The highly variable mitochondrial small-subunit ribosomal RNA gene of *Ophiostoma minus*

Mohamed HAFEZ\textsuperscript{a,b}, Georg HAUSNER\textsuperscript{a,*}

\textsuperscript{a}Department of Microbiology, University of Manitoba, Winnipeg, MB, R3T 2N2 Canada
\textsuperscript{b}Department of Botany, Faculty of Science, Suez Canal University, Suez, Egypt

\textbf{A R T I C L E  I N F O}

Article history:
Received 25 October 2010
Received in revised form 28 May 2011
Accepted 15 July 2011
Available online 23 July 2011

Corresponding Editor: Brenda Diana Wingfield

Keywords:
Blue-stain fungi
Group I and II introns
Homing endonucleases
Reverse transcriptase

\textbf{A B S T R A C T}

Mitochondrial genomes in the true fungi are highly variable both in size and organization. Most of this size variation is due to the presence of introns and intron-encoded open reading frames (ORFs). The objectives for this work were to examine the mitochondrial small-subunit ribosomal RNA (\textit{rns}) gene of \textit{Ophiostoma minus} for the presence of introns and to characterize such introns and their encoded ORFs. DNA sequence analysis showed that among different strains of \textit{O. minus} various \textit{rns} gene exon/intron configurations can be observed. Based on comparative sequence analysis and RNA secondary structure modeling group I introns with LAGLIDADG ORFs were uncovered at positions mS569 and mS1224 and group II introns were present at positions mS379 and mS952. The mS379 group II intron encoded a fragmented reverse transcriptase (RT)-like ORF and the mS952 group II intron encoded a LAGLIDADG-type ORF. Examples of intron ORF degeneration due to frameshift mutations were observed. The mS379 group II intron is the first mitochondrial group II intron to have an ORF inserted within domain II, typically RT-like ORFs are inserted in domain IV. The evolutionary dynamics of the intron-encoded ORFs have also been examined.

\textcopyright 2011 British Mycological Society. Published by Elsevier Ltd. All rights reserved.

\textbf{Introduction}

Group I and group II introns are potentially self-splicing introns that are frequently encountered within fungal mitochondrial DNA (mtDNA) genomes and these introns can be differentiated by their splicing mechanisms and secondary and tertiary RNA structures (Michel & Westhof 1990; Michel & Ferat 1995; Bonen & Vogel 2001; Fedorova & Zingler 2007). Both types of introns are considered ribozymes (Saldanha et al. 1993; Belfort et al. 2002), however splicing is assisted by intron and host genome encoded maturases/splicing factors (reviewed in Hausner 2003). Maturases are thought to stabilize the correct tertiary structure of the intron RNA to allow for the proper catalytic domains to be in contact with each other (Matsuura et al. 2001; Noah & Lambowitz 2003). Examples of host splicing cofactors are DEAD-box proteins and aminoacyl-tRNA synthetases (Halls et al. 2007; Bifano & Caprara 2008; Paulkstelis et al. 2008). The group I intron-encoded maturases are derived from intron-encoded homing endonucleases (HEs) or maturase activity can be part of some bi-functional intron-encoded HEs (Chatterjee et al. 2003; Belfort 2003). Typically group II introns encode multifunctional reverse transcriptase (RT)-like proteins that contain a maturase domain (Lambowitz & Zimmerly 2004).

Group I and group II introns are considered mobile introns as they encode proteins that facilitate mobility of these elements from an intron containing allele to a cognate allele...
lacking introns by a process referred to as homing for group I introns (Dujon 1989) or retrohoming for group II introns (Zimmerly et al. 1995a, 1995b; Eickbush 1999). In some instances the mobile introns insert into new positions by transposition events usually mediated by the intron-encoded proteins or by reverse splicing of the intron RNAs (Woodson & Cech 1989).

HEs are encoded by homing endonuclease genes (HEGs) which are embedded within group I introns, group II introns, and archael introns, as well as inteins (Stoddard 2006). HEs are named based on conserved amino acid motifs. The LAGLIDADG and GIY-YIG families of HEs are most frequently encountered among fungal mitochondrial group I introns (Stoddard 2006). Group II introns have been noted to encode HNH type HEs or in a few instances LAGLIDADG-type HEs (Michel & Ferat 1995; Toor & Zimmerly 2002; Mullineux et al. 2010). HEGs themselves can be mobile elements moving independently from their ribozyme partners (Mota & Collins 1988; Sellem & Belcour 1997). Self-splicing introns and HEGs share a mutualistic relationship, where the intron provides a neutral location for the HEG thus minimizing its effect on the host genome and the HEG provides a means of mobility and dispersal for the intron (Goddard & Burt 1999; Lambowitz et al. 1999; Belfort et al. 2002; Schäfer 2003; Stoddard 2006).

HEGs and introns are quite invasive and contribute toward the size of fungal mtDNA genomes, mtDNA polymorphisms, and they promote mtDNA rearrangements (Dujon 1989; Charter et al. 1996; Belcour et al. 1997; Salvo et al. 1998; Hamari et al. 1999; Gobbi et al. 2003). Also group I and II introns have been associated with mtDNA instabilities such as generating plasmid-like elements that are found in senescent and/or hypovirulent strains in an assortment of filamentous fungi such as Podospora anserina (Osiewacz & Esser 1984; Michel & Cummings 1985; Cummings et al. 1986, 1990; Dujon & Belcour 1989), Ophiostoma novo-ulmi (Abu-Amero et al. 1995; Sethuraman et al. 2008), and Cryphonectria parasitica (Monteiro-Vitorello et al. 2009; Baidyaroy et al. 2011a, 2011b).

