Fungal phylogenies based on mitogenomes and a study of novel introns in the *cob* and *cox3* genes of *Ophiostoma ips*

by

Nikita Patel

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Microbiology

University of Manitoba

Winnipeg, MB R3T 2N2

Abstract

Fungi are a diverse group of organisms with complex mitochondrial genomes and poorly resolved taxonomic schemes. This study examined the utility of concatenated mitochondrial amino acid sequences for inferring fungal phylogenies. Phylogenetic trees generated, supported existing taxonomic proposals based on the analysis of nuclear genes. This study also examined complex introns within *Ophiostoma ips* strains embedded in the *cob* and *cox*3 genes. Fifty-six strains were examined and 11 were selected for sequence analysis. The *O. ips cob*-490 intron was noted to be composed of three components, two group I intron are situated side by side and one group II intron is nested within the first group I intron. The *cox*3-640 intron was noted to be a side-by-side twintron composed of two group I introns located next to each other. These introns in part demonstrate why fungal mtDNAs are variable in size and complex in their architecture.

Acknowledgements

This work is dedicated to my Grandfather who taught me important life lessons in my childhood. I also dedicate this thesis and my future success to my loving husband and my three year old daughter who nurtured me with their constant love and support and made sacrifices to make this achievement possible.

My deepest gratitude to my supervisor Dr. Georg Hausner; I appreciate his commitment, faith and mentorship for my success.

I would like to thank my committee members, Dr. Deborah Court, Dr. Georg Hausner and Dr. Jeffery Marcus for their time and valuable suggestions. I am so thankful to my lab members; Alvan, Zubaer, Amal, Jigeesha, Jordan and Aamer for their company.

I am very grateful for the financial support by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Graduate Enhancement of Tri-Council Stipends (GETS) program by the Faculty of Graduate Studies (University of Manitoba) which helped me to get dedicated to study.

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

At last, thanks to the reader for making the pages most valuable.

Table of contents

List of tables

List of figures

List of abbreviations

AA	Amino Acid
ATP	Adenosine Triphosphate
atp	ATP synthetase gene
BLAST	Basic Local Alignment Search Tool
BLASTn	nucleotide BLAST
BLASTp	protein BLAST
bp	basepair
cDNA	complementary DNA
cob/ cytb	Cytochrome b gene
cox	cytochrome c oxidase gene
CDS	Coding DNA Sequence
CTAB	Cetyl-trimethyl-ammonium bromide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSB	double-strand break
EDTA	Ethylene Diamine Tetra Acetic acid
g	grams
HE	homing endonuclease
HEG	homing endonuclease gene
IEP	intron-encoded protein
EBS	Exon Binding Site
GFF	General feauture format/ gene finding format (sequences)

IBS	Intron Binding Site
IGS	internal guide sequence
ITS	internal transcribed spacer region
kb	kilo basepair
L	litre
mg	microgram
µl	microlitre
µM	micromolar
M	molar
MEA	malt extract agar
Mg	milligram
ml	millilitre
mtDNA	mitochondrial DNA
mRNA	Messenger RNA
nad	NADH dehydrogenase gene
NADH	nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
ng	nanogram
NGS	Next generation sequencing
nt	nucleotide
ORF	Open reading frame
PCR	polymerase chain reaction
PYG	peptone-yeast extract-glucose medium

rDNA	ribosomal DNA
RNA	ribonucleic acid
rRNA	ribosomal RNA
rnl	large subunit ribosomal RNA gene
rns	small subunit ribosomal RNA gene
rnp	ribonucleoprotein
rpm	revolutions per minute
rps3	ribosomal protein S3 gene
RT	reverse transcriptase
RT -PCR	reverse transcriptase - polymerase chain reaction
tRNA	transfer RNA
UV	ultraviolet
WIN(M)	culture collection at University of Manitoba, Winnipeg, MB, Canada
°C	degrees Celsius

CHAPTER 1: LITERATURE REVIEW

1. Review of Literature

1.1 The Fungi

Fungi are economically very important organisms because they include plant and animal pathogens, potential producers of antimicrobial compounds and useful enzymes. They are also very useful in bioremediation or energy (biofuels) production because of their well-known decomposing activity. Hence, continued fundamental research into the evolution, taxonomy, and physiology of fungi will always be important. Fungi are obligate aerobes (except the yeasts or rumen fungi) and are classified in four Phyla which are the Ascomycota, the Basidiomycota, the Zygomycota and the Chytridiomycota. The Ascomycota include the largest number of known fungal species.

1.2 The Ophiostomatales

Among the fungi, the order Ophiostomatales belong to the Ascomycota and contains fungi with diverse characteristics like human/plant pathogenic fungi and blue stain fungi that are found over a wide host range (De Beer & Wingfield, 2013; De Beer et al., 2013). Taxonomically this Order is very complex and to some degree its genera and their generic boundaries are still controversial. Many species of the Ophiostomatales have adapted towards dispersal by insects such as sticky asexual spores on long stalked conidiophores and, when present, the release of ascospores as "sticky masses" on the tips of long-necked ascocarps. Examples of the Ophiostomatales are *Sporothrix insectorum* (an entomopathogen), *Sporothrix schenckii* (human pathogen), *Ophiostoma pilifera* (blue stain fungi) and *Ophiostoma ulmi* (Dutch elm disease) (Abboud et al., 2018). Blue stain fungi are fungi which cause discoloration of timber by staining the wood blue/brown because of excessive production of melanin pigments (Bilto & Hausner, 2016), these fungi reduce the economic value of timber for export. Some members of this group can sometimes be pathogenic to the host tree in absence of their insect vectors. Other members form symbiotic relationships with their bark beetle

vectors and in combination can cause significant diseases [i.e. *Dendroctonus ponderosae* (Mountain Pine beetle) and *Grosmannia clavigera*]. Blue stain fungi can be found in the following genera: *Ophiostoma, Ceratocystiopsis,Grosmannia* and in their asexual counter parts such as *Leptographium,* and *Pesotum* (reviewed in Hausner et al., 2005; Zipfel et al., 2006); although recently additional genera have been circumscribed and accommodated within the Ophiostomatales (De Beer et al., 2013). The Genus *Ceratocystis* historically was placed within the Ophiostomatales, however it was demonstrated (by Hausner et al., 1992, 1993) that *Ceratocystis* is only distantly related to the Ophiostomatales and thus has been placed into the Order Microascales. Recently the Genus *Ceratocystis* has been subdivided into various genera based on molecular criteria (De Beer et al., 2013). Collectively the members of *Ophiostoma sensu lato* and *Ceratocystis sensu lato* are sometimes referred to the ophiostomatoid fungi due to their ecological similarities such as being blue stain fungi that infect the sap wood of various soft and hard woods and a reliance on insect vectors for spore dispersal. They also share various morphological features (i.e. production of sticky spore masses) due to convergent evolution occupying similar niches and insect vectors.

1.3 Species of *Ophiostoma* and *Ophiostoma ips*

Some species of the Ophiostomatales have a very close symbiotic association with certain bark beetle and a variety of other insect species which attack particular tree species, damaging roots, stems, seeds, or fruits and these beetles/ insects act as vectors for some fungi to infect the plant. The Ophiostomatales include *Ophiostoma novo-ulmi* (and members of the *Ophiosotma ulmi* species complex) and species such as *Ophiostoma ips, Ophiostoma minus and Ophiostoma piliferum*, the later three are common blue stain fungi. *Ophiostoma novo-ulmi* is a potential derivative of *Ophiostoma ulmi*, the original causative agent of Dutch Elm Disease (Brasier, 1991). But *O. novo-ulmi* is a more cold and virus tolerant species that colonizes a wider variety of bark beetle vectors

(Brasier, 1991; Bernier, 1993) and therefore has replaced *O. ulmi* in many regions. Dutch Elm Disease (DED) is the best-known example of a plant disease that has demolished urban forests such as the American elm. DED is caused by a several fungal species like *Ophiostoma ulmi* and *Ophiostoma novo-ulmi*. In contrast, *Ophiostoma ips* is a blue stain fungus primarily infecting the sap wood of soft woods and it is vectored by bark beetles and its blue staining activity reduces the economic value of timber (Linnakoski et al., 2012). *Ophiostoma ips* has been reported in North America, Europe, Japan, New Zealand, and South Africa, and appears to be vectored by a broad range of bark beetles, including species of *Ips* and *Dendroctonus* (Kim et al., 2003). *Ophiostoma ips* is not known to be pathogenic to its hosts but it is a serious sap staining fungus. This study examined the mitogenome of this fungus and focused on complex introns present in the mitochondrial DNA of *Ophiostoma ips.*

1.4 The Mitochondria

Mitochondria are eukaryotic membrane bound organelles with the primary purpose of generating energy. It is also known as "powerhouse of the cell". Mitochondria are considered as semi-autonomous organelles as they can replicate on their own and they possess their own DNA and ribosomes; although the mitochondrial genome is quite small and most products required for mitochondrial function are encoded by the nuclear genome. The origin of mitochondria in eukaryotic cells was proposed by Mereschkowsky Konstantin (1910) stating that, they are the result of an endosymbiotic event in which a prokaryote was "swallowed" up by an unicellular prokaryote followed by symbiotic relationship between two organisms, a concept has have been elaborated upon by Margulis (1970) and was reviewed by Burger et al., (2003) and Martin et al., (2015). Although mitochondria contain genetic material, it is rather limited and probably most of the ancestral genome over time got transferred to the nuclear genome explaining the synthesis of most mitochondrial

proteins in the cytoplasm (Burger et al., 2003). One of the main function of mitochondria is to produce energy via the oxidative phosphorylation in respiratory metabolism; however mitochondria also generate many biochemical intermediates via the TCA cycle and iron-sulfur clusters (in some fungi and protozoans) Mitochondria are also involved in processes like cell aging and apoptosis (Basse et al., 2010, Kolesnikova et al., 2019). In pathogenic fungi, mitochondria are reported to play a role in virulence, regulation of biofilm, hyphal growth and activation of drug resistance (Burger et al., 2003; Chatre & Ricchetti, 2014). There are multiple mitochondria in each cell and multiple mitochondrial genomes in each mitochondrion (Sandor et al., 2018). Mitochondria are not static but they are thought to constantly fuse and divide to create networks (Calderone et al., 2015). In most plants and animals, mitochondrial genome inheritance is uniparental from the maternal parent (i.e. maternal inheritance) however in fungi there are diverse patterns of mitochondrial inheritance (Xu & Wang, 2015).

1.5 Mitochondrial DNA

Mitochondrial genomes in the mammals are composed of several hundreds of presumed identical small circular DNA molecules (around 16 kb) in which genes are tightly packed, seperated by tRNAs (Dujon & Belcour, 1989). In other groups of organisms such as plants and fungi mitochondrial genomes have been noted to be considerable larger and a rich source of mobile elements. Mitochondrial genomes in fungi are often circular and composed of a single chromosome but sometimes they can be linear (Burger et al., 2003; Hausner et al., 2003). They contain a core set of genes involved in ATP production, the electron transport chain, and components required for translation. Mitochondrial genes can be classified in three categories, 1. tRNA genes, 2. rRNA genes and 3. protein coding genes. In general, fungal mitochondrial genomes contain 14 conserved protein-coding genes involved in electron transport and respiratory chain complexes (*atp6, atp8, atp9, cob,*

cox1, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*), one ribosomal protein gene (*rps3*), two genes encoding ribosomal RNA subunits – small (*rns*) and large (*rnl*) – and usually a set of tRNA genes. (Bullerwell & Lang, 2005; Gray et al., 1999). Mitochondria are essential organelles for most fungi as most are obligate aerobes, thus mtDNA genes are assumed to be essential for the efficient maintenance of cells, making mtDNA research of great interest.

1.6 Fungal mitochondrial DNA size

Fungal mitochondrial genomes are highly variable with regards to their sizes, and also with regards to organization, and the mtDNA architecture is still under explored. Variation in size is due to the presence of intergenic spacers and intervening sequences, i.e. introns. Introns are "intervening" sequences in genes that are removed during the processing of the primary transcripts to the mRNA. Mitochondrial introns in fungi can be allotted into two categories: group I and group II introns (Michel & Westhof, 1990; Lambowitz et al., 1999; Belfort et al., 2002) which are commonly referred to as mobile introns. Despite relatively conserved gene content, fungal mitochondrial DNA size can range from 12.5 kb in *Rozella allomycis* (James et al., 2013) to 235 kb in *Rhizoctonia solani* and this variation in sizes is mainly due to presence of introns (Zubaer et al., 2018; Bullerwell & Lang, 2005; Losada et al., 2014). Group I and Group II introns can be differentiated from each other by their conserved sequences, catalytic core secondary structures, and splicing mechanisms (Michel & Westhof, 1990; Lehmann & Schmidt, 2003; Deng et al., 2016; Copertino & Hallick, 1991). In fungal mitochondrial genomes, group I introns are more frequent than group II introns whereas the contrary is true for plants where group II introns dominate (Lang et al., 2007). With next generation sequencing (NGS) fungal mitogenomes are more readily available and thus can be explored as a source of molecular markers that can be applied towards fungal taxonomy.

1.7 Introns in fungal mitochondrial DNA

Introns have been classified into 4 major groups: group I, group II, spliceosomal (nuclear mRNA) and enzymatic, i.e. the tRNA introns. Group I and group II introns are found in fungal mtDNAs. As mentioned previously, mtDNA introns are the main contributors towards the size of fungal mtDNA. The presence and absence of introns is a main reason for most mtDNA polymorphism in closely related fungi (Abboud et al., 2018). Intron is usually defined as a segment of a DNA or RNA molecule that does not code for proteins and interrupts the sequence of genes; introns are removed during RNA processing of the primary transcripts. Fungal mitochondrial introns are not spliceosomal introns, instead they are self-splicing type introns and in some cases they encode open reading frames (ORFs). Also group I and group II introns appear to use intron and/or nuclear encoded maturases to enhance their splicing efficiency (Lang et al., 2007; Belfort, 2003; Hausner, 2012). Maturases are proteins that facilitate the "maturation" of transcripts and both group I and group II introns can encode proteins that act as maturases (Lang et al., 2007; Lambowitz et al., 1999). Additional proteins such as nuclear encoded RNA helicases and RNA chaperones have been co-opted to enhance the splicing of group I and group II introns (reviewed in Hausner 2012). Group I introns have a tendency to move at the DNA level involving recombination pathways whereas group II introns move via an RNA intermediate. Both types of elements are referred to as "homing introns" as they tend to insert at the same location (sequence specific) in cognate alleles that lack introns; essentially they move from intron plus alleles to intron minus alleles (reviewed by Belfort, 2003).

The finding that RNA can be a catalyst from previous studies brought new insights into RNA as a ribozyme (Thomas Cech and Sidney Altman; Nobel Prize in 1989 for the discovery of ribozymes). Ribozyme is an RNA molecule that catalyzes a biochemical reaction such as transesterification reactions that are involved in the splicing of group I and group II introns from

their respective transcripts. The discovery of ribozymes enhance the RNA world hypothesis (Alberts et al., 2002) which argues the possibility that during early stages of the evolution of life, enzymes were composed of RNA rather than protein, also RNA could serve as both: the "source of information" and "drivers of metabolism".

