Exploring the *rns* gene landscape in ophiostomatoid fungi and related taxa: Molecular characterization of mobile genetic elements and biochemical characterization of intron-encoded homing endonucleases.

By

Mohamed Hafez Ahmed Abdel-Fattah

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

in partial fulfilment of the requirements of the degree of:

DOCTOR OF PHILOSOPHY

Department of Microbiology
Faculty of Science
University of Manitoba
Winnipeg, Manitoba
Canada

ABSTRACT

The mitochondrial small-subunit ribosomal RNA (mt. SSU rRNA = *rns*) gene appears to be a reservoir for a number of group I and II introns along with the intron-encoded proteins (IEPs) such as homing endonucleases (HEases) and reverse transcriptases. The key objective for this thesis was to examine the *rns* gene among different groups of ophiostomatoid fungi for the presence of introns and IEPs. Overall the distribution of the introns does not appear to follow evolutionary lineages suggesting the possibility of rare horizontal gains and frequent loses. Some of the novel findings of this work were the discovery of a twintron complex inserted at position S1247 within the *rns* gene, here a group IIA1 intron invaded the ORF embedded within a group IC2 intron. Another new element was discovered within strains of *Ophiostoma minus* where a group II introns has inserted at the *rns* position S379; the mS379 intron represents the first mitochondrial group II intron that has an RT-ORF encoded outside Domain IV and it is the first intron reported to at position S379.

The *rns* gene of *O. minus* WIN(M)371 was found to be interrupted with a group IC2 intron at position mS569 and a group IIB1 intron at position mS952 and they both encode double motif LAGLIDADG HEases referred as I-OmiI and I-OmiII respectively. These IEPs were examined in more detail to evaluate if these proteins represent functional HEases. To express I-OmiI and I-OmiII in *Escherichia. coli*, a codon-optimized versions of I-OmiI and I-OmiII sequences were synthesized based on

differences between the fungal mitochondrial and bacterial genetic code. The optimized I-OmiI and I-OmiII sequences were cloned in the pET200/D TOPO expression vector system and transformed into *E. coli* BL21 (DE3). These two proteins were biochemically characterized and the results showed that: both I-OmiI and I-OmiII are functional HEases. Detailed data for I-OmiII showed that this endonuclease cleaves the target site two nucleotides upstream of the intron insertion site generating 4 nucleotide 3'overhangs.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis specially my advisor **Dr. Georg Hausner** for giving me the opportunity to work in his lab and supervising my PhD project. This thesis would not have been possible without his support, guidance and encouragement. I would like to thank my thesis supervision committee members, **Dr. Deborah Court**, **Dr. Ivan Oresnik** and **Dr. Michele Piercey-Normore** for their help, support, encouragement and their valuable comments. I am heartily thankful to **Dr. James Reid** for providing me with fungal cultures from his culture collections and also for his help, support, interest and valuable hints. I would like also to thank my advisor, my thesis supervision committee members as well as my external examiner **Dr. William Hintz** (University of Victoria) for their general and specific constructive comments that have greatly enhanced the value of this thesis.

I would like to thank my former and present lab mates; **Dr. Mahmood Iranpour**, **Dr. Jyothi Sethuraman**, **Dr. Sara-Taylor Mullineux**, **Shelly Rudski**, **Chen Shen**, **Iman Bilto** and **Michael Pogorzelec** for supporting me in my research work, and I would like also to thank my awesome lab mate **Tuhin Kumar Guha** for helping in I-OmiI purification. I would like to express my deepest gratitude to all the faculty members and graduate students in the Department of Microbiology specially **Dr. Linda Cameron**, **Dr. Chris Rathgeber**, **Aniel Moya-Torres** and **Munmun Nandi**. Not to forget, great

appreciation goes to **Sharon Berg**, **Madeleine Harris**, **Karen Hamill** and **Stephanie Moorhouse** for continuous help and support during the project.

I am indebted to many of my teachers and friends for shaping my personality and giving me the enthusiasm to complete this thesis and for everything, especially: **Dr. Abdel-Wahed Moustafa**, **Dr. Akram Abu Seadah**, **Dr. Essam K. Ezz Eldin**, **Dr. Ahmed Abdel-Azeem**, **Dr. Moustafa Awad**, **Mr. Khaled Salah**, **Hossam Abdel Moniem**, **Belal Saleh**, **Tamer Saad**, **Ahmed Khalaf**, **Mohamed Fawzy**, **Alaaa Rashad**, **Mohamed Ali**, **Hatim Hemeda**, **Ezz-Eldin Mohamed** and **Mohamed Mostafa**.

Financial support of this project was provided by Discovery grants from the Natural Sciences and Engineering Research Council of Canada (**NSERC**) to Dr. G. Hausner and from the **Egyptian Ministry of Higher Education and Scientific Research** through the Bureau of the Egyptian cultural and Educational Affairs in Canada to M. Hafez. I also would like to acknowledge the financial support from the Faculty of Graduate Studies, the Faculty of Science and the Graduate Students' Association at the University of Manitoba for supporting me financially to attend and present my thesis results in scientific conferences.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project and I would like to dedicate this work to **my family**, **my parents**, **my siblings**, **my wife** and my beloved daughters **Nada** and **Mariam**.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BLAST: **B**asic **L**ocal **A**lignment **S**earch **T**ool

coll: **c**o**l**ony type gene

cp: **c**hloro**p**last

CTAB: **C**etyl **T**rimethyl **A**mmonium **B**romide

cu: ***c****erato-****u****lmi* gene

DED: **D**utch **E**lm **D**isease

EBS: **E**xon **B**inding **S**ite

EF-1α: **E**longation **F**actor **1 a**lpha

ENase: **En**donucle**ase**

EtBr: **Et**hidium **Br**omide

ETS: **E**xternal **T**ranscribed **S**pacer

HEase: Homing **E**ndonucle**ase**

HEG: Homing **E**ndonuclease **G**ene

IBS: **I**ntron **B**inding **S**ite

IEP: **I**ntron-**E**ncoded **P**rotein

IGS: **I**nternal **G**uide **S**equence

IPTG: **I**so**p**ropyl β-D-1-**t**hio**g**alactopyranoside

IS: **I**nsertion **S**ite

ITS: **I**nternal **T**ranscribed **S**pacer

LB: **L**uria **B**ertani

LHEase: LAGLIDADG **H**oming **E**ndonucle**ase**

LSU: **L**arge **S**ub**u**nit

LSU rRNA: **L**arge **S**ubunit **r**ibosomal **RNA**

MEA: **M**alt **E**xtract **A**gar

MPB: **M**ountain **P**ine **B**eetles

mRNA: **m**essenger **RNA**

mt: **m**i**t**ochondrial

mtDNA: **m**itochondrial **DNA**

NCBI: National **C**enter for **B**iotechnology **I**nformation

NHEJ: Non-**H**omologous **E**nd **J**oining

nt: **n**ucleo**t**ide

ORF: **O**pen **R**eading **F**rame

PCR: **P**olymerase **C**hain **R**eaction

PYG: **P**eptone **Y**east extract **G**lucose

REase: **R**estriction **E**ndonucle**ase**

RFLP: **R**estriction **F**ragment **L**ength **P**olymorphisms

rnl: mitochondrial **l**arge subunit **r**ibosomal R**N**A gene

RNP: **R**ibo**n**ucleo**p**rotein

rns: mitochondrial **s**mall subunit **r**ibosomal R**N**A gene

rRNA: **r**ibosomal **RNA**

RT: **R**everse **T**ranscriptase

RT-PCR: **R**everse **T**ranscription **P**olymerase **C**hain **R**eaction

SDS: **S**odium **D**odecyl **S**ulphate

SOC: **S**uper **O**ptimal broth with **C**atabolite repression

SSU: **S**mall **S**ub**u**nit

SSU rRNA: **S**mall **S**ub**u**nit **r**ibosomal **RNA**

TBE: **T**ris-**B**orate **E**DTA

TE buffer: **T**ris-**E**DTA **buffer**

TPRT: **T**arget DNA-**P**rimed **R**everse **T**ranscription

tRNA: **t**ransfer **RNA**

UV: **U**ltra **V**iolet

YE: **Y**east **E**xtract

2X-YT: **2X Y**east extract **T**ryptone

GENERAL INTRODUCTION

Fungi are one of the most important groups of organisms on this planet. Fungi are involved in beneficial processes as they act: as decomposers and thus recycling the dead materials into reusable forms; as symbionts in mycorhizal and lichen associations; as sources of medical and chemical materials like antibiotics and organic acids as well as biocontrol agents that control insect pests and other plant pathogens. Some fungi are used in food production (yeasts) and certain fungi are edible (mushrooms, morels and truffles). Many fungi (*Saccharomyces cerevisiae*, *Neurospora crassa* and *Podospora anserina*) have been used extensively as model systems to understand the principles of genetics. On the other hand, fungi can cause harmful effects; a number of fungi are considered pathogenic and cause diseases to humans, animals and plants and fungi play an important role in food spoilage.

Ophiostomatoid fungi are ascomycetous fungi, and some members of this group are aggressive pathogens and cause diseases of economically important trees like Dutch Elm Diseases (DED) that has devastated many urban forests. The causative agents of DED are *O. ulmi* (Buism.) Nannf., and subspecies of *O. novo-ulmi* Brasier. Historically there have been two epidemics of DED, the first caused by *O. ulmi* and the current pandemic of this disease that is caused by *O. novo-ulmi*. The ophiostomatoid fungi also contain members (like *O. minus*) that cause a blue-staining in the wood of infected trees. Taxonomically this group of fungi is quite complex and several genera are currently designated to as ophiostomatoid fungi (see Hausner et al., 2003; Zipfel et al., 2006). Typically these fungi produce short-lived asci causing the release of the ascospores

(meiotic spores) from the perithecia and the ascospores "ooze" or are pushed out through the ostiole of usually long necked perithecia. This essentially causes sticky spore masses (droplets) to accumulate at the tip of the perithecial neck and these spores are vectored by insects such as bark beetles. Overall this group of fungi has been poorly studied with regards to their mitochondrial genomes.

Mitochondrial genomes are highly variable in size (16 – 250 kb; Palmer, 1990; Boore, 1999). The size of fungal mitochondrial DNA (mtDNA) varies greatly (19 – 175 kb; Hausner, 2003) due to intergenic spacers and the presence of self-splicing group I and II introns as well as intron encoded proteins (IEPs) (Michel and Ferat, 1995; Hausner, 2012). Self-splicing introns are autocatalytic genetic elements that can catalyze their removal from the primary RNA transcripts to allow for efficient gene expression. These self-splicing introns can be differentiated by their splicing mechanisms and secondary and tertiary RNA structures into group I and group II introns (Michel and Westhof, 1990; Michel and Ferat, 1995; Bonen and Vogel, 2001; Federova and Zingler, 2007). Both types of introns are considered ribozymes (Saldanah et al., 1993; Belfort et al., 2002); however splicing is assisted by intron and host genome encoded maturases/splicing factors (Hausner, 2003 & 2012). Self-splicing introns are widespread in the organellar genomes of plant, fungi and algae as well as bacterial genomes (Lambowitz and Zimmerly, 2004 & 2011; Hausner, 2003 & 2012). Overall very little is known about the types of introns that are present within the mtDNA or nuclear rDNA of ophiostomatoid fungi. These introns might be potential taxonomic markers or could be agents that cause genetic instabilities, as recently observed for an *rns* group II intron in *Cryphonectria*

parasicita (Baidyaroy et al., 2011). Therefore, cataloging the types of introns present among these fungi might be a valuable resource in future studies.

HEases are encoded by Homing Endonuclease Genes (HEG) which are embedded within group I introns, group II introns and archael introns, as well as inteins (Stoddard, 2006). Homing endonucleases can be classified into four major distinct families based on conserved amino acid motifs: LAGLIDADG, H-N-H, His-Cys Box, and GIY-YIG (Belfort et al., 2002; Kowalski and Derbyshire, 2002; Stoddard, 2006). Recently additional HEases have been discovered: PD-(D/E)XK (Zhao et al., 2007); vsr endonucleases (Dassa et al., 2009); Holliday junction resolvase-like (Zeng et al., 2009) and the EDxHD family (Taylor et al., 2011). The LAGLIDADG and GIY-YIG families of HEases are most frequently encountered among fungal mitochondrial group I introns (Stoddard, 2006). Typically, group II introns usually encode multifunctional proteins known as reverse transcriptases (RTs). Some group II introns have been noted to encode H-N-H-type HEases or in a few instances LAGLIDADG-type HEases (LHEases) (Michel and Ferat, 1995; Toor and Zimmerly, 2002; Mullineux et al., 2010).

Fungal mitochondrial genomes can be considered a natural reservoir for mobile elements (group I and group II) as well as IEPs (HEases and RTs). Studying the fungal mtDNA will improve our understanding of the evolutionary dynamics of these mobile elements and the potential impact these elements have had on the evolution of the fungal mtDNA (Dujon, 1989; Charter et al., 1996; Belcour et al., 1997; Salvo et al., 1998; Hamari et al., 1999; Gobbi et al., 2003). Group I and II introns have been associated with

mtDNA instabilities such as generating plasmid-like elements that are found in senescent and/or hypovirulent strains in an assortment of filamentous fungi such as *Podospora anserina* (Osiewacz and Esser, 1984; Michel and Cummings, 1985; Cummings et al., 1986; Dujon and Belcour, 1989; Cummings et al., 1990), *Ophiostoma novo-ulmi* (Abu-Amero et al., 1995; Sethuraman et al., 2008) and *Cryphonectria parasitica* (Monteiro-Vitorello et al., 2009; Baidyaroy et al., 2011).

Although mtDNA mobile elements, such as the self-splicing introns and their encoded proteins, are of interest as model systems for understanding how mobile elements can shape genomes and how mobile elements are essentially neutral elements to avoid extinction, these elements are also of interest due to their applications in biotechnology. Ribozymes and DNA cutting enzymes have many potential applications in molecular biology; therefore bioprospecting for functional native HEases within the fungal mtDNA can be fruitful work to characterize rare cutting HEases that can be used to manipulate genomes in medical and agricultural systems.

CHAPTER 1:

LITERATURE REVIEW

1.1. Ophiostomatoid fungi:

Ophiostomatoid fungi are a group of ascomycetous fungi (Sordariomycetes, Ophiostomatales, Ophiostomataceae) that often can be characterized by possessing long-necked perithecia that contain scattered evanescent (short-lived) asci. Historically these fungi were assigned into two genera: *Ophiostoma* and *Ceratocystis* (Upadhyay, 1981; reviewed in Hausner et al., 1993a). Recently additional genera have been recognized to accommodate this very diverse group of fungi (Zipfel et al., 2006). The asci are short lived and the ascospores are pushed out of the perithecial neck in the form of slimy spore droplets. These fungi produce asexual spores in either an enteroblastic (*Ophiostoma*) or holoblastic (*Ceratocystis*) fashion, again in many instances slimy masses of spores are produced at the tip of sometimes long stalked conidiophores. Overall both mitotic and meiotic spores are produced in a manner that facilitates spore dispersal by insects. The genus *Ophiostoma* contains both plant pathogens such as DED causative agents (*Ophiostoma ulmi*) and blue stain fungi like *Ophiostoma piliferum* and *Ophiostoma minus*. The majority of the ophiostomatoid fungi (including species of the following genera: *Ophiostoma*, *Grosmannia*, *Ceratocystiopsis* and *Ceratocystis*) are transmitted with the help of bark beetles (Coleoptera: Scolytidae) and this association between the fungus and the beetle is important for both partners in establishing diseases and in generating "blue-staining" of timber (Wingfield et al., 1993; Kirisits, 2004). Several phylogenetic studies based on rDNA sequence analysis showed that the two "classical"

genera of ophiostomatoid fungi, *Ophiostoma* and *Ceratocystis*, have polyphyletic origins with species of *Ophiostoma sensu lato*. They are assigned to the Ophiostomales, and species of *Ceratocystis* show affinities within the Order of the Microascales (Hausner et al., 1993a; Spatafora and Blackwell, 1994), and generally it was found that genera belonging to the ophiostomatoid fungi share morphological characteristics that are due to convergent evolution (Zipfel et al., 2006). The concept of *Ophiostoma* has also been revised and species that used to be accommodated within *Ophiostoma sensu lato* are now placed into *Grosmannia*, *Ceratocystiopsis*, and a more narrowly defined concept of *Ophiostoma* (see Zipfel et al., 2006). Features used to segregate fungi into these genera are the type of conidial (asexual) structures produced and the shape of the ascospores.

1.2. Dutch elm disease and blue-stain fungi:

Some ophiostomatoid fungi are aggressive pathogens of economically important tree species like conifers, elm and oak, while some others are causative agents of blue (sap) stain for conifers and hardwood trees (Upadhyay, 1993; Wingfield et al., 1993). Dutch elm disease is a fungal disease that affects all common forms of elm species but it has its most serious effects on *Ulmus americana*, i.e. the American Elm. The causative agents of DED are several species of ascomycetous microfungi and these are fungi transmitted with the help of bark beetles. Three DED causing species are now recognized: *Ophiostoma ulmi*, which afflicted Europe and North America during the first half of the 20th century; the second species is the more virulent species *Ophiostoma novo-ulmi*, which was first described in Europe and North America in the 1940s and has devastated elms in both areas since the late 1960s; and the third species *Ophiostoma himal-ulmi*, a species endemic to the western Himalayas. These fungi infect the sap wood

of the tree and eventually they block the conductive tissue of the elm tree, thus preventing the flow of nutrients and water, thereby killing the tree. The disease, depending on environmental conditions, can destroy a tree in as little as 3 weeks or it can take as long as 3 years (Brasier, 1979; Hintz et al., 1993; Brasier and Kirk, 2001).

Blue-stain is a discoloration observed in the wood of infected trees (i.e. pine and elm) caused by a variety of fungi, such as the ophiostomatoid fungi including species of *Ophiostoma*, *Ceratocystis*, *Pesotum* and *Leptographium* (the latter two are anamorphic states of some *Ophiostoma* species; Kirisits, 2004). Blue-stain fungi are pioneer organisms that can metabolize the simple carbohydrate compounds, fatty acids, triglycerides and other chemical compositors found in xylem vessels of the sap wood (Harrington et al., 1998; Jacobs and Wingfield, 2001). These stains are also referred to as sap stains, because the fungi affect sapwood regions of trees where food and oxygen are readily available. Blue-stain can greatly reduce the value of the wood. Millions of dollars are lost every year due to blue staining of the wood and timber products. The cause of the stain is due to the production of melanin granules within the fungal hyphal cell walls **(**Zink and Fengel, 1989 & 1990)**.** Some blue-stain fungi are relatively pathogenic and they can kill mature host trees in the absence of their bark beetle vectors; other blue-stain fungi appear to work with their insect partner on causing disease symptoms. Some blue-stain-fungi are associated with Mountain Pine Beetles (MPBs) and when these beetles attack their primary host trees the beetle release pheromones to attract other beetles to colonize this tree (Gara et al., 1984). The sharp mouthparts of the adult MPB are ideally suited for boring through the bark to make long, vertical galleries where the eggs are laid.

The spores of blue-stain fungi are carried in these mouth parts as well as on the insect's body. Blue-stain fungal spores clog up the water-conducting vessels (xylem) of the tree; in addition, the insect larvae disrupt the phloem layer. By disrupting both the xylem and phloem of pine trees, the MPB larvae and their accompanying blue-stain fungi make a deadly combination (Christiansen, 1985; Malloch and Blackwell, 1993).

O. minus is an economically important species because it is a serious agent of blue-stain of pine trees. *O. minus sensu lato* probably represents a species complex (Gorton et al., 2004) characterized by perithecia with relatively short necks; this fungus is commonly associated with the southern pine beetle *Dendroctonus frontalis* in North America, whilst in Europe it is frequently associated with the beetle vector *Tomicus piniperda* (Klepzing, 1998).

1.3. Fungal mitochondrial genome:

Mitochondrial DNA (mtDNA) is the DNA located in mitochondria of eukaryotes which contains essential protein coding genes for oxidative phosphorylation as well as components needed for translation such as transfer RNA (tRNA) and ribosomal RNA (rRNA) genes. Comparative sequence analysis showed that nuclear and mtDNA in eukaryotes are thought to have different evolutionary origins and based on the endosymbiotic theory (first introduced by Mereschkowski in 1905 and later developed and popularized by Margulis in 1970). Certain eukaryotic organelles might have been derived from free living bacteria that were taken inside a protoeukaryotic cell as endosymbionts. In general, it has been accepted that mitochondria developed from an

alpha-proteobacteria endosymbiont (reviewed in Emelyanov, 2001; Gray et al., 2001; Hausner, 2012).

Generally, we can classify the mitochondrial genes into four main groups as follows, (1) 11 genes encoding the subunits of the respiratory chain [*cox1, cox2* and *cox3*: encoding subunits of the cytochrome oxidase; *nad1, nad2, nad3, nad4, nad4L, nad4* and *nad6* encoding subunits of the NADH dehydrogenase and *cyb* encoding cytochrome b], (2) 3 genes encoding subunits of the ATP synthase [*atp6, atp8* and *atp9*], (3) 2 rRNA genes; the small subunit ribosomal RNA [mt SSU rRNA also known as *rns*] gene and the mitochondrial large subunit ribosomal RNA [mt LSU rRNA also known as *rnl*] gene. A complete set of tRNA genes are also present; however in some cases only few a tRNA genes are encoded within the mtDNA and the reminder of tRNA are encoded within the nuclear genome and tRNAs have to be imported into the mitochondria (Forget et al., 2002).

There is great variation in the sizes of mtDNA among eukaryotes; the average size of animal mtDNA is approximately 16 kb (Boore, 1999), whereas the plant mtDNA ranges upwards to 200 kb to 250 kb (Palmer, 1990). The fungal mtDNA size varies greatly in fungi from 19 kb in *Schizosaccharomyces pombe* to 175 kb in *Agaricus bisporus* (Hausner, 2003). The size variation in fungal mtDNA is mainly due to intergenic spacers and the presence or absence of elements like introns and intron encoded protein genes (Clark-Walker, 1992; Hausner, 2003 & 2012). Several fungal mitochondrial genome sequences have been completed and can easily be accessed at the

fungal mitochondrial genome project (GMGP: http://www.bch.umontreal.ca/People/lang /FMGP/seqprojects.html; Paquin et al., 1997). Recently, about 177000 new mitochondrial sequences were organized in the organellar genome database (GOBASE: http://gobase.bcm.umontreal.ca/; O'Brien, et al., 2009).

At least four distinct classes of introns have been identified based on structural and splicing characteristics: spliceosomal introns, archaeal tRNA, group I and group II introns (Haugen et al., 2004; Lang et al., 2007). Spliceosomal introns are found in nuclear protein coding genes and are removed with the help of a ribonucleoprotein spliceosome (Newman, 1994). Based on structural and functional similarities between Group II introns and spliceosomal introns, group II introns are thought to be the ancestor of spliceosomal introns (Madhani and Guthrie, 1992; Shukla and Padgett, 2002; Keating et al., 2010). Archaeal introns are found in the tRNA genes of members within the euryarchaeota as well as in the tRNA and rRNA genes of species belonging to the crenarchaeota (Lykke-Andersen et al., 1997). Archaeal introns are highly variable in sequence with few conserved nucleotides that form a bulge-helix-bulge motif at the intron exon junctions (Thompson et al., 1990; Kjems and Garrett, 1991). Some archaeal introns have been noted to encode LHEases (Dalgaard and Garrett, 1992). Archaeal introns are removed from their primary transcripts by means of RNA endonucleases (Lykke-Andersen et al., 1997; Haugen et al., 2004). With regards to organellar genomes so called group I and group II introns are more relevant and these elements are therefore discussed in more detail in the following sections.

1.4. Mobile introns:

Group I and group II introns are commonly referred to as mobile introns as they can move from an intron containing allele to a cognate allele that lacks the intron. The mobility mechanisms are usually site specific, thus the term “homing”, and mobility is mediated by proteins encoded within these introns. Intron homing is different from intron transposition, a term that is applied when an intron inserts itself at a different site; this is also referred to as ectopic integration. Group I and II introns can be distinguished from each other by their sequences, secondary and tertiary RNA structures, and splicing mechanisms. These introns are potential ribozymes catalyzing their own removal from the precursor RNA transcripts. Mobile introns essentially contain two functional components, the autocatalytic self-splicing intron RNA and an intron encoded protein (IEP), such as HEases or in group II introns a complex multifunctional RT-like protein.

Mobile introns have been noted in Eubacteria and their phages and within the Archaea. In Eubacteria group II introns frequently are encountered within other mobile elements such as conjugative transposons. With the exception of nuclear rDNA group I introns in some protozoans and fungi, group I and II introns in eukaryotes tend to be restricted to organellar genomes, although these introns are rare among the metazoan (Haugen et al., 2005). Mobile introns contribute towards genomic variability and they may influence the expression of the genes that host them.

1.5. Group I introns:

1.5.1. Distribution:

Group I introns are widespread RNAs found in a wide variety of organisms and are abundant in bacteria and eukaryotes as well as bacteriophages but so far none have been found in the archaea. Group I introns are found in both nuclear and organellar genomes of the eukaryotes; they are particularly abundant in fungal and algal organellar DNAs especially in the small subunit ribosomal RNA (SSU rRNA) gene, large subunit ribosomal RNA (LSU) gene, cytochrome oxidase subunit I gene, and within the transfer RNA (tRNA) genes of the chloroplast genomes. On the other hand, group I introns when found in nuclear genomes, are found only in rRNA genes. Group I introns are not found in animals abundantly, but recently they have been recorded in the mitochondrial genome of the sea anemone *Metridium senile* and the soft coral *Acropora tenuis* (reviewed in Belfort et al., 2002; Hausner 2003; Gissi et al., 2008).

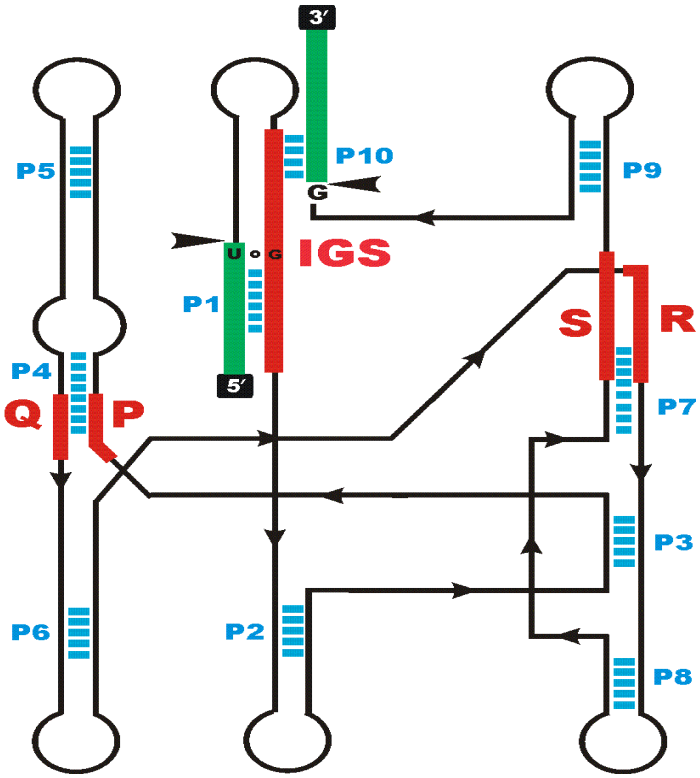
1.5.2. Structure:

Group I introns are highly variable at the primary sequence level; however they do have conserved secondary and tertiary structures and by comparative sequence analysis, it has been revealed that some of them do share short conserved sequences named P, Q, R, and S. These sequence elements had been defined within group I Introns and they do participate in forming helical regions, such that P is paired with Q and R is paired with S (Michel et al., 1982; Burke et al., 1987). Some features of the secondary structure of group I introns have been confirmed experimentally (Michel and Westhof, 1990; Woodson, 2005; Stahley and Strobel, 2005). The secondary structure of group I

introns consists of paired (P) elements designated P1 – P10 (Figure 1.1) with P1 and P10 containing the 5’ and 3’ splice sites, respectively. P2 is absent in some group I introns. The active core structure of the group I ribozyme is assembled by two helical domains P4/P6 (P4, P5 & P6), which is considered the scaffolding domain; and P3/P9 (P3, P7, P8 & P9) form the catalytic domain, which harbors the binding site for the exogenous guanosine (GTP) cofactor during the initiation of the intron splicing event. When open reading frames (ORFs) are present they are usually inserted in one of several loops that protrude from the core secondary structure, although there are examples where ORF sequences are part of the intron core sequences (Michel et al., 1982; Belfort et al., 2002; Rudski and Hausner, 2012). The ORF sequence may also extend and become part of the intron core structure and the ORF can also be fused in-frame to the upstream exon and is therefore most likely is translated as a fusion protein (Shub et al., 1988; Sethuraman et al., 2008). Based on secondary structure characteristics, nucleotide sequences within the conserved core regions, and peculiarities within the secondary structure group I introns have been classified into five categories, named class IA to IE and each class is subdivided into many subclasses (i.e. IA1, IA2 and IA3) (Saldanha et al., 1993; Suh et al., 1999; Zhou et al., 2008).

Figure 1.1. Secondary structure model of a group I intron. Intron sequences are represented by black lines, exon sequences are represented by green boxes. The ten pairing regions (P1-P10), the Internal Guide Sequence (IGS) region and the conserved sequence elements P, Q, R & S are indicated. The solid black arrowheads indicate the intron-exon junctions (5' and 3' splicing sites). This figure was generated based on the information obtained from Burke et al., (1987) and Saldanha et al., (1993).

Figure 1.1.



1.5.3. Splicing:

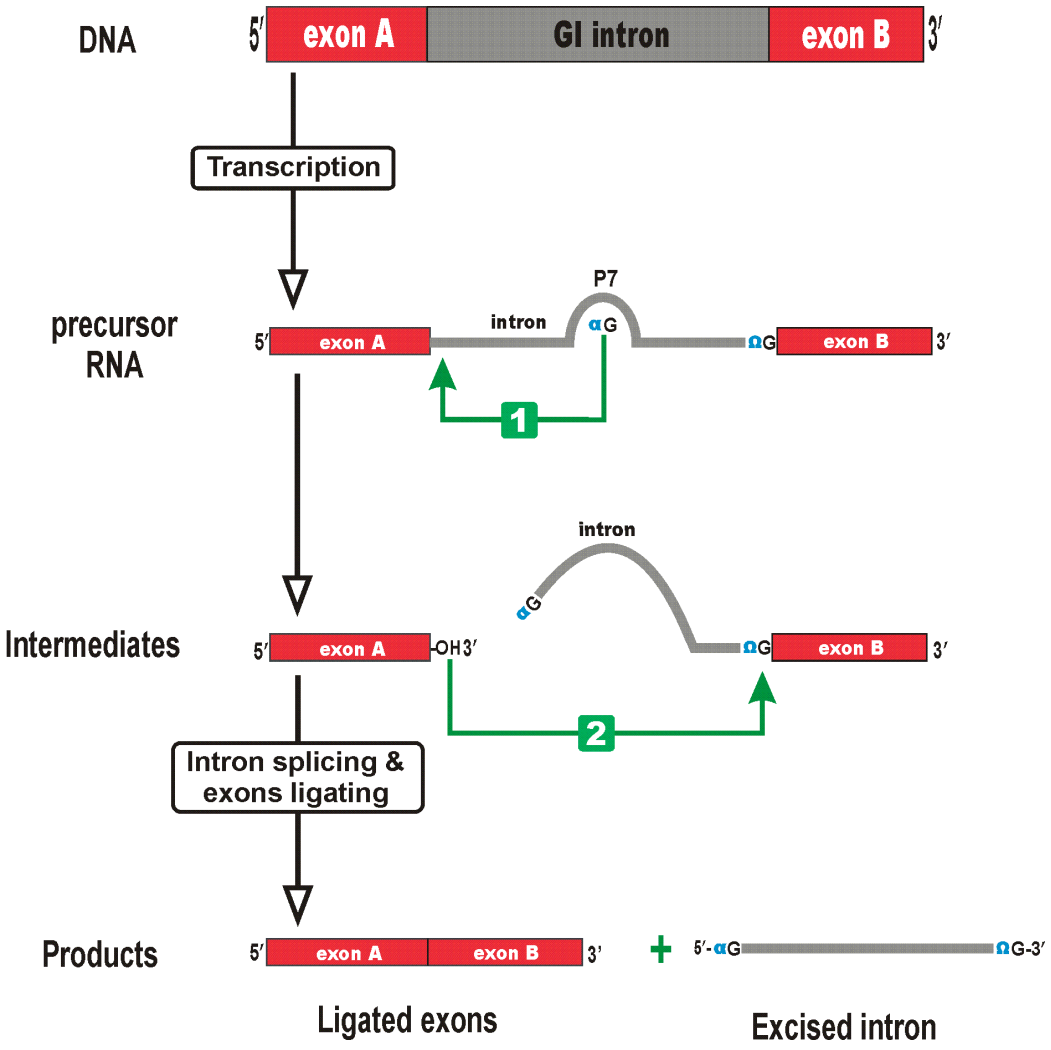
The introns are removed from the precursor RNA by an autocatalytic RNA splicing event that is mediated by the intron's RNA tertiary structure and proteins (maturases); the latter are intron and/or nuclear encoded. The reliance of group I introns on host factors for splicing further demonstrates that there is selection pressure on the group I intron, its encoded ORF and the host genome for the intron to splice accurately and quickly. In group I introns, base-pairing interactions between the 5' end of the intron and flanking exon regions define the location of the 5' and 3' splice sites. Splicing of the group I intron RNA is by a transesterification steps with an external guanosine (αG) as an initiating nucleophile; this eventually results in a linear, excised intron RNA. The αG docks onto an active G-binding site located in the P7 pairing region and with the GTPs 3'-OH group aligned to attack the 5' splice site (see Figure 1.2). The αG then attached to the 5' of the intron by 3'-5' phosphodiester bond. This is followed by conformational changes allowing the upstream exon's terminal 3' guanosine (ΩG) to swap with the αG and occupy the G-binding site to organize the second transesterification reaction (Michel et al., 1989). The 3' OH of the upstream exon attacks the 3' splice site leading to the ligation of upstream and downstream exons and the release of the intron (Saldanha et al., 1993; Cech, 1990; Cech et al., 1994). Group I intron splicing needs Mg^{2+} to stabilize RNA secondary and tertiary structure and to activate the nucleophilic attack by the 3'-OH group (Stahley and Strobel, 2005). The internal guide sequence (IGS), which is a short sequence near the 5'-end of group I intron sequence that pairs with sequences of the upstream exon to form P1 is an important sequence that can determine the 5' splice site. The 3' splice site is determined by paring of a short sequence of the downstream exon

with a portion of the IGS forming the P10 and the interactions between P9.0 and the catalytic core P3/P8 (see Figure 1.2; Cech, 1990; Michel et al., 1990).

Recently, crystal structures of several group I introns have been resolved: *Azoarcus* sp. BH72 pre-tRNA^{Ile} intron-exon complexes (Adams et al., 2004a & 2004b; Stahley and Strobel, 2005), *Tetrahymena* pre-rRNA apo enzyme (Guo et al., 2004), and the bacteriophage Twort pre-mRNA ribozyme-product complex (Golden et al., 2005). The crystal structures of these introns support the involvement of a two-metal-ion mechanism of group I intron splicing, which show similarities to the RNA-catalyzed splicing by metal-ion catalysis of the exon ligation step as seen in group II introns and in the spliceosome mediated mechanism (Stahley and Strobel, 2005).

Figure 1.2. Schematic representation of group I intron splicing. The splicing process consists of two transesterification reactions; the first reaction is initiated by exogenous guanosine (αG) which attacks the 5’ splice site. In the second transesterification reaction, the 3’-OH of the 5’ exon attacks the 3’ splice site and as a result the intron is spliced in a linear configuration and the exons are ligated. This figure was generated based on the information obtained from Saldanha et al. (1993), Cech (1990) and Cech et al. (1992).

Figure 1.2.


DNA
5′
exon A
GI intron
exon B
3′
Transcription
precursor RNA
P7
intron
αG
ΩG
exon B
1
Intermediates
OH3′
2
Intron splicing & exons ligating
Products
Ligated exons
5′-αG
ΩG-3′
Excised intron

1.5.4. Mobility:

Homing is the transfer of an intervening sequence (intron or intein) to a homologous allele that lacks the intervening sequence leading to gene conversion and the dominant transmission and inheritance of the mobile element (Dujon, 1989; Dujon et al., 1989; Belfort and Perlman, 1995); sometimes referred to as "super Mendielian inheritance" (Dujon, 1989). This process was firstly described in the *S. cerevisiae* group I intron HEase named omega (ω) or I-SceI according to the present nomenclature (Dujon, 1980). Group I intron mobility is catalyzed by the intron encoded HEases (Figure 1.3.). The HEases are usually cis-acting, and have specific target sites, with some allowance for sequence variation in their homing sites (target cleavage site). This ensures propagation against the forces of evolutionary drift which might modify the homing site within its host genome. Intron homing is initiated by the HEase introducing a double-strand break (DSB), or nick, in alleles that lack the intron. The homing process is completed by host DSB-repair pathways that use the intron/HEG-containing allele as a template to repair the DSB. The repair results in the nonreciprocal transfer of the intron/HEG composite element into the intron-minus allele, and is usually associated with co-conversion of markers flanking the intron insertion site (Saldanha et al., 1993).

Group I introns may also transpose into new sites (ectopic integration) involving RNA intermediates by reverse splicing (Figure 1.4. [a]). Reverse splicing allows a free group I intron RNA to insert into a homologous or heterologous RNA; this mode of mobility requires complementary base pairing between the intron and the exon RNA sequences. This mechanism requires the additional steps of reverse transcription of the

RNA and the integration of the cDNA into the genome by a recombination step that replaces the “intron-less copy” with the intron-plus cDNA. As reverse splicing requires less homology (4-6 nucleotides) this mechanism allows introns to spread more easily into heterologous sites (i.e. ectopic integration; Roman and Woodson 1995; Birgisdottir and Johansen, 2005). The best evidence for reverse splicing has been documented for rDNA introns where orthologous introns have been noted to be inserted in two different locations within rDNA genes (Bhattacharya et al., 2005).

Figure 1.3. Group I intron homing pathway: The black strands represent sequences of the donor allele, gray strands represent sequences of the recipient allele, red lines symbolize the intron sequence, straight and wavy lines represent DNA and RNA sequences respectively. The green pac-man symbol represent the intron-encoded endonuclease which recognizes and cleaves the target site (yellow boxes) generating 4 nt 3' overhangs, these can be chewed back by exonucleases to further stimulate the double strand break recombination DNA repair pathway, which would transfer a copy of the intron to the cleaved recipient molecule during the DSB repair process.

Figure 1.3.

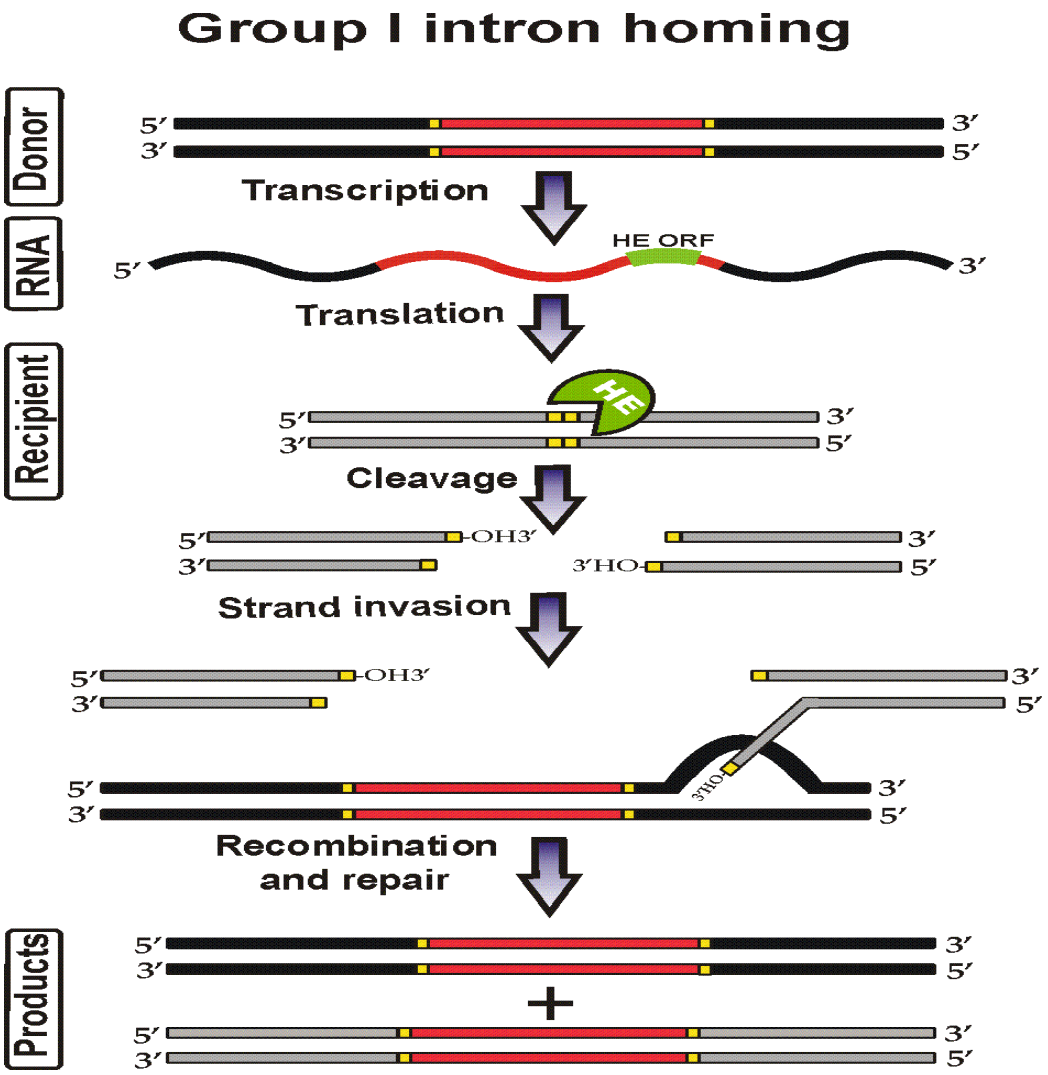
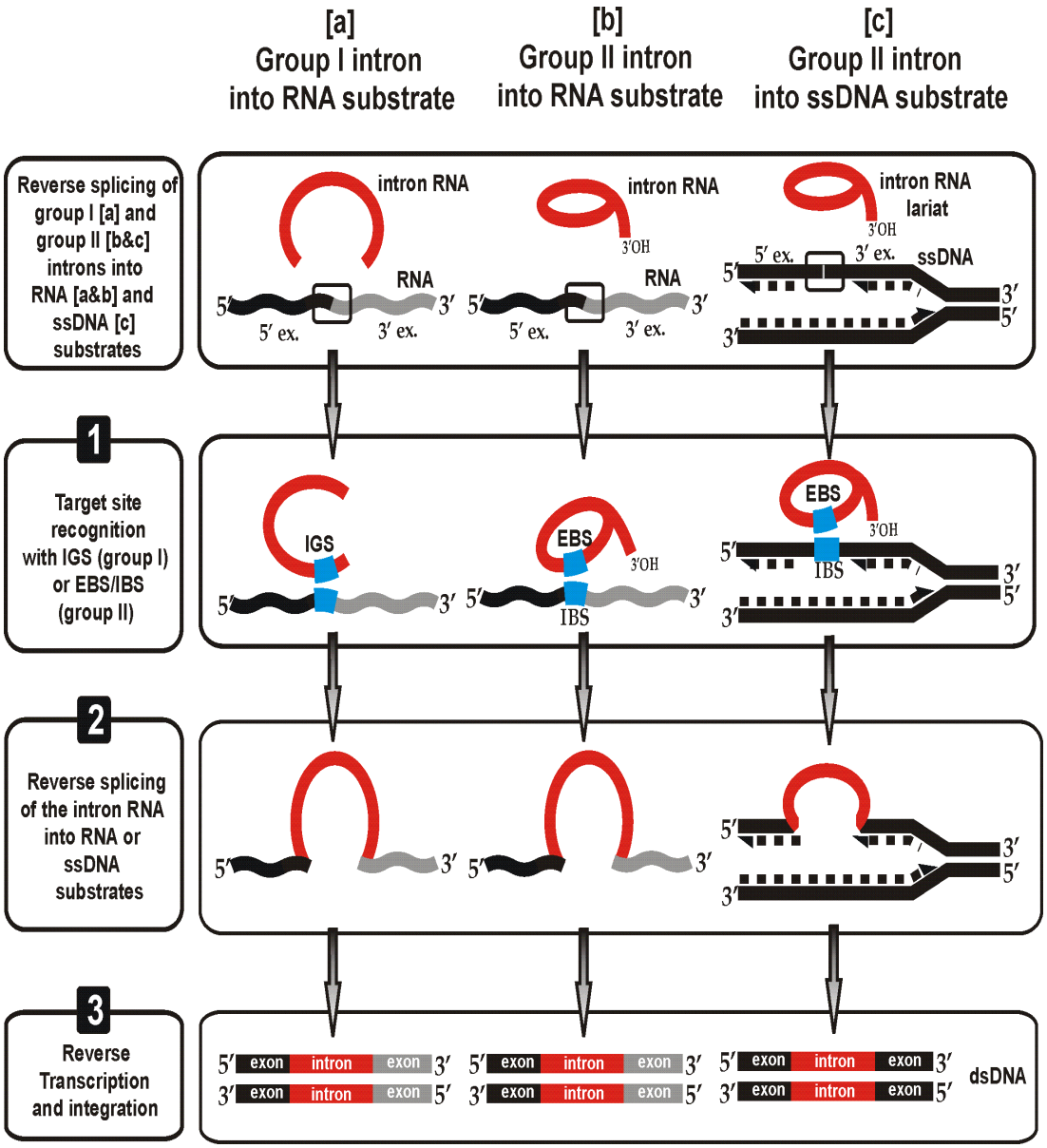

Group I intron homing
Donor
RNA
Recipient
Products
5′
3′
Transcription
HE ORF
Translation
HE
Cleavage
OH3′
3′HO
Strand invasion
Recombination
and repair

Figure 1.4. Reverse splicing of group I **[a]** and group II **[b]** intron RNAs (red lines) into RNA substrates (grey and black wavy lines). Target site recognition requires the interaction of the internal guide sequence (IGS) of the group I intron RNA with the recipient RNA or the exon binding sequence (EBS) of the group II intron RNA with the intron binding sequence (IBS) in the recipient RNA molecule. Reverse splicing requires a reverse transcription step to generate cDNA followed by homologous recombination that integrates the intron containing cDNA into the genome. **[c]** Group II intron RNA can also reverse splice into single-stranded DNA (black solid lines) as found at the DNA replication fork and again a reverse transcriptase step is required for cDNA synthesis but there is no need for endonuclease (En) activity of homologous recombination for integration of the intron to be completed (black and purple). Reverse transcription is primed by the nascent leading strand of the DNA replication fork (black dashed lines). These mechanisms have a lower homology requirement for target site recognition and therefore can promote the retrotransposition of introns into new locations within the genome.

Figure 1.4.


[a]
Group I intron
into RNA substrate
[b]
Group II intron
into RNA substrate
[c]
Group II intron
into ssDNA substrate
Reverse splicing of group I [a] and group II [b&c] introns into RNA [a&b] and ssDNA [c] substrates
intron RNA
RNA
5′ ex.
3′ ex.
3′OH
intron RNA
lariat
ssDNA
1
Target site recognition with IGS (group I) or EBS/IBS (group II)
IGS
EBS
IBS
2
Reverse splicing of the intron RNA into RNA or ssDNA substrates
3
Reverse Transcription and integration
exon
intron
dsDNA

1.6. Group II introns:

Group II introns are retroelements that are thought to be the ancestors of the spliceosomal introns and retrotransposons in eukaryotes (Copertino and Hallick, 1993; Lambowitz and Zimmerly, 2004 & 2011). Group II introns in general have conserved secondary structures at the RNA level, that can be visualized as six stem-loop domains (DI to DVI) emerging from a central wheel. The RT ORFs tend to be embedded within domain IV but in a few instances they have been observed in domain II (Simon et al., 2008; Hafez and Hausner, 2011a). Mobile group II introns typically encode a multifunctional protein, which contains a segment homologous to reverse transcriptase (RT), domain X, which has been implicated in maturase activity and the En domain which contains a potential zinc finger which has endonuclease activity. However, the En domain is absent in some fungal group II introns (Lambowitz and Zimmerly, 2004 & 2011).

1.6.1. Distribution:

Group II introns are found in bacteria and organellar genomes of various eukaryotes within plants, fungi and protists (Belfort et al., 2002; Ferat and Michel, 1993; Michel et al., 1989; Lambowitz and Zimmerly, 2011) and recently a group II inron was identified in animals for the first time inside the mitochondrial *cox1* gene of the bilaterian worm *Nephtys* sp. (Vallès et al., 2008). The latter is unexpected as animal mitochondrial genomes are small and very compact compared to those of plants, fungi and protists. Group II introns are rare in archaea and thought to be acquired from eubacteria by horizontal transfer (Dai and Zimmerly, 2003). In fungi, group II introns are found in

mitochondrial rRNA genes (Toro and Zimmerly, 2002; Hafez and Hausner, 2011a) and mitochondrial protein coding genes (Shnyreva, 1995). Unlike group I introns, group II introns have not been found in the nuclear genome of eukaryotes (Lambowitz and Zimmerly, 2011).

1.6.2. Structure:

Group II introns have a characteristic conserved RNA secondary structure consisting of six stem loop domains (DI – DVI) radiating from a central wheel (see Figure 1.5. [a] & [b]: Michel and Ferat, 1995; Qin and Pyle, 1998; Pyle and Lambowitz, 2006; Pyle, 2010; Lambowitz and Zimmerly, 2004 & 2011). The six domains interact to form a conserved tertiary structure that bring the distant 5' and 3' exon/intron junctions into a close proximity in the intron's active site along with the branch-point nucleotide residue and divalent ions to activate the appropriate bonds for catalysis (Lambowitz and Zimmerly, 2011). Domain I is the largest intronic domain and it is subdivided into several smaller stem-loop subdomains (Ia, Ib, Ic1, Ic2, … etc.). Domain I is transcribed first and thus folds firstly and acts as a scaffold for the folding of the subsequent domains; as well Domain I forms several tertiary interactions with the other domains to stabilize the intron structure and thus promote catalysis (Qin and Pyle, 1998; Fedorova and Zingler, 2007). Domain I plays an important role in recognizing both 5' and 3' exon sequences during intron splicing and mobility; DI contains the exon-binding sequences EBS1 and EBS2 which are involved in the recognition of 5' splicing site by base-pairing with the intron-binding sites IBS1 and IBS2 (complementary sequences to EBS1 and EBS2) in the 5' exon of both group IIA and IIB introns (Michel et al., 1989; Pyle, 2010). The 3' splice

site recognition in group IIA introns (Figure 1.5. [a]) is defined by base-pairing between the 1-3 nucleotides immediately upstream to the EBS1 sequence referred as delta (δ) and the first nucleotide in the 3' exon (δ^{-}); however, in group IIB introns (Figure 1.5. [b]) the 3' splice site is defined by the first nucleotide in the 3' exon (IBS3) and a nucleotide named EBS3 in an internal loop in DI (Costa et al., 2000).

Domain II contributes towards group II intron self-splicing by essential tertiary interactions: the θ – θ′ and η – ή contacts with DI and DVI respectively (Costa et al., 1997; Pyle, 2010). The θ – θ′ is a tertiary interaction between the basal stem of DII and the highly conserved terminal loop of DIC1 which stabilize the ribozyme core of the group II introns. The second tertiary interaction that involves DII is the η – ή contact which is a structurally conserved interaction among group IIA and IIB introns. In group IIA introns, the η – ή tertiary interaction is between a tetraloop located in DII and a receptor in DVI, which is in contrast to group IIB introns where the location of the tetraloop and the receptor are reversed (Costa et al., 1997). It has been proposed that the η – ή contact is an essential tertiary interaction during the conformational switch that occurs between the two steps of splicing (Chanfreau and Jacquire, 1996). It is possible for DII to serve as an insertion site for additional elements without interfering with the function of the ribozyme core because once DII is locked into position through the θ – θ′ and η – ή interactions the DII terminal loop projects away from the intron ribozyme core (Toor et al., 2010; Pyle, 2010).

Figure 1.5. Secondary structure models of group IIA **[a]** and IIB **[b]** introns. Intron sequences are in solid black lines, ORF sequences are in black dashed lines and exon sequences are in gray boxes. The positions of EBS1, EBS2 and EBS3/δ are noted. The positions of IBS1, IBS2 and IBS3/δ^{-} in the 5' and 3' exons are in blue, red and green boxes respectively. Tertiary interactions are indicated by dashed lines and Greek letters (α, β, γ, δ, ε, ζ, η, θ, κ, λ, and μ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V and VI). The solid black arrowheads indicate the intron-exon junctions (5' and 3' splicing sites). The asterisk shows the bulged adenosine nucleotide in domain VI (the branch point). Typically in group II introns ORFs are present within domain V but the RT ORF is encoded within DII in the Omi472-mS379 intron, while a LAGLIDADG ORF is encoded within DIII in the Omi371-mS952 intron (see Chapter 4; Hafez and Hausner, 2011a).

Figure 1.5. [a]

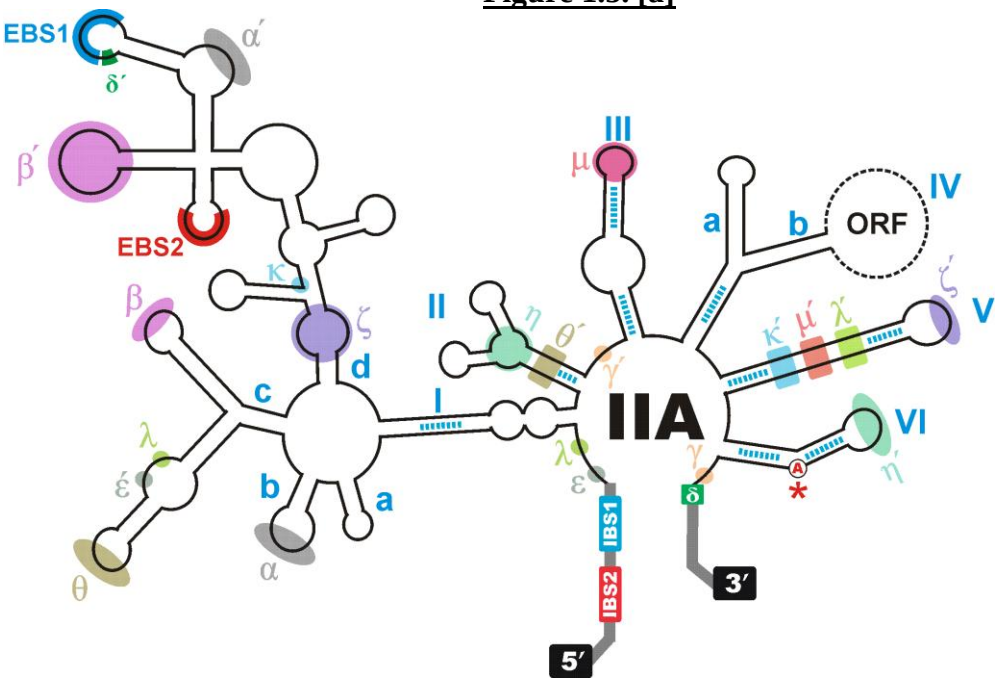
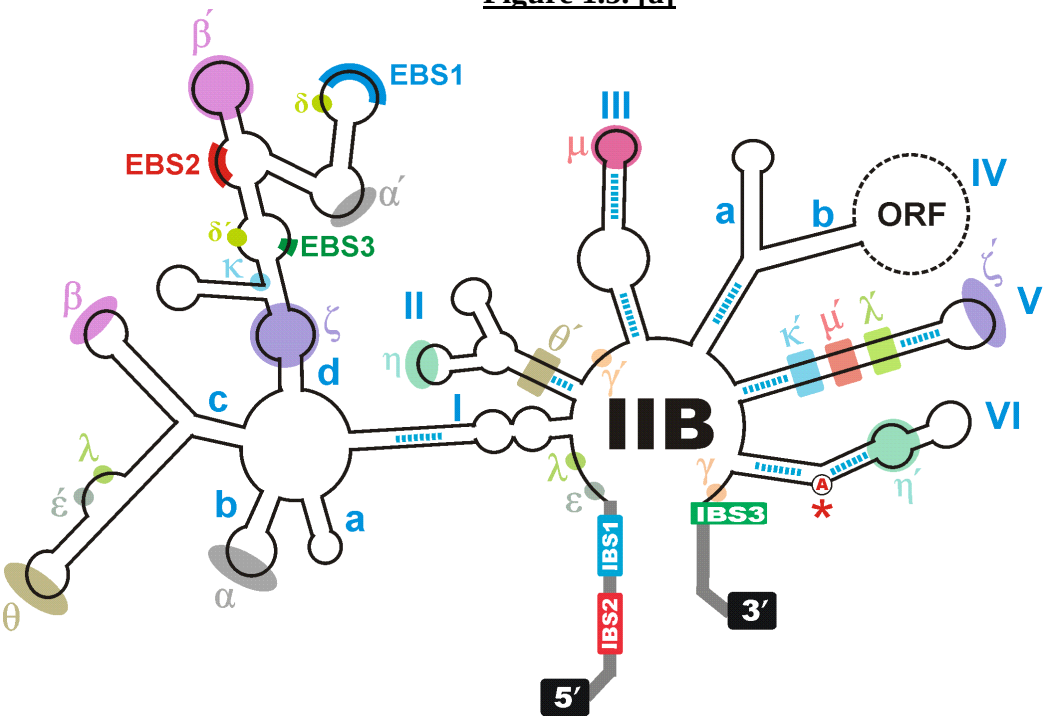


Figure 1.5. [a]



Domain III forms an important interaction (μ - μ′) with DV (Fedorova and Pyle, 2008), and also contains a highly phylogeneticly conserved internal loop that plays a major role as a catalytic effector that stimulates the catalytic activity of group II introns (Fedorova and Pyle, 2008). The basal stem of DIII can easily project away from the ribozyme core and thus this domain can carry additional insertions, as seen in some members of the group IIB1 family that encode LAGLIDADG ORFs from a terminal loop within DIII (Toor and Zimmerly, 2002; Mullineux et al., 2010; Hafez and Hausner, 2011a).

Domain IV is an ideal location for encoding ORFs because this domain is projected away from the intron core as a result of its stacking upon DIII. Domain IV also contains a high affinity binding site for the IEP at subdomain DIVa (Wank et al., 1999). Domain V is the most conserved domain in group II introns and it can form several important interactions with DI (ζ – ζ′, κ- κ′ and λ- λ′). Domain V is considered the most important component of a group II introns catalytic activity and along with DI; they are considered the minimal catalytic core requirements of group II introns (Lehmann et al., 2003; Lambowitz and Zimmerly, 2004; Pyle and Lambowitz, 2006; Toor et al., 2009). Domain VI contains the branch-point nucleotide residue (usually bulged A with a 2'-OH group), which is involved in the intron self-splicing reaction by promoting a transesterification reaction that eventually leads to the introns excision in a branched (lariat) form, stabilized with a 2'-5' phosphodiester bond (van der Veen et al., 1986; Vogel and Borner, 2002).

1.6.3. Splicing:

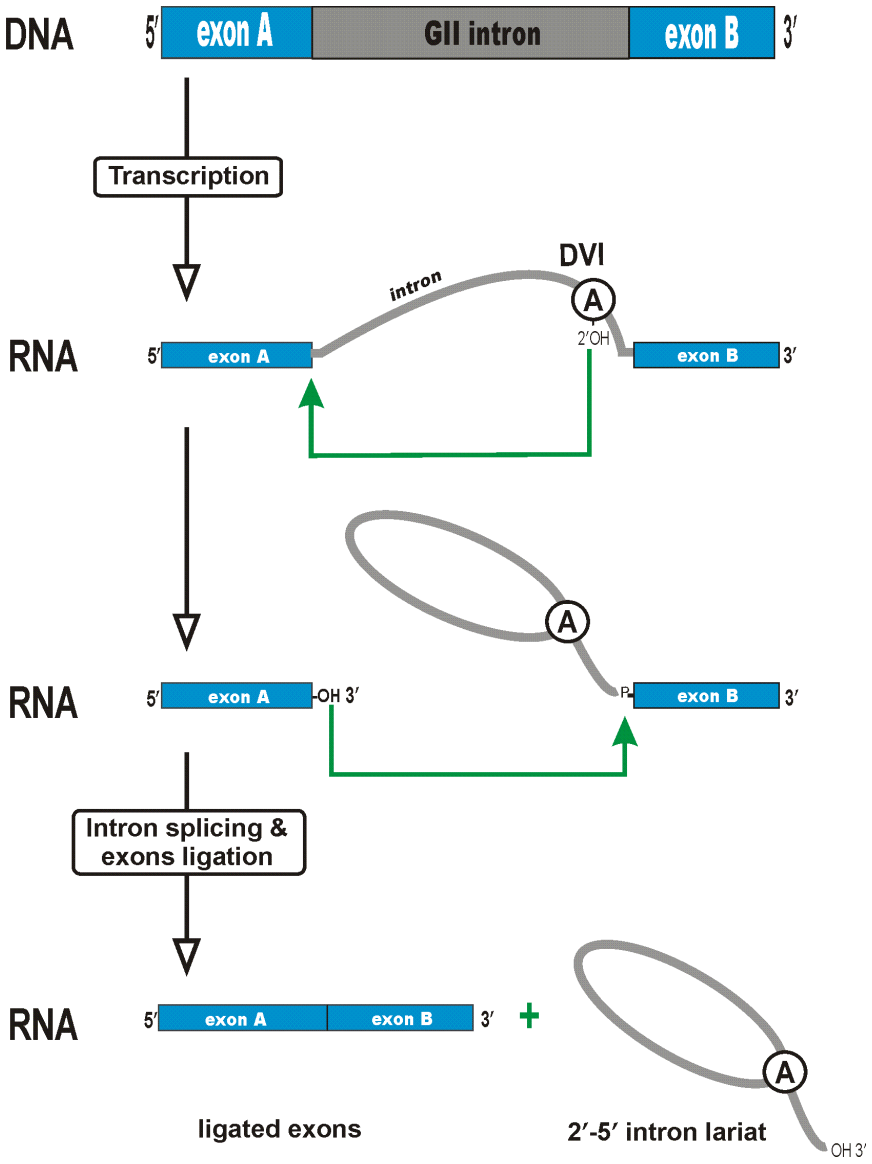
Intron-catalyzed splicing proceeds by two transestertification reactions and leads to the excision of the intron as a branched, or lariat, molecule with a characteristic 2'-5' phosphodiester bond; this is referred to as the branched pathway. There is another pathway by which group II introns are released from precursor transcripts; this pathway involves a hydrolysis step and here the introns are released in linear forms (Michel et al., 1989). In the so-called branched pathway (Figure 1.6.) during the first reaction, the 2'-OH group of a bulged adenosine residue located in DVI attacks the 5' splice site forming a lariat intermediate with a 2'-5' phosphodiester bond. In the second transestrification reaction, the 3'-OH group of the 5' exon attacks the 3' splice site leads to exon ligation and the release of the intron as a lariat molecule (Michel and Ferat, 1995; Daniels et al., 1996). The splicing process may also be initiated by 5' splice site hydrolysis and in this case the intron released is in a linear form (Michel et al., 1989; Fedorova and Zingler, 2007).

1.6.4. Mobility:

Both splicing and mobility of group II introns require the catalytic activity of the intron RNA, the intron-encoded protein, and possibly host factors. During expression of the host gene, the intronic ORF is translated and a ribonucleoprotein (RNP) particle is formed between the intron lariat and the intron-encoded protein (Figure 1.7.). The RNP recognizes a target homing site, typically an intron-less cognate allele, and the first cut is made by the 3' end (i.e a 3'-OH) of the intron RNA. This initiates a reverse splicing reaction whereby the intron RNA is inserted into the sense strand of the DNA.

Figure 1.6. The branching pathway of group II introns splicing. A bulged adenosine nucleotide located at the intron's DVI (circled A) attacks the 5' splice site and then the 3'-OH group of the released 5' exon attacks the 3' splice site. As a result the exons are ligated and the intron is spliced out as a branched (lariat) molecule with a 2'-5' phosphodiester bond stabilizing the lariat configuration. The figure was generated based on the information obtained from Saldanha et al. (1993), Michel and Ferat (1995), Costa et al. (2000) and Lambowitz and Zimmerly (2001 & 2011).

Figure 1.6.


DNA
5′
exon A
GII intron
exon B
3′
Transcription
RNA
intron
DVI
A
2′OH
5′
exon A
exon B
3′
A
RNA
5′
exon A
OH 3′
P
exon B
3′
Intron splicing &
exons ligation
RNA
5′
exon A
exon B
3′
+
A
ligated exons
2′-5′ intron lariat
OH 3′

The En domain cleaves the antisense DNA strand, generating a free 3'-OH that serves as a primer for the reverse transcription. Eventually, the host DNA repair machinery will remove the RNA and fill in any gaps. This process of "retrohoming" is mediated by a process termed target DNA-primed reverse transcription (reviewed in Lambowitz and Zimmerly, 2004 & 2011). Group II intron RNAs have been shown to retrotranspose by reverse splicing into RNA molecules (Figure 1.4. [b]), a mechanism that requires less specificity at the target site and thus allows for retro-transposition of introns to new sites within the genome (Cousineau et al., 1998; Eickbush, 1999). Group II introns can also reverse splice into single-stranded DNA (at the replication fork; Figure 1.4. [c]), facilitation their dispersal into new sites and transmission to the next generation (Zhong and Lambowitz, 2003).

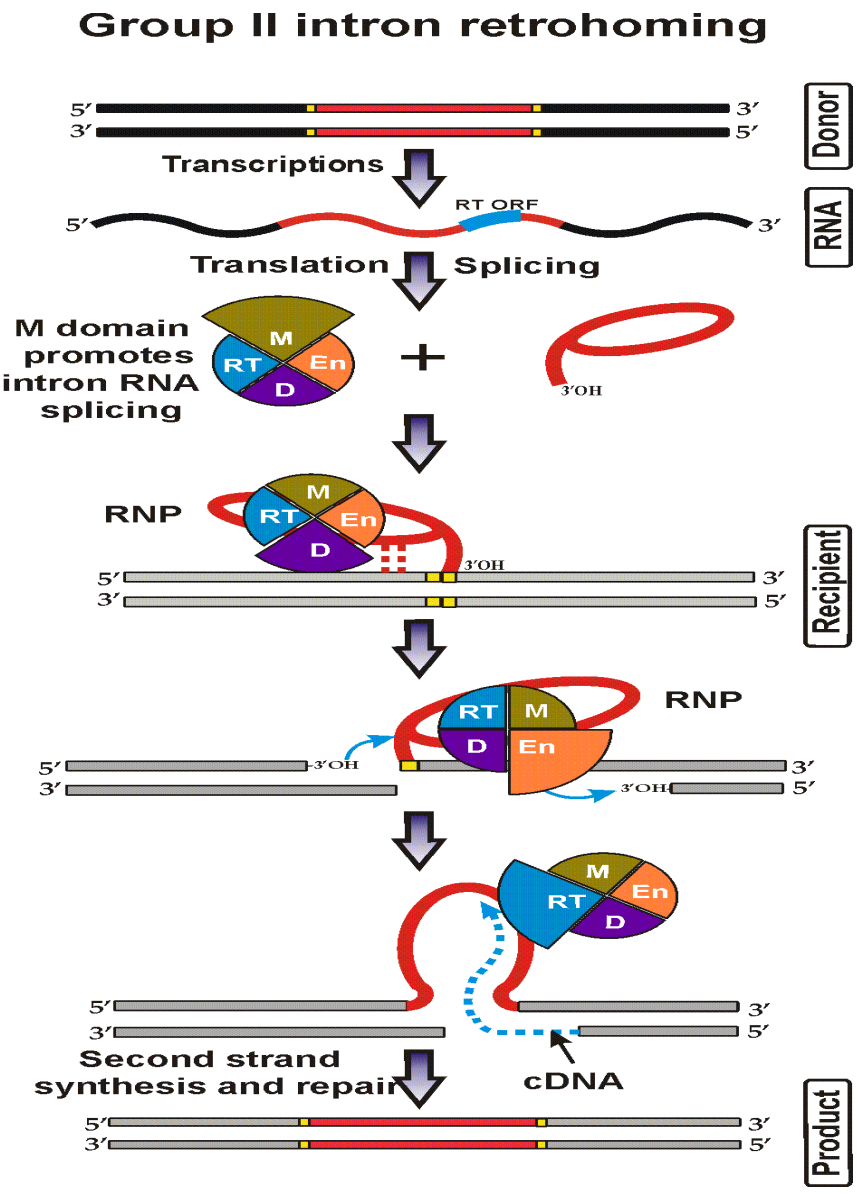
1.6.5. Evolution:

Based on structural (sequences of the key regions) and functional (identical splicing pathway) similarities between Group II introns and spliceosomal introns have made group II introns candidates for the ancestors of spliceosomal introns (Madhani and Guthrie 1992; Shukla and Padgett 2002; Keating et al., 2010). Group II introns are also thought to be the ancestor of non-**L**ong **T**erminal **R**epeats retrotransposons (non-LTR-retrotransposons) due to the similarities of their RT sequences (Xiong and Eickbush, 1990; Blocker et al., 2005) and the Target Primed Reverse Transcription (TPRT) mechanism for reverse transcription of the intron RNA template (Luan et al., 1993; Zimmerly et al., 1995a). There is also a potential connection between group II intron RTs and telomerases (Nakamura and Cech, 1998).

Figure 1.7. Group II intron retrohoming pathway: The black strands represent sequence of the donor allele, gray strands represent sequence of the recipient allele, red lines symbolize the intron sequence, straight and wavy lines representing DNA and RNA respectively and the green pac-man symbol represent the intron-encoded endonuclease which recognize and cleaves the target site (yellow boxes). The intron-encoded protein is a multi-domain protein represented by a circle with four quarters; the dashed blue line represents the cDNA. The pathway shown is for a bacterial group II intron and there are some variations among the different group II intron types.

(dsDNA = double-stranded DNA; RNP = ribonucleoprotein; RT = reverse transcriptase activity; M = maturase activity; En = endonuclease activity; D, DNA binding domain).

Figure 1.7.


Group II intron retrohoming
5′
3′
3′
5′
Donor
Transcriptions
RT ORF
5′
3′
RNA
Translation
Splicing
M domain promotes intron RNA splicing
M
RT
En
D
+
3′OH
RNP
M
RT
En
D
3′OH
5′
3′
3′
5′
Recipient
RNP
RT
M
D
En
5′
3′OH
3′
3′
3′OH
5′
M
RT
En
D
5′
3′
3′
5′
Second strand synthesis and repair
cDNA
5′
3′
3′
5′
Product

Group II introns are common in eubacteria and eukaryotic organelles but are rare in archaea. This phylogenetic distribution suggests that they evolved in bacteria and were transmitted to eukaryotic cells via endosymbiotic bacteria that gave rise to chloroplasts and mitochondria (Cavalier-Smith, 1991; Palmer and Logsdon, 1991; Ferat and Michel, 1993; Zimmerly et al., 2001). Toor et al. (2001) suggested the retroelement ancestor hypothesis as a model for the evolution of group II introns in which group II introns originated as retroelements in bacteria and the intron RNA was developed through coevolution with the IEP rather than as an independent catalytic RNAs that eventually was invaded by an ORF (i.e. the IEP was associated with the intron RNA prior to the divergence of different group II intron lineages. The retroelement ancestor hypothesis is supported by the presence of several ORF-less group II introns that contain ORF remnants (Toor et al., 2001; Dai and Zimmerly, 2003) suggesting ORF-less group II introns have been derived from those that had ORFs. This is in contrast to group I introns where ORFs can move independently from their ribozyme partners (reviewed in Hausner, 2012).

Group III introns are self-splicing RNA elements that appear to be degenerated group II introns; i.e. derived from group II introns, as they only contains DI and DVI with the conventional bulged adenosine residue, while DII to DIV are often absent (Montandon and Stutz, 1983; Christopher and Hallick, 1989). Group III introns are found in the chloroplast genome of some euglenoid protists and they are usually associated with genes involved in transcription and translation. They are typically much shorter than group I and group II introns, ranging from 95 to 110 nucleotides (Hallick et

al., 1993). The splicing of group III introns is by two transestertification reactions where the first nucleophilic attack to the 5' splice site is initiated by the bulged adenosine nucleotide in DVI and the intron is released in a lariat form (Copertino et al., 1994). Group III introns have been reported in the *psbC* gene of *Euglena gracilis* as nested introns (twintrons) that self-splice sequentially (Copertino et al., 1994).

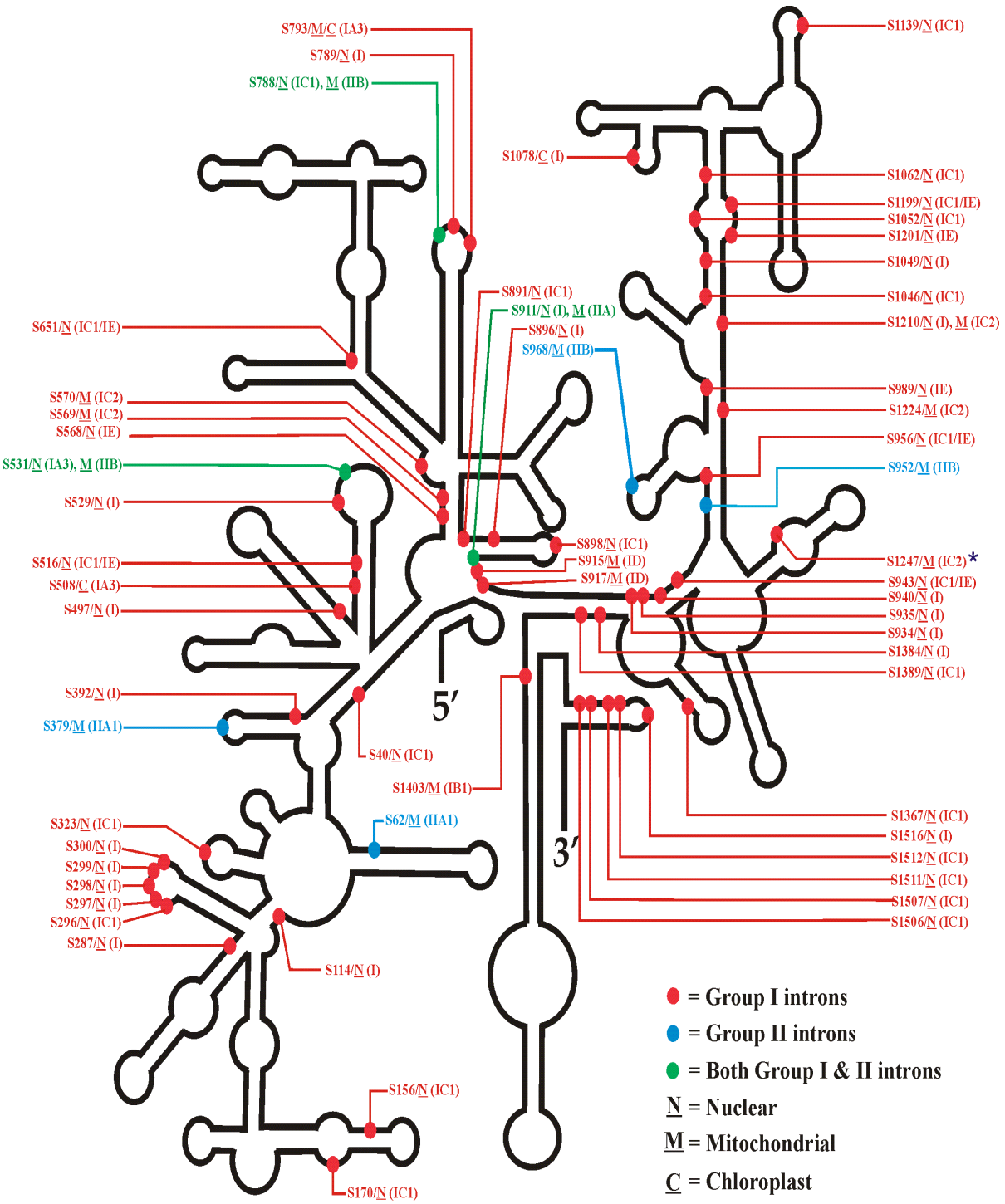
1.7. Distribution of group I and II introns in rRNA genes:

The distribution of introns in rRNA genes (SSU rRNA and LSU rRNA) is not totally random; insertion sites appear to be found within conserved sequences usually near the tRNA and mRNA binding sites which span the interface between the small and the large subunits of the ribosome, suggesting a link between intron evolution and rRNA function (Noller et al., 1981; Gerbi et al., 1982; Turmel et al., 1993; Johansen et al., 1996; Jackson et al., 2002). The relationship between intron splicing efficiency and the intron insertion position within the pre-rRNA has been studied and the results showed that, the splicing efficiency depends on the interaction between the intron and the surrounding (exon) rRNA sequences and this result may explain the preference of introns to be inserted in certain regions of rRNA (Rocheleau and Woodson, 1995). It should be noted that the small subunit RNA gene is referred to as the SSU rRNA gene or 16S / 18S RNA gene when referring to the nuclear version of this gene whereas recently *rns* has been used to refer to the mitochondrial version of the small subunit RNA gene. Based on the available information at the comparative RNA website (http://www.rna.ccbb.utexas.edu/) and the literature review, the SSU rRNA gene (including the mitochondrial and chloroplast counterparts) was found to be invaded at 59 different positions by group I and

group II introns (see Figure 1.8.), and during this thesis we characterized a novel twintron at position S1247 in the *rns* gene of *Chaetomium thermophilum* (accession number JN007486), in which a group IIA1 intron was found to be inserted within the ORF of a group IC2 intron. See Chapter 6 for a detailed analysis of the group I and II intron insertions within the mtDNA *rns* gene of the fungi.

Figure 1.8. Distribution of Group I and group II introns in the SSU rRNA sequence. The secondary structure model of SSU rRNA is represented by black thick lines. Group I (red circles) and group II (blue circles) intron insertion positions are numbered according to the *E. coli* SSU rRNA sequence. Some positions can be occupied by both group I and group II introns (green circles). The Cellular/organellar location of each intron is underlined (N = Nuclear; M = Mitochondrial; C = Chloroplast) and intron classes are given in brackets. The position S1247 was found to be invaded by group IC2 introns in some species and in some others it was found to be invaded by a twintron in which group IIA1 introns inserted in the ORF of a group IC2 intron. This figure was generated based on the information obtained from the comparative RNA website and Jackson et al., (2002) as well as results generated from this thesis.

Figure 1.8.


S793/M/C (IA3)
S789/N (I)
S788/N (IC1), M (IIB)
S1139/N (IC1)
S1078/C (I)
S1062/N (IC1)
S1199/N (IC1/IE)
S1052/N (IC1)
S1201/N (IE)
S1049/N (I)
S891/N (IC1)
S911/N (I), M (IIA)
S896/N (I)
S968/M (IIB)
S1046/N (IC1)
S1210/N (I), M (IC2)
S651/N (IC1/IE)
S570/M (IC2)
S569/M (IC2)
S568/N (IE)
S989/N (IE)
S1224/M (IC2)
S956/N (IC1/IE)
S531/N (IA3), M (IIB)
S952/M (IIB)
S529/N (I)
S898/N (IC1)
S915/M (ID)
S917/M (ID)
S1247/M (IC2)*
S516/N (IC1/IE)
S508/C (IA3)
S497/N (I)
S943/N (IC1/IE)
S940/N (I)
S935/N (I)
S934/N (I)
S1384/N (I)
S1389/N (IC1)
5′
S392/N (I)
S379/M (IIA1)
S40/N (IC1)
S1403/M (IB1)
S62/M (IIA1)
S1367/N (IC1)
S1516/N (I)
S1512/N (IC1)
S1511/N (IC1)
S1507/N (IC1)
S1506/N (IC1)
3′
S323/N (IC1)
S300/N (I)
S299/N (I)
S298/N (I)
S297/N (I)
S296/N (IC1)
S287/N (I)
S114/N (I)
S156/N (IC1)
S170/N (IC1)
= Group I introns
= Group II introns
= Both Group I & II introns
N = Nuclear
M = Mitochondrial
C = Chloroplast

1.8. Homing endonucleases:

Homing endonucleases are rare cutting enzymes encoded by homing endonuclease genes (HEGs) that are embedded within self-splicing group I, group II and archaeal introns or inteins which self-splice at protein level (protein introns) and HEGs can also be free standing ORFs between individual genes. Homing endonucleases are double-stranded DNases that target large recognition sites (14-40 bp; Stoddard, 2006). The HEGs, due to the nature of how they move, are located in the middle of their own recognition sequences and therefore alleles bearing the HEGs (HEG^+) are protected from being cleaved, and only HEG^- alleles are suitable substrate for HEase cleavage. The presence of a cleaved allele induces the homologous recombination based DSB repair process whereby the cleaved HEG^- allele is repaired by using the HEG^+ allele as the template to repair the DSB. The recombination process leads to copying the HEG across the HEG^- allele and as a result the HEG^- allele is converted into a HEG^+ allele. This process is called homing (Dujon, 1989; Lambowitz and Belfort, 1993; Belfort and Perlman, 1995; Chevalier and Stoddard, 2001).

The recognition sites of HEases are extremely rare, with none or only one of these sites present in a typical mammalian sized genome, and the recognition site is long enough to occur randomly only with a very low probability, approximately once every 7×10^{10} bp (Jasin, 1996). Homing endonucleases must display sufficient target site specificity to avoid the cleavage of essential host genes; at the same time these proteins can tolerate some sequence polymorphism in the wild-type target site in order to promote lateral transfer between closely related host species. The long recognition site of HEases

prevent nonspecific binding and cleavage within their host genome (Belfort et al., 1995; Belfort and Roberts, 1997). In addition low expression levels (i.e. tight gene regulation) for the HEase and the localization of these elements in specific compartments within the cell (mitochondria, chloroplast and nucleus) keep them cutting too frequently. By being sequestered into organelles, HEases are prevented from potentially damages the host's nuclear DNA (Jurica and Stoddard, 1999).

1.8.1. Nomenclature:

The nomenclature rules for naming HEases resembles those for the restriction enzymes. A three-letter designation contains the first letter of the host genus name followed by the first two letters of the species name followed by a Roman numeral that indicates the order in which the enzyme has been characterized to distinguish multiple enzymes from a single organism. To indicate if the endonuclease is intron encoded or intein encoded, the prefixes I- or PI- are added to the enzyme name, respectively. Another prefix, F-, was proposed for freestanding HEases. Thus, I-SceI and PI-SceI represent the first discovered intron-encoded and intein-encoded endonucleases from *S. cerevisiae* respectively (Belfort and Roberts, 1997; Roberts et al., 2003). Hybrid HEases are preceded by the prefix H-, followed by the authors' designation, for example the first synthetic HEase H-DreI was synthesized from the intron-encoded I-DmoI and I-CreI endonucleases from *Desulfurococcus mobilis* and *Chlamydomonas reinhardtii* respectively (Chevalier et al., 2002). Homing endonucleases that have been characterized biochemically are listed within REBASE (http://www.neb.com/rebase) along with restriction enzymes and methyltransferases. REBASE provides information about which

enzymes are available and which DNA sequence they recognize (Roberts and Macelis, 1997).

1.8.2. Families of homing endonucleases:

Homing endonucleases can be classified into four major distinct families; their naming is based on conserved amino acid motifs: the LAGLIDADG, H-N-H, His-Cys Box, and GIY-YIG families of HEases (Belfort et al., 2002; Kowalski and Derbyshire, 2002; Stoddard, 2006). Recently additional HEase-like proteins/families have been discovered (Table 1.1; Figure 1.9.), the PD-(D/E)XK HEases are found in cyanobacterial tRNA group I introns (Zhao et al., 2007), the Vsr (very-short patch repair) endonucleases (a predicted family of HEases) were found in phages based on environmental metagenomic data (Dassa et al., 2009), and the Holliday junction resolvase-like HEases have been identified in phages (Zeng et al., 2009). An additional HEase family has recently been designated as the EDxHD family based on detailed characterization of a group I intron ORF found within the recA gene of the *Bacillus thuringiensis* 0305φ8-36 bacteriophage (Taylor et al., 2011); this family appears to be related to the Vsr endonucleases (Dassa et al., 2009; Taylor et al., 2011). These families differ greatly structurally (in their conserved active site motif) and functionally (catalytic mechanisms) as well as their biological and genomic ranges. Homing endonucleases are widespread and are found in a wide range of organisms from all biological kingdoms, including bacteriophages, archaebacteria, eubacteria, fungi, algae and plants (see Figure 1.9; Belfort and Roberts, 1997; Jurica and Stoddard, 1999).

Table 1.1. Examples for the major homing endonuclease families.

Table 1.1.

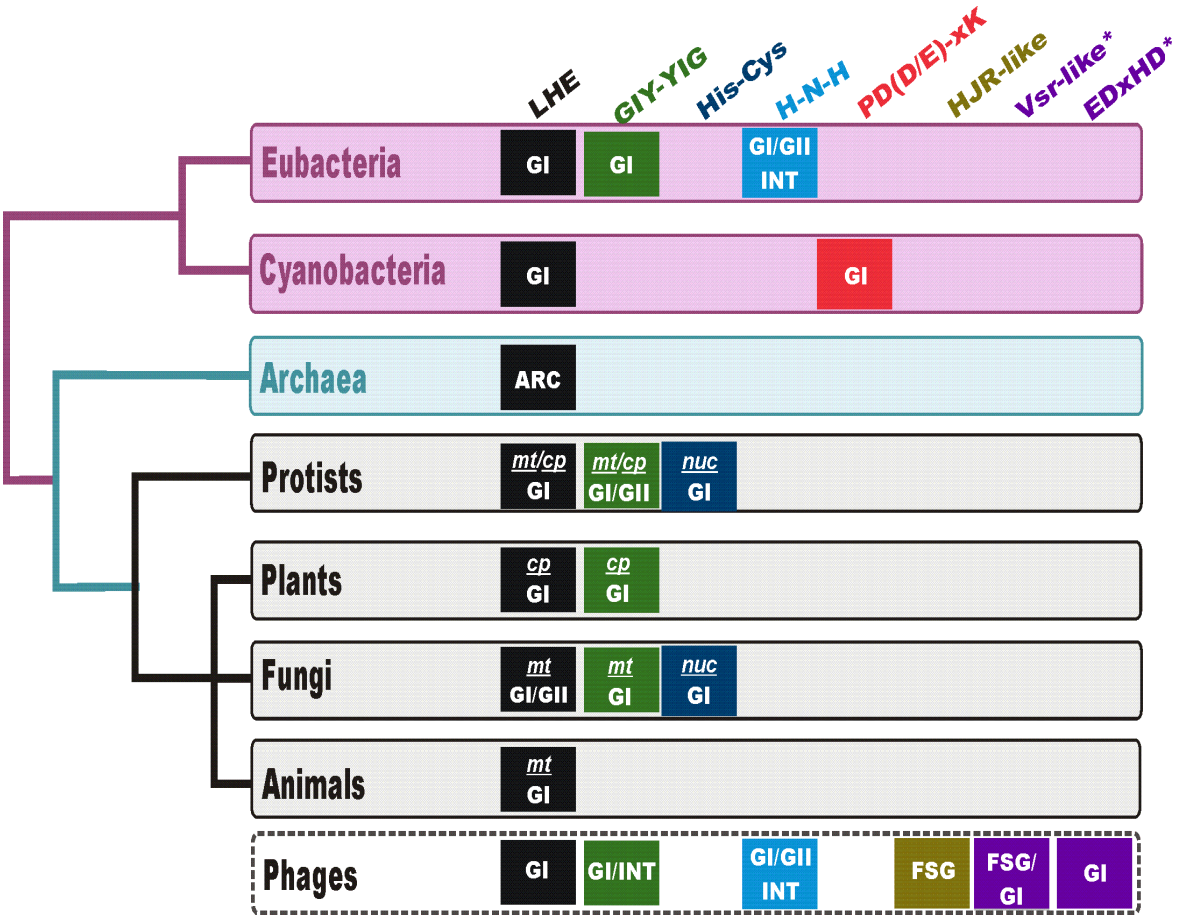
Family	Biological domain	Subcellular location	Genomic niche	Cleavage pattern	Example(s)
LAGLIDADG	Archaea Eukarya	Mitochondria Chloroplasts	Group I introns Group II introns Archaeal introns Intein	Ds* 4 nts 3' overhang	I-CreI[1] I-AniI[2] I-DmoI[3] PI-PfuI[4] I-SceI[5]
His-Cys box	Protists	Nuclear (rRNA genes)	Nuclear rDNA group I introns	Ds 4 nts 3' overhang	I-PpoI[6]
GIY-YIG	Bacteria Eukarya Phage	Mitochondria chloroplast	Free standing or Group I introns	Ds 2 nts 3' overhang	F-TevI[7] I-TevI[8]
H-N-H	Bacteria Eukarya Phage	mitochondria Phage	Free standing or Group I introns	Nicking	I-HmuI[9] I-SceV[10]
PD(D/E)-xK	Cyanobacteria	Bacteria	Group I introns	Ds 2 nts 5' overhang	I-Ssp6803I[11]
HJ Resolvase-like	Phage	Phage	Free standing	Ds 4 nts 3' overhangs	F-CphI[12]
Vsr-like	Metagenomic (suspected to be of phage origin)	Phage	Free standing Group I introns or inteins	Unknown	Putative[13]
EDxHD	Phage	Phage	Group I introns	Ds 2 nts 5' overhang	I-Bth0305I[14]

* Ds = double-stranded cut;

[1]Jurica et al., 1998; [2]Polduc et al., 2003; [3]Dalgaard et al., 1994; [4]Ichiyanagi et al., 2000; [5]Moure et al., 2003; [6]Flick et al., 1998; [7]Sharma et al., 1992; [8]Bell-Pedersen et al., 1991; [9]Shen et al., 2004; [10]Liang et al., 1996; [11]Zhao et al., 2007; [12]Zeng et al., 2009; [13]Dassa et al., 2009; [14]Taylor et al., 2011.

