

A SURVEY OF BLUE-STAIN FUNGI IN NORTHWESTERN ONTARIO AND  
CHARACTERIZATION OF MOBILE INTRONS IN RIBOSOMAL DNA

by

Shelly Marie Rudski

A Thesis submitted to the Faculty of Graduate Studies of  
The University of Manitoba  
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Microbiology  
University of Manitoba  
Winnipeg

Copyright © 2011 by Shelly Marie Rudski

## Abstract

This work presents a survey of blue-stain fungi found in Northwestern Ontario, characterization of a homing endonuclease gene within *Grosmannia piceiperda* and finally an examination of the introns and homing endonuclease genes found in the large ribosomal subunit gene in species of *Ceratocystis*; using molecular techniques and phylogenetic analysis, we studied the molecular evolution of these mobile genetic elements. The blue-stain fungi of Northwestern Ontario were identified based on phylogenetic analysis of rDNA internal transcribed spacer region sequences. This data was supplemented with morphological characteristics of the fungal cultures. The second project was an examination of a LAGLIDADG homing endonuclease and its IC2 group I intron. This intron is uniquely positioned within the group I intron-encoded *rps3* gene of the large subunit ribosomal RNA gene. The final chapter is an investigation of the large subunit ribosomal RNA gene in species of *Ceratocystis*. The 3' segment of this gene contains several novel introns and homing endonuclease genes. There is also much diversity between strains despite their close relation on the rDNA internal transcribed spacer region phylogenetic tree. Further, our data also suggest that the single motif LAGLIDADG homing endonuclease of the rDNA mL1923 intron is likely to be an ancestor to other homing endonucleases in the area. The results of these studies demonstrate the role that these elements play in the genetic diversity observed in the blue-stain fungi.

## **Acknowledgements**

This work is dedicated to my nieces; I hope you continue to show me the valuable things in life and the significance of “new beginnings”.

This thesis would not be possible without the support and commitment of my supervisor Dr. Georg Hausner. I would like to thank my committee members, Drs. Deb Court and Michele Piercey-Normore, for their insight. I am also grateful to my colleagues: Drs. Reid, Sethuraman and Summers for their advice and to my fellow lab mates: Mohamed and Kari, and Drs. Mullineux and Iranpour.

A special thanks to my honorary lab mate Suman and to my other friends and family for being there when it really counts.

# Table of Contents

<b>List of Tables</b>	viii
<b>List of Figures</b>	ix
<b>List of Abbreviations</b>	xi
<b>Chapter 1: Literature Review</b>	1
1.1 Ophiostomatoid Fungi	1
1.2 Fungal Mitochondria and Mitochondrial DNA	3
1.3 Mobile Genetic Elements in Fungal Mitochondrial DNA	7
1.4 Group I Introns	8
1.5 Group II Introns	10
1.6 Homing and Retrohoming of Group I and Group II Introns	10
1.7 Homing Endonucleases Can Mobilize Flanking DNA	12
1.8 The Homing Cycle	13
1.9 Families of Homing Endonuclease Genes	14
1.9.1 The LAGLIDADG Family of Homing Endonucleases	14
1.9.2 The GIY-YIG Family of Homing Endonucleases	16
1.10 Applications of Homing Endonucleases and Group II Introns	16
1.11 Research Objectives	17
<b>Chapter 2: General Materials and Methods</b>	20
2.1 Fungal Culturing	20
2.2 Fungal DNA Extraction	20

2.3	PCR Amplification and Purification	21
2.4	Cloning of PCR Products	25
2.5	DNA Sequencing	26
2.6	Fungal RNA Extraction and Purification	26
2.7	Reverse Transcriptase PCR Amplification and cDNA Amplification	28
2.8	Bacterial Glycerol Stocks and Precipitated Plasmid DNA Storage	28
2.9	RNA Folding	29
2.10	Bioinformatics	29
<b>Chapter 3: Molecular taxonomy of blue-stain fungi from NW Ontario</b>		31
3.1	INTRODUCTION	33
3.2	SUMMARY OF METHODS	36
3.2.1	Phylogenetic Analysis	36
3.3	RESULTS AND DISCUSSION	41
<b>Chapter 4: Additional complexity at the mitochondrial <i>rps3</i> locus within species of <i>Grosmannia</i></b>		54
4.1	INTRODUCTION	55
4.2	SUMMARY OF METHODS	57
4.2.1	Phylogenetic Analysis	57
4.2.2	Gene Synthesis of HEG	58
4.2.3	Protein Expression	58
4.2.4	Protein Purification	61

4.3	RESULTS	61
4.3.1	The IC2 Intron of <i>Grosmannia piceiperda</i>	61
4.3.2	The LAGLIDADG Homing Endonuclease of <i>Grosmannia piceiperda</i> and <i>G. laricis</i>	67
4.4	DISCUSSION	67
4.4.1	Overlap of the IC2 Intron's Open Reading Frame and Core Sequence Suggests Coevolution of the Sequences	67
4.4.2	Regulation of the IC2 Intron and Homing Endonuclease Gene	72
<b>Chapter 5: Evolutionary dynamics of introns and their open reading frames in a region of the mitochondrial <i>rnl</i> gene in species of <i>Ceratocystis</i></b>		75
5.1	INTRODUCTION	77
5.2	SUMMARY OF METHODS	80
5.2.1	Phylogenetic Analysis	80
5.3	RESULTS	81
5.3.1	The <i>rnl</i> -U7 Region	81
5.3.2	Strains Containing Only the mL1923 Insertion in the <i>rnl</i> -U7 Region	86
5.3.3	Strains Containing the mL1673 and the mL1923 Insertions in the <i>rnl</i> -U7 Region	91
5.3.4	Strains Containing the mL1787 and the mL1923 Insertions in the <i>rnl</i> -U7 Region	93
5.3.5	ITS Phylogeny Versus the <i>rnl</i> -U7 Intron Insertions	93
5.3.6	Another Group II Intron with a LAGLIDADG-Type ORF and Two Group I Introns in the U7 – U11 Inter-Region	94
5.3.7	<i>In Vivo</i> Splicing and Intron/Exon Junctions of the <i>rnl</i> Introns	95

5.3.8	Phylogeny of the mL1923 Intron-Encoded LAGLIDADG HEG	95
5.3.9	Phylogeny of the Double Motif LAGLIDADG ORF Encoded by the mL2059 Group II Intron	105
5.4	DISCUSSION	106
5.4.1	ITS Species Tree Versus <i>rnl</i> -U7 Configuration	109
5.4.2	The mL1923 Intron – ORF/Exon Configuration	109
5.4.3	GIY-YIG ORFs: “Parasites” of LAGLIDADG HEGs and HEG Transduction	111
5.4.4	More Group II Introns with LAGLIDADG ORFs	112
5.4.5	The mL1923 ORFs Have a Complex Phylogenetic History	113
5.5	CONCLUSIONS	117
	<b>References</b>	118

## List of Tables

<b>Table 2.1</b> Sequences of the PCR primer pairs used to amplify nuclear and mitochondrial DNA in this study.	23
<b>Table 3.1</b> Ophiostomatoid fungi of Northwestern Ontario (Thunder Bay area), Canada.	38
<b>Table 5.1</b> Sources, Genbank accession numbers (rDNA ITS) and the size of the <i>ml</i> -U7 region for the strains used in this study.	83

# List of Figures

<b>Figure 1.1</b> Schematic representation of the 3' segment of the mitochondrial <i>rnl</i> rDNA in <i>Grosmannia piceiperda</i> .	5
<b>Figure 3.1</b> Phylogenetic relationships among “the <i>Ophiostoma</i> set” of fungi using ITS nucleotide sequences.	43
<b>Figure 3.2</b> Phylogenetic relationships among “the <i>Pesotum</i> -like set” of fungi using ITS nucleotide sequences.	46
<b>Figure 3.3</b> Phylogenetic relationships among “the <i>Ceratocystiopsis</i> set” of fungi using ITS nucleotide sequences.	49
<b>Figure 3.4</b> Phylogenetic relationships among “the <i>Graphium</i> set” of fungi using ITS nucleotide sequences.	51
<b>Figure 4.1</b> The original homing endonuclease gene sequence from <i>Grosmannia piceiperda</i> and the modified version, which has been optimized for recombinant protein expression in <i>E. coli</i> .	59
<b>Figure 4.2</b> Schematic representations of the mitochondrial <i>rps3</i> gene in ophiostomatoid fungi.	63
<b>Figure 4.3</b> The Gpi.m.Rps3 i1 intron of <i>Grosmannia piceiperda</i> .	65
<b>Figure 4.4</b> Phylogenetic relationships of double motif LAGLIDADG amino acid sequences in <i>Grosmannia</i> spp. and related sequences from Genbank.	68
<b>Figure 4.5</b> SDS-PAGE gel showing the recombinant LAGLIDADG homing endonuclease.	70
<b>Figure 5.1</b> Overview of the mitochondrial <i>rnl</i> in species of <i>Ceratocystis</i> .	87
<b>Figure 5.2</b> ITS phylogeny and a schematic of the <i>rnl</i> -U7 region of <i>Ceratocystis</i> species used in this study.	89
<b>Figure 5.3</b> RT-PCR with total RNA isolated from <i>Ceratocystis resinifera</i> and <i>C. pinicola</i> .	96

<b>Figure 5.4</b> RT-PCR with total RNA isolated from <i>Ceratocystis polonica</i> and <i>C. resinifera</i> .	98
<b>Figure 5.5</b> RT-PCR with total RNA isolated from <i>Ceratocystis polonica</i> and <i>C. fagacearum</i> .	100
<b>Figure 5.6</b> Phylogenetic position of the <i>Ceratocystis rnl-U7</i> homing endonucleases.	103
<b>Figure 5.7</b> Phylogenetic analysis of the group II intron encoded homing endonucleases.	107
<b>Figure 5.8</b> Diagrammatic representation of the intron and homing endonuclease gene evolution in the <i>rnl-U7</i> of <i>Ceratocystis</i> .	115

## List of Abbreviations

$A_{600}$	light absorbance at wavelength 600 nm
ATP	adenosine triphosphate
<i>atp</i>	ATP synthetase gene
BLAST	Basic Local Alignment Search Tool
BLASTn	nucleotide BLAST
BLASTp	protein BLAST
bp	basepair
CBS	culture collection at Centraal Bureau voor Schimmelcultures, Utrecht, The Netherlands
cDNA	complementary DNA
cpDNA	chloroplast DNA
CTAB	hexadecyl trimethyl ammoniumbromide
<i>cox</i>	cytochrome c oxidase gene
$\Delta$	deletion
D	DNA-binding domain in group II intron
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSB	double-strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
En	endonuclease domain in group II intron
g	grams
GU	guanine (G) and uracil (U)
HE	homing endonuclease
HEG	homing endonuclease gene
His	histidine
IEP	intron-encoded protein
IPTG	isopropyl- $\beta$ -D-thiogalactopyranoside
ITS	internal transcribed spacer region
kb	kilo basepair

KCl	potassium chloride
kDa	kilo Dalton
l	litre
L	large subunit ribosomal RNA gene
LB	Luria Bertani medium
LSU	large subunit ribosomal RNA gene
µg	microgram
µl	microlitre
µM	micromolar
M	molar
MEAY	malt extract agar medium with yeast extract
mg	milligram
MgCl <sub>2</sub>	magnesium chloride
MgSO <sub>4</sub>	magnesium sulfate
min	minute
ml	millilitre
mL	mitochondrial large subunit ribosomal RNA gene
ML	maximum likelihood
mM	millimolar
mRNA	messenger RNA
mS	mitochondrial small subunit ribosomal RNA gene
mt	mitochondrial
mtDNA	mitochondrial DNA
NaCl	sodium chloride
<i>nad</i>	NADH dehydrogenase gene
NADH	nicotinamide adenine dinucleotide
NaH <sub>2</sub> PO <sub>4</sub>	sodium phosphate monobasic anhydrous
ng	nanogram
Ni-NTA	nickel-nitrilotriacetic acid
nm	nanometer
nt	nucleotide
NUMOD	nuclease associated modular DNA-binding domains

NW	Northwestern
ORF	open reading frame
P	paired region in group I intron
PCR	polymerase chain reaction
PYG	peptone-yeast extract-glucose medium
rDNA	ribosomal DNA
RNA	ribonucleic acid
rRNA	ribosomal RNA
<i>rnl</i>	large subunit ribosomal RNA gene
<i>rns</i>	small subunit ribosomal RNA gene
RNP	ribonucleoprotein
rpm	revolutions per minute
<i>rps3</i>	ribosomal protein S3 gene
RT	reverse transcriptase
RT-IEP	reverse transcriptase - intron-encoded protein
RT-PCR	reverse transcriptase - polymerase chain reaction
S	small subunit ribosomal RNA gene
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
sec	second
SOC	Super Optimal broth with Catabolite repression medium
tRNA	transfer RNA
U	universally conserved region in large subunit ribosomal RNA gene
UAMH	culture collection at University of Alberta, Microfungus Collection and Herbarium, Devonian Botanic Garden, Edmonton, AB, Canada
UV	ultraviolet
WIN(M)	culture collection at University of Manitoba, Winnipeg, MB, Canada
X	maturase domain in group II intron
X-gal	bromo-chloro-indolyl-galactopyranoside
ω	omega element
°C	degrees Celsius

# Chapter 1. Literature Review

This work focuses on presenting a survey of blue-stain fungi in Northwestern Ontario and molecular characterization of homing endonuclease genes (HEGs) and introns in the large subunit ribosomal RNA gene (*rnl*) of these organisms.

## 1.1 Ophiostomatoid Fungi

The mitochondrial (mt) genomes of fungi contain several types of mobile genetic elements, including group I introns, group II introns and HEGs. These elements are considered to be optional (nonessential) and are a source of interest for studying molecular evolution and for understanding the genetic variability of mtDNA among fungal species (Goddard and Burt 1999; Edgell 2009; Sethuraman et al. 2009a). One group of fungi, known as the ophiostomatoid group, are ascomycetes that produce sticky masses of meiotic or mitotic spores on long stalks, which is ideal for dispersing these fungi via tree-dwelling insects, such as bark beetles. Species assigned to the following genera produce meiotic (sexual) spores and are commonly referred to as the ophiostomatoid fungi: *Ophiostoma*, *Grosmannia*, *Ceratocystiopsis*, *Ceratocystis*; other asexual genera are allied to the aforementioned fungi. Some of these fungi are known to be tree pathogens and cause diseases such as Dutch elm disease (*Ophiostoma ulmi*), oak wilt (*Ceratocystis fagacearum*), vascular wilts in conifers (*Grosmannia clavigera* as an associate of the mountain pine beetle) and some are aggressive vascular stain fungi (referred to as blue-stain in sapwood; *Ceratocystis polonica*). Most of the ophiostomatoid

fungi are associated with blue-staining, which causes discoloration of the wood; although it is not harmful to the tree it does diminish the quality and therefore the economic value of the lumber (Guarro et al. 1999; Juzwik et al. 2008; Sethuraman et al. 2009a).

This group of ophiostomatoid fungi share many morphological similarities but in some cases this is due to convergent evolution, as distantly related fungi have adapted to similar habitats and spore dispersal mechanisms. Fortunately, some of these taxonomic discrepancies have recently been settled with the use of molecular data, such as the internal transcribed spacer region (ITS) found between the nuclear 18S and 26S ribosomal (r)DNA. *Ophiostoma* along with *Grosmannia* and *Ceratocystiopsis* belong to the Class Sordariomycetes, Order Ophiostomatales. Historically, *Ceratocystis* and the asexual genus *Graphium* were assigned to the Ophiostomatales but recent data that includes differences at the DNA level have instead placed these genera into the Order Microascales (Guarro et al. 1999; Marin et al. 2005; Zipfel et al. 2006; Hibbett et al. 2007; Juzwik et al. 2008).

Part of the taxonomic complications of this group is due to the different reproductive structures, which depends on the lifecycle observed in culture. The most common reproductive state observed in culture is the anamorphic (asexual or mitotic) state. During this phase of the organism's lifecycle asexual spores and characteristic mitotic reproductive structures (conidiophores) are produced. It is believed that many of these fungi are also capable of producing teleomorphic (sexual or meiotic) spores and meiotic reproductive structures (perithecia); however, the parallel sexual state is

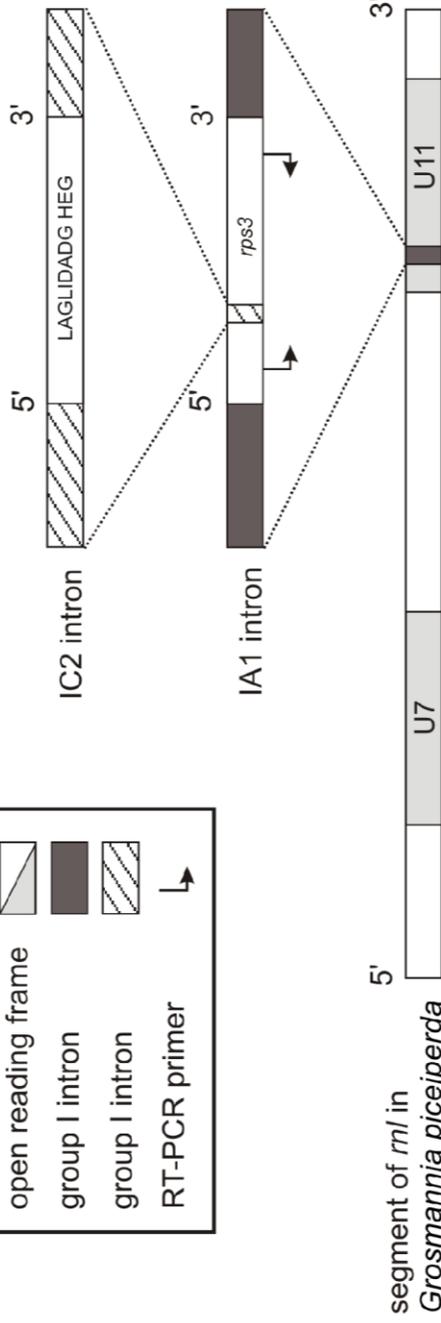
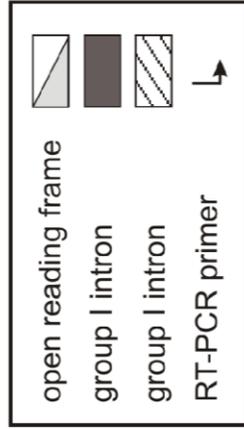
sometimes difficult to observe in culture. When a teleomorphic state cannot be observed in culture it is thought to be due to the lack of appropriate environmental conditions or because the organism has lost the ability to enter this state. Filamentous fungi have few morphological and ecological distinguishing factors and historically these asexual and sexual states have proven to be useful for species identification. In turn, with the advent of molecular distinguishing characteristics, fungal taxonomists now have even more features to help sort out this group of organisms. Reclassifying these organisms can be a difficult and tedious process considering that many of the original type-species have been lost over the years (Guarro et al. 1999; Plattner et al. 2009).

## **1.2 Fungal Mitochondria and Mitochondrial DNA**

Mitochondria are membrane-bound organelles found within most eukaryotic organisms (like fungi) and their primary purpose is to produce energy in the form of adenosine triphosphate (ATP) via oxidative phosphorylation. Although the majority of fungal DNA is found within the nuclear genome, some of the genes that are directly involved in mitochondrial function are encoded within the mtDNA genome. This mainly includes the genes involved in translation and those encoding components of the electron transport chain. The variation between fungal mtDNA genomes is partly due to non-coding regions, like intergenic spacers, and mobile genetic elements, such as group I and II introns and HEGs (Burger et al. 2003; Gibb and Hausner 2005; Sethuraman et al. 2009a).

Multiple copies of the mtDNA are present within a mitochondrion and several mitochondria are present within each cellular compartment. It is also interesting to consider that fungal mitochondria are capable of fusing upon contact; hence mixing together their contents and their mtDNA. Given that fungal colonies readily undergo anastomosis (fusion of hyphal elements), the mixing of mitochondria via cytoplasmic exchange and the mixing of mtDNA through mitochondrial fusion allows mobile elements such as group I and group II introns and HEGs to invade empty heterologous sites; the lateral transfer of these elements throughout a fungal population, or even between closely related species, seems almost inevitable (Westermann and Prokisch 2002; Burger et al. 2003; Glass et al. 2004; Haugen and Bhattacharya 2004).

Mobile genetic elements like group I and II introns and HEGs preferentially insert into conserved sequences, such as rDNA of mtDNA (Haugen and Bhattacharya 2004; Edgell 2009; Stoddard 2011). For example, the *rnl* contains 14 universally conserved regions (U1 to U14; Cummings et al. 1989); in *Grosmannia piceiperda* the *rnl*-U11 (the eleventh universally conserved region) contains an unusual gene arrangement, whereby the gene encoding ribosomal protein S3 (*rps3*) is located within a group I intron (subtype IA1). Within *rps3* there is an intron insertion that encodes a LAGLIDADG-type HEG; this intron is also classified as a group I and belongs to subtype IC2 (Figure 1.1; see Chapter 4). This complicated intron-within-intron and gene-within-gene arrangement poses interesting questions about how an intron can be spliced out from within another intron (IC2 out from within IA1) and how these three genes can be expressed while



**Figure 1.1.** Schematic representation of the 3' segment of the mitochondrial *rnl* rDNA in *Grosmannia piceiperda*. The *rnl* IA1-type group I intron (also called the mL2449 intron; dark box) encodes for the *rps3* gene. *rps3* has been invaded by an IC2-type group I intron [also called the Gpi.m.Rps3 i1 intron (striped box)], which encodes for a double motif LAGLIDADG homing endonuclease gene. RT-PCR primers used to identify the intron-exon junctions of the IC2 intron are displayed as arrows.

residing within other genes.

Another example of how introns and their open reading frames (ORFs) can influence the chromosomal landscape of the mitochondrial genome is the region upstream of *rnl*-U11 (from U7 to U11) in species of *Ceratocystis*, whereby four group I and two group II introns have inserted at various sites within approximately 1 kb of each other (see Chapter 5). Similar events can also be seen within a single species.; for instance, large numbers of insertions have been reported in the *cox1* gene of *Agaricus bisporus* (basidiomycete; 18 group I, one group II) and *Podospora anserina* (ascomycete; 15 group I; one group II) and within the mtDNA genome of *Marchantia polymorpha* (liverwort; seven group I, 25 group II). All of these examples are a chance to observe mtDNA genetic diversity generated by mobile elements and are also a demonstration of the tolerance of mtDNA to withstand a large number of intron insertions into essential genes with no observable harmful effects to the organism (Stoddard 2005; Sethuraman et al. 2009a; Férandon et al. 2010).

### **1.3 Mobile Genetic Elements in Fungal Mitochondrial DNA**

The insertion of mobile elements is generally considered to be a neutral or non-toxic event, which is how these elements are able to maintain their presence within fungal mitochondrial genomes (Stoddard 2005; Sethuraman et al. 2009a; Edgell et al. 2011). The most common occurrence in fungal mtDNA are group I introns in association with LAGLIDADG or GIY-YIG HEGs, but group II introns are also found and usually encode

reverse transcriptase (RT) intron-encoded proteins (IEPs) or LAGLIDADG homing endonucleases (HEs; Saguez et al. 2000; Lambowitz and Zimmerly 2004; Sethuraman et al. 2009a; Mullineux et al. 2011). Other mobile genetic elements related to group I and II introns, but not found in fungal mtDNA, include archaeal introns, group III introns and inteins. Group III introns are believed to be degenerate group II introns and have been found in the chloroplast genome of the protist *Euglena*. Inteins are found mainly within the genes of archaeobacteria, but are also in bacteria, viruses, bacteriophages and some eukaryotes (organelles and nuclei). Inteins (internal protein fragments) excise themselves from host proteins and rejoin the exteins (external protein fragments; exon fragments) via a peptide bond (Doetsch et al. 2000; Bonen and Vogel 2001; Chevalier and Stoddard 2001; Guhan and Muniyappa 2003; Stoddard 2005).

#### **1.4 Group I Introns**

Group I introns have a wide distribution and they have been found in premature ribosomal (r)RNA, messenger (m)RNA and transfer (t)RNA of the organellar and nuclear genomes of some eukaryotic microorganisms, and also within prokaryotes and their phages. These elements maintain a neutral presence by catalyzing their own removal from pre-RNA and rejoining the exons of the transcript. Their characteristic secondary structure typically includes 10 paired regions (P1 to P10) that form both catalytic and scaffolding domains (Haugen et al. 2004, 2005; Nielson and Johansen 2009; Raghavan and Minnick 2009; Gibb and Edgell 2010). Even though there is little primary sequence

conservation between group I introns there are small regions, for example the P7 stem and the GU nucleotides of P1, whose sequence can be exploited to help find these elements within a genome (Nielsen and Johansen 2009; Raghavan and Minnick 2009). Several examples of group I intron structures have been stored at the Group I Intron Sequence and Structure Database (GISSD; <http://www.rna.whu.edu.cn/gissd>), RNA family database (Rfam; <http://rfam.sanger.ac.uk>) and Comparative RNA Website (CRW; <http://www.rna.cccb.utexas.edu>; Nielsen and Johansen 2009). Based on similarities in secondary structure, group I introns have been categorized into subgroups IA through IE, subdivisions within the subgroup are labeled using numbers (for example subtype IC2; Michel and Westhof 1990).

The splicing of group I introns from host RNA occurs via two consecutive transesterification reactions. The first reaction occurs at the 5' splice site, where a free guanosine cofactor acts as a nucleophile and attaches to the 5' end of the intron, resulting in cleavage of the exon-intron junction. The secondary structure of the group I intron is arranged such that the host exons are in close proximity, encouraging the nucleophilic attack of the 3'OH of the 5' exon on the 3' splice site (the 3' intron-exon junction). The exons are now ligated and the excised intron is now free, circularizes and releases pyrophosphate (Michel and Westhof 1990; Nielsen and Johansen 2009; Raghavan and Minnick 2009).

## **1.5 Group II Introns**

Group II introns tend to be limited to the organelles of fungi and plants, as well as bacterial genomes and they insert into conserved regions of rDNA, tDNA and protein coding genes. The group II intron secondary structure can be conceptualized as a wheel with six spokes (domains I to VI) radiating from the centre. Very little of the group II intron primary sequence is conserved, with the exception of domain V. Based on similarities in secondary and tertiary structure and their IEPs, the group II introns have been categorized into three main families IIA, IIB and IIC (Michel and Ferat 1995; Lambowitz and Zimmerly 2004; Fedorova and Zingler 2007; Lambowitz and Zimmerly 2010; Pyle 2010). A compilation of these introns can be found at the Mobile Group II Intron Database (<http://www.fp.ucalgary.ca/group2introns>).

