The Evolutionary Dynamics of *rns* mtDNA Introns and Intron Encoded proteins of *Ophiostoma* and related taxa

By

Iman Majed Bilto

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba.

In partial fulfilment of the requirements of the degree of:

DOCTOR OF PHILOSOPHY

Department of Microbiology

Faculty of Science

University of Manitoba

Winnipeg, Manitoba

Canada

ABSTRACT

The mitochondrial small-subunit ribosomal RNA (mtSSU rRNA = *rns*) gene can be highly variable both in size and organization due to the presence of introns and intron-encoded open reading frames. Little is known about these introns and their intron encoded proteins. This study focussed on investigating introns and their intron encoded proteins present within the small-subunit ribosomal RNA gene of different ophiostomatoid fungi and related ascomycete fungi. The small-subunit ribosomal RNA genes were amplified and sequenced. The sequences of the small-subunit ribosomal RNA introns showed that the small-subunit ribosomal RNA gene hosts group I and II introns with embedded open reading frames that encode homing endonucleases or reverse transcriptases, respectively. Phylogenetic studies suggested that intron evolution is not necessarily vertical. Their distribution can be accounted for by horizontal transfers along with the loss and gain of these elements. One of the most important findings of this study is that at least some group II intron open reading frames appear to follow a similar life cycle as suggested for homing endonuclease open reading frames and group I introns. So far no model has been proposed to explain the evolutionary dynamics of the reverse transcriptase open reading frames encoded by group II introns.

In a second study, LAGLIDADG type homing endonucleases that appear to form a clade within the I-OnuI family of homing endonucleases were examined. Twelve orthologues which include members inserted at position S917 were noted to be located in five different genes, suggesting that members of this clade can recognize five different target sites. This study adds three new active members (I-CcaI, I-CcaII and I-AstI) to the well-studied I-OnuI family of LAGLIDADG homing endonucleases.

ACKNOWLEDGEMENTS

First, I would like to express my grateful to Dr. Georg Hausner for all the support he has given me. I am actually honoured to have worked in his lab and under his supervision, support, and encouragement. I would like also to thank my committee members, Dr. Deborah Court, Dr. Ann Karen Brassinga and Dr. Michele Piercey-Normore for their help, support, encouragement and their valuable comments.

I would like to thank my lab mates; Dr. Mohamed Hafez, Dr. Tuhin Kumar Guha, Chen Shen, Alvan Wai, Abdullah Zubaer and Talal Abboud for supporting me in my research work.

I also would like to acknowledge the **financial support for this project that was provided by**: The Natural Sciences and Engineering Research Council of Canada (NSERC, GH), The Islamic Development Bank, and the financial support from the Faculty of Graduate Studies (GETS), Faculty of Science Graduate Award (2016) and the Faculty of Graduate Studies, Faculty of Science and the Graduate Students Association at the University of Manitoba for providing funds for attending scientific conferences.

Table of Contents

List of Tables

List of Figures

General Introduction

Ophiostomatoid fungi are ascomycetous fungi that include species that are associated with blue stain of lumber and some members are causative agents of tree diseases. The best known example is *Ophiostoma ulmi*, the causative agent for Dutch elm disease (DED). Traditionally the ophiostomatoid fungi include the following genera of fungi: *Ophiostoma, Ceratocystis, Ceratocystiopsis* and *Grosmannia* (Zipfel *et al*., 2006). However, the genera *Ophiostoma* and *Ceratocystis* have recently been revised and subdivided into additional genera (see de Beer and Wingfield, 2013; de Beer *et al*., 2013a, b; de Beer *et al*., 2014). Blue-stain fungi are of interest as many of these fungi can cause discoloration of sapwood due to the production of melanin pigments and this reduces the economic value of lumber; it also limits the potential of the lumber to be exported.

Compared to metazoan mitochondrial genomes (mtDNA) fungal mtDNAs are variable in size even though they appear to share a common set of core genes. Fungal mitochondrial genomes are highly variable with regards to size [e.g., 19 kb in *Schizosaccharmyces pombe* or, 203 kb in *Sclerotinia borealis* (Mardanov *et al*., 2014)] and the presence and absence of introns and intron-encoded open reading frames.

Fungal mtDNA introns are self-splicing elements. Depending on the excision mechanism, there are two different intron types: Group I and group II introns. Group I and group II introns can move from an intron containing allele to a cognate allele that lacks the intron, therefore the mobility mechanism is termed "Homing" as it is site specific, and it is controlled by the intron encoded protein (IEP) . Group I and II introns can be distinguished from each other by their sequences, secondary and tertiary RNA structures, and by their splicing mechanisms. Also, group I introns usually move by utilizing a DNA intermediate, whereas group II introns are retro

elements and mobilize utilizing an RNA intermediate during part of their mobility pathway (Belfort *et al*., 2002; Lambowitz and Zimmerly, 2011).

Homing endonucleases (HEases) are referred to as rare DNA cutting enzymes as they are site specific and required long (14-40 bp) DNA target sites. HEases are encoded by homing endonuclease genes (HEGs) that can be located within self-splicing group I and in some instances within group II introns. However, HEGs can also be free standing genes. By cleaving at their cognate DNA target site, HEases generate a double-stranded break (DSB) that can be repaired by a homologous recombination event that uses the HEG^+ allele as the template. This repair process involves the cell's DNA repair machinery and the consequence is that the HEG^- allele is converted into a HEG^+ allele. The repair is associated with gene conversion that "moves the intron" sequence and in some cases flanking DNA sequences into a new location (i.e., the HEG- allele being converted into a HEG+ allele) (Belfort *et al*., 2002; Hausner, 2012).

Studying the fungal mtDNA may provide some insight into the evolution of these mobile elements and possibly their association with mtDNA instabilities that have been reported in some fungi (such as inducing senescence or hypovirulence; reviewed in Bertrand, 2000).

Fungal mtDNA are also relevant as they are a source of homing endonucleases and ribozymes (group I and group II introns), elements that have applications in biotechnology as genome editing reagents (Stoddard, 2005; Lambowitz and Zimmerly, 2004; Müller *et al*., 2016). HEases tend to share two features, high specificity and tolerance of some target site degeneracy and these make HEases potential tools in different applications, such as gene therapy, genome modifications, and pest control (reviewed in Stoddard 2005, 2011; Hafez and Hausner, 2012).

With regards to the use of HEases as genome editing applications their ability to generate a site specific DSB is the key as a DSB will trigger the cell to repair the DNA damage by either

homologous recombination (HR) or by non-homologous end joining (NHEJ). NHEJ is an error prone mechanism and can generate indels (small insertions/deletions) and this provides a means for targeted mutagenesis. HR based DNA repair allows for gene replacement strategies if a DNA repair template is provided along with the construct that allows for the expression of the genome editing reagents (i.e. HEases) (Stoddard 2005, 2011; Cox *et al*., 2015).

The objectives for this thesis are to characterize the mtDNA *rns* gene for a set of closely related fungal strains in order to assess the diversity of intron polymorphisms that can arise among closely related fungi. This may provide some insight as to the contribution mobile introns make toward mtDNA sizes and mtDNA variability among strains of the same species or among strains of different species. In addition, the second objective of this work is to provide more detailed characterization of three related HEases with regards to their activity and their possible DNA target and cleavage sites. The latter is a preamble in order to evaluate their suitability for possible future applications in biotechnology.

1. Literature review

1.1. Background

Before 1977 scientists thought that mRNA molecules that are translated to proteins are faithful copies of the gene (DNA), i.e. collinear. Eventually it was noted that genes can contain non-coding regions. Examination of genes and their corresponding mRNA by hybridizing DNA with the mRNA showed that the genes of eukaryotic cells have interruptions and these intervening DNA segments loop out from the DNA/RNA duplexes. Studies in the late 70s showed that eukaryotic mRNA molecules are shorter compared to their corresponding genes and that the mRNA miss large segments that are present within the gene. Therefore, genes can be viewed as segments that are expressed (exons, represented on the mRNA) and segments that are intervening and not translated into proteins and these are called introns (Gilbert, 1978). After the gene is transcribed the introns are removed from the transcript before the mRNA is translated into proteins (Chow *et al.*, 1977; Berget *et al.*, 1977; Berk *et al.*, 1977).

Dujon (1976) working with *Saccharomyces cerevisiae* yeast uncovered the so-called omega system (reviewed in Dujon and Belcour, 1989). His group noted that an element in the yeast large ribosomal RNA subunit gene (*rnl*) behaved in a manner that would be expected for a mobile element. Ultimately his study was on an intron in the mitochondrial (mt) DNA large ribosomal subunit gene (*rnl*) of *Saccharomyces cerevisiae* (omega+) strain which was absent in the (omega-) yeast strains. Mating omega + with omega – resulted in what has been referred to as "super Mendelian inheritance" as all progeny were converted to be omega+. Genetic and molecular studies demonstrated that this is due to the insertion of the intron into the intron-less copies of the *rnl* gene at exactly the same site (Dujon, 1981). Dujon's group also recognized that the mobility of the intron was due to the activity of the intron encoded protein (IEP), a "homing endonuclease" (HE). The movement of introns are initiated by a double strand break and

repaired by the double strand DNA break repair machinery of the cell. Co-conversion of flanking exon sequences along with the insertion of the intron was observed over a distance of up to a few hundred nucleotides beyond the intron insertion site. Another important observation was that a single substitution in the DSB region (or homing site) prevents the cleavage at the potential intron insertion site and thus prevents the insertion event (Dujon, 1981; Dujon *et al*., 1976; Szostak *et al*., 1983; Dujon *et al*., 1985; Dujon and Jacquir, 1983). This observation suggested that HEases are site specific endonucleases adapted to promote the insertion of its host intron into specific locations.

As stated above understanding intron mobility required a closer examination of the IEP which are responsible for introducing the DSB in the intron-less gene. This idea gained support when a mutation in the omega intron open reading frame (ORF) resulted in the intron to lose its ability to insert into the intron-less *rnl* gene (Jacquier and Dujon, 1985; Macreadie *et al*., 1985). Latter, it was shown that the IEP can recognise a unique sequence at the intron/exon junction region (within the intron less allele) and generate a 4-bp staggered cut at the cleavage site (Colleaux *et al*., 1988; Dujon *et al*., 1986). The IEP is unable to recognise the intron/exon junction region in the intron-plus gene as here its target site has been disrupted by the intron. Additional *in vitro* studies that involved the generation of mutations within and around the putative IEP recognition site showed that the IEP target site extends to more than 18 bp; this insures the high specificity observed with regards to the movement of the omega intron (Dujon and Belcour, 1989). Subsequently, mobile introns were described from the organellar genomes of other fungi and algae (Michel *et al*., 1982; reviewed in Hausner, 2003, 2012). From these studies, it was noticed that the same intron could be located in different and frequently unrelated species and sometimes even in different genes (Haugen *et al*., 2005; Lang *et al.*, 2007).

Combined with the observation that introns can have a random distribution among members of the same species/genus it suggests these elements can be rapidly gained and lost and are probably mobile and can be transferred laterally and on some occasions, insert into ectopic (different genes or different locations within the same gene) sites (Burke *et al*., 1984; Dujon *et al*., 1986; Dujon, 1989; Haugen *et al*., 2004).

It was noticed by Michel *et al.* (1982) that there are two types of introns: group I and group II introns (Michel *et al*., 1982; Dujon *et al*., 1989; Michel and Westhof, 1990). These introns can be differentiated based of their intron RNA secondary structure folds (Michel *et al.*, 1982). Comparative studies showed that the two types of introns potentially encode two different types of ORFs. Group I introns tend to encode homing endonucleases and group II introns can encode proteins that share homology with reverse transcriptases that are associated with retroviruses (Michel and Lang, 1985).

Belcour and Vierny (1986) suggested that there was a relationship between mtDNA intron splicing and the senescence process in *Podospora anserina*, this increased interest with regards to the phenotypic contribution introns make towards their hosts. Organellar introns have been associated with mitochondrial defects and have been studied by some groups as potential sources for hypovirulent plant pathogens that may have applications in biocontrol for certain tree diseases (see Bertrand, 2000; Baidyaroy *et al*., 2011).

Nuclear spliceosomal introns appear to be important as they allow for alternative splicing and thus the production of various transcripts (i.e. resulting in various peptides) from the same gene, these introns may also include regulatory elements controlling expression of the gene, the number and sizes of these introns may also be a way for regulating the rate of expression of nuclear genes (Gubb 1986; Palmiter *et al*., 1991; Lauderdale and Stein, 1992; Liu *et al*., 1995;

Tilgner *et al*., 2009; Hafez and Hausner, 2015). Finally, nuclear spliceosomal introns can be a source for non-coding regulatory RNA molecules (Kim and Kim, 2007; Okamura and Lai, 2008; Rearick *et al.*, 2011). It is generally assumed that group I and group II introns are probably neutral elements and have little impact on the host's genome and phenotype, however very little research has been done to assess the role of group I and II introns on gene regulation or on ogranellar genome stability. Although, as stated earlier there are a few instances where organellar introns have been associated with mitochondrial defects (Bertrand, 2000), in general organellar introns (and HEGs) are viewed to be neutral elements (Goddard and Burt, 1999) (i.e. have minimal effect on phenotype).

The origin of introns has become a rather complex topic; historically two hypotheses dominated the debate: one is referred to as "introns early" hypothesis which posits that intron evolved before the last universal common ancestor (LUCA) and that introns were important in promoting exon shuffling (Gilbert, 1978); the second hypothesis is referred to as "intron late" and it argues that introns started to evolve later during the diversification of the eukaryotes. This idea claims that early branching eukaryotes lack introns and more recently evolved eukaryotes are "intron rich" (reviewed in Koonin *et al*., 2006). Meanwhile Koonin *et al.* (2006) suggested that introns emerged during the evolution of the first eukaryote; here mobile group II introns from a bacterial endosymbiont invaded the genome of a putative Archaea host cell. However, it has been recently recognized that novel introns can continue emerge by various endogenous events (Rogozin *et al*., 2012; Koonin *et al*., 2013). Compared to spliceosomal introns, group I and group II introns are thought to be ancient and may have evolved in the RNA world, maybe in abiotic compartments that housed early replicator type systems (Koonin *et al*., 2006).

By studying intron structures based on the analysis of DNA and/or RNA sequences, and by comparing introns based on biochemical characteristics at least four distinct classes of introns are currently recognized (Alberts, 2008):

- Spliceosomal introns: introns removed by spliceosomes.
- tRNA introns: introns in some nuclear and archaeal transfer RNA genes that are removed by the means of RNA endonucleases with exons being ligated by RNA ligases.
- Self-splicing group I introns (GI).
- Self-splicing group II introns (GII).

This study will focus on mtDNA GI and GII introns in *Ophiostoma* species and related taxa of fungi. These introns are frequently encountered among fungal mitochondrial genomes and they are self-splicing elements that can behave like mobile elements (reviewed Hausner *et al*., 2014; Lambowitz and Belfort, 2015).

1.2. Fungi

The Fungi (Phylum Mycota) are eukaryotic organisms that include the molds, yeasts and mushrooms. Historically they were considered to be a distinct Kingdom. However, the term Kingdom is rarely used as the diversity of life is better represented within the three domains of life as proposed by Woese (reviewed in Woese *et al*., 1990). Among the Fungi there are many important organisms with regards to ecology and economics. Many fungi are decomposers and thus provide nutrients to the ecosystem. Many vascular plants depend on symbiotic fungi, or mycorrhiza, to supply nutrients by extending their root systems. Other fungi provide a source for drugs (antibiotics), food (mushrooms) or contribute toward the production of bread (yeast) and alcohol. Some fungi also have been used as biocontrol agents or used as medicinal "herbs" (Li *et al.*, 2013; Li *et al.*, 2015). In addition, fungi have been used as model systems for genetic and

molecular biological studies (*Saccharomyces cerevisiae*, *Neurospora crassa* and *Aspergillus nidulans* etc.) because some fungi are easy to grow and usually their sexual reproduce life cycles can occur rapidly in the laboratory. Fungi also have industrial applications such as the production of citric and lactic acids and various enzymes. Also, some fungi are of great economic concern as they can cause plant and animal diseases (Kavanagh, 2011).

1.2.1. The genus *Ophiostoma*

Ophiostoma species (Family Ophiostomtacae, Order Ophiostomatales, Class Sordariomycota, and Phylum Ascomycota) are typically characterized by possessing long-necked perithecia that contain scattered evanescent (short-lived) asci. The asci disintegrate and the ascospores are pushed out of the perithecial neck in the form of slimy spore droplets; an adaption that favours dispersal of spores by insects. The taxonomy of fungal taxa assigned to the Ophiostomatales has been controversial and historically they were assigned into either *Ophiostoma* and/or *Ceratocystis* (reviewed in Upadhyay, 1981; Hausner *et al*., 1993). These fungi produce asexual spores in either an enteroblastic (*Ophiostoma*) or holoblastic (*Ceratocystis*) fashion, again in many instances slimy masses of asexual (conidia) spores are produced at the tip of sometimes long stalked conidiophores; facilitating spore dispersal by insects. Phylogenetic studies based on rDNA sequence analysis showed species of *Ophiostoma sensu lato* can be assigned to the Ophiostomales and species of *Ceratocystis* are not closely related to *Ophiostoma* species and can be assigned to the Order of the Microascales (Hausner *et al*., 1993; Spatafora and Blackwell, 1994). The genus *Ceratocystiopsis* was originally proposed by Upadhyay (1981) to include ophiostomatoid fungi with small short-necked perithecia that produce "needle shaped" or falcate-shaped ascospores. Zipfel et al. (2006) updated the taxonomy of this group and depending on the type of conidial (asexual) structures produced and the shape

of the ascospores the Genus *Ophiostoma sensu lato* has been split into *Grosmannia*, *Ceratocystiopsis*, and *Graphilbum* (Zipfel *et al*., 2006; Reid and Hausner, 2015). *Ophiostoma* species are of interest as they are associated with various tree diseases, such as *Ophiostoma ulmi* the causative agent for Dutch elm disease (DED). Other members of this genus are blue-stain fungi, fungi associated with the discolouration of sapwood that reduces the economic value of lumber (reviewed in Wingfield *et al*., 1993).

1.2.1.1. Dutch elm disease (DED)

DED is a disease caused by members of the genus *Ophiostoma*. The disease has serious effects on *Ulmus Americana* (American Elm) and the disease is spread by the elm bark beetle (Coleoptera: Scolytidae). Current research suggests that DED started in Asia and moved westward towards Europe. From Europe it was transferred to America and eventually strains from America were transferred back to Europe. In North America there used to be large populations of elm trees but they did not have resistance against this fungus. There are three closely related species that are causative agents of DED:

- *O. ulmi* which affected Europe and North America during the first half of the 20th century;
- *O. himal-ulmi* endemic to the western Himalayas;
- *O. novo-ulmi* which is an aggressive species that was first described in Europe and North America (as races of *O. ulmi*) in the 1940s and has devastated elms in both areas since the late 1960s. *O. novo-ulmi* is now recognized to exist as two distinct subspecies: *Ophiostoma novo ulmi* subsp. *novo ulmi* (found in Europe and parts of Asia) and *Ophiostoma novo ulmi* subsp. *americana* (found in North America but has been

transferred into Europe) (Brasier and Buck, 2001; Brasier and Kirk, 2001; Paoletti *et al*., 2005, 2006).

When trees are infected by *Ophiostoma ulmi* there is usually a delay in showing symptoms (Plichta *et al*., 2016). The fungi start to colonize the sapwood and they produce cell wall degrading enzymes (Svaldi and Elgersma, 1982; Binz and Canevascini, 1996) and produce some high molecular weight toxins (Van Alfen and Turner, 1975, Takai and Richards, 1978). The fungus will enter the phloem and xylem vessels and this causes the formation of alveolar structures, tyloses and gels inside these vessels, and this eventually leads to blockages of the conductive tissues (Newbanks *et al*., 1983, Ouellette *et al*., 2004), which prevents the flow of nutrients and water and this will lead to wilt (Urban and Dvořák, 2014) and eventually death of the tree. Depending on environmental conditions, the disease can kill a tree in 3 weeks (Brasier, 1979; Hintz *et al*., 1993; Brasier and Kirk, 2001) or it can take as long as 3 years (Plichta *et al*., 2016).

1.2.1.2. Blue stain fungi

Blue stain fungi usually colonize the sap wood of either soft (e.g. conifers) or hard (e.g. elm) woods and they cause discoloration of the wood by depositing melanin type pigments. Examples of blue stain fungi are the ophiostomatoid fungi, such as species of *Ophiostoma, Ceratocystis, Pesotum, Graphilbum, Grosmannia, Ceratocystiopis, Graphium* and *Leptographium*. Blue stain fungi can metabolize substances that have accumulated inside the sapwood's resin canals, parenchyma cells and tracheid lumina; such as simple carbohydrates, fatty acids, triglycerides and other chemicals found in the xylem vessels of the sap wood (Harrington *et al*., 1998; Jacobs and Wingfield, 2001; Robinson, 1962).

Most blue stain fungi are associated with bark beetles and they can form symbiotic relationships whereby they can assist each other in infection and getting established in host trees. The bark beetles have sharp mouth parts that are used for boring through the bark and making long, vertical galleries where the eggs are laid. The spores of blue stain fungi are sometimes carried in these mouth parts as well as on the insect's body (Furniss *et al.*, 1990). The beetle may benefit, since the blue stain fungus can weaken and limit the tree's responses. In some instances the fungus is also a food source for developing insect larvae. Benefits to the fungi include spore dispersal and inoculation into suitable host trees (Malloch and Blackwell, 1993).

In some instances blue stain fungi can be destructive, for example some *Leptographium* sp. and *Ophiostoma* sp. that can produce toxic substances (Li *et al*., 2011). Some blue stain fungi can kill mature host trees in the absence of their bark beetle vectors (Solheim and Safranyik, 1997) others appear to be part of a diseases cycle that involves the bark beetle vector. Some blue stain fungi clog up the tree's water-conducting vessels (xylem) with their spores. The damage to the tree is ultimately caused to a large extent by the bark beetle where the insect larvae and the burrowing activity of the beetle disrupt the sap wood (Christiansen, 1985; Malloch and Blackwell, 1993). For example, the mountain pine beetle (MPB) depends on certain blue stain fungi to interfere with the production of resins (The "pitching out" response) that would plug up the holes beetles are generating in the bark. Once established, additional beetles can readily infect the tree and cause the rapid decline and ultimate demise of the tree. One should note that most blue stain fungi do not cause disease; they are viewed as a nuisance within the forestry sector as they can cause economic loss due to the discoloration of timber related products.

1.2.2. Fungal mitochondrial genomes

Mitochondria evolved from an alpha-proteobacteria endosymbiont during the early stages of eukaryotic evolution (Emelyanov, 2001; Gray *et al*., 2001; Koonin *et al*., 2006). Fungal mitochondrial genomes (mtDNAs) are quite variable in size but appear to share a common set of core genes. Usually mtDNA encode genes for the small subunit ribosomal RNA (mt SSU rRNA also known as *rns*) gene and the mitochondrial large subunit ribosomal RNA (mt LSU rRNA also known as *rnl*) gene, tRNA genes, genes coding for proteins involved in the respiratory chain (*cox1*, *cox2*, *cox3*, and *cob*), subunits of NADH dehydrogenase (*nad1* to *nad6*; except for some members of the Saccharomycetales), and some components of ATP synthase (*atp6, atp8,* and *atp9*) and in some instances a ribosomal protein (*rps3*, sometimes referred to as var1 or S5) (Wolf and Del Giudice, 1988; Bullerwell *et al*., 2003; Eldarov *et al*., 2011; Hausner, 2003).

Historically there were two DED pandemics, the first one was in Europe around 1920 and the second pandemic was in the late 1930s in North America. The first pandemic in Europe was caused by *O. ulmi* and the second pandemic in North America was initially caused by *O. ulmi* but this species was replaced by the more aggressive species *O. novo-ulmi*. Today *O. novo-ulmi* is the causative agent of DED in North America. In Europe one can still find *O. ulmi* and both subspecies of *O. novo-ulmi* (subsp. *novo-ulmi* and subsp. *americana*). Genetic studies showed that there are differences between the mtDNA of both species but overall it appears that *O. novo-ulmi* evolved from *O. ulmi* (Bates *et al*., 1993). Based on mtDNA RFLP analysis it was concluded that in general strains of *O. ulmi* have mtDNAs that are larger than those of *O. novo-ulmi* subsp. *novo-ulmi* and strains of *O. novo-ulmi* subsp. *americana* appear to have the smallest mtDNAs within this species complex (Bates *et al.*, 1993). Charter *et al.* (1996) suggested that the size variation they observed among members of this species complex was due to the presence and absence of mtDNA introns. This was confirmed by other studies such as Gibb and Hausner

(2005), Sethuraman *et al.* (2008) and Hafez and Hausner (2011b) which showed that mtDNA rRNA genes appear to be more intron rich among strains of *O. ulmi* and *O. novo-ulmi* subsp. *novo-ulmi*.

It has also been shown that mtDNA introns can be the cause of mitochondrial diseases in fungi that can attenuate the virulence in plant pathogens such as *O. novo-ulmi* or in the chestnut blight fungus *Cryphonectria parasitica*. Charter *et al.* (1993) and Abu-Amero *et al.* (1995) characterized a mitochondrial plasmid-like element that appeared to be associated with respiratory defects in some strains of *O. novo-ulmi*. Based on cultural observations they suggested that strains that contained this plasmid-like element appeared to be less aggressive thus in the future this element might be potentially useful as a contagious biocontrol agent. Sethuraman *et al.* (2008) examined the sequence for this plasmid-like element and concluded that this element was generated by an intrachromosomal recombination event between two related group I introns within the *rnl* gene, more specifically the mS1669 and mS2449 introns. Both of these introns are group IA1 type introns and their intron core sequences are identical, thus acting like direct repeats. Therefore the presence of related introns within mtDNAs increases the possibility of recombination events that may cause segments to be deleted or in some case inversion events. These events promote mtDNA rearrangements and may cause mtDNA diseases that may attenuate virulence in plant pathogens such as members of the *O. ulmi*/*O. novo-ulmi* species complex. For plant pathogens hypovirulence may actually allow populations to survive as killing its host in the long term will lead to the extinction of the plant pathogen.

The latter is exemplified by hypovirulent strains of *Cryphonectria parasitica* discovered among chestnut trees in the last remaining chestnut tree forests or stands in Michigan (USA) and

Southern Ontario (Canada). This fungus essentially killed off or "crippled" chestnut trees all across the Midwest of the United States with a few remaining stands in Michigan and Ontario. It turned out that these pockets of trees survived because they are infected with hypovirulent strains of the pathogens that can infect but not kill or adversely affect the trees. Instead they act as part of the natural flora that can antagonize more aggressive fungi and thus these hypovirulent strains can protect the trees from Chestnut blight (Bertrand 2000). Two genetic mechanisms can generate hypovirulence: (1) the presence of certain types of mycoviruses; (2) the presence of a group II intron (group IIA with no ORF) in the 5' terminal of the mtDNA *rns* gene. Baidyaroy *et al*., (2011) showed that in some hypovirulent strains this phenotype could be cytoplasmically transmitted in crosses with aggressive strains and the presence of a group II A intron in the *rns* gene was linked with hypovirulence; this intron was shown to splice slowly and thus the production of ribosome components were sharply reduced.

It has been suggested by Bertrand (2000) that mitochondrial instabilities may influence the fitness of plant pathogens and thus trigger hypovirulence. Mitochondrial introns appear to be associated with mitochondrial DNA stability and possible with mtDNA diseases, thus studying the mtDNA introns and their encoded ORFs of members of *Ophiostoma* and related taxa is important as it may help in identifying hypovirulent members that could be useful as biocontrol agents.

Fungal mitochondrial genomes are highly variable in size 19 kb in *Schizosaccharomyces pombe* to 203 kb in *Sclerotinia borealis* (Mardanov *et al*., 2014) they are also highly variable in organization (gene order). This variation is due to various recombination events and the presence of intergenic spacers plus introns and intron-encoded open reading frames (ORFs) (Wolf and Del

Giudice, 1988; Palmer *et al*., 2000; Mardanov *et al*., 2014, Aguileta *et al*., 2014; Wu and Hao, 2014).

1.2.3. Fungal mtDNA introns

Depending on their excision mechanism, fungal mtDNA introns can be assigned to either group I or group II introns (Saldanha *et al*., 1993; Lambowitz *et al*., 1999; Belfort *et al*., 2002; Hafez *et al*., 2013; Sethuraman *et al*., 2013 Hausner *et al*., 2014). Group I and group II introns are ribozymes that tend to be mostly restricted to organellar genomes of eukaryotes (Haugen *et al*., 2005); except for some group I introns that are found in the nuclear rDNA of some fungi and some protozoans (Hausner, 2012).

1.2.3.1. Group I introns

1.2.3.1.1. Distribution

Group I introns have been located in RNA coding genes (rRNA, tRNA) and protein coding genes. Usually they are found in fungal, plant and algal organellar DNAs: especially in ribosomal RNA (SSU and LSU rRNA) genes; genes encoding components of the cytochrome oxidase complexes; and within the transfer RNA (tRNA) genes of chloroplast genomes. They are also found in nuclear rRNA genes in fungi and some protozoans; presumably because the repetitive nature of rDNA and the nucleolus provides a suitable target region for group I intron proliferation (Hafez *et al*., 2012). They have also been observed in bacterial genomes and in bacteriophages (Saldanha *et al*., 1993; Hausner, 2012; Hafez *et al*., 2013; Sethurman *et al*., 2013; Hausner *et al.*, 2014). Recently group I and group II introns have also been recorded in the mtDNAs of metazoans such as the soft corals and sponges (Rot *et al*., 2006; Fukami *et al*. 2007; Huchon *et al*., 2015).

1.2.3.1.2. Group I intron structure

Group I introns can be highly variable with regards to their primary sequence but they do share conserved secondary and tertiary structures. Different categories of GI introns share relatively little sequence similarity but some of them do share short conserved sequences referred to as the P, Q, R, and S elements and these participate in forming helical regions (or paired regions) P is paired with Q (=P4 helix) and R is paired with S (=P7 helix) (Michel *et al*., 1982; Burke *et al*., 1987; Michel and Westhof, 1990; Saldanha *et al*., 1993). The boundaries of GI introns are typically marked by a U residue at the 3' end of the 5' exon and a G residue at the 3' end of the intron (Cech, 1988; Burke, 1988). Overall the secondary structure of group I introns can be arranged into a series of paired (P) regions or helices designated P1 to P10 with P1 and P10 containing the 5' and 3' splice sites, respectively. Some group I intron types have additional helical regions (GI D) and others can miss a paired segment, for example P2 has been noted to be absent in some group I introns. The core structure of a group I ribozyme is composed off two helical domains P4/P6 (P4, P5 & P6), referred to as the scaffolding domain; and P3/P9 (P3, P7, P8 & P9) which forms the catalytic domain, and includes the GTP binding pocket for the exogenous guanosine (GTP) cofactor required to initiate the first transesterification reaction that initiates the excision of the intron from the transcript (Michel and Westhof, 1990; Cech, 1990). Open reading frames (ORFs) can be embedded within the variable loops that connect the paired regions or in some instances ORFs have been observed to extend into the conserved intron core sequences. Many intron ORFs are fused in-frame to the upstream exon, causing the IEP to be translated as a fusion protein (Shub *et al*., 1988; Saldanha *et al*., 1993; Sethuraman *et al*., 2009; Edgell *et al*., 2011).

Group I introns have been classified based on secondary structure characteristics, nucleotide sequences within the conserved core regions, and novel features within the secondary

structure. This has resulted in five categories: Class IA to IE are further subdivided into subclasses (i.e. IA1, IA2 and IA3) (Saldanha *et al*., 1993; Suh *et al*., 1999; Zhou *et al*., 2008). For example, Group IAI introns contain two extra pairings, P7.1/P7.1a or P7.1/P7.2, between P3 and P7; group IB and IC introns typically have a large extension of P5. These extra segments in these structures are assumed to contribute towards interactions that provide additional stabilities to the intron core structures (Michel and Westhof, 1990; Michel *et al*., 1992).

1.2.3.1.3. Splicing

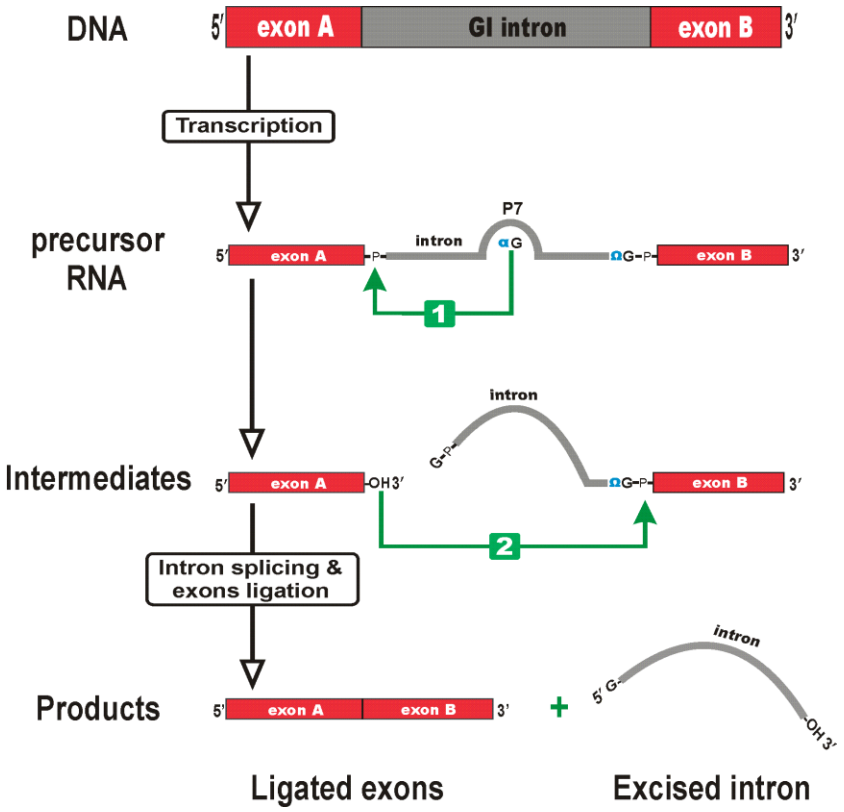
For group I introns the 5' splice site is determined by a short sequence near the 5'-end of the group I intron referred to as the internal guide sequence (IGS), which interacts with a sequence that is located in the upstream exon generating the P1 helix. The 3' splice site is determined by the pairing of a short sequence located in the downstream exon with single stranded regions within the P1 helix forming the P10. The tertiary RNA fold of the intron brings the flanking exons into close proximity with each other and this facilitates the removal of the intron and the splicing of the exons. It should also be noted that *in vivo* protein factors are probably involved in order for splicing to be more efficient by promoting the proper RNA folding or stabilizing the RNA folds of the intron RNA (reviewed by Belfort, 2003; Lang *et al*., 2007; Hausner, 2012). Group I intron splicing involves two trans-esterification steps that can happen in the absence of proteins but it requires an external guanosine (αG) as a cofactor. The αG positions itself into the G-binding pocket (site) located in the P7 pairing region. This exogenous GTPs 3'-OH group acts as a nucleophile and it will attack the phosphodiester bond at the 5'exon/intron junction site (Fig.1.1) and the trans-esterification reaction is completed when the αG binds to the 5'-end of the intron via the formation of by 5'-3' phosphodiester bond. This step requires metal ions (divalent cations) for coordinating the interactions and for catalysis

(Stahley and Strobel, 2005). This first trans-esterification reaction is followed by the 3'-OH of the 5' -exon attacking the 3' intron/exon junction phosphodiester bond, to trigger the second trans-esterification reaction (Michel *et al*., 1989) that excises the intron and ligates the exons. The intron is initially released as a linear molecule with the αG nucleoside attached at the 5' end, however this linear intron RNA can circularise via hydrolysis releasing the 5' ~ 15 nucleotides (Cech, 1990). Eventually this circular form undergoes further fragmentation steps which are assumed to prevent reverse splicing (Saldanha *et al*., 1993; Cech, 1990; Cech *et al*., 1994).

Figure 1.1: Group I intron splicing: **A**: The first reaction is initiated by exogenous guanosine (αG) which attacks the 5' splice site. **B:** The second trans-esterification reaction requires the 3'-OH of the 5' exon which attacks the 3' splice site. **C:** The intron is spliced out in a linear configuration and the exons are ligated together.

This figure based on Hausner (2012; Springer Verlag License Number: 3853820897656) and it was generated based on the information obtained from Saldanha *et al*. (1993); Cech (1990) and Hausner *et al.* (2014).