In this paper we present a detailed description of the Ophiostoma minus (Hedgcoc) Sydow et P. Sydow rns gene (=mtSSU-rRNA gene) and the introns associated with this gene. Species of Ophiostoma are of interest as this genus includes many insect-vascular forest pathogens and so-called blue-stain fungi (Harrington 1993; Kirsits 2004; Hausner et al. 2005). These fungi cause economic losses by staining timber and making it less desirable for export even facing trade embargoes or for high-end usage. Ophiostoma minus is a well known agent of blue-stain (Gorton & Webber 2000) and has been demonstrated to be a pathogen of pine (Masuya et al. 2003; Gorton et al. 2004; Benjamaa et al. 2007). There is also increased interest in mtDNA encoded ribozymes and HEs for biotechnology and human therapeutic applications (Lambowitz & Zimmerly 2004; Stoddard 2006; Marcaida et al. 2010). Overall our objectives for this work were as follows: (1) document the distribution of the rns introns in strains of O. minus, (2) examine among selected strains evidence for intron open reading frame (ORF) degeneration and, (3) characterize the introns and their ORFs by comparative sequence analysis. Ultimately this work is part off a long term effort to characterize and understand the composition and evolution of the mitochondrial genomes within the ophiostomatoid fungi (Gibb & Hausner 2005; Sethuraman et al. 2008, 2009a, 2009b).

### Materials and methods

#### DNA extraction and amplification protocols

The fungi used in this study and their sources are listed in Table 1. Strains were cultured in Petri plates containing 2 % Malt Extract Agar (MEA) supplemented with 1 g l⁻¹ yeast extract (YE) and 20 g l⁻¹ bacteriological agar. Mycelium for DNA extraction was generated as described in Hausner et al. (1992). The rns gene was amplified as previously described (Mullineux et al. 2010) with the primer mtr-1 and mtr-2 (see Suppl. Table 1 for a complete list of the primers used in the current study) using standard DNA amplification protocols utilizing the Invitrogen PCR system (Invitrogen, Burlington, Canada). The PCR conditions for the mtr-1 and mtr-2 primers were as follows: an initial denaturation at 93 °C for 1 min followed by 25 cycles of denaturation (93 °C for 1 min), annealing (55 °C for 1 min) and extension (70 °C for 3 min) followed by a final extension step at 70 °C for 10 min.

All PCR amplicons were analyzed by gel electrophoresis through 1 % agarose gels in TBE buffer (89 mM Tris-borate, 10 mM EDTA, pH 8.0). The sizes of the PCR amplicons were estimated against a 1 kb plus ladder and visualized under UV light after staining with ethidium bromide.

### Sequencing and cloning

DNA sequencing templates were prepared by purifying PCR products with the Wizard SV Gel and PCR Clean-Up system (Promega, Madison, USA). The double stranded DNA fragments were sequenced using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturers instructions. The sequencing products were denatured and resolved on a 3130 genetic analyzer (Applied Biosystems).

The rns derived PCR products were highly variable in size and the longer amplicons yielded poor sequencing results thus many of the rns derived PCR products were cloned. The amplons were cloned in pCR4 TOPO vector and transformed into Escherichia coli (DHS+) using the TOPO TA Cloning™ kit (Invitrogen), this not only improved sequencing efficiency but by using vector-specific primers this also allowed as to obtain sequences right up to the primers used to amplify the rns segments. Recombinant plasmids from positive clones were purified with the Wizard™ Plus Miniprep DNA purification system (Promega). Initially vector based primers as supplied by the TOPO cloning kit: M13 Forward, M13 Reverse, T7 (forward), and T3 (reverse) were used to obtain sequences, thereafter primers were designed based on the sequences obtained during this study as needed to complete all sequences in both directions (Suppl. Table 1).

### Sequence alignments and analysis

Individual sequences were compiled and assembled manually into contigs using the GeneDoc program v2.5.010 (Nicholas et al. 1997). The initial nucleotide sequence alignments were done with the Clustal-X program (Thompson et al. 1997) and the resulting alignments were refined by eye with the GeneDoc.
program. The ORF finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html; Genetic code setting for molds #4) was used to search for potential ORFs within the rns introns.

The online resource Basic Local Alignment Search Tool (BLAST: http://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al. 1990) was used to retrieve amino acid sequences (BLASTp) from GenBank which shared similarities to those encoded by the intron ORFs. The amino acid sequence data sets were aligned with the online multiple sequence alignment program PRALINE (Simossis & Heringa 2005; http://www.ibi.vu.nl/programs/pralinetm/). The alignments generated by PRALINE were examined and adjusted by eye with the GeneDoc program.

### Intron nomenclature and secondary structure modeling

For naming introns we followed the nomenclature proposed by Johansen & Haugen (2001), and intron insertion sites are based on corresponding nucleotide positions within the Escherichia coli SSU-rRNA sequence (GenBank accession AB035922). So for example based on this system the rns intron name: O.mi472-mS379 is based on O.mi to indicate Ophiostoma minus and the 472 is the strain number, the mS stands for the mitochondrial SSU-rRNA gene and 379 refers to the insertion site with respect to the E. coli SSU-rRNA sequence.

The secondary structures of the O. minus rns introns were predicted following the conventions for group I introns (Burke et al. 1987; Michel & Westhof 1990; Li & Zhang 2005) and group II introns (Michel & Ferat 1995; Toor et al. 2001). The online program mfold (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi; Zuker 2003) and the web server RNAweasel (http://megasun.bch.umontreal.ca/RNAweasel/; Gautheret & Lambert 2001; Lang et al. 2007) were used to identify some of the key stem-loop structural elements within the introns.

### RNA extraction and Reverse Transcription PCR (RT-PCR) analysis

The exon/intron junctions for rns introns were determined for the following strains: WIN(M)371 (for introns O.mi371-mS569 and O.mi371-mS952), WIN(M)472 (for introns O.mi472-mS379 and O.mi472-mS569), WIN(M)494 (for intron O.mi494-mS379), and WIN(M)888 (for introns O.mi888-mS379, O.mi888-mS952 and O.mi888-mS1224). For RNA extraction, cultures were grown in 125 ml conical flasks containing 50 ml Peptone Yeast-extract Glucose (PYG) liquid medium for up to 7 d at 20 °C. Fungal biomass was collected by vacuum filtration using Whatman number 1 filter paper and the wet mycelium was flash-frozen in liquid nitrogen and ground to a fine powder using a precooled mortar and pestle. The RNA was extracted using the RNeasy™ mini kit (QIAGEN, Mississauga, ON, Canada) following the protocol outlined by the manufacturer. In addition to the DNase step within RNeasy™ mini kit procedure the TURBO™ DNase kit (Ambion, Austin, TX, USA) was also applied according to the manufacturer’s protocol to ensure the removal of all DNA from the final RNA preparation. The TURBO™ DNase was inactivated by heating at 75 °C for 10 min. The Thermoscript™ RT-PCR system (Invitrogen) was used to synthesize cDNA using approximately 100 ng of template RNA. First strand synthesis was carried out with 40 pmol of primer mtsr-2 and

### Table 1 – List of strains used in the present study, area of isolation and the corresponding size of mtsr-1/2 amplicons.