Figure 1.9. Distribution of the HEase families among the different biological domains. A phylogenetic tree (modified from the rRNA universal tree of life; Woese et al., 1990) shows the biological host range of each HEase family. The genomic range of each HEase family is also indicated inside colored boxes (n = nuclear; c = chloroplast; m = mitochondrial; FSG = Free Standing Gene; ARC = archaeal; GI = group I introns; GII = group II introns; INT = Inteins). With regards to the H-N-H family, proteins with H-N-H domains such a group II intron encoded RTs or DNA binding proteins (see text) where excluded. With regards to the metazoans self-splicing introns are rare and confined to early branching metazoans (reviewed in Hausner, 2012). * The Vsr-like (putative) and EDxHD families may represent a single family of HEases.

Figure 1.9.


LHE
GIY-YIG
His-Cys
H-N-H
PD(D/E)-xK
HJR-like
Vsr-like*
EDxHD*
Eubacteria
GI
GI
GI/GII INT
Cyanobacteria
GI
GI
Archaea
ARC
Protists
mt/cp GI
mt/cp GI/GII
nuc GI
Plants
cp GI
cp GI
Fungi
mt GI/GII
mt GI
nuc GI
Animals
mt GI
Phages
GI
GI/INT
GI/GII INT
FSG
FSG/ GI
GI

1.8.2.1. LAGLIDADG HEases (LHEases):

The LAGLIDADG motif proteins represent the first identified and biochemically characterized HEG-encoded endonuclease family (Dujon, 1980; Jacquier and Dujon, 1985). This large protein family with more than 200 members has been variously termed 'DOD', 'dodecapeptide', 'decapeptide' and 'dodecamer' based on the 10-residue conserved motif (Dujon, 1989; Belfort and Roberts, 1997; Chevalier and Stoddard, 2001). Members of this family have either one or two copies of the conserved LAGLIDADG motif. LHEases that contain a single copy of this motif, such as I-CreI, act as homodimers and recognize palindromic or nearly palindromic DNA target sites (Chevalier et al., 2001). However, enzymes that have two copies of this motif in a single polypeptide chain, such as I-AniI, are monomers and recognize asymmetric DNA target sites (Boldouc et al., 2003). It was previously thought that double motif LHEases evolved via a duplication event of a single motif LHEases. This was supported by a recent study by Haugen and Bhattacharya (2004) which showed that a single motif LHEases that inserted into a preexisting group I intron in the *rnl* gene between positions L1917 and L1951 evolved into double motif LHEases and then in combination with its group I intron host spread into sites first within the *rnl* gene and subsequently invaded new sites in the *rns* genes as well as protein coding genes. It has also been shown that double motif LHEases, in combination with a group I intron that invaded protein coding genes transferred back to introns in the rRNA genes. One lineage of LHEases moved independently from its original ribozyme partner and transferred into an ORF-less preexisting group II intron to form a new family of group II introns encoding LHEases instead of the RT-like ORFs that are usually encoded by group II introns (Toor and

Zimmerly, 2002; Mullineux et al., 2010; Hafez and Hausner, 2011a). Overall this family of HEases has a very dynamic evolutionary history, which is reflected by its wide host distribution (Figure 1.9) and the many genomic niches invaded by these elements (Gimble, 2000).

The structures of several LHEases bound to their DNA targets have been elucidated (for more details: see Stoddard, 2006) and these models provide a clear picture about substrate recognition, binding and cleavage of these enzymes on their DNA targets. The structural studies indicates that LHEases consists of a core fold of mixed α/β topology (αββαββα) and the overall shape of this domain is half-cylindrical, in which the underside forms a groove due to the presence of four antiparallel β-sheets. The latter present a large number of exposed basic and polar residues allowing for DNA contact and binding.

LHEases use a flexible strategy to recognize their target sites and tolerate a number of individual polymorphisms at the target site without significant loss of the enzyme's binding affinity. Structural studies of several LHEases bound to their DNA targets showed that LHEases recognize the target site by four antiparellel β-strands in each domain providing direct and water-mediated contacts between the enzyme and the major groove of each DNA half-site (Silva et al., 1999; Ichiyanagi et al., 2000; Chevalier et al., 2002; Bolduc et al., 2003; Chevalier et al., 2003; Moure et al., 2003; Takeuchi et al., 2011). The LHEases usually make contacts to approximately 70% of all the possible hydrogen-bond donors and acceptors of the base pairs in the major groove and about 30%

of the backbone phosphate groups across the enzyme's target site (Jurica et al., 1998; Chevalier et al., 2002; Chevalier et al., 2003).

The cleavage mechanism of I-CreI (Chevalier et al., 2003) has been well studied and many of the kinetics and catalysis mechanisms of this enzyme appear to apply to other LHEases. The I-CreI is highly dependent on divalent cations for activity in which three bound divalent metal ions are coordinated by a pair of overlapping active sites with one metal ion being shared in both cleavage reactions. In contrast, the structure of the monomeric (i.e. double motif LAGLIDADG motif) HEase enzyme I-AniI when bound to its DNA target showed that only two bound metal ions were present (Moure et al., 2002). Generally, there are two different cleavage mechanisms proposed for LHEases: the sequential and concerted cleavage mechanisms: in the former, the HEase makes an incision in one strand followed by nicking in the second strand (Moure et al., 2002), while in the latter, both strands are simultaneously cleaved (Saves et al., 2002).

1.8.2.2. GIY-YIG HEases:

The GIY-YIG proteins represent the second most numerous HEase family, and they were identified for the fist time in group I introns of filamentous fungi and within the T4 bacteriophages. Members of this endonuclease family are characterized by the presence of the conserved GIY-(X10-11)-YIG amino acid motif (Michel and Dujon, 1986; Cummings et al., 1989; Kowalski et al., 1999). GIY-YIG endonucleases have been described from group I introns within mitochondrial genomes of yeasts (Tian et al., 1991) and filamentous fungi (Saguez et al., 2000), algal mitochondrial genomes (Kroymann and Zetsche, 1997) and in algal chloroplast genomes (Paquin et al., 1995) as well as bacteriophage genomes (I-TevI, I-TevII; Bell-Pedersen et al., 1990). However members of this HEase family have also been found in intergenic regions (F-TevI and F-TevII; Sharma et al., 1992) and recently one example has been noted to be encoded within a group II intron of the protozoan *Amoebidium parasiticum* (Li et al., 2011). These proteins promote the mobility of group I introns in a manner similar to that described for LAGLIDADG type HEases.

Biochemical experiments on the I-TevI HEase (the most extensively studied GIY-YIG HEase) indicate that these endonucleases acts as monomers recognizing long homing sites (37 bp). The cleavage site is found to be many nucleotides away from the intron insertion site and this endonuclease generates 2 nucleotide 3’ overhangs (Derbyshire et al., 1997). I-TevI is used as a model for GIY-YIG HEases; proteolysis and footprinting experiments have shown that I-TevI is a bipartite enzyme consisting of an N-terminal endonuclease domain and C-terminal DNA-binding domain separated by a long

flexible linker segment that functions as a molecular ruler to position the catalytic endonuclease domain at the correct distance onto the DNA substrate away from the DNA-binding domain (Derbyshire et al., 1997; Dean et al., 2002). The structure of I-TevI HEase was determined by X-ray crystallography; however the C-terminal DNA-binding domain and the N-terminal catalytic domain were analyzed separately (Van Roey et al., 2001 & 2002). The GIY-YIG nuclease motif is not unique to HEases, as the domain is associated with a variety of DNA binding enzymes/complexes that require endonuclease activity including DNA repair proteins, proteins encoded by retrotransposable elements (Pyatkov et al., 2004), and restriction endonucleases (Kaminska et al., 2008; Mak et al., 2010).

1.8.2.3. H-N-H HEases:

The H-N-H proteins or H-N-H domains have been found to be encoded within freestanding ORFs and ORFs associated with group I and group II introns and H-N-H motifs are found in the endonuclease domainS of some inteins (Dalgaard et al., 1997; Chevalier and Stoddard 2001, Chevalier et al., 2005). Typically mobile group II introns encode complex proteins with several different functional domains associated with different activities; the so-called En domain is required for cleaving one strand at a potential intron insertion site (Lambowitz and Zimmerly, 2004 & 2011). The En domain contains sequence motifs characteristic of the H-N-H endonuclease family. Biochemical and structural analysis suggests that the H-N-H proteins belong to a group of enzymes whose active sites involve a conserved ββα–Me structural motif (reviewed in Keeble et al., 2005) a feature shared with His-Cys Box endonucleases. The H-N-H endonucleases

appear to be related to plasmid-encoded endonucleases referred as colicin DNases found in *E. coli* (Chevalier and Stoddard, 2001; Zimmerly et al., 2001).

1.8.2.4. His-Cys HEases:

His-Cys box HEases have been noted to be encoded by nuclear rDNA group I introns of several fungi and protists (Haugen et al., 2004; Galburt and Jurica, 2005). The best known examples are the I-PpoI from the myxomycete *Physarum polycephalum* (Muscarella et al., 1990) and I-DirI from *Didymium iridis* (Johansen et al., 1997). Recent studies show that the His-Cys box and H-N-H endonucleases share similar DNA-binding motifs and active sites suggesting they share a common ancestor and thus should be assigned into a single "superfamily" of HEases (Galburt and Jurica, 2005).

1.9. Homing endonucleases and restriction endonucleases:

Both HEases and restriction endonucleases (REases) are sequence-specific deoxyribonucleases generating site-specific double-strand breaks (DSB) at their DNA target sites. Although HEases and REases are functionally similar, they are phylogenetically and structurally differ. Members of REase families have been found only as free standing genes in bacteria, archaea and certain phages and act as a defense mechanism against invading viruses/phages (Krüger and Bickle, 1983; Wilson, 1988). On the other hand, HEases are widely distributed and are found in all biological domains (bacteria, archaea and eukarya) and they are frequently embedded within mobile elements such as the self-splicing group I, group II and archaeal introns or inteins and they can also be found as a free standing open reading frames (ORFs) between individual genes. HEGs

offer their host introns the ability to be mobile, whereas introns provide their accompanying HEG a position within the genome that does not impart toxicity or a phenotype to the host genome. HEases can also be expressed in different cell compartments within the eukaryotic cell such as nuclei (as inteins or as components of rDNA introns), or the mitochondria and chloroplasts (Lambowitz and Belfort, 1993; Muller et al., 1993).

Based on their recognition sequence, subunit composition, cleavage position, and cofactor requirements, REases, can be divided into three major groups: Type I, II and III enzymes (Williams, 2003) whereas HEases based on conserved amino acid sequence motifs can be classified into eight families: LAGLIDADG, GIY-YIG, H-N-H, His-Cys box, PD-(D/E)XK, Vsr-like, Holliday junction resolvase-like and EDxHD (Stoddard, 2006 & 2011). The major difference between HEases and REases is the length of DNA target site; HEases recognize very long asymmetric target site ranging from 14 to 44 bp (Stoddard, 2006), whereas REases recognize relatively short target sites between 4 to 8 bp and many of them are palindromic in nature (Gowher and Jeltsch, 2001). Although HEases and REases do share a common function (cleaving double-stranded DNA), they appear to have evolved independently (Belfort and Roberts, 1997); HEGs represent selfish elements that colonize genomes and the REases represent a DNA defense strategy against foreign DNAs. However, it should be stated that some intron-encoded proteins have been co-opted to perform essential functions within host genomes such as acting as intron splicing factors, recombinases, and nuclease domains in DNA repair enzyme (reviewed in Hausner, 2012).

1.10. Applications of Homing endonucleases:

Homing endonucleases recognize large DNA binding sites (14 to 44 bps; Silva et al., 2011) which are extremely rare within a genome, thus typically only one (or none) such HEase DNA target sites exist within a mammalian-sized genome. However, unlike restriction endonucleases, these enzymes can tolerate some sequence degeneracy within their recognition site. These two unique characters (high specificity and tolerance of target site degeneracy) facilitate using HEases as a potential tools in different *in vitro*, *in vivo* or *ex vivo* applications, such genomic engineering in a variety of different genetic systems allowing for gene therapy, genome modifications, and pest control.

The use of HEases in therapeutic or genome editing applications (Figure 1.10. [a]), is mainly based on their capacity to generate a site-specific DSB, which triggers the cell to repair the DNA damage by pathways either involving homologous recombination (HR) or nonhomologous end joining (NHEJ) (Johnson and Jasin, 2001; Mladenov and Iliakis, 2011; Symington and Gautier, 2011). The DSB HR is a highly conserved mechanism that is used by many organisms to repair DNA breaks, so this DNA repair machinery is being exploited to incorporate DNA sequence modifications such as the repair of a defective gene by cutting the defective allele in the presence of a repair template that provides the corrected allele. In the past decade two different approaches have been pursued to induce highly site-specific DNA DSB: zinc finger nucleases (ZFNs; Wright et al., 2006; Klug, 2010) and HEases (Gimble 2007; Stoddard et al., 2008; Maracaida et al., 2010).

Editing the human genome can be done with viral-vector mediated transfer of genes but many of these can lead to insertional mutations (Davé et al., 2009). Therefore strategies for genome editing to correct disease-causing alleles have been developed that are more site specific. ZFNs (Urnov et al., 2005; Durai et al., 2005 ; Klug, 2010) are artificial restriction enzymes that have been generated by combining a zinc finger (ZF) binding (DNA-binding/recongnition) domain to a DNA-cleavage domain (Cathomen and Joung, 2008; Sander et al., 2007). The zinc-finger domain can be engineered to target specific sequences. Thus genome editing can be achieved by the ZFN inducing a site-specific DSB at a programmed target site within the genome, inducing homology-directed repair using a corresponding gene segment present on a gene-targeting vector that serves as a template for DSB repair and therefore promotes gene replacement (Santiago et al., 2008). The ZF module is relatively small compared to other site-specific DNA binding domains; however ZFN sequence specificity optimization is quite involved to avoid toxic cleavage events at ectopic sites. Also there is a risk that the communication between the ZF and the endonuclease domains is not always precise leading to cleavage at ectopic sites (Uronov et al., 2005; Szczepek et al., 2007; Pattanayak et al., 2011). Nevertheless the recent success in using ZFNs in modifying the HIV CCR5 coreceptor surface protein in the T lymphocytes of AIDS patients is an impressive accomplishment (Perez et al., 2008). This and other recent successes in gene repair do demonstrate the potential of site-specific endonucleases such as ZFNs and meganucleases such as HEases in genome editing applications (Silva et al., 2011). An alternative to engineered ZFNs are HEases; there is a large natural reservoir of HEases within organellar genomes of many eukaryotes, and within the genomes of bacteria, Archaea and phages (see Figure 1.9;

Barzel et al., 2011; Li et al., 2011; Takeuchi et al., 2011; Hausner, 2012). Natural HEases, genetically modified HEases and synthetic HEases have been optimized and designed or redesigned for targeting specific sequences such as for human monogenic diseases (Stoddard, 2011).

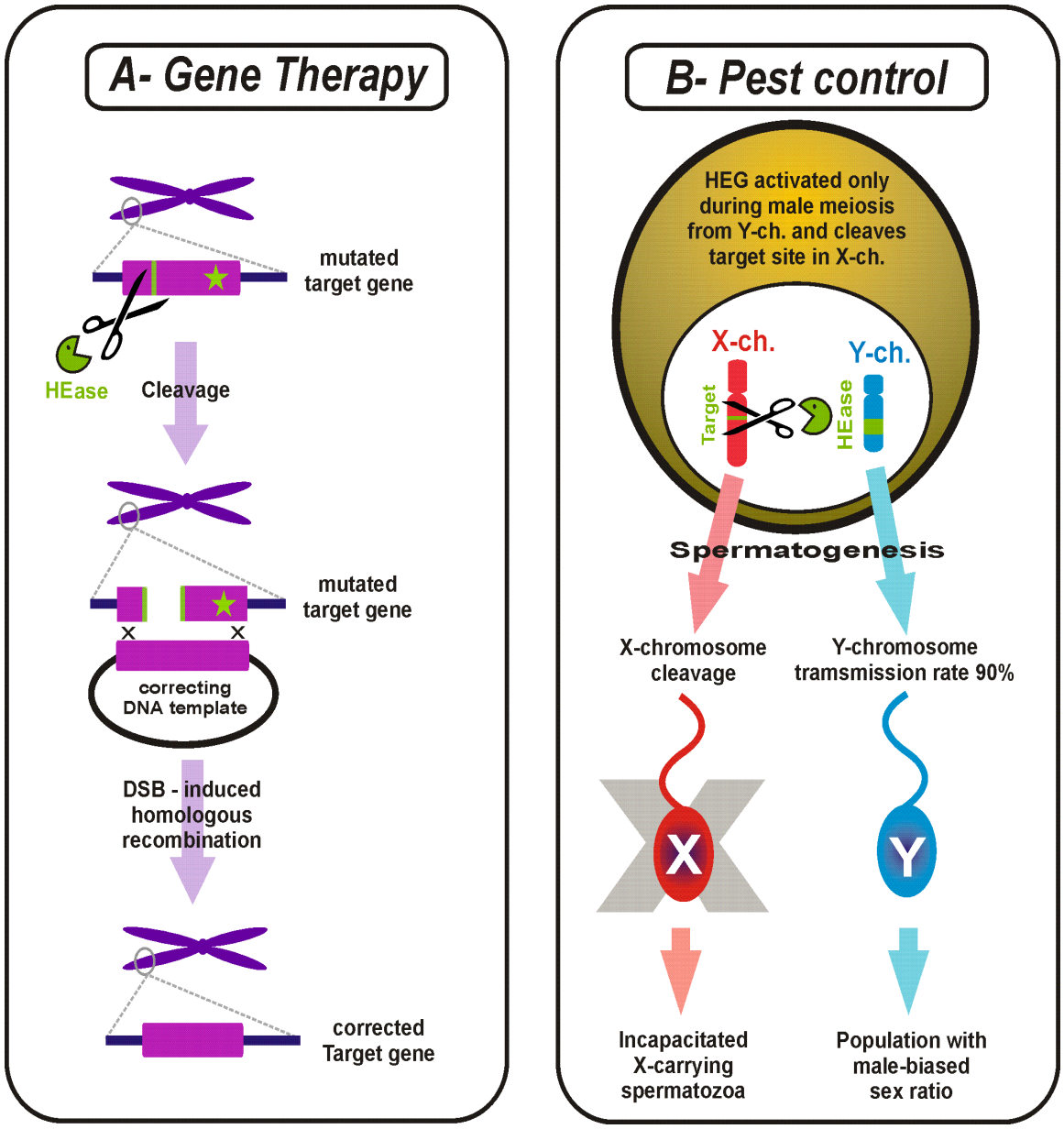
Naturally occurring LHEases are the most frequently encountered HEases (Takeuchi et al., 2011) and they tend to generate double-stranded breaks that can be either repaired as stated previously by HR, depending on mechanisms requiring a template to repair the damage, or by NHEJ (Non-Homologous End Joining) a method that can directly repair the break by rejoining the free DNA ends. However, the latter repair mechanism can result in loss of sequence information at the repair junction (Aubert et al., 2011), and in some cases promote chromosome translocation (Jeggo, 1998; Iliakis et al., 2004; Symington and Gautier, 2011). For gene targeting and precise gene replacement (repair) conditions have to be created that favour HR depending mechanisms to ensure no loss of genetic information occurs at the DSB (Muñoz et al., 2011). Some engineered variants of LHEases (Niu et al., 2008; McConell Smith et al., 2009; Chan et al., 2011) and some native members of the H-N-H family of HEases cut only a single strand of the DNA substrate and yet can still promote intron homing (Landthaler et al., 2006). Such site specific nicking HEases are viewed as promising candidates for promoting HR repair scenarios during HEase induced gene targeting and gene replacement experiments, as DSB which can generate genomic instabilities can be avoided when only one strand of the target site is disrupted (Davis and Maizels, 2011).

Although the HR based DNA repair promotes accurate gene replacements in some instances, where HEases are used to target specific sequences to induce mutations, the NHEJ mechanism is actually favored. For example recently an engineered variant of I-AniI (Y2 I-AniI) was used to successfully target and disrupt an integrated lentivirus within immortalized human cell lines by having the HEase target a sequence within the viral DNA. The DSB introduced by the HEase introduced small deletions and insertions surrounding the cleavage site during the NHEJ based repair process (Aubert et al., 2011). The long term objective of this project is to develop methods to selectively disrupt integrated HIV proviruses within latently infected cells, but this approach can also be envisioned as a genetic tool to generate mutations in eukaryotic cell lines.

Homing endonucleases hold great promise for generating a variety of genetic tools that allow for site-directed mutagenesis in a variety of organisms (Flannagan et al., 2008; Siegl et al., 2010; Wang et al., 2012). For example a system referred to as "*Delitto perfetto***"** is a genetic method for *in vivo* site-directed mutagenesis in *S. cerevisiae* where the I-SceI LHEase is used to generate a DSB at the appropriate target site and thus induce the DSB repair process which increases the frequency of targeted HR by 4,000 fold compared to experiments where DSB were not generated (Storici and Resnick, 2003).

Figure 1.10. Two areas of applications for HEases. [a] Gene replacement (gene therapy): mutated genes can be corrected if it is cleaved (scissor symbol) with a HEase (pac-man symbol) in the presence of a correcting DNA "exogenous donor". The mutated gene is repaired via DSB-induced HR using the correcting DNA as template; [b] pest control: HEases can be used to affect pest population. A HEG with a meiosis-specific promoter can be located on the Y chromosome and when expressed during gametogenesis the HEase cleaves a target site in the X chromosome leading to incapacitation of the X-carrying spermatozoa and therefore this will generate a male-biased population (Deredec et al., 2011)

Figure 1.10.


A- Gene Therapy
mutated target gene
HEase
Cleavage
mutated target gene
correcting DNA template
DSB - induced homologous recombination
corrected Target gene
B- Pest control
HEG activated only during male meiosis from Y-ch. and cleaves target site in X-ch.
X-ch.
Y-ch.
Target
HEase
Spermatogenesis
X-chromosome cleavage
Y-chromosome tramsmission rate 90%
X
Y
Incapacitated X-carrying spermatozoa
Population with male-biased sex ratio

Meganucleases are continuously engineered for targeted genome engineering with some applications relating to human gene therapy (Silva et al., 2006; Smith et al., 2006; Paques and Duchateau, 2007; Pingoud and Silva, 2007; Ashworth et al., 2010; Muñoz et al., 2012; Pessach and Notarangelo, 2011). For example derivatives of I-CreI HEases have been generated to target human genes involved in monogenic diseases (Arnold et al., 2006 & 2011). The human gene *XPC* (encoding Xeroderma Pigmentosum complementation group C) has been successfully targeted and cleaved with an engineered derivative of I-CreI with no obvious signs of genotoxicity (Arnold et al., 2006; Prieto et al., 2012). The human *RAG1* gene has been targeted by an altered I-CreI enzyme to correct mutations that give rise to severe combined immunodeficiency (SCID) phenotypes and the HR induced repair was able to induce high levels of gene correction with minimal toxicity (Grizot et al., 2009).

In addition to gene therapy, HEases have been proposed to manipulate the composition of natural populations of pests (Figure 1.10. [b]) such as invasive species, pathogens, or vectors of pathogens by decreasing population fitness and thus driving down population densities (Burt, 2003; Burt and Trivers, 2006; Henzell et al., 2008; Chan et al., 2011). The goal is to incorporate a HEG within the genome of a target species and the HEase could be programmed to target genes involved in sex determination, fertility, or key genes in the vectors of a particular pest, i.e. genes required for efficient pathogen transmission (Deredec et al., 2008 & 2011; Windbichler et al., 2008). In Australia the development of genetic tools is underway to control destructive invasive species such as

carp and cane toads by developing “daughterless” progeny with the aid of genetic constructs that include HEGs (Saunders et al., 2010).

Recently *Drosophila* and *Anopheles gambiae* have been investigated as model systems for developing approaches to develop populations with male sex ratio biases (Chan et al., 2011; Windbichler et al., 2008). For examples HEGs are applied as a potential solution for effective malaria control by targeting mosquito (*A. gambiae*) populations that vector the parasite. HEases are considered promising tools to transfer genetic modifications from engineered laboratory mosquitoes to wild-type population, due to their ability to propagate within the genome. As they are inherited in a non-Mendelian way, populations could be affected within fewer than 20 generations by a HEase-based gene drive mechanism (Windbichler et al., 2011).

Windbichler et al. (2007) demonstrated that in *A. gambiae* the HEase I-PpoI can cut a genomic rDNA site located on the X chromosome. This strategy could eliminate X-carrying spermatozoa and favour a severe male-biased sex ratio (Windbichler et al., 2007). The HEG introduced in a genome would be under the control of a meiosis-specific promoter to allow for the normal development of the heterozygous zygotes and during meiosis the HEG could be expressed and the HEase could disrupt its target site by cleavage activity in a certain fraction of gametes and resulting homozygous zygotes will die (Windbichler et al., 2008).

The application of re-engineered natural HEases in plant biotechnology is also gaining interest for developing transformation vector systems, plant genome editing, and for targeted mutagenesis (Yang et al., 2009; Gao et al., 2010; Vainstein et al., 2011; Zeevi et al., 2012). Finally natural and engineered HEases can also be used as reagents for molecular biology, similar to type II restriction enzymes. Commercially available HEases are used for linearizing large insert type cloning vectors (i.e., BACS) that have been engineered to include a specific HEase target site allowing recombinant clones to be linearized. Also rare cutting HEases are used in genomics to generate large DNA fragments suitable for pulse field electrophoretic studies (Gimble, 2005; Marcaida et al., 2010; Siegl, 2010).

1.10.1. Engineering homing endonucleases:

Sequence-specific endonucleases such as type II restriction enzymes played an important role in the rapid progress of molecular biology and biotechnology. One possible drawback in today's era of genomics is that type II restriction enzymes cut frequently, as they recognize small target sequences (4 - 8 bp), whereas HEases recognize relatively long sequence (14 - 44 bp) making them suitable for genomics and genetically modifying organisms. However, so far most applications are based on a limited number (I-CreI, I-AniI, I-SceI and I-OnuI) of well characterized HEases (Takeuchi et al., 2011; Prieto et al., 2012); therefore there is a need to bioprospect and characterize more native HEases or modify existing HEases or engineer HEases so that a wider choice of target sites can be substrate to HEases. A variety of techniques have been explored to generate

HEases and to modify recognition site specificity of HEases such as domain shuffling or applying genetic screens and selection (Arnould et al., 2011; Ashworth et al., 2010).

Domain shuffling is one of the most important strategies to alter HEase specificity, and it has been demonstrated that new active chimeric HEases can be created by fusing domains from unrelated HEases. This mixing/matching of HEase parts can result in endonucleases that recognize chimeric DNA target sites (Chevalier et al., 2002; Steuer et al., 2004). The E-DreI (now named H-DreI; Roberts et al., 2003) is a highly active artificial enzyme created by fusing domains from two different LHEases: I-DmoI and I-CreI; and H-DreI recognizes a hybrid substrate derived from each original parent donors (Chevalier et al., 2002; Epinat et al., 2003). Thus, individual domains appear to be highly modular and are responsible for recognition of individual DNA target half-sites and these modules can be shuffled and recombined to alter the DNA-binding specificity.

Another kind of chimeric endonuclease is made by combining a HEase DNA-binding module with the cleaving module of a restriction enzyme. For example it was shown that a mutated version of the LHEase I-SceI HEase, which can still bind to DNA but cannot cleave DNA when fused with the non-specific catalytic domain of the class IIS restriction enzyme FokI with the aid of a polypeptide linker, resulted in a hybrid chimeric endonuclease that cleaves double-stranded DNA at a defined site outside the original target site (Lippow et al., 2009). Another site-specific endonuclease was engineered by fusing the specific catalytic domain of the class IIP restriction enzyme

PvuII with the DNA-binding domain of the catalytically inactive variant I-SceI (Fonfara et al., 2012).

Typically for HEases to be characterized and potentially reprogramed one needs to first establish their native target sites followed by studying the structure of the protein by obtaining crystals where the enzyme was co-crystalized with its natural substrate (Kowalski and Derbyshire, 2002; Takeuchi et al., 2011). The latter, upon detailed analysis, allows for the determination of a detailed contact map which shows which amino acid moieties are actually in physical contact with the DNA target sequence. This allows for the generation of a series of mutations whereby amino acid substitutions are introduced along the DNA binding and cleavage domains of the HEase. This is followed by experiments to examine the new binding and cutting specificity of the mutated form of the HEase (Gao et al., 2010; Takeuchi et al., 2011). The recent study by Takeuchi et al., (2011) is a nice demonstration how native HEases from fungal mitochondrial genomes (Sethuraman et al., 2009a) can be re-engineered to target a variety of human genes associated with diseases. These methods are mainly based on: selection of high affinity DNA binding activity (Gimble et al., 2003) and selection for DNA cleavage activity (Seligman et al., 2002; Chen and Zhao, 2005; Doyon et al., 2006).

Controlling the cleavage activity of HEases is an extremely important feature in order to optimize these elements as tools for gene replacements. Inserting “switches” that allow the HEase to be reversibly turned on and off would increase the sensitivity of the HEase to cellular conditions and reduce potential toxic effects on the cell when

endonuclease activity is not desirable. Posey and Gimble (2002) showed that the yeast PI-SceI HEase can be turned on and off, *in vitro*, by using a successive cycles of reducing and oxidizing (redox) treatments. Two cysteine residues were introduced into flexible DNA binding loops to yield intramolecular disulfide bonds that lock the PI-SceI into a nonproductive conformation under oxidizing conditions and the enzyme activity was decreased more than 30-fold compared to the wild-type PI-SceI. In the reduced state, the loops undergo a conformational change that allows the protein to contact the DNA target site and to cleave the substrate sequence. The cleavage activity of the enzyme under reducing conditions was noted to be similar or slightly lower to that of the wild-type PI-SceI. The major drawback of the redox switch is that this system is currently limited to *in vitro* applications.

1.10.2. Future prospects:

Bioprospecting for native HEGs and the future development of engineered highly site specific HEases will make this unique group of rare cutting enzymes the molecular scissors that can be used in many applications such as characterizing genomes, or for manipulating genomes for medical, agricultural and biotechnological applications. Optimization of gene replacement strategies has to involve a better understanding of selecting appropriate cells or increasing the probability of generating genetic events where DSB are repaired by HR based mechanisms in order to avoid the introduction of genomic instabilities. *In vivo* switches may be developed that can activate and deactivate these enzymes by small molecules (ligands) or external stimuli without affecting the cellular metabolism of the cell. In most HEases the DNA-binding domains are not

distinct from the cleavage domain this makes engineering these enzymes more difficult but these enzymes may have greater selectivity for the intended target site than the engineered ZFNs or TAL (Transcription Activator-Like) effector (E) nucleases (Vainstein et al., 2011; Schleifman et al., 2008; Silanskas et al., 2012). TALE nucleases are artificial endonucleases that are the result of the fusion of a DNA binding domain (that can be redesigned) obtained from transcription activator-like proteins excreted by *Xanthomonas* bacteria to a non-specific DNA cleavage domain from the FokI endonuclease (Christian et al., 2010). These enzymes can be engineered to generate DSB at specific locations within the genome; thus similarly to HEases they can be used to edit genomes (Bogdanove and Voytas, 2011).

With the large number of yet to be discovered natural HEases there is a tremendous resource available within the organellar genomes that can be explored for finding new HEases. The characterization of these native HEases will reveal their target sites and this allows for scanning for the presence of their target sites in economically important organisms (genes) and for reprogramming native HEases to recognize new target sites such as genes involved in human monogenic diseases.

1.11. Research objectives:

1.11.1. Screening for the presence of introns and IEPs in the *rns* gene of ophiostomatoid fungi:

The *rns* gene can be considered a natural reservoir of introns and IEPs, so the objectives of this project were:

(1) PCR-based screening of the *rns* gene of several groups of ophiostomatoid fungi including strains of *O. ulmi*, *O. novo-ulmi*, and *O. minus* using universal primer pairs (mtsr-1/mtsr-2 and rns-F/rns-R) that bind at highly conserved sequences at the 5' and 3' termini of the *rns* gene;

(2) Selecting representative *rns* genes to be sequenced based on the amplicon size and their phylogenetic position;

(3) Identify and characterize potential insertions (introns and IEPs) that interrupt the *rns* gene by comparative sequence analysis;

(4) Attempt to resolve the evolutionary dynamics between the host gene, the introns and their IEPs;

(5) Evaluating the possibility of using introns as molecular markers to distinguish between different species within the Dutch Elm Disease causing species complex;

(6) and finally to provide an overview of the various types of elements that can insert within the nuclear SSU rDNA and the mtDNA *rns* gene among the ophiostomatoid fungi. The latter should prove to be useful in annotating the *rns* gene for fungi in future studies. The data are also valuable to those who are interested in prospecting for native HEases or ribozymes for potential applications in biotechnology.

1.11.2. The biochemical characterization of homing endonucleases encoded within group I and II introns:

Based on the information generated in the first project, two LHEases encoded within group I and group II introns were subjected to a detailed biochemical analysis. The *rns* gene of *O. minus* strain WIN(M)371 was found to be interrupted by two introns at positions S569 and S952; the former is a group IC2 intron while the latter is a group IIB1 intron and they both encode double motif LHEases. The objective of this project is to overexpress these two HEases (named I-OmiI and I-OmiII) in *E. coli* and purify them in order to test their endonuclease activity and if these IEPs are active LHEases and to map the cleavage sites of these enzymes. Essentially these are the first steps in assessing if these HEases could have applications in biotechnology.

CHAPTER: 2

GENERAL MATERIALS AND METHODS

2.1. Fungal strains and growth conditions:

For routine culturing the strains were maintained in Petri plates containing 2% Malt Extract Agar (MEA; Difco, Michigan, USA) supplemented with 1 gl^{-1} yeast extract (YE; Difco) and 20 gl^{-1} bacteriological agar (Fisher Scientific, location). From these cultures, agar plugs were cut containing fungal growths, which were used to inoculate 250 ml conical flasks containing 50 ml peptone yeast-extract agar (PYG) liquid medium (1 gl^{-1} peptone, 1 gl^{-1} yeast extract and 3 gl^{-1} glucose). Liquid cultures were incubated at 20 ^{o}C for 7 days to generate biomass for DNA and/or RNA extractions.

2.2 DNA extraction:

Fungal genomic DNA was extracted as previously described (Kim et al., 1990; Hausner et al., 1992). Fungal mycelia were harvested by filtering the cultures through a Whatman # 1 filter paper; mycelia were disrupted by vortexing in 3 ml of extraction buffer [1 M Tris-Cl (pH 8.0), 0.5 M $Na_2EDTA{\cdot}2H_2O$ (pH 8.0), 5 M NaCl, 10 % (w/v), cetyltrimethyl ammonium bromide (CTAB, Sigma)] and 3 ml of glass beads (Fisher Scientific), and the lysate was then incubated at 50 - 60 °C for a minumum of 2 hours in 3 ml of extraction buffer and 660 µl of 20 % (w/v) sodium dodecyl sulphate (SDS). Cell debris, denatured proteins, lipids, etc. were extracted in 6 ml chloroform and pelleted by centrifugation at 2000 rpm for 20 minutes at room temperature. The top aqueous layer was removed and the DNA was precipitated first with the addition of 2.5 volumes of ice-cold 95 % ethanol.

The tube was then stored for at least 3 hours (or overnight) at -20 °C and the DNA was pelleted by centrifugation (3000 rpm for 30 minutes) and the DNA pellet was washed with 1 ml of 70 % ethanol. The dried pellet was resuspended in 300 µl of 1X Tris-EDTA (TE) buffer [10 mM Tris-Cl (pH 7.6) and 1 mM $Na_2EDTA{\cdot}2H_2O$ (pH 8.0)].

2.3. PCR amplification:

The nuclear internal transcribed spacers (ITS1, 5.8S and ITS2) were amplified by the polymerase chain reaction (PCR) using the Invitrogen enhancer system (Invitrogen, Burlington, Canada) at the 1.5X rate containing the following ingredients (µl/reaction): 10X Taq DNA polymerase buffer (Stratagene, Agilent Technologies, La Jolla, USA) (5); 2.5 mM dNTP (4); 50 mM $MgSO_4$ (1.5); 40 pmol each of forward and reverse primers (0.5+0.5); 10X PCR enhancer solution (7.5); Taq DNA polymerase (0.25), H_2O (29.75); and genomic DNA (~10 to 100 ng) template (1). The oligonucleotide primer pairs SSU-Z/LSU-4 and SS-3/LS-2 (Alpha DNA, Montréal, Canada; see Figure 2.1.[a] and Table 2.1.) were used for amplification of the ITS region (Hausner et al., 1993b; Hausner and Wang, 2005; Mullineux and Hausner, 2009). The PCR conditions were as follows: initial denaturation at 95 ^oC for 2 minutes followed by 30 cycles of denaturation (95 ^oC for 1 min), annealing (52 ^oC for 1.5 min) and extension (70 ^oC for 1 minute) plus a final extension step at 70 ^oC for 10 minutes.

Figure 2.1. A schematic overview of **[a]** the internal transcribed spacer region (ITS1 & ITS2); **[b]** the translation elongation factor 1 alpha gene and **[c]** the mitochondrial *rns* gene, showing the initial primers used to amplify each region/gene. Figures are not drawn to scale.

Figure 2.1.

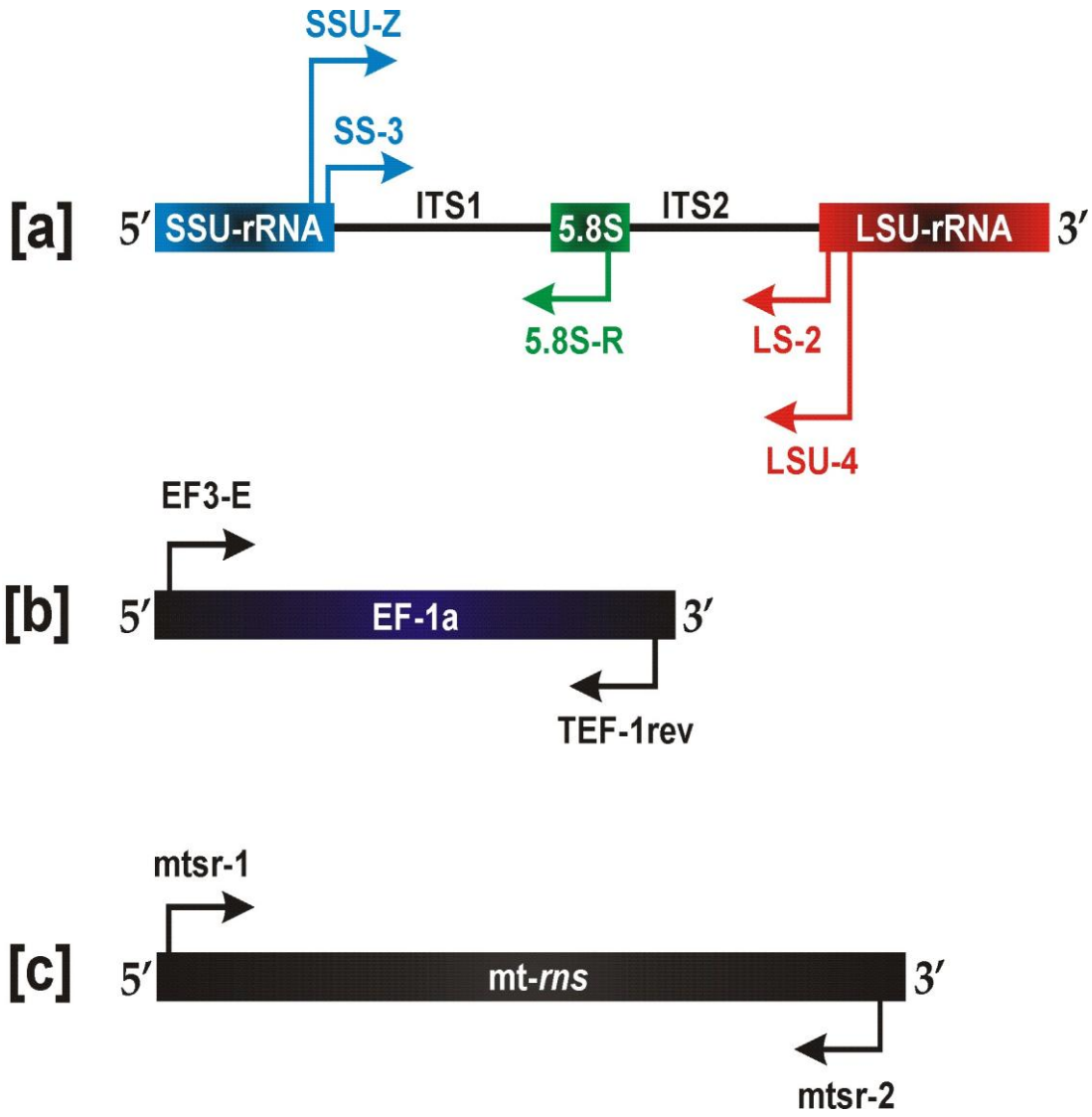

SSU-Z
SS-3
[a]
5′
SSU-rRNA
ITS1
5.8S
ITS2
LSU-rRNA
3′
5.8S-R
LS-2
LSU-4
EF3-E
[b]
5′
EF-1a
3′
TEF-1rev
mtsr-1
[c]
5′
mt-rns
3′
mtsr-2

Table 2.1. A list of the initial primers used to amplify the ITS, EF-1alpha and *rns* genes.

Table 2.1.

Forward primer		Reverse primer	
Primer name	**Sequence (5' → 3')**	**Primer name**	**Sequence (5' → 3')**
SS-3	GTCGTAACAAGGTCTCCG	LS-2	GATATGCTTAAGTCAGCG
SSU-Z	ATAACAGGTCTGTGATG	LSU-4	TTGTGCGCTATCGGTCTC
EF3-E	GTCGTYATCGGCCACGTCGA	5.8S-R	GACGCTCGGACAGGCATGCC
mtsr-1	AGTGGTGTACAGGTGAG	TEF-1rev.	GCCATCCTTGGAGATACCAGC
		mtsr-2	CGAGTGGTTAGTACCAATCC

The nuclear translation elongation factor gene 1 alpha (EF-1α) was amplified with primers EF3-E and TEF-1rev (Lim et al., 2004); the primer binding sites are shown in Figure 2.1[b] and primer sequence are described in Table 2.1. Amplifications were carried out in a reaction mix containing the following ingredients (µl/reaction): 10X Taq DNA polymerase buffer (Stratagene) (5); 50 mM $MgCl_2$ (0.5); 2.5 mM dNTP (4); 40 pmol each forward and reverse primer (0.5 + 0.5); H_2O (38.25); genomic DNA template (1); and Taq DNA polymerase (0.25). The EF-1α amplification was carried out using the following PCR conditions: initial denaturation at 94 ^oC for 2 minutes, followed by 25 cycles of denaturation (94 ^oC for 30 sec), annealing (56 ^oC for 30 sec) and extension (70 ^oC for 80 sec); followed by a final extension step at 70 ^oC for 5 minutes. The concentration of dGTP and dCTP as well as the annealing temperature were optimized as required.

The mitochondrial *rns* gene was amplified as previously described in Mullineux et al. (2010) with the primer mtsr-1 and mtsr-2 (see Figure 2.1[c] and Table 2.1) using standard DNA amplification protocols utilizing the Invitrogen PCR system (Invitrogen). Amplifications were carried out in a reaction mix containing the following ingredients (µl/reaction): 10X Taq DNA polymerase buffer (Stratagene) (5); 50 mM $MgCl_2$ (0.5); 2.5 mM dNTP (4); 40 pmol each forward and reverse primer (0.5 + 0.5); H_2O (38.25); genomic DNA template (1); and Taq DNA polymerase (0.25). The PCR conditions for the mtsr-1 and mtsr-2 primers were as follows: an initial denaturation at 93 ^oC for 1 minute followed by 25 cycles of denaturation (93 ^oC for 1 minute), annealing (55 ^oC for 1

minute) and extension (70 ^{o}C for 5 minutes) followed by a final extension step at 70 ^{o}C for 10 minutes. The annealing temperature was optimized as required.

In some instances (to optimize the PCR amplification of *rns* gene) the Invitrogen PCR system was replaced with the OneTaq Hot Start DNA Polymerase (New England Biolabs, Pickering, ON, Canada). For the latter system amplifications were carried out in a reaction mix containing the following ingredients (µl/reaction): 5X OneTaq DNA polymerase standard reaction buffer (10 µl); 10 mM dNTPs (1µ); 40 pmol each forward and reverse primer (1 µl +1 µl); H_2O (35.75 µl); genomic DNA template (1 µl); and Taq DNA polymerase (0.25 µl). The PCR conditions for the mtsr-1 and mtsr-2 primers were as follows: an initial denaturation at 94 ^{o}C for 30 seconds followed by 30 cycles of denaturation (94 ^{o}C for 30 seconds), annealing (55 ^{o}C for 1 minute) and extension (68 ^{o}C for 1 minute/kb) followed by a final extension step at 68 ^{o}C for 5 minutes.

Oligonucleotides primers utilized for both PCR amplification and DNA sequencing of the nuclear SSU rDNA gene were characterized in Hausner et al. (1993a), Hausner and Reid (2004), Gibb and Hausner (2005) and Hafez et al. (2012). The primers SSU-J and SSU-T were initially used to screen our collection for the potential of nuclear SSU rDNA insertions. For some strains the sequences were extended across the ITS regions to examine group I introns located near the 3' end of the SSU rDNA gene; here the SSU-Z and LSU-4 primers (described in Hausner and Wang, 2005) were utilized to obtain the appropriate regions. PCR conditions for the primers above are described in Hausner and Reid (2004) and Hausner and Wang (2005).

All PCR amplicons were analyzed by gel electrophoresis through 1% agarose gel in Tris-Borate EDTA (TBE) buffer (89 mM Tris-borate, 10 mM EDTA, pH 8.0). The DNA fragments were sized using the 1-kb plus ladder (Invitrogen), and nucleic acids were visualized by staining in 1X TBE buffer supplemented with 0.5 µg/ml ethidium bromide (EtBr) and exposing the stained gels with Ultra violet (UV) light.

2.4. PCR products purification:

PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, USA), the solution containing the PCR product was mixed with an equal volume of binding solution. In the case of purifying the PCR product from an agarose gel, the gel slice that contained the band of interest was dissolved in membrane binding solution (10 µl/10 mg) and vortexed, then incubated at 50 -65 °C until the gel slice was completely dissolved. The membrane binding solution that contained the PCR product was then loaded onto a mini-column and incubated for 1 minute at room temperature, and then the tube was centrifuged at 13000 rpm for 1 minute. The mini-column was washed with 700 µl of membrane wash solution and then centrifuged as above. The sample bound within the mini-column was washed a second time with 500 µl of membrane wash solution and centrifuged as above for 5 minutes. The purified PCR product was eluted from the mini-column in 50 µl of nuclease-free H_2O and stored at -20 °C for DNA sequencing.

2.5. Cloning of PCR products and purification of plasmid DNA:

The PCR products were cloned into *E. coli* (DH5α) using the TOPO TA Cloning® kit (Invitrogen) to improve sequencing efficiency. Four µL of the purified amplicon was mixed with 1 µl of the pCR®4-TOPO® vector and 1 µl salt solution (1.2 mM NaCl; 0.06 mM $MgCl_2$) and the entire reaction mix was transformed with One Shot® MAX Efficiency® DH5α Chemically Competent *E. coli* according to the recommended procedure in the TOPO-TA Cloning® Kit for Sequencing (Invitrogen). The transformed *E. coli* cells were incubated on ice for 30 minutes and then heat shocked at 42 °C for 30 – 60 seconds; thereafter 250 µl **S**uper **O**ptimal broth with **C**atabolite repression **(**SOC) medium was added and then incubated at 37 °C for 1 hour. The culture was then inoculated onto Luria Bertani (LB) agar supplemented with ampicillin (60 µg/ml) and X-GAL dissolved in dimethyl formamide (40 mg/ml). Putative transformants were identified using blue-white selection and were screened by PCR as described above except that the positive clones (*E. coli* colonies that contain the insert of interest) were used as a source of DNA template for sequencing. Plasmid DNA was harvested from 3 ml of overnight liquid (LB) cultures and purified using the Wizard® *Plus* Minipreps DNA purification system (Promega) and eluted in a final volume of 50 µl of nuclease-free water and stored at -20 °C for DNA sequencing.

2.6. DNA sequencing:

The double-stranded DNA fragments were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. The sequencing products were denatured and

resolved on a 3130 genetic analyzer (Applied Biosystems). Initially vector based primers as supplied by the TOPO cloning kit: M13 Forward, M13 Reverse, T7 (forward), and T3 (reverse) were used to obtain sequences; thereafter primers were designed as needed to complete all sequences in both directions (see Table 2.1 and Appendix 9.3. for primer list).

2.7 Sequence and phylogenetic analysis:

Individual sequences were compiled and assembled manually into contigs using the GeneDoc program v2.5.010 (Nicholas et al., 1997) and nucleotide sequence alignments were done with the Clustal-X program (Thompson et al., 1997); the Clustal-X program was also used to convert data sets into the PHYLIP file format needed for phylogenetic analysis. The ORF finder program (http://www.ncbi. nlm.nih.gov/gorf/gorf.html; Genetic code setting for molds #4) was used to search for potential ORFs within the mt-*rns* introns.

The online resource Basic Local Alignment Search Tool (BLAST: http://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al., 1990) was used to retrieve nucleotide sequences (BLASTn) and amino acid sequences (BLASTp) from GenBank (National Center for Biotechnology Information) which shared similarities to the mt-*rns* and the intron encoded ORFs, respectively. The amino acid sequence data sets were aligned with the online multiple sequence alignment program PRALINE (Simossis & Hirenga 2005; http://www.ibi.vu.nl/ programs/pralinetm/). The resulting alignments were examined and adjusted with the GeneDoc program.

Phylogenetic estimates were generated for all alignments by the programs contained within the PHYLIP package (Felsenstein, 2006) and the MrBayes program v3.1 (Ronquist and Huelsenbeck, 2003; Ronquist, 2004). In PHYLIP (Felsenstein, 2006), phylogenetic trees were obtained by analyzing the nucleotide and amino acids alignments with the DNAPARS and PROTPARS programs, respectively, both in combination with bootstrap analysis (1000 replicates, SEQBOOT) and CONSENSE to obtain the majority rule consensus trees. Phylogenetic estimates were also generated within PHYLIP using the NEIHGBOR program using distance matrices generated by DNADIST (K84 setting) and PROTDIST (setting: Dayhoff PAM250 substitution matrix; Dayhoff et al., 1978) for nucleotide and amino acid alignments, respectively.

The MrBayes program was used for Bayesian analysis and the parameters for amino acid alignments were as follows: mixed models 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 1 000 000 generations; trees were sampled every 100 generations. The first 25 % of trees generated were discarded ("burn-in") and the remaining trees were used to compute the posterior probability values. For nucleotide sequence alignments the GTR model with gamma distribution was applied and as above; four chains were run simultaneously for 1 000 000 generations with sample frequency of 100 and a "burn-in" corresponding to the first 25% of sampled trees.

Phylogenetic trees were drawn with the TreeView program (Page, 1996) using PHYLIP tree outfiles or MrBayes 50% majority rule tree files, and annotated with Corel Draw™ (Corel Corporation, Ottawa, Canada).

2.8. Intron nomenclature and secondary structure modeling:

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

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

2.9. RNA extraction and Reverse Transcription-PCR (RT-PCR):

For RNA extraction, cultures were grown in 125 ml conical flasks containing 50 ml PYG liquid medium for up to 7 days at 20 ^{o}C. Fungal biomass was collected by vacuum filtration using Whatman # 1 filter paper and the wet mycelium was flash-frozen in liquid nitrogen and ground to a fine powder using a precooled mortar and pestle. The RNA was extracted using the RNeasy® mini kit (QIAGEN, Mississauga, ON, Canada) following to the manufacturer's procedures. In addition to the DNase step within RNeasy® mini kit procedure the TURBO™ DNase kit (Ambion, Austin, TX, USA) was also applied according to the manufacturer's protocol to ensure the removal of all DNA from the final RNA preparation. The TURBO ™ DNase was inactivated by heating at 75 ^{o}C for 10 min. The ThermoScript™ RT-PCR system (Invitrogen) was used to synthesize cDNA using approximately 100 ng of template RNA. First strand synthesis was carried out with 40 pmol of primer mtsr-2 and subsequent PCR amplification was carried out with primers mtsr-1 and mtsr-2 both at 40 pmol per reaction usning the same PCR conditions described before to amplify the *rns* gene (section 2.3.). The PCR products generated by the RT-PCR reactions were sequenced as described previously.

2.10. Ancestral state reconstruction:

Estimation of the ancestral character state for the presence or absence of the mitochondrial group II intron (S379) or the nuclear group I introns (S943, S989 and S1199) among *Ophiostoma* and *Ceratocystiopsis* species were done with the MESQUITE program version 2.73 (Maddison and Maddison, 2010). A 50% majority rule consensus trees for the mitochondrial and nuclear SSU rRNA exon sequences were generated by the

MrBayes program and these trees were used as a base for reconstruction of the ancestral states. The evolutionary history of each character was traced over the 50% majority rule Bayesian tree using Maximum-Parsimony and Maximum-Likelihood ancestral state framework available in MESQUITE version 2.73. The Markov k-state 1 parameter model, which gives equal probability (or rate) for changes between any two character states, was used. In the analysis, each taxon was scored (1) or (0) for presence or absence of the intron respectively.

2.11. Over expression and Purification of I-OmiII:

2.11.1. Construction of the expression plasmid:

To express the I-OmiII in *E. coli*, a codon-optimized version of the I-OmiII sequence was synthesized commercially (Bio S & T, Montreal, QC, Canada) based on differences between the fungal mitochondrial and bacterial genetic code. The optimized I-OmiII ORF sequence was cloned into pBlueScript SK+ (insert site SmaI). The I-OmiII ORF was then amplified from the pBlueScript vector with the primers I-OmiII-F (forward; included the CACC sequence needed for the topoisomerase cloning step that follows) and I-OmiII-R (reverse) using the Platinum Taq DNA Polymerase High Fidelity kit (Invitrogen). The amplified I-OmiII ORF was then moved into the 6xHis-Tag N-terminus pET200/D-TOPO® vector supplied in the Champion™ pET Directional TOPO® expression kit (Invitrogen) following the manufacturer's recommendations to generate the pET200/D/I-OmiII construct (Figure 2.2). To enable the directional cloning of the PCR-amplified ORF, the I-OmiII-F primer was designed with a CACC 5' tail. The pET200/D/I-OmiII construct was subsequently transformed into the *E. coli* BL21 (DE3)

cell line (Invitrogen). The transformants were then analyzed by colony PCR and by treating isolated plasmid DNAs with restriction enzymes to identify potential positive clones (*E. coli*-pET200/D/I-OmiII). The construct was sequenced with the vector-specific primers T7-F and T7-R to confirm that the ORF was in frame with the N-terminal fusion 6xHis-Tag.

2.11.2. Overexpression of I-OmiII:

To express the I-OmiII protein in *E. coli*, 10 ml LB media supplemented with 100 µg/ml kanamycin and 0.25% w/v glucose was inoculated with 100 µl of *E. coli*-pET200/D/I-OmiII and incubated overnight (ON) at 37 ^{o}C. Five ml from the ON culture was inoculated into 1L of 2X-YT medium (tryptone 16 g l^{-1}, yeast extract 10 g l^{-1} and NaCl 5 g l^{-1}), and supplemented with 100 µg/ml of kanamycin and with 0.25% w/v glucose. The medium pH was adjusted to 7.0 with 5M NaOH. The culture was grown at 37 ^{o}C and induced when the OD600 nm reached ~ 0.6 – 0.8 with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The induced culture was shifted to 20 ^{o}C and incubated for 18 hours. Cells were harvested by centrifugation at 5000 rpm for 10 minutes and the pellet was frozen at -80 ^{o}C for at least 1 hour (see Figure 2.3.).

Figure 2.2. Construction of I-OmiI expression plasmid: The optimized I-OmiII ORF was amplified from the pBluescript plasmid with the primers I-OmiII-F and I-OmiII-R. The I-OmiII-F primer was designed with CACC 5′ tail to enable the directional cloning of the I-OmiII ORF into the pET200/D/TOPO plasmid and thus generating the expression plasmid pET200/D/I-OmiII.

Figure 2.2.

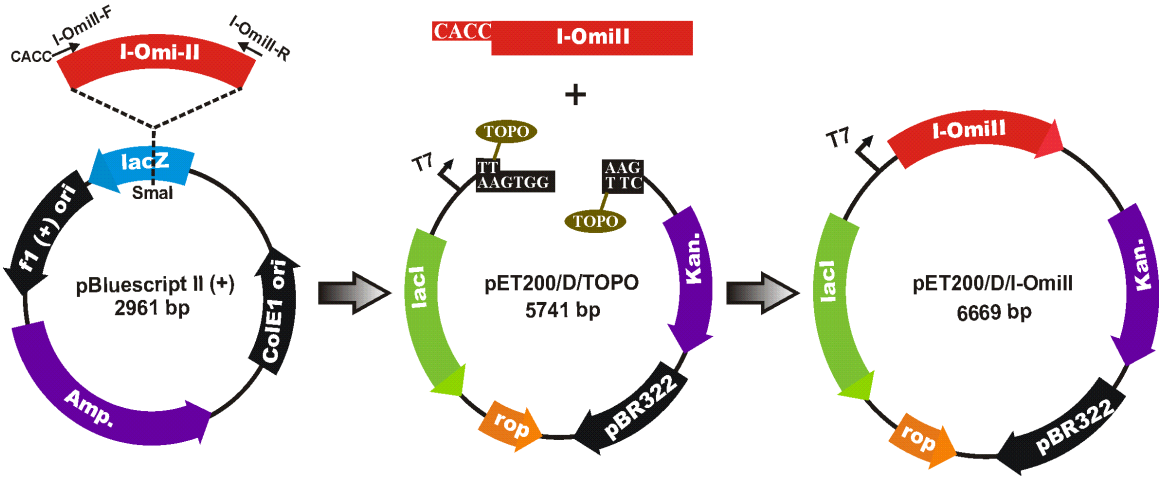
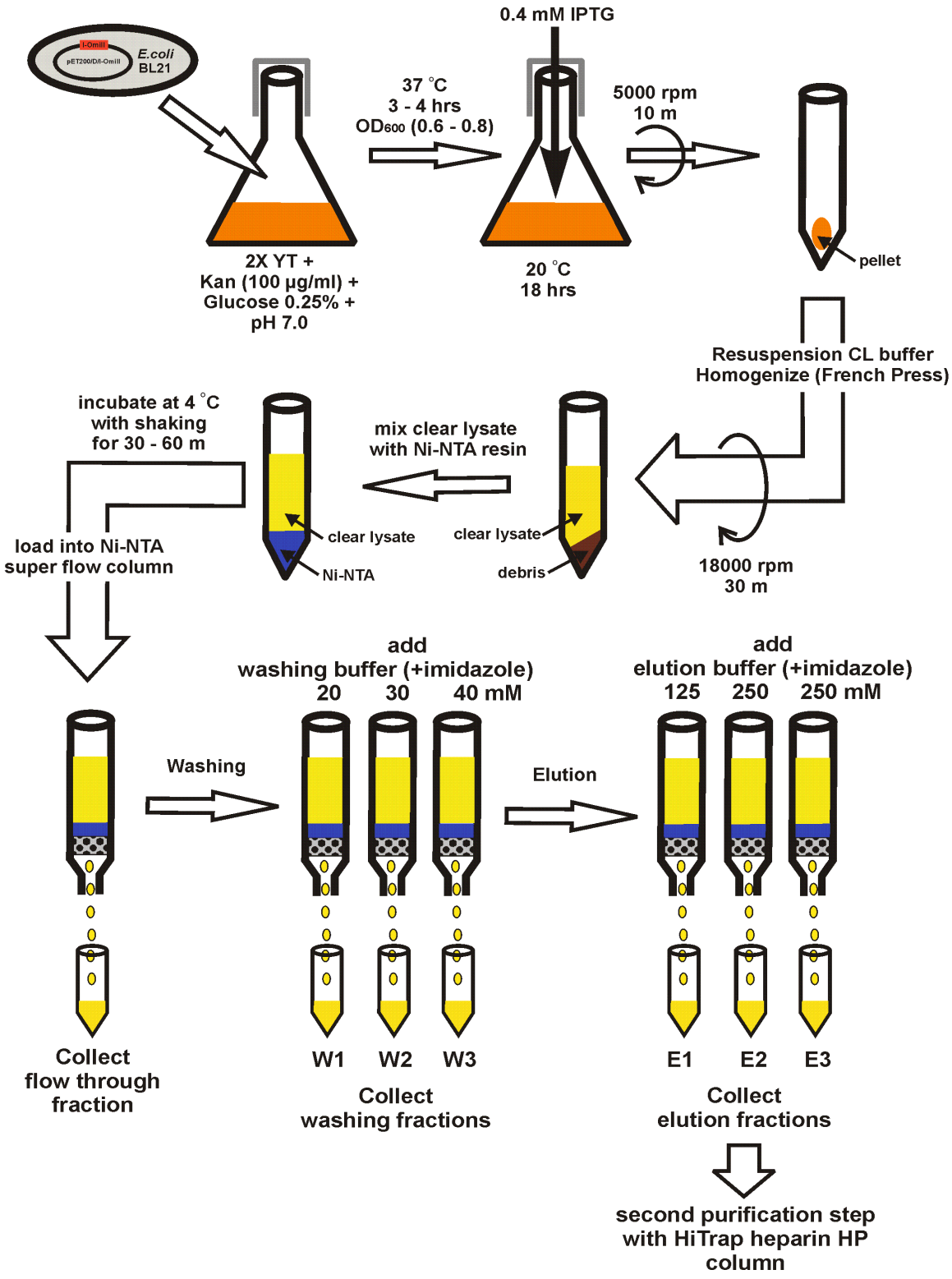

CACC
I-OmiII-F
I-Omi-II
I-OmiII-R
lacZ
SmaI
f1 (+) ori
pBluescript II (+)
2961 bp
ColE1 ori
Amp.
CACC
I-OmiII
+
TOPO
T7
TT
AAGTGG
AAG
TTC
TOPO
lacI
pET200/D/TOPO
5741 bp
Kan.
rop
pBR322
T7
I-OmiII
lacI
pET200/D/I-OmiII
6669 bp
Kan.
rop
pBR322

Figure 2.3. Overexpression and purification of the I-OmiII protein.

Figure 2.3.


E.coli
BL21
2X YT +
Kan (100 µg/ml) +
Glucose 0.25% +
pH 7.0
37 °C
3 - 4 hrs
OD600 (0.6 - 0.8)
0.4 mM IPTG
20 °C
18 hrs
5000 rpm
10 m
pellet
Resuspension CL buffer
Homogenize (French Press)
18000 rpm
30 m
clear lysate
debris
mix clear lysate
with Ni-NTA resin
clear lysate
Ni-NTA
incubate at 4 °C
with shaking
for 30 - 60 m
load into Ni-NTA
super flow column
Collect
flow through
fraction
Washing
add
washing buffer (+imidazole)
20 30 40 mM
W1 W2 W3
Collect
washing fractions
Elution
add
elution buffer (+imidazole)
125 250 250 mM
E1 E2 E3
Collect
elution fractions
second purification step
with HiTrap heparin HP
column

2.11.3. Purification of I-OmiII:

The I-OmiII protein was harvested by thawing the frozen pellet and resuspending the cell in (10 ml / 1 g wet weight) cell lysis (CL) buffer [40 mM HEPES (pH 7.5), 800 mM NaCl, 10 % (w/v) glycerol, 6 mM β-mercaptoethanol, and 250 µl / 1 g wet weight protease inhibitor cocktail (Sigma)]. Cells were homogenized using the French press two times and the lysate was centrifuged at 18000 rpm for 30 minutes at 4 ^{o}C to pellet the cell debris. The clear cell lysate (about 30 ml) was added to 3 ml of Ni-NTA resin (Qiagen) and incubated at 4 ^{o}C with shaking for 30 to 60 minutes.

The sample-Ni-NTA mixture was loaded onto a Ni-NTA super flow column (Qiagen) and the following series of washing steps were carried out; wash 1: 30 ml of washing buffer (WB) [40 mM HEPES (pH 7.5), 800 mM NaCl, 10 % (w/v) glycerol, 6 mM β-mercaptoethanol] supplemented with 20 mM of imidazole; wash 2: 30 ml of WB buffer with 30 mM of imidazole; and wash 3: 30 ml of WB buffer with 40 mM of imidazole. The protein was eluted in Elution buffers (EB) containing 40 mM HEPES, 800 mM NaCl, 20 % (w/v) glycerol, and either 125 mM or 250 mM imidazole (pH adjusted to 8 with NaOH). The sample was collected in 1 fraction (1 ml) with 125 mM imidazole and then in 2 fractions with 250 mM imidazole (1 ml each). Excess imidazole was removed by dialysis in buffer D1 (40 mM HEPES (pH 7.5), 200 mM NaCl, and 3 mM β-mercaptoethanol) using a slide-a-lyzer dialysis cassette with a 7 kDa molecular weight cut-off (Millipore, Billerica, USA) according to the manufacturer's suggested protocol.

A second purification step was carried out using a HiTrap™ heparin HP column (GE Healthcare Europe). The sample was washed with one column volume of wash buffer over a range of 200 mM to 1.5 M NaCl. In each wash fraction the NaCl concentration was increased by 100 mM; NaCl was adjusted by mixing the appropriate volumes of buffer D1 and buffer D2 (40 mM HEPES (pH 7.5), 1.5 M NaCl, and 3 mM β-mercaptoethanol). The fractions recovered with 400, 500 and 600 mM NaCl were combined and the sample was then concentrated using Amicon Ultracel centrifugal filters with 3000 Molecular Weight Cut Off (Millipore, Billerica, USA), following the manufacturer's guidelines. The sample was diluted to a final volume of 9 ml in a storage buffer [40 mM HEPES (pH 7.5), 400 mM NaCl, 1 mM DTT, and 30 % (w/v) glycerol], and centrifuged at 4000x g at 4 °C until the sample was concentrated in a final volume of 500 µl.

2.12. *In vitro* Endonuclease assay for I-OmiII:

Two plasmids, pCR4/1574 and pCR-4/371 were constructed to test the *in vitro* cleavage activity of I-OmiII. The mt-*rns* gene was amplified from *O. minus* strains WIN(M)1574 and WIN(M)371 with the primer pair mtsr-1 and mtsr-2. The resulting PCR products were cloned into the pCR-4 TOPO vector to generate the pCR4/1574 and pCR-4/371 substrate plasmids respectively. The pCR4/1574 construct contained the intact recognition site of I-OmiII (i.e. no intron at S952), while the pCR-4/371 construct served as a negative control as this *rns* gene has the I-OmiII target site interrupted with the mS952 intron. The plasmids were transformed into *E. coli* DH5α and the plasmids were purified with the Wizard Plus Minipreps DNA purification kit (Promega).

The cleavage reaction mix contained: 15 µl substrate plasmid (25 µg/ml), 5 µl Invitrogen Buffer React #3 (100 mM NaCl, 50 mM Tris–Cl, pH 7.9 and 10 mM $MgCl_2$) supplemented with 1 mM DTT, 5 µl I-OmiII (53 µg/ml) and 25 µl H_2O. Cleavage reactions were incubated at 37 ^oC and 10 µl aliquots were taken at 0, 30 and 60 minutes and the reactions were stopped with the addition of 2 µl of 200 mM EDTA (pH 8.0) and 1 µl of proteinase K (1 mg/ml) to each 10 µl aliquot, with subsequent incubation for 30 min at 37°C.

The pCR4/1574 plasmid was linearized with *Nco*I (Invitrogen) under conditions recommended by the manufacturer and the linearized plasmid was purified with the Wizard® SV Gel and PCR Clean-Up system (Promega, Madison, USA). The linearized plasmid was used as substrate in the endonuclease assays to evaluate I-OmiII cleavage activity. The cleavage reaction mix contained: 15 µl linearized pCR4/1574 plasmid (18 µg/ml), 5 µl Invitrogen Buffer React #3 supplemented with 1 mM DTT, 5 µl I-OmiII (53 µg/ml) and 25 µl H_2O. Cleavage reactions were incubated at 37 ^oC and 10 µl aliquots were taken as described above at 0, 30 and 60 minutes. The reactions were stopped with the addition of 2 µl of 200 mM EDTA (pH 8.0) and 1 µl of proteinase K (1 mg/ml) to each 10 µl aliquot, with subsequent incubation for 30 min at 37°C.

2.13. Determination of the optimum temperature for I-OmiII:

To test the effect of temperature on I-OmiII cleavage activity, the pCR4/1574 plasmid, which contains the I-OmiII recognition site, was linearized with *Nco*I. The cleavage assay performed in a 10 µl reaction mixture containing 1 µl linearized

pCR4/1574 (25 µg/ml), 1 µl Invitrogen Buffer React 3 supplemented with 1 mM DTT, 1 µl I-OmiII (53 µg/ml) and 7 µl H_2O. The cleavage reactions were incubated at a temperate range from 10 to 90 ^oC in 10 degree intervals for 2 hours. The reactions were terminated as above with the addition of 2 µl of 200 mM EDTA (pH 8.0) and 1 µl of proteinase K (1 mg/mL).

2. 14. I-OmiII cleavage site mapping:

The pCR4/1574 plasmid was digested with I-OmiII, as described above, and the now linearized plasmid was resolved on a 1 % agarose gel and the band was recovered from the gel with the Wizard® SV Gel and PCR Clean-Up system (Promega). The linearized plasmid was treated with T4 DNA polymerase (Invitrogen), under conditions that generate blunted ends, in a reaction mixture consisting of the following reagents: 40 µl linearized plasmid (25 µg/ml), 2 µl T4 DNA polymerase (5u/µl), 20 µl 5X T4 DNA polymerase buffer, 20 µl dNTP mixture (0.5 mM) and the total volume was adjusted to 100 µl with sterile distilled water. The reaction mixture was incubated at room temperature (~24 ^oC) for 20 minutes and then placed on ice for 5 minutes. The reaction was then terminated by heating at 70 ^oC for 10 minutes. The DNA was purified with the Wizard® SV Gel and PCR Clean-Up system (Promega) and about 0.25 µg of linearized plasmid DNA (in 20 µl) was treated with 2 µl of T4 DNA Ligase (1U/µl) in the presence of 10 µl 5X Ligase buffer in a total volume of 40 µl. The ligation reaction was then incubated at room temperature for 2 hours to generate the plasmid pCR4/1574-T4. The ligation reaction was diluted 5-fold and 10 µl of this dilution was used to transform *E. coli* (DH5α) using the TOPO TA Cloning® kit (Invitrogen). The pCR4/1574-T4 plasmid

was purified from transformed overnight cultures with the Wizard Plus Minipreps DNA purification kit (Promega) and sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's instructions. The sequencing products were denatured and resolved on a 3130 genetic analyzer (Applied Biosystems). The chromatogram was compared with the sequencing reaction for pCR4/1574 using the same primers for both types of constructs. Nucleotides missing in the pCR4/1574-T4 sequence when compared to the original pCR4/1574 derived sequence define the nucleotides removed by T4 DNA polymerase due to the presence of 3' overhangs in the I-OmiII digested pCR4/1574.

2.15. Overexpression and purification of I-OmiI :

Please see Appendix 9.2. for the overexpression, purification and endonucleases assays for the I-OmiI protein.

CHAPTER: 3

CHARACTERIZATION OF THE O.ul-mS952 INTRON: A POTENTIAL MOLECULAR MARKER TO DISTINGUISH BETWEEN *Ophiostoma ulmi* AND *Ophiostoma novo-ulmi* subspecies *americana*.*

3.1. ABSTRACT:

The full length *rns* gene has been characterized for *Ophiostoma novo-ulmi* subspecies *americana*. The gene was also characterized for *Ophiostoma ulmi* and a group II intron was noted in the *rns* gene of *O. ulmi*. The insertion in the *rns* gene is at position S952 and it is a group IIB1 intron that encodes a double motif LHEase from an ORF located within a loop of domain III. Secondary structure models for the *rns* RNA of *O. novo-ulmi* subsp. *americana* and *O. ulmi* were generated to place the intron within the context of the ribosomal RNA. The *in vivo* splicing of the O.ul-mS952 group II intron was confirmed with RT-PCR. A survey of 182 strains of DED causing agents showed that the mS952 intron was absent in what is considered to be the more aggressive species *O. novo-ulmi* but present in strains of the less aggressive *O. ulmi*. This observation suggests that the O.ul-mS952 intron can be used as a PCR-based molecular marker to discriminate between *O. ulmi* and *O. novo-ulmi* subsp. *americana.*

* A version of this chapter was published in World Academy of Science, Engineering and Technology (WASET) conference proceedings. Hafez M and Hausner G. **2011b**. 59:1767-1775.

3.2. INTRODUCTION:

Dutch Elm Disease (DED) is a fungal disease that has devastated many urban forests that contain *Ulmus americanus* L. (American elm) and related species. The causative agents of DED are *Ophiostoma ulmi* (Buism.) Nannf., and subspecies of *Ophiostoma novo-ulmi* Brasier. These fungi are filamentous ascomycetous microfungi belonging to the Order Ophiostomatales, Family Ophiostomataceae. These fungi are transmitted with the help of bark beetles (Family Curculionidae, Subfamily Scolytinae). DED is a vascular wilt diseases caused by the DED fungi blocking the conductive tissue of the elm tree, thus preventing the flow of nutrients and water, thereby killing the tree. Historically there have been two epidemics of DED, the first caused by *O. ulmi* and the current pandemic of this disease is caused by *O. novo-ulmi* (Brasier, 1979). *Ophiostoma novo-ulmi* based on morphological, physiological and molecular (in both nuclear and mitochondrial genomes) differences, has been segregated into O. *novo-ulmi* subsp. *novo-ulmi* (also known as the Eurasian race, EAN) and O. *novo-ulmi* subsp. *americana* (also known as the North American race, NAN) (Brasier and Kirk, 2001; Hintz et al., 1993; Konrad et al., 2002).

Several methods have been proposed to differentiate among isolates of these three biological forms (*O. ulmi, O. novo-ulmi* subsp. *novo-ulmi* and subsp. *americana*) such as colony morphology (Gibbs and Brasier, 1973), pathogenicity (Brasier, 1977), optimal growth temperature (Brasier et al., 1981), soluble protein patterns (Jeng et al., 1988; Jeng and Brasier, 1994), and fertility/genetic barriers (Gibbs and Brasier 1973; Brasier, 1977

& 1979; Brasier and Kirk, 2001). Differentiating among the various DED causing agents by the above-mentioned criteria is time consuming. Due to advances in molecular biology techniques, large numbers of highly informative DNA markers have been developed for the detection of genetic polymorphism that allows for species and/or strain designations (Hintz et al., 1989).

So far several molecular markers have been developed for recognizing the DED causing agents, such as PCR based markers based on the nuclear cerato-ulmi (*cu*) and the colony type (*col1*) genes (Konrad et al., 2002); however to differentiate between the two subspecies of *O. novo-ulmi* the PCR products have to be digested with restriction enzymes. Random amplified polymorphic DNA (RAPD) represents another PCR based technique that has been used to differentiate strains of *O. ulmi* and subspecies of O. *novo-ulmi* (Hoegger et al., 1996). Restriction fragment length polymorphism (RFLP) of nuclear and mitochondrial DNAs has also been applied in analyzing populations of *O. novo-ulmi* and related taxa (Bates et al., 1993; Hintz et al., 1991).

Previously it has been shown that the size of the mtDNA might be useful for differentiating among subspecies of *O. novo-ulmi*. In general sizes of the mtDNAs for isolates of *O. novo-ulmi* subsp. *novo-ulmi* were shown to be significantly larger than those for isolates of the more aggressive *O. novo-ulmi* subsp. *americana* (Hintz et al., 1991; Bates et al., 1993; Charter et al., 1996). In general it is thought that size variation among mtDNAs is due the sizes of intergenic spacers and introns and intron encoded open reading frames (reviewed in Hausner, 2003; Gibb and Hausner, 2005).

Mitochondrial introns usually belong to either the group I or group II category of self-splicing introns and they tend to encode proteins that can assist in intron mobility or intron splicing or in some cases the encoded protein can serve both functions (Belfort, 2003).

The objective of this study was to characterize an intron located within the mtDNA *rns* gene and to evaluate if this intron (referred to as O.ul-mS952) could be used as a molecular marker that allows for the differentiation between strains of *O. ulmi* and *O. novo-ulmi* subsp. *americana*. A PCR-based method is presented that allows for the detection of the presence or absence of the O.ul-mS952 intron using a combination of exon-specific and intron-specific primers.

3. 3. METHODS OVERVIEW:

The fungal strains used in this study, their sources and geographical origins are listed in Table 9.2. The nucleic acids were extracted as described previously in chapter 2. The *rns* gene was amplified via PCR from 182 strains of *Ophiostoma* with the primers mtsr-1 and mtsr-2 (see appendix 9.3. for a complete list of the primers used to amplify and sequence the *rns* gene of *O. ulmi* and *O. novo-ulmi*) utilizing the Invitrogen PCR system. The 182 strains analyzed include *O. ulmi* (11 strains), *O. novo-ulmi* subsp. *novo-ulmi* (1 strain) and *O. novo-ulmi* subsp. *americana* (170 strains) collected from different geographic regions (Canada, USA, England, Netherlands and Iran). Species designations were based on morphological and molecular data (Gibb and Hausner, 2005; Sethuraman

et al., 2008). A detailed PCR analysis for the presence of the O.ul-mS952 intron was performed on whole cell DNAs (i.e. nuclear plus mtDNA) from *O. ulmi* strain DAOM 171046 and *O. novo-ulmi* subsp. *americana* strain DED 02-10 using combinations of exon-specific primers (mtsr-1, mS952-F and mtsr-2) and an intron-specific primer (mS952DV-R). The primers RNS-F and RNS-R were designed based on sequences that flank the *rns* gene and these were used to amplify the complete sequence of the *rns* gene for *O. ulmi* DAOM 171046 and *O. novo-ulmi* subsp. *americana* DED 02-10. The RNS-F and RNS-R primer set allowed for the sequence characterization of the entire *rns* gene.

Ophiostoma ulmi DAOM 171046, *O. novo-ulmi* subsp. *americana* DED 02-10 and *O. novo-ulmi* subsp. *novo-ulmi* IMI 343.101 were selected for sequence characterization. The *rns* derived PCR products were converted into sequencing templates by purifying them with the Wizard® SV Gel and PCR Clean-Up system (Promega). The double-stranded DNA fragments were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following the manufacturer's instructions. The sequencing products were denatured and resolved on a 3130 genetic analyzer (Applied Biosystems) as previously described in chapter 2.

The secondary structures for the complete small subunit RNA for *O. ulmi* (DAOM 171046) and *O. novo-ulmi* subsp. *americana* (DED 02-10) were generated by comparative sequence analysis with the SSU rRNA secondary structures of *E. coli* and *Aspergillus nidulans* (GenBank accession J01393). The final secondary structures for the intron and the *rns* rRNA were drawn with CorelDraw™.

Table 3.1. A List of strains of DED fungi used in the current study and their isolation area.

Table 3.1.