Group I Intron Splicing



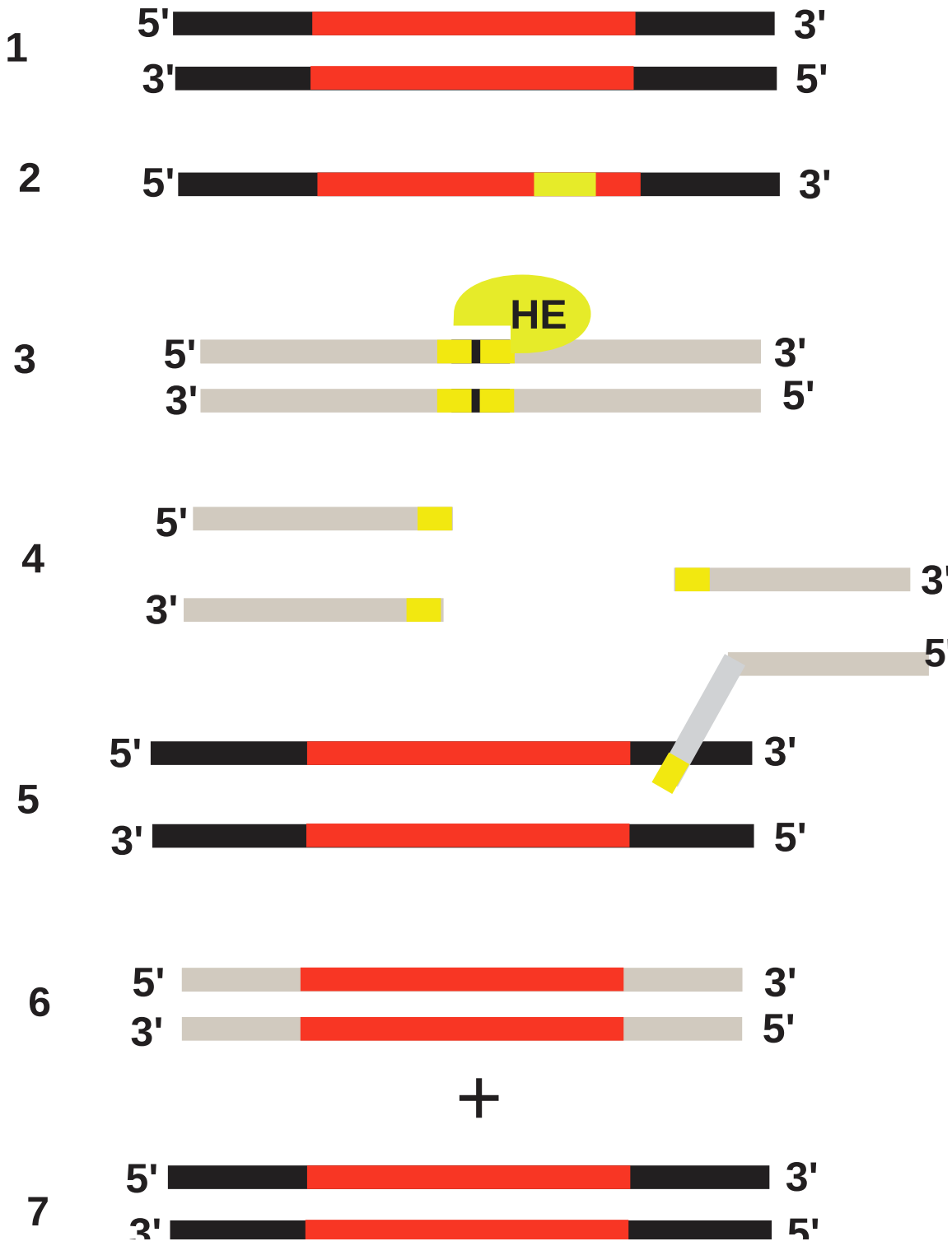
1.2.3.1.4. Group I introns are mobile elements

Group I introns can have different mobility pathways, and in one of them GI introns have the ability to propagate themselves by inserting into predetermined position (cognate alleles) into intron-less genes (i.e. intron homing). It has been observed in yeast that mtDNA group I introns demonstrate dominant transmission (or dominant inheritance) during crosses (this is referred to as "super Mendelian" inheritance) (Dujon, 1989; Dujon *et al*., 1989; Belfort and Perlman, 1995). The intron encoded homing endonucleases provide the means for their mobility (Figure 1.2). The site specific DSB cleavage initiated by the HEases or in some instances a nick, in alleles that lack the intron, can initiate the intron homing event. The host cell's DSB-repair pathways utilize the intron/HEG-containing allele as a template for repairing the DSB. The end result is the nonreciprocal transfer of the group I intron (along with its ORFs) into the intron-minus allele (Belfort *et al*., 2002; Stoddard, 2005; Hausner *et al*., 2014), and this can be associated with co-conversion of markers flanking the intron insertion site (Saldanha *et al*., 1993; Belfort *et al.*, 2002). As described previously the basic principles of intron homing were determined by Dujon's group work on the omega yeast system.

Another mobility pathway, called "reverse splicing" is a phenomenon that uses RNA as an intermediate. Here introns can be inserted into a homologous or heterologous RNA (i.e. ectopic integration). Reverse splicing is associated with the intron IGS sequence interacting with complementary sequences in other RNA sequences. This requires less homology (4-6 nucleotides) compared to the long target sites required by the intron encoded HEases. Reverse splicing requires additional steps such as the reverse transcription of the recombinant RNA and the insertion of the cDNA into the genome by a recombination step that replaces an "intron-less copy" with an intron-plus cDNA (Roman and Woodson, 1995; Bhattacharya *et al*., 2005;

Birgisdottir and Johansen, 2005). To date however, reverse splicing so far has only been demonstrated by *in vitro* experiments (Roman and Woodson, 1995), thus evidence for this more complex pathway remains circumstantial (Bhattacharya *et al*., 2005; Haugen *et al*., 2005).

Figure 1.2: Group I intron mobility: 1: Double stranded DNA (Black) and introns (Red); 2: mRNA 3: intron-encoded endonuclease (Green) recognizing its target sequence in the intronless allele (Gray); 4: cleavage of the target site generating 4 nt 3' overhangs; **5**: the double-strand break repair mechanisms involving homologous recombination; 6: DNA repair pathway is stimulated, and 7: transfer of the intron sequence to the cleaved recipient molecule during the DSB repair process (Hafez and Hausner, 2013).


1
5'
3'
3'
5'
2
5'
3'
HE
3
5'
3'
3'
5'
4
5'
3'
3'
5'
5
5'
3'
3'
5'
6
5'
3'
3'
5'
+
7
5'
3'
3'
5'

1.2.3.2. Group II introns

Group II introns have been reported from fungal mitochondria and plant chloroplast genomes. They can also be found in algal and plant mtDNAs and cpDNAs, in bacteria and in members of the Archaea (Lin *et al*., 1999; Simon *et al*., 2008; Hausner, 2012; Hausner *et al*., 2014). They have also been observed in organellar genomes from various protists (reviewed in Hausner, 2012; Kamikawa *et al*., 2016). Recent reports have also identified group II introns in some animals, for example in the mtDNAs of sponges and in one annelid species (Vallès *et al*., 2008; Huchon *et al*., 2015). So far group II introns have not been located in viruses/phages or in the nuclear genomes of eukaryotes (Palmer and Logsdon, 1991; Michel *et al*., 1989, Lin *et al*., 1999). Most of these introns are in protein coding genes but they are also found in tRNA and rRNA genes (Michel *et al*., 1989; McNeil *et al*., 2016). Group II introns are essentially retroelements and they have been proposed to be plausible ancestors of the spliceosomal nuclear introns and possible non-LTR retrotransposons in eukaryotes (Xiong Y and Eickbush, 1990; Copertino and Hallick, 1993; Xiong and Eickbush, 1990; Zimmerly *et al*., 2001; Lambowitz and Zimmerly, 2004, 2011; Koonin *et al*., 2006; Michel *et al*., 1989; Jacquier, 1990; Zimmerly and Semper, 2015; Sharp, 1991).

1.2.3.2.1. Group II intron secondary structure

Group II introns have conserved secondary and tertiary RNA structures. The secondary structure can be modelled as six stem loop domains (DI – DVI) radiating from a central wheel (Michel and Ferat, 1995; Qin and Pyle, 1998; Pyle and Lambowitz, 2006; Pyle, 2010; Lambowitz and Zimmerly, 2004, 2011; McNeil *et al*., 2016). Domain I is referred to as the scaffold domain is the largest intron domain comprising about half of the ribozyme. Domain I can be subdivided into several smaller stem-loop subdomains (Ia, Ib, Ic1, Ic2, … etc.). Domain I

being the 5’ component of the intron is transcribed first and as it folds it acts as a scaffold guiding in the folding of domains that follow DI (Qin and Pyle, 1998; Fedorova and Zingler, 2007). Sequences within Domain I facilitate multiple tertiary interactions (i.e., α–α’, β–β’, ε–ε’, ζ–ζ’, θ–θ’, λ–λ’, κ–κ’) that can H-bond to specific locations within the other domains and thus DI helps in stabilizing the overall intron RNA structure. Domain I also contains elements required for upstream (5’) exon sequence recognition: the exon-binding sequences EBS1 and EBS2. The EBS sequences can base-pair with corresponding intron-binding sites IBS1 and IBS2 located in the upstream (5’) exon of both group IIA and IIB introns (Michel *et al*., 1989; Pyle, 2010).

Domain II contains key elements that contribute towards group II intron self-splicing namely the θ – θ́ and η – ή tertiary interactions that involve contacts with D I and D VI respectively (Costa *et al*., 1997; Pyle, 2010). Additional important tertiary interactions have been noted to involve Domain III where specific nucleotides form an important interaction (μ - μ’) with DV (Fedorova and Pyle, 2008). Domain III also has been reported to contain a highly phylogenetic conserved internal loop that plays a major role as a catalytic effector that appears to stimulate the activity of group II introns (Fedorova and Pyle, 2008).

Domain V is the most conserved domain in group II introns and can be used to identify group II intron sequences (Lang et al., 2007). DV is viewed as the most important component of a group II intron, catalytic activity. DV and DI are considered the minimal catalytic core requirements of group II introns (Lehmann and Schmidt, 2003; Toor *et al*., 2009). Finally, from a catalytic perspective Domain VI contains the adenine branch-point nucleotide residue, which is involved in the intron self-splicing reaction by promoting the first transesterification reaction that leads to a second transesterification reaction that results in the introns excision in a branched

(lariat) form (Van der Veen *et al*., 1986; Vogel and Bonen, 2002; Lambowitz and Zimmerly, 2004; Pyle and Lambowitz, 2006).

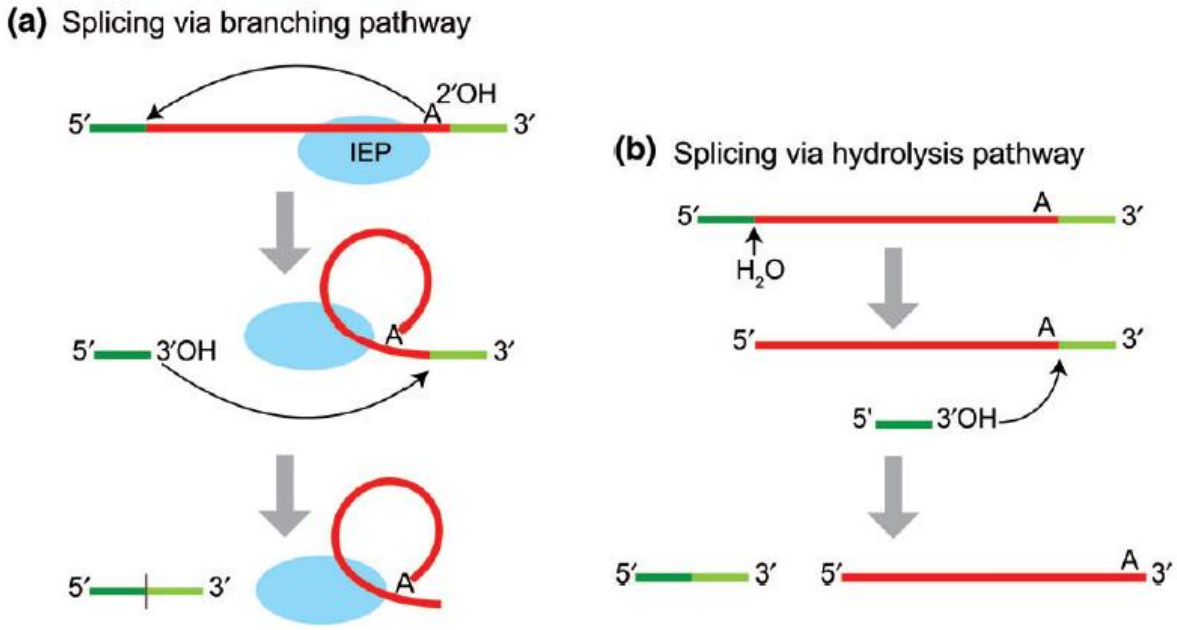
Domains II and D III have been observed to tolerate the insertion for additional sequences without interfering with the function of the ribozyme core (i.e., splicing) as these domains point away from the intron core sequences (Toor *et al*., 2010; Pyle, 2010, Toor and Zimmerly, 2002; Mullineux *et al*., 2010; Hafez and Hausner, 2011a). When present, open reading frames for reverse transcriptase are frequently embedded within Domain IV, a domain that is projected away from the intron core as a result of its being stacked upon D III (Wank *et al*., 1999).

1.2.3.2.2. Group II introns splicing

Group II introns can be spliced in two ways. The branched pathway requires two transesterification reactions and the excised intron RNA assumes a lariat configuration stabilized by a 2'-5' phosphodiester bond (Fig. 1.3a) (Michel and Ferat, 1995; Daniels *et al*., 1996; Van der Veen *et al*., 1986; Pyle and Lambowitz, 2006; Lambowitz and Belfort, 2015; Michel *et al*., 2009; Pyle, 2010; Fedorova *et al*., 2010; Marcia *et al*., 2013; McNeil *et al*., 2016).

The second pathway for group II intron splicing involves a hydrolysis step and here the intron is released in a linear form (Fig. 1.3b). This pathway is assumed to be the dominant splicing mechanisms for bacterial group II C type introns (Michel *et al*., 1989). Similar to group I introns it is assumed that group II introns *in vivo* are assisted in splicing by various protein factors that can be encoded by the intron (maturase domain of RT) or by the host genome (Lambowitz and Zimmerly, 2011).

Figure 1.3. Group II intron splicing reactions: **a**: Standard splicing reaction with two trans-esterifications and an intron released as a lariat intermediate. **b**: Splicing through hydrolysis; where the first step of splicing is initiated by a water nucleophile. This results in the intron being released in a linear form. For this Figure the copyright was obtained from John Wiley & Sons (McNeil *et al*., 2016). License Number: 3853911465390

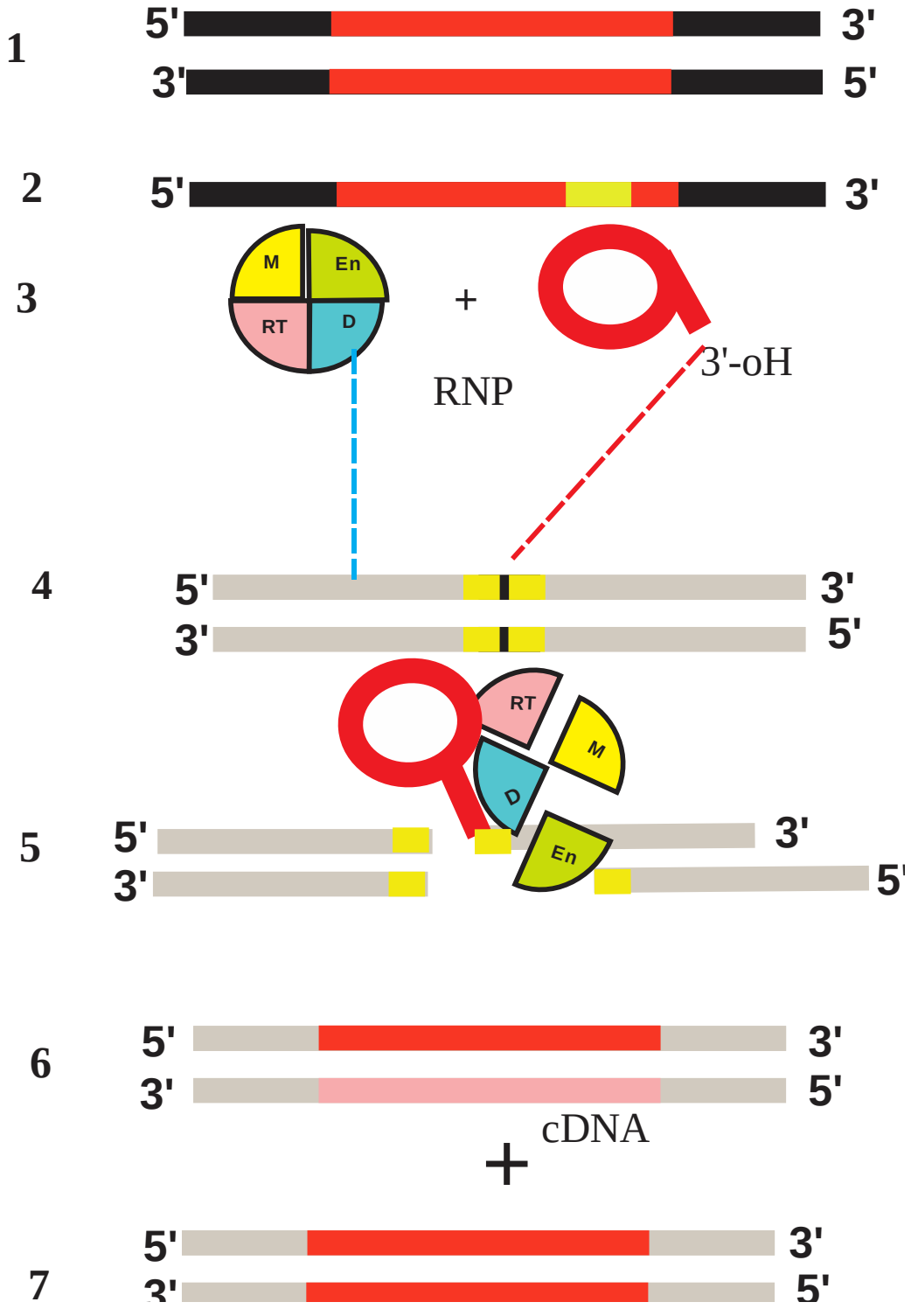

(a) Splicing via branching pathway
5′
2′OH
A
3′
IEP
5′
3′OH
A
3′
5′
3′
A
(b) Splicing via hydrolysis pathway
5′
A
3′
H2O
5′
A
3′
5′
3′OH
5′
3′
5′
A
3′

1.2.3.2.3. Group II introns mobility

Group II intron mobility is called retro-homing and it is initiated by the catalytic activity of the intron RNA and the intron-encoded protein. The intron ORF is translated during expression of the host gene and the IEP consist of four functional domains: RT (reverse transcriptase activity), maturase (X; promotes formation of splicing competent intron structures), DNA binding domain (D), and the endonuclease (EN) domain (which is not found in all group II IEPs). Upon the splicing of the intron the lariat can bind to the translated IEP to form a ribonucleoprotein complex (RNP) (Lambowitz and Zimmerly, 2004) (Figure 1.4). The RNP recognizes a 20-35 bp target homing site in an intron-less cognate allele and the EBS1, EBS2 in the intron base pair with the IBS1, IBS2 elements in the upstream exon, and the first cut is made by the 3'-OH of the intron RNA lariat (Guo *et al*., 1997). This stimulates a reverse splicing reaction for the intron RNA and this inserts (i.e. reverse splicing) the RNA into the top strand of the DNA. The EN domain of the IEP cleaves the bottom strand producing a 3'OH end that works as a primer for reverse transcriptase activity encoded within the RT domain of the IEP. This reaction will essentially generate a cDNA copy of the intron in the "bottom strand"; ultimately the intron insertion (retrohoming) event is completed with the host DNA repair system (Smith *et al*., 2005; Cousineau *et al*., 1998).

Finally it has been noted that group II introns can reverse splice into DNA or RNA as a means of moving into locations without the assistance of an intron encoded reverse transcriptase. However, eventually a reverse transcriptase is required to convert the RNA sequence into DNA (McNeil *et al.*, 2016).

Figure 1.4. Group II intron retrohoming pathway: The black strands represent the donor allele, gray strands represent the recipient allele, and red lines represent the intron sequence. Target sites are represented by the yellow boxes. The intron-encoded protein is a multi-domain protein and it is schematically represented by a circle divided into four quarters (Yellow M: The maturases, green En: Endonuclease, pink RT: Reverse transcriptase and blue D: DNA binding domain); the pink line represents the cDNA generated by the reverse transcriptase. Hafez and Hausner, 2013 was used as the reference to create the figure.


1
5'
3'
3'
5'
2
5'
3'
3
M
En
RT
D
+
RNP
3'-oH
4
5'
3'
3'
5'
RT
M
D
En
5
5'
3'
3'
5'
6
5'
3'
3'
5'
cDNA
+
7
5'
3'
3'
5'

1.2.4. Homing endonucleases

Homing endonucleases are encoded by HEGs and these are typically embedded within self-splicing group I and group II introns and archaeal tRNA introns or as components of inteins (protein introns that splice out at the protein level). HEGs can also be free standing genes. Homing endonucleases are considered rare DNA cutting enzymes that can cut double stranded DNA but they usually require long recognition sites (14-40 bp; Stoddard, 2006; 2011). These long target sites would be rare within a genome and therefore prevent nonspecific binding and cleavage within their host genomes (Belfort *et al*., 1995; Belfort and Roberts, 1997). The homing mechanisms is essentially the same as already described for group I introns that are mobilized by their encoded HEases (Dujon, 1989; Lambowitz and Belfort, 1993; Belfort and Perlman, 1995; Chevalier and Stoddard, 2001). Usually the actual HEases cleavage site may not exactly correspond to the intron insertion site. This is due to the exonuclease activity that is usually associated upon initial cutting by the HEases to extend a gap in order to stimulate the DSB repair system. The repair is associated with gene conversion that "moves the intron" sequence and in some cases flanking sequences.

Homing endonucleases have two important characteristic that appear to help in their survival. First, their long target site requirement ensures specificity which avoids cleavage of essential host genes. The second feature of these proteins is that they can tolerate some sequence polymorphism at their target site (Scalley-Kim *et al*., 2007). The latter would promote lateral transfer of HEGs (or group I introns mobilized by its HEases) between closely related host species (Scalley-Kim *et al*., 2007). It has been recently shown that the variability allowed at the HEase target site is somewhat similar to the wobble positions in protein coding DNA sequences (Scalley-Kim *et al*., 2007; Barzel *et al*., 2011). It is assumed that HEGs are under tight gene

regulation in order to keep expression levels low, and in addition it is assumed that HEGs and HEases are sequestered into organelles to prevent HEases from potentially damaging the host's genome (Jurica and Stoddard, 1999; Gibb and Edgell, 2010; Edgell *et al*., 2011).

1.2.4.1. Homing endonucleases families

Until recently, based on conserved amino acid sequence motifs, homing endonucleases were classified into four major distinct families: the LAGLIDADG, HNH, His-Cys Box, and GIY-YIG families (Belfort *et al*., 2002; Kowalski and Derbyshire, 2002; Stoddard, 2006). However, additional HEases-like proteins/families have been discovered, the PD-(D/E) XK HEases, the Vsr (very-short patch repair) endonucleases (Dassa *et al*., 2009), the Holliday junction resolvase-like HEases (Zeng *et al*., 2009), and the EDxHD family (Taylor *et al*., 2011). Homing endonucleases are found in all domains of life: archaea, eubacteria, plus eukaryotes, and even among the bacteriophages (Belfort and Roberts, 1997; Jurica and Stoddard, 1999; Hafez and Hausner, 2012; Hausner *et al.*, 2014).

1.2.4.2. Applications of Homing endonucleases

Homing endonucleases are meganucleases and they have been viewed as promising candidates for various applications that requires precise cutting of DNA, such as targeted mutagenesis, gene replacements and gene therapy, and genome modifications that can result in pest control (Stoddard, 2006; Hafez and Hausner, 2012; Stoddard 2005, 2011, 2014; Prieto *et al*., 2012; Belfort and Bonocora, 2014).

Target site modification for HEases can be achieved by genome engineering. There are different methods for HE engineering. One of the methods is by mutations of specific residues that involve the DNA-binding domain that can alter the HEases specificity. Another method involves domain shuffling depending on swapping domains of different wild-type HEases.

Recently HEases have been modified depending on the deeper understanding of the available wild-type HEases properties and by employing bioinformatics to redesign HEases with new specificities (Stoddard, 2014). In general retargeting HEases is very challenging as the DNA binding motif and cutting motif can be part of the same protein domain, so altering the DNA binding domain can mitigate cleavage activity (Stoddard 2005).

Recently the focus has been on designing modular type meganucleases, where DNA binding and DNA cutting involves different protein domains. Four different programmable macromolecular scaffolds that can bind DNA and thus generate different site-specificities can now be used for genome editing: zinc finger nucleases (ZFNs; Smith *et al*., 2000; Bibikova *et al*., 2002); transcriptional activator like (TAL) effector nucleases (TALENs) (Christian *et al*., 2010); the RNA guided clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 system (Cong *et al*., 2013; Ran *et al*., 2013); and to some extent LAGLIDADG homing endonucleases (Arnould *et al*., 2007). However, as stated earlier LADGIDADG HEases are challenging with regards to "programmability" to recognize different target sites.

HEases can be used as genome editing reagents based on their ability to generate a site specific DSB which can be repaired by homologous recombination (HR) (Gimble 2007; Stoddard *et al*., 2008; Marcaida *et al*., 2010). However, in many instances the DSB is repaired by non-homologous end joining (NHEJ) that essentially rejoins the free DNA ends. NHEJ is error prone and can result in the deletion or gain of nucleotides the repair junction (Aubert *et al*., 2011; Hafez and Hausner, 2012). The latter therefore is suitable for targeted gene mutations. In the last 15 years there has been considerable efforts undertaken with regards to characterizing natural HEases, genetically modified HEases and synthetic HEases with regards to optimizing these proteins for targeting specific sequences, such as alleles for human monogenic diseases

(Stoddard, 2005, 2011; Stoddard, 2014). Meganucleases, a term currently used to include homing endonucleases, have recently been applied in human gene therapy (Silva *et al*., 2006; Smith *et al*., 2006; Paques and Duchateau, 2007; Pingoud and Silva, 2007; Ashworth *et al*., 2010; Muñoz *et al*., 2012; Pessach and Notarangelo, 2011).

An example of utilizing HEases as a genetic tool is presented by "*Delitto perfetto"* which can be applied to promote site-directed mutagenesis in yeast (*S. cerevisiae*). Here the I-SceI LHEase is used to generate a DSB at the appropriate target site(s) and thus induces the DSB repair process which increases the frequency of targeted homologous recombination by 4,000-fold compared to experiments where DSB events were not generated (Storici and Resnick, 2003).

HEGs and thus HEases have been utilized in gene drive mechanisms that promote the spread of a HEG within a population. The goal can be to manipulate the composition of a population of a pest species by targeting genes that are beneficial with regards to fitness (Burt, 2003; Burt and Trivers, 2006; Henzell *et al*., 2008; Chan *et al*., 2011). Examples of gene drive have been achieved by incorporating a HEG within the genome of a target species such as mosquito and here the HEase was programmed to target genes involved in sex determination (Windbichler *et al*., 2008; 2011). One can envision this strategy being applied to other pests where genes are targeted that are associated with fertility, or efficient pathogen transmission (Deredec *et al*., 2008, 2011; Windbichler *et al*., 2008; Windbichler *et al*., 2011).

Genome editing reagents such as HEases are gaining popularity in plant biotechnology. HEases are applied to developing transformation vector systems, for plant genome editing, and for targeted mutagenesis (Yang *et al*., 2009; Gao *et al*., 2010; Vainstein *et al*., 2011; Zeevi *et al*., 2012). Also rare cutting HEases are used in genomics or in RFLP type analysis where there is a

need for large DNA fragments suitable for pulse field gel electrophoreses analysis (Marcaida *et al*., 2010; Siegl, 2010; Guha and Hausner, 2016).

This thesis has a focussed on characterizing the mtDNA *rns* gene of closely related fungal strains (of the *Ophiostoma piliferum* species complex) to assess the diversity of the introns and their contribution to mtDNA sizes and mtDNA variability among strains of the same species or among strains of different species.

In addition, we characterized three different (but related) HEases with regards to their cleavage activity and their possible DNA target and cleavage sites. Most intron encoded ORFs are predicted by *in silico* methods, very few are actually examined by biochemical means. This provides evidence that introns can encode functional HEases and their ability to generate DSBs and thus supports the idea that they have potential applications in biotechnology.

Chapter 2 has been published:

Bilto, I.M., *Hausner, G., 2016. The diversity of mtDNA *rns* introns among strains of *Ophiostoma piliferum*, *Ophiostoma pluriannulatum* and related species. SpringerPlus 5:1408 doi: 10.1186/s40064-016-3076-6

Original version of the paper was drafted by Bilto and final version was edited by Bilto and Hausner.

The experiments were designed by Bilto and Dr. Hausner and performed and analyzed by Bilto under Dr. Hausner's supervision. And the chapter is verbatim copy of the article.

2. The diversity of mtDNA *rns* introns among strains of *Ophiostoma piliferum*, *Ophiostoma pluriannulatum* and related species

2.1. Abstract

Ophiostoma piliferum, *Ophiostoma pluriannulatum* and related species are causative agents of blue-stain; these fungi colonize the sapwood of trees and impart a dark stain that reduces the value of the lumber. Based on previous studies, it was suspected that the mitochondrial *rns* gene within the Ophiostomatales is rich in introns. The goal of the current study is to examine the mtDNA *rns* intron landscape for these important blue-stain fungi in order to facilitate future annotation of mitochondrial genomes (mtDNA) and to potentially identify mtDNA introns that can encode homing endonucleases which may have applications in biotechnology.

Comparative sequence analysis of the mtDNA *rns* gene identified five intron insertion sites among the ophiostomatoid fungi examined. Positions mS379 and mS952 harbor group II introns, the mS379 intron encodes a reverse transcriptase, and the mS952 intron encodes a potential homing endonuclease. Positions mS569, mS1224, and mS1247 have group I introns inserted and these encode intact or eroded homing endonuclease open reading frames (ORF). Phylogenetic analysis of the intron ORFs showed that they can be found in the same insertion site in closely and distantly related species.

Based on the molecular markers examined (internal transcribed spacers of the rDNA and *rns* introns), strains representing *O. pilifera*, *O. pluriannulatum* and *Ophiostoma novae-zelandiae* could not be resolved. Phylogenetic studies suggest that introns are gained and lost and that horizontal transfer could explain the presence of related intron in distantly related fungi. With regard to the mS379 group II intron, this study shows that mitochondrial group II introns and their reverse transcriptases may also follow the life cycle previously proposed for group I introns

and their homing endonucleases.This consists of intron invasion, decay of intron ORF, loss of intron, and possible reinvasion.

2.2. Introduction

Fungal mitochondrial genomes (mtDNAs) usually encode genes for the ribosomal small and large subunit RNAs (*rns*, *rnl*), tRNA, proteins involved in the respiratory chain (*cox1*, *cox2*, *cox3*, and *cob*), subunits of the NADH dehydrogenase (*nad1* to *nad6* and *nad4L*; except for members of the Taphrinomycota and some members of the Saccharomycetales), components of the ATP synthase (*atp6*, *atp8*, and *atp9*), and in some instances ribosomal proteins (*rps3*) (Bullerwell *et al.*, 2003; Bullerwell and Lang, 2005; Procházka *et al*., 2010; Solieri, 2010; Eldarov *et al*., 2011; Freel *et al*., 2015). On a few occasions, mtDNA mutations can trigger senescence and in some fungal plant pathogens, they may also trigger hypovirulence (Bertrand, 2000). With regard to metazoans, mtDNA mutations can cause degenerative mitochondrial diseases in humans (Wallace, 2010). Fungal mtDNAs are highly variable both in size and organization due to various recombination events and the presence of intergenic spacers, introns, and intron-encoded open reading frames (ORFs) (Palmer *et al*., 2000; Mardanov *et al*., 2014, Aguileta *et al*., 2014; Wu and Hao, 2014; Freel *et al*., 2015).

Fungal mtDNA introns tend to be self-splicing elements that can catalyze their own excision from transcripts and depending on the excision mechanism, they have been divided into group I and group II introns (Saldanha *et al*., 1993; Lambowitz *et al*., 1999). Splicing of group I (GI) and II (GII) introns tend to be facilitated by a combination of intron-encoded (maturases) or host genome-encoded factors (Lang *et al*., 2007; Hausner, 2012). Some of these introns have the potential to be mobile due to the presence of intron-encoded proteins (IEPs) that promote the movement of their host introns from intron-containing alleles to cognate alleles that lack the

intron (Dujon, 1989). Homing endonucleases (HEases) are DNA-cutting enzymes encoded by homing endonuclease genes (HEGs) and these are frequently encountered as ORFs within GI introns and in some instances within GII introns (Toor and Zimmerly, 2002; reviewed in Hafez and Hausner, 2012). HEGs can also be freestanding, encoded within archaeal introns, and comprise the DNA-cutting component of inteins (Gimble, 2000; Belfort *et al*., 2002; Stoddard, 2005; Barzel *et al*., 2011). Currently, at least six families of HEases are recognized. Their naming is based on conserved amino acid motifs: the LAGLIDADG, H-N-H, His-Cys box, PD-(D/E)xK, EDxHD, and GIY-YIG families of HEases (Stoddard, 2005, 2011, 2014).

Group I and II introns are highly variable in their primary structure but both show conservation in their secondary structures. For GI introns, about 10 helical regions have been noted (P1 to P10) that stabilize the intron core in folding into a splicing-competent structure (reviewed in Hausner *et al.*, 2014). Group I introns can be assigned into various subgroups based on features related to secondary or tertiary structures and sequence peculiarities (Michel and Westhof, 1990). Group II introns tend to form secondary structures that consist of six double-helical domains (domains I to VI) radiating from a central wheel. Domain V is the most conserved component with regard to the primary sequence (reviewed in Toor *et al*., 2001). Group II introns are assigned into several different classes based on structural features and the type of interactions between intron and exon sequences (Lambowitz and Belfort, 2015). Group II introns are retroelements that encode proteins with reverse transcriptase activity (RT). Usually, GII intron mobility is promoted by a ribonucleoprotein consisting of the IEP and the spliced lariat version of the intron RNA (Lambowitz and Zimmerly, 2011). With regard to GI intron-encoded proteins, there are two families of HEases that are commonly encountered within fungal mtDNAs. These are the LAGLIDADG (LHE) and GIY-YIG families of HEases (Stoddard,

2005, 2011). It is worth noting that for some LAGLIDADG type ORFs, it has been shown that they can function as maturases or in some cases have two activities, promote splicing and mobility of their host intron (Szczepanek and Lazowska, 1996; Bolduc *et al*., 2003).

Some *Ophiostoma* species are blue stain fungi [e.g., *Ophiostoma piliferum* (Fr.) Syd. &P.Syd.] and some are plant pathogens [e.g., *Ophiostoma ulmi* (Buisman) Melin & Nannf. that cause Dutch elm disease (Wingfield *et al*., 1993)]. Blue stain fungi cause discoloration of wood and this reduces the economic value of the lumber. Currently, very little is known about the mtDNAs for species of *Ophiostoma* and the contribution of so-called mobile introns toward mtDNA stability and diversity. For *Cryphonectria parasitica* [(Murrill) M.E. Barr], there is evidence that an *rns* GII intron could be associated with inducing hypovirulence (Baidyaroy *et al*., 2011). Thus, mapping and characterizing introns may have applications with regard to attenuating virulence for pathogenic members of the genus *Ophiostoma*. Furthermore, ribozymes and HEases have been shown to have applications in biotechnology (Sullenger and Gilboa 2002; Hafez and Hausner 2012, 2015). Uncovering more HEGs and autocatalytic introns adds to the reservoir of elements that can be developed into RNA trans-cleaving agents, genome editing tools, agents for targeted mutagenesis, etc. (reviewed in Hafez and Hausner, 2012; Hausner *et al*., 2014; Stoddard, 2014; Guha and Hausner, 2016).

Previously, Hafez *et al.* (2013) surveyed the NCBI database along with sequences from species of *Ophiostoma* to assemble an *rns* intron landscape in order to identify positions that have been invaded by introns. The current study examines additional members of the genus *Ophiostoma* with a focus on *O. piliferum*, *O. pluriannulatum*, and related species to gain a better understanding of the presence of introns and their encoded ORFs.