<table>
<thead>
<tr>
<th>S</th>
<th>Species</th>
<th>Strain no.</th>
<th>Geographic origin</th>
<th>Size of mtsr-1/2 amplicons (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ophiostoma minus</td>
<td>WIN(M)292</td>
<td>Sandlunds, Forest Reserve, Manitoba, Canada</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>O. minus</td>
<td>WIM(M)371/UAMH 9805</td>
<td>Sandlands, Forest Reserve, Manitoba, Canada</td>
<td>4.4</td>
</tr>
<tr>
<td>3</td>
<td>O. minus</td>
<td>WIM(M)472</td>
<td>Taylor Lake, Alberta, Canada</td>
<td>5.4</td>
</tr>
<tr>
<td>4</td>
<td>O. minus</td>
<td>WIM(M)494/C-248 (NFRC)</td>
<td>Edmonton, Alberta, Canada</td>
<td>4.0</td>
</tr>
<tr>
<td>5</td>
<td>O. minus</td>
<td>WIM(M)495/C-262 (NFRC)</td>
<td>Edmonton, Alberta, Canada</td>
<td>4.4</td>
</tr>
<tr>
<td>6</td>
<td>O. minus</td>
<td>WIM(M)515/CBS 404.77</td>
<td>California, USA</td>
<td>7.0</td>
</tr>
<tr>
<td>7</td>
<td>O. minus</td>
<td>WIM(M)861/DAOM 29251/UAMH 9594</td>
<td>Toronto, Ontario, Canada</td>
<td>3.0</td>
</tr>
<tr>
<td>8</td>
<td>O. minus</td>
<td>WIM(M)871/DAOM 29251</td>
<td>Sandlands, Manitoba, Canada</td>
<td>4.4</td>
</tr>
<tr>
<td>9</td>
<td>O. minus</td>
<td>WIM(M)873</td>
<td>Edmonton, Alberta, Canada</td>
<td>3.0</td>
</tr>
<tr>
<td>10</td>
<td>O. minus</td>
<td>WIM(M)874/C-874 (NFRC)</td>
<td>Edmonton, Alberta, Canada</td>
<td>4.4</td>
</tr>
<tr>
<td>11</td>
<td>O. minus</td>
<td>WIM(M)875/C-845 (NFRC)</td>
<td>Edmonton, Alberta, Canada</td>
<td>3.0</td>
</tr>
<tr>
<td>12</td>
<td>O. minus</td>
<td>WIM(M)876/C-342 (NFRC)</td>
<td>Edmonton, Alberta, Canada</td>
<td>4.0</td>
</tr>
<tr>
<td>13</td>
<td>O. minus</td>
<td>WIM(M)888/ATCC 15321</td>
<td>Wyoming, Michigan, USA</td>
<td>7.0</td>
</tr>
<tr>
<td>14</td>
<td>O. minus</td>
<td>WIM(M)888A/ATCC 15321</td>
<td>Wyoming, Michigan, USA</td>
<td>7.0</td>
</tr>
<tr>
<td>15</td>
<td>O. minus</td>
<td>WIM(M)889/ATCC 22388</td>
<td>California, USA</td>
<td>7.0</td>
</tr>
<tr>
<td>16</td>
<td>O. minus</td>
<td>WIM(M)890/ATCC 11063</td>
<td>Washington, USA</td>
<td>4.0</td>
</tr>
<tr>
<td>17</td>
<td>O. minus</td>
<td>WIM(M)1213/UAMH 10159</td>
<td>Bracebridge, Ontario, Canada</td>
<td>5.4</td>
</tr>
<tr>
<td>18</td>
<td>O. minus</td>
<td>WIM(M)1275</td>
<td>Barrie, Ontario, Canada</td>
<td>4.0</td>
</tr>
<tr>
<td>19</td>
<td>O. minus</td>
<td>WIM(M)1573</td>
<td>Lakehead, Ontario, Canada</td>
<td>4.0</td>
</tr>
<tr>
<td>20</td>
<td>O. minus</td>
<td>WIM(M)1574</td>
<td>Lakehead, Ontario, Canada</td>
<td>1.2</td>
</tr>
<tr>
<td>21</td>
<td>O. minus</td>
<td>WIM(M)1575</td>
<td>Lakehead, Ontario, Canada</td>
<td>4.4</td>
</tr>
</tbody>
</table>

a WIN(M) = University of Manitoba culture collection (Winnipeg, MB, Canada); UAMH = University of Alberta Microfungus Collection & Herbarium, Devonian Botanic Garden, Edmonton, AB, Canada T6G 2E1; ATCC = American Type Culture Collection (Manassas, VA); CBS = Central Bureau voor Schimmelcultures (Utrecht, The Netherlands); NFRC = Northern Forestry Research Centre (Edmonton, AB, Canada); DAOM = Plant Research Institute, Department of Agriculture, Mycology, Ottawa, Canada.
subsequent PCR amplification was carried out under the same PCR conditions described previously for amplifying the \textit{rns} gene. The PCR products generated by the RT-PCR reactions were sequenced as described previously and the cDNA sequences were compared with the original \textit{rns} sequences in GeneDoc to establish the exon/intron junctions.