As with the group I introns, splicing for group II introns occurs via two sequential transesterification reactions, resulting in an excised intron and ligated exons. The initiating nucleophile is the 2'OH on a bulged adenosine residue in domain VI. In the second step the 3'OH of the 5' exon attacks the phosphodiester bond on the 3' splice site, resulting in ligated exons and a released intron RNA in the form of a lariat (Bonen and Vogel 2001; Chevalier and Stoddard 2001; Lambowitz and Zimmerly 2004; Fedorova and Zingler 2007).

## **1.6 Homing and Retrohoming of Group I and Group II Introns**

The act of homing is the transfer of an intron or HEG from a donor allele to a

cognate allele that lacks these elements. Homing is site specific and results in essentially splitting the invaded locus (recipient) into two parts. The DNA homing site functionally can be viewed as four overlapping sequences: the recognition, binding, cleavage and insertion regions. The process is initiated by a HE, whereby the enzyme recognizes and then cuts at a specific sequence, usually at a cognate allele that does not contain the intervening sequence. The cleaved DNA triggers the cellular double strand break (DSB) repair system, which is based on homologous recombination and therefore requires a homologous template. Mobility is successful if the HEG<sup>+</sup> allele is used as a template for repairing the damaged HEG<sup>-</sup> allele; in other words, the HEG is copied by replication into a new location (Dujon et al. 1989; Goddard and Burt 1999; Haugen et al. 2005; Stoddard 2005; Marcaida et al. 2010; Edgell et al. 2011; Stoddard 2011).

Retrohoming, also known as reverse splicing or retrotransposition, is the process of inserting RNA into a DNA allele, creating a DNA-RNA hybrid intermediate and then converting the RNA to a DNA copy via reverse transcriptase activity. This process does not require homologous recombination and is facilitated by the RT-IEP encoded within group II introns (Chevalier and Stoddard 2001; Lambowitz and Zimmerly 2004; Marcaida et al. 2010). The RT-IEP is a multifunctional protein that contains the following domains: maturase (X), DNA-binding (D), endonuclease (En) and reverse transcriptase (RT). Once a group II ribozyme has been spliced from the host transcript (assisted by the X domain of the RT-IEP) and has formed a ribonucleoprotein (RNP) complex with the RT-IEP the process of retrohoming begins. The RNP recognizes (D domain) and cuts at the

retrohoming site; first one strand is cut with the 3' OH of the intron lariat and then the second DNA strand is cut by the En domain; meanwhile the intron RNA is inserted into the target site. The RNA is then reverse transcribed by the RT domain and the original RNA template is replaced with a DNA copy by the host DNA repair system (Lambowitz and Zimmerly 2004; Robart and Zimmerly 2005; Fedorova and Zingler 2007; Edgell et al. 2011).

### **1.7 Homing Endonucleases Can Mobilize Flanking DNA**

When HEGs move to a new site they can sometimes bring along additional sequences [for instance, the HEGs that have inserted into the C-terminus of *rps3* within some ophiostomatoid fungi (Sethuraman et al. 2009a)]. The insertion of these HEGs has shifted a small portion of the *rps3* coding region downstream and disrupted the ORF with a premature stop codon. To compensate, the HEGs incorporated a duplicate of the displaced *rps3* sequence upstream to the HEG coding region. Interestingly, the nucleotide sequence is not identical to the original displaced sequence, suggesting that the sequence was brought along with the invading HEG. A similar example was observed by Paquin et al. (1994), whereby a HEG was found in an ATP synthetase subunit 6 gene (*atp6*) with a portion of foreign *atp6* sequence (fused upstream and in-frame). These events are due to gene conversion, which happens during DNA DSB repair, and is in part triggered by the exonuclease activity that increases the original gap generated by the HE cleavage at the homing site. These gene conversion events can neutralize the potential damage caused by

HEGs or introns inserting into pre-existing ORFs. In addition to correcting ORF disruptions, sometimes these extra flanking sequences are different than the host gene; for example, near the *rnl-U7* region of *Ceratocystis fagacearum* there is a degenerate GIY-YIG HEG accompanied by a portion of the NADH dehydrogenase subunit 2 gene (*nad2*). Since the HEG and *nad2* are situated within rDNA, one would theorize that the *nad2* sequence was mobilized by the GIY-YIG HEG (see Chapter 5).

## 1.8 The Homing Cycle

The mobility of HEGs and their introns can be conceptualized as undergoing a cycle, with perpetual invasion and loss (degeneration) at the homing site (Goddard and Burt 1999; Gimble 2000; Stoddard 2005; Gogarten and Hilario 2006). Goddard and Burt (1999) tested this model by working with an intron-encoded HEG known as omega ( $\omega$ ). The  $\omega$  element was surveyed in 20 species of yeast and was found to be in one of three states: functional, nonfunctional or absent (Goddard and Burt 1999). Using phylogenetic analysis they proposed a cyclic model of invasion, degeneration, loss and reinvasion that occurs approximately once every 2 million years. The model suggests that once a HEG has become established in a population (has invaded all potential sites) it will accumulate random mutations and eventually the sequence will be lost. A new cycle may begin if the HEG is able to invade a new location. Although this simple model is useful, it cannot be used in cases where the host benefits from the presence of the HEG, such as when a HE gains maturase activity, because the sequence is then subject to selection pressure

(Gogarten and Hilario 2006).

## **1.9 Families of Homing Endonuclease Genes**

HEs are ORFs that are usually embedded in group I introns, group II introns, archaeal introns or inteins and in some cases can be free-standing (without an intron or intein). These DNA-cutting enzymes are the driving force of the intron homing cycle and promote the movement of intervening sequences to new sites in the genome. They are widespread throughout the nuclear and organellar genomes of microorganisms (bacteria, fungi, protists, algae and viruses) but are yet to be found in more complex, multicellular organisms. The success of these elements is earmarked by the ability to balance sequence-specificity with sequence-flexibility; that is, remaining a neutral presence but once in a while jumping to a new site (Chevalier and Stoddard 2001; Stoddard 2005; Sethuraman et al. 2009a). The HE proteins are divided into five families based on conserved amino acid motifs known as: LAGLIDADG, GIY-YIG, HNH, His-Cys box and PD-(D/E)<sub>x</sub>K. LAGLIDADG and GIY-YIG type HEGs are currently the only families observed in the mtDNA of fungi. The HNH family is found in phage, the His-Cys box family is found in protists and the PD-(D/E)<sub>x</sub>K family is found in cyanobacteria (Chevalier and Stoddard 2001; Stoddard 2005; Zhao et al. 2007; Marcaida et al. 2010; Stoddard 2011).

### **1.9.1 The LAGLIDADG Family of Homing Endonucleases**

The LAGLIDADG homing endonuclease family of proteins have been found

mainly in organellar genomes, namely the mitochondria of fungi and protozoans and plant and algal chloroplasts, and also within the genomes of Archaeobacteria. It is currently the largest family of HEGs (Haugen and Bhattacharya 2004; Stoddard 2005; Marcaida et al. 2010). Alternative names for LAGLIDADG proteins include dodecapeptide, dodecamer, docapeptide and DOD all of which refer to the 10 conserved amino acid residues of the enzyme (Chevalier and Stoddard 2001; Stoddard 2005) and the LAGLIDADG ORF may contain a single or a double copy of the conserved motif. When single motif LAGLIDADG HEGs bind to target DNA sequences they assemble into homodimers, thus requiring symmetry in the recognition site (for example a palindromic or a pseudopalindromic sequence). In contrast, double motif LAGLIDADG HEGs bind to DNA as monomers and since their amino acid motifs are not identical their target DNA sequences are less stringent than the single motif enzymes. This flexibility has allowed the double motif LAGLIDADG HEGs to be more successful invaders of new target sites in the genome compared to their single motif HEG counterparts (Chevalier and Stoddard 2001; Haugen and Bhattacharya 2004; Stoddard 2005). Double motif LAGLIDADG HEGs are thought to have originated from gene duplication (followed by in-frame fusion of the motifs) of a single motif LAGLIDADG HEG (Gimble 2000; Haugen and Bhattacharya 2004; Gibb and Hausner 2005; Haugen et al. 2005; Sethuraman et al. 2009a).

### **1.9.2 The GIY-YIG Family of Homing Endonucleases**

The GIY-YIG homing endonucleases have been found mainly in organellar genomes, such as the mitochondria of fungi, the mitochondria and chloroplasts of algae and the genomes of phages. GIY-YIG homing endonucleases are bipartite enzymes, composed of a catalytic domain and a DNA-binding domain, joined by a linker region (Saguez et al. 2000; Stoddard 2005; Marcaida et al. 2010). Sometimes the DNA-binding domain is followed by additional domains known as NUMODs (nuclease associated modular DNA-binding domains; Stoddard 2005; Marcaida et al. 2010). There are five characteristic amino acid motifs, including GIY-YIG, that are important in structure and function; even so, there is little sequence conservation within the GIY-YIG family of HEGs (Stoddard 2005).

### **1.10 Applications of Homing Endonucleases and Group II Introns**

The heart of the homing process is essentially replacing a damaged section of DNA with a template allele; this can be thought of as a type of gene conversion event. When HEs cut at highly specific sites they rely on cellular repair mechanisms to rectify the damage. These properties can be used to the advantage of scientists trying to develop treatments for monogenic diseases (disease caused by one gene), by replacing the defective gene by homologous recombination at low toxicity levels (Stoddard 2005; Marcaida et al. 2010; Stoddard 2011). In addition to gene therapy, another potential for HEs is pest control. For example, populations of mosquitoes vectoring infectious agents

(such as malaria) could be controlled by hindering the insect's ability to reproduce or the mosquito could be manipulated to be resistant to carrying the disease (Stoddard 2005). Successful experimentation with this idea has been performed on *Drosophila*; whereby a modified HE (originally isolated from yeast) was introduced into *Drosophila* and was shown to be active by cleaving a DNA target site. Several groups have been working in the area of modifying the HE's recognition site by redesigning the enzyme's DNA contact sites (amino acid changes) and by fusing together different enzyme domains. Crystal structures have helped to elucidate how HEs interact with the target site (Stoddard 2005; Marcaida et al. 2010; Stoddard 2011). In a similar manner, the retrohoming ability of group II introns have been developed into gene knockout systems (targetrons). Targetrons are currently available for bacterial systems, but there have been recent developments towards their use in eukaryotic systems (Lambowitz and Zimmerly 2004; Lambowitz and Zimmerly 2010).

### **1.11 Research Objectives**

The work in this thesis has been divided into three chapters. One project centers on blue-stain fungi collected in Northwestern (NW) Ontario and the remaining two projects focus on the HEGs and introns found in the rDNA of these organisms.

(1) The objective of the first project was to utilize rDNA ITS region sequences to further resolve the taxonomic status of 44 ophiostomatoid fungal specimens collected from NW Ontario near Lakehead University (Thunder Bay, ON; Andersen 2009).

Initially, these fungi were placed into taxonomic groups based on morphological features (Andersen 2009). These fungi were also added to the WIN(M) culture collection and will be used in the future for bioprospecting for mobile elements (see Chapter 3).

(2) The objective of this study was to characterize the mobile genetic elements found in the *rps3* of *Grosmannia piceiperda*. Previous research by the Hausner group has been performed on the mL2449 group I intron within the *rnl*-U11 region of the mitochondrial *rnl* gene of ophiostomatoid fungi (Gibb and Hausner 2005; Sethuraman et al. 2009a). The mL2449 intron is unique because it encodes for the *rps3* gene, rather than a HEG (Figure 1.1). However, this previous work did not correctly annotate the *rps3* sequences within species of *Grosmannia* (Sethuraman et al. 2009a). The highlight of this project was the discovery that a HEG (often referred to as the A type HEG in *rps3*) which was thought to be a mobile unit on its own but is in actuality associated with a group I intron (subtype IC2). This observation provides a plausible explanation for how the A type HEG can integrate within the *rps3* N terminal domain and not cause damage to the host gene. Additional work was attempted to further characterize a recombinant form of the A type HE by over-expressing it in *E. coli*, but the protein could not be isolated in soluble form (see Chapter 4).

(3) The last chapter, an expansion of work started by Sethuraman et al. (2008), is an examination of the *rnl* locus in species of *Ceratocystis*. This project demonstrates the large amount of genetic diversity that can be generated by mobile genetic elements among closely related fungal species. As with Haugen and Bhattacharya's (2004) study, we

found phylogenetic evidence that the mL1923 single motif LAGLIDADG HEG might be the ancestor of other rDNA HEGs. This study also includes the discovery of some oddities, such as a GIY-YIG HE parasitizing a fellow LAGLIDADG HEG and a portion of a *nad2* gene fused to a GIY-YIG HEG within the *rnl* gene (see Chapter 5).

## **Chapter 2: General Materials and Methods**

### **2.1 Fungal Culturing**

All fungal strains were cultured in Petri dishes containing 2% malt extract agar with yeast extract (MEAY; 20 g malt extract, 1 g yeast extract and 20 g bacteriological agar per litre) at 20 °C for 5 - 7 days. For DNA extraction, agar plugs of fungal cultures grown on MEA were used to inoculate 50 ml of peptone yeast extract glucose medium (PYG; 1 g peptone, 1 g yeast extract and 3 g glucose per litre) in 125-ml flasks; cultures were still-grown at 20 °C for 5 - 7 days.

### **2.2 Fungal DNA Extraction**

Fungal mycelia were harvested from PYG broth by vacuum filtration on Whatman # 1 filter paper and transferred to a 15 ml Falcon tube (Corning Inc., Corning, NY) containing 3 ml extraction buffer [10 mM Tris-Cl at pH 7.6, 1 mM ethylenediaminetetraacetic acid (EDTA), 50 mM sodium chloride (NaCl), 1 % hexadecyl trimethyl ammoniumbromide (CTAB) and 0.5 % sodium dodecyl sulfate (SDS)] and 3 g of 3-mm glass beads. Mycelia were vortexed with the beads until a homogenous mixture was obtained, usually for about 15 min (alternating vortex time with incubation on ice every 5 min). An additional 3 ml of extraction buffer and 660 µl 20 % SDS was added before incubating the mixture at 55-60 °C for a minimum of 2 hours. Tubes were cooled to room temperature and mixed by inversion with 6 ml of chloroform. This mixture was incubated for 15-20 min before centrifuging at 2 000 rpm for 15-20 min. The aqueous

supernatant (typically 5 ml) was moved to a fresh tube and mixed with 2.5 volumes ice-cold 95 % ethanol to precipitate DNA. Tubes were incubated for a minimum of 2 hours at -20 °C and the DNA pelleted by centrifugation at 3 000 rpm for 30 min. Pellets were washed with 1 ml ice-cold 70 % ethanol and centrifuged again at 3 000 rpm for 30 min. All liquid was discarded and tubes were inverted at room temperature for 15-30 min before resuspending the DNA pellet in 300 µl Tris-EDTA buffer (TE; 10 mM Tris-Cl at pH 7.6 and 1 mM EDTA) and storing at -20 °C.

### **2.3 PCR Amplification and Purification**

The polymerase chain reaction (PCR) primers were obtained from Alpha DNA (Montreal, PQ, Canada) and are listed in Table 2.1. The nuclear internal transcribed spacer region (ITS) was amplified using the PCR<sub>X</sub> Enhancer System [Invitrogen, Carlsbad, CA; 50 µl reactions; 1x amplification buffer, 1-1.5x enhancer solution, 0.2 mM dNTP mix, 1.5 mM magnesium sulfate (MgSO<sub>4</sub>), 0.4 µM of each primer, 50-100 ng of whole cell DNA and 1.25 units Taq DNA polymerase in ultra pure water]. The amplification conditions of the ITS region were: initial denaturation for 2 min at 94 °C, 25-30 cycles of 1 min at 95 °C, 1.5 min at 52 °C and 3 min at 70 °C, followed by a final extension for 10 min at 70 °C.

Segments of the mitochondrial *rnl* rDNA were PCR amplified in 50 µl reactions containing 1x PCR buffer for Taq polymerase [Stratagene, Cedar Creek, TX; containing 1.5 mM

**Table 2.1.** Sequences of the PCR primer pairs used to amplify nuclear and mitochondrial DNA in this study. Forward primer listed first. Synthesized DNA refers to the fungal sequence that was modified and optimized for expression in bacteria.

<b>Location</b>	<b>Amplified Region</b>	<b>Primer Name</b>	<b>Primer Sequence (5' to 3')</b>
Nuclear DNA	Internal Transcribed Spacer (ITS)	SSU4	TTAAAGAAATTGACGGAAGGG
		LS2	GATATGCTTAAGTCAGCG
Mitochondrial DNA	Large ribosomal subunit ( <i>rnl</i> )-U7	Lsex1	GCTAGTAGAGAATACGAAGGC
		Lsex2	GACCGCATTTAACGGCCAAGG
	<i>rnl</i> -U11	IP1	GGAAAAGCTACGCTAGGG
		IP2	CTTGCGCAAATTAGCC
	Inter U7U11 region	Lsex2R	CCTTGGCCGTAAATGCGGTC
		IP1R	CCGTAGCGTAGCTTTTCC
Synthesized DNA		nF975synF	CACCTGGTGGTTTGGTAAATC
		nR975synR	TTATTAAATACGACGGGTGTTC

magnesium chloride ( $\text{MgCl}_2$ ), 0.2 mM dNTP mix, 0.75-1.5 mM  $\text{MgCl}_2$ , 0.4  $\mu\text{M}$  of each primer, 50-100 ng of whole cell DNA and 1.25 units Taq DNA polymerase in ultra pure water; some reactions contained 0.9 units Taq polymerase and 0.3 units PFU polymerase. The amplification conditions of the *rml* were: initial denaturation for 2 min at 94 °C, 20-30 cycles of 1 min at 93 °C, 1.5 min at 52.0-52.9 °C and 4.5 min at 70 °C, followed by a final extension for 10 min at 70 °C.

The synthesized A type HEG was amplified using the Platinum High Fidelity Taq Polymerase kit (Invitrogen; 50  $\mu\text{l}$  reactions; 1x amplification buffer, 0.2 mM dNTP mix, 2 mM  $\text{MgSO}_4$ , 0.4  $\mu\text{M}$  of each primer, 50-100 ng of plasmid DNA and 1.25 units Platinum High Fidelity Taq polymerase in ultra pure water). The amplification conditions were: initial denaturation for 2 min at 94 °C, 25 cycles of 1 min at 93 °C, 1.5 min at 52.9 °C and 4.5 min at 70 °C, followed by a final extension for 10 min at 70 °C.

PCR fragments were resolved by electrophoresis in 0.8-1.5 % agarose in Tris-borate-EDTA buffer (TBE; 89 mM Tris base, 89 mM boric acid, and 2 mM EDTA at pH 8.0). DNA fragments were visualized by staining the gel with ethidium bromide (0.5  $\mu\text{g}/\text{ml}$ ) and exposing it to ultraviolet (UV) light. PCR fragments were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI) according to the manufacturer's directions and eluted from the column using 30-50  $\mu\text{l}$  ultra pure water. Some DNA fragments were cut from the gel and extracted using the Wizard SV Gel and PCR Clean-Up System following the manufacturer's directions.

## 2.4 Cloning of PCR Products

Some PCR fragments were cloned using the TOPO TA Cloning Kit for Sequencing (Invitrogen) using the following procedure: purified PCR products were incubated with 1x PCR buffer for Taq polymerase (Stratagene), 0.2 mM dATP and 1 unit Taq DNA polymerase at 70 °C for 10 min to synthesize 3' A-overhangs. The PCR product was then incubated in a salt solution (200 mM NaCl, 10 mM MgCl<sub>2</sub>) and the pCR4-TOPO vector at room temperature for 5-10 min, before adding the mixture to 50 µl of DH5α-T1 chemically competent cells and incubating on ice for 30 min. Cells were heat-shocked at 42 °C for 30 sec and transferred to 250 µl Super Optimal broth with Catabolite repression medium [SOC; 0.5 % yeast extract, 2 % tryptone, 10 mM NaCl, 2.5 mM potassium chloride (KCl), 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> and 20 mM glucose] and shaken at 37 °C for 1 hour. The culture was then spread plated on Luria Bertani (LB; 10 g tryptone, 5 g yeast extract, 10 g NaCl and 20 g bacteriological agar per litre) plates supplemented with 50-60 µg/ml ampicillin and 20 mg/ml bromo-chloro-indolyl-galactopyranoside (X-gal) and incubated overnight at 37 °C.

White colonies were transferred to fresh LB plates supplemented with ampicillin and incubated overnight at 37 °C. The fresh colonies were used to inoculate LB broth supplemented with ampicillin and incubated overnight at 37 °C. Plasmids were isolated using the Wizard *Plus* SV Minipreps DNA Purification System (Promega) according to the manufacturer's directions and eluted from the column using 50 µl ultra pure water.

## **2.5 DNA Sequencing**

DNA sequences were generated using the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Austin, TX; 15-20  $\mu$ l reaction volume; 0.5x sequencing buffer, 0.25x reaction mix, 0.2 mM dNTP mix, 0.4  $\mu$ M primer and 100-500 ng DNA template in ultra pure water). Prior to adding the 0.25x reaction mix, mixtures containing plasmid were first heated at 95 °C for 10 min to denature the DNA template. Amplification conditions were as follows: 96 °C for 1 min, and then 25 cycles of 96 °C for 10 sec, 50 °C for 5 sec, 60 °C for 4 min. Amplified fragments were incubated for 15 min at room temperature in 7.35 mM EDTA and 70 % ice-cold ethanol (inverted several times to mix). DNA was pelleted by centrifugation in the cold (~ 4 °C) at 13 000 rpm for 15-30 min. The supernatant was aspirated and the DNA pellet washed twice by adding 60  $\mu$ l of 70 % ice-cold ethanol and centrifuging at 13 000 rpm for 15-30 min. Tubes were dried in the dark at 37 °C for 10-15 min (or until completely dry). Pellets were resuspended in 15-20  $\mu$ l sequencing grade formamide (Applied Biosystems), heated at 94 °C for 5 min and processed on a 3130 Genetic Analyzer (Applied Biosystems).

## **2.6 Fungal RNA Extraction and Purification**

RNA was isolated from about 50-100 mg of fungal mycelium using the RNeasy kit for total RNA isolation (Qiagen, Maryland) following the manufacturer's instructions with some modifications. Fungal cultures were grown in PYG and mycelia were harvested by vacuum filtration as described for Fungal Culturing and DNA Extraction.

Mycelia were transferred to an empty Petri dish, covered with 200  $\mu$ l RNAlater (Qiagen) and frozen at  $-60^{\circ}\text{C}$  overnight. Frozen mycelia were then transferred into a cooled mortar and ground with liquid nitrogen to a fine white powder. A 1.5-ml tube was filled to the 100- $\mu$ l mark and treated with 450  $\mu$ l RLC lysis buffer (supplemented with 4.5  $\mu$ l  $\beta$ -mercaptoethanol) and vortexed, 300  $\mu$ l 95 % ethanol was added before transferring the mixture to the spin column (included with the RNeasy kit). The column membrane was washed with 350  $\mu$ l RW1 buffer, centrifuged for 15 sec and the flowthrough discarded. On-column digestion of DNA was performed by adding 10  $\mu$ l DNase I in 70  $\mu$ l RDD buffer and incubating at room temperature for 15 min. Column membrane was then washed with 350  $\mu$ l RW1 buffer, 500  $\mu$ l RPE buffer (twice), each time discarding the flowthrough after centrifuging for 15 sec. Spin column was then transferred to a new collection tube and centrifuged for an additional 1 min at 13 000 rpm to ensure removal of residual RPE buffer (RPE buffer contains ethanol). RNA was then eluted from the membrane into a clean microcentrifuge tube by adding 30  $\mu$ l RNase-free water and centrifuging at 13 000 rpm for 1 min. Purified RNA was stored at  $-60^{\circ}\text{C}$ . Contaminating DNA was digested using the TURBO DNase kit (Applied Biosystems) following the manufacturer's directions. Briefly, a 20  $\mu$ l reaction was set-up as follows: 15  $\mu$ l purified RNA, 2  $\mu$ l TURBO DNase Buffer (10x), 2  $\mu$ l RNase-free water and 1  $\mu$ l TURBO DNase (2 units/ $\mu$ l). Reaction mixture was incubated at  $37^{\circ}\text{C}$  for 30 min and then at  $75^{\circ}\text{C}$  for 10 min to deactivate the DNase enzyme. DNA-free RNA was stored at  $-60^{\circ}\text{C}$ .

## **2.7 Reverse Transcriptase PCR Amplification and cDNA Amplification**

Complementary (c)DNA synthesis was performed using the ThermoScript RT-PCR system (Invitrogen) following the manufacturer's instructions. Briefly, a 12  $\mu$ l reaction was set-up containing 8  $\mu$ l RNase-free water, 1  $\mu$ l purified RNA, 1  $\mu$ l primer (40  $\mu$ M; Lsex2 or IP1R) and 2  $\mu$ l dNTP mix (10 mM), incubated at 65 °C for 5 min and added to a second reaction mixture of: 4  $\mu$ l cDNA synthesis buffer (5x), 1  $\mu$ l Dithiothreitol (DTT; 0.1 M), 1  $\mu$ l RNase-free water, 1  $\mu$ l RNaseOUT (40 units/ $\mu$ l), 1  $\mu$ l Reverse Transcriptase (15 units/ $\mu$ l) before incubating at 52.9 °C for 55 min, then 85 °C for 5 min. Finally, 1  $\mu$ l of RNase H was added to degrade the RNA template and incubated at 37 °C for 20 min. cDNA was stored at -60 °C and PCR amplified in a 50  $\mu$ l reaction using the Platinum Taq DNA polymerase kit (Invitrogen) as follows: 38.1  $\mu$ l RNase-free water, 5  $\mu$ l PCR buffer (10x), 1.5  $\mu$ l MgCl<sub>2</sub> (50 mM), 1  $\mu$ l dNTP mix (10 mM), 1  $\mu$ l of each primer (10  $\mu$ M; Lsex1 / Lsex2 or Lsex2R / IP1R), 2  $\mu$ l cDNA and 0.4  $\mu$ l Platinum Taq DNA polymerase (5 units/ $\mu$ l).