No.	Strain Number	Species	Location
1	DAOM 171044	*Ophiostoma ulmi*	Sault St.Marie, Ontario, Canada
2	DAOM 171045	*O. ulmi*	Mal-Hal, Quebec, Canada
3	DAOM 171046	*O. ulmi*	Westmount, Quebec, Canada
4	DAOM 171048	*O. ulmi*	Clear Lake, Iowa, USA
5	DAOM 171051	*O. ulmi*	Tennessee, USA
6	DAOM 171061	*O. ulmi*	Devon, England
7	DAOM 171063	*O. ulmi*	Denbighshire, England
8	DAOM 171064	*O. ulmi*	Thame, Oxfordshire, England
9	DAOM 171068	*O. ulmi*	Bennekom, Netherlands
10	DAOM 171069	*O. ulmi*	Amsterdam, Netherlands
11	DAOM 171071	*O. ulmi*	Baarn, Netherlands
12	IMI 343.101	*O. novo-ulmi* subsp. *novo-ulmi*	Zabok, Front Croatia.
13	DED 02-1	*O. novo-ulmi* subsp. *americana*	Manitou, Manitoba, Canada
14	DED 02-3	*O. novo-ulmi* subsp. *americana*	Manitou, Manitoba, Canada,
15	DED 02-7	*O. novo-ulmi* subsp. *americana*	Manitou, Manitoba, Canada
16	DED 02-8	*O. novo-ulmi* subsp. *americana*	Carman, Manitoba, Canada
17	DED 02-9	*O. novo-ulmi* subsp. *americana*	Steinbach, Manitoba, Canada
18	DED 02-10	*O. novo-ulmi* subsp. *americana*	Gimli, Manitoba, Canada
19	DED 02-11	*O. novo-ulmi* subsp. *americana*	Carman, Manitoba, Canada
20	DED 02-12	*O. novo-ulmi* subsp. *americana*	Selkirk, Manitoba, Canada
21	DED 02-13	*O. novo-ulmi* subsp. *americana*	S Carman, Manitoba, Canada
22	DED 02-14	*O. novo-ulmi* subsp. *americana*	Steinbach, Manitoba, Canada
23	DED 02-15	*O. novo-ulmi* subsp. *americana*	Morden, Manitoba, Canada
24	DED 02-17	*O. novo-ulmi* subsp. *americana*	Carman, Manitoba, Canada
25	DED 02-18	*O. novo-ulmi* subsp. *americana*	Arborg, Manitoba, Canada
26	DED 02-19	*O. novo-ulmi* subsp. *americana*	Gimli, Manitoba, Canada
27	DED 02-20	*O. novo-ulmi* subsp. *americana*	Trehern, Manitoba, Canada
28	DED 02-21	*O. novo-ulmi* subsp. *americana*	Morden, Manitoba, Canada
29	DED 02-22	*O. novo-ulmi* subsp. *americana*	Trehern, Manitoba, Canada
30	DED 02-23	*O. novo-ulmi* subsp. *americana*	Morden, Manitoba, Canada
31	DED 02-24	*O. novo-ulmi* subsp. *americana*	Morden Research Centre, Manitoba, Canada
32	DED 02-26	*O. novo-ulmi* subsp. *americana*	Carman, Manitoba, Canada
33	DED 02-27	*O. novo-ulmi* subsp. *americana*	Altona, Manitoba, Canada
34	DED 02-28	*O. novo-ulmi* subsp. *americana*	Steinbach, Manitoba, Canada
35	DED 02-29	*O. novo-ulmi* subsp. *americana*	Carman, Manitoba, Canada
36	DED 02-32	*O. novo-ulmi* subsp. *americana*	Morden, Manitoba, Canada
37	DED 02-33	*O. novo-ulmi* subsp. *americana*	Trehern, Manitoba, Canada
38	DED 02-34	*O. novo-ulmi* subsp. *americana*	Steinbach, Manitoba, Canada
39	DED 02-35	*O. novo-ulmi* subsp. *americana*	Steinbach, Manitoba, Canada
40	DED 02-36	*O. novo-ulmi* subsp. *americana*	Trehern, Manitoba, Canada
41	DED 02-37	*O. novo-ulmi* subsp. *americana*	Teulon, Manitoba, Canada

Table 3.1. CONTINUED

42	DED 02-38	*O. novo-ulmi* subsp. *americana*	Arborg, Manitoba, Canada
43	DED 02-40	*O. novo-ulmi* subsp. *americana*	Carman, Manitoba, Canada
44	DED 02-41	*O. novo-ulmi* subsp. *americana*	Gimli, Manitoba, Canada
45	DED 02-42	*O. novo-ulmi* subsp. *americana*	Teulon, Manitoba, Canada
46	DED 02-43	*O. novo-ulmi* subsp. *americana*	Morris, Manitoba, Canada
47	DED 02-45	*O. novo-ulmi* subsp. *americana*	Altona, Manitoba, Canada
48	DED 02-46	*O. novo-ulmi* subsp. *americana*	Stonewall, Manitoba, Canada
49	DED 02-47	*O. novo-ulmi* subsp. *americana*	Morden, Manitoba, Canada
50	DED 02-48	*O. novo-ulmi* subsp. *americana*	Steinbach, Manitoba, Canada
51	DED 02-50	*O. novo-ulmi* subsp. *americana*	Shilkirk, Manitoba, Canada
52	DED 02-51	*O. novo-ulmi* subsp. *americana*	Altona, Manitoba, Canada
53	DED 02-52	*O. novo-ulmi* subsp. *americana*	PLAP, Manitoba, Canada
54	DED 02-53	*O. novo-ulmi* subsp. *americana*	Stonewall, Manitoba, Canada
55	DED 02-54	*O. novo-ulmi* subsp. *americana*	Portage la Prairie, Manitoba, Canada
56	DED 02-56	*O. novo-ulmi* subsp. *americana*	Stonewall, Manitoba, Canada
57	DED 02-57	*O. novo-ulmi* subsp. *americana*	Portage la Prairie, Manitoba, Canada
58	DED 02-61	*O. novo-ulmi* subsp. *americana*	St. Celements, Manitoba, Canada
59	DED 02-62	*O. novo-ulmi* subsp. *americana*	Gretna, Manitoba, Canada
60	DED 02-63	*O. novo-ulmi* subsp. *americana*	Gretna, Manitoba, Canada
61	DED 02-64	*O. novo-ulmi* subsp. *americana*	Mountain Avenue, Manitoba, Canada
62	DED 02-66	*O. novo-ulmi* subsp. *americana*	Winkler, Manitoba, Canada
63	DED 02-69	*O. novo-ulmi* subsp. *americana*	Winkler, Manitoba, Canada
64	DED 02-71	*O. novo-ulmi* subsp. *americana*	La Prairie Island, Manitoba, Canada
65	DED 02-72	*O. novo-ulmi* subsp. *americana*	Winkler, Manitoba, Canada
66	DED 02-73	*O. novo-ulmi* subsp. *americana*	Winkler, Manitoba, Canada
67	DED 03-86	*O. novo-ulmi* subsp. *americana*	PTH 305, Manitoba, Canada
68	DED 03-87	*O. novo-ulmi* subsp. *americana*	PTH 305, Manitoba, Canada
79	DED 03-88	*O. novo-ulmi* subsp. *americana*	Cartwright, Manitoba, Canada
70	DED 03-89	*O. novo-ulmi* subsp. *americana*	La Riviere, Manitoba, Canada
71	DED 03-90	*O. novo-ulmi* subsp. *americana*	Snowflake, Manitoba, Canada
72	DED 03-91	*O. novo-ulmi* subsp. *americana*	Baldur, Manitoba, Canada
73	DED 03-93	*O. novo-ulmi* subsp. *americana*	Altamont, Manitoba, Canada
74	DED 03-94	*O. novo-ulmi* subsp. *americana*	Manitoba, Canada
75	DED 03-95	*O. novo-ulmi* subsp. *americana*	Wawansa, Manitoba, Canada
76	DED 03-96	*O. novo-ulmi* subsp. *americana*	Baldur, Manitoba, Canada
77	DED 03-97	*O. novo-ulmi* subsp. *americana*	Snowflake, Manitoba, Canada
78	DED 03-98	*O. novo-ulmi* subsp. *americana*	La Rivierie, Manitoba, Canada
79	DED 03-99	*O. novo-ulmi* subsp. *americana*	Miami, Manitoba, Canada
80	DED 03-100	*O. novo-ulmi* subsp. *americana*	Plum Coulee, Manitoba, Canada
81	DED 03-101	*O. novo-ulmi* subsp. *americana*	Cartwright, Manitoba, Canada
82	DED 03-102	*O. novo-ulmi* subsp. *americana*	Swan lake, Manitoba, Canada
83	DED 03-103	*O. novo-ulmi* subsp. *americana*	Crystal City, Manitoba, Canada
84	DED 03-104	*O. novo-ulmi* subsp. *americana*	Crystal City, Manitoba, Canada
85	DED 03-105	*O. novo-ulmi* subsp. *americana*	Roland, Manitoba, Canada

Table 3.1. CONTINUED

86	DED 03-106	*O. novo-ulmi* subsp. *americana*	Swan Lake, Manitoba, Canada
87	DED 03-107	*O. novo-ulmi* subsp. *americana*	Holland, Manitoba, Canada
88	DED 03-108	*O. novo-ulmi* subsp. *americana*	Holland, Manitoba, Canada
89	DED 03-109	*O. novo-ulmi* subsp. *americana*	Lowe Farm, Manitoba, Canada
90	DED 03-110	*O. novo-ulmi* subsp. *americana*	Plum Coulee, Manitoba, Canada
91	DED 03-111	*O. novo-ulmi* subsp. *americana*	Altamont, Manitoba, Canada
92	DED 03-112	*O. novo-ulmi* subsp. *americana*	Reinland Highway, Manitoba, Canada
93	DED 03-113	*O. novo-ulmi* subsp. *americana*	Miami, Manitoba, Canada
94	DED 03-114	*O. novo-ulmi* subsp. *americana*	Lowe Farm, Manitoba, Canada
95	DED 03-115	*O. novo-ulmi* subsp. *americana*	Roland, Manitoba, Canada
96	DED 03-116	*O. novo-ulmi* subsp. *americana*	Wawansa, Manitoba, Canada
97	DED 03-117	*O. novo-ulmi* subsp. *americana*	St.Claude, Manitoba, Canada
98	DED 03-118	*O. novo-ulmi* subsp. *americana*	Taymouth, York Co, New Brunswick, Canada
99	DED 03-119	*O. novo-ulmi* subsp. *americana*	Kentville, Kings Co., Nova Scotia, Canada
100	DED 03-120	*O. novo-ulmi* subsp. *americana*	Salmon River, Nova Scotia, Canada
101	DED 04-21	*O. novo-ulmi* subsp. *americana*	Moose Jaw, Saskatchewan, Canada
102	DED 04-76	*O. novo-ulmi* subsp. *americana*	Fort Qu'Appelle, Saskatchewan, Canada
103	DED 04-81	*O. novo-ulmi* subsp. *americana*	Gainsborough, Saskatchewan, Canada
104	DED 04-82	*O. novo-ulmi* subsp. *americana*	Regina, Saskatchewan, Canada
105	DED 04-95	*O. novo-ulmi* subsp. *americana*	Estevan, Saskatchewan, Canada
106	DED 04-102	*O. novo-ulmi* subsp. *americana*	Katepwa, Saskatchewan, Canada
107	DED 04-128	*O. novo-ulmi* subsp. *americana*	Regina Beach, Saskatchewan, Canada
108	DED 04-137	*O. novo-ulmi* subsp. *americana*	Pasqua Lake, Saskatchewan, Canada
109	DED 04-138	*O. novo-ulmi* subsp. *americana*	Sun Valley, Saskatchewan, Canada
110	DED 04-141	*O. novo-ulmi* subsp. *americana*	Indian Head, Saskatchewan, Canada
111	DED 04-150	*O. novo-ulmi* subsp. *americana*	Codette, Saskatchewan, Canada
112	DED 04-154	*O. novo-ulmi* subsp. *americana*	Tisdale, Saskatchewan, Canada
113	DED 04-167	*O. novo-ulmi* subsp. *americana*	Alida, Saskatchewan, Canada
114	DED 04-202	*O. novo-ulmi* subsp. *americana*	Buffalo Pound Lake, Saskatchewan, Canada
115	DED 04-257	*O. novo-ulmi* subsp. *americana*	Moosomin, Saskatchewan, Canada
116	DED 04-258	*O. novo-ulmi* subsp. *americana*	Carnduff, Saskatchewan, Canada
117	DED 04-298	*O. novo-ulmi* subsp. *americana*	Traux, Saskatchewan, Canada
118	DED 04-337	*O. novo-ulmi* subsp. *americana*	Buffalo, Saskatchewan, Canada
119	DED 05-1	*O. novo-ulmi* subsp. *americana*	Cavan, Manitoba, Canada
120	DED 05-2	*O. novo-ulmi* subsp. *americana*	The Pas, Manitoba, Canada
121	DED 05-3	*O. novo-ulmi* subsp. *americana*	The Pas, Manitoba, Canada
122	DED 05-4	*O. novo-ulmi* subsp. *americana*	The Pas, Manitoba, Canada
123	DED 05-5	*O. novo-ulmi* subsp. *americana*	Red Deer River, Manitoba, Canada
124	DED 05-6	*O. novo-ulmi* subsp. *americana*	Armit River, Manitoba, Canada
125	DED 05-9	*O. novo-ulmi* subsp. *americana*	Riverton, Manitoba, Canada
126	DED 05-10	*O. novo-ulmi* subsp. *americana*	Teulon, Manitoba, Canada
127	DED 05-11	*O. novo-ulmi* subsp. *americana*	Gimli Park, Manitoba, Canada
128	DED 05-13	*O. novo-ulmi* subsp. *americana*	Arborg, Manitoba, Canada

Table 3.1. CONTINUED

129	DED 05-14	*O. novo-ulmi* subsp. *americana*	Minnedosa, Manitoba, Canada
130	DED 05-15	*O. novo-ulmi* subsp. *americana*	Gladstone, Manitoba, Canada
131	DED 05-16	*O. novo-ulmi* subsp. *americana*	Dauphin, Manitoba, Canada
132	DED 05-17	*O. novo-ulmi* subsp. *americana*	St. Rose, Manitoba, Canada
133	DED 05-18	*O. novo-ulmi* subsp. *americana*	Birtle, Manitoba, Canada
134	DED 05-19	*O. novo-ulmi* subsp. *americana*	Rivers, Manitoba, Canada
135	DED 05-20	*O. novo-ulmi* subsp. *americana*	Rivers, Manitoba, Canada
136	DED 05-21	*O. novo-ulmi* subsp. *americana*	Neepawa, Manitoba, Canada
137	DED 05-60	*O. novo-ulmi* subsp. *americana*	Carduff, Saskatchewan, Canada
138	DED 05-61	*O. novo-ulmi* subsp. *americana*	Carduff, Saskatchewan, Canada
139	DED 05-62	*O. novo-ulmi* subsp. *americana*	Carduff, Saskatchewan, Canada
140	DED 05-63	*O. novo-ulmi* subsp. *americana*	Carduff, Saskatchewan, Canada
141	DED 05-64	*O. novo-ulmi* subsp. *americana*	Carduff, Saskatchewan, Canada
142	DED 05-85	*O. novo-ulmi* subsp. *americana*	Wawata, Saskatchewan, Canada
143	DED 05-153	*O. novo-ulmi* subsp. *americana*	Avonlea, Saskatchewan, Canada
144	DED 05-164	*O. novo-ulmi* subsp. *americana*	Tisdale, Saskatchewan, Canada
145	DED 05-165	*O. novo-ulmi* subsp. *americana*	Tisdale, Saskatchewan, Canada
146	DED 05-195	*O. novo-ulmi* subsp. *americana*	Nipawin, Saskatchewan, Canada
147	DED 05-216	*O. novo-ulmi* subsp. *americana*	Garrick, Saskatchewan, Canada
148	DED 06-1	*O. novo-ulmi* subsp. *americana*	Winnipeg, Manitoba, Canada
149	DED 06-2	*O. novo-ulmi* subsp. *americana*	Winnipeg, Manitoba, Canada
150	DED 06-3	*O. novo-ulmi* subsp. *americana*	Winnipeg, Manitoba, Canada
151	DED 06-4	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
152	DED 06-5	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
153	DED 06-6	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
154	DED 06-7	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
155	DED 06-8	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
156	DED 06-9	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
157	DED 06-10	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
158	DED 06-11	*O. novo-ulmi* subsp. *americana*	Thunderbay, Ontario, Canada
159	DAOM 171033	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
160	DAOM 171034	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
161	DAOM 171035	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
162	DAOM 171036	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
163	DAOM 171037	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
164	DAOM 171038	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
165	DAOM 171039	*O. novo-ulmi subsp. americana*	Toronto, Ontario, Canada
166	DAOM 171040	*O. novo-ulmi subsp. americana*	Sault St.Marie, Ontario, Canada
167	DAOM 171041	*O. novo-ulmi subsp. americana*	Sault St.Marie, Ontario, Canada
168	DAOM 171042	*O. novo-ulmi subsp. americana*	Sault St.Marie, Ontario, Canada
169	DAOM 171047	*O. novo-ulmi subsp. americana*	Ames, Iowa, USA
170	DAOM 171049	*O. novo-ulmi subsp. americana*	Iowa County, Iowa, USA
171	DAOM 171053	*O. novo-ulmi subsp. americana*	Ames, Iowa, USA

Table 3.1. CONTINUED

172	DAOM 171054	*O. novo-ulmi subsp. americana*	North Hampton, Massachusetts, USA
173	DAOM 171056	*O. novo-ulmi subsp. americana*	Blackburg, Virginia, USA
174	DAOM 171058	*O. novo-ulmi subsp. americana*	Tewkesbury, Glocester, England
175	DAOM 171059	*O. novo-ulmi subsp. americana*	Herfordshire, England
176	DAOM 171060	*O. novo-ulmi subsp. americana*	Basildon, Essex, England
177	DAOM 171066	*O. novo-ulmi subsp. americana*	Astra near Caspian Sea, Iran
178	DAOM 171067	*O. novo-ulmi subsp. americana*	Astra near Caspian Sea, Iran
179	DAOM 171070	*O. novo-ulmi subsp. americana*	Sloten, Friesland, Netherlands
180	DAOM 177178	*O. novo-ulmi subsp. americana*	Sault St. Marie, Ontario, Canada
181	DAOM 177179	*O. novo-ulmi subsp. americana*	Madison, Wisconsin, USA
182	DAOM 194898	*O. novo-ulmi subsp. americana*	Saskatoon, Saskatchewan, Canada

DAOM, Plant Research Institute, Department of Agriculture, Mycology, Ottawa, Canada.
IMI, International Mycological Institute, Egham, UK.
DED, isolation number for Dutch Elm Disease strains, culture inoculated as part of **WIM(M)**, University of Manitoba, Winnipeg, MB, Canada (courtesy of J. Reid).

3.4. RESULTS:

3.4.1. The *rns* gene of *O. ulmi* and *O. novo-ulmi* subspecies *americana*:

Figure 3.1.[a] represents a schematic overview of the *rns* gene, including the various primer binding sites that were developed for amplifying the gene and the mS952 intron. For screening the 182 strains of DED fungi (see Table 3.1.), the *rns* gene was amplified using the primer mtsr-1 and mtsr-2 (for complete list of the primers used in this study see Appendix 9.3.) and the PCR products ranged in size between 1.2 kb and 3.0 kb (Figure 3.1.[b]). A 1.2-kb PCR product is expected when no insertions are present within this gene; however a 3.0 kb PCR product is indicative of insertions (introns). The *O. ulmi* strain DAOM 171046 yielded a 3079 bp PCR fragment (GenBank accession HQ292075), while the *O. novo-ulmi* subsp. *americana* DED 02-10 (GenBank accession HQ292074) yielded a 1213 bp PCR amplicon (Figure 3.1.[b]). The *O. novo-ulmi* subsp. *novo-ulmi* (GenBank accession # JF837329) also yielded a 1212 bp PCR amplicon. Note sizes of the PCR products were determined by sequence analysis of representative strains.

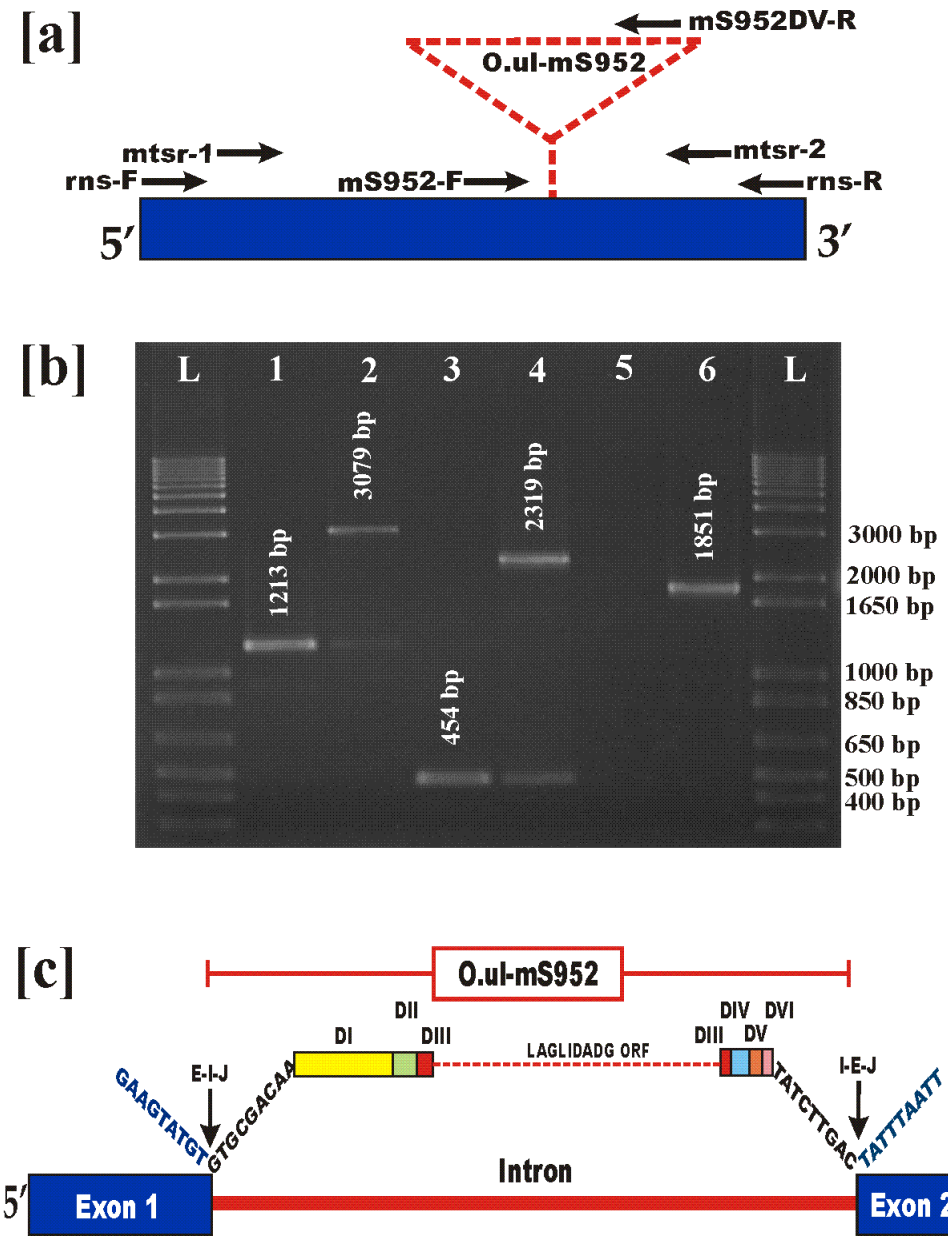
Using internal exon-specific primers (mS952-F and mtsr-2) flanking the O.ul-mS952 intron for *O. ulmi* and *O. novo-ulmi* subsp. *americana* generated 2319 and 454 bp fragments respectively. These results confirmed the mtsr-1/mtsr-2 PCR survey results, which were that the O.ul-mS952 intron is present in all *O. ulmi* strains and is absent in all *O. novo-ulmi* strains screened with mtsr-1 and mtsr-2 primers. A combination of an exon-specific primer (mS952-F) and an intron-specific primer (mS952DV-R) generated a 1851 bp PCR fragment for *O. ulmi* strains, while no amplicons were generated for *O. novo-*

ulmi strains (Figure 3.1.[b]). Based on comparing *rns* sequences that lacked insertions with those that do, an intron was found to be inserted at the position mS952 with respect to the SSU-rRNA gene of *E. coli*. The intron-exon junctions were first determined by comparative sequence analysis of *O. ulmi* and *O. novo-ulmi* mtsr-1 and mtsr-2 *rns* gene sequences (Figure 3.1.[c]).

Using internal exon-specific primers (mS952-F and mtsr-2) flanking the O.ul-mS952 intron for *O. ulmi* and *O. novo-ulmi* subsp. *americana* generated 2319 and 454 bp fragments respectively. These results confirmed the mtsr-1/mtsr-2 PCR survey results which were that the O.ul-mS952 intron is present in all *O. ulmi* strains and is absent in all *O. novo-ulmi* strains. A combination of exon-specific primer (mS952-F) and intron-specific primer (mS952DV-R) generate a 1851 bp PCR fragment for *O. ulmi* strains, while no amplicons were generated for *O. novo-ulmi* strain (Figure 3.1.[b]).

Figure 3.1. [a] A schematic representation of the *rns* gene of *O. ulmi* and O. *novo-ulmi* showing the binding sites and directions of the primers used in this study; the dotted triangle represent the position of the O.ul.mS952 intron. **[b]** PCR amplicons of the *rns* gene from *O. novo-ulmi* subsp. *americana* and *O. ulmi* (lane 1 and 2 respectively) using the primer pair mtsr-1 and mtsr-2. The presence of O.ul-mS952 intron was checked using exon-specific primers flanking the intron insertion position, mS952-F and mtsr-2, which generate a 454 bp PCR fragment in the case of O. *novo-ulmi* subsp. *americana* (lane 3) and a 2319 bp PCR fragment for *O. ulmi* (lane 4). There appears to be evidence that there is mtDNA heterogeneity within *O. ulmi* as two bands were observed, imcluding the expected 2.3 kb fragment for the intron containing *rns* allele. There is also evidence for the presence of an intron-less alleles revealed by the generation of the 454 bp PCR product (lane 4). The absence (lane 5) and the presence (lane 6) of the O.ul-mS952 intron in *O. novo-ulmi* subsp. *americana* and *O. ulmi* respectively were confirmed using the exon-specific primer mS952-F combined with the intron-specific primer mS952DV-R. Sizes of the PCR products are indicated (sizes based on sequence determination). The L lane contains the resolved 1 kb plus ladder DNA fragments (Invitrogen). **[c]** A schematic overview of the *rns* gene of *O. ulmi* showing the O.ul-mS952 group II intron which consisting of six domains (DI to DVI). The dotted line in DIII represents the LHEase ORF. The upstream and downstream exon/intron junction (E-I-J) and the intron-exon junction (I-E-J) sequences respectively are indicated.

Figure 3.1


[a]
mS952DV-R
O.ul-mS952
mtsr-1
rns-F
mS952-F
mtsr-2
rns-R
5′
3′
[b]
L
1
2
3
4
5
6
L
1213 bp
3079 bp
454 bp
2319 bp
1851 bp
3000 bp
2000 bp
1650 bp
1000 bp
850 bp
650 bp
500 bp
400 bp
[c]
O.ul-mS952
DI
DII
DIII
DIV
DV
DVI
LAGLIDADG ORF
E-I-J
I-E-J
GAAGTATGTGTGCGACAA
TATCTTGACTATTTAATT
Intron
5′
Exon 1
Exon 2
3′

3.4.2. O.ul171046-mS952 intron:

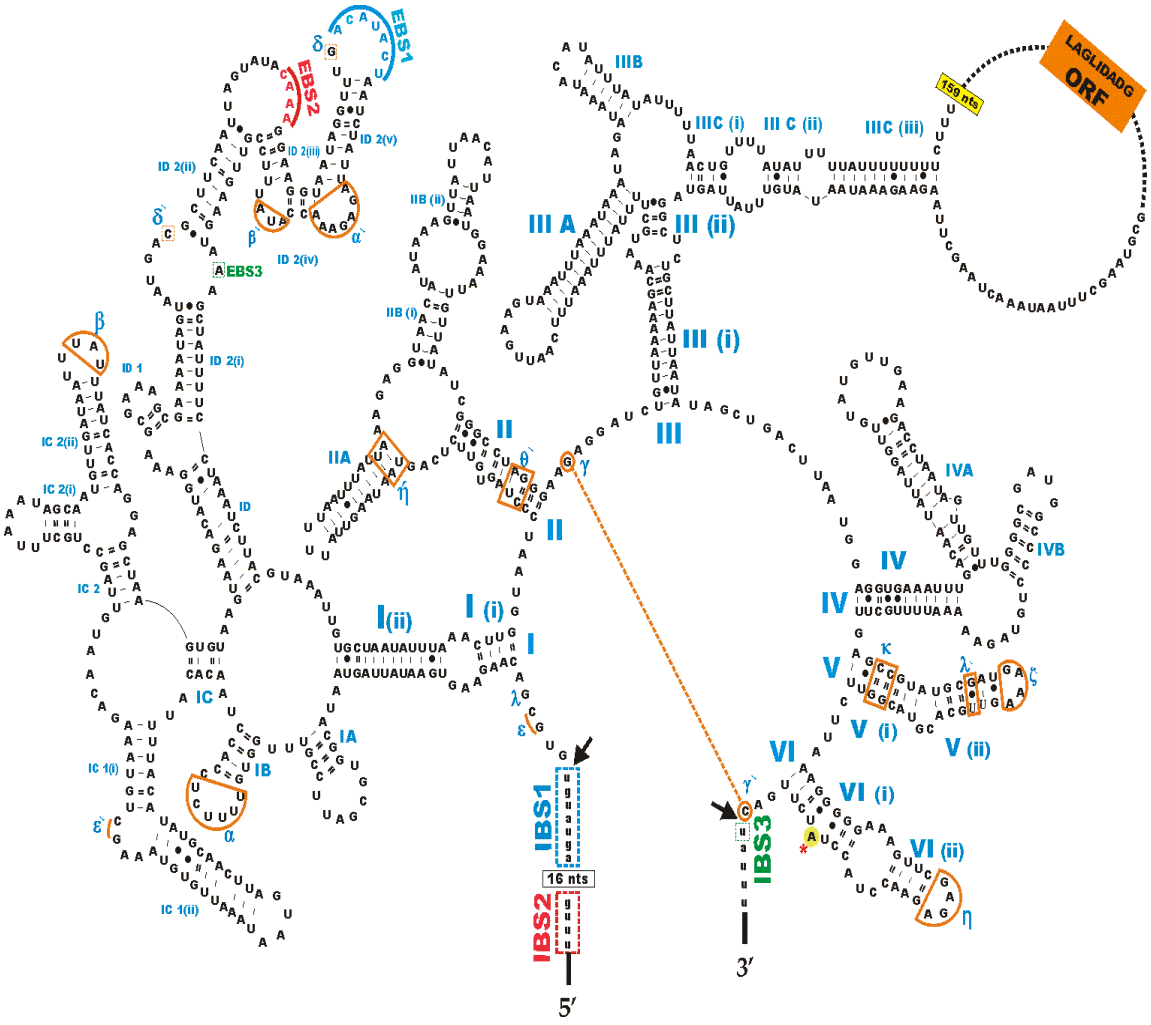
Group II introns have conserved secondary structures at the RNA level, that can be visualized as six stem-loop domains (domains I to VI) emerging from a central wheel (Michel and Ferat, 1995). The O.ul-mS952 intron is a typical group II intron (Figure 3.2.) containing the characteristic features of class B1: such as the exon binding sites (EBS1, EBS2 & EBS3) which are complementary to the intron binding sites (IBS1, IBS2 & IBS3) in the upstream and the downstream exons flanking the intron insertion site; the internal loop of DIC*i* (ε^{-}), the internal loop within DII, the linkers between domains I to VI and the absence of insertions in the 3' strand of DI*i* and DI*ii* (Michel and Ferat, 1995; Toor et al., 2001). LAGLIDADG ORF rather than the RT-ORFs, that typically are associated with group II introns, is found to be encoded within O.ul171046-mS952 intron. Folding the intron RNA indicates that the location of the LAGLIDADG ORF is within domain III (Figure 3.2.).

3.4.3. Mitochondrial *rns* RNA secondary structure model:

Figure 3.3. provides a secondary structure model for the 16S rRNA consisting of four well-defined domains denoted I, II, III & IV (Noller and Woese, 1981; Kochel and Kuntzel, 1981). The O.ul-mS952 intron insertion site was characterized within this secondary structure model and the intron is located in a stem region at the lower half of domain III of the 16S rRNA molecule.

Figure 3.2. The secondary structure of the O.ul171046-mS952 group II B1 intron RNA. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The positions of EBS1, EBS2 and EBS3 are noted. The positions of IBS1, IBS2 and IBS3 in the 5′ and 3′ exons are boxed with dotted lines. Tertiary interactions are indicated by dashed lines and Greek letters (ε, λ, α, β, θ, κ, ζ, and γ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V and VI). The solid black arrowheads indicate the intron-exon junctions (5′ and 3′ splicing sites). The asterisk shows the bulged adenosine nucleotide in domain VI (the branch point). The LAGLIDADG ORF is encoded within DIII.

Figure 3.2.


EBS1
EBS2
EBS3
ID 2(v)
ID 2(iii)
ID 2(ii)
ID 2(iv)
ID 2(i)
ID 1
ID
IC 2(ii)
IC 2(i)
IC 2
IC 1(i)
IC 1(ii)
IC
IA
IB
IIB
IIB (i)
IIB (ii)
IIA
II
I
I (i)
I(ii)
IIIB
III A
III (ii)
III (i)
III
IIIC (i)
III C (ii)
IIIC (iii)
150 nts
LAGLIDADG
ORF
IVA
IVB
IV
V
V (i)
V (ii)
VI
VI (i)
VI (ii)
IBS1
IBS2
16 nts
IBS3
5′
3′

3.4.4. *In vivo* splicing of the O.ul-mS952 intron:

Total RNA was isolated from the *O. ulmi* strain DAOM 171046 and analyzed by RT-PCR, as a means of determining the *in vivo* splicing activity of the O.ul-mS952 intron (Figure 3.4.). Genomic DNA of *O. ulmi* DAOM 171046 was used as a control for RT-PCR experiments. The DNA amplification of the *rns* gene from the genomic DNA was carried out using the mtsr-1 and mtsr-2 primers. PCR yielded a product of 3081 bp representing the *rns* exon sequence (1211 bp) as well as the O.ul-mS952 intron sequence (1868 bp). The mtsr-1/mtsr-2 amplification using cDNA as a template resulted in a 1211 bp fragment; the latter is expected if the O.ul-mS952 has been spliced out. Sequence analysis of the cDNA confirmed that the intron was indeed spliced out; the sequence data also confirmed the predicted exon/intron junction sequences previously based on comparative sequence analysis (Figure 3.1.).

Figure 3.3. A secondary structure model for the *rns* RNA of *O. novo-ulmi* subsp. *americana* showing the four structural domains (I to IV). The circled nucleotides indicates the differences between *O. novo-ulmi* subsp. *americana* and the *O. ulmi rns* RNA. The O.ul-mS952 intron insertion site is also indicated and it is located in a stem region in domain III.

Figure 3.3.

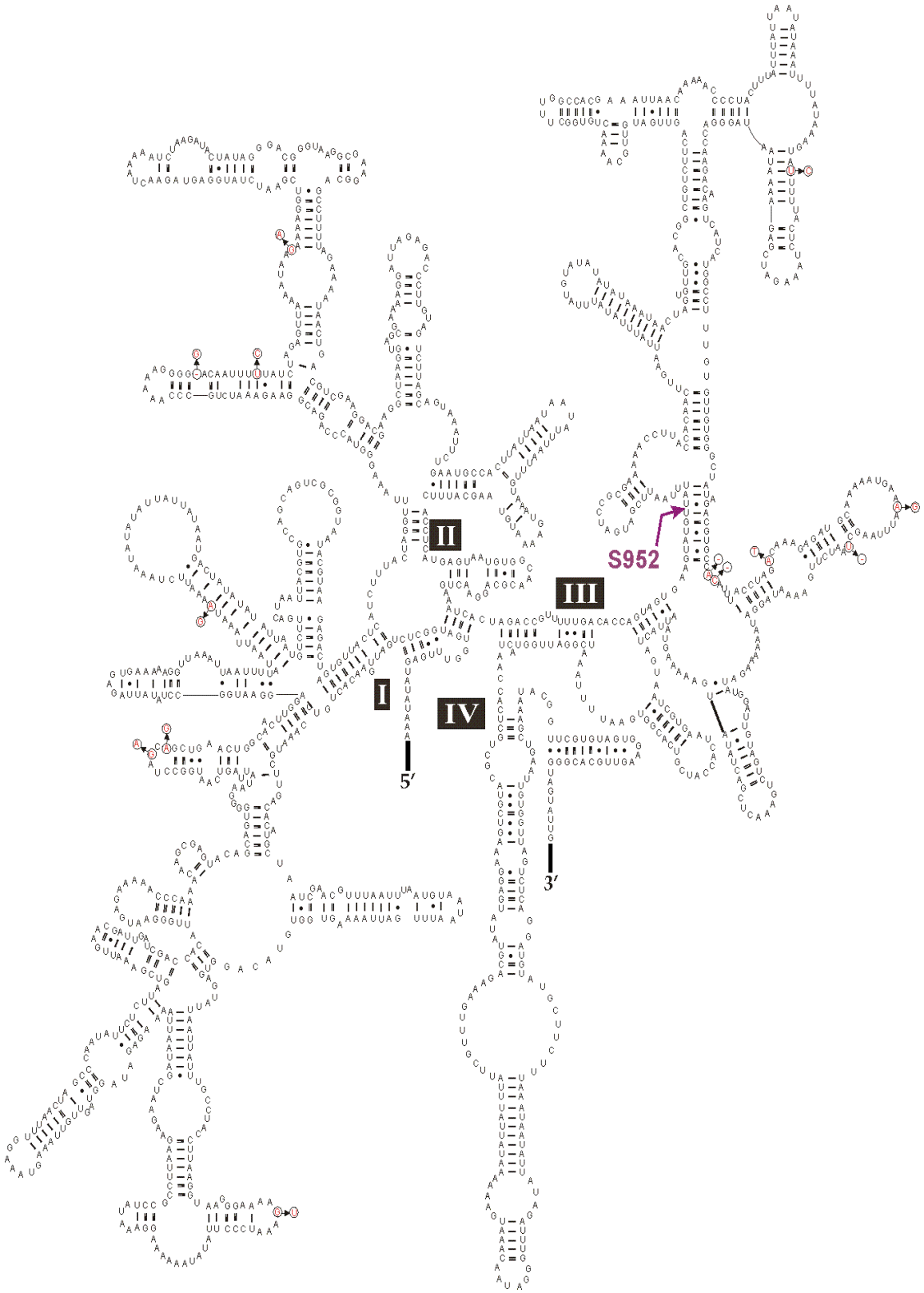
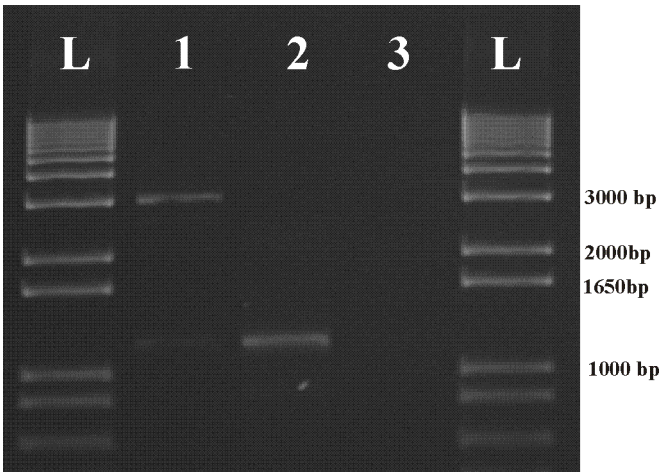


Figure 3.4. RT-PCR analysis of the *rns* transcript for demonstrating the *in vivo* splicing of the O.ul171046-mS952 intron. The *rns* transcripts for the *O. ulmi* DAOM171046 was analyzed by RT-PCR. The standard PCR reaction using genomic DNA as a template is shown in lane 1 and it generated a 3.0 kb fragment. The amplicon length for the cDNA was 1.2 kb (lane 2) indicating that the O.ul-mS952 intron was splicing out. The negative controls for RT-PCR (standard PCR using DNA free whole cell RNA as a template without the RT step) assay yielded no bands (lane 3) showing that the RNA was free of any DNA. The lane marked "L" contains the 1 kb plus ladder (Invitrogen).

Figure 3.4.



3.4.5. The mS952 intron ORF family:

Phylogenetic analysis of mS952 intron ORF sequences showed that the *O. ulmi* mS952 intron ORF sequence is related to sequences found in species that belong to the related asexual genus *Leptographium* (Figure 3.5.). Using the most distantly related sequence from *Cryphonectria parasitica* as the outgroup showed that the *O. ulmi* sequence is positioned at the basal node from which all the *Leptographium* mS952 ORF sequences can be derived. This is expected as *Ophiostoma* is a genus that is closely related to *Grosmannia* (a sexual genus that is defined by forming asexual reproductive structures that can be assigned to *Leptographium*) (Zipfel et al., 2006). Based on the currently available sequences there is no strong evidence for horizontal movement of the mS952 intron within the ophiostomatoid fungi.

3.5. DISCUSSION:

3.5.1. The mS952 intron distinguishes between *O. ulmi* and *O. novo-ulmi* subsp. *americana*:

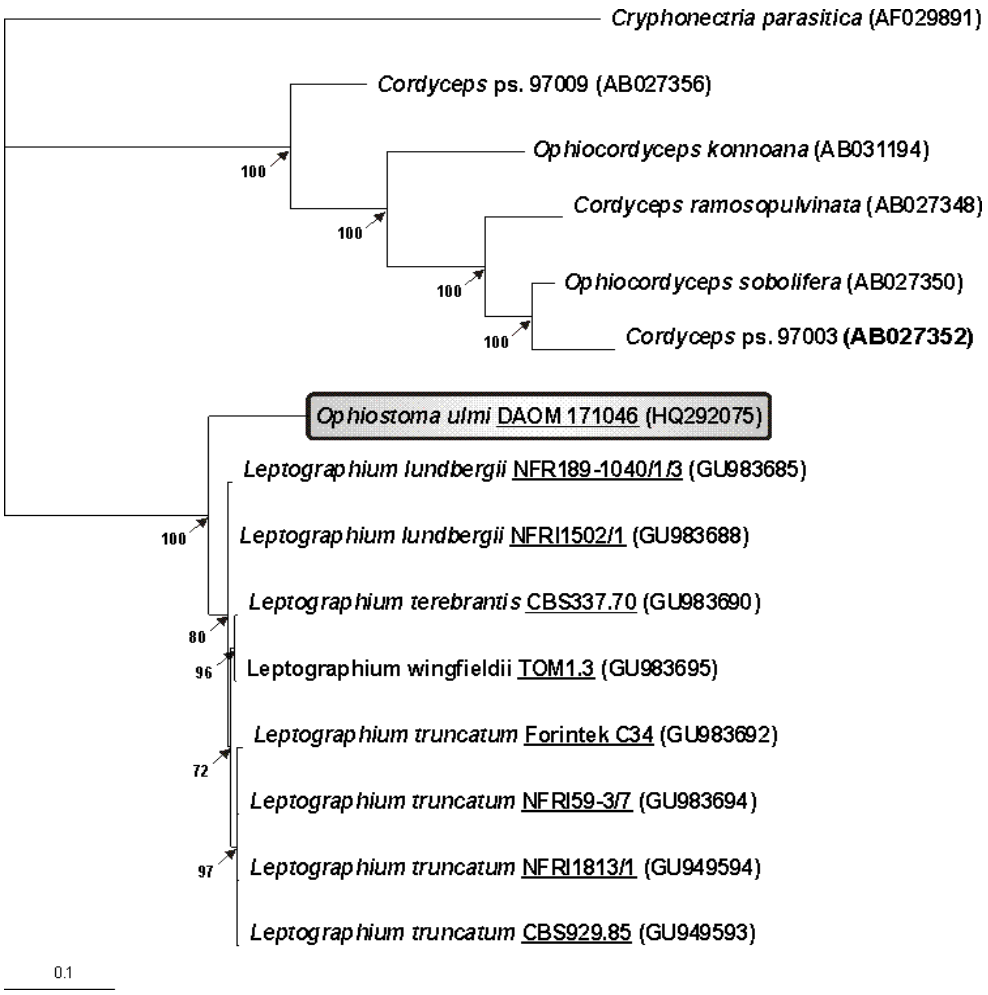
Ophiostoma ulmi is viewed as the first recognized DED causing agent that was introduced into North America in the late 1930s (reviewed in Brasier, 2000). More recently in North America *O. ulmi* appears to have been replaced by *O. novo-ulmi* subsp. *americana,* but in Europe and Asia all three forms of the DED causing agents can still be found (Brasier and Buck, 2001). Although these fungi are viewed as distinct taxa they seem to be able to mate under certain conditions and hybrids have been recovered from various locations in Europe; the latter is of concern as hybridization between these

species/subspecies might be one mechanism whereby more virulent forms of DED causing agents can arise (Konrad et al., 2002; Brasier, 2000). Indeed it is proposed that *O. novo-ulmi* itself might be a product of a hybridization event between *O. ulmi* and another closely related *Ophiostoma* species (Brasier, 1991).

Insertions or deletions of introns within the mtDNA genes are probably the major source of DNA-based polymorphisms between *O. ulmi* and the *O. novo-ulmi* subspecies (Charter et al., 1996; Gibb and Hausner, 2005). Due to sequence conservation of the mtDNA genes, PCR based primers can be readily designed to allow for surveying genes for the presence of optional introns. The goal of this work was to characterize an *rns* intron and to evaluate if this intron is stable enough to be a potential marker that could be useful for distinguishing among the causal agents of DED. Although we assayed for the presence or absence of introns using a PCR approach (see Figure 3.1.); we have to admit that a "minus" result does not unambiguously identify taxa that lack introns. Each mitochondrion can contain many copies of the mt chromosome. PCR prefentially amplifies smaller DNA fragments thus if there is heterogeneity among the mtDNAs rare intron-plus alleles could be missed. Therefore we utilized one primer combination (mS952-F and mS952DV-R) where the forward primer is exon based and the reverse primer is intron based. This combination should capture instances were an intron containing *rns* gene was missed as its PCR derived product might be outcompeted by intron-less *rns* PCR derived product as the shorter fragments could be amplified preferentially when using exon-based primers.

Figure 3.5. Phylogenetic tree showing the phylogenetic position of the *O. ulmi* mS952 group II intron ORF nucleotide sequence within the mS952 intron ORF family. The tree includes sequences from *Leptographium* and *Cordyceps* spp., and the tree is rooted with the *C. parasitica* mS952 intron ORF sequence. The tree topology is based on Bayesian analysis and the numbers at the nodes represent the % posterior probability values as obtained from a 50% majority rule Bayesian consensus tree. The branch lengths are based on Bayesian analysis and are proportional to the number of substitutions per site. Strain numbers are underlined, and the GenBank sequence accession numbers are listed in brackets.

Figure 3.5.


Cryphonectria parasitica (AF029891)
Cordyceps ps. 97009 (AB027356)
Ophiocordyceps konnoana (AB031194)
Cordyceps ramosopulvinata (AB027348)
Ophiocordyceps sobolifera (AB027350)
Cordyceps ps. 97003 **(AB027352)**
Ophiostoma ulmi DAOM 171046 (HQ292075)
Leptographium lundbergii NFR189-1040/1/3 (GU983685)
Leptographium lundbergii NFRI1502/1 (GU983688)
Leptographium terebrantis CBS337.70 (GU983690)
Leptographium wingfieldii TOM1.3 (GU983695)
Leptographium truncatum Forintek C34 (GU983692)
Leptographium truncatum NFRI59-3/7 (GU983694)
Leptographium truncatum NFRI1813/1 (GU949594)
Leptographium truncatum CBS929.85 (GU949593)
100
100
100
100
100
80
96
72
97
0.1

There appears to be evidence that there is mtDNA heterogeneity within individual strains of *O. ulmi* as two bands were observed when exon based primers were used (Figure 3.1.[b]); however we did not observe evidence for additional bands within *O. novo-ulmi* subsp. *americana*. In addition the primer combination: mS952-F and mS952DV-R failed to detect any evidence for the presence of introns within *O. novo-ulmi* subsp. *americana*. Based on the dara, the mS952 intron is found in *O. ulmi* but not in *O. novo-ulmi* subsp. *americana.*

The absence of the intron in all *O. novo-ulmi* subsp. *americana* strains suggests that this intron could be a useful marker in distinguishing these two economically important species. In areas were both species still exist, a molecular marker might be an alternative to cultural methodologies for distinguishing these two species. Due to the potential for the various DED agents to hybridize, a monitoring strategy for the presence of *O. ulmi* might be valuable in risk assessment for predicting future pandemics of DED. In experimental approaches where these fungi are maintained in the same laboratory or where hybrids might have to be generated or identified, molecular markers might be very valuable in facilitating the analysis of inter-species crosses or heterokaryon formation.

In a previous study on the mtDNA large ribosomal submit gene (*rnl*) gene (Sethuraman et al., 2008), we noted that there was evidence that the *rnl*-U7 group I intron (mL1699) could potentially be transferred between *O. novo-ulmi* subsp. *americana* and *O. ulmi*. This demonstrates that group I introns like group II introns are mobile elements that can readily cross species barriers, in particular between closely related species such

as members of the *O. ulmi* species complex. However, we failed to detect any mS952 introns within *O. novo-ulmi* subsp. *americana* suggesting that this intron so far failed to be transferred between the two species in the specimens examined. It is known that for efficient splicing both group I and II introns usually required host genome encoded splicing factors/maturases, and one could speculate that the mS952 intron requires a factor for efficient splicing that is not available within the *O. novo-ulmi* subsp. *americana* genetic background. In order for a potentially mobile element to move horizontally to a new species it has to find a compatible genomic environment that allows the element to be non-toxic to the new host. Splicing deficiency could be costly to the host. For example, in a recent study on a mtDNA *rns* group II intron in *Cryphonectria parasitica* (Chestnut blight fungus) splicing deficiency was linked to growth abnormalities and hypovirulence (Baidyaroy et al., 2011).

3.5.2. The *rns* RNA secondary structure model and the O.ul-mS952 intron:

Sequence characterization has shown that the *O. ulmi rns* intron is a group II intron that encodes a LAGLIDADG type ORF in domain III. Similar introns have been previous described and they are novel as typically group II introns encode RT ORFs not HEases (Toor and Zimmerly, 2002; Monteiro-Vitorello et al., 2009; Mullineux et al., 2010). The RT-PCR results confirm that this element is an intron and that it is spliced *in vivo*.

We modeled the *rns* RNA secondary structure to evaluate the potential functional constraints on the intron insertion on the rRNA. In general group I or II introns occur at a

particular position within the rRNA due to vertical inheritance of the allele containing an intron, homing or transposition (ectopic integration) in an intron-less allele/site, or due to horizontal transfer of the intron into a specific conserved target site (Jackson et al., 2002). Usually mobile introns insert into alleles that lack the intron by mechanisms that are site specific. Therefore, introns evolved strategies that optimize their dispersal, such as targeting highly conserved sequences present in every member of a population or even in other species. The distribution of introns in rRNA genes is not totally random; insertion sites appear to be found within conserved sequences usually near the tRNA and mRNA binding sites which span the interface between the small and the large subunits of the ribosome, suggesting a link between intron evolution and rRNA function (Gerbi et al., 1982). Generally, domain III is located at the "head" of the small (30S) subunit of the ribosome to which ribosomal proteins such as S2, S3, S7, S9, S10, S13, S14 and S19 can bind (Moore et al., 1986). The O.ul-mS952 intron is inserted in the 5' end of the lower half of the 16S rRNA domain III region which represents the interaction area with ribosomal protein S7 (Dragon and Brakier-Gingras, 1993; Robert et al., 2000). Ribosomal protein S7 plays an important role in ribosome function as this protein is responsible for initiating the assembly of the 30S subunit. S7 is also one of the major ribosomal proteins to cross-link with tRNA molecules during the decoding process (Hosaka, 1997).

The presence of intron RNA in the mature rRNA could affect the ribosome assembly by altering the secondary structure or by blocking the tertiary or quaternary contacts and ribosomal protein interactions (Woodson and Cech, 1991). Unspliced or

slowly splicing introns would prevent or interfere with ribosome subunit assembly and the formation of ribosomes which are deficient or not competent with regards to translational activity (Nikolcheva and Woodson, 1997). So given the location of the mS952 intron in domain III of the SSU any splicing deficiency could be detrimental or in some way affect the virulence of the strain.

CHAPTER: 4

THE HIGHLY VARIABLE MITOCHONDRIAL SMALL-SUBUNIT RIBOSOMAL RNA GENE OF O*phiostoma minus*.*

4.1. ABSTRACT:

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

* A version of this chapter was published in Fungal Biology. Hafez M. and Hausner G. **2011a**. 115: 1122-1137.

4.2. INTRODUCTION:

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

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

usually mediated by the intron encoded proteins or by reverse splicing of the intron RNAs (Woodson and Cech, 1989).

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

Homing endonuclease genes and introns are quite invasive and contribute towards the size of fungal mtDNA genomes, mtDNA polymorphisms, and they promote mtDNA rearrangements (Dujon, 1989; Charter et al. 1996; Belcour et al., 1997; Salvo et al., 1998; Hamari et al., 1999; Gobbi et al., 2003). Also group I and II introns have been associated with mtDNA instabilities such as generating plasmid-like elements that are found in

senescent and/or hypovirulent strains in an assortment of filamentous fungi such as *Podospora anserina* (Osiewacz and Esser, 1984; Michel and Cummings, 1985; Cummings et al., 1986; Dujon and Belcour, 1989; Cummings et al., 1990), *Ophiostoma novo-ulmi* (Abu-Amero et al., 1995; Sethuraman et al., 2008) and *Cryphonectria parasitica* (Monteiro-Vitorello et al., 2009; Baidyaroy et al., 2011).

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

4.3. RESULTS:

Methods relevant to Chapter 4 are presented in Chapter 2 (Materials and Methods).

4.3.1. The *rns* introns in strains of *O. minus*:

Amplification of a segment of the *rns* gene from 21 strains of *O. minus* using the mtsr-1 and mtsr-2 primer generated PCR products ranging in size from 1.2 kb to 7.0 kb (Figure 4.1. & Table 4.1.). A 1.2 kb fragment is expected when no insertions are present, so the 3.0, 4.0, 4.4, 5.4 and 7.0 kb amplicons indicate the presence of introns. Representatives of the various amplicon size classes [WIN(M)1574, WIN(M)873, WIN(M)494, WIN(M)371, WIN(M)472, and WIN(M)515] were selected for detailed DNA sequence analysis (See Appendix 9.3. for a complete primers list used to sequence the *O. minus rns* gene and Appendix 9.4. for GenBank accession numbers). Based on comparing *rns* sequences that lacked introns with those that contain introns, four intron insertion sites could be identified: mS379, mS569, mS952 and mS1224 (Table 4.2., Figure 4.2. [a] & [b]). Introns were named according to the nomenclature for group I introns inserted in the ribosomal RNA genes (Johansen and Haugen, 2001). Intron insertion site in *O. minus rns* gene and the corresponding insertion site in *E. coli* SSU rRNA gene were also indicated. The letters and numbers used in the nomenclature were underlined (see Table 4.2.).

The mS379 intron characterized in strains WIN(M)472, WIN(M)494 and WIN(M)515 is a class A group II intron, based on the presence of the conserved domain V (D V) sequence. The intron encodes a putative RT-like protein that has been fragmented into two segments by frame shift mutations (Figure 4.2. [a]). The mS569

intron found in strains WIN(M)371 and WIN(M)472 is a class C2 group I intron and it encodes a LHEase-like sequence (Table 4.2.). The third intron at position mS952 as characterized in strains WIN(M)371, WIN(M)515 and WIN(M)873 is a typical class B group II intron and it encodes a LAGLIDADG HE-like sequence.

Figure 4 1. PCR amplicons from different strains of *O. minus* using the mtsr-1 and mtsr-2 primers that target the *rns* gene (see Table 4.1. for list of strains corresponding to lane numbering). Sizes of representative PCR products are indicated. The L lane contains the resolved 1 kb plus (Invitrogen) ladder. See Table 4.1. for strain names.

Figure 4.1.

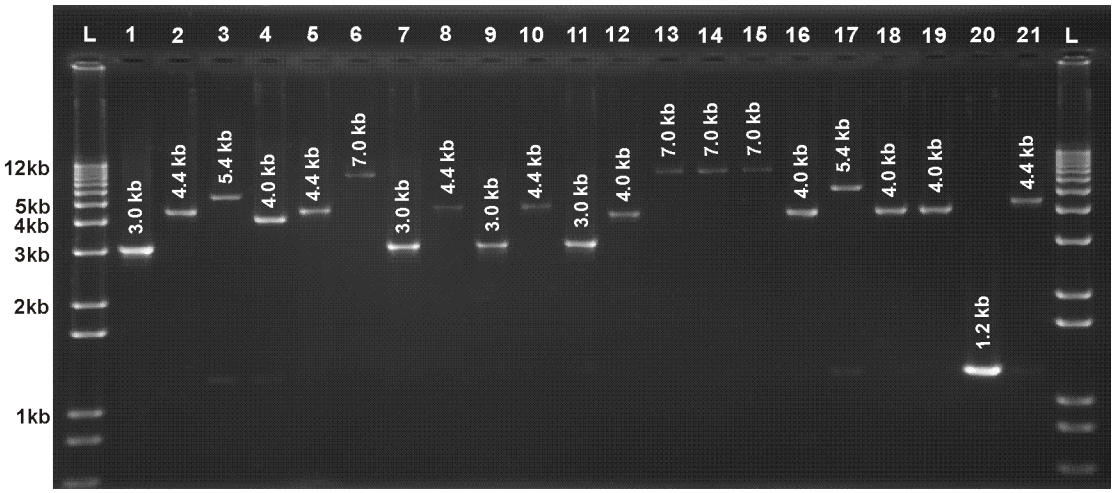

L 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 L
12kb
5kb
4kb
3kb
2kb
1kb
3.0 kb
4.4 kb
5.4 kb
4.0 kb
4.4 kb
7.0 kb
3.0 kb
4.4 kb
3.0 kb
4.4 kb
3.0 kb
4.0 kb
7.0 kb
7.0 kb
7.0 kb
4.0 kb
5.4 kb
4.0 kb
4.0 kb
1.2 kb
4.4 kb

Table 4.1. List of *O. minus* strains used in the present study, location of collection and the corresponding *rns* gene length of each strain.

Table 4.1.

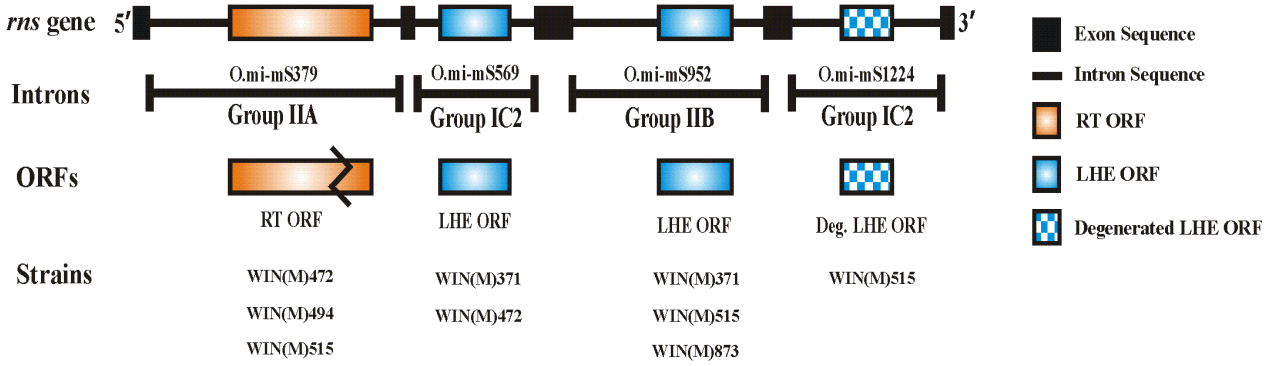
S	Species	Strain No.	Location	*rns* Length (kb)
1	*O. minus*	WIN(M)292	Sandilands, Forest Reserve, Manitoba, Canada	3.0
2	*O. minus*	WIN(M)371	Forest Besrve, Manitoba,Canada	4.4
3	*O. minus*	WIN(M)472	Taylor lake, Alberta, Canada	5.4
4	*O. minus*	WIN(M)494 / C-248 (NFRC)	Edmonton, Alberta, Canada	4.0
5	*O. minus*	WIN(M)495 / C-262 (NFRC)	Edmonton, Alberta, Canada	4.4
6	*O. minus*	WIN(M)515 / CBS 404.77	California. USA	7.0
7	*O. minus*	WIN(M)861 / DAOM 29251	Toronto, Ontario, Canada	3.0
8	*O. minus*	WIN(M)871 / DAOM 29251	Sanchlando, Manitoba, Canada	4.4
9	*O. minus*	WIN(M)873	Edmonton, Alberta, Canada	3.0
10	*O. minus*	WIN(M)874 / C-874 (NFRC)	Edmonton, Alberta, Canada	4.4
11	*O. minus*	WIN(M)875 / C-845 (NFRC)	Edmonton, Alberta, Canada	3.0
12	*O. minus*	WIN(M)876 / C-342 (NFRC)	Edmonton, Alberta, Canada	4.0
13	*O. minus*	WIN(M)888 / ATCC 15321	Wyoming, Michigan, USA	7.0
14	*O. minus*	WIN(M)888A / ATCC 15321	Wyoming, Michigan, USA	7.0
15	*O. minus*	WIN(M)889 / ATCC 22388	California. USA	7.0
16	*O. minus*	WIN(M)890 / ATCC 11065	Washington, UAS	4.0
17	*O. minus*	WIN(M)1213	Bracebridge, Ontario, Canada	5.4
18	*O. minus*	WIN(M)1275	Barrie, Ontario, Canada	4.0
19	*O. minus*	WIN(M)1573	Lakehead, Ontario, Canada	4.0
20	*O. minus*	WIN(M)1574	Lakehead, Ontario, Canada	1.2
21	*O. minus*	WIN(M)1575	Lakehead, Ontario, Canada	4.4

WIN(M) = University of Manitoba culture collection (Winnipeg, MB, Canada).
ATCC= American Type Culture Collection (Manassas, VA).
CBS= Central Bureau voos Schimmelcultures (Utrecht, The Netherland).
NFRC= Northern Forestry Research Center (Edmonton, AB, Canada).
DAOM= Plant Research Institute, Department of Agriculture, Mycology, Ottawa, Canada.

Figure 4.2. [a] A schematic overview of the *O. minus rns* gene showing the two intron types (group I and II) and the intron encoded ORFs (RTs and LHEases). The introns are located at positions mS379, mS569, mS952 and mS1224 with reference to the *E. coli* SSU rRNA sequence**.** The RT ORF is fragmented into two segments by frame shift mutations. Examples of strains are listed for which the complete mtsr-1/2 PCR product sequences were obtained. **[b]** Sequences that define the exon/intron junctions for the group I (O.mi-mS569 and O.mi-mS1224) and group II introns (O.mi-mS379 and O.mi-mS952). For the group II introns the IBS1, IBS2, IBS3 and δ^{-} sequences are indicated.

Figure 4.2.

[a]


rns gene 5′
3′
Introns
O.mi-mS379
Group IIA
O.mi-mS569
Group IC2
O.mi-mS952
Group IIB
O.mi-mS1224
Group IC2
ORFs
RT ORF
LHE ORF
LHE ORF
Deg. LHE ORF
Strains
WIN(M)472
WIN(M)494
WIN(M)515
WIN(M)371
WIN(M)472
WIN(M)371
WIN(M)515
WIN(M)873
WIN(M)515
Exon Sequence
Intron Sequence
RT ORF
LHE ORF
Degenerated LHE ORF

[b]

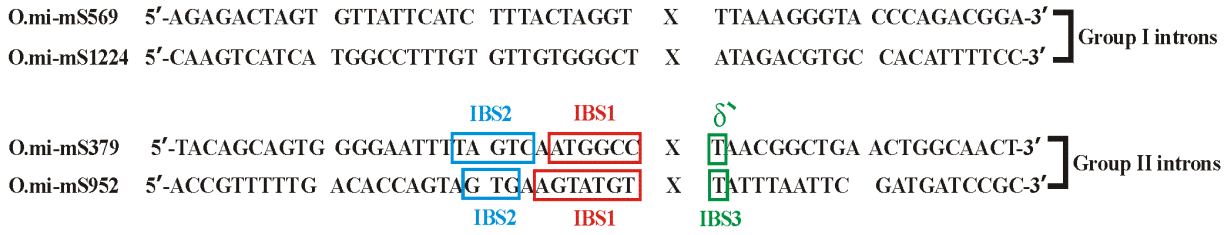

O.mi-mS569 5′-AGAGACTAGT GTTATTCATC TTTACTAGGT X TTAAAGGGTA CCCAGACGGA-3′
O.mi-mS1224 5′-CAAGTCATCA TGGCCTTTGT GTTGTGGGCT X ATAGACGTGC CACATTTTCC-3′
Group I introns
IBS2
IBS1
δ`
O.mi-mS379 5′-TACAGCAGTG GGGAATTTTA GTCAATGGCC X TAACGGCTGA ACTGGCAACT-3′
O.mi-mS952 5′-ACCGTTTTTG ACACCAGTAG TGAAGTATGT X TATTTAATTC GATGATCCGC-3′
Group II introns
IBS2
IBS1
IBS3

Table 4.2. Ribosomal RNA group I and II introns nomenclature.

Table 4.2.

Intron name	Host species	Strain #	Host gene	Insertion site (in *O. minus*)	Insertion site (in *E. coli*)
O.mi472-mS379	***O. mi****nus*	WIN(M)**472**	**m**t-**S**SU-rRNA	284-285	**379**-380
O.mi494-mS379	*O. minus*	WIN(M)**494**	mt-SSU-rRNA	284-285	379-380
O.mi515-mS379	*O. minus*	WIN(M)**515**	mt-SSU-rRNA	284-285	379-380
O.mi371-mS569	*O. minus*	WIN(M)**371**	mt-SSU-rRNA	434-435	569-570
O.mi472-mS569	*O. minus*	WIN(M)**472**	mt-SSU-rRNA	433-434	569-570
O.mi371-mS952	*O. minus*	WIN(M)**371**	mt-SSU-rRNA	777-778	952-953
O.mi515-mS952	*O. minus*	WIN(M)**515**	mt-SSU-rRNA	778-779	952-953
O.mi873-mS952	*O. minus*	WIN(M)**873**	mt-SSU-rRNA	777-778	952-953
O.mi515-mS1224	*O. minus*	WIN(M)**515**	mt-SSU-rRNA	1028-1029	1224-1225

The fourth intron inserted at position mS1224 in strain WIN(M)515, is a class C2 group I intron, and contains a highly degenerated/fragmented LAGLIDADG ORF due to frame shift mutations and the presence of premature stop codons.

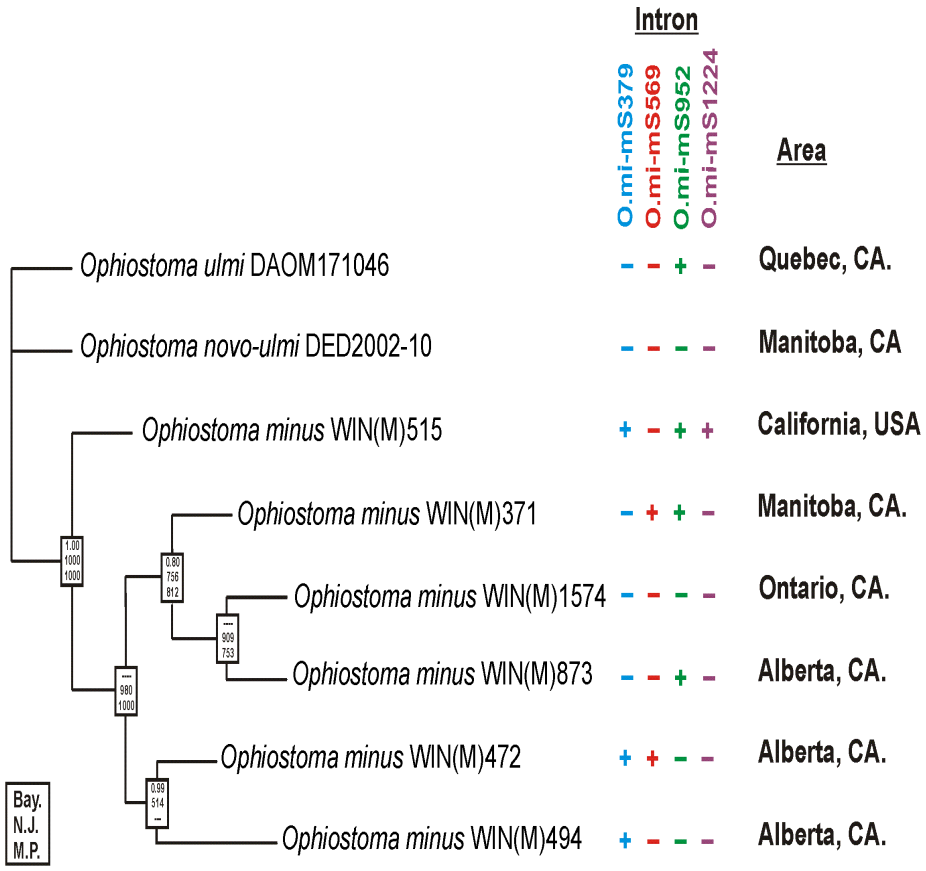
An *O. minus* species tree was generated by combining the *rns*, ITS, and partial EF-1α nucleotide sequences for those strains that were selected for more detailed analysis of the *rns* gene (Figure 4.3.). The species tree was compared to the distribution of the various introns uncovered in this study and there was no obvious connection between relatedness among the various strains and their intron composition. The strain with the most introns WIN(M)515 appears to be the most divergent member among the tested *O. minus* strains (See appendix 9.3 for complete list of GenBank accession numbers of the *rns*, ITS, and partial EF-1α nucleotide sequences used in this phylogenetic analysis)

4.3.2. Group I introns:

The group I introns found in the *O. minus rns* gene belong to class C2 due to the presence of P5a and P5b in the P5 paring region and the absence of a P2 region (Comparative RNA Web Site (CRW), http://www.rna.ccbb.utexas.edu, Cannone et al., 2002). The P, Q, R and S sequences (Michel and Westhof, 1990) could be identified but P2 is absent (Figure 4.4. [a] & [b]). The start codon of the LAGLIDADG ORF is located at the stem of P5b, while the stop codon is found within the loop of P9.1 region. The mS1224 intron is longer (1394 bp) than the mS569 introns (1323 bp), and the P9 region of the mS1224 intron is a more complex fold and contains a highly fragmented LAGLIDADG type ORF (Figure 4.5.).

Figure 4.3. The phylogenetic tree showing the relatedness among various *O. minus* strains based on a combined *rns*, ITS and EF-1α nucleotide sequence data set. The tree topology is based on NJ analysis. The *O. ulmi* and *O. novo-ulmi* subspecies *americana* sequences data served as the outgroup. The upper number represents the posterior probabilities (PP) values obtained by generating a Bayesian 50% majority rule consensus tree. The middle number is the bootstrap (BS) support based on NJ analysis, while the last number represents the BS values based on Parsimony analysis. Nodes that received less than 50% support (BS or PP) were collapsed. The branch lengths as shown are based on NJ analysis and are proportional to the number of substitutions per site. Distribution of *rns* introns and geographic origin of strains are shown on the right.

Figure 4.3.



* See appendix 9.4 for complete list of GenBank accession numbers of the *rns*, ITS, and partial EF-1α nucleotide sequences used in this phylogenetic analysis.

Figure 4.4. Secondary structure of **[a]** O.mi472-mS569 and **[b]** O.mi371-mS569 group I (class C2) introns found in the *O. minus* strain WIN(M)472) and WIN(M)371 *rns* gene respectively. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The ten paring regions (P1-P10) are indicated and the P10 region is boxed. The solid black arrowheads indicate the intron-exon junctions (5′ and 3′ splicing sites). The LAGLIDADG ORF has a putative start codon within the P5b region and the ORF extends into P9.1 loop.

Figure 4.4. [a]

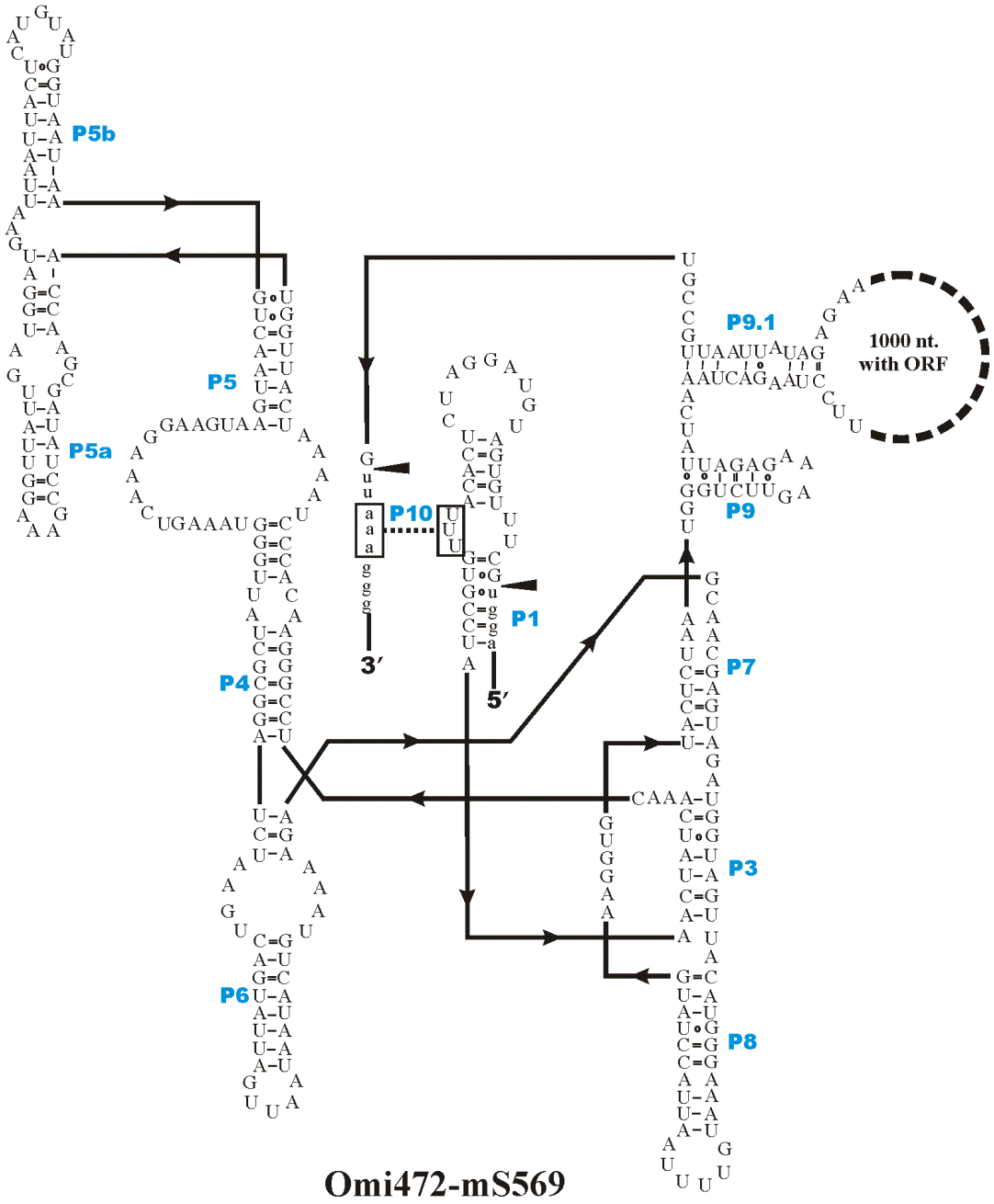

P5b
P5a
P5
P4
P6
P10
P1
P9.1
P9
P7
P3
P8
1000 nt.
with ORF
3′
5′
Omi472-mS569

Figure 4.4. [b]

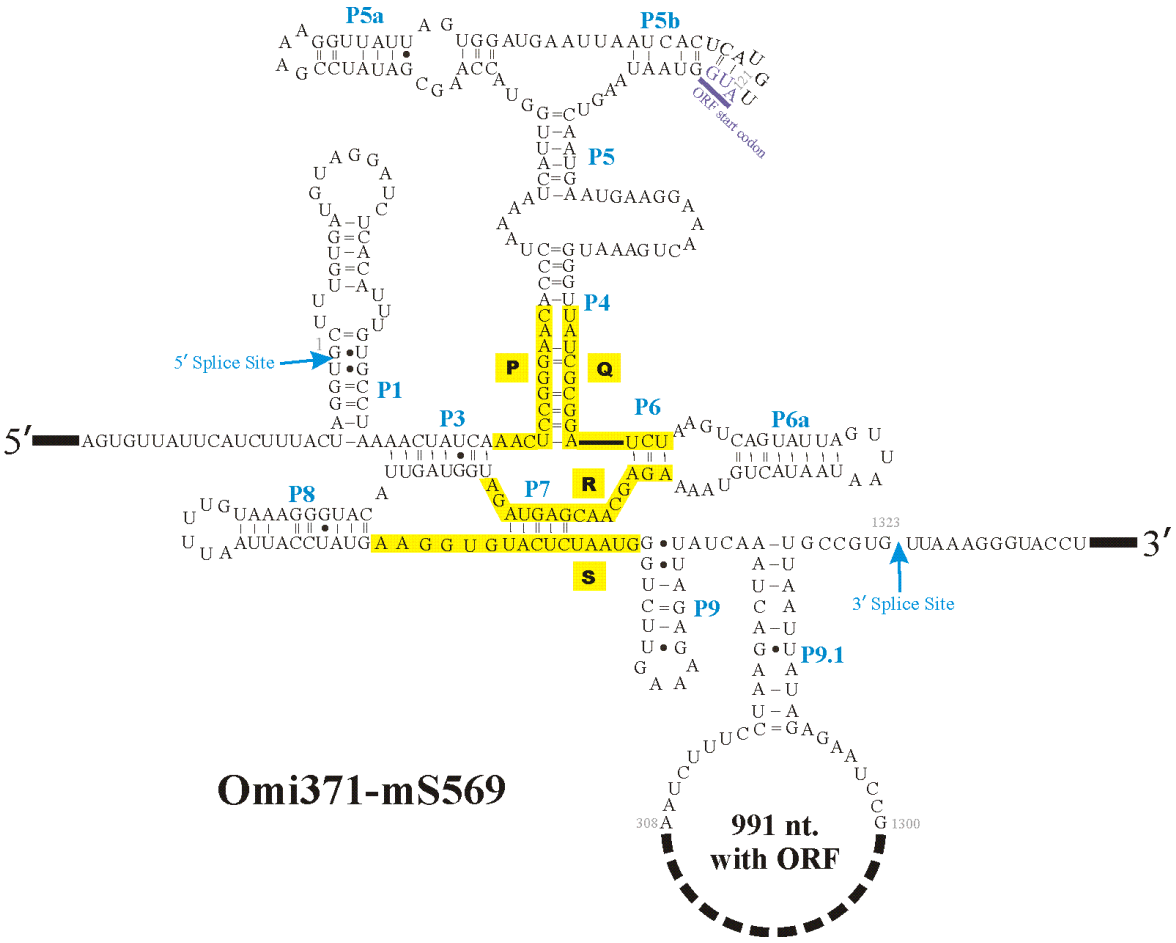
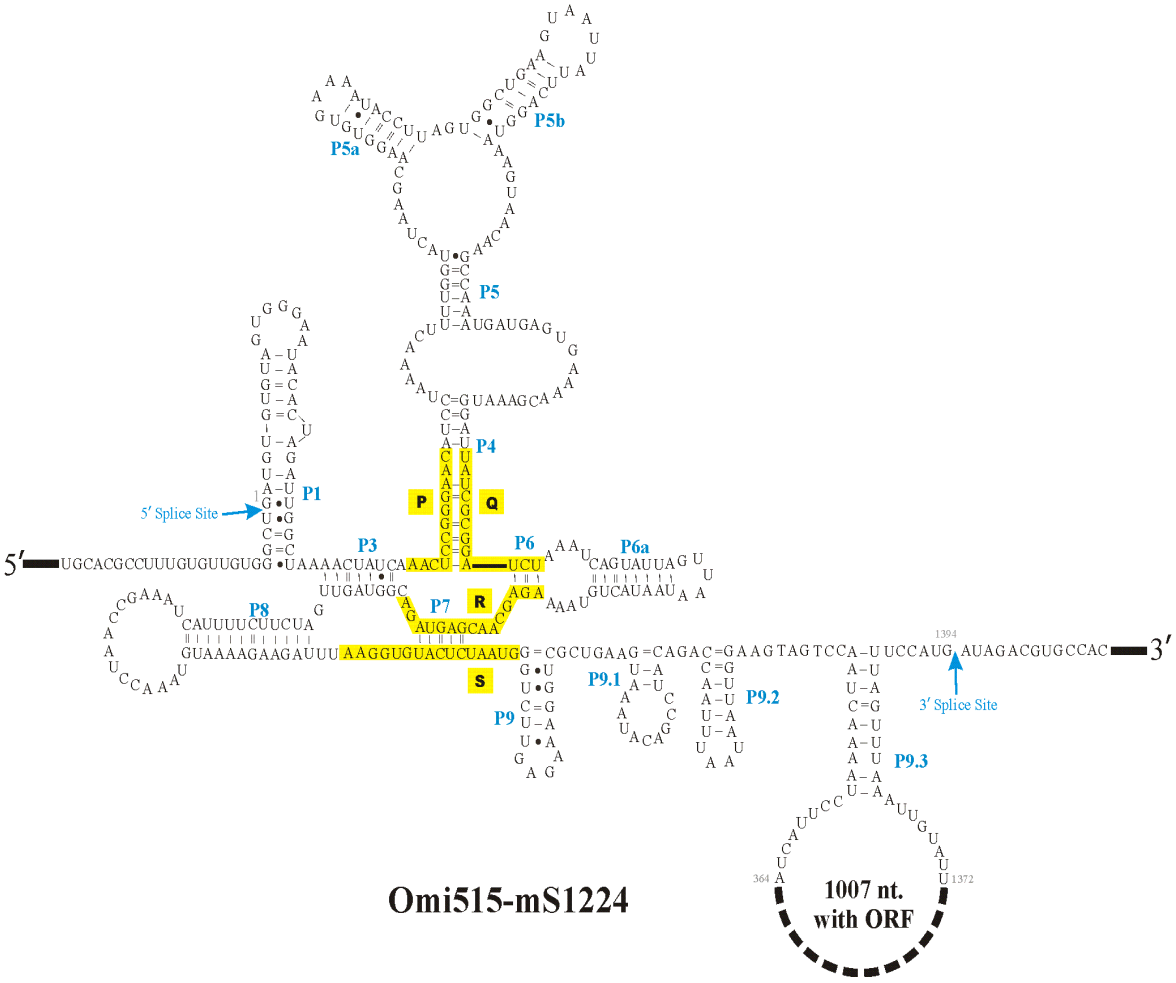

P5a
P5b
P5
P4
P1
P3
P6
P6a
P7
P8
P9
P9.1
P
Q
R
S
5′ Splice Site
3′ Splice Site
ORF start codon
5′
3′
AGUGUUAUUCAUCUUUACU–AAAACUAUCAAACU–A
UCU
UUGAUGGU
AUGAGCAA
AAGA
AAAU
UCAGUAUUAG
GUCAUAAU
UAAAGGGUAC
AUUACCUAUG
AAGGUGUACUCUAAUG
UAUCAA–UGCCGUG
UUAAAGGGUACCU
1323
308
1300
991 nt.
with ORF
Omi371-mS569

Figure 4.5. Secondary structure of O.mi515-mS1224 group IC2 intron found in the *O. minus* strain WIN(M)515 *rns* gene. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The ten paring regions (P1-P10) are indicated and the P10 region is boxed. The solid black arrowheads indicate the intron-exon junctions (5′ and 3′ splicing sites). The LAGLIDADG ORF is highly degenerated.

Figure 4.5.


P5a
P5b
P5
P4
P1
5′ Splice Site
P
Q
P3
P6
P6a
P8
P7
R
S
P9
P9.1
P9.2
P9.3
3′ Splice Site
1007 nt.
with ORF
Omi515-mS1224

4.3.3. Group II introns:

Both introns at mS379 (Figure 4.6.) and mS952 (Figure 4.7.) are typical group II introns containing the characteristic features of class A and class B respectively, such as the exon binding sites (EBS1, EBS2 & EBS3/δ), which are complementary to the intron binding sites (IBS1, IBS2 & IBS3/δ^{-}) in the upstream and the downstream exons flanking the intron insertion site, the internal loop of DIC*i* (ε^{-}), the internal loop of DII, the linkers between domains I to VI and the absence of insertions in the 3' strand of DI*i* and DI*ii* (Michel and Ferat, 1995; Toor et al., 2001). What distinguishes this group II intron from others are the type and position of the encoded ORF; group II introns usually encode an RT-like ORF in the loop of domain IV. The mS379 intron encodes a RT-like ORF in the loop of domain II (Figure 4.6 [a], [b] & [c]), while the mS952 intron encodes a potential LHEase within the loop of domain III (Figure 4.7 [a], [b] & [c]).

4.3.4. *In vivo* splicing of the *rns* introns:

To confirm the predicted exon/intron junctions derived by comparative sequence analysis, RT-PCR analysis was applied to RNA from strains WIN(M)371, WIN(M)472, WIN(M)494 and WIN(M)888 (Figure 4.8.). We were not able to extract RNA from strain WIN(M)515 so we substituted it with strain WIN(M) 888, a strain that has the same *rns* gene configuration as WIN(M)515. The *rns* genes from these strains contain representatives of all introns encountered during this study: WIN(M) 371 contains mS569 and mS952, WIN(M) 472 contains m379 and mS569, WIN(M) 494 contains mS379, and WIN(M) 888 contains mS379, mS952 and mS1224 (Figure 4.2. and Table 4.2.).

Figure 4.6. Secondary structures of the O.mi472-mS379 [a]; O.mi494-mS379 [b] and O.mi515-mS952 [c] group IIA1 intron RNAs. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The positions of EBS1, EBS2 and δ are noted. The positions of IBS1, IBS2 and δ- in the 5′ and 3′ exons are boxed with dotted lines. Tertiary interactions are indicated by dashed lines and Greek letters (ε, λ, α, β, θ, κ, ζ, and γ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V and VI). The solid black arrowheads indicate the intron-exon junctions (5′ and 3′ splicing sites). The asterisk shows the bulged adenosine nucleotide in domain VI (the branch point). The RT ORF is encoded within DII.

Figure 4.6. [a]

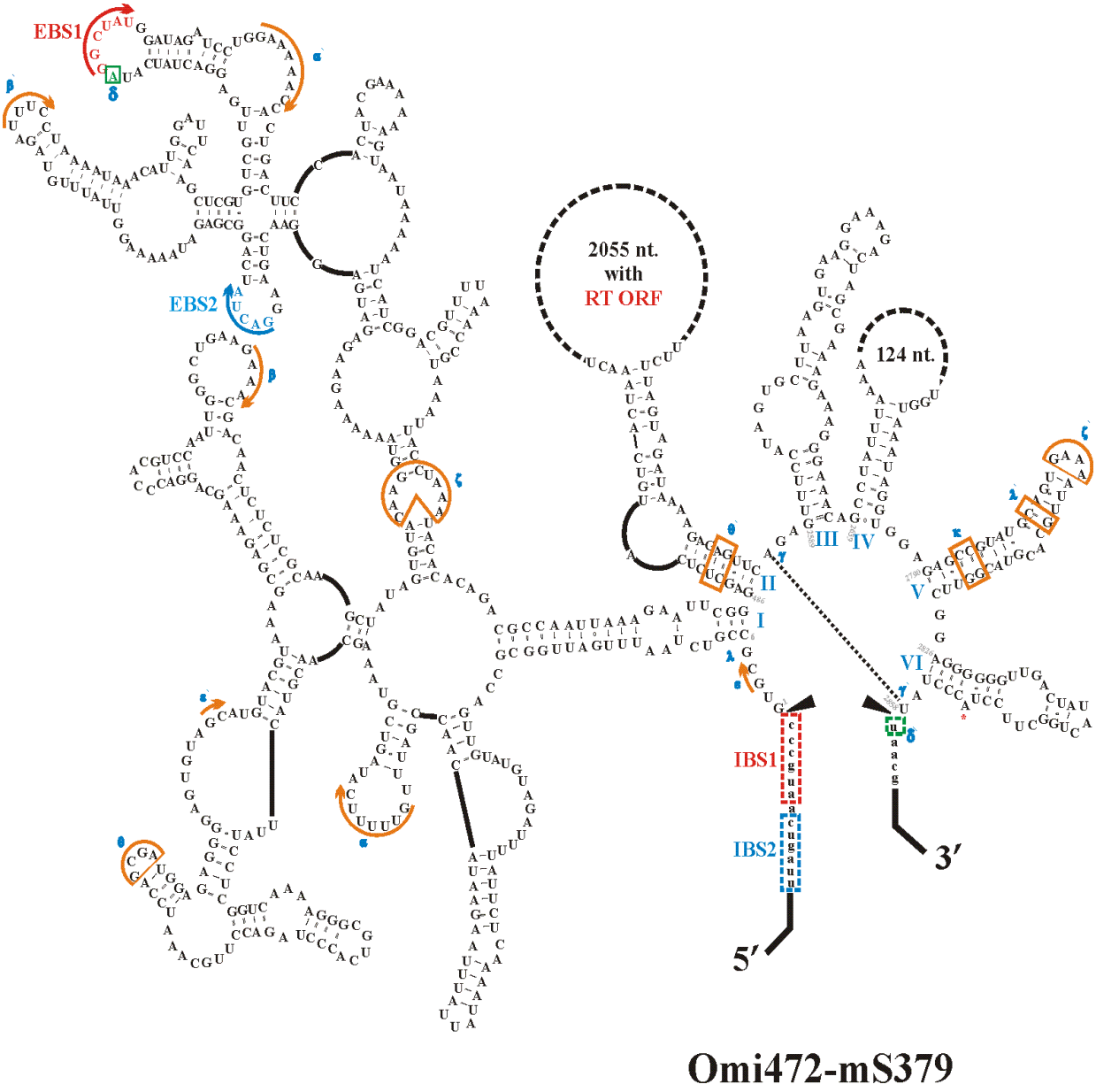

EBS1
EBS2
2055 nt.
with
RT ORF
124 nt.
I
II
III
IV
V
VI
IBS1
IBS2
5′
3′
Omi472-mS379

Figure 4.6. [b]

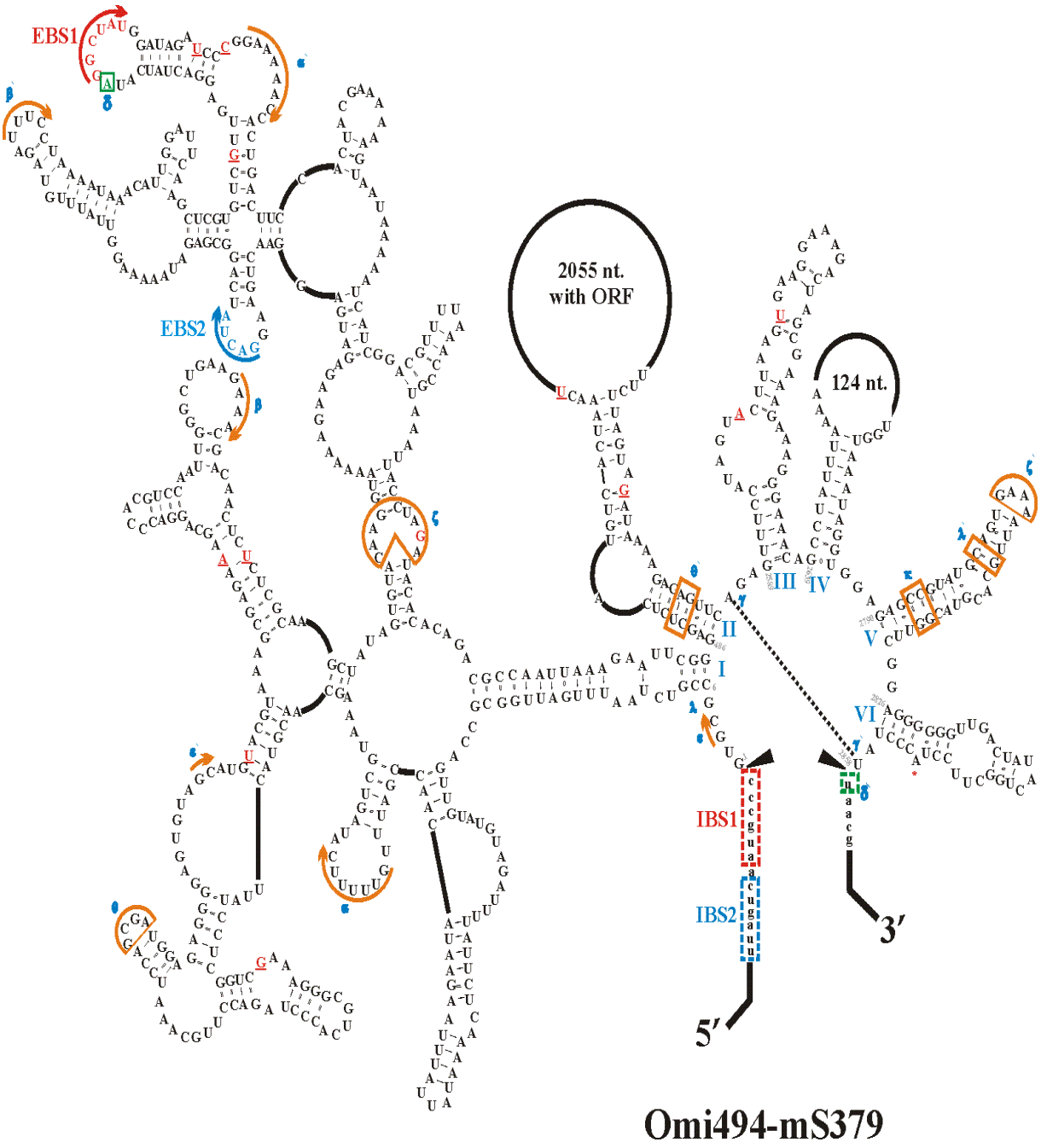

EBS1
EBS2
2055 nt.
with ORF
124 nt.
I
II
III
IV
V
VI
IBS1
IBS2
5′
3′
Omi494-mS379

Figure 4.6. [c]

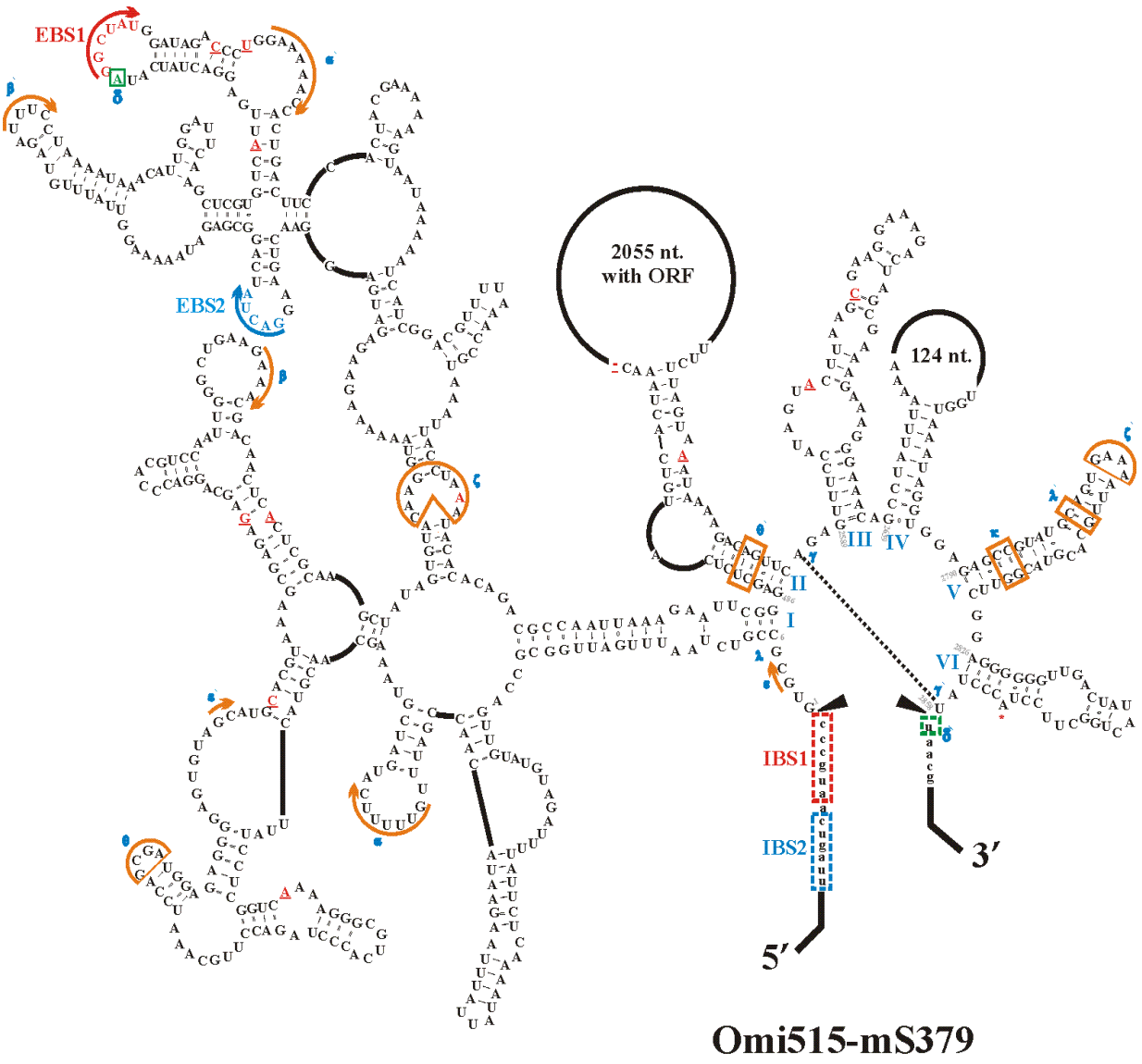
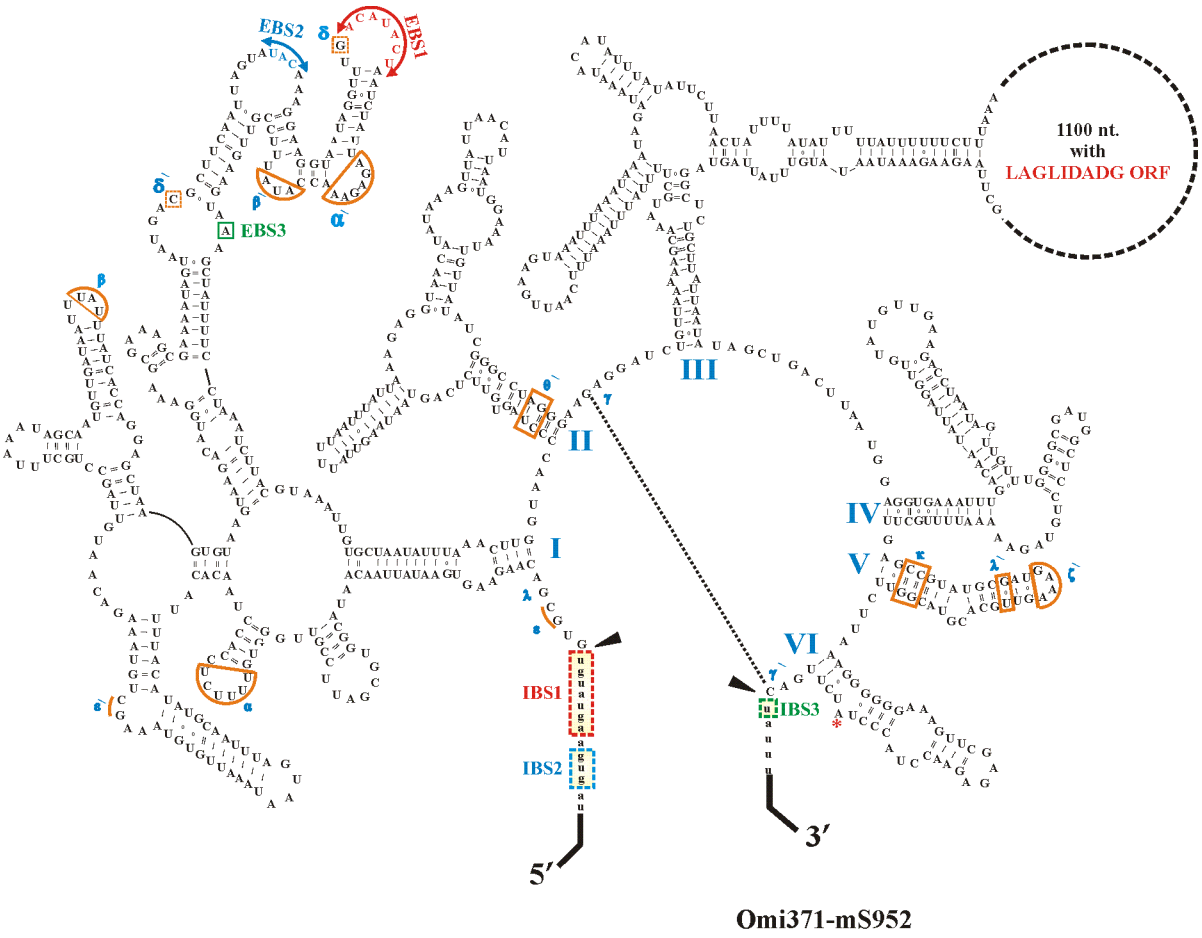

EBS1
EBS2
2055 nt.
with ORF
124 nt.
III
IV
V
VI
II
I
IBS1
IBS2
5′
3′
Omi515-mS379

Figure 4.7. Secondary structures of the O.mi371-mS952 **[a];** O.mi873-mS952 **[b]** and O.mi515-mS952 **[c]** group IIB1 intron RNAs. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The positions of EBS1, EBS2 and EBS3 are noted. The positions of IBS1, IBS2 and IBS3 in the 5′ and 3′ exons are boxed with dotted lines. Tertiary interactions are indicated by dashed lines and Greek letters (ε, λ, α, β, θ, κ, ζ, and γ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V and VI). The solid black arrowheads indicate the intron-exon junctions (5′ and 3′ splicing sites). The asterisk shows the bulged adenosine nucleotide in domain VI (the branch point). The LAGLIDADG ORF is encoded within DIII.