In summary, fungal mitochondrial introns are mobile ribozymes that can splice in and out of genome (Saldanha et al., 1993; Sandor et al., 2018). Group I and II introns are not only ribozymes that catalyse their own splicing but are also mobile genetic elements (reviewed in Lambowitz & Belfort, 1993). Previous studies have shown that some mtDNA introns are composed of several intron modules: twintrons (intron nested within an intron) and tandem introns (side-by-side introns) (Hafez et al., 2013; Deng et al., 2016). These so called complex or nested introns show the dynamic nature of these mobile elements (Hafez & Hausner, 2015).

1.8 Intron Encoded Proteins

Intron Encoded Proteins (IEPs) are proteins encoded within the intron RNA. IEPs usually promote movement of intron from the host alleles to cognate alleles who are lacking the intron (Dujon, 1989) and hence, these encoded elements are required for intron mobility; however it is suspected that many IEPs are potentially bifunctional and also serve as RNA maturases promoting the proper splicing of intron RNAs (reviewed by Belfort, 2003). Thus, some IEPs serve two for one purpose, having both maturase and HE activity (Lang et al, 2007; Hausner, 2012). Group I introns usually encode homing endonucleases (HE) and group II introns encode reverse transcriptases. However, there are a few examples of group II introns that encode homing endonucleases (Toor & Zimmerly, 2002) and these group II introns appear to have a mobility pathway similar to group I introns (Mullineux et al., 2010). Some group II introns (matR and matK) encode proteins that have

only maturase activities (Hausner et al., 2006) which recognize intron specific elements to encourage correct folding of precursor RNA leading to catalysis but many group II introns encode RTs that have a maturase domain that binds to the intron RNA and promotes proper folding (Belfort et al., 2002).

Homing Endonucleases (HEs) can be categorized based on conserved amino acid motives, for example with regards to fungal mtDNA introns HEs with LAGLIDADG, GIY-YIG domains are frequently encountered (Stoddard, 2014). Some group I introns have been noted to encode mitochondrial ribosomal proteins (rps3) and unusual ORFs that resemble nuclear N-acetyltransferase genes (Wai et al., 2019). The reverse transcriptases (RT) encoded within group II introns are multi domain proteins: maturase domain, DNA cutting domain, DNA binding domain, and RT domain (Zimmerly et al. 2001). The RT is involved in the actual group II splicing event and in the intron retro homing mechanism.

1.9 IEP function

Intron encoded proteins (endonucleases and reverse transcriptases) have been studied extensively as they have variety of applications in biotechnology as gene editing tools (Candales et al., 2012; Lehmann & Schmidt, 2003; Hafez & Hausner, 2012; Guha et al., 2017). IEPs promote the mobility of their host introns and can assist in splicing of their host introns (Szczepanek & Lazowska, 1996; Lang et al., 2007).

1.10 Homing Endonucleases (HEs)

Homing endonucleases are small (<40 kDa) proteins that are sometimes intron encoded, and they are highly site-specific and it has been reported that HEs have long DNA recognition sequences (usually 14 bp to 40 bp) (Stoddard, 2006), thus they are “rare cutting” enzymes. Homing

endonuclease domains are also found in inteins and there are many free-standing versions of HEs found in phage and bacterial genomes (Stoddard, 2005). Instances have been documented where homing endonucleases move independently from their intron hosts (reviewed in Hausner 2012), so HEs are mobile elements that can mobilize only themselves or they can mobilize flanking regions during homing or transposition into new sites. HEs are DNA endonucleases that assist homing of intervening sequence by catalyzing single or double-strand break inside their target sequence (Dujon, 1989; Hafez & Hausner, 2012).

1.11 HEGs

The genes that code for Homing endonuclease are called HEGs (Homing endonuclease genes). HEGs are located within terminal loops of group I introns and the D3/D4 domains of some group II introns. In general they are located in loops and domains that do not interfere with intron RNA folding and splicing of the host intron (Toor & Zimmerly, 2002). Presently at least seven families of HEs are known (Fang et al., 2018)and they are found in all three domains of life and viruses/phages. The naming of each category is based on conserved amino acid motifs, the LAGLIDADG, H-N-H, His- Cysbox, PD-(D/E)xK, EDxHD, DHHRN and GIY-YIG families of HEs (Stoddard, 2011; 2014). HEGs can move on their own or as part of a composite element that includes a group I (or II) intron. In some instances HEGs invade new locations in the host genome (transposition) and sometimes they invade other introns by inserting in the resident HEG that already had “occupied” the intron; here it is thought that HEGs invade new sites without actually increasing the intron or HEG load on the host genome as one HEG parasitize another HEG (Guha et al., 2018). Overall HEGs are argued to be “neutral elements” that in order to persist in genomes have to minimise their impact on the host genome (Goddard & Burt, 1999).

1.12 Intron Homing

Homing is a site-specific mobility event where a mobile intervening sequence is horizontally transferred, usually to a homologous allele of the host gene lacking the intron/HEG. Intron transposition events are different as here the intron/HEG inserts into a new gene or site.

An alternative mobility mechanism for group I introns involved RNA reverse splicing. Here the intron RNA reverse splices into an mRNA molecule; it should be noted that group I intron splicing is a reversible process whereby the intron can splice out and splice back into its original transcript. Degradation of the intron RNA tends to favour the accumulation of spliced products reducing the chance of reverse splicing. The reverse spliced product can be reverse transcribed and via recombination inserted into the genome, this mechanism can be site specific, guided by the introns internal guide sequences (group I introns) or exon binding sequences (group II introns) or it can lead to the insertion of the intron into a new location (reviewed in Hausner 2012).

Group I intron mobility requires a double-strand break that can be generated by the HE near the intron-insertion site in an allele that does not have any insertion. It is assumed that the region that was cut by the HE is further processed by exonucleases that will widen the gap; this will induce the host encoded DNA repair system. The homing process involves the double-strand repair mechanism whereby homologous recombination uses the intron containing allele is used as the repair template and thus is transferred into the gap. Group II introns perform like retro-elements where mobility of the intron is carried out by an RNA intermediate and requires reverse transcriptase activity. In general, mobility of group I and group II introns is referred to as homing or retro-homing respectively. However, these elements can insert into new locations (called ectopic integration) which is termed as transposition or retro- transposition (Hausner, 2003).

1.13 Mobile elements

Mobile elements are DNA sequences that can move around the genome, changing their copy numbers or simply shifting their location, examples include DNA transposable elements such as insertion elements, transposons, integrons, inteins, mobile introns; plasmids and bacteriophage elements. The sum total of all mobile genetic elements in a genome is often referred as the "mobilome" (Muszewska et al., 2019). All higher eukaryotes have genomes containing copies of mobile elements. Mobile elements play a role in insertional mutagenesis along with altered gene expression and potentially influencing genetic recombination. Forty percentage of the human genome is made up of mobile elements (Deininger & Roy-Engel, 2002). Mobile genetic elements in bacteria have in cases distributed a diverse collection of virulence related genes and thereby play an essential role in the evolution of bacterial pathogens. (Davis & Waldor, 2002)

1.14 Intron Splicing

Group I and II introns are spliced at the RNA level requiring various factors which are intron (maturases) and genome encoded (RNA chaperones such as DEAD box proteins) (Lang et al., 2007; Lambowitz & Zimmerly, 2011; Hausner, 2012). In case of group I introns, the 5' splice site is defined by the Internal guide sequence (IGS) located at the 5' end of group I introns, and the 3' splice site is determined by coupling of a short sequence from the downstream exon that can interact with the upstream IGS thus bringing into close proximity the upstream and downstream exons with the active site of the group I intron (P7). Group I introns are composed of paired regions (P1 to P10) and loops. The P7 segment can recruit a free GTP (with 3'OH) that can serve as a nucleophile that attacks the upstream exon/intron boundary initiating the splicing pathway that eventually releases the intron in a linear form and joins the flanking exons. Thus, the tertiary fold of the intron RNA brings the flanking exons together facilitating splicing of the exons and removal of the intron. Group I intron splicing needs an exogenous GTP and Mg^{2+} as a cofactor.

Group II introns do not require an exogenous GTP instead they need a bulged adenine with the 2'OH group (in domain 6) that can serve as the nucleophile initiating the splicing reaction. Again folding of the intron RNA is important in order for the various interactions required for splicing to occur. Group II intron splicing is catalyzed by two transesterification reaction and results in elimination of the intron as a branched or lariat molecule, this pathway is called the branched pathway. The other splicing pathway, suspected to be the most common in bacteria, for group II intron is the hydrolysis pathway that includes a hydrolysis step and the intron is released in a linear form.

1.15 Group I introns

Group I introns have wide phylogenetic distribution having been found in the genomes of Eubacteria, bacteriophages, eukaryotic organelles and nuclei (confined to rDNA) (reviewed in Lambowitz & Belfort, 1993; Hausner, 2012). In addition group I introns have been recently reported to be widespread in the Archaea (Nawrocki et al., 2018). Group I introns are rarely encountered among bacteria (reviewed in Hausner et al., 2014). According to the reviews (Lambowitz & Belfort, 1993; Lambowitz et al., 1999), 30% of group I introns are estimated to contain internal ORFs and a significant number of them are assumed to be mobile.

Group I introns fold into characteristic, well-conserved secondary and tertiary structures. Group I introns often contain open reading frames that assist the intron in its own mobility. Interest in group I intron is due to their self-splicing activity and they are mobile genetic elements. Group I introns tend to be inserted in phylogenetically conserved regions of RNA and protein coding genes of mitochondria in some groups of eukaryotes. Group I introns tend to have two conserved nucleotides at their boundaries, that is a Thymine residue in the upstream exon and they end with a

Guanosine residue (Paquin & Shub, 2001).

Most fungal mtDNA group I introns code for homing endonucleases that provides mobility via recombination (they form a cut at intron less allele and use intron containing allele as the repair template). Loops are important in group I intron structures as they can contain ORFs . For Group I introns about 10 helical paired regions have been reported (P1 – P10) that stabilize the intron core to fold into a splicing competent structure (reviewed in Hausner et al., 2014). The P1 and P10 stems are short-lived interactions, involving base pairing of nucleotides from both the up- and downstream exons and the intron thus bringing the 5' and 3' splice site into close proximity and thus facilitating the splicing (Paquin & Shub, 2001). However, that can be some variability with regards to the various components, for example P2 segment is absent in some group I introns. Group I introns are grouped into various subtypes (A to E) based on features of their RNA fold and presence or absence of key sequences or RNA-RNA interactions (i.e. paired regions) (Michel &Westhof, 1990; Hafez & Hausner, 2014).

1.16 Group II introns

Group II introns are found in mitochondrial DNA of fungi, mitochondrial and chloroplast genomes of plants, algae and in the mtDNAs of some protozoans and metazoans (soft corals and sponges) and they have been noted in bacterial and archaeal genomes (Gimble, 2000; Hafez & Hausner, 2012; Lambowitz & Belfort, 2015). Recently they have been found in Proteobacteria and Cyanobacteria which are closest living relatives to the endosymbiotic ancestors of mitochondria and plastids (Ferat & Michel, 1993; Knoop & Brennicke, 1994). Group II introns are frequently found in chloroplast and mtDNA of plants, to a lesser extend in fungal mtDNA.

Group II introns are reported to form secondary structures consisting of six double helical domains (domains I to VI) forming a "six fingered wheel" like arrangement (Bilto & Hausner,

2016). Group II introns are retroelements encoding ORFs that can mediate reverse splicing and reverse transcriptase activity (Kennell et al., 1993; Augustin et al., 1990) and DNA endonuclease activities (Zimmerly et al., 1995). Group II intron mobility occurs by target DNA primed reverse transcription (Zimmerly et al., 1995). Group II introns move by forming ribonucleoprotein particles (RNP) consisting of the IEP and the spliced lariat version oftheintron RNA (Lambowitz & Zimmerly, 2011). The RNP can scan for target sites by means of recognizing sequences that are complementary to the so-called exon binding sequences present in domain II of the group II intron RNA. The RNP can generate a nick in the sense strand allowing for reverse splicing of the intron RNA into the DNA target site. The endonuclease domain will cleave the antisense strand and the freed 3' end will served as the primer for the reverse transcriptase activity generating the cDNA version of the intron. The host repair system will ultimately replace the reverse spliced intron RNA with DNA.

Group II introns have been engineered into useful bacterial genome editing agents referred to as targetrons. Some bacterial group II introns can be configured to target specific sequences for insertions, thus they can be applied for "insertional mutagenesis" strategies. (Enyeart et al., 2014; Lambowitz & Belfort, 2015; Belfort & Lambowitz, 2019).

2. Objectives

Currently within the fungi, taxonomic proposals are based on morphological and molecular characters. With regards to molecular characters most studies rely on nuclear genes for generating fungal phylogenies. Many of these tend to be poorly resolved and thus new molecular markers need to be generated. In this study I will examine the utility of mitochondrial genomes for generating robust fungal phylogenetic trees.

Objective 1: To generate a phylogenetic tree using mitochondrial derived amino acid sequences obtained from all major fungal groups.

Fungal mitochondrial genomes are highly variable in size and this is due in part because of mobile introns. Some introns can be composed of several "intron modules" and these are referred to as complex introns (or nested introns). Exploring these introns may provide insight on how introns can mobilize into new locations or into pre-existing introns and these configurations maybe of interest with regards to understanding gene regulation as splicing of these complex introns may offer a platform for alternative splicing or modulating gene expression.

Objective 2: Exploring the mtDNA *cob* and *cox3* genes to find novel introns in the mitogenome of various strains of *Ophiostoma ips*.

CHAPTER 2: METHODOLOGY

2.1Computational Methods

2.1.1. Collection of data

The mitochondrial protein sequences were obtained from NCBI (https://www.ncbi.nlm.nih.gov) and MitoFun (Ntertili et al., 2013, http://mitofun.biol.uoa.gr/) sites. Selecting 'organelle' and specifying 'fungi' and 'mitochondria' type in the search criteria provided a list of fungal mitochondrial genomes within the NCBI organelle genome database. MitoFun is a curated database for fungal mitochondrial genomes and in some instances the amino acids sequences were already compiled. Thirteen different mitochondrial protein coding sequences (amino acid sequences) were chosen for following genes: *atp6, atp8, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5* and *nad6*. The *atp9* and *rps3* genes were not chosen because of their absence in some fungi.

