Exploring the mtDNA *rnl* and *nad4* genes in *Ophiostoma* species for novel introns and homing endonucleases

by

Chen Shen

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillmentof the requirements for the Degree of

Master of Science

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba

Abstract

Fungal mitochondria are variable in size due to the presence of potential mobile elements such as group I and group II introns and homing endonuclease genes (HEGs). In this work the mitochondrial large ribosomal subunit gene (mt-*rnl*) of *Ophiostoma ulmi* and related species have been screened for the presence of introns and intron encoded proteins. Five introns have been noted in different regions of the *rnl* gene of O. *ulmi* and related species. Based on this *rnl* survey and *rnl* data from Genbank, an *rnl* intron landscape for ascomycetous and basidiomyctous fungi was generated by using bioinformatic based analysis. A total number of 23 possible intron insertion sites were found in the *rnl* gene of ascomycetous and basidiomycetous fungi. The results also indicate that regions of the *rnl* gene are more prone to intron invasion then others. The second project dealt with the evolution of mitochondrial ribosomal protein S5 (*rps3*) gene within the filamentous ascomycetes fungi. Within members of this group of fungi the *rps3* gene typically is a component of the mL2449 group I intron but there are free-standing forms of *rps3*. The study examined if these free standing forms evolved only once due to an as of yet unknown recombination event or if the *rps3* gene was transferred from the mL2449 intron to a new mtDNA locus several times during the evolution of the filamentous ascomycetes fungi. The third project was to sequence and characterize the intron and HEG found in the mitochondrial NADH dehydrogenase 4 (*nad4*) gene of an undescribed species of *Pesotum.* A 1.4 kb group IC2 intron has been identified in the *nad4* gene of *Pesotum* strain WIN (M)1630. Overall the three studies demonstrate the invasive nature of introns and their associated ORFs and the potential of these introns to influence gene structure and size variation among the fungal mtDNAs.

Acknowledgements

I would like to thank my supervisor, Dr. G. Hausner, for his support, encouragement, and eminent guidance. I could not have completed this project without him. I attribute the level of this thesis to his immense knowledge, research insight, and expertise.

I wish to thank all my committee members, Dr. D. Court and Dr. M. Piercey-Normore who have reviewed my work in past, and for their valuable comments and suggestions.

I am also grateful to all my colleagues: Tuhin, Iman, Alvan, and Mohamed, for all their technical support and friendship.

A special thanks to my parents, my boyfriend and my other friends for being there with me when I really need them.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BLAST: Basic Local Alignment Search Tool

bp: base pair

DED: Dutch Elm Disease

EtBr: Ethidium Bromide

HEase: Homing Endonuclease

HEG: Homing Endonuclease Gene

IEP: Intron-Encoded Protein

LHE: LAGLIDADG Homing Endonuclease

LSU : Large Subunit

LSU rRNA: Large Subunit ribosomal RNA

mRNA: messenger RNA

mt: mitochondrial

mtDNA: mitochondrial DNA

NADH: *Nicotinamide Adenine Dinucleotide* plus Hydrogen flavin adenine dinucleotide

nad4: NADH dehydrogenase subunit 4

NCBI: National Center for Biotechnology Information

nt: nucleotide

ORF: Open Reading Frame

PCR: Polymerase Chain Reaction

rnl: mitochondrial large subunit ribosomal RNA gene

rns**:** mitochondrial small subunit ribosomal RNA gene

rps3: ribosomal protein S3 gene

Rps3: ribosomal protein S3

rRNA: ribosomal RNA

RT: Reverse Transcriptase

RT-PCR: Reverse Transcription Polymerase Chain Reaction

SOC: Super Optimal broth with Catabolite repression

SSU: Small Subunit

SSU rRNA: Small Subunit ribosomal RNA

tRNA: transfer RNA

Chapter 1: Literature review

1.1 Ophiostomatoid fungi

Fungi, which have an extracellular digestion system, are distinct from animal, plants, and bacteria, and they play an important part in the ecosystem. Some fungal species are of economic importance, such as blue stain or sap stain fungi that cause cosmetic discoloration of timber. Blue stain fungi include species that are found in the following genera: *Ophiostoma, Ceratocystiopsis*, *Ceratocystis, Grosmannia* and in their asexual counter parts such as *Leptographium,* and *Pesotum* (reviewed in Hausner et al., 2005; Zipfel et al., 2006). The activity of these species can lead to minor wood degradation, and although usually they do not seem to alter the structural integrity of the wood, the aesthetic problem of "brown- blue" staining reduces the value of the wood (lumber/timber), thus causing significant economic losses (Seifert 1993). The genus *Ophiostoma,* which belong to the order Ophiostomatales, includes both potential plant pathogens, and many blue-stain fungi (Upadhay 1993).

During the first part of the 20th century, the numbers of elm trees in North America and Europe have been reduced due to the devastating Dutch elm disease (DED) (Brasier 1995). DED is caused by the ascomycete microfungus *Ophiostoma ulmi,* and this fungus is spread by bark beetles. The bark beetles bore into an infected tree, pick up fungal spores then move to another tree, bore into an uninfected tree and thus transfer the fungal spores. Many species of blue stain fungi appear to have a specific association with certain bark beetle species and these plus the fungi usually are associated with certain plant host species (Linnakoski et al., 2009).Three fungal species have been recognized that can cause DED: *Ophiostoma ulmi,* which was the cause of the

first pandemic of DED outbreak in North America and Europe during the early 20th century. Since the 1960s the more aggressive and virulent *Ophiostoma novo-ulmi* (includes two subspecies: *O. novo-ulmi* subspecies *novo-ulmi* and *O. novo-ulmi* subspecies *americana*), is the major cause of DED in both Europe and North America. *Ophiostoma himal-ulmi*, a third but overall less well studied species, is endemic to the western Himalaya region (Brasier 1995, Spooner 2005).

Ophiostomatoid fungi share many morphological similarities (reviewed in Upadhyay 1993; Hausner et al., 1993; Zipfel et al., 2006); therefore taxonomic discrepancies to resolve species and generic status have been settled with molecular data [i.e. internal transcribed spacer region (ITS) between the nuclear 18S and 26S ribosomal (r) DNA or nuclear ribosomal small and large subunit sequences]. *Ophiostoma*, *Grosmannia*, and *Ceratocystiopsis* all belong to the order Ophiostomatales, whereas *Ceratocystis* has been assigned to the order Microascales based on molecular data (Hausner et al., 1992; Hausner et al., 1993; Guarro et al., 1999; Marin et al., 2005; Zipfel et al., 2006; Seifert et al., 2013). Part of the taxonomic confusion is due to convergent evolution of sexual reproductive structures and various asexual modes of spore production. The reproductive stages: anamorphic state (asexual or mitotic) and teleomorphic state (sexual or meiotic) in the past have been assigned different taxonomic stastus (i.e. names).

The rules that govern the naming of the fungi have recently been changed as a new nomenclature system has been released. According to Hawksworth (2011), one key principle of the 2011 IBC (International Botanical Congress) in Melbourne meeting was the proposal that the nomenclature of algae, fungi and plants (ICD) will not use separate names for anamorphs and teleomorphs. i.e. “one fungus one name”; so *Pesotum* spp should now be referred to as *Ophiostoma* spp.

In this thesis, species of the *O. ulmi* species complex (*O. ulmi, O. novo- ulmi, O. himal- ulmi*) and strains of *Pesotum* will be examined with regards to intron distribution among the mitochondrial large subunit gene (*rnl*). *Pesotum spp.,* which also belong to the order Ophiostomatales, are restricted to asexual states of *Ophiostoma* species (Okada et al., 2000) or species that are phylogenetically linked to *Ophiostoma* (i.e. related based on molecular data). And as stated above strains of *Pesotum* can now be referred to as *Ophiostoma* spp.; however for this thesis, the more traditional anamorphic and teleomorphic generic designations will be used.

1.2 Fungal mitochondrial genomes

Mitochondria are membrane-bound organelles found within most eukaryotic organisms, with the primary purpose to generate energy in the form of adenosine triphosphate (ATP). Mitochondria are also involved in other tasks such as signaling, cellular differentiation, cell death, as well as the control of the cell cycle and cell growth (McBride et al., 2006). Mitochondrial DNA (mtDNA) is the genetic material located inside the mitochondria and this genome is comprised of a variety of genes that usually encode; (1) rRNAs genes; the small subunit ribosomal RNA (mt SSU rRNA also known as *rns*) and the large subunit ribosomal RNA (mt LSU rRNA also known as *rnl*) gene, (2) transfer RNAS (tRNAs), (3) proteins involved in the electron transport chain, (4) subunits of the Nicotinamide Adenine Dinucleotide plus Hydrogen flavin adenine dinucleotide (NADH) dehydrogenase complex, (5) protein components of ATP synthase, and in some fungi, genes encoding ribosomal proteins such as ribosomal protein 3 (rps3; reviewed in Bullerwell et al., 2003a; Hausner 2003).

Fungal mtDNAs can range in size from 18.9 kilobase pair (kbp) in *Schizosaccharomyces*

pombe to 175 kbp in *Agaricus bisporus* (Hintz et al., 1985; Hausner 2003). Size variation can exist within and between species, and this is mostly due to non-coding regions, like intergenic spacers, and the presence or absence of mobile elements such as group I and group II introns and homing endonuclease (HE) genes (HEGs) (reviewed in Hausner 2003). Mitochondrial DNA variability is sometimes used for estimating genetic diversity within and among fungal species (Gardes et al., 1991; Hausner et al., 1993).

Recently, systematic sequencing of fungal mitochondrial genomes has generated mtDNA sequences from representatives of the four principle divisions of fungi: Ascomycota, Basidiomycota, Zygomycota and Chytridiomycota (Bullerwell et al., 2003a). In filamentous ascomycete fungi, mitochondrial inheritance can be uniparental and sometimes biparental. Biparental mitochondrial inheritance has been observed in Basidiomycetes fungi and also in some yeast (Xu 2005). In Chytridiomycetes, the mtDNA appears to be uniparentally inherited (Borkhardt and Olson, 1983).

1.3 Mitochondrial ribosomal proteins

Mitochondrial genes can be usually classified into three major categories: rRNA, tRNA and protein coding genes (Adams and Palmer 2003). The presence of mitochondrial genes coding for ribosomal proteins shows some variation among species. The ribosome is comprised of a large and complex molecular machinery that is composed of the small and large subunit RNAs, and serves as the primary site for translation where mRNA is translated into protein. The small subunit of the ribosome helps in the interactions between mRNAs and tRNAs (Wool et al., 1995).

So far three mitochondrial ribosomal proteins have been described among the fungi: Var1, S5 and Rps3. The relationship between these genes is that they appear to encode proteins with some sequence similarities, especially in the Carboxy-terminal region of these proteins; this suggests these proteins are probably homologs (Bullerwell et al., 2000). The Var1 (variable locus 1) gene was first noted in the yeast mtDNA and the VAR1 protein was noted in the mitochondrial ribosomes (Groot et al., 1979; Terpstra et al., 1979). In the same year (1979), the S5 ribosomal protein was described in *N. crassa* (Lambowitz et al., 1979; reviewed in Bullerwell et al., 2003a). The name rps3 (ribosomal protein S3) was not used untill 1996, it was mentioned during the characterization of the mitochondrial DNA of the Chytridiomycete fungi, *A. macrogynus* by Paquin and Lang (1996).

The *rps3* homologs from the fungi have diverged considerably from the rps3 sequences found in prokaryotes, eukaryotic nuclei, and plant chloroplast genomes (Sethuraman et al., 2009a). It has been observed that the rps3 protein can act as a moon-lighting protein (Wool 1996). In some instances it was shown that it is (1) a structural component of the ribosome, (2) an enzyme that is involved in DNA repair, (3) has DNA endonuclease activity and is involved in repair of oxidative DNA damage in *Drosophila* and in mammals (reviewed in Wool 1996; Bullerwell et al., 2003b; Sethuraman et al., 2009b).

1.4 Rps3 ribosomal protein

The Rps3 ribosomal protein in many Ascomycete fungi is an essential cellular protein (LaPolla and Lambowitz 1981) which can be encoded by a group I intron inserted within the *rnl* gene (reviewed by Hausner 2003). This was first shown in *Neurospora crassa* where a group I

intron inserted within the mtDNA *rnl* gene encodes an ORF for a putative ribosomal protein originally named S5 (Burke and RajBhandary 1982). More recently this intron encoded ORF and the freestanding mtDNA *var1* gene were recognized as *rps3* (ribosomal protein 3) homologues (Bullerwell et al., 2000).

With regards to some filamentous ascomycetes (e.g., *Penicillium chrysogenum*, *Neurospora crassa*, etc) , the *rps3* gene has been found to be encoded within a group I intron (mL2449) that is inserted in the U11 region of the mitochondrial large subunit rDNA (*rnl*) gene at the position that corresponds to *Escherichia coli* LSU position 2449 (Johansen and Haugen 2001; Sethuraman et al., 2009a), and in a few examples these intron encoded *rps3* ORFs have been invaded by homing endonuclease genes (Hausner et al., 1999; Gibb and Hausner 2005; Sethuraman et al., 2009b).

In general it is assumed that during eukaryotic evolution, ribosomal protein coding genes were transferred from the mitochondrial genome to the nuclear genomes. In animals, the mtDNA ribosomal protein genes are all located in the nucleus, while in plants some of ribosomal protein coding genes still remain in mitochondrion (reviewed in Lang, Gary, and Burger 1999; Timmis et al., 2004). The mitochondrial *rps3* locus found in plants has undergone several changes with respect to its intron content. For example in the liverwort, *Marchantia*, no introns are present in the mitochondrial *rps3*, while in most angiosperms *rps3* harbors at least one conserved intron. However, according to Kubo's study (2000), *Beta vulgaris* which belongs to the angiosperms, the mitochondrial *rps3* appears to be intron-free.

1.5 Mobile elements in fungal mitochondrial genomes

Group I and II introns can be detected in fungal mtDNA; however the majority of fungal mtDNA introns belong to the group I intron family (Hausner 2003). Both types of introns are potential ribozymes and are capable of self-splicing (in the absence of proteins) from pre-mRNA transcripts via different mechanism (see Fig 1.1). Self-splicing requires the formation of defined secondary and tertiary RNA catalytic structures (Michel et al., 1982). However, group I and II introns can employ help from either intron or host encoded maturases to stabilize the RNA folds, thus facilitating splicing (reviewed in Lambowitz and Belfort 1993; Hausner 2012). Group I and II introns are sometimes referred to as selfish mobile elements; they contribute to the overall size of the mtDNA genome but seem to offer neither advantages nor disadvantages to the host genome (Chevalier and Stoddard 2001). Group I and group II introns are located within highly conserved regions/sequences such as the *rnl* gene (Sethuraman et al, 2009b).

1.6 Group I introns

Some group I introns have short (8-12 bp) conserved sequence motifs referred to as P-, Q-, R-, and S- sequence motifs. These motifs are involved in the formation of loop and stem regions of the RNA secondary structure. Group I intron can be classified into classes IA-IE (can be further divided into e.g. IA1, IC2, etc.) based on the secondary structure, nucleotide sequences within the conserved core regions, and peculiarities within the secondary structure (Michel and Westhof 1990; Suh et al., 1999; Woodson 2005). Several examples of group I intron structures

have been deposited at the Group I Intron Sequence and Structure Database (GISSD; http://www.rna.whu.edu.cn/gissd/ ; Zhou et al., 2008).

The introns are removed from its host RNA transcript by intron splicing, the splicing process for group I intron involves two consecutive transesterfication reactions. The first transesterfication reaction happens at the 5' splice site where a free guanosine (GTP) cofactor acts as a nucleophile and causes the cleavage of the upstream exon-intron junction; the GTP attaches to the 5' end of the intron. The 5' and 3' splice sites are defined by the base-pairing interactions between the 5' end of the intron and the flanking downstream exon region (reviewed by Hausner 2003). The second reaction occurs at the 3' splice site, where the 3' OH of the upstream exon attacks the 3' exon-intron junction, resulting in a linear, excised intron RNA and the ligated exons (see Fig 1.1a; Michel and Westhof 1990; Hausner 2003).

Group I introns are mobile elements, the mobility of group I introns is catalyzed by a homing endonuclease (HE), which usually is encoded by an open reading frame (ORF) that is embedded within the intron (reviewed in Dujon 1989; Hausner 2003). HEGs can move to other locations via an event called "homing"; here a HEG containing allele will be transferred into the cleaved target site present within in a recipient cognate allele (HEG minus allele). Essentially the HE introduces a double-strand break (DSB) at its target site. The cleaved DNA triggers the host DSB repair system which uses the intron plus HEG containing allele as a template to repair the DSB in the cognate intron-minus allele (Dujon 1989).

The group I introns also can invade DNA sequences by a mechanism called reverse splicing, also known as retrohoming or retrotransposition; here an "RNA- mediated" mobility mechanism is involved. There has been experimental (*in vitro*) support that group I intron RNA can successfully integrate into foreign RNA via reverse splicing (Roman and Woodson 1998; Roman

et al., 1999). The integrated intron RNA plus the "host" transcript have to be reverse transcribed into DNA and inserted into the organellar genome via recombination (reviewed in Hausner 2003). The reverse splicing mechanism might be less efficient but has the advantage of requiring a shorter target site; the intron insertion site would be based on matching sequences that resemble the introns internal guide sequence (IGS) (Saldanha et al., 1993; Roman et al., 1999).

1.7 Group II introns

Group II introns are self-splicing RNAs that are believed to be the ancestors of nuclear pre-mRNA introns or small nuclear (sn)RNAs (Jacquier 1990; Strobel 2013). Group II introns consist of an RNA component and in some cases a reverse transcriptase (RT) ORF. Group II introns also have conserved secondary structures at the RNA level, they can be characterized by the presence of six typical loop-stem domains, also known as domain I to VI. The reverse transcriptase ORF, if present, usually is located in the loop of domain IV (reviewed by Hausner 2003). Group II introns can be assigned into several major families II1, IIB and IIC, etc., based on the similarities in secondary and tertiary structure and their intron encoded proteins (IEPs; Michel and Feral1995; Lambowitz and Zimmerly 2011; Pyle 2010). Detailed information on group II introns can be found at the Mobile Group II Intron Database (http://www.fp.ucalgary.ca/group2introns/; Dai et al., 2003).

Both splicing and mobility activity of group II introns require the catalytic activity of the intron RNA, the intron-encoded protein, and frequently host factors (Zimmerly et al., 1995a; Lambowitz and Zimmerly 2011). The splicing of group II introns also involves two transesterfication reactions (see Fig 1.1 b). The splicing process is initiated by the 2' OH group

on a bulged adenosine residue in domain VI, the next step involves the 3' OH of the 5' exon attacking the phosphodiester bond at the 3' splice site, which results in the ligation of the exons and the release of the intron RNA in a lariat configuration (Bonen and Vogel 2001; Hausner 2003; Lambowitz and Zimmerly 2004; Fedorova and Zingler 2007). The mobility of group II introns is achieved by a process called retrohoming, which involves the formation of a ribonucleoprotein (RNA) particle between the intron lariat and the IEP. The RNP recognizes an intron-less cognate allele and generates a single stranded DNA cleavage (nick) by using the 3' OH of the intron RNA; this initiates a reverse splicing reaction which allows the intron RNA to be inserted into the sense DNA strand. The endonuclease (En) domain of the reverse transcriptase cuts the antisense DNA strand, thus a 3'-OH end is generated that serves as a primer for the reverse transcriptase (RT). Finally the intron RNA will be removed by the host DNA repair mechanism (reviewed in Belfort et al., 2002; Hausner 2003; Zimmerly et al., 1995a, b).

1.8 Homing endonuclease

Homing endonuclease genes (HEGs) encode DNA endonucleases that recognize and cleave at or near rather long target sites (14-44 bp; Chevalier and Stoddard 2001). HEGs are mobile elements, which can be considered as "neutral" elements, similar to group I and group II introns; they can remain neutral, do not interrupt host gene function while "jumping" into a new location. HEGs can be encoded by group I or group II intron, and thus HEGs can promote the mobility for their host introns. Typically HEases introduce double-strand breaks (DBS) in alleles which lack the endonuclease coding (HEG-minus strain) sequence, and this trigger the DBS

repair process that will “move” the HEG into the HEG-minus locus. This mechanism is known as “homing”; and is a process that completed by the host-DSB repair pathway, which simply uses the HEG containing strand as a template to repair the double strand break (see Fig 1.2) i.e. similar to group I intron homing. This process gives HEGs the great potential to manipulate genes, to generate new alleles, and further contribute to the molecular evolution of the mtDNA (Hausner 2003; Poggeler and Kempken 2004; Sethuraman et al. 2009a,b).

HEGs and mobile introns are neutral elements that go through a cycle of intron invasion, followed by ORF degeneration resulting in both ORF and possibly intron loss (Goddard and Burt 1999). During evolution, the HEG-containing intron can become fixed in a population; this results in the loss of function of HEG as there is no selection pressure to maintain the function of the HEG sequence, eventually leading to the deletion of the intron (Hausner 2003). Therefore, the HEGs and mobile introns have to continuously invade cognate alleles or move into new sites (ecotopic integration) or they will be eliminated from a population.

Historically HE proteins have been classified into four major families with the naming based on conserved amino acid motifs; the four major families are as follows: H-N-H, HIS-CYS box, LAGLIDADG, and GIY-YIG families. The HNH family is found in phages, the HIS-CYS box family is found in protists (Kowalski and Derbyshire 2002; Stoddard 2005, Marcaida et al., 2010; Stoddard 2011); the LAGLIDADG and GIY-YIG families have been reported from bacterial and organelle genomes. More recently one additional class of HEase has been described, the PD-(D/E) XK HE protein, these are found in bacterial tRNA group I introns (Stoddard 2005). There have been additional reports of new HEases recovered from phage and bacterial genomes and these are reviewed in Hafez and Hausner (2012).

1.9 The LAGLIDADG Homing endonuclease family

The LAGLIDADG homing endonuclease family is the most diverse HEG family; it has been found mainly in organellar genomes, such as in the mitochondrial genomes of fungi and protozoans, chloroplast genomes of plants and algae, and also within the genomes of Eubacteria and the Archaea. These enzymes are encoded within introns or form a component of inteins (Dalgaard et al., 1997; Haugen and Bhattacharya 2004; Stoddard 2005; Marcaida et al., 2010). Descendants of LAGLIDADG homing endonuclease also include the yeast HO mating type switch endonuclease (Jin et al., 1997). So these genes have in some cases been coopted by their host genomes to perform essential functions. Based on the conservation of a 10-residue sequence motif, the LAGLIDADG HEs also have been referred to as dodecapeptide, dodecamer, decapeptide, and DOD protein family (Dujon 1989; Dojun et al., 1989; Chevalier and Stoddard 2001). Members of LAGLIDADG family can contain either one or two LAGLIDADG motifs in their ORF. The version that has only the single motif, such as I-CreI (Thompson et al., 1992; Wang et al., 1997), can bind to target DNA sequences and need to assemble into homodimers, and they require a palindromic or near- palindromic symmetry in their DNA recognition sites. The LAGLIDADG HEs with double motifs can act as monomers, as they process a pair of structurally similar nuclease domains in a single peptide chain; thus symmetrical recognition sites are not required for the double motifs LAGLIDADG HE (Stoddard 2005). This makes double motif enzymes more efficient in recognizing a wide variety of different target sites and it has been noted that double motif forms have been more successful in invading more sites compared to their single motif ancestors (Haugen et al., 2005).

1.10 The GIY-YIG endonuclease family

The GIY-YIG endonuclease family is the second most numerous HEs family; it has been first identified in intron-encoded ORFs in filamentous fungi and in the bacteriophage T4 (Michel and Dujon 1986; Cummings et al., 1989a). Overall the GIY-YIG ORFs have been found in group I introns in fungal mitochondria, and in the mitochondrial and chloroplast genomes of algae (Chevalier and Stoddard 2001). GIY-YIG HEs only have one motif in their N-terminal region which interacts with the DNA at the cutting site. The best studied GIY-YIG endonuclease is I-TevI, which is encoded by a group I intron found in the thymidylate synthase (*td*) gene of bacteriophage T4 (Edgell et al., 2010). It is a monomeric enzyme, composed of a catalytic domain and a DNA-binding domain; the two parts are connected by a linker region. These endonucleases can recognize rather long DNA binding sites and usually cleave the DNA many base pairs away from the actual intron insertion sites (Chevalier and Stoddard 2001; Stoddard 2005).

1.11 Applications of HEases

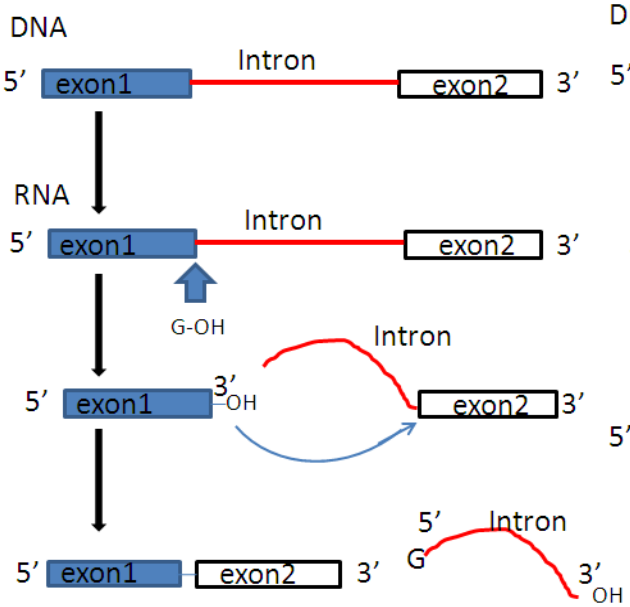
HEases can generate site-specific DSB, and trigger the host DSB repair system to essentially replace/repair a damaged section of DNA with a template allele (if provided). This principle can be used to propose novel strategies for genome editing such as gene therapy or gene replacement (reviewed in Stoddard 2011). When HEases cut at highly specific sites they

require cellular repair mechanisms to correct the damaged DNA. This can be used to the advantage of the scientist trying to develop treatments for monogenic disease, by cutting the defected allele in the presence of a repair template that provides the correct allele, and replacing the defected gene by homologous recombination (Stoddard 2005; Stoddard 2011).