Figure 4.7. [a]



Omi371-mS952

Figure 4.7. [b]

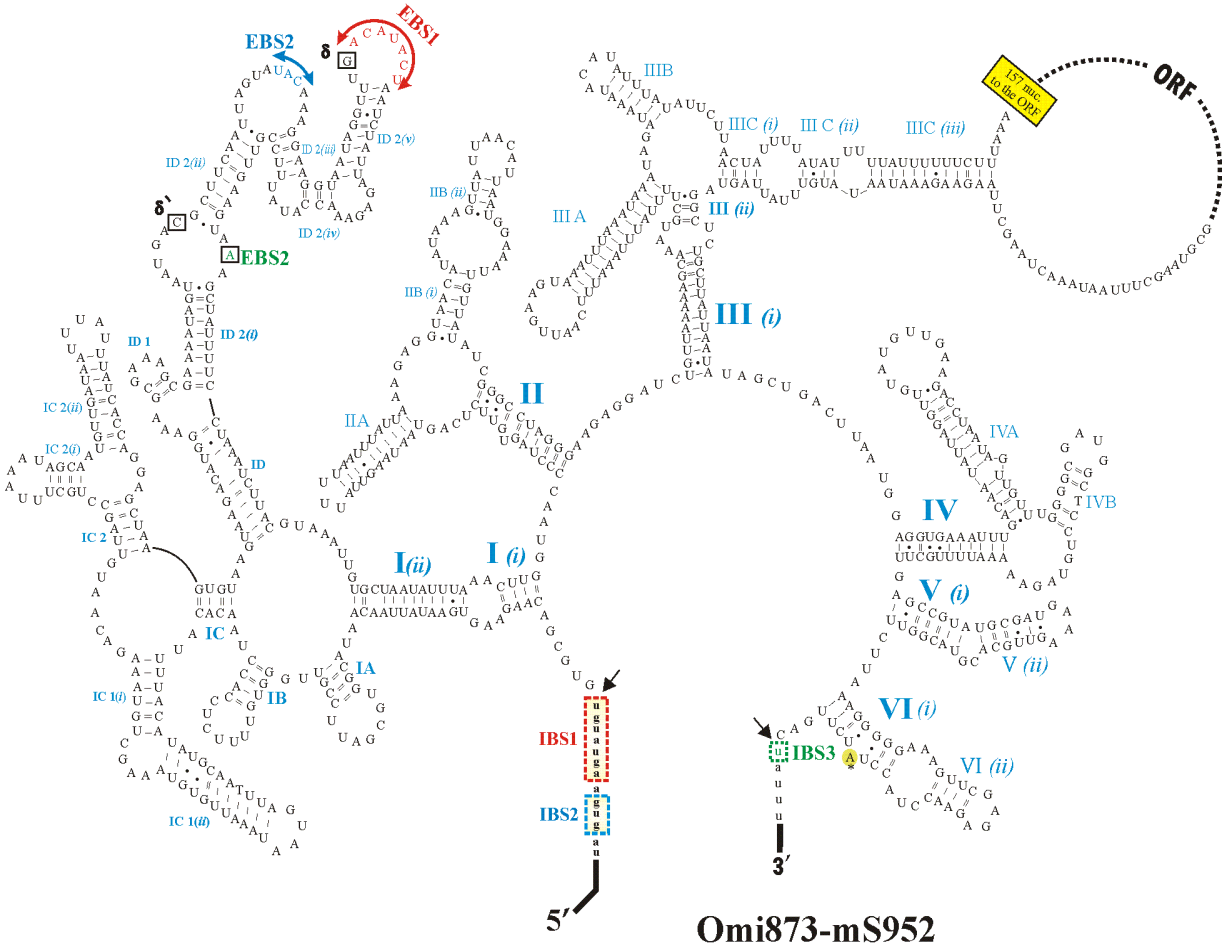
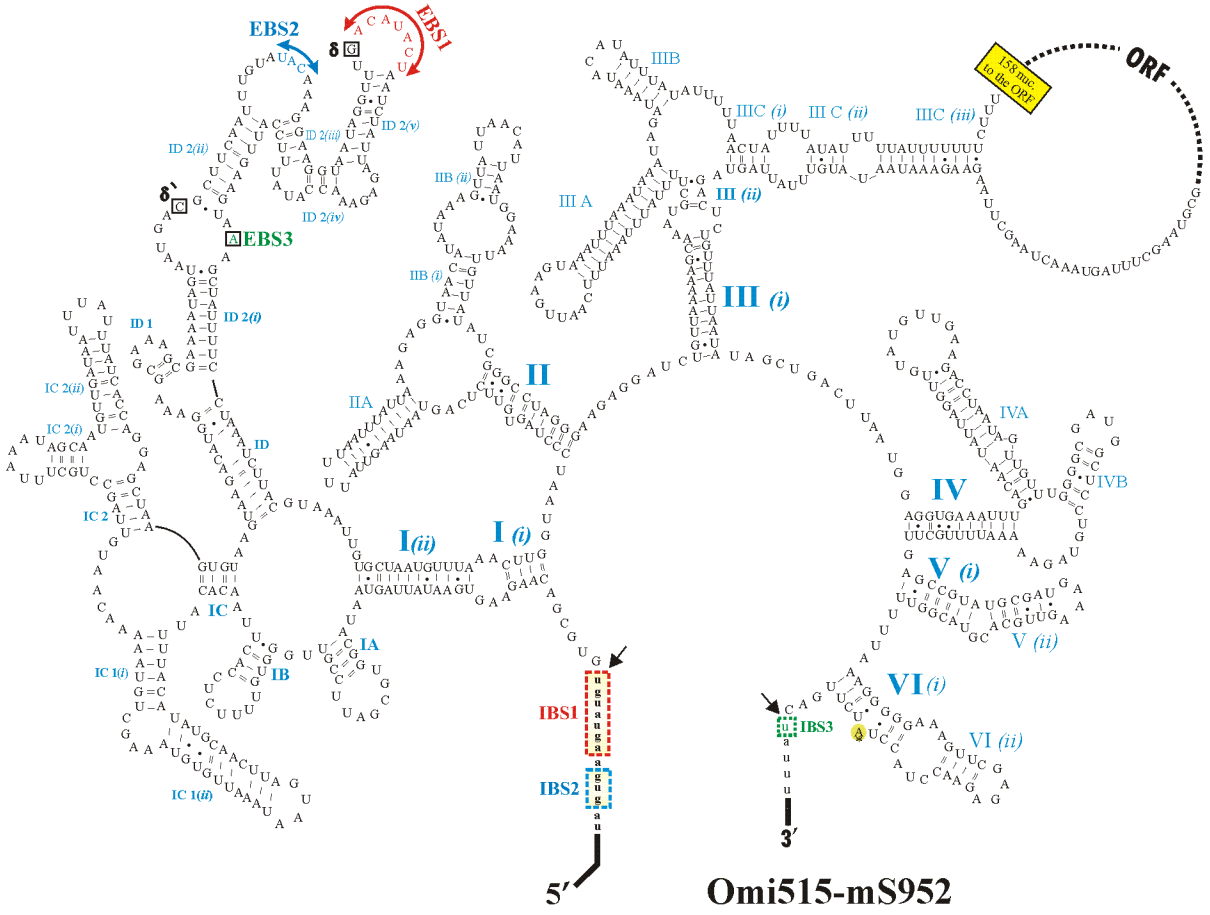

EBS2
EBS1
δ
δ'
ID 2(ii)
ID 2(iii)
ID 2(v)
ID 2(iv)
EBS2
IIB (ii)
IIB (i)
III A
IIIB
III C (ii)
IIIC (i)
IIIC (iii)
III (ii)
III (i)
ID 1
ID 2(i)
IC 2(ii)
IC 2(i)
IC 2
ID
II
IIA
I(ii)
I (i)
IC
IA
IB
IC 1(i)
IC 1(ii)
IBS1
IBS2
IBS3
IVA
IVB
IV
V (i)
V (ii)
VI (i)
VI (ii)
157 nuc. to the ORF
ORF
3′
5′
Omi873-mS952

Figure 4.7. [c]


EBS2
EBS1
EBS3
ID 1
ID 2(i)
ID 2(ii)
ID 2(iii)
ID 2(iv)
ID 2(v)
ID
IC
IC 2
IC 2(i)
IC 2(ii)
IC 1(i)
IC 1(ii)
IA
IB
IIA
IIB
IIB (i)
IIB (ii)
II
I (i)
I(ii)
IIIB
III A
III (i)
III (ii)
IIIC (i)
III C (ii)
IIIC (iii)
158 nuc. to the ORF
ORF
IBS1
IBS2
IBS3
IVA
IVB
IV
V (i)
V (ii)
VI (i)
VI (ii)
5′
3′
Omi515-mS952

Sequence analysis of the cDNA confirmed that the predicted introns were all spliced out and are thus missing from the mature rRNA and the sequence data confirmed the predicted exon/intron junction sequences (Figure 4.2. [b]).

4.3.5. Phylogenetic analysis of the LAGLIDADG ORFs:

Double motif LHEases were found within the *O. minus* mS569, mS952, and mS1224 introns. A LHEase-like 370 amino acid ORFs is encoded within the mS569 group I intron and a 306 amino acid LHEase-like ORF is located in the mS952 group II intron. The mS952 and mS1224 intron ORFs from strain WIN(M)515 were found to be highly degenerated/fragmented due to numerous mutations. These ORFs were omitted from the phylogenetic analysis (Figure 4.9.) as unambiguous reconstruction of the putative intact (or near intact) amino acid sequence was not possible. Related sequences to the *rns* intron encoded ORFs were extracted from GenBank with BLASTp. In total 43 double motif LAGLIDADG ORF amino acid sequences were compiled, aligned and subjected to phylogenetic analysis in order to determine the possible origins of the *O. minus rns* intron ORFs (Figure 4.9.).

The *O. minus* mS952 ORF belongs to a family of double motif LAGLIDADG ORFs that appear to share a common ancestor with LAGLIDADG ORFs that are encoded by group I introns located within protein coding genes and rRNA genes (node 1, Figure 4.9.). The tree topology shows that group I encoded ORFs branch basal to the mS952 encoded ORF, which suggests that the LAGLIDADG ORF most likely transferred from a group I intron to the *rns* mS952 group II intron (node 2, Figure 4.9.).

Figure 4.8. RT-PCR analysis for demonstrating *in vivo* splicing of *O. minus* introns and determining the intron/exon junctions. The *rns* transcripts for the following strains were analyzed by RT-PCR: WIN(M)371, 472, 494 and 888 (lanes 2, 5, 8 and 11 respectively). The standard PCR reactions using genomic DNA as templates are shown in lanes 1, 4, 7 and 10 for strains WIN(M)371, 472, 494 and 888 respectively. The PCR products from the genomic DNA controls generated 4.4, 5.4, 4.0 and 7.0 kb bands (lanes 1,4,7 and 11 respectively) as expected from previous data (see Figure 4.1. and Table 4.1.). The amplicon lengths for all cDNAs were 1.2 kb indicating that the introns were all spliced out. The negative controls for all RT-PCR (standard PCR using DNA free whole cell RNA as a template) assays yielded no bands (lanes 3, 6, 9 and 12) showing that the RNA was free of any DNA. The lane marked "L" contains the 1 kb plus ladder (Invitrogen).

Figure 4.8.

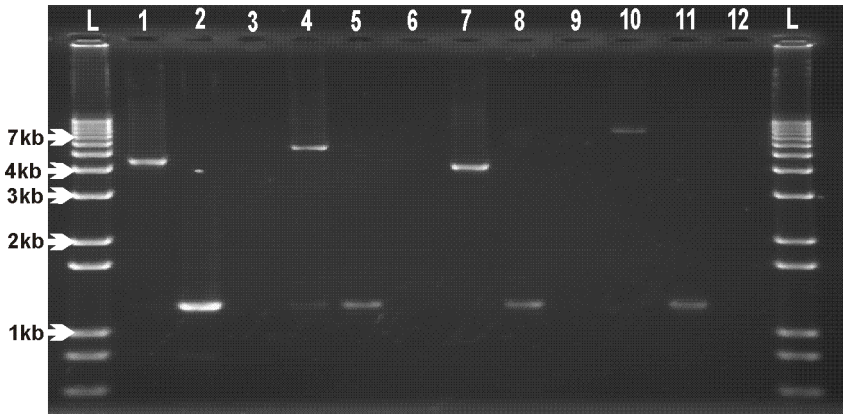
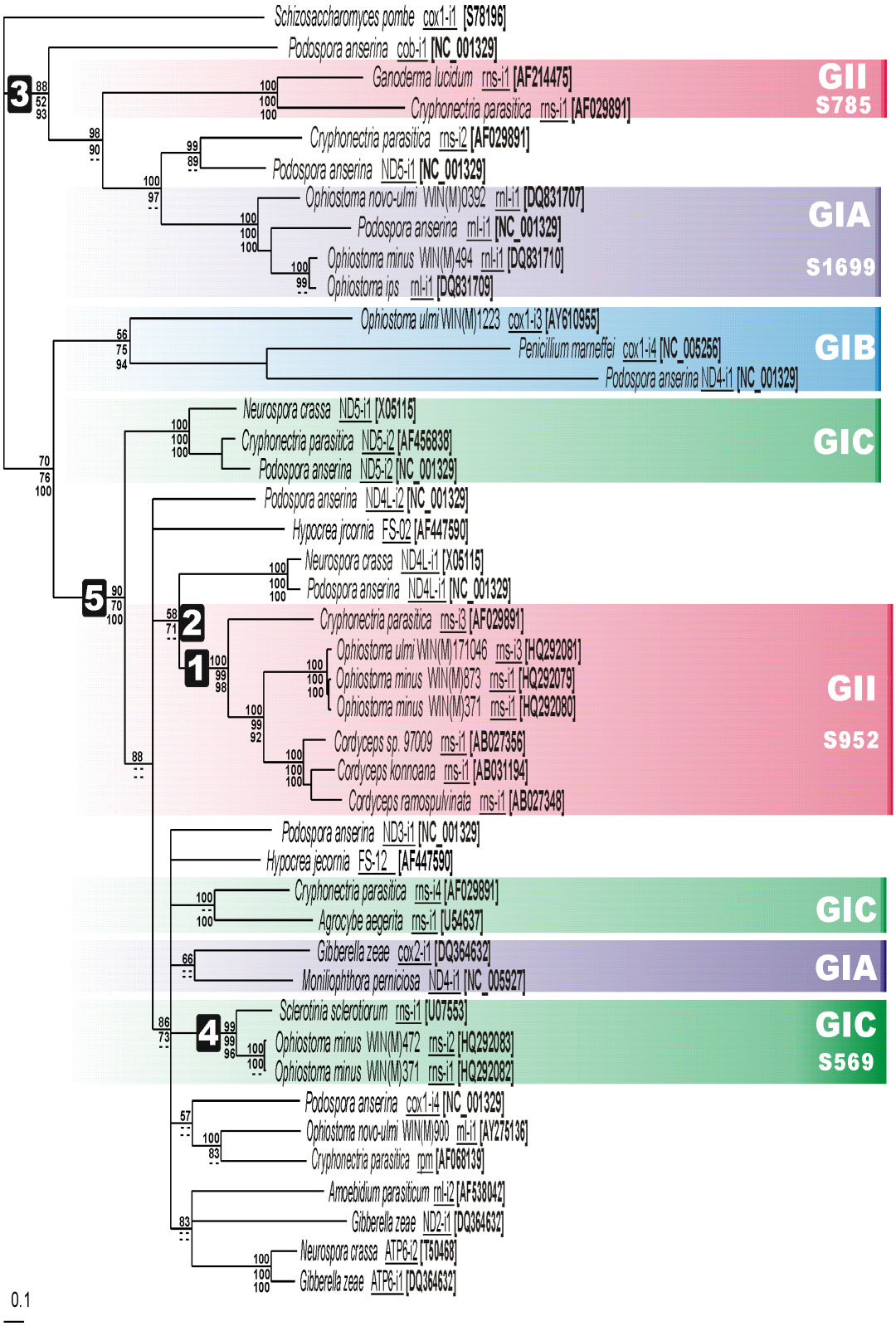


Figure 4.9. Phylogenetic tree of the LAGLIDADG HEG-like elements related to the *O. minus rns* group I and group II intron-encoded ORFs. The tree topology is based on Bayesian analysis and the numbers at the nodes represent the following: top number is the PP value as obtained from a 50% majority Bayesian consensus tree, the middle number is the bootstrap (BS) support based on NJ analysis, and the last number represents the BS support based on Parsimony analysis. Nodes that received less than 50% support (BS or PP) reduced to polychotomies. The branch lengths are based on Bayesian analysis and are proportional to the number of substitutions per site. Host gene and intron number are underlined. Sequence accession numbers are listed in brackets.

Figure 4.9.


Schizosaccharomyces pombe cox1-i1 [S78196]
Podospora anserina cob-i1 [NC_001329]
Ganoderma lucidum rns-i1 [AF214475]
Cryphonectria parasitica rns-i1 [AF029891]
GII
S785
Cryphonectria parasitica rns-i2 [AF029891]
Podospora anserina ND5-i1 [NC_001329]
Ophiostoma novo-ulmi WIN(M)0392 rnl-i1 [DQ831707]
Podospora anserina rnl-i1 [NC_001329]
Ophiostoma minus WIN(M)494 rnl-i1 [DQ831710]
Ophiostoma ips rnl-i1 [DQ831709]
GIA
S1699
Ophiostoma ulmi WIN(M)1223 cox1-i3 [AY610955]
Penicillium marneffei cox1-i4 [NC_005256]
Podospora anserina ND4-i1 [NC_001329]
GIB
Neurospora crassa ND5-i1 [X05115]
Cryphonectria parasitica ND5-i2 [AF456838]
Podospora anserina ND5-i2 [NC_001329]
GIC
Podospora anserina ND4L-i2 [NC_001329]
Hypocrea jecornia FS-02 [AF447590]
Neurospora crassa ND4L-i1 [X05115]
Podospora anserina ND4L-i1 [NC_001329]
Cryphonectria parasitica rns-i3 [AF029891]
Ophiostoma ulmi WIN(M)171046 rns-i3 [HQ292081]
Ophiostoma minus WIN(M)873 rns-i1 [HQ292079]
Ophiostoma minus WIN(M)371 rns-i1 [HQ292080]
GII
S952
Cordyceps sp. 97009 rns-i1 [AB027356]
Cordyceps konnoana rns-i1 [AB031194]
Cordyceps ramosopulvinata rns-i1 [AB027348]
Podospora anserina ND3-i1 [NC_001329]
Hypocrea jecornia FS-12 [AF447590]
Cryphonectria parasitica rns-i4 [AF029891]
Agrocybe aegerita rns-i1 [U54637]
GIC
Gibberella zeae cox2-i1 [DQ364632]
Moniliophthora perniciosa ND4-i1 [NC_005927]
GIA
Sclerotinia sclerotiorum rns-i1 [U07553]
Ophiostoma minus WIN(M)472 rns-i2 [HQ292083]
Ophiostoma minus WIN(M)371 rns-i1 [HQ292082]
GIC
S569
Podospora anserina cox1-i4 [NC_001329]
Ophiostoma novo-ulmi WIN(M)900 rnl-i1 [AY275136]
Cryphonectria parasitica rpm [AF068139]
Amoebidium parasiticum rnl-i2 [AF538042]
Gibberella zeae ND2-i1 [DQ364632]
Neurospora crassa ATP6-i2 [T50468]
Gibberella zeae ATP6-i1 [DQ364632]
0.1

There are other lineages of LAGLIDADG ORFs that appear to have invaded group II introns, see node 3 (Figure 4.9.). The latter appear to be unrelated to any of the *O. minus rns* intron ORFs. The *O. minus* mS569 intron ORFs appear to be related to another intron ORF inserted at the same position in the *rns* gene of *Sclerotinia sclerotiorum* (node 4, Figure 4.9.). Overall based on the limited number of sequences available, the ORF sequences evolved according to their respective insertion sites. Overall the *O. minus* intron ORFs are part of a large clade (node 5, Figure 4.9.) that includes free-standing ORFs and group I intron ORFs inserted in protein coding genes (*nad2, nad3, nad4, nad4l, nad5, cox1, cox2 and atp6*) as well as rRNA genes (*rns* and *rnl*).

4.3.6. Characterization of mS379 intron encoded RT ORFs:

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

acid sequence could be assembled that like those described previously, showed similarities a group II intron encoded RT-like ORFs

4.3.7. Phylogenetic analysis of group II intron encoded RT proteins:

In order to determine the evolutionary position of the *O. minus* mS379 ORFs among other group II intron encoded ORFs a data set was compiled by extracting related sequences from GenBank with a BLASTp search using the O.mi472-mS379 ORF amino acid sequence as a query (Figure 4.10.). The phylogenetic analysis of the aligned amino acid sequence data set composed of 63 sequences was restricted to segments that could be unambiguously aligned such as the RT domains (RT-0 to RT-7) and the X-maturase domains (Figure 4.11.) (Lambowitz and Zimmerly, 2004).

The data set contained 34 mitochondrial, six chloroplast and 23 bacterial group II intron related RT-like ORF sequences. The alignment consists of 571 amino acid positions. The non-Long Terminal Repeat element (non-LTR) from *Drosophila melanogaster* was used as an outgroup because this class of retroelements is mechanistically and phylogenetically linked to mobile group II introns (Eickbush, 1999). Phylogenetic trees generated from the RT sequence alignment placed the group II RT sequences into several groupings. The following clades were observed: the plant matR ORFs (node 1, Figure 4.10.), the euglenoid RT ORFs (node 2, Figure 4.10.), six bacterial subclasses (A, B, C, D, E and F; node 3, Figure 4.10.), and the mitochondrial RT ORFs (node 4, Figure 4.10.) (Martínez-Abarca and Toro, 2000; Zimmerly et al., 2001; Simon et al., 2008 & 2009).

Figure 4.10. A phylogenetic tree showing the relatedness among a set of group II intron RT ORFs. The tree topology is based on Bayesian analysis. The tree was rooted with non-LTR RT from *Drosophila melanogaster* (accession number: X51967.). Numbers at the nodes represented the following: upper number represents the PP values obtained from Bayesian 50% majority consensus tree, middle number the bootstrap support based on NJ analysis, and the last number is the BS support obtained from Parsimony analysis. Nodes that received less than 50% support (BS or PP) were collapsed into polychotomies. The major categories of RT ORF such as matR, euglenoid, mitochondrial and bacterial are indicated. Host gene and intron number are underlined. Sequence accession numbers are listed in brackets.

Figure 4.10.

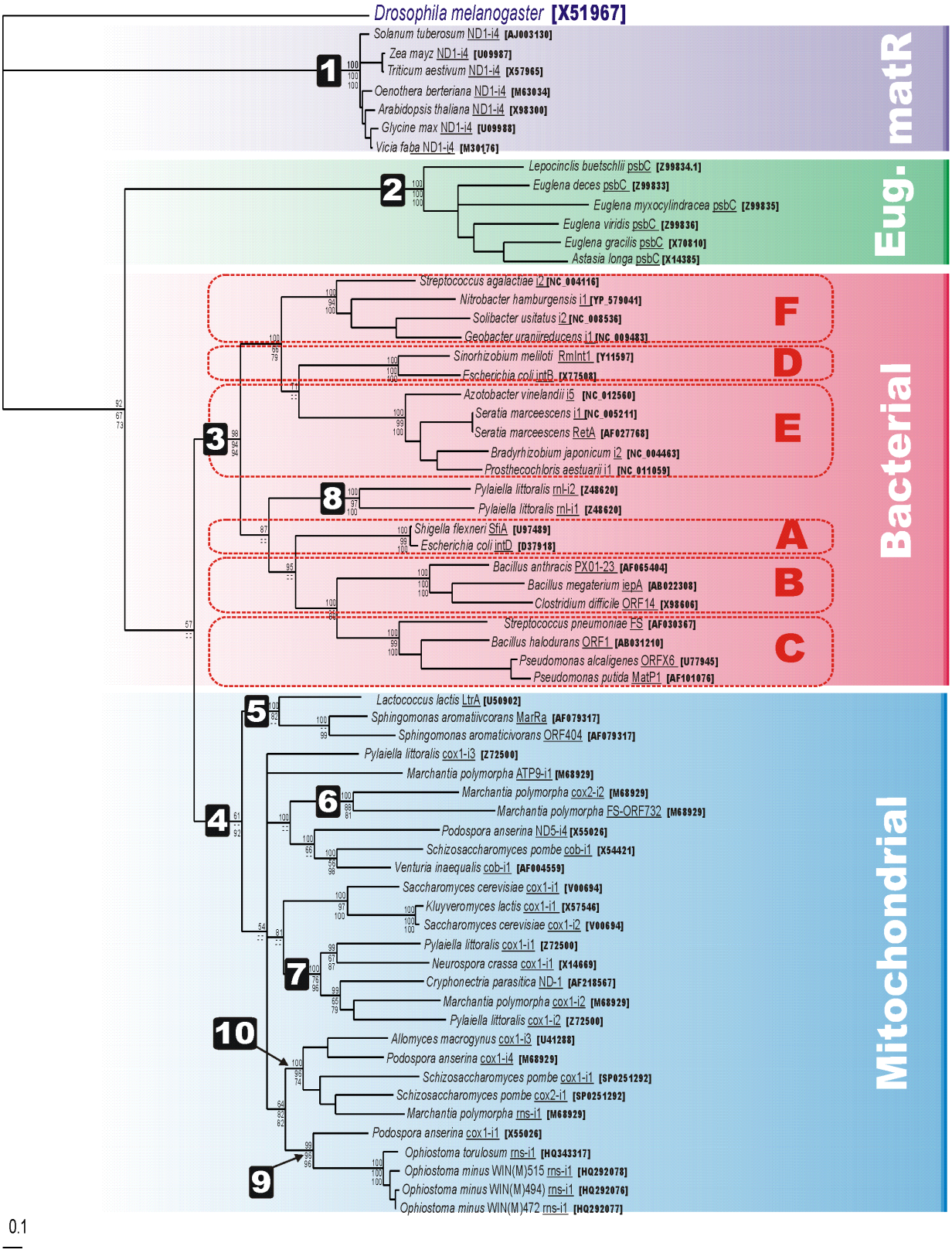
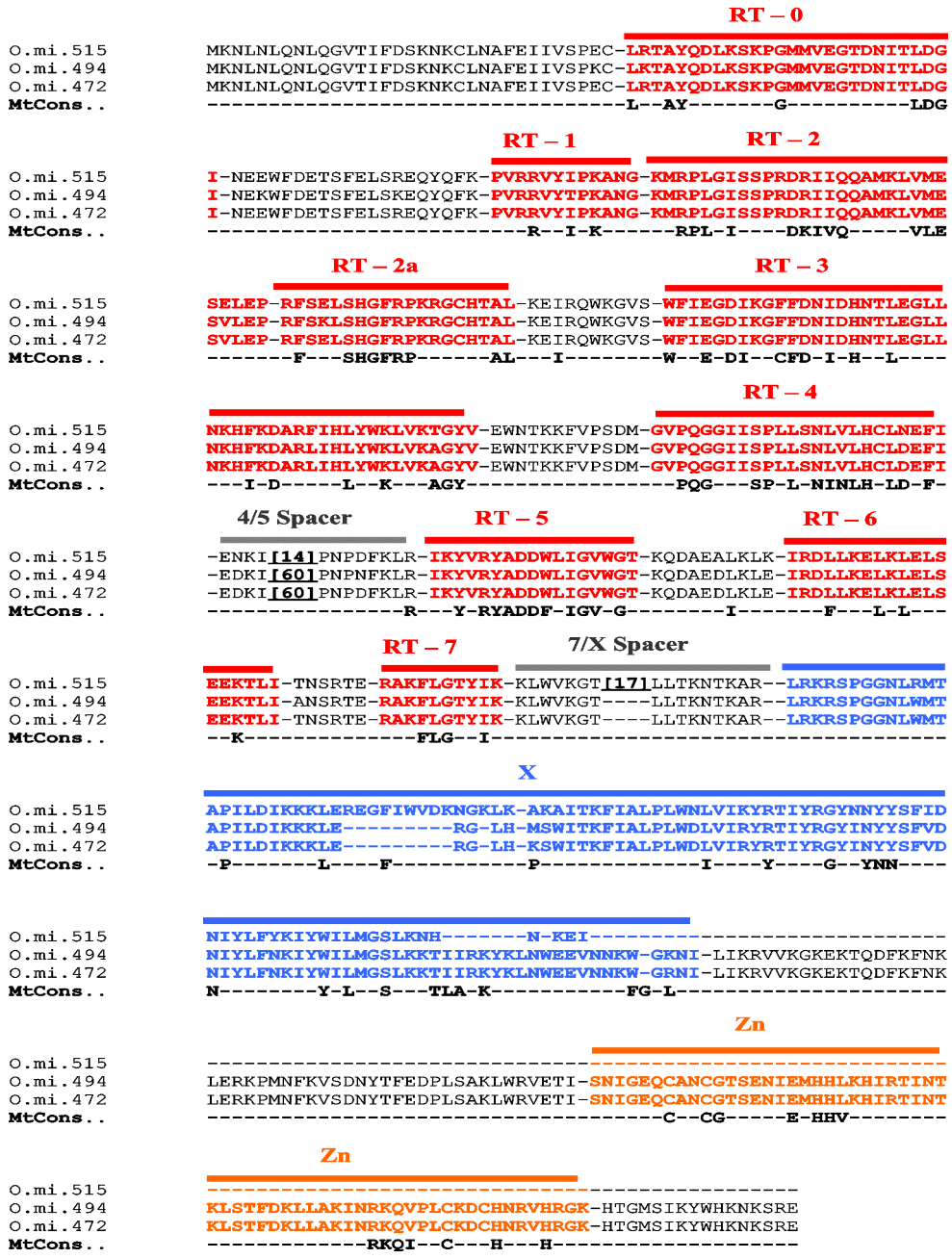

Drosophila melanogaster [X51967]
Solanum tuberosum ND1-i4 [AJ003130]
Zea mayz ND1-i4 [U09987]
Triticum aestivum ND1-i4 [X57965]
Oenothera berteriana ND1-i4 [M63034]
Arabidopsis thaliana ND1-i4 [X98300]
Glycine max ND1-i4 [U09988]
Vicia faba ND1-i4 [M30176]
matR
Lepocinclis buetschlii psbC [Z99834.1]
Euglena deces psbC [Z99833]
Euglena myxocylindracea psbC [Z99835]
Euglena viridis psbC [Z99836]
Euglena gracilis psbC [X70810]
Astasia longa psbC [X14385]
Eug.
Streptococcus agalactiae i2 [NC_004116]
Nitrobacter hamburgensis i1 [YP_579041]
Solibacter usitatus i2 [NC_008536]
Geobacter uraniireducens i1 [NC_009483]
F
Sinorhizobium meliloti RmInt1 [Y11597]
Escherichia coli intB [X77508]
D
Azotobacter vinelandii i5 [NC_012560]
Seratia marceescens i1 [NC_005211]
Seratia marceescens RetA [AF027768]
Bradyrhizobium japonicum i2 [NC_004463]
Prosthecochloris aestuarii i1 [NC_011059]
E
Pylaiella littoralis rnl-i2 [Z48620]
Pylaiella littoralis rnl-i1 [Z48620]
Shigella flexneri SfiA [U97489]
Escherichia coli intD [D37918]
A
Bacillus anthracis PX01-23 [AF065404]
Bacillus megaterium iepA [AB022308]
Clostridium difficile ORF14 [X98606]
B
Streptococcus pneumoniae FS [AF030367]
Bacillus halodurans ORF1 [AB031210]
Pseudomonas alcaligenes ORFX6 [U77945]
Pseudomonas putida MatP1 [AF101076]
C
Bacterial
Lactococcus lactis LtrA [U50902]
Sphingomonas aromatiivcorans MarRa [AF079317]
Sphingomonas aromaticivorans ORF404 [AF079317]
Pylaiella littoralis cox1-i3 [Z72500]
Marchantia polymorpha ATP9-i1 [M68929]
Marchantia polymorpha cox2-i2 [M68929]
Marchantia polymorpha FS-ORF732 [M68929]
Podospora anserina ND5-i4 [X55026]
Schizosaccharomyces pombe cob-i1 [X54421]
Venturia inaequalis cob-i1 [AF004559]
Saccharomyces cerevisiae cox1-i1 [V00694]
Kluyveromyces lactis cox1-i1 [X57546]
Saccharomyces cerevisiae cox1-i2 [V00694]
Pylaiella littoralis cox1-i1 [Z72500]
Neurospora crassa cox1-i1 [X14669]
Cryphonectria parasitica ND-1 [AF218567]
Marchantia polymorpha cox1-i2 [M68929]
Pylaiella littoralis cox1-i2 [Z72500]
Allomyces macrogynus cox1-i3 [U41288]
Podospora anserina cox1-i4 [M68929]
Schizosaccharomyces pombe cox1-i1 [SP0251292]
Schizosaccharomyces pombe cox2-i1 [SP0251292]
Marchantia polymorpha rns-i1 [M68929]
Podospora anserina cox1-i1 [X55026]
Ophiostoma torulosum rns-i1 [HQ343317]
Ophiostoma minus WIN(M)515 rns-i1 [HQ292078]
Ophiostoma minus WIN(M)494) rns-i1 [HQ292076]
Ophiostoma minus WIN(M)472 rns-i1 [HQ292077]
Mitochondrial
0.1

Figure 4.11. Alignment of the RT ORF amino acid sequences of the *O. minus* (strains WIN(M) 515, 494 and 472) and the consensus sequences of the mitochondrial group II type RT ORFs (as given in Zimmerly et al., 2001). Conserved RT domains (RT-0 to RT-7, X and Zn) are indicated above the alignment.

Figure 4.11.


RT – 0
O.mi.515 MKNLNLQNLQGVTIFDSKNKCLNAFEIIVSPEC-LRTAYQDLKSKPGMMVEGTDNITLDG
O.mi.494 MKNLNLQNLQGVTIFDSKNKCLNAFEIIVSPKC-LKTAYQDLKSKPGMMVEGTDNITLDG
O.mi.472 MKNLNLQNLQGVTIFDSKNKCLNAFEIIVSPEC-LRTAYQDLKSKPGMMVEGTDNITLDG
MtCons.. ----------L--AY-------G----------LDG
RT – 1
RT – 2
O.mi.515 I-NEEWFDETSFELSREQYQFK-PVRRVYIPKANG-KMRPLGISSPRDRIIQQAMKLVME
O.mi.494 I-NEKWFDETSFELSKEQYQFK-PVRRVYTPKANG-KMRPLGISSPRDRIIQQAMKLVME
O.mi.472 I-NEEWFDETSFELSREQYQFK-PVRRVYIPKANG-KMRPLGISSPRDRIIQQAMKLVME
MtCons.. ----------R--I-K------RPL-I----DKIVQ-----VLE
RT – 2a
RT – 3
O.mi.515 SELEP-RFSELSHGFRPKRGCHTAL-KEIRQWKGVS-WFIEGDIKGFFDNIDHNTLEGLL
O.mi.494 SVLEP-RFSKLSHGFRPKRGCHTAL-KEIRQWKGVS-WFIEGDIKGFFDNIDHNTLEGLL
O.mi.472 SVLEP-RFSELSHGFRPKRGCHTAL-KEIRQWKGVS-WFIEGDIKGFFDNIDHNTLEGLL
MtCons.. -------F---SHGFRP------AL---I--------W--E-DI--CFD-I-H--L----
RT – 4
O.mi.515 NKHFKDARFIHLYWKLVKTGYV-EWNTKKFVPSDM-GVPQGGIISPLLSNLVLHCLNEFI
O.mi.494 NKHFKDARLIHLYWKLVKAGYV-EWNTKKFVPSDM-GVPQGGIISPLLSNLVLHCLDEFI
O.mi.472 NKHFKDARLIHLYWKLVKAGYV-EWNTKKFVPSDM-GVPQGGIISPLLSNLVLHCLDEFI
MtCons.. ---I-D-----L--K---AGY----------PQG---SP-L-NINLH-LD-F-
4/5 Spacer
RT – 5
RT – 6
O.mi.515 -ENKI[14]PNPDFKLR-IKYVRYADDWLIGVWGT-KQDAEALKLK-IRDLLKELKLELS
O.mi.494 -EDKI[60]PNPNFKLR-IKYVRYADDWLIGVWGT-KQDAEDLKLE-IRDLLKELKLELS
O.mi.472 -EDKI[60]PNPDFKLR-IKYVRYADDWLIGVWGT-KQDAEDLKLE-IRDLLKELKLELS
MtCons.. ----------R---Y-RYADDF-IGV-G--------I-------F---L-L---
RT – 7
7/X Spacer
O.mi.515 EEKTLI-TNSRTE-RAKFLGTYIK-KLWVKGT[17]LLTKNTKAR--LRKRSPGGNLRMT
O.mi.494 EEKTLI-ANSRTE-RAKFLGTYIK-KLWVKGT----LLTKNTKAR--LRKRSPGGNLWMT
O.mi.472 EEKTLI-TNSRTE-RAKFLGTYIK-KLWVKGT----LLTKNTKAR--LRKRSPGGNLWMT
MtCons.. --K----------FLG--I-----------------------------------
X
O.mi.515 APILDIKKKLEREGFIWVDKNGKLK-AKAITKFIALPLWNLVIKYRTIYRGYNNYYSFID
O.mi.494 APILDIKKKLE---------RG-LH-MSWITKFIALPLWDLVIRYRTIYRGYINYYSFVD
O.mi.472 APILDIKKKLE---------RG-LH-KSWITKFIALPLWDLVIRYRTIYRGYINYYSFVD
MtCons.. -P-------L----F----------P----------I----Y----G--YNN----
O.mi.515 NIYLFYKIYWILMGSLKNH-------N-KEI---------
O.mi.494 NIYLFNKIYWILMGSLKKTIIRKYKLNWEEVNNKW-GKNI-LIKRVVKGKEKTQDFKFNK
O.mi.472 NIYLFNKIYWILMGSLKKTIIRKYKLNWEEVNNKW-GRNI-LIKRVVKGKEKTQDFKFNK
MtCons.. N--------Y-L--S---TLA-K----------FG-L----------
Zn
O.mi.515 ----------
O.mi.494 LERKPMNFKVSDNYTFEDPLSAKLWRVETI-SNIGEQCANCGTSENIEMHHLKHIRTINT
O.mi.472 LERKPMNFKVSDNYTFEDPLSAKLWRVETI-SNIGEQCANCGTSENIEMHHLKHIRTINT
MtCons.. ----------C--CG-----E-HHV--------
Zn
O.mi.515 ----------
O.mi.494 KLSTFDKLLAKINRKQVPLCKDCHNRVHRGK-HTGMSIKYWHKNKSRE
O.mi.472 KLSTFDKLLAKINRKQVPLCKDCHNRVHRGK-HTGMSIKYWHKNKSRE
MtCons.. ----------RKQI--C---H---H----------

We noted examples of possible horizontal transfer of RT ORFs between unrelated groups of organisms. For example among fungal mitochondrial group II intron ORFs one finds interspersed ORFs from bacterial (node 5, Figure 4.10.), liverwort (node 6, Figure 4.10.) and brown algal sources. For example the *cox1*-i1 and *cox1*-i2 RT ORFs from the *Pyaiella littoralis* (a brown algae, stramenopiles) group with the *Neurospora crassa cox1*-i1 and *Marchantia polymorpha* cox1-i2 RT ORFs respectively (node 7, Figure 4.10.). Also the *Pyaiella littoralis rnl*-i1 and *rnl*-i2 RT ORFs grouped within the bacterial domain of the RT ORF tree (node 8, Figure 4.10.). The heterogeneous distribution of closely related RT ORFs in different taxonomically unrelated groups is indicative of horizontal transfers.

The O.mi-mS379 ORFs formed a sister group with the P.an.cox1-i1 RT (node 9, Figure 4.10.) and the node support values are 96 and 99 % for bootstrap support and posterior probability respectively, and this cluster shares a common node with a set of group II intron ORFs mostly inserted within the *cox*1 gene (*Allomyces macrogynus* cox1-i3, *Podospora anserina* cox1-i4, *Schizosaccharomyces pombe* cox1-i1 and cox1-i2) or the *Marchantia polymorpha rns* gene (*rns*-i1) (node 10, Figure 4.10.).

4.4. DISCUSSION:

4.4.1. *Ophiostoma minus rns* gene has two IC2 Group I intron insertion sites:

The secondary structure of group I introns consist of base-paring elements designated as P1 to P9. Another conserved structure is P10 which forms by base pairing

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

4.4.2. Group II intron:

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

ORFs, encoded within a potentially mobile element, there are numerous examples of horizontal transfer within and between the different groups of organisms (Zimmerly et al., 2001).

The O.mi-mS379 RT ORFs formed a sister group with the P.an.cox1-i1 RT suggesting the transfer of a group II intron from a protein coding gene into an rRNA gene. Typically mobile introns invaded cognate intronless alleles; however group II intron RNAs have been shown to retrotranspose by reverse splicing into RNA molecules (Bonen and Vogel, 2001), a mechanism that requires less specificity at the target site and thus allows for retrotransposition of introns to new sites within the genome (Lambowitz and Zimmerly, 2010). The *Marchantia polymorpha* (M.po.) *rns*-i1 RT ORF failed to group closely with the O. mi-mS379 ORFs; however the M.po-*rns*-i1 is inserted between positions S474/S475 (based on the *E. coli* reference sequence). Thus this group II intron represents a different lineage than the S379 introns.

4.4.3. Group II introns with RT-ORF embedded within domain II:

Group II introns have conserved secondary structures at the RNA level, that can be visualized as six stem-loop domains (DI to DVI) emerging from a central wheel (Michel and Ferat, 1995). When RT-like ORFs are present they typically are embedded within DIV. However in some bacterial group II introns, RT type ORFs have been observed in DII (Simon et al., 2008) and some LAGLIDADG type ORFs are inserted in DIII or DIV (Toor and Zimmerly, 2002). Some group II intron encoded proteins extend upstream and are fused to the upstream exon; this results in the generation of a fusion protein upon

translation which probably is resolved by proteolysis (Michel and Ferat, 1995). Here we report the first example of a mitochondrial group II intron that has the entire ORF embedded within domain II.

Initially, attempts were made to fold mS379 in a mode that would position the ORF into DIV, but that approach prevented recovering structures that contained the characteristic secondary and tertiary interactions expected for group II introns. By moving the ORF into DII a secondary structure could be generated that allowed for the recognition of the majority of secondary (DI to DVI) and tertiary (ε, λ, α, β, θ, κ, ζ, and γ) interactions expected for group II introns. It is quite possible that DII can carry additional insertions like the RT-ORF because once DII is locked into position by the tertiary interactions (θ- θ^{-}; see Figure 4.6.) and a series of contacts with the S-turn in DIII, the loop of DII (that contains the ORF) projects away from the ribozyme core (Pyle, 2010).

One can only speculate on how the ORF shifted into the DII position; maybe the ancestral version of this intron was encoded within a protein coding gene and the RT-ORF extended from DIV upstream and was fused to the 5′ flanking exon. A configuration seen in the yeast cox1 group II intron ORFs (Lambowitz and Zimmerly, 2004) suggested that, overtime, sequences downstream of the ORF position changed and could substitute for DIII thus allowing the ORF to "slide" into the DII position. Alternative sequences could fold into a DIV configuration, and DV, the most conservative component of the group II ribozyme, remained the same. As group II introns appear to coevolve with their encoded RT ORFs (Toor et al., 2001) the above "ORF

sliding" model is a possibility as the group II RT phylogeny shows that the O.mi-mS379 ORFs are related to the *Podospora anserina cox1*i1 group II intron ORF, an RT ORF that is fused to the upstream exon as seen in the yeast *cox1* introns (Mohr et al., 1993; Lambowitz and Belfort, 1993).

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

4.4.4. LAGLIDADG type ORFs in group I and group II introns:

Group II introns that have LAGLIDADG type ORFs such as the mS952 intron have been described from several fungal taxa (Toor and Zimmerly, 2002; Monteiro-Vitorello et al., 2009; Mullineux et al., 2010) suggesting that these introns are wide spread. HEGs are quite invasive and can invade sites that are neutral, thus avoiding toxicity to the host

genome. Therefore, self-splicing introns are ideal targets for invasion by HEGs. Most likely the mS952 intron arose by the HEG originated from a group I intron source invading an ORFless group II intron. Recently Mullineux et al. (2010) demonstrated that in a *Leptographium* species the mS952 group II intron encoded HEG expresses a functional HE that cleaves the *rns* gene two nucleotides upstream of the mS952 position in an intronless allele, thus potentially facilitating the mobility of its host intron by a mechanism that is similar to that for group I introns.

During this study we also encountered group I encoded LAGLIDADG type ORFs that had degenerated (mS1224) due the accumulation of mutations that introduced frame shifts and premature stop codons. This is expected as many HEGs are viewed as neutral elements that are not subject to selection; thus they can accumulate non-adaptive mutations that eventually will lead to their complete loss from the genome (Goddard and Burt, 1999). Thus in order for HEGs to persist they have to continuously invade new sites either via vertical or horizontal transfers or gain new functions that might make them an essential component of the genome (Gogarten and Hilario, 2006).

4.4.5. Evolutionary dynamics of LHEase and RT ORFs:

Based on recent studies (Haugen and Bhattacharya, 2004) it is thought that single motif LHEases that inserted into a preexisting group I intron in the *rnl* gene between positions L1917 and L1951 evolved into double motif LHEases and then in combination with their group I intron host spread into sites first within the *rnl* gene and subsequently invaded new sites in the *rns* gene as well as protein coding genes (see Figure 4.12.). The

mS569 ORF belongs to a clade of intron ORFs that are encoded in introns located in rDNA and protein coding genes a distribution predicted by Haugen and Bhattacharya (2004) model for the dispersal of rDNA introns into other rDNA sites and protein coding loci.

Figure 4.12. Evolutionary model for the spread of LAGLIDADG ORFs. A schematic diagram that summarizes findings by Haugen and Bhattacharya (2004) suggesting that *rnl* single motif LAGLIDADG ORFs represent the oldest intron ORFs and after the evolution of double motif LAGLIDADG ORF via a gene duplication and fussion event (red thin arrow); the introns spread into other rDNA sites and into protein coding genes (black thin arrows). Our phylogenetic data (Figure 4.9.) would suggest that the *O. minus* group I introns represent a reinvasion of rDNA by group I introns originating within protein coding genes. The mS952 group II intron ORF appears to be an example of a "host switch" where a group I intron encoded double motif LAGLIDADG ORF moved from a group I intron into a group II introns in the *rns* and *rnl* genes (thick blue arrows).

Figure 4.12.

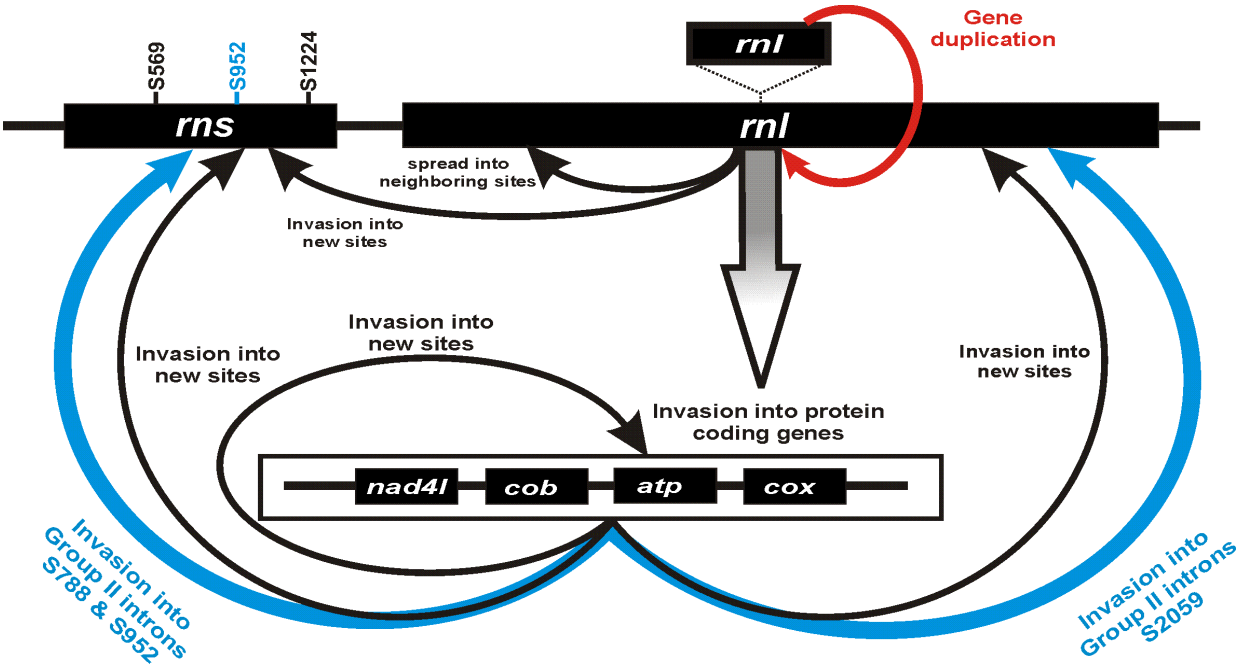

S569
S952
S1224
rns
rnl
rnl
Gene duplication
spread into neighboring sites
Invasion into new sites
Invasion into new sites
Invasion into new sites
Invasion into new sites
Invasion into protein coding genes
nad4l
cob
atp
cox
Invasion into Group II introns S788 & S952
Invasion into Group II introns S2059

Figure 4.13. Evolutionary model for the spread of RT ORFs. The spread of Group II RT ORFs and changes within the structure of the RT ORFs. Models include recent findings by Zimmerly et al. (2001) and Simon et al. 2009. Figure 4.13. adapted from Zimmerly et al. (2001). The RT component of the group II intron ORF can be divided into 8 subdomains (0-7) which is followed by the X and Zn (also referred to as En) domains. The reverse transcriptase activity is characterized by the YADD amino acid sequence motif contained within subdomain 5. The X (also referred to as M) domain provides maturase activity and the En (or Zn domain) provides endonuclease activity; there is a D domain (not shown) located between the X and Zn domains in some bacterial group II introns and it is associated with DNA binding (Lambowitz and Zimmerly, 2004 & 2010). The aspect of this model that applies to this work is the dynamics of the fungal mtDNA group II intron ORFs, as shown in the phylogeny in Figure 4.10. These introns horizontally transfer (H.T. = blue arrows) among fungi, between brown algae and fungi and potentially from fungi to bacteria. In euglenoid chloroplasts and mitochondria of higher plants the intron-ORF unit appear to have lost mobility and transfers vertically (V.T. = orange arrows). The *O. minus* mS379 group II intron ORF, however, appears to be degenerating due to the presence of frame shift mutations suggesting it lost its mobility.

Figure 4.13.

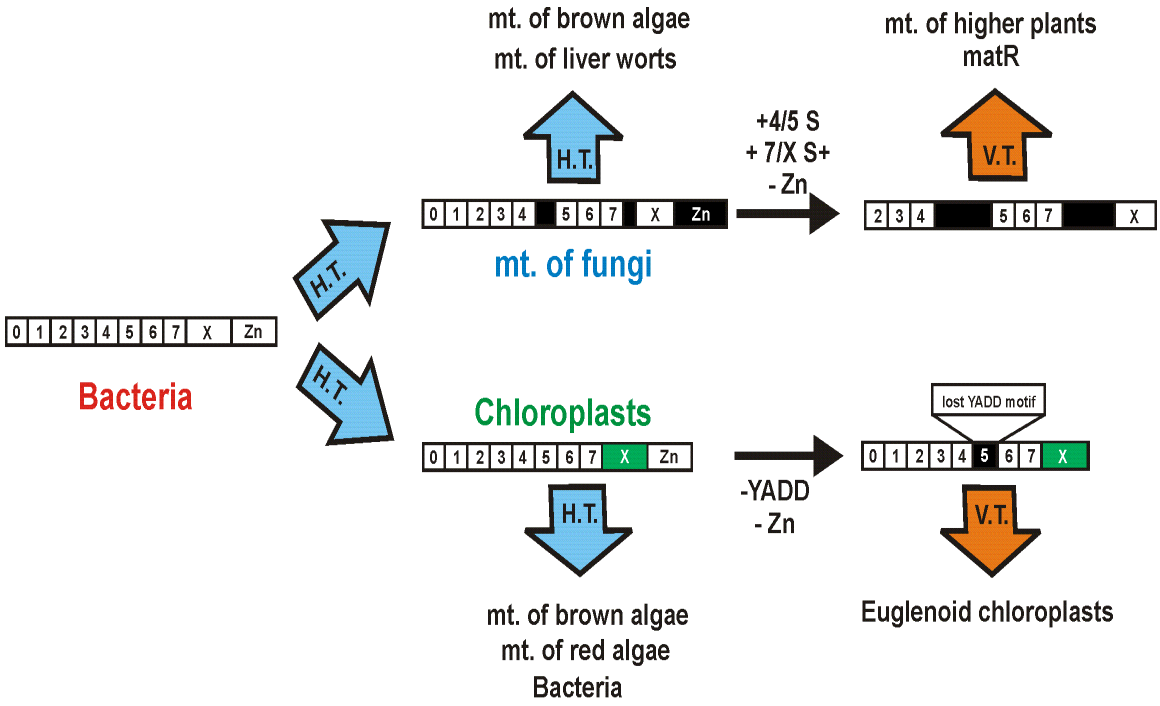
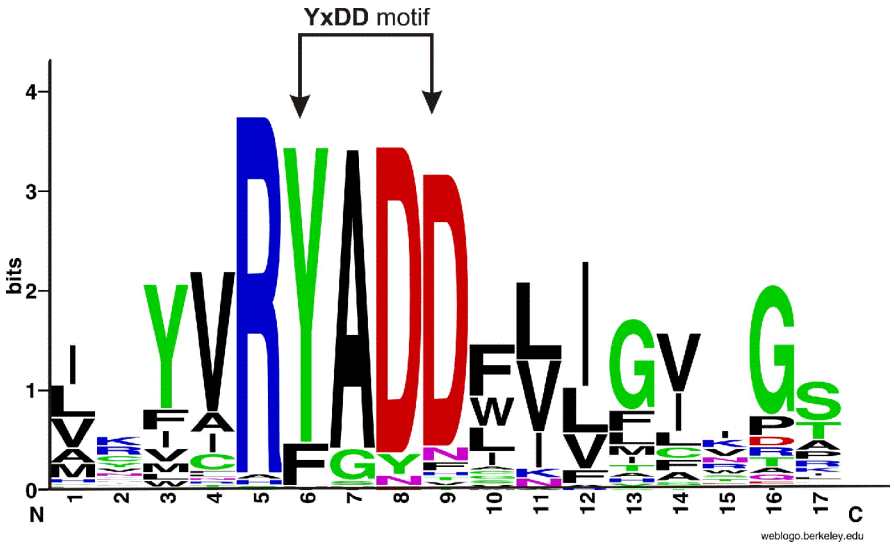

mt. of brown algae
mt. of liver worts
mt. of higher plants
matR
H.T.
+4/5 S
+ 7/X S+
- Zn
V.T.
0 1 2 3 4 5 6 7 X Zn
2 3 4 5 6 7 X
mt. of fungi
H.T.
0 1 2 3 4 5 6 7 X Zn
Bacteria
H.T.
Chloroplasts
lost YADD motif
0 1 2 3 4 5 6 7 X Zn
-YADD
- Zn
0 1 2 3 4 5 6 7 X
H.T.
V.T.
mt. of brown algae
mt. of red algae
Bacteria
Euglenoid chloroplasts

Figure 4.14. Amino acid sequence logo of the reverse transcriptase subdomain 5 (RT-5). The alignment analyzed consisted of 64 RT ORFs originating form mitochondrial, chloroplast and bacterial sources. The logo identifies the YxDD conserved motif which was replaced by the FxDD motif in some mitochondrial RT ORFs and lost in the euglenoid RT ORFs. The logo was generated by the online program WebLogo version 2.8.2. (Crooks et al., 2004).

Figure 4.14.


YxDD motif
bits
0
1
2
3
4
N
C
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
weblogo.berkeley.edu

Thus, the double motif LHEases in combination with a group I intron invaded protein coding genes and also transferred back to rRNA genes such as we observed for the mS569 or mS1224 introns and their ORFs. Based on those previous studies and our observations we also suggest that one lineage of LHEases moved independently from its original ribozyme partner (most likely a group I intron) and transferred into DIII of an ORF-less preexisting group II introns located at positions S788 and S952 (Figure 4.12.).

Based on more exhaustive studies by Zimmerly et al. (2001) and Simon et al. (2009) it appears that group II RT ORFs probably originated in bacteria and extensively transferred horizontally (see Figure 4.13.). The group II introns and their RT ORFs were probably introduced to eukaryotes via the endosymbionts that gave rise to mitochondria and chloroplast. The chloroplast RT ORFs appear to spread horizontally but in the euglenoid chloroplast the RT-ORF YADD amino acid motif (see Figure 4.14.), and the Zn domains were lost and thus the intron-ORF unit appear to have lost mobility and transfers vertically (Zimmerly et al., 2001; Simon et al., 2009). It has been noted in previous reports that within the mitochondrial lineage the ORFs also underwent changes (Zimmerly et al. 2001; Figure 4.13.) and evidence for both vertical and horizontal transmission has been presented in this and other studies (Zimmerly et al., 2001; Lambowitz and Zimmerly, 2004) showing that these elements, like other mobile elements, persist by continuously invading new sites. For example the O.mi-mS379 intron was probably derived from a group II intron that most likely retrotransposed from the *cox1* gene into the *rns* gene. However, the current ORF configuration in the O.mi-

mS379 intron suggests degeneration and most likely this element has lost its mobility and is therefore vertically transferred.

4.5. CONCLUSIONS:

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

CHAPTER: 5

The mS379 INTRON: A NOVEL GROUP IIA1 INTRON INTERRUPTS THE MITOCHONDRIAL *RNS* GENE OF SOME SPECIES OF *Ophiostoma*.

5.1. ABSTRACT:

During a survey of mitochondrial *rns* genes for ophiostomatoid fungi, a novel group II intron was found at position S379. Secondary structure modeling of the intron sequences indicates that it is a group IIA1 with an eroded RT-like ORF. The mS379 intron is novel as it represents the first mitochondrial group II intron that has an RT-ORF encoded outside Domain IV and it is the first intron reported at position S379. Based on the limited distribution of the mS379 intron among species of *Ophiostoma* and other fungi and signs of ORF degeneration or loss, it is suggested that this intron might represent a "snap shot" of a mobile intron that is in the process of elimination.

5.2. INTRODUCTION:

Fungal mitochondrial genomes are highly variable in size and this is in part due to the presence of self-splicing group I and group II introns (Michel and Ferat, 1995; Hausner, 2012). Group II introns represent one category of autocatalytic introns that are widespread in the organellar genomes of plants, fungi and algae as well as bacterial genomes (Lambowitz and Zimmerly, 2004 & 2011). Group II introns are essentially large ribozymes that can catalyze their removal from the primary transcripts to allow for efficient gene expression (Lehmann and Schmidt, 2003), although many group II introns require intron-encoded (maturases) and/or host genome-encoded protein factors to facilitate the splicing process (Ostheimer et al., 2003; Hausner, 2012). Group II introns potentially influence important aspects of organellar gene expression (Bonen and Vogel, 2001; Hausner et al., 2006; Petersen et al., 2011; Hausner, 2012); also group II introns have been associated with mitochondrial genome instabilities in fungi such as senescence in *Podospora anserina* (Dujon, 1989; Griffiths, 1992) and mitochondrial induced hypovirulence in *Cryphonectria parasitica* (Baidyaroy et al., 2011).

During a recent study (Chapter 4) we noted that in strains of *Ophiostoma minus* (Hedgc.) Syd. & P. Syd. the mtDNA *rns* gene appears to be a reservoir for group I and group II introns (Hafez and Hausner, 2011a). In particular we noted a novel group II intron, referred to as mS379 based on its *rns* insertion site with respect to the *E. coli* SSU rRNA gene. Here we examine additional examples of the mS379 intron from *Ophiostoma torulosum* (Butin & G. Zimm.) G. Hausner, J. Reid & Klassen and *Ophiostoma hyalothecium* (R.W. Davidson) G. Hausner, J. Reid & Klassen. Many species of

Ophiostoma are blue-stain fungi that can discolor timber and thus reduce its economic value. So far very little information is available on their mitochondrial genomes and their intron complements. The mS379 intron represents a new intron/ORF arrangement for organellar group II introns and therefore warrants more detailed characterization.

Group II introns show limited conservation at the DNA sequence level but they are defined by highly conserved secondary structures; at the RNA level they can be visualized as six double-helical domains (Domains I to Domain VI, i.e. DI to DVI) radiating from a central wheel (Michel and Ferat, 1995). Based on features of their secondary structures, group II introns can be assigned to three subgroups: IIA, IIB and the bacterial classes IIC/IIE. Based on variants within the IIA and IIB structures several subdivisions have been designated such as IIA1, IIA2, IIB1 and IIB2 (Michel et al., 1989; Michel and Ferat, 1995; Michel et al., 2007 & 2009; Toor et al., 2001; Lambowitz and Zimmerly, 2011).

Domain I (DI) is the largest group II intron domain and it is essential for self-splicing activity as it is involved in exon recognition and in other tertiary interactions. Domain I is transcribed and folded first, thus forming the basic scaffold for the assembly of catalytic and ORF-containing domains that follow the DI sequence (Qin and Pyle, 1998; Pyle et al., 2007; Steiner et al., 2008). Domain II (DII) forms two essential tertiary contacts with DI (θ–θ′) and DVI (η–ή); DII contains a conserved internal loop structure that is important in stimulating the catalytic activity of group II introns (Fedorova and Pyle, 2008). Less is known about the function of DIII (Fedorova and Zingler, 2007;

Federova et al., 2010) but this domain can house ORFs such as LHEase ORFs (Toor and Zimmerly, 2002; Mullineux et al., 2010).

Domain IV usually projects away from the intron's catalytic center and contains primary binding sites for maturases and/or other intron splicing factors that promote splicing reactions (Wank et al., 1999). Domain IV can encode ORFs that contribute to intron mobility and splicing (Lambowitz and Zimmerly, 2004). Typically group II IEPs are multifunctional proteins containing the following domains: (1) reverse transcriptase (RT), (2) maturase (X), (3) DNA binding (D) and the (4) endonuclease (En) domain. The IEP binds to the intron RNA to stabilize the catalytically active structure. Domain V is the most conserved domain; and DVI contains the bulged adenosine nucleotide with the 2' OH group that initiates the splicing reaction and this adenine is the branch point that allows for the 5'-2' linkage that stabilizes the intron lariat upon its release (Michel and Ferat, 1995; Lehmann and Schmidt, 2003; Pyle and Lambowitz, 2006; Michel et al., 2009). Although RT-like ORFs, if present, are typically encoded within a loop in domain IV (Lambowitz and Zimmerly, 2011) there are a few instances where RT-like ORFs have been observed in domain II or domain III of some bacterial group IIE1 introns (Simon et al., 2008).

In this study we examine in detail the S379 group II intron that encodes an RT-like ORF within DII. These RT-like ORFs demonstrates a cycle of degeneration typically observed within group I intron encoded ORFs (Goddard and Burt, 1999) except that we

could not find examples for ORF rejuvenation via reinvasion of the intron into intronless alleles.

5.3. METHODS OVERVIEW:

Two species of *Ophiostoma* were examined in this study and they were *Ophiostoma hyalothecium* [ATCC 28825 (from the American Type Culture Collection; Manassas, VA, USA); WIN(M)848 (from the University of Manitoba culture collection; Winnipeg, MB, Canada); geographic origin: Wyoming, USA] and *Ophiostoma torulosum* [CBS770.71 (from the Central Bureau voos Schimmelcultures; Utrecht, The Netherlands); WIN(M)860; geographic origin: Hanover-Münden, Germany]. Methods for DNA extraction and PCR amplification of the *rns* gene (with primers mtsr-1 and mtsr-2) were previously described in Hafez and Hausner (2011a). The mS379 intron was amplified from different strains with the primer pair mtsr-1 and S379-R that flanking the intron (see Appendix 9.3. for complete primers list).

The PCR amplicons were purified with the Wizard® SV Gel and PCR Clean-Up system and cloned into *E. coli* (DH5α) using the TOPO TA Cloning® kit (Invitrogen) to improve sequencing efficiency. The FindModel program was used to evaluate the nucleotide sequence alignments, the GTR model with gamma distribution was identified as the best model. The MrBayes program was used for Bayesian analysis of the *rns* data. The 50% majority rule tree generated by MrBayes program based on the aligned *rns* data set was used as a base for reconstruction of the ancestral states. Estimation of the

ancestral character state for the presence or absence of the mS379 intron among *Ophiostoma* species was done with the MESQUITE program version 2.73. Secondary structure models of the S379 intron and DII were developed using *mfold* program (Zuker 2003). Finally, the *in vivo* splicing of mS379 intron was determined by RT-PCR as described in Chapter 2.

5.4. RESULTS:

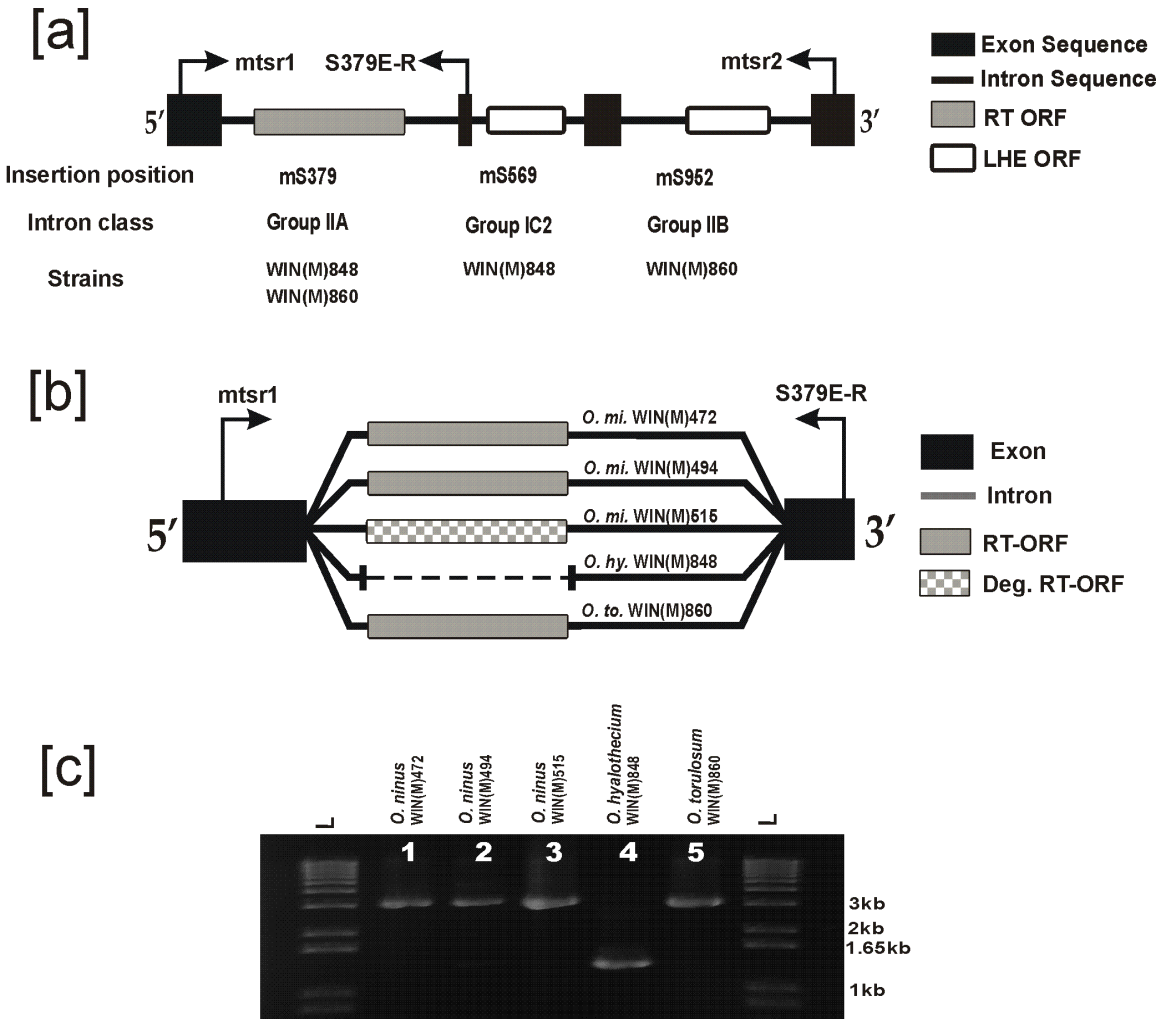
5.4.1. The *rns* gene and the mS379 intron in *O. hyalothecium* and *O. torulosum*:

For naming introns we followed the nomenclature proposed by Johansen and Haugen (2001), and intron insertion sites are based on corresponding nucleotide positions within the *E. coli* SSU rRNA sequence (accession number AB035922). DNA sequence analysis showed that the sizes of mtsr-1/mtsr-2 PCR products for *O. hyalothecium* and *O. torulosum* were 3416 bp and 5597 bp respectively. The mitochondrial *rns* genes of the two species examined (Figure 5.1. [a]) were interrupted with introns at the following positions: S379 and S569 for *O. hyalothecium* (GenBank accession: JN871711) and S379 and S952 for *O. torulosum* (GenBank accession: HQ343317). The mS569 intron is a group I intron that belongs to the IC2 subgroup and it encodes a potential LHEase ORF. In contrast the mS379 and mS952 introns are group II introns with mS379 encoding a potential reverse transcriptase-like (RT) ORF and mS952 encoding a double motif LAGLIDADG type HE ORF. The mS569 and mS952 introns have been characterized in detail previously (Monteiro-Vitorello et al., 2009; Mullineux et al., 2010 & 2011; Hafez and Hausner, 2011a); however the mS379 intron was noted previously in three strains of *Ophiostoma minus* (Figure 5.1. [b]) [WIM(M)472 = UAMH9775, University of Alberta

Microfungus Collection & Herbarium (Edmonton, AB, Canada); WIM(M)494 = UAMH5035; WIM(M)515 = CBS 404.77; Hafez and Hausner, 2011a).

Figure 5.1. **[a]** A schematic overview of the *rns* genes of *O. hyalothecium* and *O. torulosum* showing the relative position of primer binding sites and the various intron types (group I and II) and intron encoded ORFs (RTs and LHEases). The introns are inserted at positions S379, S569 and S952 with reference to the *E. coli* SSU rRNA sequence**. [b]** A schematic overview of the mS379 intron and flanking *rns* exons showing the relative positions of the primers binding sites and the RT-ORF status (present = gray box, degenerated = gray and white box or absent = dashed line) of the different strains that have the mS379 intron. **[c]** PCR amplicons of the mS379 intron from different strains using the mtsr-1 and S379E-R primers. The band length varied from 1.3 kb (RT-ORF absent; lane 4) to 3.2 kb (RT-ORF present; lanes 1, 2, 3 & 5). The amplicons were resolved in a 1% agarose gel and the gel was stained with ethidium bromide. The L lanes contain the 1 kb plus ladder.

Figure 5.1.


[a]
mtsr1
S379E-R
mtsr2
5′
3′
Exon Sequence
Intron Sequence
RT ORF
LHE ORF
Insertion position
mS379
mS569
mS952
Intron class
Group IIA
Group IC2
Group IIB
Strains
WIN(M)848
WIN(M)860
WIN(M)848
WIN(M)860
[b]
mtsr1
S379E-R
O. mi. WIN(M)472
O. mi. WIN(M)494
O. mi. WIN(M)515
O. hy. WIN(M)848
O. to. WIN(M)860
5′
3′
Exon
Intron
RT-ORF
Deg. RT-ORF
[c]
L
O. ninus WIN(M)472
O. ninus WIN(M)494
O. ninus WIN(M)515
O. hyalothecium WIN(M)848
O. torulosum WIN(M)860
L
1
2
3
4
5
3kb
2kb
1.65kb
1kb

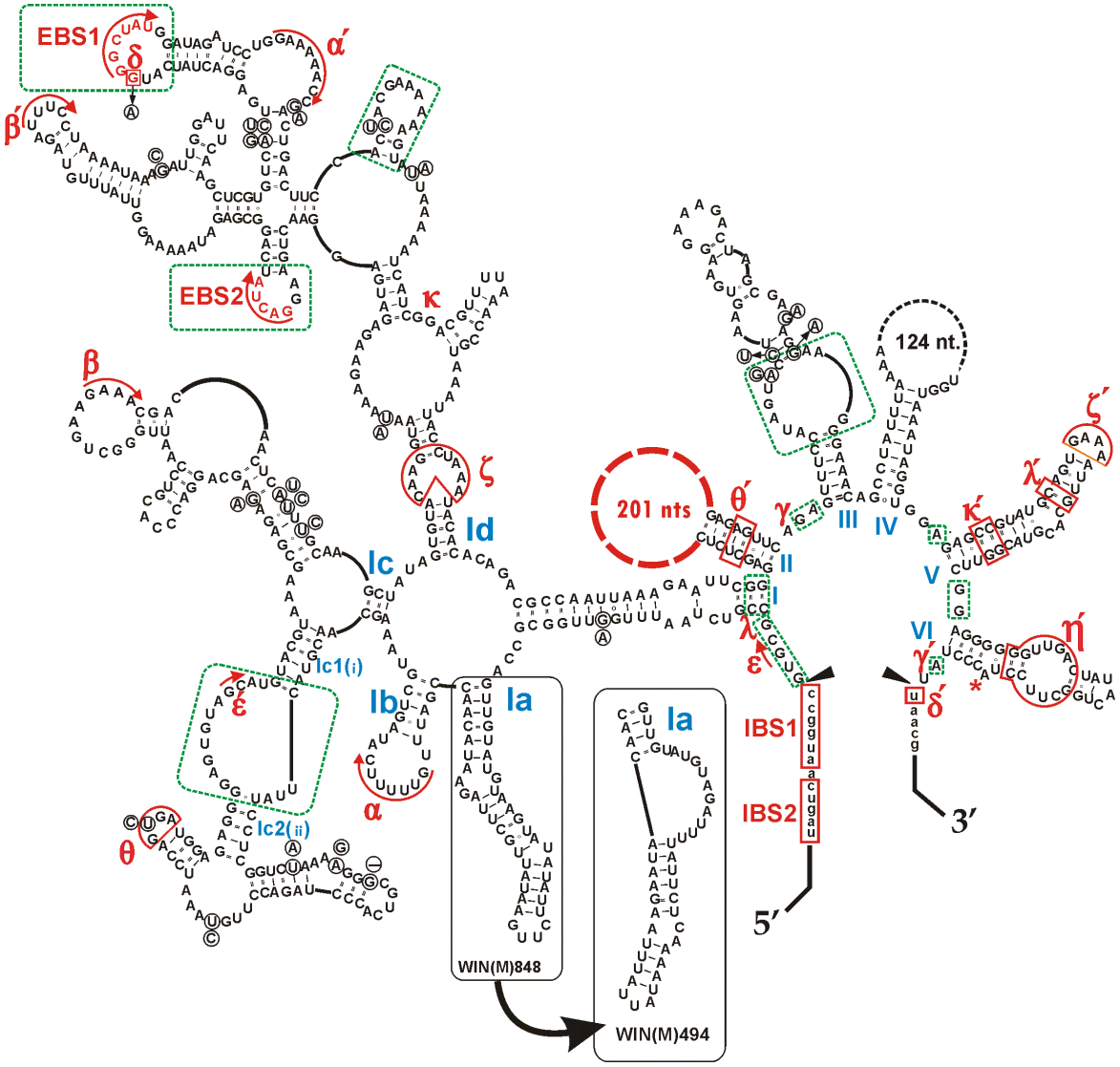
The mS379 intron is a group II intron (class A1) and is the only *rns* group II intron so far identified among *Ophiostoma* species that encodes an RT-like ORF; in addition the ORF position within domain II of this intron RNA is unique. The intron was previously found in three *O. minus* strains and is now also reported from *O. hyalothecium* WIN(M)848 and *O. torulosum* WIN(M)860 (Figure 5.1. [a] & [b]). However the mS379 intron in *O. hyalothecium* is 978 bp long compared to 2858 bps in *O. torulosu*; the size variation is due to the complete lack of an ORF within the *O. hyalothecium* mS379 intron (Figure 5.1. [b]). The absence or presence of introns and intron encoded ORFs contribute towards mtDNA polymorphism among the ophiostomatoid fungi. The mS379 intron is a group II intron based on the presence of the conserved domain V (DV) sequence. The A-type designation is based on secondary structural features (Michel and Ferat, 1995; Toor et al., 2001) identified within mS379 intron RNA fold (Figure 5.2.).

5.4.2. RT-ORF:

Based on the program ORF finder segments of an RT-like ORF were detected within the *O. torulosum* mS379 intron; however the ORF sequence appears to be degenerating. Adjustments were needed to correct certain mutations in order to generate a 614 amino acid ORF that contains all the components expected for a mtDNA group II intron-encoded protein with RT and maturase activity (Lambowitz and Zimmerly, 2004). The ORF was reconstructed as follows, first a frame shift mutation was corrected by removing a T insertion at nucleotide position 2257 (accession number HQ343317), plus two substitution mutations were corrected [G → C] and [A → T] at positions 2274 and 2304 respectively.

Figure 5.2. Secondary structure of the mS379 group II (class A1) intron RNA for *O. hyalothecium* WIN(M)848. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The positions of the exon binding sites EBS1, EBS2 and δ are noted. The positions of the intron binding sites IBS1, IBS2 and δ^- in the 5' and 3' exons are boxed. Tertiary interactions are indicated by Greek letters (ε, λ, α, β, θ, η, κ, ζ, and γ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V and VI). The solid black arrowheads indicate the intron-exon junctions (5' and 3' splicing sites). The asterisk shows the bulged adenosine nucleotide (the branch point) in domain VI. Group IIA-type features are boxed with green-dashed lines. The circled nucleotides indicates the differences between *O. hyalothecium* WIN(M)848 and *O. minus* WIN(M)494 mS379 intron sequences, while differences between their respective DIa are shown in a boxed insert. Detailed secondary structures of domain II in both ORF-less (WIN(M)848 and ORF-containing WIN(M)494 introns are given in Figure 5.3.

Figure 5.2.


EBS1
EBS2
α′
β′
δ
β
κ
ζ
Id
Ic
Ic1(i)
Ic2(ii)
ε
θ
Ib
α
Ia
WIN(M)848
Ia
WIN(M)494
201 nts
θ′
γ
III
IV
124 nt.
ζ′
χ′
κ′
V
VI
γ′
η′
δ′
*
II
I
λ
ε′
IBS1
IBS2
5′
3′

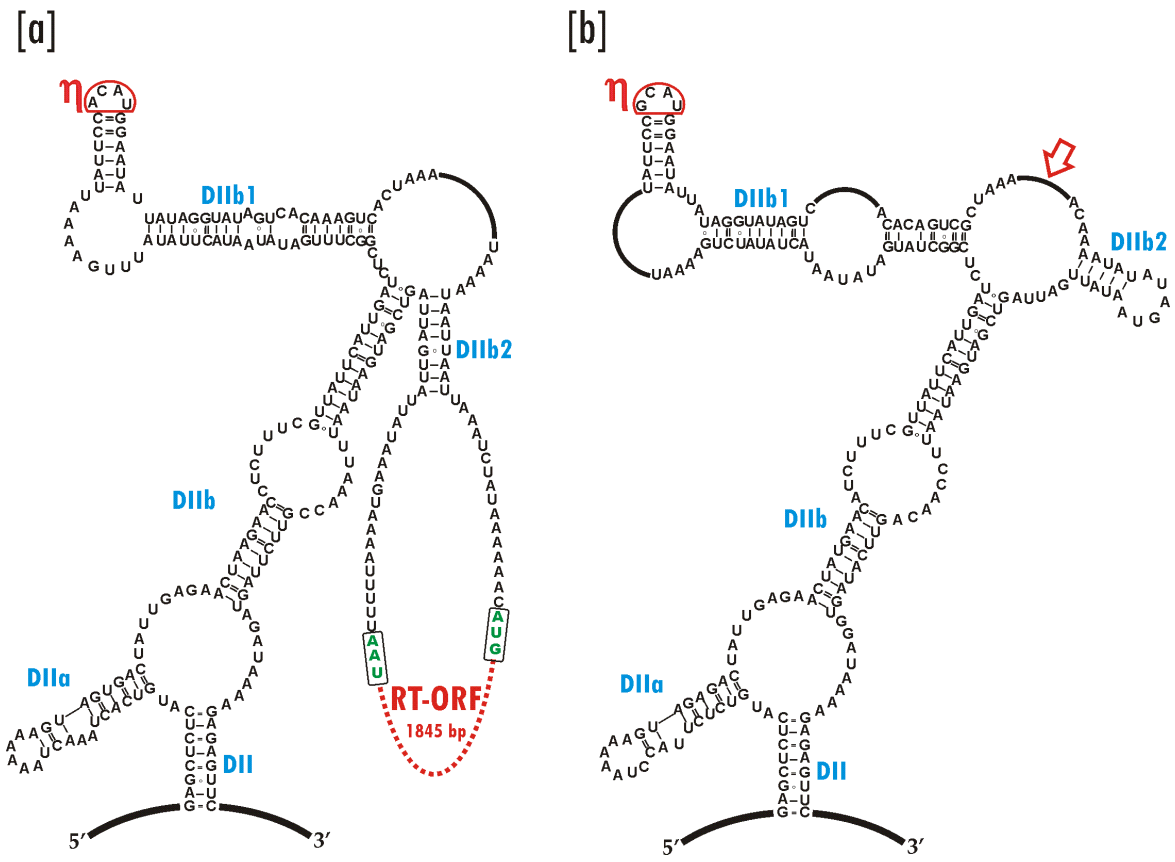
For the mS379 intron of *O. hyalothecium* ORF finder failed to detect an ORF or any ORF fragments. Sequence alignments of this sequence against those of *O. torulosum* and previously described mS379 sequences from *O. minus* also failed to detect any ORF remnants such as possible start or stop codons.