## **2.8 Bacterial Glycerol Stocks and Precipitated Plasmid DNA Storage**

Bacterial glycerol stocks and precipitated plasmid DNA were stored at -60 °C and prepared as follows: bacterial colonies were transferred to a tube containing 5 ml LB supplemented with antibiotic (100  $\mu$ g/ml ampicillin or kanamycin) and incubated overnight at 37 °C with shaking. Aliquots of 50  $\mu$ l were then added to 250  $\mu$ l LB supplemented with 36 % glycerol. Plasmid DNA was purified as described in the Cloning

of PCR Products section. Per 50 µl of plasmid: 12.5 µl of 5 M NaCl and 160 µl ice-cold 95 % ethanol were added and mixed by inversion. Mixture was incubated for a minimum of 1 hour at -20 °C and centrifuged in the cold (~ 4 °C) at 13 000 rpm for 25 min. Supernatant was aspirated and DNA pellet was washed by adding 80 µl ice-cold 70 % ethanol and centrifuging for an additional 25 min at 13 000 rpm. Supernatant was discarded and the pellets were allowed to dry at room temperature before storage.

## **2.9 RNA Folding**

All DNA sequences were submitted to the online program RNA Weasel (Lang et al. 2007; <http://megasun.bch.umontreal.ca/RNAweasel/>) and scanned for sequences that are characteristic of group I and group II introns. In cases where RNA Weasel detected a potential intron signature, the nucleotide sequence was submitted to the online program Mfold (Zuker 2003; <http://www.bioinfo.rpi.edu/~zukerm/rna/>), which was used to model the intron's secondary structure. Final structures were drawn using Corel Draw (v14.0; Corel Corporation Limited).

## **2.10 Bioinformatics**

Raw sequence data obtained from mitochondrial (mt)DNA loci were compiled manually into contiguous sequences using GeneDoc (v2.7.000; Nicholas et al. 1997). All nucleotide sequences were examined for the presence of open reading frames (ORFs; such as homing endonuclease genes) by using the online program ORF Finder

(<http://www.ncbi.nlm.nih.gov/gorf/gorf.html/>; setting: genetic code for mtDNA of molds). Searches for sequences related to intron encoded ORFs were conducted using BLASTp (<http://www.ncbi.nlm.nih.gov/BLAST/>; Altschul et al. 1990) and sequences obtained from this study and those obtained from GenBank were aligned with the PRALINE multiple amino acid sequence alignment program (Simossis and Heringa 2003; 2005). The alignments were adjusted by eye with the alignment editor function contained within GeneDoc. The ribosomal (r)DNA ITS data were also archived and assembled within the GeneDoc program and ITS nucleotide sequences related to those obtained from this study were gathered via BLASTn. The ITS data sets were aligned using Clustal-X (v2.0.12; Thompson et al. 1997) and when required the alignments were adjusted manually with GeneDoc. Phylogenetic trees were drawn with TreeView (Page 1996) and Corel Draw.

## **Chapter 3: Molecular taxonomy of blue-stain fungi from NW Ontario**

This chapter complements the morphological work by S. Andersen at Lakehead University (Thunder Bay, ON, Canada; Hutchison Research Group). Morphological descriptions of the ophiostomatoid fungi were contributed by the Hutchison Research Group and the molecular data (ITS sequences) were generated by the Hausner Research Group. For detailed information about how and where these cultures were collected see the following reference.

Andersen S. 2009. An examination of bark beetles and their associated blue stain fungi on boreal jack pine (*Pinus banksiana*) and white spruce (*Picea glauca*) in the Thunder Bay region. [M.Sc.F. thesis]. (Thunder Bay, ON): Lakehead University.

## **ABSTRACT**

Some members of the ophiostomatoid fungi cause blue-stain in sapwood and are of economical interest in the lumber industry. Very little survey-related research has been performed in the forests of Northwestern Ontario to determine the types of ophiostomatoid fungi present. Sixty-five putative strains of ophiostomatoid fungi collected from this area were examined and the ribosomal internal transcribed spacer region (ITS) was sequenced for forty four representative strains. Taxonomic classifications were established for some of these strains using the ITS sequence data, in combination with ITS data extracted from Genbank and morphological characteristics. The study revealed some strains that are related to the blue-stain fungus *Leptographium fruticetum*, which has never been reported in this area before, in addition to three strains that may be related to a fungus described from Japan, *Ophiostoma rectangulosporium*. This study also indicates the presence of an exotic species, *Leptographium wingfieldii*, that is thought to have been recently introduced into Canada. This investigation provides an account of some of the blue-stain currently present in Northwestern Ontario and preliminary identification of strains provides a foundation for future taxonomic and molecular studies.

### 3.1 INTRODUCTION

Ophiostomatoid fungi are ascomycetes, forcible ascospore discharge is absent, have deliquescent asci (thin walled and short-lived spore sacks) and form sticky ascospore droplets at the apex of their perithecial necks. These fungi also tend to produce slimy/sticky conidia on long-stalked conidiophores (asexual spore-producing structures). Taxonomically the following genera are included in the ophiostomatoid group: *Ceratocystis*, *Sphaeronaemella*, *Cornuvesica*, *Ophiostoma*, *Grosmannia* and *Ceratocystiopsis* (Hausner and Reid 2004; Zipfel et al. 2006). These genera are associated with a plethora of conidial states (asexual states) that have been assigned to asexual genera such as *Garbanaudia*, *Chalara*, *Pesotum*, *Sporothrix*, *Leptographium*, *Hyalorhinochlaeniella* etc. (Upadhyay 1981; Hausner et al. 2000; Hausner and Reid 2004; Zipfel et al. 2006). Over the last 20 years genetic data has shown that the ophiostomatoid fungi are polyphyletic and that *Ceratocystis*, *Sphaeronaemella* and *Cornuvesica* belong to the Microascales (Hausner et al. 1993a; Hausner and Reid 2004), while *Ophiostoma*, *Grosmannia* and *Ceratocystiopsis* are genera that can be placed into the Ophiostomataceae (Ophiostomatales; Hausner et al. 2000; Zipfel et al. 2006). Species that are assigned to *Ceratocystiopsis* in many ways resemble species of *Ophiostoma*, except they tend to have small dark perithecia with short perithecial necks and falcate, sheathed ascospores (Upadhyay 1981; Zipfel et al. 2006; Plattner et al. 2009) and a lower tolerance to cycloheximide than *Ophiostoma* species (Harrington 1981; Hausner et al. 1993a,

1993b). Species of *Grosmannia* can be distinguished from members of *Ophiostoma* and *Ceratocystiopsis* by the presence of a conidial state that can be assigned to the genus *Leptographium* Lagerb. & Melin.

Collectively these fungi have evolved to be dispersed by bark-beetles, thus their asexual and, if present, sexual reproductive structures produce masses of slimy/sticky spores. Identification of species among these fungi is very difficult due to convergent evolution affecting the few morphological features available to describe cultures of these fungi. Furthermore, in the laboratory these fungi often fail to fruit, leading to even fewer characteristics available for species identification in culture.

Many ophiostomatoid fungi are blue-stain fungi and some are known to be tree pathogens, such as the Dutch elm disease causing agent *Ophiostoma ulmi* and oak wilt that is caused by *Ceratocystis fagacearum* (reviewed in Wingfield et al. 1993). Blue-stain is caused by the discoloration of the infected wood. These fungi produce secondary metabolites (melanin) involving Polyketide Synthase (Tanguay et al. 2006); this results in reduced market value of the timber. Due to the economic impact of blue-stain fungi on timber and the virulent nature of some ophiostomatoid fungi it is beneficial to perform surveys on Canadian forests (Hausner et al. 2005).

This study mostly focuses on ophiostomatoid fungi collected from bark beetle galleries of various conifer species and from different bark beetles in the Thunder Bay region of Ontario. Sixty-five putative ophiostomatoid strains collected from Northwestern (NW) Ontario were sent to our laboratory. With the aid of cultural and morphological

characteristics and internal transcribed spacer (ITS) sequence data our goal was to taxonomically classify these fungal strains. It has been previously shown that the ITS regions of nuclear ribosomal (r)DNA can be used to help identify fungal species (Guarro et al. 1999).

The cultures were selected by the Hutchison Research Group (Lakehead University) based on conidial states (asexual structures) and those cultural features that are typically associated with species of *Ophiostoma* or closely related species (such as *Grosmannia* or *Ceratocystiopsis*). Not all of the collected strains produced perithecia (sexual state structures) and some even lacked well-defined conidial states in culture making even preliminary identification difficult. The fungal strains provided to us were assigned to morphological groups based on features such as: *Pesotum*-like, *Sporothrix*-like, *Hyalorhinocladiella*-like, *Leptographium*-like, *Graphium*-like and “unknown conidial states”. *Graphium*-like fungi superficially resemble *Pesotum* states, as both produce synnematosus type conidiophores [i.e. the stalk of the conidiophores consists of “bundles” of hyphae that have aggregated together and at the hyphal tips produce conidia (mitotic/asexual) spores]. However, *Graphium* and *Pesotum* can be differentiated based on the mode of conidiogenesis (reviewed in Okada et al. 2000). Conidiogenesis requires detailed microscopic analysis, sometimes supported by electron microscopic examination of conidiophores. *Graphium* is an asexual genus associated with members of the Microascales, while *Pesotum* is an asexual genus associated with species of the genus *Ophiostoma*; therefore, these two asexual genera can be readily distinguished based on

molecular data such as sequences of the ITS region.

The first objective was to obtain ITS sequence data for strains representative of the various morphological subgroups within the Lakehead collection and then assemble a data set that contained ITS sequences from GenBank. The second objective was to undertake a phylogenetic analysis of the ITS data, which allowed for assigning the unknown strains to genera (and even to the species level).

## **3.2 SUMMARY OF METHODS**

The rDNA ITS regions of 45 fungi were PCR amplified and sequenced (Table 3.1). Related ITS sequences were retrieved from Genbank and relationships between sequences were established based on the results of phylogenetic and morphological data. For the following methods see Chapter 2: fungal culturing, DNA extraction, PCR, cloning, sequencing and bioinformatics.

### **3.2.1 Phylogenetic Analysis**

Only those segments of the multiple sequence alignment where all sequences could be aligned unambiguously were retained for phylogenetic analyses. Phylogenetic estimates were generated by the programs contained within the PHYLIP package (Felsenstein 1985; 2006) and the MrBayes program v3.1 (Ronquist and Huelsenbeck 2003; Ronquist 2004). In PHYLIP (version 3.69c), phylogenetic trees were

**Table 3.1.** Ophiostomatoid fungi of Northwestern Ontario (Thunder Bay area), Canada.

<b>Taxon<sup>a</sup></b>	<b>Source<sup>b</sup></b>	<b>Substrate, Isolation Method, Locality<sup>c</sup></b>	<b>Accession Numbers</b>
<i>Ceratocystiopsis</i> sp. H.P. Upadhyay & W.B. Kendr.	WIN(M) <sup>1</sup> 1578 = LJH <sup>2</sup> 62	White Spruce; Adult Washing; Jack Haggerty Forest	HM363150
	WIN(M) 1644 = LJH 55	Jack Pine; Adult Washing; Silver Mountain	HM363151
<i>Ceratocystiopsis manitobensis</i> (J. Reid & Georg Hausner) Zipfel, Z.W. Beer & M.J. Wingf.	WIN(M) 1641 = LJH 46	Jack Pine; Adult Washing; Jack Haggerty Forest	HM363152
	WIN(M) 1643 = LJH 54	Jack Pine; Adult Washing; Silver Mountain	HM363153
<i>Ceratocystiopsis minima</i> (Olchow. & J. Reid) H.P. Upadhyay	WIN(M) 1628 = LJH 50	Jack Pine; Larval Washing; Quackenbush Woodlot	HM363154
<i>Ceratocystiopsis minuta</i> (Siemaszko) H.P. Upadhyay & W.B. Kendr.	CBS <sup>3</sup> 117042 = WIN(M) 1523	Norway spruce; NA <sup>4</sup> ; Kobernausserwald, Upper Austria, Austria	HM363147
<i>Graphium</i> sp. Corda	WIN(M) 1589 = LJH 53	Jack Pine; Adult Washing; Silver Mountain	HM363148
	WIN(M) 1631 = LJH 22	Black Spruce; Beetle Frass; Jack Haggerty Forest	HM363149
<i>Leptographium</i> sp. Lagerb. & Melin	WIN(M) 1577 = LJH 61	White Spruce; Beetle Frass; Jack Haggerty Forest	HM363155
	WIN(M) 1595 = LJH 4	Tamarack; Fruiting Body; Jack Haggerty Forest	HM363156
	WIN(M) 1618 = LJH 23	Black Spruce; Fruiting Body; Jack Haggerty Forest	HM363157
<i>Leptographium fruticetum</i> Alamouti, J.J. Kim & C. Breuil	WIN(M) 1576 = LJH 60	White Spruce; Fruiting Body; Jack Haggerty Forest	HM363158
	WIN(M) 1583 = LJH 26A	Black Spruce; Adult Washing; Jack Haggerty Forest	= HM363158
	WIN(M) 1599 = LJH 8	Tamarack; Fruiting Body; Jack Haggerty Forest	HM363159

	WIN(M) 1600 = LJH 9	Tamarack; Beetle Frass; Jack Haggerty Forest	= HM363159
	WIN(M) 1632 = LJH 24	Black Spruce; Fruiting Body; Jack Haggerty Forest	= HM363159
	WIN(M) 1633 = LJH 25B	Black Spruce; Adult Washing; Jack Haggerty Forest	= HM363158
<i>Leptographium wingfieldii</i> M. Morelet	WIN(M) 1586 = LJH 36	Jack Pine; Fruiting Body; Silver Mountain	HM363160
	WIN(M) 1587 = LJH 37	Jack Pine; Stained Wood; Silver Mountain	= HM363160
<i>Ophiostoma abietinum</i> Marm. & Butin	WIN(M) 1579 = LJH 63	White Spruce; Adult Washing; Jack Haggerty Forest	HM363169
	WIN(M) 1585 = LJH 34	Jack Pine; Adult Washing; Silver Mountain	= HM363169
	WIN(M) 1596 = LJH 5	Tamarack; Beetle Frass; Jack Haggerty Forest	HM363163
	WIN(M) 1597 = LJH 6	Tamarack; Fruiting Body; Jack Haggerty Forest	= HM363163
	WIN(M) 1636 = LJH 39	Jack Pine; Beetle Frass; Silver Mountain	= HM363169
<i>Ophiostoma allantosporum</i> (H.D. Griffin) M. Villarreal	WIN(M) 1592 = LJH 1	Tamarack; Adult Washing; Jack Haggerty Forest	HM363162
	WIN(M) 1593 = LJH 2	Tamarack; Adult Washing; Jack Haggerty Forest	= HM363162
	WIN(M) 1634 = LJH 26B	Black Spruce; Adult Washing; Jack Haggerty Forest	HM363168
	WIN(M) 1638 = LJH 41	Jack Pine; Stained Wood; Silver Mountain	HM363170
<i>Ophiostoma flexuosum</i> H. Solheim	WIN(M) 1582 = LJH 65	White Spruce; Stained Wood; Quackenbush Woodlot	HM363171
	WIN(M) 1594 = LJH 3	Tamarack; Fruiting Body; Jack Haggerty Forest	= HM363171
<i>Ophiostoma floccosum</i> Math.-Käärik	WIN(M) 1639 = LJH 42	Jack Pine; Adult Washing; Jack Haggerty Forest	HM363172
	WIN(M) 1642 = LJH 47	Jack Pine; Adult Washing; Quackenbush Woodlot	HM363173
<i>Ophiostoma ips</i> (Rumbold) Nannf.	WIN(M) 1598 = LJH 7	Tamarack; Beetle Frass; Jack Haggerty Forest	HM363174

<i>Ophiostoma montium</i> (Rumbold) Arx	WIN(M) 1635 = LJH 30	Jack Pine; Stained Wood; Silver Mountain	HM363175
<i>Ophiostoma pulvinisporum</i> X.D. Zhou & M.J. Wingf.	WIN(M) 1584 = LJH 27	Black Spruce; Stained Wood; Jack Haggerty Forest	HM363176
	WIN(M) 1623 = LJH 49	Jack Pine; Stained Wood; Jack Haggerty Forest	= HM363176
<i>Ophiostoma rectangulosporium</i> Ohtaka, Masuya & Yamaoka	WIN(M) 1602 = LJH 11	Balsam Fir; Larval Washing; Jack Haggerty Forest	HM363177
	WIN(M) 1603 = LJH 12	Balsam Fir; Fruiting Body; Jack Haggerty Forest	HM363164
	WIN(M) 1629 = LJH 51	Jack Pine; Stained Wood; Quackenbush Woodlot	= HM363164
<i>Pesotum</i> sp. J.L. Crane & Schokn.	WIN(M) 1590 = LJH 57	Jack Pine; Adult Washing; Jack Haggerty Forest	HM363161
	WIN(M) 1619 = LJH 31	Jack Pine; Adult Washing; Silver Mountain	HM363165
	WIN(M) 1620 = LJH 32	Jack Pine; Beetle Frass; Silver Mountain	HM363166
	WIN(M) 1630 = LJH 21	Balsam Fir; Fruiting Body; Jack Haggerty Forest	HM363167
<i>Thysanophora penicillioides</i> <sup>1</sup> (Roum.) W.B. Kendr.	WIN(M) 1581 = LJH 64	NA	HM363178

<sup>a</sup> <sup>1</sup> identity based on BLAST (megaBLAST; <http://www.ncbi.nlm.nih.gov/BLAST/>; Altschul et al. 1990)

<sup>b</sup> <sup>1</sup> WIN(M) = University of Manitoba (Winnipeg) Collection, Canada; <sup>2</sup> LJH = Leonard J. Hutchison collection, Lakehead University, Ontario, Canada; <sup>3</sup> CBS = Centraal Bureau voor Schimmelcultures, Utrecht, The Netherlands

<sup>c</sup> <sup>4</sup> NA = no data available

obtained by analyzing nucleotide sequence alignments with the DNAPARS (DNA parsimony algorithm) program. Bootstrap analysis (SEQBOOT) was used to provide an estimate of the confidence levels for the major nodes present within the phylogenetic trees and CONSENSE was used to generate a majority rule consensus tree (Felsenstein 1985). Phylogenetic estimates were also generated within PHYLIP using the NEIGHBOR program, using distance matrices generated by DNADIST (K84 setting).

The MrBayes program was used for Bayesian analysis and the parameters were as follows: the GTR model with gamma distribution with 4 gamma rate. The Bayesian inference of phylogenies was initiated from a random starting tree and four chains were run simultaneously for 5 000 000 generations; trees were sampled every 1000 generations. The first 25 % of trees generated were discarded ("burn-in") and the remaining trees were used to compute the posterior probability values.

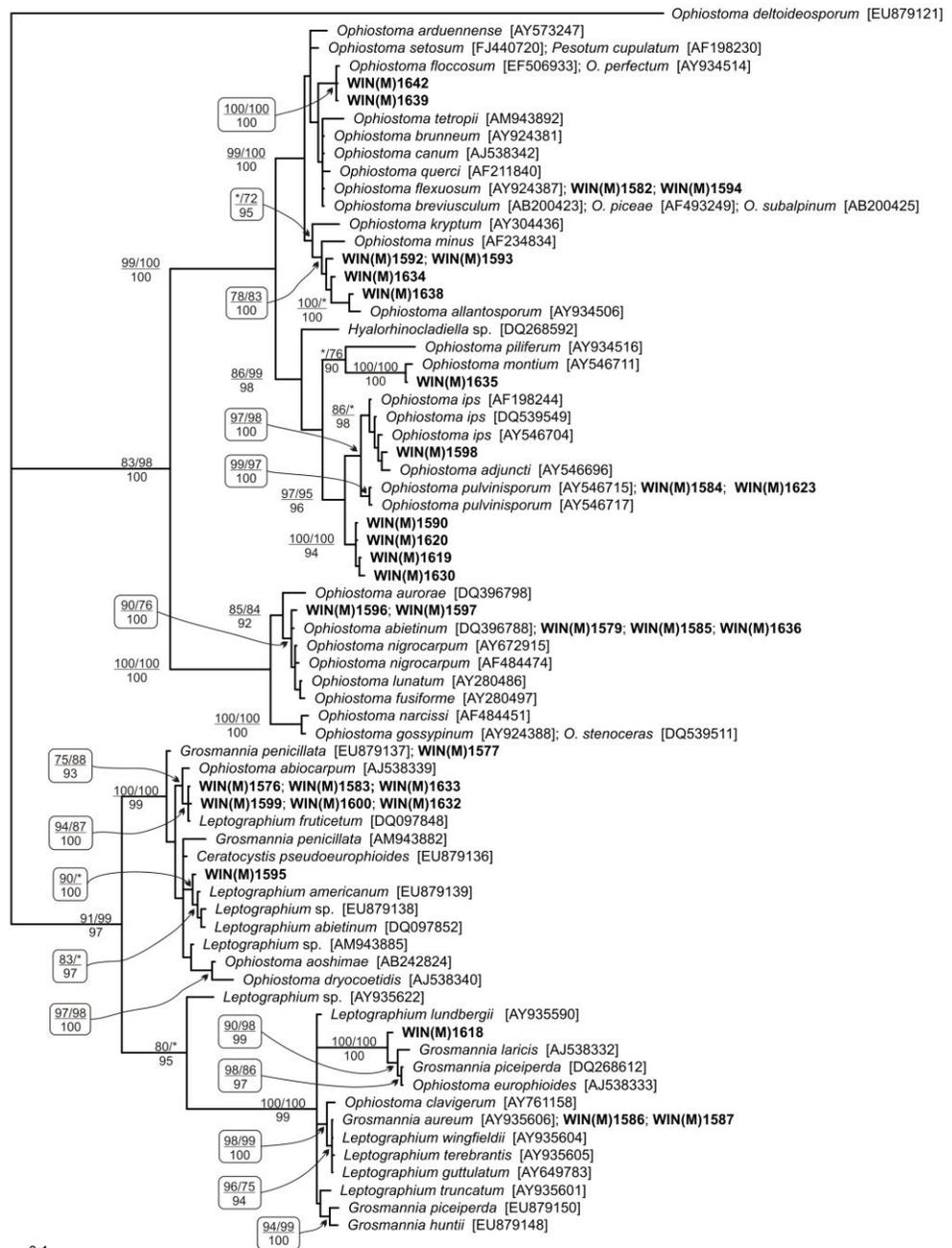
### **3.3 RESULTS AND DISCUSSION**

The ITS region was sequenced for 44 strains from the Lakehead collection and for one strain from the Hausner Research Group (WIN(M) 1523; Table 3.1) and 99 sequences were obtained from GenBank. The ITS sequences were too divergent to be treated as one large dataset, therefore the ITS data was separated into four ITS alignments referred to as: (1) the *Ophiostoma* set, (2) the *Pesotum*-like set, (3) the *Ceratocystiopsis* set, and (4) the *Graphium* set.

The majority of strains collected near Thunder Bay appear to be allied to either

*Ophiostoma* or *Grosmannia* (Figure 3.1). In culture these strains produced *Sporothrix*-like, *Pesotum*-like and *Leptographium*-type conidial states. Some strains could potentially be new taxa, as no significant matching GenBank entries could be identified [see WIN(M) 1577; or 1590, 1619, 1620, 1630; or 1595; or 1618] while others, based on the ITS data and isolation data, could be potentially assigned to a species. For example, WIN(M) 1579, 1585, 1596, 1597, and 1636 with *Sporothrix*-like anamorphs appear to be *O. abietinum*, and WIN(M) 1576, 1583, 1599, 1600, 1632 and 1633 match *L. fruticetum*. *O. abietinum* is a common species that has been isolated from several conifer species and bark beetles and from various locations such as Canada, USA, Mexico, New Zealand and South Africa (Marmolejo and Butin 1990; De Beer et al. 2003; Zhou et al. 2006). *L. fruticetum* represents a recently described species isolated from spruce-infesting bark beetle *Ips perturbatus* (northern spruce engraver beetle) collected from felled spruce trees and logs in northern British Columbia and Yukon Territory (Massoumi Alamouti et al. 2006). This study thus extends the range of *L. fruticetum* into NW Ontario. Another interesting match are the ITS sequences of WIN(M) 1586 and 1587 with those of *L. wingfieldii* and *Grosmannia aurea*. Based on a previous study and morphology (see Hausner et al. 2005) these strains most likely represent *L. wingfieldii*, a species that appears to have been introduced into Canada recently along with its bark beetle vector *Tomicus piniperda*, probably from Europe (see Hausner et al. 2005; Jacobs et al. 2004).