**Phylogenetic analysis**

Phylogenetic trees were generated for all LAGLIDADG ORF and group II intron RT ORF amino acid sequence alignments by distance, parsimony, and Bayesian methods. For distance and parsimony methods programs contained within the PHYLIB package were used (Felsenstein 2006). Phylogenetic trees were generated by distance methods by first generating distance matrices PROTDIST (JTT setting) for amino acid sequence alignments. The resulting distance matrices were used by the NEIGHBOUR program to generate Neighbor-joining (NJ) trees. In order to evaluate support for the nodes observed in the NJ phylogenetic trees bootstrap (BS) analysis (Felsenstein 1985) was conducted by generating 1000 BS replicates by SEQBOOT, the BS replicates were analyzed by PROTDIST (JTT setting), the resulting matrices were analyzed by NEIGHBOUR and resulting tree files were analyzed by the CONSENS program in order to obtain a majority rule consensus trees. For parsimony analysis phylogenetic trees were obtained with the PROTPARS program, and when combined with BS analysis the pseudo-replicate data set (1000 replicates, SEQBOOT) was directly analyzed with PROTPARS and the resulting tree file was analyzed with the CONSENS program in order to obtain the majority rule consensus trees.

The MrBayes program (v3.1) was used for Bayesian analysis (Ronquist & Huelsenbeck 2003; Ronquist 2004) and the parameters for amino acid alignments were as follows: mixed models and gamma distribution with four gamma rate parameters. The models used for generating the final 50 % majority rule trees were estimated by the program itself: RT-Blosum for the RT amino acid data set and Cprev for the LAGLIDADG data set these models were consisted with the models recommended by ProtTest (Abascal et al. 2005). However the top model recommended by ProtTest was JTT for both the LAGLIDADG ORF and group II RT data sets, therefore the final phylogenetic trees were generated using the JTT model. The Bayesian inference of phylogenies was initiated from a random starting tree and four chains were run simultaneously for 1,000,000 generations; trees were sampled every 100 generations. The first 25 % of trees generated were discarded (‘burn-in’) and the remaining trees were used to compute the posterior probability (PP) values. Phylogenetic trees were drawn with the TreeView program (Page 1996) using PHYLIB and/or MrBayes tree outfiles. When appropriate the intron type and insertion sites were plotted onto the LAGLIDADG ORF tree.

**Results**

**The \textit{rns} introns in strains of Ophiostoma minus**

Amplifying a segment of the \textit{rns} gene from 21 strains of \textit{O. minus} using the mtsr-1 and mtsr-2 primers generated PCR products ranging in size from 1.2 kb to 7.0 kb (Table 1). A 1.2 kb fragment is expected when no insertions are present, so the 3.0, 4.0, 4.4, 5.4, and 7.0 kb amplicons are indicative for the presence of introns. Representatives of the various amplicon size classes [WIN(M)1574, WIN(M)873, WIN(M)494, WIN(M)371, WIN(M)472, and WIN(M)515] were selected for detailed DNA sequence analysis (GenBank accession numbers are listed in Suppl. Table 2). Based on comparing \textit{rns} sequences that lacked introns with those that do, four intron

---

**Fig 1** — [A] A schematic overview of the \textit{O. minus} \textit{rns} gene showing the various intron types (group I and II) and the intron-encoded ORFs (RT; LHE). The introns are located at positions mS379, mS569, mS952, and mS1224 with reference to the \textit{E. coli} SSU-rRNA sequence. The mS379 intron-encoded RT ORF is fragmented into two segments due to frameshift mutations. Examples of strains are listed for which the complete mtsr-1/2 PCR product sequences were obtained. [B] Sequences that define the exon/intron junctions for the group I (O.mi-mS569 and O.mi-mS1224) and group II introns (O.mi-mS379 and O.mi-mS952). For the group II introns the IBS1, IBS2, IBS3, and IBS4 sequences are indicated. The symbol X denotes the insertion site for the intron.
Table 2 – Names and features of the *O. minus* mtDNA *rns* introns characterized during this study.

<table>
<thead>
<tr>
<th>Intron name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>O. minus strain no.</th>
<th>Insertion site (with respect to <em>E. coli</em>)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Intron class</th>
<th>Intron length (nucleotide)</th>
<th>ORF type</th>
<th>ORF length (amino acids)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O.mi472-mS379</td>
<td>WIN(M)472</td>
<td>379–380</td>
<td>IIA</td>
<td>2858</td>
<td>RT&lt;sup&gt;b&lt;/sup&gt;</td>
<td>616&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>O.mi494-mS379</td>
<td>WIN(M)494</td>
<td>379–380</td>
<td>IIA</td>
<td>2859</td>
<td>RT&lt;sup&gt;b&lt;/sup&gt;</td>
<td>616&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>O.mi515-mS379</td>
<td>WIN(M)515</td>
<td>379–380</td>
<td>IIA</td>
<td>2858</td>
<td>RT&lt;sup&gt;b&lt;/sup&gt;</td>
<td>616&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>O.mi371-mS569</td>
<td>WIN(M)371</td>
<td>569–570</td>
<td>IC2</td>
<td>1323</td>
<td>LHE</td>
<td>370</td>
</tr>
<tr>
<td>O.mi472-mS569</td>
<td>WIN(M)472</td>
<td>569–570</td>
<td>IC2</td>
<td>1323</td>
<td>LHE</td>
<td>370</td>
</tr>
<tr>
<td>O.mi371-mS952</td>
<td>WIN(M)371</td>
<td>952–953</td>
<td>IIB</td>
<td>1867</td>
<td>LHE</td>
<td>306</td>
</tr>
<tr>
<td>O.mi515-mS952</td>
<td>WIN(M)515</td>
<td>952–953</td>
<td>IIB</td>
<td>1870</td>
<td>LHE</td>
<td>306</td>
</tr>
<tr>
<td>O.mi873-mS952</td>
<td>WIN(M)873</td>
<td>952–953</td>
<td>IIB</td>
<td>1867</td>
<td>LHE</td>
<td>306</td>
</tr>
<tr>
<td>O.mi515-mS1224</td>
<td>WIN(M)515</td>
<td>1224–1225</td>
<td>IC2</td>
<td>1394</td>
<td>LHE&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup> Intron naming based on the nomenclature proposed by Johansen & Haugen (2001).
<sup>b</sup> ORF type based on BLASTp analysis showing similarities to members of group II intron-encoded RT-like/maturase superfamily.
<sup>c</sup> Intron contains highly degenerated/fragmented LHE ORF due to frameshift mutations.
<sup>d</sup> ORFs encoding 616 amino acids were obtained by correcting for frameshift mutations.