5.4.3. Domain II:

Domain II was examined in more detail (Figure 5.3.) as this domain encodes the RT-like ORFs in previously described versions of the mS379 intron (Hafez and Hausner, 2011a) and *O. torulosum*. Two tertiary interactions were identified within the mS379 intron's DII (θ – θ́ and η – ή) and these are involved in group II intron self-splicing event. The θ–θ́ is a tertiary interaction between the basal stem of DII and the terminal loop of DIc1(ii)b (Figure 5.2.). The η – ή is another tertiary interaction between DII and DVI where the tetraloop is located in DII (DIIb1) and the receptor is located in DVI (Figure 5.2. and 5.3.). The RT-ORFs found in the *O. minus* (Figure 5.3. [a]) and *O. torulosum* mS379 introns are inserted in the DIIb2 loop, while in the case of the ORF-less *O. hyalothecium* mS379 intron the terminal loop of DIIb2 is a relatively small loop (Figure 5.3. [b]). ORF containing versions of the mS379 intron such as that of *O. minus* and *O. torulosum* contain a bit of extra upstream and downstream sequences flanking the ORF when it was compared with the ORF-less version in *O. hyalothecium* (Figure 5.3.).

Figure 5.3. Detailed predicted secondary structures for domain II of the *O. minus* (WIN(M)494) [a] and the *O. hyalothecium* (WIN(M)848) [b] mS379 introns. The RT-ORF is inserted in the loop of DIIb2 in the *O. minus* (WIN(M)494) mS379 intron, while the DIIb2 loop of the *O. hyalothecium* version of this intron does not contain any ORF sequences and the arrow indicates the ORF insertion site with respect to the *O. minus* (WIN(M)494) mS379 intron. In the latter the start and stop codons of the RT-ORF are indicated in boxes.

Figure 5.3.



5.4.4. *In vivo* splicing of the mS379 introns:

In order to confirm the predicted exon/intron junctions derived by comparative sequence analysis, RT-PCR analysis was applied to RNA extracted from *O. hyalothecium* WIN(M)848 (Figure 5.4.); *O. minus* WIN(M)494 RNA was included in this analysis as a control and as a representative of an ORF containing version of this intron. Amplifying the *rns* gene from *O. minus* WIN(M)494 and *O. hyalothecium* WIN(M)848 with the primer pair mtsr-1/mtsr-2 using genomic DNA templates generated

4.0 and 3.4 kb PCR fragments respectively. Using the same primers with cDNA as starting material generated a 1.2 kb bands (Figure 5.4.) for both tested species. This confirmed that the predicted introns were all spliced out and sequence analysis of the RT-PCR derived amplicons confirmed the predicted exon/intron junction sequences (Figure 5.1.). The RT-PCR results also showed that both ORF-containing (*O. minus*) and ORF-less (*O. hyalothecium*) versions of these group II introns were spliced out.

5.4.5. Reconstruction of ancestral character states for the mS379 intron:

The Maximum Likelihood model suggest that there is a high probability that the mS379 intron was present in a common ancestor (Figure 5.5., node 1) for the following three clades: the *ulmi/novo-ulmi/piliferum* clade (Figure 5.5., node 7), the *minus/torulosum* clade (Figure 5.5., node 5), and the *hyalothecium/ossiformis* clade (Figure 5.5., node 3). This would support an argument for one initial invasion event followed by vertical inheritance of this intron among the tested members of *Ophiostoma*. The absence of this intron would therefore be best explained by the loss of the intron instead of several independent invasion events.

Figure 5.4. RT-PCR analysis for demonstrating *in vivo* splicing of mS379 intron and determining the intron/exon junctions. The *rns* transcripts for the following strains were analyzed by RT-PCR: *O. minus* WIN(M)494 and *O. hyalothecium* WIN(M)848 (lanes 3 and 6 respectively). Standard PCR reactions using genomic DNA as templates are shown in lanes 2 and 5 for strains WIN(M)494 and WIN(M)848 respectively. The PCR products from the genomic DNA controls generated 4.0 and 3.4 kb bands (lanes 2 and 5 respectively) as expected from previous data (see Figure 5.1.). The amplicon lengths for all cDNAs were 1.2 kb (lanes 3 and 6) indicating that all introns were all splicing out. The negative controls for all RT-PCR (standard PCR using DNA free whole cell RNA as a template) assays yielded no bands (lanes 4 and 7) showing that the RNA was free of any DNA. Lanes 1 and 8 contains the 1 kb plus ladder (Invitrogen).

Figure 5.4.

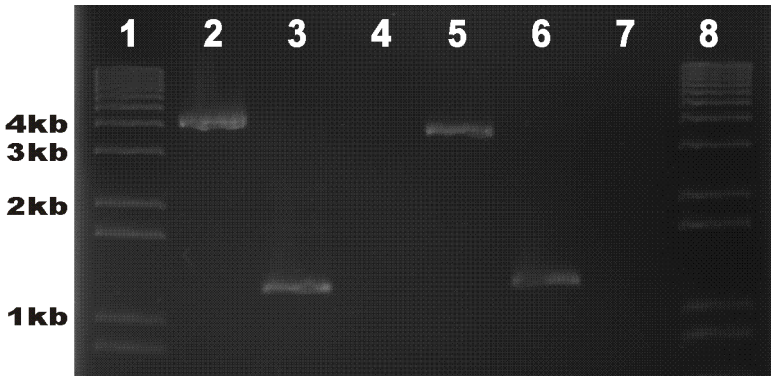
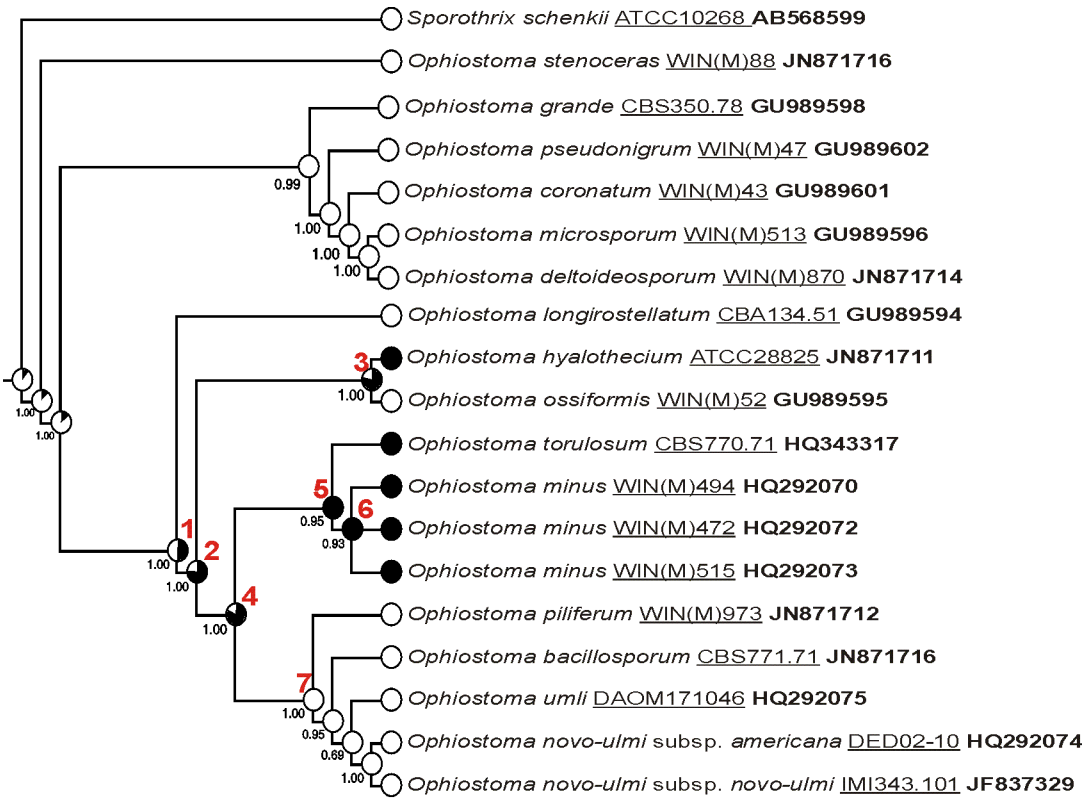


Figure 5.5 . Ancestral state reconstruction for the mS379 intron. The character state was traced along the Bayesian majority-rule consensus tree based on *Ophiostoma rns* nucleotide sequences. The numbers at the nodes indicate the Posterior Probability (PP) values obtained from a 50% majority consensus tree generated by the MrBayes program. The ancestral states for the presence or absence of mS379 intron were reconstructed by Mesquite version 2.74 using Maximum Likelihood criteria. Some of the key nodes are numbered (1 to 7) and the proportional likelihood values calculated by Mesquite are indicated. Character states are: absence of mS379 intron = 0, indicated by white circles; presence mS379 intron = 1, indicated by black circles. The slices in the circular pie charts indicate relative support for different ancestor states (present = black; absent = white). Strain numbers (underlined) and GenBank accession numbers (bold) are listed next to species names.

Figure 5.5.



Character: mS379 i
Marginal prob. recon. with model Mk1 (est.) [rate 4.68679714 [est.]] -log L.: 9.79956912 (Opt.: width 1.0) Reporting likelihoods as Proportional Likelihoods; Threshold when decisions made: 2.0 Calc. by Maximum likelihood reconstruct (Generic categorical) (id# 1061)
0
1

Proportional likelihood values

Node #	0 = absent	1 = present
1	0.49562823	0.50437177
2	0.24065593	0.75934407
3	0.21453726	0.78546274
4	0.17867199	0.82132801
5	0.00710581	0.99289419
6	0.00000405	0.99999595

5.5. DISCUSSION:

Here we characterize the mS379 intron in *O. torulosum* WIN(M)860 and *O. hyalothecium* WIN(M)848; previously this intron was reported from three *O. minus* strains WIN(M)472, WIN(M)494 and WIN(M)515 (Hafez and Hausner, 2011a). What makes this intron unique is that the RT-ORF is located in DII as usually RT-like ORFs in organellar group II introns are located in DIV. In addition, based on the previous publications (Jackson et al., 2002) and the comparative RNA website (http://www.rna.ccbb.utexas.edu/), this is the first record for an intron to be inserted at position S379. The *O. hyalothecium* version of this intron is unique because of its complete lack of an ORF. Otherwise the *O. torulosum* intron is similar to those observed in *O. minus* where the ORFs are present but shows evidence of erosion due to the presence of mutations that introduce frameshifts and premature stop codons.

5.5.1. Domain II:

Domain II contributes towards group II intron self-splicing by essential tertiary interactions the θ – θ′ and η – ή contacts with DI and DVI respectively (Costa et al., 1997; Pyle, 2010). The θ – θ′ is a tertiary interaction between the basal stem of DII and the highly conserved terminal loop of DIC1, which stabilizes the ribozyme core of the group II introns (Figure 5.2.). The second tertiary interaction that involves DII is the η – ή contact, which is a structurally conserved among group IIA and IIB introns. In group IIA introns (such mS379 intron), the η – ή tertiary interaction is between a tetraloop located in DII (the terminal loop of DIIa in mS379 intron) and a receptor in DVI, which is in contrast to group IIB introns where the location of the tetraloop and the receptor are

reversed (Costa et al., 1997). It has been proposed that the η – ή contact is an essential tertiary interaction during the conformational switch that occurs between the two steps of splicing (Chanfreau and Jacquire, 1996). It is possible for DII to serve as an insertion site for additional elements without interfering with the function of the ribozyme core because once DII is locked into position through the θ – θ́ and η – ή interactions the DII terminal loop can easily project away from the intron ribozyme core (Toor et al., 2010; Pyle, 2010).

As stated previously expanding DII to accommodate ORFs has so far only been observed in bacterial group IIE1 introns (Simon et al., 2008) and here ORF start codons were located in DII or DIII. The mS379 intron represents the first organellar group II intron that encodes a complete RT-ORF within the DIIb2 loop of DII. The RT-PCR results demonstrate that the intron is indeed spliced and that both the RT-ORF containing and the ORF-less versions of the mS379 intron appear to be spliced out *in vivo*. The position of the ORF within domain II is unusual and two scenarios have been previously proposed (Hafez and Hausner, 2011a), first the domain II ORF is due to "ORF sliding" of a domain IV encoded ORF into domain II (Figure 5.6. [a]) or by a twintron-like ancestor whereby an ORF containing group II intron ectopically integrated within the domain II of a possible ORF-less group II intron (Figure 5.6. [b]).

5.5.2. The mS379 intron is in the process of elimination:

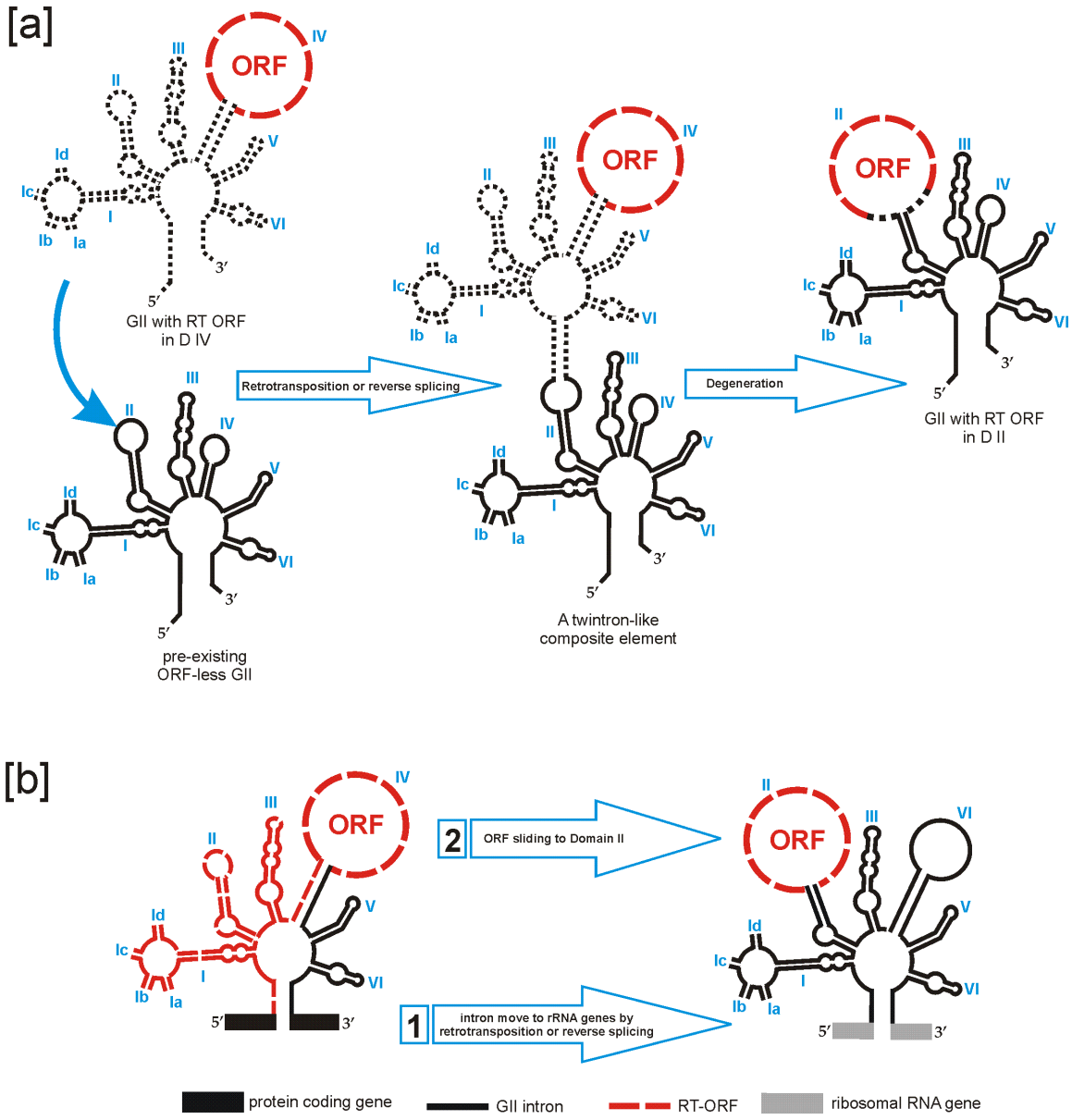
The maximum-likelihood method for tracing the origin of the mS379 intron among *Ophiostoma* species suggests that the intron was inserted once within an ancestor

that gave rise to the various clades (see Figure 5.5., node 1) but the intron was lost in some lineages. This would suggest a pattern of vertical transmission with frequent loss of the intron among the examined *Ophiostoma* species. The origin of the RT-ORF is however difficult to trace as we only found one member that lacked the ORF and other examples contain ORFs that show signs of erosion and are probably non-functional due to the presence of frameshift mutations (Figure 5.6.).

Toor et al. (2001) presented a survey of group II introns including ORF-less versions and suggested that ORF-less introns evolved from ORF containing version as group II intron structures appear to co-evolve with their RT-like encoded proteins. Also many "ORF-less" introns still harbor small but detectable remnants of ORFs such as the start or stop codons (Toor et al., 2001; Simon et al., 2008).

Figure 5.6. Proposed evolutionary models for explaining the presence of RT-ORFs within Domain II of the mS379 intron. **[a]** A group II intron (dotted line) encoding an RT-ORF (dashed line) within Domain IV could have either reverse spliced or retrotransposed (ectopic integration) into domain II of a pre-existing ORF-less group II intron (solid black line) thus generating a twintron-like composite element but over time the internal intron lost most of its ribozyme coding components leaving only the RT-ORF sequence within domain II. **[b]** The ancestral version of the mS379 intron was encoded within a protein coding gene (black box) and the RT-OTF (dashed red line) extended from domain IV upstream and was fused with to the upstream exon. This in frame fusion of intron ORFs with the upstream exons has been observed in several group II introns encoded within the *cox*1 gene in a variety of fungi (Paquin and Lang 1996; Lambowitz and Zimmerly 2004); this arrangements allows the intron ORF to be translated along with the upstream exon. Once the intron invaded (either by retrotransposition or reverse splicing) a ribosomal gene the RT-ORF had to acquire its own regulatory sequences that allow for its translation and therefore the ORF shifted into DII as gradually sequences downstream of the ORF position changed and could substitute for DIII thus allowing the ORF to "slide" into the DII position. Alternative sequences could fold into a DIV configuration and DV, the most conservative component of the group II ribozyme, remained the same.

Figure 5.6.


[a]
ORF
GII with RT ORF
in D IV
Retrotransposition or reverse splicing
pre-existing
ORF-less GII
A twintron-like
composite element
Degeneration
GII with RT ORF
in D II
[b]
2
ORF sliding to Domain II
1
intron move to rRNA genes by
retrotransposition or reverse splicing
protein coding gene
GII intron
RT-ORF
ribosomal RNA gene

As many of these elements are potentially neutral the lack of natural selection leads to the accumulation of mutations within the ORF and eventually can lead to the complete elimination of the ORF. However, the loss of efficient splicing could lead to a detrimental phenotype for the host genome; therefore features of the secondary structure are conserved and trans-acting factors are probably available that compensate for the lack of intron-encoded maturates (reviewed in Hausner et al., 2012). In this study we describe an ORF-less group II intron in *O. hyolothecium* with no evidence of residual sequences of an ORF (see Figure 5.3.) and versions of the same intron with ORFs in various states of "decay" (Figure 5.1.). One would assume that for mobility the ancestor of this intron would have contained a functional RT-like ORF which provided all the features needed for retrohoming into cognate alleles that lack the intron or for retrotransposition into ectopic (new) sites. NCBI database searches (blastn or blastp) with the mS379 intron sequence or ORF amino acid sequence or the *rns* nucleotide sequence revealed no additional examples for this intron or *rns* genes that contained a group II intron at this position. This suggests that this intron might be rare, even in recent studies on *rns* genes among ophiostomatoid fungi we have only encountered this intron in three different species (Figure 5.1. [b] & [c]) among approximately 50 species examined so far (Mullineux et al., 2011; Hafez and Hausner, 2011a; unpublished data). The *rns* gene may represent a reservoir for various types of mobile introns (Hafez and Hausner, 2011a) with the *rns* gene being both the recipient and the donor of introns. The mS379 probably represents a recent addition to the *rns* "intron landscape" but appears to be in the process of elimination a trend that has also been recently noted for various introns located within the fungal mitochondrial *cox*1 gene (Férandon et al., 2010a & 2010b).

CHAPTER: 6

THE mtDNA *RNS* LANDSCAPE IN ASCOMYCYTES FUNGI: A NATURAL RESERVOIR FOR INTRONS, TWINTRONS AND HOMING ENDONUCLEASES.

6.1. ABSTRACT:

Comparative sequence analysis of the mitochondrial *rns* gene among species of *Ophiostoma, Grosmannia, Ceratocystiopsis* and related taxa provides an overview of the types of introns that have invaded this gene. The *rns* gene appears to be a reservoir for a number of group I and group II introns along with intron associated open reading frames such as HEases and reverse transcriptases. This study uncovered a novel twintron complex where a group IIA1 intron invaded the open reading frame embedded within a group IC2 intron. Overall the distribution of the introns does not appear to follow evolutionary lineages suggesting the possibility of horizontal gains and frequent loses. The work tries to provide an overview of the various types of introns that can invade the *rns* gene and thus be a resource for those who try to annotate fungal mitochondrial genomes.

6.2. INTRODUCTION:

The gene coding for the mitochondrial small subunit ribosomal RNA (*rns*) is a highly conserved essential gene that like its nuclear, Eubacterial and Archaean counterparts has been exploited as an evolutionary chronometer for rDNA based phylogenetic analysis. In 1988 it was reported that all ascomycete *rns* genes are continuous (i.e. no introns; Wolf and Del Giudice, 1988); however since that time a variety of introns have been noted to be inserted within fungal mtDNA *rns* genes (Haugen and Bhattacharya, 2004; Haugen et al., 2004; Lang et al., 2007; Hafez and Hausner, 2011a). The initial focus of this study was to examine the *rns* intron landscape for species of *Ophiostoma, Grosmannia* and *Ceratocystiopsis*. These genera contain plant pathogens and blue-stain fungi; the latter can discolor timber and thus reduce its economic value (Wingfield et al., 1993) and so far very little information is available on their mitochondrial genomes and their intron complements. We expanded the study to examine the types of introns inserted within the *rns* gene in other members of the Ascomycetous fungi.

The mitochondrial genomes of fungi contain two types of introns, group I and group II introns. Group I and group II introns are self-splicing elements and therefore they minimize their impact to the genes that host them. Many group I introns contain open reading frames (ORFs) that either promote their mobility and/or assist in their splicing activity such as HEases and/or maturases (Belfort et al., 2002). HEases are enzymes that can initiate intron mobility by introducing double-stranded breaks (DSBs) at specific target sites, which activates the cellular DSB-repair pathway (Belfort et al.,

2002). Maturase proteins assist the intron with folding the intron RNA into a splicing competent ribozyme (Caprara and Waring, 2005). Intron-encoded maturases are derived from HEases and there are instances of bifunctional HEases that have maturase and site specific endonuclease activities (reviewed in Belfort, 2003). There are two families of HEases commonly encountered among fungal mtDNAs; they are named according to conserved amino acids motifs, the LAGLIDADG and GIY-YIG proteins (Stoddard, 2006 & 2011). Homing endonuclease genes (HEGs) have also been shown to be mobile elements that can move independently from their ribozyme counterparts (Mota and Collins, 1988). Mobile introns and HEGs are of interest as they have applications in biotechnology as rare cutting site specific endonucleases (Takeuchi et al., 2011; Stoddard, 2011; Hausner, 2012).

Group II introns typically encode reverse transcriptase-like (RT) ORFs, which in contrast to group I intron ORFs, appear to be co-evolve with their ribozyme counterpart (Toor et al. 2001). This is in part driven by the dependence of the group II introns on intron-encoded maturases that can bind to the various structural variants of group II RNA transcripts, plus the inability of the reverse transcriptase ORF to move on its own (Toor et al., 2001). Group II introns are mobile elements; however they move by a mechanism that involves an RNA intermediate and requires the RT to form a ribonucleoprotein particle (RNP) with the spliced intron lariat to form a "site-specific endonuclease" that will allow the element to retrohome into cognate alleles that lack the intron (Lambowitz and Zimmerly, 2004 & 2011). Some group II introns encode LAGLIDADG type HEGs and it appears that these HEGs have the potential to mobilize their group II intron host by

a mechanism invoked to explain HEase facilitated homing of group I intron into alleles that lack the intron (Mullineux et al., 2010).

HEGs and introns are potential invasive elements and therefore contribute towards the size of fungal mtDNA genomes, mtDNA polymorphisms, and they can promote mtDNA rearrangements (Dujon, 1989; Charter et al., 1996; Belcour et al., 1997; Salvo et al., 1998; Gobbi et al., 2003). Also some of these elements have been associated with mtDNA instabilities in a variety of fungi (Osiewacz and Esser, 1984; Michel and Cummings, 1985; Cummings et al., 1986; Dujon and Belcour, 1989; Cummings et al., 1990; Abu-Amero et al., 1995). Group II introns have been associated with mitochondrial genome instabilities leading to processes such as senescence in *Podospora anserina* (Dujon, 1989; Griffiths, 1992) and mitochondrial induced hypovirulence in *Cryphonectria parasitica* (Baidyaroy et al., 2011).

This study was designed to provide an overview of the various types of elements that can insert within the *rns* gene and this should prove to be useful in annotating the *rns* gene for fungi in future studies. The data are also valuable to those who are interested in prospecting for native HEases or ribozymes for potential applications in biotechnology (Long et al., 2003; Bagheri and Kashani-Sabet, 2004; Takeuchi et al., 2011).

6.3. RESULTS:

Methods relevant to Chapter 6 are presented in Chapter 2 (Materials and Methods).

6.3.1. The *rns* gene landscape:

Amplifying the mtDNA *rns* gene from different species and strains of ophiostomatoid fungi using the primers mtsr-1 and mtsr-2 generated short (1.2 kb) and long (>1.2 kb) PCR products (see Chapter 2, section 2.3, for PCR amplification protocols of the *rns* gene). These length variations were attributed to the absence or presence of insertions (i.e. introns) within the *rns* gene; sequences were obtained for these PCR products as described in Chapter 2. BLAST was used to retrieve nucleotide sequences (BLASTn) from NCBI data bases belonging to ascomycetous fungi and by comparative sequence analysis of ascomycete *rns* genes eleven positions were noted that have insertions. The insertion sites (positions) were identified based on sequence comparison with the *E. coli* SSU rRNA gene and the insertion sites were as follows: S62, S379, S531, S569, S788, S915, S917, S952, S1210, S1224 and S1247 (Figure 6.1.).

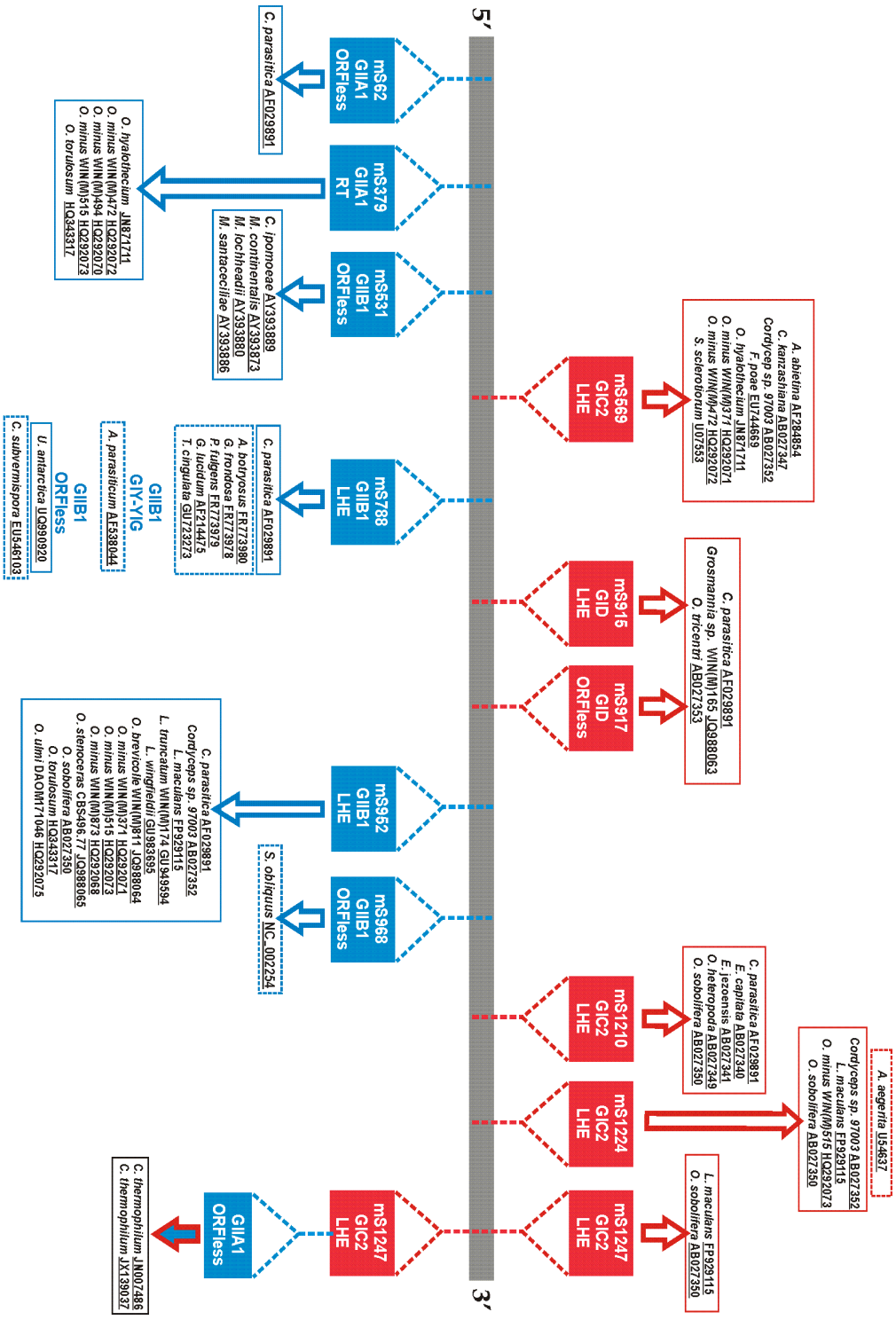
Group I and group II introns were found and examples for both intron types were observed that either had an ORF sequence embedded or not. The introns were categorized based on sequence and structural information and comparison with previously published reports (Nikoh and Fukatsu, 2000; Marinoni and Lachance, 2004; Monteiro-Vitorello et al., 2009; Baidyaroy et al., 2011; Hafez and Hausner, 2011a & 2011b) and the web server RNA weasel program (Lang et al., 2007) was used to confirm the intron designations. Group I introns were recorded at the following positions (in brackets are intron type

designations): S569 (IC2), S915 (ID), S917 (ID), S1210 (IC2), S1224 (IC2), and S1247 (IC2) while group II introns were recorded in the following positions: S62 (IIA1), S379 (IIA1), S531 (IIB1), S788 (IIB1) and S952 (IIB1).

A unique twintron was found at position S1247 (group IIA1 intron inserted in a group IC2 intron) in *Chaetomium thermophilum* (Table 6.1). During the NCBI data search we also examined partial *rns* sequences (not included in the *rns* phylogeny) for insertions and we included noteworthy insertions found in algae and protists (Figure 6.1.).

Figure 6.1. Schematic diagram depicting the *rns* introns landscape (gray line) for the ascomycetous fungi and some selected eukaryotes. The insertions of twelve introns including a possible twintron are shown and the intron insertion sites, intron class [group I (GI) or group II (GII)] and type of intron encoded proteins (RT = reverse transcriptase ORF, LAGLIDADG or GIY-YIG HE ORFs) are indicated. Representative sequences with their GenBank accession numbers are boxed for each intron.

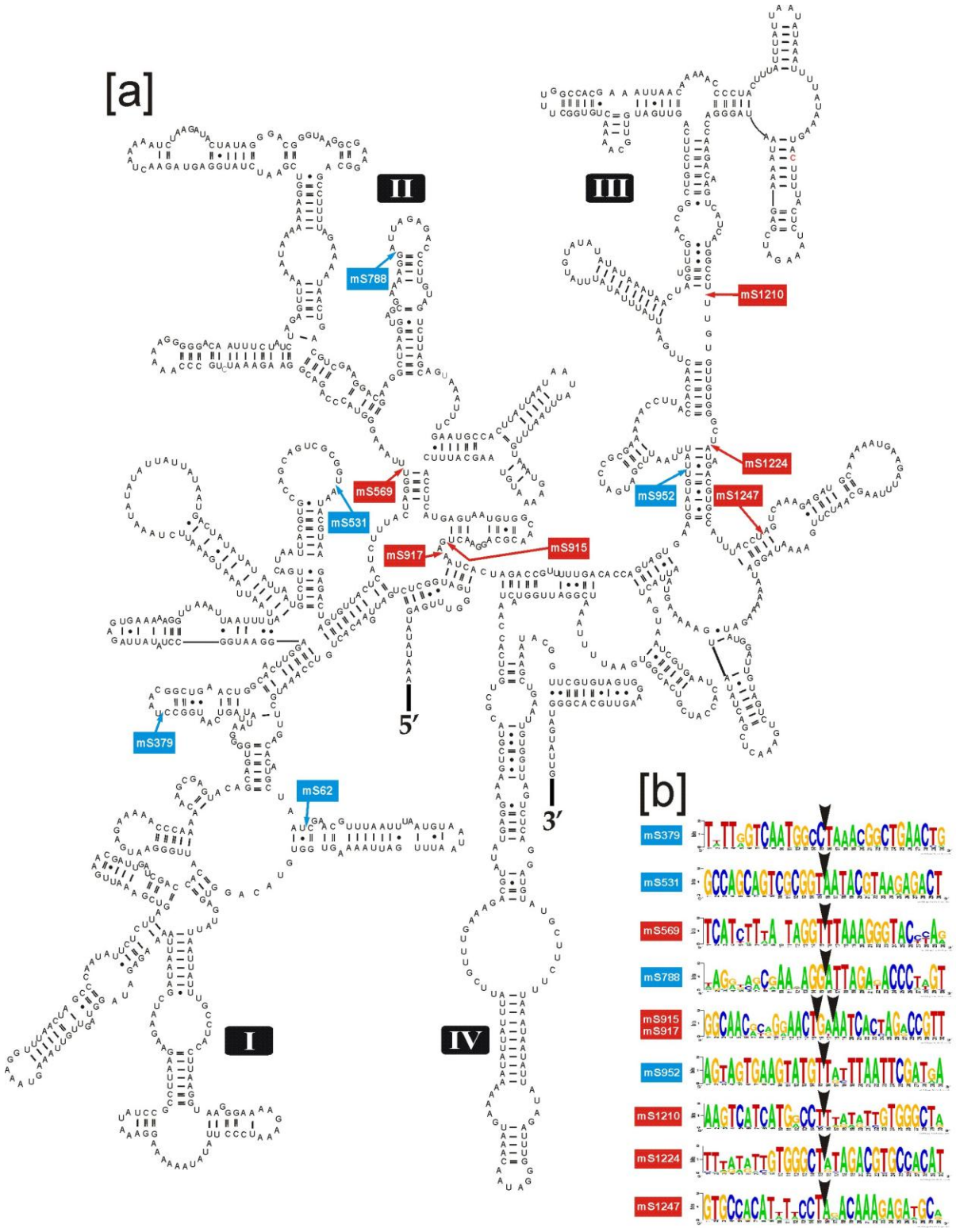
Figure 6.1.


5′
3′
mS62 GIIA1 ORFless
mS379 GIIA1 RT
mS531 GIIB1 ORFless
mS569 GIC2 LHE
mS788 GIIB1 LHE
mS915 GID LHE
mS917 GID ORFless
mS952 GIIB1 LHE
mS968 GIIB1 ORFless
mS1210 GIC2 LHE
mS1224 GIC2 LHE
mS1247 GIC2 LHE
mS1247 GIC2 LHE
GIIA1 ORFless
GIIB1 GIY-YIG
GIIB1 ORFless
C. parasitica AF029891
O. hyalothecium JN871711
O. minus WIN(M)472 HQ292072
O. minus WIN(M)494 HQ292070
O. minus WIN(M)515 HQ292073
O. torulosum HQ343317
C. parasitica AF029891
C. ipomoeae AY393889
M. conthenalis AY393873
M. lochheadii AY393880
M. santacciliae AY393886
C. parasitica AF029891
A. botryosus FR773980
G. frondosa FR773978
P. fulgens FR773979
G. lucidum AF214475
T. cingulata GU723273
A. parasiticum AF538044
L. wingfieldii GU983695
C. parasitica AF029891
Cordyceps sp. 97003 AB027352
L. maculans FP929115
O. sobolifera AB027350
S. obliquus NC_002254
C. parasitica AF029891
Cordyceps sp. 97003 AB027352
L. maculans FP929115
O. sobolifera AB027350
L. brevicolle WIN(M)811 JQ888064
O. minus WIN(M)371 HQ292071
O. minus WIN(M)515 HQ292073
O. minus WIN(M)873 HQ292068
O. stenoceras CBS496.77 JQ988065
O. sobolifera AB027350
O. torulosum HQ343317
O. ulmi DAOM171046 HQ292075
U. antarctica UQ990920
C. subvermispora EU546103
C. parasitica AF029891
Grosmannia sp. WIN(M)165 JQ888063
O. tricentri AB027353
C. parasitica AF029891
E. jezoensis AB027340
E. capitata AB027341
O. heteropoda AB027349
O. sobolifera AB027350
L. maculans FP929115
O. sobolifera AB027350
A. abietina AF284854
C. kanzashiana AB027347
Cordyceps sp. 97003 AB027352
F. poae EU744669
O. hyalothecium JN871711
O. minus WIN(M)371 HQ292071
O. minus WIN(M)472 HQ292072
S. sclerotiorum U07553
A. aegerita U54637
Cordyceps sp. 97003 AB027352
L. maculans FP929115
O. minus WIN(M)515 HQ292073
O. sobolifera AB027350
C. thermophilum JN007486
C. thermophilum JX139037

A secondary structure model for the complete mitochondrial small subunit RNA for *O. novo-ulmi* subsp. *americana* DED 2002-10 (Figure 6.2. [a]) was generated in order to place the intron insertions within the context of the *rns* rRNA secondary structure. The *rns* rRNA is composed of four well-defined domains denoted I, II, III & IV (Noller and Woese, 1981; Köchel and Kuntzel, 1981; Hafez and Hausner, 2011b). The intron insertion sites were characterized within this secondary structure model and DNA sequence logos (Figure 6.2. [b]) were generated to assess the level of sequence conservation in the mt *rns* exon sequences flanking the intron insertion sites. The sequence logos were based on 30 nucleotides with 15 nucleotides flanking upstream and downstream of the intron insertion sites. In general the DNA sequence logos showed that intron insertion sites correspond to highly conserved *rns* sequences. Based on the *rns* secondary structure (Figure 6.2. [a]) we noted that the following group I introns are inserted within unpaired regions (such as loops) S569, S915, S917, S1210, and S1224. Only S1247 is inserted within a stem segment within the *rns* RNA structure. Group II introns appear to be inserted into unpaired regions (S379, S531, and S788) and stem segments (S62 and S952), with one group II intron inserted within the ORF of the S1247 group I intron. With regards to the *rns* secondary structure, intron insertions occurred in domains I through III, but no insertions were recorded for the *rns* domain IV.

Figure 6.2. [a] Nucleotide sequence and secondary structure model for the *O. novo-ulmi* subsp. *americana* mitochondrial *rns* RNA (mt. SSU rRNA) (GenBank accession number HQ292074) showing the four structural domains indicated by Roman numbers (I to IV). Watson-Crick base pairing is indicated by dashes (A - U) and double dashes (G = C), while wobble base pairing is indicated by circles (G ◦ U). Group I and group II intron insertion positions relative to the *E. coli* SSU rRNA sequence are indicated by red and blue arrows respectively. **[b]** Nucleotide sequence logos are shown for the exon sequences that flank the intron insertion positions (15 upstream and 15 downstream nucleotides). The black arrowhead indicates the insertion position of each intron. The sequence logos were generated from 61 intronless SSU rRNA sequences by the online program WebLogo version 2.8.2. (Crooks et al., 2004)

Figure 6.2.


[a]
II
III
I
IV
mS788
mS1210
mS569
mS1224
mS952
mS1247
mS531
mS917
mS915
mS379
mS62
5′
3′
[b]
mS379
mS531
mS569
mS788
mS915
mS917
mS952
mS1210
mS1224
mS1247

6.3.2. Phylogenetic analysis based on *rns* sequences:

The *rns* sequences from ophiostomatoid fungi and related ascomycetous fungal taxa were subjected to phylogenetic analysis. For this analysis partial *rns* sequences were excluded. The *rns* sequence of *S. cerevisiae* was selected as the outgroup for this analysis. Figure 6.3. shows the 50% majority rule Bayesian consensus tree for 61 taxa (sequences) based on an unambiguously aligned data set that covered 1644 nucleotide positions. The *rns* phylogeny provided moderate to strong support (see Figure 6.3.) for the monophyly of the following orders: Ophiostomatales, Sardariales, Hypocreales and Eurotiales. Within the Ophiostomatales the *rns* phylogeny appears to resolve three clades composed of species that can be assigned to the following genera: *Grosmannia, Ceratocystiopsis* and *Ophiostoma*. However, there are "orphans" such as *O. microsporum, O. deltoideosporum, O. coronatum, C. pseudonigra, O. grande* and *O. stenoceras* along with *Sporothrix schenckii*. The latter five species do share an internal node from which eventually a core group of *Ophiostoma* species can be derived (see node 5 in Figure 6.3.). *Ophiostoma microsporum* groups next to the node (node 6 in Figure 6.3) that supports the monophyly of species belonging to the genus *Ceratocystiopsis*, and *O. deltoideosporum* is placed on a single branch that is part of a trichotomy that includes species that can be assigned to *Grosmannia* (see node 7; Figure 6.3.) and *Ceratocystiopsis*. Albeit, the node support for this trichotomy is poor. *Ophiostoma microsporum* and *O. deltoideosporum* are ophiostomatoid species that based on classical criteria and recent recommendations would have been expected to belong to the genus *Ophiostoma sensu* (Zipfel et al., 2006). *Ophiostoma longisporum* grouped next to the node that supports the monophyly of the tested strains representing *Ceratocystiopsis*; also

the node supporting the monophyly of members of *Ceratocystiopsis* inclusive of *O. longisporum* received strong support (> 95% for both ML and Bayesian analysis). The generic status of *O. longisporum* has not yet been fully resolved (reviewed in Hafez et al., 2012).

A NCBI data search and the *rns* phylogeny were used to contrast the distribution of the various types of introns inserted within the *rns* gene among various members of the Ascomycota (Table 6.1., Figure 6.1. and Figure 6.3.). In general it can be stated that there is no obvious connection between species relatedness and intron distribution. For example mS1224 was found in three distantly related phylogentic lineages, two members of the Hypocreales, one strain of *Ophiostoma minus* and in *Leptosphaeria maculans* (Pleosporales). Similarly the mS569 intron was noted in *Cordyceps* sp. 97003 and among three members of the *Ophiostoma* clade. Within the Ophiostomatales mS379, mS788 and mS952 are the most frequently encountered introns (Table 6.1.). Only the mS379 group II intron is unique to the ophiostomatoid fungi compared to the currently available ascomycete *rns* sequences. This group II intron was noted to either lack an ORF or have a degenerated ORF suggesting that this intron might be in the process of being eliminated from the *rns* gene. Forty-three *rns* sequences were analyzed for species that belong to the Ophiostomatales of those 30 had no insertions and 13 contained one or up to three introns. The most frequently encountered intron was mS952, a group II intron that encodes a LAGLIDADG ORF. The mS952 intron was also observed within *rns* sequences among species belonging to the Pleosporales, Hypocreales, and Diporthales.

Among the sequences examined the *rns* gene of *Cryphonectria parasitica* contained the largest number of introns (6 introns) if one considers the KFC9 strain of *C. parasitica*, a hypovirulent variant of the Chestnut blight fungus (Baidyaroy et al., 2011). The KFC9 strain contains an additional ORF-less group II intron at position mS62, an intron that appears to be unique to this strain and appears to be implicated in inducing hypovirulence (Baidyaroy et al., 2000 & 2011). This would suggest that this intron imparts a phenotype and thus it might be under selection and this would limit its spread within or among different species.

Previously Monteiro-Vitorello et al. (2009) reported four introns for the *rns* gene of the EP155 strain of *C. parastica*: mS788, mS915, mS952 and mS1210. The mS915 intron was reexamined (Michel F., personal communications) and it was noted that the original annotation was incorrect. The original mS915 intron annotation missed the fact that actually two ID intron cores are present. The first ID intron encodes a double motif LAGLIDADG type ORF and the second ID intron lacks an ORF and is separated from the upstream ID intron by two nucleotides. So with the recognition of the mS917 intron the *rns* gene of standard reference strain of *C. parasitica* EP155 would contain a total of five introns.

Figure 6.3. Phylogenetic tree based on the aligned *rns* data set for some ascomycetous fungi. The tree topology is based on Bayesian analysis, solid circles, squares and triangles represent Posterior Probability (PP) supportive values (95-100%), (85-94%) and (50-84%) respectively as obtained from a 50% majority Bayesian consensus tree. Open circles, squares and triangles represent bootstrap (BS) support values (95-100%), (85-94%) and (50-84%) respectively, based on ML analysis. The branch lengths are based on Bayesian analysis and are proportional to the number of substitutions per site. Strain numbers (underlined) and GenBank accession numbers (in square brackets) are listed next to species names. The tree is rooted with *rns* (SSU rRNA) sequences from the yeast fungi *S. cerevisiae* (GenBank accession AJ011856) and *Candida glabrata* (GenBank accession AJ511533). The presence of introns in the *rns* gene (mtDNA SSU rRNA) are denoted by colored (+) symbols.

Figure 6.3.

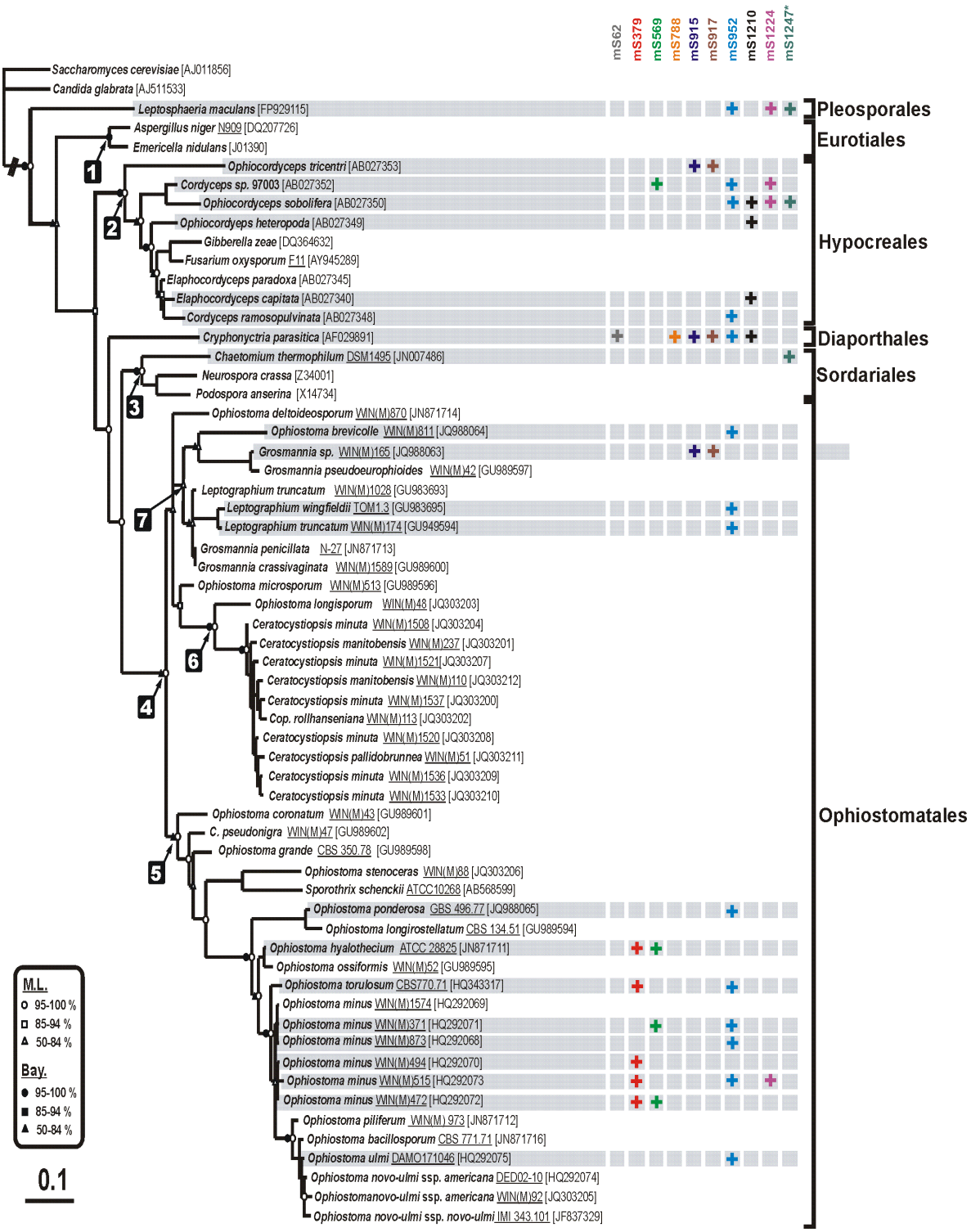

mS62
mS379
mS569
mS788
mS915
mS917
mS952
mS1210
mS1224
mS1247*
Saccharomyces cerevisiae [AJ011856]
Candida glabrata [AJ511533]
Leptosphaeria maculans [FP929115]
Aspergillus niger N909 [DQ207726]
Emericella nidulans [J01390]
Ophiocordyceps tricentri [AB027353]
Cordyceps sp. 97003 [AB027352]
Ophiocordyceps sobolifera [AB027350]
Ophiocordyeps heteropoda [AB027349]
Gibberella zeae [DQ364632]
Fusarium oxysporum F11 [AY945289]
Elaphocordyceps paradoxa [AB027345]
Elaphocordyceps capitata [AB027340]
Cordyceps ramosopulvinata [AB027348]
Cryphonyctria parasitica [AF029891]
Chaetomium thermophilum DSM1495 [JN007486]
Neurospora crassa [Z34001]
Podospora anserina [X14734]
Ophiostoma deltoideosporum WIN(M)870 [JN871714]
Ophiostoma brevicolle WIN(M)811 [JQ988064]
Grosmannia sp. WIN(M)165 [JQ988063]
Grosmannia pseudoeurophioides WIN(M)42 [GU989597]
Leptographium truncatum WIN(M)1028 [GU983693]
Leptographium wingfieldii TOM1.3 [GU983695]
Leptographium truncatum WIN(M)174 [GU949594]
Grosmannia penicillata N-27 [JN871713]
Grosmannia crassivaginata WIN(M)1589 [GU989600]
Ophiostoma microsporum WIN(M)513 [GU989596]
Ophiostoma longisporum WIN(M)48 [JQ303203]
Ceratocystiopsis minuta WIN(M)1508 [JQ303204]
Ceratocystiopsis manitobensis WIN(M)237 [JQ303201]
Ceratocystiopsis minuta WIN(M)1521[JQ303207]
Ceratocystiopsis manitobensis WIN(M)110 [JQ303212]
Ceratocystiopsis minuta WIN(M)1537 [JQ303200]
Cop. rollhanseniana WIN(M)113 [JQ303202]
Ceratocystiopsis minuta WIN(M)1520 [JQ303208]
Ceratocystiopsis pallidobrunnea WIN(M)51 [JQ303211]
Ceratocystiopsis minuta WIN(M)1536 [JQ303209]
Ceratocystiopsis minuta WIN(M)1533 [JQ303210]
Ophiostoma coronatum WIN(M)43 [GU989601]
C. pseudonigra WIN(M)47 [GU989602]
Ophiostoma grande CBS 350.78 [GU989598]
Ophiostoma stenoceras WIN(M)88 [JQ303206]
Sporothrix schenckii ATCC10268 [AB568599]
Ophiostoma ponderosa GBS 496.77 [JQ988065]
Ophiostoma longirostellatum CBS 134.51 [GU989594]
Ophiostoma hyalothecium ATCC 28825 [JN871711]
Ophiostoma ossiformis WIN(M)52 [GU989595]
Ophiostoma torulosum CBS770.71 [HQ343317]
Ophiostoma minus WIN(M)1574 [HQ292069]
Ophiostoma minus WIN(M)371 [HQ292071]
Ophiostoma minus WIN(M)873 [HQ292068]
Ophiostoma minus WIN(M)494 [HQ292070]
Ophiostoma minus WIN(M)515 [HQ292073
Ophiostoma minus WIN(M)472 [HQ292072]
Ophiostoma piliferum WIN(M) 973 [JN871712]
Ophiostoma bacillosporum CBS 771.71 [JN871716]
Ophiostoma ulmi DAMO171046 [HQ292075]
Ophiostoma novo-ulmi ssp. americana DED02-10 [HQ292074]
Ophiostomanovo-ulmi ssp. americana WIN(M)92 [JQ303205]
Ophiostoma novo-ulmi ssp. novo-ulmi IMI 343.101 [JF837329]
Pleosporales
Eurotiales
Hypocreales
Diaporthales
Sordariales
Ophiostomatales
1
2
3
4
5
6
7
M.L.
95-100 %
85-94 %
50-84 %
Bay.
95-100 %
85-94 %
50-84 %
0.1

Table 6.1. List of species and strains for mtDNA *rns* sequences along with the sequence GenBank accession numbers.

Table 6.1.

Organism	Strain/Isolate	Order	NCBI Accession #	Intron insertions sites											
				S62	S379	S5531	S5569	S7888	S9115	S9117	S9552	S9968	S12110	S12247	S12447
Agrocybe aegerita (V. Brig.) Singer	SM47	Agaricales	**U54637**											**X**	
Aleurodiscus botryosus Burt	NA	Russulales	**FR7739**					**X**							
Amoebidium parasiticum Cienkowski	JAP-7-2	Amoebidiales	**AF538044**					**X**							
Ascocalyx abietina (Lagerb.) Schläpf.-Bernh.	JA72-0001	Helotiales	**AF284856**				**X**								
Aspergillus niger Tiegh.	N909	Eurotiales	**DQ207726**												
Candida glabrata (H.W. Anderson) S.A. Mey. & Yarrow	ATCC 2001	Saccharomycetales	**AJ511533**												
Candida ipomoeae Lachance, C.A. Rosa, Starmer & J.M. Bowles	UWO(PS)91-672.1	Saccharomycetales	**AY393889**			**X**									
Ceratocystiopsis manitobensis (J. Reid, Eyjólfsdóttir & G. Hausner) Zipfel, Z.W. de Beer & M.J. Wingf.	UAMH9813 WIN(M) 237	Ophiostomatales	**JQ303201**												
Ceratocystiopsis minuta (Siemaszko) H.P. Upadhyay & W.B. Kendr.	CBS145.59 WIN(M) 1508	Ophiostomatales	**JQ303204**												
C. minuta	RJ 5095 WIN(M)1533	Ophiostomatales	**JQ303210**												
C. minuta	CBS 116796 WIN(M) 1536	Ophiostomatales	**JQ303209**												
C. minuta	CBS 116763 WIN(M) 1537	Ophiostomatales	**JQ303200**												
Ceratocystiopsis pallidobrunnea (Olchow. & J. Reid) H. P. Upadhyay	WIN(M) 51	Ophiostomatales	**JQ303211**												
Ceratocystiopsis rollhanseniana (J. Reid, Eyjólfsdóttir & Georg Hausner) Zipfel, Z.W. de Beer & M.J. Wingf.	UAMH 9797 WIN(M) 110	Ophiostomatales	**JQ303212**												
C. rollhanseniana	UAHM 9774 WIN(M) 113	Ophiostomatales	**JQ303202**												
Ceratocystiopsis sp.	MPB1 WIN(M) 1520	Ophiostomatales	**JQ303208**												
Ceratocystiopsis sp.	MPB2 WIN(M)1521	Ophiostomatales	**JQ303207**												

Table 6.1. CONTINUED

Ceratocystis ossiformis Olchow. & J. Reid	WIN(M) 52	Ophiostomatales	**GU989595**												
Ceriporiopsis gilvescens	BRNM 709970	Russulales	**EU546103**					**X**							
Chaetomium thermophilum La Touche	DSM 1495	Sordariales	**JN007486**												**X**
C. thermophilum	UAMH 2024	Sordariales	**JX139037**												**X**
Cordyceps kanzashiana Kobayasi & Shimizu	NA	Hypocreales	**AB027347**				**X**								
Cordyceps ramosopulvinata Kobayasi & Shimizu	NA	Hypocreales	**AB027348**								**X**				
Cordyceps sp. 97003	NA	Hypocreales	**AB027352**				**X**				**X**			**X**	
Cryphonectria parasitica (Murrill) M.E. Barr	Ep155 ATCC 38755	Diaporthales	**AF029891**	**X**				**X**	**X**	**X**	**X**		**X**		
Elaphocordyceps capitata (Holmsk.) G.H. Sung, J.M. Sung & Spatafora	NA	Hypocreales	**AB027340**										**X**		
Elaphocordyceps jezoensis (S. Imai) G.H. Sung, J.M. Sung & Spatafora	NA	Hypocreales	**AB027341**										**X**		
Elaphocordyceps paradoxa (Kobayasi) G.H. Sung, J.M. Sung & Spatafora	NA	Hypocreales	**AB027345**												
Emericella nidulans (Eidam) Vuill.	NA	Eurotiales	**J01390**												
Fusarium oxysporum Schltdl.	F11	Hypocreales	**AY945289**												
Fusarium poae (Peck) Wollenw.	TBig2ai	Hypocreales	**EU744669**				**X**								
Ganoderma lucidum (Curtis) P. Karst.	ATCC 64251	Polyporales	**AF214475**					**X**							
Gibberella zeae (Schwein.) Petch	NA	Hypocreales	**DQ364632**												
Grifola frondosa (Dicks.) Gray	NA	Polyporales	**FR773978**					**X**							
Grosmannia crassivaginata (H.D. Griffin) Zipfel, Z.W. de Beer & M.J. Wingf.	UAMH 7004 WIN(M)1589	Ophiostomatales	**GU989600**												
Grosmannia penicillata (Grosmann) Goid.	N-27	Ophiostomatales	**JN871713**												
Grosmannia pseudoeurophioides (Olchow. & J. Reid) Zipfel, Z.W. de Beer & M.J. Wingf.	WIN(M) 42	Ophiostomatales	**GU989597**												
Grosmannia sp.	WIN(M) 165	Ophiostomatales	**JQ988063**						**X**	**X**					
Leptographium truncatum (M.J. Wingf. & Marasas) M.J. Wingf.	J.R.88-449 WIN(M) 1028	Ophiostomatales	**GU983693**												
Leptographium truncatum	NFRI 1813/1 WIN(M) 174	Ophiostomatales	**GU949594**								**X**				

Table 6.1. CONTINUED

Leptographium wingfieldii M. Morelet	TOM1.3	Ophiostomatales	**GU983695**								**X**				
Leptosphaeria maculans (Sowerby) P. Karst.	v23.1.3	Pleosporales	**FP929115**								**X**			**X**	**X**
Metschnikowia continentalis Lachance, C.A. Rosa, Starmer, Schlag-Edl., J.S.F. Barker & J.M. Bowles	UFMG96-179	Saccharomycetals	**AY393873**			**X**									
Metschnikowia lochheadii Lachance	UWO(PS)00-133.2	Saccharomycetals	**AY393880**			**X**									
Metschnikowia santaceciliae Lachance, J.M. Bowles & Starmer	UWO(PS)01-517a1	Saccharomycetals	**AY393886**			**X**									
Neurospora crassa Shear & B.O. Dodge	FGSC #2489 74-OR23-1A	Sordariales	**Z34001**												
Ophiocordyceps heteropoda (Kobayasi) G.H. Sung, J.M. Sung, Hywel-Jones & Spatafora	NA	Hypocreales	**AB027349**										**X**		
Ophiocordyceps sobolifera (Hill ex Watson) G.H. Sung, J.M. Sung, Hywel-Jones & Spatafora	NA	Hypocreales	**AB027350**								**X**		**X**	**X**	**X**
Ophiocordyceps tricentri (Yasuda) G.H. Sung, J.M. Sung, Hywel-Jones & Spatafora	NA	Hypocreales	**AB027353**						**X**	**X**					
Ophiostoma stenoceras Math.-Käärik [=*Ophiostoma stenoceras* (Robak) Melin & Nannf.]	NFRI 60-15 B-23 WIN(M) 88	Ophiostomatales	**JQ303206 JN871716**												
Ophiostoma bacillosporum (Butin & Zimmerman) de Hoog & Scheffer	CBS 771.71	Ophiostomatales	**JN871715**												
Ophiostoma brevicolle (R.W. Davidson) de Hoog & R.J. Scheff.	CBS 150.78 WIN(M) 811	Ophiostomatales	**JQ988064**								**X**				
Ophiostoma coronatum (Olchow. & J. Reid) M. Villarreal	WIN(M)43	Ophiostomatales	**GU989601**												
Ophiostoma deltoideosporum (Olchow. & J. Reid) Georg Hausner, J. Reid & Klassen	WIN(M) 870	Ophiostomatales	**JN871714**												
Ophiostoma grande Samuels & E. Müll.	CBS 350.78	Ophiostomatales	**GU989598**												

Table 6.1. CONTINUED

Ophiostoma hyalothecium (R.W. Davidson) Georg Hausner, J. Reid & Klassen	ATCC 28825	Ophiostomatales	**JN871711**		**X**		**X**								
Ophiostoma longirostellatum (B.K. Bakshi) Arx & E. Müll.	CBS 134.51	Ophiostomatales	**GU989594**												
Ophiostoma longisporum (Olchow. & J. Reid) Georg Hausner, J. Reid & Klassen	WIN(M) 48	Ophiostomatales	**JQ303203**												
Ophiostoma microsporum (R.W. Davidson) Arx	CBS 412.77 WIN(M) 513	Ophiostomatales	**GU989596**												
Ophiostoma minus (Hedgc.) Syd. & P. Syd.	WIN(M) 1574	Ophiostomatales	**HQ292069**												
O. minus	WIN(M) 873	Ophiostomatales	**HQ292068**								**X**				
O. minus	C-248 (NFRC) WIN(M) 494	Ophiostomatales	**HQ292070**		**X**										
O. minus	UAMH 9805 WIN(M) 371	Ophiostomatales	**HQ292071**				**X**				**X**				
O. minus	WIN(M) 472	Ophiostomatales	**HQ292072**		**X**		**X**								
O. minus	CBS 404.77 WIN(M) 515	Ophiostomatales	**HQ292073**		**X**						**X**			**X**	
Ophiostoma novo-ulmi ssp. *americana* Brasier	DED 2002-10	Ophiostomatales	**HQ292074**												
Ophiostoma novo-ulmi ssp. *americana*	DED 2002-92	Ophiostomatales	**JQ303205**												
Ophiostoma novo-ulmi ssp. *novo-ulmi* Brasier	IMI 343.101	Ophiostomatales	**JF837329**												
Ophiostoma piliferum (Fries) Sydow and Sydow	UAMH 9561 WIN(M) 973	Ophiostomatales	**JN871712**												
Ophiostoma ponderosae (T.E. Hinds & R.W. Davidson) Georg Hausner, J. Reid & Klassen	CBS 496.77	Ophiostomatales	**JQ9883065**								**X**				
Ophiostoma pseudonigrum (Olchow. & J. Reid) Georg Hausner & J. Reid	WIN(M) 47	Ophiostomatales	**GU989602**												
Ophiostoma torulosum Butin & G. Zimm.) Georg Hausner, J. Reid & Klassen	CBS 770.71	Ophiostomatales	**HQ343317**		**X**						**X**				
Ophiostoma ulmi (Buisman) Nannf.	DAOM 171046	Ophiostomatales	**HQ292075**								**X**				
Podospora anserina (Rabenh.) Niessl	NA	Sordariales	**X14734**												
Pycnoporellus fulgens (Fries) Donk	NA	Polyporales	**FR773979**					**X**							

Table 6.1. CONTINUED

Saccharomyces cerevisiae Meyen ex E.C. Hansen	FY1679	Saccharomycetals	**AJ011856**												
Scenedesmus obliquus (Turpin) Kützing	KS3-2	Sphaeropleales	**NC_002254**									**X**			
Sclerotinia sclerotiorum (Lib.) de Bary	NA	Helotiales	**U07553**				**X**								
Sporothrix schenckii Hektoen & C.F. Perkins	ATCC 10268	Ophiostomatales	**AB568599**												
Trametes cingulata Berk.	ATCC 26747	Polyporales	**GU723273**					**X**							
Usnea antarctica Du Rietz	AFTOL-ID 813	Lecanorales	**DQ990920**					**X**							

* The presence of introns is represented by (X) symbol.
NA = data not available
UAMH = University of Alberta Microfungus Collection & Herbarium, Devonian Botanic Garden, Edmonton, AB, Canada, T6G 2E1;
RJ = R. Jankowiak collection (University of Agriculture in Cracow, Cracow, Poland)
MPB = Mountain Pine Beetle collection, /Department of Forestry, University of British Columbia, Canada;/
NFRI = Norwegian Forest Research Institute, AS, Norway; B-23 is an older designation used in NFRI for the NFRI 6-15 strain
WIN(M) = WIN(M) = University of Manitoba (Winnipeg) Collection
CBS = Centraal Bureau voor Schimmelcultures, Utrecht, The Netherlands
ATCC = American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108, USA.

6.3.3. Notes on the *rns* insertions:

mS62:

The mS62 intron is an ORF-less group IIA1 intron, as stated above it so far has only been observed in a hypovirulent strain of *C. parasitica* (KFC9). This intron, Cpa-mS62 (also known as InC9) is associated with splicing deficiency and has been linked to growth abnormalities and a cytoplasmically transmissible form of hypovirulence (Baidyaroy et al., 2011).

mS371:

The S379 position in some member of *Ophiostoma* was found to be invaded with a group IIA1 intron with an eroded reverse transcriptase-like open reading frame. The mS379 intron is unique because it has an open reading frame encoded within DII (Hafez and Hausner, 2011a). An ORF-less version of the mS379 intron was discovered in the closely related species *O. hyalothecium* (GenBank accession number JN871711; see Chapter 5 of this thesis for more detailed analysis of this intron**)**.

mS531:

The mS531 insertion is a 626 bp ORF-less group IIB1 intron (Li et al., 2011) which so far has only been recorded in the *rns* gene of the yeast fungus *Candida ipomoeae* and related taxa (Table 6.1.; Marinoni and Lachance, 2004; Li et al., 2011).

mS569:

The mS569 intron is a group IC2 intron that encodes a double motif LHEase (for secondary structure models of mS569, see Chapter 4 & Hafez and Hausner, 2011a). The mS569 intron was recorded (sizes given in brackets) in several strains of *O. minus* and *O. hyalothecium* (1327 bp), *Cordyceps* sp. 97003 (1323 bp), *A. abietina* (1395 bp), *F. poae* (1100 bp), *S. sclerotiorum* (1380 bp) and *C. kanzashiana* (1336 bp).

mS788:

The mS788 is a group IIB1 intron that has been found in two distantly related ascomycetes fungi: *C. parasitica* (Monteiro-Vitorello et al., 2009) and the lichen fungus *Umbilicaria antarctica* (Miadlikowska et al., 2006), the former (2182 bp) encodes a double motif LHEase while the later (1076 bp) is an ORF-less group II intron. Several ORF-less and ORF-containing examples of this intron were also found in some basidiomycetous fungi (Li et al., 2011). Another group IIB1 intron was also found at this position in the protozoan *Amoebidium parasiticum* (Lang et al., 2002) but this intron (1344 bp) encodes a GIY-YIG HEase. The latter is a nice demonstration of how HEGs can mobilize into various genomic niches such as group I intron or in this case a group II intron.

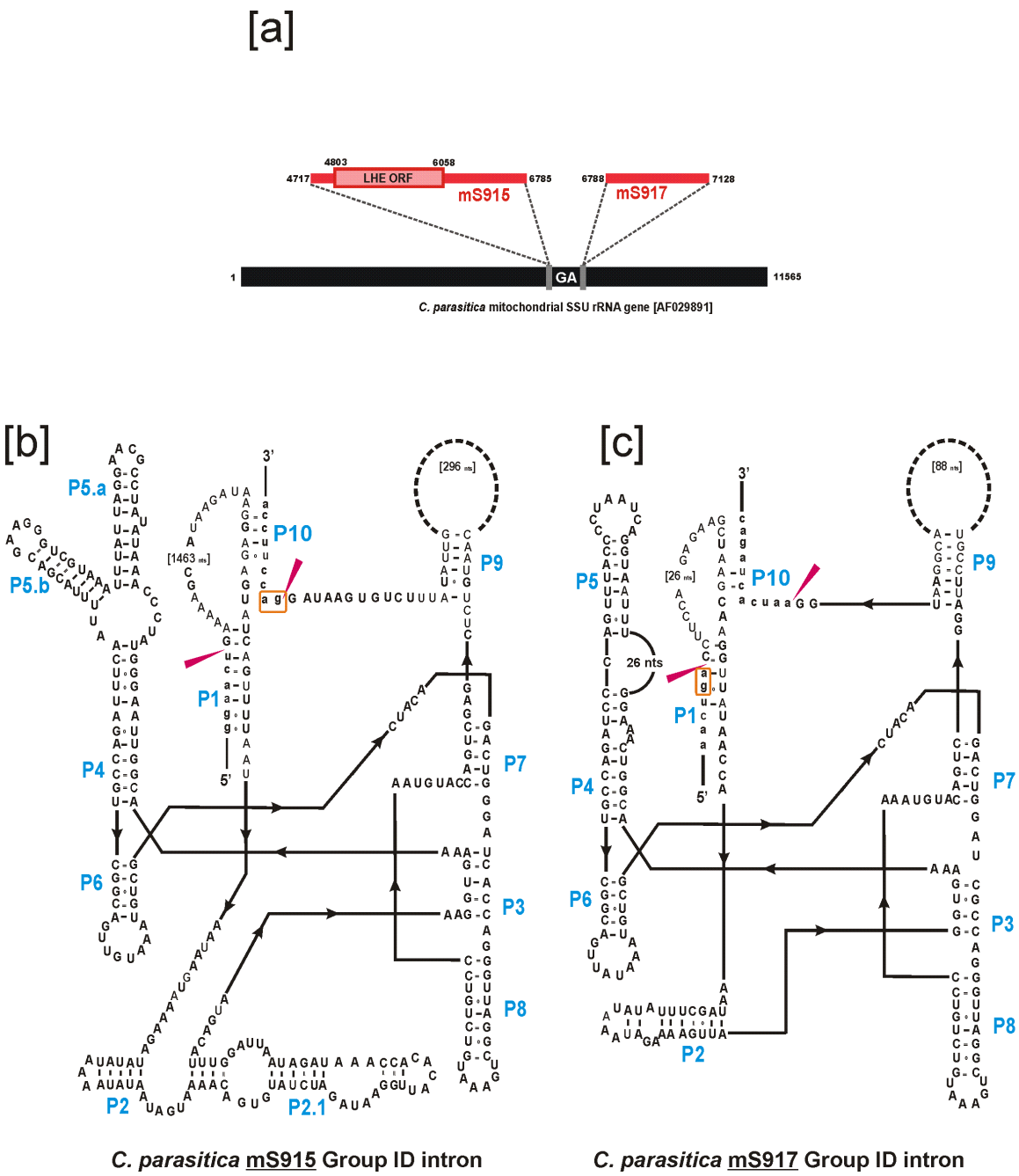
mS915/mS917:

This combination of group ID type introns has so far been identified in three unrelated species. As stated previously these introns were misinterpreted in the *C. parasitica rns* sequence as one intron. The presence of two ID core sequences that can form two individual introns is shown in Figure 6.4. [a]. The mS915 intron (Figure 6.4. [b]) encodes a double motif LHEase while the mS917 intron is an ORF-less ID intron (Figure 6.4. [c]).

Additional examples of this intron pair are the mS915/mS917 insertions in *O. tricentri* and a *Grosmannia* sp. strain [WIN(M)165]. However, in the latter species the RNA weasel program only detected one group ID intron core. But based on comparative sequence analysis among these three species and their mS915/mS917 sequences clearly shows that: *O. tricentri* contains two group ID introns exactly like *C. parasitica* although the mS917 intron is slightly degenerated in *O. tricentri*, while in *Grosmannia* sp. strain WIN(M)165 the mS917 intron is highly degenerated; i.e. has deviated from the expected consensus sequences referred to as the P, Q, R, S elements, which could not be identified. The *rns* sequence alignment that included all sequences studies in Figure 6.3. predicts that the two introns are separated by a very short two nucleotide exon (5'-GA-3').

Figure 6.4. The mS915 and mS917 group ID introns. **[a]** A schematic representation of the *C. parasitica* mitochondrial *rns* (SSU rRNA) gene showing the mS915 and mS917 group ID introns which are separated by a two-nucleotide exon (5'-GA-3'). The mS915 intron encodes a double motif LHEase ORF. Panel **[b]** and **[c]** show the **s**econdary structures for the mS915 and mS917 group ID introns respectively. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The ten pairing regions (P1-P10) are indicated. The solid red arrowheads indicate the intron-exon junctions (putative 5' and 3' splicing sites). Designation of intron types is based on the conventions for group I introns (Burke et al., 1987; Michel and Westhof, 1990; Suh et al., 1999; Li and Zhang, 2005) and the web server RNA weasel.

Figure 6.4.


[a]
4803
6058
4717
LHE ORF
6785
mS915
6788
mS917
7128
1
GA
11565
C. parasitica mitochondrial SSU rRNA gene [AF029891]
[b]
[c]
P5.a
P5.b
P10
P1
P4
P6
P9
P7
P3
P8
P2
P2.1
P5
26 nts
C. parasitica mS915 Group ID intron
C. parasitica mS917 Group ID intron

mS952:

The mS952 is a group IIB1 intron that encodes a putative double motif LHEase rather than the expected group II intron RT-ORF. This intron has been well characterized and appears to be quite common in the mitochondrial *rns* gene of ascomycetous and basidiomycetous fungi (Toor and Zimmerly, 2002; Mullineux et al., 2010 & 2011; Hafez and Hausner, 2011a). It has been shown in one instance that this intron contains an ORF in Domain III that encodes a functional homing endonuclease that could facilitate the movement of this group II intron in a DNA based manner typical for the mobility of group I introns (Mullineux et al., 2010).

mS1210:

The mS1210 is a group IC2 intron that encodes a double motif LHEase. The mS1210 intron was reported previously in *C. parasitica* (1335 bp) (Monteiro- Vitorello et al., 2009), and can also be detected in four species within the order Hypocreales: *E. capitata* (1309 bp), *E. jezoensis* (1331 bp), *O. heteropoda* (1335 bp) and *O. sobolifera* (1312 bp), (Nikoh and Fukatsu, 2000).

mS1224:

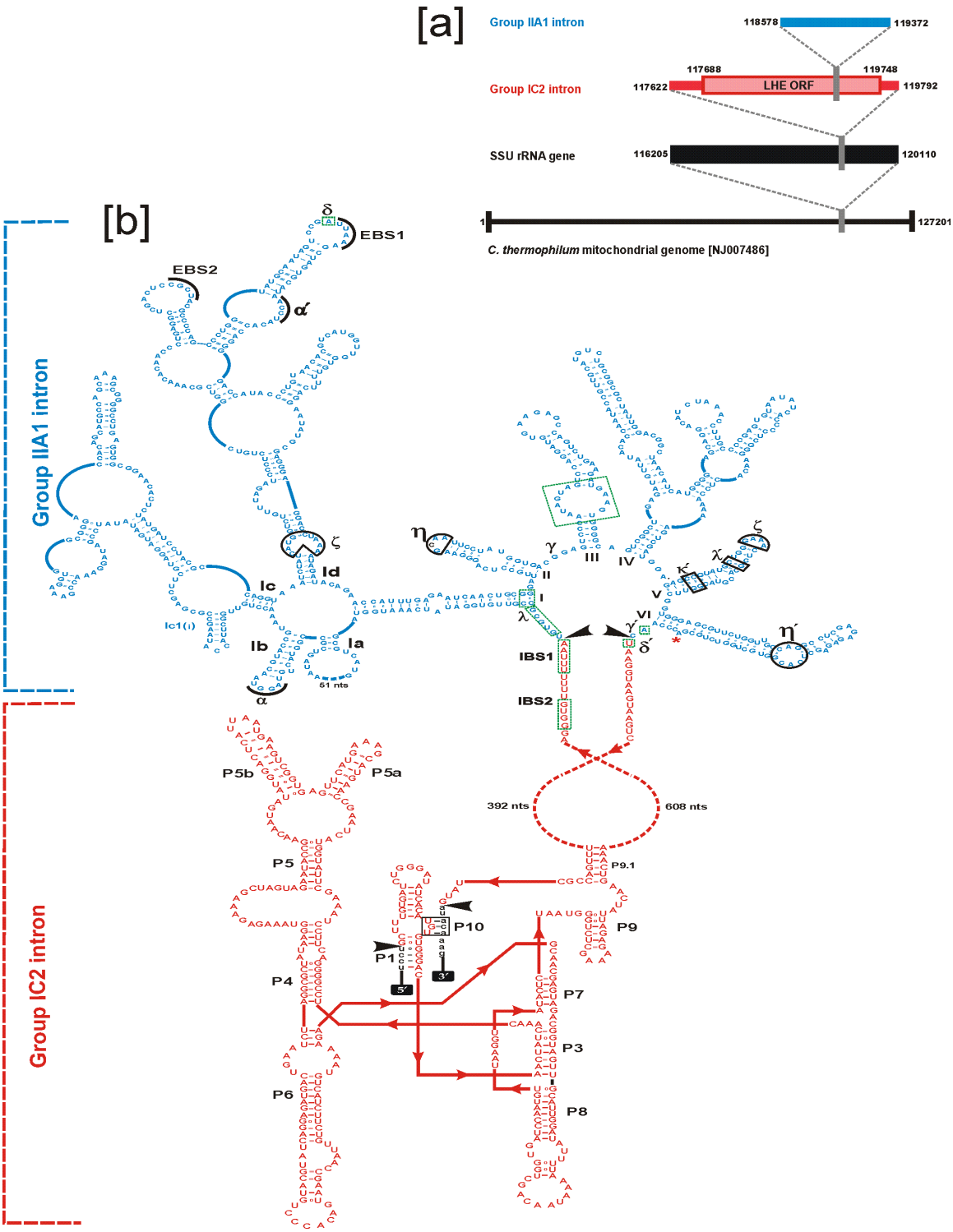
The mS1224 is a group IC2 intron which encodes a double motif LHEase. This intron was found in four ascomycetous fungi: *Leptosphaeria maculans* (1399 bp), *Cordyceps* sp. strain 97003 (1351 bp), *Ophiostoma minus* WIN(M)515 (1394 bp) and *Ophiocordyceps sobolifera* (1343 bp). It also found in the SSU rRNA gene of the basidiomycetous fungus *Agrocebe aegerita* (Gonzalez et al., 1997).

mS1247:

The mS1247 position can be occupied by a group IC2 intron, which encodes a double motif LHEase. This intron was found in two ascomycetous fungi: *Ophiocordyceps sobolifera* (1371 bp) and *Leptosphaeria maculans* (1426 bp). Another example of the mS1247 insertion was found in *Chaetomium thermophilum* DSM 1495, but here a group II intron (795 bp) was found to have invaded the double motif LHEase ORF encoded within the group I intron (1376 bp). This gives rise to a unique twintron element (2171 bp) (Figure 6.5.), where an ORF-less group II intron has inserted within a group IC2 intron ORF located within a big loop in the group I intron's P9.1 region. Another example of this twintron was found in *C. thermophilum* UAMH 2024. The insertion of the group II intron fragments the group I intron resident ORF, but it is worth noting that removal of the group II intron would reconstitute the LAGLIDADG ORF into a continuous reading frame. The IBS1 and IBS2 sequences contact elements for the EBS1 and EBS2 sequences required for forming an active splicing complex are located within the LAGLIDADG sequence upstream of the intron insertion site. This intron arrangement could be a new composite mobile element composed of a host group I intron and a tenant group II intron with both being mobilized by the resident introns encoded protein. Clearly this requires experimental investigation.

Figure 6.5. The mS1247 twintron. **[a]** A schematic representation of the *C. thermophilum* mitochondrial genome contig (black solid line) showing the *rns* (SSU rRNA) gene (black box) which is interrupted by a group IC2 intron (red line) that encodes a double motif LHEase ORF (pink box). The group IC2 intron is also interrupted by an ORF less group IIA1 intron (blue line) to form a unique twintron element. **[b]** A detailed secondary structure model of the mS1247 twintron, the group IC2 intron is labeled as indicated in Figure 6.4. For the group IIA1 intron, the positions of EBS1, EBS2 and δ are noted. The positions of the IBS1, IBS2 and δ^- in the 5' and 3' "exons" (found in the group IC2 ORF sequence) are boxed with dotted lines. Tertiary interactions are indicated by dashed lines and Greek letters (α, β, γ, ε, ζ, η, κ, and λ). The six major structural domains are indicated by Roman numbers (I, II, III, IV, V and VI). The solid black arrowheads indicate the intron-"exon" junctions (5' and 3' splicing sites). The asterisk shows the bulged adenosine nucleotide in domain VI (the branch point).

Figure 6.5.


[a]
Group IIA1 intron
Group IC2 intron
SSU rRNA gene
LHE ORF
118578
119372
117688
119748
117622
119792
116205
120110
1
127201
C. thermophilum mitochondrial genome [NJ007486]
[b]
Group IIA1 intron
Group IC2 intron
EBS1
EBS2
Id
Ic
Ic1(i)
Ib
Ia
61 nts
IBS1
IBS2
392 nts
608 nts
P5b
P5a
P5
P4
P6
P1
P10
P9
P9.1
P7
P3
P8

6.4. DISCUSSION:

6.4.1. Taxonomic implications:

In this study we examined the *rns* sequence for species that can be assigned to the following genera of the Ophiostomatales: *Ceratocystiopsis* H.P. Upadhyay & W.B. Kendr., *Grosmannia* Goid. and *Ophiostoma* Syd. & P. Syd. These genera include economically important tree pathogens and blue-stain fungi (see Wingfield et al., 1993; Hausner and Reid, 2004; Hausner et al., 2005). Members of these genera lack forcible ascospore discharge, have deliquescent asci and form sticky ascospore droplets at the apex of their perithecial necks. These fungi in many instances also tend to produce their slimy/sticky conidia on long-stalked conidiophores. Species assigned to *Ceratocystiopsis* 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 & 1993c).

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. Zipfel et al. (2006), based on partial LSU rDNA and β-tubulin sequences, provided some justification to recognize the above three genera. Hafez et al. (2012), using the nuclear SSU rDNA gene, showed that Zipfel's proposal has merit but that some species cannot easily be accommodated within either of these three genera. The mtDNA *rns* gene analysis shows that three clades can be readily identified that

would correspond to the circumscription of the three genera discussed above but again some species do not readily align within the three currently recognized genera. For example species that are based on morphological criteria are to be assigned to *Ophiostoma*, *O. microsporum* and *O. deltoideosporum* failed to group within the *Ophiostoma* clade and species such as *O. coronatum*, *C. pseudonigra*, *O. grande*, and *O. stenoceras* although allied to *Ophiostoma* only group with the *Ophiostoma* clade on branches with deeper nodes with poor to moderate bootstrap support (Figure 6.3.). These data suggest that the current taxonomic arrangement defines paraphyletic genera and in the future new genera might have to be proposed to assign species that currently do not fit. However, one species that is still assigned to *Ophiostoma*, *O. longisporum*, based on the *rns* phylogeny should be transferred to *Ceratocystiopsis*.

Previously Plattner et al. (2009) suggested that *O. longisporum* species should be excluded but based on nuclear SSU rDNA data (Hafez et al., 2012) and the *rns* phylogeny presented here this species is best accommodated within *Ceratocystiopsis*. This is also in agreement with morphological criteria as *O. longisporum* has elongated falcate-like sheathed ascospores and micronematous conidiophores (Upadhyay, 1981), features typical for members of *Ceratocystiopsis*. The distribution of insertions within the *rns* genes as previously noted did not follow the evolutionary lineages of the species that contain them so no taxonomic conclusions could be reached based on intron distribution.

6.4.2. Novel introns and intron combinations:

6.4.2.1. mS915 and mS917 introns are separated by a two nucleotide exon:

The *rns* gene of *C. parasitica* has been previously investigated by Monteiro-Vitorello and co-workers (2009) and they suggested that there is one group I intron (Cpa-mS915) inserted at position mS915. However, further analysis of the Cpa-mS915 sequence with the RNA weasel program indicated the presence of two group ID introns instead of one. Sequence analysis of *rns* gene alignment indicates that these two introns are separated by a two nucleotide exon. This would be the shortest exon reported so far in the literature. Previously mtDNA short exons have been reported within the *coxI* gene that is estimated to be three nucleotides long (Cummings et al., 1989; Férandon et al., 2010a). The presence of this short exon (5'-GA-3') separating the mS915 and mS917 introns was confirmed by RT-PCR and sequence analysis of the cDNA product. Such a short exon raises some interesting questions, such as how do these introns splice? As group I intron splicing involves flanking "exon sequences" one has to assume that the two introns (915 and 917) splice sequentially. One has to assume that the mS915 intron, which encodes a double motif LHEase (that could act as a maturase) splices first. The mS915 intron during splicing uses part of the mS917 intron sequence to form the P10 paring region which defines the 3' splice site. Once the mS915 intron has been removed the mS917 intron can use the now available upstream "exon" sequence for forming the P1 helix and the internal guide sequence necessary to define the 5' splice site. Exon-intron junctions of these two introns need further investigation by studying the RNA splicing *in vitro* allowing for the recovery of splicing intermediates for these two introns. What is also worth noting is that these two introns so far have always been reported as a

pair, so it is possible that the mS915 intron encoded protein mobilizes both of these introns as a "composite" type element and further that the IEP acts as a maturase for both of these introns. Clearly these speculations require experimental investigation.

6.4.2.2. The mS1247 insertion is a novel twintron complex:

Twintrons are introns within introns and are spliced by sequential splicing reactions in which the internal intron(s) splice first to reconstitute the external intron. The term twintron (**Twin** in**tron**) was coined for the first time by Copertino and Hallick (1991) when they noted a twintron within the chloroplast *psbF* gene of *Euglena gracilis*. The *psbF* twintron is a group II intron inserted into another group II intron. The majority of these twintrons have been characterized within the *Euglena* chloroplast genome but these elements have also been found in cryptomonad algae (*Pyrenomonas salina*; Maier et al., 1995), and group I based twintrons (group I inserted within a group I intron) have been described in *Didymium iridis* (Einvik et al., 1998). Since the discovery of the *psbF* twintron, several categories of twintron have been characterized. A twintron can be simple (external intron interrupted by 1 internal intron), or complex (external intron interrupted by multiple internal introns). Typically, the internal and external introns comprising a twintron element are from the same category; group I internal to group I (Einvik et al., 1998), group II internal to group II (Copertino and Hallick, 1991; Copertino et al., 1993), and group III internal to group III (Copertino et al., 1992 & 1993). Mixed twintrons (consisting of introns belonging to different categories) were characterized from the *Euglena gracilis rps3* gene in which an internal group II intron is found to interrupt an external group III intron (Copertino et al., 1991 & 1993). In

Rhodomonas salina (=*Pyrenomonas salina*) twintrons (nested group II/group III introns) were identified where the internal intron lost its splicing capacity, essentially merging with the outer intron forming "one" splicing unit (Khan and Archibald 2008). This illustrates the potential of generating new intron variants by introns invading other introns. To our knowledge, the mS1247 twintron represents the first mixed twintron consisting of group II intron as an internal intron and a group I intron as an external intron and also it represent the first mitochondrial twintron.