2.1.2. Customizing and organizing the data

Most nucleotide sequences collected were available in annotated formats. However, some sequences had to be annotated with the aid of MFannot (Beck and Lang, 2010) and manually validated by BLAST or ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/; Genetic code setting for molds #4; Sayers et al., 2010). In many instances "annotated sequences" as presented in GenBank had to be reassessed as above to ensure proper annotations. Using ORF finder nucleotide sequences for the genes of interest were converted to protein sequences. The amino acids sequences were concatenated in the following gene order (*atp6, atp8, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5*and *nad6*). The multiple concatenated protein sequence dataset was aligned using MAFFT (Multiple Alignment of Fast Fourier Transform) (Katoh & Standley, 2013).

2.1.3. Alignment of sequences

In general, sequences were aligned with MAFFT or MUSCLE program and the alignments would be observed for features and conservation in AliView program (Larsson, 2014). Raw sequence data obtained from mitochondrial DNA loci from NCBI or from my own DNA sequencing of PCR products derived from various *O.ips* strains were compiled and adjusted manually into contiguous sequences using GeneDoc (v2.7.000; Nicholas et al., 1997) and AliView. All nucleotide sequences were examined for the presence of open reading frames (ORFs; such as homing endonuclease genes) by using the online program ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html/; setting: genetic code for mtDNA of molds). Intron exon junctions of the gene were initially obtained by MFannot and verified by multiple sequence alignments (MSA) for gene and aligning it to the CDS of the same gene of related sequences available in NCBI using BLAST search. Sequence alignments were performed using MAFFT and MUSCLE program available in AliView. Intron encoded ORF sequence search was conducted using BLASTp (http://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al. 1990) and collected amino acid sequences were aligned using PRALINE multiple amino acid sequence alignment program (Simossis & Heringa 2003; 2005).

2.1.4. Phylogenetic analysis

The phylogenetic tree was generated by W-IQ-TREE program (Trifinopoulos et al., 2016) in which the aligned amino acid data set was used to generate phylogenetic tree using maximum likelihood criteria and by applying default settings. Only exon-derived sequences were maintained and the bootstrap option (1000 iterations) was implemented to estimate node support values. For large data sets W-IQ-TREE program was applied but in some instances MrBayes (version 3.2.6; Ronquist &

Huelsenbeck, 2003) was used to generate phylogenetic trees. [For Bayesian inference, analysis was initiated from a random starting tree and four chains were run simultaneously for 5 000 000 generations; trees were sampled every 1000 generations. The first 25% of trees generated were discarded ("burn-in") and the remaining trees were used to compute a majority rule consensus tree showing the posterior probability values].

Molecular phylogenetic analysis of intron-encoded ORFs were carried out using programs implemented in MEGA 7 (Kumar et al., 2016); the substitution models were as predicting by the "best model option" in MEGA and the Maximum Likelihood method was selected for inferring the tree topology. The model applied was as follows: Jones et al. w/freq. model [1] with discrete Gamma distribution to model evolutionary rate differences among sites (6 categories (+*G*, parameter = 1.3781)). The analysis for the *cob*-490 ORF data set involved 96 amino acid sequences. All positions with less than 90% site coverage were eliminated. There were a total of 270 positions in the final dataset.

2.1.5 *In silico* Intron prediction

Preliminary analysis of sequences from NCBI using MFannot and RNAweasel gave indication of introns being present in the *cob* (*cytb*) and *cox3* genes in *Ophiostoma ips* (GenBank accession number: NTMB01000349.1). Specifically in two regions complex introns were noted: *cob-490* and *cox3-640*. Introns are named according to their insertions sites with reference to the *Saccharomyces cerevisiae* homologues. In addition *cob* and *cox3* sequences from various fungi were aligned using MAFFT (Katoh & Standley, 2013). The E-INS-I algorithm was selected within MAFFT as it allows for short conserved sequences separated by long gaps, which facilitates CDS segments identification

by proper alignment of genes without the intron and with the genes having intron; a strategy that allows for prediction intron/exon junctions based on homology.

2.1.6. Intron RNA *in silico* folding

The DNA sequences were submitted to the online program RNAweasel (Lang et al., 2007; http://megasun.bch.umontreal.ca/RNAweasel/) and this program scanned the sequences for elements that are characteristic for group I and/or group II introns. In cases where RNAweasel detected a potential intron signature fold, the nucleotide sequence was submitted to the online program Mfold (Zuker, 2003; http://www.bioinfo.rpi.edu/~zukerm/rna/), which was used to model (fold) the intron's secondary structure. Intron folding utilizes the RNAweasel output for constraints but folding is also based on comparative approaches by comparing previously published RNA folds with the data at hand (Michel & Westfhof, 1990). Final intron folds were manually drawn using CorelDRAW Graphics Suite X6 (v14.0; Corel Corporation Limited).

2.1.7. Internal transcribed spacer (ITS) analysis

The internal transcribed spacer region (ITS1-5.8S-ITS2) nucleotide sequences obtained were aligned in AliView (MUSCLE option). This data set was generated for the *Ophiostoma ips* project to verify the identity of the fungal strains. Blastn was utilized to match sequences obtained from the strains studied with sequences deposited in GenBank.

2.2. Molecular Biology Methods

2.2.1. Culturing the fungi

All fungal strains were cultured in media plates containing 3% malt extract agar supplemented with yeast extract (MEAY; 30 g malt extract, 1 g yeast extract and 20 g bacteriological agar per 1 liter) at 20^{o}C for 8-10 days. For DNA extraction, fungal mycelium clumps were scraped from the MEA plates and these were transferred to 250 ml flasks containing 50 ml of Peptone Yeast Extract Glucose (PYG) medium (20 g of peptone, 5 g of yeast extract and 20 g of glucose per liter). The inoculated fungal cultures were still grown at 20^{o}C for 7-10 days. The fungal strains of *Ophiostoma ips* utilized in this study are listed in Table 4.

2.2.2. Fungal DNA extraction

Fungal mycelia were harvested from PYG broth by vacuum filtration using Buchner funnel and Whatman® #1 qualitative filter paper then transferred to a 15 ml Falcon tube (Corning Inc., Corning, NY) and stored at -20^{o}C. Frozen mycelia were crushed with acid washed sand and 2X CTAB buffer (2-5 ml) in a mortar and pestle in an extraction medium (1X: 1.5X: 2X proportion of mycelia: acid washed sand: CTAB buffer) until the formation of slimy slurry. The DNA extraction protocol is a modification of the CTAB based procedure described in Hausner et al. (1992). The slurry was collected into fresh 15 ml of Falcon tube and centrifuged at 3500 rpm for 5 minutes to sediment sand and cellular debris. The top aqueous layer was transferred to a new 15 ml Falcon centrifuge tube and mixed with 4 µL of RNase A (100 mg/ml; QIAGEN) to remove RNA and incubated in a 55-65°C

water bath for 1h or overnight with gentle mixing the tube. Tube was cooled to room temperature and chloroform extraction was carried out by adding chloroform (Thermo Fisher Scientific) to the tube in 1:1 ratio (equal volume) (to remove RNase, protein and cell debris). The tubes were mixed by inverting thoroughly until formation of an emulsion and centrifuged at 3500 g for 15 min to separate aqueous layer from organic layer. The chloroform extraction step was repeated twice followed by centrifugation until the aqueous layer was clear. Aqueous upper layer was transferred to new tube, added with 2.5 to 3 volumes of 95% ice-cold ethanol and mixed by gently inverting the tube. The tube was stored at -20°C for 2-3 hrs to enhance DNA precipitation. The DNA pellet was obtained by centrifugation at 3500 rpm for 20 minutes. The supernatant was discarded and the DNA pellet was rinsed with 70 % Ethanol subsequently the tubes were spun again and afterwards the tubes were kept open in an inverted position at room temperature for 15-30 minutes to evaporate residual ethanol. Finally the DNA pellets were resuspended in 200-300 µl DEPC treated autoclaved water and stored at -20°C.

2.2.3. Agarose gel electrophoresis and DNA quantification

The presence of DNA was confirmed by preparing agarose gel electrophoresis using ethidium bromide (EtBr) added to a 1% Agarose gel [5µl of EtBr (10 mg/ml) in 100 ml of gel] in Tris-borate-EDTA (TBE) buffer (TBE; 89 mMTris base, 89 mM boric acid, and 2 mM EDTA at pH 8.0). DNA samples were loaded onto the gel by adding (~2 to 33 µl) DNA sample premixed with 3 µl of 6X loading dye. The standard Invitrogen 1kb plus DNA ladder (ThermoFisher Scientific, Waltham, USA) was used for gel electrophoresis to serve as MW markers. The gel was visualized by exposing it to UV light using the Axygen gel documentation system (Corning, Corning, USA). The extracted

DNA was quantified with a NanoDrop 2000c UV-Vis Spectrophotometer (Thermo Fisher Scientific) and the DNA qualities were determined on the basis of the 260/280 and 260/230 OD ratio.

2.2.4. PCR Amplification

The primers for polymerase chain reaction (PCR) and DNA sequencing were synthesized by Alpha DNA (225 Bridge CP 4023, Montreal, Quebec, Canada, H3C 0J7) and are listed in Table 2.

2.2.4.1. PCR conditions for *Ophiostoma ips* complex intron project

Two sets of PCR primers (for *cob* and *cox3*) were initially designed for use with all *O. ips* strains and related taxa based on the *O. ips* sequence available in NCBI GenBank (Accession number: NTMB01000349.1). These primers were located in the exon sequences flanking the mitochondrial *cob*-490 (*cob* nucleotide position 490) and *cox3*-640 (*cox*3 nucleotide position 640) [nucleotide positions relative to *Saccharomyces cerevisiae* (Genbank accession number: CP006539.1, KP263414.1)]. The primer name and sequences are listed in Table 2.

PCR amplification of selected segments was carried out for each strain according to standard PCR protocols(Sambrook & Russell, 2001) using Gene DireX Taq polymerase (GenedireX, USA) or custom made Taq Polymerase (Jack Taq) according to manufacturer's instructions and using either BioRAD T100 thermal cycler (Bio-Rad Laboratories. Berkeley,USA) or Techne TC 512 thermocycler. The applied PCR cycles for all amplifications are listed in Table 3. The size of PCR amplified DNA fragments were determined by running samples in 1% agarose gel electrophoresis prior to purification of PCR products for generating sequencing templates.

2.2.5. PCR product purification

The selected amplified PCR products were purified using the protocol supplied in the QIAquick PCR purification Kit (QIAGEN, Hilden, Germany). The concentration and purity of the DNA samples was measured using Nanodrop 2000C Spectrophotometer mentioned above.

2.2.6. DNA sequencing of purified PCR products

The Sanger sequencing of purified DNA samples was facilitated by the Research Institute in Oncology & Hematology (RIOH sequencing) (675 McDermot Avenue, Winnipeg, Manitoba, Canada, R3E 0V9). Because of the relatively long lengths of some of the amplified DNA regions (up to 4.4 kb), multiple sequencing runs were required. For such samples, nested primers for subsequent sequencing runs were designed based on newly generated sequences (primer walking), allowing for the sequences of the amplified *cob*-490 and *cox3*-640 regions to be assembled as contigs (after sequence coverage of the whole region was attained). Chromatograms produced from Sanger sequencing results were visualized and converted to nucleotide sequences by SnapGene Viewer (Anonymous 2017), and sequences were assembled into contigs using the alignment program AliView (Larsson, 2014) and CAP3 (http://doua.prabi.fr/software/cap3, Huang & Madan, 1999).

2.2.7. Fungal RNA extraction and purification

The RNA was isolated from about 50-100 mg (wet weight) of fungal mycelium using the RNeasy Plant Mini Kit for total RNA isolation (QIAGEN, Hilden, Germany) following the manufacturer's instructions with some modifications as follow. Fungal mycelia were grown in PYG broth for 7 to

10 days and harvested by vacuum filtration. The mycelia were transferred to sterile small Petri plate, covered with 200µl of RNA*later* (QIAGEN) and frozen overnight at -80°C.

Frozen mycelia were transferred into a cooled (at -20°C for overnight) mortar and ground with a pestle in liquid nitrogen to a fine white powder. The lysis mix was prepared in a pre-cooled Eppendorf tube by mixing 450 µl of RLT buffer provided in kit with 4.5 µl of β-mercaptoethanol. Mortar was pre-cooled by adding liquid nitrogen in the Mortar several times. Frozen mycelium was placed into the cooled mortar and liquid nitrogen was added and the material was ground with a pestle. Liquid nitrogen was added during the process as needed and material was ground until a fine powder was obtained. The walls of the mortar were scrapped to collect all the 'powder' using a plastic inoculating loop. The ground up mycelium with the aid of a steel spatula was placed into the lysis mix containing Eppendorf tube. This tube was mixed by vortexing to generate the lysate for the RNA extraction steps. The RNA extraction was carried out with the RNA Easy Plant Mini Kit (QIAGEN) where the lysate was transferred to a QIA shredder column and centrifuged at 13,000 rpm for 2 minutes. The flow through was transferred to new 1.5 ml micro centrifuge tube. Chilled ethanol (96%) was added to the supernatant (half volume of the supernatant). The tube was immediately vigorously mixed. The sample was transferred to an RNeasy spin column (pink). The column was centrifuged at 10,000 rpm for 15 seconds(s). Seven hundred (700) µl of RW1 buffer was added to RNeasy spin column and centrifuged at 10,000 rpm for 15 s to wash the membrane. The flow through was discarded. Five hundred (500) µl of RPE buffer was added to RNeasy spin column, centrifuged at 10,000 rpm for 15 s, flow through was discarded. Again, 500 µl of RPE buffer was added to RNeasy spin column, centrifuged at 10,000 rpm for 2 min, flow through was discarded. Column was placed in new 1.5 ml tube and centrifuged for 1 min. Approximately 30 to 50

µl of RNase free water was added and the tube was centrifuged for 1 minute to elute pure RNA. The RNA was stored at -80 °C.

2.2.8. Turbo-DNase treatment for pure RNA

The contaminating DNA was digested using the TURBO-DNase kit (Applied Biosystems) following the manufacturer's directions. Briefly, a 20 µl reaction was set-up as follows: 15 µl purified RNA, 2 µl TURBO-DNase Buffer (10x), 2 µl RNase-free water and 1 µl TURBO DNase (2 units/µl). Reaction mixture was incubated at 37 °C for 30 min, followed by incubation at 75 °C for 10 min to deactivate the DNase enzyme. This whole process was repeated twice to confirm the complete absence of DNA. The presence of DNA was tested by regular PCR using gene specific primers and genomic DNA as a positive control. The TurboDNase treated DNA free RNA was stored at -80 °C.

2.2.9. Complementary DNA (cDNA) synthesis using ThermoScript Kit

Complementary (c)DNA synthesis was performed using the ThermoScript RT-PCR system (Invitrogen) following the manufacturer's instructions. Briefly, a 12 µl reaction was set-up containing 7 µl RNase-free water, 2 µl purified RNA, 1 µl primer (40 µM; Lsex2 or IP1R) and 2 µl dNTP mix (10 mM) and the mixture was incubated at 65 °C for 5 min. After incubation, following reagents were added to previously made mixture; 4 µl cDNA synthesis buffer (5x), 1 µl dithiothreitol (DTT; 0.1 M), 1 µl RNase-free water, 1 µlRNaseOUT (40 units/µl), 1 µl Reverse Transcriptase (15 units/µl, Thermoscript). After that the reaction mixture was incubated at 50 °C for 1 hour, then 85 °C

for 5 min. Finally, 1 µl of RNase H was added to degrade the RNA template and for this step the tube was incubated at 37 °C for 20 min. The cDNA was stored at -80 °C.