HEases also have great potential in pest control. A recent study has shown that populations of mosquitoes vectoring infectious agents could be controlled by hindering the insect's reproduction ability (Stoddard 2005). Successful experiments have been done by introducing modified HEase into *Drosophila* population; the modified HEase could disrupt its target site (segment on the X chromosome) by its cleavage activity thus controlling the insects population (Windbichler et al., 2007, 2008 and 2011; reviewed in Hafez and Hausner 2012). Due to the special features of the HEases, they can be used as DNA-binding or -cutting reagents for molecular biology. HEases can be modified by various genetic engineering strategies and thus they can be used as potential tools in many biological applications such as genomic engineering, gene replacement/ therapy, pest control, and site-specific mutagenesis (reviewed in Hafez and Hausner 2012).

Group II introns also have great potential in biotechnology. The retrohoming ability of group II introns has been developed into a gene knockout system referred to as targetrons. Targetrons can be engineered and customized for various applications and are currently available for bacterial systems (reviewed in Lambowitz and Zimmerly 2004, 2011).

A). Group I introns



B). Group II introns

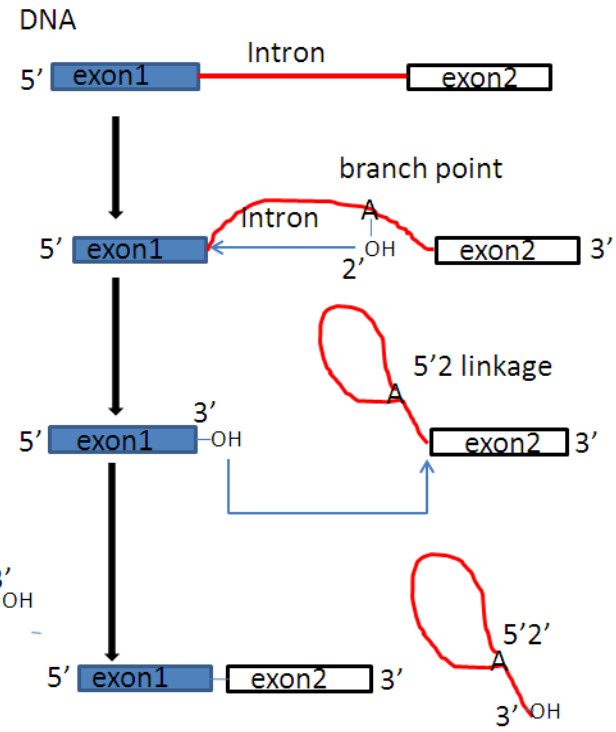


Figure 1.1 Schematic representation for splicing mechanisms for group I (A) and group II introns (B). Red line indicates intron RNA and in the case of group II intron the formation of the lariat. See text for more details.

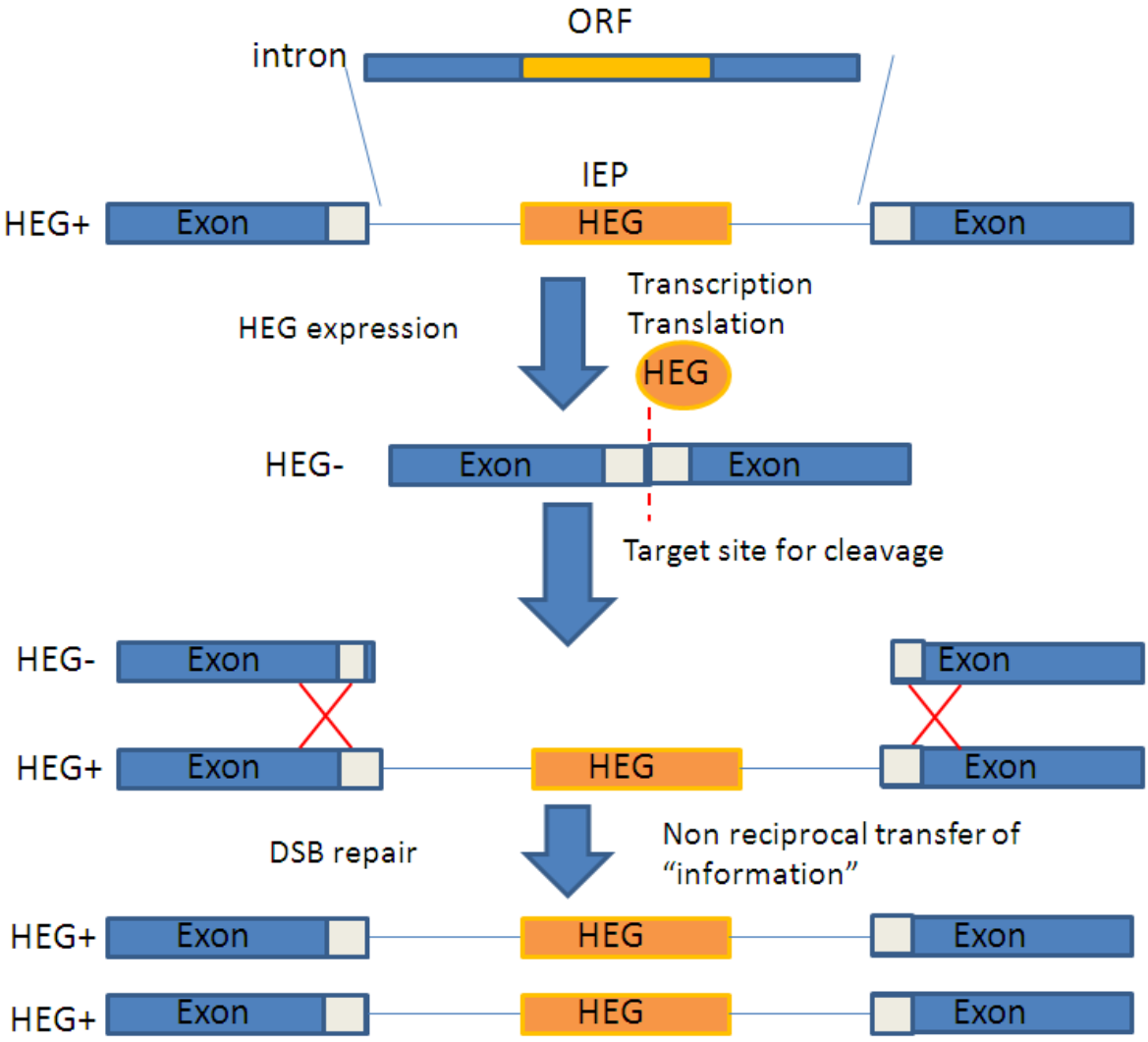

ORF
intron
IEP
HEG+
Exon
HEG
Exon
HEG expression
Transcription
Translation
HEG
HEG-
Exon
Exon
Target site for cleavage
HEG-
Exon
Exon
HEG+
Exon
HEG
Exon
DSB repair
Non reciprocal transfer of
"information"
HEG+
Exon
HEG
Exon
HEG+
Exon
HEG
Exon

Figure 1.2. Homing of the homing endonuclease gene (or intron). After protein expression/production, the HEase will recognize its cleavage target site within the HEG minus allele; this introduces a double strand break in a genome at the cognate allele that lacks the endonuclease coding sequence/ or intron; this activity triggers the double strand repair process, which will essentially "move" the HEG into the HEG minus intron locus (i.e., DSB repair uses the HEG containing strand as template to repair the DSB). Note: HEG plus = HEG containing strand; HEG minus = HEG minus strand; gray box = HEG target site.

Chapter 2: General materials and methods

2.1 Fungal DNA extraction

Fungal cultures were grown in 50 ml liquid PYG medium [Peptone 1 g/l, glucose 3 g/l, and yeast extract 1 g/l] and after 48 to 96 hours fungal mycelia were collected by using the vacuum filtration method. The harvested mycelia were transferred into 15 ml Falcon tubes (Corning Inc., Corning, NY). The mycelia were combined with 3 g of glass beads (1mm, Fisher Scientific) and 3 ml extraction buffer [1 mM EDTA, 5M NaCl, 0.5% SDS, 1% CTAB (Cetyl Trimethyl Ammonium Bromide) and 10 mM Tris-Cl at PH 7.6]. Thereafter the 15 ml Falcon tube was vortexed for about 5 - 15 min; periodically the tubes were placed on ice for a few minutes. Finally, an additional 3 ml of extraction buffer plus 660 µl of 20% SDS were added and the tubes were incubated at 55 °C to 60 °C in a water bath for 2 hours. Six ml chloroform was added after the tubes were cooled to room temperature. The mixture was centrifuged at 2000 rpm for 20 min in order to separate cell debris, glass beads and denatured proteins. In order to recover the nucleic acids the aqueous (i.e. top) layer was transferred to a new 15 ml tube and mixed with 2.5 volumes of ice-cold 95% ethanol. After a short incubation at -20 °C the mixture was centrifuged at 3000 rpm for 30 min at room temperature. The supernatant was discarded and the DNA pellets were washed with 1 ml ice-cold 70% ethanol and the tubes were again centrifuged at 3000 rpm for 30 min. The 70 % ethanol layer was discarded and the tubes were inverted the DNA pellets were dried at room temperature. The DNA pellets were resuspended in 300 µl TE buffer pH 7.6 (1M Tris-CL pH 7.6 and 0.1M EDTA) and stored at -20 °C.

2.2 PCR amplification

All primers used in this project were synthesized by Alpha DNA ™ (Montreal, Quebec) and were supplied in a lyophilized form. DEPC (diethylpyrocarbonate) treated water was used to dissolve these primerS to a final concentration of 400 μM (or 400 picomoles/μl). Further 1/10 and 1/100 dilutions were required in order to make primer stocks needed for PCR amplification and DNA sequencing, respectively. Names and sequences of primers used during this study are listed in Table 2.0.

The "One Taq Hot start" DNA Polymerase system [New England BioLabs Inc. 25 μl reactions; One Taq standard reaction buffer (5×), 10 mM dNTPs, 10 μM forward primer, 10 μM reverse primer, template DNA and One Taq DNA Polymerase in nuclease-free water] was used to amplify the large ribosomal subunit (*rnl*) region. PCR amplification reactions for the NADH dehydrogenase subunit 4 (*nad4*) gene and components of the *rnl* gene (U= universally conserved: U7, U11, and U7U11 inter regions) were made up to a total volume of 50 μl. For the PCR reaction the actual components are listed in Table 2.1

Usually for the *rnl* regions studied the PCR conditions were as follows: with the thermal cycler lid temperature set at 105 °C, an initial denaturation at 94 °C for 2 min was followed by 30 cycles of denaturing (1 min at 93 °C), annealing (1 min at 50 °C), and extension (3 min at 72°C). The samples were incubated for a final extension at 72 °C for 5 min, then tubes were incubated "on hold" set at 10 °C.

However, two thermocycler programs were used for investigating the absence or presence of introns in the mtDNA *rnl* U7, U11 and U7U11 inter-regions. For the *rnl* U7 region, the PCR conditions were as follows: an initial denaturation at 94 °C for 2 min was followed by 25 cycles of denaturing (1 min at 93 °C), annealing (1 min 30 sec at 53.1 °C), and extension (4

min 30 sec at 70 °C).The samples were incubated for a final extension at 70 °C for 10 min, then placed "on hold" at 10 °C. For the *rnl* U11 region and rnl U7U11 inter-region, the PCR conditions were as follows: initial denaturation at 94 °C for 2 min was followed by 25 cycles of denaturing (1 min at 93 °C), annealing (1 min 30 sec at 52.9 °C), and extension (4 min 30 sec at 70 °C).The PCR samples were incubated for a final extension at 70 °C for 10 min, then placed "on hold" at 10 °C.

PCR amplification condition for the *nad4* gene was as follows: with the lid temperature at 105 °C, an initial denaturation at 94 °C for 2 min was followed by 25 cycles of denaturing (1 min at 93 °C), annealing (1 min 30 sec at 52.1 °C), and extension (4 min 30 sec at 70°C).The sample were incubated for a final extension at 72 °C for 5 min, then placed "on hold" at 10 °C.

PCR products were separated in 1% agarose submarine gels in 1×TBE buffer under electrophoresis at 100V for an hour. DNA fragments were sized by using the 1-kb DNA ladder (Invitrogen). Agarose gels were stained in Ethidium Bromide (EtBr 0.5 μg/ml) and later were exposed to ultraviolet (UV) light to observe the size of the DNA bands.

2.3 Cloning of PCR product

Some of the PCR fragments were cloned in order to optimize DNA sequencing conditions by using the TOPO TA Cloning Kit for Sequencing (Invitrogen). PCR conditions were modified with the addition of an extra 10 minute extension at 72 °C to enhance the synthesis of the 3' A-overhang. The ligation reaction consisted of the PCR product being incubated along with the TOPO vector in the salt solution (Invitrogen, 200 mM NaCl, 10mM $MgCl_2$) at room temperature for 5 minutes. The manufacturer (Invitrogen) supplied One Shot Chemical

Transformation protocol was applied in order to transform the cloned PCR fragments into the *E. coli* (DH5α –T1) cell line. Conditions for transformation were as follows: The ligation reaction mix was added into a vial that contained 50 µl DH5α –T1 chemically competent cells. The tube was then incubated on ice for 30 min; thereafter the cells were heat-shocked for 1 minute at 42°C. Cells were transferred back onto the ice and 250 µl of S.O.C (Super Optimal broth with Catabolite repression) medium [S.O.C; 0.5% yeast extract, 2% tryptone, 2.5 mM KCl, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose] was added. The cells were allowed to recover at 37°C with agitation for 1 hour. Screening for potential recombinants was done by spreading 20 µl of culture on Luria Bertani (LB; 10 g tryptone, 5 g yeast extract, 10 g NaCl and 20 g bacteriological agar per litre) plates containing 60 µg/ml ampicillin and 20 mg/ml bromo-chloro-indolyl-galactopyranoside (X-gal). The plates were incubated overnight at 37 °C.

Two types of colonies developed on the LB plates; the white colonies were the successfully transformed ones. White colonies were selected and transferred to new LB plates supplemented with ampicillin and X-gal and incubated at 37 °C overnight. Newly grown single colonies were used to inoculate 5 ml LB broth tubes supplemented with ampicillin and incubated with agitation at 37°C overnight. Cells were harvest by centrifugation of 1.5 ml of the overnight culture and plasmids were purified using the Wizard *Plus* SV Minipreps DNA Purification System (Promega) following the manufacturer's protocols.

2.4 Bacterial Glyceral stocks for long term storage at -80 °C

Bacterial colonies were transferred to a tube containing 5 ml LB supplemented with 100 µl/ml ampicillin and incubated overnight at 37 °C with shaking. Aliquots of 50 µl of O/N (overnight) culture were then taken and added to 250 µl of LB supplemented with 36% glycerol.

These bacterial glycerol stocks were transferred to -60 °C (or -80 °C) for long term storage.

2.5 DNA Sequencing

For DNA sequencing, PCR products ware purified using the Wizard SV PCR clean-up kit (Promega, California, USA) in accordance with the manufacturer's instructions. Initially DNA sequences were generated using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Austin, TX). Each sequencing reaction consisted to 10.5 µl sdH_2O; 0.75µl cleaned PCR product, 0.75 µl 4µM primer, 1.5µl BigDye V3.1 buffer, and 1.5 µl BigDye V3.1 reaction mix (fluorescent dye-labeled dideoxy NTPs) making a final volume of 15 µl. The cycle sequencing program utilized was as suggested by the Applied Biosystems BigDye Terminator V3.1 manual (Foster City, California, USA). The Cycle sequencing PCR products were purified as to the following protocol: 4.2 µl of 125 µM EDTA and 53 µl of 95% ethanol were added to each cycle sequencing reaction in order to precipitate the DNA. The tubes were inverted several times to mix the ingredients and incubated at room temperature for 15 mins. After the incubation, the reaction mixtures were centrifuged at 4 °C for 15 mins. The resulting DNA pellets were washed with 60 µl of 70% of ethanol (70% wash). The DNA pellets were dried at room temperature until all the ethanol was evaporated. Formamide (15 µl) was added to dissolve the DNA pellets and then the DNA was denatured at 94 °C for 5 min. The resuspended DNA pellets were transferred to a 96-well microtitre type sequencing plate and the plate was placed onto the loading platform in the ABI tm Prism 3130 Genetic Analyzer. The machine was operated according to the protocols provided by the manufacturer. In addition, when a large number of samples were on hand they were sent to the DNA Technologies Unit, NRC, Saskatoon,

Saskatchewan for cycle sequencing. In all cases the results, chromatograms, were examined in CHROMAS (http://chromas.software.informer.com/2.4/) and the sequences were compared in the GeneDoc program version 2.7 (Nicholas et al. 1997). The online CAP 3 (http://pbil.univ-lyon1.fr/cap3.php) program was used for building contigs.

2.6 Sequence manipulations, phylogenetic analysis and RNA folding

A data set comprised of mtDNA ribosomal protein 3 (rps3) amino acid sequences was compiled by conducting blastq queries (Altschul et al. 1990; http://blast.ncbi.nlm.nih.gov/Blast.cgi) and by examining the organellar (mtDNA) NCBI data base. In same instances "raw contigs" deposited in NCBI were annotated with the MFannot program (Beck and Lang 2010; http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl). Amino acid sequences extracted from NCBI were formatted into the fasta format and aligned initially with the ClustalX program (Thompson and Gibson 1997). The MAFFT program was used to refine the alignment (Katoh et al., 2002; http://mafft.cbrc.jp/alignment/server/), and the data was also aligned with the online program PRALINE (Simossis and Heringa 2005). Finally the amino acid alignment was further refined with the Gendoc program (Nicholas et al. 1997).

For the characterization of the *O. novo-ulmi nad4* gene the ORF Finder program (http://www.ncbi.nlm.nih.gov/projects/gorf/) was used to find potential open reading frames (ORFs) (genetic code 4: mtDNA). Nucleotide sequences obtained during this study were also compared to those in the NCBI database by using the online program BLASTn (Altschul et al. 1990; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Online program CAP3 (Huang and Madan 1999;

http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::cap3) was used to assemble *nad4* contig.

For the *rnl* sequences and its various introns the online RNAweasel program (http://megasun.bch.umontreal.ca/RNAweasel/) was used to identify the potential type/category of intron. The online RNA folding program mFOLD (Zuker 2003; http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) combined with output from the RNAweasel program and the goal to optimize H-bonding were used to generate a secondary RNA structure for a group I intron. In addition RNA folds previously published by Burke et al., (1987), Cech et al., (1994), and Michel and Westhof (1990) were examined as a reference.

For the aligned rps3 data set phylogenetic trees were generated with a variety of programs. Phylogenetic trees were drawn with the neighbor joining, parsimony, and maximum likelihood options as implemented in MEGA5: Molecular Evolutionary Genetics Analysis program (Tamura et al., 2011), and with the MrBayes program v3.1 (Ronquist and Huelsenbeck 2003; Ronquist 2004). Phylogenetic trees were visualized with the treeview program (Page 1996) and the trees were annotated with Corel Draw (v14.0; Corel Corporation Limited, Ottawa).

Table 2.0 Oligonucleotides (primers) utilized during this study

Location	**Amplified region**	**Primer name**	**Primer sequence (5' to 3')**
Mitochondrial DNA	Large ribosomal subunit (rnl)	Ou-rnl-F1	GGGGAACCTTCCTCAAAGAC
		Ou-rnl-R4	CATCGGATATCTTTGCAAC
		Ou-rnl-F2	GTTGCAAAGATATCCGATG
		Ou-rnl-R3	CTTGTTCTGGGCTGTTTCCC
		Ou-rnl-F3	GGGAAACAGCCCAGAACAAG
		Ou-rnl-R2	CACCATAAGCTTTAGGCGA
		Ou-rnl-F4	CGCGTAAGCGGGGCATTAGATC
		Ou-rnl-R1	GCTTCATACTTATATGCCTTCAG
		Ou-rnl-F1a	GAGATTCTATKATTAGCGWGA
		Ou-rnl-R1a	GAAAACCAGCTAATACAGTCAG
		Ou-rnl-F1b	GTACCGTGAGGGAARARMC
		Ou-rnl-R1b	GCCTACTRTTATAGGTTTCGCAG
		Ou-rnl-F1'	GGATAATCCGCAGCCAAGTTCCA
		Ou-rnl-R1'	CAAATACCTATGTGGGTAGTTCAG
		Ou-rnl-F2'	GGTGTATGTTTATACCCTAATAGTG
		Ou-rnl-R2'	AGTAGTCGCCTGTCGTATTAAGTC
	Rnl-U7	LSUex1	GCTAGTAGAGAATACGAAGGC
		LSUex2	GACCGCATTTAACGGCCAAGG
	Rnl-U11	IP1	GGAAAAGCTACGCTAGGG
		IP2	CTTGCGCAAATTAGCC

	Rnl-U7U11	LSUex2R IP1-R	CCTTGGCCGTTAAATGCGGTC CCCTAGCGTAGCTTTTCC
Mitochondrial DNA	NADH Dehydrogenase subunit 4 (ND4)	ND4-F ND4-R 1630PND4-F1 1630PND4-R1 1630PND4-F2 1630PND4-R2 1630ND4Fa 1630ND4Ra 1630ND4Fb 1630ND4Rb 1630PND4-F1a 1630PND4-R1a 1630PND4-Fb1 1630PND4-Rb1 ND4OctF1 ND4OctF2 ND4OctR	GAAAGTATATTACCDCCTTTA GGTGHACCACAATTACCTAAA GTCTACATATACCGTATAAAC GTTGTACCCATAATTGAAGAC CAGATATGTGGGTGATCGCGG GTTCTGTATCTCATGCTGCAG CCTTCTTCAGTTAAATGTTC CATATTGTTCAACACAGTAGAG GGGATATCAAATCCTAGT TCCAAGACATTGTTG GTAGGCCTTCTTCAGTTAAATG CCGTCTCCACTAGTTAAACC GTGGTAGAGTAGAACTAA GGCTCACGTTGAATCTCC CTACTCTACCACAATTTAAAG GTAGGCCTTCTTCAG GTTTCCCAGCGACTAGAC

Table 2.1 General components present within the PCR reactions

Component	Volume (µl)
Sd H_2O	37.25
200 µM dNTP	4.0
10× PCR reaction buffer	5.0
50 mM $MgCl_2$	1.5
Primer 1	0.5
Primer 2	0.5
Taq polymerase	0.25
Genomic DNA	1.0

Chapter 3: The occurrence of potential mobile elements in the mtDNA *rnl* gene in the Ophiostomatales

3.1 Introduction

The gene coding for the large subunit ribosomal RNA is a highly conserved gene that is present in all domains of life and in the chloroplast and mitochondrial organelles. However with regards to comparative DNA sequence analysis, information about this gene is limited to the large subunit RNA having a large number of highly variable sites and indels (insertions and deletions) which make these sequences more difficult to align. With regard to the fungi it has been suggested that characterizing mitochondrial DNAs (mtDNAs) might be a possible solution for studying phylogenetic relationships among fungi. This is in part due to the small size of the mtDNA and the relatively high percentage of coding sequences including the rRNA coding sequences (Bullerwell et al., 2003b). However, in recent years it has become apparent that the mtDNAs is highly variable in size and gene arrangements, in part; these arise due to the presence and absence of mobile introns and their associated ORFs (Cummings et al., 1990; Charter et al., 1996; Youssar et al., 2013; Joardar et al., 2012).

Recent studies on the ophiostomatoid fungi such as species of *Ceratocystis* and *Ophiostoma* indicated that the mtDNA *rnl* gene might be structurally quite complex due to the presence of introns (Gibb and Hausner 2005; Sethuraman et al., 2008; 2009a,b; 2013). Previous studies focused on partial sequences that investigated the U7 (U= universal) and U11 regions of the *rnl* gene; herein the focus is on the entire *rnl* gene of the mtDNA *rnl* gene of the *Ophiostoma ulmi* species complex. These species are economically important as they are the causative agents of Dutch Elm Disease, a wilt disease that has devastated many urban forests in North America.

Typically the *rnl* gene would be expected to be around 3 kb (Cannone et al., 2002) but from preliminary data and access to the mtDNA genome for a strain of *Ophiostoma novo-ulmi* subspecies *novo-ulmi* (courtesy of Dr. Ken Dewar, McGill University, Department of Human Genetics and Genome Quebec Innovation Centre; Forgetta et al., 2013) it was suspected that among *O. ulmi* strains and related species the *rnl* gene could range in size from 4kb to 8kb and maybe even above 10 kb. Therefore, the strategy was to study this gene "piece by piece" using different PCR primer sets (Fig 3.1) in order to cover and eventually amplify the entire *rnl* gene.

The presence of potential mobile elements (introns) in the mtDNA *rnl* gene of strains collected across Europe and North America of *O. ulmi s.l* was investigated through a PCR survey. It was hoped that this survey would allow for estimating the relative frequencies of some of the *rnl* introns and to see if potential patterns emerge with regards to intron distribution among the strains examined from various geographic locations. In addition mtDNA introns are of interest as they have been associated with mitochondrial diseases (Abu-Amero et al., 1995; Baidyaroy et al., 2011; Monteiro-Vitorello et al., 2009; Dujon and Belcour 1989) and it has been suggested that mitochondrial genome instabilities may influence the fitness of plant pathogens (Bertrand, 2000).

This study was extended by collecting *rnl* sequences from other ascomycetous fungi from Genbank. Various bioinformatics tools were applied in order to align these sequences and to align the dataset against the *E. coli rnl* homolog reference sequence (Johansen and Haugen 2001; Cannone et al., 2002). The *E. coli* sequence serves as a guide for naming introns (see Johansen and Haugen 2001) as mtDNA rDNA introns are named according to their insertion site with respect to the *E. coli* rLSU nucleotide sequence position. This *rnl* survey generated an intron landscape and this study also demonstrated that the *rnl* gene appears to be a reservoir for a many group I and group II introns along with their encoded ORFs. The purpose of this study is twofold:

(1) as an aid in annotating *rnl* genes in fungal mtDNA sequencing projects; and (2) to provide a map for potential intron insertion sites in *rnl* gene among ascomycetous and some selected basidiomycetous fungi. The latter is potentially useful for those that are bioprospecting for ribozymes and homing endonucleases (Hafez et al., 2014).