The most parsimonious explanation of twintron formation is the insertion of one or more introns into another intron by ectopic integration (Copertino and Hallick, 1991; Copertino et al., 1991), so for the mS1247 intron an ancestral mobile group II intron (that assumes it encoded a functional RT ORF) inserted within the group I intron. However the internal group II intron eventually lost its ORF. The formation of the mS1247 twintron could be also be explained by the possible transposition of a group II intron into the mS1247 group I intron via reverse splicing. The internal group II intron in the mS1247 twintron contains exon-binding sites (EBSs) complementary to their respective intron-binding sites (IBSs) which are located in the external group I intron ORF. These base-pairing interactions are required for splicing, reverse splicing and for inserting into DNA target sites during intron homing. Thus one can envision a process were reverse splicing of a group II intron into an intact *rns* precursor transcript was followed by reverse transcription of the invaded transcript and the cDNA via homologous recombination replaced the original *rns* gene that lacked the group II intron. The end result was the permanent insertion of the group II intron within the group I intron at mS1247. Scenarios

of reverse splicing and retrotransposition or cDNAs recombining with the original mtDNA loci resulting in intron-loss or gain have been suggested to occur in the mitochondria or repetitive loci such as the nuclear rDNA (Lambowitz, 1989; Eickbush, 1999).

6.4.3. The *rns* intron landscape:

During the current study, twelve insertions have been characterized within the *rns* gene of ascomycetous fungi and some eukaryotes (Figure 6.1.); for this reason the *rns* gene appears to be a natural reservoir for introns (ribozymes) and IEPs (HEases) that over time moved into other loci (see Haugen and Bhattacharya 2004). The mitochondrial genome is highly variable due to the presence of introns and IEPs and the current study provides an updated overview of the introns and twintrons that can interrupt the *rns* gene. From a practical point of view this survey also suggests that the *rns* gene might be a good target for bioprospecting for ribozymes and endonucleases that can be biochemically characterized and used in several biotechnological applications (Hausner, 2003; Gimble, 2005; Lambowitz et al., 2005; Marcaida et al., 2010; Lambowitz and Zimmerly, 2011; Hafez and Hausner, 2011a).

The sequence logos for intron insertion sites confirmed that group I and II introns insert in highly conserved regions; this is probably in part due to the fact that IEP (such as HEases) and the group II EBS require large recognition sites thus favouring conserved regions within the rDNA as targets. It has also been noted that introns at the same insertion site are more closely related to each other than to introns inserted at different

positions; this confirms the idea that composite mobile elements, due to the specificities of their IEPs are predisposed to invade primarily one target site.

It has been suggested that rRNA sequences might "attract" introns and their ORFs due to their conserved nature and "universality". This provides mobile elements opportunity to move horizontally to avoid extinction by saturation of all insertion sites too quickly and thus being eroded by the accumulation of neutral mutations. These elements are subjected to drift as they are not expected to be subject to selection (Goddard and Burt, 1999). However it has been noted by Jackson et al. (2002) that there are possible constraints as to where introns insert or not. For example, very few intron positions correspond with sites which interact with ribosomal proteins; instead insertion sites tend to be in close proximity to tRNA binding sites (Jackson et al., 2002). Also possible insertion sites are constrained by the folding of the flanking rRNA exon regions as group I and II introns require the formation of a precise structure that involves some flanking exon sequences and these have to be available and not be competed away by rRNA folding. Conversely rRNA sequences are not to interact with certain intron sequences and thereby interfere with the proper folding of the intron RNAs.

6.4.4. Spotty distribution of group I and II introns in the *rns* gene:

The species tree (Figure 6.3.) was compared to the distribution of the various introns uncovered in this study and there was no obvious connection between relatedness among the various species/strains and their intron composition. Overall theses introns have a spotty distribution, which can be explained by intron gains by vertical and/or

horizontal transfers followed by rapid intron loss (Figure 6.3.). The mobility of group I and II introns (homing and retrohoming respectively) to homologous sequences is initiated by the IEP. Group I and II introns can also move to new (ectopic) sites via reverse splicing which requires less homology (4 – 6 nucleotides). The reverse splicing step is then followed by reverse transcription to generate cDNA followed by homologous recombination that integrates the intron-containing cDNA into the genome. As reverse splicing requires less homology, this mechanism allows introns to spread more easily into heterologous sites. One possible scenario that could explain the precise loss of an intron is when an intron-containing host gene transcript after being processed (i.e. the intron was removed from the pre-mRNA) undergoes reverse transcription followed by a recombination event that replaces the genomic intron-containing version of the gene with the intron-minus cDNA (Roman and Woodson, 1995; Birgisdottir and Johansen, 2005; Hafez et al., 2012).

6.4.5. Unusual group II introns in the *rns* gene:

6.4.5.1. Group II introns encoding HEases: The intron ORF host switches:

Typically group II introns encode RT-like ORFs, but recently a new subfamily of group II introns was found to encode LAGLIDADG ORFs typical of group I introns. LAGLIDADG ORFs are wide spread elements that can be inserted in group I introns, archaeal introns and inteins. Several examples of this new intron-ORF combination have been reported from the mtDNA *rns* gene (mS952 and mS788) and LSU rRNA (mL2059) genes (Toor and Zimmerly, 2002). Based on the secondary structure models, so far all the LAGLIDADG-encoding introns are group IIB1 introns and the LAGLIDADG ORF is

inserted in the loop of domain III (mS952), or domain IV (mS788 and mL2059) which is the insertion site of most RT-like ORFs of group II introns (Michel and Ferat, 1995; Toor and Zimmerly, 2002; Mullineux et al., 2010; Hafez and Hausner, 2011a & 2011b).

The origins of group II intron LAGLIDADG ORFs are uncertain. The closest relatives of these ORFs are fungal group I intron encoded LAGLIDADG ORFs suggesting that these ORFs invaded group II introns within the fungal mitochondria (Toor and Zimmerly, 2002). In a phylogenetic analysis of 43 fungal double motif LAGLIDADG ORFs, a group I-encoded LAGLIDADG ORF (interrupts the mitochondrial ND5 gene) branched basal to the mS952 group II- encoded LAGLIDADG ORFs, which suggested that group II introns' LAGLIDADG ORFs most likely originated from mtDNA group I intron ORFs (see Chapter 4; Hafez and Hausner, 2011a). The LAGLIDADG encoding group II intron ORFs appear to be an example of a "host switch" where a group I intron-encoded LAGLIDADG ORF moved from group I intron into a group II intron. Two mS952 group II intron encoded LAGLIDADG ORFs (I-LtrII; Mullineux et al., 2010 and I-OmiII; Chapter 7 & Hafez and Hausner, 2011c) were biochemically characterized and are found to be active and functional HEases. They both recognize and cleave the target site 2 nucleotides upstream of the mS952 position in the intron-minus allele generating 4 nucleotide 3' overhangs and thus the HEases could facilitate the mobility of this intron by a homing mechanism similar to that for group I introns. Another example of an unusual group IIB1 intron that encodes a GIY-YIG ORF was found in *Amoebidium parasiticum* at position mS788 (Li et al., 2011). Overall these observations confirm that HEGs are mobile elements that can move independently of

their ribozyme counterparts; however they tend to invade niches that are "neutral" such as self-splicing introns or inteins.

6.4.5.2. Group II introns with novel RT-ORF location:

Another unusual group II intron is found at position mS379. What makes this intron unique is that the RT-ORF is located in DII as usually RT-like ORFs in organellar group II introns are located in DIV. This intron was the only group II intron in the *rns* gene of the fungi examined in this study that encoded an RT-ORF and this is considered to be unique as this intron represents the first organellar group II intron that has an RT-ORF encoded outside domain IV. The mS379 intron was characterized in *Ophiostoma torulosum* WIN(M)860 and *Ophiostoma hyalothecium* WIN(M)848, and previously this intron was reported from three *O. minus* strains WIN(M)472, WIN(M)494 and WIN(M)515 (see Chapter 4 and Chapter 5; Hafez and Hausner 2011a & 2011b). The *O. hyalothecium* version of this intron is unique because of its complete lack of an ORF. Otherwise the *O. torulosum* intron is similar to those observed in *O. minus* where the ORFs show evidence of erosion due to the presence of mutations that introduce frameshifts and premature stop codons. In general all versions of this intron had degenerated RT ORFs, or in one example a complete loss of the ORF, suggesting that this intron might be in the process of being eliminated (see Chapter 5 for more discussion on this matter) from the *rns* gene as it most likely has lost its mobility-promoting IEP.

6.4.6. Mutualism between mobile introns and host genes:

Group I and II introns are sometimes described as molecular parasites that can interrupt genes within the nuclear rDNA (for group I introns) and organellar genomes in all domains of life. They are self-splicing elements (ribozymes) that are spliced at the RNA level during the transcript maturation process and thereby the introns do not disrupt their host genes. Group I and II introns can also promote their own mobility within and between genomes by encoding ORFs (HEases or RTs) and therefore these elements can be considered selfish elements with their primary purpose being to increase their copy number by invading empty target sites. In general mobile introns and their IEPs are considered neutral elements that impart no phenotype on the host organism, but recently it has been shown that some HEases can affect the expression of their host gene and potentially influence cellular metabolism (Edgell et al., 2011). The so called "social networking" between mobile introns, their ORFs and their host genes can be clearly illustrated by the synergistic co-evolution between the following three "parties": 1. the bacteriophage thymidylate synthase (TS) as a host gene, 2. the *td* group I intron as the self-splicing component, and 3. the intron-encoded I-TevI HEase as the agent that can mobilize the intron (Gibb and Edgell, 2010; reviewed in Stoddard and Belfort, 2010). It has been shown that the efficient splicing of the *td* intron requires delayed translation of the I-TevI due to the extension of I-TevI ORF into the intron's catalytic core and therefore the translation of the I-TevI prior to intron splicing will affect the timing of the *td* intron splicing and therefore reduces the host TS synthesis (Gibb and Edgell, 2010). An example of the intron ORF overlapping with the group I intron core was found during this study within the O.mi-mS569 group IC2 intron in *O. minus* (Chapter 4; Hafez and

Hausner, 2011a) where the predicted LAGLIDADG ORF (370 amino acid residues) is located in the loop of P9.1 region and extends downstream into the intron catalytic core with the start codon located in P5b region (see Table 1 and "Edgell et al. 2011" for more examples of ORFs overlap with core sequences of group I introns). So the interactions between the introns, their ORFs and the host gene in which they are embedded might be for more complex than was previously appreciated.

Proper intron splicing is a critical and required step for mRNA maturation and the production of functional host gene products. Any deficiency in intron splicing will contribute to the disruption of the production of the host genes final product. It has been shown that splicing deficiency of the Cpa-mS62 group IIA1 intron disrupts the maturation of the *rns* transcript and also appears to decrease the pathogenicity of certain *Cryphonectria parasitica* (Chestnut blight) strains (Baidyaroy et al., 2011). The term hypovirulence is used in describing strains of fungal pathogens wthat have reduced virulence and thus these strains do not damage the infected trees. Hypovirulence can be induced by the presence of certain mycoviruses or genetic mutations (Bertrand, 2000). Hypovirulent strains could be used as biocontrol agents as these strains could convert virulent strains to hypovirulence after anastomosis (hyphal fusion) whereby agent(s) that cause/induce hypovirulence can be transferred (Bertrand, 2000). In nature hypovirulence might be an adaptive trait as it prevents an aggressive pathogen to become extinct by eliminating its host species.

CHAPTER 7:

BIOCHEMICAL CHARACTERIZATION OF THE INTRON-ENCODED I-OmiI AND I-OmiII LHEases FROM THE BLUE-STAIN FUNGUS *Ophiostoma minus*.

7.1. ABSTRACT:

The mitochondrial small subunit ribosomal RNA (*rns*) gene of the ascomycetous fungus *Ophiostoma minus* was found to be invaded with several group I and group II introns. In the *O. minus* WIN(M)371 *rns* gene, group IC2 and group IIB1 introns are inserted at positions mS569 and mS952 respectively. The mS569 intron encodes a double motif LHEase (I-OmiI) from an ORF located partially within the P9.1 region and extending downstream into the P5b region where the start codon is located. The mS952 intron also encodes a double motif LHEase (I-OmiII) from an ORF embedded within domain III of the group II intron. Codon-optimized versions of these ORFs, I-OmiI and I-OmiII, sequences were synthesized based on differences between the fungal mitochondrial and bacterial genetic code. The optimized ORFs were cloned into expression vectors and transformed into *E. coli* BL21 star (DE3) for protein expression. Characterization of the protein showed that: I-OmiII is a functional HEase that cleaves the target site two nucleotides upstream (sense strand) of the intron insertion site, generating 4 nucleotide 3′ overhangs. The endonuclease activity of I-OmiII was tested using a variety of possible substrates and the optimum temperature of I-OmiII was also estimated. The *in vitro* cleavage assays of I-OmiI indicate that this enzyme is an active

HEase but the pET expression system and the Ni-NTA purification method might not be suitable to express and purify the enzyme in a soluble form in sufficient concentration to allow for complete biochemical characterization.

7.2. INTRODUCTION:

HEases are encoded by HEGs which are embedded within group I introns, group II introns and archaeal introns, as well as inteins. HEases are named based on conserved amino acid motifs. The LAGLIDADG and GIY-YIG families of HEases are most frequently encountered among fungal mitochondrial group I introns (Stoddard, 2006). Enzymes possessing the LAGLIDADG motif cleave DNA generating four-base 3'-OH overhangs at their DNA target sites. The DNA recognition sequences are generally asymmetrical and long (14-40 bp; Stoddard, 2006; Belfort & Roberts, 1997). Double motif LHEases act as monomers, and possess a pair of structurally similar nuclease domains on a single peptide chain, and are therefore not constrained to symmetric DNA targets (Agaard et al., 1997; Dalgaard et al., 1997; Lucas et al. 2001) like their single motif LHEase ancestors (Haugen and Bhattacharya 2004). HEases usually target cognate alleles that lack HEGs or introns. HEases have specific target sites, but with some allowance for sequence variation in their recognition sites. This potentially allows for the propagation of HEases against the forces of evolutionary drift which might modify the homing site within its host gene.

HEases require long DNA recognition sites and therefore cut infrequently within a genome; this makes them useful for DNA engineering and genomics. HEases can be engineered to cleave at desired locations and therefore HEases can become site specific tools that can be used to target specific genes for gene replacement (Stoddard, 2011). Strategies are also being developed that would allow for therapeutic applications of HEases that target genes associated with human diseases (Barzel et al., 2011). Recently the value of native HEases has been demonstrated by Takeuchi et al. (2011) on work based on HEGs inserted within the mtDNA *rps3* gene; these HEGs were first characterized in ophiostomatoid fungi such as *Ophiostoma novo-ulmi* subspecies *americana* and *Letographium truncatum* (Gibb and Hausner, 2006; Sethuraman et al., 2009a & 2009b). In the current study we examine potential HEases encoded within *rns* introns in the blue-stain fungus *Ophiostoma minus*. This species is an economically important agent of blue-stain (sap-stain) to various pine species (Gorton and Webber, 2000) and this fungus has been demonstrated to be a pathogen of pine (Masuya et al., 2003; Gorton et al., 2004; BenJamaa et al., 2007). Blue-stain is a discoloration in the wood of infected trees or timber and this stain can greatly reduce the value of the wood (Zink and Fengel 1990).

Our goals in the current study were to characterize recently discovered HEG-like sequences within the *rns* gene of *O. minus* (Chapter 4 & Hafez and Hausner, 2011a). The HEGs were noted in introns ms569 and mS952 and the corresponding proteins are named I-OmiI and I-OmiII respectively. This work involved the over expression of the intron

encoded proteins and its purification, and to examine the potential endonuclease activity of these enzymes. In addition we mapped the HEase cleavage sites.

7.3. METHODS OVERVIEW:

Codon optimized versions of I-OmiI and I-OmiII were cloned into pET200/D-TOPO expression vector and cloned into *E. coli* BL21 (DE3) and IPTG used to induce the expression of both proteins. The 6xHis-tagged I-OmiI and I-OmiII proteins were purified and the endonuclease activity for each enzyme was tested *in vitro* and finally the cleavage site of I-OmiII was mapped. For detailed methodology see Chapter 2 (for I-OmiI and I-OmiII), and Appendix 9.2. (for preliminary work on the overexpression and purification of I-OmiI).

7.4. RESULTS:

7.4.1. *O. minus rns* and the O.mi-mS569 and mS952 introns:

The *rns* gene of *O. minus* WIN(M)371 (accession number HQ292071) was amplified and sequenced previously (see Chapter 4), and the gene was found to be interrupted by group I and group II introns at positions mS569 and mS952 respectively (Figure 7.1.). The mS569 and mS952 introns encode double motif LHEases (I-OmiI and I-OmiII respectively), the former encodes a 370-amino acid protein (accession number HQ 292082), while the later encodes a 306-amino acid protein (accession number HQ292080). The RT-PCR analysis was applied to study the splicing of *O. minus* WIN(M)371 mS569 and mS952 introns *in vivo* and the results showed that both introns can spliced out from the pre-RNA (Figure 7.2.).

Figure 7.1. A schematic overview of the *O. minus* WIN(M)371 *rns* gene showing the various intron types. The introns are located at positions S569 and S952 with reference to the *E. coli* SSU rRNA sequence**.** The O.mi-mS569 is a group IC2 intron, while the O.mi-S952 intron is a group IIB1 intron and they both encode I-OmiI and I-OmiII double motif LHEases respectively.

Figure 7.1.

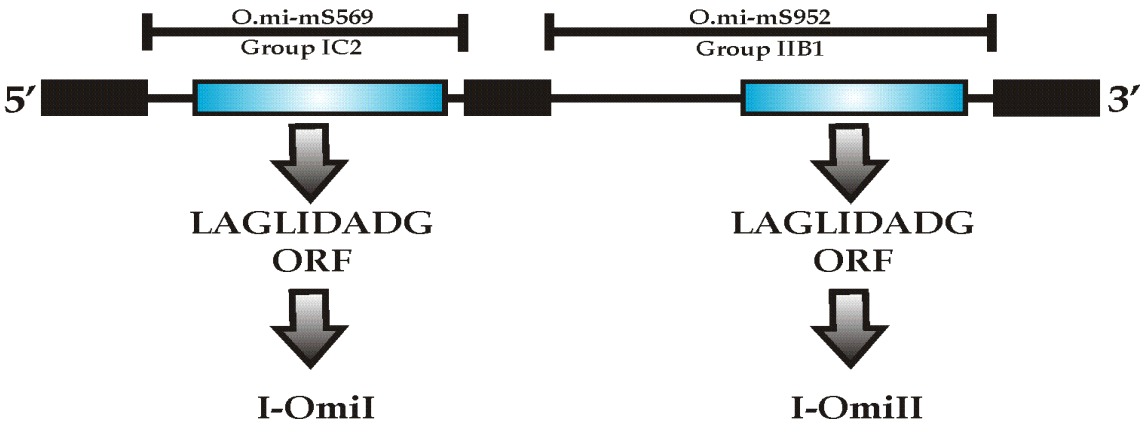
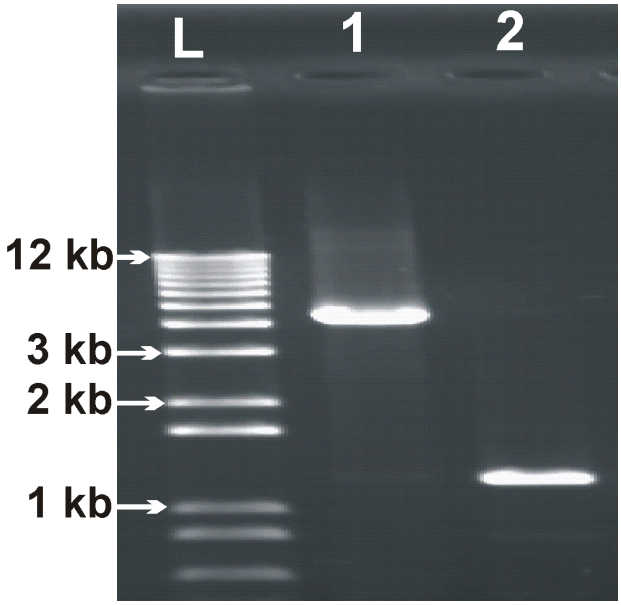

O.mi-mS569
Group IC2
O.mi-mS952
Group IIB1
5′
3′
LAGLIDADG
ORF
LAGLIDADG
ORF
I-OmiI
I-OmiII

Figure 7.2. RT-PCR analysis for demonstrating the *in vivo* splicing of the O.mi-mS569 and O.mi-mS952 introns. The *rns* transcripts for the *O. minus* WIN(M)371 were analyzed by RT-PCR. The standard PCR reaction using genomic DNA as a template is shown in lane 1 and it generated a 4.4-kb fragment. The amplicon length for the cDNA was 1.2 kb (lane 2) indicating that both O.mi-mS569 and O.mi-mS952 introns were spliced out. The lane marked "L" contains the 1 kb plus ladder (Invitrogen).

Figure 7.2.



7.4.2. Expression and purification of I-OmiI:

I-OmiI was over expressed in *E. coli* BL21 and the results showed that, I-OmiI was only found in the insoluble fraction; this may be due to misfolding and aggregation of the protein. Several methods have been applied to increase the solubility of I-OmiI, such as slowing down the growth of *E. coli* to allow for the proper folding of the enzyme or preventing the expressed proteins from aggregating together, or by treating the cell lysate with DNase in the presence of high salt concentration in order to dissociate the protein from the DNA. Overall it was determined that using sarkosyl (0.5-2 % w/v) was the most efficient method for solubilizing the protein from inclusion bodies (Figure 7.3. & Figure 9.2.5.; see Appendix 9.2. for more details on the various approached utilized to improve the expression and purification of this protein from *E. coli*).

7.4.3. *In vitro* cleavage assay of I-OmiI:

Briefly it can be stated that the endonuclease activity of I-OmiI partially solubilized from inclusion bodies was tested against the substrate plasmid pCR4/1574 (i.e. contains I-OmiI target site). The *in vitro* cleavage assay results (Figure 7.4.) showed that I-OmiI appears to be an active HEase that recognized a site in the pCR4/1574 substrate plasmid and therefore generated a linearized molecule with a length of 5.2 kb. Another experiment to test the endonuclease activity of I-OmiI against the substrate plasmid pCR4/1574 was carried out using the 60% fraction of the ammonium sulphate crude cell lysate, which contained the over expressed I-OmiI protein (see Figure 9.2.6., Appendix 9.2.), and the results again showed that I-OmiI potentially is an active HEase.

Figure 7.3. Purification of I-Omi-I solubilized from inclusion bodies with sarkosyl (0.5:2 % w/v). The SDS-PAGE (10%) gel shows the results of the I-OmiI purification by Ni-NTA resin (Qiagen). Lanes are indicated as follows: S1 = supernatant without sarkosyl; S2 = supernatant with 0.5% sarkosyl; M = protein marker (BioRad); FT = Flow through; E = elution fractions (E1 and E2, 125 mM imidazole; E3 and E4, 250 mM imidazole). I-OmiI was eluted from S2 fraction. The red arrow indicates the 43 kDa I-Omi-I protein (the expected size of I-Omi-I is 40 kDa, but the size increases due to the N-terminal fusion tag).

Figure 7.3.

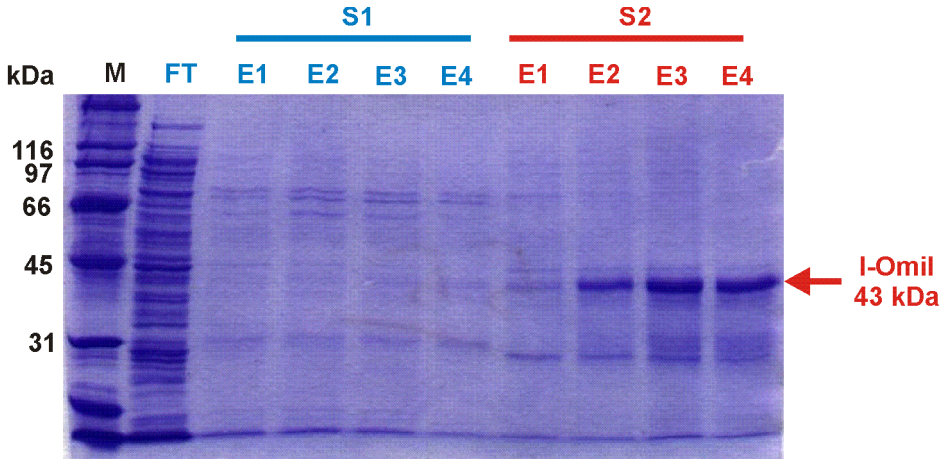
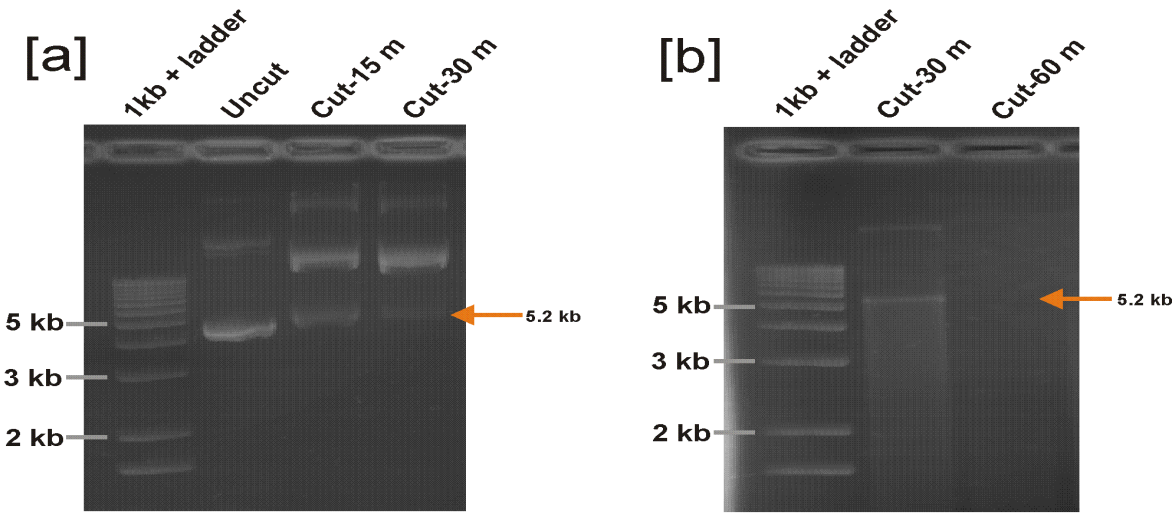


Figure 7.4. *In vitro* endonuclease assay for I-OmiI. The enzyme was incubated with [a] high template concentration (25ng/µl) for 15 and 30 minutes and [b] low template concentration (5ng/µl) for 30 and 60 minutes.

Figure 7.4.


[a]
1kb + ladder
Uncut
Cut-15 m
Cut-30 m
5 kb
3 kb
2 kb
5.2 kb
[b]
1kb + ladder
Cut-30 m
Cut-60 m
5 kb
3 kb
2 kb
5.2 kb

7.4.4. Expression and purification of I-OmiII:

After the induction period cells were harvested and lysed to prepare the clear lysate, which was used to purify the I-OmiII protein with the Ni-NTA resin. Purification of I-OmiII was then monitored by SDS-PAGE (Figure 7.5.). The clear lysate fraction displayed several bands and a major protein of 37 kDa was detected. The expected molecular weight of I-OmiII is 34 kDa, and due to the presence of the N-terminal fusion tag the size of I-OmiII has increased to 37 kDa. The 37 kDa band dominates the elution fractions (Figure 7.5. [a]), but in order to remove the non-specific bands, a second purification step was carried out using a HiTrap heparin HP column (GE Healthcare Europe). The applied sample was eluted with NaCl starting at 200 mM up to 900 mM NaCl in 100 mM intervals. The I-OmiII protein was eluted in the 400, 500 and 600 mM NaCl fractions (Figure 7.5. [b]). These three fractions were combined and concentrated with the Amicon® Ultra centrifugal filter (Millipore Corporation, Billerica, MA, USA).

7.4.5. *In vitro* cleavage assay I-OmiII:

I-OmiII endonuclease activity was tested *in vitro* by incubating I-OmiII with circular (Figure 7.6. [a]) and linearized versions of the substrate plasmid (Figure 7.6. [b]). The endonuclease activity of I-OmiII was tested at three time points (0, 30 and 60 min), and the results showed that I-OmiII completely cleaves and linearized the substrate plasmid (pCR4/1574) after 60 min, I-OmiII also cleaved the substrate plasmid linearized with *Nco*I enzyme into two fragments (2194 and 2973 bp).

Figure 7.5. Purification of I-OmiII. **[a]** SDS-PAGE (10%) gel of I-OmiII purification by Ni-NTA resin (Qiagen). Lanes are indicated as follows: M = marker; CL = clear lysate; FT = flow through; W wash, and E = elution. I-OmiII was washed with a narrow range of imidazole (20, 30 and 40 mM) and eluted in three fractions E1 (125 mM imidazole), E2 and E3 (250 mM imidazole). The red arrow marked the 37 kDa I-OmiII protein (the expected size of I-OmiII is 34 kDa, but the size increase due to the N-terminal fusion tag). **[b]** 10% SDS-PAGE gel for the second purification step of I-OmiII. The elution fractions (E1-E3) from the first purification step were combined and applied to a HiTrap heparin HP column (GE Healthcare) and then washed with one column volume wash buffer over a range of 200 mM to 900 mM NaCl as indicated by the red left-facing triangle. The I-OmiII eluted in three fractions (400, 500 and 600 mM NaCl).

Figure 7.5.

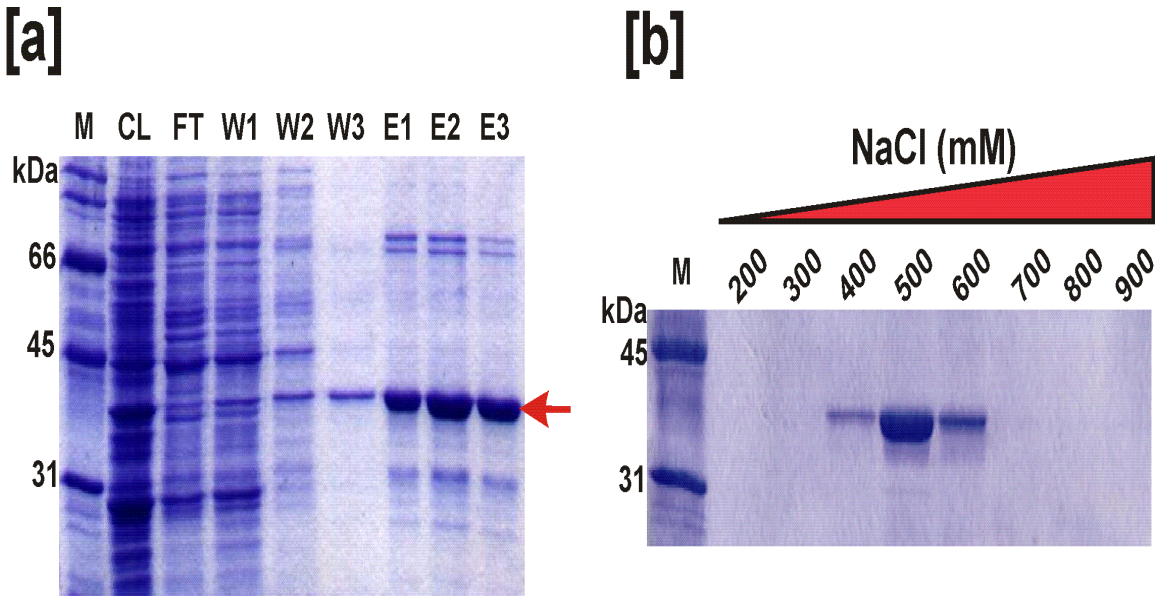
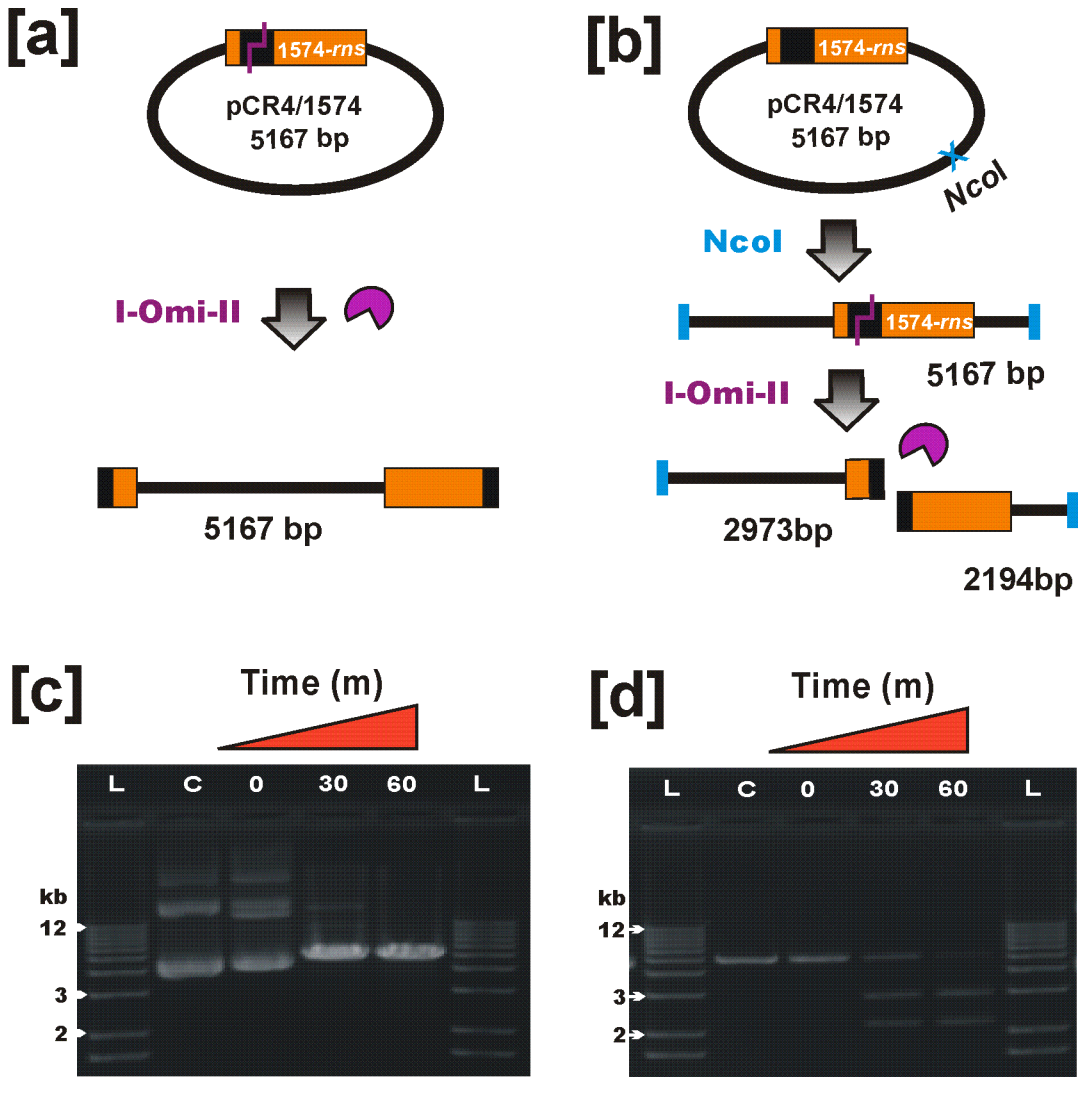

[a]
M CL FT W1 W2 W3 E1 E2 E3
kDa
66
45
31
[b]
NaCl (mM)
M 200 300 400 500 600 700 800 900
kDa
45
31

Figure 7.6. *In vitro* cleavage assay with I-OmiII. Schematic representation **[a]** and gel image **[c]** for the ability of I-OmiII (purple pac-man symbol) to cleave/linearize the pCR4/1574 plasmid, which contains the mt-*rns* gene (orange box) with the recognition site of I-OmiII (black box). The circular pCR4/1574 plasmid incubated with I-OmiII for 0, 30 and 60 minutes. I-OmilII completely linearized the plasmid after 1 hour generating a 5167 bp fragment. Schematic representation **[b]** and gel image **[d]** for the ability of I-OmiII to cleave the pCR4/1574 plasmid that was previously linearized with *Nco*I. The linearized plasmid was also incubated with I-OmiII for 0, 30 and 60 minutes and the protein cleaved the template into two fragments (2194 and 2973 bp).

Figure 7.6.


[a]
1574-rns
pCR4/1574
5167 bp
I-Omi-II
5167 bp
[b]
1574-rns
pCR4/1574
5167 bp
NcoI
NcoI
1574-rns
5167 bp
I-Omi-II
2973bp
2194bp
[c]
Time (m)
L C 0 30 60 L
kb
12
3
2
[d]
Time (m)
L C 0 30 60 L
kb
12
3
2

The endonuclease activity of I-OmiII was also tested in comparison with the *Nco*I REase, the pCR4/1574 was incubated with I-OmiII and *Nco*I (separately) and the results indicated that I-OmiII completely cleaved the substrate plasmid as efficiently as *Nco*I (Figure 7.7.; lanes 3 and 4 respectively). In contrast, no cleavage was observed when I-OmiII was incubated with the pCR4/371 plasmid (linearized with *Nco*I); this plasmid is a negative control as the I-OmiII recognition site is interrupted with the mS952 intron (Figure 7.8.).

7.4.6. I-OmiII Cleavage site mapping:

Usually, LHEases cleave the target site generating 4 nucleotide 3′ overhangs. The ability of T4 DNA polymerase to catalyze the repair of 5' overhangs and the hydrolysis of 3' overhangs can be used to determine the cleavage patterns produced by endonucleases. Sequencing of the religated I-OmiII cleaved, T4 DNA polymerase treated pCR4/1574 plasmid showed that 4 nucleotides had been removed, GTAT (sense strand); these four nucleotides therefore represent the 3' 4 nt overhangs generated by the I-OmiII endonuclease. With regards to the sense strand this HEase cleaves 2 nucleotides upstream of the mS952 intron insertion site (Figure 7.9.).

7.4.7. Optimum temperature for I-OmiII endonuclease activity:

The ability of I-OmiII to cleave the linearized substrate plasmid (pCR4/574) was tested at different temperatures (from 10 ^{o}C to 90 ^{o}C). I-OmiII is active and cleaves the recognition site in a wide temperature range (from 10 ^{o}C to 50 ^{o}C), and the highest activity was recorded when incubating the enzyme at 40 ^{o}C (Figure 7.10.).

Figure 7.7. Comparison between the endonuclease activity of the LHEase I-OmiII and the REase *Nco*I. The pCR4/1574 plasmid which contains the recognition site of both I-OmiII (black box), and *Nco*I (blue X) was incubated with *Nco*I and/or I-OmiII for 60 minutes at 37 ^{o}C. Lanes 1 and 6 contain the 1 kb plus ladder; lane 2 contains undigested pCR4/1574; lane 3 contains pCR4/1574 digested with *Nco*I REase; lane 4 contains pCR4/1574 digested with I-OmiII; lane 5 contains pCR4/1574 digested with both *Nco*I and I-OmiII.

Figure 7.7.

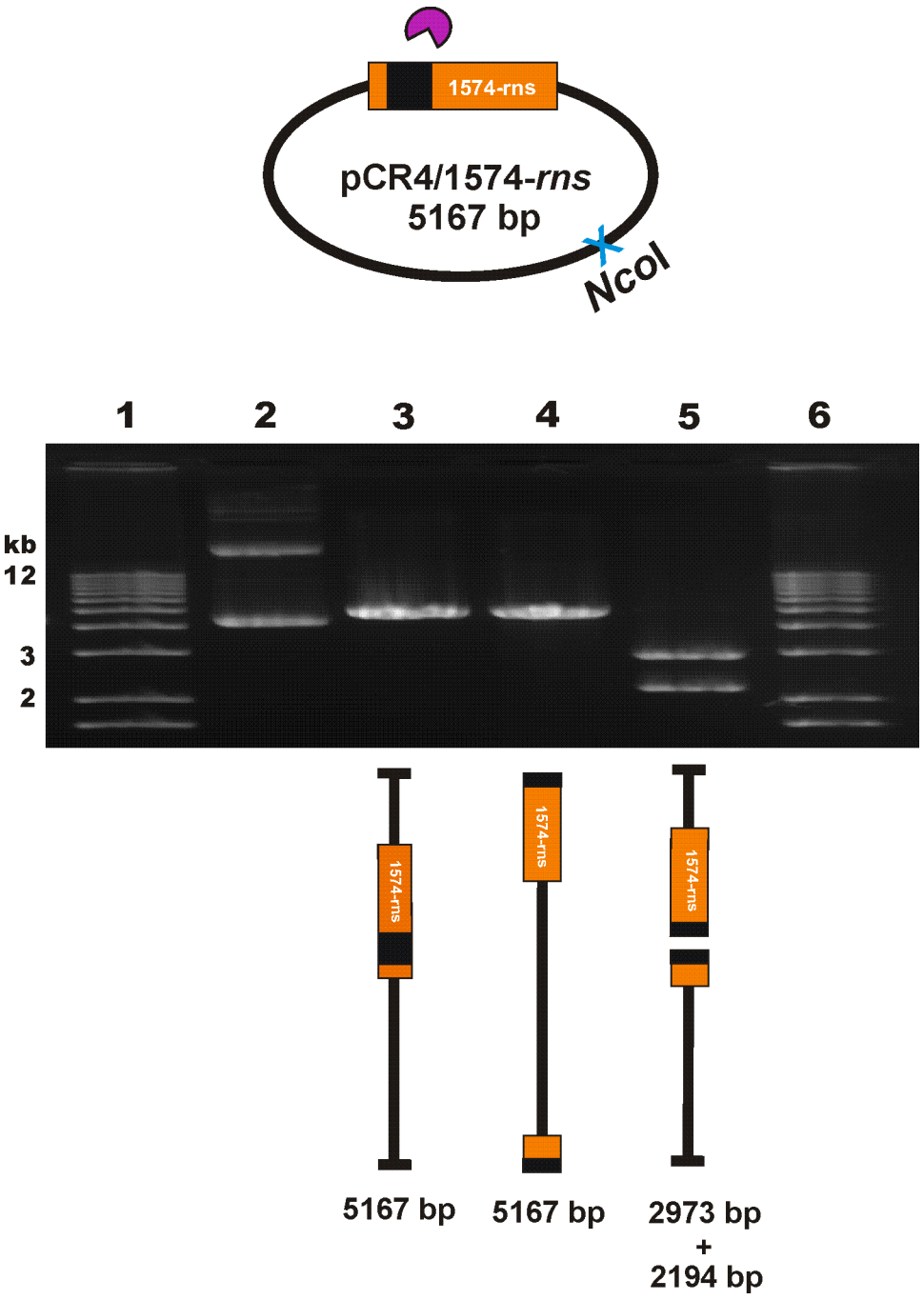

1574-rns
pCR4/1574-rns
5167 bp
NcoI
1
2
3
4
5
6
kb
12
3
2
1574-rns
1574-rns
1574-rns
5167 bp
5167 bp
2973 bp
+
2194 bp

Figure 7.8. A schematic representation **[a]** and gel image **[b]** of the *in vitro* cleavage assay with I-OmiII. The plasmid pCR4/371 contains the *rns* gene amplified from *O. minus* WIN(M)371 (orange bar). The I-OmiII recognition site is interrupted with mS952 intron (black box with x mark). The pCR4/371 was initially linearized with *Nco*I to generate a 8357 bp fragment (lane C) and then incubated with I-OmiII (purple pac-man symbol) for 0, 30 and 60 m (lanes 0, 30 & 60 respectively). I-OmiII failed to cleave the linearized plasmid because the enzyme's recognition site is interrupted by the mS952 intron.

Figure 7.8.

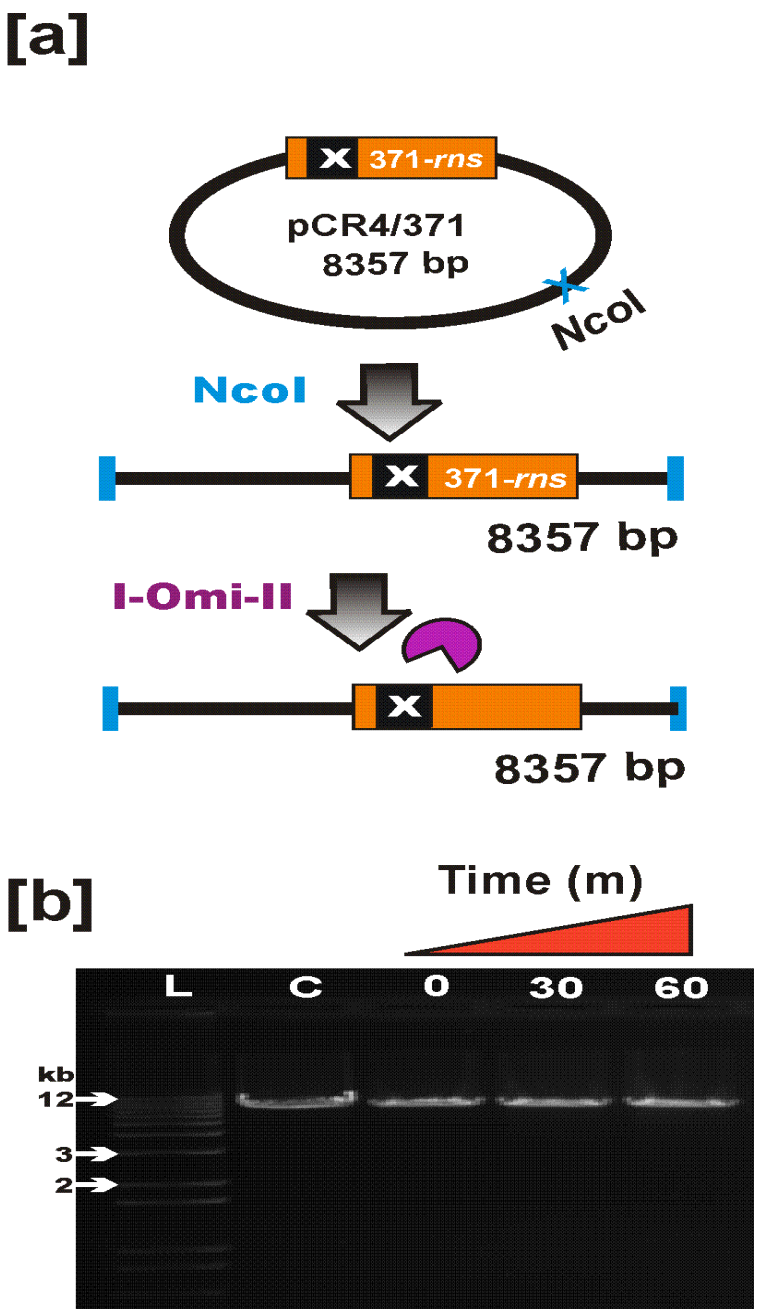

[a]
371-rns
pCR4/371
8357 bp
NcoI
NcoI
371-rns
8357 bp
I-Omi-II
8357 bp
[b]
Time (m)
L
C
0
30
60
kb
12
3
2

Figure 7.9. [a] Cleavage site mapping of I-OmiII by DNA sequencing of the pCR4/1574 and pCR4/1574-T4 plasmids. The pCR4/1574-T4 plasmid was obtained from pCR4/1574 which was cleaved with I-OmiII (purple pac-man symbol) generating 4 nucleotide 3′ overhangs and then blunt ended with T4 DNA polymerase. **[b]** Schematic representation of the *rns* sequence that flanks the mS952 intron insertion site (I.S.; represented by blue double head arrow). The cleavage site of I-OmiII (C.S.; represented by dashed line) is found to be 2 and 6 nucleotides upstream of the intron insertion site for the sense and antisense strands respectively.

Figure 7.9.

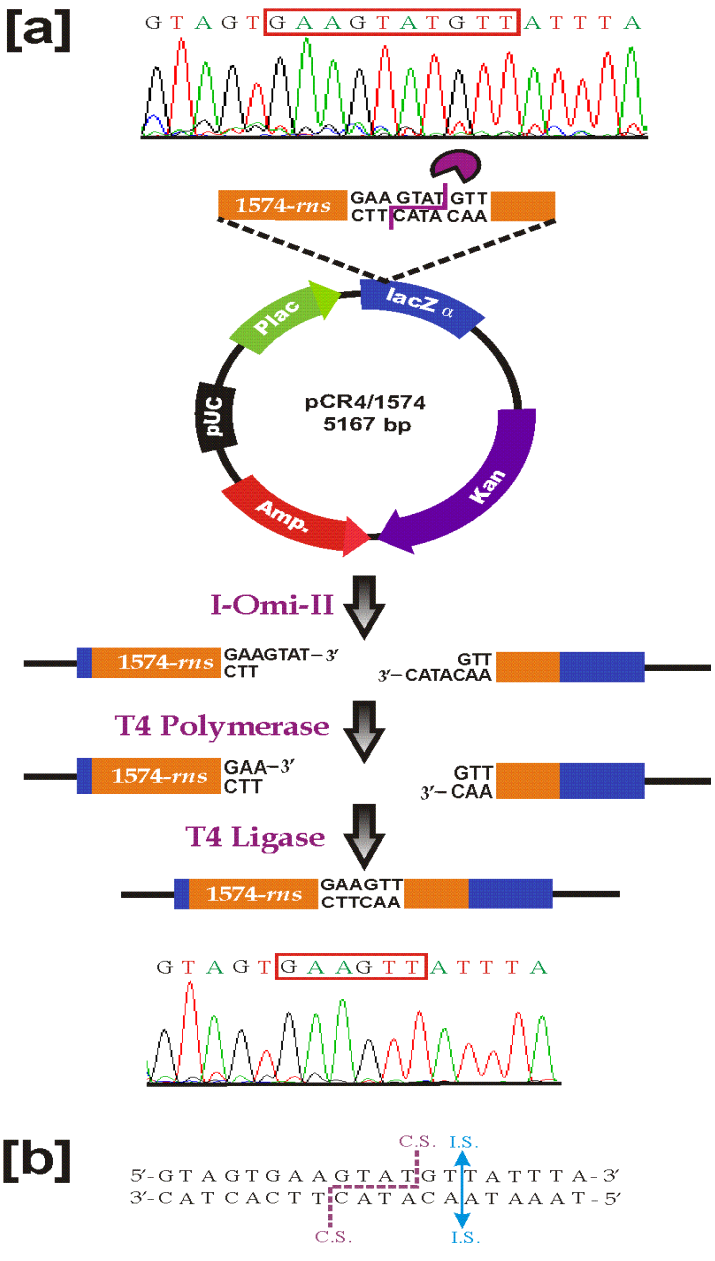
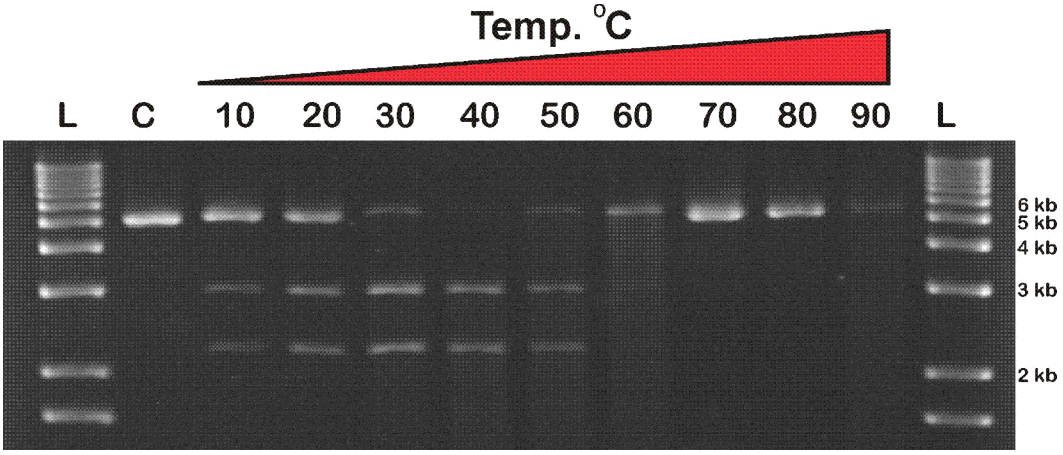

[a]
G T A G T G A A G T A T G T T A T T T A
1574-rns
GAA GTAT GTT
CTT CATA CAA
lacZ α
Plac
pUC
pCR4/1574
5167 bp
Amp.
Kan
I-Omi-II
1574-rns
GAAGTAT–3′
CTT
GTT
3′–CATACAA
T4 Polymerase
1574-rns
GAA–3′
CTT
GTT
3′–CAA
T4 Ligase
1574-rns
GAAGTT
CTTCAA
G T A G T G A A G T T A T T T A
[b]
C.S.
I.S.
5′-G T A G T G A A G T A T G T T A T T T A-3′
3′-C A T C A C T T C A T A C A A T A A A T-5′
C.S.
I.S.

Figure 7.10. Effect of temperature on I-OmiII endonuclease activity. The C (= control) lane contains the linearized substrate plasmid (pCR4/1574) without being treated with I-OmiII, while the L lanes contain 1 kb plus DNA ladder (Invitrogen).

Figure 7.10.



7.5. DISCUSSION:

The *rns* gene of *O. minus* WIN(M)371 was found to be interrupted by two introns, the mS569 group IC2 intron and the mS952 group IIB1 intron and they both encode double motif LHEases. The study demonstrated that the blue-stain fungi are a reservoir for native HEGs that appear to be active and thus could be used in various biotechnological applications.

7.5.1. Comments on I-OmiI HEase:

Preliminary work on the overexpression and purification of I-OmiI (see Appendix 9.2.) indicate that this enzyme could be an active HEase, but the pET expression system and the Ni-NTA purification method might not be suitable to express and purify the enzyme in a soluble form in sufficient concentration to allow for complete biochemical characterization. During this study it was repeatedly noticed that the I-OmiI did not "stick" to the Ni NTA column, suggesting that the 6xHis-Tag on the protein was not effective. This could be the result of misfolding of the N-terminal segment of the protein and thus the 6xHis-Tag is not accessible or available for binding to any ligand. Using a C-terminal 6xHis-Tag I-OmiI protein also did not solve this problem, as the C-terminus 6xHis-Tag enzyme also did not bind to the Ni-NTA resin exactly like the N-terminus 6xHis-tagged protein. In order to overcome this problem, other expression systems might have to be explored. For example the pMAL protein fusion and purification system (New England BioLabs) might be an alternative system; there are claims that it is more suitable for proteins that are difficult to purify due to solubility issues or an ineffective His-Tag.

7.5.2. Comments on I-OmiII HEase:

This study showed that the mS952 IEP is an active endonuclease that cleaves 2 nt upstream of the intron insertion site in a cognate allele that lacks the intron. This is the second example of an mS952 group II intron encoding an active LAGLIDADG type HEase. Previously is was demonstrated that in the mitotic species (asexual) *Leptographium trancatum* the mS952 group II intron encodes an active LAGLIDADG type HEase that did not appear to act as a maturase for the intron, but was capable of providing a means for mobility for this composite mobile element (Mullineux et al., 2010). This study now demonstrates that the mS952 intron in a sexual species also encodes an active HEase; this is important as for future work involving genetic manipulations and demonstrating mobility of this intron in crosses in a host species with a meiotic life cycle is absolutely essential. *Ophiostoma minus* might therefore become a suitable model system for studying group II introns that gained group I intron ORFs (Toor and Zimmerly, 2002; Mullineux et al., 2010). Goddard and Burt (1999) suggested a life cycle of HEGs consisting of 3 stages: horizontal transmission and fixation, followed by degeneration and eventually a precise loss. So in more detail the cycle starts with a HEG invading a site possibly within a self-splicing intron leading to the formation of a mobile intron/HEG composite element which spreads within a population by the help of the HEase activity of the HEG.

Eventually the HEase facilitates the invasion of all possible homing/target sites, at this point the HEG is no longer under selection pressure and starts to accumulate mutations and thus it will degenerate. In general mobile introns are viewed as neutral

elements in order to avoid damaging the host genome; however that means that natural selection will not act on this element and thus there is no selection for fitness. The degeneration stage facilitates the removal of the HEG and the intron from the genome. This event actually regenerates a possible homing site that will allow the homing cycle to be repeated. So, the presence of an active HEG (i.e. encoding an active protein) could be viewed as an indicator for a more recent DNA transfer event and thus one can speculate that the mS952 group II intron might still be in the invasion part of the HEG life cycle and maybe this group II introns has adapted towards a DNA-based mobility mechanism by acquiring a HEase ORF. Li et al. (2011) studied mitochondrial group IIB1 introns with unusual 5' terminal nucleotides insertions (1-33 nts inserted between the 5' exon and the consensus sequence, GUGYG, in the intron's 5' end), and the conclusion of this study was that these introns are adapted to DNA based mobility by encoding LHEase from within their DIII or DIV component of the group II introns.

7.5.3. The role of I-OmiII in mS952 intron mobility:

Mapping the cleavage site of I-OmiII indicated that the enzyme recognizes and cleaves the intron-less version of the *O. minus rns* gene 2 nucleotides upstream to the mS952 intron insertion site, as observed for similar LHEases (I-LtrII: Mullineux et al., 2010). This observation might reflect an expected role of I-OmiII protein in the mobility of its hosting intron (mS952), and the relationship between the mS952 group II intron and I-OmiII HEase could be a mutually beneficial relationship. The mS952 intron (as a self-splicing element) provides a neutral insertion site for the I-OmiII HEG, which would avoid disrupting the host gene; at the same time, I-OmiII initiates the mobility of mS952

intron (along with the I-OmiII ORF). More often, HEGs are associated with group I introns and inteins and due to the broad distribution of HEGs among unrelated elements it has been proposed that HEGs and their hosts (introns or inteins) have evolved independently (Dalgaard et al., 2009). So most likely I-OmiII invaded a group II intron, maybe by transferring from a group I intron to an ORF-less group II intron.

7.5.4. I-OmiI and I-OmiII as genome editing tools:

As described in Chapter 1, HEases can be engineered to target certain genes in human (gene therapy; Marcaida et al., 2010), plants (genome editing and targeted mutagenesis; Gao et al., 2010) or insects (pest control; Windbichler et al., 2011). However, enzyme engineering is a lengthy process and is expensive. Another draw back of engineered/manufactured HEases is the possibility of off-target effects which are very difficult to predict for synthetic HEases (Szczepek et al., 2007). On the other hand, with naturally occurring (native) HEases, the off-target effects (which could be toxic) may be more predictable (Barzel et al., 2011). Tapping the natural reservoir of native HEases could be an easier alternative to engineered HEases (Barzel et al., 2011; Takeuchi et al., 2011), but these HEases should be studied in more detail, including co-crystallizing these proteins with their native substrates. These experiments may show exactly how the proteins bind to their target sites and which amino acids might be involved in target site specificity. This may also provide a "blue print" for understanding the protein scaffold for DNA binding specificity and cleavage activity and may provide insight into what amino acids could be changed to retarget the HEase. Also data bases are currently being assembled that show HEases and their target sites and these can be screened against

sequences representing genes of interest that contain segments that are identical or highly similar to HEase target sites (i.e. can these HEases eventually be used in their native form or modified form to be used in biotechnology such as controlling pests or in gene replacement type experiments).

Both I-OmiI and I-OmiII could be powerful tools for targeting human genes involved in monogenic diseases. So as stated above using currently available data bases for endonucleases (Roberts et al. 2010) or NCBI databases using the I-OmiI and I-OmiII target sequences as queries, one can start to formulate strategies for targeting sequences in pathogenic organisms, vectors of pathogens or sequences in human genes involved in diseases. Essentially the next task is to scan for sequences that are identical or highly similar to the target sites for both I-OmiI and I-OmiII (see appendix 9.8. for more details about these sequence similarities). Also future efforts should include structural analysis of these proteins when bound to the native substrate; the crystal structure of I-OmiI and I-OmiII would provide a contact map showing the exact amino acid/DNA sequence interactions providing valuable information on strategies that could be used to modify the binding and cleavage activities of these DNA cutting enzymes.

CHAPTER: 8

GENERAL CONCLUSION

The *rns* genes of several groups of ophiostomatoid fungi have been examined, sequenced, and annotated in order to assess if these genes are a natural reservoir of mobile introns and intron-encoded proteins. In part, this work may also contribute towards a better understanding of the evolution of fungal mtDNAs. This study also provides useful information about the evolutionary dynamics of mobile introns and intron encoded-proteins, especially LHEases (Haugen and Bhattacharya, 2004). An overview of the various types of introns (group I and II) and intron-encoded proteins (HEases and RTs) that can invade the *rns* gene was generated and these observations may be a resource for those who try to annotate fungal mitochondrial genomes. This work also demonstrated that the *rns* gene is a good target for bioprospecting for potentially active HEases and ribozymes that have applications in biotechnology (Hausner, 2003; Gimble, 2005; Lambowitz et al., 2005; Marcaida et al., 2010; Lambowitz and Zimmerly, 2010).

8.1. Major findings of this thesis:

8.1.1. Characterization of a molecular marker to distinguish between *O. ulmi* and *O. novo-ulmi* subsp. *americana*:

In areas were both *O. ulmi* and *O. novo-ulmi* still exist, a molecular marker might be an alternative to cultural methodologies for distinguishing these two species. Due to the potential for the various DED agents to hybridize, monitoring for the presence of *O. ulmi* might be valuable in risk assessment for predicting future pandemics of DED.

In this study, we evaluated if the O.ul-mS952 intron could be used as a molecular marker that allows for the differentiation between strains of *O. ulmi* and *O. novo-ulmi* subsp. *americana*. The mS952 intron was absent in what is considered to be the more aggressive species, *Ophiostoma novo-ulmi*, but present in strains representing the less aggressive *Ophiostoma ulmi*. This study did not show any evidence for the introgression of the O.ul-mS952 intron into the tested *O. novo-ulmi* subsp. *americana* populations. Previously, Sethuraman et al. (2008) presented some evidence that some introns in the mtDNA *rnl* gene can move between these two species. So possibly there is a factor missing that prevents the mS952 intron from establishing itself within the mtDNA of *O. novo-ulmi* subsp. *americana.* In the future a series of forced crosses between these two species including strains with and without the mS952 intron should be initiated to investigate if this intron can cross the species barrier.

8.1.2. Characterization of novel introns and twintrons:

Comparative sequence analysis of the *rns* gene among different species of ophiostomatoid fungi and related taxa provided an overview of the intron landscape for this gene among species of *Ophiostoma* and quite possible other ascomycetes fungi. The *rns* gene appears to be a reservoir for a number of group I and II introns along with intron associated ORFs such as HEases and RTs. Some of the novel findings of this work including the discovery of a twintron complex inserted at position S1247 within the *rns* gene, here a group IIA1 intron invaded the open reading frame embedded within a group IC2 intron. Another new element was discovered within strains of *Ophiostoma minus*, where a group II introns has inserted at the *rns* position S379; the mS379 intron

represents the first mitochondrial group II intron that has an RT-ORF encoded outside Domain IV and it is the first intron reported at position S379. In general the *rns* gene appears to be the target for many mobile elements; this could be due to the conservative nature of rDNA sequences, which avoids the loss of target sites due to drift. Mobile elements such as group I and II introns need target sites that are conserved within species and among species to allow for horizontal transfers. The latter is key to the long term survival of mobile introns and their encoded HEGs.

8.1.3. Biochemical characterization of two novel LHEases: I -OmiI and I-OmiII:

The *rns* gene in strain WIN(M)371 was found to be interrupted with a group IC2 intron at position mS569 and a group IIB1 intron at position mS952 and they both encoded double motif LHEases referred as I-OmiI and I-OmiII respectively. These potential HEases were biochemically characterized and the results showed that both I-OmiI and I-OmiII are functional HEases. Detailed data for I-OmiII showed that this endonuclease cleaves the target site two nucleotides upstream (sense strand) of the intron insertion site generating 4 nucleotide 3′ overhangs. These findings confirm the original premise that the *rns* gene might be a good source of active HEases. The work also demonstrates that the *rns* genes and their introns and HEGs form a dynamic segment of the mtDNAs; with still active HEases being present in this gene it makes it possible for these introns to be mobile and invade *rns* genes that lack these introns and HEGs.

During the current study we noticed that the distribution of introns in the *rns* gene is not totally random and the introns appear to select insertion sites within highly conserved sequences. It has been previously reported that introns tend to insert near the tRNA and mRNA binding sites, suggesting a link between intron evolution and rRNA function (Jackson et al., 2002). Mobile introns and associated HEGs are also considered neutral elements that do not impart toxicity or a selectable phenotype to the host genome as HEGs offer their hosting introns the ability to be mobile, and the self-splicing mobile introns provide a neutral insertion site within the host genome. On the other hand, an intron splicing deficiency would be costly to the host; for example splicing deficiency of a mtDNA *rns* group II intron in *C. parasitica* was linked to growth abnormalities and hypovirulence (Baidyaroy et al., 2011). Examples of intron ORF degeneration due to frame shift mutations and the presence of premature stop codons were also observed during this study in a variety of introns suggesting that indeed some introns might be in the process of being eliminated from the *rns* gene. The example of the ms1247 twintron also illustrates the "co-operative" evolution of these elements towards one another whereby the insertion of the group II intron probably does not interfere with the expression of the host intron's ORF. Also, examples where intron ORF sequences overlap with intron core sequences is suggestive of complex co-evolution as the maintenance of the ORF is now under constraint to maintain also the ribozyme component of this composite element.

8.2. Future prospects:

8.2.1. *In vitro* intron splicing for mS915 and mS917 introns:

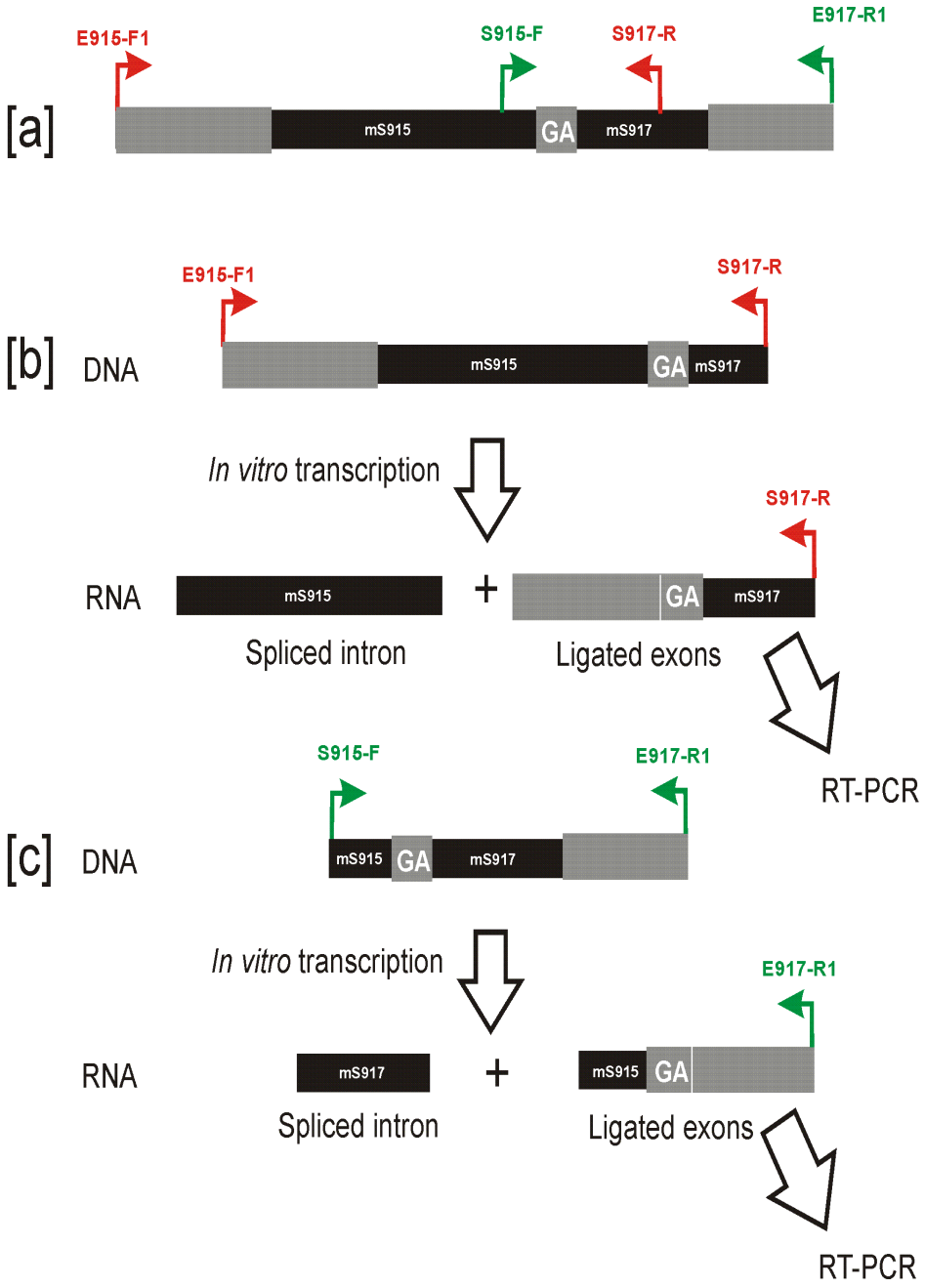
In Chapter 6, two group ID introns (mS915 and mS917) have been identified in the *rns* gene of *Cryphonectria parasitica*. Comparative sequence analysis of the *rns* gene derived from different ascomycetous fungi indicated that these two introns are separated by a very short two nucleotide exon (5'-GA-3'). This would be the shortest exon reported so far in the literature. Previously mtDNA short exons have been reported within the *coxI* genes that are estimated to be three nucleotides long (Cummings et al. 1989; Férandon et al. 2010a). The RT-PCR results for the *C. parasitica rns* gene only confirmed the presence of a two-nucleotide exon but it did not resolve the exact exon-intron junctions for the mS915 and mS917 introns. Currently the exon-intron junctions have been resolved based on secondary structure predictions for the mS915 and mS917 introns (see Figure 6.4., Chapter 6). To physically determine the exon-intron junctions one needs to investigate the actual splicing intermediates.

Studying intron splicing *in vitro* will help to pin point the 5'-GA-3' exon and to identify the exon-intron junctions more accurately. The problem is that there are several GA nucleotide pairs in this regions of the *rns* gene; therefore a method is needed that can discriminate which GA is maintained as an exon. Assuming that the mS915 intron splices first (as it encodes a double motif LHEase and the intron needs the GA exon to form the P10 helix) a transcript intermediate should exist with only the mS917 intron. To study the mS915 and mS917 intron splicing *in vitro*, each intron could be analyzed separately this requires that constructs are made for each intron region that includes about

400 upstream and 400 downstream nucleotides (see Figure 8.1.). The segments can be amplified by PCR and then cloned into suitable plasmids featuring a T7 RNA polymerase promoter. *In vitro* transcription of each PCR fragment could be performed by T7 RNA polymerase followed by splicing reactions that can be resolved by denaturing PAGE. In addition cDNAs can be synthesized for the transcripts by RT-PCR. Finally the cDNA products can be analyzed by sequencing to confirm the exon-intron junctions for each intron (Salman et al., 2012).

Figure 8.1. A schematic representation of a segment of the *C. parasitica rns* gene [a], to show the priming sites of the two primer pairs E915-F1/S-917-R and S915-F/E917-R, that were designed to amplify the mS915 [b] and mS917 [c] introns respectively along with upstream and down stream flanking sequences (see Appendix 9.3. for primer sequences).

Figure 8.1.


[a]
E915-F1
S915-F
S917-R
E917-R1
mS915
GA
mS917
[b]
DNA
E915-F1
S917-R
mS915
GA
mS917
In vitro transcription
RNA
mS915
+
GA
mS917
S917-R
Spliced intron
Ligated exons
RT-PCR
[c]
DNA
S915-F
E917-R1
mS915
GA
mS917
In vitro transcription
RNA
mS917
+
mS915
GA
E917-R1
Spliced intron
Ligated exons
RT-PCR

8.2.2. Characterize the mS1247 twintron:

A novel twintron element was discovered in the *rns* gene of the ascomycetous fungus *C. thermophilum* strain DSM 1495 (see Chapter 6 for more details). This twintron is inserted at position mS1247 and this element consists of an ORF-less group IIA that interrupts the double motif LAGLIDADG ORF of a group IC2 intron. The mS1247 twintron represents the first described mixed twintron consisting of a group II intron as an internal intron and a group I intron as an external intron and it also represents the first mitochondrial twintron. A secondary structure model for this twintron element was predicted (see Figure 6.5., Chapter 6) and the internal group II intron in the mS1247 twintron contains exon-binding sites (EBSs) complementary to their respective intron-binding sites (IBSs), which are located in the external group I intron ORF.

To confirm the predicted secondary structure model for this twintron, the splicing of the internal and external introns should be further investigated by studying the *in vivo* and *in vitro* splicing of the mS1247 twintron components (see section 2.8, Chapter 2 for *in vivo* splicing detailed methods and section 8.2.1., Chapter 8 for proposed *in vitro* experiment). The *C. thermophilum* mS1247 group IC2 intron component also encodes a putative double motif LHEase (I-CthI-P; the name based on Belfort and Roberts, 1997; P for putative). This HEase could be expressed in *E. coli* and purified to test its endonuclease activity experimentally starting with the same methods that were followed for both I-OmiI and I-OmiII (see Chapter 2 for detailed methodology). This HEase could be involved in providing the twintron with a mechanism for mobility.

8.2.3. *In vivo* cleavage assay of I-OmiI and I-OmiII:

In vitro cleavage assay of any HEases requires the purification of this HEase enzyme in a pure form, but some HEases are very difficult to express and purify as they tend to aggregate within inclusion bodies when expressed in *E. coli*. Also some HEases have been reported to induce cell lysis, which also limits the yields of active protein (Jurica and Stoddard, 1999). For the above mentioned reasons, studying the endonuclease activity *in vivo* (i.e. in *E. coli* cell) could be an alternative to confirm the endonuclease activity of I-OmiI. The latter HEase has proven to be problematic during this study and still needs further characterization.

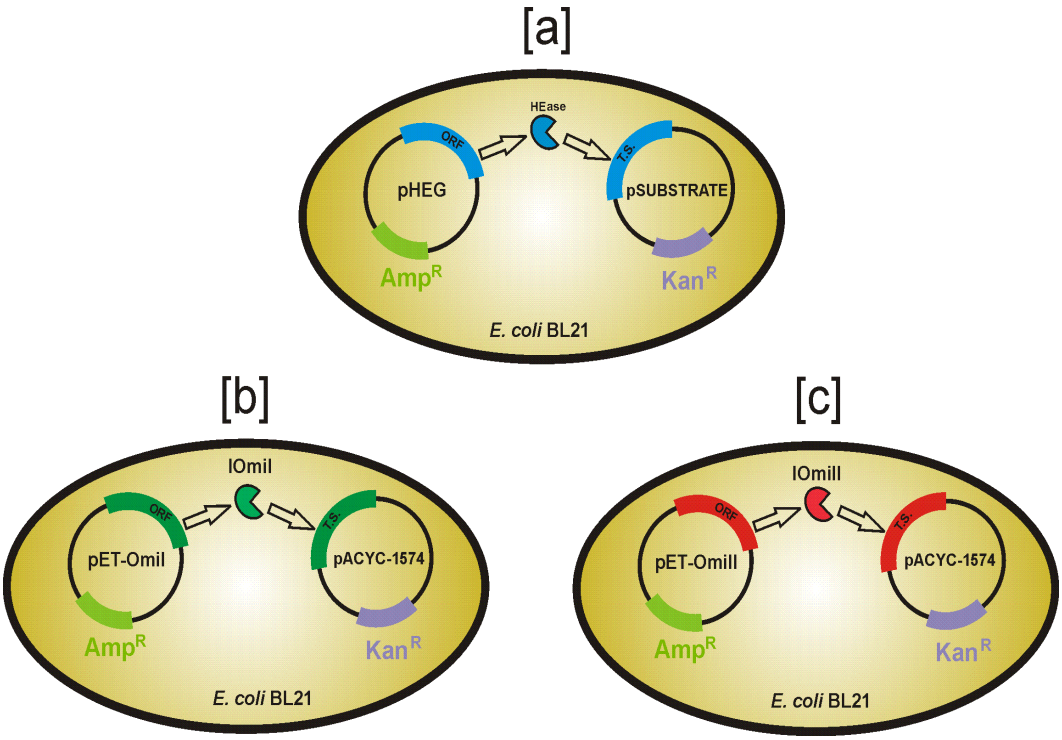
Seligman et al. (1997) presented an *in vivo* selection system that links the survival of *E. coli* cells with HEase-mediated DNA cleavage activity and sequence specificity. In this system, *E. coli* cells are co-transformed with two plasmids, one plasmid contains the HEase ORF (pHEG) and the other plasmid contains the target site of the tested enzyme (pSUBSTRATE) and this substrate plasmid also contains an antibiotic resistant gene (Kan^R). If the HEase is produced in an active form then binding to and cleavage of the target site can be tested as the loss of the Kan^R marker inhibits the ability of *E. coli* cells to grow on kanamycin-containing medium.

In order to test the *in vivo* endonuclease activity of both I-OmiI and I-OmiII, the substrate plasmid pACYC/1574 has already been constructed, which contains the target sites of both I-OmiI and I-OmiII. *E. coli* BL21 cells have been co-transformed with this substrate plasmid and with either one of the HEase expressing plasmids: pET200-OmiI or

pET200-OmiII (contains I-OmiI and I-OmiII ORFs respectively) (Figure 8.2.; see Appendix 9.7. for more information about these bacterial clones). The conditions to optimize the *in vivo* endonuclease assay for these two enzymes are obviously a priority for future efforts.

Figure 8.2. [a] A schematic representation of an *in vivo* endonuclease assay system showing the *E. coli* BL21 cells that are co-transformed with the HEG plasmid (pHEG) which contains the tested HEase ORF (ORF), and the substrate plasmid (pSUBSTRATE) which contains the target site of the tested HEase (T.S.). The two plasmids contain two different antibiotic resistance genes like ampicillin (Amp^R) and kanamycin (Kan^R). [b] and [c] represent the *in vivo* endonuclease assay systems for both I-OmiI and I-OmiII respectively.

Figure 8.2.


[a]
HEase
ORF
T.S.
pHEG
pSUBSTRATE
AmpR
KanR
E. coli BL21
[b]
IOmiI
ORF
T.S.
pET-OmiI
pACYC-1574
AmpR
KanR
E. coli BL21
[c]
IOmiII
ORF
T.S.
pET-OmiII
pACYC-1574
AmpR
KanR
E. coli BL21

8.2.4. X-Ray crystallography of I-OmiII with its native substrate DNA:

Finally to capitalize on the potential of I-OmiII with regards to biotechnology applications the actual DNA binding and cleavage sites need to be fully characterized. This can be resolved by obtaining a crystal structure of I-OmiII HEase when bound to its native target site. Resolving the crystal structure of I-OmiII would represent the first group II-encoded HEase to be characterized in such detail. This work would provide important information about the minimal recognition site of I-OmiII and it would show which amino acid positions are in physical contact with the target DNA, i.e. a contact map. Data from the crystal structure of I-OmiII would also facilitate the engineering of this enzyme to potentially recognize other sequences, as the contact map would identify which amino acids could be targeted for mutagenesis in order to change the target site specificity of the HEase.

CHAPTER 9:

APPENDICES

APPENDIX: 9.1.

IDENTIFICATION OF GROUP I INTRONS WITHIN THE NUCLEAR SSU rRNA GENE IN SPECIES OF *Ceratocystiopsis* AND RELATED TAXA. *

9.1.1. ABSTRACT:

During a recent phylogenetic study, group I introns were noted that interrupt the nuclear SSU rDNA gene in species of *Ceratocystiopsis*. Group I introns were found to be inserted at the following rDNA positions: S943, S989, and S1199. The introns have been characterized and phylogenetic analysis of the host gene and the corresponding intron data suggest that for S943 vertical transfer and frequent loss appear to be the most parsimonious explanation for the distribution of nuclear SSU rDNA introns among species of *Ceratocystiopsis*. The SSU rDNA data do suggest that a recent proposal of segregating the genus *Ophiostoma sensu lato* into *Ophiostoma sensu stricto*, *Grosmannia*, and *Ceratocystiopsis* has some merit but may need further amendments, as the SSU rDNA suggests that *Ophiostoma s. str.* may now represent a paraphyletic grouping.

* A version of this Chapter was published in Fungal Biology. Hafez et al. (2012); Vol. **116**: pp. 98 – 111.

9.1.2. INTRODUCTION:

Group I introns are self-splicing elements that occur in bacteria, bacteriophages and in the organelles of fungi, plants, protists, and early branching metazoans (sea anemones, sponges and soft corals) (Belfort et al., 2002; Hausner, 2003; Gissi et al., 2008). Group I introns are also found in nuclear ribosomal genes (rDNA) in a wide variety of fungi, algae and protists with lichen fungi being the richest source for group I introns (de Wachter, 1992; Hibbett, 1996; Bhattacharya et al., 1996; Bhattacharya et al., 2002; Lickey et al., 2003; Haugen et al., 2005; Gutiérrez et al., 2007; Feau et al., 2007; Hoshina & Imamura, 2009).

Insertions within the nuclear rDNA usually occur at highly conserved sequences and they are relatively common among the fungi and have been reported from several rDNA positions (Gargas et al., 1995; Cannone et al., 2002); in addition spliceosomal introns have also been discovered in the rDNA of ascomycetes (Bhattacharya et al., 2000). Among the intron rich members of the lichen fungi examples of vertical transmission and horizontal spread of introns have been documented; introns also appear to move to new rDNA sites by reverse splicing into novel rRNA sites (Dujon, 1989; Woodson & Cech, 1989; Grube et al., 1999; Bhattacharya et al., 2000; Martin et al., 2003; Bhattacharya et al., 2005; Haugen et al., 2005). In addition rDNA introns can have a sporadic distribution among phylogenetically closely related fungi suggesting that introns can be gained and lost relatively rapidly (Nikoh & Fukatsu, 2001; Haugen et al., 2005).

Group I introns show minimal primary sequence conservation, but they do have conserved secondary and tertiary structures and these elements are autocatalytic by catalyzing their own excision from an RNA molecule; they can therefore be viewed as ribozymes. Almost all group I introns contain pairing regions referred to as P1 to P10, along with sequence segments that connect these helical regions. The P4-P6 and P3-P9 paired helical domains make up the catalytic core components and the P1 and P10 helices are the substrate domain (includes the internal guide sequence) wherein the 5' and 3' splice sites are juxtaposed to each other (Cech et al., 1994; Woodson, 2005; Lindstrom & Pistolic, 2005). The P3-P7-P9 helix contains the GTP binding pocket; here the 3'OH of an exogenous GTP initiates the splicing reaction that involves two transesterification reactions resulting in the splicing of the exons and the release of the intron RNA (Raghavan & Minnick, 2009). Based on secondary structure characteristics, group I introns were classified initially into two subdivisions: IA and IB (Michel et al., 1982; Cech et al., 1994). However, based on both nucleotide sequences within the conserved core regions and variations within the secondary structure, group I intron classification has been further refined. Currently at least five classes are recognized (IA to IE), and these can be subdivided further e.g., IA1, IC3 etc. (Michel & Westhof, 1990; Suh et al., 1999; Li & Zhang, 2005). Over 20 000 group I introns have been identified in a variety of organisms, and the secondary structures of some group I introns and a list of rDNA intron insertions sites have been compiled in the Comparative RNA Web Site (R. Gutell; http://www.rna.ccbb.utexas.edu/, Cannone et al., 2002) and the Group I intron sequence and structure database (Zhou et al., 2008).

Nuclear group I introns usually lack potential open reading frames (ORFs). However, some group I introns belonging to the IC1 and IE subgroups have been noted in some fungi to encode HEases belonging to the His-Cys HEase family (Haugen et al., 2004). Nuclear rDNA group I introns range in size from 250 to 600 nucleotides (nts), although there are examples that exceed 1000 nts (Haugen et al., 2004). There are also highly compact group I introns at 62 to 78 nts and these introns appear to be IC1 introns that only maintained the P1, P7 and P10 paired regions; however they still have self-splicing capacity (Harris & Rogers, 2008).

During a recent study to investigate phylogenetic relationships among selected ophiostomatoid-like fungi, the SSU rDNA gene sequences were examined for members of the fungal genera *Ceratocystiopsis* H.P. Upadhyay & W.B. Kendr., *Grosmannia* Goid. and *Ophiostoma* Syd. & P. Syd. Among these fungi are many economically important tree pathogens and blue-stain fungi (see Wingfield et al., 1993; Hausner & Reid, 2004; Hausner et al., 2005). Members of these genera lack forcible ascospore discharge, have deliquescent asci and form sticky ascospore droplets at the apex of their perithecial necks. These fungi also tend to produce their slimy/sticky conidia on long-stalked conidiophores. Many of these morphological features are adaptations for arthropod dispersal, and therefore these characters are under strong selection pressure for optimizing dispersal and competition for insect vectors and substrates (Harrington, 1993; Farrell et al., 2001; Six and Wingfield, 2011). Species 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 & 1993c). 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. This SSU rDNA based study allowed as to investigate the evolutionary dynamics of nuclear SSU rDNA group I introns within the SSU rDNA gene of the ophiostomatoid fungi. So far no introns have been reported for this group of fungi thus this study is the first report on the occurrence and distribution of introns within this economically important group of fungi.

The objective of this study is to examine the group I intron variability in the nuclear SSU rDNA gene in species of *Ophiostoma, Grosmannia* and *Ceratocystiopsis* and related taxa. This study will try to assess the stability of the introns found within this group of organisms and examine the evolutionary biology of the encountered group I introns and potentially provide insights into the mechanisms that are involved in their diversification and persistence within the SSU rDNA gene.

9.1.3. METHODS OVERVIEW

The nuclear SSU rRNA gene was amplified from different *Ceratocystiopsis* strains and related taxa (See Table 9.1.1. for complete list of species and strains). Methods for DNA extraction and PCR amplification of the nuclear SSU rRNA gene (with primers SSU-J and SSU-T) were previously described in Chapter 2 and in Hafez et al., (2012).

Table 9.1.1. List of species/strains and the nuclear SSU rRNA sequence GenBank accession numbers.

Table 9.1.1.