2.2.10. Reverse Transcriptase PCR amplification and cDNA amplification

The cDNA was further PCR amplified in a 50 µl reaction volume in a PCR tube using the Platinum Taq DNA polymerase kit (Invitrogen) as follows: 38.1 µl RNase-free water, 5 µl PCR buffer without $MgCl_2$ (10x), 1.5 µl $MgCl_2$ (50 mM), 1 µl dNTP mix (10 mM), 1 µl of forward primer (10 µM), 1 µl of reverse primer (10 µM), 2 µl cDNA and 0.4 µl of Platinum Taq DNA polymerase (5 units/µl). For the positive control, cDNA was replaced by genomic DNA. The tube was mixed gently and incubated at 94 °C for 2 min, followed by 40 cycles of PCR cDNA amplification with the optimized conditions for each sample (see Table 3). For RT PCR assays two controls were applied: (1) one tube was prepared with standard PCR reagents (DNA as a template and no RT) for a positive control and one tube was prepared using RNA as a template (no RT treatment) as a negative control. The PCR products were purified as described above and sequenced. Sequences for cDNAs were compared with sequences representing the genomic version of the loci and sequences absent from cDNAs were assumed to be intronic sequences.

CHAPTER 3: Applications for fungal mitochondrial genomes:

Fungal phylogeny using mtDNA amino acid sequences

3.1. Abstract

Phylogenetic trees can provide information about fungal evolution and taxonomy. Many fungal groups are still poorly resolved with regards to their taxonomic status based on morphological data and rDNA studies. Hence this study was focused on generating a tree with representative fungal strains from various fungal taxonomic groupings. Databases such as NCBI and MitoFun were explored to extract fungal mitochondrial genomes in order to obtain concatenated mitochondrial amino acid sequences for the protein coding genes. In addition to the phylogenetic analysis this study also collected information such as mitogenome sizes, GC% contents and their taxonomic positions. This study will provide foundational information regarding the use of mitogenome data for fungal taxonomy and mitogenome evolution.

3.2. Introduction

Fungi are economically important microorganisms because of their pathogenicity, ability to produce antimicrobial compounds/enzymes, ecological significance as decomposers, applications in food production and biotechnology. Hence, there is always a need of ongoing research on basic fungal evolution and physiology. The fungal mitochondrial DNA can be linear or circular and there are about a hundred of copies of mitochondrial DNA per organelle and there can be several mitochondria per cell. The mtDNA is thought to replicate through a rolling circle type mechanism (Maleszka et al., 1991). The mitochondrial DNA of fungi can be highly variable in size because of the presence of introns and intergenic regions. The mitochondria possess a core set of genes which are required for oxidative phosphorylation, energy production, and translation. The molecular data like fungal nuclear and mitochondrial sequences have been often used to generate fungal phylogeny for several decades (Hausner et al., 1992; James et al., 2006; Hibbett et al., 2007). Phylogenetic analysis is based on the premise that related species contain similar DNA, RNA, and protein sequences, while more distantly related species will not. Fungal classification is useful because that facilitates grouping of related organisms into taxons and one may be able to make predictions with regards to certain characteristics on the basis of the taxonomic position a particular species occupies.

A phylogenetic tree based on mitochondrial core gene amino acid sequences might be beneficial to provide additional insight into fungal evolution besides trees based on nuclear markers. Nuclear markers like ITS regions are very popular with regards to molecular identification and phylogenic tree construction for fungi. However, ITS sequences are relatively short, thus limiting with regards to the number of informative sites, and they appear to evolve too fast to be of use for higher level taxonomic studies. Hence, ITS sequences among different fungal phylums (or even genera) cannot be aligned. Also, 16S and 18S ribosomal sequences have been used to generate

phylogenetic trees but there is lack of satisfactory resolution (i.e. node support values can be quite low). Various molecular markers for fungal identification and phylogenetic trees have been developed over the last 20 years and those include RNA polymerase, calmodulin, actin, elongation factor, and beta tubulin sequences. Because, many fungal phylogenies that are currently available at present are not well resolved at many levels (lack of high statistical support) there is a need for more markers. Recent genomic data is being explored for resolving taxonomic issues but comparative genomics can be challenging for larger genomes and there is still an active debate on what components of fungal genomes should be applied for resolving taxonomic questions (Spatafora et al., 2017). Therefore the smaller genome housed within the mitochondria might be an alternative to single nuclear markers, or multigene phylogenies based on nuclear genomes. The size of fungal mitogenomes can range from ~ 12 kb to > 230 kb and this makes them plausible targets for next generation sequencing (NGS). In addition mitogenomes are "side products" of fungal genome products so more fungal mitogenomes appear in public databases such as NCBI and JGI. (MycoCosm, Grigoriev et al., 2014).

The availability of more fungal mitogenomes has given this study impetus to explore the possibility of using mitochondrial core gene amino acid sequences for generating a phylogenetic tree. The 13 core gene sequences of each organism were collected that included two subunits of ATP synthase (*atp6, atp8*), apocytochrome b (*cob*), three subunits of cytochrome c oxidase (*cox1, cox2, cox3*) and seven subunits of reduced nicotinamide adenine dinucleotide ubiquinone oxidoreductase (*nad1, nad2, nad3, nad4, nad4L, nad5* and *nad6*). There were two goals for this study;

1. To generate a comprehensive phylogenetic tree using mitochondrial protein sequences from the four major groups of fungi

2. To get some insights of the similarities or the differences among fungal mtDNAs with regards to size, GC content and gene content.

3.3. Materials and Methods

3.3.1. Sequence compilation for generating a phylogenetic tree

The project started by building on earlier work of compiling mtDNA genomic sequences. Sequences were extracted from NCBI and MitoFun (http://mitofun.biol.uoa.gr/). Thirteen different mitochondrial protein coding (CDS) sequences for all examined fungal sequences available in NCBI and Mitofun were collected. The analysis was based on concatenated amino acid sequences of 13 mitochondrial proteins encoded by the following genes (*atp6, atp8, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5,* and *nad6*) that were collected and arranged in alphabetical order of gene names.The whole compiled dataset was divided into three parts, Ascomycota, Basidiomycota, Zygomycota and Chytridiomycota. All of the collected sequences were combined in one file. Some mtDNA sequences that were not available as annotated files from NCBI were annotated in this study using MFannot and these annotations were manually checked for accuracy by using features of the BLAST suite of programs (NCBI) during the collection process. The *atp*9 and *rps*3 genes were not included in our phylogenetic analysis as they are not universal in their occurrence in fungal mitochondrial genomes. Total 205 amino acid sequences of different fungal strains were collected. In addition to collecting sequences, sizes of genomes, GC content, and features such as gene content etc. were tabulated.

3.3.2. Alignment of sequences

Concatenated amino acid sequences for the sampled mtDNA protein sequences were aligned with the MAFFT alignment program (https://mafft.cbrc.jp/alignment/server/) using the E_INS_I option. The aligned sequences were observed in the GENDOC program (Nicholas et al., 1997) and here the alignments were adjusted if required.

3.3.3. Phylogenetic analysis

The phylogenetic tree was generated by the W-IQ-TREE program (Trifinopoulos et al., 2016) in which the aligned amino acid sequence data set was used to generate phylogenetic tree using maximum likelihood criteria and by applying default settings. This program was used as the data set was too large to be analyzed by programs compiled within MEGA or by MrBayes.The bootstrap option (1000 iterations) was implemented to estimate node support values. Sequences from *Allomyces macrogynus*, a species producing motile spores which is viewed as an early branching fungus was used as the outgroup to generate the tree.

3.3.4. Generating synteny maps from fungal sequences

In order to generate synteny maps "gff formatted sequences" were collected from NCBI. The pattern of gene order was observed and noted for members of each fungal Order. The protein and rDNA coding genes were used to generate synteny maps. The Table 1 summarizes the number of sequences selected for each fungal Order group. The synteny maps were obtained for all Orders of the Ascomycota. The diversity of fungi surveyed generated a great diversity of gene maps, thus the

project had to focus on selected Orders which appears to have conservation among their gene syntenies. Significant observations were noted for four fungal Orders (Hypocreales, Glomerallales, Microascales and Ophiostomatales) and these were compared in more detail. The CREx (Common interval Rearrangement Explorer) analysis program (http://pacosy.informatik.uni-leipzig.de/crex) was used to obtain synteny maps and to find out the consensus gene blocks.

3.4. Results

3.4.1. Trends observed in phylogenetic tree

The mtDNA based fungal phylogeny (Figure 1 and Supplementary figure S1) reflected some known trends of fungal evolutionary history, with all fungi grouping within their respective phyla and within their recognized orders (James et al., 2006; Schoch et al., 2009). The node support values were consentingly high for the majority of nodes suggesting mtDNA may provide stable tree topologies for resolving deeper fungal phylogenetic relationships. The tree was routed with *Allomyces macrogynus* fungus because it is viewed as a basal fungus by forming motile spores (Bullerwell et al., 2003; Bullerwell & Gray, 2004). Trees were also generated with MrBayes but it took seven days to perform the analysis on a PC desktop computer using a LINUX operating system,overall topology of the trees generated from this program were identical to that generated with W-IQ-TREE.

The study also collected GC data and it was noted that GC values ranged from 16.5% to 46.7% (Table 1). The data also showed that mtDNA size varies greatly among the fungi (~12 kb to

>200 kbp) and this is in part due to the presence of intergenic spacers and in some cases large numbers of introns (group I and group II type introns).

In addition, gene content was also recorded and the presence of the *rps*3 gene was somewhat sporadic (absence/ presence) among members of the Chytridiomycota and the Zygomycota while among the Ascomycota and Basidiomycota the *rps*3 gene was present in most members except for early branching members of the Ascomycota. In the filamentous ascomycetes fungi, it was frequently noted that a group IA intron located within the *rnl* (large ribosomal subunit) gene encoded *rps*3 gene. The *atp*9 gene also appears to be absent in some members of the Ascomycota. The fungal members of the Pleosporales lacked the *atp*8 gene. *Saccharomyces cerevisiae* and other members of the Saccharomycetes did not have the *nad* set of genes. Some members of Zygomycota and all the members of Chytridiomycota (which are considered as lower fungi) did not appear to have the *rps*3 gene. (Although sequence similarity can be low and annotation programs can miss the presence of this gene; (Bullerwell et al., 2000).

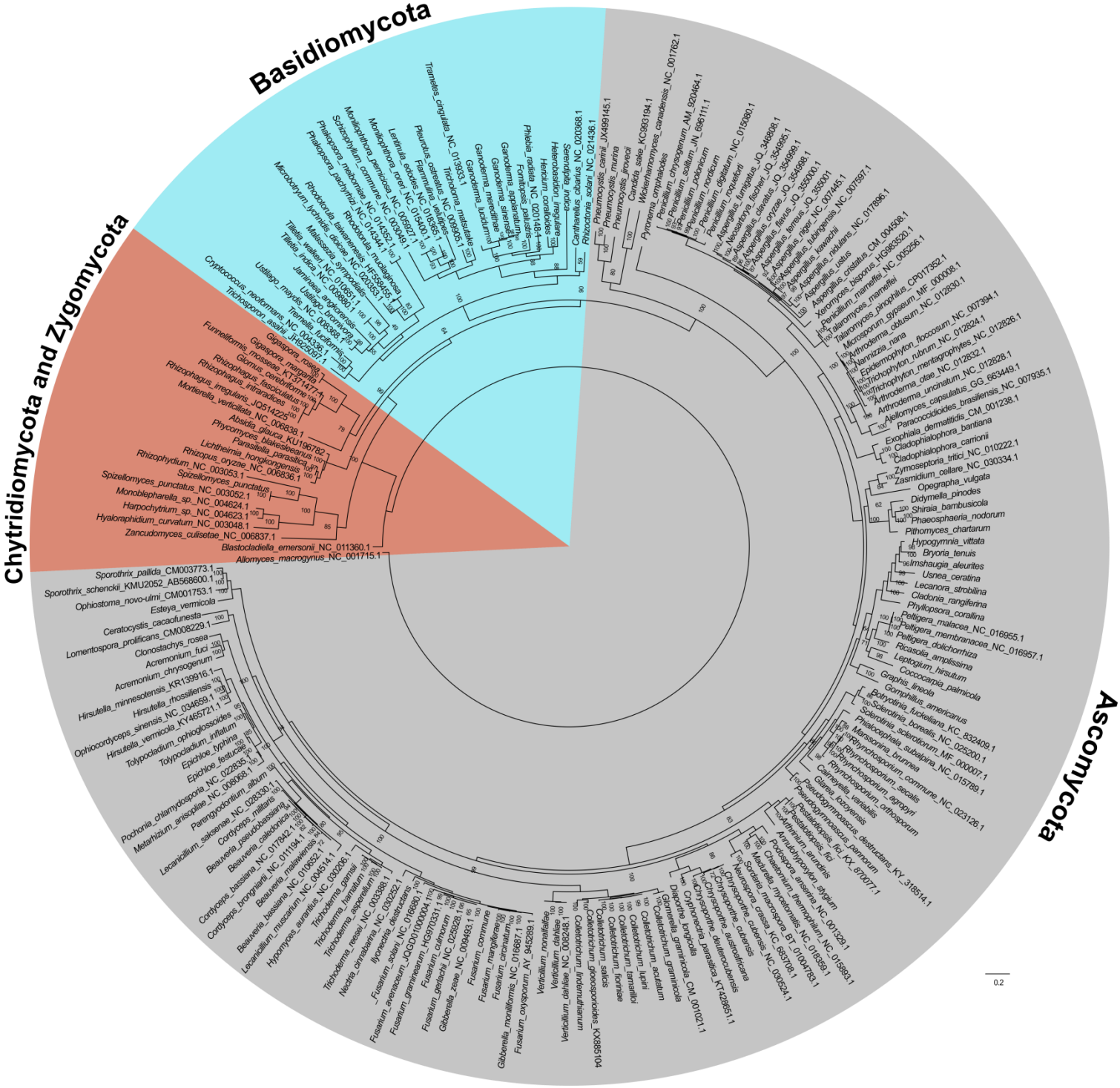

Basidiomycota
Chytridiomycota and Zygomycota
Ascomycota
0.2

Figure 1. Phylogenetic tree based on concatenated mitochondrial amino acid sequences generated by W-IQ-TREE applying default settings and using 1000 bootstrap iteration that provide node support values. The bar of 0.2 is a scale bar and it represents the probability of substitution rate. The grey and blue color indicates the Ascomycota and the Basidiomycota phylum respectively. The orange color represents the Cytridiomycota and the Zygomycota phyla combinely.