Strains WIN(M) 1590, 1619, 1620 and 1630, as stated above, could not be tied to a specific taxon but they were originally described as *Graphium*-like. This identification

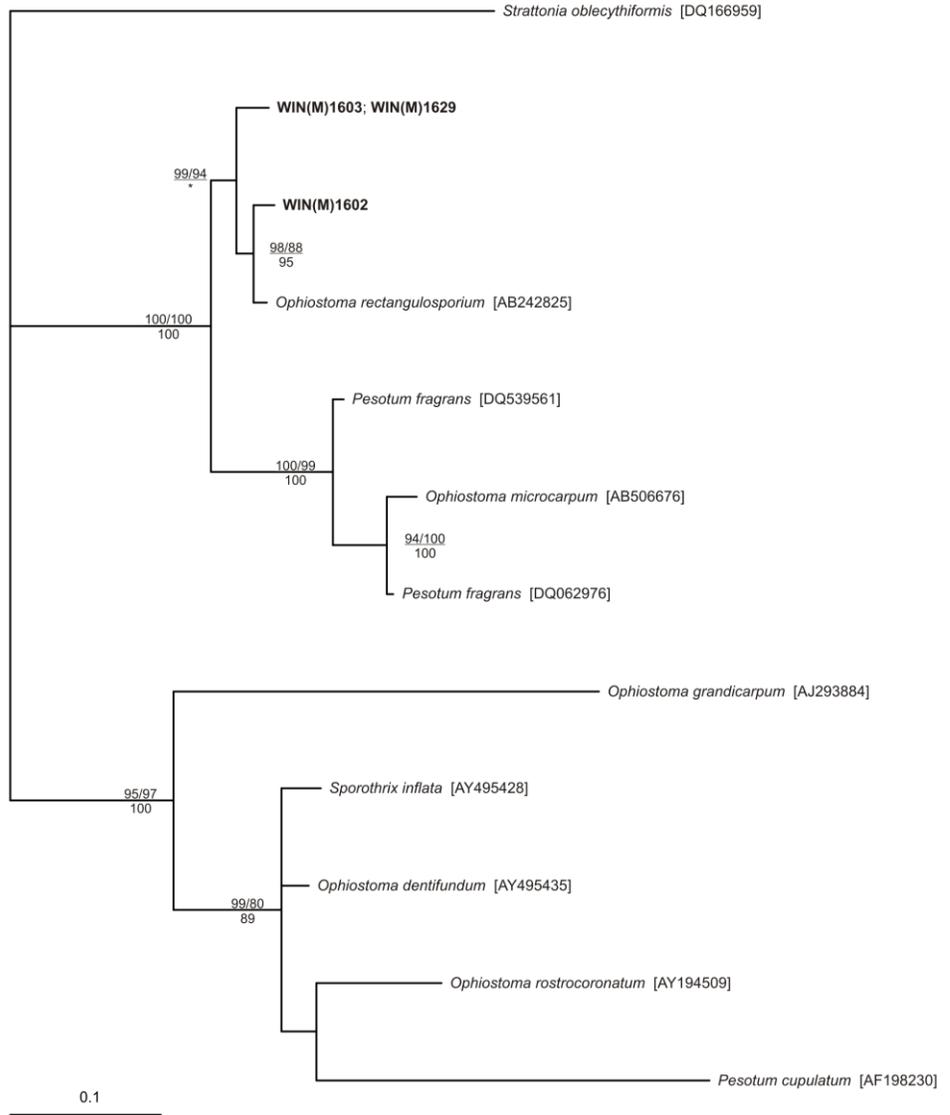


0.1

**Figure 3.1.** Phylogenetic relationships among “the *Ophiostoma* set” of fungi using ITS (ITS1, 5.8S, ITS2) nucleotide sequences. Numbers above the branches are based on bootstrap support values for neighbor joining and parsimony; numbers below are posterior probability values. Only levels of support  $\geq 70\%$  are labeled, \* indicates the node was either below 70% or collapsed in the consensus. *O. deltoideosporum* is the outgroup and accession numbers [ ] are provided for those sequences obtained by megaBLAST searches.

is doubtful since *Graphium* species are not closely related to *Ophiostoma* or *Grosmannia*; thus, these species are better classified as *Pesotum* sp. Strains WIN(M) 1584 and 1623 yield ITS sequences that match those of *O. pulvinisporum*, a species first described from Mexico (Zhou et al. 2004). Strain WIN(M) 1598 appears to be related to *O. ips*, a very commonly encountered blue-stain fungus in North America (Upadyay 1981). Strain WIN(M) 1635 is allied to *O. montium*, a species that is difficult to distinguish morphologically from *O. ips* but can be distinguished based on rDNA data (Hausner et al. 1993). Two strains, WIN(M) 1582 and 1594, had *Pesotum*-like conidial states that match *O. flexuosum*, a species that is usually associated with a *Sporothrix* state (Solheim 1986). Based on the ITS data other interesting connections that have been established are for strains WIN(M) 1639 and 1642 with *O. perfectum* or *O. floccosum*, for strains WIN(M) 1592, 1593, 1634, 1638 with *O. allantosporum* and for WIN(M) 1598 with *O. ips*. However, at this stage due to a lack of cultural and morphological data these connections cannot be used to definitively assign species names to the strains listed but the data may provide strong clues as to what these strains should be compared with in the future following more intensive cultural methods to provide morphological features and more detailed sequence analysis (such as sequencing additional variable loci).

A small subset of Lakehead *Sporothrix*-like strains and *Ophiostoma* ITS sequences were highly divergent from the other ITS sequences collected in this study and were thus analyzed within their own data set (Figure 3.2). Three strains WIN(M) 1602, 1603 and 1629 appear to be allied to *O. rectangulosporium*. This species was previously

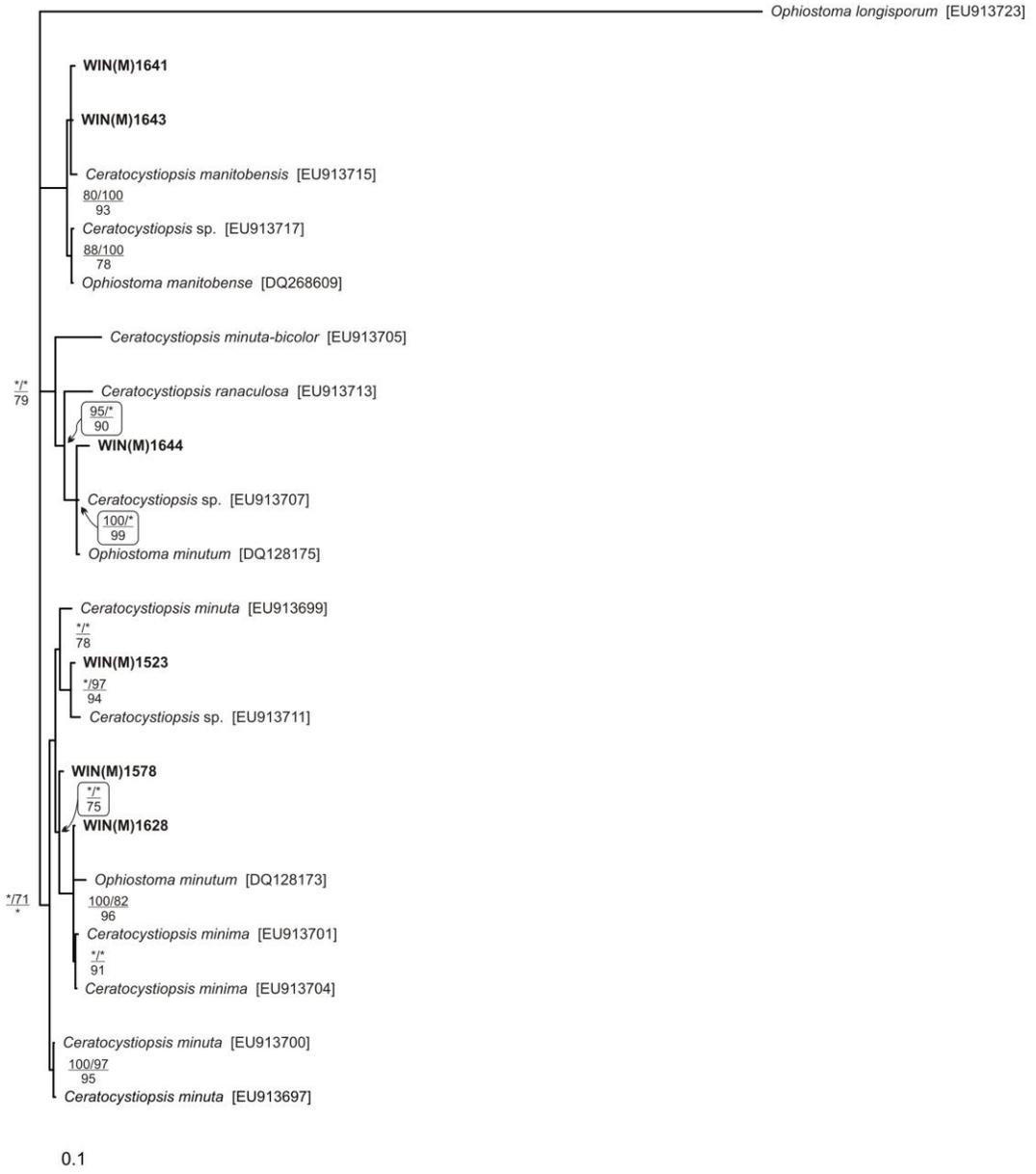


**Figure 3.2.** Phylogenetic relationships among “the *Pesotum*-like set” of fungi using ITS (ITS1, 5.8S, ITS2) nucleotide sequences. Numbers above the branches are based on bootstrap support values for neighbor joining and parsimony; numbers below are posterior probability values. Only levels of support  $\geq 70\%$  are labeled, \* indicates the node was either below 70% or collapsed in the consensus. *Strattonia oblecythiformis* is the outgroup and accession numbers [ ] are provided for those sequences obtained by megaBLAST searches.

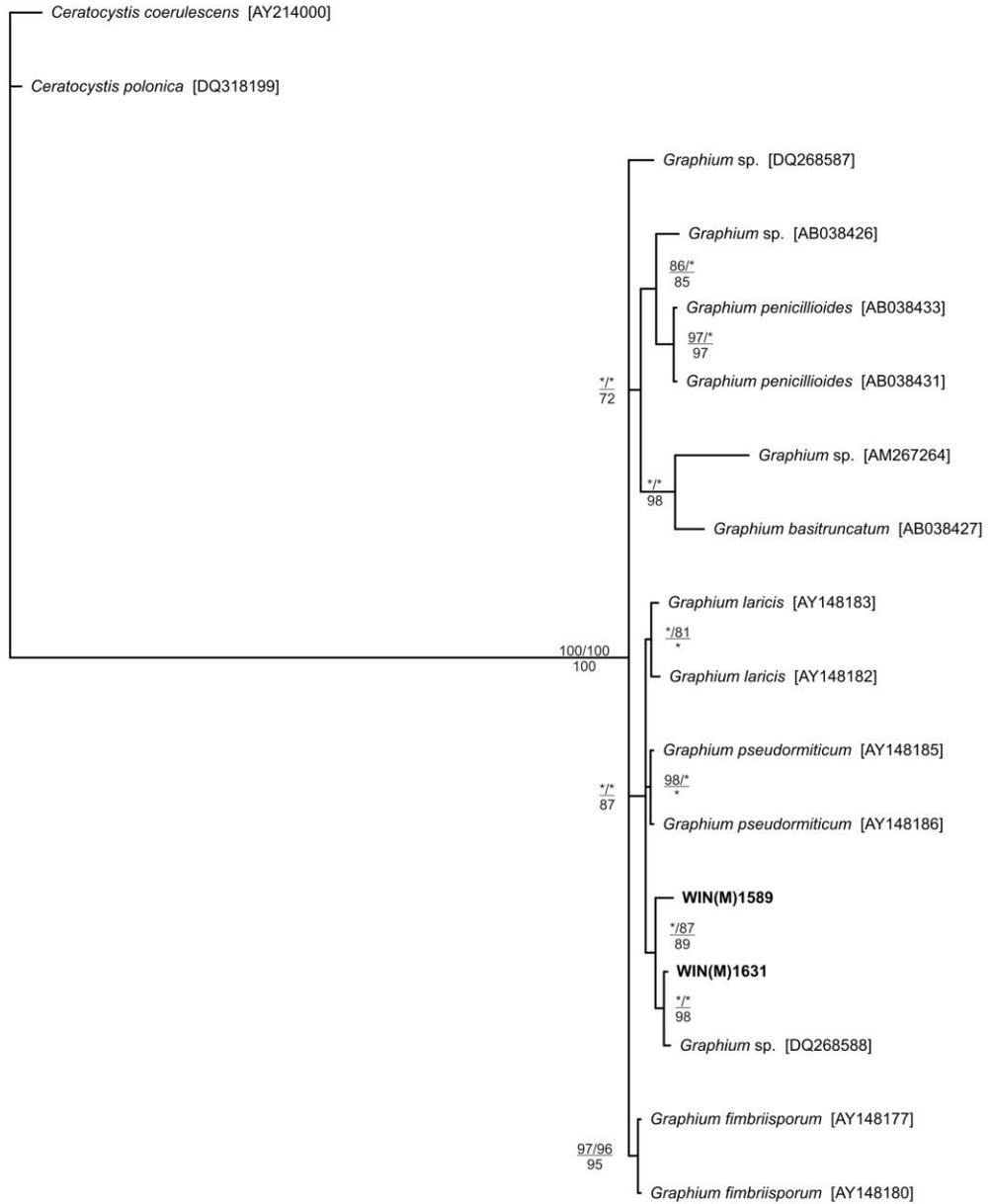
described from Japan and has been noted to be unusual as it did not produce a conidial state (Ohtaka et al. 2006).

Five strains from the Lakehead collection could be assigned to the *Ceratocystiopsis* ITS data set (Figure 3.3). Two isolates, WIN(M) 1641 and 1643, appear to match the ITS sequence of *Ceratocystiopsis manitobensis*, a species previously reported from Manitoba. The WIN(M) strain 1628 appears to match the ITS sequence for *Ceratocystiopsis minima*, another species that has previously been reported from Manitoba (Olchowecki and Reid 1974). The Genbank strain recorded under the name *Ophiostoma minutum* (DQ128173) has most likely been misidentified and should be assigned to *Ceratocystiopsis minima* (see Plattner et al. 2009 on misidentified strains of putative *O. minutum* and *Ceratocystiopsis minuta* strains). The ITS of strain WIN(M) 1523 (CBS 117042) was sequenced as a representative of *Ceratocystiopsis minuta* (Reid and Hausner 2010). Two Lakehead strains [WIN(M) 1644 and 1578], which produced only *Hyalorhinocladia*-like conidial states, have ITS sequences allied to species of *Ceratocystiopsis*. This finding is not surprising since species of *Ceratocystiopsis* usually have this type of conidial state.

Within the Lakehead collection two strains, WIN(M) 1589 and 1631, based on the ITS region could be assigned to the genus *Graphium* (Figure 3.4). Specifically, the two strains appear to be allied to an undescribed *Graphium* species that was reported as an associate of the northern spruce engraver beetle *Ips perturbatus* in western Canada (Massoumi Alamouti et al. 2007).



**Figure 3.3.** Phylogenetic relationships among “the *Ceratocystiopsis* set” of fungi using ITS (ITS1, 5.8S, ITS2) nucleotide sequences. Numbers above the branches are based on bootstrap support values for neighbor joining and parsimony; numbers below are posterior probability values. Only levels of support  $\geq 70\%$  are labeled, \* indicates the node was either below 70% or collapsed in the consensus. *Ophiostoma longisporum* is the outgroup and accession numbers [ ] are provided for those sequences obtained by megaBLAST searches.



0.1

**Figure 3.4.** Phylogenetic relationships among “the *Graphium* set” of fungi using ITS (ITS1, 5.8S, ITS2) nucleotide sequences. Numbers above the branches are based on bootstrap support values for neighbor joining and parsimony; numbers below are posterior probability values. Only levels of support  $\geq 70\%$  are labeled, \* indicates the node was either below 70% or collapsed in the consensus. *Ceratocystis coerulescens* is the outgroup and accession numbers [ ] are provided for those sequences obtained by megaBLAST searches.

This study demonstrates that ITS sequences are useful criteria in assigning unknown strains of fungi that lack sexual features, and in some cases poorly defined asexual features, to potential taxonomic groupings. In this work fungi associated with bark beetles and/or bark beetle galleries could be assigned to *Ophiostoma*, *Ceratocystiopsis* and *Grosmannia*. In some cases the data provided evidence to allow for species designation, while in other instances species could be identified that could be closely related to strains collected in the Thunder Bay area. This study may provide evidence for new records for certain species in NW Ontario (such as the presence of *L. fruticetum*) and provides evidence that the range of the exotic (introduced) fungus *L. wingfieldii* has spread from Central Ontario (Hausner et al. 2005) to NW Ontario, suggesting that this fungus may already be present in Eastern Manitoba. The data also helped in differentiating *Pesotum* from *Graphium* species, genera that based on morphology are sometimes difficult to separate. Overall, this study suggests that there are many different species of ophiostomatoid fungi present in NW Ontario. As these fungi are of economic significance, due to their blue-stain ability and in some instances pathogenicity, this study suggests that more detailed surveys of ophiostomatoid fungi in NW Ontario are warranted and should be accompanied with molecular work.

## Chapter 4: Additional complexity at the mitochondrial *rps3* locus within species of *Grosmannia*

### ABSTRACT

In the mitochondrial *rnl* of some filamentous ascomycetes fungi a group I intron (mL2449) encodes for the *rps3* gene. In *Grosmannia piceiperda* and *G. laricis* the N-terminal region of the intron-encoded *rps3* open reading frame has been invaded by an IC2-type group I intron (Gpi.m.Rps3 i1). This IC2-type intron disrupts the recipient *rps3* and fragments it into two open reading frames. The Gpi.m.Rps3 i1 intron encodes a putative double motif LAGLIDADG open reading frame, which is fused in-frame to the upstream *rps3* exon sequence. Reverse transcriptase PCR confirmed that this intron is spliced *in vivo* and as a result this could allow for the expression of a functional Rps3 protein.

## 4.1 INTRODUCTION

The mitochondrial (mt) genomes of fungi contain numerous types of mobile genetic elements, including group I introns, group II introns and homing endonuclease genes (HEG). Group I and group II introns are typically spliced from transcripts, which minimizes the chances of damaging the gene they are located within. Many group I introns contain open reading frames (ORFs) for intron-encoded proteins (IEP) such as homing endonucleases (HE) or maturases. HEs are enzymes important for initiating intron mobility by making a double strand break (DSB) at a specific locus; this activates the cellular DSB-repair pathway. Maturase proteins assist the intron with folding into a ribozyme (Caprara and Waring 2005). There are five families of HEGs, categorized according to conserved amino acid motifs: LAGLIDADG, GIY-YIG, His-Cys box, HNH, and PD-(D/E)<sub>x</sub>K (Stoddard 2005; Zhao et al. 2007; Stoddard 2011); only the first two HEGs are found in the mtDNA of fungi. Group II intron ORFs can encode LAGLIDADG HEs or, more frequently, reverse transcriptase (RT)-like proteins (Stoddard 2005; Lambowitz and Zimmerly 2004). These RT elements mobilize their intron counterparts by reverse splicing them into new target sites in the host's genome. In addition to encoding different proteins, group I and group II introns have distinct secondary structures and splicing mechanisms that facilitate their removal from host transcripts. HEGs have also been shown to be mobile elements that can move independently from their ribozyme counterparts (Mota and Collins 1988).

The *rnl*-U11 region (U11 = eleventh universally conserved region; Cummings et al. 1989) within the ophiostomatoid fungi contains a group I intron (subtype IA1) known as mL2449, which contains an unusual ORF encoding for the ribosomal protein S3 (Rps3; Burke and RajBhandary 1982; Gibb and Hausner 2005; Sethuraman et al. 2009a, 2009b). A detailed survey among 85 species of *Ophiostoma*, *Grosmannia* and related genera showed that the intron-encoded *rps3* ORF has been invaded by HEGs at three different positions. These insertions are referred to as A, B, and C type respective to their order of insertion (Sethuraman et al. 2009a). The B and C type insertions are common in this group of fungi but only two examples of the A type were noted, in strains of *G. piceiperda* and *G. laricis*.

The B and C type HEGs have inserted near the C-terminus of the *rps3* ORF, displacing the remaining *rps3* sequence downstream of the HEG insertion. These HEGs are inserted in-frame with the host gene and contain terminal stop codons, effectively creating a mono-ORFic hybrid or fusion gene. It was speculated that these fusion proteins are resolved by proteolytic activity to generate functional copies of the Rps3 protein and of the HE (Gibb and Hausner 2005; Sethuraman et al. 2009a). On the other hand, the A type HEG inserts near the N-terminal of *rps3* generating a bi-ORFic complex (two putative ORFs). The first ORF is due to the stop codon introduced by the A type HEG, while the second stop codon is already present in the *rps3* ORF (Sethuraman et al. 2009a). The introduction of this HEG stop codon truncates the *rps3* coding region, rendering it into a pseudogene. It was speculated that a functional Rps3 protein could be

assembled by expressing the two ORFs separately using an alternative start codon followed by interaction of the N- and C-terminal segments, or possibly by an alternative splicing event to generate a continuous *rps3* transcript. The objective of the work is to re-examine the A type HEG insertions discovered by Sethuraman et al. (2009a), gaining a better understanding of *rps3* expression in *G. piceiperda* and also to attempt to express a recombinant form of the HE.

## **4.2 SUMMARY OF METHODS**

For the following methods see Chapter 2: fungal culturing, DNA extraction, PCR, cloning, sequencing, RNA extraction, RT-PCR, RNA folding and bioinformatics.

### **4.2.1 Phylogenetic Analysis**

Only those segments of the multiple sequence alignment where all sequences could be aligned unambiguously were retained for phylogenetic analyses. Phylogenetic estimates were generated by the programs contained within the PHYLIP package (Felsenstein 1985; 2006) and the MrBayes program v3.1 (Ronquist and Huelsenbeck 2003; Ronquist 2004). In PHYLIP (version 3.69c), phylogenetic trees were obtained by analyzing alignments with PROTPARS (protein parsimony algorithm) for amino acid sequences. Bootstrap analysis (SEQBOOT) was used to provide an estimate of the confidence levels for the major nodes present within the phylogenetic trees and CONSENSE was used to generate a majority rule consensus tree (Felsenstein 1985).

Phylogenetic estimates were also generated within PHYLIP using the NEIGHBOR program, using distance matrices generated by PROTDIST (setting: Dayhoff PAM250 substitution matrix; Dayhoff et al. 1978).

The MrBayes program was used for Bayesian analysis and the parameters for amino acid alignments were as follows: mixed model and gamma distribution with 4 gamma rate parameters. The Bayesian inference of phylogenies was initiated from a random starting tree and four chains were run simultaneously for 5 000 000 generations; trees were sampled every 1000 generations. The first 25 % of trees generated were discarded ("burn-in") and the remaining trees were used to compute the posterior probability values.

#### **4.2.2 Gene Synthesis of HEG**

Due to differences in fungal mtDNA and *E. coli* DNA the A type HEG was synthesized to achieve an optimal genetic code (Bio S&T Inc., Montreal, PQ, Canada; Figure 4.1). For example, the tryptophan codon UGA in fungal mtDNA is a stop codon in *E. coli*.

#### **4.2.3 Protein Expression**

The synthesized A type HEG was amplified and cloned into the pET200/D-TOPO vector (Invitrogen) following the manufacturer's directions. To express the recombinant A type HEG, an overnight culture was diluted 1:10 with 50 ml Luria-Bertani broth (LB)

*Grosmannia piceiperda* mtDNA LAGLIDADG HEG sequence

TGATGATTTGGGAAGTCATCTTTTATATGGGATAAACTATCAAATTCGGGGAACGTCCTA  
AAGTCATTGGTACCAAGCAATACCCGAAAGGTTATTTGTGGATGAAGTAATTACTCATGT  
ATGGTAATAAGTCAATGAATGAACGAAAGTGAAATGGATAATCGCGGATCTAAATCAGTA  
TTGTTTAATAACCTTACTGTAAAAGAGCAACGAGTAAAAGGTAGTTATTGTATAAATGGT  
ATACAATTAAGATGTACTCTAGTGGGTTTCGAAAGAAATTATCAAGTCAAATCCCATCT  
AATCAAATAATAAATAAATCAATAAGATATATCTCAAATTCAGCTACGGTTACTCCTTTA  
ATAGATCCTTGATTTATAACAGGATTTGCTGATGCTGAAAGTTCCTTTGTAGTTTCAATT  
AAGAGAAAATAAAAAAATCAAATGTGGTTGAAATGTAGTAACTAGATTCCAAATAGCTTTA  
TCTCAAAGGATTTAGCTTTATTAGAACGTATTAAGGCTATTTCAAAGATGCCGGGAAT  
ATTTATATAAAAAGCGATAAAGTGTGCGTTGATTGACACGTAACCTCGGTAAGGATTTA  
AAGATAATCTTAGACCATTTTGATAAATACCTTTAAAACTGAGAAATTAGCTGATTAC  
ATACTGTTTAAAGAAGTGTTTAATATAATTTTAACTAAACAACATCTAACAGTTGAAGGT  
ATACAGAAAATTGTAGCAATTAGAGCATCAATAAATAAAGGTTTATATGGTGAATTGAAA  
GCTGCTTTTCCGAATATTATTCCCTGTTCAAAGACCTAAAATTGATGATAGATTTATTATC  
GACATCCAACCTTGATGAGTAGCAGGTTTACTGAAGGGGAAGGATGTTTTAGTGTGTG  
GTTACGAATTCGCCTTCTACTAAAAGTGGATTTTCGGCAAGTTTGATTTTTCAAATAACT  
CAACATTCAGGGATATTGTATTAATGCAAATATAATAAAATTTTTAGGTTGTGGTAGA  
ATACATAAGAGATCTAAGGAGGAAGCTGTCGATATTTTAGTAACTAAATTTTCAGATTTG  
ACTGAGAAAGTTATCCCGTTTTTTGAAAGTATACCTTTGCAAGGTTTAAACTTAAAAAC  
TTTACGGATTTCTCTAAAGCGGCTGATATAATAAAGTTAAAGGACACTTAACTCCGAAA  
GGTTTAGATAAAAATATTACAAATAAAATTAGGAATGAACACAAGAAGAATTTAA

Optimized HEG sequence for recombinant protein expression in *E. coli*

TGGTGGTTTGGTAAATCTAGCTTCATCTGGGATAAACTGAGCAATAGCGGCAACGTTCTG  
AAATCTCTGGTTCCGTCTAACACGCGTAAAGTAATCTGTGGTTGGTCTAACTATAGCTGC  
ATGGTTATCAGCCAGTGGATGAATGAGAGCGAGATGGACAATCGTGGTAGCAAATCCGTA  
CTGTTTAAACAACCTGACCGTTAAGGAACAGCGTGTGAAGGGCTCTTACTGCATCAACGGT  
ATCCAGCTGCGTTGTACCCTGGTTGGCTTTGAACGCAACTACCAGGTGAAGATCCCGAGC  
AACCAGATCATTAAACAAGTCTATTCGTTATATTTCCAACCTCTGCAACCGTTACTCCGCTG  
ATCGACCCGTGGTTCATTACCGGTTTCGCGGACGCCGAGAGCTCCTTTGTAGTATCCATC  
AAACGTAACAAAAAGATTAAGTGCGGTTGGAACGTGGTACTCGCTTCCAGATTGCGCTG  
TCCCAGAAAGATCTGGCCCTGCTGGAACGCATTAATCCTACTTCAAGGACGCTGGTAAC  
ATCTACATCAAATCCGACAAAGTATCTGTTGATTGGCACGTTACCTCTGTGAAAGACCTG  
AAAATCATTTCTGGACCATTTTCGATAAGTACCCGCTGAAAACCGAAAAACTGGCGGACTAC  
ATCCTGTTCAAAGAAGTTTTCAACATTATCCTGACCAAACAACATCTGACTGTCGAAGGC  
ATCCAGAAAATCGTGGCAATCCGCGCGTCTATCAACAAGGGTCTGTATGGTGAAGTAAA  
GCTGCTTTCCCAAATATTATTCCAGTGCAACGTCGAAAATTGATGACCGTTTCATCATC  
GATATCCAACCGTGGTGGGTAGCAGGCTTCACCGAAGGTGAGGGCTGTTTTAGCGTTGTT  
GTGACTAACTCCCCGTCCACTAAATCTGGTTTTAGCGCGTCCCTGATTTTTTCAGATCACT  
CAGCACTCCCGGATATTGTCTGATGCAGAACATCATCAAATTCCTGGGCTGCGGTTCG  
ATCCACAAACGTTCTAAAGAAGAAGCAGTCGACATCCTGGTCACGAAATTCTCCGATCTG  
ACGGAAAAAGTTATCCCGTTCTTCGAAAGCATCCCGCTGCAGGGCCTGAAACTGAAAAAC  
TTTACCGATTTCTCTAAAGCCGCTGATATCATTAAAGTAAAGGCCACCTGACCCCTAAA  
GGTCTGGACAAAATCCTGCAGATCAAATGGGCATGAACACCCGTCGTATTTAA

**Figure 4.1.** The original homing endonuclease gene sequence from *Grosmannia piceiperda* and the modified version, which has been optimized for recombinant protein expression in *E. coli*.

supplemented with 100 µg/ml kanamycin and 1% glucose and grown at 28 °C to A<sub>600</sub> 0.5-0.8. Expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and growing the culture for an additional 2 hours.