Species	Strain	Introns					Accession Number
		S516 IE	S943 IC1	S989 IE	S1199 IE	1512 ?	
Ceratocystiopsis sp.	WIN(M)1520[a] = MPB 1[b]						**HQ634836**
Ceratocystiopsis sp.	WIN(M)1521 = MPB 2						**HQ634837**
Ceratocystiopsis sp.	WIN(M)1522 = MPB 3						**HQ634838**
Ceratocystiopsis brevicomi Hsiau & T.C. Harr	WIN(M)1452 = CBS 333.97[c]						**HQ202311**
Ceratocystiopsis collifera Marm. & Butin	WIN(M)908 = CBS 129.89						**HQ634832**
Ceratocystiopsis concentrica (Olchow & J. Reid) H.P. Upadhyay	WIN(M)53 = JR71-21[d]		**X**				**HQ634849**
Ceratocystiopsis manitobensis (J. Reid & Georg Hausner) Zipfel, Z.W. de Beer & M.J. Wingf.	WIN(M)237 = UAMH 9813[e]		**X**				**HQ634850**
Ceratocystiopsis minima (Olchow. & J. Reid) H.P. Upadhyay	WIN(M)1501		**X**				**HQ202312**
Ceratocystiopsis minima	WIN(M)1462 =CBS 182.86		**X**				**HQ634851**
Ceratocystiopsis minima	WIN(M)61 = JR69-37		**X**				**HQ634852**
Ceratocystiopsis minima	WIN(M)85 = JR71-03		**X**				**HQ634856**
Ceratocystiopsis minuta (Siemaszko) H.P. Upadhyay & W.B. Kendr.	WIN(M)1453 = CBS 441.94		**X**				**HQ634853**
Ceratocystiopsis minuta	WIN(M)1533 = RJ 5095[f]						**HQ634826**
Ceratocystiopsis minuta	WIN(M)1532; RJ705 = UAMH 11218						**HQ634827**
Ceratocystiopsis minuta	WIN(M)1535 = RJ191						**HQ634828**
Ceratocystiopsis minuta	WIN(M)1534 = RJ689						**HQ634829**
Ceratocystiopsis minuta	WIN(M)1537 = CBS 116963		**X**		**X**		**HQ634854**
Ceratocystiopsis minuta	WIN(M)1536 = CBS 116796						**HQ634830**
Ceratocystiopsis minuta	WIN(M)1523 = CBS 117042		**X**				**HQ634857**
Ceratocystiopsis minuta-bicolor (R.W. Davidson) H.P. Upadhyay	WIN(M)480 = JR87-6		**X**				**HQ634848**
Ceratocystiopsis pallidobrunnea (Olchowecki & J. Reid) H.P. Upadhyay	WIN(M)51 = JR69-14		**X**	**X**	**X**		**HQ634842**
Ceratocystiopsis parva (Olchow. & J. Reid) Zipfel, Z.W. de Beer & M.J. Wingf.	WIN(M)59 = JR71-21		**X**				**HQ595735**
Ceratocystiopsis ranaculosus J.R. Bridges & T.J. Perry	WIN(M)919						**HQ634840**
Ceratocystiopsis rollhanseniana (J. Reid, Eyj_olfsd_ottir & Georg Hausner) Zipfel, Z.W. de Beer & M.J. Wingf.	WIN(M)113		**X**		**X**		**HQ595736**
Ceratocystiopsis rollhanseniana	WIN(M)110 = UAMH 9797						**HQ634834**
***Ceratocystiopsis.* crassivaginata**	WIN(M)1458 = CBS 512.83						**HQ595740**
Exophiala calicioides (Fr.) G. Okada & Seifert	WIN(M)717 = JR87-16		**X**				**HQ202314**
Graphium pseudormiticum M. Mouton & M.J. Wingf.	WIN(M)1569 = DAOM 234026[g]				**X**		**HQ634858**
Graphium pseudormiticum	WIN(M)1571 = DAOM 234028				**X**		**HQ634859**
Graphium pseudormiticum	WIN(M)1570 = DAOM 234027				**X**		**HQ634860**
Graphium sp. (novo sp.)	WIN(M)1490 = JR 87-3d						**HQ202315**
Grosmannia crassivaginata (H.D. Griffin) Zipfel, Z.W. de Beer & M.J. Wingf.	WIN(M)918 = UAMH 7004						**HQ634833**
Grosmannia crassivaginata	WIN(M)184						**HQ634835**
Grosmannia davidsonii (Olchow. & J. Reid) Zipfel, Z.W. de Beer & M.J. Wingf.	WIN(M)60 = JR71-30						**HQ595732**
Grosmannia davidsonii	WIN(M)1495 = MCC 871[h]						**HQ634815**
Grosmannia davidsonii	WIN(M)1494 = MCC 870						**HQ634816**

Table 9.1.1. CONTINUED

Grosmannia davidsonii	WIN(M)1132						**HQ634819**
Grosmannia penicillata (Grosmann) Goid.	WIN(M)27 = NOR60-21[i]						**HQ634822**
Grossmania piceaperda (Rumbold) Goid.	WIN(M)980 = UAMH 9788				**X**		**HQ595733**
Leptographium sp.	WIN(M)984			**X**			**HQ595734**
Leptographium sp.	WIN(M)1429						**HQ595737**
Meria laricis Vuill.	WIN(M)1525 = CBS 298.52						**HQ634817**
Meria laricis	WIN(M)1526 = CBS 283.59		**X**			**X**	**HQ634844**
Meria laricis	WIN(M)1527 = CBS 281.59		**X**			**X**	**HQ634845**
Meria laricis	WIN(M)1528 = CBS 282.59						**HQ634818**
Ophiostoma brevicolle (R.W. Davidson) de Hoog & R.J. Scheff.	WIN(M)811 = CBS 150.78						**HQ634823**
Ophiostoma crenulatum (Olchow. & J. Reid) Georg Hausner & J. Reid	WIN(M)58 = JR 70-17				**X**		**HQ634855**
Ophiostoma longisporum (Olchow. & J. Reid) Georg Hausner, J. Reid & Klassen	WIN(M)48						**HQ634831**
Ophiostoma minus (Hedgc.) Syd. & P. Syd.	WIN(M)861 = DAOM 29251						**HQ634820**
Ophiostoma minus	WIN(M)871 = JR 23Rp3-367TA						**HQ634821**
Ophiostoma retusum (R.W. Davidson & T.E. Hinds) Georg Hausner, J. Reid & Klassen	ATCC 22324[j]						**HQ634841**
Ophiostoma sp.	WIN(M)1391						**HQ634824**
Pesotum fragrans (Math.-K€a€arik) G. Okada & Seifert	WIN(M)1396 = JR87-4H				**X**		**HQ634846**
Pesotum sp.	WIN(M)1426 = JR87-4C				**X**		**HQ634847**
Pesotum sp.	WIN(M)1394 = JR87-10C						**HQ634839**
Sarcotrochila macrospore Ziller & A. Funk	WIN(M)1538 = CBS 274.74	**X**	**X**	**X**	**X**		**HQ634843**

[a] **WIN(M)** = University of Manitoba (Winnipeg) Collection.
[b] **MPB** = Mountain Pine Beetle collection, Department of Forestry, University of British Columbia, Canada.
[c] **CBS** = Centraal Bureau voor Schimmelcultures, Utrecht, The Netherlands.
[d] **JR** = J. Reid collection (in WIN(M) see above).
[e] **UAMH** = University of Alberta Microfungus Collection & Herbarium, Devonian Botanic Garden, Edmonton, AB, Canada T6G 2E1.
[f] **RJ** = R. Jankowiak collection (University of Agriculture in Cracow, Cracow, Poland).
[g] **DAOM** = Plant Research Institute, Department of Agriculture, Mycology, Ottawa, Canada.
[h] **MCC** = Culture collection of H. Masuya, Forestry and Forest Products Research Institute, Ibaraki, Japan.
[i] **NOR** = NFRI, Norwegian Forest Research Institute, AS, Norway.
[j] **ATCC** = American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108, USA.

The PCR products were purified with the Wizard SV Gel and PCR clean-up system (Promega, 2800 Woods Hollow Road, Madison, WI 53711) in order to obtain DNAs suitable for DNA sequence analysis. The PCR products were sequenced using the cycle-sequencing protocols performed according to the manufacturer's recommendations (Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404). Automated fluorescent DNA sequence analysis was performed using an ABI Prism 310 Genetic Analyzer (University of Calgary Core DNA service, Faculty of Medicine, University of Calgary, 3350 Hospital Drive NW, Calgary, AB, Canada T2N 4N1). Initially the original forward and reverse primers utilized for the PCR reactions were used to start the sequences for the PCR products; from there primers were designed as needed to sequence both strands of the PCR amplicons.

For phylogenetic analysis, nucleotide sequences were initially aligned with Clustal X program and the alignments were refined manually with the aid of the GeneDoc program. Phylogenetic estimates were generated using parsimony, likelihood-based and Bayesian approaches, by using components contained within the PHYLIP package, PHYML online web server, and the MrBayes program v3.1 respectively. The models applied for Likelihood approaches were based on evaluating the nucleotide sequence alignments with the FindModel program. The program MESQUITE version 2.73 was used estimate the ancestral character states for the presence and absence of the S943, S989 and S1199 introns. Secondary structure models were developed using the web server RNAweasel and the online program *mfold* program. Finally, the *in vivo* splicing of mS379 intron was determined by RT-PCR as described in Chapter 2.

9.1.4. RESULTS:

9.1.4.1. SSU rDNA introns:

Although we assayed for the presence or absence of introns using a PCR approach (see Figure 9.1.1.) we have to admit that a "minus" result does not unambiguously identify taxa that lack introns. PCR potentially amplifies smaller DNA fragments; thus, if there is heterogeneity within the rDNA tandem repeats rare intron-plus alleles could be missed (Simon et al., 2005a, 2005b). So we are reporting the intron distribution for the "dominant" version of the SSU rDNA gene for the taxa examined during this study.

Among the ophiostomatoid fungi studied we noted three nuclear SSU rDNA group I introns (Figure 9.1.1.) inserted at positions 943, 989 and 1199 (Figure 9.1.2. [a] & [b]). Based on structural features the S943 intron (Figure 9.1.3. b) is an IC1 type and the S989 and S1199 introns (Figure 9.1.3. [c] & [d] respectively) can be categorized as IE type group I introns. Among the ophiostomatoid fungi, the SSU rDNA gene of *Cop. pallidobrunnea* WIN(M)51 had all three introns present. The same three introns were also found in the very distantly related species *Sarcotrochila macrospora* Ziller & A. Funk, belonging to the Helotiales. The S943 intron was the most frequently encountered intron, 13 times among our tested 25 *Ceratocystiopsis* strains, this compares to one instance of the S989 intron [in *Cop. Pallidobrunnea*] and three examples of the S1199 intron [*Cop. pallidobrunnea, Cop. minuta* (WIN(M)1537), and *Cop. rollhanseniana* (WIN(M)113)].

Figure 9.1.1. Phylogenetic tree based on the nuclear SSU rDNA data set for *Ceratocystiopsis* and related ascomycetous taxa. The tree topology is based on Bayesian analysis, solid circles, squares and triangles represents Posterior Probability (PP) supportive values (90-100%), (80-89%) and (70-79%) respectively as obtained from a 50% majority Bayesian consensus tree. Open circles, squares and triangles represent bootstrap (BS) support (90-100%), (80-89%) and (70-79%) respectively based on ML analysis. Nodes supported with BS ≥ 86 % based on Parsimony analysis are marked with asterisks. Nodes that received less than 70% support (BS or PP) were collapsed. The branch lengths are based on Bayesian analysis and are proportional to the number of substitutions per site. Strain numbers (underlined) and GenBank accession numbers are listed next to species names. The tree is rooted with SSU rRNA sequences from the Basidiomycetous fungi *Fomes fomentarius* (L.) Fr. and *Phlebia radiata* Fr.

Figure 9.1.1.

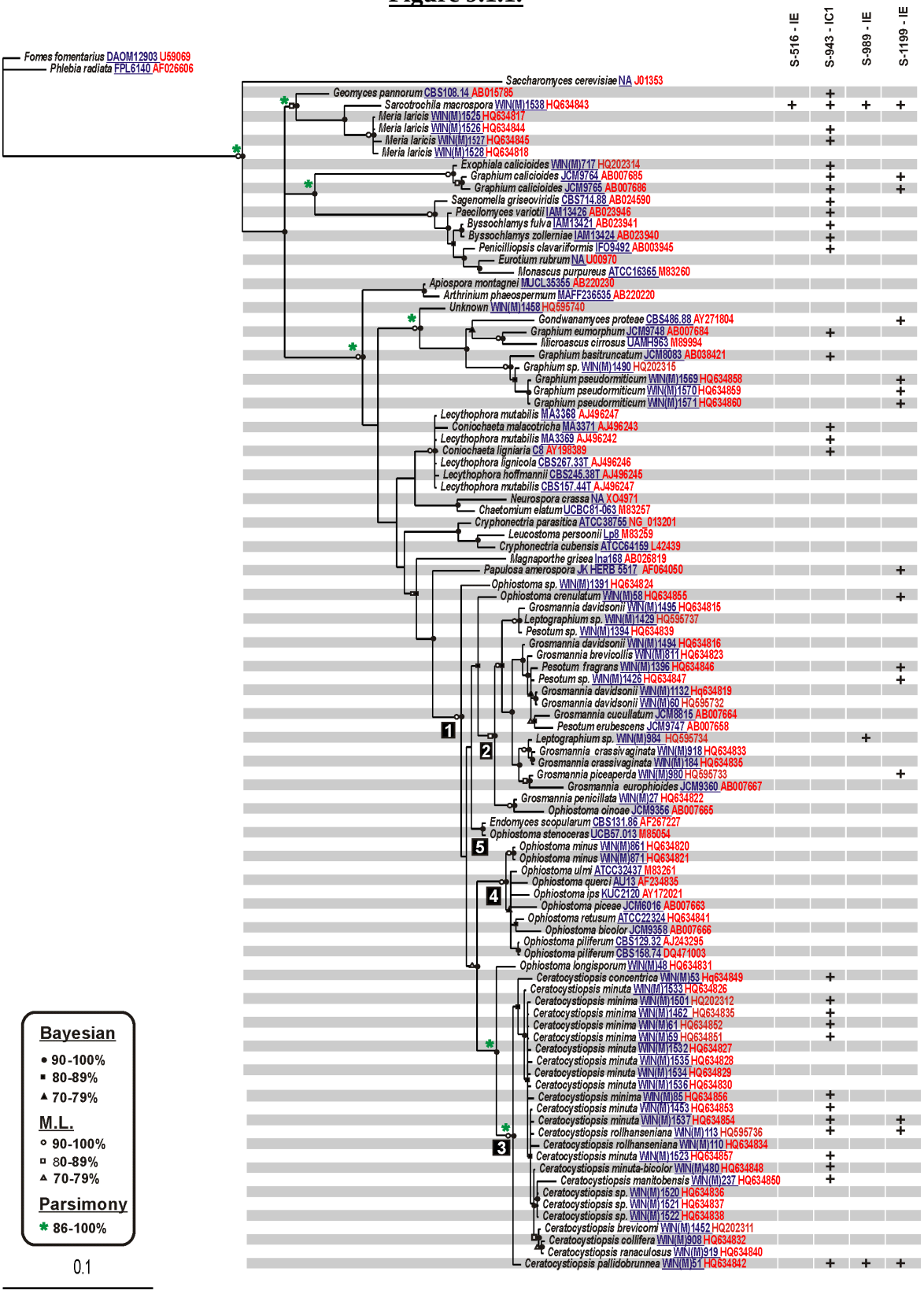
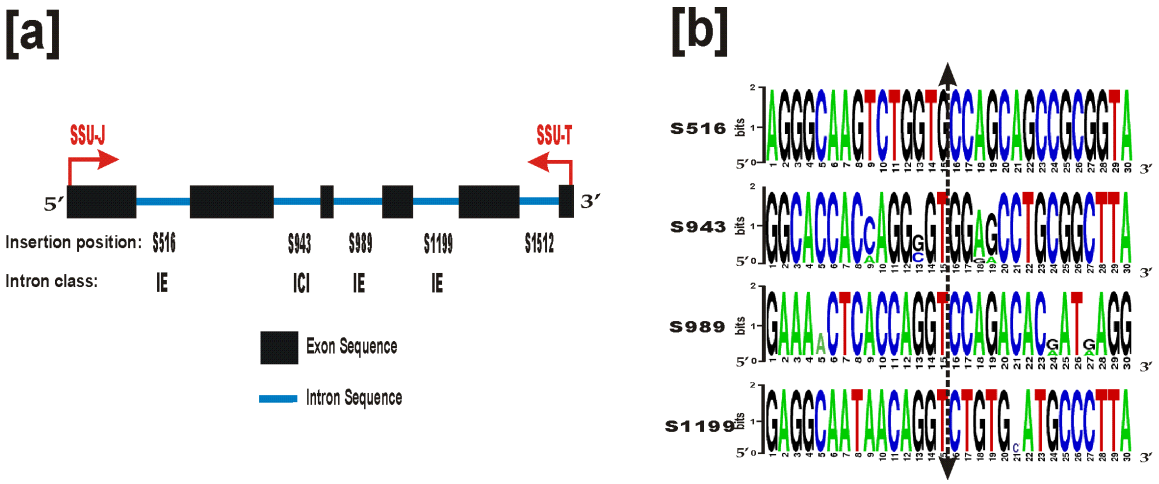

S-516 - IE
S-943 - IC1
S-989 - IE
S-1199 - IE
Fomes fomentarius DAOM12903 U59069
Phlebia radiata FPL6140 AF026606
Saccharomyces cerevisiae NA J01353
Geomyces pannorum CBS108.14 AB015785
Sarcotrochila macrospora WIN(M)1538 HQ634843
Meria laricis WIN(M)1525 HQ634817
Meria laricis WIN(M)1526 HQ634844
Meria laricis WIN(M)1527 HQ634845
Meria laricis WIN(M)1528 HQ634818
Exophiala calicioides WIN(M)717 HQ202314
Graphium calicioides JCM9764 AB007685
Graphium calicioides JCM9765 AB007686
Sagenomella griseoviridis CBS714.88 AB024590
Paecilomyces variotii IAM13426 AB023946
Byssochlamys fulva IAM13421 AB023941
Byssochlamys zollerniae IAM13424 AB023940
Penicilliopsis clavariiformis IFO9492 AB003945
Eurotium rubrum NA U00970
Monascus purpureus ATCC16365 M83260
Apiospora montagnei MUCL35355 AB220230
Arthrinium phaeospermum MAFF236535 AB220220
Unknown WIN(M)1458 HQ595740
Gondwanamyces proteae CBS486.88 AY271804
Graphium eumorphum JCM9748 AB007684
Microascus cirrosus UAMH963 M89994
Graphium basitruncatum JCM8083 AB038421
Graphium sp. WIN(M)1490 HQ202315
Graphium pseudormiticum WIN(M)1569 HQ634858
Graphium pseudormiticum WIN(M)1570 HQ634859
Graphium pseudormiticum WIN(M)1571 HQ634860
Lecythophora mutabilis MA3368 AJ496247
Coniochaeta malacotricha MA3371 AJ496243
Lecythophora mutabilis MA3369 AJ496242
Coniochaeta ligniaria C8 AY198389
Lecythophora lignicola CBS267.33T AJ496246
Lecythophora hoffmannii CBS245.38T AJ496245
Lecythophora mutabilis CBS157.44T AJ496247
Neurospora crassa NA XO4971
Chaetomium elatum UCBC81-063 M83257
Cryphonectria parasitica ATCC38755 NG_013201
Leucostoma persoonii Lp8 M83259
Cryphonectria cubensis ATCC64159 L42439
Magnaporthe grisea Ina168 AB026819
Papulosa amerospora JK HERB 5517 AF064050
Ophiostoma sp. WIN(M)1391 HQ634824
Ophiostoma crenulatum WIN(M)58 HQ634855
Grosmannia davidsonii WIN(M)1495 HQ634815
Leptographium sp. WIN(M)1429 HQ595737
Pesotum sp. WIN(M)1394 HQ634839
Grosmannia davidsonii WIN(M)1494 HQ634816
Grosmannia breviocollis WIN(M)811 HQ634823
Pesotum fragrans WIN(M)1336 HQ634846
Pesotum sp. WIN(M)1426 HQ634847
Grosmannia davidsonii WIN(M)1132 Hq634819
Grosmannia davidsonii WIN(M)60 HQ595732
Grosmannia cucullatum JCM8815 AB007664
Pesotum erubescens JCM9747 AB007658
Leptographium sp. WIN(M)984 HQ595734
Grosmannia crassivaginata WIN(M)918 HQ634833
Grosmannia crassivaginata WIN(M)184 HQ634835
Grosmannia piceaperda WIN(M)980 HQ595733
Grosmannia europhioides JCM9360 AB007667
Grosmannia penicillata WIN(M)27 HQ634822
Ophiostoma olnoae JCM9356 AB007665
Endomyces scopularum CBS131.86 AF267227
Ophiostoma stenoceras UCB57.013 M85054
Ophiostoma minus WIN(M)861 HQ634820
Ophiostoma minus WIN(M)871 HQ634821
Ophiostoma ulmi ATCC32437 M83261
Ophiostoma querci AU13 AF234835
Ophiostoma ips KUC2120 AY172021
Ophiostoma piceae JCM6016 AB007663
Ophiostoma retusum ATCC22324 HQ634841
Ophiostoma bicolor JCM9358 AB007666
Ophiostoma piliferum CBS129.32 AJ243295
Ophiostoma piliferum CBS158.74 DQ471003
Ophiostoma longisporum WIN(M)48 HQ634831
Ceratocystiopsis concentrica WIN(M)63 Hq634849
Ceratocystiopsis minuta WIN(M)1533 HQ634826
Ceratocystiopsis minima WIN(M)1501 HQ202312
Ceratocystiopsis minima WIN(M)1462 HQ634835
Ceratocystiopsis minima WIN(M)61 HQ634852
Ceratocystiopsis minima WIN(M)59 HQ634851
Ceratocystiopsis minuta WIN(M)1532 HQ634827
Ceratocystiopsis minuta WIN(M)1535 HQ634828
Ceratocystiopsis minuta WIN(M)1534 HQ634829
Ceratocystiopsis minuta WIN(M)1536 HQ634830
Ceratocystiopsis minima WIN(M)85 HQ634856
Ceratocystiopsis minuta WIN(M)1453 HQ634853
Ceratocystiopsis minuta WIN(M)1537 HQ634854
Ceratocystiopsis rollhanseniana WIN(M)113 HQ595736
Ceratocystiopsis rollhanseniana WIN(M)110 HQ634834
Ceratocystiopsis minuta WIN(M)1523 HQ634857
Ceratocystiopsis minuta-bicolor WIN(M)480 HQ634848
Ceratocystiopsis manitobensis WIN(M)237 HQ634850
Ceratocystiopsis sp. WIN(M)1520 HQ634836
Ceratocystiopsis sp. WIN(M)1521 HQ634837
Ceratocystiopsis sp. WIN(M)1522 HQ634838
Ceratocystiopsis brevicorni WIN(M)1452 HQ202311
Ceratocystiopsis collifera WIN(M)908 HQ634832
Ceratocystiopsis ranaculosus WIN(M)919 HQ634840
Ceratocystiopsis pallidobrunnea WIN(M)51 HQ634842
1
2
3
4
5
Bayesian
• 90-100%
▪ 80-89%
▲ 70-79%
M.L.
○ 90-100%
▫ 80-89%
△ 70-79%
Parsimony
* 86-100%
0.1

Figure 9.1.2. [A] A schematic representation of the nuclear SSU rDNA showing the initial primers (SSU-J and SSU-T) used to amplify the gene and the various types of group I introns found during this study. The introns are located at positions S516, S943, S989 and S1199 with reference to the *E. coli* SSU rDNA sequence. **[B]** Nucleotide sequence logos are shown for the exon sequences that flank the intron insertion sites (15 upstream and 15 downstream nucleotides). The double headed arrow indicates the insertion position for the following introns: S516, S943, S989 and S1199. The logos were generated by the online program WebLogo version 2.8.2 (Cooks et al., 2004; http://weblogo.berkeley.edu/logo.cgi).

Figure 9.1.2.

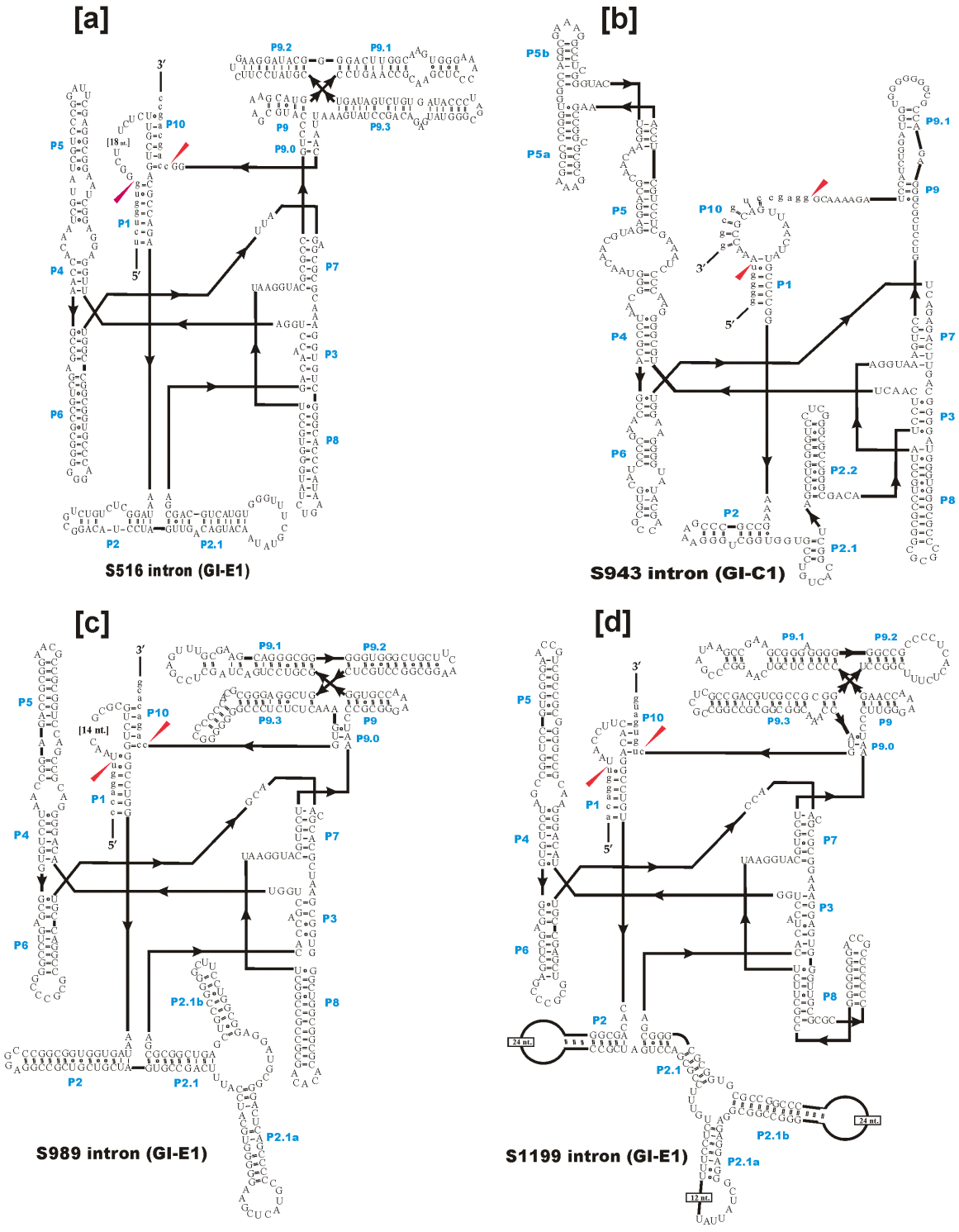

[a]
SSU-J
SSU-T
5′
3′
Insertion position: S516 S943 S989 S1199 S1512
Intron class: IE ICI IE IE
Exon Sequence
Intron Sequence
[b]
S516
S943
S989
S1199
bits

RT-PCR was performed on *Cop. pallidobrunnea* rRNA (Figure 9.1.4.), as this offered an opportunity to examine representatives for all three introns with regards to their ability to splice *in vivo* and to define the intron/exon junctions. The latter confirmed the intron/exon junctions as defined by comparing SSU rDNA sequences that lacked introns with those that have them (Figure 9.1.2. [b]; Figure 9.1.4.).

The S943 intron has been described from a wide variety of fungi (Gargas et al., 1995; Perotto et al., 2000; Feau et al., 2007). Additional examples of the S943 intron were noted among members of the Eurotiomycetes, two members of the Microascales, and three strains belonging to the Coniochaetales (Sordariomycetes); however, none were noted among the tested members of *Grosmannia* and *Ophiostoma*. The S989 intron was only encountered three times during this study with examples in *Cop. pallidobrunnea*, *Leptographium* sp. WIN(M)984 (*Grosmannia* clade) and *S. macrospora*. The S1199 intron also has a spotty distribution but it was observed in three *Ceratocystiopsis* species, three members of the *Grosmannia* clade, in *Ophiostoma crenulatum*, in *Papulosa amerospora*, three members of the Micoascales, two members of the Chaetothyriales (Eurotiomycetes) and this intron is also present in *S. macrospora*. Finally the S516 group IE (Figure 9.1.3. [a]) and S1512 introns were only encountered within *S. macrospore* and two strains of *Meria laricis* respectively. The S1512 intron did not appear to follow any of the conventional group I intron folds so we could not assign it to any category of group I intron. Typically group I introns inserted into the S1512 position belong to the IC1 category (http://www.rna.ccbb.utexas.edu/). The S1512 intron had been previously noted by Gutiérrez et al. (2007) in members of the lichen family Parmeliaceae.

Figure 9.1.3. Secondary structures for the following group I introns: [a] S516, [b] S943, [c] S989 and [d] S1199. Intron sequences are in upper-case letters and exon sequences are in lower-case letters. The ten pairing regions (P1-P10) are indicated. The solid red arrowheads indicate the intron-exon junctions (putative 5' and 3' splicing sites). Designation of intron types is based on previous reports by Michel & Westhof (1990) and Suh et al. (1999).

Figure 9.1.3.

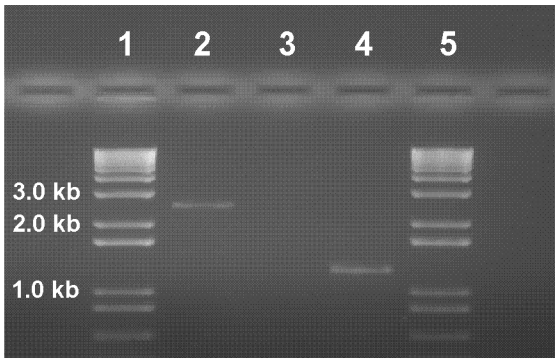

[a]
S516 intron (GI-E1)
[b]
S943 intron (GI-C1)
[c]
S989 intron (GI-E1)
[d]
S1199 intron (GI-E1)

9.1.4.2. Intron vs. host gene phylogenies:

We focused on the evolutionary dynamics of the S943 and S1199 group I introns, as several examples were found in the SSU rDNA in some members of the ophiostomatoid fungi. The other introns were only noted in a few instances, thus, not providing enough of a data set to warrant further analysis. Phylogenetic trees for the host SSU rDNA sequences when mirrored with the phylogeny based on the S943 intron data set (Figure 9.1.5. [a] & [b] respectively) had similar topologies. For the S943 intron we focussed on those found within *Ceratocystiopsis* species, as we could not generate unambiguous alignments for the intron core sequences when non-ophiostomatoid sequences were included. The S943 intron has an extremely wide distribution (Nikoh & Fukatsu 2001; Haugen et al. 2005) and that would suggest that it is an ancient intron. The presence of the S943 intron in the basal branching member of *Ceratocystiopsis* (*Cop. pallidobrunnea*) combined with the SSU rDNA tree being similar to the S943 intron tree suggest that this intron most likely is vertically transferred among species of *Ceratocystiopsis* and the random distribution noted (Figure 9.1.1.) is most likely due to frequent loss.

Figure 9.1.4. RT-PCR analysis of the SSU rDNA for *Cop. pallidobrunnea* WIN(M)51 to demonstrate *in vivo* splicing of the S943, S989 and S1199 introns and to confirm the exon/intron junctions. Lanes 1 and 5 contains the 1 kb plus DNA ladder (Invitrogen), lane 2 represents a standard PCR reaction using primers SSU-U and SSU-T with whole DNA as a template, while lane 4 contains amplicons derived from the RT generated cDNA template. The negative control in lane 3 (standard PCR using whole cell RNA as a template) yielded no bands confirming that the RNA samples was DNA-free. The genomic DNA generated a 2.7 kb PCR fragment indicating the presence of the S943, S989 and S1199 introns, while the cDNA template generated a 1.2 kb PCR fragment indicating that all introns were spliced out.

Figure 9.1.4.



The S1199 intron is widely distributed among the ascomycetes fungi (Gibb & Hausner 2003) and we noted several examples of this intron among the sequences examined (Figure 9.1.1.). When the host gene tree was mirrored with the intron-based tree (Figure 9.1.5. [c] & [d] respectively) the topologies were essentially similar, failing to support strong arguments for horizontal movement of this intron among the taxa examined in this study. Ancestral state analysis (Figure 9.1.6.) appears to confirm the results for the S943 intron; however it fails to support the notion that presence of S1199 can be explained by vertical transmission and frequent loss.

9.1.4.3. SSU rDNA phylogenetic analysis:

The phylogenetic analysis showed some well-supported clades with regards to fungi that are commonly referred to as ophiostomatoid fungi (i.e. members of *Ophiostoma*, *Ceratocystiopsis*, and *Grosmannia*) (Figure 9.1.1., node 1). Species that can be assigned to *Grosmannia* based on the presence of a *Leptographium* anamorph and *Leptographium* species can be included within one clade (Figure 9.1.1. node 2), although this grouping also includes *Pesotum* sp. (WIN(M) 1426 and 1394), *Pesotum erubescens* (Math.-Käärik) G. Okada and *Pesotum fragrans* (Math.-Käärik) M. Morelet. Species that can be assigned to *Ceratocystiopsis* (Figure 9.1.1.; node 3) formed a monophyletic group that received strong node support. The SSU rDNA sequence of *Ophiostoma longisporum* was placed outside of the node that joins all *Ceratocystiopsis* species. Another well-supported clade (Figure 9.1.1., node 4), referred to from now on as the "*Ophiostoma* group", includes representatives of *Ophiostoma minus*, *Ophiostoma ips*, (Rumbold)

Nannf., *Ophiostoma piliferum* (Fr.) Syd. & P. Syd., *Ophiostoma retusum*, *Ophiostoma bicolor* R.W. Davidson & D.E. Wells.

The SSU rDNA sequences of *Ophiostoma stenoceras* (Robak) Melin & Nannf. and *O. crenulatum* and an undescribed species of *Ophiostoma* (WIN(M)1391; Reid and Hausner unpublished data) do not appear to have a close link to the "*Ophiostoma* group" (Figure 9.1.1.). A strain of *Endomyces scopularum* Helfer, a yeast-like fungus, appears to be monophyletic with *O. stenoceras* (Figure 9.1.1., node 5); however, it should be noted that *E. scopularum* should be viewed as a *Sporothrix* sp. (see Suh et al. 2001).

9.1.5. DISCUSSION:

9.1.5.1. SSU rDNA Intron distribution among the ophiostomatoid fungi:

As less is known about the occurrence of rDNA group I introns in non-lichenized fungi we focussed on the ophiostomatoid fungi, and based on the taxon sampling represented in this study, we can conclude that species of *Ceratocystiopsis* appear to have more group I introns within their SSU rDNA gene compared to species of *Ophiostoma* and *Grosmannia.*

Figure 9.1.5. Phylogenetic relationships between group I introns and their corresponding host SSU rDNAs. [a] SSU rDNA based phylogeny versus the S943 group I intron phylogeny [b]; based on the P1 to P8 conserved stem sequences of the S943 intron. [c] SSU rDNA phylogeny compared to the S1199 group I intron phylogeny [d]; the intron alignment consists of the P1 to P8 conserved stem sequences. For the group I intron alignments the P9 stem sequences were excluded because they could not be aligned unambiguously. The tree topology [a,b,c & d] is based on the Majority rule consensus tree generated by Parsimony analysis (DNAPARS), open circles, squares and triangles represents nodes received bootstrap (BS) and/or Posterior probability (PP) values (95-100%), (80-94%) and (50-79%) respectively as obtained from a 50% consensus tree. GenBank accession numbers for the SSU rDNA sequences are listed along with the species name.

Figure 9.1.5.

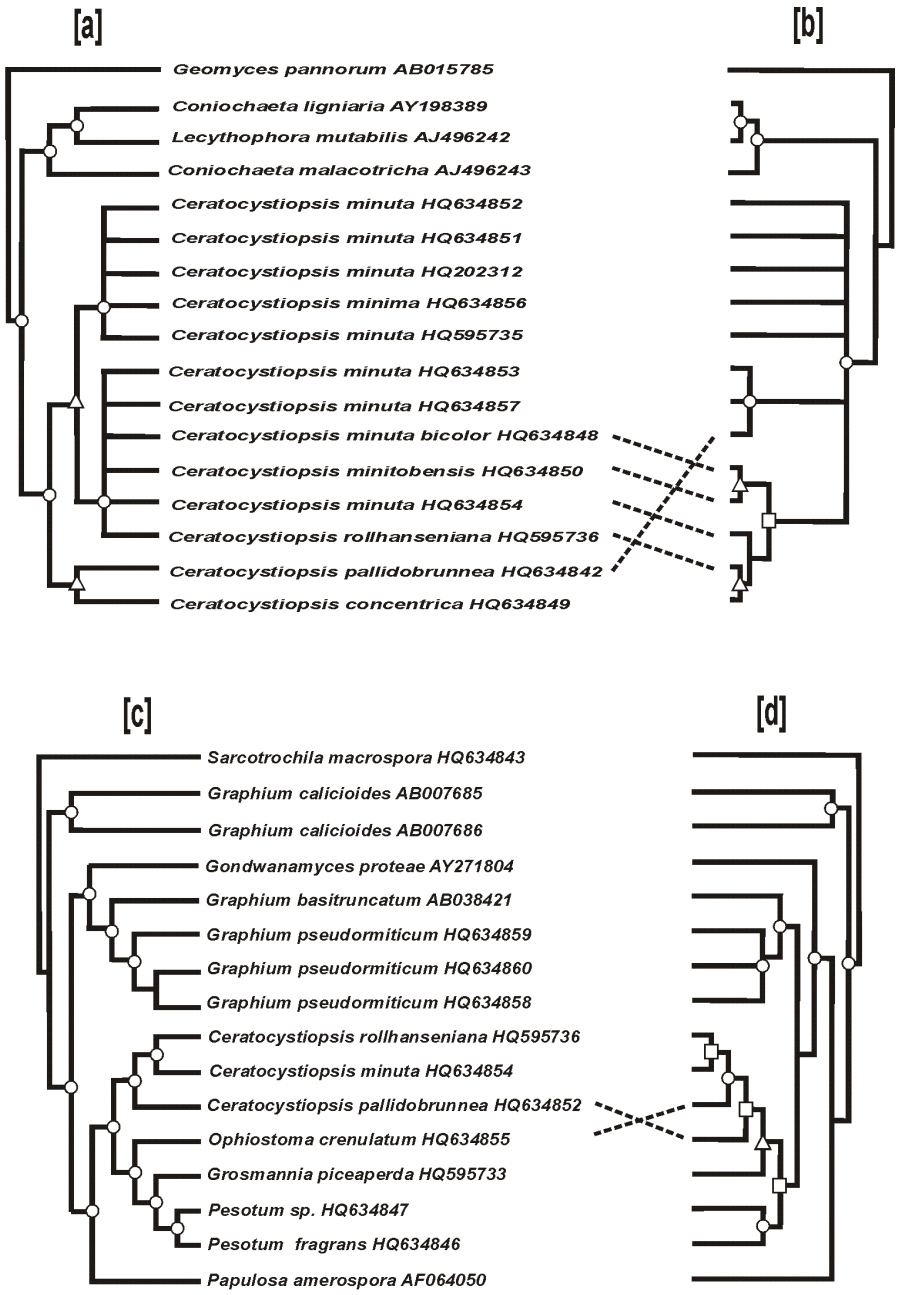
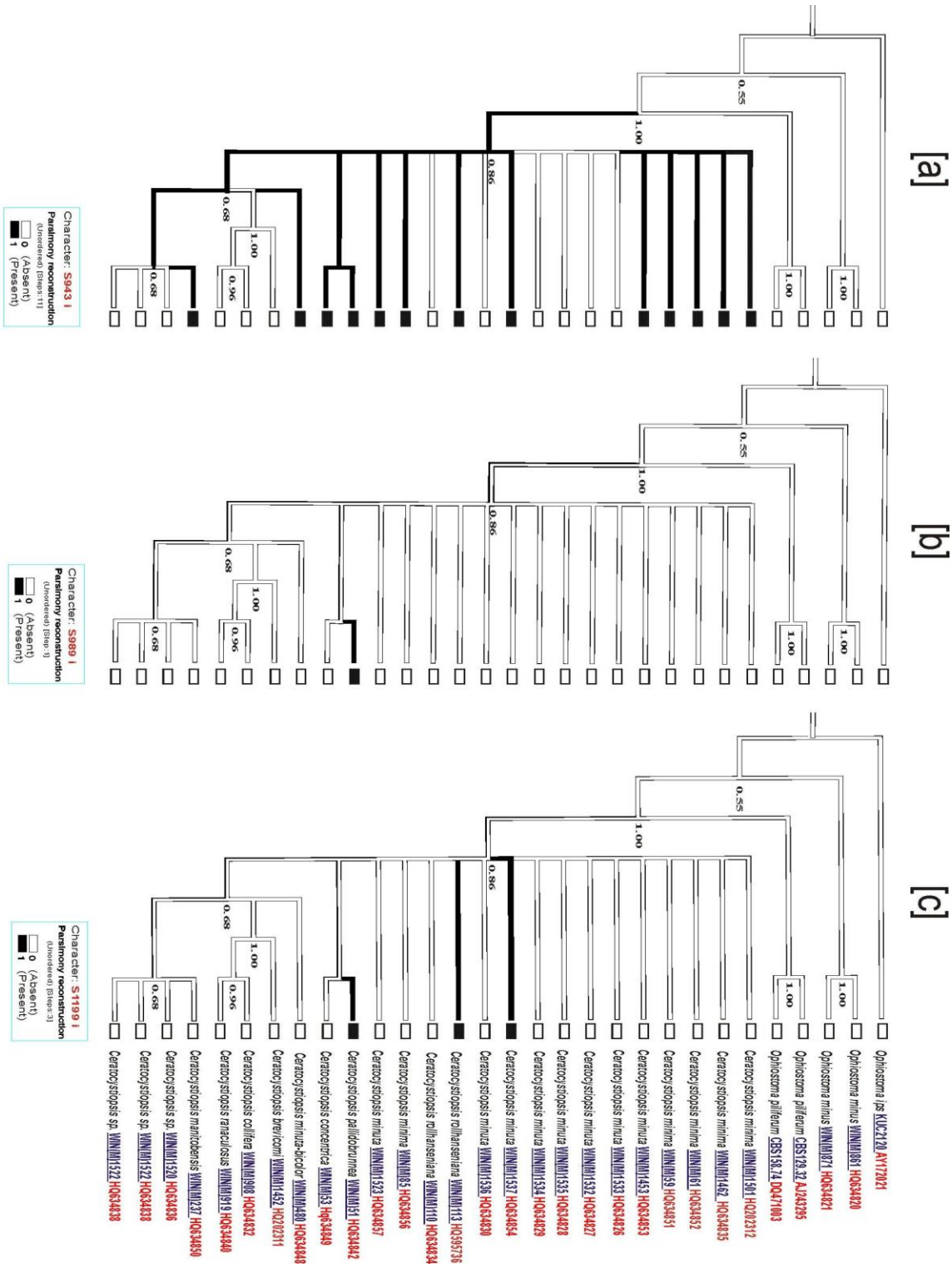

[a]
[b]
Geomyces pannorum AB015785
Coniochaeta ligniaria AY198389
Lecythophora mutabilis AJ496242
Coniochaeta malacotricha AJ496243
Ceratocystiopsis minuta HQ634852
Ceratocystiopsis minuta HQ634851
Ceratocystiopsis minuta HQ202312
Ceratocystiopsis minima HQ634856
Ceratocystiopsis minuta HQ595735
Ceratocystiopsis minuta HQ634853
Ceratocystiopsis minuta HQ634857
Ceratocystiopsis minuta bicolor HQ634848
Ceratocystiopsis minitobensis HQ634850
Ceratocystiopsis minuta HQ634854
Ceratocystiopsis rollhanseniana HQ595736
Ceratocystiopsis pallidobrunnea HQ634842
Ceratocystiopsis concentrica HQ634849
[c]
[d]
Sarcotrochila macrospora HQ634843
Graphium calicioides AB007685
Graphium calicioides AB007686
Gondwanamyces proteae AY271804
Graphium basitruncatum AB038421
Graphium pseudormiticum HQ634859
Graphium pseudormiticum HQ634860
Graphium pseudormiticum HQ634858
Ceratocystiopsis rollhanseniana HQ595736
Ceratocystiopsis minuta HQ634854
Ceratocystiopsis pallidobrunnea HQ634852
Ophiostoma crenulatum HQ634855
Grosmannia piceaperda HQ595733
Pesotum sp. HQ634847
Pesotum fragrans HQ634846
Papulosa amerospora AF064050
Nodes received at least one BS and/or PP value ranges between:
95%-100%
80%-94%
50%-79%

Figure 9.1.6. Ancestral state reconstruction for the S943, S989 and S1199 introns. Majority-rule senescence Bayesian tree of different *Ceratocystiopsis* strains based on the SSU rDNA nucleotide sequence. The numbers at the nodes indicate the Posterior Probability (PP) values obtained from 50% majority consensus tree generated by Mr Bayes. The ancestral states of the [a] S943 [b] S989 and [c] S1199 introns were reconstructed by Mesquite version 2.74 using Parsimony ancestral states. Character states: 0 = absent (white lines and squares); 1 = present (black lines and squares).

Figure 9.1.6.


[a]
[b]
[c]
Character: S943 i
Parsimony reconstruction
(Unordered) [Steps: 11]
0 (Absent)
1 (Present)
Character: S989 i
Parsimony reconstruction
(Unordered) [Steps: 3]
0 (Absent)
1 (Present)
Character: S1199 i
Parsimony reconstruction
(Unordered) [Steps: 3]
0 (Absent)
1 (Present)
Ophiostoma ips KUC2129 AY172021
Ophiostoma minus WINIM1881 HQ634820
Ophiostoma minus WINIM1871 HQ634821
Ophiostoma piliferum CBS129.32 AJ243295
Ophiostoma piliferum CBS158.74 DQ470983
Ceratocystiopsis minima WINIM1501 HQ282312
Ceratocystiopsis minima WINIM1462 HQ634835
Ceratocystiopsis minima WINIM1801 HQ634852
Ceratocystiopsis minima WINIM159 HQ634851
Ceratocystiopsis minuta WINIM1453 HQ634853
Ceratocystiopsis minuta WINIM1533 HQ634826
Ceratocystiopsis minuta WINIM1532 HQ634827
Ceratocystiopsis minuta WINIM1535 HQ634828
Ceratocystiopsis minuta WINIM1534 HQ634829
Ceratocystiopsis minuta WINIM1537 HQ634854
Ceratocystiopsis minuta WINIM1536 HQ634830
Ceratocystiopsis rollhanseniana WINIM113 HQ595736
Ceratocystiopsis rollhanseniana WINIM1110 HQ634834
Ceratocystiopsis minima WINIM1865 HQ634856
Ceratocystiopsis minuta WINIM1523 HQ634857
Ceratocystiopsis pallidobrunnea WINIM151 HQ634842
Ceratocystiopsis concentrica WINIM153 Hq634849
Ceratocystiopsis minuta-bicolor WINIM1490 HQ634848
Ceratocystiopsis brevicomi WINIM1452 HQ282311
Ceratocystiopsis colifera WINIM1908 HQ634832
Ceratocystiopsis ranaculosus WINIM1919 HQ634840
Ceratocystiopsis manitobensis WINIM1237 HQ634850
Ceratocystiopsis sp. WINIM1520 HQ634836
Ceratocystiopsis sp. WINIM1522 HQ634838
Ceratocystiopsis sp. WINIM1522 HQ634838

In general the introns are not stable markers that could be used in taxonomic applications; for example, among 8 strains indentified as *Cop. minuta* three stains [WIN(M)1453, 1537, 1523] had the S943 introns but 5 strains lacked introns (WINM) 1532, 1533, 1534, 1535, 1536) and one strain of *Cop. minuta* (Win(M) 1537), in addition to the S943 intron, had the S1199 intron. However, as mentioned previously *Cop. minuta* could represent a species complex. All *Cop. minima* strains had the S943 intron but one strain of *Cop. rollhanseniana* (WIN(M)113) had both the S943 and S1199 introns whereas the second strain of this species WIN(110) had no introns.

The presence of multiple introns (S943, S989, S1199) within *Cop. pallidobrunnea* is of interest as: (1) it is the only examined member of the *Ceratocystiopsis* clade that has three introns including the S989 intron; and (2) it appears to be a basal member of the genus *Ceratocystiopis* (this paper Figure 9.1.1. and Plattner et al., 2009). In this study the only other fungus allied with the ophiostomatoid fungi that has the S989 intron is a strain of *Leptographium* (WIN(M) 984). This rather spotty distribution of some group I introns does raise questions with regards to lateral transfers but at this stage we need to examine more members with *Grosmannia* to get a better understanding of the evolutionary dynamics of this particular group I intron.

For both comparisons between the SSU rDNA tree and S943 or S1199 trees the topologies between the host tree and the intron trees were somewhat discrepant (Figure 9.1.5.). Considering how closely related species of *Ceratocystiopsis* are (see Plattner et al. 2009) and the weak statistical support obtained for many nodes within the intron trees

we feel the discrepancies noted were more likely the result of the limited resolution within the intron sequence phylogenetic analysis than evidence for lateral transfer of group I introns among species of *Ceratocystiopsis*. However, we cannot rule out the possibility of these introns moving laterally among members of this genus. Using the S943 and S1199 introns as discreet characters in ancestral state analysis provides support that S943 is an ancestral character for the members of the *Ceratocystiopsis* clade, however for the S1199 intron this analysis failed to provide any support for the hypothesis that this intron is ancestral to this group. The latter therefore suggests that the S1199 intron could have been gained independently by members of this genus.

Overall, our data appear to be similar to what have been observed in some other systems (Nikoh & Fukatsu 2001; Bhattacharya et al. 2002), which is that the evolutionary history of group I introns within the SSU rDNA gene among ophiostomatoid fungi is driven by rare gains (such as lateral transfers) followed by rapid loss(es) in descendant lineages, thus resulting in what appears to be sporadic intron distributions (Figure 9.1.1.). However, it appears that for some introns (S1199) frequent gain cannot be excluded as a possibility of the spotty distribution of some introns.

The distribution of nuclear group I introns can be the result of a variety of processes such as intron gain by horizontal and/or vertical transmission and intron loss due to precise excision. Reverse splicing allows a free group I intron RNA to insert into a homologous or heterologous RNA; this mode of mobility requires complementary base pairing between the intron and the exon RNA sequences. This model of transposition

would explain how group I introns, as encountered in this study, that lack intron-encoded proteins can persist in populations. This mechanism requires the additional steps of reverse transcription of the RNA and the integration of the cDNA into the rDNA locus by a recombination step that replaces the intron-less copy with the intron-plus cDNA. Although reverse splicing has not yet been demonstrated in genetic crosses, circumstantial evidence would suggest that nuclear rDNA group I introns can move by this mechanism (Bhattacharya et al. 2002 & 2005). As reverse splicing requires less homology (4-6 nucleotides), this mechanism allows introns to spread more easily into heterologous sites, although in this study we did not see any examples of ectopic integration.

A possible scenario for precise intron loss can be envisioned by the reverse transcription of a processed (i.e. spliced) version of the SSU rRNA followed by a recombination event whereby the intron-minus cDNA replaces the intron-plus genomic version of the SSU rDNA. This mechanism could remove introns from one rDNA unit, but the subsequent mechanism(s) whereby the now rDNA-intron-minus unit replaces the other intron hosting rDNA units within the repetitive rDNA gene family are unknown. One would assume that genetic processes that drive concerted evolution would be involved in this transition.

9.1.5.2. Group I introns in nuclear rDNA:

Ribosomal RNA genes are under functional constraints and thus are expected to evolve slowly, ideal for ribozyme elements that need to maintain their core sequences to maintain auto-splicing abilities or potentially face extinction, as they might become toxic to the host gene (Goddard et al. 2006; Lynch et al. 2006; Hausner, 2012). Also repetitive DNA is ideal for mobile introns as intron-minus and intron-plus alleles could exist within the same space; thus rDNA offers the presence of multiple targets for group I introns to invade.

In fungi, like most eukaryotes, ribosomal RNA genes are organized in tandem repeats with copy numbers for the repeats varying from 45 (*Aspergillus nidulans*) to 180 (*Saccharomyces paradoxus*) in taxa examined so far (Ganley & Kobayashi, 2007). A recent study showed that among members of the Saccharomycetaceae, *Aspergillus nidulans* and *Cryptococcus neoformans* rDNA repeats showed remarkably low levels of sequence variation, strongly suggesting that concerted evolution operates within this repetitive gene family, leading to rapid homogenization among its individual repeats (Ganley & Kobayashi, 2007). Homogenization in part is thought to be accomplished by a continual turnover of repeat copies by unequal recombination (reviewed in Elder & Turner 1995; Liao, 1999; Eickbush & Eickbush, 2007). The spread or loss of rDNA introns within the rDNA tandem repeats may be accelerated because of concerted evolution. An intron may gain a foothold within rDNA not because it has a highly efficient way to rapidly invade all available sites but because by chance one rDNA copy

that gained an intron eventually becomes the dominant version. Conversely, an intron might be lost as an intron-less version over time replaces rDNA repeats with introns.

The presence of nuclear rDNA group I introns raises some interesting questions as to why they are found in fungi, protozoans and algae but so far have not been noted in plant and metazoan nuclear rDNA. Fungi are organisms that represent an "open system" (no true cell walls separating cells) with limited separation between somatic and germ line tissue; thus, there are few barriers separating somatic compartments from those involved in generating meiotic or mitotic spores. Such a system may allow for transmitting mobile genetic agents by transient hyphal contacts or via intermediates such as viruses or genetic transfer systems similar to those within *Agrobacterium tumefaciens* (Andersson, 2009). Also as discussed previously concerted evolution may be a mechanism that allows introns to spread through the rDNA repeats. One can also speculate that rDNA tolerates group I introns, as rDNA is transcribed by RNA polymerase I, which must be tolerant to transcribing structural RNAs (such as rRNAs). Finally, in fungi repeat-induced point mutations (RIP) are a mechanism that guards against mobile elements spreading within the genome; however, RIP does not appear to operate in the nucleolus (specifically the nuclear organizer regions; reviewed in Hane & Oliver 2010). The rDNA tandem repeats are therefore not affected by this mechanism and mobile elements such as group I introns have a refuge within this genomic niche. In other regions of the genome RIP mutates duplicated DNA sequences during sexual reproduction, potentially inactivating copies of repeated sequences and thus limiting the spread of mobile elements (Selker, 2002; Diguistini et al. 2011). All the above factors

probably contribute towards the evolutionary dynamics of nuclear rDNA group I introns within the fungi.

9.1.5.3. SSU rDNA and taxonomic implications for *Ophiostoma sensu lato*:

The taxonomy of the Ophiostomatales and the family Ophiostomataceae is very complex (Hausner & Reid 2004; Zipfel et al. 2006) and historically has been subject of vigorous debates (Upadhyay 1981; Hausner et al. 1993a & 1993b; Spatafora & Blackwell 1994). Recently Zipfel et al. (2006), based on partial LSU rDNA and β-tubulin sequences, provided some justification to subdivide the genus *Ophiostoma* by resurrecting the genus *Grosmannia* for "*Ophiostoma* species" with *Leptographium* anamorphs and for maintaining the genus *Ceratocystiopsis* for ophiostomatoid species with short perithecial necks and falcate ascospores. What now remains in *Ophiostoma* sensu [Zipfel et al. (2006)] could broadly be defined as members of the "ips" and "pilifera" spore groups, as defined by Olchowecki and Reid (1974).

Our study showed that some *Pesotum* species are allied with members of *Grosmannia*. Okada *et al*. (2000), based on SSU rDNA data, placed *P. fragrans* next to *Ophiostoma penicillatum* (Grosmann) Siemaszko (now *Grosmannia penicillata* (Grosmann) Goid.). Zipfel et al. (2006) showed that *Grosmannia galeiformis* (B.K. Bakshi) Zipfel, Z.W. de Beer & M.J. Wingf., a species with both synnematous and mononematous anamorphs (i.e. *Pesotum*-like and *Leptographium*-like conidial states), belongs to the *Grosmannia* clade, so in the future more species with *Pesotum* conidial states might be assigned to this genus. Previous studies that included SSU rDNA

sequences representing different *Raffaelea* Arx & Hennebert emend. T.C. Harr. species (Gebhardt et al. 2005; Kolařík & Hulcr 2008; Harrington et al. 2010) suggests, that these asexual species are also phylogenetically linked to species of *Grosmannia*.

Ceratocystiopsis, a genus that includes species with falcate ascospores, short perithecial necks and, if present, anamophs assignable to *Hyalorhinocladiella*, formed a monophyletic clade. It has to be stated that the taxonomy of *Cop. minuta*, the type species for *Ceratocystiopsis*, is rather complex and, as currently circumscribed, probably includes several cryptic species (Plattner et al. 2009). To help resolve this situations strain WIN(M) 1532 [= UAMH 11218, = R. Jankowiak 705] was grown and dried on wood chips and has then been designated as the epitype (Reid and Hausner, 2010). *Ophiostoma longisporum*, based on ITS, partial β-tubulin and LSU rDNA sequences, was excluded from *Ceratocystiopsis* by Plattner et al. (2009) but in that study this species did group basally to other members of this genus. This study, albeit based on SSU rDNA sequences only, again confirms a basal position for *O. longisporum* next to the *Ceratocystiopsis* clade.

The SSU rDNA data suggest that there is a problem with *Ophiostoma sensu* Zipfel et al. (2006). Although our analysis shows strains centred around *O. ips* and *Ophiostoma piceae* (Münch) Syd. & P. Syd., including *O. minus* and *O. piliferum* (referred to as the *Ophiostoma* group), form a monophyletic grouping but several *Ophiostoma* species failed to group with the three currently accepted genera that comprise what used to be *Ophiostoma sensu lato*. Kolařík & Hulcr (2008) also showed in

their SSU rDNA analysis that *O. stenoceras* groups apart from both *Grosmannia* species and *Ophiostoma* species, forming a monophyletic group with *Sporothrix schenckii* Hektoen & C.F. Perkins and *E. scopularum.* The SSU rDNA data does suggest that the proposal by Zipfel et al. (2006) has some merit but it also shows that the status of *Ophiostoma* s.s. may need more consideration and additional genera may have to be proposed to accommodate those species not allied to the ips/pilifera/piceae group.

APPENDIX: 9.2.

PRELIMINARY WORK ON THE EXPRESSION AND PURIFICATION OF I-OmiI

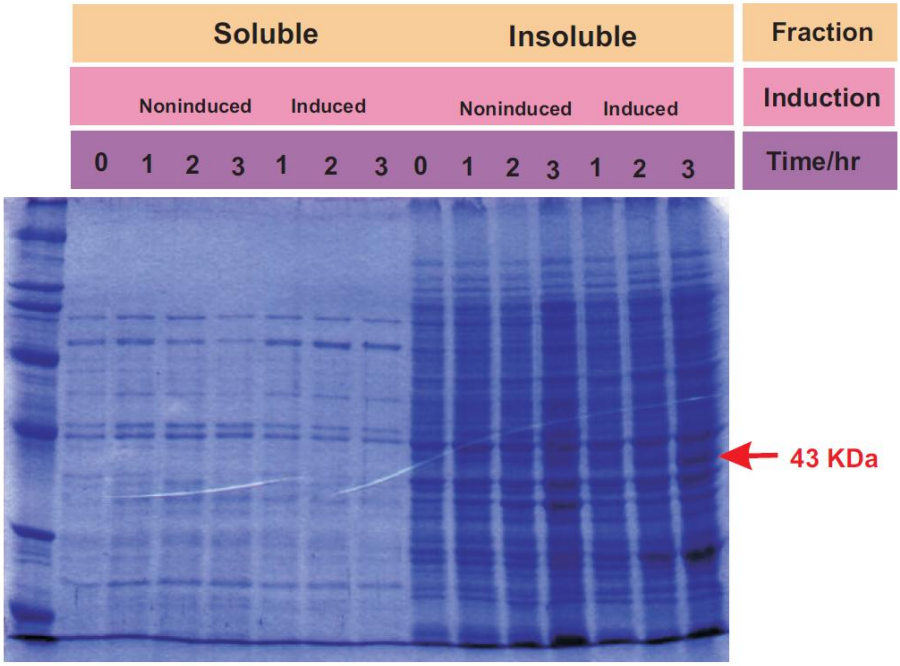
9.2.1. Construction of the expression plasmid:

A codon-optimized version of the I-OmiI sequence was synthesized commercially (Bio S & T, Montreal, QC, Canada) based on differences between the fungal mitochondrial and bacterial genetic code. The I-OmiI ORF was then amplified from the pBlueScript vector with primers I-OmiI-F (forward; included the CACC sequence needed for the topoisomerase cloning step that follows) and I-OmiI-R (reverse) using the Platinum Taq *DNA* Polymerase High Fidelity kit (Invitrogen). The amplified I-OmiI ORF was then moved into the 6xHis-Tag N-terminus pET200/D-TOPO® vector supplied in the Champion™ pET Directional TOPO® expression kit (Invitrogen) following the manufacturer's recommendations to generate the pET200/D/I-OmiI construct. The pET200/D/I-OmiI construct was subsequently transformed into the *E. coli* BL21 (DE3) cell line (Invitrogen). The transformants were then analyzed by colony PCR and by treating isolated plasmid DNAs with restriction enzymes to identify potential recombinant clones (*E. coli*-pET200/D/I-OmiI). One potential construct was sequenced with the vector-specific primers T7-F and T7-R to confirm that the ORF was in frame with the N-terminal fusion tag.

9.2.2. Overexpression of I-OmiI:

Initially, attempts were made to over express I-OmiI using the same expression conditions as were applied for I-OmiII (see section 2.10.2., Chapter 2). The *E. coli*-pET200/D/I-OmiI cells had been induced with 0.3 mM IPTG and were grown at 28 ^{o}C for 1, 2 and 3 hours. Both, the soluble and insoluble protein fractions were analyzed by SDS-PAGE (see Figure 9.2.1.). The results showed that, I-OmiI was only found in the insoluble fraction; this may be due to miss folding and aggregation of the protein.

Figure. 9.2.1. Overexpression of I-OmiI at 28 ^{o}C and 0.3 mM IPTG.

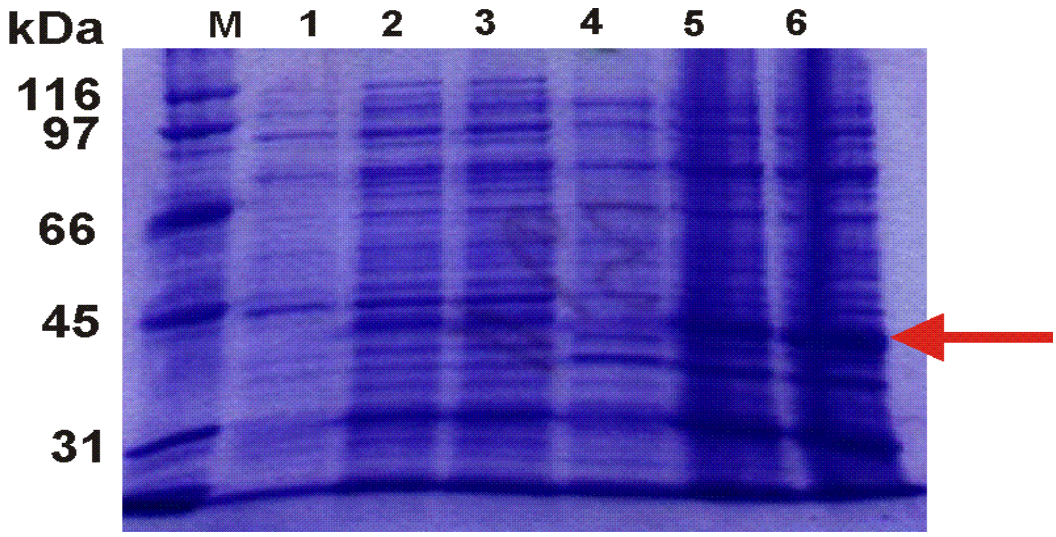


9.2.3. Increasing the solubility of I-OmiI:

In order to increase the solubility of the I-OmiI protein during the protein expression process and thus facilitate its purification thereafter, we tried the following conditions:

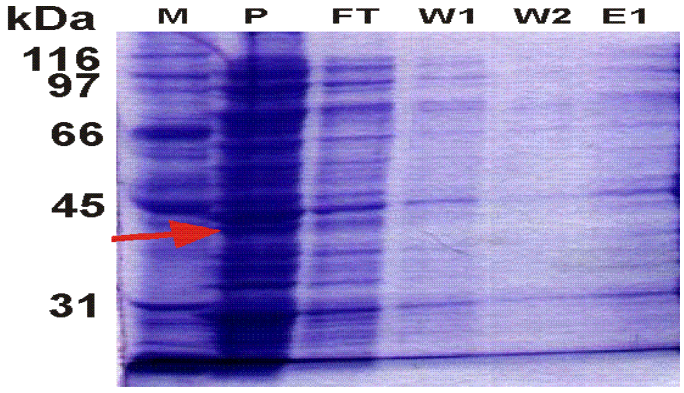
1. Slow down the growth of *E. coli*: According to the literature to allow for correct protein folding of eukaryotic protein when expressed in bacterial cells and to avoid precipitating of this protein in inclusion bodies, the growth conditions should be slowed down along with slowing down the expression of the foreign protein (Singh et al., 2009). Therefore, conditions were adjusted to slow down growth during the expression of I-OmiI by decreasing the incubation temperature (18 – 15 ^{o}C) and decreasing the IPTG concentration to 0.2 mM (see Figure 9.2.2.). The clear lysate (soluble fraction) was applied to a Ni-NTA column to purify I-OmiI and the results showed that I-OmiI was lost in the flow through fraction and it did not bind to the Ni-NTA resin.

Figure 9.2.2. SDS-PAGE for soluble (1-3) and insoluble (4-6) *E. coli* protein fractions after 15 hour incubation at 15 ^{o}C (2 & 5) and 18 ^{o}C (3 & 6). IPTG used at a final concentration of 0.2 mM to induce the protein expression. Lanes 1 & 4 contain the soluble and insoluble protein fractions respectively at zero induction time. The red arrow indicates the 43 kDa I-OmiI protein.



2. Treating the cell lysate with DNase: Like other DNA-binding enzymes, one would expect that I-OmiI could bind to DNA when cells are disrupted. To release the protein from DNA, the DNA can be removed by DNase in the presence of high salt concentration in order to dissociate the protein from the DNA. So DNase was applied during the early stage of protein purification in order to remove the contaminating DNA (Kang et al., 2007). TURBO DNase was added to the cell lysate during the sonication step and after that, 1/10 of sample volume of 5M NaCl was added to bring the salt concentration to 0.5M NaCl in the homogenate. Cell debris was spun down and the clears lysate was applied to a Ni-NTA column to purify I-OmiI (see Figure 9.2.3.). SDS-PAGE results showed that the I-OmiI protein was still found in the pellet (i.e. not soluble) and only a small amount of the enzyme was moved into the clear lysate, but the soluble fraction of this protein was lost with the flow through and thus no protein actually stuck to the Ni-NTA resin.

Figure 9.2.3. SDS-PAGE for I-OmiI purification after treating the cell lysate with TURBO DNase. M = Marker; P = Pellet; FT = Flow Through; W = Washing Fraction, E = Elution Fraction.



3. Changing the position of the His-Tag: Previous results showed that I-OmiI did not "stick" to the Ni-NTA column, suggesting that the His-Tag on the protein was not effective. This could be the result of misfolding of the N-terminal segment of the protein and thus the His-Tag is not accessible or available for binding to any ligand. A version of the I-OmiI ORF was amplified from the pBlueScript SK+/I-OmiI plasmid with the primers I-OmiI-F and I-OmiI-Rhis (see appendix 9.3. for primer sequence). The reverse primer I-OmiI-Rhis was designed to add a C-terminal 6X His-Tag to the I-OmiI protein. The PCR-amplified I-OmiI ORF was inserted into the expression vector pET200/D-TOPO® to generate the expression construct pET200/D/I-OmiI-C, which was transformed to *E. coli* BL21 for expression using methods as described previously for I-OmiII (see section 2.10, Chapter 2 for more details). The SDS-PAGE gel for I-OmiI expression indicates that adding a C-terminal 6xHis-Tag to the protein did not increase the solubility of I-OmiI and also did not affect the protein binding to Ni-NTA resin during the purification as the protein washed off in the flow through.

4. Solubilizing I-OmiI from inclusion bodies: Many recombinant eukaryotic proteins can aggregate and/or be packaged into inclusion bodies when expressed in bacterial cells (Przybycien et al., 1994). Analyzing the soluble and insoluble protein fractions after expressing I-OmiI in *E. coli* indicates that the enzyme is insoluble (and maybe packaged into inclusion bodies). Sarkosyl was used to purify natively folded I-OmiI from inclusion bodies following the protocol below (Tao et al., 2010):

1. Thaw bacterial pellet from a ~250 ml O/N culture at room temperature for 10 minutes.
2. Resuspend the bacterial pellet (P1) in 6 ml cell lysis buffer E (see appendix 9.5.3. for buffer composition) plus protease inhibitors. To each 6 ml of buffer E add 36.5 µl of 1M DTT and add 40 µl of 50 mg/ml lysozyme.
3. Sonicate the cell suspension with 10 short burst of 10 sec followed by intervals of 30 sec for cooling.
4. Centrifuge at 8500 rpm for 40 minutes (4 ^{o}C).
5. Analyse the supernatant (S1) by SDS-PAGE (Fraction 1: F1).
6. Resuspend the pellet (P2) in 7 ml buffer E supplemented with different concentrations of sarkosyl (0.5% to 2%).
7. Centrifuge at 10000 rpm for 30 minutes (4 ^{o}C).
8. Analyse the supernatant (S2) by collecting fractions (F2 to F8) for each sarkosyl concentration used (F = 3 µl sample + 12 µl H_20 + 5 µl of 2X sample buffer) from the supernatant in order to evaluate if the induced protein has remained soluble or been sequestered into inclusion bodies.

9. Run a 10% SDS-PAGE (see Figure 9.2.4 for schematic representation of the methodology and Figure 9.2.5. for SDS-PAGE).

The SDS-PAGE results (Figure 9.2.5) showed that sarkosyl can solubilize I-OmiI from inclusion bodies (cell debris/pellet) and there was no significant difference between the different sarkosyl concentrations used.

In order to purify I-OmiI, the F1 (supernatant 1 with no sarkosyl) and F2 fraction (supernatant 2 with the lowest sarkosyl concentration, 0.5%) were applied to a Ni-NTA column and the SDS-PAGE results (see Figure 7.3., Chapter 7) showed that I-OmiI was eluted in four fractions (E1 & E2: 125 mM imidazole; E3 & E4: 250 mM imidazole); these four fractions were then pooled together and dialyzed to remove the excess imidazole. However, the protein was partially precipitated inside the dialysis cassette when the sarkosyl concentration decreased. The dialyzed fraction was then centrifuged to remove the precipitated protein and the supernatant was concentrated with Amicon Ultra filtration tubes.

Figure 9.2.4. Methodology for solubilizing I-OmiI from inclusion bodies.

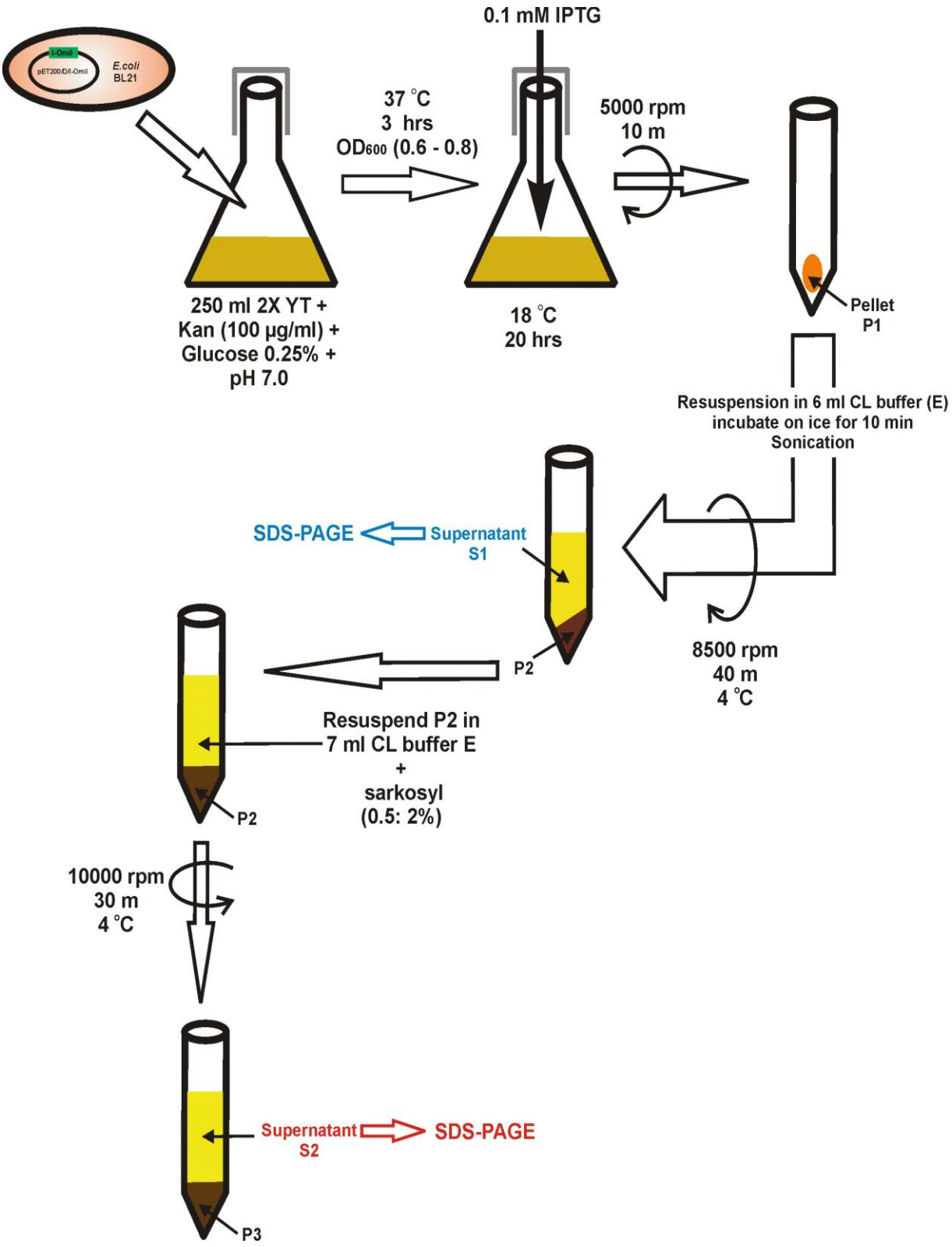
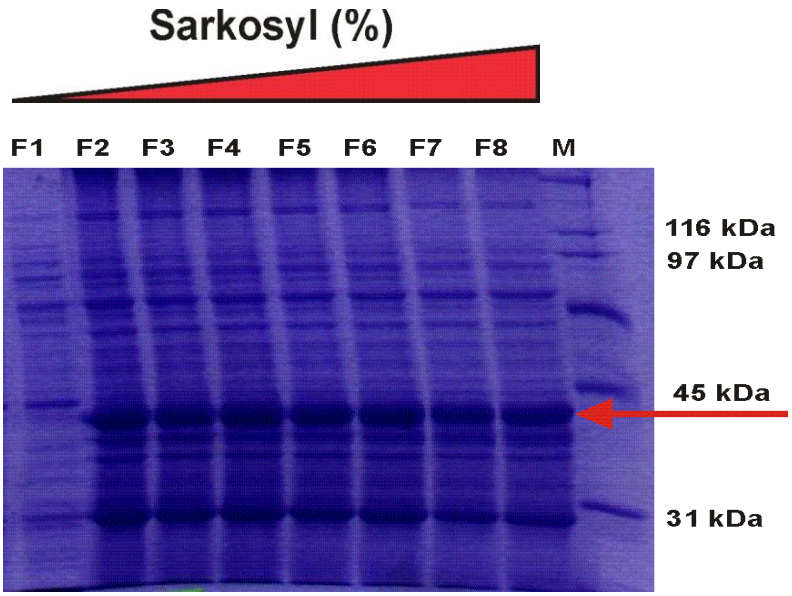


Figure 9.2.5. SDS-PAGE showing the results of solubilizing I-OmiI from inclusion bodies with sarkosyl (0.5:2%).



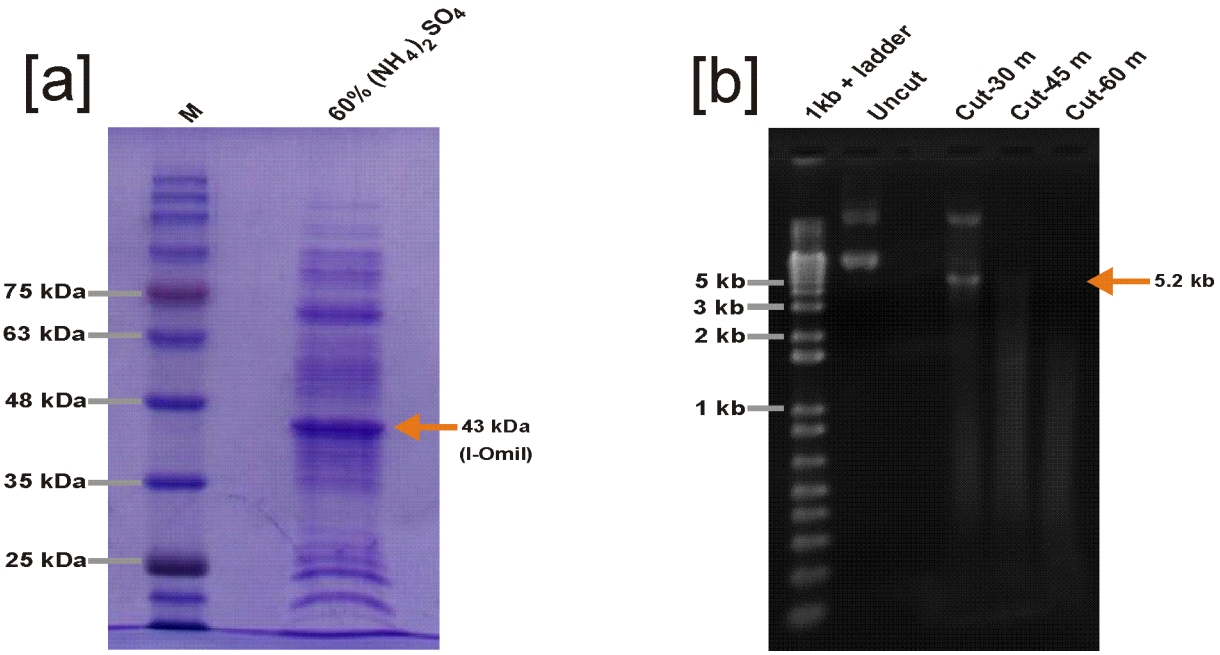
9.2.4. Preliminary endonuclease assays with crude I-OmiI preparations:

The endonuclease activity of partially purified I-OmiI was tested against the substrate plasmid pCR4/1574 (i.e. contains I-OmiI target site). The cleavage reaction mix contained: 4 µL of substrate plasmid (0.5 – 0.1 µg), 5 µl Invitrogen Buffer React #3 (100 mM NaCl, 50 mM Tris–Cl, pH 7.9 and 10 mM $MgCl_2$) supplemented with 1 mM DTT, 4 µL the crude I-OmiI preparation and 7 µl H_2O. Cleavage reactions were incubated at 37 ^oC for 30 and 60 minutes and the reactions were stopped with the addition of 2 µl of 200 mM EDTA (pH 8.0) and 1 µl of proteinase K (1 mg/mL) to each 10 µl aliquots, followed by an incubation period for 30 min at 37°C. The gel electrophoresis results (Figure 7.4., Chapter 7) showed that I-OmiI appears to be an active HEase that recognized the target site in the pCR4/1574 substrate plasmid and therefore generated a linearized molecule with a length of 5.2 kb.

The *in vitro* endonuclease assay of I-OmiI solubilized from inclusion bodies by sarkosyl indicates that this enzyme is an active HEase. To confirm the endonuclease activity of I-OmiI, the crude lysate of *E. coli* cells transformed with the pET200/D/I-OmiI plasmid, was used to assay for endonuclease activity such as cleaving the pCR4/1574 plasmid. The cleavage reaction mix contained: 2 µL of substrate plasmid (about 300 ng), 5 µl of Invitrogen Buffer React #3 (100 mM NaCl, 50 mM Tris–Cl, pH 7.9 and 10 mM $MgCl_2$) supplemented with 1 mM DTT, 1 µL crude cell lysate (60% cut of the ammonium sulphate; see Figure 9.2.6.[a]) and 12 µl H_2O. Cleavage reactions were incubated at 37 ^oC for 30, 45 and 60 minutes and the reactions were stopped with the addition of 2 µl of 200 mM EDTA (pH 8.0) and 1 µl of proteinase K (1 mg/mL) to each

10 µl aliquots, followed by incubation periods of 30, 45 & 60 min at 37°C. The gel electrophoresis results (Figure 9.2.6 [b]) showed that I-OmiI appears to cleave the substrate plasmid.

Figure 9.2.6. [a] SDS-PAGE gel of the 60% cut of the $(NH_4)_2SO_4$ fraction to the crude lysate. I-OmiI (43 kDa) is the most dominant protein in this fraction. M = BLUeye prestained protein ladder (GeneDireX), [b] Endonuclease assay for I-OmiI activity within the crude cell lysate.



APPENDIX: 9.3.

COMPLETE LIST OF THE PRIMERS USED IN THIS STUDY.

Primer name	Gene/Region	Direction	Sequence (5' → 3')
mtsr-1	*rns*	Forward	AGTGGTGTACAGGTGAG
M13 -F	*rns*	Forward	GTAAAACGACGGCCAG
Rns-gp2F1	*rns*	Forward	GAGTAACGTGGCAACACGGAAACTG
Rns-gp2F2	*rns*	Forward	CGTGGCAACACGGAAATCAC
rns-F	*rns*	Forward	GAGTTTGGTGATGGCTCTGATTGAACACTG
rns-ulmi-F3	*rns*	Forward	CAGGGATTAGAGACCCTTGTAGTCTTAG
rns-ulmi-F4	*rns*	Forward	CTTTAAAGCGCTACCGTCTTGATTC
rnsulmi-F5	*rns*	Forward	GCCCCAGGGCGAAAGGAAAATTAGTTA
rns-ulmi-F6	*rns*	Forward	GTGGTTTAGTAGACGCAGAAGGAAG
T7-F	*rns*	Forward	ATTAACCCTCACTAAAGGGA
371-F3	*rns*	Forward	CCTAAAACTATCAAACTCCGGGAACACCC
371-F4	*rns*	Forward	GAAATGGGTTATCGCGGAT
371-F5	*rns*	Forward	GGTGTAGGTAAAATCCATAGCTCTGGG
371-F6	*rns*	Forward	GAGCATTTAAATCAAGACGGTTTG
472-F2	*rns*	Forward	CCAACTGACTTCCACTAC
472-F3	*rns*	Forward	GACGCCAATTAAAGAATTCGGG
472-F4	*rns*	Forward	GGAATAAGCTCTCCTAGAGATAGG
472-F5	*rns*	Forward	CTCAGAGTTATCACATGGATTCC
472-F7	*rns*	Forward	CAAACTCTAGAACAGAAAGAGCC
472-F8	*rns*	Forward	CCATATATAAGAAACCCGGTATACC
494-F3	*rns*	Forward	CCCGGAAAAACCAACTGACTTCCAC
494-F4	*rns*	Forward	CTCTCCTAGAGATAGGATAATAC
494F5	*rns*	Forward	CTTCTTTGACAATATAGACCATAAC
494F6	*rns*	Forward	GAACCTAGTACTCCACTGCTTAGATG
515-F2	*rns*	Forward	GGATTCAAGCTCGTGTCATTGAGG
515-F3	*rns*	Forward	CCTAAATACACACAGACGCC
515-F4	*rns*	Forward	GGAATCAGCTCTCCTAGAGATAGG
515-F5	*rns*	Forward	GACATTAAAGGCTTCTTTGACAATATAG
515-F6	*rns*	Forward	CGAAGTCTACAATTCCCTAATCCAGAC
515-F8	*rns*	Forward	CCATAACCAAATTTATCGCTCTGCC
515-F9	*rns*	Forward	GGGAGGAGGTTAACAACAAATGAGG
515-F10	*rns*	Forward	GAGTACATGGAGGTAAGCATACTGGG
515-F11	*rns*	Forward	GCGAAGGAAGACTAGCGAAAGGAAGGG
mtsr-2	*rns*	Reverse	CGAGTGGTTAGTACCAATCC
M13-R	*rns*	Reverse	CAGGAAACAGCTATGAC
Rns-gp2-R1	*rns*	Reverse	CATTAACTGGAAACAGCCGTGCAAC
rns-R	*rns*	Reverse	CCACTACACGAACCGTATTTCGAC
rns-ulmi-R3	*rns*	Reverse	ACGTGCAACTTTCATCGCATACGGCTC
rns-ulmi-R4	*rns*	Reverse	CTCTTTTTTTAAACTTTCAGATAAACC
T3	*rns*	Reverse	TAATACGACTCACTATAGGG
371-R3	*rns*	Reverse	CCTTTCGCCCAGGGGACTAAGGTAG
371-R4	*rns*	Reverse	GATCCTCTTCCCTAGGCCCG
371-R5	*rns*	Reverse	CACATACTTCACTACTGGTGTC
472-R2	*rns*	Reverse	CTAAGACTACAAGGGTCTCTAATCC
472-R3	*rns*	Reverse	CTCCTCCATAGATTCGACC

472-R4	*rns*	Reverse	GGTAGCCTGTATTGGATTTACTAG
472-R5	*rns*	Reverse	GGGATATTAACAGCGAAGTCTG
472-R6	*rns*	Reverse	CTGATTTGATAATCTCTTTCAAGACCC
494-R2	*rns*	Reverse	GTTGCCACATTACTCATGAGGTGAAATG
494-R3	*rns*	Reverse	GCCTTCGTCCTTCGACGTCGG
494-R4	*rns*	Reverse	GGCTGTTTCCCTTTCTTTCGCTAG
494-R5	*rns*	Reverse	CTGTTTAAATTATTCATCGACTAATC
515-R3	*rns*	Reverse	GATCTTTCTCTCTAATATGTAAACAAG
515-R4	*rns*	Reverse	GATAAAAGAACCTTCCGCGTCAG
515-R5	*rns*	Reverse	CTGTATTTACTTCAGCGACCTTTCTC
515-R6	*rns*	Reverse	GTAGGGGTTTTTGTTAATTACATGG
515-R8	*rns*	Reverse	CTAACTTCTACTCTACCAC
515-R9	*rns*	Reverse	CTTAAATACCAACTGTACAGCTTTACC
515-R10	*rns*	Reverse	CTTCCTTCTGCATCTACTAAACCAC
515-R11	*rns*	Reverse	CCTTTCGCCCAGAGGAC
SSU-Z	ITS	Forward	ATAACAGGTCTGTGATG
SS-3	ITS	Forward	GTCGTAACAAGGTCTCCG
LSU-4	ITS	Reverse	TTGTGCGCTATCGGTCTC
LS-2	ITS	Reverse	GATATGCTTAAGTCAGCG
5.8S-R	ITS	Reverse	GACGCTCGGACAGGCATGCC
EF3E	EF-1α	Forward	GTCGTYATCGGCCACGTCGA
TEF-1rev	EF-1α	Reverse	GCCATCCTTGGAGATACCAGC
Ms952-F	952 intron	Forward	TTTGACACCAGTAGTGAAGTATG
Ms952DV-R	952 intron	Reverse	AACCGTACGTGCAACTTTCATCGCATACGGC
S379-R	379 intron	Reverse	CCGCGACTGCTGGCACGTAATTAGTC
SSU-J	SSU-Rrna	Forward	CTGGTTGATCCTGCCAGTAG
SSU-T	SSU-Rrna	Reverse	ACGGAACCTTGTTACGACT
T7-F	I-OmiI/II	Forward	TAATACGACTCACTATAGGG
T7-R	I-OmiI/II	Reverse	TAGTTATTGCTCAGCGGTGG
I-OmiI-F	I-OmiI	Forward	CACCATGGTGATTAGCCAATGGATG
I-OmiI-R	I-OmiI	Reverse	TCATTAGAATTCGCGACCGCGG
I-OmiI-R2	I-OmiI	Reverse	CAACCTTCGCCATCAATGAAACC
I-OmiII-F	I-OmiII	Forward	CACCATGCAGTCTCTGAAAAATAAC
I-OmiII-R	I-OmiII	Reverse	TCATTAGATGCGGTTGGTGTTC
I-OmiII-R2	I-OmiII	Reverse	GATCAGGTCGATAACCATGCAAAACAGG
I-OmiI-Rhis	I-OmiI	Reverse	TCATTAATGATGATGATGATGATGGAATTC GCGACCGCGGTTCAT

APPENDIX: 9.4.

GENEBANK ACCESSION NUMBERS OF ALL GENES/REGIONS SEQUENCED DURING THIS STUDY.