Fungal Phylum	Fungal Order	Number of sequences per Order	Size range of mtDNA (kb)	Range of G/C% content
Ascomycota	Diaporthales	5	53.4–190.8	27.7–44.1
	Microascalales	2	23.9–103.1	26.8
	Eurotiales	22	24.6–77.6	24.6–28.2
	Glomerellales	12	26.1–55.1	24.9–28.2
	Helotiales	11	27.1–70.3	26.3–29.7
	Hypocreales	40	24.5–103.8	25.9–33.1
	Lecanorales	7	32–84.2	28.4–38.9
	Ophiostomatales	4	27.1–66.3	24.4–25.1
	Ostropales	2	24.9-28.3	28.4-29.5
	Peltigerales	6	51.1-120.9	24.8-29.8
	Pleosporales	5	39-68.9	25.2-31.9
	Pneumocystis	2	24.6-35.6	25.5-29.8
	Xylariales	4	48.9-133.7	28.4-29.9
	Chaetothyriales	3	26-26.8	24.5-25.9
	Onygenales	10	23.5-71.3	23.4-24.7
	Saccharomycetales	2	26.2-27.7	18.1-25.7
	Pneumocystidales	1	26.1	29.8
	Arthoniales	1	38.9	30.8
	Capnodiales	1	23.7	27.8
	Pseudeurotiaceae	1	26.9	28.1
Basidiomycota	Agaricales	7	49.7-121.3	16.5-31.9
	Cantharellales	2	23.5-58.6	26.8-35.9
	Polyporales	7	60.6-156.3	24-31.2
	Pucciniales	2	31.8-32.5	34.5-34.9
	Russulales	2	72.9-114.2	18-22.8
	Tilletiales	2	59.4-65.1	28.8
	Tremellales	3	24.8-35.1	28.9-34.9
	Ustilaginales	3	56.8-177.5	31.2-36
	Microstromatales	1	30	32.2
	Microbotryales	1	107.8	33.7
	Malasseziales	1	38.6	32.1
	Sebacinales	1	63.7	26.3
	Sporidiobolales	2	47	40.4
Zygomycota	Glomerales	5	59.6-134.9	37.2-46.7
	Mucorales	4	31.8-83.4	26.1-34.1
	Herpellales	1	58.6	18.5
	Mortierellales	1	58.7	27.8
	Phycomyces	1	62.1	37.9
	Gigaspora	2	97-97.4	44.8-45
Chytridiomycota	Spizellomycetales	1	58.8	31.8
	Blastocladiales	2	36.5-57.5	35.1-39.5
	Monoblepharidales	3	24.2-60.4	36.2-43.2
	Rhizophydiales	1	68.8	22.9

Table 1.Summary of findings such as mitochondrial DNA size and GC content for the certain number of fungal members for a specific fungal Order.

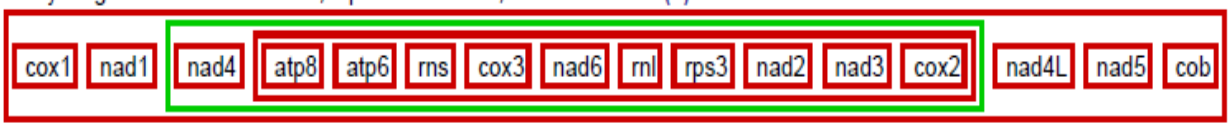
3.4.2. Gene synteny among four Orders of the Ascomycota

There are some reports on gene syntenies across the fungi with regards to the mtDNA gene order but the results are highly variable due to diversity of fungi sampled. In this study we attempted to look for gene synteny in specific groups (Orders) with the intention to find blocks of genes which appear to be linked together. We did find certain blocks of genes among four orders of the Ascomycota. The reason for linkage could be that sets of genes share a common promoter; thus they are always expressed together. Genes can be transposed to new locations but they either have a promoter or they get incorporated into a different transcriptional unit. It has been noted that at higher taxonomic ranks fungal mtDNA gene synteny can be quite variable (Aguileta et al., 2014 ; Wu & Hao, 2014) probably due to recombination events mediated by short repeats (Auileta et al., 2014) and/or by the presence of homing endonucleases that due to their nicking and cutting activity can promote recombination (Wu & Hao, 2014).

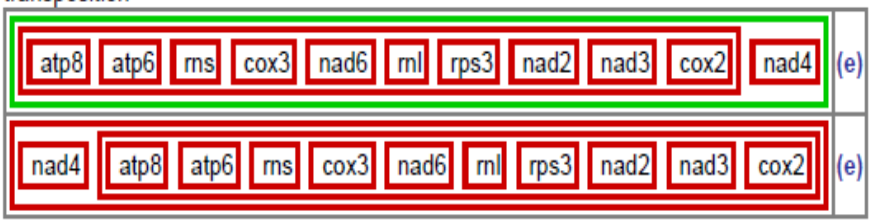
Hypocreales → Glomerellales, Ophiostomatales, Microascalales

- family diagram for Hypocreales (e)

- family diagram for Glomerellales, Ophiostomatales, Microascalales (e)

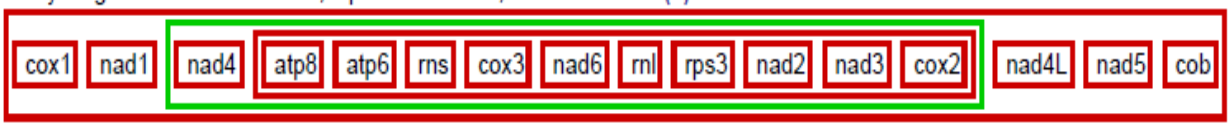


- scenario:
  - transposition



Glomerellales, Ophiostomatales, Microascalales → Hypocreales

- family diagram for Glomerellales, Ophiostomatales, Microascalales (e)



- family diagram for Hypocreales (e)

- scenario:
  - transposition

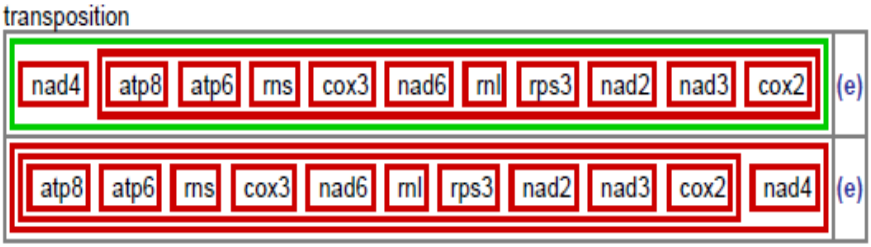


Figure 2: A Family diagram showing the mtDNA gene synteny of four Orders of Ascomycota group. It is showing how the arrangement of genes in the Hypocreales is different from those observed in the other three fungal Orders. The synteny figure highlights that *nad4* gene arrangement is different and the potential reason for that is transposition of that gene that resulted in different space allocation for the *nad4* gene. Red blocks show the gene blocks that are similar in all orders because they are conserved. The green blocks show the gene(s) variation that could be a result of transposition events whereby a gene was relocated to a new site.

3.5. Discussion

Exploring in more details fungal mitochondrial DNA is important in order to gain a better understanding for fungal evolution, RNA splicing (introns), catalytic RNA (ribozymes), transposition/recombination, nuclear mitochondrial interactions and exchange of genes (movements of mtDNA genes to the nuclear genome ; e.g. *rps*3, some ATPase genes etc.), intracellular communication (between nucleus and organelles) and mtDNA as a resource for the development of enzymes that have applications in biotechnology (homing endonucleases and ribozymes) In addition fungal mitochondrial genomes are essential for most strictly aerobic fungi as they are needed to maintaining mitochondrial integrity, an essential organelle for energy production and various biosynthetic pathways. Recent advances in high throughput sequencing technologies (i.e. NGS) provide opportunities to explore mitochondrial genome evolution.The mtDNA-based fungal phylogeny is in general agreement with currently accepted trends in fungal evolutionary history (Marcet-Houben & Gabaldón, 2009; Ebersberger et al., 2012). The fungal species sampled in this study grouped within their expected Orders/Families and overall node support values were consistently high (~ 90%) for the majority of nodes recovered in the phylogenetic analysis. This suggests that mtDNA may provide stable tree topologies for resolving deeper fungal phylogenetic

relationships. This could be valuable when assessing taxonomic status for unknown fungi that have few morphological characters or are highly derived by having adapted to very specific environments such as plant and animal pathogens or symbionts of insects or plants/algae. Currently, most fungal phylogenetic investigations are based on the analysis of one or a few nuclear loci, mtDNA offers a larger pool of genes that can be analyzed. Also, because of generous availability of sequencing technologies and computational biology tools, total more than 800 fungal genomes have been sequenced (For example; http://genome.jgi.doe.gov/fungi) (Spatafora et al., 2017). Hence, the phylogenies based on mitochondrial genes might be suitable alternative for resolving problems in terms of fungal taxonomy and evolution.

3.6. Research Importance and Future Goals

This study can be the foundation for a large scale review of fungal mitochondrial genomes. A more detailed analysis would be useful that could identify potential genetic factors such as repeats and mobile introns/elements that could promote mtDNA rearrangements that can explain the variable gene orders observed among the fungi. In addition, further work can be done on "missing" mtDNA genes (*atp*9 and in some cases *rps*3) that can provide a link between the nuclear and the mitochondrial genome. In a recent study by Wai et al., (2019) it was shown that *rps*3 can move from the mitochondrial genome to the nuclear genome. The same study also showed that *rps*3 in some cases moved from intergenic spacers (i.e. *rps*3 free standing) to be inserted into introns (*rnl*) and possible from an intronic location back to intergenic locations. Further detailed analysis can be done to assess the contribution of mtDNA introns, intron encoded ORFs, inserted plasmid components, variable gene content, intergenic spacers, and repeats, on mitogenome organization and size variations (Hausner, 2002; Hausner, 2012).

CHAPTER 4: A study exploring novel complex introns in the *cob* and *cox3* genes located in the mitochondrial genome of various strains of *Ophiostoma ips*

This Chapter is in preparation for a publication and Jordan Perillo will be a co-author. The RNA folding was to large extent performed by Perillo as part of this MBIO 4030 Honours Project. We also acknowledge Alvan Wai for his insights on folding intron RNAs.

4.1. Abstract

The mitochondrial genomes of fungi contain mobile group I and group II introns that are self-splicing and their transpositions are facilitated by the homing endonuclease or reverse transcriptase proteins encoded by them. In rare cases, these introns are found in twintron or tandem intron arrangements because of the insertion of one intron within or beside an intron catalyzed by intron encoded proteins. Hence, additional models are needed to better understand the formation, structures and splicing of these complex introns. The purpose of this study was to annotate and model the fourth intron in the mitochondrial cytochrome b gene (*cob* I4) and the second intron in the mitochondrial cytochrome oxidase III gene (*cox3* I2) of *Ophiostoma ips* (GenBank accession number: NTMB01000349.1) by means of computational modeling and comparative sequence analysis. The intron exon boundaries were also determined by RT- PCR and sequencing the whole intron. Thus, this study provides *in vivo* and *in vitro* studies to evaluate the presence of complex introns and a plausible model for the mode of splicing of these two complex introns. For *cox3* I2, a tandem intron model containing two homing endonuclease ORFs was proposed, splicing was observed to happen as a composite unit, and the downstream intron was presumed to be the native intron. For *cob* I4, a three-component self-splicing intron model is proposed , this complex intron was predicted to encoded two homing endonuclease ORFs. The information gathered from this study might be useful for biotechnological application purposes and will give valuable insights into the evolution of fungal mitochondrial genomes.

Keywords: *Ophiostoma ips*, fungal mitochondrial DNA, novel complex introns, RNA modeling, Homing endonucleases.

4.2. Introduction

The fungus *Ophiostoma ips* is vectored by bark beetles and causes blue stain on sapwood in conifers (Kim et al., 2003; Pastirčáková et al., 2018). The mitochondrial genomes of ophiostomatoid fungi are of great interest because of the presence of self-splicing mobile introns (ribozymes) and the open reading frames (ORFs) that they encode (Hafez et al., 2013). Introns and HEGs are considered to be neutral or non toxic components for the fungal mtDNA. The host genome appears to be quite tolerant to these mobile elements, although these elements may influence regulatory events for gene expression and require factors for intron splicing.

The proteins coded by the intron encoded ORFs are referred to as intron encoded proteins (IEPs), and their functions can vary in accordance with the classification of the intron that hosts the ORF (Hafez et al., 2013; Wai et al., 2019). Introns are mainly divided into two groups; group I and group II introns, which differ in their splicing mechanisms, conserved sequences and catalytic core secondary structures (Michel & Westhof, 1990; Landthaler & Shub, 1999; Candales et al., 2011). Group I IEPs are usually LAGLIDADG or GIY-YIG type of homing endonucleases, named based on conserved amino acid motifs. Group II IEPs are usually reverse transcriptases;both types of IEPs promote site-specific intron transposition and have applications in biotechnology (Mullineux et al., 2010; Hafez & Hausner, 2012; Hafez et al., 2013).

Sometimes it is difficult to characterize fungal mitochondrial introns due to the ability of IEPs to transpose introns into pre-existing introns (Hafez & Hausner, 2015; Deng et al., 2016). The twintron (intron within an intron) and tandem (side- by-side) intron models have been reported previously in the literature (Copertino & Hallick, 1991; Hafez & Hausner, 2015; Deng et al., 2016) and these types of complex introns encode IEP in both types of arrangements. This study includes for the first time a report for a ‘Trintron’ (a complex intron made of components derived from three

introns). Hence, there is always a need to gain a better understanding of such complex introns with regards to their RNA folds and the possible splicing mechanisms.

In terms of splicing for twintrons,a mechanism has been proposed where intron components can splice individually but the internal components splices first allowing the "outer" components to achieve a splicing competent RNA fold (Copertino & Hallick, 1991) and this model might be applicable to other complex intron models. However, a recent review showed that many different types of "nested" intron arrangements exist and there might be many mechanisms whereby they can be spliced out (Hafez & Hausner, 2015). In general, ribozyme-type introns maintain secondary structures that require the formation of conserved intron helices and these have to be maintained in even composite arrangements (Michel & Westhof, 1990; Deng et al., 2016) allowing for splicing competent folds to be generated.

Group I intron secondary RNA structures form up to ten helical regions (P1 to P10) (Michel & Westhof, 1990), whereas group II introns are expected to fold into six helical domains (I -VI) radiating from a central linker sequence (Lehmann & Schmidt, 2003). Helical regions in both intron types are the structural elements necessary for achieving ribozyme activity and ultimately self-splicing (Winter et al., 1990; Michel et al., 1992; Lehmann & Schmidt, 2003; Hausner et al., 2014), although splicing appears to be assisted by intron encoded factors (maturases) or host factors such as RNA helicases or RNA chaperones (Hausner, 2012).