#### **4.2.4 Protein Purification**

A mini-protein purification was performed by pelleting 5 ml of bacterial cells at 6 000 rpm for 2 min. Cell pellets were resuspended in lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole at pH 8.0) and lysozyme added to a final concentration of 1 mg/ml and incubated on ice for 30 min. Cells were lysed by vortexing and centrifuged at 13 000 rpm for 10 min to remove cell debris. Next, 20 µl of a 50 % slurry of nickel-nitrilotriacetic acid (Ni-NTA; Qiagen, Hilden) was added to the supernatant, mixed gently for 30 min at 4 °C and centrifuged at 3500 rpm for 10 sec. The pelleted resin was washed twice with 100 µl wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole at pH 8.0). The 6xHis-tagged protein was eluted three times from the column with 20 µl elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole at pH 8.0). Protein fractions and lysed whole cells were resolved by 12% sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE).

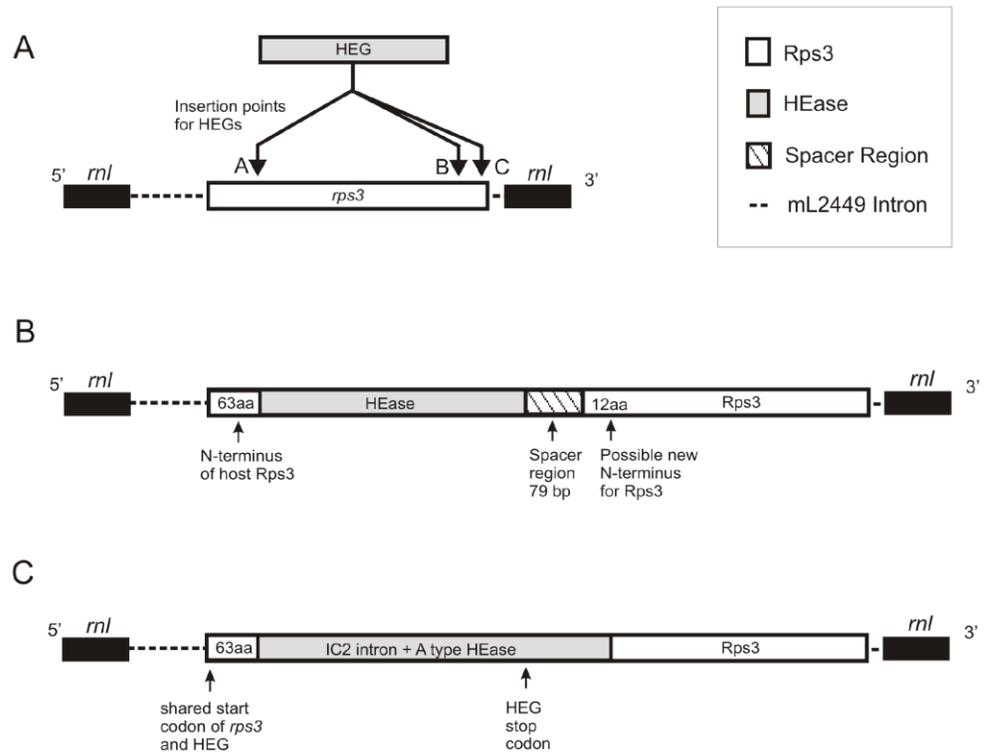
### **4.3 RESULTS**

#### **4.3.1 The IC2 Intron of *Grosmannia piceiperda***

The *rps3*-encoding mL2449 intron has been characterized by Gibb and Hausner

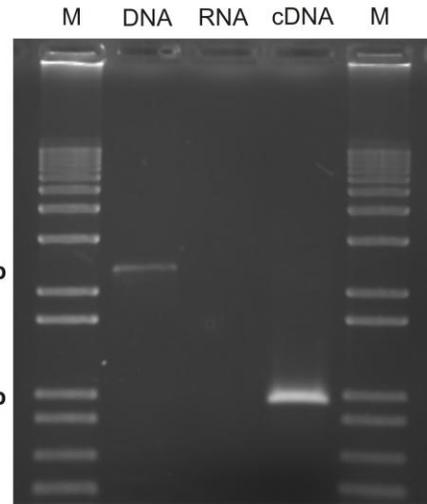
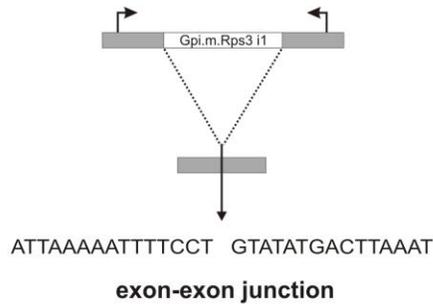
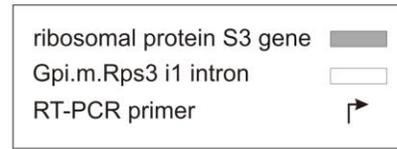
(2005) and Sethuraman et al. (2009a). This group I intron is of the subtype IA1 and encodes a protein important in ribosome function, known as Rps3. In some *Grosmannia* species, the *rps3* ORF contains a group I intron-encoded double motif LAGLIDADG HEG (group I subtype IC2; see Figure 1.1 in Chapter 1 and Figure 4.2). In addition to this LAGLIDADG HEG, two other types of HEGs were characterized by Sethuraman et al., designated the A, B and C type HEGs based on their site of insertion into *rps3*. It was originally thought that all three types of HEGs inserted independent of introns; however, upon reexamination of the *rps3* ORF of *G. piceiperda* with the RNAweasel program the IC2 group I intron was discovered (Lang et al. 2007). We sequenced a second strain of *G. piceiperda* WIN(M) 975 (also known as strain UAMH 9784) and noted that it yielded the same results as was previously obtained for WIN(M) 979 (UAMH 9787; Sethuraman et al. 2009a).

The *rps3* locus of *G. piceiperda* WIN(M) 975 was amplified under standard PCR conditions and resulted in an amplicon of around 2.4 kb. In contrast, when the same primers were used to generate complementary (c)DNA, this region was only found to be around 900 bp (Figure 4.3 A). This demonstrates that about 1.5 kb of the 2.4 kb *rps3* transcript is removed during maturation. In addition, *rps3* sequences of closely related species lacking the IC2 intron, such as *G. aurea* (strain CBS 438.69) and *G. huntii* [strain WIN(M) 492], were aligned to those of *G. piceiperda* to further confirm the exon-intron junction sites. This intron has been named Gpi.m.Rps3 i1, as it is the first group I intron inserted within the mitochondrial *rps3* of *G. piceiperda* (Figure 4.3 B).

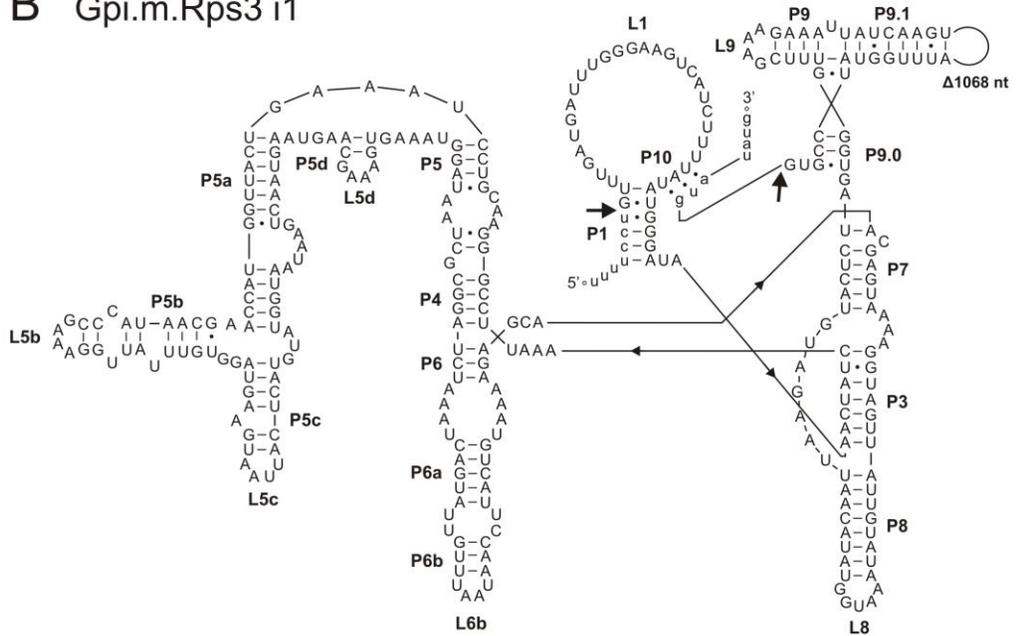


**Figure 4.2.** Schematic representations of the mitochondrial *rps3* gene in ophiostomatoid fungi. **(A)** Insertion sites of the A, B and C type homing endonuclease genes. **(B)** The A type homing endonuclease insertion in *Grosmannia piceiperda* and *G. laricis* as described by Sethuraman et al. (2009a). **(C)** The A type homing endonuclease insertion in *G. piceiperda* and *G. laricis* as described in this work.

A



B Gpi.m.Rps3 i1



**Figure 4.3.** The Gpi.m.Rps3 i1 intron of *Grosmannia piceiperda*. **(A)** RT-PCR on total RNA isolated from *G. piceiperda*. Diagram shows the primers (bent arrows) flanking the Gpi.m.Rps3 i1 intron (white box) in the *rps3* gene (grey box). Flanking *rps3* exon sequence is displayed beneath the schematic. Lanes of the agarose gel marked M contain 1-kb plus DNA marker. Positive control (DNA) contains genomic DNA template for PCR. Negative control (RNA) contains RNA template for PCR, to screen for the absence of DNA contamination. Lane marked cDNA is the intron-less cDNA amplicon of *rps3*. **(B)** Secondary structure of the Gpi.m.Rps3 i1 intron. Intron nucleotides (nt) are presented as uppercase letters, flanking exons are lowercase letters and large arrows show intron-exon junctions. P denotes paired regions, L denotes loops and  $\Delta 1068$  nt is the intron-encoded open reading frame (the double motif LAGLIDADG homing endonuclease gene) and is not shown.

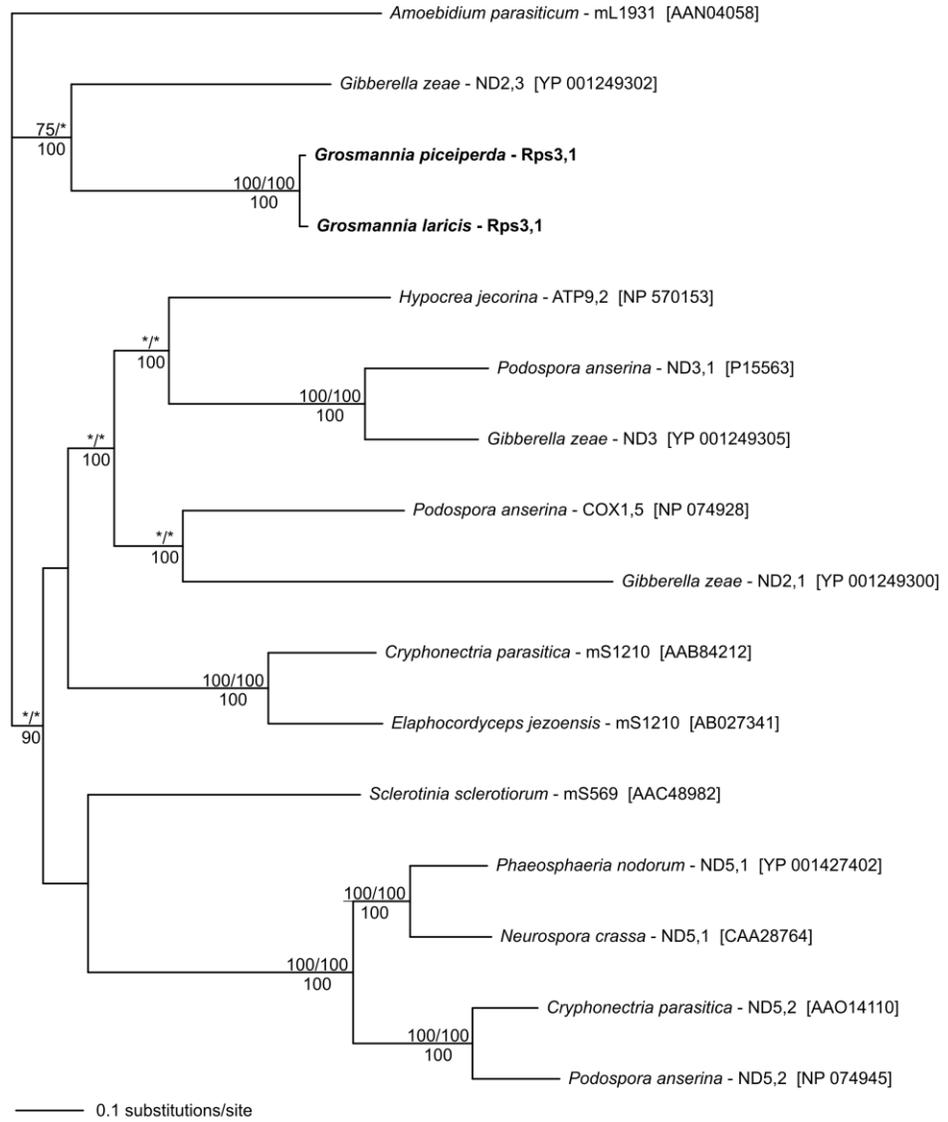
### **4.3.2 The LAGLIDADG Homing Endonuclease of *Grosmannia piceiperda* and *G. laricis***

Upon closer examination and re-sequencing of the *G. laricis* WIN(M) 1461 A type HEG we found that this ORF is not degenerate, as was previously suspected by Sethuraman et al (2009). Originally it was thought that a frameshift mutation generated a nonsense mutation; instead we now know that, with the exception of seven amino acids, the HEG sequence is identical to *G. piceiperda*. In spite of their similarity to one another the A type HEGs show low homology to other HE sequences found in Genbank. Nonetheless, it is interesting to note that all of the HEGs in Figure 4.4 are associated with IC2 type introns and most are within protein coding genes like *rps3*. The recombinant form of the A type HEG was expressed in *E. coli* but when attempts were made to purify the enzyme, it remained in the insoluble fraction (Figure 4.5).

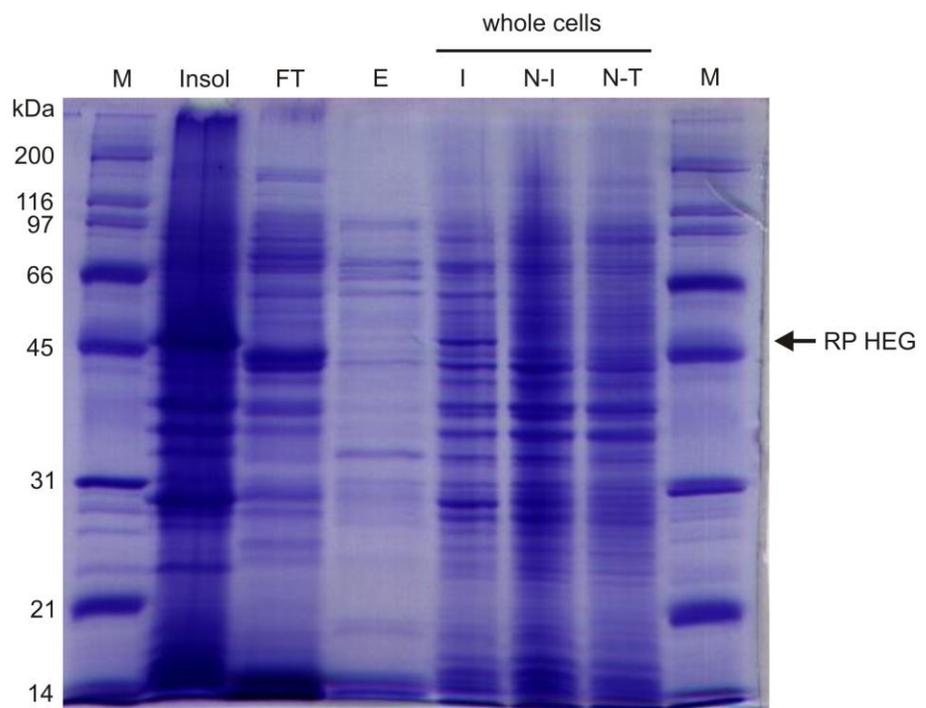
## **4.4 DISCUSSION**

### **4.4.1 Overlap of the IC2 Intron's Open Reading Frame and Core Sequence Suggests Coevolution of the Sequences**

IC2-type introns are very compact and their ORF sequences overlap with the majority of the intron core sequence. The 5' end of the A type HEG constructs nine (P1 to P9.1) of the ten intron paired regions and this large degree of overlap suggests coevolution of two individual mobile genetic elements (Nielson and Johansen 2009). Oftentimes ORFs are found in the looped regions of introns or genes to minimize



**Figure 4.4.** Phylogenetic relationships of double motif LAGLIDADG amino acid sequences in *Grosmannia* spp. (bold) and related sequences from Genbank (accession numbers in [ ]). *Amoebidium parasiticum* is the outgroup. Numbers above the branches are based on bootstrap support values for neighbor joining and parsimony; numbers below are posterior probability values. Only levels of support  $\geq 70\%$  are labeled, \* indicates the node was either below 70% or collapsed in the consensus.



**Figure 4.5.** SDS-PAGE gel showing the recombinant LAGLIDADG homing endonuclease (RP-HEG). Lanes marked M contain BioRad broad range protein ladder. Lanes containing insoluble protein (Insol), flowthrough (FT) and elution (E) fractions are marked according. The remaining lanes contain aliquots of cells induced with IPTG (I), cells not induced (N-I) and cells not transformed with the recombinant LAGLIDADG homing endonuclease gene (N-T). Gel stained with Coomassie Brilliant Blue.

interference with secondary and tertiary structure (Sethuraman et al. 2009a; Edgell et al. 2011). This feature also allows ORFs to insert into different types of introns and even to degenerate over time without affecting intron structure.

Since the IC2 intron is essentially the same as the ORF this could be a possible explanation for why all the associated introns in Figure 4.4 are of the IC2 subtype. Rather than the HEG carrying the necessary sequences to coexist with the intron, a recently proposed process, termed ‘core creep’, suggests the ORF mutates post-invasion; for instance, selection of an alternative initiation codon at *rps3* (Edgell et al. 2011). In addition, the compactness of the IC2 intron helps to explain why it was not initially found by Sethuraman et al. (2009a; Figure 4.2). Furthermore, the “spacer region” after the HEG stop codon is actually the IC2 intron and the entire unit was found to be spliced from the *rps3* transcript (Figure 4.3 A).

#### **4.4.2 Regulation of the IC2 Intron and Homing Endonuclease Gene**

Since the secondary and tertiary structures of introns are key for self-splicing from the host transcript, compromising these structures may be detrimental to the organism. Any circumstance that risks the integrity of the host gene, transcript or protein could be toxic to the organism (Edgell et al. 2011). *rnl* rRNA and Rps3 have been shown to be associated with ribosome structure and function (LaPolla and Lambowitz 1981). The overall complexity of the situation suggests that carefully coordinated events must occur in order to express each ORF and splice each intron from the master transcript, while

balancing simultaneous transcription and translation events in the fungal mitochondria (Sethuraman et al. 2009a; Gibb and Edgell 2010; Edgell et al. 2011).

Expression of the A type HEG must occur before the IC2 intron splices from *rps3*. Similarly, *rps3* must be expressed before the mL2449 intron splices and before the *rnl* transcript can mature. Further complications include intron secondary structure inhibiting translation of the ORFs and the resulting Rps3-HE fusion protein. Since the HEG is fused in-frame to the upstream *rps3* exon one would assume that the HE is translated as a fusion protein, thus benefiting from the *rps3* ORF's start codon and regulatory regions that permit translation (Gibb and Edgell 2010). The Rps3 N terminus is either proteolytically removed from the HE or it remains attached and does not alter the function of the HE (Sethuraman et al. 2009a; Gibb and Edgell 2010). The HE may have maturase activity that is needed to facilitate the efficient removal of the IC2 intron from the *rps3* precursor transcript, thus generating the *rps3* mRNA that can be translated to yield a functional Rps3 (Stoddard 2005; Edgell et al. 2011; Stoddard 2011). This scenario of stepwise splicing has been evoked to explain the expression of mtDNA genes interrupted by several introns that encode putative proteins whose reading frames are fused to the upstream exon regions (Grivell 1995). The insertion of the IC2 intron within the N terminus of the *rps3* ORF probably requires a cascade of splicing events that allow for the expression of the HE and for the Rps3 protein.

This study describes a LAGLIDADG HE and its IC2 group I intron, both of which have inserted into the *rps3* gene of some *Grosmannia* species (see Figure 1.1 in Chapter

1). Although the HE could not be purified for usage in endonuclease experiments, the IC2 intron was demonstrated by RT-PCR to splice from the *rps3* transcript (Figure 4.3). Introns and HEGs are generally considered to be neutral and non-toxic components of fungal mtDNA. Host genomes seem to cope well with these mobile elements, despite balancing regulatory events for gene expression and intron splicing.

## **Chapter 5: Evolutionary dynamics of introns and their open reading frames in a region of the mitochondrial *rnl* gene in species of *Ceratocystis***

This chapter this is a compilation of data generated by myself and members of the Hausner Research Group: J. Sethuraman and K. Wosnitza.

### **ABSTRACT**

A region of approximately 1 kb within the mtDNA *rnl* gene has been examined for the presence of potential optional elements within selected species of the genus *Ceratocystis*. Examples of single and double motif LAGLIDADG homing endonuclease genes were noted within group I and group II introns. A GIY-YIG type homing endonuclease gene was found situated within a LAGLIDADG type open reading frame and a second degenerate GIY-YIG open reading frame was fused to a sequence that is related to the mtDNA *nad2* gene in a group I intron. The great diversity of observed elements that appear to be randomly distributed among closely related species illustrates the potential for introns and their open reading frames to invade new locations and thus generating mtDNA genetic diversity. The analysis of intronic protein-encoding sequences revealed that LAGLIDADGs, which usually are associated with group I introns, were horizontally transferred at least four times into group II B1 type introns. Also, the analysis provides an example of how double motif LAGLIDADG open reading frames are derived from the duplication and fusion of a single motif type LAGLIDADG open reading

frames.

## 5.1 INTRODUCTION

In a previous study the mitochondrial (mt)DNA large subunit ribosomal RNA gene (*rnl*) within members of the genus *Ophiostoma* was examined for the presence of introns (Sethuraman et al. 2008). During that study it was noted that the *rnl*-U7 region (U7 = seventh universally conserved region; Cummings et al. 1989) can be polymorphic due to the presence or absence of a group I intron (IA1 type) corresponding to position L1699 of the *E. coli rnl* (Sethuraman et al. 2008). In contrast, the *rnl*-U11 region always contained a group I intron (mL2449) encoding the ribosomal protein S3 gene (*rps3*); in a few *Ophiostoma* species it was noted that the *rps3* gene was invaded by homing endonuclease genes (HEGs; Gibb and Hausner 2005; Sethuraman et al. 2009a, 2009b). In the current study, the *rnl*-U7 region for species of *Ceratocystis* is examined in order to assess if introns have invaded the mL1669 site or if other types of insertions have occurred that generate mtDNA polymorphisms. The genera *Ophiostoma* and *Ceratocystis* contain both insect-associated plant pathogens and wood-staining species. Although morphologically, members of these two genera are quite similar, molecular sequence data suggests that they are examples of convergent evolution and they do not share a recent common ancestor (Hausner et al. 1992; 1993).

Although group I and II introns are self-splicing, sometimes the process is assisted by intron-encoded maturases (Belfort et al. 2002; Belfort 2003; Caprara and Waring 2005); in some instances the intron-encoded proteins also act as homing endonucleases

(HEs), facilitating intron-mobility (Dujon 1989; Belfort et al. 2002; Stoddard 2005).

Some HEGs themselves are mobile elements that can move from an open reading frame (ORF)-containing intron to an "ORF-less" intron (Mota and Collins 1988; Sellem and Belcour 1997). Self-splicing introns provide the HEGs with a neutral location, thus reducing the chance of the HEG damaging the host genome (Goddard and Burt 1999; Schäfer 2003).