2.3. Materials and methods

2.3.1. Maintenance of fungal cultures and DNA extraction

Fungal strains examined in this study are listed in Table1. All reagents, unless noted were obtained from ThermoFisher Scientific Canada. Fungi were grown on malt extract agar (MEA; per liter: 1 g of yeast extract, 30 g of malt extract, and 20 g of agar) plates. In order to generate biomass for DNA extraction, fungi were grown in peptone yeast glucose broth (PYG; per liter: 1 g of peptone, 1 g of yeast extract, and 3 g of D-glucose). An Erlenmeyer flask containing fifty ml of PYG was inoculated with agar blocks (~1x1 mm) derived from an agar plate culture and the liquid cultures were incubated in the dark at 20 °C for 4 to 6 days. The protocol used for DNA extraction was previously described by Hausner *et al*. (1992). Briefly, fungal mycelia were harvested by vacuum filtration using Whatman filter paper #1 and collected in 15 ml centrifuge tubes. Six ml of extraction buffer [100 mM Tris-HCl (pH 8.0), 20 mM $Na_2EDTA.2H_2O$ (pH 8.0), 1.5 M NaCl, 2 % (w/v) cetyltrimethylammonium bromide (CTAB)] was added along with 3 g of glass beads and 660 µl of 20 % (w/v) sodium dodecyl sulfate (SDS). This mixture was vortexed for about 2 min. Thereafter, the lysate was incubated at 55 °C for 2 hours. Cell debris and contaminants were removed by chloroform extraction and centrifugation for 20 min at 2000 rpm. The top aqueous layer was recovered and 2.5 volume of ice cold 95 % ethanol was added to precipitate the nucleic acids. Finally, the nucleic acids were recovered by centrifuging for 30 min at 3000 rpm. The DNA pellets were washed with 1 ml of 70% ethanol and resuspended in 300 µl TE buffer [10 mMTris-HCl (pH 7.6) and 1 mM $Na_2EDTA.2H_2O$ (pH 8.0)].

Table 1: Strains used in this study

	Species	Strain number	*rns* length (kb)
1	*Ophiostoma pluriannulatum*	WIN(M)1530 (= NZPS 1555)	1.2
2	*Ophiostoma pluriannulatum*	WIN(M)1531 (= NZPS 1552)	1.2
3	*Ophiostoma pluriannulatum*	WIN(M)1529 (= NZPS 1553)	1.2
4	*Ophiostoma pluriannulatum*	WIN(M)1572(= DAOM 175754)	1.2
5	*Ophiostoma perfectum*	WIN(M)823 (= CBS 636.66)	7
6	*Ophiostoma piliferum*	WIN(M)1543	1.2
7	*Ophiostoma californicum*	WIN(M)505	4.6
8	*Ophiostoma carpenteri*	WIN(M)853	5
9	*Ophiostoma carpenteri*	WIN(M)855 = UAMH 9695	5
10	*Ophiostoma subannulatum*	WIN(M)539	4.6
11	*Ophiostoma novae-zelandiae*	WIN(M)869 (= UAMH 9559)	1.2
12	*Ophiostoma pluriannulatum*	WIN(M)1561	1.2
13	*Ophiostoma pluriannulatum*	WIN(M)455 (= ATCC 8714)	1.2
14	*Pesotum* sp.	WIN(M)163	1.2
15	*Ophiostoma piliferum*	WIN(M)1548	1.2
16	*Ophiostoma novae-zelandiae*	WIN(M)864 (= UAHM 9557)	1.2
17	*Ophiostoma pluriannulatum*	WIN(M)1549	1.2
18	*Ophiostoma pluriannulatum*	WIN(M)1552	1.2
19	*Ophiostoma novae-zelandiae*	WIN(M)863 = UAMH 9556	1.2
20	*Ophiostoma piliferum*	WIN(M)971	1.2
21	*Ophiostoma piliferum*	WIN(M)972	4.6

22	*Ophiostoma piliferum*	WIN(M)973	1.2
23	*Ophiostoma piliferum*	UAHM 7459	1.2
24	*Ophiostoma piliferum*	UAMH 7233	1.2

ATCC = American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108, USA; **CBS** = Centraal Bureau voor Schimmelcultures, Utrecht, The Netherlands; **DAOM**, Cereal and Oilseeds Research, Agriculture & Agri-Food Canada, Ottawa, Ont., Canada; NZPS = from Colette Breuil (University of British Columbia) collected by Roberta Farrell, University of Waikato, New Zealand; **UAMH** = University of Alberta Microfungus Collection & Herbarium, Devonian Botanic Garden, Edmonton, AB, Canada, T6G 2E1; **WIN(M)** = University of Manitoba (Winnipeg) Collection

Table 2: Primers used for amplifying segments of the mtDNA *rns* gene (Designed by Hafez and Hausner)

Primer name	**Sequence (5′ to 3′)**
rns-F0	GAGTTTGGTGATGGCTCTG
rns-F1	GCTGCCAGCAGTCGCGG
rns-F2	GGATTAGAGACCCTTGTAG
rns-F3	ACACCAGTAGTGAAGTATG
rns-R0	CCACTACACGAACCGTATTTC
rns-gp2-R1	CATTAACTGGAAACAGCCGTGCAAC
rns-R2	CTACAAGGGTCTCTAATCC
rns-R3	CCGCGACTGCTGGCACG
mtsr-1	AGTGGTGTACAGGTGAG
mtsr-2	CGAGTGGTTAGTACCAATCC

2.3.2. PCR amplification of the nuclear rDNA ITS region and *rns* gene and DNA sequencing

Primers and conditions for the polymerase chain reaction (PCR) for obtaining the nuclear internal transcribed spacer (ITS) rDNA regions and mtDNA *rns* sequences were previously described by Hausner and Wang (2005) and Hafez and Hausner (2011a), respectively. Briefly, the ITS (ITS1 and ITS2) regions were amplified using the following primers: SSU-Z/LSU-4 and SSU-3/LSU-2 (Hausner and Wang 2005). Primers utilized for the amplification of the mtDNA *rns* gene were previously described by Hafez and Hausner (2011) and Hafez *et al*. (2013); additional primers were designed for sequencing in order to extend reads and to close contigs (Table 2). All PCR products where analyzed on 1% agarose gels by electrophoresis in TBE buffer [89 mM Tris-borate, 10 mM EDTA (pH 8.0)]. The 1-Kb Plus DNA Ladder (Invitrogen) was included as molecular weight markers. Gels were stained with ethidium bromide (0.5 µg/ml in 1X TBE buffer) and examined under UV light. PCR products were cleaned up using the Gene aid kit (Froggo Bio, 230 Canarctic Drive, Toronto, ON, Canada) following the protocol provided by the manufacturer. PCR fragments were sequenced by cycle sequencing utilizing the BigDye® Terminator v3.1cycle sequencing kit (Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404). Automated fluorescent DNA sequence analysis was performed by the Manitoba Institute of Cell Biology DNA Sequencing Facility [675 McDermot Ave., Cancer Care Manitoba (CCMB) Building].

2.3.3. Preliminary DNA sequence analysis

Sequencing results were assembled as FASTA-formatted files and applied to the sequence assembly program CAP3 (http://doua.prabi.fr/software/cap3). The resulting contigs were used as templates for designing new primers. Once completed, the *rns* sequences were

applied to the basic local alignment search tool (BLAST; http://www.ncbi.nlm.nih.gov/Blast.cgi/) to identify similar sequences. Sequences were compiled and aligned by the online multiple sequence alignment program, MAFFT (http://mafft.cbrc.jp/alignment/server/; setting: E-INS-I; Katoh and Standley, 2013). The aligned sequences were further examined with the GeneDoc program (Nicholas *et al*., 1997) to determine intron/exon boundaries. This was accomplished by aligning sequences with introns against those that lack introns.

2.3.4. Intron folding

Folding of GI and GII introns was based on previous work by Michel and Westhof (1990), Toor *et al*. (2001), Lambowitz and Zimmerly (2004) and Hafez and Hausner (2011), along with input from the online program RNAweasel (http://megasun.bch.umontreal.ca/RNAweasel/). The literature was also consulted with regard to comparing introns identified in this study with introns and their putative folds, characterized in other studies (Hafez *et al*., 2013 and citations within). Intron folds compiled at the RNA comparative web site (http://www.rna.icmb.utexas.edu/) were also consulted. Intron sequences (along with flanking exon sequences) were applied to RNAweasel (Lang *et al*., 2007). This program can identify and, in many cases, classify introns. The program can also predict the intron core sequences as it relates to the secondary structure (Lang *et al*., 2007). The mfold online program (http://unafold.rna.albany.edu/?q=mfold; Zuker, 2003) was also used to assist in folding sections of the introns. ORF Finder (NCBI; set on genetic code 4) was used to identify potential ORFs within intron sequences. The final intron RNA folds were manually drawn using CorelDRAW Graphics Suite X6 (Corel Corporation, Ottawa).

2.3.5. Phylogenetic studies

Nucleotide and amino acid alignments were analyzed with programs contained within the MEGA 6 program package (Tamura *et al.*, 2011). For all aligned data sets, the most suitable models for phylogenetic analysis were chosen based on the "best model" option as implemented in MEGA 6. The ITS data were aligned with ClustalX (Thompson *et al*., 1997) and the alignment was adjusted with GeneDoc (Nicholas *et al*., 1997). The ITS alignment was analyzed with neighbor joining (NJ, Maximum Composite Likelihood model), parsimony (PARS), and the maximum likelihood (ML, T92+G model) methods. In all cases, the bootstrap option was implemented (1000 replicates) in order to assess node support values. The ITS data were also analyzed with Mr. Bayes (F81 model) running 5000000 generations and removing (burn-in command) 40% of sampled trees to compute the majority-rule consensus tree.

Datasets for intron-encoded ORFs were enriched by extracting sequences from NCBI databases using sequences obtained from this study as queries in blastp searches. Datasets were aligned with MAFFT and manual adjustments were made, if necessary, with GeneDoc. Amino acid (aa) alignments were analyzed with three tree building methods [neighbour joining (NJ), parsimony (PARS), and maximum likelihood (ML)] and the bootstrap option was implemented (2000 replicates) in order to assess the level of support for the tree topologies generated by the respective methods. With regard to NJ analysis, the maximum composite likelihood method and its defaults were selected along with the complete deletion of gaps option. For the PARS method, the complete deletion option was also selected. In the ML method, the best model was first determined with the "best model" function as implemented in MEGA. Therefore, for ML analysis, the Whelan and Goldman plus Freq. model was selected and the complete deletion option was selected to remove segments of the alignments that contained gaps. All outgroups were chosen based on being the most distant member in the dataset.

2.4. Results

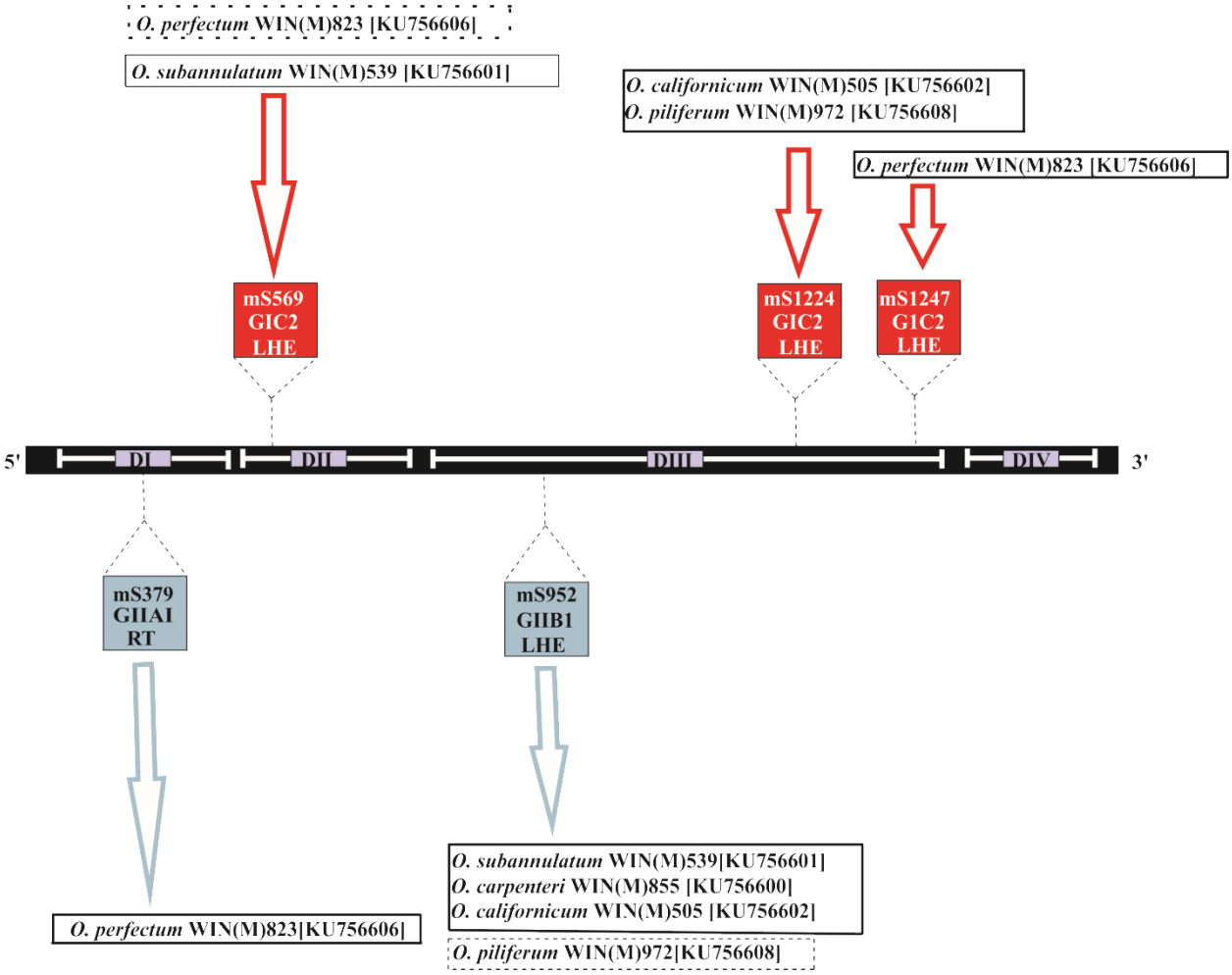
2.4.1. Introns within the mtDNA *rns* gene

A PCR-based survey revealed polymorphism with regard to the size of the *rns* gene among the strains of *Ophiostoma* examined in this study. Based on previous reports (Hafez and Hausner, 2011), PCR products of 1.2 kb are representative of *rns* genes that do not have insertions and *rns* derived amplicons greater than 1.2 kb have insertions (introns). Among the 23 strains surveyed, only the following strains, based on the size of *rns*-derived PCR amplicons, appear to have insertions within the *rns* gene: *Ophiostoma californicum* WIN(M)505 [(DeVay, R.W. Davidson & W.J. Moller) G. Hausner, J. Reid & Klassen], *Ophiostoma pluriannulatum* WIN(M)539 [(Hedgc.) Syd.& P. Syd.], *Ophiostoma carpenteri* WIN(M)855 [J. Reid & Georg Hausner], *Ophiostoma perfectum* WIN(M)823 [(R.W. Davidson) de Hoog], and *O. piliferum* WIN(M)972. The sizes of the amplicons were as follows: *O. perfectum* WIN(M)823 yielded a 7 kb PCR amplicon while *O. pluriannulatum* WIN(M)539, *O. piliferum* WIN(M)972, *O. californicum* WIN(M)505, and *O. carpenteri* WIN(M)855 yielded *rns*-derived PCR amplicons of about 5kb (see Table 1).

As summarized in (Fig. 2.1) DNA sequence analysis showed that the *rns* gene of *O. perfectum* WIN(M)823 contains three introns at the following positions (with respect to the *E. coli* 16S rRNA gene; Johansen and Haugen, 2001): mS379, mS569, and mS1247.The naming of introns is based on the convention of Johansen and Haugen (2001), where “m” stands for mitochondria, “S” for small ribosomal subunit gene, and the number indicates the position of the intron with reference to the *E. coli* 16S rRNA sequence. *O. piliferum* WIN(M)972 and *O. californicum* WIN(M)505 have introns inserted at positions mS952 and mS1224. *O. carpenteri*

WIN(M)855 has one intron at position mS952. *O. pluriannulatum* WIN(M)539 also has an intron at mS952 in addition to an intron inserted at mS569.

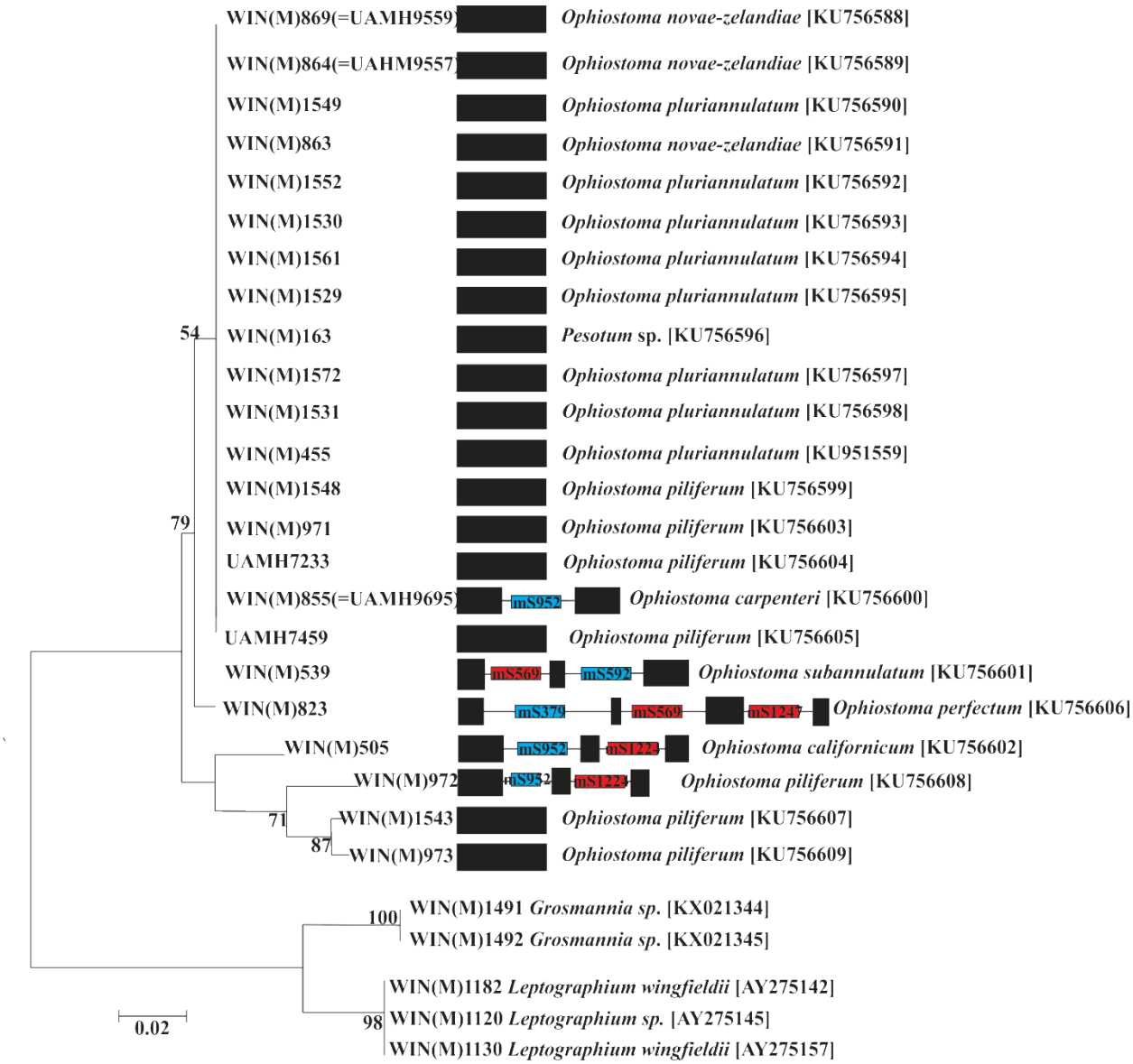
Figure 2.1. Schematic diagram depicting the mtDNA *rns* gene and its introns insertion sites for *Ophiostoma* species. The diagram shows five insertion sites and the intron classes [GroupI (GI) and Group II (GII)], intron types and the intron-encoded proteins are indicated (RT = reverse transcriptase, LHE = LAGLIDADG type endonuclease). The *rns* strucutral domains are indicated by Roman numbers (I-IV) (see Hafez *et al*., 2013). The red boxes represent GI introns and the blue boxes represent GII introns. Names surrounded by boxes drawn with dashed lines indicate the presence of degraded LHEs, whereas those in boxes with solid lines represent introns with intact LHEs. GenBank accession numbers are also indicated.


O. perfectum WIN(M)823 [KU756606]
O. subannulatum WIN(M)539 [KU756601]
O. californicum WIN(M)505 [KU756602]
O. piliferum WIN(M)972 [KU756608]
O. perfectum WIN(M)823 [KU756606]
mS569 GIC2 LHE
mS1224 GIC2 LHE
mS1247 G1C2 LHE
5'
DI
DII
DIII
DIV
3'
mS379 GIIAI RT
mS952 GIIB1 LHE
O. perfectum WIN(M)823[KU756606]
O. subannulatum WIN(M)539[KU756601]
O. carpenteri WIN(M)855 [KU756600]
O. californicum WIN(M)505 [KU756602]
O. piliferum WIN(M)972[KU756608]

2.4.2. ITS region analysis

Phylogenetic analysis of the nuclear ITS rDNA region was used to infer relatedness of the investigated *Ophiostoma* species (Fig. 2.2). The ITS region was sequenced for *O. piliferum* strains along with related taxa such as *Ophiostoma novae-zelandiae, O. californicum*, *Ophiostoma subannulatum*, *O. carpenteri*, *O. perfectum*, and *O. pluriannulatum* (GenBank accession numbers: KU756588, KU756609, KX021344, KX021345). The rDNA ITS region of the studied species ranged between 518-640 bp and the alignment showed the presence of numerous indels within the ITS1 and ITS2 segments. Sequences for strains representing *Leptographium* and *Grosmannia* were included as outgroups. All *Ophiostoma* species were clearly separated from the outgroup species. The taxonomy of *O. piliferum* is a complex issue (Hausner *et al*., 1993; Schroeder *et al*., 2001; Hausner *et al*., 2003; de Beer *et al*., 2013) and it has been suggested that *O. novae-zelandiae* could be a synonym of *O. pluriannulatum* (Thwaites *et al*., 2005). The ITS phylogenetic analysis failed to accommodate the various species within their own clades (Fig. 2.2). In addition, although the intronless species appear to be grouped together based on the ITS sequences, the node support values for the tree do not support monophyly for this cluster.

Figure 2.2. Phylogeny based on nuclear ITS rDNA region for *Ophiostoma* species examined in this study and the schematic representation of the intron/exon combinations noted in the corresponding mtDNA *rns* genes.The phylogenetic tree is based on an ITS region alignment comprising 23 strains that belong to the genus *Ophistoma*. The black boxes represent exons of the *rns* gene and the black lines represent the coresponding introns. The blue boxes are for GII introns and the red boxes are for G1 introns (see Fig.1.1). Tree topology is based on Mr. Bayes (MB) analysis and percentages at the nodes are node support values based in posterior probabilities (F81 model; 5000000 generations and burn-in of 40%). Names of organisms and GenBank accession numbers are provided. The branch lengths are based on MB analysis and are proportional to the mean number of substitutions per site (see scale bar).

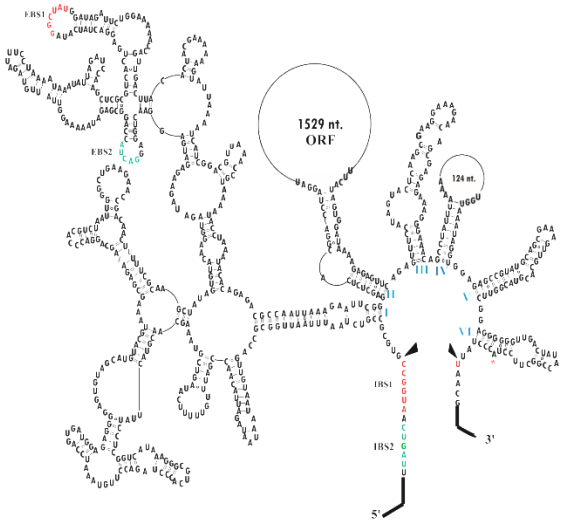

WIN(M)869(=UAMH9559)
Ophiostoma novae-zelandiae [KU756588]
WIN(M)864(=UAHM9557)
Ophiostoma novae-zelandiae [KU756589]
WIN(M)1549
Ophiostoma pluriannulatum [KU756590]
WIN(M)863
Ophiostoma novae-zelandiae [KU756591]
WIN(M)1552
Ophiostoma pluriannulatum [KU756592]
WIN(M)1530
Ophiostoma pluriannulatum [KU756593]
WIN(M)1561
Ophiostoma pluriannulatum [KU756594]
WIN(M)1529
Ophiostoma pluriannulatum [KU756595]
54
WIN(M)163
Pesotum sp. [KU756596]
WIN(M)1572
Ophiostoma pluriannulatum [KU756597]
WIN(M)1531
Ophiostoma pluriannulatum [KU756598]
WIN(M)455
Ophiostoma pluriannulatum [KU951559]
WIN(M)1548
Ophiostoma piliferum [KU756599]
79
WIN(M)971
Ophiostoma piliferum [KU756603]
UAMH7233
Ophiostoma piliferum [KU756604]
WIN(M)855(=UAMH9695)
mS952
Ophiostoma carpenteri [KU756600]
UAMH7459
Ophiostoma piliferum [KU756605]
WIN(M)539
mS569
mS592
Ophiostoma subannulatum [KU756601]
WIN(M)823
mS379
mS569
mS1247
Ophiostoma perfectum [KU756606]
WIN(M)505
mS952
mS1223
Ophiostoma californicum [KU756602]
WIN(M)972
mS952
mS1223
Ophiostoma piliferum [KU756608]
71
WIN(M)1543
Ophiostoma piliferum [KU756607]
87
WIN(M)973
Ophiostoma piliferum [KU756609]
100
WIN(M)1491 Grosmannia sp. [KX021344]
WIN(M)1492 Grosmannia sp. [KX021345]
WIN(M)1182 Leptographium wingfieldii [AY275142]
98
WIN(M)1120 Leptographium sp. [AY275145]
WIN(M)1130 Leptographium wingfieldii [AY275157]
0.02

2.4.3. The *rns* introns of *Ophiostoma perfectum*

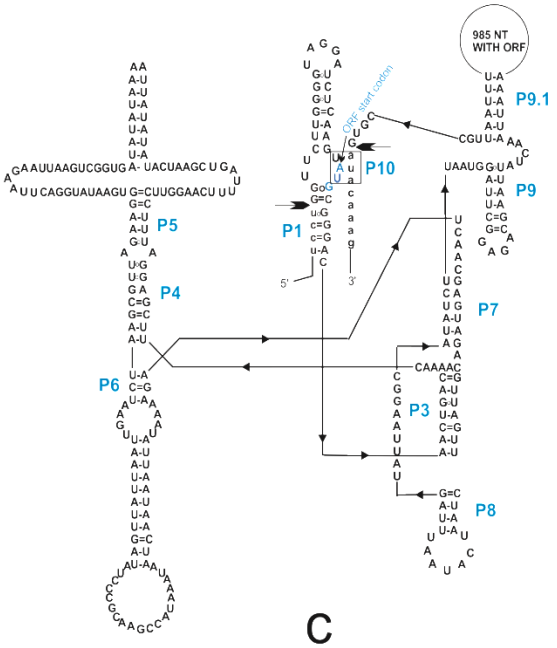
The *rns* gene of *O. perfectum* WIN(M)823 was observed to be the most complex with regard to intron arrangements (Fig.2.1). It includes a GII intron encoding a reverse transcriptase (RT) located at position mS379 and GI introns encoding LADLIDADG type ORFs positioned at mS569 and mS1247. These introns were investigated in more detail by characterizing the potential secondary folds (Fig. 2.3 a, b, c) and by evaluating the evolutionary relationships of these intron ORFs. The mS379 intron is a GII intron, type A, based on the inter domain joiners that are represented by δ and δ' interactions, which stabilize the tertiary structure (Fig. 2.3 a) (Lambowitz and Belfort, 2015). The intron fold shows that the exon binding sites (EBS1, EBS2) within domain I are complementary to intron binding sites (IBS1, IBS2) in the upstream *rns* exon (Toor *et al*, 2001). This intron has an ORF embedded within domain II and the ORF encodes a RT-like protein (510aa). The RT ORF appears to be complete with four domains that are typically associated with GII introns: RT domain, maturase domain (X), DNA binding domain (D), and an endonuclease domain (En) (Lambowitz and Zimmerly, 2004).

The mS569 intron, based on structural features, is a GI C2 type (Michel and Westhof, 1990) intron (Fig. 2.3 b) as it contains P5a and P5b, and no P2. The mS569 intron contains a fragmented LHE ORF due to premature stop codons. The start codon is located within the stem of P5b and the original stop codon appears to be located within the P9.1 loop. Another GI type C2 intron has been noted at position mS1247 and it also has a degraded LHE but with a putative start codon located within the P1 stem (Fig. 2.3 c).

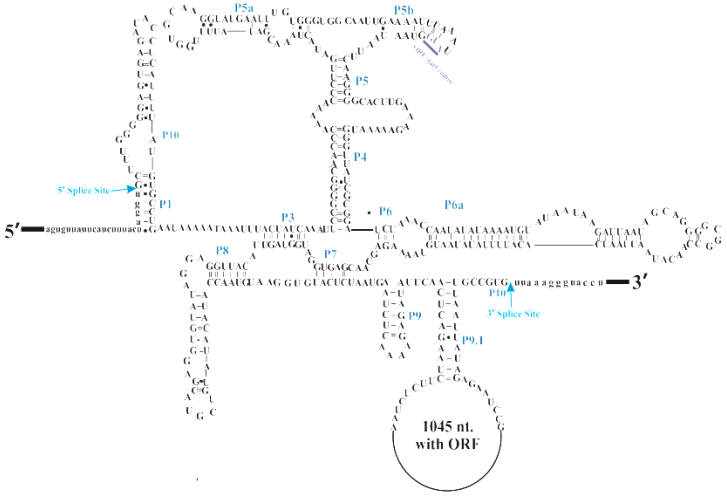
Figure 2.3. The predicted RNA folds for the *O. perfectum rns* introns. (a) Group II intron class A1 inserted at mS379; (b) a group I intron class C2 inserted at mS569; (c) group I intron class C2 inserted at mS1247. Domains (I to VI) and the exon and intron binding sequence (EBS and IBS) segments are shown for the group II introns. For the group I introns the helices (P1 to P10) and conserved sequence elements are labelled. The group II intron (a) encodes a reverse transcriptase-type ORF within domain II. The two group I introns encode LAGLIDADG type ORFs; however, these ORFs appear to be degenerated due to the presence of premature stop codons (see text).


1529 nt.
ORF
124 nt.
EBS1
EBS2
IBS1
IBS2
5'
3'

a


985 NT
WITH ORF
P1
P3
P4
P5
P6
P7
P8
P9
P9.1
P10
ORF start codon
5'
3'

c

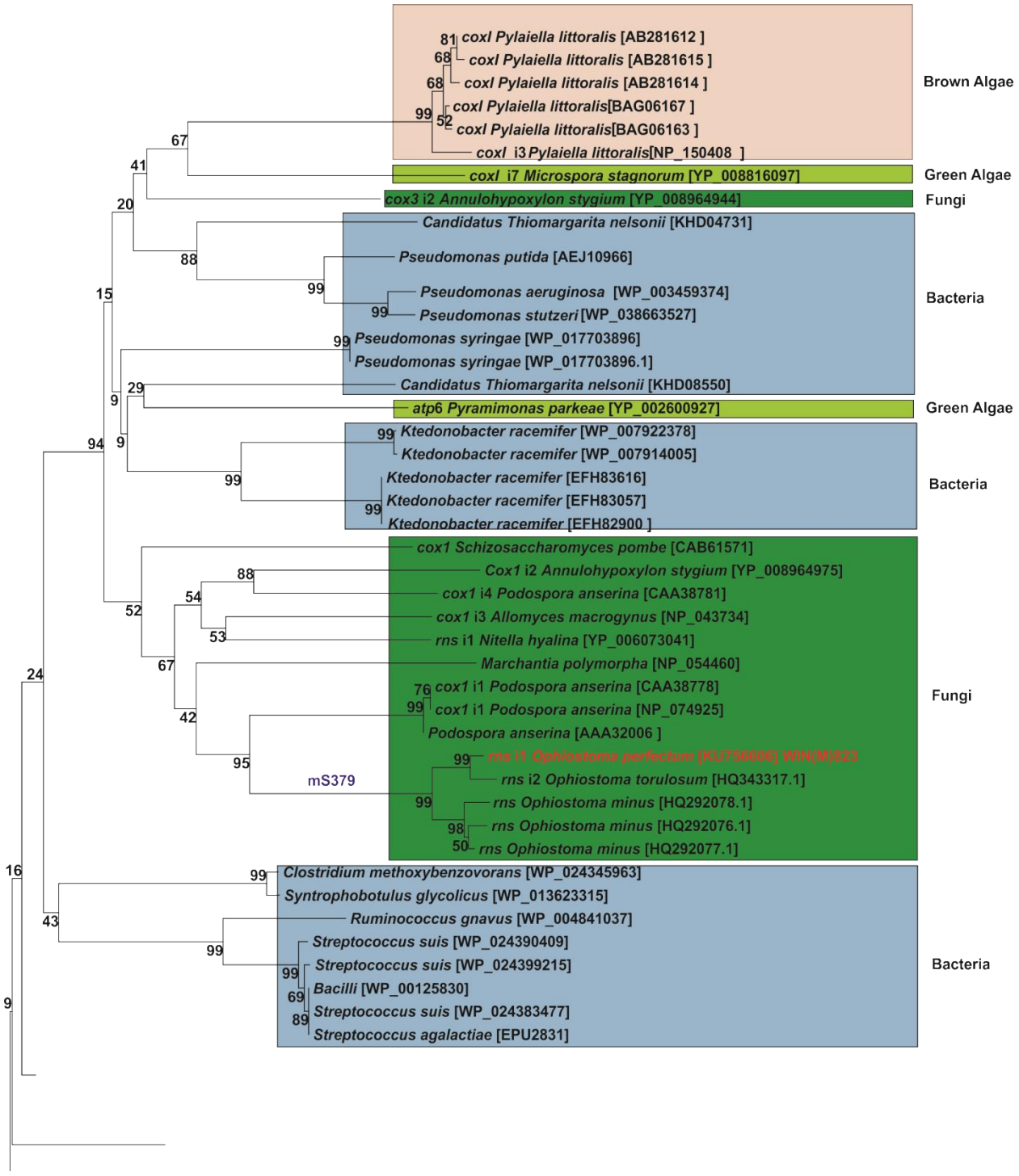

P5a
P5b
P5
P4
P10
P1
P3
P6
P6a
P8
P7
P9
P9.1
P10
5' Splice Site
3' Splice Site
1045 nt.
with ORF

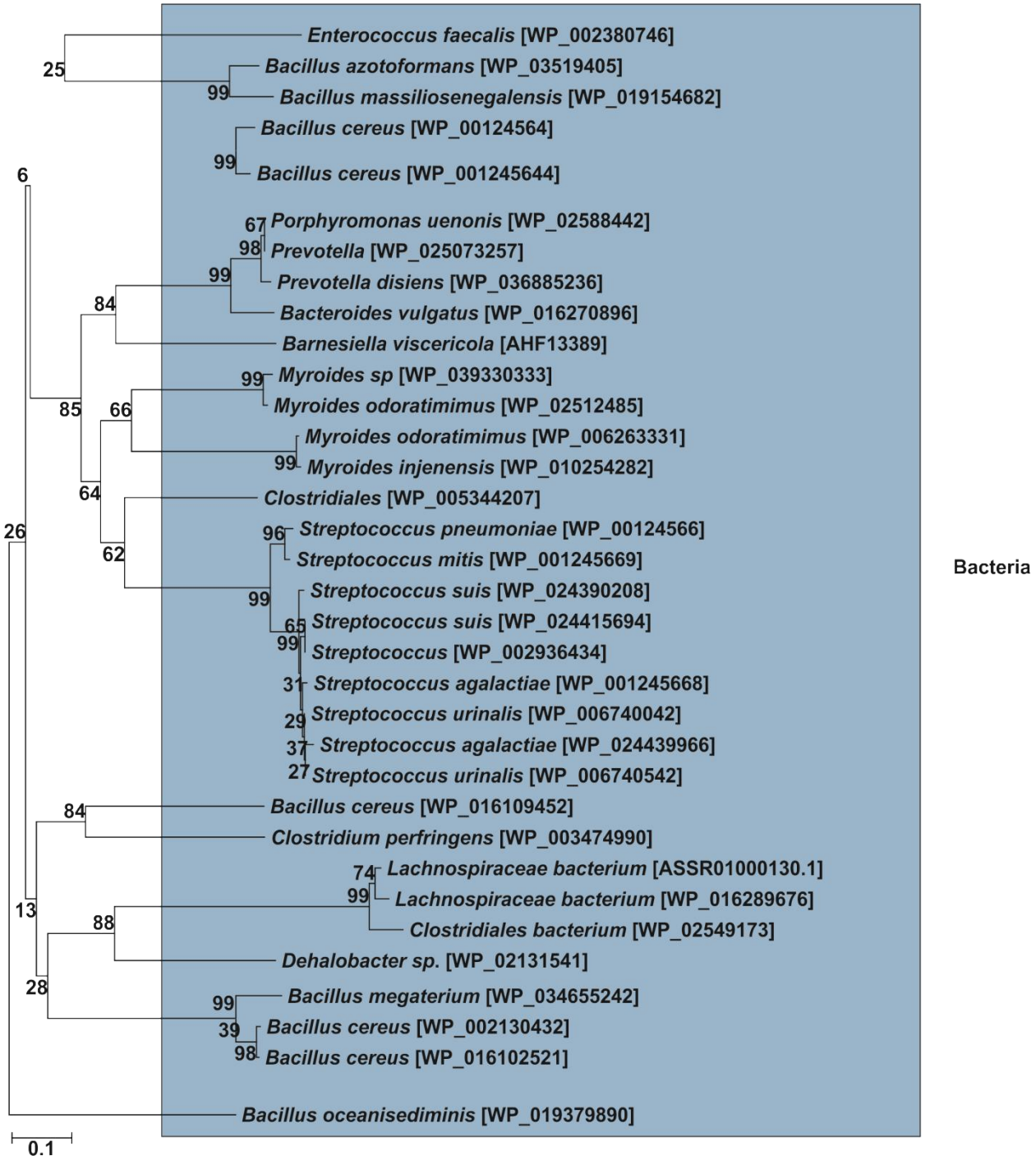
b

2.4.4. Phylogenetic analysis of *rns* intron-encoded open reading frames.

The ORF Finder program identified an intact GIIA RT ORF (510 aa) in the *O. perfectum* mS379 intron. The phylogenetic tree generated for the RT dataset included various fungal and bacterial group II ORFs, with *Bacillus oceanisediminis* serving as the outgroup (Fig. 2.4). The trees generated by NJ, ML and PARS were similar in topology and showed that the mS379 *O. perfectum* ORF is related to the mS379 ORF found in *Ophiostoma torulosum* [(Butin & G. Zimm.) Georg Hausner, J. Reid & Klassen] with strong bootstrap support (99%) and both of these sequences are related to the mS379 ORFs from the *Ophiostoma minus* [(Hedgc.) Syd.& P. Syd.] *rns* genes. The phylogenetic tree also shows that mS379 RT ORFs are related to an RT ORF that exists in the *cox1* gene of *Podospora anserina*. There are additional RT ORFs, mostly encoded within *cox1* (except of the *Nitella hyalina rns* intron-encoded RT ORF introns) but these appear to show monophyly with the above grouping with a low node support value (< 50%). A node that received high bootstrap support (94%) suggests monophyly for RT ORFs encoded within bacterial genomes such as *Pseudomonas syringae, Pseudomonas putida, Pseudomonas aeruginosa, Ktedonobacter racemifer*, and the *cox1* and *cox3* intron-encoded RT ORFs from various fungi, brown algae (*Pylaiella littoralis*) and liverwort (*Marchantia polymorpha*) along with the fungal mS379 ORFs.