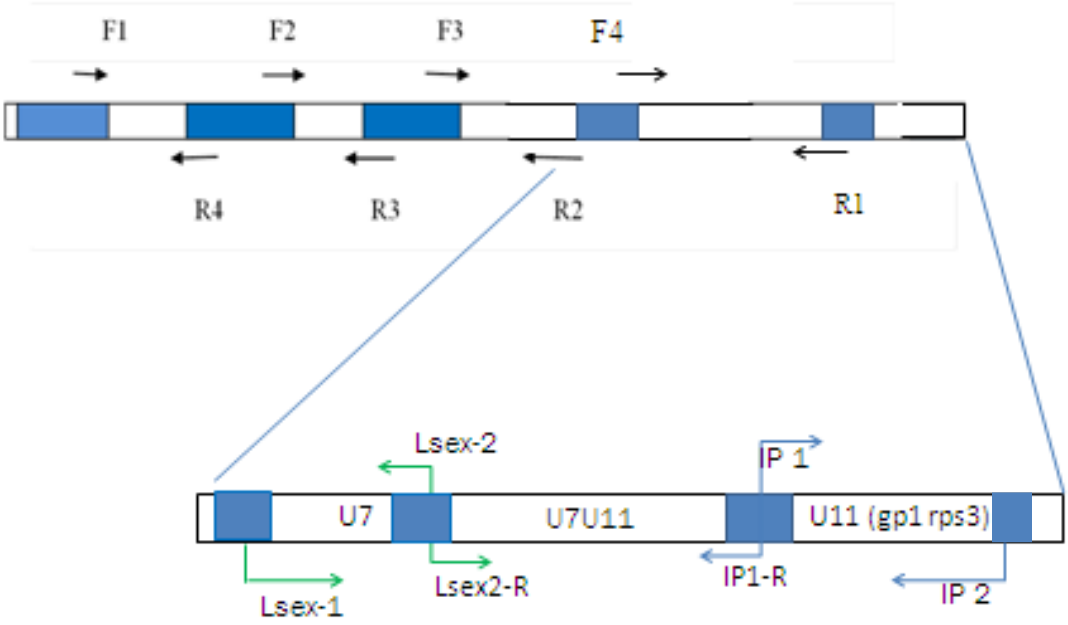

F1
F2
F3
F4
R4
R3
R2
R1
Lsex-2
IP 1
U7
U7U11
U11 (gp1 rps3)
Lsex-1
Lsex2-R
IP1-R
IP 2

Figure 3.1 Overview of the *rnl* gene and relative positions of PCR primers utilized to amplify sections of the gene in this study. Exon (blue boxes) and intron (white boxes) predictions based on *rnl* gene sequences extracted from GenBank and based on the mtDNA sequence provided by Dr. Ken Dewar. Intron 1: amplified by primers F1R4; Intron 2: amplified by primers F2R3; Intron 3: amplified by primers F3R2; Introns 4 and 5: region of *rnl* referred to U7, U7U11, and U11, were amplified with the Lsex1 and Lsex2 and IP1 and IP2 primer sets (see Sethuraman et al., 2008; 2009a,b).

3.2 Material and Methods

Strains used in this study

Some strains were obtained from the DAOM culture collection (National Mycological Herbarium (DAOM) and the Canadian Collection of Fungal Cultures (CCFC); Ottawa, ON, Canada). DED strains from the Thunder Bay (ON, Canada) region were provided by Dr. L.J. Hutchison (Lakehead University) in 2006; strains from Saskatchewan (Canada) were collected during 2004 and 2005 during Saskatchewan DED surveys, and Manitoba Department of the Environment provided infected branches collected during the annual 2002, 2003, 2004, and 2005 DED surveys by the Manitoba Conservation Authority. Fungi were recovered from infected DE branch material as previously described in Sethuraman et al., (2008). Living strains of the materials analyzed are maintained in the University of Manitoba (Winnipeg) Collection = WIN(M).

DNA extraction and PCR amplifications of *rnl* segments

The Fungal DNA extraction and DNA purification protocols and the PCR procedures were as described in Chapter 2 (page 20). Primers used in the *rnl* PCR survey are listed in Chapter 2, Table 2.1. For selected strains PCR products were sequenced and those protocols are described in Chapter 2.

Bioinformatics

A set of mtDNA *rnl* genes was collected from the online database Genbank (http://www.ncbi.nlm.nih.gov/genbank/) and initially the sequences were compiled in the GENEDOC program (Nicholas et al. 1997). In addition a "seed" 23S rRNA alignment was obtained from "The comparative RNA web (CRW) site" (Cannone et al., 2002; http://www.rna.ccbb.utexas.edu/DAT/3C/Alignment/; Table 1: Three domain alignment for 23 S rRNA). The MAFFT program (http://mafft.cbrc.jp/alignment/server/; Kuraku et al., 2013) was used to align the *rnl* sequence against the *E. coli* rDNA LSU sequence (or 23 S); the following setting was used E-INS-i (Very slow; recommended for <200 sequences with multiple conserved domains and long gaps). The online programs RNAweasal and mFOLD were used to predict RNA secondary structures for selected introns from *Ophiostoma* species. Detailed descriptions of methods are provided in Chapter 2 (page 24).

3.3 Results and Discussion

PCR survey for mtDNA *rnl* introns in strains of ophiostomatoid fungi

The *rnl* gene sequence for *Ophiostoma novo-ulmi* subsp. *novo-ulmi rnl* gene was extracted from Genbank (accession number AMZD01000161; Forgetta et al., 2013) supplied to us in 2010 by Ken Dewar and based on this sequence (template sequence) primers were designed (table 2.0).

Based on the naming system of Johansen and Haugen (2001) the following introns were

noted in the *O. novo-ulmi* subsp. *novo-ulmi rnl* gene: mL740 mL963, mL1092 and mL2449. PCR amplicons derived from *O. ulmi* species complex whole cell DNAs generated by amplifying the mL740, mL963 and mL1092 introns with primers (see Fig. 3.1) located in the flanking exon regions resulted in 1.5, 1.9 and 1.8 kb DNA fragments, respectively. The 3' end of the *rnl* gene for members of the *O. ulmi* species complex have been previously characterized by Gibb and Hausner (2005), and Sethuraman et al. (2008) and all members of this species complex contain the mL2449 intron that encodes the rps3 ORF. In addition some members of this species complex have an intron inserted within the *rnl*-U7 region (mL1669) that encodes a double-motif LAGLIDADG type ORF (Sethuraman et al., 2008).

O. novo-ulmi subsp. *novo-ulmi rnl*i1 or mL740

A total of 152 strains represented *O. ulmi*, subspecies of *O. novo-ulmi* were examined with regards to detecting the presence of introns encountered in previous studies and introns noted in the *O. novo-ulmi* subsp. *novo-ulmi rnl* gene (Table 3.1). The PCR assay noted that 100 strains tested positive for rnli1.

Bioinformatic analyses were applied to the *O. novo-ulmi* subsp. *novo-ulmi rnl* gene. The rnli1 corresponding to *E. coli* LSU rDNA position 740; the ORF finder program showed the presence of a short putative ORF of 471 bp. The online program RNAweasel shows that rnli1 is a group I C1 type intron. The predicted ORF encodes a 156 amino acid polypeptide that contains a conserved copy of the GIY-YIG endonuclease motif, which suggests that this ORF is a potential HEG (Belfort and Roberts, 1997). A BLASTp database search using the 156 amino acid sequence as a query showed similarities to putative GIY-YIG type homing endonuclease-

like amino acid sequences such as the one reported in *rnl* gene of *Cordyceps brongniartii* (YP_002213592; Ghikas et al., 2010); a related form of the GIY-YIG type ORF has also been found in *atp6* gene of *Ceratocystis cacaofunesta* (YP_007507082) and in the *rnl* gene of *Peltigera malacea* (YP_005351165).

Strains representing *Ophiostoma himal-ulmi, Ophiostoma minus*, and *Ophiostoma piceae* were also included in this mtDNA *rnl* intron survey for comparative purposes. Only one strain of *Ophiostoma piceae* (CBS 263.35) yielded a positive PCR signal for the presence of an intron at mL740.

Table 3.1 List of strains and summary of PCR survey results with regards to insertions within the mtDNA *rnl* gene

No.	* Strain	Species designation	** rnl-U7 ml1669	rnl-U11 ml2449	Intron1 mL740	Intron2 mL963	Intron3 mL1092	Location
1	WIN(M)894	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Richot, Manitoba, Canada
2	WIN(M)895	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Morris, Manitoba, Canada
3	WIN(M)897	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Gretna, Manitoba, Canada
4	WIN(M)898	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Virden, Manitoba, Canada
5	WIN(M)899	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Reverton, Manitoba, Canada
6	WIN(M)900	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Manitou, Manitoba, Canada
7	WIN(M)901	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Morden, Manitoba, Canada
8	WIN(M)902	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Carman, Manitoba, Canada
9	WIN(M)903	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Winnipeg, Manitoba, Canada
10	IMI 343.101	*O. novo-ulmi* subsp. *novo-ulmi*	S	L	+	_	+	Zabok, Front Croatia
11	DAOM171044	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Sault St.Marie, Ontario, Canada
12	DAOM171047	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Ames, Iowa, USA
13	DAOM171048	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Clear Lake, Iowa, USA
14	DAOM171063	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Denbighshire, England
15	DAOM171064	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Thame, Oxfordshire, England
16	DAOM171068	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Bennekom, Netherlands
17	DAOM171033	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Toronto, Ontario, Canada

18	DAOM171034	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Toronto, Ontario, Canada
19	DAOM171035	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Toronto, Ontario, Canada
20	DAOM171036	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	Toronto, Ontario, Canada
21	DAOM171037	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Toronto, Ontario, Canada
22	DAOM171038	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Toronto, Ontario, Canada
23	DAOM171039	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Toronto, Ontario, Canada
24	DAOM171041	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Sault St.Marie, Ontario, Canada
25	DAOM171042	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Sault St.Marie, Ontario, Canada
26	DAOM171045	*O. ulmi*	L	L	+	+	_	Mal-Hal, Quebec, Canada
27	DAOM171046	*O. ulmi*	L	L	+	+	_	Westmount, Quebec, Canada
28	DAOM171051	*O. ulmi*	L	L	+	_	_	Tennessee, USA
29	DAOM171053	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Ames, Iowa, USA
30	DAOM171054	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	North Hampton, Massachusetts, USA
31	DAOM171056	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	Blackburg, Virginia, USA
32	DAOM171058	*O. novo-ulmi* subsp. *americana*	L	S	+	_	_	Tewkesbury, Glocester, England
33	DAOM171059	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Herfordshire, England
34	DAOM171060	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Basildon, Essex, England
35	DAOM171061	*O. ulmi*	L	L	+	_	_	Devon, England

36	DAOM171066	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Astra near Caspian Sea, Iran
37	DAOM171069	*O. ulmi*	L	L	+	_	_	Amsterdam, Netherlands
38	DAOM171070	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Sloten, Friesland, Netherlands
39	DAOM171071	*O. ulmi*	L	L	+	+	_	Baarn, Netherlands
40	DAOM171079	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	Madison, Wisconsin, USA
41	DAOM194898	*O. novo-ulmi* subsp. *americana*	L	L	+	_	_	Saskatoon, Saskatchewa, Canada
42	DED 02-1	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Manitou, Manitoba, Canada
43	DED 02-3	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Manitou, Manitoba, Canada
44	DED 02-5	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Manitou, Manitoba, Canada
45	DED 02-6	*O. novo-ulmi* subsp. *americana*	S	L	+	_	_	Carman, Manitoba, Canada
46	DED 02-7	*O. novo-ulmi* subsp. *americana*	S	L	+	_	_	Steinbach, Manitoba, Canada
47	DED 02-8	*O. novo-ulmi* subsp. *americana*	L	S	+	_	_	Gimli, Manitoba, Canada
48	DED 02-9	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Carman, Manitoba, Canada
49	DED 02-11	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Carman, Manitoba, Canada
50	DED 02-12	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Steinbach, Manitoba, Canada
51	DED 02-13	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Carman, Manitoba, Canada
52	DED 02-16	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Arborg, Manitoba, Canada
53	DED 02-17	*O. novo-ulmi* subsp. *americana*	S	L	+	_	_	Carman, Manitoba, Canada
54	DED 02-18	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Arborg, Manitoba, Canada

55	DED 02-19	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Gimli, Manitoba, Canada
56	DED 02-21	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Morden, Manitoba, Canada
57	DED 02-22	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Trehern, Manitoba, Canada
58	DED 02-23	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Morden, Manitoba, Canada
59	DED 02-24	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Morden, Morden Research Centre, Manitoba, Canada
60	DED 02-26	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Carman, Manitoba, Canada
61	DED 02-28	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Steinbach, Manitoba, Canada
62	DED 02-29	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Carman, Manitoba, Canada
63	DED 02-32	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Morden, Manitoba, Canada
64	DED 02-33	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Trehern, Manitoba, Canada
65	DED 02-34	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Steinbach, Manitoba, Canada
66	DED 02-35	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Steinbach, Manitoba, Canada
67	DED 02-36	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Trehern, Manitoba, Canada
68	DED 02-37	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Teulon, Manitoba, Canada
69	DED 02-38	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Arborg, Manitoba, Canada
70	DED 02-39	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Carman, Manitoba, Canada
71	DED 02-41	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Gimli, Manitoba, Canada

72	DED 02-42	*O. novo-ulmi* subsp. *americana*	S	S	–	–	–	Teulon, Manitoba, Canada
73	DED 02-45	*O. novo-ulmi* subsp. *americana*	S	S	–	–	–	Altona, Manitoba, Canada
74	DED 02-46	*O. novo-ulmi* subsp. *americana*	S	S	–	–	–	Stonewall, Manitoba, Canada
75	DED 02-50	*O. novo-ulmi* subsp. *americana*	S	S	–	–	–	Selkirk, Manitoba, Canada
76	DED 02-55	*O. novo-ulmi* subsp. *americana*	S	S	–	–	–	Portage la Prairie, Manitoba, Canada
77	DED 02-61	*O. novo-ulmi* subsp. *americana*	S	S	+	–	–	St. Celements, Manitoba, Canada
78	DED 02-62	*O. novo-ulmi* subsp. *americana*	S	S	–	+	–	Gretna, Manitoba, Canada
79	DED 02-63	*O. novo-ulmi* subsp. *americana*	S	S	+	+	–	Gretna, Manitoba, Canada
80	DED 02-64	*O. novo-ulmi* subsp. *americana*	S	S	+	–	–	Mountain Avenue, Manitoba, Canada
81	DED 02-69	*O. novo-ulmi* subsp. *americana*	S	S	+	–	–	Winkler, Manitoba, Canada
82	DED 02-71	*O. novo-ulmi* subsp. *americana*	L	S	–	+	–	La Prairie Island, Manitoba, Canada
83	DED 02-72	*O. novo-ulmi* subsp. *americana*	S	S	+	+	–	Winkler, Manitoba, Canada
84	DED 02-73	*O. novo-ulmi* subsp. *americana*	S	S	+	+	–	Winkler, Manitoba, Canada
85	DED 02-74	*O. novo-ulmi* subsp. *americana*	S	S	–	+	–	Selkirk, Manitoba, Canada
86	DED 02-76	*O. novo-ulmi* subsp. *americana*	L	S	–	+	–	Stonewall, Manitoba, Canada
87	DED 03-86	*O. novo-ulmi* subsp. *americana*	S	S	+	–	–	PTH 305, Manitoba, Canada
88	DED 03-88	*O. novo-ulmi* subsp. *americana*	L	S	–	–	–	Cartwright, Manitoba, Canada

89	DED 03-89	*O. novo-ulmi* subsp. *americana*	L	S	_	+	_	La Riviere, Manitoba, Canada
90	DED 03-90	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Snowflake, Manitoba, Canada
91	DED 03-91	*O. novo-ulmi* subsp. *americana*	L	S	+	_	_	Baldur, Manitoba, Canada
92	DED 03-93	*O. novo-ulmi* subsp. *americana*	L	S	+	_	_	Altamont, Manitoba, Canada
93	DED 03-96	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Baldur, Manitoba, Canada
94	DED 03-100	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Plum Coulee, Manitoba, Canada
95	DED 03-102	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	Swan lake, Manitoba, Canada
96	DED 03-104	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Crystal City, Manitoba, Canada
97	DED 03-106	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	Swan lake, Manitoba, Canada
98	DED 03-108	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Holland, Manitoba, Canada
99	DED 03-113	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Miami, Manitoba, Canada
100	DED 03-114	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Lowe Farm, Manitoba, Canada
101	DED 03-115	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Roland, Manitoba, Canada
102	DED 03-116	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Wawansa, Manitoba, Canada
103	DED 03-117	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	St.Claude, Manitoba, Canada
104	DED 03-119	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Kentville, Kings Co, Nova Scotia, Canada
105	DED 03-120	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	Salmon River, Nova Scotia, Canada

106	DED 04-21	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Moose Jaw, Saskatchewan, Canada
107	DED 04-81	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	Gainsborough ,Saskatchewan, Canada
108	DED 04-82	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Regina, Saskatchewan, Canada
109	DED 04-95	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	Estevan, Saskatchewan, Canada
110	DED 04-102	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Katepwa, Saskatchewan, Canada
111	DED 04-137	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Pasqua Lake, Saskatchewan, Canada
112	DED 04-138	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Sun Valley, Saskatchewan, Canada
113	DED 04-141	*O. novo-ulmi* subsp. *americana*	L	L	_	+	_	Indian Head, Saskatchewan, Canada
114	DED 04-150	*O. novo-ulmi* subsp. *americana*	S	L	_	_	_	Codette, Saskatchewan, Canada
115	DED 04-167	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Alida, Saskatchewan, Canada
116	DED 04-202	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Buffalo Pound Lake, Saskatchewan, Canada
117	DED 04-257	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Moosomin, Saskatchewan, Canada
118	DED 04-258	*O. novo-ulmi* subsp. *americana*	L	L	_	_	_	Carnduff, Saskatchewan, Canada
119	DED 04-298	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Traux, Saskatchewan, Canada
120	DED 04-338	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Carnduff, Saskatchewan, Canada
121	DED 05-01	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Cavan, Manitoba, Canada
122	DED 05-02	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	The Pas, Manitoba, Canada

123	DED 05-03	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	The Pas, Manitoba, Canada
124	DED 05-04	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	The Pas, Manitoba, Canada
125	DED 05-05	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Red Deer River, Manitoba, Canada
126	DED 05-06	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	Armit River, Manitoba, Canada
127	DED 05-10	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	Teulon, Manitoba, Canada
128	DED 05-16	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Carlyle, Saskatchewa, Canada
129	DED 05-60	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Carduff, Saskatchewa, Canada
130	DED 05-61	*O. novo-ulmi* subsp. *americana*	NA	S	+	+	_	Carduff, Saskatchewa, Canada
131	DED 05-63	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Carduff, Saskatchewa, Canada
132	DED 05-64	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Carduff, Saskatchewa, Canada
133	DED 05-85	*O. novo-ulmi* subsp. *americana*	L	S	+	+	_	Wawata, Saskatchewa, Canada
134	DED 05-153	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Avonlea, Saskatchewa, Canada
135	DED 05-164	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Tisdale,Saskatchewan, Canada
136	DED 05-165	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Tisdale, Saskatchewa, Canada
137	DED 05-195	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Nipawin, Saskatchewa, Canada
138	DED 05-216	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Garrick, Saskatchewa, Canada
139	DED 05-14	*O. novo-ulmi* subsp. *americana*	S	S	+		_	Minnedosa, Manitoba, Canada
140	DED 05-15	*O. novo-ulmi* subsp. *americana*	S	S	+	+	_	Gladstone, Manitoba, Canada

141	DED 05-16	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	Dauphin, Manitoba, Canada
142	DED 05-17	*O. novo-ulmi* subsp. *americana*	S	S	_	+	_	St. Rose, Manitoba, Canada
143	DED 05-20	*O. novo-ulmi* subsp. *americana*	S	S	+	_	_	River, Manitoba, Canada
144	DED 05-21	*O. novo-ulmi* subsp. *americana*	S	L	+	+	_	Neepawa, Manitoba, Canada
145	DED 06-01	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Winnipeg, Manitoba, Canada
146	DED 06-02	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Winnipeg, Manitoba, Canada
147	DED 06-03	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Winnipeg, Manitoba, Canada
148	DED 06-05	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Thunder Bay, Ontario, Canada
149	DED 06-06	*O. novo-ulmi* subsp. *americana*	S	L	_	_	_	Thunder Bay, Ontario, Canada
150	DED 06-07	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Thunder Bay, Ontario, Canada
151	DED 06-08	*O. novo-ulmi* subsp. *americana*	S	L	_	_	_	Thunder Bay, Ontario, Canada
152	DED 06-09	*O. novo-ulmi* subsp. *americana*	L	S	_	_	_	Thunder Bay, Ontario, Canada
153	DED 06-10	*O. novo-ulmi* subsp. *americana*	S	S	_	_	_	Thunder Bay, Ontario, Canada
154	WIN(M)1223	*O. ulmi*	L	S	+	_	_	Crediton, Devon, UK
155	WIN(M)1224	*O. ulmi*	S	S	_	_	_	StRomuald, Quebec, Canada
156	WIN(M)1225	*O. novo-ulmi* subsp. *americana*	L	L	+	+	_	Tewkesbury, Gloucestershire, UK
157	WIN(M)1226	*O. novo-ulmi* subsp. *americana*	S	L	+	_	+	Crediton, Devon, UK
158	WIN(M)292	*O. minus*	S	S	_	_	_	Sandilands, Forest Reserve, Manitoba, Canada

159	WIN(M)371	*O. minus*	S	L	_	_	_	Forest Besrve, Manitoba, Canada
160	WIN(M)460	*O. pseudominus*	S	S	_	_	+	Lake Cowichan, Vancouver Island, BC, Canada
161	WIN(M)461	*O. pseudominus*	S	S	_	_	+	Lake Cowichan, Vancouver Island, BC, Canada
162	WIN(M)472	*O. minus*	L	S	_	_	_	Taylor lake, Alberta, Canada
163	WIN(M)861	*O. minus*	L	S	_	_	_	Toronto, Ontaria, Canada
164	WIN(M)874	*O. minus*	L	S	_	+	_	Edmonton, Alberta, Canada
165	WIN(M)888	*O. minus*	L	S	_	_	_	Wyoming, Michigan, USA
166	WIN(M)889	*O. minus*	S	S	_	+	_	California, USA
167	WIN(M)1573	*O. minus*	L	L	_	_	_	Lakehead, Ontario, Canada
168	WIN(M)1574	*O. minus*	S	S	_	_	_	Lakehead, Ontario, Canada
169	CBS 108.21	*O. picea (ex-type)*	S	S	_	_	_	Germany
170	CBS 263.35	*O. picea*	S	S	+	_	_	Germany
171	CBS 374.67	*O. himal ulmi*	L	L	_	_	_	Kashmir, India

*CBS, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; DAOM, Cereal and Oilseeds Research, Agriculture & Agri-Food Canada, Ottawa, ON, Canada; DED, isolation number for Dutch elm disease strains; cultures generated from material provided by Manitoba Conservation and Saskatchewan DED surveys; strains now part of WIN(M); FSC, Fredericton Stock Culture Collection; WIN(M), University of Manitoba, Winnipeg, MB, Canada (courtesy of J. Reid).

Notes:

**Mt-rnl-U7 => S, short fragment of 0.35 kb (assuming no HEG and intron present); => L, long fragment of 1.6 kb (assuming intron present; see Sethuraman et al. 2008).

Mt-rnl-U11 => S, short fragment of 1.6 kb (intron present that encodes rps3); => L, long fragment of 2.6 kb (assume intron present with rps3 plus a HEG; see Sethuraman et al. 2009a).

*rnl*i1 +, present if PCR fragment 1.5 kb; - , absence of insertion

*rnl*i2 +, present if PCR fragment 1.9 kb; - , absence of insertion

*rnl*i3 +, present if PCR fragment 1.8 kb; - , absence of insertion

O. novo-ulmi subsp. *novo-ulmi rnl*i2 (mL963)

Based on PCR amplifications targeting the mL963 intron revealed that among *Ophiostoma* sp. 73 strains out of 171 (Table 3.1) yielded an amplicon of about 1.9 kb. Sequence analysis showed that this intron is inserted with respect to the *E. coli* 23S rDNA at position 963. RNAweasel failed to identify this element as an intron, but comparative sequence analysis suggested this region represents an IC1 group I intron. The ORFfinder program upon closer examinations revealed the presence of a highly degenerated ORF of about 483 bps that encodes a 160 amino acid peptide. One can assume that the degenerated ORF maybe is related to the ORF in mL740 and the degeneration may due to the presence of premature stop codons, probably due to multiple substitutions.

Related ORFs based on a BLASTp database search identified putative GIY-YIG type homing endonuclease-like proteins in *Cordyceps brongniartii* (EU100743; Ghikas et al., 2010) and in *rnl* gene of *Peltigera malacea* (AEK48293). Strains representing *Ophiostoma himal-ulmi, Ophiostoma minus,* and *Ophiostoma piceae* were also included in this survey and an insertion has been noted in a strain of *Ophiostoma minus* (WIN(M)889); therefore one can assume that this insertion might be related to the mL963 group I intron found in the genome.

O. novo-ulmi subsp. *novo-ulmi rnl*i3 (mL1092)

A third intron was noted in *O. novo-ulmi* subsp. *novo-ulmi*. This intron is positioned at 1092 with regards to the *E. coli* LSU rDNA; the online program RNAweasel showed that this intron is a group I C2 type. The ORFfinder program indicates the presence of a putative ORF. The ORF appears to encode a 449 amino acid polypeptide that contains two conserved copies of the

LAGLIDADG motifs (double motif), suggesting this ORF is a potential homing endonuclease or a maturase (Belfort and Roberts, 1997; Belfort et al., 2002). Based on BLASTp analysis related ORFs were noted in the *Fusarium graminearum* (YP_001249305), and *Grosmannia piceaperda* (ACV41152) inserted within the *nad3, rps3* genes, respectively.

The PCR survey suggests that this intron at mL1092 is not frequently encountered among the ophiostomatoid fungi studied herein. Only four strains (out of 171) were noted to yield a PCR product indicative of the presence of an insertion in this segment of the *rnl* gene (Table 3.1). Intron mL1092 is absent in all tested strains of *O.novo-ulmi* subsp. *americana,* but it was detected in one of the *O.novo-ulmi* subsp. *americana* strains (WIN(M)1226) and in two strains of *Ophiostoma pseudominus* (WIN(M)460 and WIN(M)461).

In general no obvious patterns were noted with regard to the presence (or absence) of *rnl* introns (mL740, mL963, mL1092) among strains of *O. novo-ulmi* subsp. *americana*. The random distribution of introns among strains of the *O. ulmi* species complex might be indicative of rapid gain and /or loss of introns due to the potential of horizontal transfer of introns and the rapid loss of introns (and HEGs) as a result of the life cycles of these mobile elements as predicted by Goddard and Burt's (1999). However, it is possible that the ancestor of the *O. ulmi* species complex was intron-rich and introns were vertically transferred but either due to chance (drift) or selection pressure introns are lost among members of this species complex. It has been noted that in general among the more recently derived subspecies of *O. novo-ulmi* there appears to be a reduction in the size of the mitochondrial genome (Bates et al., 1993). The latter is due to the loss of optional insertions such as introns (Bates et al., 1993; Gibb and Hausner 2005; Paoletti et al., 2005; Paoletti et al., 2006).