Group I and II introns are RNA elements that fold into characteristic secondary and tertiary structures which allows these introns to act as ribozymes and undergo self-splicing from the primary transcript (Burke 1989; Cech 1990; Lambowitz and Zimmerly 2004). Usually group I and II introns, due to their ability to self-splice, are cryptic and are thus considered to be silent or neutral genomic elements. However, because of the strong selection against non-functional gene products the loss of self-splicing ability could result in their elimination from the host genome. It is now known that along with intron-encoded maturases nuclear-encoded factors participate in the splicing of group I and II introns (Mohr et al. 2001).

HEs are classified into five families based on the presence of conserved amino acid sequence motifs. However, among the fungi the most frequently encountered group I intron-encoded maturases or endonucleases that belong to the LAGLIDADG and GIY-YIG families (Stoddard 2005). Group II introns typically encode reverse transcriptase-like proteins, although a few group II introns have been identified that encode LAGLIDADG-type proteins (Toor and Zimmerly 2002; Monteiro-Vitorello et al. 2009).

LAGLIDADG type endonucleases can either contain one or two of the conserved LAGLIDADG amino acid motifs. The double motif HEs act as monomers and single motif HEs function as homodimers (Gimble 2000; Guhan and Muniyappa 2003; Stoddard 2005). Double motif HEs may have originated from a gene duplication event of a single motif HEG followed by a fusion event (Lambowitz et al. 1999; Haugen and Bhattacharya 2004). The GIY-YIG type endonucleases form the second largest family of homing endonucleases; they are highly variable in their sequences and have been noted as intron-encoded proteins within fungal mtDNA, phage genomes, chloroplast genomes, eubacterial and archaeal genomes (Van Roey and Derbyshire 2005) but they have also been noted to be encoded by free-standing genes (Stoddard 2005).

Mitochondrial genomes of the filamentous ascomycete fungi exhibit remarkable variation in conformation, size, gene organization and arrangements (Hausner 2003). The presence or absence of introns and HEGs in particular can cause tremendous variability among gene arrangements and length heterogeneity of the mtDNA within members of the same species (Salvo et al. 1998). Understanding the plasticity and evolution of mtDNA gene arrangements requires the comparison of closely related fungal species; here we have chosen a segment of the *rnl* gene within species of *Ceratocystis*. This work shows the success of introns and their ORFs in colonizing mtDNA rDNA in *Ceratocystis* and the results provide examples of how these elements can be significant factors in shaping the mtDNA of fungi.

## **5.2 SUMMARY OF METHODS**

For the following methods see Chapter 2: fungal culturing, DNA extraction, PCR, cloning, sequencing, RNA extraction, RT-PCR, RNA folding and bioinformatics.

### **5.2.1 Phylogenetic Analysis**

Only those segments of the multiple sequence alignment where all sequences could be aligned unambiguously were retained for phylogenetic analyses. Phylogenetic estimates were generated by the programs contained within the PHYLIP package (Felsenstein 1985; 2006) and the MrBayes program v3.1 (Ronquist and Huelsenbeck 2003; Ronquist 2004). In PHYLIP (version 3.69c), phylogenetic trees were obtained by analyzing alignments with either PROTPARS (protein parsimony algorithm) or DNAPARS programs for amino acid or nucleotide sequences. Bootstrap analysis (SEQBOOT) was used to provide an estimate of the confidence levels for the major nodes present within the phylogenetic trees and CONSENSE was used to generate a majority rule consensus tree (Felsenstein 1985). Phylogenetic estimates were also generated within PHYLIP using the NEIGHBOR program, using distance matrices generated by PROTDIST (setting: Dayhoff PAM250 substitution matrix; Dayhoff et al. 1978) or DNADIST (K84 setting).

The MrBayes program was used for Bayesian analysis and the parameters for amino acid alignments were as follows: mixed model and gamma distribution with 4

gamma rate parameters. The Bayesian inference of phylogenies was initiated from a random starting tree and four chains were run simultaneously for 5 000 000 generations; trees were sampled every 1000 generations. The first 25 % of trees generated were discarded ("burn-in") and the remaining trees were used to compute the posterior probability values. For nucleotide sequence alignments the GTR model with gamma distribution was applied to the data set and as above four chains were run simultaneously for 1 000 000 generation with sample frequency of 100 and a "burn-in" corresponding to the first 25% of sampled trees.

The internal transcribed spacer region (ITS) data was also analyzed with the Tree-Puzzle (TP) program, a maximum likelihood (ML) phylogenetic analysis using quartets and parallel computing (Schmidt et al. 2002). The settings for the quartet puzzling algorithm were 10 000 puzzling steps, transition/transversion parameter estimated from data set, HKY evolutionary model (Hasegawa et al. 1985).

## **5.3 RESULTS**

### **5.3.1 The *rnl-U7* Region**

PCR amplification of the *rnl-U7* segment (with Lsex1 and Lsex2 primers) of 23 strains (Table 5.1) representing various members of the Order Microascales, including species of *Ceratocystis* and related taxa (Hausner and Reid 2004), revealed four size classes of PCR products: 0.3 kb, 1.3 kb, 3.5 kb and 4.5 kb. A previous study (Sethuraman et al. 2008) has shown that a 300 bp product is indicative of the *rnl-U7*

**Table 5.1.** Sources, Genbank accession numbers (rDNA ITS) and the size of the *rnl*-U7 region for the strains used in this study.

**TABLE 5.1**

<b>Species</b>	<b>Isolate/Source WIN(M) numbers</b>	<b>GenBank Accession (ITS region)</b>	<b>Size of <i>rnl</i>-U7 PCR products (kb)</b>
<i>Ceratocystis adiposa</i> (E.J. Butler) C. Moreau	CBS <sup>1</sup> 127.27	=DQ318195*	3.5
<i>Ceratocystis fagacearum</i> (Bretz) Hunt	WIN(M) <sup>2</sup> 892 =ATCC <sup>3</sup> 24790	DQ318193	1.3
<i>Ceratocystis norvegica</i> J. Reid	WIN(M) 87	DQ318194	0.3
	WIN(M) 196	=DQ318194	0.3
<i>Ceratocystis paradoxa</i> (Dade) C. Moreau	WIN(M) 925 =CBS 107.22	DQ318203	0.3
<i>Ceratocystis pinicola</i> T.C. Harr & M. J. Wingf.	WIN(M) 98 =UAMH <sup>4</sup> 9550	DQ318196	1.3
	WIN(M) 1437 =CBS 100200	DQ318198	0.3
<i>Ceratocystis polonica</i> Siernaszko	WIN(M) 143	=DQ318200	3.5
	WIN(M) 199	=DQ318200	0.3
	WIN(M) 325 =UAMH 9783	DQ318200	3.5

<i>Ceratocystis resinifera</i> T.C. Harr. & M. J. Wingf.	WIN(M) 79 =NFRI <sup>5</sup> 66-157, UAMH 9644		1.3
	WIN(M) 1409	DQ318205	4.5
	WIN(M) 1411	DQ318197	
<i>Canariomyces notabilis</i> Arx	UAMH 4906		0.3
<i>Faurelina fimigena</i> Locq.-Lin.	UAMH 9396		0.3
<i>Gondwanamyces proteae</i> (M.J. Wingf., P.S. van Wyk & Marasas) G.J. Marais & M.J. Wingf.	CBS 486.88		0.3
<i>Kernia pachypleura</i> Malloch & Cain	WIN(M) 253 =UAHM 9790	DQ318208	0.3
<i>Lophotrichus bartlettii</i> (Masse & E.S. Salmon) Malloch & Cain	UAMH 8395		0.3
<i>Microascus cirrosus</i> Curzi	UAMH 963		0.3
<i>Petriella guttulata</i> G.L. Barron & Cain	UAMH 8399		0.3
<i>Pithoascus platysporus</i> Arx & Veenb.-Rijks	UAMH 9259		0.3

<i>Sphaeronaemella fimicola</i>	WIN(M) 818	0.3
Marchal	WIN(M) 1402	5.0

<sup>1</sup>CBS = CBS-KNAW Fungal Biodiversity Center, Utrecht, Netherlands

<sup>2</sup>WIN(M) = culture collection of J. Reid, Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada

<sup>3</sup>ATCC = American Type Culture Collection, Manassas, Virginia, USA

<sup>4</sup>UAMH = University of Alberta Microfungus Collection and Herbarium, Devonian Botanic Garden, Edmonton, Alberta, Canada

<sup>5</sup>NFRI = Norwegian Forest Research Institute, Ås, Norway

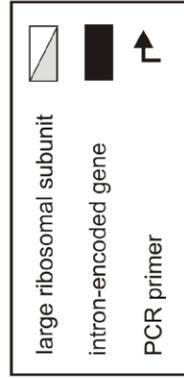
\**Ceratocystis major* CBS 138.34 rDNA ITS sequence identical to *C. adiposa*

region with no insertions whereas the larger amplicons were indicative of insertions. Based on this PCR survey representative strains with different exon/intron arrangements were further characterized by DNA sequence analysis. Strains without *rnl-U7* insertions were *C. norvegica* WIN(M) 87 and 196, and *C. paradoxa* WIN(M) 925. An overview of the various introns and intron-encoded ORFs uncovered during this study is provided by Figure 5.1.

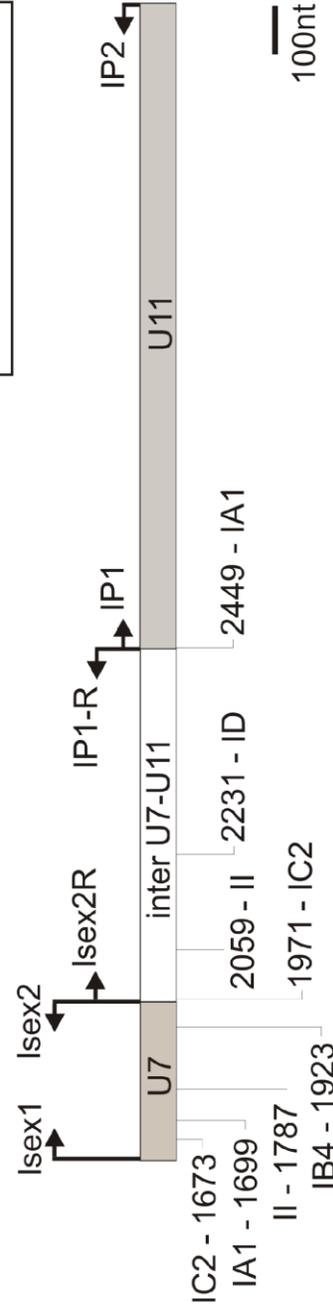
### 5.3.2 Strains Containing Only the mL1923 Insertion in the *rnl-U7* Region

Comparative sequence analysis for those strains that yielded a 1.3 kb Lsex-1/Lsex-2 PCR product: *C. resinifera* [WIN(M) 79], *C. pinicola* [WIN(M) 98] and *C. fagacearum* [WIN(M) 892] along with representative *rnl-U7* sequences obtained from Genbank (*Hypocrea jecorina* AF447590, *Gibberella zeae* DQ364632, *Cordyceps bassiana* EU100742) showed the presence of an insertion at position mL1923 [based on *E. coli* large subunit (LSU) rRNA gene; Figure 5.1 A and B; Figure 5.2 A]. The size of the inserted element corresponded to 1053 bp, 1179 bp and 893 bp in *C. resinifera* [WIN(M) 79], *C. pinicola* [WIN(M) 98] and *C. fagacearum* respectively. Analysis of these sequences with the RNA weasel (Lang et al. 2007; <http://megasun.bch.umontreal.ca/RNAweasel/>) and Mfold programs indicated that these insertions represent group I introns belonging to the class IB4 type.

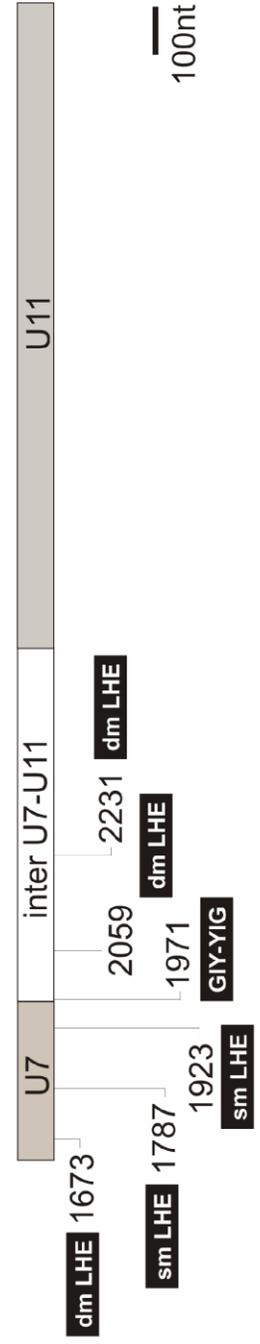
The ORF finder program identified putative ORFs of 224, 241 and 200 amino acids within the mL1923 intron of *C. resinifera* [WIN(M) 79], *C. pinicola* [WIN(M) 98]



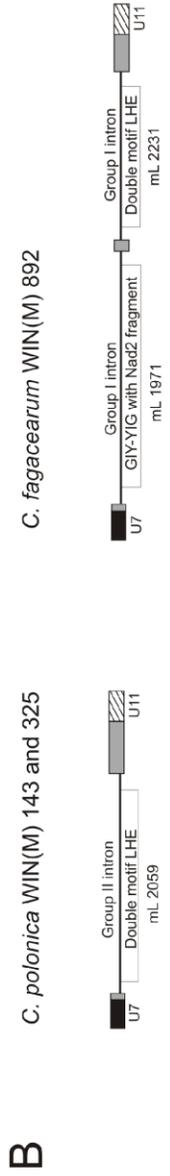
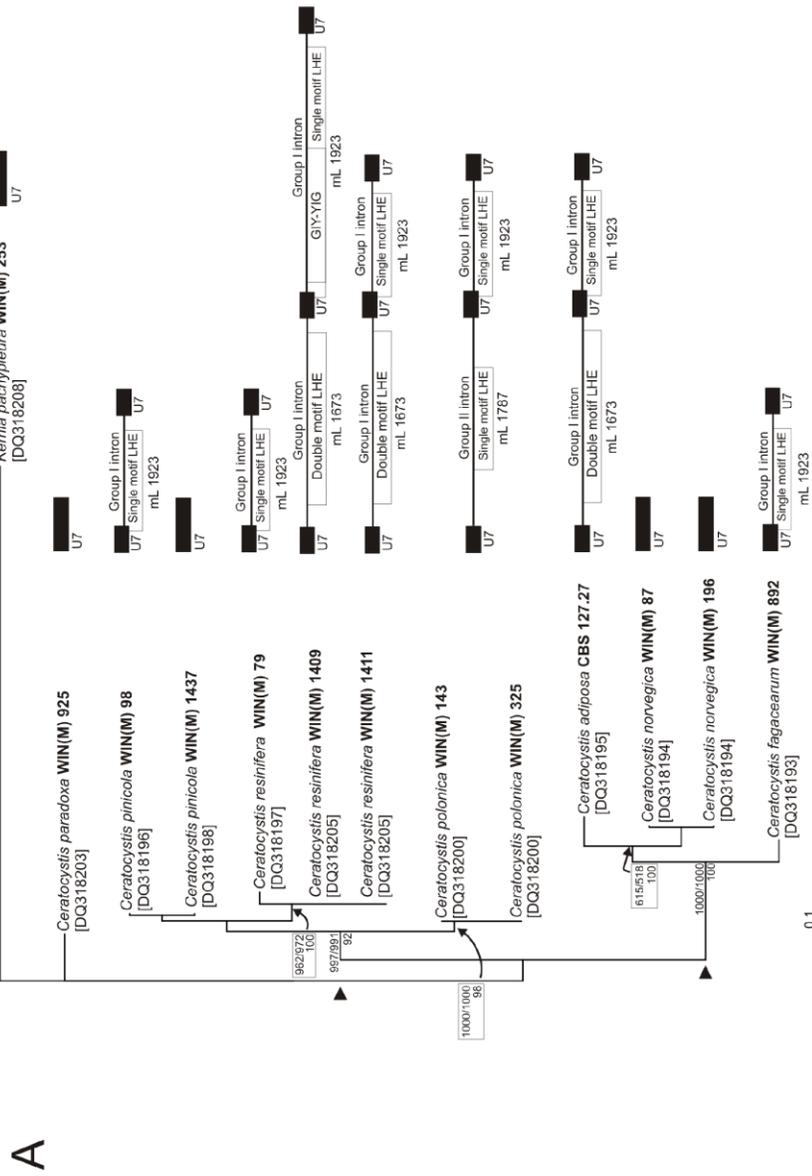
A



B



**Figure 5.1.** Overview of the mitochondrial *rnl* (white and grey boxes) in species of *Ceratocystis*. **(A)** Intron insertion sites and intron subtypes discovered in this work and discussed by Haugen and Bhattacharya (2004). PCR primers used to amplify regions of the *rnl* are represented by arrows. **(B)** Intron insertion sites and their encoded open reading frames (dark boxes) discussed in this study.



**Figure 5.2.** ITS phylogeny and a schematic of the *rnl-U7* region of *Ceratocystis* species used in this study. **(A)** Phylogenetic tree based on 1157 positions of the ITS sequence of 12 strains belonging to the species *Ceratocystis*. *Kernia pachypleura*, a closely related species to *Ceratocystis*, was used as the outgroup. The two clades are represented by ►. The numbers at the nodes indicate the values obtained using bootstrap analysis in combination with Neighbor-Joining and Parsimony respectively. The number below the line indicates the posterior probability values obtained from the 50% majority consensus tree generated using Bayesian inference. Only the nodes that yielded statically significant numbers are denoted. The branch lengths shown are proportional to the numbers of nucleotide substitutions per site (see scale bar). Schematics of the *rnl-U7* (dark boxes) for each strain are situated next to the tree to showing the diversity of insertions (lines and white boxes). The single motif LAGLIDADG homing endonuclease (LHE) open reading frame overlaps the mL1923 intron core sequence. **(B)** The *rnl-U7-U11* inter-region (grey boxes) of two *Ceratocystis* species, showing the introns and their associated open reading frames.

and *C. fagacearum* respectively. All of these ORFs showed the presence of a single copy of the LAGLIDADG sequence motif suggesting that these ORFs represent potential HEs or maturases (Belfort and Roberts 1997; Belfort et al. 2002). A BLASTp database search using these ORF amino acid sequences as queries showed that these ORFs are similar to other single motif LAGLIDADG homing endonucleases found in mtDNA or chloroplast (cp)DNA *rnl* genes at position L1923. The E-value for the BLASTp search results with regards to L1923 ORFs was in the range of  $1^{-17}$  to  $1^{-15}$ . Examples of cpL1923 intron ORFs are found within members of the Class Chlorophyceae (Turmel et al. 1997).

### **5.3.3 Strains Containing the mL1673 and the mL1923 Insertions in the *rnl*-U7**

#### **Region**

Representative *Ceratocystis* strains that yielded Lsex-1/Lsex-2 PCR product of 3.5 kb and 4.5 kb were analyzed and sequences were obtained for the *rnl*-U7 region for *C. polonica* [WIN(M) 143 and 325; 3.5 kb], *C. adiposa* (CBS 127.27; 3.5 kb) and *C. resinifera* [WIN(M) 1409, 1411; 3.5 kb and 4.5 kb respectively; Figure 5.2 A]. Sequence analysis revealed the presence of an additional insertion to mL1923 at position mL1673, such as in *C. adiposa* and *C. resinifera*. Sequence analysis showed that the *C. adiposa* and *C. resinifera* insertions at mL1673 is a 2116 bp group I intron that belongs to the IC2 class of group I introns. The ORF finder program in combination with BLASTp identified a putative ORF of 308 amino acids with similarities to double motif LAGLIDADG type ORFs that also contain a second set of conserved sequence motifs referred to as  $r_A$  and  $r_B$

(Cummings et al. 1989).

With regards to *C. adiposa* and *C. resinifera* [WIN(M) 1411] the mL1923 intron is similar to the L1923 intron described above. However, for *C. resinifera* strain WIN(M) 1409 there was a significant difference; here, the ORF finder program in combination with BLASTp identified a 382 amino acid sequence that includes a potential single motif LAGLIDADG element plus a GIY-YIG endonuclease sequence (299 amino acids). The WIN(M) 1409 mL1923 intron ORF appears to be the result of a resident LAGLIDADG ORF being split into two pieces due to the insertion of a GIY-YIG endonuclease gene. This GIY-YIG is fused in-frame to the upstream 273 bp N-terminus coding segment of the resident LAGLIDADG HEG.

Based on a comparison of the mL1923 intron sequence of *C. resinifera* strain WIN(M) 1409 with the following strains *C. resinifera* [WIN(M) 79 and 1411], *C. pinicola* [WIN(M) 98], *C. polonica*, *C. fagacearum* and *C. adiposa*, it appears that the GIY-YIG HEG introduced a novel 15 bp sequence at its 5' terminus and a unique 20 bp segment at the 3' terminus (Figure 5.2 A). This 20 bp region is followed by the C-terminal remnant of the displaced resident LAGLIDADG ORF. Sequence comparison shows that the displaced LAGLIDADG sequence is identical to that of the intact LAGLIDADG ORFs present in the other *Ceratocystis* mL1923 introns described above. However, this sequence is likely to be non-coding since the GIY-YIG ORF contains a stop codon.

### 5.3.4 Strains Containing the mL1787 and the mL1923 Insertions in the *rnl-U7*

#### Region

The *rnl-U7* region within the two strains of *C. polonica* [WIN(M) 143 and 325] contains two insertions (Figure 5.2 A). One is the mL1923 group I intron, as seen in the other species, and the second insertion is at position mL1787. The mL1787 intron of 2427 bp was determined to be a group II B1 intron based on the RNA Weasel and Mfold programs. The group II intron at mL1787 encodes a single motif LAGLIDADG type HEG of 176 amino acids based on the ORF Finder program and BLASTp analysis.

### 5.3.5 ITS Phylogeny Versus the *rnl-U7* Intron Insertions

According to the ITS phylogeny, the *Ceratocystis* species examined during this study group into two clades (Figure 5.2 A). The first clade includes strains of *C. pinicola*, *C. resinifera* and *C. polonica* and has unexpected variation within the *rnl-U7* region. For example, *C. resinifera* WIN(M) 79 has only the mL1923 intron, whereas the other two have two introns (mL1673 and mL1923); adding complexity, the strain WIN(M) 1409 has a GIY-YIG HEG interrupting the mL1923 intron ORF. The two *C. pinicola* strains in this study also showed variation. Strain WIN(M) 1437 is void of introns in the *rnl-U7*, whereas strain WIN(M) 98 harbours the mL1923 intron.

The second clade observed in the rDNA ITS phylogeny includes strains of *C. fagacearum*, *C. adiposa*, and *C. norvegica*. Among these three species, three different

intron/exon arrangements were observed for the *rnl-U7* region: (1) no introns in *C. norvegica*, (2) one intron (mL1923) in *C. fagacearum*, and (3) two introns (mL1673 and mL1923) in *C. adiposa*. The latter has an *rnl-U7* arrangement similar as found in *C. resinifera* WIN(M) 1411 of clade 1.

### **5.3.6 Another Group II Intron with a LAGLIDADG-Type ORF and Two Group I Introns in the U7 – U11 Inter-Region**

For two *Ceratocystis* species, *C. polonica* and *C. fagacearum*, the entire sequence *rnl-U7* to *rnl-U11* was amplified and sequenced (Figure 5.1 and 5.2). For all strains, the *rnl-U11* region contained the mL2449 group I intron encoding the ribosomal protein S3 (Rps3; Bullerwell et al. 2000, 2003); this was expected based on previous work and this intron has been previously described in several filamentous ascomycetes (Burke and Rajbhandary 1982; Sethuraman et al. 2009a ; reviewed in Sethuraman et al. 2009b). Sequencing the segment between the *rnl-U7* and U11 regions for strains of *C. polonica* revealed the presence of an insertion at position mL2059. This intron, based on its RNA fold, is a group II B1 intron and based on ORF finder and BLASTp analysis encodes a double motif LAGLIDADG homing endonuclease-like gene.

A different arrangement was observed in *C. fagacearum* for this region; here two insertions were noted downstream of the U7 region. First a group I intron with a GIY-YIG type ORF could be identified at position mL1971. This intron, based on its RNA fold, is an IC2 type. One unique feature of this insertion is its ORF, which appears to be a

mosaic of a *nad2* gene segment and a degenerated GIY-YIG type sequence. The second insertion at mL2231 is a ID subtype group I intron with a degenerated LAGLIDADG-type ORF (the sequence contains premature stop codons).

### **5.3.7 *In Vivo* Splicing and Intron/Exon Junctions of the *rnl* Introns**

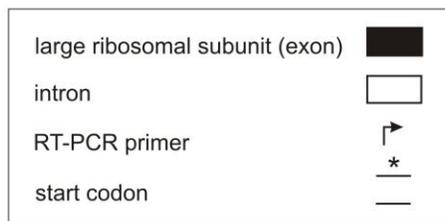
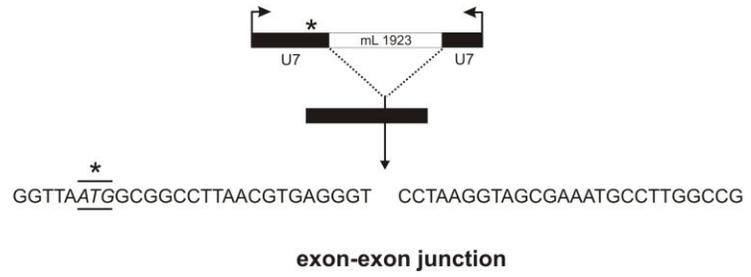
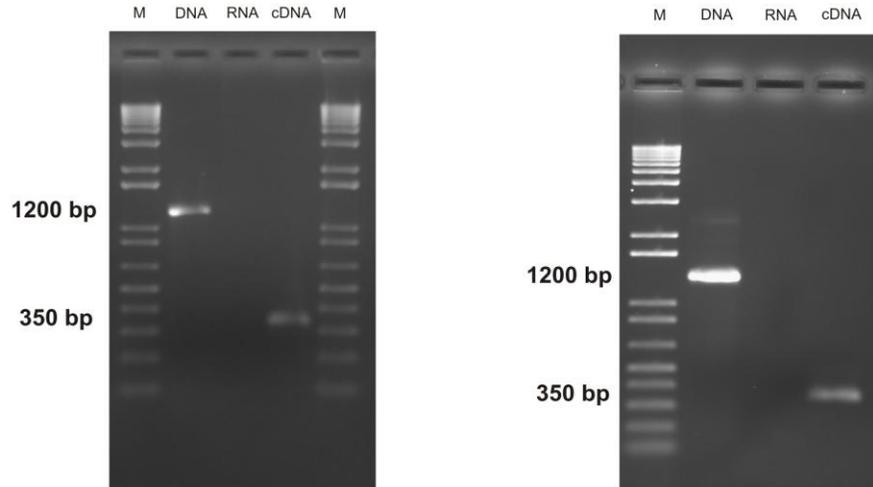
For every type of putative intron encountered during this study at least one representative was examined for the potential to be spliced out *in vivo* from their precursor transcripts (Figures 5.3, 5.4 and 5.5). The RT-PCR products were sequenced, so that the intron/exon junctions could be determined by comparing the complementary (c)DNA sequences with the original *rnl* genomic sequences. All intron sequences were found to be absent from the cDNAs. The single motif LAGLIDADG ORF encoded by the mL1923 intron includes the structural components of the 5' intron sequence and 25 bp of the ORF extends into the upstream exon (*rnl*-U7 region; Figure 5.3). If the exon/intron boundaries for mL1923 are correct, then translation of the intron-encoded ORF is required before the intron splices from the host mRNA transcript.