Preliminary analysis of the mitochondrial genome of *Ophiostoma ips* (GenBank accession number: NTMB01000349.1) indicated the presence of novel complex introns in the cytochrome b (*cob*) and cytochrome oxidase III (*cox3*) genes. Hence a total 55 *Ophiostoma* strains were selected for screening for these introns. The purpose of this study was to annotate and model the fourth intron

present in the mitochondrial *cob* gene (*cob* I4) and the second intron present in the mitochondrial *cox3* gene (*cox3* I2), of *Ophiostoma ips* (GenBank accession number: NTMB01000349.1), using various computational tools along with DNA sequencing and homology-based comparative sequence analysis.

4.3. Summary of methods

Please refer to Chapter 2 for detailed descriptions of the following methods: fungal culturing, DNA extraction, PCR amplification, RT-PCR, intron survey, DNA sequencing, intron annotation and modeling. For details with regards to primers sequences and PCR conditions applied to this study refer to Table 2 and Table 3.

No.	Species and strain (WIN_M) number	Amplified region with description	Primer Name	Orientation	Primer Sequence (5'→3')
1	*Ophiostoma ips*	*cob*-490, screening* primers	Ipscytb490-F1 Ipscytb490-R1 Ipscytb490-F2 Ipscytb490-R2	Forward Reverse Forward Reverse	GTGCTATACCTTGAATTGG GATCCAGCTGTATCATGAAG GCTACAGTGATTACAAACC CCCCTAAAGCATTAGGCATG
2	*Ophiostoma ips* (1635)	*cob*-490 of 1635 species,	1635 IPSCOB-F3 1635 IPSCOB-R3 1635 IPSCOB-F4 1635 IPSCOB-R4 COBF-1635w2 COBR-1635w2	Forward Reverse Forward Reverse Forward Reverse	GGATTTACATGTTCAGAAG TATGTAATCTCTGTTTAATTTG TCTGGATGGTACAAAACAG GTAGGTAGTTCATGTAAATCTG CTGGATGGTACAAAACAGG GGTAGTTCATGTAAATCTGC
3	*Ophiostoma ips*	*cob*-490 1.3 kb intron	IPS1.3COB-F3 IPS1.3COB-R3	Forward Reverse	GTAGGATTCACWTSTTCAG ACTCACTCATTAGTCGTTGAACG
4	*Ophiostoma ips*	*cob*-490 second walk sequencing primers	IPSCOB-F3 IPSCOB-R3 IPSCOB-F4 IPSCOB-R4	Forward Reverse Forward Reverse	GGATTCACATGTTCAGAAGG TATCATAGGATGATGACCTG TTACATACAAATTTATTTGAAGC TATAACTGATTGTACATCAG
5	*Ophiostoma ips* (1639)	*cob*-490 of 1639 species	COBF-1639w2 COBR-1639w2	Forward Reverse	GATTCACATCTTCAGAAGG CCGAATTAAGTTCCGCTAC
6	*Ophiostoma ips*	*cob*-490	IPSCOB-F5 IPSCOB-R5 IPSCOB-F6 IPSCOB-R6	Forward Reverse Forward Reverse	GGATATTGAATTAGGAACG TGCTATACCAAGAATATGGAAG GTCTATGCAGAGAGTTCAG ACGCGAATTATAATGTACAG
7	*Ophiostoma ips*	*cob*-490, RT PCR primers	RTCOB-7F RTCOB-11Fgp2 RTCOB-6 RTCOB-8R	Forward Forward Forward Reverse	GACATAGTCTAATTATATTTG GCGTAAGCGAAGATGTGG CGCGTTATAGTAAATTATGC CAAATATAATTAGACTATGTC

8	*Ophiostoma ips*	*cox3*-640, screening* primers	Ipscox3I2-FL Ipscox3I2-R1 Ipscox3I2-F2 Ipscox3I2-R2	Forward Reverse Forward Reverse	CAAGAAGTTGAATATGATC CAGTATGCAATAGCACCTTC CCAGCAGTATGAGGAGGATTAG CCAGTCATACAACATCTAC
9	*Ophiostoma ips*	*cox3*-640	IPSCOX3-F3 IPSCOX3-R3 IPSCOX3-F4 IPSCOX3-R4	Forward Reverse Forward Reverse	TACAATCTAATATCGAGGG TTAATCTGTATTTATTAATACG GAAGCTATTACATTGTTAGA TTCCTTCTAATGTAGTATG
10	*Ophiostoma ips*(1487)	*cox3* 640	IPS1487cox3F2	Forward	TTGATTAGCTGGATTTATAG
11	*Ophiostoma ips* (1488)	*cox3*-640	COX3F-1488W3	Forward	CAAATTATACAGTCGCTAATAC
12	*Ophiostoma ips* (1478)	*cox3*-640	COX3F-1478W2 COX3R-1478W2	Forward Reverse	GCTATTACATTGTTAGAATC CTTCTAATGTAGTATGATAAC
13	ITS amplification	ITS region	SSU Z$^{\Delta}$ LSU 4$^{\Delta}$ SS3$^{\Delta}$ LS2$^{\Delta}$ LR3$^{\Delta}$ ITS1$^{\Delta}$_F ITS4_R$^{\Delta}$	Forward Reverse Forward Reverse Reverse Forward Reverse	ATAACAGGTCTGTGATG TTGTGCGCTATCGGTCTC GTCGTAACAAGGTCTCCG GATATGCTTAAGTTCAGCG CCGTGTTTCAAGACGGG TCCGTAGGTGAACCTGCGG TCCTCCGCTTATTGATATGC

Table 2. List of Primers used for gene amplification by PCR and RT-PCR including regions like ITS, *cox*3-640 and *cob*-490. Screening* primers are primers designed to amplify the entire intron. The others were nested primers used for sequencing. The primers with $^{\Delta}$symbol were designed and used for ITS amplification to confirm fungal identity.

No.	PCR amplification goal	Cycle steps	Cycle temperature (ºC)	Cycle duration minutes (min), seconds (s)	Number (#) of cycles (of Denaturation, annealing and Extension)
1.	ITS amplification *Ophiostoma ips*	Initial denaturation	94	5 m	
		Denaturation	94	30 s	32
		Annealing	56	1 m	
		Elongation	72	2 m	
		Final extension	72	5 m	
2.	*Ophiostoma ips* cox3 640 twintron amplification	Initial denaturation	94	5 m	
		Denaturation	94	30 s	34
		Annealing	50	1 m	
		Elongation	68	3 m	
		Final extension	72	5 m	
3.	*Ophiostoma ips* cob 490 trintron amplification	Initial denaturation	95	5 m	
		Denaturation	95	30 s	34
		Annealing	55	1 m	
		Elongation	68	4 m	
		Final extension	72	5 m	
4.	*Ophiostoma ips cox3* 640 cDNA amplification by one step cox3 platinum RT PCR kit	cDNA synthesis	50	30 m	
		pre-denaturation	94	2 m	
		Denaturation	94	30 s	40
		Annealing	50	1 m	
		Elongation	68	3 m	
		Final extension	68	5 m	

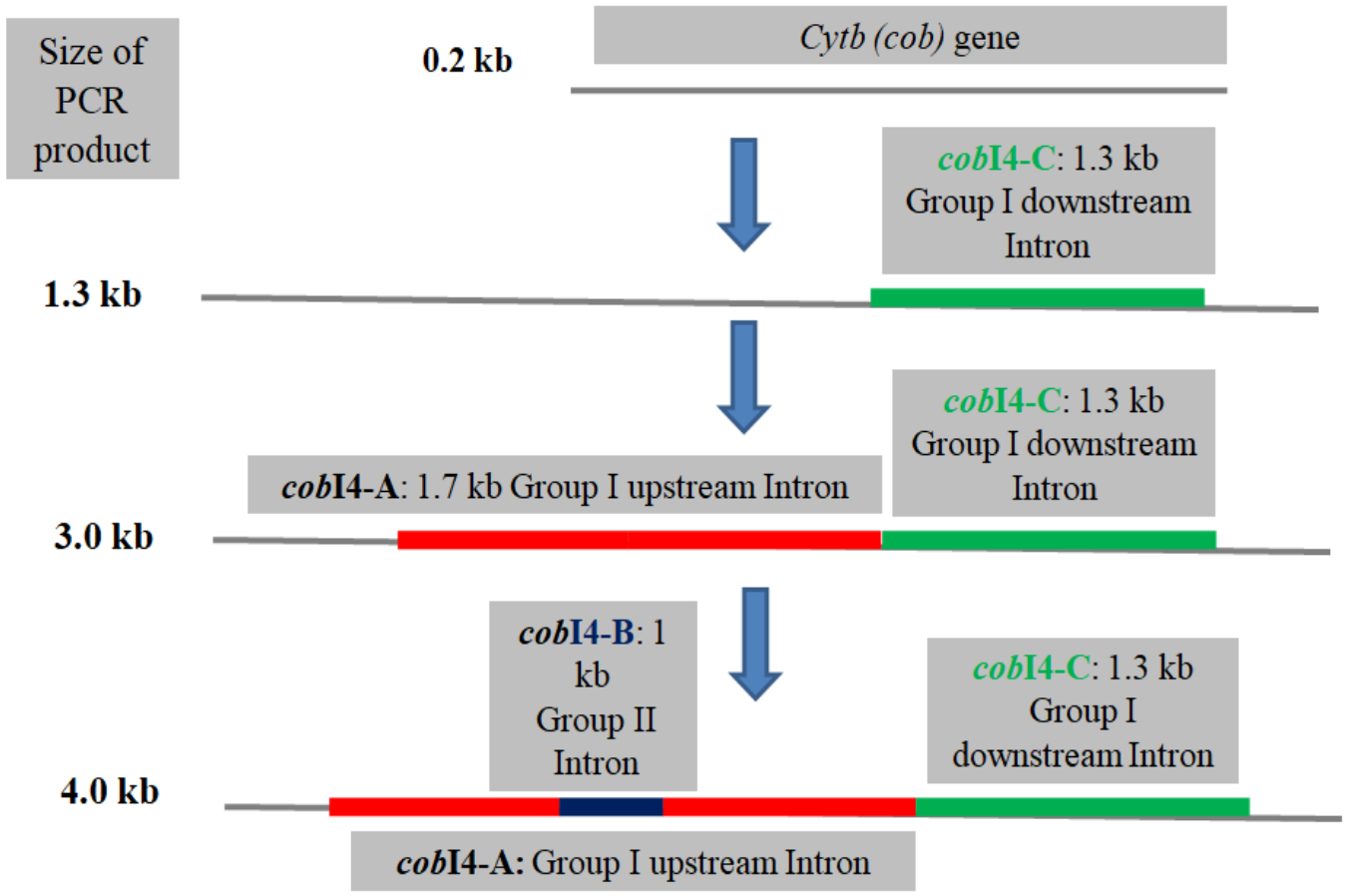
5.	*Ophiostoma ips cob* 490 cDNA amplification by one step platinum RT PCR kit	cDNA synthesis	55	30 m	40
		pre-denaturation	94	2 m	
		Denaturation	94	30 s	
		Annealing	55	1 m	
		Elongation	68	4 m	
		Final extension	68	5 m	

Table 3. A summary of the steps (cycling conditions) used for the PCR programs applied to amplify the *cob*-490 and *cox*3-640 introns and ITS regions. ITS sequences were amplified for fungal identification and RT-PCR amplifications were performed to confirm intron exon boundaries and potential splicing pattern of these introns.

4.4. Results

4.4.1. *cob* I4 annotation and modeling

The insertion site for *cob* I4 was confirmed as being *cob*-490 relative to the *S. cerevisiae cob* sequence (GenBank accession number: CP006539.1) with the total length of the intron being 3367 nucleotides (~3.4 kb). Similar sequences were extracted from NCBI data bases with the BLAST suite of programs. The sequences immediately upstream and downstream to *cob* I4 were designated as exon sequence that was determined using MFannot analysis (Beck & Lang, 2010) and alignment and sequence comparisons with intron-less versions of the mitochondrial *cob* genes. In order to examine how common complex introns are in *O. ips* and related species a total of 56 *Ophiostoma* strains were selected and screened for the presence of complex introns in the *cox3* and *cob* gene. The Table 4 shows the strains examined and the results obtained. The *cob* data showed there are four possible states for this region: no intron (0.2 kb PCR product), and 1.3, 3.0 or 4.0 kb products. The latter three would suggest intron insertions of various complexities. For the *cox*3 data we noted three types of PCR products, 0.2 kb products were expected sizes if there are not insertions (i.e. no intron); 1.4 and 3 kb products would suggest instances with only one intron being present (1.4 kb) and sites with more complex arrangements (3 kb).



cob*I4 490 "Trintron" in *Ophiostoma ips

Figure 3: Schematic presentation of intron gain and possible splicing pattern for the *cob* I4 490 intron.

