Fungal Genetics and Biology 53 (2013) 71–83

Contents lists available at SciVerse ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

The mtDNA *rns* gene landscape in the Ophiostomatales and other fungal taxa: Twintrons, introns, and intron-encoded proteins

Mohamed Hafez [a,1], Anna Majer [a], Jyothi Sethuraman [a], Shelly M. Rudski [a], François Michel [b], Georg Hausner [a,*]

[a] Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada R3T 2N2
[b] Centre de Génétique Moléculaire du CNRS, 91190 Gif-Sur-Yvette Cedex, France

ARTICLE INFO

Article history:
Received 6 October 2012
Accepted 15 January 2013
Available online 9 February 2013

Keywords:
mtDNA
rRNA
Group I and II introns
Homing endonucleases
Reverse transcriptase

ABSTRACT

Comparative sequence analysis of the mitochondrial small subunit ribosomal RNA (*rns*) gene among species of *Ophiostoma*, *Grosmannia*, *Ceratocystiopsis* and related taxa provides an overview of the types of introns that have invaded this gene within the ophiostomatoid fungi. The *rns* gene appears to be a reservoir for a number of group I and group II introns along with intron-associated open reading frames such as homing endonucleases and reverse transcriptases. This study uncovered two twintrons, one at position mS917 where a group ID intron encoding a LAGLIDADG ORF invaded another ORF-less group ID intron. Another twintron complex was detected at position mS1247 here a group IIA1 intron invaded the open reading frame embedded within a group IC2 intron. Overall the distribution of the introns does not appear to follow evolutionary lineages suggesting the possibility of rare horizontal gains and frequent losses. Results of this study will make a significant contribution to the understanding of the complexity of the mitochondrial intron landscape, and offer a resource to those annotating mitochondrial genomes. It will also serve as a resource to those that bioprospect for ribozymes and homing endonucleases.

1. Introduction

The gene coding for the mitochondrial small subunit ribosomal RNA is a highly conserved essential gene that like its nuclear, chloroplast, Eubacterial and Archaean counterparts has been exploited as an evolutionary chronometer for phylogenetic studies. In 1988 it was reported that all ascomycetous mtDNA *rns* genes are continuous (i.e. no introns; Wolf and Del Giudice, 1988), however, since then a variety of introns have been noted to be inserted within fungal *rns* genes (Haugen and Bhattacharya, 2004; Haugen et al., 2005; Lang et al., 2007; Hafez and Hausner, 2011a). The mitochondrial genomes of fungi contain two types of introns, group I and group II introns, these are potential self-splicing elements and therefore they minimize their impact to the genes that host them. However, it should be noted that many group I and group II introns for efficient *in vivo* splicing require intron- or host-encoded factors that assist in the splicing process (Wank et al., 1999; Bonen and Vogel, 2001; Mohr et al., 2001; Matsuura et al., 2001; Mohr and Lambowitz, 2003; Haugen et al., 2005; Lang et al., 2007; Fedorova et al., 2010; Hausner, 2012). Many mtDNA introns encode open reading frames (ORFs) for so-called intron-encoded proteins (IEPs). For group I introns, IEPs either promote intron mobility and/or assist in intron splicing activity by acting as homing endonucleases (HEs) and/or HE-derived maturases, respectively (Belfort, 2003). In group II introns IEPs typically are reverse transcriptase-like (RT) proteins (Lambowitz and Zimmerly, 2011).

Homing endonucleases are enzymes that can initiate intron mobility by introducing double stranded breaks (DSBs) at specific target sites, which activates the cellular DSB-repair pathway (Belfort et al., 2002). Intron-encoded maturase proteins assist the intron RNA with folding into a splicing-competent ribozyme (Caprara and Waring, 2005). There are two families of HEs commonly encountered among fungal mtDNAs; named according to conserved amino acids motifs, these are the LAGLIDADG and GIY-YIG proteins (Stoddard, 2005, 2011). Homing endonuclease genes (HEGs) have also been shown to be mobile elements that can move independently from their ribozyme counterparts (Mota and Collins, 1988). Mobile introns and their IEPs are of interest as they have applications in biotechnology as rare cutting, site-specific endonucleases (Takeuchi et al., 2011).

Typical group II introns are retro elements that encode proteins with RT activity. Group II intron mobility is promoted by a ribonucleoprotein consisting of the IEP and the spliced lariat version of the intron RNA (Michel and Lang, 1985; Michel and Ferat, 1995; Lambowitz and Zimmerly, 2011). However, some group II introns encode LAGLIDADG type HEs (LHEs) related to IEPs associated with

* Corresponding author. Fax: +1 204 474 7603.
E-mail address: hausnerg@cc.umanitoba.ca (G. Hausner).
[1] Present address: Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, 2900 Boulevard Edouard Montpetit, Montréal, Québec, Canada H3C 3J7.

1087-1845/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.fgb.2013.01.005

group I introns (Toor and Zimmerly, 2002). It has been suggested that these LHEs have the potential to mobilize their group II intron host by a DNA-based mechanism invoked to explain the HE-facilitated homing of group I introns into cognate alleles that lack the intron (Mullineux et al., 2010, 2011; Li et al., 2011; Pfeifer et al., 2012).