### **5.3.8 Phylogeny of the mL1923 Intron-Encoded LAGLIDADG HEG**

The mL1923 single motif LAGLIDADG ORF and the mL1673 double motif LAGLIDADG ORF were used as queries in successive BLASTp searches and related endonuclease sequences were extracted from Genbank. It has been shown that double

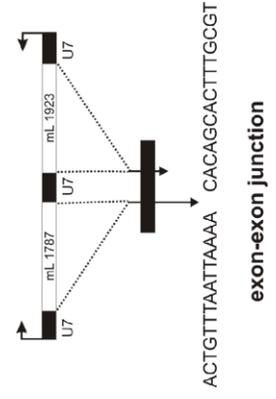
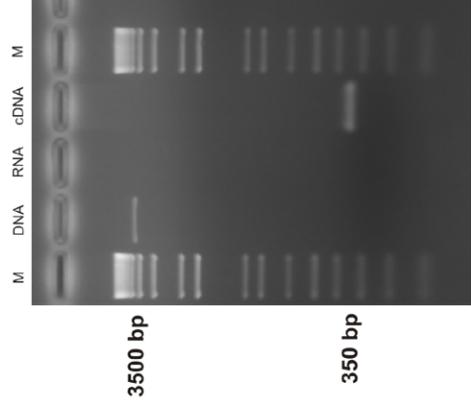
***C. resinifera* WIN(M)79**

***C. pinicola* WIN(M)98**

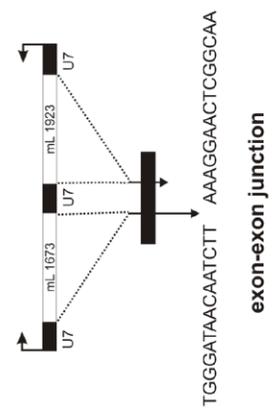
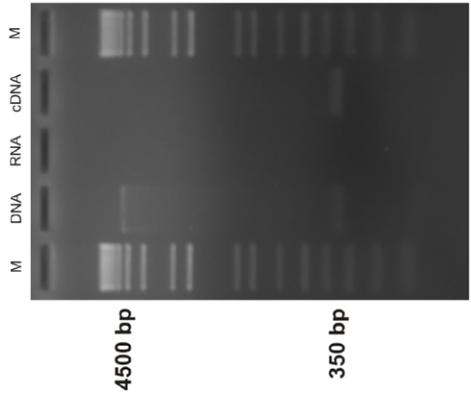


**Figure 5.3.** RT-PCR with total RNA isolated from *Ceratocystis resinifera* and *C. pinicola*. Diagram shows the *rnl*-U7 region (dark boxes) and the RT-PCR primers (bent arrows) flanking the intron (white box). The core sequence of the mL1923 intron overlaps with its open reading frame (\*). Lanes of the agarose gels marked M contain 1-kb plus DNA marker. Positive control (DNA) contains genomic DNA template for PCR. Negative control (RNA) contains RNA template for PCR, to screen for the absence of DNA contamination. Lane marked cDNA is the intron-less cDNA amplicon of *rnl*.

***C. polonica* WIN(M)325**

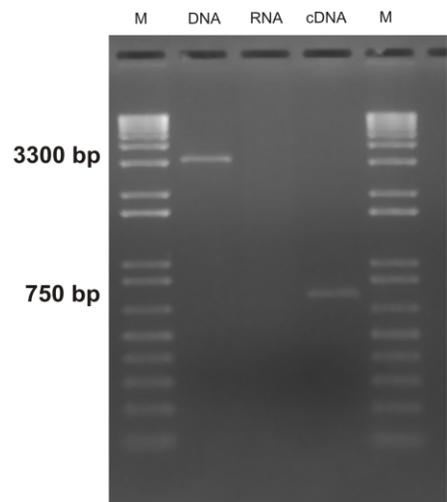


***C. resinifera* WIN(M)1409**



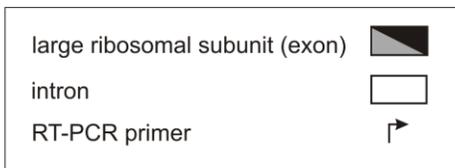
**Figure 5.4.** RT-PCR with total RNA isolated from *Ceratocystis polonica* and *C. resinifera*. Diagram shows the *rnl-U7* region (dark boxes) and the RT-PCR primers (bent arrows) flanking the introns (white boxes). Lanes of the agarose gels marked M contain 1-kb plus DNA marker. Positive control (DNA) contains genomic DNA template for PCR. Negative control (RNA) contains RNA template for PCR, to screen for the absence of DNA contamination. Lane marked cDNA is the intron-less cDNA amplicon of *rnl*. See Figure 5.3 for the intron-exon junction of the mL1923 intron.

***C. polonica* WIN(M)325**

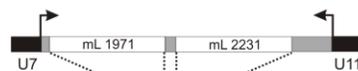
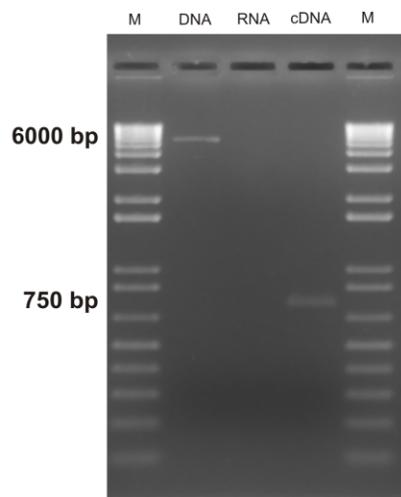


TCTAGCTGGACGAGA AGACCCTATGCAGCT

**exon-exon junction**



***C. fagacearum* WIN(M)892**



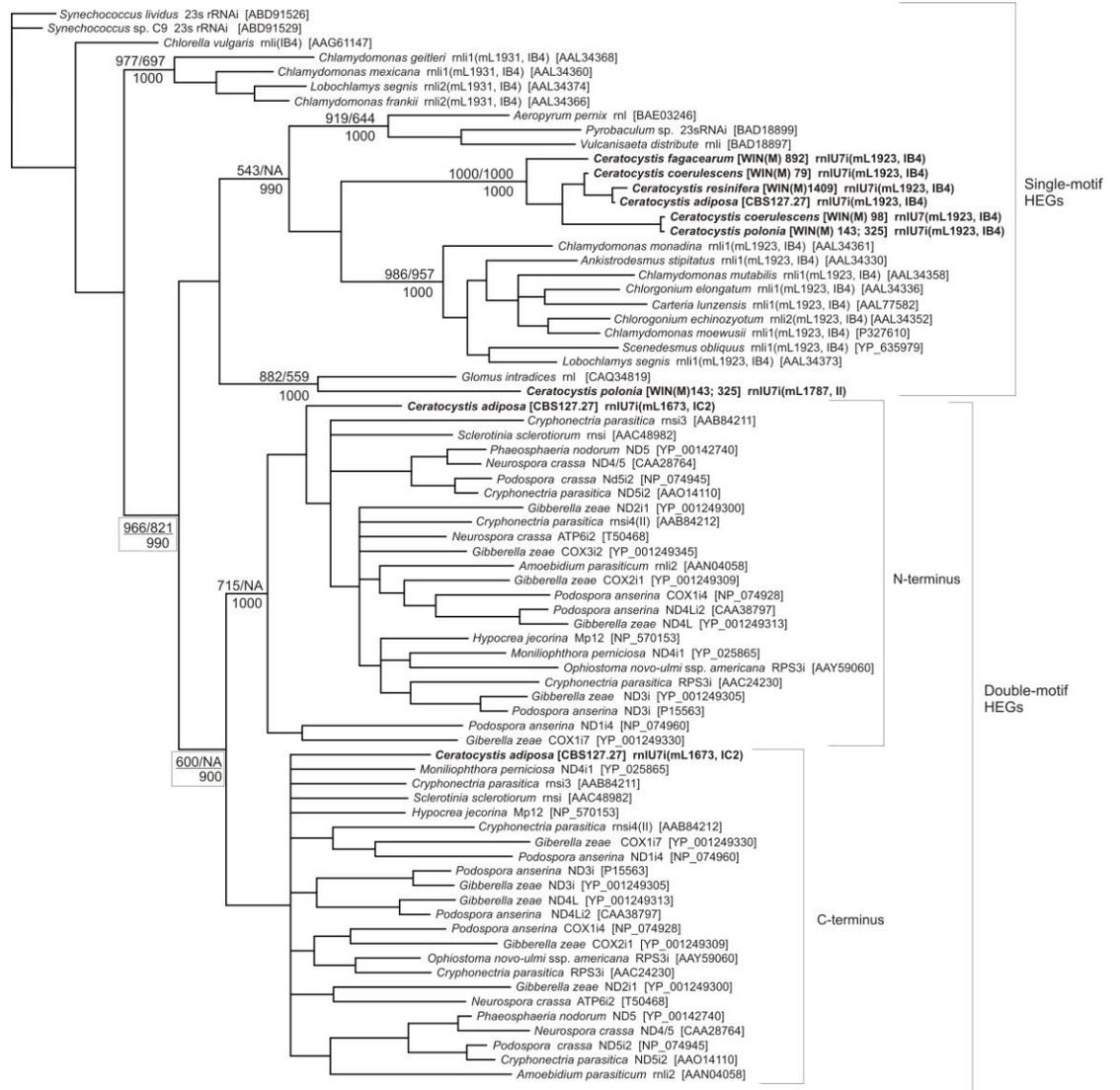
AATGCGGTCTTGCAT <sup>1</sup> GAATGATGTAACGAT

TGTTAGGAGACAGTT <sup>2</sup> TATGCGGGGCACAGA

**Figure 5.5.** RT-PCR with total RNA isolated from *Ceratocystis polonica* and *C. fagacearum*. Diagram shows the *rnl*-interU7-U11 region (grey boxes) and the RT-PCR primers (bent arrows) flanking the introns (white boxes). Lanes of the agarose gels marked M contain 1-kb plus DNA marker. Positive control (DNA) contains genomic DNA template for PCR. Negative control (RNA) contains RNA template for PCR, to screen for the absence of DNA contamination. Lane marked cDNA is the intron-less cDNA amplicon of *rnl*.

motif type LAGLIDADG HEGs may have evolved from sequential events of gene duplication and gene fusion, involving an ancestral single motif type LAGLIDADG HEG (Haugen and Bhattacharya 2004). Therefore, we treated the double motif LAGLIDADG ORFs as two separate parts for the phylogenetic analysis: the N-terminus with LAGLIDADG motif 1 and the C-terminus component with the LAGLIDADG motif 2 (see Haugen and Bhattacharya 2004; Sethuraman et al. 2008).

A single motif LAGLIDADG ORF present in the *Synechococcus lividus* 23S RNA was used as the outgroup for the phylogenetic analysis. The topology of the resulting phylogenetic tree (Figure 5.6) showed that our *Ceratocystis rnl*-U7 introns mL1923 and mL1673 encoded HEGs that are part of a HEG family referred to as “Clade1” by Haugen and Bhattacharya (2004). This earlier study focused on the spread of HEGs into rDNA and suggests that the ancestor for Clade1 rDNA double motif type HEGs arose from a duplication event involving a single motif HEG encoded within the pool of L1917 - L1951 introns. This is in agreement with our analysis, as the topology of the tree places the L1923 and L1931 ORFs basal to the position of the *Ceratocystis* mL1673 double motif HEGs. The mL1787 group II intron-encoded single motif LAGLIDADG appears to be related to the single motif LAGLIDADG present in the mL1923 intron. Specifically, the *C. polonica* mL1787 group II intron single motif LAGLIDADG ORF is related to another single motif LAGLIDADG ORF that is also located within a group II intron at the same *rnl* insertion site within the distantly related species *Glomus intradices* (Figure 5.6).



0.1

**Figure 5.6.** Phylogenetic position of the *Ceratocystis rnl-U7* homing endonucleases. Double motif LAGLIDADG type open reading frames were separated into N-terminus and C-terminus components and treated as separate sequences. The tree suggests that the single motif homing endonucleases at position L1923 could be part of the lineage that eventually gave rise to the double motif homing endonucleases present in position L1671 along with other double motif HEGs that invaded other rDNA sites and protein coding genes. The double motif HEGs appear to be derived from an ancient duplication event thus the N and C terminal region of these proteins form separate clades expected for paralogs. The numbers at the nodes indicate the level of support based on bootstrap analysis in combination with Neighbor-Joining and Parsimony respectively. The third number below the line represents the posterior probability values obtained from the 50% majority consensus tree generated using Bayesian analysis. Numbers are only provided for those nodes that were statistically significant i.e., posterior probability values of  $\geq 99\%$  and for bootstrap support the numbers should be  $\geq 95\%$ . The branch lengths shown are proportional to the number of substitutions per site (see scale bar).

In a comprehensive study on rDNA LAGLIDADG HEGs by Haugen and Bhattacharya (2004), the Clade1 type double motif HEGs were encoded by S569, S1224, S1210, L1931, L1939 and L1949 introns. In our study, the mL1673 intron-encoded double motif LAGLIDADG HEG represents a new entry to Haugen and Bhattacharya's Clade1 list. As previously found with mL2449 HEGs, the mL1673 intron ORFs group with other double motif LAGLIDADG HEGs located within rDNA [small ribosomal subunit (*rns*) and *rnl*], protein coding genes (*nad3*, *nad4*, *nad5*, *cox1*, *atp6* etc.) and potentially free standing mtDNA HEGs (for example, Mp12 from *Hypocrea jecorina* mtDNA in Figure 5.6; Sethuraman et al. 2009a). In general, the double motif LAGLIDADG HEGs were noted to be inserted within introns located in various *rns* or *rnl* sites and protein coding genes. However, the single motif HEGs are more restricted with regards to the insertion sites. With the exception of the mL1787 intron ORF, the *Ceratocystis rnl*-U7 single motif HEGs are confined to the mL1923 intron, the same site occupied by their more distantly related cp-*rnl* counterparts (cpL1923). Also, the mL1923 intron present in both chloroplasts and mitochondria (*Ceratocystis*), belongs to IB4 group I introns.

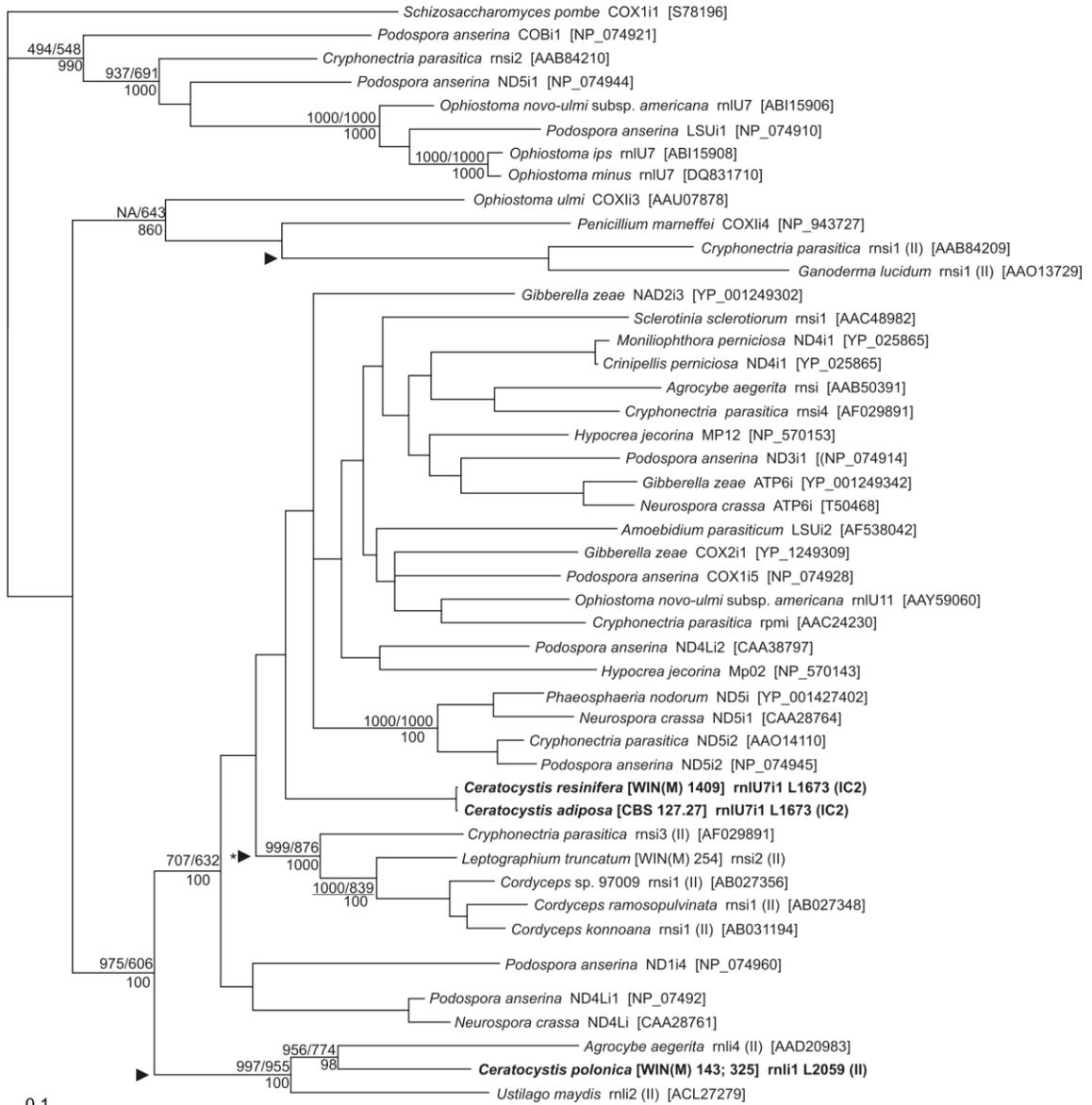
### **5.3.9 Phylogeny of the Double Motif LAGLIDADG ORF Encoded by the mL2059 Group II Intron**

The phylogenetic position of the mL2059 ORF was examined in more detail amongst other double motif LAGLIDADG ORFs encoded by either group I or II introns

(Figure 5.7). This HEG is related to group II intron ORFs located within the *rnl* genes of *Agrocybe aegerita* and *Ustilago maydis*. Within this phylogenetic tree we noted three clades for group II intron ORFs, showing that several evolutionary lineages exist where group I intron-encoded ORFs were transferred into group II introns. Two of the clades included group II introns ORFs inserted within the mtDNA *rns* gene (mS785 and mS952) and the third clade consisted of three members inserted within the mtDNA *rnl* gene (mL2059).

## 5.4 DISCUSSION

HEGs and introns are quite invasive and contribute towards the size of fungal mtDNA genomes, mtDNA polymorphisms and mtDNA rearrangements (Dujon and Belcour 1989; Charter et al. 1996; Sellem et al. 1996; Hamari et al. 1999; Gogarten and Hilario 2006). These elements are of interest since intron-associated mtDNA instabilities have been observed in yeast, as well as in an assortment of filamentous fungi (Cummings et al. 1986; Cummings et al. 1990; Gillham et al. 1994; Abu-Amero et al. 1995; Baidyaroy et al. 2011); thus, a detailed description of the mt-*rnl* gene and its introns (see Figure 5.1 for summary) in species of *Ceratocystis* is presented. Additionally, there is increasing interest in mtDNA encoded ribozymes and HEGs for biotechnology and human therapeutic applications (Lambowitz and Zimmerly 2004; Stoddard 2005, 2011; Marcaida et al. 2010).



0.1

**Figure 5.7.** Phylogenetic analysis of the group II intron encoded homing endonucleases. The tree suggests the possibility of (indicated by ►) three different lineages for the movement of homing endonuclease genes from a group I intron to group II introns. The clade that includes two group II intron encoded homing endonuclease lineage appears to be derived from the Clade 1 of rDNA type LAGLIDADG HEGs (\*) as explained in Haugen and Bhattacharya (2004). The numbers at the node indicates the level of support obtained and only the values that were statistically significant are shown. The numbers above the line represents the support obtained by bootstrap in combination with the Neighbor-Joining and Parsimony analysis whereas the number below the line indicates the posterior probability values obtained from the 50% majority consensus tree generated using Bayesian analysis. NA indicates that a node did not receive significant values with the particular phylogeny reconstruction method. The branch lengths shows are proportional to the number of substitutions per site (see scale bar).

#### **5.4.1 ITS Species Tree Versus *rnl-U7* Configuration**

There is a lack of congruency between the ITS phylogeny and the *rnl-U7* introns. The intron arrangement shows a somewhat random distribution, even among the same species of *Ceratocystis* (see *C. pinicola* and *C. resinifera* in Figure 5.2 A). This might be indicative of horizontal transfer of introns and their ORFs between different species of *Ceratocystis*. Given the limited number of strains examined, another explanation could be vertical inheritance of all these introns from a recent ancestor, followed by periodic random loss; however, this is unlikely for the mL1787 group II intron and its LAGLIDADG ORF or for the GIY-YIG ORF inserted within the mL1923 intron ORF. Overall, the *rnl-U7* region of *Ceratocystis* is an example of the amount of mitochondrial genomic variability that can be generated by the presence or absence of mobile introns and their associated HEGs. The diversity of *rnl-U7* insertions between strains of the same species indicates that these introns may not be suitable markers for species identification. It is interesting to note that despite the large number of introns present, splicing efficiency must be maintained in order to maintain a functional host gene. The presence of degenerate ORFs suggests that either some of these introns are extremely efficient in self-splicing or there is a reliance on host factors to expedite removal of introns.

#### **5.4.2 The mL1923 Intron – ORF/Exon Configuration**

For the mL1923 intron, the overlap of the ORF with the intron core creates a situation where intron splicing is competing with ORF translation and thus might be a

strategy to down-regulate the expression of the HE ORF. As the HEG is an optional component with limited target sites, large production of the HE protein would be wasteful of host resources. Another possibility for regulation of translation is that proper folding and splicing of the intron is facilitated by a ribosome. Translatability of 5' exons is essential for self-splicing of the td group I intron in bacteriophage T4 (Semrad and Schroeder 1998). An active ribosome on the pre-mRNA blocks exon segments from interfering with the proper folding of the intron core (i.e. exon sequences competing for intron sequences to fold with intron segments). Ribosomes have been implicated in the splicing of ribozyme-containing introns (group I and group II introns) in several systems where there is no separation between transcription and translation (Sandegren and Sjöberg 2007; Michel et al. 2007). Finally, if this ORF has maturase activity then it would have to be translated before the intron can be efficiently removed from the *rnl* precursor transcript.

The arrangement of the mL1923 intron is also of interest in terms of the splicing competency and the potential for mobility (which requires a functional ORF). These events could potentially compete with each other, since the ORF sequence merges with the intron core sequence. Intron-encoded ORFs can extend into the ribozyme core regions, but once established there must be strong selection pressure to maintain both the ribozymes's catalytic function as well as the reading frame for the HE (Lazarevic et al. 1998; Bonocora and Shub 2001; Carbone et al. 1995; Haugen et al. 2007; Edgell et al. 2011). Host factors may reduce the functional constraints on the intron sequence, thus

allowing the intron core structure to degenerate somewhat and permit the ORF to displace some of the intron's sequence (Mohr et al. 2001).

### **5.4.3 GIY-YIG ORFs: “Parasites” of LAGLIDADG HEGs and HEG Transduction**

The *Podospora curvicolla* I-PcI GIY-YIG HEG, like the *C. resinifera* GIY-YIG HEG inside the mL1923 intron, is located within a LAGLIDADG ORF. Both HEGs disrupt a LAGLIDADG coding sequence, except that the I-PcI HEG is inserted after the first motif of the double-motif LAGLIDADG ORF, whereas *C. resinifera* has a single motif LAGLIDADG ORF (Figure 5.2 A). The *P. curvicolla* I-PcI HE has been demonstrated to be an active DNA-cutting enzyme (Saguez et al. 2000). Both the I-PcI and the *C. resinifera* GIY-YIG ORF may represent a lineage of GIY-YIG HEGs that have specialized in inserting in-frame within LAGLIDADG-type ORFs. Expression of these HEGs may require alternate splicing (Sellem and Belcour 1994; 1997) or proteolysis to resolve the fusion protein (Pel and Grivell 1993; van Dyck et al. 1998; Arlt et al. 1998). The GIY-YIG coding region may benefit from being fused to the upstream LAGLIDADG coding component as it gains access to the cis-acting regulatory sequences required for expression. The insertion of a mobile element within another mobile element renders the resident element inactive and may actually be beneficial to the host genome. Insertion within another element ensures that the newly acquired element is neutral with regards to its impact on the host gene and genome, by not adding to the proliferation of mobile elements within a genome (Saguez et al. 2000).

A second example of a GIY-YIG HEG, albeit degenerated due to frameshift

mutations, was discovered within the *ml-U7* mL1971 group I intron. This ORF had a partial *nad2* sequence fused upstream to the remnant of the degenerated HEG (Figure 5.2 B). The GIY-YIG HEG possibly originated from a *nad2* gene and inserted into the mL1971 intron; since HEGs mobilize via the double strand break repair system, sequences flanking the HEG ORF can be moved along with the HEG into a new position (reviewed in Hausner 2003) due to gene convergence. Examples where HE activity mobilized sequences flanking HEGs into new location have been reported previously (Paquin et al. 1994; Sethuraman et al. 2009a). In some instances these sequences allow the free-standing HEGs to repair the damage they cause during insertion, when they displace segments of protein encoding genes (reviewed in Sethuraman et al. 2009a), but this is unlikely for this HEG since the GIY-YIG inserted into a group I intron (a neutral event) instead of into a host gene.