O. novo-ulmi subsp. *novo-ulmi rnl*i4 (mL2449)

Previously it was reported that some strains of *O. ulmi* and *O. novo-ulmi* subsp. *americana* contain an intron located at mL1669; however, this intron was not encountered within the subspecies *novo-ulmi*. *Ophiostoma novo-ulmi* subsp. *novo-ulmi* is a species known to occur in Europe and possibly parts of Asia (Brasier 1991); but unfortunately very few strains are available for studies in North America. Therefore, one cannot comment on the possible absence of this intron from this species. However, the subspecies *novo-ulmi rnl* sequence examined during this study did show the presence of the mL2449 intron.

The *rnl*-U11 intron (=mL2449) has been reported to be variable in size because in addition to encoding the rps3 protein (Burke and Rajbhandary, 1982) it can sometimes contain additional elements. In strains of *Cryphonectria parasitica*, *O. novo-ulmi* subsp. *americana*, *Leptographium* spp., and *Grosmannia* spp. the rps3 coding region was shown to be invaded by HEG-like/maturase-like elements (Hausner et al., 1999; Gibb and Hausner, 2005; Sethuraman et al., 2009a; Rudski and Hausner 2012).

It was attempted to use primers, designed on available *rnl* sequences to amplify *rnl* genes from other member of the *O. ulmi* complex (such as strain DAOM 171051). However, only partial sequences were obtained due to repeated failure of sequencing reactions. In the future new sequencing strategies need to be developed in order to sequence other *rnl* genes from members of this species complex.

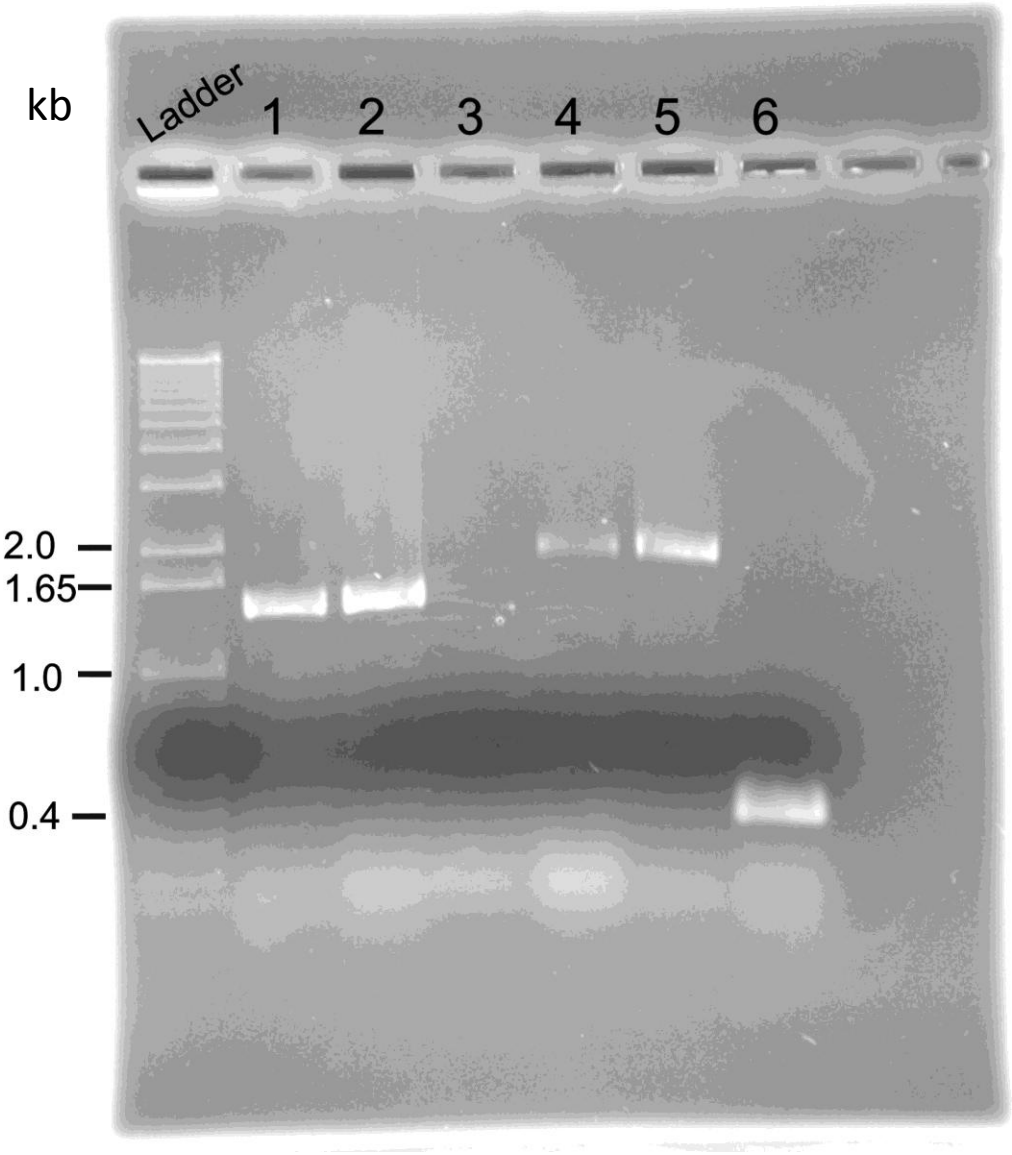

kb
Ladder
1
2
3
4
5
6
2.0
1.65
1.0
0.4

Figure 3.2. Representative PCR amplification products obtained from members of *O. ulmi s.l.* and related taxa for primers targeting *rnl* regions that typically harbor insertions (see Table 3.1). The PCR products for *O. novo-ulmi* subsp. *novo-ulmi* (strain IMI 343.101) are in lanes 1, 3, and 5; PCR products for *O.ulmi* (strain DAOM 171045) are at lane 2, 4, and 6. Products in lanes 1 and 2 represent the rnli1region; lanes 3 and 4 represent the rnli2 region; lanes 5 and 6 represent the rnli3 region. The PCR products for lane 1 and lane 2 are around 1.5 kb, which indicates the presence of an insertion at rnli1 region for both *O. novo-ulmi* subsp. *novo-ulmi* (IMI 343.101) and *O.ulmi* (DAOM 171045). Only one product (~1.9 kb) shown in lane 4 and no product at lane 3 indicate insertion was present at rnli2 region for IMI 343.101 but not DAOM 171045. *O. ulmi* (DAMO 171051) yieled a product around 1.8 kb (lane 5) at rnli3 region, while *O. novo-ulmi* subsp. *novo-ulmi* (IMI 343.101) gave a product around 0.4 kb (lane 6); this suggests lack of inserts/introns at rnli3 region for *O. novo-ulmi* subsp. *novo-ulmi*.

RNA Secondary Structure models for the mtDNA *rnl* introns as noted in *O. novo-ulmi* subsp. *novo-ulmi*

The *rnl*i1, *rnl*i2, and *rnl*i3 sequences found in *O. novo-ulmi* subsp. *novo-ulmi* contained the characteristic features expected for group I introns (Burke et al., 1987; Cech, 1988; Michel and Westhof, 1990). In previous studies the mL1669 intron, found in some *O. ulmi* and *O. novo-ulmi* strains and the mL2449 intron, found in all strains of the *O. ulmi* species complex, were characterized (Gibb and Hausner, 2005; Sethuraman et al., 2008). So here I focused on those introns not previously characterized. The online program RNAweasel was used to identify the intron core sequences (P3, P4, P6, P7, and P8) for the three *rnl* introns. Eventually the intron secondary structures were generated by the mFOLD program. With regards to the *E. coli* 23S rDNA these introns, as previously stated, can be designated as mL740, mL963, and mL1092.

For the mL740 intron a putative ORF was identified that belongs to the GIY-YIG HE family; it was located between the P7 and P9 conserved regions (Fig 3.3 A). A presumed degenerated GIY-YIG type ORF was also detected in the mL963 intron; here the putative ORF was located in the P9 region (Fig 3.3 B), and it has a large degree of overlap with the intron P9 region. The presence of the conserved P7 sequence suggests that the mL740 and mL963 introns belong to the IC1 type of group I introns (Michel and Westhof, 1990). The mL1092 intron based on the presence of conserved sequences that comprise the P1 and P4 elements identifies this intron as a member of the IC2 type of group I introns (Michel and Westhof, 1990). A LAGLIDADG type ORF was identified in the mL1092 intron, this ORF is located in the P9.1 loop region of the intron structure (Fig 3.3 C). Typically intron encoded ORFs are found in the looped regions, which presumably minimize interference with intron RNA secondary and tertiary structures in order to maintain a splicing competent configuration (Edgell et al., 2011).

The secondary and tertiary structures of the intron RNAs are the key for these introns to be able to self-splice from the host transcripts; any features that compromise the RNA structure may cause a splicing defect and this might be lethal or result in mitochondrial defects (Edgell et al., 2011; Hausner, 2012). It has been noted that situations exist where the ORF sequences overlap with the intron core structures (see mL740) and this requires coordinated events that ensure that the HE ORF can be expressed and translated and that the intron can splice from the host gene transcript (Gibb and Edgell 2010; Edgell et al., 2011). In general, to ensure long term survival of introns and their open reading frames these composite elements must minimize their impact on the host gene yet be able to invade cognate alleles that lack insertions (i.e. homing) or be able to invade new sites (i.e. transposition).

Fig 3.3 A

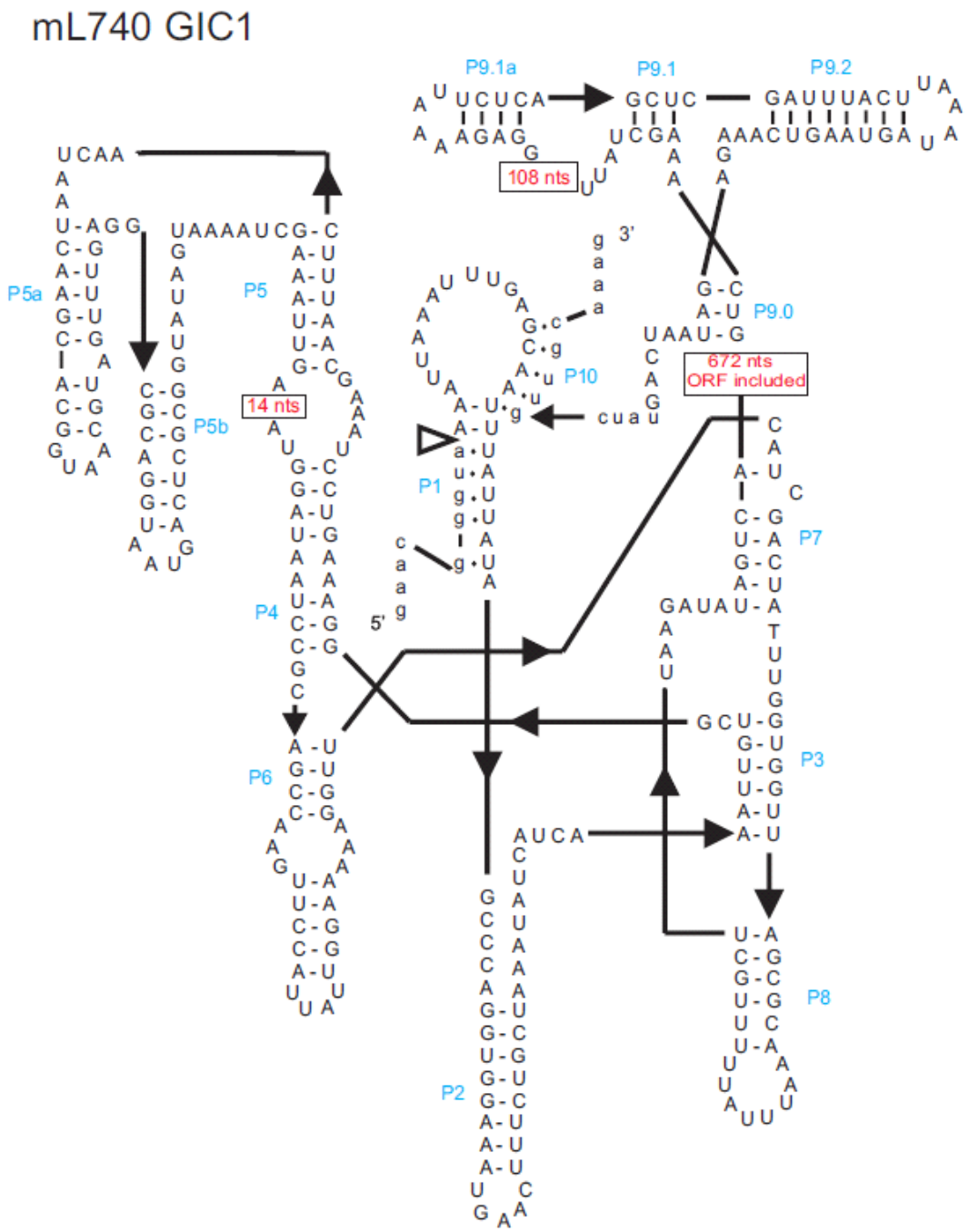

mL740 GIC1
P9.1a
P9.1
P9.2
108 nts
P5a
P5
P5b
14 nts
P10
P9.0
672 nts
ORF included
P1
P7
P4
P6
P3
P2
P8

Fig 3.3 B

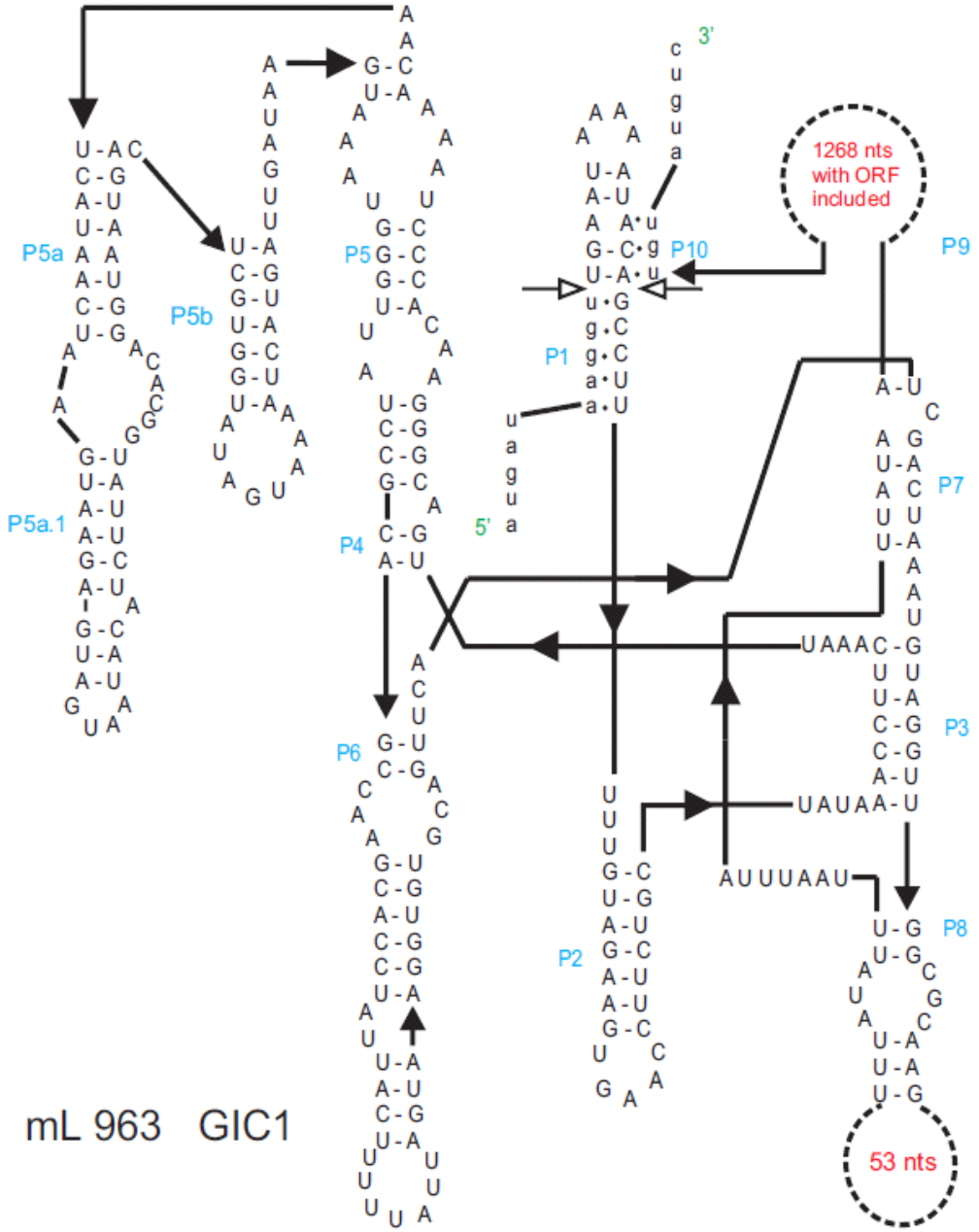

P5a
P5a.1
P5b
P5
P4
P6
P1
P2
P3
P7
P8
P9
P10
5'
3'
1268 nts
with ORF
included
53 nts
mL 963 GIC1

Fig 3.3 C

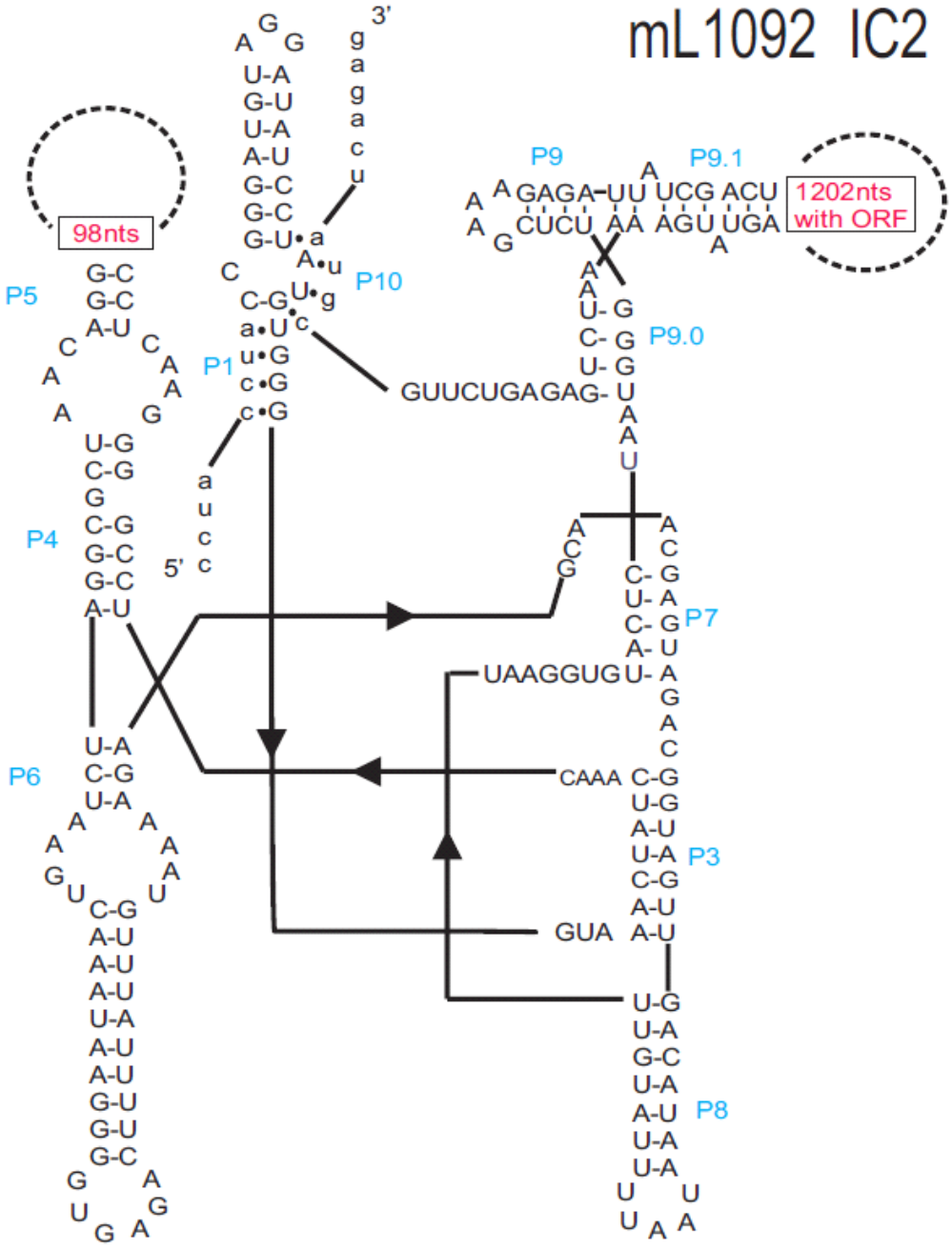

mL1092 IC2
98nts
1202nts with ORF
P1
P3
P4
P5
P6
P7
P8
P9
P9.0
P9.1
P10
GUUCUGAGAG-
UAAGGUG
CAAA
GUA
5'
3'

Figure 3.3. A) Predicted secondary structures of the mL740 GIC1 intron RNA. Intron nucleotides (nt) are presented as uppercase letters, flanking exons sequences (5' to 3') are in lowercase letters and large arrows show intron-exon junctions, that have been designated based on secondary structure modeling and sequence homology with other *rnl* genes (Cummings et al., 1989). The position of the ORF is indicated with red letters in the text box. Nucleotides forming the P10 region are labeled. B) Secondary structure of the mL963 GIC1 intron RNA. C) Secondary structure of the mL1092 GIC2 intron RNA.

The mtDNA *rnl* gene landscape in ascomycetous and some basidiomycetous fungi

The initial focus of this study was to examine the *rnl* intron landscape for a variety of ascomycetous and basidiomyctous fungi. So *rnl* gene sequences were extracted from Genbank (by blastn and by examining the NCBI organellar genome site). It was noted that there was considerable length variation among *rnl* genes and this is due to the presence of introns (and if applicable intron-associated ORFs). The *rnl* sequences were compiled and by comparative sequence analysis a total of 23 possible intron insertion sites were noted in the *rnl* genes of ascomycetous and basidiomycetous fungi in this study. The *rnl* insertion sites (positions) were identified based on sequence comparison with the *E. coli* LSU rDNA gene (Johansen and Haugen 2001) and the insertion sites were noted to be as follows: L442, L680, L740, L780, L810, L963, L1092, L1282, L1671, L1699, L1787, L1923, L1931, L1971, L2029/2030, L2048, L2059, L2231, L2330, L2449, L2451, L2499/2500, and L2584 (see Fig. 3.4 and Table 3.2).

Overall among the fungal *rnl* intron complement both group I and group II introns were noted, and most of the intron contained ORFs, although a few group I introns were noted that lacked an ORF. The online program RNAweasel (Lang et al., 2007) was used to confirm the intron designations with regards to intron subtypes (Michel and Westhof 1990). Group II introns were found in the following positions: L1787 (IIB1) and L2059 (IIB1) and Group I introns were recorded in the following positions: L442 (IA3), L680 (IC2), L740 (IC1), L780 (IA3), L810 (IC1), L963 (IC1), L1092 (IC2), L1282 (IA), L1671 (IC2), L1699 (IA), L1923 (IB4), L1931 (IB4), L1971 (IC2), L2029/2030 (ID), L2048 (IB4), L2231 (ID), L2449 (IB), L2451 (IA), L2499/2500 (IB), and L2584 (IA) (Table 3.2).

A special intron has been found in position L2330 (*Peltigera membranace* JN088165;

with a 394 nt insertion); this insertion appears to be ORFless, but a highly degenerated ORF might be present (i.e. only short pieces can be found identified by the ORF finder program) and was not identified by the online RNA weasel program and the sequence and secondary structure does not match any currently available intron structures. One must assume that this insertion is probably removed by some splicing event that may be facilitated by host factors that assist what appears to be a degenerated intron core.

The *rnl* intron insertion sites have been mapped onto the *E. coli* 23 S secondary structure model (http://rna.ucsc.edu/rnacenter/ribosome_images.html). The objective was to examine if intron insertion sites show a biased distribution, such as insertion within stem or loop regions within the 23 S secondary structure. Indeed some introns are present within stem regions but most of the insertions are in loop or single stranded regions (see Table 3.1a). One has to assume that intron insertions are neutral and do not interfere with the expression and processing of rRNAs so intron(s) contained within the stem region(s) are spliced efficiently. With regard to the *E. coli* 23 S secondary structure model, some domains such as domain II, VI and V, appear to have more insertions than others; for example, six insertions have been found in domain IV while no insertions have been found in domain VI (see Table 3.1b). Again one would assume that insertions are located where they do not interfere with rRNA folding and function.

This study shows that many filamentous ascomycetous fungi have an intron present at position mL2449. In many fungi the mtDNA *rnl* L2449 group I intron encodes a mitochondrial ribosomal protein previously referred to as S5 (Cummings et al., 1989b; Hausner et al., 1999; Hausner 2003); this S5 gene has been recognized as a homolog of the ribosomal protein 3 gene (*rps3*) (Bullerwell et al., 2000). In most fungi the *rps3* gene is either absent or free

Table 3.1 Location of *rnl* introns with regards to the 23 S secondary structure model (as designated for *E. coli*, (http://rna.ucsc.edu/rnacenter/ribosome_images.html)).

a)

***E. coli* LSU region**	**Introns**
Stem	L442, L680, L740, L810, L1923, L2048, L2231
Loop	L780, L963, L1092, L1282, L1671, L1699, L1787, L1931, L1971, L2029/2030, L2059, L2330, L2449, L2451, L2499/2500, L2584

b)

***E. coli* LSU domain**	**Introns**
Domain I	L442
Domain II	L680, L740, L780, L810, L963, L1092
Domain III	L1282
Domain IV	L1671, L1699, L1787, L1923, L1931, L1971
Domain V	L2029/2030, L2048, L2059, L2231, L2330, L2449, L2451, L2499/2500, L2584
Domain VI	NA

standing but among most filamentous ascomycetous fungi rps3 is encoded by the mL2449 intron. In some ascomycetous fungi the *rps3* gene has been relocated from the mL2449 intron and now it is in a free-standing form; a more detailed study on *rps3* and its relationship with the mL2449 intron is presented in Chapter 4 (page 70).