Number	WIN_M collection number	Fungal identity	Geographic Origin	*cob* I4 intron Band Size (kb)	*cox3* I2 intron Band Size (kb)
1	1478	*Ophiostoma ips*	TB	4	3
2	1479	*Ophiostoma ips*	TB	4	3
3	1480	*Ophiostoma ips*	TB	4	3
4	1481	*Cladosporium sp.*	TB	N/A	N/A
5	1484	*Cladosporium sp.*	JP	N/A	N/A
6	1486	*Ophiostoma ips*	BC	4	3
7	1487	*Ophiostoma ips*	US	0.2	1.4
8	1488	*Ophiostoma ips*	US	4	3
9	1489	*Ophiostoma ips*	US	4	3
10	83-d	*Ophiostoma ips*	NO	N/A	N/A
11	83-q	*Ophiostoma ips*	NO	4	N/A
12	87-A	*Ophiostoma ips*	NO	4	0.2
13	88-508	*Ophiostoma ips*	NZ	4	0.2
14	88-141	*Ophiostoma ips*	NZ	4	N/A
15	88-138	*Ophiostoma ips*	NZ	4	N/A
16	88-135	*Ophiostoma ips*	NZ	4	N/A
17	88-134	*Ophiostoma ips*	NZ	N/A	N/A
18	88-131	*Ophiostoma ips*	NZ	N/A	N/A
19	88-105	*Ophiostoma ips*	NZ	4	N/A
20	88-100b	*Ophiostoma ips*	NZ	4	N/A
21	92	*Ophiostoma ips*	NO	4	N/A
22	96	*Ophiostoma ips*	NO	4	N/A
23	114	*Ophiostoma ips*	NO	4	N/A
24	182	*Ophiostoma ips*	NO	4	N/A
25	391	*Ophiostoma ips*	NO	4	N/A
26	1576	*Ophiostoma sp.*	TB	N/A	N/A
27	1577	*Ophiostoma sp.*	TB	N/A	N/A
28	1579	*Ophiostoma sp.*	TB	N/A	N/A
29	1582	*Ophiostoma sp.*	TB	0.2	0.2
30	1583	*Ophiostoma sp.*	TB	N/A	N/A
31	1584	*Ophiostoma sp.*	TB	4	N/A
32	1585	*Ophiostoma sp.*	TB	0.2	N/A
33	1586	*Ophiostoma sp.*	TB	N/A	N/A
34	1587	*Ophiostoma sp.*	TB	N/A	N/A
35	1590	*Ophiostoma sp.*	TB	0.2	N/A
36	1592	*Ophiostoma sp.*	TB	N/A	N/A
37	1593	*Ophiostoma sp.*	TB	1.3	N/A
38	1594	*Ophiostoma sp.*	TB	1.3	0.2
39	1595	*Ophiostoma sp.*	TB	N/A	N/A
40	1596	*Ophiostoma sp.*	TB	N/A	N/A

41	1597	*Ophiostoma sp.*	TB	N/A	N/A
42	1598	*Ophiostoma sp.*	TB	1.3	0.2
43	1599	*Ophiostoma sp.*	TB	N/A	N/A
44	1600	*Ophiostoma sp.*	TB	N/A	N/A
45	1618	*Ophiostoma sp.*	TB	N/A	N/A
46	1619	*Ophiostoma sp.*	TB	0.2	N/A
47	1620	*Ophiostoma sp.*	TB	0.2	N/A
48	1623	*Ophiostoma sp.*	TB	N/A	N/A
49	1632	*Ophiostoma sp.*	TB	N/A	N/A
50	1633	*Ophiostoma sp.*	TB	N/A	N/A
51	1634	*Ophiostoma sp.*	TB	N/A	N/A
52	1635	*Ophiostoma sp.*	TB	3	N/A
53	1636	*Ophiostoma sp.*	TB	N/A	N/A
54	1638	*Ophiostoma sp.*	TB	N/A	N/A
55	1639	*Ophiostoma sp.*	TB	1.3	N/A
56	1642	*Ophiostoma sp.*	TB	0.2	N/A

Table 4. A list of *Ophiostoma* strains used in the study. The table also includes the sizes of mitochondrial *cob*-490 and *cox*3-640 introns (if present) in these strains that were determined by PCR amplification of the intron regions and agarose gel electrophoresis. The numbers represent DNA band sizes observed on gel under U.V. light visualization after 1% agarose gel electrophoresis. A band size of 0.2 kb corresponds to no intron being present in the corresponding gene insertion site; N/A corresponds to a failed amplification of the region. *cox*3 and *cob* bands/introns shaded in grey color were sequenced by Sanger sequencing.

WIN(M) = Culture collection of J. Reid (G. Hausner), Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada. Fungal strains were sampled from following geographic locations; TB = Thunder Bay (Ontario, Canada), JP = Japan, BC = British Columbia (Canada), US = United States of America, NO = Norway, NZ = New Zealand.

						Aligned intron presence and type[a]		
Number	Organism	GenBank accession number or strain	Query Cover (%)[e]	Identity (%)[f]	Intron size (kb)[g]	*cob* I4-A[b]	*cob* I4-B[c]	*cob* I4-C[d]
1	*Ophiostoma ips*	NTMB01000349.1	---------	---------	3.367	IA	IIB	IA
2	*Ophiostoma ips*	1478 (pending)	100	100.00	3.367	IA	II	IA
3	*Ophiostoma ips*	1479 (pending)	100	99.97	3.367	IA	II	IA
4	*Ophiostoma ips*	1480 (pending)	100	100.00	3.367	IA	II	IA
5	*Cladosporium sp.*	1481 (pending)	100	99.97	3.367	IA	II	IA
6	*Ophiostoma ips*	1486 (pending)	100	99.97	3.367	IA	II	IA
7	*Ophiostoma ips*	1488 (pending)	≥95	98.88	≥3.208	IA	II	IA
8	*Ophiostoma ips*	1489 (pending)	100	99.85	3.368	IA	II	IA
9	*Ophiostoma ips*	1593 (pending)	≥99	83.94	≥1.195	IA	----------	----------
10	*Ophiostoma ips*	1598 (pending)	≥98	99.02	≥1.214	IA	----------	----------
11	*Ophiostoma ips*	1639 (pending)	≥98	88.80	≥1.189	IA	----------	----------
12	*Ophiostoma novo-ulmi*	KY084294.1	99	91.23	1.189	IA	----------	----------
13	*Ophiostoma novo-ulmi*	MG020143.1	99	91.23	1.189	IA	----------	----------
14	*Ophiostoma novo-ulmi*	KY084297.1	99	91.23	1.189	IA	----------	----------
15	*Ophiostoma novo-ulmi*	KY084295.1	99	91.23	1.189	IA	----------	----------
16	*Ophiostoma novo-ulmi*	KY084296.1	99	91.15	1.189	IA	----------	----------
17	*Ophiostoma ulmi*	KY084300.1	99	90.99	1.189	IA	----------	----------
18	*Ophiostoma ulmi*	KY084298.1	99	90.99	1.189	IA	----------	----------
19	*Ophiostoma ulmi*	KY084299.1	99	90.91	1.189	IA	----------	----------
20	*Leptographium lundbergii*	KY082962.1	89	83.99	1.180	IA	----------	----------
21	*Leptographium truncatum*	KY082963.1	89	83.72	1.180	IA	----------	----------
22	*Nectria cinnabarina*	KT731105.1	100	82.46	1.222	IA	----------	----------
23	*Fusarium oxysporum*	LT906345.1	99	81.38	1.231	IA	----------	----------
24	*Fusarium oxysporum*	LT906346.1	99	81.07	1.231	IA	----------	----------
25	*Talaromyces stipitatus*	JQ354994.1	88	82.54	1.168	IA	----------	----------
26	*Fusarium oxysporum*	LT906347.1	99	80.99	1.231	IA	----------	----------
27	*Fusarium oxysporum*	EU035604.1	99	80.93	1.232	IA	----------	----------
28	*Fusarium circinatum*	JX910419.1	99	81.00	1.232	IA	----------	----------
29	*Ustilagino ideavirens*	JN204426.1	100	80.68	1.223	IA	----------	----------
30	*Epichloe hybrida*	KX066187.1	88	80.79	1.163	IA	----------	----------
31	*Epichloe festucae*	KX066186.1	88	80.79	1.163	IA	----------	----------
32	*Fusarium venenatum*	LN649234.1	99	81.73	1.235	IA	----------	----------
33	*Fusarium oxysporum*	LT571433.1	99	80.91	1.231	IA	----------	----------
34	*Epichloe festucae*	CP031392.1	88	80.60	1.163	IA	----------	----------
35	*Hirsutella rhossiliensis*	NC_030164.1	85	79.85	1.244	IA	----------	----------
36	*Hirsutella rhossiliensis*	MG979071.1	85	79.85	1.244	IA	----------	----------
37	*Hirsutella rhossiliensis*	KU203675.1	85	79.85	1.244	IA	----------	----------
38	*Fusarium graminearum*	BK010547.1	83	79.11	1.267	IA	----------	----------
39	*Fusarium graminearum*	BK010543.1	83	79.11	1.267	IA	----------	----------
40	*Fusarium graminearum*	BK010542.1	83	79.11	1.267	IA	----------	----------
41	*Raffaelea albimanens*	PCDJ01000011.1	92	77.65	2.655	IA	----------	IA
42	*Chaetomium thermophilum*	JN007486.1	81	78.96	1.449	----------	----------	IA
43	*Monilia mumecola*	JN204425.1	82	78.89	1.459	----------	----------	IA
44	*Monilia yunnanensis*	HQ908793.1	81	78.89	1.456	----------	----------	IA

Table 5. *Ophiostoma ips* (GenBank accession number: NTMB01000349.1) mitochondrial *cob* I4 annotation and comparison to similar intron sequences (and their sources) used for comparative sequence analysis. [a]Refers to an intron being present and sequences aligned and similar to the corresponding *O. ips* (GenBank accession number: NTMB01000349.1) *cob* I4 intron component. [b]*cob* I4 upstream intron component (*cob* I4-A) in *O. ips* (GenBank accession number: NTMB01000349.1). [c]*cob*I4 middle intron component (*cob* I4-B) in *O. ips* (GenBank accession number: NTMB01000349.1). [d]*cob* I4 downstream intron component (*cob* I4-C) in *O. ips* (GenBank accession number: NTMB01000349.1). [e]Refers to blastn (Altschul et al., 1990) query cover (percent of query aligned to organism sequence). [f]Refers to blastn (Altschul et al., 1990) identity (percent of aligned query that is identical to organism sequence). [g]Intron size was determined by comparative sequence analysis with intron-less *cob* genes. IA represents group I intron type A, IIB represents group II intron type B. Intron type was determined using RNAweasel (Lang et al., 2007; Beck & Lang, 2009), comparative sequence analysis, and GenBank database annotations. Query cover and identity were calculated using the blastn (Altschul et al., 1990) algorithm (with an expect threshold of $1x10^{-6}$) relative to *Ophiostoma ips'* (GenBank accession number: NTMB01000349.1) *cob* I4 (organisms in blue), *cob* I4-A only (organisms in red), *cob* I4-A and *cob* I4-C only (organism in green), and *cob* I4-C only (organisms in purple) as the queries. Values in grey are approximate due to incomplete sequencing of their corresponding organism/strain. Relative to *Saccharomyces cerevisiae* (GenBank accession number: CP006539.1), all organisms' introns were inserted at exon nucleotide position 490 in the mitochondrial *cob* gene.

Based on the PCR surveys and comparative sequence analysis (Table 5) of PCR products obtained from various strains representing various possible introns arrangements for the *cob* I4 for several *O.ips* strains it was annotated as being composed of three intron components. The components are as follows: an upstream group IA intron (*cob* I4-A) corresponding to *cob* I4 nucleotides 1-1153 and 1956-2035, a middle group IIB intron (*cob* I4-B) inserted in the P8 loop of *cob* I4-A corresponding to *cob* I4 nucleotides 1154-1955, and a downstream group IA1 intron (*cob* I4-C) corresponding to *cob* I4 nucleotides 2036-3367 (see Figure 4). The presence of the three introns, *cob* I4-A's and *cob* I4-C's core pairing sequences (P4-P7), *cob* I4-B's domain V sequence, and the intron classifications were initially determined by RNAweasel and MFannot (Lang et al., 2007; Beck & Lang, 2009).

A schematic overview of the *cob*I4 intron is shown in Figure 4. As tabulated in Table 5 the complex intron appears to consist of three distinct modules that appear to contain all the necessary components for splicing. The group I intron components contain ORFs that could encode double motif LAGLIDADG type homing endonuclease proteins. There was a short sequence separating the two group I intron sequences that could not be defined based on blastn analysis. This so called "inter-intron sequence" would be used as a "pseudoexon" by the upstream intron component for the formation of the P10 helix (Figure 5 part A) or for the downstream intron in its P1 formation (Figure 5 part C). "Pseudoexon" is a term to describes intronic sequences that be utilized during splicing by serving as "exon" sequences; ultimately "pseudoexon" sequence are removed (Hausner, unpublished). The P1 and P10 helices are essential in aligning sequences that are to be spliced out or spliced together. The upstream intron can also form a P10 interaction with the downstream exon.

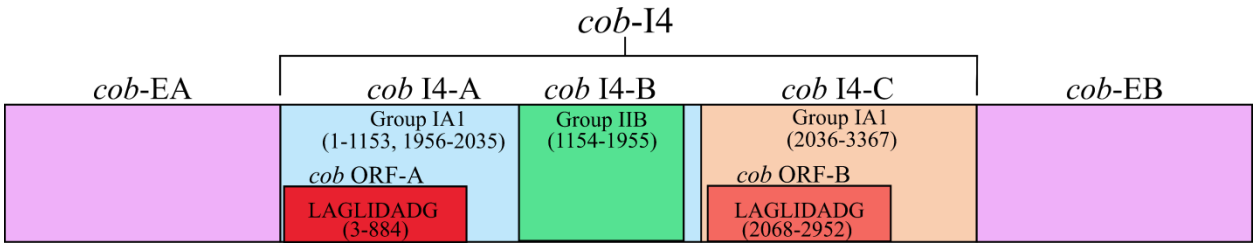


Figure 4. *cob* I4 490 intron schematic diagram.*cob* I4 = the entire complex intron at *cob* 490 position. *cob* I4-A = cob I4's upstream group IA intron, *cob*-EA = upstream exon, *cob* ORF-A = *cob* I4-A's ORF, *cob* I4-B = *cob*I4's middle group IIB intron, *cob* I4-C = *cob* I4's downstream group IA intron, *cob* ORF B = *cob* I4-C's ORF, *cob*-EB = downstream exon. LAGLIDADG represents type of homing endonuclease ORF encoded by group I intron. The numbers in brackets represent the position and length of each intron element relative to the start of *cob* I4.

Regarding splicing requirements, the *cob* I4-A's catalytically required P1-P10 pairing regions (Michel & Westhof, 1990; Golden & Cech, 1996) could only form if *cob* I4-B (B- is a group II intron) was inserted in *cob* I4-A's P8 loop. In addition, *cob* I4-B's exon binding sequences (EBS) and intron binding sequences (IBS) had to be contained in one region (*cob* I4-A's P8 loop), since matching (and correctly positioned) IBS and EBS sequences could only be found if there were no unexpected large sequence insertions between them (like an inserted intron). Furthermore, *cob* I4-C's P1-P10 catalytically required paring regions (Michel & Westhof, 1990; Golden & Cech, 1996) could only form if *cob* I4-C was present tandem to *cob* I4-A and *cob* I4-B. Regarding alignment with similar intron sequences, similar sequences only aligned with specific sequence regions in *cob* I4, providing information as to how many different introns were present, and where the intron boundaries were; in Table 5 sequences for organisms 1-40 aligned to only an upstream segment (*cob*

I4-A) of *cob* I4, the sequence for organism 41 aligned to an upstream and downstream segment (*cob* I4-A and *cob* I4-C) of *cob* I4, and sequences for organisms 42-44 aligned to only a downstream segment (*cob* I4-C) of *cob* I4. This also shows that potentially related introns exist within fungal mitochondrial genomes and in *O. ips* the complex intron was generated by introns inserting into a pre-existing intron.