Fig 2 – Secondary structure of the Omi472-m5569 group I (IC2 class) intron found in the *O. minus* strain WIN(M)472 *rns* gene. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The ten pairing regions (P1–P10) are indicated and the P10 region is boxed. The solid black arrowheads indicate the intron–exon junctions (5’ and 3’ splicing sites). The LAGLIDADG ORF has a putative start codon within the P5b region and the ORF extends into the P9.1 loop.
Insertion sites could be identified: mS379, mS569, mS952, and mS1224 (Table 1, Fig 1A and B).

The mS379 intron characterized in strains WIN(M)472, WIN(M)494, and WIN(M)515 is a class A group II intron, based on the presence of the conserved domain (D) V sequence. The intron encodes a putative RT-like protein that has been fragmented into two segments by frameshift mutations (Fig 1). The mS569 intron found in strains WIN(M)371 and WIN(M)371 and

**Figure 3** — Secondary structures of the Omi472-mS379 [A] and Omi371-mS952 [B] group II intron (class IIA1 and IIB1 respectively) RNAs. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The positions of EBS1, EBS2, EBS3, and δ are noted. The positions of IBS1, IBS2, IBS3, and δ in the 5' and 3' exons are boxed with dotted lines. Tertiary interactions are indicated by dashed lines and Greek letters (ε, λ, α, β, θ, κ, γ and δ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V, and VI). The solid black arrowheads indicate the intron–exon junctions (5' and 3' splicing sites). The asterisk shows the bulged adenosine nucleotide in domain VI (the branch point). The RT ORF is encoded within D II in the Omi472-mS379 intron [A], while the LAGLIDADG ORF is encoded within D III in the Omi371-mS952 intron [B].
WIN(M)472 is a class C2 group I intron and it encodes a LAGLIDADG HE (LHE)-like sequence (Table 2). The third intron at position mS952 as found in strains WIN(M)371, WIN(M)515, and WIN(M)873 is a typical class B group II intron and it encodes a LHE-like sequence. The fourth intron inserted at position mS1224 in strain WIN(M)515, is a class C2 group I intron, and contains a highly degenerated/fragmented LAGLIDADG ORF due to frameshift mutations and the presence of premature stop codons.

**Group I introns**

The group I introns found in the Ophiostoma minus rns gene belong to class C2 due to the presence of P5a and P5b in the P5 paring region and the absence of the P2 region (Comparative RNA Web Site (CRW), http://www.rna.ccb.utexas.edu, Cannone et al. 2002). The P, Q, R, and S sequences (Michel & Westhof 1990) could be identified but P2 is absent (Fig 2, Suppl. Fig 1A). The start codon of the LAGLIDADG ORF is located at the stem of P5b while the stop codon is found within the loop of the P9.1 region. The mS1224 intron is longer (1394 bp) than the mS569 introns (1323 bp), and the P9 region of the mS1224 intron is a more complex fold and contains a highly fragmented LAGLIDADG-type ORF (Suppl. Fig 1B).

**Group II introns**

Both introns at mS379 and mS952 are typical group II introns containing the characteristic features of class A and class B respectively: such as the exon binding sites (EBS1, EBS2, & EBS3/0) which are complementary to the intron binding sites (IBS1, IBS2, & IBS3/0) in the upstream and the downstream exons flanking the intron insertion site (Fig 1B; Fig 3); the internal loop of D Ii (‘C’), the internal loop of D II, the linkers between domains I and VI and the absence of insertions in the 3’ strand of D Ii and D lii (Michel & Ferat 1995; Toor et al. 2001). What distinguishes these group II introns from others are the type and position of the encoded ORFs, group II introns usually encodes RT-like ORF in the loop of domain IV. The mS379 intron encodes a RT-like ORF in the loop of domain II (Fig 3A; Suppl. Fig 2), while the mS952 intron encodes a potential LHEs within the loop of domain III (Fig 3B; Suppl. Fig 3).

**In vivo splicing of the rns introns**

To confirm the predicted exon/intron junctions derived by comparative sequence analysis RT-PCR was applied to RNA from strains WIN(M)371, WIN(M)472, WIN(M)494, and WIN(M)888 (Fig 4). We were not able to extract RNA from strain WIN(M)515 so we substituted it with strain WIN(M)888, a strain that has the same rns gene configuration as WIN(M)515. The rns genes from these strains contain representatives of all introns encountered during this study: WIN(M)371 contains mS569 and mS952, WIN(M)472 contains mS379 and mS569, WIN(M)494 contains mS379, and WIN(M)888 contains mS379, mS952, and mS1224 (see Fig 1 and Table 2). Sequence analysis of the cDNA confirmed that the predicted introns were all spliced out and are thus missing from the mature rRNA and the sequence data confirmed the predicted exon/intron junction sequences (Fig 1B).

**Phylogenetic analysis of the LAGLIDADG ORFs**

Double motif LHEs were found within the Ophiostoma minus mS569, mS952, and mS1224 introns. A LHE-like 370 amino acid ORFs is encoded within the mS569 group I intron and a 306 amino acid LHE-like ORF is located in the mS952 group II intron. The mS952 and mS1224 intron ORFs from strain WIN(M)515 were found to be highly degenerated/fragmented due to numerous mutations. These ORFs were omitted from the phylogenetic analysis as unambiguous reconstruction of the putative intact (or near intact) amino acid sequence was not possible. Sequences related to the rns intron-encoded ORFs were extracted from GenBank with BLASTp. In total 43 double motif LAGLIDADG ORF amino acid sequences were compiled, aligned and subjected to phylogenetic analysis in order to determine the possible origins of the O. minus rns intron ORFs (Fig 5).