The *rnl* gene is conserved, and thus an ideal gene to be invaded by elements that need to recognize a specific target sequence in order for insertion into cognate alleles. This allows for these elements to spread and persist in a population. For both group I and group II introns the IEPs (intron encoded proteins) require long recognition sites thus favouring insertion into conserved regions such as within the rDNA regions. Generally it has been noted that introns located at the same insertion site are more closely related to each other than to introns inserted at different positions (Haugen et al., 2005).

Group I introns move by DNA based recombination methods medicated by their encoded homing endonucleases. Group II introns typically are mobilized via mechanisms that involves an RNA intermediate and the activity of the group II intron encoded reverse transcriptase (Lambowitz and Zimmerly 2011). In the system presented here, the intron RNA is spliced from its host transcript and the intron lariat forms a complex with the intron encoded reverse transcriptase (RT) to form a cDNA that can be inserted into a new location (Zimmerly et al., 1995a, b). It has also been postulated that mobile introns can move into new locations by reverse splicing. For both group I and II introns it has been postulated that intron RNA can reverse splice into a transcript and that an RT reverse transcribes the RNA into a cDNA, which is then inserted into a new location. These mechanisms may explain how introns over time have invaded so many different locations within the *rnl* gene among so many different organisms (i.e. 23 possible intron insertions detected in these studies). However, there is only circumstantial evidence that

reverse splicing could occur; so far experimental (*in vivo*) evidence for this mechanisms is lacking.

It is worth noting that the group II introns reported during this study on the *rnl* gene encode LAGLIDADG type ORFs, which are usually expected to be located in group I introns. Toor and Zimmerly (2002) noted that group II introns can encode group I type ORFs. Additional examples of group II introns with LAGLIDADG ORFs have been described by Hafez and Hausner (2011), Mullineux et al., (2010), and Sethuraman et al., (2013). It has been shown by Mullineux et al. (2010) that group II intron encoded LAGLIDADG type ORFs can express functional homing endonucleases; thus these types of group II introns might utilize mechanisms of mobility that are similar to those of group I introns.

The mitochondrial genome is highly variable due to the presence of introns and intron encoded HEases and the current study provides an updated overview of the potential mobile elements that can interrupt the *rnl* gene. From an applied point of view, this may suggest that the *rnl* gene is a good target for bioprospecting for ribozymes and endonucleases, elements that have applications in biotechnology (Gimble, 2005; Lambowitz et al., 2005; Hausner, 2003; Hafez and Hausner 2012).

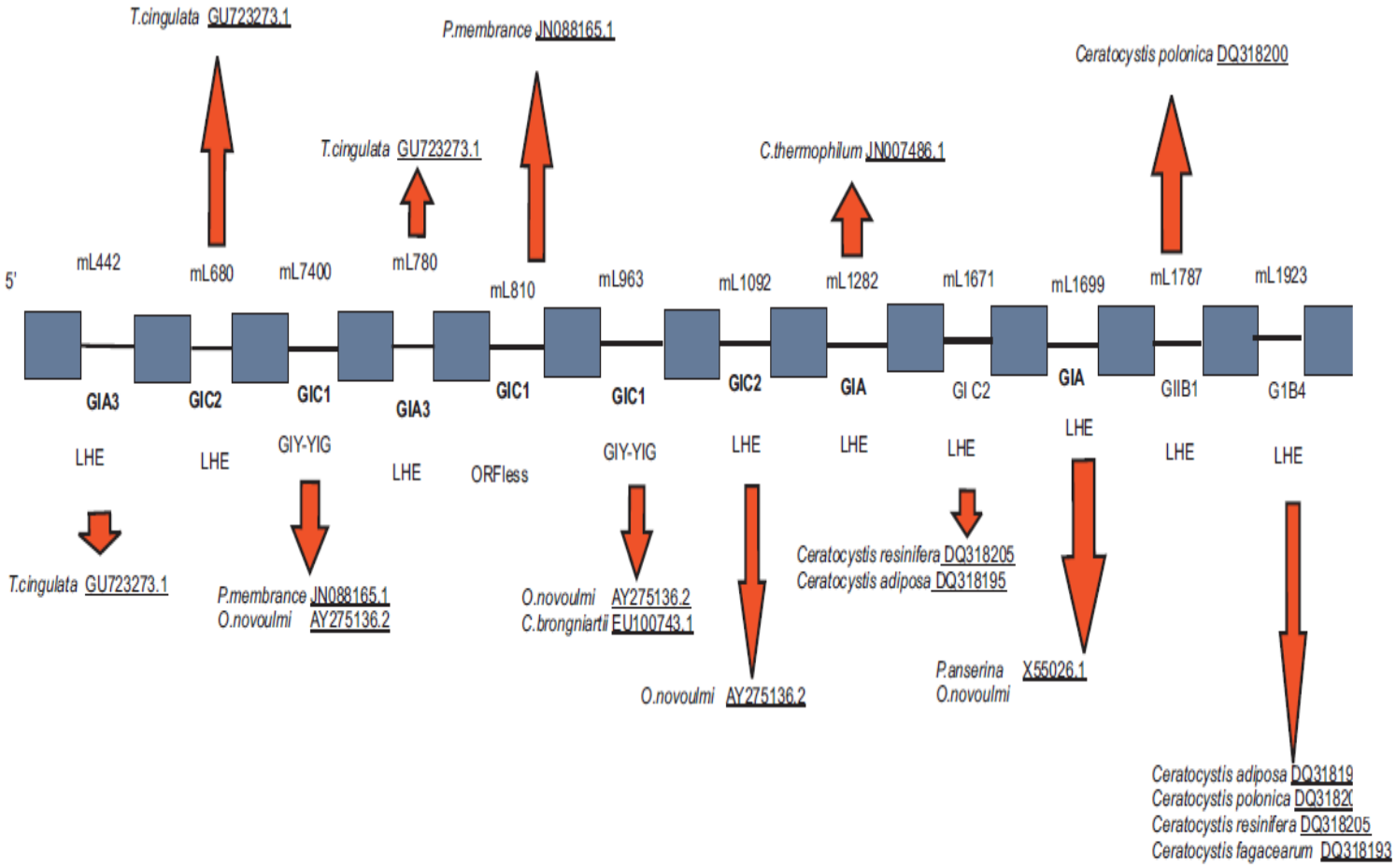
Table 3.2 List of introns and intron types noted within *rnl* genes for ascomycetous and some basidiomycetous fungi.

Intron insertion site	**Intron type**	**Examples**	**Accession number**	**ORF**
mL 442	GIA3	*Trametes cingulate*	GU723273	LHE
mL 680	GIC2	*Trametes cingulate*	GU723273	LHE
mL 740	GIC1	*Peltigra membrance* *Ophiostoma novo-ulmi*	JN088165 AY275136	GIY-YIG
mL 780	GIA3	*Trametes cingulate*	GU723273	LHE
mL 810	GIC1	*Peltigra membrance*	JN088165	ORFless
mL 963	GIC1	*Ophiostoma novo-ulmi* *Cordyceps brongniartii*	AY275136 EU100743	GIY-YIG
mL 1092	GIC2	*Ophiostoma novo-ulmi*	AY275136	LHE
mL 1282	GIA	*Chaetomium thermophilum var. thermophilum*	JN007486	LHE
mL 1671	GIC2	*Ceratocystis resinifera* *Ceratocystis adiposa*	DQ318205 DQ318195	LHE
mL 1699	GIA	*Podospora anserina* *Ophiostoma novo-ulmi*	X55026 AY275136	LHE
mL 1787	GIIB1	*Ceratocystis polonica*	DQ318200	LHE
mL 1923	GIB4	*Ceratocystis adiposa* *Ceratocystis polonica* *Ceratocystis resinifera* *Ceratocystis fagacearum*	DQ318195 DQ318200 DQ318205 DQ318193	LHE
mL 1931	GIB4	*Phakopsora pachyrhizi* *Phakopsora meibomiae*	GQ332420 GQ338834	ORFless
mL 1971	GIC2	*Ceratocystis fagacearum*	DQ318193	GIY-YIG
mL 2029	GID	*Ceratocystis cacaofunesta*	JX185564	LHE
mL 2048	GIB4	*Trametes cingulate*	GU723273	LHE
mL 2059	GIIB1	*Agrocybe aegerita* *Ceratocystis polonica*	AF087656 DQ318200	LHE
mL 2231	GID	*Ceratocystis fagacearum*	DQ318193	LHE
mL 2330		*Peltigra membrance*	JN088165	ORFless ,deg LAG
mL 2449	GIB	*Ceratocystis cacaofunesta* *Chaetomium thermophilum var. thermophilum* *Peltigra membrance* *Ophiostoma novoulmi* *Cordyceps brongniartii* *Arthroderma obtusum* *Aspergillus tubingensis*	JX185564 JN007486 JN088165 AY275136 EU100743 FJ385029 DQ217399	Rps3 (in a few instances LHE)

		Aspergillus niger *Trichophyton rubrum* *Verticillium dahliae* *Podospora anserina* *Trichophyton mentagrophytes* (also see Sethuraman et al., 2009b)	DQ207726 FJ385026 DQ351941 X55026 FJ385027	
mL 2451	GIA	*Trametes cingulata*	GU723273	LHE
mL 2499	GIB	*Ceratocystis cacaofunesta*	JX185564	LHE
mL 2584	GIA	*Chaetomium thermophilum var. thermophilum*	JN007486	LHE

GI = group I intron; GII = group II intron; A-D = intron subtype

LHE = LAGLIDADG type ORF; GIY-YIG = GIY-YIG type ORF


T.cingulata GU723273.1
P.membrance JN088165.1
Ceratocystis polonica DQ318200
T.cingulata GU723273.1
C.thermophilum JN007486.1
5'
mL442
mL680
mL7400
mL780
mL810
mL963
mL1092
mL1282
mL1671
mL1699
mL1787
mL1923
GIA3
GIC2
GIC1
GIA3
GIC1
GIC1
GIC2
GIA
GI C2
GIA
GIIB1
G1B4
LHE
LHE
GIY-YIG
LHE
ORFless
GIY-YIG
LHE
LHE
LHE
LHE
LHE
LHE
T.cingulata GU723273.1
P.membrance JN088165.1
O.novoulmi AY275136.2
O.novoulmi AY275136.2
C.brongniartii EU100743.1
O.novoulmi AY275136.2
Ceratocystis resinifera DQ318205
Ceratocystis adiposa DQ318195
P.anserina X55026.1
O.novoulmi
Ceratocystis adiposa DQ31819
Ceratocystis polonica DQ3182(
Ceratocystis resinifera DQ318205
Ceratocystis fagacearum DQ318193

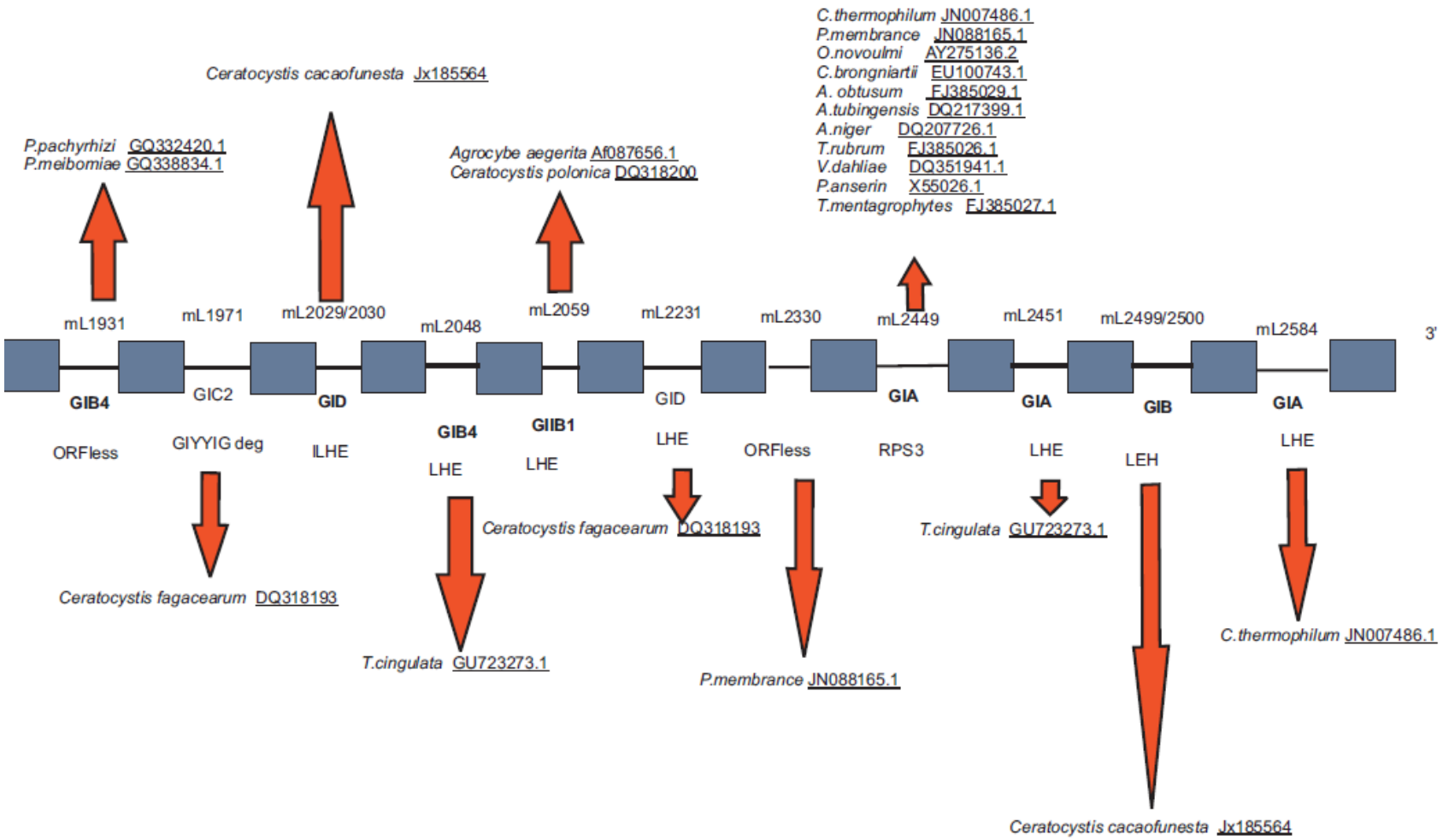

C.thermophilum JN007486.1
P.membrance JN088165.1
O.novoulmi AY275136.2
C.brongniartii EU100743.1
A. obtusum FJ385029.1
A.tubingensis DQ217399.1
A.niger DQ207726.1
T.rubrum FJ385026.1
V.dahliae DQ351941.1
P.anserin X55026.1
T.mentagrophytes FJ385027.1
Ceratocystis cacaofunesta Jx185564
P.pachyrhizi GQ332420.1
P.meibomiae GQ338834.1
Agrocybe aegerita Af087656.1
Ceratocystis polonica DQ318200
mL1931
mL1971
mL2029/2030
mL2048
mL2059
mL2231
mL2330
mL2449
mL2451
mL2499/2500
mL2584
3'
GIB4
GIC2
GID
GIB4
GIIB1
GID
GIA
GIA
GIB
GIA
ORFless
GIYYIG deg
ILHE
LHE
LHE
LHE
ORFless
RPS3
LHE
LEH
LHE
Ceratocystis fagacearum DQ318193
T.cingulata GU723273.1
Ceratocystis fagacearum DQ318193
P.membrance JN088165.1
T.cingulata GU723273.1
C.thermophilum JN007486.1
Ceratocystis cacaofunesta Jx185564

Figure 3.4 The *rnl* intron landscape. The *rnl* introns are represented by gray lines. This figure was generated by aligning the collected *rnl* sequences with the *E.coli* 23S reference sequence. This allowed for designating insertion sites. Most putative introns are Group I introns, but some positions show group II intron insertions (eg. L1787and L2059). The intron insertion sites, intron types and type of intron encoded proteins are indicated, for more detail also consult Table 3.2. Gray boxes indicate exons; black lines in between the boxes indicate introns. Size of the arrows does not indicate the frequency of presence of introns.

Chapter 4: Evolutionary History of the mtDNA *rps3* gene among the Ascomycota

4.1 Introduction

The mitochondrial ribosomal protein S3 (*rps3*) gene within the fungi is very diverse in its sequence, location and organization. Among many ascomycetes the *rps3* gene appears to be located inside the mL2449 intron. However, among other fungi and the Saccharomycetales the *rps3* gene is either missing from the mtDNA or a free standing gene (i.e. not associated with an intron) (Sethuraman et al., 2009a). Previously it was shown that in one ascomycete fungus, *Phaeosphaeria nodorum*, the loss of the mL2449 intron in large ribosomal subunit gene (*rnl*) was associated with the appearance of a free standing version of the *rps3* gene (Sethuraman et al., 2009a).

In a previous study, it was shown that the *P. nodorum rps3* gene appears to have a long ORF encoding a 771 amino acid peptide (Hane et al., 2007; YP_001427396; Sethuraman et al., 2009a). However, there were some anomalies with regards to this ORF. The amino acid positions 45-417 within the *P. nodorum rps3* ORF align with other mL2449 introns (in particular with the *Penicillium marneffei*; AAQ54923) encoded Rps3 protein sequences, but the amino acid positions 418-771 have no similarity with positions in other Rps3 homologs. The amino acid positions 432-487, appears to be related to part of the cytochrome oxidase subunit 1 protein of *P. nodorum*. Although the ORF appears to be free-standing, the *rps3* gene based on phylogenetic analysis is related to the intron encoded versions of *rps3* (Sethuraman et al., 2009a). Based on this observation one can speculate that a recombination event relocated the *rps3* gene from the original mL2449 intron position and somehow a *cox1* gene segment was incorporated within the

new *rps3* locus. This study reexamines the evolution of the rps3 locus among members of the Ascomycota to examine if free-standing forms of rps3 evolved only once due to an as of yet unknown recombination event or if this event happened several times during the evolution of the filamentous ascomycetes fungi.

4.2 Material and methods

Bioinformatics

All the Rps3 amino acid sequences in this study were collected from the online database GenBank (NCBI). Table 4.1 lists GenBank accession numbers for all Rps3 amino acid sequences obtained for this study. With regards to *rnl* gene sequences the ORF finder program (http://www.ncbi.nlm.nih.gov/projects/gorf/) was used to search for potential ORFs within the mL2449 intron nucleotide sequences. The online source BLASTp (Altschul et al., 1990) was used to retrieve additional Rps3 sequences. Amino acid sequences were compiled and aligned with the online program PRALINE multiple sequence alignment program (Simossis and Heringa 2003) and the MAFFT program (= Multiple alignment program for amino acid or nucleotide sequences; Version 7; Katoh et al., 2002-2013). The amino acid sequence alignment was further refined with GeneDoc. The online annotation program MFannot (Gautheret and Lambert, 2001; Lang et al., 2007; http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) was used to identify the *rps3* region in some fungal mtDNA sequences, which either were not annotated yet or have been poorly annotated.

Table 4.1 Organism name and GenBank accession numbers for Rps3 sequences analyzed during this study.

Organism	**Strain number (if available)**	**GenBank number**
Aspergillus nidulans	FGSC A4	YP_006303579
Ashbya gossypii	ATCC 10895	NP_987081
Beauveria brongniartii	CBS 128.53	FJ895613
Bipolaris maydis		KB733493
Botryotinia fuckeliana		KC832409
Candida glabrata	CBS138	NP_818776
Ceratocystiopsis brevicomi	WIN(M) 1452	FJ717840
Ceratocystis cacaofunesta		YP_007507039
Ceratocystis coerulescens	WIN(M) 98	FJ895608
Ceratocystis fimbriata	WIN(M) 931	FJ895616
Ceratocystis polonica	WIN(M) 325	KF366612
Ceratocystiopsis minuta-bicolor	WIN(M) 480	FJ717855
Ceratocystiopsis parva	WIN(M) 59	FJ717854
Ceratocystiopsis rollhanseniana	WIN(M) 113	FJ717853
Chaetomium thermophilum var. thermophilum	DSM 1495	YP_004769778
Cornuvesica falcata	UAMH 9702	FJ895609
Cryphonectria parasitica		AAC24230
Cyberlindnera mrakii		RMAR_CYBMR
Epidermophyton floccosum		YP_313620
Exophiala dermatitidis	NIH/UT8656	EHY51767
Fusarium oxysporum		AAW67496
Fusarium oxysporum		YP_006341038
Gelasinospora tetrasperma	ATCC 11345	FJ895607
Glarea lozoyensis		KF169905
Glomerella graminicola	M1.001	EFQ24852
Gondwanamyces proteae	CBS 484.88	FJ895610
Grosmannia aurea	WIN(M) 809	ACV41162
Grosmannia europhioides	WIN(M) 449	ACV41156
Grosmannia europhioides	WIN(M) 1430	ACV41151
Grosmannia europhioides	WIN(M) 1431	ACV41151
Grosmannia laricis	WIN(M) 1461	ACV41166
Grosmannia penicillata	WIN(M) 136	ACV41148
Grosmannia penicillata	WIN(M) 27	ACV41146
Grosmannia piceiperda	WIN(M) 979	ACV41152
Grosmannia pseudoeurophioides	WIN(M) 42	ACV41163

Kazachstania servazzii		NP_861467
Kernia pachypleura	WIN(M) 253	FJ895612
Kluyveromyces lactis		YP_054503
Kluyveromyces thermotolerans		YP_184729
Lecanicillium muscarium		NP_775396
Leptographium pithyophilum	WIN(M) 1454	FJ607137
Leptographium procerum	WIN(M) 1250	FJ717850
Leptographium truncatum	WIN(M) 254	FJ717852
Leptographium truncatum	WIN(M) 1434	FJ717849
Leptographium truncatum	WIN(M) 1435	FJ717835
Microsporum canis		YP_002970890
Neurospora crassa	OR74A	RMS5_NEUCR
Ophiostoma distortum	WIN(M) 847	FJ717845
Ophiostoma himal-ulmi	CBS 374.67	FJ717862
Ophiostoma ips	WIN(M) 923	FJ717857
Ophiostoma ips	WIN(M) 1487	FJ717858
Ophiostoma megalobrunneum	WIN(M) 509	FJ717856
Ophiostoma minus	WIN(M) 888	FJ717859
Ophiostoma minus	WIN(M) 861	FJ717860
Ophiostoma novo-ulmi subsp.americana	WIN(M) 900	AY275136
Ophiostoma nigrum	CBS 163.61	FJ717846
Ophiostoma tetropii	WIN(M) 111	FJ717843
Ophiostoma tetropii	WIN(M) 451	FJ717844
Ophiostoma torulosum	WIN(M) 730	FJ717861
Ophiostoma ulmi	WIN(M) 1223	FJ717838
Paracoccidioides brasiliensis		YP_537113
Peltigera malacea		YP_005351166
Peltigera membranacea		YP_005351208
Penicillium chrysogenum		BAA02977
Penicillium digitatum		YP_004221875
Penicillium solitum (reconstructed)		NC_016187
Penicillium urticae		P47907
Phaeosphaeria nodorum	SN15	YP_001427397
Phialocephala subalpina		YP_004733050
Podospora anserina		NP_074911
Saccharomyces castellii		NP_644679
Saccharomyces cerevisiae		AAA67536
Saccharomyces douglasii		RMAR_SACDO
Sacotrochila macrospora	WIN(M) 1538	FJ895611
Sphaeronaemella fimicola	WIN(M) 1402	FJ895614
Sphaeronaemella fimicola	WIN(M) 818	FJ895615

Sporothrix sp.	WIN(M) 924	FJ717834
Sordaria fimicola	ATCC 6739	FJ717863
Talaromyces marneffei		AAQ54923
Talaromyces stipitatus		AFD95916
Trichoderma reesei		NP_570148
Trichosporon asahii var. asahii	CBS 2479	EJT45004
Wickerhamomyces canadensis		NP_038223

CBS = Central Bureau voor Schimmelcultures, Utrecht, The Netherlands
WIN(M) = University of Manitoba (Winnipeg) Collection
UAMH = University of Alberta Microfungus Collection and Herbarium, Devonian Botanic Garden, Edmonton, AB, Canada
ATCC = American Type Culture Collection, Manassas, VA, USA

Phylogenetic analysis

MEGA 5 (Tamura et al. 2011) and the MrBayes program v3.1 (Ronquist and Huelsenbeck 2003; Ronquist 2004) were used for phylogenetic analysis of the aligned Rps3 data set. Phylogenetic trees were to obtain with the neighbor joining (NJ), parsimony, and maximum likelihood (ML) options as implemented in MEGA5. The bootstrap method with 1000 replications was applied to the NJ, parsimony, and ML analyses. For ML analysis, the program called "find the best DNA/protein models" (ML) in MEGA5 was used to obtained the best fit model that applies to the data set to be analyzed. Jones-Taylor-Thornton (JTT) with frequencies (F) and gamma (G) distribution with five gamma rate parameters was selected for ML analysis based on the aforementioned analysis. The JTT model was also applied to the NJ analysis.

The MrBayes program was used for Bayesian analysis and the parameters applied for analysis of the amino acid alignment was as follows: JTT model and gamma distribution with four 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 deleted ("burn-in") and the remaining trees were used to compute the majority rule consensus tree and the posterior probability values.

Phylogenetic trees were drawn with the TreeView program (Page 1996) using the MrBayes tree file, and the trees were further refined (annotated) with Corel Draw (Corel Corporation and Corel Corporation Limited, Ottawa, ON).

4.3 Result and Discussion

The rps3 phylogenetic tree (Fig. 4.1) was rooted with the Rps3 sequence from *Trichosporon asahii* var. *asahii* (Basidiomycota). So far Rps3 sequences among the Basidiomycota have been noted to be highly variable and the *rps3* gene within this group of fungi is free standing (Hausner, unpublished work). The resulting tree had the var1 sequences from various species of the Saccharomycetales branching near the root of the tree. The *var1* genes (= rps3, see Bullerwell et al., 2000) are free-standing versions of the *rps3* gene and *var1* genes so far have not been noted to be associated with introns (Butow et al., 1985).

What follows on the tree are mitochondrial Rps3 sequences from filamentous ascomycetes fungi; previous studies have shown that among filamentous ascomycetes, the rps3 gene is encoded within a group I intron (mL2449) in the U11 region of the mitochondrial large subunit (*rnl*) gene (Bullerwell et al., 2000; Burke and Rajbhandary 1982; Sethuraman et al., 2009a,b). The node that unites this group of intron encoded versions of Rps3 received high support in all phylogenetic methods applied to the data set. In general the Rps3 sequences group into clades that corresponds to the taxonomic groups expected for the fungi (Webster and Weberz 1980; St-Germain and Summerbell 1996). The Rps3 data grouped all analyzed members of Saccharomycetales into a single clade. The data also grouped Sordariales and Pleosporales into single clade.