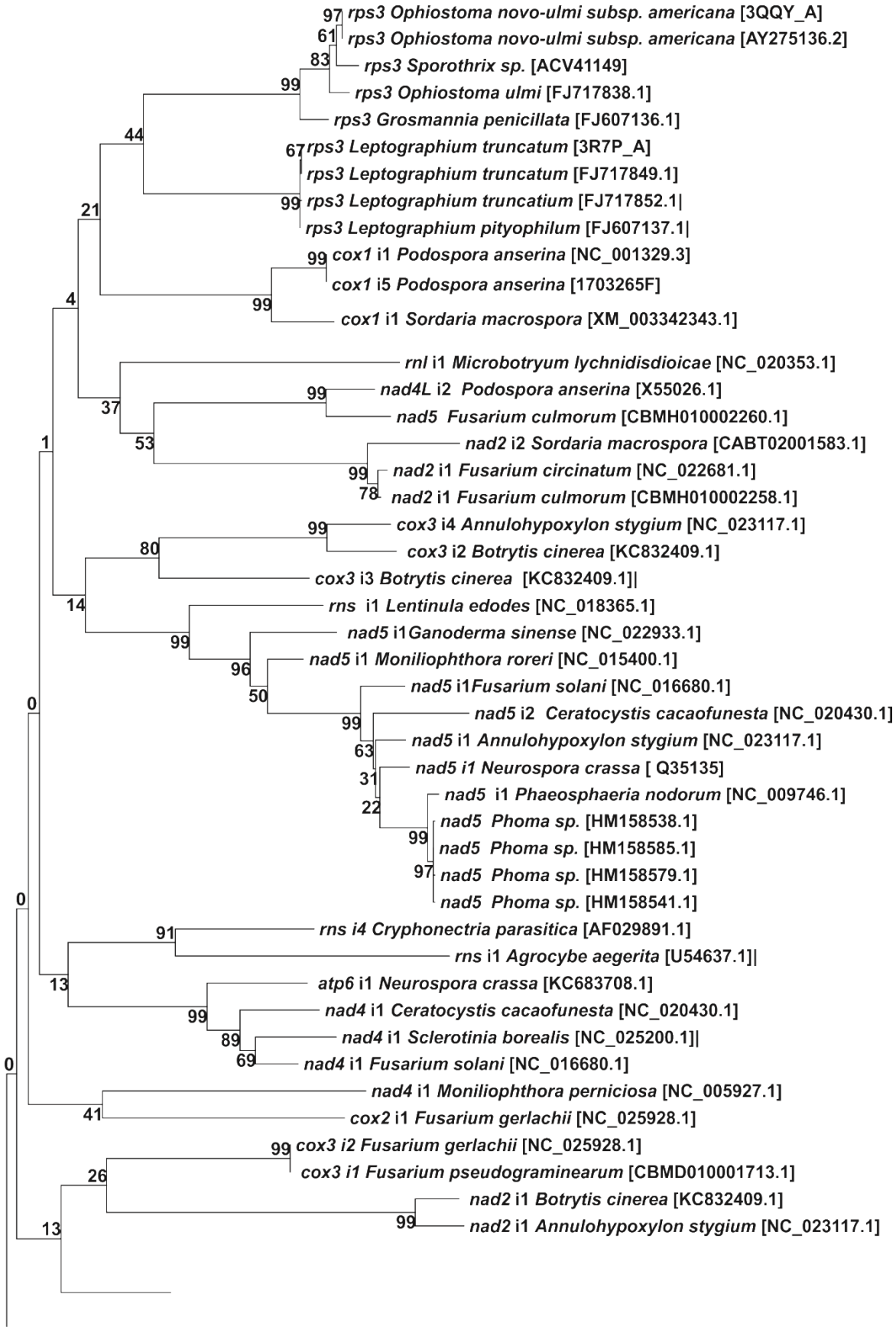
Figure 2.4. A phylogenetic tree showing the relatedness of mS379 intron ORFs among group II intron-encoded reverse transcriptase amino acid sequences. Tree topology is based on ML analysis and percentages at the nodes are node support values based on bootstrap analysis (1000 replicates). Names of organisms, host genes, intron number and location/position (when available) and GenBank accession numbers are provided. The branch lengths are based on ML (Whelan and Goldman plus Freq. model) analysis and are proportional to the mean number of substitutions per site (see scale bar).

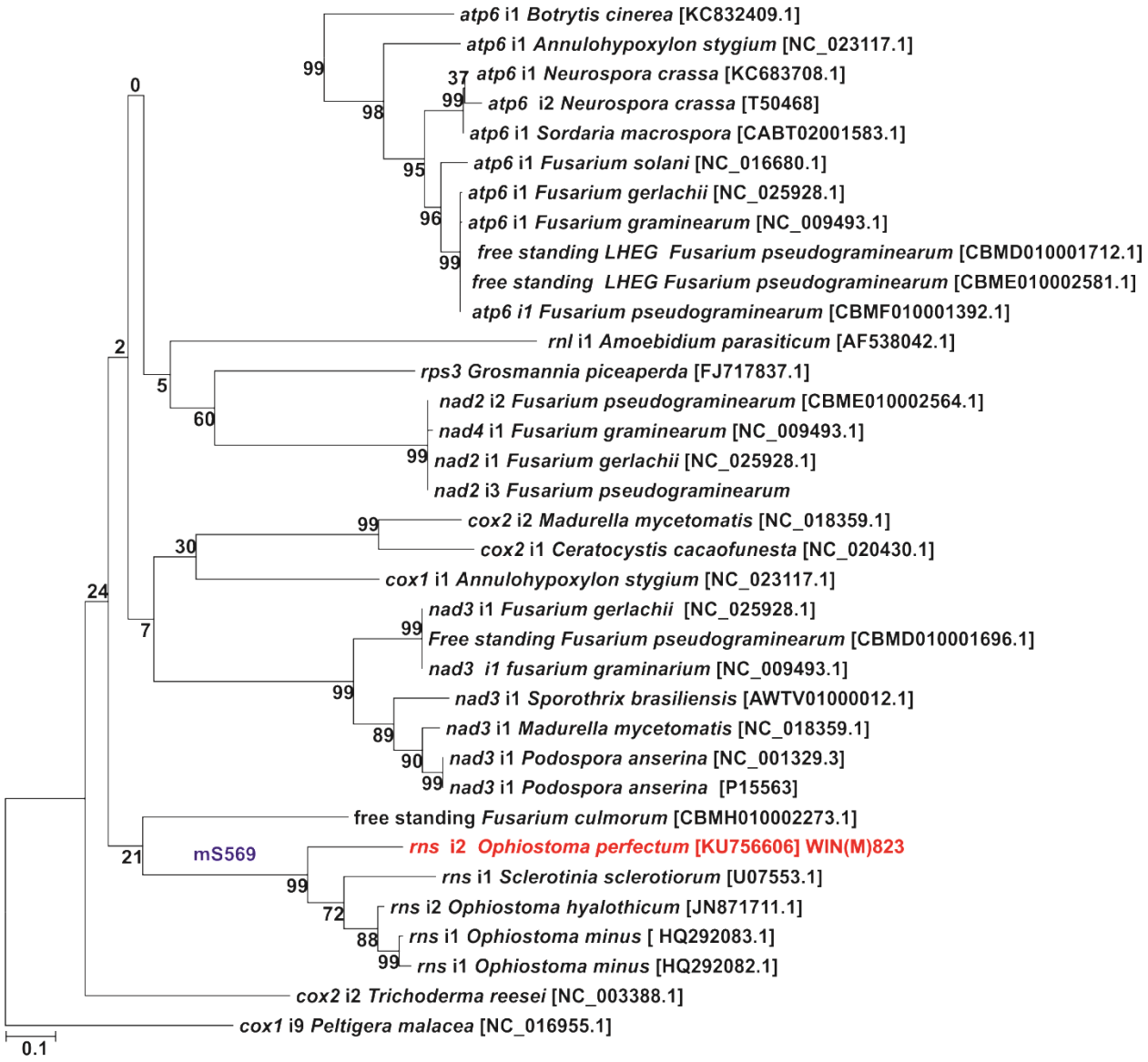

coxI Pylaiella littoralis [AB281612]
coxI Pylaiella littoralis [AB281615]
coxI Pylaiella littoralis [AB281614]
coxI Pylaiella littoralis[BAG06167]
coxI Pylaiella littoralis[BAG06163]
coxI i3 Pylaiella littoralis[NP_150408]
Brown Algae
coxI i7 Microspora stagnorum [YP_008816097]
Green Algae
cox3 i2 Annulohypoxylon stygium [YP_008964944]
Fungi
Candidatus Thiomargarita nelsonii [KHD04731]
Pseudomonas putida [AEJ10966]
Pseudomonas aeruginosa [WP_003459374]
Pseudomonas stutzeri [WP_038663527]
Pseudomonas syringae [WP_017703896]
Pseudomonas syringae [WP_017703896.1]
Candidatus Thiomargarita nelsonii [KHD08550]
Bacteria
atp6 Pyramimonas parkeae [YP_002600927]
Green Algae
Ktedonobacter racemifer [WP_007922378]
Ktedonobacter racemifer [WP_007914005]
Ktedonobacter racemifer [EFH83616]
Ktedonobacter racemifer [EFH83057]
Ktedonobacter racemifer [EFH82900]
Bacteria
cox1 Schizosaccharomyces pombe [CAB61571]
Cox1 i2 Annulohypoxylon stygium [YP_008964975]
cox1 i4 Podospora anserina [CAA38781]
cox1 i3 Allomyces macrogynus [NP_043734]
rns i1 Nitella hyalina [YP_006073041]
Marchantia polymorpha [NP_054460]
cox1 i1 Podospora anserina [CAA38778]
cox1 i1 Podospora anserina [NP_074925]
Podospora anserina [AAA32006]
rns i1 Ophiostoma perfectum [KU756606] WIN(M)823
rns i2 Ophiostoma torulosum [HQ343317.1]
rns Ophiostoma minus [HQ292078.1]
rns Ophiostoma minus [HQ292076.1]
rns Ophiostoma minus [HQ292077.1]
Fungi
mS379
Clostridium methoxybenzovorans [WP_024345963]
Syntrophobotulus glycolicus [WP_013623315]
Ruminococcus gnavus [WP_004841037]
Streptococcus suis [WP_024390409]
Streptococcus suis [WP_024399215]
Bacilli [WP_00125830]
Streptococcus suis [WP_024383477]
Streptococcus agalactiae [EPU2831]
Bacteria


Enterococcus faecalis [WP_002380746]
Bacillus azotoformans [WP_03519405]
Bacillus massiliosenegalensis [WP_019154682]
Bacillus cereus [WP_00124564]
Bacillus cereus [WP_001245644]
Porphyromonas uenonis [WP_02588442]
Prevotella [WP_025073257]
Prevotella disiens [WP_036885236]
Bacteroides vulgatus [WP_016270896]
Barnesiella viscericola [AHF13389]
Myroides sp [WP_039330333]
Myroides odoratimimus [WP_02512485]
Myroides odoratimimus [WP_006263331]
Myroides injenensis [WP_010254282]
Clostridiales [WP_005344207]
Streptococcus pneumoniae [WP_00124566]
Streptococcus mitis [WP_001245669]
Streptococcus suis [WP_024390208]
Streptococcus suis [WP_024415694]
Streptococcus [WP_002936434]
Streptococcus agalactiae [WP_001245668]
Streptococcus urinalis [WP_006740042]
Streptococcus agalactiae [WP_024439966]
Streptococcus urinalis [WP_006740542]
Bacillus cereus [WP_016109452]
Clostridium perfringens [WP_003474990]
Lachnospiraceae bacterium [ASSR01000130.1]
Lachnospiraceae bacterium [WP_016289676]
Clostridiales bacterium [WP_02549173]
Dehalobacter sp. [WP_02131541]
Bacillus megaterium [WP_034655242]
Bacillus cereus [WP_002130432]
Bacillus cereus [WP_016102521]
Bacillus oceanisediminis [WP_019379890]
0.1
Bacteria

The aligned dataset that contained the mS569 intron ORFs (Fig. 2.5), as found in both *O. perfectum* WIN(M)823 and *O. pluriannulatum* WIN(M)539, showed that the ORF is degraded in *O. perfectum* due to some frame shift mutations but otherwise nearly identical to the intact sequence found in *O. pluriannulatum*. The *O. perfectum/O. pluriannulatum* mS569 ORFs grouped at high node support values with *rns* intron-encoded ORFs located within mS569 introns in the following fungi: *Sclerotinia sclerotiorum*, *O. hyalothecium* and *O. minus*. This phylogenetic tree with regards to the deeper nodes overall received poor node support values, so no further conclusions could be made about the mS569 ORFs and their relationships to other intron ORFs.

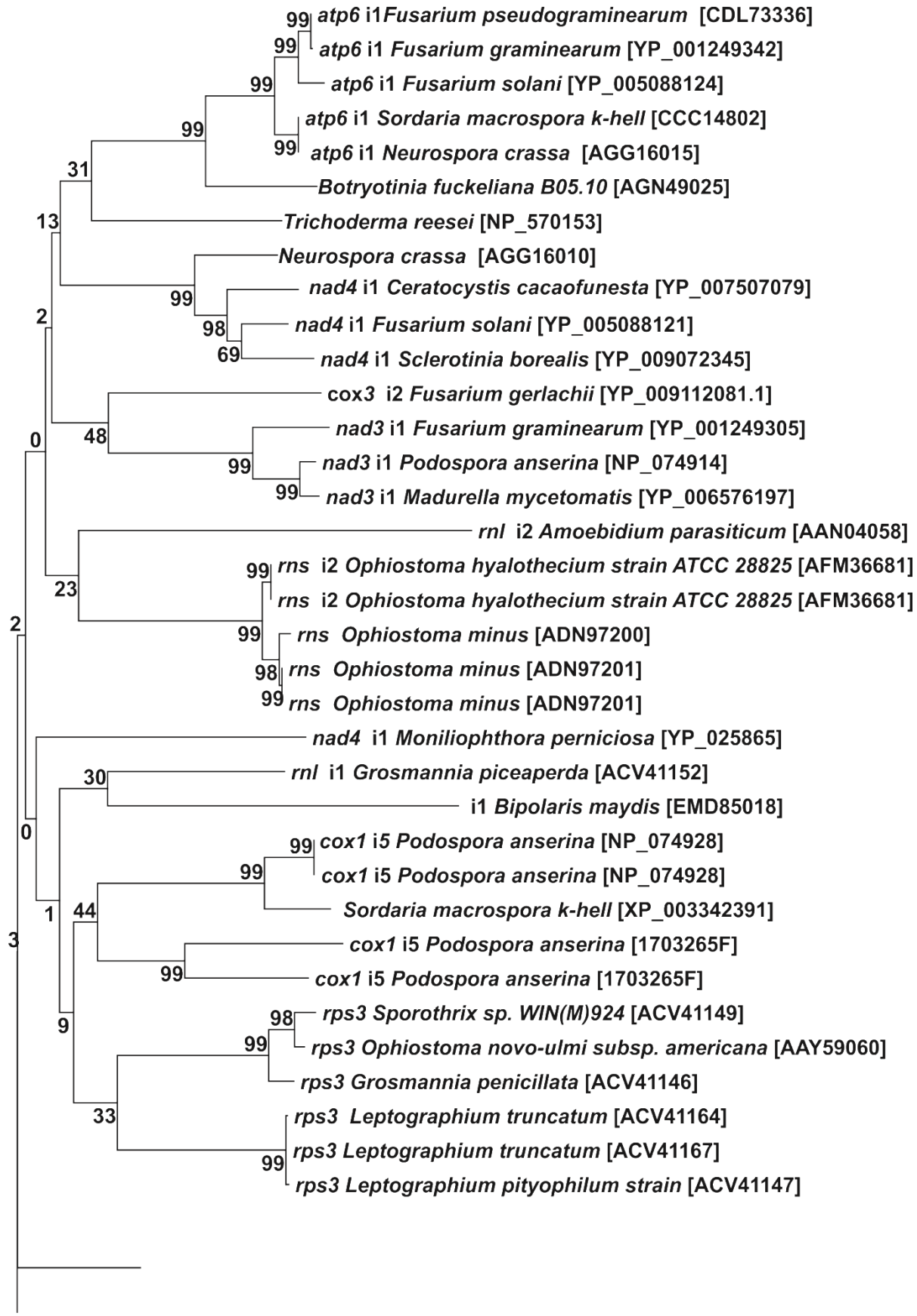
Figure 2.5. A phylogenetic tree showing the relatedness of mS569 intron-encoded ORF a.a squences among a set of GI intron-encoded LAGLIDADG ORFs. Tree topology is based on ML analysis and percentages at the nodes are node support values based on bootstrap analysis (1000 replicates). Names of organisms, host genes, intron number and location/position (when available) and GenBank accession numbers are provided. The branch lengths are based on ML (Whelan and Goldman plus Freq. model) analysis and are proportional to the mean number of substitutions per site (see scale bar). A data set was assembled by extracting related sequences from GenBank depending on the *O. perfectum* WIN(M)823 sequence data as a query. The alignment was analyzed with three programs contained within MEGA6 (ML, NJ, and PARS). *Peltigera malacea* was used as an outgroup.

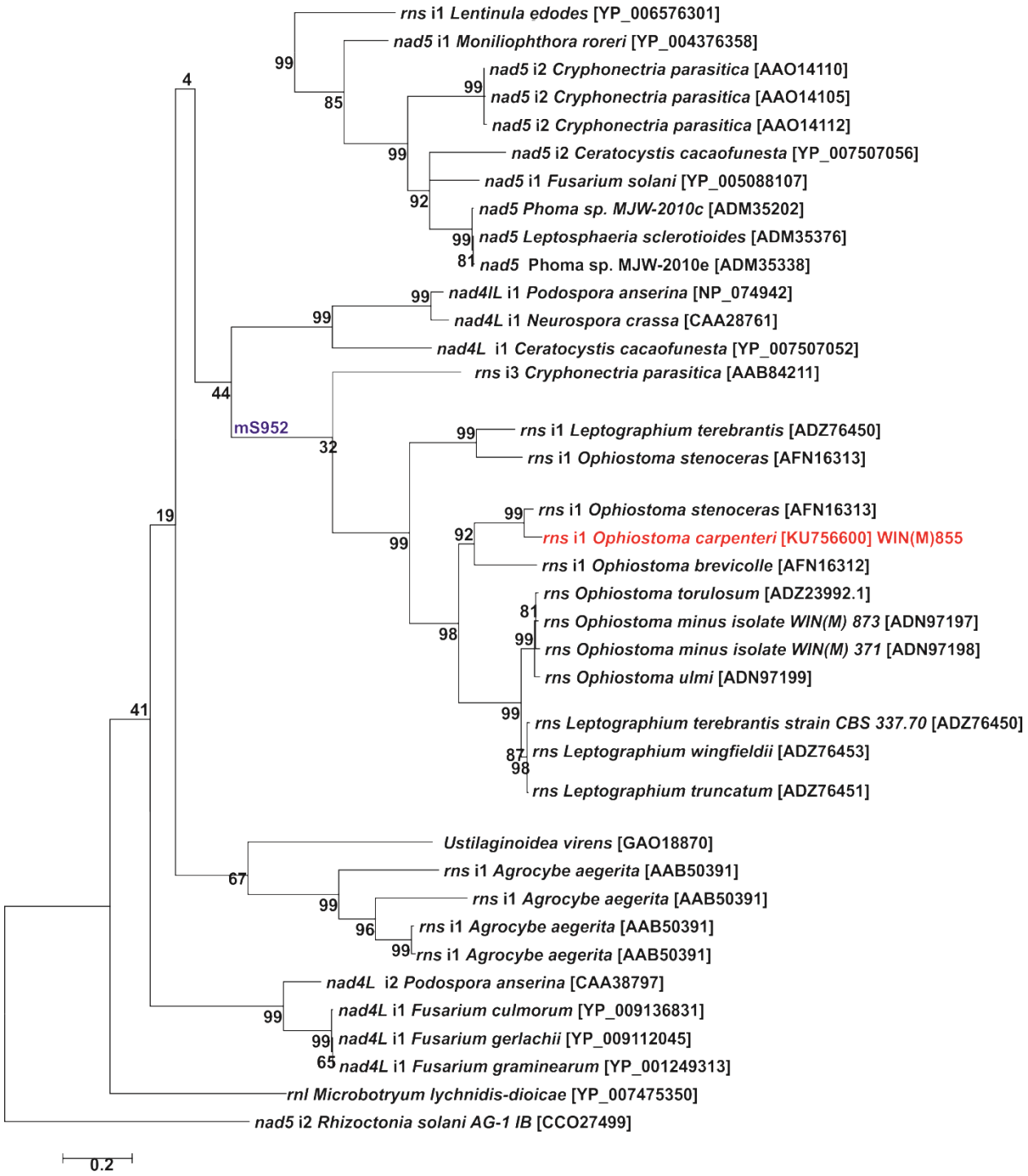

rps3 Ophiostoma novo-ulmi subsp. americana [3QQY_A]
rps3 Ophiostoma novo-ulmi subsp. americana [AY275136.2]
rps3 Sporothrix sp. [ACV41149]
rps3 Ophiostoma ulmi [FJ717838.1]
rps3 Grosmannia penicillata [FJ607136.1]
rps3 Leptographium truncatum [3R7P_A]
rps3 Leptographium truncatum [FJ717849.1]
rps3 Leptographium truncatium [FJ717852.1|
rps3 Leptographium pityophilum [FJ607137.1|
cox1 i1 Podospora anserina [NC_001329.3]
cox1 i5 Podospora anserina [1703265F]
cox1 i1 Sordaria macrospora [XM_003342343.1]
rnl i1 Microbotryum lychnidisdioicae [NC_020353.1]
nad4L i2 Podospora anserina [X55026.1]
nad5 Fusarium culmorum [CBMH010002260.1]
nad2 i2 Sordaria macrospora [CABT02001583.1]
nad2 i1 Fusarium circinatum [NC_022681.1]
nad2 i1 Fusarium culmorum [CBMH010002258.1]
cox3 i4 Annulohypoxylon stygium [NC_023117.1]
cox3 i2 Botrytis cinerea [KC832409.1]
cox3 i3 Botrytis cinerea [KC832409.1]|
rns i1 Lentinula edodes [NC_018365.1]
nad5 i1Ganoderma sinense [NC_022933.1]
nad5 i1 Moniliophthora roreri [NC_015400.1]
nad5 i1Fusarium solani [NC_016680.1]
nad5 i2 Ceratocystis cacaofunesta [NC_020430.1]
nad5 i1 Annulohypoxylon stygium [NC_023117.1]
nad5 i1 Neurospora crassa [Q35135]
nad5 i1 Phaeosphaeria nodorum [NC_009746.1]
nad5 Phoma sp. [HM158538.1]
nad5 Phoma sp. [HM158585.1]
nad5 Phoma sp. [HM158579.1]
nad5 Phoma sp. [HM158541.1]
rns i4 Cryphonectria parasitica [AF029891.1]
rns i1 Agrocybe aegerita [U54637.1]|
atp6 i1 Neurospora crassa [KC683708.1]
nad4 i1 Ceratocystis cacaofunesta [NC_020430.1]
nad4 i1 Sclerotinia borealis [NC_025200.1]|
nad4 i1 Fusarium solani [NC_016680.1]
nad4 i1 Moniliophthora perniciosa [NC_005927.1]
cox2 i1 Fusarium gerlachii [NC_025928.1]
cox3 i2 Fusarium gerlachii [NC_025928.1]
cox3 i1 Fusarium pseudograminearum [CBMD010001713.1]
nad2 i1 Botrytis cinerea [KC832409.1]
nad2 i1 Annulohypoxylon stygium [NC_023117.1]


atp6 i1 Botrytis cinerea [KC832409.1]
atp6 i1 Annulohypoxylon stygium [NC_023117.1]
atp6 i1 Neurospora crassa [KC683708.1]
atp6 i2 Neurospora crassa [T50468]
atp6 i1 Sordaria macrospora [CABT02001583.1]
atp6 i1 Fusarium solani [NC_016680.1]
atp6 i1 Fusarium gerlachii [NC_025928.1]
atp6 i1 Fusarium graminearum [NC_009493.1]
free standing LHEG Fusarium pseudograminearum [CBMD010001712.1]
free standing LHEG Fusarium pseudograminearum [CBME010002581.1]
atp6 i1 Fusarium pseudograminearum [CBMF010001392.1]
rnl i1 Amoebidium parasiticum [AF538042.1]
rps3 Grosmannia piceaperda [FJ717837.1]
nad2 i2 Fusarium pseudograminearum [CBME010002564.1]
nad4 i1 Fusarium graminearum [NC_009493.1]
nad2 i1 Fusarium gerlachii [NC_025928.1]
nad2 i3 Fusarium pseudograminearum
cox2 i2 Madurella mycetomatis [NC_018359.1]
cox2 i1 Ceratocystis cacaofunesta [NC_020430.1]
cox1 i1 Annulohypoxylon stygium [NC_023117.1]
nad3 i1 Fusarium gerlachii [NC_025928.1]
Free standing Fusarium pseudograminearum [CBMD010001696.1]
nad3 i1 fusarium graminarium [NC_009493.1]
nad3 i1 Sporothrix brasiliensis [AWTV01000012.1]
nad3 i1 Madurella mycetomatis [NC_018359.1]
nad3 i1 Podospora anserina [NC_001329.3]
nad3 i1 Podospora anserina [P15563]
free standing Fusarium culmorum [CBMH010002273.1]
mS569
rns i2 Ophiostoma perfectum [KU756606] WIN(M)823
rns i1 Sclerotinia sclerotiorum [U07553.1]
rns i2 Ophiostoma hyalothicum [JN871711.1]
rns i1 Ophiostoma minus [HQ292083.1]
rns i1 Ophiostoma minus [HQ292082.1]
cox2 i2 Trichoderma reesei [NC_003388.1]
cox1 i9 Peltigera malacea [NC_016955.1]
0.1

The mS952 GII intron LHE ORF from *O. carpenter* was used to extract related sequences from GenBank. The mS952 ORF appears to be a double motif LHE in *O. carpenter* (322 aa) and *O. pluriannulatum* WIN(M)539 (286 aa) and for *O. californicum* WIN(M)505 (323 aa) but for *O. pluriannulatum* WIN(972) (173 aa) the ORF only has one LAGLIDADG domain. The latter single domain ORF may be an indication that this ORF has eroded due to mutations that generated premature stop codons. The phylogenetic tree shows that the *O. carpenter* GII intron-encoded LHE ORF (representing the other mS952 ORFs uncovered during this study) is found in a variety of fungi including several members of the Ophiostomatales: *Ophiostoma stenoceras* [(Robak) Melin & Nannf.], *Ophiostoma brevicolle* [(R.W. Davidson) de Hoog & R.J. Scheff.], *O. torulosum, Leptographium* species and *O. minus.* A LHE sequence from the *nad5* i2 ORF from *Rhizoctonia solani* was used as the outgroup (Fig. 2.6).

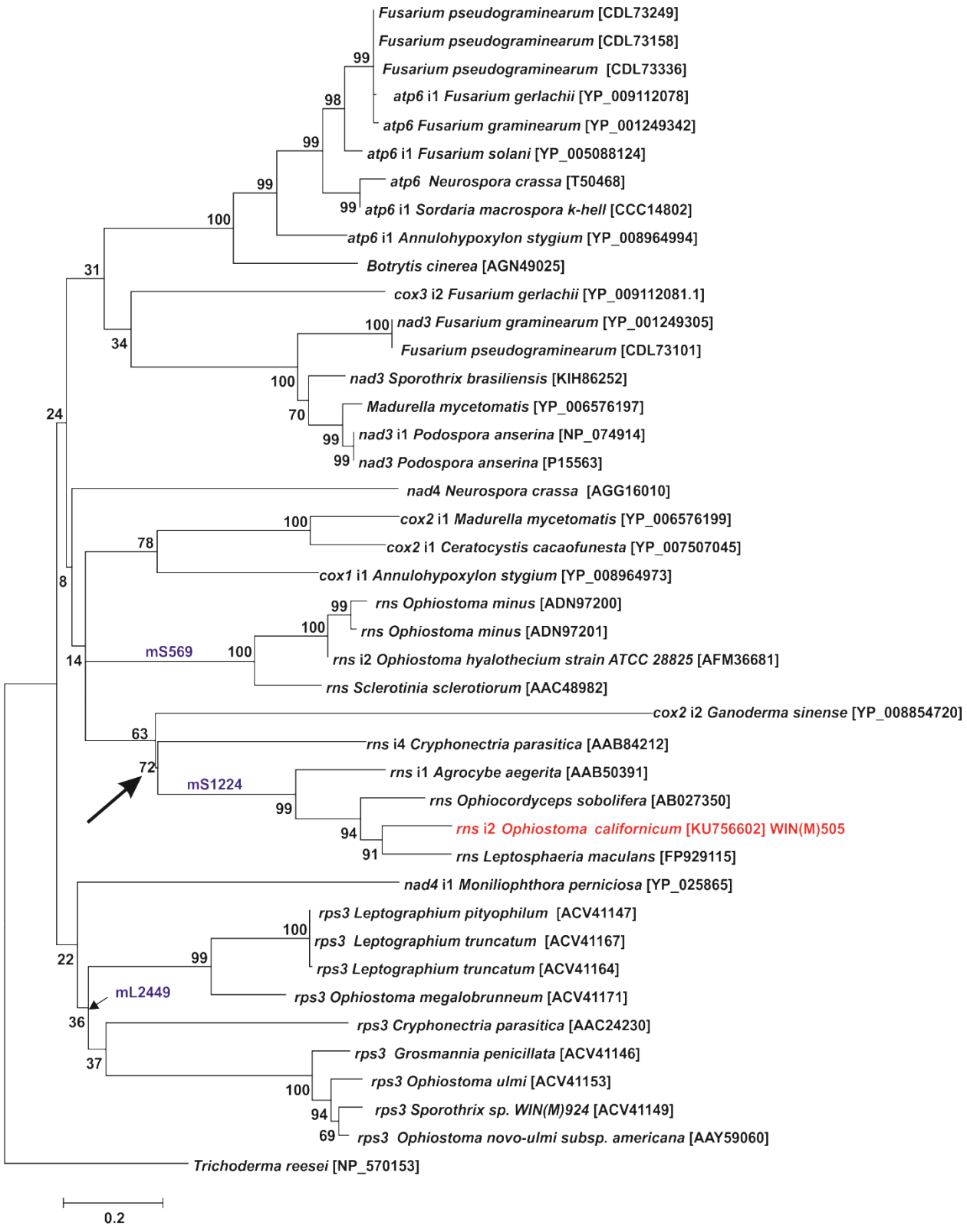
Figure 2.6. A phylogenetic tree showing the relatedness mS952 GII intron-encoded LAGLIDADG ORFs among group I intron ecoded LAGLIDADG ORFs. Tree topology is based on ML analysis of the mS952 LHE ORF a.a sequence and percentages at the nodes are node support values based on bootstrap analysis (1000 replicates). Names of organisms, host genes, intron number and location/position (when available) and GenBank accession numbers are provided. The branch lengths are based on ML (Whelan and Goldman plus Freq. model) analysis and are proportional to the mean number of substitutions per site (see scale bar).


atp6 i1Fusarium pseudograminearum [CDL73336]
atp6 i1 Fusarium graminearum [YP_001249342]
atp6 i1 Fusarium solani [YP_005088124]
atp6 i1 Sordaria macrospora k-hell [CCC14802]
atp6 i1 Neurospora crassa [AGG16015]
Botryotinia fuckeliana B05.10 [AGN49025]
Trichoderma reesei [NP_570153]
Neurospora crassa [AGG16010]
nad4 i1 Ceratocystis cacaofunesta [YP_007507079]
nad4 i1 Fusarium solani [YP_005088121]
nad4 i1 Sclerotinia borealis [YP_009072345]
cox3 i2 Fusarium gerlachii [YP_009112081.1]
nad3 i1 Fusarium graminearum [YP_001249305]
nad3 i1 Podospora anserina [NP_074914]
nad3 i1 Madurella mycetomatis [YP_006576197]
rnl i2 Amoebidium parasiticum [AAN04058]
rns i2 Ophiostoma hyalothecium strain ATCC 28825 [AFM36681]
rns i2 Ophiostoma hyalothecium strain ATCC 28825 [AFM36681]
rns Ophiostoma minus [ADN97200]
rns Ophiostoma minus [ADN97201]
rns Ophiostoma minus [ADN97201]
nad4 i1 Moniliophthora perniciosa [YP_025865]
rnl i1 Grosmannia piceaperda [ACV41152]
i1 Bipolaris maydis [EMD85018]
cox1 i5 Podospora anserina [NP_074928]
cox1 i5 Podospora anserina [NP_074928]
Sordaria macrospora k-hell [XP_003342391]
cox1 i5 Podospora anserina [1703265F]
cox1 i5 Podospora anserina [1703265F]
rps3 Sporothrix sp. WIN(M)924 [ACV41149]
rps3 Ophiostoma novo-ulmi subsp. americana [AAY59060]
rps3 Grosmannia penicillata [ACV41146]
rps3 Leptographium truncatum [ACV41164]
rps3 Leptographium truncatum [ACV41167]
rps3 Leptographium pityophilum strain [ACV41147]


rns i1 Lentinula edodes [YP_006576301]
nad5 i1 Moniliophthora roreri [YP_004376358]
nad5 i2 Cryphonectria parasitica [AAO14110]
nad5 i2 Cryphonectria parasitica [AAO14105]
nad5 i2 Cryphonectria parasitica [AAO14112]
nad5 i2 Ceratocystis cacaofunesta [YP_007507056]
nad5 i1 Fusarium solani [YP_005088107]
nad5 Phoma sp. MJW-2010c [ADM35202]
nad5 Leptosphaeria sclerotioides [ADM35376]
nad5 Phoma sp. MJW-2010e [ADM35338]
nad4IL i1 Podospora anserina [NP_074942]
nad4L i1 Neurospora crassa [CAA28761]
nad4L i1 Ceratocystis cacaofunesta [YP_007507052]
rns i3 Cryphonectria parasitica [AAB84211]
mS952
rns i1 Leptographium terebrantis [ADZ76450]
rns i1 Ophiostoma stenoceras [AFN16313]
rns i1 Ophiostoma stenoceras [AFN16313]
rns i1 Ophiostoma carpenteri [KU756600] WIN(M)855
rns i1 Ophiostoma brevicolle [AFN16312]
rns Ophiostoma torulosum [ADZ23992.1]
rns Ophiostoma minus isolate WIN(M) 873 [ADN97197]
rns Ophiostoma minus isolate WIN(M) 371 [ADN97198]
rns Ophiostoma ulmi [ADN97199]
rns Leptographium terebrantis strain CBS 337.70 [ADZ76450]
rns Leptographium wingfieldii [ADZ76453]
rns Leptographium truncatum [ADZ76451]
Ustilaginoidea virens [GAO18870]
rns i1 Agrocybe aegerita [AAB50391]
rns i1 Agrocybe aegerita [AAB50391]
rns i1 Agrocybe aegerita [AAB50391]
rns i1 Agrocybe aegerita [AAB50391]
nad4L i2 Podospora anserina [CAA38797]
nad4L i1 Fusarium culmorum [YP_009136831]
nad4L i1 Fusarium gerlachii [YP_009112045]
nad4L i1 Fusarium graminearum [YP_001249313]
rnl Microbotryum lychnidis-dioicae [YP_007475350]
nad5 i2 Rhizoctonia solani AG-1 IB [CCO27499]
0.2