Accession #	Gene/region	Species	Strain
HQ292068	*rns*	*O. minus*	WIN(M) 873
HQ292069	*rns*	*O. minus*	WIN(M) 1574
HQ292070	*rns*	*O. minus*	WIN(M) 494
HQ292071	*rns*	*O. minus*	WIN(M) 371
HQ292072	*rns*	*O. minus*	WIN(M) 472
HQ292073	*rns*	*O. minus*	WIN(M) 515
HQ292074	*rns*	*O. novo-ulmi* subsp *americana*	DED 2002-10
HQ292075	*rns*	*O. ulmi*	DAOM 171046
HQ292076	RT	*O. minus*	WIN(M) 494
HQ292077	RT	*O. minus*	WIN(M) 472
HQ292078	RT	*O. minus*	WIN(M) 515
HQ292079	LHEase	*O. minus*	WIN(M) 873
HQ292080	LHEase	*O. minus*	WIN(M) 371
HQ292081	LHEase	*O. ulmi*	DAOM 171046
HQ292082	LHEase	*O. minus*	WIN(M) 371
HQ292083	LHEase	*O. minus*	WIN(M) 472
HQ292084	ITS	*O. novo-ulmi* subsp *americana*	DED 2002-10
HQ292085	ITS	*O. ulmi*	DAOM 171046
HQ292086	ITS	*O. minus*	WIN(M) 1574
HQ292087	ITS	*O. minus*	WIN(M) 873
HQ292088	ITS	*O. minus*	WIN(M) 494
HQ292089	ITS	*O. minus*	WIN(M) 371
HQ292090	ITS	*O. minus*	WIN(M) 472
HQ292091	ITS	*O. minus*	WIN(M) 515
HQ292092	EF-1alpha	*O. novo-ulm* subsp *americana i*	DED 2002-10
HQ292093	EF-1alpha	*O. ulmi*	DAOM 171046
HQ292094	EF-1alpha	*O. minus*	WIN(M) 1574
HQ292095	EF-1alpha	*O. minus*	WIN(M) 873
HQ292096	EF-1alpha	*O. minus*	WIN(M) 494
HQ292097	EF-1alpha	*O. minus*	WIN(M) 371
HQ292098	EF-1alpha	*O. minus*	WIN(M) 472
HQ292099	EF-1alpha	*O. minus*	WIN(M) 515
HQ634815	SSU rRNA	*Grosmannia davidsonii*	WIN(M)1495

APPENDIX: 9.4. CONTINUED

HQ634816	SSU rRNA	*Grosmannia davidsonii*	WIN(M)1494
HQ634817	SSU rRNA	*Meria laricis*	WIN(M)1525
HQ634818	SSU rRNA	*Meria laricis*	WIN(M)1528
HQ634819	SSU rRNA	*Grosmannia davidsonii*	WIN(M)1132
HQ634820	SSU rRNA	*Ophiostoma minus*	WIN(M)861
HQ634821	SSU rRNA	*Ophiostoma minus*	WIN(M)871
HQ634822	SSU rRNA	*Grosmannia penicillata*	WIN(M)27
HQ634823	SSU rRNA	*Ophiostoma brevicolle*	WIN(M)811
HQ634824	SSU rRNA	*Ophiostoma sp.*	WIN(M)1391
HQ634825	SSU rRNA	*Ophiostoma sp.*	WIN(M)1392
HQ634826	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1533
HQ634827	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1532
HQ634828	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1535
HQ634829	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1534
HQ634830	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1536
HQ634831	SSU rRNA	*Ophiostoma longisporum*	WIN(M)48
HQ634832	SSU rRNA	*Ceratocystiopsis collifera*	WIN(M)908
HQ634833	SSU rRNA	*Grosmannia crassivaginata*	WIN(M)918
HQ634834	SSU rRNA	*Ceratocystiopsis rollhanseniana*	WIN(M)110
HQ634835	SSU rRNA	*Graphium sp.*	WIN(M)1589
HQ634836	SSU rRNA	*Ceratocystiopsis sp.*	WIN(M)1520
HQ634837	SSU rRNA	*Ceratocystiopsis sp.*	WIN(M)1521
HQ634838	SSU rRNA	*Ceratocystiopsis sp.*	WIN(M)1522
HQ634839	SSU rRNA	*Pesotum sp.*	WIN(M)1394
HQ634840	SSU rRNA	*Ceratocystiopsis ranaculosus*	WIN(M)919
HQ634841	SSU rRNA	*Ophiostoma retusum*	ATCC 22324
HQ634842	SSU rRNA	*Ceratocystiopsis pallidobrunnea*	WIN(M)51
HQ634843	SSU rRNA	*Sarcotrochila macrospora*	WIN(M)1538
HQ634844	SSU rRNA	*Meria laricis*	WIN(M)1526
HQ634845	SSU rRNA	*Meria laricis*	WIN(M)1527
HQ634846	SSU rRNA	*Pesotum fragrans*	WIN(M)1396
HQ634847	SSU rRNA	*Pesotum sp.*	WIN(M)1426
HQ634848	SSU rRNA	*Ceratocystiopsis minuta-bicolor*	WIN(M)480
HQ634849	SSU rRNA	*Ceratocystiopsis concentrica*	WIN(M)53
HQ634850	SSU rRNA	*Ceratocystiopsis manitobensis*	WIN(M)237
HQ634851	SSU rRNA	*Ceratocystiopsis minima*	WIN(M)1462
HQ634852	SSU rRNA	*Ceratocystiopsis minima*	WIN(M)61
HQ634853	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1453
HQ634854	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1537
HQ634855	SSU rRNA	*Ophiostoma crenulatum*	WIN(M)58

APPENDIX: 9.4. CONTINUED

HQ634856	SSU rRNA	*Ceratocystiopsis minimum*	WIN(M)85
HQ634857	SSU rRNA	*Ceratocystiopsis minuta*	WIN(M)1523
HQ634858	SSU rRNA	*Graphium pseudormiticum*	WIN(M)1569
HQ634859	SSU rRNA	*Graphium pseudormiticum*	WIN(M)1571
HQ634860	SSU rRNA	*Graphium pseudormiticum*	WIN(M)1570
JF837329	*rns*	*O. novo-ulmi* subsp *novo-ulmi*	IMI 343.101
JN871711	*rns*	*O. hyalothecium*	ATCC 28825
JN871712	*rns*	*O. piliferum*	WIN(M)973
JN871713	*rns*	*G. penicillata*	N-27
JN871714	*rns*	*O. deltoideosporum*	WIN(M)870
JN871715	*rns*	*O. bacillosporum*	CBS 771.71
JN871716	*rns*	*O. stenoceras*	WIN(M)88
JQ303200	*rns*	*Ceratocystiopsis minuta*	WIN(M)1537
JQ303201	*rns*	*Ceratocystiopsis manitobensis*	WIN(M) 237
JQ303202	*rns*	*Ceratocystiopsis rollhanseniana*	WIN(M)113
JQ303203	*rns*	*Ophiostoma longisporum*	WIN(M)48
JQ303204	*rns*	*Ceratocystiopsis minuta*	WIN(M)1508
JQ303205	*rns*	*O. novo-ulmi* ssp. *americana*	DED 2002-92
JQ303206	*rns*	*Ophiostoma albidum*	WIN(M)88
JQ303207	*rns*	*Ceratocystiopsis* sp.	WIN(M)1521
JQ303208	*rns*	*Ceratocystiopsis* sp.	WIN(M)1520
JQ303209	*rns*	*Ceratocystiopsis minuta*	WIN(M)1536
JQ303210	*rns*	*Ceratocystiopsis minuta*	WIN(M)1533
JQ303211	*rns*	*Ceratocystiopsis pallidobrunnea*	WIN(M)51
JQ303212	*rns*	*Ceratocystiopsis rollhanseniana*	WIN(M)110
JQ988063	*rns*	*Grosmannia sp*	WIN(M)165
JQ988064	*rns*	*Ophiostoma brevicolle*	WIN(M)811
JQ988065	*rns*	*Ophiostoma stenoceras*	CBS 496.77
JX139037	*rns*	*Chaetomium thermophilum*	UAMH 2024
JX139038	*cDNA-rns*	*Chaetomium thermophilum*	UAMH 2024

APPENDIX: 9.5.

OPTIMIZED DNA AND AMINO ACID SEQUENCES OF

I-OMII AND I-OMIII

9.5.1. Optimized DNA sequence of I-OmiI:

5'-
caccATGGTGATTAGCCAATGGATGAAGGAAACCGAAATGGGCTACCGTGGCTCTAAAAGCGT
CCTGGTTAATAACACTGTCAAGGAACAGCGTGTGGACGGTTCTTATATGGGCAACTGTTTCCG
TCTGCCTATGCTGCGCTGCACGCTGATGGGTCTGGAACGCGATTATCAGATTCGCATCCTGTCC
AAAGGTCTGTACAAAAAAGAATGCAAGTTCTACTCTACTCTGAACAACAACATTAATCCGTGG
TTCCTGACCGGTTTCATTGATGGCGAAGGTTGCTTCAAAATCAGCCTGACCAAAGTGAACCGT
GCAATCGGCTGGCGTGTACAGCTGTTCTTCCAGATCAACCTGCACGAAAAAGATCGTGCGCTG
CTGGAATCCATCAAAAACTATCTGGGCGTTGGTAAAATCCATAGCTCTGGCAAAAATATCCTG
CAGTACCGTATTCAGACTTTCGACGAGTTCACTATCATCATCCGCCACATGGAAAAATACCCA
CTGATCTCCCAGAAACGTGCAGATTTCGAGCTGTTTAAGAAAGCGTACGACCTGGTTATGAAG
AACGAACATCTGAACCAAGACGGTCTGCTGAAAATCGTAAGCCTGAAAGCCTCCCTGAACCT
GGGCCTGTCTGATGATCTGAAACTGGCATTTCCGAACGTGATCCCGGCTACCCGCTTCACCGA
CTTTGCCGTCAACATCCCGGACCCGCAGTGGCTGGCGGGCTTCGCGTCCGCTGAAGGTTGTTT
TATGGTGGGTATCAAAAAATCTTCCAAAAGCAACACGGGTTATCAGGTTTACGTTATCTTTAT
CATTACCCAGCACATTCGTGATGAGCTGCTGATGAAAGGTATTCTGGACTACCTGGACTGTGG
TCGTCTGGCTCGTAAGCGTGACGTATATGAATACCAGGTTTCCAAGTTTAGCGATGTAGTTGA
CAAAATTCTGGGTTTCTTCGATAAATACCCGATCCTGGGCGAAAAAGCGAAAGACCTGCTGGA
TTTCTGCATTGTGTCTGCTCTGATGAAATCTAAAGACCACCTGACCGAGGTTGGTGTTGCAAA
AATCCGTAAGATCAAAGAGGGCATGAACCGCGGTCGCGAATTC**taatga**
-3'

9.5.2. Amino acid sequence of I-OmiI based on the optimized DNA sequence:

MVISQWMKETEMGYRGSKSVLVNNTVKEQRVDGSYMGNCFRLPMLRCTLMGL
ERDYQIRILSKGLYKKECKFYSTLNNNINPWFLTGFIDGEGCFKISLTKVNRAIGW
RVQLFFQINLHEKDRALLESIKNYLGVGKIHSSGKNILQYRIQTFDEFTIIIRHMEK
YPLISQKRADFELFKKAYDLVMKNEHLNQDGLLKIVSLKASLNLGLSDDLKLAF
PNVIPATRFTDFAVNIPDPQWLAGFASAEGCFMVGIKKSSKSNTGYQVYVIFIITQ
HIRDELLMKGILDYLDCGRLARKRDVYEYQVSKFSDVVDKILGFFDKYPILGEKA
KDLLDFCIVSALMKSKDHLTEVGVAKIRKIKEGMNRGREF

9.5.3. Optimized DNA sequence of I-OmiII:

5' –
caccATGCAGTCTCTGAAAAATAACATTGAATACCTGAATTGGTATATTTGCGGCCTGGTCGAT
GCGGAGGGTTCTTTCGGCGTTAACGTTGTTAAACACGCGACCAACAAAACTGGCTACGCAGTC
CTGACCTACTTCGAGATCGCTATGAACTCCAAGGACAAACAGCTGCTGGAGCTGATTAAGAA
AACCTTCGACCTGGAGTGCAATATCTACCACAATCCGAGCGATGATACCCTGAAATTCAAAGT
GTCCAACATCGAACAGATCGAAAACAAAATTATCCCGTTCTTCAAAAAGTATACGCTGTTCTC
TCAGAAGTCCGGCGACTTTATCCTGTTTTGCATGGTTATCGACCTGATCAAAAACAAAGAACA
CCTGACCCTGGACGGTCTGATCAAGATTCTGAACATCAAAGCGGCGATGAACCTGGGTCTGAG
CGAAAACCTGAAGAAAGAATTCTCCCTGTCTTCTAAGGGCGTACTGAAACGTCCTGCACCGGA
TCTGTCTAACCTGAACAAACGTTGGCTGGCTGGTTTCATTGAAGGTGAGGCATGTTTCTTCGTT
TCCATCTACAACTCTCCAAAAAGCAAACTGGGTAAAGCTGTACAACTGGTTTTCAAAATCACT
CAGCATATCAAAGACAAGATCCTGATTGAATCTGTAGTGGAACTGCTGAACTGTGGTCGTGTG
GAAGTGCGCAAATCTAACGAAGCCTGCGATTTCACGGTTACCAGCATTAAAGAAATCGAAAA
ATACATTATTCCGTATTTTAACGAGTACCCGCTGATCGGTCAGAAACTGTATAACTACGAAGA
TTTTAAACTGATCTTTAACATGATGAAAACTAAAGACCACCTGACTGAAGACGGCCTGAGCAA
AATCATCGAAATCCGTAACAAAATGAACACCAACCGCATC**taatga**
- 3'

9.5.4. Amino acid sequence of I-OmiII based on the optimized DNA sequence:

MQSLKNNIEYLNWYICGLVDAEGSFGVNVVKHATNKTGYAVLTYFEIAMNSKD
KQLLELIKKTFDLECNIYHNPSDDTLKFKVSNIEQIENKIIPFFKKYTLFSQKSGDFI
LFCMVIDLIKNKEHLTLDGLIKILNIKAAMNLGLSENLKKEFSLSSKGVLKRPAPD
LSNLNKRWLAGFIEGEACFFVSIYNSPKSKLGKAVQLVFKITQHIKDKILIESVVEL
LNCGRVEVRKSNEACDFTVTSIKEIEKYIIPYFNEYPLIGQKLYNYEDFKLIFNMKT
KDHLTEDGLSKIIEIRNKMNTNRI

APPENDIX: 9.6.

MEDIA AND BUFFERS COMPOSITION

9.6.1. Media for bacterial and fungal growth:

Luria Bertani (LB) medium:
(Recipe for 1 litter)
10g Bacteriological peptone
5g Yeast extract
5g NaCl
H_2O to 1L

Super optimal broth with Catabolite repression (SOC) medium:
(Recipe for 1 litter)
20g Bacteriological tryptone
5g Yeast extract
0.584g NaCl
0.186g KCl
2.408g MgSO4
3.603g Glucose
H_2O to 1L

Malt-extract agar (MEA) medium:
(Recipe for 1 litter)
20g Malt extract
1g yeast extract
20g agar
H_2O to 1L

Peptone/Yeast-extract/Glucose (PYG) medium:
(Recipe for 1 litter)
1g peptone
1g Yeast extract
3g Glucose
H_2O to 1L

2x Yeast-extract Tryptone (2x YT) medium:
(Recipe for 1 litter)
16g Tryptone
10g Yeast extract
5g NaCl
Adjust pH to 7.0 with 5N NaOH
H_2O to 1L

Terrific Broth (TB) medium:
(Recipe for 1 litter)
12g Tryptone
24g Yeast extract
4ml Glycerol
100 ml (0.17M KH_2PO_4 / 0.72M K_2HPO_4 solution)
H_2O to 1L

9.6.2. DNA buffers:

DNA extraction buffer:
(Recipe for 100 ml)
11 ml 1M Tris (pH 8.0)
11 ml 0.5M EDTA (pH 8.0)
30.8ml NaCl 5M
11ml 10% CTAB
H_2O to 100ml

Tris-Borate EDTA (TBE) buffer:
89 mM Tris-borate
10 mM EDTA, pH 8.0

Tris-EDTA (TE) buffer:
10 mM Tris (pH 8.0)
1 mM EDTA

9.6.3. Cell lysis buffers:

Lysis Buffer A
50 mM potassium phosphate, pH 7.8
400 mM NaCl
100 mM KCl
10% glycerol
0.5% Triton X-100
10 mM imidazole

Lysis Buffer B
500 mM HEPES
100 mM NaCl

Lysis Buffer C
40 mM HEPES (pH 7.5)
800 mM NaCl
10% glycerol

Lysis Buffer D
50 mM Tris
100 mM NaCl
5% Glycerol
1 mM $MgCl_2$
5 mM DTT

Lysis Buffer E
40mM HEPES (pH 7.5)
800mM NaCl
10% glycerol
6mM BME
Protease inhibitor cocktail

9.6.4. Protein purification buffers:

Buffer NWB5 (Wash buffer for native conditions)
40 mM HEPES
800 mM NaCl
10% Glycerol
6 mM BME
5 mM Imidazole

Buffer NWB10 (Wash buffer for native conditions)
40 mM HEPES
800 mM NaCl
10% Glycerol
6 mM BME
10 mM Imidazole

Buffer NWB20 (Wash buffer for native conditions)
40 mM HEPES
800 mM NaCl
10% Glycerol
6 mM BME
20 mM Imidazole

Buffer NWB30 (Wash buffer for native conditions)
40 mM HEPES
800 mM NaCl
10% Glycerol
6 mM BME
30 mM Imidazole

Elution Buffer NEB125/250Mm
40 mM HEPES
800 mM NaCl
20% glycerol
125/250 mM imidazole

9.6.5. Protein dialysis buffers:

Dialysis Buffer A
40 mM Hepes (pH 7.5)
200 mM NaCl
3 mM b-Mercapto-ethanol

Buffer B
40 mM Hepes (pH7.5)
1.5 M NaCl
3 mM b-Mercapto-ethanol

9.5.6. Protein storage buffer
40 mM Hepes (pH 7.5)
400 mM NaCl
0.5 mM DTT
10% (w/v) glycerol

9.6.7. SDS –PAGE gel formulations:

Running (Resolving) Gel:

	7.5%		10%		12.5%		15.%	
	1 gel	2 gels	1 gel	2 gels	1 gel	2 gels	1 gel	2 gels
dH_2O (ml)	1.75	3.5	1.6	3.2	1.25	2.5	0.9	1.8
Acry.:Bis-acry., 30:0.8 % (ml)	1.0	2.0	1.33	2.67	1.66	3.33	2.0	4.0
1.5 M Tris-Cl, pH 8.8. (ml)	1.0	2.0	1.0	2.0	1.0	2.0	1.0	2.0
10% SDS (µl)	40	80	40	80	40	80	40	80
10% APS (µl)	40	80	40	80	40	80	40	80
TEMED µl)	4	8	4	8	4	8	4	8
Total Volume (ml)	4	8	4	8	4	8	4	8

Stacking (Loading) Gel:

	4%		6%	
	1 gel	2 gels	1 gel	2 gels
dH_2O (ml)	1.5	3.0	1.3	2.6
Acry.:Bis-acry., 30:0.8 % (ml)	0.33	0.67	0.5	1.0
0.5 M Tris-Cl, pH 6.8 (ml)	0.62	1.25	0.62	1.25
10% SDS (µl)	25	50	25	50
10% APS (µl)	25	50	25	50
TEMED (µl)	2.5	5	2.5	5
Total Volume (ml)	2.5	5	2.5	5

APPENDIX: 9.7.
LIST OF THE BACTERIAL CLONES PREPARED DURING THIS STUDY.

Code	Cell line	Plasmid/construct	Antibiotic marker
MH-1	*E. coli* DH5α	pCR4-371	Ampicillin (Amp.)
MH-2	*E. coli* DH5α	pCR4/1574	Amp.
MH-3	*E. coli* DH5α	pBSK-OmiI	Amp.
MH-4	*E. coli* DH5α	pBSK-OmiII	Amp.
MH-5	*E. coli* DH5α	pET200-OmiI	Kanamycin (Kan.)
MH-6	*E. coli* DH5α	pET200-OmiII	Kan.
MH-7	*E. coli* DH5α	pRK-415	Tetracyclin (Tc)
MH-8	*E. coli* DH5α	pCR4/1574*	Amp.
	* 1-Cleaved with I-OmiII, 2- Treated with T4 polymerase, 3- Ligated with T4 ligase.		
MH-9	*E. coli* PM191	pACYC184	Chloramphenicol (Cp)
MH-10	*E. coli* BL21 (DE3)	pET200-OmiI	Kan.
MH-11	*E. coli* BL21 (DE3)	pET200-OmiII	Kan.
MH-12	*E. coli* BL21 (DE3)	pET200-OmiI	Kan.
MH-13	*E. coli* BL21 (DE3)	pET200-lacZ	Kan.
MH-14	*E. coli* BL21 (DE3)	pET200-OmiII + pACYC184	Kan. + Cp
MH-15	*E. coli* BL21 (DE3)	pET200-OmiII + pACYC1574	Kan. + Cp
MH-16	*E. coli* Top10	pET200-OmiI	Kan.
MH-17	*E. coli* Top10	pET200-OmiII	Kan.
MH-18	*E. coli* DH5α	pRK-371	Amp.
MH-19	*E. coli* DH5α	pRK-1574	Tc
MH-20	*E. coli* DH5α	pACYC1574	Cp
MH-21	*E. coli* BL21 (DE3)	pET200-OmiII + pRK-415	kan. + Tc
MH-22	*E. coli* BL21 (DE3)	pET200-OmiI + pACYC-184	Kan. + Cp
MH-23	*E. coli* BL21 (DE3)	pET200-OmiI + pACYC-1574	Kan. + Cp
MH-24	*E. coli* BL21 (DE3)	pET200-OmiII + pACYC184	Kan. + Cp
MH-25	*E. coli* BL21 (DE3)	pET200-OmiII + pACYC1574	Kan. + Cp

APPENDIX: 9.8.
HUMAN SEQUENCES THAT ARE HIGHLY SIMILAR TO I-OmiI AND I-OmiII TARGET SITES.

9.8.1. Human sequences that are highly similar to the I-OmiI target site.

>ref|NT_008470.19| Homo sapiens chromosome 9 genomic contig, GRCh37.p5 Primary Assembly
Length=62237592

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
syntaxin-17

Score = 36.2 bits (18), Expect = 0.31
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Minus

```
Query  1         ATCTTTACTAGGTTTAAA  18
                 ||||||||||||||||||
Sbjct  31876718  ATCTTTACTAGGTTTAAA  31876701
```

>ref|NW_001839236.2| Homo sapiens chromosome 9 genomic contig, alternate assembly
HuRef SCAF_1103279188289, whole genome shotgun sequence
Length=2001 3418

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
syntaxin-17

Score = 36.2 bits (18), Expect = 0.31
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Plus

```
Query  1         ATCTTTACTAGGTTTAAA  18
                 ||||||||||||||||||
Sbjct  17068527  ATCTTTACTAGGTTTAAA  17068544
```

>ref|NT_011295.11| Homo sapiens chromosome 19 genomic contig, GRCh37.p5 Primary Assembly
Length=15894584

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features flanking this part of subject sequence:
106713 bp at 5' side: zinc finger protein 98
124447 bp at 3' side: zinc finger protein 492

Score = 34.2 bits (17), Expect = 1.2
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Minus

```
Query  5         TTACTAGGTTTAAAGGG  21
                 |||||||||||||||||
Sbjct  13974557  TTACTAGGTTTAAAGGG  13974541
```

>ref|NT_113891.2| Homo sapiens chromosome 6 genomic contig, GRCh37.p5 alternate locus group ALT_REF_LOCI_2
Length=4795371

Features flanking this part of subject sequence:
61622 bp at 5' side: POU domain, class 5, transcription factor 1 isoform 2
39984 bp at 3' side: class I histocompatibility antigen, Gogo-C*0203 alpha cha...

Score = 34.2 bits (17), Expect = 1.2
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Plus

```
Query  6         TACTAGGTTTAAAGGGT  22
                 |||||||||||||||||
Sbjct  2710104   TACTAGGTTTAAAGGGT  2710120
```

>ref|NT_008470.19| Homo sapiens chromosome 9 genomic contig, GRCh37.p5 Primary Assembly
Length=62237592

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
syntaxin-17

Score = 36.2 bits (18), Expect = 0.31
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Minus

```
Query  1         ATCTTTACTAGGTTTAAA  18
                 ||||||||||||||||||
Sbjct  31876718  ATCTTTACTAGGTTTAAA  31876701
```

>ref|NW_001839236.2| Homo sapiens chromosome 9 genomic contig, alternate assembly
HuRef SCAF_1103279188289, whole genome shotgun sequence
Length=2001418

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
syntaxin-17

Score = 36.2 bits (18), Expect = 0.31
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Plus

```
Query  1         ATCTTTACTAGGTTTAAA  18
                 ||||||||||||||||||
Sbjct  17068527  ATCTTTACTAGGTTTAAA  17068544
```

>ref|NT_011295.11| Homo sapiens chromosome 19 genomic contig, GRCh37.p5 Primary Assembly
Length=15894584

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features flanking this part of subject sequence:
106713 bp at 5' side: zinc finger protein 98
124447 bp at 3' side: zinc finger protein 492

Score = 34.2 bits (17), Expect = 1.2
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Minus

```
Query  5         TTACTAGGTTTAAAGGG  21
                 |||||||||||||||||
Sbjct  13974557  TTACTAGGTTTAAAGGG  13974541
```

>ref|NT_113891.2| Homo sapiens chromosome 6 genomic contig, GRCh37.p5 alternate locus group ALT_REF_LOCI_2
Length=4795371

Features flanking this part of subject sequence:
61622 bp at 5' side: POU domain, class 5, transcription factor 1 isoform 2
39984 bp at 3' side: class I histocompatibility antigen, Gogo-C*0203 alpha cha...

Score = 34.2 bits (17), Expect = 1.2
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Plus

```
Query  6         TACTAGGTTTAAAGGGT  22
                 |||||||||||||||||
Sbjct  2710104   TACTAGGTTTAAAGGGT  2710120
```

9.8.2. Human sequences that are highly similar to the I-OmiII target site.

```
>ref|NT_009952.14| Homo sapiens chromosome 13 genomic contig, GRCh37.p5 Primary
Assembly
Length=25443670

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Features flanking this part of subject sequence:
   113969 bp at 5' side: ras-related protein Rap-2a precursor
   391353 bp at 3' side: importin-5

 Score = 36.2 bits (18),  Expect = 0.31
 Identities = 18/18 (100%), Gaps = 0/18 (0%)
 Strand=Plus/Minus

Query  4         AGTGAAGTATGTTATTTA  21
                 ||||||||||||||||||
Sbjct  11320358  AGTGAAGTATGTTATTTA  11320341
```

```
>ref|NT_011109.16| Homo sapiens chromosome 19 genomic contig, GRCh37.p5 Primary
Assembly
Length=31387201

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Features flanking this part of subject sequence:
   556203 bp at 5' side: hypothetical protein LOC100507527
   203574 bp at 3' side: zinc finger protein 507

 Score = 32.2 bits (16),  Expect = 4.9
 Identities = 16/16 (100%), Gaps = 0/16 (0%)
 Strand=Plus/Minus

Query  7        GAAGTATGTTATTTAA  22
                ||||||||||||||||
Sbjct  4908381  GAAGTATGTTATTTAA  4908366
```

```
>ref|NT_008705.16| Homo sapiens chromosome 10 genomic contig, GRCh37.p5 Primary
Assembly
Length=39094935

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Features flanking this part of subject sequence:
   102359 bp at 5' side: coiled-coil domain-containing protein 3 precursor
   5179 bp at 3' side: optineurin

 Score = 32.2 bits (16),  Expect = 4.9
 Identities = 16/16 (100%), Gaps = 0/16 (0%)
 Strand=Plus/Minus

Query  5         GTGAAGTATGTTATTT  20
                 ||||||||||||||||
Sbjct  13085944  GTGAAGTATGTTATTT  13085929
```

```
>ref|NT_022139.13| Homo sapiens chromosome 2 genomic contig, GRCh37.p5 Primary Assembly
Length=1439476

                                                    Sort alignments for this subject sequence by:
                                                      E value  Score  Percent identity
                                                      Query start position  Subject start position
 Features flanking this part of subject sequence:
   706156 bp at 5' side: doublecortin domain-containing protein 2C

 Score = 32.2 bits (16),  Expect = 4.9
 Identities = 16/16 (100%), Gaps = 0/16 (0%)
 Strand=Plus/Plus

Query  4       AGTGAAGTATGTTATT  19
               ||||||||||||||||
Sbjct  945287  AGTGAAGTATGTTATT  945302
```

>ref|NT_011520.12| Homo sapiens chromosome 22 genomic contig, GRCh37.p5 Primary Assembly
Length=29755346

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features flanking this part of subject sequence:
36257 bp at 5' side: piwi-like protein 3
7554 bp at 3' side: small G protein signaling modulator 1 isoform 1

Score = 34.2 bits (17), Expect = 1.2
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Minus

```
Query  4         AGTGAAGTATGTTATTT  20
                 |||||||||||||||||
Sbjct  4585308   AGTGAAGTATGTTATTT  4585292
```

>ref|NT_007933.15| Homo sapiens chromosome 7 genomic contig, GRCh37.p5 Primary Assembly
Length=77412220

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
hypothetical protein LOC79974 isoform 1
hypothetical protein LOC79974 isoform 2

Score = 34.2 bits (17), Expect = 1.2
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Minus

```
Query  6         TGAAGTATGTTATTTAA  22
                 |||||||||||||||||
Sbjct  58804282  TGAAGTATGTTATTTAA  58804266
```

>ref|NT_167187.1| Homo sapiens chromosome 8 genomic contig, GRCh37.p5 Primary Assembly
Length=31697033

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features flanking this part of subject sequence:
246069 bp at 5' side: LON peptidase N-terminal domain and RING finger protein 1
4699 bp at 3' side: hypothetical protein LOC57604 isoform 1

Score = 32.2 bits (16), Expect = 4.9
Identities = 16/16 (100%), Gaps = 0/16 (0%)
Strand=Plus/Minus

```
Query  3       TAGTGAAGTATGTTAT  18
               ||||||||||||||||
Sbjct  717159  TAGTGAAGTATGTTAT  717144
```

>ref|NW_001842360.1| Homo sapiens chromosome X genomic contig, alternate assembly
HuRef SCAF_1103279188416, whole genome shotgun sequence
Length=28681338

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features flanking this part of subject sequence:
988148 bp at 5' side: melanoma-associated antigen B5
541559 bp at 3' side: DDB1- and CUL4-associated factor 8-like protein 2

Score = 32.2 bits (16), Expect = 4.9
Identities = 16/16 (100%), Gaps = 0/16 (0%)
Strand=Plus/Plus

```
Query  7         GAAGTATGTTATTTAA  22
                 ||||||||||||||||
Sbjct  18753906  GAAGTATGTTATTTAA  18753921
```

REFERENCES

Abu-Amero SN, Charter NW, Buck KW and Brasier CM. **1995**. Nucleotide-sequence analysis indicate 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. Current Genetics, 28: 54-59.

Adams PL, Stahley MR, Gill ML, Kosek AB, Wang J and Strobel SA. **2004a**. Crystal structure of a group I intron splicing intermediate. RNA, 10:1867-1887.

Adams PL, Stahley MR, Kosek AB, Wang J and Strobel SA. **2004b**. Crystal structure of a self splicing group I intron with both exons. Nature, 430:45-50.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. **1990**. Basic local alignment search tool. Journal of Molecular Biology, 215: 403-410.

Andersson JO. **2009**. Horizontal gene transfer between microbial eukaryotes. Methods in Molecular Biology, 532: 473-487.

Arnould S, Chames P, Perez C, Lacroix E, Duclert A, Epinat JC, Stricher F, Petit AS, Patin A and Guillier S. **2006**. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. Journal of Molecular Biology, 355: 443-458.

Arnould S, Delenda C, Grizot S, Desseaux C, Pâques F, Silva GH and Smith J. **2011**. The I-*Cre*I meganuclease and its engineered derivatives: applications from cell modification to gene therapy. Protein Engineering Design and Selection. Publications, 24(1-2): 27-31.

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

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

Baidyaroy D, Hausner G, Hafez M, Michel F, Fulbright D and Bertrand H. **2011**. Detection of a 973 bp insertion within the mtDNA *rns* gene in a mitochondrial hypovirulent strain of *Cryphonectria parasitica* isolated from nature. Fungal Genetics and Biology, 48:775-783.

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

Bates MR, Buck KW and Brasier CM. **1990**. Molecular variation in the Dutch elm disease fungus. In Molecular Evolution (ed. Clegg MT & O'Brian SJ), UCLA Symposium on Molecular and Cellular Biology New Series (122): Wiley-Liss: New York. pp. 171-178.

Bates MR, Buck KW and Brasier CM. **1993**. Molecular relationships of the mitochondrial DNA of *Ophiostoma ulmi* and the NAN and EAN races of *O . novo-ulmi* determined by restriction fragment length polymorphisms. Mycological Research, 97: 1093-1100.

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

Belfort M. **2003**. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes and Development, 17: 2860-2863.

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

Belfort M and Perlman PS. **1995**. Mechanisms of intron mobility. Journal of Biological Chemistry, 270(51): 30237-30240.

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

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

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

Bell-Pedersen D, Quirk S, Bryk M and Belfort M. **1991**. I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. Proceedings of the National Academy of Sciences USA, 88: 7719-7723.

BenJamaa ML, Lieutier F, Yart A, Jerraya A and Khouja ML. **2007**. The virulence of phytopathogenic fungi associated with the bark beetles *Tomicus piniperda* and *Orthotomicus erosus* in Tunisia. Forest Pathology, 37: 51-63.

Bertrand H. **2000**. Role of mitochondrial DNA in the senescence and hypovirulence of fungi and potential for plant disease control. Annual Review of Phytopathology, 38: 397-422.

Bhattacharya D, Friedl T and Damberger S. **1996**. Nuclear-encoded rDNA group I introns: origin and phylogenetic relationships of insertion site lineages in the green algae. Molecular Biology and Evolution, 13: 978-989.

Bhattacharya D, Friedl T and Helms G. **2002**. Vertical evolution and intragenic spread of lichen-fungal group I introns. Journal of Molecular Evolution, 55: 74-84.

Bhattacharya D, Lutzoni F, Reeb V, Simon D, Nason J and Fernandez F. **2000**. Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. Molecular Biology and Evolution, 17: 1971-1984.

Bhattacharya D, Reeb V, Simon DM and Lutzoni F. **2005**. Phylogenetic analysis suggests reverse splicing spread of group I introns in fungal ribosomal DNA. BMC Evolutionary Biology, 5: 68.

Bifano AL and Caprara MG. **2008**. A DExH/D-box protein coordinates the two steps of splicing in a group I intron. Journal of Molecular Biology, 383: 667-682.

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

Blocker FJ, Mohr G, Conlan LH, Qi L, Belfort M and Lambowitz AM. **2005**. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. RNA, 11: 14-28.

Bogdanove, AJ and Voytas DF **2011**. TAL effectors: customizable proteins for DNA targeting. Science, 333(6051): 1843-1846.

Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring, RB and Stoddard BL. **2003**. Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. Genes and Development, 17(23): 2875-2888.

Bonen L and Vogel J. **2001**. The ins and outs of group II introns. Trends in Genetics, 17: 322-331.

Boore JL. **1999**. Animal mitochondrial genomes. Nucleic Acids Research, 27: 1767-1780.

Brasier CM. **1977**. Inheritance of pathogenecity and cultural characters in *Ceratocystis ulmi*: hybridization of protoperithecial and non-aggressive strains. Transactions of the British Mycological Society, 68: 45-52.

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

Brasier CM. **1991**. *Ophiostoma novo-ulmi* sp. *novo-ulmi*, causative agent of current Dutch elm disease pandemics. Mycopathologia, 115: 151-161.

Brasier CM. **2000**. The rise of the hybrid fungi. Nature 405: 134-135.

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

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

Brasier CM, Lea J and Rawlings MK. **1981**. The aggressive and non-aggressive strains of *Ceratocyctis ulmi* have different optima growth. Transactions of the British Mycological Society, 76: 213-218.

Burke JM, Belfort M, Cech TR, Davies RW, Schweyen RJ, Shub DA, Szostak JW and Tabak HF. **1987**. Structural conventions for group-I introns. Nucleic Acids Research, 15: 7217-7222.

Burt A. **2003**. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proceedings of the Royal Society of London Series B-Biological Sciences, 270: 921-928.

Burt A and Trivers R. **2006**. Genes in Conflict: The Biology of Selfish Genetic Elements. Cambridge, MA: Belknap Press.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Müller KM, Pande N, Shang Z, Yu N and Gutell RR. **2002**. The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and

Other RNAs. BMC Bioinformatics, (3): 2. [Correction: BMC Bioinformatics, 3: 15.]

Cathomen T and Joung JK. **2008**. Zinc-finger nucleases: the next generation emerges. Molecular Therapy, 16(7): 1200-1207.

Cavalier-Smith T. **1991**. Intron phylogeny: A new hypothesis. Trends in Genetics, 7: 145-148.

Cech TR. **1988**. Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene, 73: 259-271.

Cech TR. **1990**. Self-splicing of group-I introns. Annual Review of Biochemistry, 55: 599-629.

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

Chan SH, Stoddard BL and Xu SY. **2011**. Natural and engineered nicking endonucleases -from cleavage mechanism to engineering of strand-specificity. Nucleic Acids Research, 39(1): 1-18. [Erratum in: Nucleic Acids Research, 2011, 39(5):1966.]

Chanfreau G and Jacquier A. **1996**. An RNA conformational change between the two chemical steps of group II self-splicing. European Molecular Biology Organization Journal, 15: 3466-3476.

Charter NW, Buck KW and 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*. Mycological Research, 100: 368-372.

Chatterjee P, Brady KL, Solem A, Ho Y and Caprara MG. **2003**. Functionally distinct nucleic acid binding sites for a group I intron encoded RNA maturase/DNA homing endonuclease. Journal of Molecular Biology, 329: 239-251.

Chen Z and Zhao H. **2005**. A highly sensitive selection method for directed evolution of homing endonucleases. Nucleic Acids Research, 33(18): e154.

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

Chevalier B, Turmel M, Lemieux C, Monnat RJ Jr and Stoddard BL. **2003**. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. Journal of Molecular Biology, 329(2): 253-269.

Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ and Stoddard BL. **2002**. Design, activity, and structure of a highly specific artificial endonuclease. Molecular Cell, 10(4): 895-905.

Chevalier BS, Monnat RJ Jr and Stoddard, BL. **2001**. The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. Nature Structural Biology, 8(4): 312-316.

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

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ and Voytas DF. **2010**. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics, 1862: 757-761.

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

Christopher DA and Hallick RB. **1989**. *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. Nucleic Acids Research, 17(19): 7591-608.

Clark-Walker GD. **1992**. Evolution of mitochondrial genomes in fungi. International Review of Cytology, 141: 89-127.

Copertino DW, Hall ET, Van Hook FW, Jenkins KP and Hallick RB. **1994**. A group III twintron encoding a maturase-like gene excises through lariat intermediates. Nucleic Acids Research, 22: 1029-1036.

Copertino DW and Hallick RB. **1993**. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. Trends in Biochemical Sciences, 18: 467-471.

Costa M, Dème E, Jacquier A and Michel F. **1997**. Multiple tertiary interactions involving domain II of group II self-splicing introns. Journal of Molecular Biology, 267: 520-536.

Costa M, Michel F and Westhof E. **2000**. A three-dimensional perspective on exon

binding by a group II self-splicing intron. European Molecular Biology Organization Journal, 19: 5007-5018.

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

Crooks GE, Hon G, Chandonia JM and Brenner SE. **2004**. WebLogo: A sequence logo generator. Genome Research, 14: 1188-1190.

Cummings DJ, McNally KL, Domenico JM and Matsuura ET. **1990**. The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. Current Genetics, 17: 375-402.

Cummings DJ, Michel F and McNally KL. **1989**. DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I mitochondrial gene from *Podospora anserina*: a gene with sixteen introns. Current Genetics, 16: 381-406.

Cummings DJ, Turker MS and Domenico JM. **1986**. Mitochondrial excision-amplification plasmids in senescent and long-lived cultures of *Podospora anserina*. In: Wickner RB, Hinnebusch A, Lambowitz AM, Gonsalus IC, Hollaender A (eds), Extrachromosoma1 elements in lower eukaryotes, New York, Plenum Press, pp. 129-146.

Dai L and Zimmerly S. **2003**. ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. RNA, 9: 14-19.

Dalgaard JZ and Garrett RA. **1992**. Protein-coding introns from the 23S rRNA-encoding gene form stable circles in hyperthermophilic arhaeon *Pyrobaculum organotrophum*. Gene, 121: 103-110.

Dalgaard JZ, Garrett RA and Belfort M. **1994**. Purification and characterization of two forms of I-DmoI, a thermophilic site-specific endonuclease encoded by an archaeal intron. Journal of Biological Chemistry, 269: 28885-28892.

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

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

Dassa B, London N, Stoddard BL, Schueler-Furman O and Pietrokovski S. **2009**. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. Nucleic Acids Research, 37(8): 2560-73.

Davé UP, Akagi K, Tripathi R, Cleveland SM, Thompson MA, Yi M, Stephens R, Downing JR, Jenkins NA and Copeland NG. **2009**. Murine leukemias with retroviral insertions at Lmo2 are predictive of the leukemias induced in SCID-X1 patients following retroviral gene therapy. PLoS Genetics, 5(5): e1000491.

Davis L and Maizels N. **2011**. DNA nicks promote efficient and safe targeted gene correction. PLoS One, 6(9): e23981.

Dayhoff MO, Schwartz RM and Orcutt BC. **1978**. A model of evolutionary change in proteins. In: Dayhoff MO, (ed), Atlas of protein sequence and structure. Washington DC, National Biomedical Research Foundation, pp. 345-352.

De Wachter R, Neefs JM, Goris A and Van de Peer Y. **1992**. The gene coding for small ribosomal subunit RNA in the basidiomycete *Ustilago maydis* contains a group I intron. Nucleic Acids Research, 20: 1251-1257.

Dean AB, Stanger MJ, Dansereau JT, Van Roey P, Derbyshire V and Belfort M. **2002**. Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. Proceedings of the National Academy of Sciences USA, 99: 8554-8561.

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

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

Deredec A, Godfray HC and Burt A. **2011**. Requirements for effective malaria control with homing endonuclease genes. Proceedings of the National Academy of Sciences USA, 108(43): 874-880.

Diguistini S, Wang Y, Liao NY, Taylor G, Tanguay P, Feau N, Henrissat B, Chan SK, Hesse-Orce U, Alamouti SM, Tsui CK, Docking RT, Levasseur A, Haridas S,

Robertson G, Birol I, Holt RA, Marra MA, Hamelin RC, Hirst M, Jones SJ, Bohlmann J and Breuil C. **2011**. Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. Proceedings of the National Academy of Sciences USA, 108: 2504-2509.

Doyon JB, Pattanayak V, Meyer CB and Liu DR. **2006**. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. Journal of the American Chemical Society, 128(7): 2477-2484.

Drager RG and Hallick RB. **1993**. A complex twintron is excised as four individual introns. Nucleic Acids Research, 21: 2389-2394.

Dragon F and Brakier-Gingras L. **1993**. Interaction of *Escherichia coli* ribosomal protein S7 with 16S rRNA. Nucleic Acids Research, 21: 1199-1203.

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

Dujon B. **1989**. Group I introns as mobile genetic elements: facts and mechanistic speculation - a review. Gene, 82: 91-114.

Dujon B and Belcour L. **1989**. Mitochondrial DNA instabilities and rearrangements in yeasts and fungi. In: Berg DE, Howe MM (eds), Mobile DNA. Washington DC, The American Society for Microbiology, pp. 861-878.

Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, Perlman PS and Vogt VM. **1989**. Mobile introns: definition of terms and recommended nomenclature. Gene, 82(1): 115-118.

Durai S, Mani M, Kandavelou K, Wu J, Porteus MH and Chandrasegaran S. **2005**. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucleic Acids Research, 33(18): 5978-5990.

Edgell DR. **2009**. Selfish DNA: homing endonucleases find a home. Current Biology, 19: R115-117.

Edgell DR, Chalamcharla VR and Belfort M. **2011**. Learning to live together: mutualism between self-splicing introns and their hosts. BMC Biology, 11: 9-22.

Edgell DR, Gibb EA and Belfort M. **2010**. Mobile DNA elements in T4 and related phages. Virology Journal, 7:290.

Eickbush TH. **1999**. Mobile introns: retrohoming by complete reverse splicing. Current Biology, 9: R11-14.

Eickbush TH and Eickbush DG. **2007**. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics, 175: 477-485.

Elder JF and Turner BJ. **1995**. Concerted evolution of repetitive DNA-sequences in eukaryotes. Quarterly Review of Biology, 70: 297-320.

Emelyanov VV. **2001**. Evolutionary relationship of Rickettsiae and mitochondria. FEBS Letters, 501(1):11-8.

Epinat JC, Arnould S, Chames P, Rochaix P, Desfontaines D, Puzin C, Patin A, Zanghellini A, Pâques F and Lacroix E. **2003**. A novel engineered

meganuclease induces homologous recombination in yeast and mammalian cells. Nucleic Acids Research, 31(11): 2952-2962.

Farrell B, Sequeira A, O'Meara BC, Normark B, Chung J and Jordal B. **2001**. The evolution of agriculture in beetles (Curculionidae: Scolytinae and Platypodini). Evolution, 55: 2011-2027.

Feau N, Hamelin RC and Bernier L. **2007**. Variability of nuclear SSU-rDNA group introns within *Septoria* species: incongruence with host sequence phylogenies. Journal of Molecular Evolution, 64: 489-499.

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

Fedorova O, Solem A and Pyle AM. **2010**. Protein-facilitated folding of group II intron ribozymes. Journal of Molecular Biology, 397: 799-813.

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

Felsenstein FJ. **2006**. PHYLIP (Phylogeny Inference Package). Version 3.6a. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Férandon C, Moukha S, Callac P, Benedetto JP, Castroviejo M and Barroso G. **2010a**. The *Agaricus bisporus cox1* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. PLoS One, 5: e14048.

Férandon C, Moukha S, Callac P, Benedetto JP, Castroviejo M and Barroso G. **2010b.** The largest reservoir of mitochondrial introns is a relic of an ancestral split gene. Nature Precedings: hdl:10101/npre.2010.4373.1.

Ferat JL and Michel F. **1993**. Group II self-splicing introns in bacteria. Nature, 364: 358-

361.

Flannagan RS, Linn T and Valvano MA. **2008**. A system for the construction of targeted unmarked gene deletions in the genus *Burkholderia*. Environmental Microbiology, 10(6): 1652-1660.

Flick KE, Jurica MS, Monnat RJ Jr and Stoddard BL. **1998**. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. Nature, 394: 96-101.

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

Forget L, Ustinova J, Wang Z, Huss VA and Lang BF. **2002**. *Hyaloraphidium curvatum*: A linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. Molecular Biology and Evolution, 19: 310–319.

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

Ganley ARD and Kobayashi T. **2007**. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Research, 17: 184-191.

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

Gara RI, Geiszler DR and Littke WR. **1984**. Primary attraction of the mountain pine beetle to lodgepole pine in Oregon. Annals of the Entomological Society of America 77: 333-334.

Gargas A, DePriest PT and Taylor JW. **1995**. Positions of multiple insertions in SSU rDNA of lichen-forming fungi. Molecular Biology and Evolution, 12: 208-218.

Gautheret D and Lambert A. **2001**. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. Journal of Molecular Biology, 313: 1003-1011.

Gebhardt H, Weiss M and Oberwinkler F. **200**5. *Dryadomyces amasae*: a nutritional fungus associated with ambrosia beetles of the genus *Amasa* (Coleoptera: Curculionidae, Scolytinae). Mycological Research, 109: 687-696.

Gerbi SA, Gourse RL and Clark CG. **1982**. Conserved regions within ribosomal DNA: locations and some possible functions. In: The Cell Nucleus, Volume X (Busch H, Rothblum L, eds.) Academic Press, New York. pp. 351-386.

Gibb EA and 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. Molecular Microbiology, 78(1): 35-46.

Gibb EA and Hausner G. **2003**. A group I intron-like sequence in the nuclear SSU gene of the ophiostomatoid fungus *Gondwanamyces proteae*. Mycological Research, 107: 1442-1450.

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

Gibbs JN and Brasier CM. **1973**. Correlation between cultural characters and pathogenecity in Ceratocystis ulmi isolates from Britain, Europe and North America. Nature, 241: 381-383.

Gimble FS. **2000**. Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiology Letters, 185(2): 99-107.

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

Gimble FS. **2007**. Engineering home endonucleases to modify complex genomes. Gene Therapy and Regulation, 3: 33-50.

Gimble FS, Moure CM and Posey KL. **2003**. Assessing the plasticity of DNA target site recognition of the PI-SceI homing endonuclease using a bacterial two-hybrid selection system. Molecular Biology, 334(5): 993-1008.

Gissi C, Iannelli F and Pesole G. **2008**. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity, 101: 301-320.

Gobbi E, Firrao G, Carpanelli A, Locci R and Van Alfen NK, **2003**. Mapping and characterization of polymorphism in mtDNA of *Cryphonectria parasitica*: evidence of the presence of an optional intron. Fungal Genetics and Biology, 40: 215–224.

Goddard MR and Burt A. **1999**. Recurrent invasion and extinction of a selfish gene. Proceedings of the National Academy of Sciences USA, 96: 13880-13885.

Goddard MR, Leigh J, Roger AJ and Pemberton AJ. **2006**. Invasion and persistence of a selfish gene in the Cnidaria. PLoS One, 1: e3.

Gogarten JP and Hilario E. **2006**. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. BMC Evolutionary Biology, 6: 94. doi:10.1186/1471-2148-6-94.

Golden BL, Kim H and Chase E. **2005**. Crystal structure of a phage Twort group I ribozyme-product complex. Nature Structural & Molecular Biology, 12: 82-89.

Gorton C, Kim SH, Henricot B, Webber J and Breuil C. **2004**. Phylogenetic analysis of the blue stain fungus *Ophiostoma minus* based on partial ITS rDNA and β-tubulin gene sequences. Mycological Research, 108: 759-765.

Gorton C and Webber JF. **2000**. Re-evaluation of the status of the blue stain fungus and bark beetle associate *Ophiostoma minus*. Mycologia, 92: 1071-1079.

Gowher H and Jeltsch A. **2001**. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG sites. Journal of Molecular Biology, 309: 1201-1208.

Gray MW, Burger G and Lang BF. **2001**. The origin and early evolution of mitochondria. Genome Biology, 2: 1018.1-1018.5.

Grizot S, Smith J, Daboussi F, Prieto J, Redondo P, Merino N, Villate M, Thomas, S, Lemaire L, Montoya G, Blanco FJ, Pâques F and Duchateau P. **2009**. Efficient

targeting of a SCID gene by an engineered single chain homing endonuclease. Nucleic Acids Research, 37: 5405-5419.

Grube M, Gutmann B, Arup U, de los Rios A, Mattsson J and Wedin M. **1999**. An exceptional group-I intron-like insertion in the SSU rDNA of lichen mycobionts. Current Genetics, 35: 536-5341.

Guo F, Gooding AR and Cech TR. **2004**. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. Molecular Cell, 16: 351-362.

Gutiérrez G, Blanco O, Divakar PK, Lumbsch HT and Crespo A. **2007**. Patterns of group I intron presence in nuclear SSU rDNA of the lichen family Parmeliaceae. Journal of Molecular Evolution 64,: 181-195.

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

Hafez M and Hausner G. **2011b**. Characterization of the O.ul-mS952 intron: a potential molecular marker to differentiate between *Ophiostoma ulmi* and *Ophiostoma novo-ulmi* subsp. *americana*. World Academy of Science Engineering and Technology, 59: 1767-1775.

Hafez M and Hausner G. **2011c**. Characterization of a double motif LAGLIDADG homing endonuclease from the blue stain fungus *Ophiostoma minus*. The 61st Canadian Society of Microbiologists conference, St. John's, Newfoundland, Canada. 20-23 June 2011.

Hafez M, Iranpour M, Mullineux S-T, Sethuraman J, Wosnitza K, Lehn P, Kroeker J, Loewen PC, Reid J and Hausner G. **2012**. Identification of group I introns

within the SSU rDNA gene in species of *Ceratocystiopsis* and related taxa, Fungal Biology, 116: 98-111.

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A and Stutz E. **1993**. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Research, 21: 3537-3544.

Halls C, Mohr S, Del Campo M, Yang Q, Jankowsky E and Lambowitz AM. **2007**. Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and –independent mechanisms, and general RNA chaperone activity. Journal of Molecular Biology, 365: 835-855.

Hamari Z, Pfeiffer I, Ferenczy L and Kevei F. **1999**. Interpretation of variability of mitochondrial genomes in the species *Aspergillus carbonarius*. Antonie van Leeuwenhoek, 75: 225-231.

Hane JK and Oliver RP. **2010**. *In silico* reversal of repeat-induced point mutation (RIP) identifies the origins of repeat families and uncovers obscured duplicated genes. BMC Genomics, 11: 655.

Harrington TC. **1993**. Biology and taxonomy of fungi associated with bark beetles. In: Schowalter TD, Filip GM, (eds), Beetle-Pathogen Interactions in Conifer Forests, Academic Press, pp. 37-58.

Harrington TC, Aghayeva DN and Fraedrich SW. **2010**. New combinations in *Raffaelea*, *Ambrosiella*, and *Hyalorhinocladiella*, and four new species from the redbay ambrosia beetle, *Xyleborus glabratus*. Mycotaxon, 111: 337-361.

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

Harris L and Rogers SO. **2008**. Splicing and evolution of an unusually small group I intron. Current Genetics, 54: 213-222.

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

Haugen P, Reeb V, Lutzoni F and Bhattacharya D. **2004**. The evolution of homing endonuclease genes and group I introns in nuclear rDNA. Molecular Biology and Evolution, 21: 129-140.

Haugen P, Simon DM and Bhattacharya D. **2005**. The natural history of group I introns. Trends in Genetics, 21: 111-119.

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

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

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

Hausner G, Iranpour M, Kim JJ, Breuil C, Davis CN, Gibb EA, Reid J, Loewen PC and Hopkin AA. **2005**. Fungi vectored by the introduced bark beetle *Tomicus piniperda* in Ontario, Canada and comments on the taxonomy of

Leptographium lundbergii, *L. terebrantis*, *L. truncatum* and *L. wingfieldii*. Canadian Journal of Botany, 83: 1222-1237.

Hausner G, Klassen GR and Reid J. **1993b**. Unusually compact ribosomal RNA gene cluster in *Sphaeronaemella fimicola*. Current Genetics, 23: 357-359.

Hausner G, Olsen R, Johnson I, Simone D, Sanders ER, Karol KG, McCourt RM and Zimmerly S. **2006**. Origin and Evolution of the Chloroplast *trnK* (*matK*) Intron: a Model for Evolution of Group II Intron RNA Structures. Molecular Biology and Evolution, 23: 380-391.

Hausner G and Reid J. **2004**. The nuclear small subunit ribosomal genes of *Sphaeronaemella helvellae*, *Sphaeronaemella fimicola*, *Gabarnaudia betae* and *Cornuvesica falcata*: phylogenetic implications. Canadian Journal of Botany, 82: 752-762.

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

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

Hausner G, Reid J and Klassen GR. **1993c**. *Ceratocystiopsis*: a reappraisal based on molecular criteria. Mycological Research, 97: 625-633.

Hausner G and Wang X. **2005**. Unusual compact rDNA gene arrangements within some members of the Ascomycota: evidence for molecular co-evolution between

ITS1 and ITS2. Genome, 48: 648-660.

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

Hibbett DS. **1996**. Phylogenetic evidence for horizontal transmission of Group 1 introns in the nuclear ribosomal DNA of mushroom-forming fungi. Molecular Biology and Evolution 13: 903-917.

Hintz WE, Jeng RS, Hubbes MM and Horgen PA. **1991**. Identification of three populations of *Ophiostoma ulmi* (Aggressive subgroup) by mitochondrial DNA restriction-site mapping and nuclear DNA fmgerprinting. Experimental Mycology, 15: 316-325.

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

Hintz, WE, Anderson JB and Horgen PA. **1989**. Relatedness of three species of Agclricus inferred from restriction fragment length polymorphism analysis of the ribosomal DNA repeat and mitochondrial DNA. Genome, 32: 173-178.

Hoegger PJ, Binz T and Heiniger U. **1996**. Detection of genetic variation between Ophiostoma ulmi and the NAN and EAN races of O. novo-ulmi in Switzerland using RAPD markers. European Journal of Forest Pathology, 26: 57-68.

Hosaka H, Nakagawa A, Tanaka I, Harada N, Sano K, Kimura M, Yao M and Wakatsuki S. **1997**. Ribosomal protein S7: a new RNA-binding motif with structural similarities to a DNA architectural factor. Structure, 5: 1199-1208.

Hoshina R and Imamura N. **2009**. Phylogenetically close group I introns with different positions among *Paramecium bursaria* photobionts imply a primitive stage of intron diversification. Molecular Biology and Evolution, 26: 1309-1319.

Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K and Morikawa K. **2000**. Crystal structure of an archaeal-intein-encoded homing endonuclease PI-PfuI. Journal of Molecular Biology, 300(4): 889-901.

Iliakis G, Wang H, Perrault AR, Boecker W, Rosidi B, Windhofer F, Wu W, Guan J, Terzoudi G and Pantelias G. **2004**. Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenetic and Genome Research, 104: 14-20.

Jackson SA, Cannone JJ, Lee JC, Gutell RR and Woodson SA. **2002**. Distribution of rRNA introns in the three-dimensional structure of the ribosome. Journal of Molecular Biology, 323: 35-52.

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

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

Jasin M. **1996**. Genetic manipulation of genomes with rare-cutting endonucleases. Trends in Genetics, 12: 224-228.

Jeggo PA. **1998**. DNA breakage and repair. Advances in Genetics (38): 185-218.

Jeng RS, Bernier L and Brasier CM. **1988**: A comparative study of cultural and electrophoretic characteristics Eurasian and North American races of *Ophiostoma ulmi*. Canadian Journal of Botany, 66: 1325-1333.

Jeng RS and Brasier CM. **1994**: Two dimensional mapping of mycelial polypeptides of *Ophiostoma ulmi* and *O. novo-ulmi*, causal agent of Dutch elm disease. Canadian Journal of Botany, 72: 370-377.

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

Johansen S, Elde M, Vader A, Haugen P, Haugli K and Haugli F. **1997**. *In vivo* mobility of a group I twintron in nuclear DNA of the myxomycete Didymium iridis. Molecular Microbiology, 24: 737-745.

Johansen S, Muscarella DE and Vogt VM. **1996**. Insertion elements in ribosomal DNA. In Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis (Zimmermann RA & Dahlberg, AE, eds), CRC Press, Boca Raton, FL, pp. 89-108.

Johnson RD and Jasin M. **2001**. Double-strand-break-induced homologous recombination in mammalian cells. Biochemical Society Transactions, 29: 196-201.

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

Jurica MS and Stoddard BL. **1999**. Homing endonucleases: structure, function and evolution. Cellular and Molecular Life Sciences, 55: 1304-1326.

Kang J, Lee MS and Gorenstein DG. **2007**. Application of RNase in the purification of RNA-binding proteins. Analytical Biochemistry, 365(1):147-148.

Kaminska KH, Kawai M, Boniecki M, Kobayashi I and Bujnicki JM. **2008**. Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. BMC Structural Biology, 8: 48.

Keating KS, Toor N, Perlman PS and Pyle AM. **2010**. A structural analysis of the group II intron active site and implications for the spliceosome. RNA, 16:1-9.

Keeble, AH, Maté MJ and Kleanthous C. **2005**. H-N-H endonucleases. In: Belfort M, Derbyshire V, Stoddard BL, Wood DL (eds) Homing endonucleases and inteins. New York, NY, Springer. pp. 49-66.

Khan H and Archibald JM. **2008**. Lateral transfer of introns in the cryptophyte plastid genome. Nucleic Acids Research, 36: 3043-3053.

Kim WK, Mauthe W, Hausner G and Klassen GR. **1990**. Isolation of high molecular weight DNA and double-stranded RNAs from fungi. Canadian Journal of Botany, 68: 1898-1902.

Kirisits T. **2004**. Fungal associates of European bark beetles with special emphasis on the ophiostomatoid fungi. In: Lieutier F, Day KR, Battisti A, Gre´goire JC, Evans HF (eds) Bark and wood boring insects in living trees in Europe, a synthesis. Kluwer, The Netherlands, pp. 181–235.

Kjems J and Garrett RA. **1991**. Ribosomal RNA introns in archaea and evidence for RNA conformational changes associated with splicing. Proceedings of the National Academy of Sciences USA, 88(2):439-443.

Klepzig KD. **1998**. Competition between a biological control fungus, *Ophiostoma piliferum*, and symbionts of the southern pine beetle. Mycologia, 90: 69-75.

Klug A. **2010**. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. Quarterly Reviews of Biophysics, 431: 1-21.

Köchel HG and Küntzel H. **1981**. Nucleotide sequence of the *Aspergillus nidulans* mitochondrial gene coding for the small ribosomal subunit RNA: homology to *E. coli* 16S rRNA. Nucleic Acids Research, 9(21):5689-5696.

Kolařík M and Hulcr J. **2008**. Mycobiota associated with the ambrosia beetle *Scolytodes unipunctatus* (Coleoptera: Curculionidae, Scolytinae). Mycological Research, 113: 44-60.

Konrad H, Kirisits T, Riegler M, Halmschlager E and Stauffer C. **2002**. Genetic evidence for natural hybridization between the Dutch elm disease pathogens *Ophiostoma novo-ulmi* ssp. *novo-ulmi* and *O. novo-ulmi* ssp. *americana*. Plant Pathology, 51: 78-84.

Kowalski JC, Belfort M, Stapleton MA, Holpert M, Dansereau JT, Pietrokovski S, Baxter SM and Derbyshire V. **1999**. Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. Nucleic Acids Research, 27(10): 2115-2125.

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

Kroymann J and Zetsche K. **1997**. The apocytochrome-b gene in *Chlorogonium elongatum* (Chlamydomonadaceae): an intronic GIY-YIG ORF in green algal mitochondria. Current Genetics, 31(5): 414-418.

Krüger DH and Bickle TA. **1983**. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiological Review, 47 (3): 345-60.

Lambowitz AM and Belfort M. **1993**. Introns as mobile genetic elements. Annual Review of Biochemistry, 62: 587-622.

Lambowitz AM, Caprara MG, Zimmerly S and 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 (eds), The RNA World (2nd edn.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 451-485.

Lambowitz AM, Mohr G and Zimmerly S. **2005**. Group II intron homing endonucleases: ribonucleoprotein complexes with programmable target specificity. In: Belfort M, Derbyshire V, Stoddard BL, Wood DL (eds), Homing endonucleases and inteins. New York, NY, Springer, pp. 121-145.

Lambowitz AM and Zimmerly S. **2004**. Mobile group II introns. Annual Review of Genetics 38: 1-35.

Lambowitz AM and Zimmerly S. **2011**. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harbor Perspectives in Biology, 3(8):a003616. doi: 10.1101/cshperspect.a003616.

Landthaler M, Shen BW, Stoddard BL and Shub DA. **2006**. I-BasI and I-HmuI: Two

phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. Journal of Molecular Biology, 358: 1137-1151.

Lang BF, Laforest MJ and Burger G. **2007**. Mitochondrial introns: a critical view. Trends in Genetics 23: 119-125.

Lehmann K and Schmidt U. **2003**. Group II introns: structural and catalytic versatility of large natural ribozymes. Critical Reviews in Biochemistry and Molecular Biology, 38: 249-303.

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

Li ZJ and Zhang Y. **2005**. Predicting the secondary structures and tertiary interactions of 211 group I introns in IE subgroup. Nucleic Acids Research, 33: 2118-2128.

Liang F, Romanienko PJ, Weaver DT, Jeggo PA and Jasin M. **1996**. Chromosomal double-strand break repair in Ku80-deficient cells. Proceedings of the National Academy of Sciences USA, 93 (17): 8929-8933.

Liao D. **1999**. Concerted evolution: Molecular mechanism and biological implications. American Journal of Human Genetics, 64: 24-30.

Lickey EB, Hughes KW and Petersen RH. **2003**. Variability and phylogenetic incongruence of an SSU nrDNA group I intron in *Artomyces*, *Auriscalpium*, and *Lentinellus* (Auriscalpiaceae: Homobasidiomycetes). Molecular Biology and Evolution, 20: 1909-1916.

LimYW, Alamouti SM, Kim JJ, Lee S and Breuil C. **2004**. Multigene phylogenies of *Ophiostoma clavigerum* and closely related species from bark beetle-attacked *Pinus* in North America. FEMS Microbiology, 237: 89-96.

Lindstrom SC and Pistolic J. **2005**. Detection of a Group I (IE) fungal intron in the green algal genus *Urospora* (Ulvophyceae). Journal of Phycology, 41: 359-365.

Lippow SM, Aha PM, Parker MH, Blake WJ, BaynesBM and Lipovsek D. **2009**. Creation of a type IIS restriction endonuclease with a long recognition sequence. Nucleic Acids Research, 37: 3061-3073.

Luan DD, Korman MH, Jakubczak JL and Eickbush TH. **1993**. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. Cell, 72: 595-605.

Lykke-Andersen J, Aagaard C, Semionenkov M and Garrett RA. **1997**. Archaeal introns: splicing, intercellular mobility and evolution. Trends in Biochemical Sciences, 9:326-31.

Lynch M, Koskella B and Schaack S. **2006**. Mutation pressure and the evolution of organelle genomic architecture. Science, 311:1727-1730.

Maddison WP, Maddison DR. **2010**. Mesquite: a modular system for evolutionary analysis. Version 2.73, http://mesquiteproject.org.

Madhani HD and Guthrie C. **1992**. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell, 71: 803–817.

Mak AN, Lambert AR and Stoddard BL. **2010**. Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. Structure, 18(10): 1321-31.

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

Marcaida MJ, Muñoz IG, Blanco FJ, Prieto J and Montoya G. **2010**. Homing endonucleases: from basics to therapeutic applications. Cellular and Molecular Life Sciences, 67: 727-748.

Margulis L. **1970**. Origin of Eukaryotic Cells, Yale University Press.

Martín MP, Coucheron DH and Johansen S. **2003**. Structural features and evolutionary considerations of group IB introns in SSU rDNA of the lichen fungus *Teloschistes*. Fungal Genetics and Biology, 40: 252-260.

Martínez-Abarca F and Toro N. **2000**. Group II introns in the bacterial world. Molecular Microbiology, 38: 917-926.

Masuya H, Kubono T and Ichihara Y. **2003**. *Ophiostoma ssiori* sp. nov. (Ophiostomatales, Ascomycetes) isolated from a bark beetle in *Prunus* species. Bulletin of the National Science Museum, Tokyo, Ser. B., 29: 35-43.

Matsuura M, Noah JW and Lambowitz AM. **2001**. Mechanism of maturase-promoted group II intron splicing. European Molecular Biology Organization Journal, 20: 7259e7270.

McConnell SA, Takeuchi R, Pellenz S, Davis L, Maizels N, Monnat RJ Jr and Stoddard BL. **2009**. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proceedings of the National Academy of Sciences USA, 106(13): 5099-5104.

Mereschkowsky C. **1905.** About nature and origin of the chromatophores in the plant kingdom. Biol. Centralbl., 25: 593-604.

Michel F, Costa M, Doucet AJ and Ferat JL. **2007**. Specialized lineages of bacterial group II introns. Biochimie, 89: 542-553.

Michel F, Costa M and Westhof E. **2009**. The ribozyme core of group II introns: a structure in want of partners. Trends in Biochemical Sciences, 34: 189-199.

Michel F and Cummings DJ. **1985**. Analysis of class I introns in a mitochondrial plasmid associated with senescence of *Podospora anserina* reveals extraordinary resemblance to the *Tetrahymena* ribosomal intron, Current Genetics, 10: 69-79.

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

Michel F and Ferat JL. **1995**. Structure and activities of group II introns. Annual Review of Biochemistry, 64: 435-61.

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

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

Michel F, Netter P, Xu MQ and Shub DA. **1990**. Mechanism of 3' splice site selection by the catalytic core of the sunY intron of bacteriophage T4: the role of a novel base-pairing interaction in group I introns. Genes & Development, 4: 777-788.

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

Michel F and Westhof E. **1990**. Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. Journal of Molecular Biology, 216: 585–610.

Mladenov E and Iliakis G. **2011**. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. Mutation Research, 711(1-2): 61-72.

Mohr G, Perlman PS and Lambowitz AM, **1993**. Evolutionary relationships among group 11 intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucleic Acids Research, 21: 4991e4997.

Montandon PE and Stutz E. **1983**. Nucleotide sequence of a *Euglena gracilis* chloroplast genome region coding for the elongation factor Tu; evidence for a spliced mRNA. Nucleic Acids Research, 11 (17): 5877-92.

Monteiro-Vitorello CB, Hausner G, Searles DB, Gibb EA, Fulbright DW and Bertrand H. **2009**. The *Cryphonectria parasitica* mitochondrial *rns* gene: Plasmid-like

elements, introns and homing endonucleases. Fungal Genetics and Biology, 46: 837-848.

Moore PB, Capel M, Kjeldgaard M and Engelman DM. **1986**. In Structure, Function and Genetics of Ribosomes (eds) Hardesty B, Kramer G, Springer, New York, pp. 87−100

Mota EM and Collins RA. **1988**. Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. Nature, 332: 654–656.

Moure CM, Gimble FS and Quiocho FA. **2002**. Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. Natural Structural Biology, 9(10): 764-770.

Moure CM, Gimble FS and Quiocho FA. **2003**. The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. Journal of Molecular Biology, 334: 685-695.

Mueller MW, Allmaier M, Eskes R and Schweyen RJ. **1993**. Transposition of group II intron aI1 in yeast and invasion of mitochondrial genes at new locations. Nature, 366(6451): 174-176.

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

Mullineux ST, Willows K and Hausner G. 2011. Evolutionary dynamics of the mS952 intron, a novel mitochondrial group II intron encoding a LAGLIDADG homing endonuclease gene. Journal of Molecular Evolution, 72: 433-449.

Mullineux T and Hausner G. **2009**. Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*. Fungal Genetics and Biology, 46: 855-867.

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

Muscarella DE, Ellison EL, Ruoff BM and Vogt VM. **1990**. Characterization of I-Ppo, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. Molecular and Cellular Biology, 10: 3386-3396.

Nakamura TM and Cech TR. **1998**. Reversing time: origin of telomerase. Cell, 92: 587-590.

Newman A. **1994**. Small nuclear RNAs and pre-mRNA splicing. Current Opinion in Cell Biology, 3: 360-367.

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

Nikoh N and Fukatsu T. **2001**. Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. Molecular Biology and Evolution, 18: 1631-1642.

Nikolcheva T and Woodson SA. **1997**. Association of a group I intron with its splice junction in 50S ribosomes: Implications for intron toxicity. RNA, 3: 1-12.

Niu Y, Tenney K, Li H and Gimble FS. **2008**. Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. Journal of Molecular Biology, 382(1): 188-202.

Noah JW and Lambowitz AM. **2003**. Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. Biochemistry, 42: 12466e12480.

Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov A M, et al. **1981**. Secondary structure model for 23 S ribosomal RNA. Nucl. Acids Research, 9: 6167–6189.

Noller HF and Woese CR. **1981**. Secondary structure of 16S ribosomal RNA. Science, 212(4493): 403-411.

O'Brien EA, Zhang Y, Wang E, Marie V, Badejoko W, Lang, BF and Burger G. **2009**. Nucleic Acids Research, 37: D946-950.

Okada G, Jacobs K, Kirisits T, Louis-Seize GW, Seifert KA, Sugita T, Takematsu A and Wingfield MJ. **2000**. Epitypification of *Graphium penicillioides* Corda, with comments on the phylogeny and taxonomy of *Graphium*-like synnematous fungi. Studies in Mycology, 45: 169-188.

Olchowecki A and Reid J. **1974**. Taxonomy of the genus *Ceratocystis* in Manitoba. Canadian Journal of Botany, 52: 1675-1711.

Osiewacz HD and Esser K. **1984**. The mitochondrial plasmid of *Podospora anserina*: a mobile intron of a mitochondrial gene. Current Genetics, 8: 299-305.

Page RDM. **1996**. TREEVIEW: an application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences, 12: 357–358.

Palmer JD. **1990**. Contrasting modes and tempos of genome evolution in land plant organelles. Trends in Genetics, 6: 115-120.

Palmer JD and Logsdon JM Jr. **1991**. The recent origins of introns. Current Opinion in Genetics & Development, 1: 470–477.

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

Paquin B, Laforest MJ, Forget L, Roewer I, Wang Z, Longcore J and Lang BF. **1997**. The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression. Current Genetics, 31: 380-395.

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

Pattanayak V, Ramirez CL, Joung JK and Liu DR. **2011**. Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection. Nature Methods, 8(9): 765-770.

Paukstelis PJ, Chen JH, Chase E, Lambowitz AM and Golden BL. **2008**. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. Nature, 451: 94-97.

Perez EE, Wang J, Miller JC, Jouvenot Y, Kim KA, Liu O, Wang N, Lee G, Bartsevich VV, Lee YL, Guschin DY, Rupniewski I, Waite AJ, Carpenito C, Carroll RG, Orange JS, Urnov FD, Rebar EJ, Ando D, Gregory PD, Riley JL, Holmes MC

and June CH. **2008**. Establishment of HIV-1 resistance in $CD4^+$ T cells by genome editing using zinc-finger nucleases. Nature Biotechnology, 26: 808-816.

Perotto S, Nepote-Fus P, Saletta L, Bandi C and Young JP. **2000**. A diverse population of introns in the nuclear ribosomal genes of ericoid mycorrhizal fungi includes elements with sequence similarity to endonuclease-coding genes. Molecular Biology and Evolution, 17: 44-59.

Pessach IM and Notarangelo LD. **2011**. Gene therapy for primary immunodeficiencies: looking ahead, toward gene correction. Journal of Allergy and Clinical Immunology, 127(6): 1344-1350.

Petersen K, Schöttler MA, Karcher D, Thiele W and Bock R. **2011**. Elimination of a group II intron from a plastid gene causes a mutant phenotype. Nucleic Acids Research, 39:5181-5192.

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

Plattner A, Kim J-J, Reid J**,** Hausner G, Woon YL, Yamaoka Y and Breuil C. **2009**. Resolving taxonomic and phylogenetic incongruence within the species *Ceratocystiopsis minuta*. Mycologia**,** 101: 878-887.

Polduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring, RB and Stoddard BL. **2003**. Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing cofator. Genes and Development, 17: 2875-2888.

Posey KL and Gimble FS. **2002**. Insertion of a reversible redox switch into a rare-cutting DNA endonuclease. Biochemistry, 41(7): 2184-2190.

Prieto J, Molina R and Montoya G. **2012**. Molecular scissors for in situ cellular repair. Critical Reviews in Biochemistry and Molecular Biology, doi:10.3109/10409238.2011.652358.

Przybycien TM, Dunn JP, Valax P and Georgiou G. **1994**. Secondary structure characterization of beta-lactamase inclusion bodies. Protein Engineering, 7:131-136.

Pyatkov KI, Arkhipova IR, Malkova NV, Finnegan DJ and Evgen'ev MB. **2004**. Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. Proceedings of the National Academy of Sciences of the USA, 101: 14719-14724.

Pyle AM. **2010**. The tertiary structure of group II introns: implications for biological function and evolution. Critical Reviews in Biochemistry and Molecular Biology, 45: 215-232.

Pyle AM, Fedorova O and Waldsich C. **2007**. Folding of group II introns: a model system for large, multidomain RNAs? Trends in Biochemical Sciences, 32:138-145.

Pyle AM and Lambowitz AM. **2006**. Group II introns: ribozymes that splice RNA and invade DNA. In Gesteland R, Cech TR, Atkins JF (eds) The RNA World, 3rd ed, Cold Spring Harbor: Cold Spring Harbor Press. pp. 469-505.

Qin PZ and Pyle AM. **1998**. The architectural organization and mechanistic function of group II intron structural elements. Current Opinion in Structural Biology, 8:301-308.

Raghavan R and Minnick MF. **2009**. Group I introns and inteins: disparate origins but convergent parasitic strategies. Journal of Bacteriology, 191: 6193-6202.

Reid J and Hausner G. **2010**. The epitypification of *Ophiostoma minutum*, now *Ceratocystiopsis minuta.* Mycotaxon, 113: 463-474.

Robart AR and Zimmerly S. **2005**. Group II intron retroelements: function and diversity. Cytogenetics and Genome Research, 110: 589-597.

Robert F, Gagnon M, Sans D, Michnick S and Brakier-Gingras L. **2000**. Mapping of the RNA recognition site of Escherichia coli ribosomal protein S7. RNA, 6, 1649-1659.

Roberts RJ and Macelis D. **1997**. REBASE-restriction enzymes and methylases. Nucleic Acids Research, 25(1): 248-262.

Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SKh, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumport RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Krüger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG and Xu SY. **2003**. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Research, 31(7):1805-1812.

Rocheleau GA and Woodson SA. **1995**. Enhanced self-splicing of *Physarum polycephalum* intron 3 by a second group I intron. RNA, 1: 183-193.

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

Ronquist F. **2004**. Bayesian inference of character evolution. Trends in Ecology and Evolution, 19: 475-481.

Ronquist F and Huelsenbeck JP. **2003**. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Rudski S and Hausner G. **2012**. The mtDNA rps3 locus has been invaded by a group I intron in some species of *Grosmannia*. Mycoscience, DOI10.1007/s10267-012-0183-2.

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

Saldanha R, Mohr G, Belfort M and Lambowitz AM. **1993**. Group I and group II introns. The Federation of American Societies for Experimental Biology Journal, 7: 15-24.

Salman V, Amann R, Shub DA and Schulz-Vogt HN. **2012**. Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. Proceedings of the National Academy of Sciences USA, 109(11): 4203-4208.

Salvo JL, Rodeghier B, Rubin A and Troischt T, **1998**. Optional introns in mitochondrial DNA of *Podospora anserina* are the primary source of observed size polymorphisms. Fungal Genetics and Biology, 23: 162-168.

Sander JD, Zaback P, Joung JK, Voytas DF and Dobbs D. **2007**. Zinc Finger Targeter (ZiFiT): An Engineered Zinc Finger/Target Site Design Tool. Nucleic Acids Research, 35: W599-W605.

Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A and Collingwood TN. **2008**. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proceedings of the National Academy of Sciences of the USA, 105(15): 5809-5814.

Saunders S, Cooke B, McColl K, Shine R and Peacock T. **2010**. Modern approaches for the biological control of vertebrate pests: An Australian perspective. Biological control, 52: 288-295.

Saves I, Ozanne V, Dietrich J and Masson JM. **2000**. Inteins of Thermococcus fumicolans DNA polymerase are endonucleases with distinct enzymatic behaviors. The Journal of Biological Chemistry, 275(4): 2335-2341.

Schäfer B. **2003**. Genetic conservation versus variability in mitochondria: the architecture of the mitochondrial genome in the petite-negative yeast *Schizosaccharomyces pombe*. Current Genetics, 43: 311-326.

Schleifman EB, Chin JY and Glazer PM. **2008**. Triplex-mediated gene modification. Methods in Molecular Biology, 435: 175-190.

Seligman LM, Chisholm KM, Chevalier BS, Chadsey MS, Edwards ST, Savage JH and Veillet AL. **2002**. Mutations altering the cleavage specificity of a homing endonuclease. Nucleic Acids Research, 30(17): 3870-3879.

Selker EU. **2002**. Repeat-induced gene silencing in fungi. Advances in Genetics, 46: 439-450.

Sellem CH and Belcour L. **1997**. Intron open reading frames as mobile elements and evolution of a group I intron. Molecular Biology and Evolution, 14: 518-526.

Sethuraman J, Majer A, Friedrich NC, Edgell DR and 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. Molecular Biology and Evolution, 26: 2299-2315.

Sethuraman J, Majer A, Iranpour M and Hausner G. **2009b**. Molecular evolution of the mtDNA encoded *rps3* gene among filamentous ascomycetes fungi with an emphasis on the ophiostomatoid fungi. Journal of Molecular Evolution**,** 69: 372-385.

Sethuraman J, Okoli VC, Majer A, Corkery T and Hausner G. **2008**. The sporadic occurrence of group I intron-like element in the mtDNA *rnl* gene of *Ophiostoma novo-ulmi* subspecies *americana.* Mycological Research, 112: 564-582.

Sharma M, Ellis RL, Hinto and DM. **1992**. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. Proceedings of the National Academy of Sciences USA, 89(14): 6658-6662.

Shen BW, Landthaler M, Shub DA and Stoddard BL. **2004**. DNA binding and cleavage by the H-N-H homing endonuclease I-HmuI. Journal of Molecular Biology, 342(1): 43-56.

Shnyreva AV. **1995**. Mitochondrial introns of fungi and their role in evolution. Genetika, 31(7): 869-876.

Shub DA, Gott JM, Xu MQ, Lang BF, Michel F, Tomaschewski J, Pedersen-Lane J and Belfort M. **1988**. Structural conservation among three homologous introns of bacteriophage T4 and the group I introns of eukaryotes. Proceedings of the National Academy of Sciences USA, 85:1151-1155.

Shukla GC and Padgett RA. **2002**. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. Molecular Cell, 9: 1145-1150.

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

Silanskas A, Zaremba M, Sasnauskas G and Siksnys V. **2012**. Catalytic activity control of restriction endonuclease - triplex forming oligonucleotide conjugates. Bioconjugate Chemistry, 23(2): 203–211.

Silva G, Poirot L, Galetto R, Smith J, Montoya G, Duchateau P and Pâques F. **2011**. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. Current gene therapy, 11(1): 11-27.

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

Silva GH, Dalgaard JZ, Belfort M and Van Roey PV. **1999**. Crystal structure of the thermostable archaeal intron-encoded endonuclease I-DmoI. Journal of Molecular Biology, 286, 1123-1136.

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

Simon DM, Hummel CL, Sheely SL, Bhattacharya D. **2005a**. Heterogeneity of intron presence or absence in rDNA genes of lichen species *Physcia aipolia* and *P. stellaris*. Current Genetics, 47: 389-399.

Simon DM, Moline J, Helms G, Friedl T and Bhattacharya D. **2005b**. Divergent histories of rRNA group I introns in the lichen family Physciaceae. Journal of Molecular Evolution, 60: 434-446.

Simon DM, Kelchner SA and Zimmerly S. **2009**. A broad scale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. Molecular Biology and Evolution, 26: 2795-2808.

Simossis VA and Heringa J. **2005**. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and structure information, Nucleic Acids Research, 33: W289-W294.

Six DL and Wingfield MJ. **2011**. The role of phytopathogenicity in bark beetle-fungus symbioses: A challenge to the classic paradigm. Annual Review of Entomology, 56: 255-272.

Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, Chames P, Prieto J, Redondo P, Blanco FJ, Bravo J, Montoya G, Pâques F and Duchateau P. **2006**. A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Research, 34(22): e149.

Spatafora JW and Blackwell M. **1994**. Polyphyletic origins of ophiostomatoid fungi. Mycological Research, 98: 1-9.

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

Steiner M, Karunatilaka KS, Sigel RK and Rueda D. **2008**. Single molecule studies of group II intron ribozymes. Proceedings of the National Academy of Sciences USA, 105:13853-13858.

Steuer S, Pingoud V, Pingoud A and Wende W. **2004**. Chimeras of the homing endonuclease PI-SceI and the homologous *Candida tropicalis* intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. Chembiochem, 5(2): 206-213.

Stoddard B and Belfort M. **2010**. Social networking between mobile introns and their host genes. Molecular Microbiology, 78(1): 1-4.

Stoddard BL. **2006**. Homing endonuclease structure and function. Quarterly Review of Biophysics, 38: 49-95.

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

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

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

Suh S-O, Kurtzman CP and Blackwell M. **2001**. The status of *Endomyces scopularum*--a filamentous fungus and two yeasts. Mycologia, 93: 317-322.

Suh S-Q, Jones KG and Blackwell M. **1999**. A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetes fungus: Evidence for a new major class of group I introns. Journal of Molecular Evolution, 48: 493-500.

Symington LS and Gautier J. **2011**. Double-strand break end resection and repair choice. Annual Review of Genetics, 45: 247-271.

Szczepek M, Brondani V, Buchel J, Serrano L, Segal DJ and Cathomen T. **2007**. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. Nature Biotechnology, 25: 786-793.

Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, Gloor GB, Scharenberg AM, Edgell DR and Stoddard BL. **2011**. Tapping natural reservoirs of homing

endonucleases for targeted gene modification. Proceedings of the National Academy of Sciences USA, 108(32): 13077-13082.

Tao H, Liu W, Simmons BN, Harris HK, Cox TC and Massiah MA. **2010**. Purifying natively folded proteins from inclusion bodies using sarkosyl, Triton X-100, and CHAPS. Biotechniques, 48(1):61-4.

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F and Higgins DG. **1997**. The Clustal-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 24: 4876-4882.

Thompson LD and Daniels CJ. **1990**. Recognition of exon-intron boundaries by the *Halobacterium volcanii* tRNA intron endonuclease. Journal of Biological Chemistry, 265(30): 18104-18111.

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

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

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

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

Toor N, Keating KS and Pyle AM. **2009**. Structural insights into RNA splicing. Current Opinion in Structural Biology, 19: 260-266.

Toro N, Molina-Sanchez M and Fernandez-Lopez M. **2002**. Identification and characterization of bacterial class E group II introns. Gene, 299: 245-250.

Turmel M, Mercier JP and Cote MJ. **1993**. Group I introns interrupt the chloroplast *psa*B and *psb*C and the mitochondrial *rrn*L gene in Chlamydomonas. Nucleic Acids Research, 21: 5242-250.

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

Upadhyay HP. **1993**. Classification of the ophiostomatoid fungi. In: Wingfield MJ, Seifert KA, Webber JA (eds). *Ceratocystis* and *Ophiostoma*: taxonomy, ecology and pathogenicity. APS Press, St. Paul, Minnesota. pp. 7-13.

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD and Holmes MC. **2005**. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature, 435(7042): 646-651.

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

Vallès Y, Halanych KM and Boore JL. **2008**. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS ONE, 3: e1488.

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

Van Roey P, Meehan L, Kowalski JC, Belfort M and Derbyshire V. **2002**. Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. Nature Structural and Molecular Biology, 9: 806-811.

Van Roey P, Waddling CA, Fox KM, Belfort M and Derbyshire V. **2001**. Interwined structure of the DNA-binding domain of intron endonuclese I-*Tev*I with its substrate. European Molecular Biology Organization Journal, 20: 3631-3637.

Vogel J and Börner T. **2002**. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. Nature Structural & Molecular Biology, 21: 3794-380.

Wang H, Liu X, Liu S, Yu Y, Lin J, Lin J, Pang X and Zhao, J. **2012**. Development of a markerless gene replacement system in *Acidithiobacillus ferrooxidans* and construction of a *pfk*B mutant. Applied and Environmental Microbiology, 78(6): 1826-18235.

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

Williams RJ. **2003**. Restriction endonucleases: classification, properties, and applications. Molecular Biotechnology, 23 (3): 225-243.

Wilson GG. **1988**. Cloned restriction-modification systems--a review. Gene, 74(1): 281-289.

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

Windbichler N, Papathanos PA, Catteruccia F, Ranson H, Burt A and Crisanti A. **2007**. Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. Nucleic Acids Research, 35: 5922-5933.

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

Wingfield MJ, Seifert KA and Webber JF. **1993**. *Ceratocystis* and *Ophiostoma*. Biology, Taxonomy and Ecology. American Phytopathological Society Press, St. Paul.

Woese CR, Kandler O and Wheelis ML. **1990**. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences USA, 87(12): 4576-4579.

Woodson SA. **2005**. Structure and assembly of group I introns. Current Opinion in Structural Biology, 15: 324-330.

Woodson SA and Cech TR. **1989**. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell, 57: 335-345.

Woodson SA and Cech TR. **1991**. Alternative secondary structures in the 59 exon affect both forward and reverse self-splicing of the *Tetrahymena* intervening sequence RNA. Biochemistry, 30:2042-2050.

Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, Eichtinger M, Fu F, Porteus MH, Dobbs D, Voytas DF and Joung J.K. **2006**. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nature Protocols, 1: 1637-1652.

Xiong Y and Eickbush T. **1990**. Origin and evolution of retroelements based upon their reverse transcriptase sequences. European Molecular Biology Organization Journal, 9: 3353-3362.

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

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

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

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. European Molecular Biology Organization Journal, 26(9):2432-42.

Zhong J and Lambowitz AM. **2003**. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. European Molecular Biology Organization Journal, 22:4555-4565.

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

Zimmerly S, Guo H, Perlman PS and Lambowitz AM. **1995a**. Group II intron mobility occurs by target DNA-primed reverse transcription. Cell, 82: 545-554.

Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS and Lambowitz AM. **1995b**. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell, 83: 529-538.

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

Zink P and Fengel D, **1989**. Studies on the colouring matter of bluestain fungi. Part 2 Electron microscopic observations of the hyphae walls. Holzforschung 43: 371-374.

Zink P and Fengel D. **1990**. Studies on the colouring matter of bluestain fungi. Part 3. Spectroscopic studies on fungal and synthetic melanins. Holzforschung, 44: 163-168.

Zipfel RD, de Beer ZW, Jacobs K, Wingfield BD and Wingfield MJ. **2006**. Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. Studies in. Mycology, 55**:** 75-97.

Zuker M. **2003**. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31: 3406-3415.