The O. minus mS952 ORF belongs to a family of double motif LAGLIDADG ORFs that appear to share a common ancestor with LAGLIDADG ORFs that are encoded by group I introns located within protein coding genes and rRNA genes (node 1 and 2, Fig 5). The tree topology shows that group I encoded ORFs branch basal to the mS952 encoded ORF which suggests that the LAGLIDADG ORF most likely transferred from a group I intron to the rns mS952 group II intron (node 2, Fig 5). Although there are other lineages of LAGLIDADG ORFs that appear to have invaded group II introns, see node 3 (Fig 5). The latter lineage (mS785) appeared to be unrelated to any of the O. minus rns intron ORFs. The O. minus mS569 intron ORFs appear to be related to another intron ORF inserted at the same position in the rns gene of Sclerotinia sclerotiorum (node 4, Fig 5). Overall based on...
Fig 5 – Phylogenetic tree of the LAGLIDADG HEG-like elements related to the O. minus rns group I and group II intron-encoded ORFs. The tree topology is based on Bayesian analysis and the numbers at the nodes represent the following: top number is the PP value as obtained from a 50% majority Bayesian consensus tree, the middle number is the BS support based on NJ analysis, and the last number represents the BS support based on Parsimony analysis. Nodes that received less than 50% support (BS or PP) were reduced to polytomies. The branch lengths are based on Bayesian analysis and are proportional to the mean number of substitutions per site. Host gene and intron number are underlined. GenBank accession numbers are listed in brackets.

The highly variable mtSSU-rRNA gene of O. minus
Fig 6 – A phylogenetic tree showing the relatedness among a set of group II intron RT ORFs. The tree topology is based on Bayesian analysis. The tree was rooted with the non-LTR RT from Drosophila melanogaster (GenBank accession number: X51967). Numbers at the nodes represented the following: upper number represents the PP values obtained from Bayesian 50% majority consensus tree, middle number the BS support based on NJ analysis, and the last number is the BS support.
the limited number of sequences available, the ORF sequences evolved according to their respective insertion sites. Overall the O. minus intron ORFs are part of a large clade (node 5, Fig 5) that includes free-standing ORFs and group I intron ORFs inserted in protein coding genes (nad2, nad3, nad4, nad4l, nad5, cox1, cox2, and atp6) as well as rRNA genes (rns and rnl).

Characterization of the mS379 intron-encoded RT ORFs

The O.mi494-mS379 ORF is degenerated and thus the ORF finder program identified two separate putative ORFs, encoding the RT domain (412 amino acid) and the maturase domain (152 amino acid) respectively. However, a contiguous 616 amino acid RT ORF can be generated by correcting for a frameshift mutation (removal of a T insertion) at the nucleotide position 1900 (GenBank accession HQ292070), the mutation caused the introduction of a premature stop codon (TAA). Similarly the mS379 ORFs were degenerated in all strains studied. The degenerated ORF encoded by the O.mi472-mS379 intron could be regenerated into a 616 amino acid sequence by correcting the frameshift mutation as in the O.mi494-mS379 ORF plus a deletion mutation (A) at position 1708 and a substitution mutation at position 1711 [G → A] (accession HQ292072). However, the RT-like ORF from O.mi515-mS379 was highly degenerated and could not be regenerated into a complete continuous ORF instead only a fragmented 447 amino acid sequence could be assembled that like those described previously showed similarities a group II intron-encoded RT-like ORFs.

Phylogenetic analysis of group II intron-encoded RT proteins

In order to determine the evolutionary position of the Ophiostoma minus mS379 ORFs among other group II intron-encoded ORFs a data set was compiled by extracting related sequence from GenBank with a BLASTp search using the O.mi472-mS379 ORF amino acid sequence as a query (Fig 6). The phylogenetic analysis of the aligned amino acid sequence data set composed of 63 sequences was restricted to segments that could be unambiguously aligned such as the RT domains (RT-0 to RT-7) and the X-maturase domains (Suppl. Fig 4) (Lambowitz & Zimmerly 2004). The data set contained 34 mitochondrial, six chloroplast and 23 bacterial group II intron related RT-like ORF sequences. The alignment consists of 571 amino acid position. The non-Long Terminal Repeat element (non-LTR) from Drosophila melanogaster was used as an outgroup because this class of retroelements is mechanistically and phylogenetically linked to mobile group II introns (Eickbush 1999).

Phylogenetic trees generated from the RT sequence alignment placed the group II RT sequences into several groupings. The following clades were observed: the plant matR ORFs (node 1, Fig 6), the euglenoid (node 2, Fig 6) RT ORFs, six bacterial subclasses (A, B, C, D, E, and F; node 3, Fig 6), and the mitochondrial RT ORFs (node 4, Fig 6) (Martinez-Abarca & Toro 2000; Zimmerley et al. 2001; Simon et al. 2008, 2009). We noted examples of horizontal transfer of RT ORFs between unrelated groups of organisms. For example among fungal mitochondrial group II intron ORFs one finds interspersed ORFs from bacterial (node 5, Fig 6), liverwort (node 6, Fig 6) and brown algal sources. For example the cox1-i1 and cox1-i2 RT ORFs from the Pyliella litoralis (a brown algae, stramenopiles) group with the Neurospora crassa cox1-i1 and Marchantia polymorpha cox1-i2 RT ORFs respectively (node 7, Fig 6). Also the P. litoralis rnl-i1 and rnl-i2 RT ORFs grouped within the bacterial domain of the RT ORF tree (node 8, Fig 6). The heterogeneous distribution of closely related RT ORFs in different taxonomically unrelated groups is indicative of horizontal transfers.

The O.mi-mS379 ORFs formed a sister group with the P.an.cox1-i1 RT (node 9, Fig 6) and the node support values are 96% and 99% for BS support and PP respectively, and this cluster shares a common node with a set of group II intron ORFs mostly inserted within the cox1 gene (Allomyces macrognus cox1-i3, Podospora anserina cox1-i4, Schizosaccharomyces pombe cox1-i1 and cox1-i2) or the M. polymorpha rns gene (rns-i1) (node 10, Fig 6).

Discussion

Ophiostoma minus rns gene has two IC2 group I intron insertion sites

The secondary structure of group I introns consist of base-pairing elements designated as P1–P9. Another conserved structure is P10 which forms by base pairing the 3' exon sequence with the internal guide sequence (IGS) contained in the P1 loop. The P2 element is not found in all group IC introns. There are also a set of conserved internal sequence elements found in some group I introns (P, Q, R, and S) where the P element pairs with the Q sequence and the R sequence pairs with the S element (Burke et al. 1987; Cech 1988). During this study we found two possible insertion sites for IC2 group I introns among the examined strains of O. minus. So far only a few rns introns have been characterized within the fungi, the mS569 intron has only been described from Sclerotinia sclerotiorum and an undescribed species of Cordyceps; whereas there is only a single report for an mS1224 IC2 type intron in Cordyceps sobolifera (group I intron sequence and structure database, Zhou et al. 2008). More examples of these types of introns have to be described in order to evaluate their origins.