HEGs and introns are invasive elements and therefore contribute towards the size of fungal mtDNA genomes, mtDNA polymorphisms, and they can promote mtDNA rearrangements (Dujon, 1989; Charter et al., 1996; Belcour et al., 1997; Gibb and Hausner, 2005; Sethuraman et al., 2009; Hafez and Hausner, 2011a). Also some of these elements have been associated with mtDNA instabilities in a variety of fungi (Michel and Cummings, 1985; Cummings et al., 1986; Dujon and Belcour, 1989; Abu-Amero et al., 1995; Sethuraman et al., 2008). Group II introns have been associated with mitochondrial genome instabilities such as senescence in *Podospora anserina* (Dujon, 1989; Griffiths, 1992) and mitochondrial-induced hypovirulence in *Cryphonectria parasitica* (Baidyaroy et al., 2011).

The initial focus of this study was to examine the *rns* intron landscape for species of *Ophiostoma*, *Grosmannia* and *Ceratocystiopsis*; collectively these fungi are sometimes referred to as the ophiostomatoid fungi (Wingfield et al., 1993). These genera contain plant pathogens such as the causative agents of Dutch Elm Disease (*Ophiostoma ulmi* and *Ophiostoma novo-ulmi*) and blue-stain fungi (e.g. *Ophiostoma minus*); the latter can discolor timber and thus reduce its economic value (Harrington, 1993; Hausner et al., 2005). Some blue-stain fungi such as *Grosmannia clavigera* are fungal associates of the mountain pine beetle (*Dendroctonus ponderosae*) and this fungus is a pathogen of lodgepole pine (*Pinus contorta*) (Fisher et al., 2012). So far very little information is available on the mitochondrial genomes and their intron complements for the ophiostomatoid fungi. We expanded the study to examine the types of introns inserted within the *rns* gene in other members of the Ascomycota, the results of this study should be useful in annotating mtDNA genomes and the *rns* genes for fungi.

2. Material and methods

2.1. DNA extraction and amplification protocols

The fungi used in this study and their sources are listed in Supplementary 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. Procedures for the extraction of DNA and for the amplification of the *rns* gene were described previously in Hafez and Hausner (2011a). The *rns* gene was amplified by PCR with the primers mtsr-1 and mtsr-2 (Mullineux et al., 2010) using standard DNA amplification protocols utilizing the Invitrogen PCR system (Invitrogen, Burlington, Canada). In some instances primers rns-5′ (5′-GAGTTTGGTGATGGCTCTG-3′) and rns-3′ (5′-CCACTACACGAACCGTATTTC-3′) were used to amplify the complete sequence of the *rns* gene; these primers bind to the 5′ and 3′ ends of the *rns* gene and thus allowed for the sequence characterization of the entire *rns* gene. All PCR amplicons were analyzed by gel electrophoresis through 1% agarose gel in TBE buffer (89 mM Tris-borate, 10 mM EDTA, pH 8.0). The DNA fragments were sized using the 1-kb plus ladder (Invitrogen), and nucleic acids were visualized by staining with ethidium bromide (0.5 μg/ml) and exposing the stained gels with UV light.

2.2. Cloning and sequencing

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). In some instances the *rns* derived PCR products were cloned into *Escherichia coli* (DH5α) using the TOPO TA Cloning® kit (Invitrogen) to improve sequencing efficiency. Recombinant plasmids from positive clones were purified with the Wizard® Plus Minipreps 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, additional primers were designed as needed to complete sequences in both directions [see Hafez and Hausner (2011a) for a complete list of the primers used to sequence the *rns* gene].

2.3. Sequence analysis

Individual sequences were compiled and assembled manually into contigs using the GeneDoc program v2.5.010 (Nicholas et al., 1997) and nucleotide sequences were aligned with the Clustal-X program (Thompson et al., 1997). Twenty-three *rns* sequences were obtained during this study and combined with *rns* sequences obtained from a previous studies (Mullineux et al., 2011; Hafez and Hausner, 2011a,b) and data extracted from NCBI data bases (see Supplementary Table 1). 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* intron sequences.

The online resource Basic Local Alignment Search Tool (BLAST: http://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al., 1997) was used to retrieve nucleotide sequences (BLASTn) from GenBank which shared similarities to the *rns* gene and the introns that interrupt this gene. For phylogenetic analyses, only those segments of the alignment that could be aligned unambiguously were retained, also intron sequences were deleted.

Phylogenetic estimates were generated using likelihood-based and Bayesian approaches, by using PHYML (Guindon et al., 2010), and the MrBayes program v3.1 (Ronquist, 2004) respectively. The models applied for Likelihood approaches were based on evaluating the nucleotide sequence alignments with the FindModel program (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html), which identified the GTR with gamma distribution as the best model. The PHYML program within the HIV sequence database (http://www.hiv.lanl.gov/content/sequence/PHYML/interface.html) was used to obtain a tree topology based on ML analysis; here the nucleotide sequences were analyzed using the GTR model and by applying bootstrap analysis (1000 replicates). The MrBayes program was used for Bayesian analysis (Ronquist and Huelsenbeck, 2003) applying the GTR model with gamma distribution to the *rns* data set; four chains were run simultaneously for 2,000,000 generation; 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 values. The phylogenetic tree was drawn with the TreeView program (Page, 1996) using the MrBayes 50% majority rule tree file.

2.4. Intron nomenclature and secondary structure models

Introns were named according to the proposed nomenclature by Johansen and Haugen (2001), and intron insertion sites are based on corresponding nucleotide positions within the *E. coli* SSU-rRNA sequence (GenBank accession number AB035922). The secondary structures of the *rns* introns were predicted following the