Among the mL2449 intron encoded Rps3 sequences sampled it was noted that two clades contained Rps3 sequences that did not appear to be inserted within an intron. Free-standing *rps3* genes were found in *Phaeosphaeria nodorum* (as reported previously by Sethuraman et al., 2009a) and in *Leptosphaeria maculans, Bipolaris maydis, Marssonina brunnea,* and *Phialocephala subalpine* (Duo et al., 2012; YP_004733050). In all five cases it was noted that

the *rnl* genes lacked the mL2449 intron.

Some of the sequences included here in this study were not yet available in the previous study by Sethuraman et al., (2009a). Also, some of the more recently deposited sequences were essentially generated as part of genome projects and the mtDNA sequences were not annotated or only partially annotated. Therefore the *Leptosphaeria maculans* (Rouxel et al., 2011; XM_003836958), *Bipolaris maydis* (Ohm et al., 2012; KB733493), and *Marssonina brunnea* (Zhu et al. 2012; JN204424) mtDNA sequences had to be annotated first (using the online ORF finder and MFannot programs) before the Rps3 ORFs and the *rnl* genes could be identified. The Rps3 ORF was noted that it encodes a 490 amino acid polypeptide for *Bipolaris maydis*, the amino acid sequence aligns with other mL2449 intron encoded rps3 sequences (such as *Penicillium chrysogenum*; BAA02977). The *Leptosphaeria maculans* Rps3 ORF encodes a 416 amino acid peptide the amino acid sequence showed a 95% identity with other mL2449 intron encoded protein sequence in particular with the *Penicillium chrysogenum* (BAA02977). The *Marssonina brunnea* mtDNA Rps3 ORF encodes a 506 amino acid peptide and this ORF aligns with other mL2449 intron encoded Rps3 protein sequence such as the *Glarea lozoyensis* Rps3 sequence (AGN74485). The *Phialocephala subalpine* mtDNA Rps3 ORF encodes a 546 amino acid peptide and it shows a 95% identity level with other intron (mL2449) encoded Rps3 protein sequences, in particular with *Botryotinia fuckeliana* (AGN49003). Also as stated previously all four free-standing mitochondrial Rps3 ORFs plus the previously characterized *P. nodorum* Rps3 sequence belong to the intron (mL2449) encoded Rps3 protein clade. Or in other words these free-standing variants of rps3 are derived from intron encoded rps3 sequences. In all five examples the *rnl* genes were examined for the possible presence of the mL2449 intron (or remnants of this intron) and the *rps3* nucleotide sequences were also examined with regards to

flanking mtDNA sequences (500 bps upstream and downstream of the ORF) to see if there is any evidence for the presence of "residual" intron like sequences. For both (mL2449 introns and residual intron-like sequence flanking the free standing rps3 ORFs) the results were negative.

Among many ascomycete fungi the *rps3* gene is usually associated with an *rnl* group I intron (mL2449). The mL2449 group I intron belongs to the class IA1 and is transcribed along with the host gene and it is assumed the intron is spliced out from the host transcript; thus the intron would not perturb the function of the host gene. Therefore, the intron is interpreted to be neutral and would be the effect of any ORFs located within the intron. Typically group I introns encode homing endonucleases; which promote the mobility of the intron and therefore ensure the long term survival of both, intron and HEG, within a fungal population. Based on this observation and regards to the mL2449 intron, one must assume that encoding a host factor that is required for ribosome function ensures the intron is maintained, but once the Rps3 ORF is lost the intron becomes redundant and due to neutral evolution mutations would accumulate (due to lack of selection) quickly leading to the elimination of the intron (Goddard and Burt 1999). It is noteworthy that the *Neurosproa crassa* mL2449 intron has been extensively characterized by various research groups (Lambowitz et al., 1993; Bertrand et al., 1982) as it does not appear to readily splice and it requires various nuclear encoded factors in order to be efficiently removed from the rRNA precursor (Akins and Lambowitz 1987; Bertrand et al., 1982; reviewed in Hausner 2012). This unusual splicing of *Neurosproa crassa* mL2449 intron would suggest that maintaining this intron comes at a cost to the host cell. If the Rps3 ORF is lost there would be not benefit in maintaining this intron.

Within the filamentous ascomycetes the mL2449 intron encoded Rps3 version appear to be the "standard" arrangement for this locus. So how did the free standing versions get

established? Previous work showed that sometimes the intron encoded Rps3 coding region has been invaded by or the Rps3 ORF has been fused to LAGLIDADG type HEG (Hausner et al., 1999; Gibb and Hausner 2005; Sethuraman et al., 2009a; Rudski and Hausner 2012) maybe the Rps3 ORF was relocated due to an ectopic integration of the HEG (plus the attached Rps3 sequence) into a mtDNA intergenic spacer region. This essentially generated a duplication of the *rps3* gene and thus the intron encoded version is now redundant and eventually was lost. The HEG associated with the *rps3* gene degenerated thus only leaving behind the now free standing *rps3* gene. The rapid degeneration of HEGs has been noted by Goddard and Burt (1999) as these elements are not selected for and thus mutations can quickly accumulate.

Reverse transcriptase (RT) activity may be another explanation for the relocation of the *rps3* gene. Reverse transcriptase can switch template from one transcript to another (referred to as template switching) during cDNA synthesis (Luo and Taylor 1990; Temin 1993; Zaphiropoulos 2002; Cocqet et al,. 2006), and this event can lead to the fusing of transcripts or transcript segments and RTs can promote the relocation/insertion of the newly made transcripts within genomes. However, this would require ectopic retrohoming, which might be a rather rare event. RT activity and template switching may explain the previous observation of a free standing *rps3* gene that is fused to a short *cox1* segment in *Phaeosphaeria nodorum* (Sethuraman et al., 2009a). The *rps3* encoded "intron" RNA may be reverse transcribed and the RT switched onto a cox1 transcript, thus a mosaic cDNA may be generated and relocated into a new location. The retrotransposition events, including ectopic integrations, may promote changes in the mtDNA gene landscape.

Finally with regards to the five fungi that showed free standing versions of "intron derived" rps3 ORFs, *M. brunnea* and *P. subalpina* belong to order Helotiales, and *B. maydis*, *L.*

maculans and *P. nodorum* belong to the order Pleosporales. These two orders are not closely related; thus the relocation event that gives rise to the free-standing forms of the mitochondrial rps3 ORF happened at least twice among the various fungal lineages.

During this study we also noted two anomalies. First for the *Penicillium solitum* mtDNA genome (JN696111) it was reported that no intron was located within the *rnl* gene (Eldarov et al., 2012). However, upon closer inspection it was noted that an intron was present at position L2499 and based on various translation program two coding regions were identified that when they are combined, they have the potential to encode the rps3 protein. This could be due to missannotation of the sequence by the authors and the fragmented Rps3 ORF might be a sequencing error or indeed an indication that the ORF is degenerating due to a frame shift mutation. As all *Pennicillium* species and related taxa so far have been reported to have functional versions of the intron encoded Rps3 ORF (Naruse et al., 1993; Yamamoto et al., 1995; Sun et al., 2011) one must suspect that indeed the original report on *P. solitum* is probably an error and the data need to be reexamined or the *rnl* gene resequenced.

The second unexpected finding was a sequence supposedly derived from *Microbacterium laevaniformans* (Actinobacteria) that was annotated as Rps3 (WP_005050872). This gene encodes a 342 amino acid polypeptides, and the amino acid sequence aligns with other fungal mL2449 intron encoded Rps3 sequences. One explanation is that this could be due to contamination of the bacterial culture with a filamentous ascomecetes, or some sequence mix up, or a rather unusual horizontal gene transfer event from a fungus to a bacteria. In bacterium, horizontal gene transfer (HGT) is widely recognized as the mechanism responsible for the widespread distribution of antibiotic resistance genes, gene clusters encoding biodegradative pathways, pathogenicity determinants, and also for speciation and sub-speciation in bacteria

(Cruz and Davies, 2000). HGT has also been reported between bacteria and eukaryotes with regards to signaling domain homologues (Ponting et al., 1999). So there is a possibility of HGT explaining the origin of the bacterial "fungal" like Rps3 protein.

Finally a recent paper by Gutiérrez et al. (2014) shows the value of an *rnl* landscape in facilitating mtDNA annotations. The authors of this paper characterized the mtDNA of *Colletotrichum lindemuthianum* (common bean anthracnose pathogen) and they appear to misannotate the *rnl* gene (Fig. 1, in Gutiérrez et al., 2014), showing the gene ending before the Rps3 ORF. Reexaming their data (Genbank accession number: KF953885) with blastn and the ORFfinder program showed that this fungus has an *rnl* gene that is interrupted by the mL2449 intron, the intron encoding the Rps3 protein.

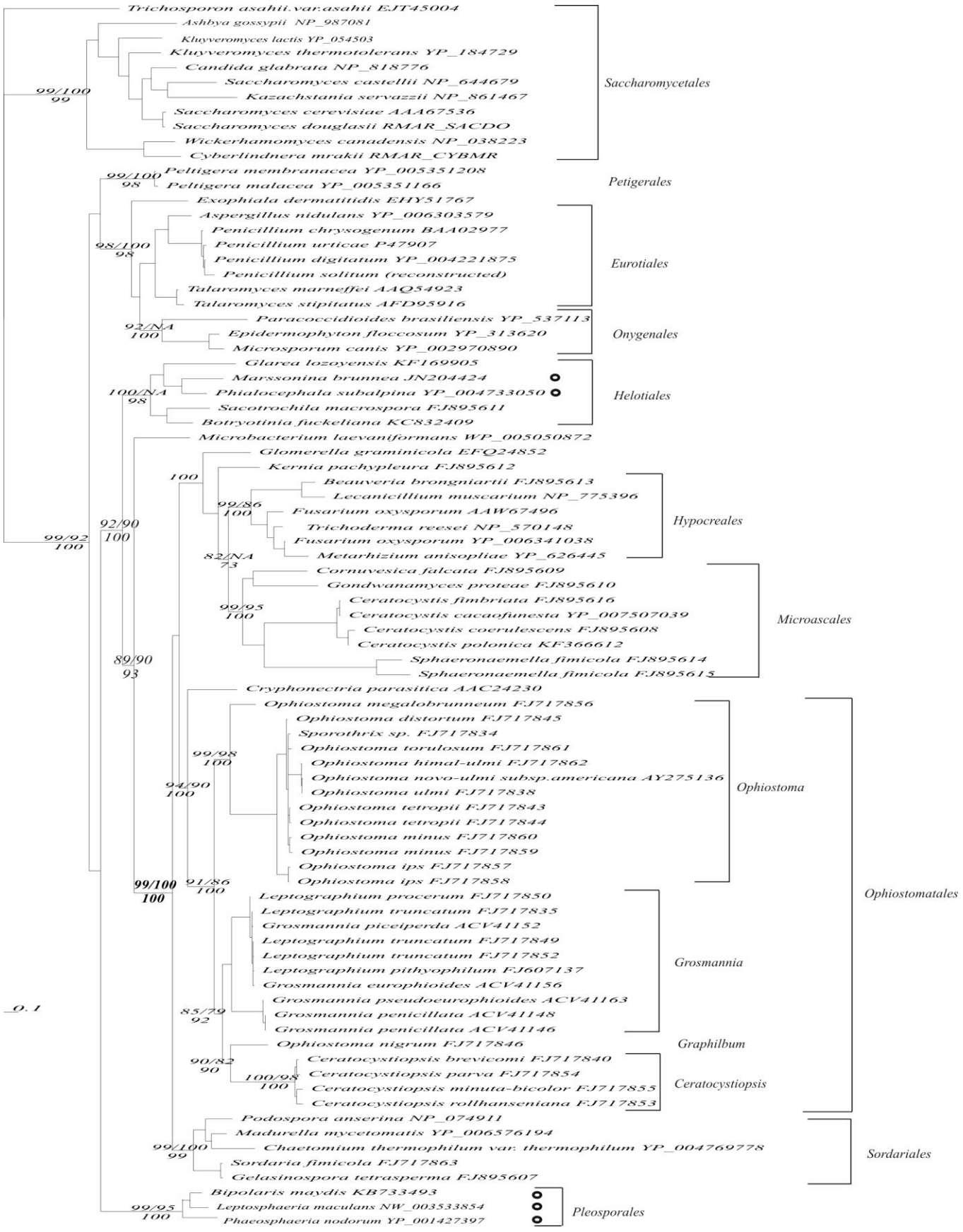

Trichosporon asahii.var.asahii EJT45004
Ashbya gossypii NP_987081
Kluyveromyces lactis YP_054503
Kluyveromyces thermotolerans YP_184729
Candida glabrata NP_818776
Saccharomyces castellii NP_644679
Kazachstania servazzii NP_861467
Saccharomyces cerevisiae AAA67536
Saccharomyces douglasii RMAR_SACDO
Wickerhamomyces canadensis NP_038223
Cyberlindnera mrakii RMAR_CYBMR
Saccharomycetales
99/100
99
Peltigera membranacea YP_005351208
Peltigera malacea YP_005351166
Petigerales
99/100
98
Exophiala dermatitidis EHY51767
Aspergillus nidulans YP_006303579
Penicillium chrysogenum BAA02977
Penicillium urticae P47907
Penicillium digitatum YP_004221875
Penicillium solitum (reconstructed)
Talaromyces marneffei AAQ54923
Talaromyces stipitatus AFD95916
Eurotiales
98/100
98
Paracoccidioides brasiliensis YP_537113
Epidermophyton floccosum YP_313620
Microsporum canis YP_002970890
Onygenales
92/NA
100
Glarea lozoyensis KF169905
Marssonina brunnea JN204424
Phialocephala subalpina YP_004733050
Sacotrochila macrospora FJ895611
Botryotinia fuckeliana KC832409
Helotiales
100/NA
98
Microbacterium laevaniformans WP_005050872
Glomerella graminicola EFQ24852
Kernia pachypleura FJ895612
Beauveria brongniartii FJ895613
Lecanicillium muscarium NP_775396
Fusarium oxysporum AAW67496
Trichoderma reesei NP_570148
Fusarium oxysporum YP_006341038
Metarhizium anisopliae YP_626445
Hypocreales
100
99/86
100
92/90
100
99/92
100
82/NA
73
Cornuvesica falcata FJ895609
Gondwanamyces proteae FJ895610
Ceratocystis fimbriata FJ895616
Ceratocystis cacaofunesta YP_007507039
Ceratocystis coerulescens FJ895608
Ceratocystis polonica KF366612
Sphaeronaemella fimicola FJ895614
Sphaeronaemella fimicola FJ895615
Microascales
99/95
100
89/90
93
Cryphonectria parasitica AAC24230
Ophiostoma megalobrunneum FJ717856
Ophiostoma distortum FJ717845
Sporothrix sp. FJ717834
Ophiostoma torulosum FJ717861
Ophiostoma himal-ulmi FJ717862
Ophiostoma novo-ulmi subsp.americana AY275136
Ophiostoma ulmi FJ717838
Ophiostoma tetropii FJ717843
Ophiostoma tetropii FJ717844
Ophiostoma minus FJ717860
Ophiostoma minus FJ717859
Ophiostoma ips FJ717857
Ophiostoma ips FJ717858
Ophiostoma
99/98
100
94/90
100
99/100
100
91/86
100
Leptographium procerum FJ717850
Leptographium truncatum FJ717835
Grosmannia piceiperda ACV41152
Leptographium truncatum FJ717849
Leptographium truncatum FJ717852
Leptographium pithyophilum FJ607137
Grosmannia europhioides ACV41156
Grosmannia pseudoeurophioides ACV41163
Grosmannia penicillata ACV41148
Grosmannia penicillata ACV41146
Grosmannia
85/79
92
Ophiostoma nigrum FJ717846
Graphilbum
Ceratocystiopsis brevicomi FJ717840
Ceratocystiopsis parva FJ717854
Ceratocystiopsis minuta-bicolor FJ717855
Ceratocystiopsis rollhanseniana FJ717853
Ceratocystiopsis
90/82
90
100/98
100
Ophiostomatales
Podospora anserina NP_074911
Madurella mycetomatis YP_006576194
Chaetomium thermophilum var. thermophilum YP_004769778
Sordaria fimicola FJ717863
Gelasinospora tetrasperma FJ895607
Sordariales
99/100
99
Bipolaris maydis KB733493
Leptosphaeria maculans NW_003533854
Phaeosphaeria nodorum YP_001427397
Pleosporales
99/95
100
0.1

Figure 4.1 Rps3 phylogenetic trees generated by the MrBayes program. Five free-standing *rps3* genes were found in *Phaeosphaeria nodorum, Leptosphaeria maculans, Bipolaris maydis, Marssonina brunnea,* and *Phialocephala subalpine*. This may prove that the appearance of the free standing *rps3* is not a single event, this event happened at least twice during the evolution of the filamentous ascomycetes. Black circles indicate fungi that have a free-standing *rps3*gene.

Chapter 5: NADH dehydrogenase subunit 4 (*nad4*) gene in a strain of *Pesotum* sp.

5.1 Introduction

Some of the subunits of the NADH dehydrogenase (i.e. *nad1, nad2, nad3, nad4, nad4L, nad5* and *nad6*) are encoded within the mitochondrial genome. So far most of the *nad4* genes characterized in fungi are not intron-rich; however, group II introns have been found in the *nad4* and *nad7* genes in flowering plants (Pruchner et al., 2001).

Characterizing ribozymes and HEs is important as these elements have applications in biotechnology (Hausner 2003; Hafez and Hausner, 2012). These elements also contribute towards the size of fungal mtDNA and may contribute towards mtDNA rearrangements and polymorphism (Charter et al., 1996; Belcour et al., 1997; Salvo et al., 1998; Hamari et al., 1999; Gobbi et al., 2003; Hausner 2012). Currently little information is available with regards to the *nad4* gene and its intron complement for species of *Pesotum*. Many *Pesotum* species are of economic concern as they cause blue-stain in timber (Hausner et al., 2005; and see Chapter 1). The goal of this study is to find potential mobile elements (introns) inserted in *Pesotum spp*. *nad4* genes, along with the characterization of the intron and HEG via *in silico* methods.

5.2 Material and methods

DNA extraction

DNA was purified from fungal material as described in Chapter 2 page 19.

RNA extraction and purification

RNA was extracted from fungal mycelium using the RNeasy kit (Qiagen, Maryland) following the manufacturer's instructions. First, mycelium was generated by growing the fungus in liquid PYG medium for about 5 days at 20 °C. Thereafter the mycelium was harvested by vacuum filtration. The mycelium was transferred into an empty Perti plate, and 200 µl of RNA later (Qiagen) was added to the harvested mycelium before being stored store at -60 °C. Frozen fungal tissue was transferred into a pre-cooled mortar and ground up with a pestle in liquid nitrogen to generate a fine white powder. RNA was extracted from the ground up powder as described in the RNeasy extraction kit; essentially the procedure followed the plant RNA extraction protocol.

Complementary (c) DNA synthesis from the RNA was performed using the ThermoScript RT–PCR system (Invitrogen) following the manufacturer's instructions. A total reaction volume of 12 µl was set up in a 200 µl microcentrifuge tube which contained 1 µl of 10 µM ND4-R (reverse) primer, 1 µl (~100 µg) purified template RNA, 2 µl dNTP Mix (10 mM), and 8µl RNase-free water. The mixture was denatured by incubating at 65 °C for 5 min. To this mixture was added: 4 µl cDNA synthesis buffer (5x), 1 µl Dithiothreitol (DTT; 0.1 M), 1 µl RNaseOUT (40U/ µl), 1 µl RNase-free water, and 1 µl ThermoScript Reverse Transcript (15U/

µl). This reaction was incubated at 52.9 °C for 55 min, and followed by incubation at 85 °C for 5 min. Finally, 1 µl of RNase H was added to degrade the RNA template and the reaction was incubated at 37 °C for 20 min. The cDNA was stored at -60 °C. In order to amplify processed *nad4* transcripts (mRNA), the Platinum Taq DNA polymerase kit (Invitrogen) was used for the standard PCR reaction; and a 50 µl reaction mix contained the following: 5 µl PCR buffer (10x) minus Mg^{+2}, 1.5 µl $MgCl_2$ (50 mM), 1 µl dNTP mix (10 mM), 1 µl of both ND4-F and ND4-R primers (10 µM), 0.4 µl Platium Taq (5U/ µl), 2 µl of the cDNA reaction, and 38.1 of µl DEPC-treated water. The reaction was first incubated at 94 °C for 2 min, this was followed by 30 cycles of PCR amplification (PCR condition: with the lid temperature at 105 °C, an initial denaturation at 94 °C for 2 min was followed by 25 cycles of denaturing (1 min at 93 °C), annealing (1 min 30 sec at 52.1 °C), and extension (4 min 30 sec at 70°C). The samples were incubated for a final extension at 72 °C for 5 min, then placed "on hold" at 10 °C.

Cloning of PCR products

The n*ad4* gene was amplified using the ND4-F (forward) and ND4-R (reverse) primers (Table 2.1). The *nad4* derived PCR products were cloned with the TOPO TA cloning kit (Invitrogen), as described in Chapter 2, page 21.

Restriction enzyme assay

Restriction enzymes were used to confirm if the colonies selected contained the recombinant plasmid. *Eco*R1 has two cleavage sites in the pCR™4-TOPO® plasmid vector; the restriction

sites are upstream and downstream of where the PCR product is expected to be inserted. A successful digest, indicating a recombinant, should yield two fragments, one is the size of the plasmid vector around 4 kb, and the other fragment should be 2.8 kb, i.e. the size of the *nad4* derived PCR product.

DNA Sequencing

Details are described in Chapter 2 page 23. Initially sequences where obtained by using primers to amplify the *nad4* gene segment along with vector based primers; thereafter primers were designed as needed to sequence the *nad4* segment in both directions.

Snap Cool

The *nad4* project encountered some PCR and DNA sequencing problems; the designed primers for sequencing did not amplify products for the *nad4* PCR template or on the cloned version of the *nad4* template. The absence of amplicons may have been due to DNA secondary structures, such as the formation of "hair- pin" structures or the present of GC- rich regions in the template DNA. The "snap cool" process was used to potentially resolve this problem. The PCR product was denatured at 98 °C for 10 min, and immediately after the denaturing step the tube was placed into a – 80 °C 95% ethanol bath. This step potentially prevents the reformation of double-stranded DNA and discourages the formation of secondary structures.

Bioinformatics

The details with regards to analyzing sequences, i.e. building sequence contigs, sequence alignments and phylogenetic analysis, are described in Chapter 2 page 24.

5.3 Results and discussion

In order to study the possible presence of intron insertions in the *nad4* gene in members of the genus *Pesotum*, 8 strains were selected from the culture collection (Table 5.1). The *nad4* gene was amplified for all strains from whole cell DNA by using the ND4-F and ND4-R primers. PCR results showed that strain WIN(M)1630 appeared to generate a PCR product around 2.8 kb while other strains only generated PCR products with a size around 0.85 kb. Based on published *nad4* gene data, a 0.85 kb product was expect if no insertions are present within the *nad4* gene. These results indicate that there is at least one potential element inserted into the strain WIN(M)1630 *nad4* gene. The 2.8 kb PCR product was cloned into a Topo vector and initial sequences were obtained with the vector based T3 and T7 primers. A cDNA assay was performed to confirm the presence of an intron; however attempts to obtain cDNA were not successful. Based on DNA sequence analysis the ND4-F/ND4-R PCR product was 2724 nt long.

Sequencing results showed that the *nad4* gene contains an intron and the intron encodes a double motif LAGLIDADG type ORF. The intron appears to belong to the group I C2 type based on structural features (Fig 5.0). The group I intron in the *nad4* gene is 1.4 kb. The intron/exon junctions were identified based on sequence homology, i.e. the WIN(M)1630 *nad4*

gene was aligned with other *nad4* gene (that lack introns) extracted from GenBank.

The mtDNA *nad4* gene has rarely been noted among the fungi to contain introns. A set of *nad4* genes was selected from GenBank to examine the potential occurrence of intron insertions; only few *nad4* genes with intron insertions were found. For example, the *Neurospora crassa nad4* gene (GenBank accession: KC683708) has one intron, and the ORF finder program identified a double motif LAGLIDADG type ORF within the intron sequence, but the RNA weasel program could not identify the intron type; this intron maybe is related to the IC2 intron in *Pesotum* but it might be highly degenerated. It is worth noting that some genes such as *cox1* or rRNA genes (Férandon et al., 2010; Nikoh and Fukatsu 2001), have been invaded by a multitude of introns but other genes (such as *nad4*) appear to be "intron-poor". The *nad4* gene is highly conserved at the nucleotide level so it is surprising that this gene appears to contain few introns. Typically mobile introns invade highly conserved sequences (Belfort et al., 2002). A possible reason for a lack of introns in some genes might be that the insertion of mobile elements interferes with gene expression.

Table 5.1 Strains used in the *nad4* project. N/A means that no PCR products were obtained.