Group II intron

The position of the matR members (plant mtDNA ND1-i4) and the euglenoid RT ORFs within the group II RT ORF family was not resolved in our analysis, previous reports suggest that matR sequences are derivatives of mitochondrial group II RT-like sequences (Zimmerly et al. 2001; Hausner et al. 2006).
The chloroplast euglenoid lineage of RT ORFs are a set of highly divergent sequences typically associated with group III (degenerated group II introns) introns present within the psbC gene, and because of their high divergence they usually generate unstable tree topologies (Zimmerly et al. 2001). The data set does however confirm the recent findings of Toro et al. (2002) and Simon et al. (2008, 2009) with regards to the existence of the bacterial subclasses E and F respectively. As expected for group II intron ORFs, encoded within a potentially mobile element, there are numerous examples of horizontally transfer within and between the different groups of organisms (Zimmerly et al. 2001).

The O.mi-mS379 RT ORFs formed a sister group with the P.an.coxl-11 RT suggesting the transfer of a group II intron from a protein coding gene into an rRNA gene. Typically mobile introns invaded cognate intronless alleles, however group II intron RNAs have been shown to retrotranspose by reverse splicing into RNA molecules (Bonen & Vogel 2001), a mechanism that requires less specificity at the target site and thus allows for retrotransposition of introns to new sites within the genome (Lambowitz & Zimmerly 2010). The Marchantia polymorpha (M.po.) rns-11 RT ORF failed to group closely with the O.mi-mS379 ORFs, however the M.po-rns-11 is inserted between position S474/S475 (based on the Escherichia coli reference sequence) thus this group II intron most likely represents a different lineage then the mS379 introns.

**Group II introns with RT ORF embedded within domain II**

Group II introns have conserved secondary structures at the RNA level, that can be visualized as six stem-loop domains (D I–D VI) emerging from a central wheel (Michel & Ferat 1995). When RT-like ORFs are present they typically are embedded within D IV. However in some bacterial group II introns RT type ORFs have been observed in D II (Simon et al. 2008) and some LAGLIDADG-type ORFs are inserted in D III or D IV (Toor & Zimmerly 2002). Some group II intron-encoded proteins extend upstream and are fused to the upstream exon, this results in the generation of a fusion protein upon translation which probably is resolved by proteolysis (Michel & Ferat 1995). Here we report the first example of mitochondrial group II intron that has the entire ORF embedded within domain II.

Initially attempts were made to fold mS379 in a mode that would position the ORF into D IV but that prevented recovering structures that contained the characteristic secondary and tertiary interactions expected for group II introns. By moving the ORF into D II a secondary structure could be generated that allowed for the recognition of the majority of secondary (D I–D VI) and tertiary (i, λ, α, β, δ, ε, γ, and γ) interactions expected for group II introns. It is quite possible that D II can carry additional insertions like the RT ORF because once D II is locked into position by the tertiary interactions (0–0’; see Fig 3A and Suppl. Fig 2) and a series of contacts with the S-turn in D III, the loop of D II (that contains the ORF) projects away from the ribozyme core (Pyle 2010).

One can only speculate on how the ORF shifted into the D II position, maybe the ancestral version of this intron was encoded within a protein coding gene and the RT ORF extended from D IV upstream and was fused to the 5’ flanking exon. A configuration seen in the yeast coxl group II intron ORFs (Lambowitz & Zimmerly 2004) over time sequences downstream of the ORF position changed and could substitute for D III thus allowing the ORF to ‘slide’ into the D II position. Alternative sequences could fold into a D IV configuration and D V the most conservative component of the group II ribozyme remained the same. As group II introns appear to coevolve with their encoded RT ORFs (Toor et al. 2001) the above ‘ORF sliding’ model is a possibility as the group II phylogeny shows that the O.mi-mS379 ORFs are related to the Podospora anserina coxl-1 group II intron ORF, an RT ORF that is fused to the upstream exon as seen in the yeast coxl introns (Mohr et al. 1993; Lambowitz & Belfort 1993).

An alternative explanation might be a scenario whereby a group II intron inserted into domain II of an ORF less group II intron thus generating a ‘twintron’ like composite element, but over time the internal intron lost most of its ribozyme coding components leaving only the ORF sequence. Twintrons composed of group III introns, essentially degenerated group II introns (reviewed in Robart & Zimmerly 2005), and group II with group III introns have been described (Drager & Hallick 1993) and examples of twintrons have also been reported from crytomonas alga such as Rhodomonas salina (Pyrenomonas salina) where the internal intron lost its splicing capacity, essentially merging with the outer intron forming ‘one’ splicing unit (Khan & Archibald 2008). This illustrates the potential of generating new intron variants by introns invading other introns and this might also be an explanation for the generation of group II introns with repositioned RT ORFs.

**LAGLIDADG-type ORFs in group I and group II introns**

Group II introns that have LAGLIDADG-type ORFs such as the mS952 intron have been described from several fungal taxa (Toor & Zimmerly 2002; Monteiro-Vitorello et al. 2009; Mullineux et al. 2010) suggesting that these introns are wide spread. HEGs are quite invasive and can invade sites that are neutral, thus avoiding toxicity to the host genome. Therefore, self-splicing introns are ideal targets for invasion by HEGs. Most likely the mS952 intron arose by the HEG originated from a group I intron source invading an ORF less group II intron. Recently Mullineux et al. (2010) demonstrated that in a Leptographium species the mS952 group II intron-encoded HEG expresses a functional HE that cleaves the rns gene two nucleotides upstream of the mS952 position in an intronless allele, thus potentially facilitating the mobility of its host intron by a mechanism that is similar to that for group I introns.