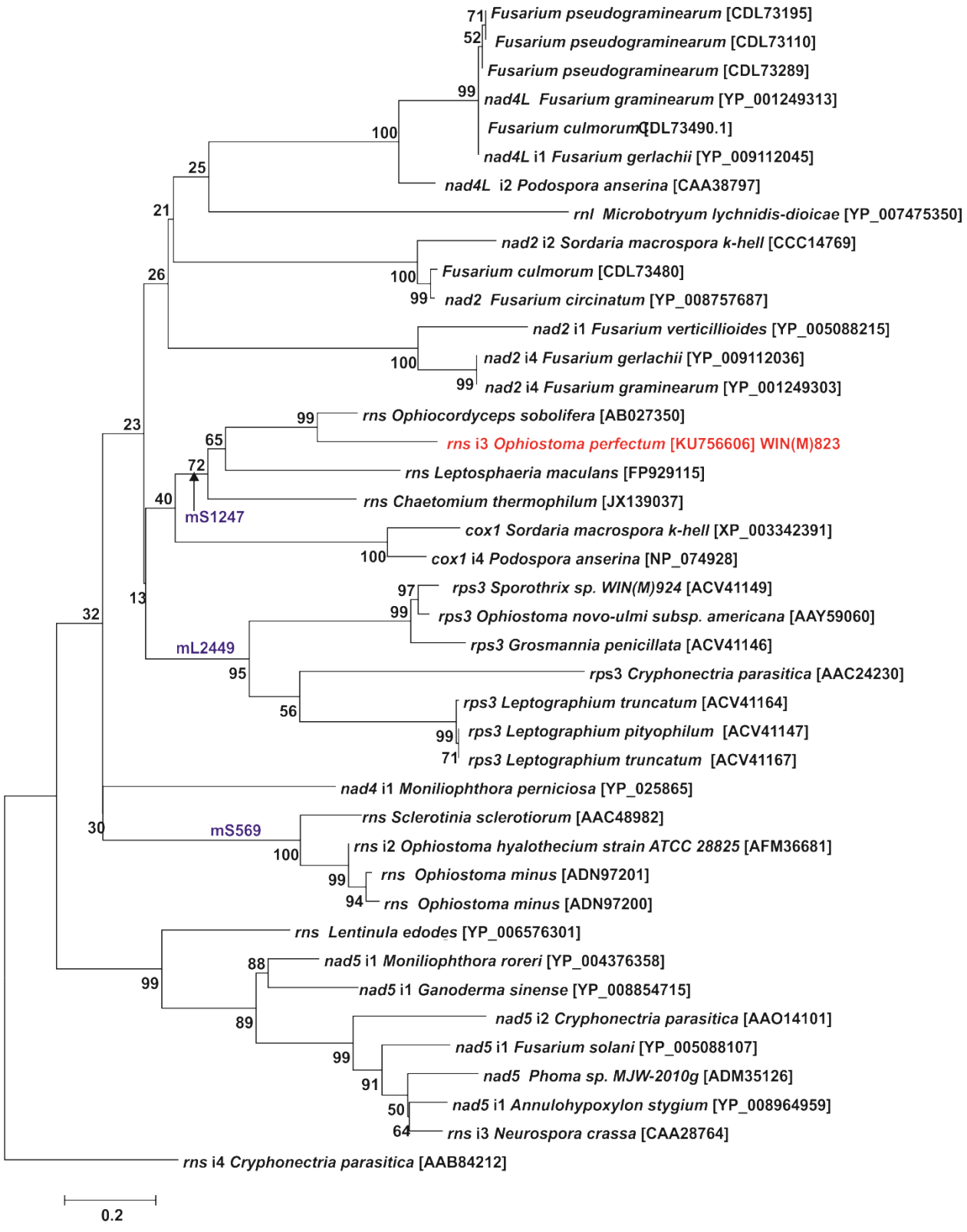
The phylogenetic analysis for the mS1224 ORFs (Fig. 2.7) recovered during this study from *O. californicum* WIN(M)505 (319aa) and *O. piliferum* WIN(M)972 (441aa) suggests that these ORFs are closely related to the ms1224 intron ORFs that exists in *Ophiocordyceps sobolifera* [AB027350], *Agrocybe aegerita* [AAB50391] and *Leptosphaeria maculans* [FP929115]. The clade is supported with high node support values. The *rns* i4 intron (=mS1210) ORF from *C. parasitica* joins the above clade at a node that received only moderate support (72%). The remaining nodes representing deeper branching patterns within this tree received only poor support values and thus, distant relationships cannot be extracted from this analysis.

Figure 2.7. A phylogenetic tree showing the relatedness of the mS1224 intron ORFs among other GI intron-encoded LAGLIDADG ORFs. Tree topology is based on ML analysis of the a.a sequence of the mS1224 ORF and percentages at the nodes are node support values based on bootstrap analysis (1000 replicates). Names of organisms, host genes, intron number and location/position (when available) and GenBank accession numbers are provided. The branch lengths are based on ML (Whelan and Goldman plus Freq. model) analysis and are proportional to the mean number of substitutions per site (see scale bar).


Fusarium pseudograminearum [CDL73249]
Fusarium pseudograminearum [CDL73158]
Fusarium pseudograminearum [CDL73336]
atp6 i1 Fusarium gerlachii [YP_009112078]
atp6 Fusarium graminearum [YP_001249342]
atp6 i1 Fusarium solani [YP_005088124]
atp6 Neurospora crassa [T50468]
atp6 i1 Sordaria macrospora k-hell [CCC14802]
atp6 i1 Annulohypoxylon stygium [YP_008964994]
Botrytis cinerea [AGN49025]
cox3 i2 Fusarium gerlachii [YP_009112081.1]
nad3 Fusarium graminearum [YP_001249305]
Fusarium pseudograminearum [CDL73101]
nad3 Sporothrix brasiliensis [KIH86252]
Madurella mycetomatis [YP_006576197]
nad3 i1 Podospora anserina [NP_074914]
nad3 Podospora anserina [P15563]
nad4 Neurospora crassa [AGG16010]
cox2 i1 Madurella mycetomatis [YP_006576199]
cox2 i1 Ceratocystis cacaofunesta [YP_007507045]
cox1 i1 Annulohypoxylon stygium [YP_008964973]
rns Ophiostoma minus [ADN97200]
rns Ophiostoma minus [ADN97201]
mS569
rns i2 Ophiostoma hyalothecium strain ATCC 28825 [AFM36681]
rns Sclerotinia sclerotiorum [AAC48982]
cox2 i2 Ganoderma sinense [YP_008854720]
rns i4 Cryphonectria parasitica [AAB84212]
rns i1 Agrocybe aegerita [AAB50391]
mS1224
rns Ophiocordyceps sobolifera [AB027350]
rns i2 Ophiostoma californicum [KU756602] WIN(M)505
rns Leptosphaeria maculans [FP929115]
nad4 i1 Moniliophthora perniciosa [YP_025865]
rps3 Leptographium pityophilum [ACV41147]
rps3 Leptographium truncatum [ACV41167]
rps3 Leptographium truncatum [ACV41164]
mL2449
rps3 Ophiostoma megalobrunneum [ACV41171]
rps3 Cryphonectria parasitica [AAC24230]
rps3 Grosmannia penicillata [ACV41146]
rps3 Ophiostoma ulmi [ACV41153]
rps3 Sporothrix sp. WIN(M)924 [ACV41149]
rps3 Ophiostoma novo-ulmi subsp. americana [AAY59060]
Trichoderma reesei [NP_570153]
0.2

Another dataset were constructed by using the mS1247 intron ORF sequence from *O. perfectum* (*rns* i3 ORF) as a query in blastp. The *Amoebidium parasiticum rnl*-i2 ORF was used as the outgroup in this analysis (Fig. 2.8). The mS1247 ORF from *O. perfectum* is fragmented (three fragments) due to frameshift mutations but 418 aa could be assembled. The *O. perfectum* mS1247 ORF grouped with mS1247 ORFs from *Ophiocordyceps sobolifera* [AB027350], *Leptosphaeria maculans* [FP929115] and *Chaetomium thermophilum* [JX139037]. However, the node support values are moderate for the clade which included all four mS1247 ORF sequences; but the node linking the *O. sobolifera* with *O. perfectum* received strong support (99%). With regard to deeper nodes, the phylogenetic tree received poor node support values so no further conclusions could be reached as to possible relationships of the mS1247 with other LAGLIDADG type ORFs.

Figure 2.8. A phylogenetic tree showing the relatedness of mS1247 intron ORFs among GI intron-encoded LAGLIDADG ORFs. Tree topology is based on ML analysis and percentages at the nodes are node support values based on bootstrap analysis (1000 replicates). Names of organisms, host genes, intron number and location/position (when available) and GenBank accession numbers are provided. The branch lengths are based on ML (Whelan and Goldman plus Freq. model) analysis and are proportional to the mean number of substitutions per site (see scale bar).


71 Fusarium pseudograminearum [CDL73195]
52 Fusarium pseudograminearum [CDL73110]
Fusarium pseudograminearum [CDL73289]
99 nad4L Fusarium graminearum [YP_001249313]
Fusarium culmorum [CDL73490.1]
100
nad4L i1 Fusarium gerlachii [YP_009112045]
25
nad4L i2 Podospora anserina [CAA38797]
21
rnl Microbotryum lychnidis-dioicae [YP_007475350]
nad2 i2 Sordaria macrospora k-hell [CCC14769]
26
100 Fusarium culmorum [CDL73480]
99 nad2 Fusarium circinatum [YP_008757687]
nad2 i1 Fusarium verticillioides [YP_005088215]
100
nad2 i4 Fusarium gerlachii [YP_009112036]
99 nad2 i4 Fusarium graminearum [YP_001249303]
99 rns Ophiocordyceps sobolifera [AB027350]
23
65
rns i3 Ophiostoma perfectum [KU756606] WIN(M)823
72
rns Leptosphaeria maculans [FP929115]
40
rns Chaetomium thermophilum [JX139037]
mS1247
cox1 Sordaria macrospora k-hell [XP_003342391]
100
cox1 i4 Podospora anserina [NP_074928]
97
rps3 Sporothrix sp. WIN(M)924 [ACV41149]
13
32
99
rps3 Ophiostoma novo-ulmi subsp. americana [AAY59060]
rps3 Grosmannia penicillata [ACV41146]
mL2449
95
rps3 Cryphonectria parasitica [AAC24230]
56
rps3 Leptographium truncatum [ACV41164]
99
rps3 Leptographium pityophilum [ACV41147]
71 rps3 Leptographium truncatum [ACV41167]
nad4 i1 Moniliophthora perniciosa [YP_025865]
rns Sclerotinia sclerotiorum [AAC48982]
30
mS569
100
rns i2 Ophiostoma hyalothecium strain ATCC 28825 [AFM36681]
99
rns Ophiostoma minus [ADN97201]
94 rns Ophiostoma minus [ADN97200]
rns Lentinula edodes [YP_006576301]
88
nad5 i1 Moniliophthora roreri [YP_004376358]
99
nad5 i1 Ganoderma sinense [YP_008854715]
89
nad5 i2 Cryphonectria parasitica [AAO14101]
99
nad5 i1 Fusarium solani [YP_005088107]
91
nad5 Phoma sp. MJW-2010g [ADM35126]
50
nad5 i1 Annulohypoxylon stygium [YP_008964959]
64 rns i3 Neurospora crassa [CAA28764]
rns i4 Cryphonectria parasitica [AAB84212]
0.2

2. 5. Discussion

2.5.1. *Ophiostoma* species relationships

Among the strains of *Ophiostoma* species sampled in this study, many lacked introns within the *rns* gene, but a few members did contain introns [*O. piliferum, O.carpenteri, O. pluriannulatum, O. perfectum and O. californicum*]. This is suggestive of either rapid gain or rapid loss of introns. Gain of introns could be achieved by horizontal transfer or by vertical transmission of these introns, random loss of introns could be achieved by the original intron containing allele being replaced by a cDNA version of the gene (Hausner, 2012). These patterns of inheritance/maintenance of introns among the fungi have been observed among other *rns* introns (Haugen and Bhattacharya, 2004; Haugen *et al*., 2004).

With regard to *O. piliferum*, *O. pluriannulatum* and *O. novae-zelandiae*, these species, based on morphological studies, tend to be easily confused with one another (Thwaites *et al*., 2005). Phylogenetic analysis, based on ITS sequences, also failed to distinguish among these species. Previous work, based on nuclear large ribosomal RNA subunit data, also found that currently available strains of *O. piliferum* arranged into at least two distinct clades, suggesting that *O. piliferum*, as currently circumscribed, accommodates two or more species (Hausner and Reid, 2003). Although it has been proposed that the rDNA ITS region could be used as a barcode marker for the fungi (Schoch *et al*., 2012; Blaalid *et al*., 2013), this study shows that, with regard to some members of the Ophiostomatales, the ITS region may not be suitable for this purpose. Thus, more molecular markers might have to be applied to resolve these species or they may indeed be synonyms as suggested by Thwaites *et al*. (2005), but for the purpose of this study, the ITS data suggest we are dealing with closely related species/strains. This allows us to

potentially observe more recent events with regard to the evolutionary dynamics of introns and their encoded proteins among members of the *O. piliferum* complex.

The distribution of *rns* introns also did not relate to species status, an expected result as previous studies involving Ascomycota mtDNA introns also suggested that intron-rich sequences may actually not be useful for developing DNA barcodes (Santamaria *et al.*, 2009), i.e. the presence and absence of introns is too erratic to provide stable markers. Only a few examples exist to show that mtDNA introns could be useful in species indentification; for example the *cox1*-ai3β may have some application in identifying a subgroup within *Saccharomyces cerevisiae* (Wolters *et al.*, 2015). The mS952 intron has been noted to be absent in strains of *Ophiostoma novo-ulmi* subspecies *americana*, but it has been noted to be present in strains of *Ophiosotoma ulmi* (Gibb and Hausner, 2005; Hafez and Hausner, 2011b). Movement of introns along with their IEPs is facilitated by their encoded proteins that target specific sites in cognate intronless alleles (Colleaux *et al*., 1986; Dujon, 1989) or by the ability of GI and GII introns to reverse splice into transcripts that need to be reverse transcribed into cDNA, followed by a recombinaton event that inserts the cDNA into the genome. Reverse splicing, for GI introns in particular. requries less sequence recognition and thus, it could be an efficient method for invading new sites (Roman *et al*., 1999; Bhattacharya *et al*., 2002, 2005).

2.5.2. The *Ophiostoma rns* gene structure

This study focused on the *rns* gene of *O. piliferum* and related taxa. In general, it was noted that the *rns* gene is variable in size (polymorphic) due to the presence/absence of introns and associated ORFs. Also, introns appear to be inserted at the same positions among related and distantly related fungi, which is a reflection of the homing mechanism promoted by the intron-encoded homing endonucleases. Homing endonucleases recognize long target sites and therefore,

require conserved regions for their long-term survival and for moving laterally from one species to another (Stoddard, 2005; Belfort *et al.*, 2002). On rare occasions, they move into new sites (ectopic integration) either by transposition or by means of reverse splicing of the intron RNA [a mechanism that has not yet been experimentally demonstrated and requires reverse transcriptase activity (Bhattacharya *et al*., 2005; Hausner *et al*., 2014)].

2.5. 3. The *Ophiostoma rns* introns

2.5.3.1. The mS379 intron

The analysis of sequence data have shown that the *O. perfectum* WIN(M)832 *rns* gene is 7 kb in length and has three introns inserted in highly conserved regions within the *rns* gene. The mS379 intron is a GII intron, type AI that was previously reported in three *O. minus* strains by Hafez and Hausner (2011a). In these three strains, the authors noted that the mS379 ORFs were at various stages of degeneration due to the presence of premature stop codons. The *O. perfectum* WIN(M)823 intron at mS379 encodes a complete RT ORF that is inserted into domain II of the GII intron. The latter is unusual as organellar GII intron ORFs are typically located in domain IV (Toor *et al.*, 2001; Hafez *et al*., 2013) found related introns with a degraded ORF inserted into domain II of the mS379 GII intron in *O. torulosum* and these authors described one example of an mS379 intron that completely lacked an ORF in *O. hyalothecium*. Goddard and Burt (1999) predicted, based on their work on the yeast omega intron (mL2449/2450 GI intron in the *rnl* gene), a life cycle of invasion and degeneration for HEGs and their hosting GI introns. Our study suggests that RT ORFs in GII introns can follow a similar life cycle; i.e. invasion into an empty spot, followed by slow degeneration initially of the ORF, and eventually complete loss of the ORF, and presumably, complete loss of the intron, regenerating a potential site available for reinvasion.

A node in the mS379 tree supports monophyly for the *Ophiostoma* mS379 RT ORF within a grouping that includes sequences from brown algae, bacteria, plants and fungi. Thus, it appears this intron ORF has a complex evolutionary history that could include horizontal transfers. It has been shown that GII intron ORFs probably co-evolve with their host introns but these composite elements can evolve independently from the host genomes that encode them due to lateral transfers (Zimmerly *et al*., 2001; Toor *et al*., 2001; Toro and Nisa-Martínez, 2014).

2.5.3.2. The mS569 intron

The second intron in *O. perfectum* WIN(M)823 is the mS569 GI intron, type C2,with a LHE ORF. This intron is also found in *O. pluriannulatum* WIN(M)539. Hafez *et al.* (2013) found the same intron in *O. torulosum* and *O. hyalothecium*. This intron is a nice example of a phenomenon referred to as core creep (Edgell *et al*., 2011). Essentially, the ORF extended ("creeped") towards the upstream exon and in some instances fused with it .The extention of the ORF could be happen due to mutations that generate additional coding amino acids at the 5' segment and the illimination of termination codon(s) upstream of the original ORF (Edgell *et al*., 2011). This exonization event (intron sequences becoming coding sequences) also requries that the intron folding capacity is not altered due to changes that allow for the ORF to fuse to the upstream exon. This means that the ORF sequence overlaps with the intron core sequences. This might facilitate intron-encoded ORFs that are present in protein-coding genes to be more efficiently translated but among *rns* introns, it may not provide a major advantage. However, it might be a remnant of the intron's origin, assuming it transferred from a protein-coding gene into the *rns* gene.

2.5.3. 3. The mS952 intron

In this study, four closely related species have the same intron positioned at mS952: *O. carpenteri* WIN(M)855, *O. pluriannulatum* WIN(M)927, *O. subannulatum* WIN(M)539, and *O. californicum* WIN(M)505. This GII intron was previously found in other *Ophiostoma* species by Mullineux *et al*. (2010, 2011), within various strains of *O. minus and O. ulmi* (Hafez and Hausner, 2011a, b), and it was first noted in other fungi by Toor and Zimmerly (2002). The wide distribution of this intron suggests that it is probably, predominately, vertically inherited with some evidence of loss and possible horizontal gene transfer (HGT) (Mullineux *et al*., 2011). Also, the mS952 LHE ORF that is characteristic for group I introns, might be an indicator of this GII intron being potentially mobilized by a GI intron type pathway; i.e. DNA-based.

2.5.3.4. The mS1224 intron

In a previous survey, the mS1224 intron was not seen in *Ophistoma* species but in this study, strains of *O. pluriannulatum* WIN(M)972 and *O. californicum* WIN(M)505 have a GI intron, type C2, at postion S1224 and a LHE as an IEP. This intron and ORF are also found in rather distantly related fungi, *Leptosphaeria maculans, Ophiocordyceps sobolifera, Agrocybe aegerita* and *Cryphonectria parasitica* (Hafez *et al*., 2013).

2.5.3.5. The mS1247 intron

The third intron found in *O.* perfectumWIN(M)823 is located at position mS1247. The mS1247 intron is a GI, type C2, encoding a LHE ORF. So far within *Ophiostoma*, the mS1247 intron has only been observed in *O. perfectum*. Hafez *et al*. (2013) had reported this intron in *Ophiocordyceps sobolifera and Leptosphaeria maculans* and they also noted that this site can be occupied by a twintron or nested introns, which is an intron inserted within an intron (Copertino and Hallick, 1991; Michel *et al.*, 1989; Hafez and Hausner, 2015). For example, in *Chaetomium thermophilum* strain UAMH 2024, a GI intron with a double motif LHE has been invaded by an

ORF-less GII intron. The latter arrangment offers a platform for designing HEases with an internal regulatory element (Guha and Hausner, 2016).

2. 6. Conclusions

With regard to the *rns* intron landscape for members of the *O. piliferum* complex, this study showed that five potential intron sites exist and these are occupied by GI (S569, S1224, S1247) and GII introns (S379, S952). Among intron-encoded ORFs, the LAGLIDADG family appears to have invaded four of these introns, with the mS379 GII intron maintaing an RT ORF. Intron distribution did not correspond to species desigantion, so mtDNA *rns* introns may not be useful as molecular taxonomic markers. However, intron landscapes provide information for those that annotate mtDNAs or for those searching for sources of polymorphisms. This study identfied examples of a phenomena referred to as "core creep" where intron encoded ORFs have fused to the upstream exons (mS569) previously observed in introns embedded within protein coding genes but not frequently reproted for rDNA introns.

In addition, ribozymes and intron-encoded ORFs (HEases and RTs) have applications in biotechnology, genome editing, or as functional genomics tools (Stoddard, 2005; Belfort and Bonocora, 2014; Enyeart *et al*., 2014; Hafez and Hausner, 2015; Qin *et al*., 2016). Intron landscapes may identify LADLIDADG HEases and host introns at insertion sites that may be similar to sequences present in genes associated with pathogenicity and/or monogenic diseases. These HEases could be developed into gene targeting tools (Hafez and Hausner, 2012).

Phylogenetic investigations suggest that the intron ORFs evolve rapidly and thus, most trees had poorly supported topologies. In general, among members of the Ophiostomatales, introns appear to be gained and lost frequently. The intron ORFs trees topologies do not follow the expected host species/genome trees, which hints at the possibilities of horizontal transfers of

these elements among distantly related lines. This work also provided some insight into the evolution of GII introns and their RT ORFs; the mS379 intron appears to follow the intron/HEG life cycle model of invasion, decay, loss and possible reinvasion as proposed by Goddard and Burt (1999).

A version of Chapter 3 has been submitted (Canadian Journal of Microbiology):

Bilto IM, Guha, T.K, Wai A, Hausner G. Three new active members of the I-OnuI family of homing endonucleases. Can. J. Micro. (In Press) doi: 10.1139/cjm-2017-0067.

Bilto and Dr. Hausner designed the experiments, Guha, T.K. provided assistance and ideas with regards to protein purification, Wai, A. provided the I AstI analysis and Bilto performed the experiments and analysis under supervision of Dr. Hausner.

Original version of the paper was drafted by Bilto and final version was edited by Bilto and Hausner.

3. Orthologues of the mS917 intron-encoded protein of *Ceratocystis cacaofunesta* have invaded other mitochondrial genes.

3.1. Abstract

Phylogenetic analysis of mS917 LAGLIDADG type homing endonuclease sequences identified orthologs that appear to have inserted into at least five different mtDNA genes. These orthologs are all inserted within group I D type introns. *In vitro* characterization of three mS917 related orthologos located within introns located in the *rns* (*C. cacaofunesta,* I-CcaI), *rnl* (*C. cacaofunesta,* I-CcaII) and *cox3* (*A. stygium*. I-AstI) genes showed that they are functional homing endonucleases that cleave their cognate target sites. The *C. cacaofunesta,* I-CcaI protein could be overexpressed and purified in a consistent manner and was therefore selected for further characterization. The endonuclease activity of I-CcaI was tested at various temperatures and using various substrates and its minimum DNA recognition sequence was estimated to be 26 nts. This set of homing endonucleases may provide some insight on how these types of mobile elements can migrate into new locations in order to prevent their extinction. The study also provides additional homing endonucleases that could be applied in biotechnology.

3.2. Introduction

Homing endonucleases (HEases) require long DNA target sites (~12 to 40 bp) and are therefore, highly specific DNA cleavage enzymes (Belfort and Roberts, 1997; Chevalier and Stoddard, 2001; Lambert *et al*., 2016). They are encoded by homing endonuclease genes (HEGs) that are embedded within elements such as Group I and sometimes Group II and archaeal introns (Dujon, 1989; Belcour *et al*., 1997; Toor and Zimmerly, 2002; Mullineux *et al*., 2010; Tocchini-Valentini *et al*. 2011). HEGs can also be freestanding and they can form the DNA-cutting component of inteins (Gimble, 2000; Hafez and Hausner, 2012). HEases can promote the mobility of themselves and of the genetic elements that encode them such as introns or inteins. In addition, some introns encode HEases that act as maturases; i.e. they appear to help the intron in

splicing from their primary transcripts (Dujon and Belcour, 1989; Szczepanek and Lazowska, 1996; Belfort *et al.*, 2002; Bolduc *et al*., 2003; Hausner, 2012). Earlier studies showed that HEGs are mobile genetic elements that can generate double-stranded DNA breaks in a cognate allele resulting in the cell's DNA double-strand break repair mechanism by using the HEG containing DNA as a template (Dujon, 1989; Belfort *et al*., 2002; Burt and Trivers, 2006). Homing endonuclease activity results in repair that involves the non-reciprocal transfer of the HEG into new sites, and it also can be associated with co-conversion of markers flanking the HEG (or intron) insertion site (Mueller *et al*., 1996a, b; Parker *et al*., 1999; Belfort *et al*., 2002; Muñoz *et al*., 2012).

Homing endonucleases can be assigned into six families based on the presence of conserved amino acid motifs. With regards to inteins, group I and II introns, four families are the most relevant (Stoddard 2005): LAGLIDADG (LHE), GIY-YIG, H-N-H and His-Cys box (Stoddard, 2005; Skowronek and Bujnicki, 2007; Hafez and Hausner, 2012). The LAGLIDADG and GIY-YIG families are most frequently encountered in fungal mitochondrial group I introns (Haugen, *et al*., 2005; Stoddard, 2005; Hausner, 2012). LHEs generate staggered cuts with 4 nucleotides 3′ overhangs at their cleavage site (Belfort and Robert, 1997; Chevalier and Stoddard, 2001). LHEs in their active configuration can act as homodimers when the peptide contains one LAGLIDADG motif or they can act as monomers when the peptide contains two LAGLIDADG motifs (Haugen and Bhattacharya, 2004; Dalgaard *et al*., 1997; Lucas *et al.*, 2001). Some HEases have been shown to tolerate some sequence degeneracy within their recognition sites which gives HEGs (and their host elements) the flexibility to invade new sites and thus provides the possibility of HEGs to avoid elimination by genetic drift (Scalley-Kim *et al*., 2007; Barzel *et al*., 2011). Specificity with some allowance for variations among alleles

makes HEases potential tools in genome editing such as gene replacements, targeted mutagenesis, gene drive mechanisms for pest control, and in designing vectors systems that require unique endonuclease target sites (Hafez and Hausner, 2012).

Previously, Monteiro-Vitorello *et al*. (2009) reported an intron at mS915 in *Cryphonectria parasitica*. Hafez *et al*. (2013) reannotated this intron and noted that it is inserted at S917. The later study also showed that this intron is a nested (or twintron) group ID intron where the internal group ID intron encodes a double motif LAGLIDAG-type ORF and it is inserted into an ORF-less external group ID intron. The authers noted that orthologues of this HEG can be found in other fungal species where these orthologues were observed to be encoded within group ID introns inserted in the *rnl, cox3, nad5* and *nad6* genes. Exploration of these HEases may provide some clues on how HEGs can move into different sites thereby promoting their spread within a genome, within a population, or between different species. The objectives for the current study were to identify and characterize active members of the mS917 HEases family, focusing on three members inserted into different genes: the mS917 HE version that is encoded within an *rns* intron in *Ceratocystis cacaofunesta* (GenBank accession: YP_007507084.1); a HE encoded within an *rnl* group ID intron also found in *C. cacaofunesta* (Ambrosio *et al.*, 2013; GenBank accession: YP_007507038.1); and a HE that is encoded within the *cox3* intron in *Annulohypoxylon stygium* (GenBank accession: NC_023117.1). This study describes new additional endonucleases that are engineered naturally and are able to recognise novel recognition sites. Currently most available HEases have been engineered from approximately five protein scaffolds (see Hafez and Hausner 2012), which limits the availability of HEases for various applications. This also makes HEases very expensive as genome editing tools. Identifying and characterizing native HEases is therefore an alternative to reengineering

existing HEases. The HEases characterized in this work can be added to the catalog of currently available HEases that may have various biotechnology applications (Stoddard, 2014; Hafez *et al*., 2014).

The previous chapter explored the *rns* genes for selected members of the genus *Ophiostoma*; this study examines in more detail the *rns* encoded mS917 ORF and its encoded HEase. It has been previously noted (Stoddard 2005; Belfort et al. 2012) that most intron encoded ORFs are never actually tested for activity. This study shows that intron encoded LHEases display properties that shows that many group I introns encode homing endonucleases.

3.3. Materials and methods

3.3.1. Phylogenetic analysis of homing endonuclease sequences

The LAGLIDADG datasets were based on extracting sequences from the NCBI data base using the mS917 intron encoded HE amino acid (aa) sequence of *C. cacaofunesta* (GenBank accession: YP_007507084.1) as a query in blastp. A total of 52 sequences were extracted from GenBank, and the datasets were aligned with MAFFT (Katoh and Standley, 2013). Manual adjustments were made to the alignment, if necessary, with GeneDoc (Nicholas *et al*., 1997). The aa alignment was analyzed with three tree building methods, as implemented in the MEGA (version 6) program (Tamura *et al*., 2011), neighbour joining (NJ), maximum parsimony (PARS), and maximum likelihood (ML). The bootstrap option was implemented (2000 replicates) in order to assess the level of support for the tree topologies generated by the respective methods. With regards to NJ analysis, distances were calculated with the maximum composite likelihood method and its default settings were selected along with the complete deletion of gaps option. For the PARS method, the complete deletion option was also selected. In the ML analysis, the best model was first determined with the “best model” program as

implemented in MEGA. The Whelan and Goldman (2001) plus Freq. model was selected for the ML analysis and the complete deletion option was again selected to remove segments of the alignments that contained gaps.

3.3.2. HEase naming nomenclature

HEases were named according to the proposed nomenclatures by Belfort and Roberts (1997). HEases encoded by group I introns have the prefix (I-) followed by abbreviations for the genus and species name. The final Roman numeral distinguishes multiple enzymes that have been characterized for that organism. The HEases examined in this study are therefore named as follows: I-CcaI for the intron encoded protein (IEP) encoded by the *rns* i1 of *C. cacaofunesta*, I-CcaII for the IEP encoded by the *rnl*i1 of *C. cacaofunesta*, I-AstI for theIEP encoded by the *cox3*i3 of *A. stygium*.

3.3.3. Construction of expression vectors and substrate plasmids

The HE sequences were obtained from GenBank for *C.cacaofunesta* YP_007507084.1 (*rns*), *C. cacaofunesta* YP_007507038.1 (*rnl*) and *A. stygium* NC_023117.1 (*cox3*). To allow for efficient expression in *Escherichia coli (E. coli)*, the genetic code for these HE open reading frames (ORFs) were codon optimized for *E. coli*. The pET28b+ vector was used for assembling the HE expression constructs (Genscript, New Jersey, USA). The HE ORFs were inserted at the BamHI/NdeI restriction enzyme site and the vector provided an N-terminal 6X Histidine (His)-tag and the T7 promoter. The three expression vectors were named as follows: pI-CcaI, pI-CcaII and pI-AstI. The constructs were transferred into chemically competent *E.coli* BL21 (λDE3) cells (New England Biolabs) for protein expression followed by purification of the His-tagged protein on a Ni-NTA column as previously described (Hafez *et al.*, 2014).

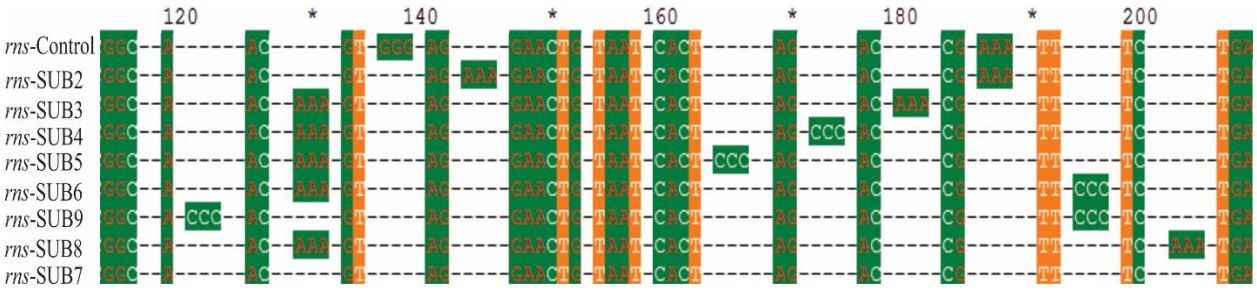
Suitable substrates for testing the above HEases were designed as follows. The corresponding host gene sequences were examined and 100 nucleotides of exon sequences flanking the intron insertion sites (~200 nts) were combined, synthesised and inserted into the pUC57 (2.7 kb) vector (GenScript, New Jersey, USA). These substrates were named as follows: p*rns*-SUB for testing the *C. cacaofunesta rns* i1 IEP (I-CcaI), p*rnl*-SUB for testing the *C. cacaofunesta rnl* i1 IEP (I-CcaII), and p*cox3*-SUBfor testing the *A. stygium cox3* i3 IEP (I-AstI).

All plasmids were maintained in *E. coli* DH5α and the plasmids were purified with the Presto^{TM} Mini Plasmid Kit (FroggaBio, North York, Ontario). Transformed cells were kept at -80˚C in 80% glycerol stocks (Sambrook *et al*., 1989) and purified plasmids were stored at -20˚C.

3.3.4. Minimum recognition sequence and substrate plasmids for the: I-CcaI

In addition to the p*rns*-SUB, a set of 8 substrate plasmids (p*rns*-SUB2 to 9) were designed to determine the minimum recognition sequence required for the HE activity for the mS917 IEP (I-CcaI) of *C. cacaofunesta.* Essentially, the original substrate sequence was modified by shortening the potential recognition sequence two nucleotides at a time on either side of the HE cleavage site by inserting nucleotide triplets that disrupt the potential HE recognition sequence (Bae *et al*., 2009; see Fig. 3.1).

Figure 3.1. Substrates based on modifications on the p*rns*-SUB plasmid: The "x" designates the central four nucleotide recognition motif. The insertion of triplets (AAA, GGG, or CCC) defines various length versions of the potential minimum recognition sequence for I-CcaI.


