetes (plasmodial slime molds). This research investigated the evolutionary aspects of these catalytic RNA introns, focusing on RNA sequence conservation and structural variability. We conducted phylogenetic assessments of three nuclear group I intron families (S516, S1389, and L2066), mainly from myxomycetes ribosomal DNA (rDNA). The results revealed that each family has a conserved ribozyme core, highly variable peripheral segments, distinct RNA sequence variations, and complex inheritance patterns. These findings led to the proposal of a revision of the Goddard-Burt cyclic evolutionary model for nuclear group I introns. This research expands our understanding of the molecular evolution of catalytic RNA introns, the homing endonuclease gene-spliceosomal introns and the regulation of homing endonuclease mRNA expression within an rDNA context. Overall, this work supplements the field of catalytic RNA intron evolution by providing an in-depth insight into the unique features and evolutionary dynamics of nuclear group I intron families in myxomycetes. This research also provides additional insight that may broaden our understanding of RNA sequence variations and evolution, which again can be helpful in advancing new biotechnological methods, developing RNA biomarkers, and the continuation of the development of RNA-based medical treatments.