During this study we also encountered group I encoded LAGLIDADG-type ORFs that had degenerated (mS1224) due the accumulation of mutations that introduced frameshifts and premature stop codons. This is expected as many HEGs are viewed as neutral elements that are not subject to selection thus they can accumulate non-adaptive mutations that will eventually lead to their complete loss from the genome (Goddard & Burt 1999). Thus in order for HEGs to persist they have to continuously invade new sites either via vertical or horizontal transfers or gain new functions that might make them an essential component of the genome (Gogarten & Hilario 2006).
Fig 7 – Evolutionary models for the spread of LAGLIDADG ORFs [A] and group II intron RT ORFs [B]. [A] A schematic diagram that summarizes findings by Haugen & Bhattacharya (2004) suggesting that rnl single motif LAGLIDADG ORFs represent the oldest intron ORFs and after the evolution of double motif LAGLIDADG ORF via a gene duplication event the introns spread into other rDNA sites and into protein coding genes. Our phylogenetic data (Fig 5) would suggest that the *O. minus* group I introns represent a reinvasion of rDNA by group I introns originating within protein coding genes. The mS952 group II intron ORF appears to be an example of a 'host switch' where a group I intron encoded double motif LAGLIDADG ORF moved from a group I intron into a group II intron. [B] The spread of Group II RT ORFs and changes within the structure of the RT ORFs. Models include recent findings by Zimmerly et al. (2001) and Simon et al. (2009). [B] adapted from Zimmerly et al. (2001). The RT component of the group II intron ORF can be divided into eight domains (0–7) which is followed by the X and Zn (also referred to as En) domains. The RT activity is characterized by the YADD amino acid sequence motif contained within domain 5. The X (also referred to as M) domain provides maturase activity and the En (or Zn domain) provides endonuclease activity; there is a D domain (not shown) located between the X and Zn domains in some bacterial group II introns and it is associated with DNA binding (Lambowitz & Zimmerly 2004, 2010). The RT ORF components changed over time in some lineages such as expansion of spacer (S) segments and in some lineages the loss of the Zn domain or the loss of the YADD motif. The aspect of this model that applies to this work is the dynamics of the fungal mtDNA group II intron ORFs, as shown in the phylogeny in Fig 6, these introns can move horizontally among fungi, between brown algae and fungi and potentially from fungi to bacteria. The *O. minus* mS379 group II intron ORF however appears to be degenerating due to the presence of frameshift mutations suggesting it lost its mobility.
Evolutionary dynamics of LHE and RT ORFs

Based on recent studies (Haugen & Bhattacharya 2004) it is thought that a single motif LHEs that inserted into a preexisting group I intron in the rnl gene between positions L1917 and L1951 evolved into double motif LHEs and then in combination with their group I intron host spread into sites first within the rnl gene and subsequently it invaded new sites in the rns gene as well as protein coding genes (see Fig 7A). The mS569 ORF belongs to a clade of intron ORFs that are encoded in introns located in rDNA and protein coding genes a distribution predicted by Haugen & Bhattacharya (2004) model for the dispersal of rDNA introns into other rDNA sites and protein coding loci. So the double motif LHEs in combination with a group I intron invaded protein coding genes and also transferred back to rRNA genes such as we observed for the mS569 or mS1224 introns and their ORFs. Based on those previous studies and our observations we also suggest that one lineage of LHEs moved independently from its original ribosome partner (most likely a group I intron) and transferred into D III of an ORF less preexisting group II introns located at position S952 (Fig 7A). Other lineages of LHEs moved into group II introns that are located at mS785 and mL2059 (Toor & Zimmerly 2002; Monteiro-Vitorello et al. 2009; Fig 7A).

Based on more exhaustive studies by Zimmerly et al. (2001) and Simon et al. (2009) it appears that group II RT ORFs probably originated in bacteria and extensively transferred horizontally (see Fig 7B). The group II introns and their RT ORFs were probably introduced to eukaryotes via the endosymbionts that gave rise to mitochondria and chloroplast. The chloroplast RT ORFs appear to spread horizontally but in the euglenoid chloroplast the RT ORF YADD amino acid motif (see Suppl. Fig 5), and the Zn domains were lost and thus the intron ORF unit appear to have lost mobility and transfers vertically (Zimmerly et al. 2001; Simon et al. 2009). It has been noted in previous reports that within the mitochondrial lineage the ORFs also underwent changes (Zimmerly et al. 2001; Fig 7B) and evidence for both vertical and horizontal transmission has been presented in this and other studies (Zimmerly et al. 2001; Lambowitz & Zimmerly 2004) showing that these elements like other mobile elements persist by continuously invading new sites. For example the O.mi-mS379 intron was probably derived from a group II intron that most likely retrotransposed from the cox1 gene into the rns gene. However, the current ORF configuration in the O.mi-mS379 intron suggests degeneration and most likely this element has lost its mobility and is therefore vertically transferred.

Conclusions

Overall this study demonstrated that the rns gene in Ophiostoma minus has been invaded by group I and group II introns, thus generating DNA polymorphisms. The mtDNA variations caused by putative mobile introns might not be stable enough to be useful in species identification but these polymorphisms could be useful in distinguishing O. minus strains from one other in combination with other nuclear or mitochondrial markers. This study indicates that the rns gene within a species (or among closely related species) can be highly variable and that group I and group II introns and their encoded ORFs contribute toward the variability and complexity of the mitochondrial genomes among the ophiostomatoid fungi. The study also shows O. minus strains might be good candidates for bioprospecting for potentially active HEs and ribozymes that have applications in biotechnology (Hausner 2003; Gimble 2005; Lambowitz et al. 2005; Marcaida et al. 2010; Lambowitz & Zimmerly 2010).

Acknowledgments

This work is supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to G.H. We are deeply indebted to Drs. François Michel and Steven Zimmerly for their suggestions on how to fold the O.mi-mS379 group II intron. We also greatly thank Dr. James Reid for providing us with fungal cultures. M.H. is supported by the Egyptian Ministry of Higher Education and Scientific Research. We also would like to acknowledge the comments of the anonymous reviewers that helped in improving the manuscript.

Supplementary data

Supplementary data related to this article can be found online at doi:10.1016/j.funbio.2011.07.007.

References