120
*
140
*
160
*
180
*
200
rns-Control
rns-SUB2
rns-SUB3
rns-SUB4
rns-SUB5
rns-SUB6
rns-SUB9
rns-SUB8
rns-SUB7

3.3.5. Expression and purification of recombinant proteins (HEases)

Five ml of Terrific broth (TB) [12 g tryptone, 24 g yeast extract, 4 ml of 100 % glycerol, all in 900 ml H_2O was autoclaved followed by the addition of sterile 100 ml of 0.17 M KH2PO4/ 0.72 M K2HPO4], supplemented with 100 µg/ml kanamycin was inoculated with 100 µl of *E. coli* BL21 (λDE3) transformed cells (with the appropriate expression vector) and the cultures were incubated overnight (O/N) with shaking at 37˚C. *E. coli* BL21 (λDE3) cells transformed with the pET28b+ vector were used as the negative control in the protein expression study. Two ml of the O/N culture was used to inoculate a 250 ml TB flask and incubated at 37˚C until the OD600 reached about 0.5. At that point, 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added for the induction of protein expression and the cultures were moved to various temperatures (16˚C, room temperature (RT: ~23.5˚C), 28˚C, 37˚C) with shaking (O/N = 16 h) in order to determine optimal HE protein expression conditions. For large scale protein expression, four large fluted conical flasks containing 500 ml of TB media were prepared and the recombinant proteins were expressed based on the pre-determined conditions as gathered from the small-scale expression trials. Cells were harvested by centrifuging for 7000 x g for 10 min and the resulting pellets were frozen at -80˚C. Cells were thawed on ice and resuspended with 5 ml washing/binding buffer (WB) [50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 10% (w/v) glycerol, 6 mM β-mercaptoethanol]. Cells were lysed with a French Press at 1200 psi, and the homogenized material was centrifuged at 17000 x g for 20 min at 4˚C, the supernatant was passed onto a Ni-NTA Superflow resin column (Qiagen, Toronto, Ontario). The column was washed first with the 30-ml washing buffer (WB) and this was followed by a series of washings: 30 ml WB + 25 mM imidazole, 30 ml WB+ 50 mM imidazole, and 30 ml WB + 100 mM imidazole. Finally, the HE protein was eluted with WB + 250 mM imidazole in 10 fractions (1

ml each). All the washing fractions and the elution fractions were collected and analyzed on a SDS polyacrylamide gel (12.5%). Electrophoresis was used to evaluate HE protein purity. Samples were dialysed using the dialysis buffer (50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 6 mM β-mercaptoethanol) using the Slide-A-LyzerTM dialysis cassette (Millipore, Billerica, Massachusetts, USA) with a 10kDa molecular weight cut-off was used to remove the imidazole. Dialysed samples were concentrated to 1 ml with the Amicon concentrator (Model 8050) using a Millipore YM-10 membrane. Usually, the purified, concentrated protein (HE) was immediately assayed for its endonuclease activity before storing. However, for long term storage, 200 µl of protein storage buffer [50 mM Tris–HCl (pH 8.0), 400 mM NaCl, 0.5 mM DTT, 10 % (w/v) glycerol] was added to the purified protein and the samples were stored at -80˚C.

3.3.6. Endonuclease assays

The purified HEases were challenged with the appropriate DNA substrate plasmids in order to assess their endonuclease potential. Methodology is based on those previously presented in Hafez *et al.* (2014). Briefly, the endonuclease reaction mixtures contained: 25 µl of substrate plasmid (25 ng/ml), 5 µl Invitrogen Buffer React® #3 (100 mM NaCl, 50 mM Tris–HCl (pH 8.0) and 10 mM $MgCl_2$) supplemented with 1mM DTT, 5 µl HE protein (9µg) and H_2O to reach a final volume of 50 µl, The endonuclease reactions were incubated at 37˚Cand 10 µl, aliquots were withdrawn at the following time points: 0, 30, 60, 90, and 120 min. The endonuclease reactions were stopped by adding 2 µl of 200 mM EDTA (pH8.0) and 1 µl proteinase K (1mg/ml) to the reaction aliquots. Finally, the reaction products were resolved on a 1% agarose gel ran at 80 volts and gels were stained with Ethidium bromide (0.5 g/ml).

3.3.7. Mapping of the I-CcaI, I-CcaII, and I-AstI cleavage sites

The *in vitro* cleavage site mapping strategy was based on Bae *et al*., (2009) and has been previously described (Hafez *et al*., 2014; Guha and Hausner, 2014). The substrate plasmids were digested with the corresponding HE described above and the linearized substrates were cut from the gel and purified with the Wizard-SV Gel and PCR Clean-Up system (Promega; ThermoFisher Scientific). Forty µl of linearized substrate was treated with T4 DNA polymerase (T4 DNA pol) to generate blunt ends by removing the characteristic 4 nts 3′-OH overhangs generated by LAGLIDADG-type HEases (Bae *et al*., 2009). The T4 DNA pol treatment reaction contained the following components: 40 µl of linearized substrate plasmid (25 µg/ml), 2 µl of T4 DNA pol (5 units/µl; Invitrogen), 20 µl of 5X T4 DNA polymerase buffer, 20 µl dNTP mixture (0.5mM) and H_2O to achieve the final volume of 100 µl. The reaction was incubated at room temperature (24˚C) for 20 min thereafter the tubes were placed on ice for 5 min and finally the reactions were terminated by transferring to 70˚C for 10 min. The DNA was purified with the Wizard-SV Gel and PCR Clean-Up system (Promega). The blunted substrate DNA was relegated in the following reaction mixture: 20 µl (0.25 µg) of T4 DNA polymerase treated DNA, 2 µl of T4 DNA ligase (1 u/µl; Invitrogen), 10 µl of 5X ligase buffer and the addition of H_2O to achieve a final volume of 50 µl., ligation reactions were incubated at room temperature for 2 hours. Thereafter, the ligation mixtures were diluted 5-fold and a 10 µl aliquot of the dilution was transformed into chemically competent *E. coli* DH5α cells. The transformed cells were plated onto LB Agar plates supplemented with ampicillin (100 µg/ml). Single colonies were picked from these plates and these were used to inoculate 5 ml LB broth (100 µl/ml ampicillin) tubes which were incubated for ~18 hours at 37˚C. The plasmids were purified by the Wizard® Plus Minipreps DNA Purification System (Promega) and they were sent to the DNA Technologies Unit (NRC, Saskatoon, Saskatchewan) for cycle sequencing using the M13 forward and reverse

primers. Untreated substrate plasmids were sent along with the HE and T4 DNA polymerase treated substrate plasmids. This allowed for examining what sequences had been removed by the T4 DNA polymerase treatment. The sequencing results were manually aligned with the GeneDoc program (version 2.7; Nicholas *et al*., 1997). Comparison of untreated samples with HE plus T4DNA polymerase treated samples revealed the four nucleotides that were removed and thus allowed for the determination of the actual HE cleavage site.

3.3.8. Temperature requirement and the minimal DNA recognition sequence for I-CcaI

The I-CcaI HE (mS917 IEP) was evaluated with regards to its activity at different temperatures. Endonucleases assays were set up as described previously except the reactions were incubated for 1 hour at 20˚C, 30˚C, 37˚C, 40˚C, and 50˚C. In addition, the I-CcaI HE was further characterized with regards to its minimum recognition sequence requirement. Here the HE was challenged with various substrate plasmids containing sequences modified to provide various lengths of the putative recognition sequence. The preparation of the substrate plasmids (p*rns*-SUB2 to 9 have been described above) and the endonuclease assays were performed at 37˚C for 2 hours as described above.

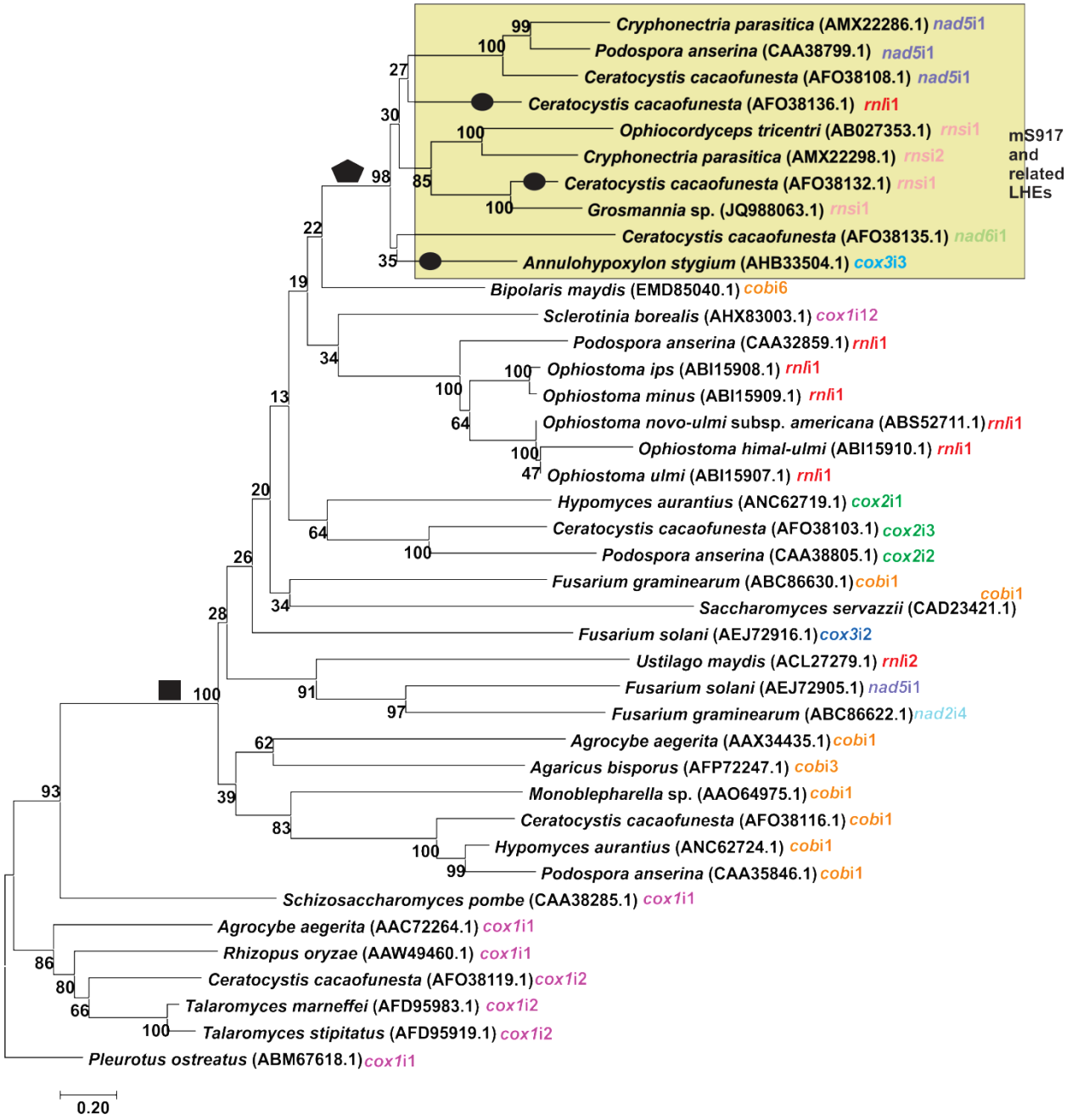
3.4. Results

3.4.1. Phylogeny of mS917 HEases and related LAGLIDADG type ORFs

Fifty-two LAGLIDADG-type HE sequences were extracted from NCBI by using the mS917 IEP (I-CcaI) sequence from *C. cacaofunesta* (YP_007507084.1) as a query. Phylogenetic analysis with three different programs yielded similar tree topologies (Fig. 3.2) showing that 12 sequences, all encoded within group ID introns appear to be derived from a common ancestor. This grouping includes five mS917 IEPs along with four IEPs encoded within the *nad5* gene, one IEP encoded within the *rnl* gene, one IEP from the *cox3* gene, and one IEP from the *nad6* gene.

This set of orthologues appears (bootstrap support of node >95 %) to have inserted into at least five different genes, including rRNA and protein coding genes. It is also worth noting that the mS917 introns in *Ophiocordyceps tricentri* and *C. parasitica* are complex and appear to be nested introns where the internal group ID intron encodes the LHE and external group ID does not contain an ORF (Hafez *et al*., 2013). Originally, the goal was to characterize a member from each host gene (including *nad5* and *nad6*) and from the nested mS917 intron arrangements; however, we could only overexpress IEPs and show activity for IEPs from *C. cacaofunesta* YP_007507084.1 (*rns*), *C. cacaofunesta*YP_007507038.1 (*rnl*), and *A. stygium* NC_023117.1 (*cox3*).

Figure 3.2. Phylogenetic tree inferred using the Neighbor-Joining method for the aa sequence of the HEease similar to the mS917 ORFs: The optimal tree with the sum of branch length = 26.11071834 is shown. The percentages at the nodes are based on bootstrap analysis (1000 replicates). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method (Jones *et al*., 1992) and are in the units of the number of amino acid substitutions per site. The analysis involved 40 amino acid sequences of LHEases and ambiguous alignment positions were removed for each sequence pair. There were a total of 781 positions in the final datasets. The node marked with black octagon supports the monophyly of the mS917 clade of homing endonucleases. The black circles indicate the position of those members that were tested in this study for activity. The node marked by the black square supports the monophyly of the I-OnuI family of homing endonucleases.

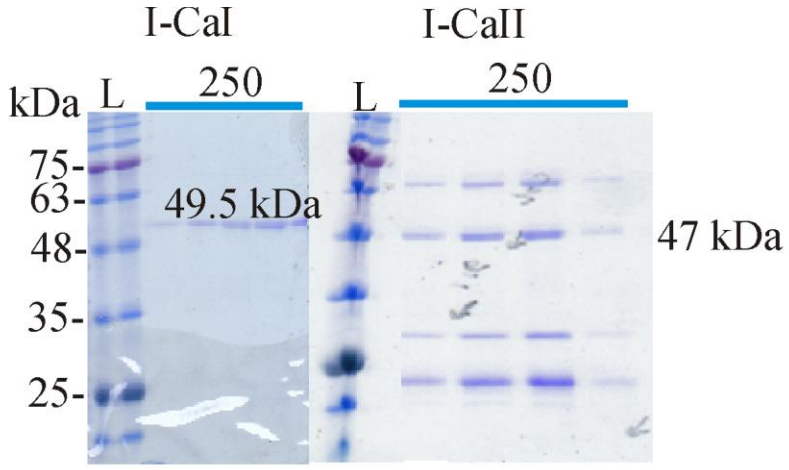

Cryphonectria parasitica (AMX22286.1) nad5i1
Podospora anserina (CAA38799.1) nad5i1
Ceratocystis cacaofunesta (AFO38108.1) nad5i1
Ceratocystis cacaofunesta (AFO38136.1) rnli1
Ophiocordyceps tricentri (AB027353.1) rnsi1
Cryphonectria parasitica (AMX22298.1) rnsi2
Ceratocystis cacaofunesta (AFO38132.1) rnsi1
Grosmannia sp. (JQ988063.1) rnsi1
Ceratocystis cacaofunesta (AFO38135.1) nad6i1
Annulohypoxylon stygium (AHB33504.1) cox3i3
mS917 and related LHEs
Bipolaris maydis (EMD85040.1) cobi6
Sclerotinia borealis (AHX83003.1) cox1i12
Podospora anserina (CAA32859.1) rnli1
Ophiostoma ips (ABI15908.1) rnli1
Ophiostoma minus (ABI15909.1) rnli1
Ophiostoma novo-ulmi subsp. americana (ABS52711.1) rnli1
Ophiostoma himal-ulmi (ABI15910.1) rnli1
Ophiostoma ulmi (ABI15907.1) rnli1
Hypomyces aurantius (ANC62719.1) cox2i1
Ceratocystis cacaofunesta (AFO38103.1) cox2i3
Podospora anserina (CAA38805.1) cox2i2
Fusarium graminearum (ABC86630.1) cobi1
Saccharomyces servazzii (CAD23421.1) cobi1
Fusarium solani (AEJ72916.1) cox3i2
Ustilago maydis (ACL27279.1) rnli2
Fusarium solani (AEJ72905.1) nad5i1
Fusarium graminearum (ABC86622.1) nad2i4
Agrocybe aegerita (AAX34435.1) cobi1
Agaricus bisporus (AFP72247.1) cobi3
Monoblepharella sp. (AAO64975.1) cobi1
Ceratocystis cacaofunesta (AFO38116.1) cobi1
Hypomyces aurantius (ANC62724.1) cobi1
Podospora anserina (CAA35846.1) cobi1
Schizosaccharomyces pombe (CAA38285.1) cox1i1
Agrocybe aegerita (AAC72264.1) cox1i1
Rhizopus oryzae (AAW49460.1) cox1i1
Ceratocystis cacaofunesta (AFO38119.1) cox1i2
Talaromyces marneffei (AFD95983.1) cox1i2
Talaromyces stipitatus (AFD95919.1) cox1i2
Pleurotus ostreatus (ABM67618.1) cox1i1
0.20

3.4.2. HEases protein overexpression in *E. coli* and purification

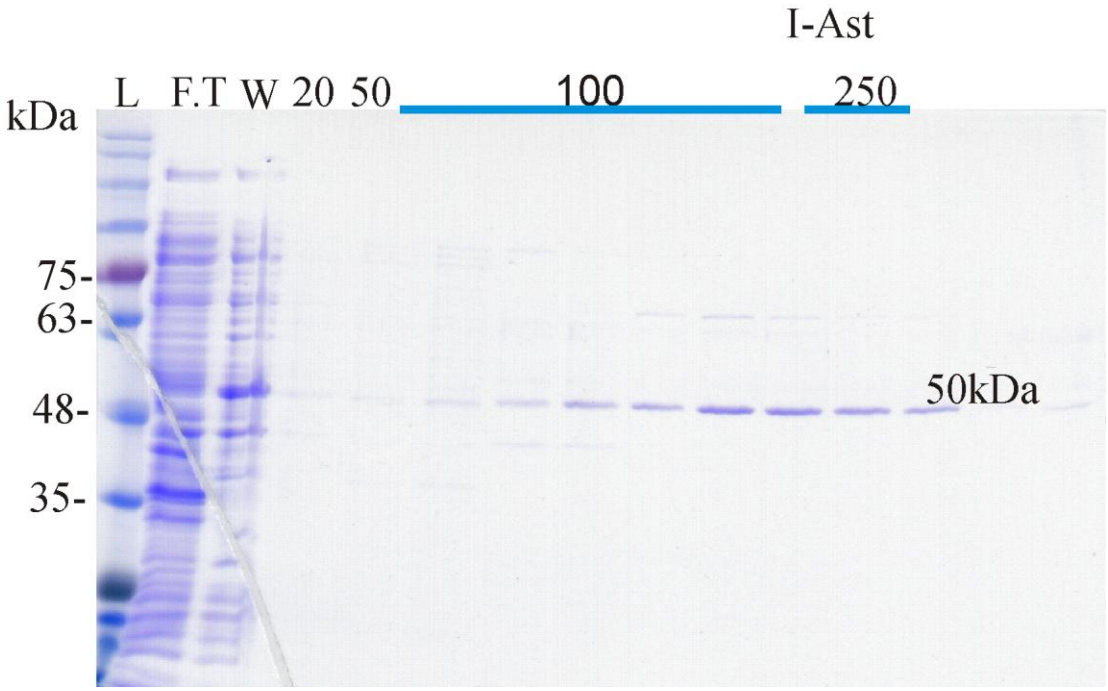
In total, 12 mS917 HEases and related LAGLIDADG HEases were identified and protein overexpression was attempted for three LHEs ORFs embedded within introns located in the *rns*, *rnl* and *cox3* genes. All of the studied ORFs were double-motif LAGLIDADG HEGs and their lengths ranged from 433-551 aa. Protein expression in *E. coli* BL21 (λDE3) was induced with 0.5 mM IPTG and cultures were incubated at room temperature for 16 hours following induction. Therefore, this particular HE was further analysed with regards to temperature preference and the minimal sequence required for DNA recognition and cleavage activity.

Figure 3.3. Protein purification: (a) SDS–PAGE (12.5 %) gel of I-CcaI and I-CcaII purification by Ni-NTA resin (Qiagen). Lanes are indicated as follows: M = molecular weight markers that contain the BLUeye prestained protein ladder (GeneDireX; FroggaBio, Toronto, Ontario). The column was first washed with low concentrations of imidazole (25, 50 and 100 mM). The displayed bands represent I-CcaI (49.5 kDa) and I-CcaII (47 kDa) proteins respectively when eluted with 250 mM imidazole. Note that the I-CcaII protein could only be partially purified. (b) SDS–PAGE (12.5%) gel for the I-AstI purification by Ni-NTA resin (Qiagen). Lanes are indicated as follows: M = molecular weight markers; FT = flow through; W = wash and E = elution by (250 mM imidazole). Lane L contains the BLU eye prestained protein ladder (GeneDireX; FroggaBio, Toronto, Ontario).

a


I-CaI
I-CaII
kDa
L
250
L
250
75-
63-
48-
35-
25-
49.5 kDa
47 kDa

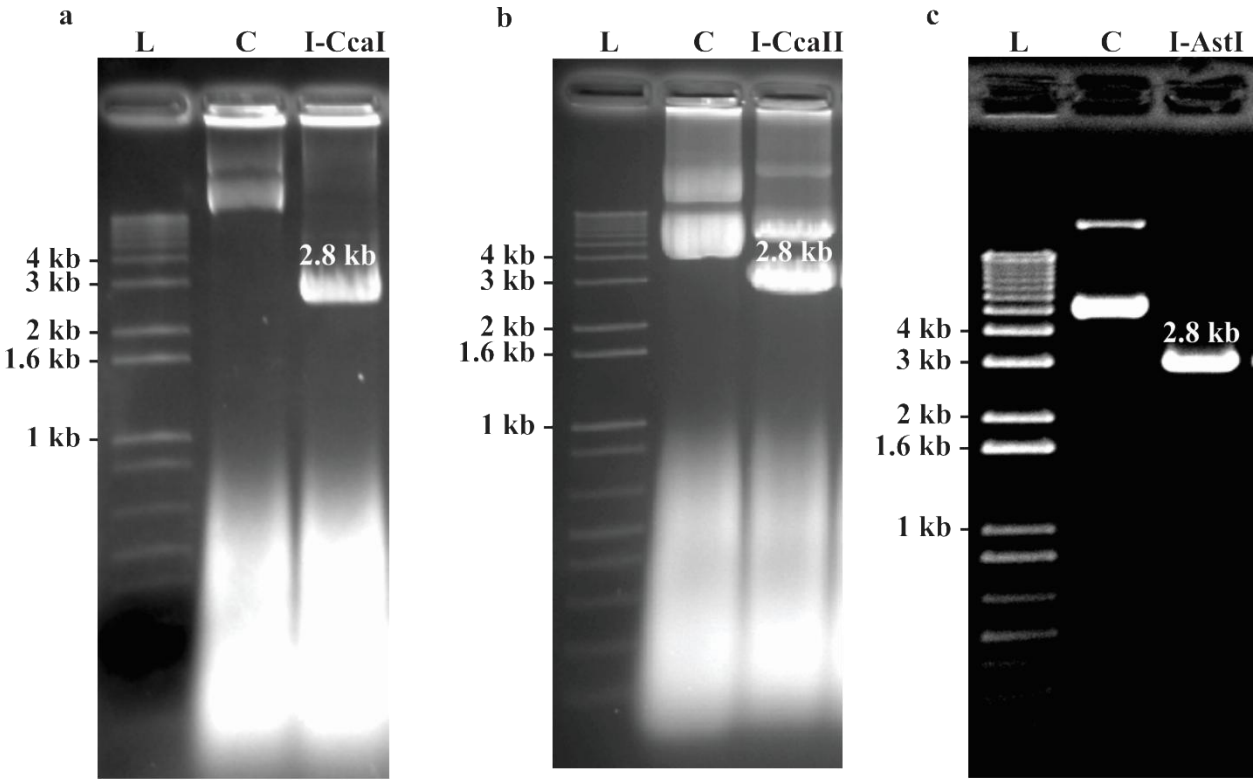
b


I-Ast
kDa
L
F.T
W
20
50
100
250
75-
63-
48-
35-
50kDa

3.4.3. Endonuclease activity of the mS917 IEP orthologues

Endonuclease assays were performed by incubating the purified LHEs (I-CcaI, I-AstI, and I-CcaII) with the appropriate plasmid substrates (p*rns*-SUB, p*cox3*-SUB and p*rns*-SUB). The HE activity was tested at 37˚C at different time periods (0, 30, 60, 90, and 120 min) and the best completely linearized substrates (2.8 kb) were already observed at 30 min (Fig. 3.4) for I-CcaI, I-AstI, and I-CcaII. The control assays, i.e. untreated (no HE) substrates (p*rns*-SUB, p*cox3*-SUB, p*rns*-SUB), showed no cleavage.

Figure 3.4. Gel images for the *in vitro* endonuclease assays: (a) I-CcaI was incubated with the circular p*rns*-SUB plasmid at 37°C for 30 min. (b) I-CcaII was incubated with the circular *rnl*-SUB plasmid for 30 min. (C) I-AstI was incubated with the circular *cox3*-SUB plasmid at room temperature overnight. Lanes denoted with "L" contain the 1 kb plus DNA ladder (Thermo Fisher Scientific); Lanes denoted with "C" contain the untreated plasmid substrate.



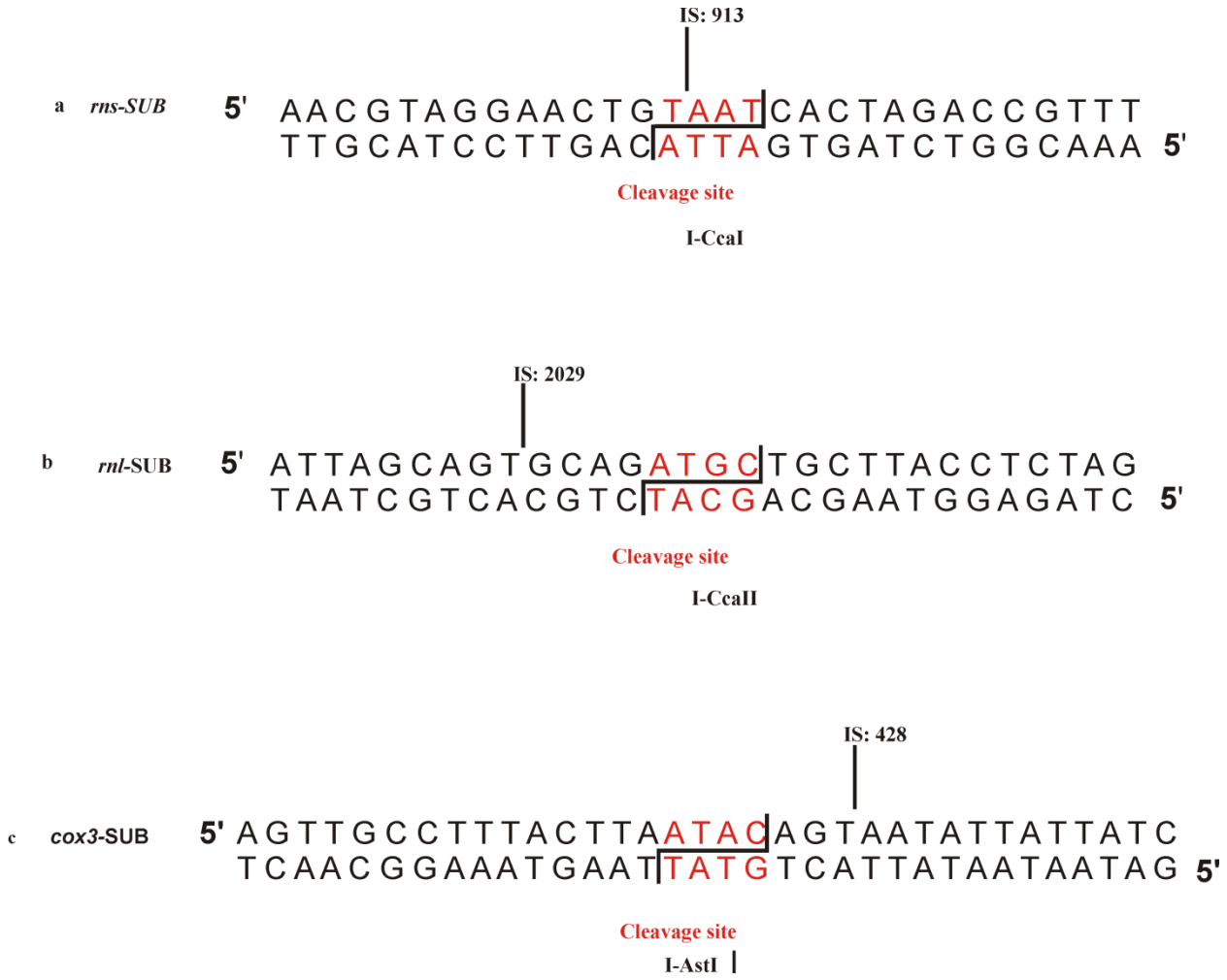
p*rns*-SUB

p*rnl*-SUB

p*cox3*-SUB

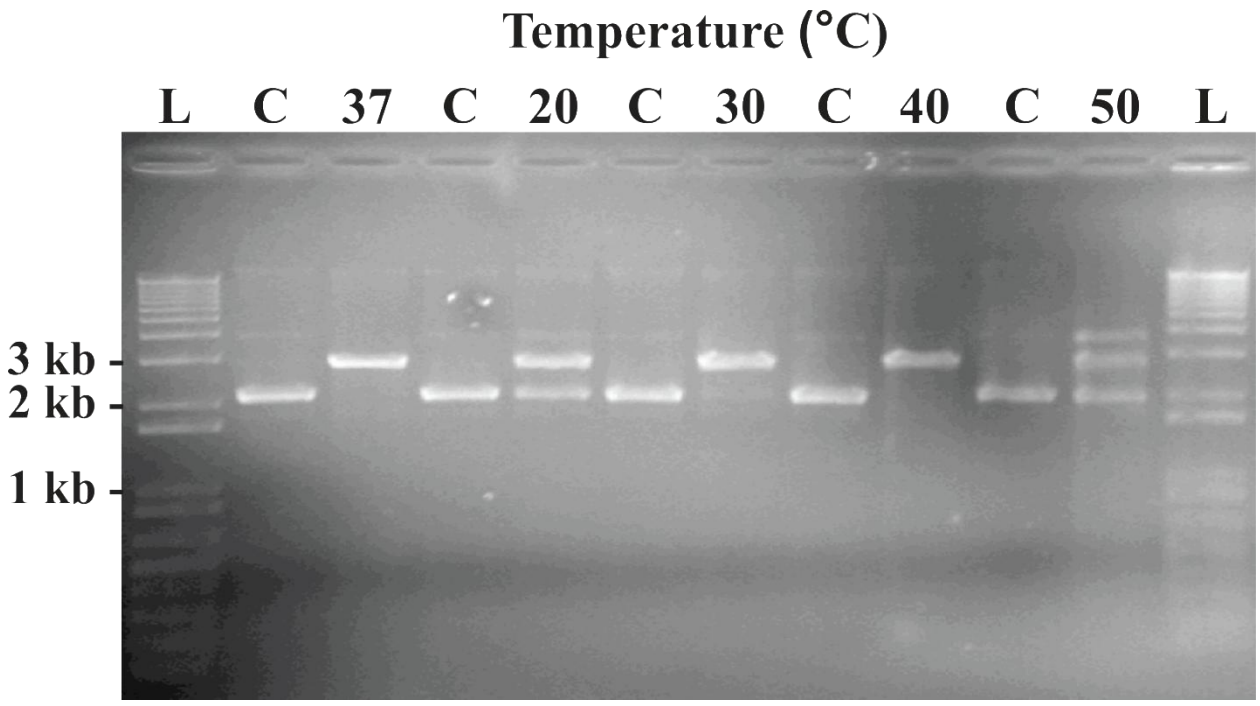
The *in vitro* cleavage mapping assay for I-CcaI, I-AstI, and I-CcaII confirmed that these LHEs are active and generate staggered ends with 4 nt overhang at the 3′ end (Fig. 3.5). The HE cleaved and T4 DNA pol treated substrates (p*rns*-SUB, p*cox3*-SUB, p*rns*-SUB) showed that the following 4 nts were removed 5′-TAAT 3′, 5′-ATAC-3′, and 5′-ATGC-3′ by I-CcaI, I-AstI, and I-CcaII respectively. These 4 nts at the cleavage site are referred to as the "central four recognition motif" (CRM, reviewed in Lambert *et al*., 2016).

Figure 3.5. Cleavage site mapping results: Schematic for the *rns* (a), *rnl* (b), *cox3* (c) substrates and the cleavage sites for I-CcaI, I-CcaII, and I-AstI respective. The staggered cuts and the four central recognition motifs for the HEases are indicated. The four nucleotides 3′ overhangs generated by each of the enzyme and removed by T4 DNA polymerase are shown in red. The intron insertion sites were designated based on reference sequences, for the rns and rnl genes the insertion sites are based on *E. coli* rDNA (AB035922.1) as proposed by Johansen and Haugen (2001). The cox3 intron insertion site is based on comparison with the *S. cerevisiae cox3* sequence (KP263414.1).


IS: 913
a rns-SUB
5' AACGTAGGAACTGTAATCACTAGACCGTTT
TTGCATCCTTGACATTAGTGATCTGGCAAA 5'
Cleavage site
I-CcaI
IS: 2029
b rnl-SUB
5' ATTAGCAGTGCAGATGCTGCTTACCTCTAG
TAATCGTCACGTCTACGACGAATGGAGATC 5'
Cleavage site
I-CcaII
IS: 428
c cox3-SUB
5' AGTTGCCTTTACTTAATACAGTAATATTATTATC
TCAACGGAAATGAATTATGTCATTATAATAATAG 5'
Cleavage site
I-AstI

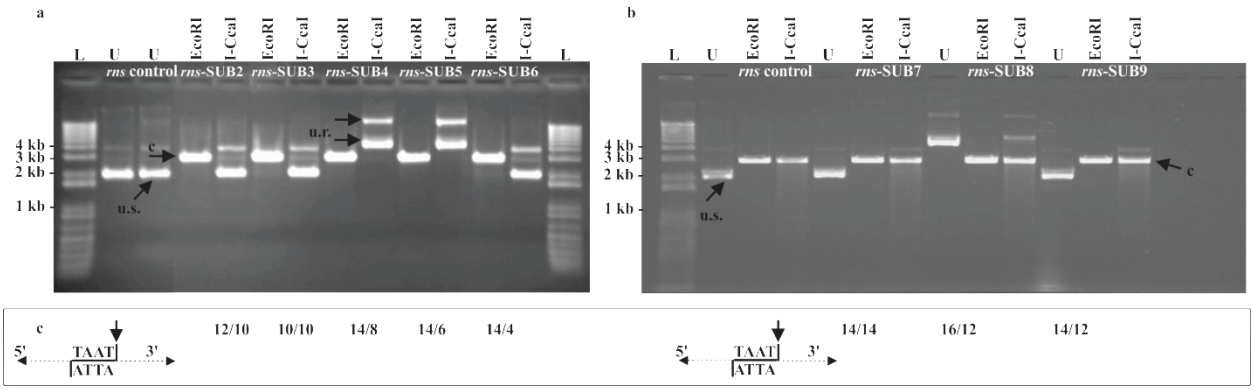
I-CcaI activity was tested at different temperatures (20˚C, 30˚C, 37˚C, 40˚C, and 50˚C) (Fig. 3.6). The enzyme showed activity at 30˚C, 37˚C, 40˚C as the expected linearized product can be observed on the agarose gel. However, assays performed at 20˚C and 50˚C yielded multiple products indicating partial or no activity, as the substrate plasmid may not have been cut and thus appears in its various supercoiled forms on the agarose gel (Fig. 3.6).

Figure 3.6. Effect of temperature on I-CcaI endonuclease activity: The plasmid substrate was incubated with I-CcaI at different temperatures (range from 20°C to 50°C). The "C" (control) lane contains the untreated substrate plasmid, while the lanes denoted with "L" contain the 1 kb plus DNA ladder (ThermoFisher Scientific).


Temperature (°C)
L C 37 C 20 C 30 C 40 C 50 L
3 kb
2 kb
1 kb

I-CcaI was further characterized to estimate the minimal recognition sequence; so this LHE was challenged with 9 different synthesized substrates containing variations of the known target sequence and the results showed (Fig. 3.7) that the enzyme lost its activity when the target sequence was less than 14 nts upstream of the actual cleavage site and less than 12 nts downstream from the cleavage site; i.e. the enzyme required a DNA recognition site that is estimated to be about 26 nts in length.