#### **5.4.4 More Group II Introns with LAGLIDADG ORFs**

LAGLIDADG type HEGs are viewed as rather promiscuous elements that have been noted to exist sometimes as free-standing ORFs or, more frequently, found to successfully colonize group I introns, inteins, archaebacterial introns, and group II introns (Lambowitz et al. 1999; Gimble 2000; Toor and Zimmerly 2002). Overall, LAGLIDADG ORFs are most commonly found in group I introns and in some cases these ORFs have evolved into maturases to assist their host intron in self-splicing (Belfort 2003).

A partial sequence of a group II intron from *Trimorphomyces papilionaceus* was

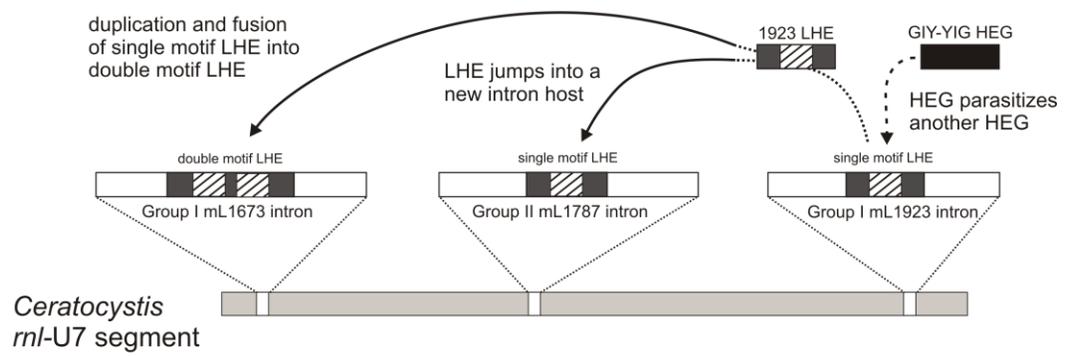
the first to hint at the possibility of LAGLIDADG ORFs invading group II introns (Michel and Ferat 1995). Toor and Zimmerly (2002) recognised 11 fungal mtDNA group II introns that encoded complete or degenerated LAGLIDADG ORFs; however, the authors noted that these group II encoded LAGLIDADG ORFs were not closely related to other LAGLIDADG ORFs, thus their origins remained uncertain. Although this work cannot pinpoint the exact origin of group II intron encoded LAGLIDADG ORFs, it is apparent that group I intron-encoded double-motif LAGLIDADG ORFs have invaded group II introns on several occasions (Figure 5.7); all group II introns with such ORFs fit the IIB1 RNA structures (Toor and Zimmerly 2002). Previous work suggested that at least three families of LAGLIDADG HEGs exist that switched intron hosts (from group I to group II; Monteiro-Vitorello et al. 2009). The current study shows that there is another potential family of LAGLIDADG ORFs (Figures 5.6), represented by the single motif LAGLIDADG ORF encoded by the mL1787 intron. This HEG has transferred into a group II intron independent of the other three “intron host jumps” previously described (see Figure 5.6).

#### **5.4.5 The mL1923 ORFs Have a Complex Phylogenetic History**

Fungal mitochondrial genomes are in a constant flux due to the presence of mobile elements. Among the ascomycetes fungi, the *mL-U7* region has previously been shown to have optional elements inserted (Cummings et al. 1989; Sethuraman et al. 2008; Figure 5.1 A); for example, group I introns and their associated ORF in the mL1699

region. However, within members of *Ceratocystis* only three locations were noted to have introns present: mL1673, mL1787 and mL1923 (Figure 5.1 and Figure 5.8). The presence of optional elements in the mL1673 and mL1787 sites had not been reported previously. The phylogenetic analysis presented in this study and in Haugen and Bhattacharya (2004) suggests that the L1923 ORFs could be the lineage that eventually gave rise to the double motif ORFs inserted within mL1673 introns.

Haugen and Bhattacharya (2004) demonstrated that some rDNA HEGs could have evolved from a duplication event of a single motif LAGLIDADG HEG. After this duplication event, the resulting double motif HEGs are less stringent with their target sites and are more successful at spreading into new rDNA sites and protein coding genes. In our work we can trace such a proposed model for rDNA HEGs, as we noted that the single motif LAGLIDADG ORF encoded within mL1923 is basal to the N- and C-terminus components of the double motif LAGLIDADG ORF located within the mL1673 intron (Figure 5.6). We noted that the ancestral L1923 ORF is found within rDNA IB4 type group I intron, similar to related single motifs ORFs found in similar locations in the early branching members of this family of HEGs, but the derivatives of the L1923 ORF after a duplication event have invaded other types of introns such as IC2 or group II introns (Figure 5.6 and 5.7). Figure 5.8 summarizes the possible movement of LAGLIDADG ORFs from the mL1923 ancestral position into new sites such as a group II intron (mL1787) or after the ORF duplicated to form double motif LAGLIDADG ORF into a group I intron (mL1673).



**Figure 5.8.** Diagrammatic representation of the intron (white boxes) and homing endonuclease gene (dark and striped boxes) evolution in the *rnl-U7* (light grey box) of *Ceratocystis*. The single motif LAGLIDADG homing endonuclease gene (LHE) of mL1923 moved from a group I intron to a group II intron. In a separate event, the mL1923 LHE moved to another group I intron, but this time the LAGLIDADG motif duplicated and fused in-frame, creating a double motif LHE gene. As observed in *C. resinifera*, the mL1923 LHE can also be parasitized by another homing endonuclease gene, effectively inactivating it.

## 5.5 CONCLUSIONS

This survey for insertions within a few *Ceratocystis* strains provides a “snapshot” of the sporadic distribution of mobile elements within one small conserved region; in this case, a portion of the *rnl*. One would assume that the insertions we observed are non-toxic to the host genomes. The diversity of intron and HEG combinations seen within the *rnl*-U7 region and the presence of degenerated HEGs within mL1971 and mL2231 can be attributed to the proposed life cycle of the HEs and mobile introns (Goddard and Burt 1999; Sandegren and Sjöberg 2004; Gogarten and Hilario 2006) which predicts that HEGs and their intron hosts undergo a cycle of invasion followed by degeneration. This cycle can be escaped by horizontal transfer into a new genome, finding a new homing site or by gaining a new function (Gogarten and Hilario 2006). Overall, our study is consistent with the “endonuclease-gene invasion” hypothesis (see Belfort and Roberts 1997; Belfort 2003), which argues that HEGs in the course of their evolution have invaded self-splicing elements numerous times. The observation that different families of LAGLIDADG ORFs exist within group II introns provides more evidence on the invasive nature of HEGs.

## 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(1):54-9.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403-410.
- Bonen L, Vogel J. 2002. The ins and outs of group II introns. *Trends Genet.* 17(6):322-31.
- Andersen S. 2009. An examination of bark beetles and their associated blue stain fungi on boreal jack pine (*Pinus banksiana*) and white spruce (*Picea glauca*) in the Thunder Bay region. [M.Sc.F. thesis]. (Thunder Bay, ON): Lakehead University.
- Arlt H, Steglich G, Perryman R, Guiard B, Neupert W, Langer T. 1998. The formation of respiratory chain complexes in mitochondria is under the proteolytic control of the m-AAA protease. *EMBO J.* 17(16):4837-47.
- Baidyaroy D, Hausner G, Hafez M, Michel F, Fulbright D, and 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):775-83.
- Belfort M, Roberts RJ. 1997. Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* 25(17):3379-88.
- 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, editors. *Mobile DNA II*. Washington (DC): American Society of Microbiology Press. p. 761–83.
- Belfort M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev.* 17:2860–3.
- Bonen L, Vogel J. 2002. The ins and outs of group II introns. *Trends Genet.* 17(6):322-31.
- Bonocora RP, Shub DA. 2001. A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. *Mol Microbiol.* 39(5):1299-306.

- Bullerwell CE, Burger G, Lang BF. 2000. A novel motif for identifying *rps3* homologs in fungal mitochondrial genomes. *Trends Biochem Sci.* 25(8):363-5.
- Bullerwell CE, Leigh J, Seif E, Longcore JE and Lang BF. 2003. Evolution of the Fungi and their mitochondrial genomes. In: Arora DK, Khachatourians GG, editors. *Applied Mycology and Biotechnology. Volume III: Fungal Genomics.* Elsevier Science, New York. p. 133-59.
- Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends Genet.* 19(12):709-16.
- Burke JM, RajBhandary UL. 1982. Intron within the large rRNA gene of *N. crassa* mitochondria: a long open reading frame and a consensus sequence possibly important in splicing. *Cell.* 31(3):509-20.
- Burke JM. 1989. Selection of the 3'-splice site in group I introns. *FEBS Lett.* 250(2):129-33.
- Caprara MG, Waring RB. 2005. Group I introns and their maturases: uninvited, but welcome guests. *Nucl Acids Mol Biol.* 16:103-19.
- Carbone I, Anderson JB, Kohn LM. 1995. A group-I intron in the mitochondrial small subunit ribosomal RNA gene of *Sclerotinia sclerotiorum*. *Curr Genet.* 27(2):166-76.
- Charter NW, Buck KW, 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.* 100:368-72.
- Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* 29(18):3757-74.
- Cummings DJ, Domenico JM, Nelson J. 1989. DNA sequence and secondary structures of the large subunit rRNA coding regions and its two class I introns of mitochondrial DNA from *Podospira anserina*. *J Mol Evol.* 28:242-55.
- Cummings DJ, McNally KL, Domenico JM, Matsuura ET. 1990. The complete DNA sequence of the mitochondrial genome of *Podospira anserina*. *Curr Genet.* 17:375-402.
- Dayhoff MO, Schwartz RM, Orcutt BC. 1978. A model of evolutionary change in proteins. In: Dayhoff MO, editor. *Atlas of protein sequence and structure.* Washington (DC): National Biomedical Research Foundation. Suppl. 3:345-52.
- De Beer ZW, Wingfield BD, Wingfield MJ. 2003. The *Ophiostoma piceae* complex in the Southern Hemisphere: a phylogenetic study. 107(4):469-76.

- Doetsch NA, Favreau MR, Kuscuoglu N, Thompson MD, Hallick RB. 2001. Chloroplast transformation in *Euglena gracilis*: splicing of a group III twintron transcribed from a transgenic *psbK* operon. *Curr Genet.* 39(1):49-60.
- Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, Perlman PS, Vogt VM. 1989. Mobile introns: definition of terms and recommended nomenclature. *Gene.* 82(1):115-8.
- Dujon B, Belcour L. 1989. Mitochondrial DNA instabilities and rearrangements in yeasts and fungi. In: Berg DE, Howe MM, editors. *Mobile DNA.* Washington (DC): American Society of Microbiology. p. 861–78.
- Edgell DR. 2009. Selfish DNA: homing endonucleases find a home. *Curr Biol.* 19(3):R115-7.
- Edgell DR, Chalamcharla VR, Belfort M. 2011. Learning to live together: mutualism between self-splicing introns and their hosts. *BMC Biol.* 9:22.
- Fedorova O, Zingler N. 2007. Group II introns: structure, folding and splicing mechanism. *Biol Chem.* 388(7):665-78.
- Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 39:783–91.
- Felsenstein, J. 2006. PHYLIP (Phylogeny Inference Package) version 3.6c [computer program]. Distributed by the author. Department of Genetics, University of Washington, Seattle, Wash.
- Férandon C, Moukha S, Callac P, Benedetto JP, Castroviejo M, Barroso G. 2010. The *Agaricus bisporus coxI* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *PLoS One.* 5(11):e14048.
- 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):1112–26.
- 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.* 78:35-46.
- Gillham NW, Boynton JE, Hauser CR. 1994. Translational regulation of gene expression in chloroplasts and mitochondria. *Annu Rev Genet.* 28:71-93.
- Gimble FS. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett.* 185(2):99-107.

- Glass NL, Rasmussen C, Roca MG, Read ND. 2004. Hyphal homing, fusion and mycelial interconnectedness. *Trends Microbiol.* 12(3):135-41.
- Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* 96(24):13880-5.
- Gogarten JP, Hilario E. 2006. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol.* 6:94.
- Grivell LA. 1995. Nucleo-mitochondrial interactions in mitochondrial gene expression. *Crit Rev Biochem Mol Biol.* 30(2):121-64.
- Guarro J, Gené J, Stchigel AM. 1999. Developments in fungal taxonomy. *Clin Microbiol Rev.* 12(3):454-500.
- Guhan N, Muniyappa K. 2003. Structural and functional characteristics of homing endonucleases. *Crit Rev Biochem Mol Biol.* 38(3):199-248.
- Hamari Z, Pfeiffer I, Ferenczy L, Kevei F. 1999. Interpretation of variability of mitochondrial genomes in the species *Aspergillus carbonarius*. *Antonie Van Leeuwenhoek.* 75(3):225-31.
- Harrington, TC. 1981. Cycloheximide sensitivity as a taxonomic character in *Ceratocystis*. *Mycologia.* 73(6):1123-9.
- Hasegawa, M., Kishino, H., Yano, T.-A., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22, 160-174.
- Haugen P, Reeb V, Lutzoni F, Bhattacharya D. 2004. The evolution of homing endonuclease genes and group I introns in nuclear rDNA. *Mol Biol Evol.* 21(1):129-40.
- Haugen P, Bhattacharya D. 2004. The spread of LAGLIDADG homing endonuclease genes in rDNA. *Nucleic Acids Res.* 32(6):2049-57.
- Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. *Trends Genet* 21(2):111-9.
- Haugen P, Bhattacharya D, Palmer JD, Turner S, Lewis LA, Pryer KM. 2007. Cyanobacterial ribosomal RNA genes with multiple, endonuclease-encoding group I introns. *BMC Evol Biol.* 7:159.
- Hausner G, Reid J, and Klassen GR. 1992. Do galeate-ascospore members of the Cephaloascaceae, Endomycetaceae and Ophiostomataceae share a common phylogeny? *Mycologia.* 84(6): 870–81.

- Hausner G, Reid J, Klassen GR. 1993a. On the phylogeny of *Ophiostoma*, *Ceratocystis* s.s., *Microascus*, and relationships within *Ophiostoma* based on partial ribosomal DNA sequences. *Can J Bot.* 71:1249-65.
- Hausner G, Reid J, Klassen GR. 1993b. *Ceratocystiopsis* - a reappraisal based on molecular criteria. *Mycol Res.* 97(5):625-33.
- Hausner G, Reid J, Klassen GR. 2000. On the phylogeny of members of *Ceratocystis* s.s. and *Ophiostoma* that possess different anamorphic states, with emphasis on the anamorph genus *Leptographium*, based on partial ribosomal DNA sequences. *Can J Bot.* 78:903-6.
- Hausner G. 2003. Fungal mitochondrial genomes, plasmids and introns. In: Arora DK, Khachatourians GG, editors. *Applied Mycology and Biotechnology*, Vol. III: fungal genomics. New York: Elsevier Science. p. 101–31.
- Hausner G, Reid J. 2004. The nuclear small subunit ribosomal genes of *Sphaeronaemella helvellae*, *Sphaeronaemella fimicola*, *Gabarnaudia betae* and *Cornuvesica falcata*: phylogenetic implications. *Can J Bot.* 82: 752-62.
- Hausner G, Iranpour M, Kim JJ, Breuil C, Davis CN, Gibb EA, Reid J, Loewen PC, Hopkin AA. 2005. Fungi vectored by the introduced bark beetle *Tomicus piniperda* in Ontario, Canada, and comments on the taxonomy of *Leptographium lundbergii*, *Leptographium terebrantis*, *Leptographium truncatum*, and *Leptographium wingfieldii*. *Can J Bot.* 83(10):1222-37.
- Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, et al. 2007. A higher-level phylogenetic classification of the Fungi. *Mycol Res.* 111(5):509–47.
- Jacobs K, Bergdahl DR, Wingfield MJ, Halik S, Seifert KA, Bright DE, and Wingfield BD. 2004. *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. *Mycol Res.* 108: 411-18.
- Juzwik J, Harrington TC, MacDonald WL, Appel DN. 2008. The origin of *Ceratocystis fagacearum*, the oak wilt fungus. *Annu Rev Phytopathol.* 46:13-26.
- 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: Gesteland RF, Cech TR, Atkins JF, editors. *The RNA world*. New York: Cold Spring Harbor Laboratory Press. p. 451–85.

- Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. *Annu Rev Genet.* 38:1-35.
- Lambowitz AM, Zimmerly S. 2010. Group II Introns: Mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol* (In press). doi: 10.1101/cshperspect.a003616
- Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet.* 23:119-25.
- LaPolla RJ, Lambowitz AM. 1981. Mitochondrial ribosome assembly in *Neurospora crassa*. Purification of the mitochondrially synthesized ribosomal protein, S-5. *J Biol Chem.* 256(13):7064-7.
- Lazarevic V, Soldo B, Düsterhöft A, Hilbert H, Mauël C, Karamata D. 1998. Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPbeta. *Proc Natl Acad Sci USA.* 95(4):1692-7.
- Marcaida MJ, Muñoz IG, Blanco FJ, Prieto J, Montoya G. 2010. Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci.* 67(5):727-48.
- Marin M, Preisig O, Wingfield BD, Kirisits T, Yamaoka Y and Wingfield MJ. 2005. Phenotypic and DNA sequence data comparisons reveal three discrete species in the *Ceratocystis polonica* species complex. *Mycol Res.* 109(10): 1137–48.
- Marmolejo JG, Butin H. 1990. New conifer-inhabiting species of *Ophiostoma* and *Ceratocystiopsis* Ascomycetes Microascales from Mexico. 42:193-9.
- Massoumi Alamouti S, Kim JJ, Humble LM, Uzunovic A, Breuil C. 2007. Ophiostomatoid fungi associated with the northern spruce engraver, *Ips perturbatus*, in western Canada. *Antonie Van Leeuwenhoek.* 91(1):19-34.
- Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol.* 216(3):585-610.
- Michel F, Ferat JL. 1995. Structure and activities of group II introns. *Annu Rev Biochem.* 64:435-61.
- Michel F, Costa M, Doucet AJ, Ferat JL. 2007. Specialized lineages of bacterial group II introns. *Biochimie.* 89(4):542-53.
- Mohr G, Rennard R, Cherniack AD, Stryker J, Lambowitz AM. 2001. Function of the *Neurospora crassa* mitochondrial tyrosyl-tRNA synthetase in RNA splicing. Role of the idiosyncratic N-terminal extension and different modes of interaction with different group I introns. *J Mol Biol* 307:75-92

- 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):837-48.
- Mota EM, Collins RA. 1988. Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. *Nature.* 332(6165):654-6.
- Mullineux ST, Willows K, Hausner G. 2011. Evolutionary Dynamics of the mS952 Intron: A Novel Mitochondrial Group II Intron Encoding a LAGLIDADG Homing Endonuclease Gene. *J Mol Evol.* (In press). doi: 10.1007/s00239-011-9442-7
- Nicholas KB, Nicholas HB Jr, Deerfield DW. 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEWNEWS.*4:14.
- Nielsen H, Johansen SD. 2009. Group I introns: Moving in new directions. *RNA Biol.* 6(4):375-83.
- Ohtaka N, Masuya H, Yamaoka Y, Kaneko S. 2006. Two new *Ophiostoma* species lacking conidial states isolated from bark beetles and bark beetle-infested *Abies* species in Japan Source. *Can J Bot.* 84(2):282-93.
- Okada G, Jacobs K, Kirisits T, Louis-Seize GW, Seifert KA, Sugita T, Takematsu A, Wingfield MJ. 2000. Epitypification of *Graphium penicillioides* Corda, with comments on the phylogeny and taxonomy of Graphium-like synnematosus fungi. *Stud Mycol.* 45:169–88.
- Olchowecki A, Reid J. 1974. Taxonomy of the genus *Ceratocystis* in Manitoba. *Can J Bot.* 52:1675–711.
- Page RD. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci.* 12:357-58.
- Paquin B, Laforest M-J, Lang F. 1994. Interspecific transfer of mitochondrial genes in fungi and creation of a homologous hybrid gene. *Proc Natl Acad Sci USA.* 91(25):11807-10
- Pel HJ, Grivell LA. 1993. The biology of yeast mitochondrial introns. *Mol Biol Rep.* 18(1):1-13.
- Plattner A, Kim JJ, Reid J, Hausner G, Lim YW, Yamaoka Y, Breuil C. 2009. Resolving taxonomic and phylogenetic incongruence within species *Ceratocystiopsis minuta*. *Mycologia.* 101(6):878-87.

- Pyle AM. 2010. The tertiary structure of group II introns: implications for biological function and evolution. *Crit Rev Biochem Mol Biol.* 45(3):215-32.
- Raghavan R, Minnick MF. 2009. Group I introns and inteins: disparate origins but convergent parasitic strategies. *J Bacteriol.* 191(20):6193-202.
- Reid J, Hausner G. 2010. The epitypification of *Ophiostoma minutum*, now *Ceratocystiopsis minuta*. *Mycotaxon.* 113:463-74.
- Robart AR, Zimmerly S. 2005. Group II intron retroelements: function and diversity. *Cytogenet Genome Res.* 110(1-4):589-97.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 19:1572-4.
- Ronquist F. 2004. Bayesian inference of character evolution. *Trends Ecol Evol* 19:475-81.
- Saguez C, Lecellier G, Koll F. 2000. Intronic GIY-YIG endonuclease gene in the mitochondrial genome of *Podospira curvicolla*: evidence for mobility. *Nucleic Acids Res.* 28(6):1299-306.
- Salvo JL, Rodeghier B, Rubin A, Troischt T. 1998. Optional introns in mitochondrial DNA of *Podospira anserina* are the primary source of observed size polymorphisms. *Fungal Genet Biol.* 23:162-8.
- Sandegren L, Sjöberg BM. 2004. Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J Biol Chem.* 279(21):22218-27.
- Sandegren L, Sjöberg BM. Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. *J Bacteriol.* 189(3):980-90.
- Schäfer B. 2003. Genetic conservation versus variability in mitochondria: the architecture of the mitochondrial genome in the petite-negative yeast *Schizosaccharomyces pombe*. *Curr Genet.* 43:311-26.
- Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, 18(3): 502-504. doi:10.1093/bioinformatics/18.3.502. PMID:11934758.
- Sellem CH, Belcour L. 1994. The in vivo use of alternate 3'-splice sites in group I introns. *Nucleic Acids Res.* 22:1135-7.

- Sellem CH, d'Aubenton-Carafa Y, Rossignol M, Belcour L. 1996. Mitochondrial intronic open reading frames in *Podospora*: mobility and consecutive exonic sequence variations. *Genetics*. 143:777–788.
- Sellem CH, Belcour L. 1997. Intron open reading frames as mobile elements and evolution of a group I intron. *Mol Biol Evol*. 14:518–526.
- Semrad K, Schroeder R. 1998. A ribosomal function is necessary for efficient splicing of the T4 phage thymidylate synthase intron in vivo. *Genes Dev*. 12(9):1327-37.
- 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, Majer A, Friedrich N, Edgell D, Hausner G. 2009a. Genes-within-genes: Multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within a *rnl* group I intron of *Ophiostoma* and related taxa. *Mol Biol Evol*. 26(10):2299-315.
- Sethuraman J, Majer A, Iranpour M, Hausner G. 2009b. Molecular evolution of the mtDNA encoded *rps3* gene among filamentous ascomycetes fungi with an emphasis on the Ophiostomatoid fungi. *J Mol Evol*. 69(4):372-85.
- Simossis VA, Heringa J. 2003. The PRALINE online server: optimising progressive multiple alignment on the web. *Comput Biol Chem*. 27:511-9.
- Simossis VA, Heringa J. 2005. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res*. 33(Web Server issue): W289-94.
- Solheim H. 1986. Species of Ophiostomataceae isolated from *Picea abies* infested by the bark beetle *Ips typographus*. *Nordic Journal of Botany*. 6(2):199-207.
- 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.
- Tanguay P, Loppnau P, Morin C, Bernier L, Breuil C. 2006. A spontaneous albino mutant of *Ceratocystis resinifera* results from a point mutation in the polyketide synthase gene, PKS1. *Can J Microbiol*. 52(6):501-7.

- 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* 25:4876-82.
- Toor N, Zimmerly S. 2002. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA*. 8:1373-7.
- Turmel M, Otis C, Côté V, Lemieux C. 1997. Evolutionarily conserved and functionally important residues in the I-CeuI homing endonuclease. *Nucleic Acids Res.* 25(13):2610-9.
- Upadhyay HP, 1981. A monograph of *Ceratocystis* and *Ceratocystiopsis*. University of Georgia Press: Athens, G.A.
- Van Dyck L, Neupert W, Langer T. 1998. The ATP-dependent PIM1 protease is required for the expression of intron-containing genes in mitochondria. *Genes Dev.* 12:1515-24.
- Van Roey P, Derbyshire V. 2005. GIY-YIG homing endonucleases - Beads on a string. In: Belfort M, Stoddard BL, Wood DW, Derbyshire V, editors. *Homing Endonucleases and Inteins*. Volume 16. p. 67-83.
- Westermann B, Prokisch H. 2002. Mitochondrial dynamics in filamentous fungi. *Fungal Genet Biol.* 36(2):91-7.
- Wingfield MJ, Seifert KA, Webber JF (eds). 1993. *Ceratocystis* and *Ophiostoma*. Biology, Taxonomy and Ecology. American Phytopathological Society Press, St. Paul.
- Zhao L, Bonocora RP, Shub DA and Stoddard BL. 2007. The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J.* 26(9):2432-42.
- Zhou X, de Beer ZW, Cibrian D, Wingfield BD, Wingfield MJ. 2004. Characterisation of *Ophiostoma* species associated with pine bark beetles from Mexico, including *O. pulvinisporum* sp. nov. *Mycol Res.* 108(6):690-8.
- Zipfel RD, Wilhelm de Beer, Z, Jacobs K, Wingfield BD and Wingfield MJ. 2006. Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. *Stud Mycol.* 55: 75-97.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406-3415.