Strains	**WIN Numbers**	**PCR product size (kb)**
Pesotum	1384	0.85
Pesotum	1387	0.85
Pesotum	1388	N/A
Pesotum	1390	N/A
Pesotum	1602	N/A
Pesotum	1603	N/A
Pesotum	1629	N/A
Pesotum	1630	2.8

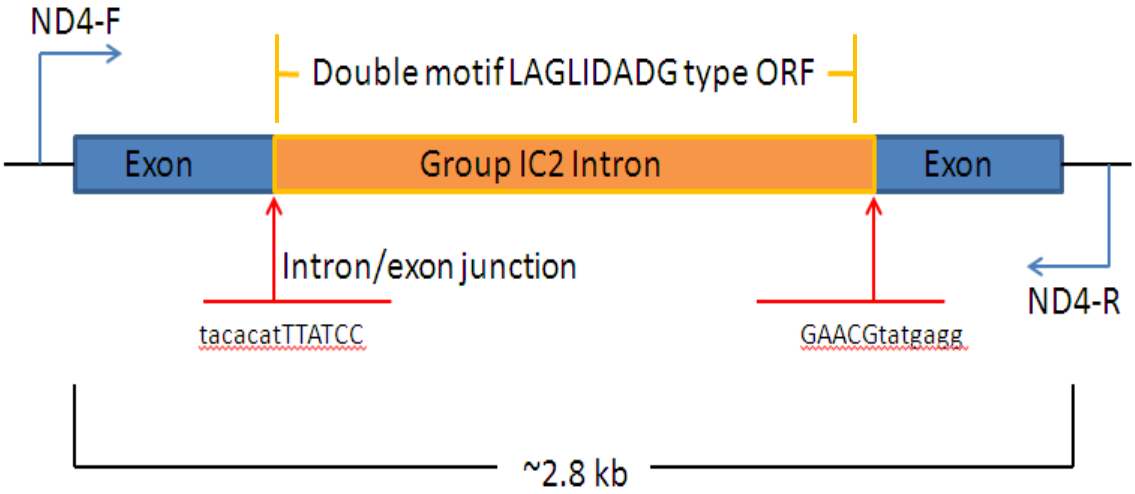

ND4-F
Double motif LAGLIDADG type ORF
Exon
Group IC2 Intron
Exon
Intron/exon junction
ND4-R
tacacatTTATCC
GAACGtatgagg
~2.8 kb

Figure 5.0. A schematic for the *Pesotum spp*. WIN(M)1630 *nad4* gene. The n*ad4* gene was amplified using the ND4-F and ND4-R primers. The n*ad4* gene is estimated to be 2724 nt with a group IC2 intron insertion of around 1.4 kb. A sequence encoding a double motif LAGLIDADG type protein is embedded within the intron, this HE encoding sequence appears to be fused in frame with the upstream exon nad4 coding sequence. The exon/intron junction sequences are indicated and they were identified by analyzing a *nad4* gene alignment, see text for more details. Blue boxes indicate exons; the orange box indicates the intron. Exon sequences are presented in lowercase letters, intron sequences are in uppercase letters.

A Group IC2 intron in the *nad4* gene

The intron found in WIN(M)1630 *nad4* gene is around 1.4 kb in length; and this intron contained the characteristic features of a group I C2 type intron (Burke et al., 1987; Cech, 1988). The online program RNAweasel was used to identify the intron core (P3, P4, P6, P7, and P8) components, the initial intron secondary structure was generated by the mFOLD program (Zuker 2003; http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form). Based on previous studies and structure presented by Cech (1988) and Michel and Westhof (1990) the *nad4* intron RNA fold was generated (Fig 5.1). The putative ORF was identified by the online program ORF finder, the intron ORF encodes a double motif LAGLIDADG type HE and the beginning of the HE ORF is fused with upstream (5') exon sequence. In addition the ORF sequence overlaps with the intron core structure, the overlap of this IC2 intron's open reading frame with the intron core sequences suggest coevolution of these sequences. One explanation is that splicing of the intron is important to allow for the nad4 gene to be expressed and maintenance of the HE ORF sequence is essential for intron mobility. Therefore the intron and HE ORF sequences have to co-evolve in order for both functions (splicing and mobility) to be preserved.

In group I introns the ORFs are usually found in the looped regions of the folded intron in order to minimize interruption of the RNA secondary and tertiary structures (Edgell et al., 2011). This feature also allows ORFs to insert into different types of introns (group I and group II) and even if the ORF degenerates over time it would not affect the intron RNA structure. The intron observed in this study is unusual as some components of the ORF extend into the intron core (stem) sequences. This fits into a concept proposed by Edgell et al. (2011), a process termed "core creep", whereby a HEG invades an intron and over time it incorporates the intron sequences into its ORF until it is fused to the upstream exon. The HEG would benefit from this

arrangement as it can enhance its expression by being co-translated with the host gene protein. One explanation is that a protease can release the HE protein from the precursor fusion protein.

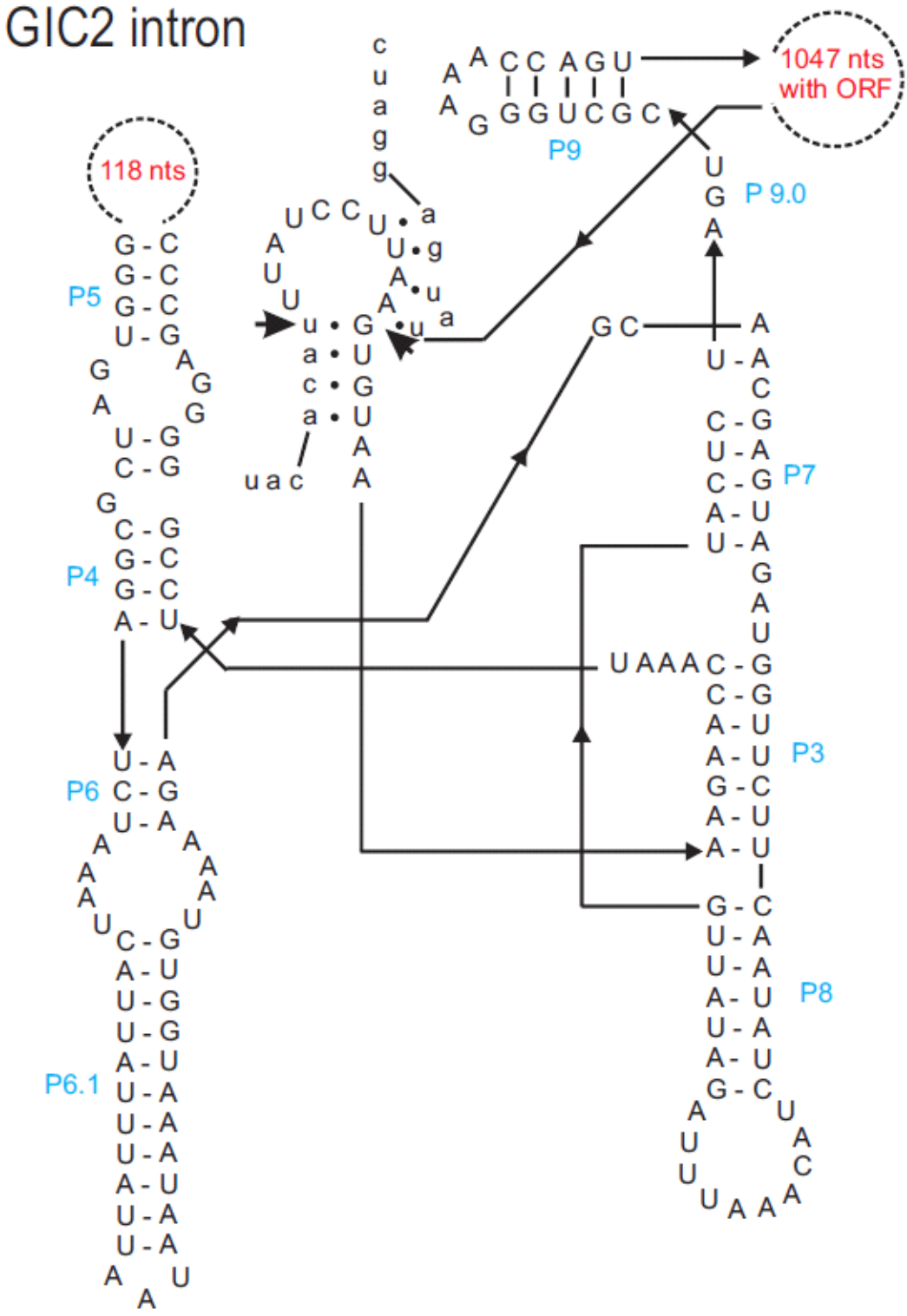

GIC2 intron
1047 nts with ORF
118 nts
P9
P 9.0
P5
P4
P7
P3
P8
P6
P6.1

Figure 5.1. GIC2 intron secondary structure (the IC2 type intron found in *Pesotum spp*. WIN(M)1630 *nad4* gene). Usually IC2 type introns lack the P2 region and this intron structure fits the classic IC2 model (Michel and Westhof 1990). Parts of the intron are labeled with blue letter (P denotes paired regions), intron nucleotides (nt) are presented as uppercase letters, flanking exon sequences are in lowercase letters. The P9 loop contains part of the ORF sequence but the N terminal coding component of the LAGLIDADG coding sequence extends towards to the 5' end of the intron. The large arrow head indicates the intron-exon junctions. A LAGLIDADG open reading frame overlap with most of the intron structure and details are discussed in the text.

Phylogenetic study

The *Pesotum spp*. strain WIN(M)1630 *nad4* intron ORF amino acid sequence was used as a query to search for related sequences in GenBank by using the BLASTp program. This data base search identified 21 related endonuclease sequences. These were compiled and aligned by the online program COBALT (http://www.ncbi.nlm.nih.gov/tools/cobalt/; Papadopoulos and Agarwala 2007) and the alignment was edited with GeneDoc. The purpose of this study was to examine the phylogeny of intron associated LAGLIDADG type ORFs that are related to the *nad4* intron encoded ORFs. In general, maybe due to poor sequence conservation and high rates of divergence (Stoddard 2005), all phylogenetic methods applied to this study yielded tree topologies with the deeper nodes being poorly supported (Fig. 5.2). In general only the central portion of the intron ORF sequence (~40 amino acids) had 90% similarity with other selected LAGLIDADG type ORFs. The phylogenetic study shows that the WIN(M)1630 *nad4* intron sequence is not closely related to other previously described intron ORFs found in *nad4* genes or other mtDNA genes in other organisms. *Trichoderma reesei* was chosen as the outgroup for this phylogenetic study as it is a free standing type HEG and this fungus is distantly related to *Pesotum*.

Although most of the nodes in the phylogenetic tree were poorly supported, two nodes appear to define clades that show moderate to high levels of bootstrap support. For example a set of *nad4* intron encoded ORFs found in *Podospora anserina, Neurospora crassa,* and *Ceratocystis cacaofunesta* etc; form a clade with moderate support (85 % based on ML analysis). Another clade was noted where *nad3* intron encoded LAGLIDADG type ORFs in *Fusarium graminearum, Podospora anserina, and Madurella mycetomatis* are grouped together by a node with bootstrap support at 100 % level based on ML analysis. However the *Pesotum* spp.

WIN(M)1630 *nad4* intron ORF fails to group with either of the two clades described above.

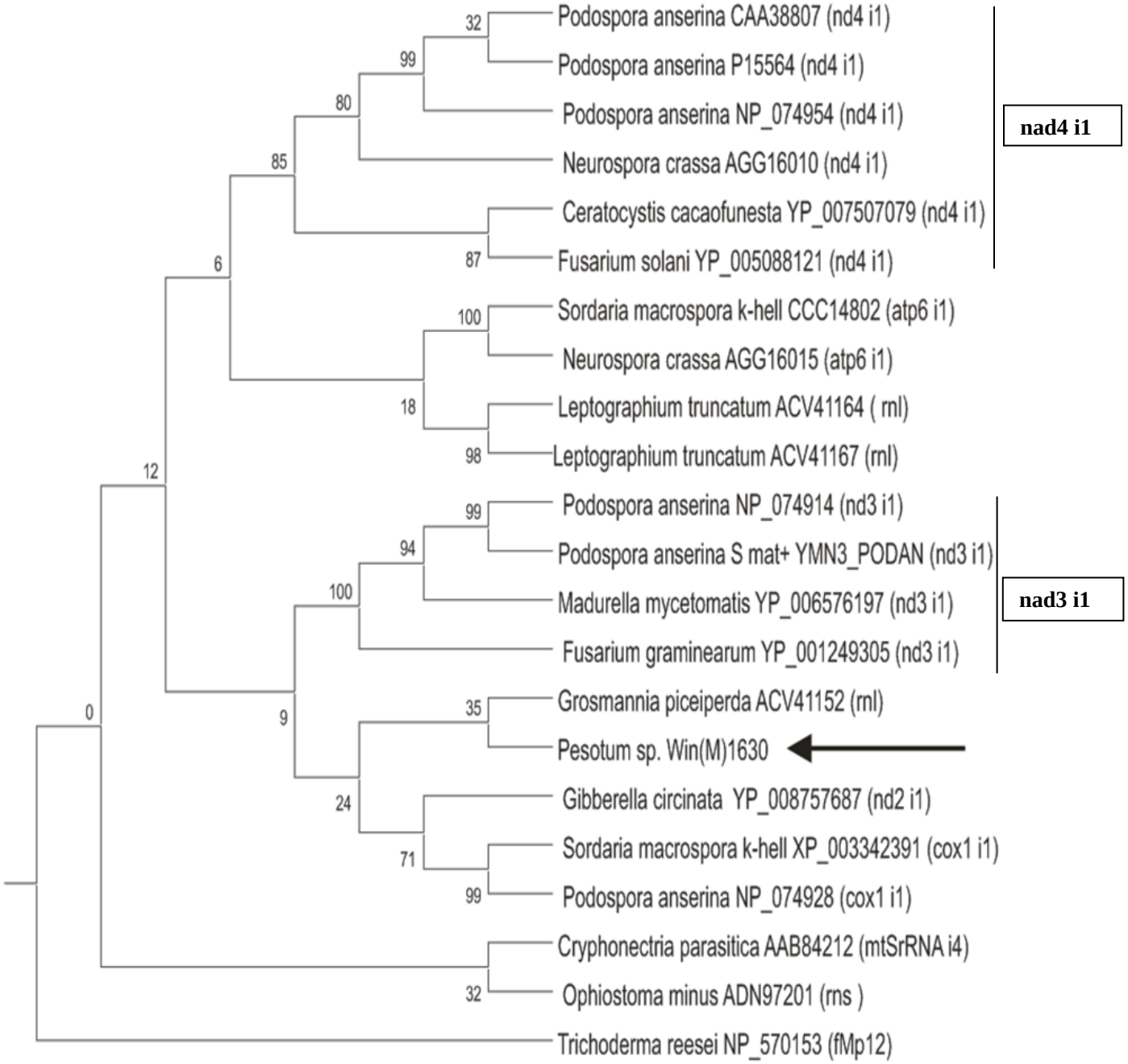


Figure 5.2. The *nad4* intron encoded ORF phylogenetic tree generated by the MEGA5 program (ML analysis). Two clades appear to receive moderate (nad4 i1 ORF clade at the top of the tree) or strong support (nad3 i1 ORF clade). Tree topology shown is based on a ML majority rule consensus tree obtained by analysis of 1000 pseudo-replicates generated by the bootstrap option within the MEGA program. Big arrow indicates the *Pesotum* sp. WIN (M) 1630 *nad4* intron sequence.

From the BLASTp results and the phylogenetic tree it appears that that double motif

LAGLIDADG homing endonuclease can insert into a wide variety of genes (*nad2*, *nad3*, *nad4*, *atp6*, *rns*, *rnl* etc.). It has been noted that HEases that contain the conserved LAGLIDADG motif are the most widespread HE family within the organelle genomes; they occur in several different types of mobile elements (i.e. group I and group II intron etc). This suggests that compared to other HE types (GIY YIG, HNH etc.) the LAGLIDADG HEGs are more versatile in adapting to new genomic niches (Gimble 2000).

The members of the LAGLIDADG family can contain either one or two LAGLIDADG motifs in their ORF. The version that has only the single motif, binds to target DNA sequences and assembles into homodimers, and requires a palindromic or near-palindromic symmetry in the DNA recognition site. The LAGLIDADG HEs with double motifs can act as monomers, as they process a pair of structurally similar nuclease domains in a single peptide chain, thus symmetrical recognition sites are not required for the double motif LAGLIDADG HE (Stoddard 2005). This makes double motif enzymes more efficient in recognizing a wide variety of different target sites. Previous studies suggest that the successful spread of LAGLIDADG HEGs into different introns and genes was driven by a gene duplication of a single motif LAGLIDADG type ORF and a fusion event generating the double motif type LAGLIDADG type ORFs. This was followed by the rapid evolution of the DNA-binding domain in the double motif version allowing the double motifs to be more flexible in their association with the DNA template and thus tolerating more sequence divergence (Lucas et al., 2001; Chevalier et al., 2003). The wide "host range" of HEGs reflects how they recognize a larger number of target sequences and thus have invaded a large number of genes.

According to Haugen and Bhattacharya (2004) HEGs can move from rDNA coding genes to protein coding genes and vice versa. Also HEGs or mobile introns can move by vertical or

horizontal transmission, which is important for their long term survival in populations (Goddard and Burt 1999). The insertion of mobile elements typically requires specific recognition sites within the host genome, but in many instances some HE can tolerate a few nucleotide changes at the HE recognition site in order to avoid elimination due to drift (Gimble 2000; Chevalier et al., 2003).

In the future other attempts should be made to characterize the *Pesotum nad4* intron by RT-PCR, thus providing physical evidence for the intron/exon junctions. Also it would be desirable to generate a codon optimized version of the *Pesotum nad4*i ORF and over-express the protein in *E. coli*. This would allow for characterizing the homing endonuclease protein with regards to its cleavage activity and potentially map its cleavage site (Sethuraman et al. 2009a,b).

.

Conclusion

Fungal mitochondria have been shown to be a home for many mobile elements (such as group I and group II introns) thus generating size variations and mtDNA polymorphisms (Charter et al., 1996; Belcour et al., 1997; Salvo et al., 1998; Hamari et al., 1999; Gobbi et al., 2003; Hausner 2003; Gibb and Hausner 2003). Screening of different regions within the *rnl* gene of *O. ulmi* and related taxa identified five group I introns inserted in different regions of the *rnl* gene. Comparative sequence analysis suggests that with regards to the *O. ulmi* complex, five possible insertions can be present but only one is always present. The mt-*rnl* U11 intron (i.e. mL2449), which encodes the ribosomal protein S3, *rps3*, appears to be essential, while introns encoding HEase type ORFs appear to be optional. The mL2449 intron encodes a gene (rps3) that is required for the assembly of the 30 S ribosomal subunit and presumed essential; thus the intron has ensured its survival by encoding an essential host protein.

An *rnl* intron landscape of ascomycetous and basidiomyctous fungi was generated based on the *rnl* gene survey and data pulled from Genbank. The *rnl* sequences were compiled and by comparative sequence analysis a total number of 23 possible intron insertion sites were found in the *rnl* gene of ascomycetous and basidiomycetous fungi. Group I introns appear to dominate this landscape with only two group II introns having been noted. In addition most introns appear to encode either LAGLIDADG or GIY type ORFs, ORFs assumed to encode either homing endonucleases that can promote intron mobility or maturases that assist in intron RNA splicing. However, some ORFs appear to be degenerate (due to the presence of premature stop codons) and other ORFs can only be presumed to encode functional proteins.

Among the 23 introns inserted in the *rnl* gene of ascomycetous and basidiomycetous fungi, the "*rps3* intron" (mL2449) was studied in more detail. Among most ascomycetes the *rps3*

gene appears to be located inside the mL2449 intron. However, a previous study showed that in one ascomycetes fungus, *Phaeosphaeria nodorum*, the loss of the mL2449 intron in the large ribosomal subunit gene (*rnl*) was associated with the appearance of a free standing version of the *rps3* gene (Sethuraman et al., 2009a). In this study, based on data available from Genbank, additional examples were found of freestanding rps3 genes; these were found in *Leptosphaeria maculans, Bipolaris maydis, Marssonina brunnea,* and *Phialocephala subalpine*. A phylogenetic study showed that these fungi (with free standing forms of *rps3*) group into two distinct clades. This indicates the free-standing forms of *rps3* evolved several times (or at least twice) during the evolution of the filamentous ascomycetes due to an as of yet unknown recombination event. The recombination event might involve a HEG transduction event (*rps3* gene moved along by a linked HEG), an RNA mediated mobility event (reverse splicing), or the rps3 protein has the ability to nick or cut DNA thus acting like a HEase. Sethuraman et al. (2009a) presented some evidence that based on observations in the literature that Rps3 might be a moonlighting protein with endonuclease activity involved in DNA repair.

In the *nad4* gene study involving the *Pesotum* WIN(M)1630 strain, a group I C2 type intron was uncovered with a double motif LAGLIDADG type ORF. The phylogenetic study shows that the WIN(M)1630 *nad4* intron sequence is not closely related to other previously described intron ORFs found in *nad4* genes, suggesting that this ORF originated from a different gene. Frequently introns located in similar positions within the same gene share a common ancestor but there are many exceptions as these elements can invade ectopic sites in order to maintain themselves in populations (Goddard and Burt, 1999).

Overall, this work demonstrates that mobile elements help to modify and shape the fungal mitochondrial genome. Introns and HEGs are capable of inserting into conserved regions within

different species allowing for both vertical and horizontal transmission, the flexibility of HEGs to move between different genes allows for these elements to escape the postulated prediction that HEGs, due to lack of selection, would eventually become extinct (Goddard and Burt, 1999).

Future directions

This study uncovered some new introns and HEGs, only a few of them have been studied in detail via comparative sequence analysis and none have yet been physically characterized. Therefore, some issues that should be addressed next are:

1) Using next generation sequencing methods to get the complete mitochondrial *rnl* sequence of several species and strains that make up the *O. ulmi* species complex. Comparing closely related mtDNA genomes may show more clearly how mobile introns can reshape or alter mitochondrial genomes.

2) Many of the HEGs found herein should be studied in detail in terms of showing if they encode active endonucleases. This would involve over expression of the HEGs and their encoded proteins in systems (such as *E. coli*) that allow for over expression and purification of foreign proteins.

3) Further characterization of the *Pesotum nad4* intron by RT-PCR, thus providing physical evidence for the intron/exon junctions is needed. Also it would be desirable to generate a codon optimized version of the *Pesotum nad4*i ORF and over-express the protein in *E. coli*. This would allow for characterization of the homing endonuclease protein with regards to its cleavage activity and potential mapping of its cleavage site.

Significance

Mitochondria are essential organelles in most eukaryotes; therefore factors that affect mtDNA integrity can have a profound effect on the organism. Mitochondrial DNA instabilities have been associated with mitochondrial dysfunction and respiratory defects (reviewed in Bertrand 2000). Mobile elements such as plasmids and mobile introns have been noted to be involved in causing mitochondrial defects; this has relevance in understanding phenomena such as senescence in some fungi (*Podospora anserina*) and hypovirulence (*Cryphonectria parasitica*), and possible mechanisms involved in horizontal gene transfer (Osiewacz et al., 1989; Kück et al., 1985; Bertrand, 2000; Baidyaroy et al., 2000, 2011a,b; Monteiro-Vitorello et al., 2009; Hausner 2003, 2012). In addition mobile introns, which are composed of two components, a ribozyme and a coding region that can promote mobility, are a source of novel enzymes that have applications in biotechnology (Stoddard 2005, 2011; Hafez and Hausner, 2012; Belfort and Bonocora, 2014).

References

Abu-Amero, S. N., Charter, N. W., Buck, K. W., & Brasier, C. M. (1995). Nucleotide-sequence analysis indicates that a DNA plasmid in a diseased isolate of *Ophiostoma novo-ulmi* is derived by recombination between two long repeat sequences in the mitochondrial large subunit ribosomal RNA gene. Current Genetics, 28(1), 54-59.

Adams, K. L., Palmer, J. D. (2003) Evolution of mitochondrial gene content: gene loss and transfer to nucleus. Molecular Phylogenetics and Evolution, 29(3), 380- 395.

Akins, R. A., & Lambowitz, A. M. (1987). A protein required for splicing group I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell, 50(3), 331-345.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403-410.

Baidyaroy, D., Hausner, G., Hafez, M., Michel, F., Fulbright, D., & 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.

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

Baidyaroy, D., Huber, D. H., Fulbright, D. W., & Bertrand, H. (2000). Transmissible mitochondrial hypovirulence in a natural population of *Cryphonectria parasitica*. Molecular Plant-Microbe Interactions, 13(1), 88-95.

Bates, M. R., Buck, K. W., & Brasier, C. M. (1993). Molecular relationships between *Ophiostoma ulmi* and the NAN and EAN races of *O. novo-ulmi* determined by restriction fragment length polymorphisms of nuclear DNA. Mycological Research, 97(4), 449-455.

Beck, N., & Lang, B. F. (2010). MFannot, organelle genome annotation websever.

Belcour, L., Rossignol, M., Koll, F., Sellem, C. H., & 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(4), 308-317.

Belfort, M., & Roberts, R. J. (1997). Homing endonucleases: keeping the house in order. Nucleic Acids Research, 25(17), 3379-3388.

Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B., & Lambowitz, A. M. (2002). Mobile introns: pathways and proteins. Mobile DNA II, 761-783.

Belfort, M., & Bonocora, R. P. (2014). Homing Endonucleases: From Genetic Anomalies to Programmable Genomic Clippers. In Homing Endonucleases(pp. 1-26). Humana Press.

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(1), 397-422.

Bertrand, H., Bridge, P., Collins, R. A., Garriga, G., & Lambowitz, A. M. (1982). RNA splicing in *Neurospora* mitochondria. Characterization of new nuclear mutants with defects in splicing the mitochondrial large rRNA. Cell,29(2), 517-526.

Brasier, C. M., & Mehrotra, M. D. (1995). *Ophiostoma himal-ulmi* sp. *nov*., a new species of Dutch elm disease fungus endemic to the Himalayas. Mycological Research, 99(2), 205-215.

Brasier, C. M. (1979). Dual origin of recent Dutch elm disease outbreaks in Europe. Nature, 281, 78 – 80.

Borkhardt, B., & Olson, L. W. (1983). Paternal inheritance of the mitochondrial DNA in interspecific crosses of the aquatic fungus *Allomyces*. Current genetics, 7(5), 403-404.

Bonen, L., & Vogel, J. (2001). The ins and outs of group II introns. Trends in Genetics, 17(6), 322-331.

Bullerwell, C. E., Leigh, J., Seif, E., Longcore, J. E., & Lang, B. F. (2003a). Evolution of the fungi and their mitochondrial genomes. Applied Mycology and Biotechnology, 3, 133-159.

Bullerwell, C. E., Forget, L., & Lang, B. F. (2003b). Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. Nucleic Acids Research, 31(6), 1614-1623.

Bullerwell, C. E., Burger, G., & Lang, B. F. (2000). A novel motif for identifying *rps3* homologs in fungal mitochondrial genomes. Trends in Biochemical Sciences, 25(8), 363-365.

Butow, R. A., Perlman, P. S., & Grossman, L. I. (1985). The unusual *varl* gene of yeast mitochondrial DNA. Science, 228(4707), 1496-1501.

Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J.W., & Tabak, H. F. (1987). Structural conventions for group I introns. Nucleic Acids Research, 15(18), 7217-7221.

Burke, J. M., & Rajbhandary, U. L. (1982). Intron within the large rRNA gene of *N. crassa* mitochondria: a long open reading frame and a consensus sequence possibly important in splicing. Cell, 31(3), 509-520.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., Pande, N., Shang, Z., Yu, N., & Gutell, R. R. (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(1), 2.

Cech, T. R. (1988). Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene, 73(2), 259-271.