Figure 3.7. Gel images for the I-CcaI cleavage assays with various substrates (see Fig. 3.1). I-CcaI was incubated for 30 min with nine different plasmids (control p*rns*-SUB and p*rns*-sub2 to p*rns*-sub9) containing various length versions of the target region for I-CcaI. Lanes denoted with "L" contain the 1 kb plus DNA ladder (Invitrogen); Lanes denoted with "U" contain the untreated (no HE) corresponding plasmid substrate (p*rns*-sub2 to p*rns*-sub9); Lanes denoted with "EcoRI" contain positive controls where the substrates were incubated with the EcoRI restriction enzyme in order to show the migration of the cleaved substrate plasmids. Panel (a) shows the results for substrates that failed to be cut by I-CcaI and panel (b) shows the results for those substrates that were cleaved by I-CcaI. Substrate plasmids that are untreated (U) or have been treated with EcoRI are the negative and positive controls, respectively. Panel (c) below the gels indicates how many nucleotides upstream and downstream (upstream/downstream) of the cleavage site (indicted by arrow) are present within the corresponding substrate plasmid. Uncut plasmids appear behave as supercoiled and relaxed plasmids thus their migration patterns on the gel are quite variable, migrating faster or slower in comparison to linearized versions. Note: c = position of cut plasmids, u.r. notes the position of uncut relaxed plasmids, and u.s. denotes the position of uncut supercoiled plasmids.


a
L U U EcoRI I-CcaI EcoRI I-CcaI EcoRI I-CcaI EcoRI I-CcaI EcoRI I-CcaI L
rns control rns-SUB2 rns-SUB3 rns-SUB4 rns-SUB5 rns-SUB6
4 kb
3 kb
2 kb
1 kb
c
u.s.
u.r.
b
L U EcoRI I-CcaI U EcoRI I-CcaI U EcoRI I-CcaI U EcoRI I-CcaI
rns control rns-SUB7 rns-SUB8 rns-SUB9
4 kb
3 kb
2 kb
1 kb
u.s.
c
c
5'
TAAT
ATTA
3'
12/10
10/10
14/8
14/6
14/4
5'
TAAT
ATTA
3'
14/14
16/12
14/12

3.5. Discussion

3.5.1. Phylogeny of mS917 HEases and related LAGLIDADG type ORFs

In this study fifty-two LHE sequences were extracted from NCBI and one clade (bootstrap support >95%) could be identified that included 12 sequences (Fig. 3.1). Members of this group (from now on referred to as the mS917 clade) have inserted into at least five different genes which suggest that these LHEs have evolved to recognize five different target sites. In order to avoid extinction, HEGs have to be able to move into new target sites or gain new function (Goddard and Burt, 1999). The fact that members of this clade have been able to invade new insertion sites is also relevant to the application of LHEs as rare cutting DNA enzymes in biotechnology. Considerable efforts have been made to redesign homing endonucleases to cut at different target sites. This study presents examples where closely related LHEs have naturally adapted to different target sites. This warrants further investigation in the future to establish what features/amino acid changes might be involved in allowing LHEs to insert into, i.e. recognize new target sites.

The phylogenetic tree (Fig. 3.1) shows that the 917 LHE orthologues belong to the I-OnuI family of HEases. I-OnuI was originally described and characterized by Gibb and Hausner (2005) and Sethuraman *et al*. (2009) and extensively studied in detail by Takeuchi *et al*. (2011). The latter study showed that I-OnuI can be engineered to recognize and cut genes such as the MAO-B gene involved in mono-genic human diseases (involved in neurodegenerative disorders including Parkinson's disease) that contains a DNA sequence that differs from the native I-OnuI target site by only five bp positions (Takeuchi *et al*., 2011). In a recent major study presenting a survey of HEGs belonging to the I-OnuI family, Lambert *et al*. (2016) demonstrated the potential utility of this family of LHEs with regards to applications in biotechnology.

One of the limitations of LHEs, in particular for members of the I-OnuI family, with regards to retargeting them for genome editing applications usually is the conserved nature of the 4 bp sequence comprising the central recognition motif (CRM, i.e., cleavage site). The three members of the mS917 clade characterized in this study displayed some variability with regards to the CRM (5′- TAAT-3′ for I-CcaI, 5′-ATAC-3′ for I-AstI, and 5′-ATGC-3′ for I-CcaII) and this would make the mS917 clade an attractive HE protein scaffold that could be modified with regards to targeting a wider variety of potential target sites (Lambert *et al*., 2016). In addition, harnessing the DNA cutting domains of these proteins and combining them with more programmable DNA binding domains currently applied to synthetic endonucleases such as TALENS or ZFNs might be another promising direction for utilizing homing endonucleases that belong to the I-OnuI family (Hafez and Hausner, 2012; Wolfs *et al*., 2014; Boissel *et al*., 2014; Romano *et al*., 2016).

3.5.2. The expression and purification of LHEs

The study showed that I-CcaI, I-CcaII, and I-AstI can be overexpressed in *E. coli* and purified utilizing the N-terminal His-tag in sufficient amounts to demonstrate their activity as endonucleases and to map their cleavage sites. HEases are important because of their high specificity and they can be found as components of mobile elements (i.e. group I introns) (Hafez and Hausner, 2012; Stoddard, 2014). Fungal mitochondrial genomes have been shown to be a resource with regards to recovering new HEases that have new target specificities (Stoddard, 2005, 2011; Baxter *et al*., 2012; Jacoby *et al*., 2012; Taylor *et al*., 2012; Chan *et al*., 2013; Hafez *et al*., 2014; Ferandon *et al*., 2010). Overall, few HEases have been biochemically characterized so far (Marcaida *et al*., 2010; Prieto *et al*., 2012) and this limits the application of HEases with regards to applying them to recognise a wide variety of different target sites. Finding native

HEases is valuable as it offers a starting point from which one can either directly apply the HE sites that fortuitously match the HEases native target site or it offers a protein scaffold that can be reengineered to recognize related sequences in genes that are of economic importance. The overexpression of LHE ORFs in *E. coli* was a challenge with regards to several members of the mS917 group. Although sequences were codon optimized for expression in *E. coli*, I-Cpa (mS917 nested version; Genbank accession number: AAB84210.1), I-CcaIII (*nad5*; Genbank accession number: YP_007507054.1) and I-CcaIV (*nad6*; Genbank accession number: YP_007507089.1) failed to express or purify via Ni-NTA columns. In future efforts, these proteins have to be examined with regards to their surface hydrophobicity, adjustments in their amino acid sequence that may decrease the surface hydrophobicity of these proteins which may allow for improved expression and recovery (Lambert *et al*., 2016). In addition, different affinity tags, such as glutathione S-transferase (GST) or the maltose binding protein (MBP), could be applied to LHEs as it is possible that the N-terminal 6X His-tags are not accessible in some forms due to improper folding and thus they cannot bind to the Ni-NTA column.

3.5.3. Endonuclease and cleavage assays

As stated previously, three members of the mS917 clade could be tested with regards to their DNA target sites; the most consistent HE in our analyses with regards to overexpression and purification was I-CcaI. It was noted that the optimum activity was at 37˚C and activity was highly reduced or lost at 20˚C and 50˚C, respectively. In previous studies, it was shown that endonuclease activity for LHEs tends to be around 37˚C (see I-SceI group and I-CreI; Fonfara *et al*., 2011, Dürrenberger and Rochaix 1993). In contrast, the I-DmoI LHE endonuclease, recovered from the thermophile *Desulfurococcus mobilis*, shows optimal activity at 65˚C (Dalgaard *et al*., 1993; Silva and Belfort, 2004).

Challenging the I-CcaI with 8 different synthesized substrates allowed for gaining better insights into the length of the DNA sequence required by the LHE, which appears to be around 22 nts. This is in the expected range for LHEs belonging to the I-OnuI family with regards to their target site length requirements, which tend to be 26 nts (Lambert *et al*., 2016).

3.5.4. Comparing the I-CcaI, I-CcaII, and I-AstI with other members of the I-OnuI HE family

Twelve sequences from the mS917 clade were compared with seven amino acid sequences representing the I-OnuI family that were studied by Lambert *et al*., (2016). Recent studies showed that the central four nucleotides within the cleavage recognition motif are not in direct contact with the LHE but they are still very important for controlling enzyme activity as this region undergoes bending during cleavage (Curuksu *et al*., 2009; Lambert *et al*., 2016). A comparison between the DNA substrates of LHEs studied herein and the recently characterized members of the I-OnuI family of HEases (Lambert *et al*., 2016) (Fig. 3.8) showed that I-CcaI and I-AstI fit the standard model where the CRM elements are AT-rich. Lambert *et al*., (2016) found that within the CRM sequence (the four positions are defined as -2 -1 +1 +2), the -1,+1 positions are critical for the cleavage activity with regards to the I-OnuI family of HEases and they also noted that G:C are strongly stacked and need high energy to unstack so that might explain why these four bp CRM segments tend to be AT rich. The results of our study showed that the central four nucleotides were 5′-TAAT-3′, 5′-ATGC-3′, and 5′-ATAC-3′ with regards to the cleavage sites for I-CcaI, I-CcaII and I-AstI, respectively. With regards to I-CcaII at the +1 position of the CRM is a G. This would, as stated earlier, extend the range of possible target sites this enzyme could be optimized for. However, with the same token, it may make this enzyme less effective as

according to Lambert *et al*. (2016), a G:C nt in the +1 site of the CRM could affect the enzyme's activity, as bending and melting G:C base pairs is energetically more difficult.

Figure 3.8. An alignment for the central four recognition motifs (CRMs in red) for members of the I-OnuI family of homing endonucleases (Lambert *et al.*, 2016) A sequence logo (Crooks *et al.*, 2004) is presented below the alignment to indicated sequence conservation among the various CRMs. See text for more details.

CRM sequence

-2 -1 +1 +2

I-CcaI	A	G	G	A	A	C	T	G	T	A	A	T	C	A	C	T	A	G	A
I-CcaII	C	A	G	T	G	C	A	G	A	T	**G**	C	T	G	C	T	T	A	C
I-AstI	T	T	T	A	C	T	T	A	A	T	A	C	A	G	T	A	A	T	A
I-AabMI	G	G	T	A	C	C	C	T	T	T	A	A	A	C	C	T	A	C	T
I-CpaMI	A	G	C	C	C	A	C	A	A	T	A	T	T	A	A	G	G	C	C
I-GpeMI	T	T	C	C	G	C	T	T	A	T	T	C	A	A	C	C	C	T	T
I-GzeII	T	G	G	G	T	A	C	C	A	T	A	T	T	G	G	T	A	C	A
I-LtrI	A	T	G	C	T	C	C	T	A	T	A	C	G	A	C	G	T	T	T
I-LtrWI	A	G	T	A	G	T	G	A	A	G	T	A	T	G	T	T	A	T	T
I-OnuI	T	T	T	C	C	A	C	T	T	A	T	T	C	A	A	C	C	T	T
I-PanMI	G	C	T	C	C	T	C	A	T	A	A	T	C	C	T	T	A	T	C
I-SmaMI	A	T	C	C	T	C	C	A	T	T	A	T	C	A	G	G	T	G	T

However, it should be noted that there are LHEs that appear to tolerate CRM elements that are GC-rich. For example, the intein-encoded PI-PfuI and PI-SceI cleave at 5′-GGGA-3′ (Komori *et al*., 1999) and 5′-GTGC-3′ (Gimble and Thorner, 1992; Bremer *et al*., 1992), respectively. Also, the group I intron-encoded I-CreI cleaves at 5′-GTGA-3′ (Thompson *et al*., 1992; Jurica *et al*., 1998). Beylot and Spassky (2001) noted that in the case of PI-SceI, the G at the CRM+1 position can be methylated, which may affect the energy required for DNA bending and strand separation (Severin *et al*., 2013), but more work appears to be needed to resolve how LHEs deal with GC-rich regions.

3.6. Conclusions

This current study identified a group of native LHEs that have spread into five different genes suggesting they have been able to adapt to five different target sites. The ability of LHEs to invade new sites is relevant in understanding how LHEs avoid extinction (Goddard and Burt 1999) and can persist long term within fungal mitochondrial genomes. Based on the recent work by Lambert *et al.,* (2016) minor amino acid changes can cause a shift where the LHE will bind and cut a different DNA sequence; this is probably exemplified by the LHEs investigated in this study. This study examined a set of phylogenetically related LHEases that have inserted into at least five different target sites. The mS917 LHE clade may warrant further characterization, such as structural analysis to understand what changes at the amino acid level may have contributed towards their ability to shift into different target sites (genes). This information might be useful in developing strategies for engineering LHEs to target sites located within genes of economic importance, such as alleles associated with human diseases.

The LAGLIDADG family of HEases has been extensively studied over the years as potential genome editing tools (Gimble, 2000; Stoddard, 2005, 2011, 2014). Therapeutic

applications demand high precision in gene modification activity and HEases are considered to be compact target-specific ‘molecular scissors’ with little known issues with regards to off-target activities (Stoddard, 2014; Takeuchi *et al*., 2014; Lambert *et al*., 2016). However, one potential drawback for this class of enzymes is the non-modular configuration; the DNA recognition and cleavage functions can be combined within the same protein domain (Hafez and Hausner, 2012). Therefore, engineering of LHE has been challenging and recently more focus has been placed on alternate modular type synthetic endonucleases such as TALEN endonucleases or Zinc finger endonucleases or the currently popular RNA guided CRISPR/Cas9 system (Cox *et al*., 2015). However, a recent study involving members of the I-OnuI family of LHEases, showed that there are multiple points across the LAGLIDADG protein that can be involved in holding metal ions in suitable positions to facilitate cleavage, a finding that may help in reengineering LHEs to new target sites without mitigating the cleavage activity of the enzyme (Lambert *et al*., 2016). These results along with the development of new technologies such as yeast surface display-SELEX hold promise for LHE to be engineered more efficiently in the near future (Takeuchi *et al*., 2015; Lambert *et al*., 2016; Jacoby *et al*., 2016; Romano *et al*., 2016)

4. General Conclusions

This study began a focused on the *rns* gene of the ophiostomatoid fungi and related taxa. The *rns* gene was sequenced and then annotated to determine the positions of introns and intron-encoded proteins (IEP) in order to study the evolution of the fungal mitochondrial introns and intron-encoded proteins. Previously very little was known about the *rns* genes and *rns* introns within *O. piliferum* and related species. This study identified different types of introns (group I and II) and intron-encoded proteins (such as HEases and RTs) within the mitochondrial *rns* gene of *Ophiostoma* and related taxa. The study also provided information about intron insertion sites, intron RNA secondary structures, and some details about the IEP families and their position within their host introns. Frequently, intron encoded open reading frames show evidence of degeneration due to premature stop codons like the GI intron at mS569 and mS1427 insertion sites.

This observation tends to confirm the Goddard and Burt (1999) model of the homing endonuclease life cycle. HEases and their insertion sites identified in this study could have important applications in biotechnology, as HEases are rare cutting endonucleases that have been applied in genetic engineering and genome editing (Hausner, 2003; Lambowitz and Zimmerly, 2010; Stoddard, 2005, 2011, 2014).

The second component of this work focused on a set of homing endonuclease genes that are related to IEPs inserted within the S917 intron of the *rns* gene in some fungi. This position can be occupied by a group I D intron that encodes a double motif LADGLIDADG type HEase. This HEG is particularly interesting in that orthologues for this element were also noted in the *rnl* gene and protein coding genes, suggesting that this HEG has invaded different sites and thus appears to be quite adaptable in invading other genes.

4.1. Major findings of this thesis

4.1.1. The evolutionary dynamics of introns and their encoded proteins among members of the *O. piliferum* complex.

Members of the *O. piliferum* complex are easily confused with one another based on morphological features due to the limiting number of characters and convergent evolution of some traits due to adaptation towards insect dispersal (reviewed in Hausner *et al*. 2003). Phylogenetic analysis, based on ITS sequences, also failed to distinguish among members of this species complex but it does indicating the presence of closely related species/strains. This allows one to potentially observe more recent events with regard to the evolutionary dynamics of mitochondrial introns and their encoded proteins among members of the *O. piliferum* complex.

Some interesting highlights were the *Ophiostoma perfectum* mS379 RT ORF that appears to be related to RT sequences found in brown algae, bacteria, plants and other fungi. This means this group II intron ORF has a complex evolutionary history that very likely includes several horizontal transfers. Another intron noted in this study was the mS952 GII intron. Its distribution and phylogenetic analysis suggests that this intron is predominately vertically inherited, with some evidence of loss and possible horizontal gene transfer (HGT). The mS952 intron is noteworthy as it is a group II intron that encodes a LAGLIDADG type ORF commonly found encoded within group I introns. We presume that this intron has adapted to be mobilized by a DNA based mobility mechanism driven by its LHE. This is unlike typical group II introns that are mobilized via an RNA intermediate that is converted into cDNA, and catalyzed in part by a reverse transcriptase. This work also noted the mS1224 GI intron which so far has not been reported for members of *Ophistoma.* However, this intron was noted to be present in rather distantly related fungi suggesting again the possibility of HGT or demonstrating the dynamic nature of intron loss and gain among fungal mtDNAs.

4.1.2. GII intron ORFs appear to follow a similar life cycle as suggested for LHEs and group I introns as proposed by Goddard and Burt (1999)

The mS379 GII intron (type AI) was previously reported from: 1 - *O. minus* strains as an intron with either complete or degraded RT ORFs, the latter due to the presence of premature stop codons; 2 - *O. torulosum* as an intron with a degraded RT ORF; 3 - *O. hyalothecium* where the mS379 intron completely lacked the RT ORF; and from the current study, 4 - *O. perfectum* WIN(M)823 where the mS379 intron encodes a complete RT ORF. The results suggest that this group II intron RT ORF follows a similar life cycle as noted for GI introns and their HEases ORFs as suggested by Goddard and Burt (1999), a life cycle of invasion and degeneration for HEGs and their hosting GI introns. Previously group II introns (ribozyme component) were noted to co-evolve with their ORFs and that ORFless group II introns evolved from ORF-containing introns (Toor *et al.,* 2001) but no overall model for group II intron evolution has currently been proposed. Our study suggests that RT ORFs in GII introns can follow a similar life cycle; i.e. invasion into an empty spot, followed by slow degeneration initially of the ORF, and eventually complete loss of the ORF, and presumably, complete loss of the intron there by regenerating a potential site available for reinvasion.

4.1.3. Closely related LHEs have adapted to different target sites

This study identified a clade (bootstrap support >95%) that included 12 LHE orthologues. Members of this group have inserted into at least 5 different genes which could imply that these LHEs have changed their DNA binding affinity and now can recognize five different target sites. The ability of LHEs to invade new sites is relevant in understanding how LHEs avoid extinction and can persist long term within fungal mitochondrial genomes. This clade belongs to the I-OnuI family of LHEs that has been touted by Stoddard and others (Takeuchi *et al*., 2011; Lambert *et*

al., 2016; Baxter *et al*., 2012; Chan *et al*., 2013; Taylor *et al*., 2012, Baxter *et al*., 2014) to be a valuable resource for the development of meganucleases that can be applied as genome editing tools. This study identifies three new active members to the I-OnuI family of LHEs. Also, the mS917 clade members display some plasticity with regards to being able to adapt to different DNA target sites; this might be a useful property with regards to these enzymes being retargeted to new target sites that are related to genes involved in diseases or genes of economic importance.

4.2. Future prospects

4.2.1. Identifying the contact map for members of the mS917 clade

Members of the mS917 clade deserved to be further investigated with regards to their ability to bind and cleave DNA target sites. This could be accomplished in part by examining the structure of the LHE in contact with its substrate with X-ray crystallography. Similar to a recent study by Lambert *et al*. (2016) this may help in defining the components of LHEs involved in DNA recognition and cleavage; features that need to be known to retarget these enzymes to new sites. In the short term one could use the data from the work presented in this thesis and examine if economically important genes and genes related to human genetic diseases contain sequences that could be cut by the characterized members of the mS917 clade.

4.2.2. Bioprospecting for more mS917 HEases

Screening for more members of the mS917 HE clade by sequencing more fungal mtDNAs and by analyzing mtDNA sequences deposited in GenBank and at JGI. Blastp could be used to screen NCBI and JGI (http://genome.jgi.doe.gov/; Joint Genome Institute and MyoCosm; http://genome.jgi.doe.gov/programs/fungi/index.jsf) for more members of this HEases group. This can assist in finding more examples of these HEases that moved into new/different genes

that could be evaluated for activity and could be used to build up a "catalog" of mS917 clade HEases that might cut different DNA sequences.

4.2.3. Future prospects for Meganucleases:

Meganucleases (such as LHEs) are gene targeting nucleases (Lambert *et al*., 2016) and they are recognized as "highly specific molecular scissors" (Stoddard, 2005, 2011, 2014). However, redesigning meganucleases (i.e. homing endonucleases) is difficult; due to the non-modular configuration of meganucleases (Hafez and Hausner, 2012). In recent years there has been a focus on modular type synthetic endonucleases such as TALENS or Zn-finger endonucleases or the currently popular RNA guided CRISPR/Cas9 system (Cox *et al*., 2015). For these it is possible to alter the DNA binding site without compromising the DNA cleavage domains and cleavage activities (i.e. DNA target site specificity or endonuclease activity). For example, the CRISPR/Cas9–RNA guided system – can be easily retargeted without major protein engineering (as would be required for meganucleases) but CRISPR/Cas9 is not quite as precise (i.e., there can be off target activities) and the PAM sequences are critical requirement. PAM stands for Protospacer adjacent motif and depending on the source of Cas9 it can be a 2-6 base pair DNA sequence immediately following the DNA sequence to be cleaved by Cas9. Although PAM sequences are short; they need to be taken into account when CRISPR/Cas9 reagents are designed to target a specific gene. For targeted mutagenesis this might be less of a concern but for potential therapeutic applications off- target activity by CRISPR/Cas9 would be of major concern. Meganucleases may still have applications were any off-target activity is not acceptable, here the extra effort and costs of engineering LHEs might be warranted. Recently it has been shown that TALENS or ZN-fingers can be combined with cleavage domains derived from LHEs (such as megaTALS; Boissel *et al*., 2014) and thus LHEs are potential protein

scaffolds for engineering new meganucleases or “hybrids” that utilize a variety of modular DNA binding strategies with new DNA cutting specificities (Stoddard *et al*., 2008; Marcaida *et al*., 2010; Taylor and Stoddard, 2012; Hafez and Hausner, 2012). Therefore, bioprospecting for native HEGs with novel DNA binding and cleavage sites may still be of value with regards to generating new genome editing tools.

5. References

Abu-Amero SN, Charter NW, Buck KW, Brasier CM. 1995. Nucleotide-sequence analysis indicates that a DNA plasmid in a diseased isolate of *Ophiostoma novo-ulmi* is derived by recombination between two long repeat sequences in the mitochondrial large subunit ribosomal RNA gene. Curr Genet 28: 54–59.

Aguileta G, De Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, Gabaldón T. 2014. High variability of mitochondrial gene order among fungi. Genome Biol Evol 6(2): 451-465.

Alberts B. 2008. Molecular biology of the cell. New York: Garland Science.

Ambrosio AB, do Nascimento LC, Oliveira BV, Teixeira PJ, Tiburcio RA, Toledo Thomazella DP, Leme AF, Carazzolle MF, Vidal RO, Mieczkowski P, Meinhardt LW, Pereira GA, Cabrera OG. 2013. Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome. BMC Genet 14: 91-107.

Arnould S, Perez C, Cabaniols J-P, Smith J, Gouble A, Grizot S, Epinat J-C, Duclert A, Duchateau P, Paques F. 2007. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. J Mol Biol 371: 49–65.

Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL, Baker D. 2010.Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. Nucleic Acids Res 38(16): 5601-5608.

Aubert M, Ryu BY, Banks L, Rawlings DJ, Scharenberg AM, Jerome KR. 2011.Successful targeting and disruption of an integrated reporter lentivirus using the engineered homing endonuclease Y2 I-AniI. PLoS One 6(2): e16825.

Bae H, Kim KP, Song JM, Kim JH, Yang JS, Kwon ST. 2009. Characterization of intein homing endonuclease encoded in the DNA polymerase gene of *Thermococcus marinus*. FEMS Microbiol Lett 297: 180-188.

Baidyaroy D, Hausner G, Hafez M, Michel F, Fulbright DW, Bertrand H . 2011. A 971-bp insertion in the *rns* gene is associated with mitochondrial hypovirulence in a strain of *Cryphonectria parasitica* isolated from nature. Fungal Genet Biol 48(8):775-783.

Barzel A, Privman E, Peeri M, Naor A, Shachar E, Burstein D, Lazary R, Gophna U, Pupko T, Kupiec M. 2011. Native homing endonucleases can target conserved genes in humans and in animal models. Nucleic Acids Res 39(15):6646-6659.

Bates R, Buck W, Brasier M. 1993. Molecular relationships between *Ophiostoma ulmi* and the NAN and EAN races of *O. novo-ulmi* determined by restriction fragment length polymorphisms of nuclear DNA. Mycological Res 97: 449--455.

Baxter S, Lambert A, Kuhar R, Jarjour R, Kulshina N, Parmeggiani F, Danaher P, Gano J, Baker D, Stoddard B, Scharenberg. 2012.Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases. Nucleic Acid Res 40 (16):7985-8000.

Baxter SK, Scharenberg AM, Lambert AR. 2014.Engineering and flow-cytometric analysis of chimeric LAGLIDADG homing endonucleases from homologous I-OnuI-family enzymes. Methods Mol Biol 1123:191-221.

Belcour L, Vierny C.1986. Variable DNA splicing sites of a mitochondrial intron: relationship to the senescence process in *Podospora.* EMBO 5(3): 609-614.

Belcour L, Rossignol M, Koll F, Sellem CH, Oldani C. 1997. Plasticity of the mitochondrial genome in *Podospora*. Polymorphism for 15 optional sequences: Group-I, group-II introns, Intronic ORFs and an intergenic region. Curr Genet 31:308-317.

Belfort M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev 17(23):2860-2863. Belfort M, Bonocora RP.2014. Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods MolBiol1123:1-26.

Belfort M and Perlman PS. 1995. Mechanisms of intron mobility. J Biol Chem 270(51): 30237-30240.

Belfort M, Roberts RJ. 1997. Homing endonucleases: keeping the house in order. Nucleic Acids Res 25: 3379-3388.

Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM. 2002. Mobile introns: pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (eds), Mobile DNA II.A S M Press, Washington DC pp. 761-783.

Belfort M, Reaban ME, Coetzee T, Dalgaard JZ. 1995. Prokaryotic introns and inteins: a panoply of form and function. J Bacteriol 177(14): 3897-3903.

Bell-Pedersen D, Quirk S, Clyman J, Belfort M. 1990. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res 18(13): 3763-3770.

Bertrand H. 2000.Role of mitochondrial DNA in the senescence and hypovirulence of fungi and potential for plant disease control. Annu Rev Phytopathol38:397-422.

Beylot B, Spassky A. 2001. Chemical probing shows that the intron-encoded endonuclease I-SceI distorts DNA through binding in monomeric form to its homing site. J Biol Chem 276(27): 25243-25253.

Bibikova M, Golic M, Golic KG, Carroll D. 2002.Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161:1169–1175.

Binz T, Canevascini G.1996.Xylanases from the Dutch elm disease pathogens *Ophiostoma ulmi* and *Ophiostoma novo-ulmi*. Physiol Mol Plant Pathol 49:159-175.

Berget SM, Moore C, Sharp PA.1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA 74(8): 3171-3175.

Bertrand H. 2000. Role of mitochondrial DNA in the senescence and hypovirulence of fungi and potential for plant disease control. Annu Rev Phytopathol38: 397-422.

Bhattacharya D, Friedl T, Helms G. 2002. Vertical evolution and intragenic spread of lichen-fungal group I introns. J Mol Evol 55: 74-84

Bhattacharya D, Reeb V, Simon DM, Lutzoni F.2005. Phylogenetic analyses suggest reverse splicing spread of group I introns in fungal ribosomal DNA. BMC EvolBiol5:68-79.

Binz T, Canevascini G.1996. Differential production of extracellular laccase in the Dutch elm disease pathogens *Ophiostoma ulmi* and *O. novo-ulmi*. Mycol Res 100(9): 1060-1064.

Birgisdottir AB, Johansen S. 2005.Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. Nucleic Acids Res 33: 2042-2051.

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H. 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. Mol Ecol Resource 13(2):218-224

Boissel S, Jarjour J, Astrakhan A, Adey A, Gouble A, Duchateau P, Shendure J, Stoddard BL, Certo MT, Baker D, Scharenberg AM. 2014. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. Nucleic Acids Res 42(4):2591-2601.

Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring, RB, Stoddard BL. 2003.Structural and biochemical analyses of DNA and RNA binding by a

bifunctional homing endonuclease and group I intron splicing factor. Genes Dev 17(23): 2875-2888.

Brasier CM. 1979.Dual origin of recent Dutch elm disease outbreaks in Europe. Nature 281: 78-79.

Brasier CM, Buck KW.2001. Rapid evolutionary changes in a globally invading fungal pathogen (Dutch elm disease). Biol Invasions 3: 223–233.

Brasier CM, Kirk SA. 2001. Designation of the EAN and NAN races of *Ophiostoma novo-ulmi*as subspecies. Mycol Res 105: 547-554.

Bremer MC, Gimble FS, Thorner J, Smith CL. 1992. VDE endonuclease cleaves *Saccharomyces cerevisiae* genomic DNA at a single site: physical mapping of the VMA1 gene. Nucleic Acids Res 20(20):5484.

Bullerwell CE, Forget L, Lang BF. 2003. Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. Nucleic Acid Res 31(6):1614-1623.

Bullerwell CE, Lang BF. 2005. Fungal evolution: the case of the vanishing mitochondrion. Curr Opin Microbiol 8(4): 362-369

Burke JM. 1988. Molecular genetics of group I introns: RNA structure and protein factors required for splicing. Gene 73:273-294.

Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW, Tabak HF.1987.Structural conventions for group-I introns. Nucleic Acids Res15: 7217-7222.

Burt A. 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proc R Soc A270: 921-928.

Burke JM, Breitenberger C, Heckman JE, Dujon B, Raj B.1984. *Cytochrome b* gene of *Neurospora* mitochondria: partial sequence and location of the introns at sites different from those in *S. cerviciae* and *A. nidulanse*. J Biol Chem 259:504-511

Burt A, Trivers R. 2006. Genes in conflict: The biology of selfish genetic elements. Cambridge, MA: Belknap Press pp. 1-18.

Chan YS, Takeuchi R, Jarjour J, Huen DS, Stoddard BL, Russell S. 2013. The design and in vivo evaluation of engineered I-OnuI-based enzymes for HEG gene drive PLoS One 8(9): e74254.

Cech, TR.1988. Conserved Sequences and structures of group I introns: Building an active site for RNA catalysis. Gene 73: 259-271.

CechTR. 1990. Self-splicing of group-I introns. Annu Rev Biochem 55: 599-629.

Cech TR, Damberger SH Gutell ER. 1994. Representation of the secondary and tertiary structure of group I introns. Nature Struct Biol 1: 273-280.

Chan YS., Takeuchi R., Jarjour J., Huen DS., Stoddard BL, Russell S. 2013. The Design and in vivo Evaluation of Engineered I-OnuIBased Enzymes for HEG Gene Drive. Pols one . 8(9): e74254.

Chan SH, Stoddard BL, Xu SY. 2011. Natural and engineered nicking endonucleases -from cleavage mechanism to engineering of strand-specificity. Nucleic Acids Res 39(1): 1-18.

Charter W, Buck W, Brasier CM. 1993. De-novo generation of mitochondrial DNA plasmids following cytoplasmic transmission of a degenerative disease in *Ophiostoma novoulmi*. Curr Genet 24: 505–514.

Charter W, Buck W, Brasier CM. 1996. Multiple insertions and deletions determine the size differences between the mitochondrial DNAs of the EAN and NAN races of *Ophiostoma*

novo-ulmi. Mycol Res 100: 368–372.

Chevalier B, Monnat RJ, Stoddard BL. 2005.The LAGLIDADG homing endonuclease family. In: Belfort M, Derbyshire V, Stoddard BL, Wood DL, (eds). Homing endonucleases and inteins. New York, NY: Springer, pp. 33-47

Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 29: 3757-3774.

Chow LT, Roberts JM, Lewis JB, Breoker TR.1977. A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA: DNA hybrids. Cell 11(4):819-36.

Christiansen E.1985. *Ceratocystis polonica* inoculated in Norway spruce: Blue-staining in relation to inoculum density, resinosis and tree growth. Eur J Forest Pathol 15: 160-167.

Colleaux L, D'Auriol L, Betermier M, Cottarel G, Jaquier A, Galibert F, Dujon B.1986. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. Cell 44(4):521-533

Colleaux L, d'Auuriol L, Galibert F, Dujon B. 1988. Recognition and cleavage site of the intron encoded *omega* transposase. Proc Natl Acad Sci USA 85: 6022-6026.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823.

Copertino DW, Hallick RB.1991. Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene.EMBO J 10(2):433-442.

Copertino DW, Hallick RB. 1993. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA intronsTrends Biochem Sci 18 (12): 467-471.

Costa M, Dème E, Jacquier A, Michel F.1997. Multiple tertiary interactions involving domain II of group II self-splicing introns. J Mol Biol267: 520-536.

Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell 94:451-462.

Cummings DJ, Michel F, McNally KL. 1989. DNA sequence analysis of the 24.5 kilo base pair cytochrome oxidase subunit I mitochondrial gene from *Podospora anserina*: a gene with sixteen introns. Curr Genetics16: 381-406.

Cox DB, Platt RJ, Zhang F. 2015. Therapeutic genome editing: prospects and challenges. Nat Med 21(2): 121-131.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186:757–761

Curuksu J, Zacharias M, Lavery R, Zakrzewska K. 2009. Local and global effects of strong DNA bending induced during molecular dynamics simulations. Nucleic Acids Res 37(11):3766-3773.

Dalgaard JZ, Garrett RA, Belfort M. 1993. A site-specific endonuclease encoded by a typical archaeal intron. Proc Natl Acad Sci USA 90(12):5414-5417.

Dalgard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS. 1997. Statistical modeling and analysis of the LAGLIDADG family of site specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Res 25: 4626-4638.

Daniels DL, Michels WJ Jr, Pyle AM. 1996. Two competing pathways for self-splicing by group II introns: a quantitative analysis of *in vitro* reaction rates and products. J Mol Biol 256: 31-49.

Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S. 2009.Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. Nucleic Acids Res 37(8): 2560-2573.

De Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ. 2014. Redefining Ceratocystis and allied genera. Stud Mycol. 79: 187-219.

De Beer ZW, Seifert KA, Wingfield MJ. 2013. The ophiostomatoid fungi: their dual position in the Sordario mycetes. In: Ophiostomatoid fungi: Expanding frontiers. Seifert KA, De Beer ZW, Wingfield MJ. (Eds.), CBS-KNAW Fungal Biodiversity Centre, Utrecht (The Netherlands). Biodiversity Series 12: 1-19.

De Beer ZW, Seifert KA, Wingfield, MJ. 2013 b. a nomenclator for ophiostomatoid genera and species in the Ophiostomatales and Microascales. In: Ophiostomatoid fungi: Expanding frontiers. (Eds.) Seifert KA, DeBeer ZW, Wingfield MJ, CBS-KNAW Fungal Biodiversity Centre, Utrecht (The Netherlands). Biodiversity Series 12: 245-322.

De Beer ZW, Wingfield MJ. 2013. Emerging lineages in the Ophiostomatales. In: Ophiostomatoid fungi: Expanding Frontiers. Seifert KA, de Beer ZW, Wingfield MJ. (Eds.), CBS-KNAW Fungal Biodiversity Centre, Utrecht (The Netherlands). Biodiversity Series 12: 21-46.

Derbyshire V, Kowalski JC, Dansereau JT, Hauer CR, Belfort, M. 1997. Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. J Mol Biol 265(5): 494-506.

Deredec A, Burt A, Godfray HC. 2008. The population genetics of using homing endonuclease genes in vector and pest management. Genetics 179(4): 2013-2026.

Deredec A, Godfray HC, Burt A. 2011. Requirements for effective malaria control with homing endonuclease genes. Proc Natl Acad Sci USA 108(43): 874-880.

Dujon B. 1980.Sequence of the intron and flanking exons of the mitochondrial 21SrRNA gene of yeast strains having different alleles at the omega and rib-1 loci. Cell 20(1): 185-197.

Dujon B. 1981. Mitochonderial genetics and functions. In: Strathern J, Jones E, Broach J (ed), Molecular Biology of the Yeast *Saccharomyces*: life cycle and inheritance. Cold Spring Harbor Laboratory Press, pp. 505-635

DujonB. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculation. Gene 82: 91-114.

Dujon B, Belcour L. 1989.Mitochonderial DNA instabilities and rearrangements in yeast and fungi. In: Douglas B, Howe M. Mobile DNA. Academic press, New York, pp. 861-878.

Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, Perlman PS, Vogt VM. 1989. Mobile introns: definition of terms and recommended nomenclature. Gene82(1): 115-118.

Dürrenberger F, Rochaix JD. 1993. Characterization of the cleavage site and the recognition sequence of the I-CreI DNA endonuclease encoded by the chloroplast ribosomal intron of *Chlamydomonas reinhardtii*. Mol Gen Genet 236(2-3): 409-414.