Cech, T. R., Damberger, S. H., & Gutell, R. R. (1994). Representation of the secondary and tertiary structure of group I introns. Nature Structural & Molecular Biology, 1(5), 273-280.

Charter, N. W., Buck, K. W., & Brasier, C. M. (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(3), 368-372.

Chevalier, B. S., & Stoddard, B. L. (2001). Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Research, 29(18), 3757-3774.

Chevalier, B., Monnat Jr, R. J., & Stoddard, B. L. (2005). The LAGLIDADG homing endonuclease family. In Homing endonucleases and inteins (pp. 33-47). Springer Berlin Heidelberg.

Chevalier, B., Turmel, M., Lemieux, C., Monnat Jr, R. J., & Stoddard, B. L. (2003). Flexible DNA Target Site Recognition by Divergent Homing Endonuclease Isoschizomers *I-CreI* and *I-MsoI*. Journal of Molecular Biology, 329(2), 253-269.

Cocquet, J., Chong, A., Zhang, G., & Veitia, R. A. (2006). Reverse transcriptase template switching and false alternative transcripts. Genomics, 88(1), 127-131.

de la Cruz, F., & Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. Trends in Microbiology, 8(3), 128-133.

Cummings, D. J., Michel, F., & McNall, K. L. (1989a). 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(5-6), 381-406.

Cummings, D. J., Domenico, J. M., & Nelson, J. (1989b). DNA sequence and secondary structures of the large subunit rRNA coding regions and its two class I introns of mitochondrial DNA from *Podospora anserina*. Journal of Molecular Evolution, 28(3), 242-255.

Cummings, D. J., McNally, K. L., Domenico, J. M., & Matsuura, E. T. (1990). The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. Current Genetics, 17(5), 375-402.

Dalgaard, J. Z., Klar, A. J., Moser, M. J., Holley, W. R., Chatterjee, A., & Mian, I. S. (1997). Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. Nucleic Acids Research, *25*(22), 4626-4638.

Dai, L., Toor, N., Olson, R., Keeping, A., & Zimmerly, S. (2003). Database for mobile group II introns. Nucleic Acids Research, 31(1), 424-426.

Duò, A., Bruggmann, R., Zoller, S., Bernt, M., & Grünig, C. R. (2012). Mitochondrial genome evolution in species belonging to the *Phialocephala fortinii* sl-*Acephala applanata* species complex. BMC Genomics, 13(1), 166.

Dujon, B. (1989). Group I introns as mobile genetic elements: facts and mechanistic speculations – a review. Gene, 82(1), 91–114.

Dujon, B., Belfort, M., Butow, R. A., Jacq, C., Lemieux, C., Perlman, P.S. & Vogt, V. M. (1989). Mobile introns: definition of terms and recommended nomenclature. Gene, 82(1), 115–118.

Dujon, B., & Belcour, L. (1989). Mitochondrial DNA instabilities and rearrangements in yeasts and fungi. Mobile DNA, 861-878.

Edgell, D., Chalamcharla, V., & Belfort, M. (2011). Learning to live together: mutualism between self-splicing introns and their hosts. BMC Biology, 9(1), 22.

Edgell, D. R., Gibb, E. A., & Belfort, M. (2010). Mobile DNA elements in T4 and related phages. Virol J, 7(290), 422X-7.

Eldarov, M. A., Mardanov, A. V., Beletsky, A. V., Dzhavakhiya, V. V., Ravin, N. V., & Skryabin, K. G. (2012). Complete mitochondrial genome of compactin-producing fungus *Penicillium solitum* and comparative analysis of *Trichocomaceae* mitochondrial genomes. FEMS Microbiology Letters, 329(1), 9-17.

Fedorova, O., & Zingler, N. (2007). Group II introns: structure, folding and splicing mechanism. Biological chemistry, 388(7), 665.

Férandon, C., Moukha, S., Callac, P., Benedetto, J. P., Castroviejo, M., & Barroso, G. (2010). The *Agaricus bisporus cox1* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. PloS One, 5(11), e14048.

Forgetta, V., Leveque, G., Dias, J., Grove, D., Lyons Jr, R., Genik, S., Wright, C., Singh, S., Peterson, N., Zianni, M., Kieleczawa, J., Steen, R., Perera, A., Bintzler, D., Adams, S., Hintz, W., Jacobi, V., Bernier, L., Levesque, R., & Dewar, K. (2013). Sequencing of the Dutch Elm Disease fungus genome using the Roche/454 GS-FLX Titanium system in a comparison of multiple genomics core facilities. Journal of Biomolecular Techniques: JBT, 24(1), 39.

Gardes, M, Mueller, G.M., Fortin, J.A., & Kropp, B.R. (1991). Mitochondrial DNA polymorphisms in *Laccaria bicolor, L. laccata, L. proxima* and *L. amethystina*. Mycological Research ,95(2), 206–216.

Gautheret, D., & Lambert, A. (2001). Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. Journal of Molecular Biology, 313(5), 1003-1011.

Ghikas, D. V., Kouvelis, V. N., & Typas, M. A. (2010). Phylogenetic and biogeographic implications inferred by mitochondrial intergenic region analyses and ITS1-5.8 S-ITS2 of the entomopathogenic fungi *Beauveria bassiana* and *B. brongniartii*. BMC Microbiology, 10(1), 174.

Gibb, E. A., & 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(10), 1112-1126.

Gibb, E. A., & Edgell, D. R. (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.

Gimble, F. S. (2005). Engineering homing endonucleases for genomic applications. In Homing Endonucleases and Inteins (pp. 177-192). Springer Berlin Heidelberg.

Gimble, F. S. (2000). Invasion of a multitude of genetic niches by mobile endonuclease genes. FEMS Microbiology Letters, 185(2), 99-107.

Gobbi, E., Firrao, G., Carpanelli, A., Locci, R., & Van Alfen, N. K. (2003). Mapping and characterization of polymorphism in mtDNA of *Cryphonectria parasitica* : evidence of the presence of an optional intron. Fungal Genetics and Biology, 40(3), 215-224.

Goddard, M. R., & Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. Proceedings of the National Academy of Sciences, 96(24), 13880-13885.

Groot, G. S., Mason, T. L., & Van Harten-Loosbroek, N. (1979). *Var 1* is associated with the small ribosomal subunit of mitochondrial ribosomes in yeast. Molecular and General Genetics MGG, 174(3), 339-342.

Guarro, J., Gené, J., & Stchigel, A. M. (1999). Developments in fungal taxonomy. Clinical Microbiology Reviews, 12(3), 454-500.

Gutiérrez, P., Alzate, J., Yepes, M. S., & Marín, M. (2014). Complete mitochondrial genome sequence of the common bean anthracnose pathogen *Colletotrichum lindemuthianum*. Mitochondrial DNA, (0), 1-2.

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

Hafez, M., Guha, T. K., Shen, C., Sethuraman, J., & Hausner, G. (2014). PCR-Based Bioprospecting for Homing Endonucleases in Fungal Mitochondrial rRNA Genes. In Homing Endonucleases (pp. 37-53). Humana Press.

Hamari, Z., Pfeiffer, I., Ferenczy, L., & Kevei, F. (1999). Interpretation of variability of mitochondrial genomes in the species *Aspergillus carbonarius*. Antonie van Leeuwenhoek, 75(3), 225-231.

Hane, J. K., Lowe, R. G., Solomon, P. S., Tan, K. C., Schoch, C. L., Spatafora, J. W., Crousc, P. W., Kodirad, C., Birrend, B.W., Galagand, J. E., Torrianie, S. F., McDonalde, B. A., & Oliver, R. P. (2007). Dothideomycete–plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. The Plant Cell Online, 19(11), 3347-3368.

Haugen, P., & Bhattacharya, D. (2004). The spread of LAGLIDADG homing endonuclease genes in rDNA. Nucleic Acids Research, 32(6), 2049-2057.

Haugen, P., Simon, D. M., & Bhattacharya, D. (2005). The natural history of group I introns. Trends in Genetics, 21(2), 111-119.

Hausner, G., Reid, J., Klassen, G.R. (1993). On the phylogeny of *Ophiostoma*, *Ceratocystis* s.s., and *Microascus*, and relationships within *Ophiostoma* based on partial ribosomal DNA sequences. Canadian Journal of Botany , 71(9), 1249–1265.

Hausner, G., Reid, J., & Klassen, G. R. (1992). Do galeate-ascospore members of the Cephaloascaceae, Endomycetaceae and Ophiostomataceae share a common phylogeny?. Mycologia, 870-881.

Hausner, G., Monteiro-Vitorello, C. B., Searles, D. B., Maland, M., Fulbright, D. W., & Bertrand, H. (1999). A long open reading frame in the mitochondrial LSU rRNA group-I intron of *Cryphonectria parasitica* encodes a putative S5 ribosomal protein fused to a maturase. Current Genetics, 35(2), 109-117.

Hausner, G. (2003). Fungal mitochondrial genomes, plasmids and introns. Applied Mycology and Biotechnology, 3, 101-131.

Hausner, G., Eyjólfsdóttir, G.G., Reid, J. (2003). Three new species of *Ophiostoma* and notes on *Cornuvesica falcata*. Canadian Journal of Botany, 81, 40–48.

Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E. R., Karol, K. G., McCourt, R. M., & 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(2), 380-391.

Hausner, G. (2012). Introns, mobile elements, and plasmids. In Organelle genetics (pp. 329-357). Springer Berlin Heidelberg.

Hawksworth, D. L. (2011). A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. IMA Fungus: The Global Mycological Journal, 2(2), 155.

Hintz, W. E., Mohan, M., Anderson, J. B., & Horgen, P. A. (1985). The mitochondrial DNAs of Agaricus: heterogeneity in *A. bitorquis* and homogeneity in *A. brunnescens*. Current Genetics, 9(2), 127-132.

Hintz, W., Pinchback, M., Bastide, P., Burgess, S., Jacobi, V. ,Hamelin, R., Breuil, C., & Bernier, L. (2011). Functional categorization of unique expressed sequence tags obtained from the yeast-like growth phase of the elm pathogen *Ophiostoma novo-ulmi*. BMC Genomics, 12(1), 431.

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Research, 9(9), 868-877.

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

Jin, Y., Binkowski, G., Simon, L.D. & Norris, D. (1997). Ho endonuclease cleaves MAT DNA in vitro by an inefficient stoichiometric reaction mechanism. Journal of Biological Chemistry, 272, 7352–7359.

Joardar, V., Abrams, N. F., Hostetler, J., Paukstelis, P. J., Pakala, S., Pakala, S. B., Zafar, N., Abolude, O. O., Payne, G., Andrianopoulos, A., Denning, D. W., & Nierman, W. C. (2012). Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genome size variability. BMC Genomics, 13(1), 698.

Johansen, S. T. E. I. N. A. R., & Haugen, P. E. I. K. (2001). A new nomenclature of group I introns in ribosomal DNA. RNA, 7(7), 935-936.

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research, 30(14), 3059-3066.

Kowalski, J. C., & Derbyshire, V. (2002). Characterization of homing endonucleases. Methods, 28(3), 365-373.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., & Mikami, T. (2000). The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNACys (GCA). Nucleic Acids Research, 28(13), 2571-2576.

Kück, U., Osiewacz, H. D., Schmidt, U., Kappelhoff, B., Schulte, E., Stahl, U., & Esser, K. (1985). The onset of senescence is affected by DNA rearrangements of a discontinuous mitochondrial gene in *Podospora anserina*. Current genetics, 9(5), 373-382.

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Research, 41(W1), W22-W28.

Kowalski, J. C., & Derbyshire, V. (2002). Characterization of homing endonucleases. Methods, 28(3), 365-373.

LaPolla, R. J., & Lambowitz, A. M. (1981). Mitochondrial ribosome assembly in *Neurospora crassa*. Purification of the mitochondrially synthesized ribosomal protein, S-5. Journal of Biological Chemistry, 256(13), 7064-7067.

Lambowitz, A. M., LaPolla, R. J., & Collins, R. A. (1979). Mitochondrial ribosome assembly in *Neurospora*. Two-dimensional gel electrophoretic analysis of mitochondrial ribosomal proteins. The Journal of Cell Biology, 82(1), 17-31.

Lambowitz, A. M., & Belfort, M. (1993). Introns as mobile genetic elements. Annual Review of Biochemistry, 62(1), 587-622.

Lambowitz, A. M., & Zimmerly, S. (2004). Mobile group II introns. Annu. Rev. Genet., 38, 1-35.

Lambowitz, A. M., & Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. Cold Spring Harbor Perspectives in Biology, 3(8), a003616.

Lambowitz, A. M., Mohr, G., & Zimmerly, S. (2005). Group II intron homing endonucleases: ribonucleoprotein complexes with programmable target specificity. In Homing endonucleases and inteins (pp. 121-145). Springer Berlin Heidelberg.

Lang, B. F., Gray, M. W., & Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. Annual Review of Genetics, 33(1), 351-397.

Lang, B. F., Laforest, M. J., & Burger, G. (2007). Mitochondrial introns: a critical view. Trends in Genetics, 23(3), 119-125.

Linnakoski, R., De Beer, Z. W., Rousi, M., Solheim, H., & Wingfield, M. J. (2009). *Ophiostoma denticiliatum* sp. *nov*. and other *Ophiostoma* species associated with the birch bark beetle in southern Norway. Persoonia-Molecular Phylogeny and Evolution of Fungi, 23(1), 9-15.

Lucas, P., Otis, C., Mercier, J. P., Turmel, M., & Lemieux, C. (2001). Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. Nucleic Acids Research, 29(4), 960-969.

Luo, G. X., & Taylor, J. (1990). Template switching by reverse transcriptase during DNA synthesis. Journal of Virology, 64(9), 4321-4328.

Marcaida, M. J., Muñoz, I. G., Blanco, F. J., Prieto, J., & Montoya, G. (2010). Homing endonucleases: from basics to therapeutic applications. Cellular and Molecular Life Sciences, 67(5), 727-748.

Marin, M., Preisig, O., Wingfield, B. D., Kirisits, T., Yamaoka, Y., & Wingfield, M. J. (2005). Phenotypic and DNA sequence data comparisons reveal three discrete species in the *Ceratocystis polonica* species complex. Mycological Research, 109(10), 1137-1148.

McBride, H. M., Neuspiel, M., & Wasiak, S. (2006). Mitochondria: more than just a powerhouse. Current Biology, 16(14), R551-R560.

Michel, F., Jacquier, A., & Dujon, B. (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie, 64(10), 867-881.

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

Michel, F., & Feral, J. (1995). Structure and activities of group II introns. Annual Review of Biochemistry, 64(1), 435-461.

Michel, F., & Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. Journal of Molecular Biology, 216(3), 585-610.

Monteiro-Vitorello, C. B., Hausner, G., Searles, D. B., Gibb, E. A., Fulbright, D. W., & Bertrand, H. (2009). The *Cryphonectria parasitica* mitochondrial *rns* gene: Plasmid-like elements, introns and homing endonucleases. Fungal Genetics and Biology, 46(11), 837-848.

Mullineux, S. T., Costa, M., Bassi, G. S., Michel, F., & Hausner, G. (2010). A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. RNA, 16(9), 1818-1831.

Naruse, A., Yamamoto, H., & Sekiguchi, J. (1993). Nucleotide sequence of the large mitochondrial rRNA gene of *Penicillium chrysogenum*. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 1172(3), 353-356.

Nielsen, H., & Johansen, S. D. (2009). Group I introns: moving in new directions. RNA Biology, 6(4), 375-383.

Nicholas, K. B., Nicholas, H. B. J., & Deerfield, D. W. (1997). GeneDoc: analysis and visualization of genetic variation. Embnew. News, 4(1).

Nikoh, N., & 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(9), 1631-1642.

Okada, G., Jacobs, K., Kirisits, T., Louis-Seize, G.W., Seifert, K.A., Sugita, T., Takematsu, A., & Wingfield, M.J. (2000). Epitypification of *Graphium penicillioides Corda*, with comments on the phylogeny and taxonomy of *Graphium*-like synnematous fungi. Studies in Mycology, (45), 169-186.

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., Bradford J. Condon,Copeland, A.C., Dhillon, B., Glaser, F., Hesse, C. N., Kosti, I., LaButti, K., Lindquist, E. A., Lucas, S., Salamov, A. A., Bradshaw, R. E., Ciuffetti, L., Hamelin, R. C., Kema, G. H., Lawrence, C., Scott, J. A., Spatafora, J. W., Turgeon, B. G., de Wit, P., Zhong, S., Goodwin, S. B., & Grigoriev, I. V. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen *Dothideomycetes* fungi. PLoS Pathogens, 8(12), e1003037.

Osiewacz, H. D., Hermanns, J., Marcou, D., Triffi, M., & Esser, K. (1989). Mitochondrial DNA rearrangements are correlated with a delayed amplification of the mobile intron (plDNA) in a long-lived mutant of *Podospora anserine.* Mutation Research/DNAging, 219(1), 9-15.

Papadopoulos, J. S., & Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics, 23(9), 1073-1079.

Paquin, B., & Franz Lang, B. (1996). The mitochondrial DNA of *Allomyces macrogynus*: the complete genomic sequence from an ancestral fungus. Journal of Molecular Biology, 255(5), 688-701.

Paoletti, M., Buck, K. W., & Brasier, C. M. (2005). Cloning and sequence analysis of the MA TB (MAT-2) genes from the three Dutch elm disease pathogens, *Ophiostoma ulmi*, *O. novo-ulmi* and *O. himal-ulmi*. Mycological Research, 109(9), 983-991.

Paoletti, M., Buck, K. W., & Brasier, C. M. (2006). Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote *Ophiostoma novo-ulmi*. Molecular Ecology, 15(1), 249-262.

Pruchner, D., Nassal, B., Schindler, M., & Knoop, V. (2001). Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Molecular Genetics and Genomics, 266(4), 608-613.

Pöggeler, S., & Kempken, F. (2004). Mobile genetic elements in mycelial fungi. In Genetics and Biotechnology (pp. 165-198). Springer Berlin Heidelberg.

Ponting, C. P., Aravind, L., Schultz, J., Bork, P., & Koonin, E. V. (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. Journal of molecular biology, 289(4), 729-745.

Pyle, A. M. (2010). The tertiary structure of group II introns: implications for biological function and evolution. Critical Reviews in Biochemistry and Molecular Biology, 45(3), 215-232.

Roman, J., & Woodson, S. A. (1998). Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. Proceedings of the National Academy of Sciences, 95(5), 2134-2139.

Roman, J. U. D. I. B. E. L. L. E., Rubin, M. N., & Woodson, S. A. (1999). Sequence specificity of in vivo reverse splicing of the *Tetrahymena* group I intron. RNA, 5(1), 1-13.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(12), 1572-1574.

Ronquist, F. (2004). Bayesian inference of character evolution. Trends in Ecology & Evolution, 19(9), 475-481.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Howlett, B. J., & et al. (2011). Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nature Communications, 2, 202.

Rudski, S. M., & Hausner, G. (2012). The mtDNA *rps3* locus has been invaded by a group I intron in some species of *Grosmannia*. Mycoscience, 53(6), 471-475.

Saldanha, R., Mohr, G., Belfort, M., & Lambowitz, A. M. (1993). Group I and group II introns. The FASEB Journal, 7(1), 15-24.

Salvo, J. L., Rodeghier, B., Rubin, A., & Troischt, T. (1998). Optional Introns in Mitochondrial DNA of *Podospora anserina* Are the Primary Source of Observed Size Polymorphisms. Fungal Genetics and Biology, 23(2), 162-168.

Sethuraman, J., Okoli, C. V., Majer, A., Corkery, T. L., & Hausner, G. (2008). The sporadic occurrence of a group I intron-like element in the mtDNA rnl gene of *Ophiostoma novo-ulmi* subsp. *americana*. Mycological Research, 112(5), 564-582.

Sethuraman, J., Majer, A., Friedrich, N. C., Edgell, D. R., & Hausner, G. (2009a). Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of *Ophiostoma* and related taxa. Molecular Biology and Evolution, 26(10), 2299-2315.

Sethuraman, J., Majer, A., Iranpour, M., & Hausner, G. (2009b). Molecular evolution of the mtDNA encoded *rps3* gene among filamentous ascomycetes fungi with an emphasis on the ophiostomatoid fungi. Journal of Molecular Evolution, 69(4), 372-385.

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

Seifert, K. A. (1993). Sapstain of commercial lumber by species of *Ophiostoma* and *Ceratocystis*. *Ceratocystis* and *Ophiostoma*. Taxonomy, Ecology and Pathogenicity, 141-151.

Seifert, K. A., de Beer, Z. W., & Wingfield, M. J. (2013). The Ophiostomatoid Fungi: Expanding Frontiers. CBS-KNAW Fungal Biodiversity Centre.

Simossis, V. A., & Heringa, J. (2005). PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Research, 33(suppl 2), W289-W294.

Spooner, B. & Roberts, P. (2005). Fungi. Collins New Naturalist series No. 96. HarperCollins Publishers, London.

St-Germain, G., & Summerbell, R. (1996). Identifying filamentous fungi: a clinical laboratory handbook. Star Publishing Company.

Stoddard, B. L. (2005). Homing endonuclease structure and function. Quarterly Reviews of Biophysics, 38(01), 49-95.

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

Strobel, S. A. (2013). Biochemistry: Metal ghosts in the splicing machine. Nature, 503(7475), 201-202.

Suh, S. O., Jones, K. G., & Blackwell, M. (1999). A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of group I introns. Journal of Molecular Evolution, 48(5), 493-500.

Sun, X., Li, H., & Yu, D. (2011). Complete mitochondrial genome sequence of the phytopathogenic fungus *Penicillium digitatum* and comparative analysis of closely related species. FEMS Microbiology Letters, 323(1), 29-34.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10), 2731-2739.

Temin, H. M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proceedings of the National Academy of Sciences, 90(15), 6900-6903.

Terpstra, P., Zanders, E., Butow, R.A. (1979). The association of *var1* with the *38 S* mitochondrial ribosomal subunit in yeast. Journal of Biologibal Chemistry, 254, 12653–12661.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y., & Martin, W. (2004). Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nature Reviews Genetics, 5(2), 123-135.

Thompson, A. J., Yuan, X., Kudlicki, W. &Herrin, D. L.(1992). Cleavage and recognition pattern of a doublestrand- specific endonuclease (I-CreI) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. Gene, 119, 247–251.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The

CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research,25(24), 4876-4882.

Toor, N., & Zimmerly, S. (2002). Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA, 8(11), 1373-1377.

Upadhyay, H. P. (1993). Classification of the ophiostomatoid fungi.*Ceratocystis* and *Ophiostoma:* Taxonomy, ecology and pathogenicity. The American Phytopathological Society. St. Paul, MN. USA. pp, 7-13.

Uzunovic, A., Dale, A., Stirling, R., & Sidhu, A. (2013). The Effect of Fungal Preinfection on the Efficacy of Mold and Sapstain Control Products. Forest Products Journal, 63(1).

Wang, J., Kim, H.-H., Yuan, X. & Herrin, D. L. (1997). Purification, biochemical characterization and protein- DNA interactions of the I-CreI endonuclease produced in *Escherichia coli.* Nucleic Acids Research 25, 3767–3776.

Webster, J., & Weber, R. (1980). Introduction to fungi (Vol. 667). Cambridge: Cambridge University Press.

Windbichler, N., Menichelli, M., Papathanos, P. A., Thyme, S. B., Li, H., Ulge, U. Y., Hovde, B. T., Baker, D., Monnat, R. J., Burt, A., & Crisanti, A. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature, 473(7346), 212-215.

Windbichler, N., Papathanos, P. A., Catteruccia, F., Ranson, H., Burt, A., & Crisanti, A. (2007). Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. Nucleic Acids Research, 35(17), 5922-5933.

Windbichler, N., Papathanos, P. A., & 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, M. J., Seifert, K. A., & Webber, J. F. (1993). *Ceratocystis* and *Ophiostoma*: Taxonomy, ecology and pathogenicity. American Phytopathological Society (APS).

Woodson, S. A. (2005). Structure and assembly of group I introns. Current Opinion in Structural Biology, 15(3), 324-330.

Wool, I. G. (1996). Extraribosomal functions of ribosomal proteins. Trends in Biochemical Sciences, 21(5), 164-165.

Wool, I. G., Chan, Y. L., & Glück, A. (1995). Structure and evolution of mammalian ribosomal proteins. Biochemistry and Cell Biology, 73(11-12), 933-947.

Xu, J. (2005). The inheritance of organelle genes and genomes: patterns and mechanisms. Genome, 48(6), 951-958.

Yamamoto, H., Naruse, A., Ohsaki, T., & Sekiguchi, J. (1995). Nucleotide sequence and characterization of the large mitochondrial rRNA gene of *Penicillium urticae*, and its comparison with those of other filamentous fungi. Journal of Biochemistry, 117(4), 888-896.

Youssar, L., Grüning, B. A., Günther, S., & Hüttel, W. (2013). Characterization and Phylogenetic Analysis of the Mitochondrial Genome of *Glarea lozoyensis* Indicates High Diversity within the Order Helotiales. PloS One, 8(9), e74792.

Zaphiropoulos, P. G. (2002). Template switching generated during reverse transcription?. FEBS letters, 527(1), 326.

Zeng, X. C., & Wang, S. X. (2002). Evidence that BmTXKß–BmKCT cDNA from Chinese scorpion *Buthus martensii Karsch* is an artifact generated in the reverse transcription process. FEBS Letters, 520(1), 183-184.

Zhou, Y., Lu, C., Wu, Q. J., Wang, Y., Sun, Z. T., Deng, J. C., & Zhang, Y. (2008). GISSD: group I intron sequence and structure database. Nucleic Acids Research, 36(suppl 1), D31-D37.

Zhu, S., Cao, Y. Z., Jiang, C., Tan, B. Y., Wang, Z., Feng, S., Zhang, L., Su, X. H., Brejova, B., Vinar, T., Xu, M., Wang, M. X., Zhang, S. G., Huang, M. R., Wu, R., & Zhou, Y. (2012). Sequencing the genome of *Marssonina brunnea* reveals fungus-poplar co-evolution. BMC Genomics, 13(1), 382.

Zimmerly, S., Guo, H., Eskest, R., Yang, J., Perlman, P. S., & Lambowitz, A. M. (1995a). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell, 83(4), 529-538.

Zimmerly, S., Guo, H., Perlman, P. S., & Lambowltz, A. M. (1995b). Group II intron mobility occurs by target DNA-primed reverse transcription. Cell, 82(4), 545-554.

Zipfel, R. D., de Beer, Z. W., Jacobs, K., Wingfield, B. D., & Wingfield, M. J. (2006). Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. Studies in Mycology, 55(1), 75-97.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31(13), 3406-3415.