Edgell DR, Chalamcharla VR, Belfort M. 2011.Learning to live together: mutualism between self-splicing introns and their hosts. BMC Biol 9:22-31.

Eldarove MA, Mardanove AV, Beletsky AV, Ravin NV, Skryabin KG.2011. Complete sequence and analysis of the mitochondrial genome of the methylotrophic yeast *Hansenula polymorpha* DL-1. FEMS Yeast Res 11(6):464-472

Emelyanov VV. 2001. Evolutionary relationship of Rickettsiae and mitochondria. FEBS Lett 501(1):11-18.

Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM.2014. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. Mob DNA 5(1):2-21.

Fedorova O, Pyle AM.2008. A conserved element that stabilizes the group II intron active site. RNA 14: 1048-1056.

Fedorova O, Solem A, Pyle AM.2010. Protein-facilitated folding of group II intron ribozymes. J Mol Biol 397(3):799-813.

Fedorova O, Zingler N.2007. Group II introns: structure, folding and splicing mechanism. Biol Chem 388: 665-678.

Ferandon C, Moukha S, Callac P, Benedetto JP, Castroviejo M, Barrosso G. 2010. The Agaricusbisporus *cox1* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. Plos One 5(11): e14048.

Fonfara I, Curth U, Pingoud A, Wende W. 2012. Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. Nucleic Acids Res 40(2):847-860.

Freel KC, Friedrich A, Schacherer J.2015. Mitochondrial genome evolution in yeasts: an all-encompassing view. FEMS Yeast Res. 15(4):1-9

Fukami H, Chen CA, Chiou CY, Knowlton N.2007. Novel group I introns encoding a putative homing endonuclease in the mitochondrial *cox1* gene of *Scleractinian corals*. J Mol Evol 64:591–600.

Furniss MM, Solheim H, Christiansen E.1990. Transmission of Blue-Stain Fungi by Ips typographus (Coleoptera: Scolytidae) in Norway Spruce. Ann Entomol Soc Am J 83(4): 712-716.

Galburt EA, Jurica MS. 2005.The His-Cys box homing endonuclease family. In: Belfort M, Derbyshire V, Stoddard B, Wood D (ed.) Homing Endonucleases and Inteins. Berlin: Springer-Verlag, pp. 85-102

Gao H, Smith J, Yang M, Jones S, Djukanovic V, Nicholson MG, West A, Bidney D, Falco SC, Jantz D, Lyznik LA. 2010. Heritable targeted mutagenesis in maize using a designed endonuclease. Plant J 61(1): 176-187.

Gibb EA, Edgell DR. 2010. Better late than early: delayed translation of intron-encoded endonuclease I-TevI is required for efficient splicing of its host group I intron. Mol Microbiol 78(1):35-46.

Gibb EA, Hausner G. 2005.Optional mitochondrial introns and evidence for a homing-endonuclease gene in the mtDNA *rnl* gene in *Ophiostoma ulmi* s. lat. Mycol Res 109 (10):1112-1126

Gilbert W. 1978. Why genes in pieces? Nature 271:501.

Gimble FS. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiol Lett 185: 99-107.

Gimble FS. 2005. Engineering homing endonucleases for genomic applications. In: Belfort M, Derbyshire V, Stoddard BL, Wood DL (eds), Homing endonucleases and inteins. New York, NY, Springer pp. 177-192

Gimble FS. 2007. Engineering home endonucleases to modify complex genomes. Gene Ther Regul 3: 33-50.

Gimble FS, Thorner J. 1992. Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. Nature 357(6376): 301-306.

Goddard MR, Burt A. 1999.Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci USA 96(24): 13880-13885.

Gray MW, Burger G, Lang BF. 2001. The origin and early evolution of mitochondria. Genome Biol 2(1018): 1-5.

Griffiths AJF. 1992. Fungal senescence. Annu Rev Genet 26:351-372.

Guo H, Zimmerly S, Perlman PS, Lambowitz AM.1997. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. EMBO 16(22): 6835-6848.

Gubb D.1986. Intron-delay and the precision of expression of homoeotic gene products in *Drosophila*. Dev Genet 7(3): 119-131.

Guha TK, Hausner G. 2014. A homing endonuclease with a switch: characterization of a twintron encoded homing endonuclease. Fungal Genet Biol 65:57-68.

Guha KT, Hausner G. 2016. Using group II introns for attenuating the in vitro and in vivo expression of a homing endonuclease. PLoS one 11(2):e0150097

Hafez M, Guha T, Hausner G. 2014. I-OmiI and I-OmiII: two intron-encoded homing endonucleases within the *Ophiostoma minusrns* gene. Fungal Biol 118(8):721-31.

Hafez M, Hausner G.2011a.The highly variable mitochondrial small subunit ribosomal RNA gene of *Ophiostoma minus*. Fungal Biol 115: 1122-1137.

Hafez M, Hausner G.2011b. Characterization of the *O.ulmi* S952 intron: a potential molecular marker to distinguish between *Ophiostoma ulmi* and *Ophiostoma novo-ulmi* subsp. *americana*. World Acad Sci Eng Technol 59:1767-1775.

Hafez M, Hausner G. 2012. Homing endonucleases: DNA scissors on a mission. Genome 55(8):553-569.

Hafez, M, Hausner G. 2013. Intron homing. In: SM Hughes, ed. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press, 2013:127-32

Hafez M, Hausner G. 2015. Convergent evolution of twintron-like configurations: One is never enough. RNA Biol 12(12): 1275-1288.

Hafez M, Majer A, Sethuraman J, Rudski SM, Michel F, Hausner G. 2013. The mtDNA *rns* gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. Fungal Genet Biol 53:71-83.

Harrington TC, Steimel J, Kile G. 1998. Genetic variation in three *Ceratocystis* species with outcrossing, selfing and asexual reproductive strategies. Eur J Forest Pathol 28: 217-226.

Haugen P, Bhattacharya D. 2004. The spread of LAGLIDADG homing endonuclease genes in rDNA. Nucleic Acids Res 32: 2049-2057.

Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. Trends Genet 21(2):111-119.

Heugen P; Reeb V; Lutzoni F; Bhattacharya D. 2004. The evolution of homing endonuclease genes and group I introns in nuclear rDNA 21(1): 129-140.

Hausner G, Hafez M, Edgell DR.2014. Bacterial group I introns: mobile RNA catalysts. Mobile DNA 5:8-20.

Hausner G. 2003. Fungal mitochondrial genomes, plasmids and introns. In: Arora DK, Khachatourians GG (eds). Applied mycology and biotechnology: Fungal genomics. New York. Elsevier Science. pp. 101-131

Hausner G. 2012. Introns, mobile elements and plasmids. In: Bullerwell CE (ed) Organelle Genetics: Evolution of Organelle Genomes and Gene Expression. Springer Verlag, Berlin, pp. 329-358.

Hausner G, Eyjólfsdóttir GG, Reid J. 2003. Three new species of *Ophiostoma* and notes on *Cornuvesica falcata*. Can J Bot 81:40-48.

Hausner G, Hafez M, Edgel DR. 2014. Bacterial group I introns: mobile RNA catalysts. Mob DNA 5(1):8-20.

Hausner G, Reid J.2003. Notes on *Ceratocystis brunnea* and *Ophiostoma* based on partial ribosomal DNA sequence data. Can J Bot 81:865-876.

Hausner G, Reid J, Klassen GR.1992. Do galeate-ascospore members of the Cephaloascaceae, Endomycetaceae and Ophiostomataceae share a common phylogeny? Mycologia 84:870-881.

Hausner G, Reid J, Klassen GR. 1993.On the phylogeny of *Ophiostoma*, *Ceratocystis*s, *Microascus*, and relationships within *Ophiostoma* based on partial ribosomal DNA sequences. Can J Botany 71: 1249-1265.

Hausner G, Wang X. 2005. Unusual compact rDNA gene arrangements within some members of the Ascomycota: evidence for molecular co-evolution between ITS1 and ITS2. Genome 48(4):648-660.

Henzell RP, Cooke BD, Mutze GJ. 2008. The future of biological control of pest populations of European rabbits *Oryctolagus cuniculus*. Wildlife Res 35: 633-650.

Hintz WE, Jeng RS, Yang QD, Hubbes PA, Horgen PA. 1993. A genetic survey of the pathogenic fungus *Ophiostoma ulmi* across a Dutch elm disease front in western Canada. Genome 36: 418-426.

Huchon D, Szitenberg A, Shefer S, Ilan M, Feldstein T.2015. Mitochondrial group Iand group II introns in the sponge orders Agelasida and Axinellida. BMC Evol Biol 15:278-292.

Jacobs K, Wingfield MJ. 2001. *Leptographium* species: Tree pathogens, insect associates, and agents of blue stain. APS Press, St. Paul., USA.

Jacoby K, Lambert AR, Scharenberg AM. 2016. Characterization of homing endonuclease binding and cleavage specificities using yeast surface display SELEX (YSD-SELEX). Nucleic Acids Res, doi: 10.1093/nar/gkw864.

Jacoby K; Metzger M; Shen BW; Certo MT; Jarjour J; Stoddard B; Scharenberg AM. 2012. Expanding LAGLIDADG endonuclease scaffold diversity by rapidly surveying evolutionary sequence space. Nucleic Acids Res 40(11): 4954–4964

Jacquier A.1990. Self-splicing group II and nuclear pre-mRNA introns: how similar are they? Cell 15(9): 351-354.

Jacquier A, Dujon B.1985.An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell 41(2): 383-394.

Johansen JS, Haugen P.2001. A new nomenclature of group I introns in ribosomal DNA. RNA 7:935-936

Jurica MS, Monnat RJ Jr, Stoddard BL. 1998. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. Mol Cell 2(4):469-476.

Jurica MS, Stoddard BL. 1999. Homing endonucleases: structure, function and evolution. *Cell Mol Life Sci* 55(10): 1304-1326.

Kamikawa R, Shiratori T, Ishida K, Miyashita H, Roger AJ. 2016. Group II intron-mediated trans-splicing in the gene-rich mitochondrial genome of an enigmatic eukaryote, diphylleiarotans. Genome Biol Evol 8(2):458-466.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772-780.

Kavanagh K.2011. Fungi: Biology and Applications. Wiley-Blackwell, UK, Keeble AH, Maté MJ, Kleanthous C. 2005. HNH endonucleases. In: Belfort M, Derbyshire V, Stoddard BL, Wood DL (eds). Homing endonucleases and inteins. New York, NY, Springer, pp. 49-66.

Kendrick, B. 2001. The fifth Kingdom 3rd Ed. Focus Publishing/R. Pullins Co. Newburyport, MA, USA.

Kim YK, Kim VN. 2007. Processing of intronic microRNAs. EMBO J 26(3): 775-783.

Komori K, Fujita N, Ichiyanagi K, Shinagawa H, Morikawa K, Ishino Y.1999. PI-PfuI and PI-PfuII, intein-coded homing endonucleases from *Pyrococcus furiosus*. I. Purification and identification of the homing-type endonuclease activities. Nucleic Acids Res 27(21):4167-4174.

Koonin EV. 2006. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct 1:22-45.

Koonin EV, Csuros M, Rogozin IB. 2013. Whence genes in pieces: reconstruction of the exon-intron gene structures of the last eukaryotic common ancestor and other ancestral eukaryotes. Wiley Interdiscip Rev RNA 4(1):93-105.

Koonin EV, Senkevich TG, Dolja VV.2006. The ancient virus world and evolution of cells. Biol Direct 1:29-56.

Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM, Derbyshire V. 1999. Configuration of the catalytic GIY-YIG domain of intron

endonuclease I-TevI: coincidence of computational and molecular findings. Nucleic Acids Res 27(10): 2115-2125.

Kowalski JC, Derbyshire V.2002. Characterization of homing endonucleases. Methods 28(3): 365-373.

Lambert AR, Hallinan JP, Shen BW, Chik JK, Bolduc JM, Kulshina N, Robins LI, Kaiser BK, Jarjour J, Havens K, Scharenberg AM, Stoddard BL. 2016. Indirect DNA sequence recognition and its impact on nuclease cleavage activity. Structure 24(6): 862-873.

Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. Annu Rev Biochem 62: 587-622.

Lambowitz AM, Belfort M. 2015. Mobile bacterial group II introns at the crux of eukaryotic evolution. Microboil Spectr 3(1): MDNA3-0050-2014. doi: 10.1128/microbiolspec. MDNA3-0050-2014.

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS. 1999. Group I and group II ribozymes as RNPs: Clues to the past and guides to the future. In: Gestel RF, Cech TR, Atkins JF (eds), The RNA World. Cold Spring Harbor Laboratory Press, NY, pp. 451-485

Lambowitz AM, ZimmerlyS. 2004. Mobile group II introns. Annu Rev Genet 38: 1-35.

Lambowitz AM, Zimmerly S. 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol 3(8): a003616.

Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. Trends Genet 23(3):119-25.

Lauderdale JD, Stein A. 1992. Introns of the chicken ovalbumin gene promote nucleosome alignment in vitro. Nucleic Acids Res 20(24): 6589-6596.

Lehmann K, Schmidt U. 2003. Group II introns: structural and catalytic versatility of large natural ribozymes. Crit Rev Bioch Mol Biol 38: 249-303.

Li CF, Costa M, Bassi G, Lai YK, Michel F. 2011. Recurrent insertion of 5'-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs. RNA 17(7): 1321-1335.

Li J, Zhang J, Chen H, Chen X, Lan J, Liu C. 2013. Complete mitochondrial genome of the medicinal mushroom *Ganoderma lucidum*. PLoS One 8(8): e72038.

Li Y, Hu XD, Yang RH, Hsiang T, Wang K, Liang DQ, Liang F, Cao DM, Zhou F, Wen G, Yao YJ. 2015. Complete mitochondrial genome of the medicinal fungus *Ophiocordyceps sinensis*. Sci Rep 5: 13892.

Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, Fujii CY, Mason T, Bowman CL, Barnstead M. 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. Nature 402:761-768.

Liu K, Sandgren EP, Palmiter RD, Stein A. 1995. Rat growth hormone gene introns stimulate nucleosome alignment in vitro and in transgenic mice. Proc Natl Acad Sci USA 92(17): 7724-7728.

Losada L; Pakala SB; Fedorova ND; Joardar V; Shabalina SA; Hostetler J; Pakala SM; Zafar N; Thomas E; Rodriguez-Carres M; Dean R; Vilgalys R; Nierman WC; Cubeta MA. 2014. Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. FEMS Microbiol Lett 352: 165–173

Lucas P, Otis C, Mercier JP, Turmel M, Lemieux C. 2001. Rapid evolution of the DNA-binding site in LAGLIDAG homing endonuclease. Nucleic Acids Res 29: 960-969.

Malloch D, Blackwell M.1993. Dispersal biology of the Ophiostomatoid fungi. In: Wingfield MJ, Seifert KA, Webber JF (eds.) *Ceratocystis* and *Ophiostoma*. Taxonomy, Ecology, and Pathogenicity. APS St. Paul, MN, pp.195-206

Marcaida MJ, Muñoz IG, Blanco FJ, Prieto J, Montoya G. 2010. Homing endonucleases: from basics to therapeutic applications. Cell Mol Life Sci 67: 727-748.

Macreadie MJ; Scott RM; Zinn AR; Butow RA. 1985. Transportation of an intron in yeast mitochondria requires a protein encoded by that intron. Cell 41: 395-402.

Marcia M, Somarowthu S, Pyle AM. 2013. Now on display: a gallery of group II intron structures at different stages of catalysis. Mob DNA 4(1):14-27.

Mardanov AV, Beletsky AV, Kadnikov VV, Ignatov AN, Ravin NV. 2014. The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. PLoS One 9(9):e107536.

McNeil BA, Semper C, Zimmerly S. 2016.Group II introns: versatile ribozymes and retroelements. RNA 7(3): 341-355.

Michel F, Dujon B. 1986. Genetic exchanges between bacteriophage T4 and filamentous fungi? Cell 46(3): 323-330.

Michel F, Ferat JL. 1995. Structure and activities of group II introns. Annu Rev Biochem 64: 435-461.

Michel F, Jacquier A, Dujon B. 1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochem 64: 867-881.

Michel F, Jaeger L, Westhof E, Kuras R, Tihy F, Xu MQ, Shub DA. 1992. Activation of the catalytic core of a group I intron by a remote 3' splice junction. Genes Dev 6: 1373-1385.

Michel F; Lang BF. 1985. Mitochondrial class II introns encode proteins related to the reveres transcriptases of retroviruses. Nature 316: 641-643.

Michel F, Umesono K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns. Gene 82: 5-30.

Michel F, Westhof E.1990. Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216: 585-610.

Monteiro-Vitorello CB, Hausner G, Searles DB, Gibb EA, Fulbright DW, Bertrand H. 2009. The *Cryphonectria parasitica* mitochondrial *rns* gene: plasmid-like elements, introns and homing endonucleases. Fungal Genet Biol 46(11):837-848.

Moure CM, Gimble FS, Quiocho FA. 2002. Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Nature Struct Biol 9(10): 764-770.

Müller S, Appel B, Balke D, Hieronymus R., Nübel C. 2016. Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? F1000 Res. 5

Mueller JE, Clyman J, Huang YJ, Parker MM, Belfort M. 1996 a. Intron mobility in phage T4 occurs in the context of recombination-dependent DNA replication by way of multiple pathways. Genes Dev 10 (3): 351-364.

Mueller JE, Smith D, Belfort M. 1996 b. Exon coconversion biases accompanying intron homing: battle of the nucleases. Genes Dev 10(17): 2158 -2166.

Mullineux ST, Costa M, Bassi GS, Michel F, Hausner G. 2010.A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. RNA 16: 1818-1831.

Monteiro-Vitorello CB, Hausner G,Searles DB, Gibb EA, Fulbright DW, Bertrand H. 2009. The *Cryphonectria parasitica* mitochondrial *rns* gene: Plasmid-like elements, introns and homing endonucleases. Fungal Genet Biol 46: 837-848.

Muñoz NM, Beard BC, Ryu BY, Luche RM, Trobridge GD, Rawlings DJ, Scharenberg AM, Kiem HP. 2012. Novel reporter systems for facile evaluation of I-SceI-mediated genome editing. Nucleic Acids Res 40(2): e14.

Newbanks D, Bosch A, Zimmermann MH. 1983. Evidence for xylem dysfunction by embolization in Dutch elm disease. Phytopathol 73:1060-1063.

Nicholas KB, Nicholas HB, Deerfield DW. 1997. GeneDoc: analysis and visualization of genetic variation. EMB News 4: 14.

Okamura K, Lai EC. 2008. Endogenous small interfering RNAs in animals. Nat Rev Mol Cell Biol 9(9): 673-678.

Ouellette GB, Rioux D, Simard M, Cherif M. 2004. Ultrastructural and cytochemical studies of host and pathogens in some fungal wilt diseases: retro- and introspection towards a better understanding of DED. Invest Agrar Sist Recur For 13:119-145.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. 2000. Dynamic Evolution of Plant Mitochondrial Genomes: Mobile Genes and Introns and Highly Variable Mutation Rates. Proc Natl Acad Sci USA 79(13): 6960-6966.

Palmer JD, Logsdon JM Jr. 1991.The recent origins of introns. Curr Opin Genetics Dev 1: 470-477.Palmiter RD, Sandqren EP, Avarbock MR, Allen DD, Brinster RL. 1991. Heterologous introns can enhance expression of transgenes in mice.Proc Natl AcadSci USA 88(2):478-482.

Paoletti M, Buck KW, Brasier CM. 2005. Cloning and sequence analysis of the MAT-B (MAT-2) genes from the three Dutch elm disease pathogens, *Ophiostoma ulmi*, *O. novo-ulmi*, and *O. himal-ulmi*. Mycol Res 109(9):983-991.

Paoletti M, Buck KW, Brasier CM. 2006.Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote *Ophiostoma novo-ulmi*. Mol Ecol 15(1): 249-262.

Pâques F, Duchateau P. 2007. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr GeneTher 7(1): 49-66.

Paquin B, O'Kelly CJ, Lang BF.1995. Intron-encoded open reading frame of the GIY-YIG subclass in a plastid gene. Curr Genetics 28(1): 97-99.

Parker MM, Belisle M, Belfort M. 1999. Intron homing with limited exon homology. Illegitimate double-strand-break repair in intron acquisition by phage T4. Genetics 153(4): 1513-1523.

Pessach IM, Notarangelo LD. 2011. Gene therapy for primary immune deficiencies: looking ahead, toward gene correction. J Allergy Clin Immunol 127(6): 1344-1350.

Pingoud A, Silva GH. 2007. Precision genome surgery. Nature Biotechnol 25(7):743-744.

Plichta R, Urban J, Gebauer R, Dvořák M, DŭrkovicĴG. 2016. Long-term impact of *Ophiostoma novo-ulmi* on leaf traits and transpiration of branches in the Dutch elm hybrid ‘Dodoens’. Tree Physiol 36(3): 335-344.

Prieto, J., Molina, R., and Montoya, G. 2012. Molecular scissors for in situ cellular repair. Crit. Rev. Biochem. Mol. Biol 47(3): 207– 221.

Procházka E, Poláková S, Piskur J, Sulo P. 2010. Mitochondrial genome from the facultative anaerobe and petite-positive yeast *Dekkera bruxellensis* contains the NADH dehydrogenase subunit genes. FEMS Yeast Res 10(5): 545-557.

Pyle AM. 2010. The tertiary structure of group II introns: implications for biological function and evolution. Crit Rev Biochem Mol Biol 45: 215-232.

Pyle AM, Lambowitz AM. 2006. Group II introns: ribozymes that splice RNA and invade DNA. In: Gesteland R, Cech TR, Atkins JF (eds) The RNA World, Cold Spring Harbor Press, pp. 469-505

Qin PZ, Pyle AM. 1998. The architectural organization and mechanistic function of group II intron structural elements. Curr Opin Struct Biol 8:301-308.

Qin Y, Yao J, Wu DC, Nottingham RM, Mohr S, Hunicke-Smith S, Lambowitz AM.2016. High-throughput sequencing of human plasma RNA by using thermostable group IIintron reverse transcriptases. RNA 22(1):111-128.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154:1380–1389.

Rearick D, Prakash A, McSweeny A, Shepard SS, Fedorova L, Fedorov A. 2011.Critical association of ncRNA with introns. Nucleic Acids Res 39(6): 2357-2366.

Reid J, Hausner G.2015. A new *Graphilbum* species from western hemlock (*Tsugaheterophylla*) in Canada. Mycotaxon 130: 399-419.

Robinson RC. 1962. Blue stain fungi in lodgepol pine (*Pinus contorta*Dougl. Var. *Latifolia engelm*. Infested by the mountain pine beetle. Can J Bot 40(4): 609-614.

Rogozin IB, Carmel L, Csuros M, Koonin EV. 2012. Origin and evolution of spliceosomal introns. Biol Direct 7:11-39.

Roman J, Rubin MN, Woodson SA. 1999. Sequence specificity of in vivo reverse splicing of the *Tetrahymena* group I intron. RNA5(1):1-13.

Roman J, Woodson SA. 1995. Reverse splicing of the *Tetrahymena* IVS: Evidence for multiple reaction sites in the 23S rRNA. RNA 1: 478-490.

Romano IGS, Paul B, Sather BD, Younan PM, Sommer K, Kowalski JP, Hale M, Stoddard B, Jarjour J, Astrakhan A, Kiem HP, Rawlings DJ. 2016. Efficient modification of the CCR5 locus in primary human T cells with mega TAL nuclease establishes HIV-1 eesistance. Mol Ther Nucleic Acids 5(8):e352.

Rot C, Goldfarb I, Ilan M, Huchon D.2006. Putative cross-kingdom horizontal gene transfer in sponge (*Porifera*) mitochondria. BMC Evol Biol 6:71-82.

Saldanha R, Mohr G, Belfort M, Lambowitz AM. 1993. Group I and group II introns. FASEB J 7: 15-24.

Sambrook J, Fritsch E F, Maniatis T. 1989. Molecular cloning: A laboratory manual. Cold Spring Harbor, New York pp

Santamaria M, Vicario S, Pappadà G, Scioscia G, Scazzocchio C, Saccone C. 2009. Towards barcode markers in Fungi: an intron map of Ascomycota mitochondria. BMC Bioinformatics 10(Suppl 6):S15.

Saguez C, Lecellier G, Koll F. 2000. Intronic GIY-YIG endonuclease gene in the mitochondrial genome of *Podospora curvicolla*: evidence for mobility. Nucleic Acids Res 28(6): 1299-1306.

Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G. 2009. Genes within genes: Multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an *rnl* group I Intron of *Ophiostoma* and related taxa. Mol Biol Evol 26(10): 2299-2315.

Sethuraman J, Okoli CV, Majer A, Corkery TL, Hausner G. 2008. The sporadic occurrence of a group I intron-like element in the mtDNA rnl gene of *Ophiostoma novo-ulmi subsp. Americana.* Mycol Res 112(5):564-82.

Sethuraman J, Rudski SM, Wosnitza K, Hafez M, Guppy B, Hausner G. 2013. Evolutionary dynamics of introns and their open reading frames in the U7 region of the mitochondrial *rnl* gene in species of *Ceratocystis*. Fung Biology 117(11-12): 791-806.

Severin PM, Zou X, Schulten K, Gaub HE. 2013. Effects of cytosine hydroxymethylation on DNA strand separation. Biophysical1 04(1):208-215.

Scalley-Kim M, McConnell-Smith A, Stoddard BL. 2007. Coevolution of a homing endonuclease and its host target sequence. J Mol Biol 372(5):1305-1319.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 2012. Fungal Barcoding Consortium; Fungal Barcoding Consortium Author List. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 109(16):6241-6246.

Schroeder S, Kim SH, Cheung WT, Sterflinger K, Breuil C. 2001. Phylogenetic relationship of *Ophiostoma piliferum* to other sapstain fungi based on the nuclear rRNA gene. FEMS Microbiol Lett 195(2):163-167.

Sharp PA. 1991. Five easy pieces. Science 254(5032): 663.

Siegl T, Petzke L, Welle E, Luzhetskyy A. 2010. I-SceI endonuclease: a new tool for DNA repairs studies and genetic manipulations in *Streptomyces*. Appl Microbiol Biotechnol 87(4): 1525-1532.

Silva GH, Belfort M. 2004. Analysis of the LAGLIDADG interface of the monomeric homing endonuclease I-DmoI. Nucleic Acids Res 32(10):3156-3168.

Silva GH, Belfort M, Wende W, Pingoud A. 2006. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. J Mol Biol 361(4): 744-754.

Simon DM, Clarke NAC, McNeil BA, Johnson I, Pantuso D, Dai L, Chai D, Zimmerly S. 2008.Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA 14:1704-1713.

Skowronek KJ, Bujnicki JM. 2007. Restriction and homing endonucleases. In: Polaina J, MacCabe AP. Industrial enzymes. Structure, function and applications. Springer-Verlag, Berlin, pp. 357-378.

Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M. 2005. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. Genes Dev 19:2477-2487.

Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D .2000. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res 28:3361–3369.

Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Pâques F, Duchateau P. 2006. A combinatorial approach to create

artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res 34(22): e149.

Solieri L. 2010. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. Trends Microbiol 18(11):521-530.

Solheim H, Safranyik L. 1997. Pathogenicity of the spruce beetle associated blue-stain fungi, *Ceratocystis rufipenni* and *Leptographium abietinum* to Sitka spruce. Can J For Res 27:1336-1341.

Spatafora JW, Blackwell M. 1994. Polyphyletic origins of ophiostomatoid fungi. Mycol Res 98: 1-9.

Stahley MR, Strobel SA. 2005. Structural evidence for a two-metal-ion mechanism of group I intron splicing. Science 309(5740):1587-1590.

Stoddard BL. 2005. Homing endonuclease structure and function. Q Rev Biophys 38(1): 49-95.

Stoddard BL. 2011. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. Structure 19(1): 7-15.

Stoddard BL. 2014. Homing endonucleases from mobile group I introns: discovery to genome engineering. Mob DNA 5(1):7-23.

Stoddard BL, Scharenberg AM, Monnat RJ. 2008. Advances in engineering homing endonucleases for gene targeting: Ten years after structures. In: Bertolotti R, Ozawa K, Hackensack, NJ (eds). Progress in Gene Therapy 3: Autologous and Cancer Stem Cell Gene Therapy. World Scientific Press. pp. 135-167

Storici F, Resnick MA. 2003. *Delitto perfetto* targeted mutagenesis in yeast with oligonucleotides. Genet Eng 25: 189-207.

Shub DA, Gott JM, Xu MQ, Lang BF, Michel F, Tomaschewski J, Pedersen-Lane J, Belfort M.

1988. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proc Natl Acad Sci USA 85:1151-1155.

Taylor G; Stoddard B. 2012. Structural, functional and evolutionary relationships between homing endonucleases and proteins from their host organisms. Nucleic Acids Res 40 (12): 5189–5200.

Suh S-Q, Jones KG, Blackwell M. 1999. A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetes fungus: Evidence for a new major class of group I introns. J Mol Evol48: 493-500.

Sullenger BA, Gilboa E. 2002. Emerging clinical applications of RNA. Nature 418:252-258

Svaldi R, Elgersma DM. 1982. Further studies on the activity of cell wall degrading enzymes of aggressive and non-aggressive isolates of *Ophiostoma ulmi*. Eur J Pathol 12:29-36.

Szczepanek T, Lazowska J. 1996. Replacement of two non-adjacent amino acids in the *S. cerevisiae* bi2 intron-encoded RNA maturase is sufficient to gain a homing-endonuclease activity. EMBO 15(14): 3758-3767.

Szostak, JW; Orr-Weaver TL; Rothstein, RJ; Stahl FW. 1983. The double strand break repair model for recombination. Cell 33: 25-35.

Takai S, Richards WC. 1978. Cerato-ulmin, a wilting toxin of *Ceratocystis ulmi*: isolation and some properties of cerato-ulmin from the culture of *C. ulmi*. J Phytopathol 91:129-146.

Takeuchi R, Choi M, Stoddard BL. 2014. Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. Proc Natl Acad Sci USA 111(11):4061-4066.

Takeuchi R, Choi M, Stoddard BL. 2015. Engineering of customized meganucleases via in vitro compartmentalization and in cellulo optimization. Methods Mol Biol 1239:105-132.

Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, Gloor GB, Scharenberg AM, Edgell DR, Stoddard BL. 2011. Tapping natural reservoirs of homing endonucleases for targeted gene modification. Proc Natl Acad Sci USA108(32): 13077-13082.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular 574 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and 575 maximum parsimony methods. Mol Biol Evol 28(10):2731-2739.

Taylor GK, Heiter DF, Pietrokovski S, Stoddard BL. 2011. Activity, specificity and structure of I-Bth0305I: a representative of a new homing endonuclease family. Nucleic Acids Res 39(22):9705-9719.

Taylor GK, Petrucci LH, Lambert AR, Baxter SK, Jarjour J, Stoddard BL. 2012. LAHEDES: the LAGLIDADG homing endonuclease database and engineering server. Nucleic Acids Res 40: 110-116.

Tian GL, Michel F, Macadre C, Slonimski PP, Lazowska J. 1991. Incipient mitochondrial evolution in yeasts. II. The complete sequence of the gene coding for cytochrome b in *Saccharomyces douglasii* reveals the presence of both new and conserved introns and discloses major differences in the fixation of mutations in evolution. J Mol Biol 218(4): 747-760.

Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R. 2009. Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol 16(9) 996-1001.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG.1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25(24):4876-4882.

Thompson AJ, Yuan X, Kudlicki W, Herrin DL. 1992. Cleavage and recognition pattern of a double-strand-specific endonuclease (I-CreI) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. Gene 119(2):247-51.

Thwaites J M, Farrell Rl, Duncan SM, Reay Sd, Blanchette RA, Hadar E, Hadar Y, Harrington TC, McNew D. 2005. Survey of potential sapstain fungi on *Pinusradiata* in New Zealand. N Z J Bot 43(3):653-663.

Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. 2011. Evolution of introns in the archeal world. Proc Natl Acad Sci USA 108: 4782-4787.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA 7:1142-1152.

Toor N, Keating A, Fedorova O, Rajashankar K, Wang J, Pyle AM. 2010. The tertiary architecture of a group II intron. RNA 16:57-69.

Toor N, Keating KS, Pyle AM. 2009. Structural insights into RNA splicing. Curr Opin Struct Biol 19: 260-266.

Toor N, Zimmerly S. 2002. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA 8: 1373-1377.

Toro N, Nisa-Martínez R. 2014. Comprehensive phylogenetic analysis of bacterial reverse transcriptases. PLoS One 9(11): e114083.

Urban J, Dvořák M. 2014. Sap flow-based quantitative indication of progression of Dutch elm disease after inoculation with *Ophiostoma novo-ulmi*. Trees 28:1599-1605.

Upadhyay HP. 1981. A monograph of C*eratocystis* and *Ceratocystiopsis*. University of Georgia Press: Athens, G.A.

Vainstein A, Marton I, Zuker A, Danziger M, Tzfira T. 2011. Permanent genome modifications in plant cells by transient viral vectors. Trends Biotechnol 29(8): 363-369.

Vallès Y, Halanych KM, Boore JL. 2008. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS One 3(1):0001488.

Van Alfen NK, Turner NC. 1975. Influence of a *Ceratocystis ulmi* toxin on water relations of elm (*Ulmus americana*). Plant Physiol 55:312-316.

Van der Veen R, Arnberg AC, van der Horst G, Bonen L, Tabak HF, Grivell LA. 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *In vitro*. Cell 44: 225-234.

Vogel J, Börner T. 2002. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. Nat Struct Mol Biol 21: 3794-380.

Wallace DC. 2010. Mitochondrial DNA mutations in disease and aging. Environ Mol Mutagen 51(5):440-450.

Wank H, Sanfilippo J, Singh RN, Matsuura M, Lambowitz AM. 1999. A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. Mol Cell 4: 239-250.

Webster J, Webster RWS. 2007. Introduction to fungi 3rd Edition. Cambridge University Press, New York, USA.

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat R Jr, Burt A, Crisanti A. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature 473(7346): 212-215.

Windbichler N, Papathanos PA, Crisanti A. 2008. Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae*. PLoS Genetics 4(12): e1000291.

Wingfield MJ, Seifert KA, Webber JF. 1993. *Ceratocystis* and *Ophiostoma*: taxonomy, ecology and pathogenicity. American Phytopathological Society Press, St. Paul, MN.

Whelan S, Goldman NA. 2001. General empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691-699.

Woese CR; Kandler A; Wheelis ML. 1990. Towards a natural system of organisms: proposal foe domains archaea, bacteria, and eucarya. Proc Natl Acad Sci USA 87: 4576-4579.

Wolfs JM, DaSilva M, Meister SE, Wang X, Schild-Poulter C, Edgell DR. 2014. MegaTevs: single-chain dual nucleases for efficient gene disruption. Nucleic Acids Res 42(13):8816-8829.

Wolf K, Del Giudice L. 1988. The vaiable mitochonderial genome of ascomycetes: organization, mutational alterations, and expression. Adv Genet 25:185-308.

Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. BMC Genomics 16:451

Wu B, Hao W. 2014. Horizontal transfer and gene conversion as an important drivingforce in shaping the landscape of mitochondrial introns. G3 (Bethesda)4(4):605-612.

Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon theirreverse transcriptase sequences. EMBO 9(10):3353-3362.

Yang M, Djukanovic V, Stagg J, Lenderts B, Bidney D, Falco SC, Lyznik LA. 2009. Targeted mutagenesis in the progeny of maize transgenic plants. Plant Mol Biol 70(6): 669-679.

Zeevi V, Liang Z, Arieli U, Tzfira T. 2012. Zinc finger nuclease and homing endonuclease-mediated assembly of multigene plant transformation vectors. Plant Physiol 158(1): 132-144.

Zeng Q, Bonocora RP, Shub DA. 2009. A free-standing homing endonuclease targets an intron insertion site in the *psbA* gene of cyanophages. Curr Biol 19(3):218-222.

Zhou Y, Lu C, Wu Q-J, Wang Y, Sun Z-T, Deng J-C, Zhang Y. 2008. GISSD: Group I intron sequence and structure database. Nucleic Acids Res 36: D31-D37.

Zimmerly S, Hausner G, Wu X. 2001. Phylogenetic relationships among group II introns ORFs. Nucleic Acids Res 29: 1238-1250.

Zimmerly S, Semper C. 2015. Evolution of group II introns. Mobile DNA 6:7.

Zimmerly S, Wu L. 2015. An unexplored diversity of reverse transcriptases in bacteria. Microbiol Spectr3 (2): MDNA3-0058-2014.

Zipfel RD, de Beer ZW, Jacobs K, Wingfield BD, Wingfield MJ. 2006. Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. Stud Mycol55: 75-97.

Zuker M. 2003. M fold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13):3406-3415.