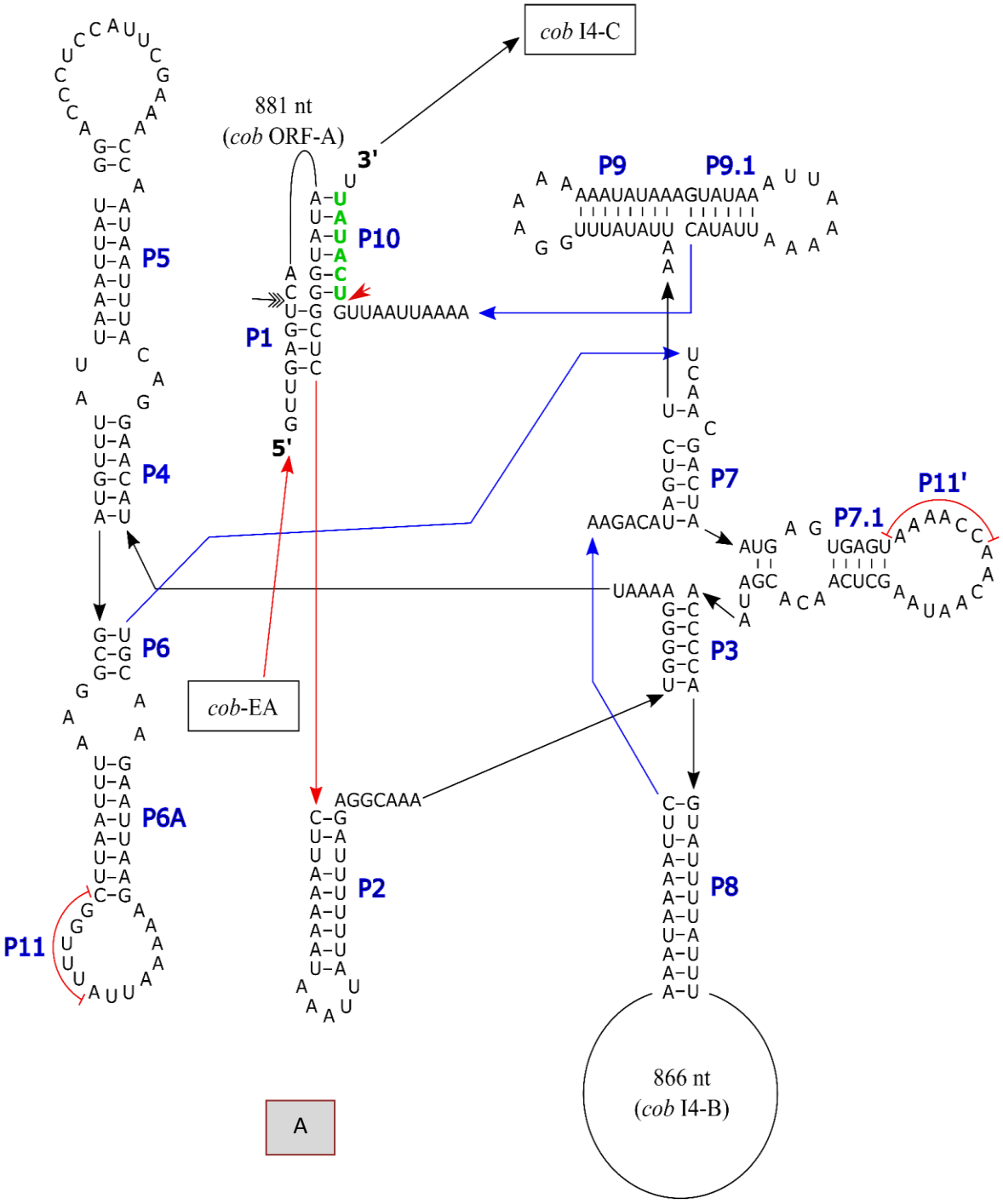
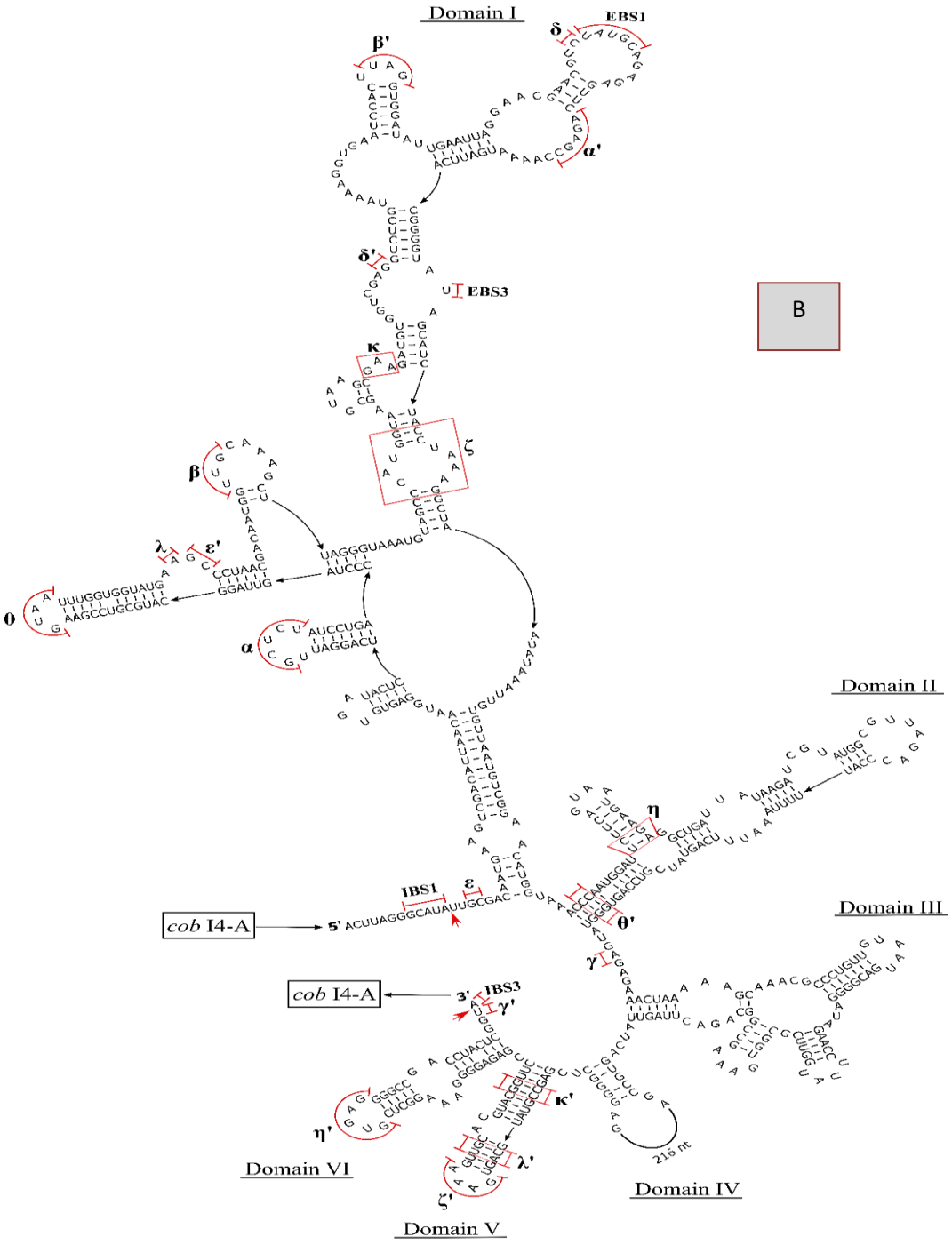
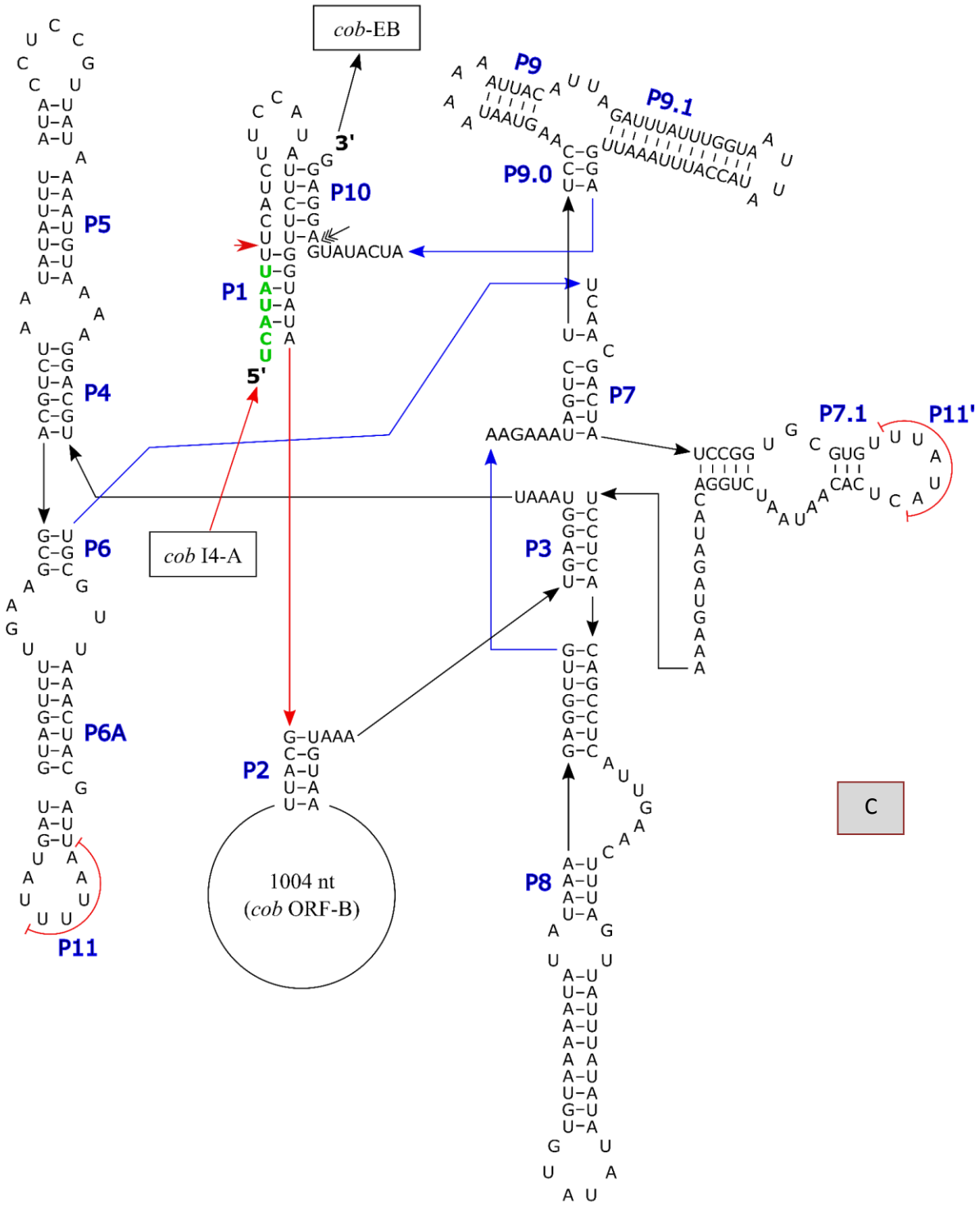

cob I4-C
881 nt
(cob ORF-A)
3'
5'
P1
P2
P3
P4
P5
P6
P6A
P7
P8
P9
P9.1
P10
P11
P11'
P7.1
cob-EA
A
866 nt
(cob I4-B)

Figure 5 (A, B, C). RNA folds for the various intron modules generated with mfold and RNAweasel.*cob* I4 RNA secondary structure model composed of three introns. Pairing regions (P1- P11) labeled by blue text and tertiary interactions shown by red lines. Hached black arrows indicate functional intron exon boundaries; small red arrows indicate functional intron- intron boundaries. Green sequence is a pseudo-exon (exon mimicking) sequence, which is actually annotated as within cobI4's downstream group IA intron (*cob* I4-C), but is shown outside cob I4-C boundaries due to splicing requirements. Red sequence is actually annotated as downstream exon sequence (*cob*-EB), but acts functionally as intron sequence for splicing and is mimicked by pseudoexon sequence.

(A) *cob* I4's upstream group IA1 intron (*cob* I4-A) RNA secondary structure model. *cob*-EA, upstream exon; *cob* ORF-A, *cob* I4-A's ORF; *cob* I4-B, *cob* I4's middle group IIB intron. **(B)** *cob*I4-B RNA secondary structure model. IBS, intron binding sequence; EBS, exon binding sequence. Helical domains I-VI branching from a central linker sequence ("six fingered hand") shown. Potential tertiary interactions (Greek letters) are indicated. **(C)** *cob*I4-C RNA secondary structure model. *cob* ORF-B, *cob* I4-C's ORF.


Domain I
EBS1
β'
δ
α'
δ'
EBS3
κ
ζ
β
λ
ε'
θ
α
Domain II
η
IBS1
ε
cob I4-A
5' ACUUAGGGCAUAUUGCGAC
θ'
Domain III
γ
cob I4-A
3'
IBS3
γ'
η'
κ'
λ'
216 nt
Domain VI
Domain IV
Domain V
B


cob-EB
3'
5'
P1
P2
P3
P4
P5
P6
P6A
P7
P8
P9
P9.1
P9.0
P10
P11
P11'
P7.1
cob I4-A
1004 nt
(cob ORF-B)
C

The whole *cob* I4 intron contained two ORFs in which *cob* I4-A was annotated to encode a double motif LAGLIDADG homing endonuclease ORF (*cob* ORF-A) in its P1 loop, corresponding to *cob* I4 nucleotides 3-884 as shown in the schematic. *cob* I4-C was annotated to encode a double motif LAGLIDADG homing endonuclease ORF (*cob* ORF-B) in its P2 loop, corresponding to *cob* I4 nucleotides 2068-2952 (see Figure 4, showing schematic). The proposed RNA secondary structure fold of *cob* I4 (Figure 5) also corroborated the three-component intron model and its introns' classifications (group IA, group IIB, and group IA).

In this intron model, the exon sequence downstream to *cob* I4 is used to form *cob* I4-C's P10 was not the exon sequence immediately following the complex intron designation; *cob* I4-C's P10 was formed with exon sequence (5' AGGAG 3') seven nucleotides downstream to the beginning of annotated exon sequence. Furthermore, a sequence within *cob* I4 appeared be exon-mimicking (i.e., a pseudo-exon) sequence (5' UCAUAUU 3'), as it was nearly identical to seven nucleotides of exon sequence (5' UCAUAUG 3') immediately downstream to *cob* I4. This pseudo-exon sequence was noted to be involved with the P10 pairing region in *cob* I4-A and the P1 pairing region in *cob* I4-C. The exon sequence used for *cob*-I4-C's P10 and the pseudo-exon sequence were modeled as such since this placed a ωG at the terminal nucleotide position of *cob* I4-C.

Group I intron P1 and P10 regions normally consist of immediate upstream and downstream exon sequences binding to a central intron internal guide sequence, which brings the flanking exon sequences close together for ligation (Michel & Westhof, 1990; Hedberg & Johansen, 2013).

4.4.2. *cox3*I2 annotation and modeling

The insertion site for *cox3* I2 was confirmed as being *cox3*-640 relative to the *S. cerevisiae* *cox3* sequence (GenBank accession number: KP263414.1) with the total length of the intron being 2964 nucleotides (~3 kb), and sequences immediately upstream and downstream to *cox3* I2 were designated as exon sequence. This was determined using MFannot (Beck & Lang, 2010), and alignment with intron-less mitochondrial *cox3* sequences (Table S2: organisms 24-56 of list of intron-less sequences).

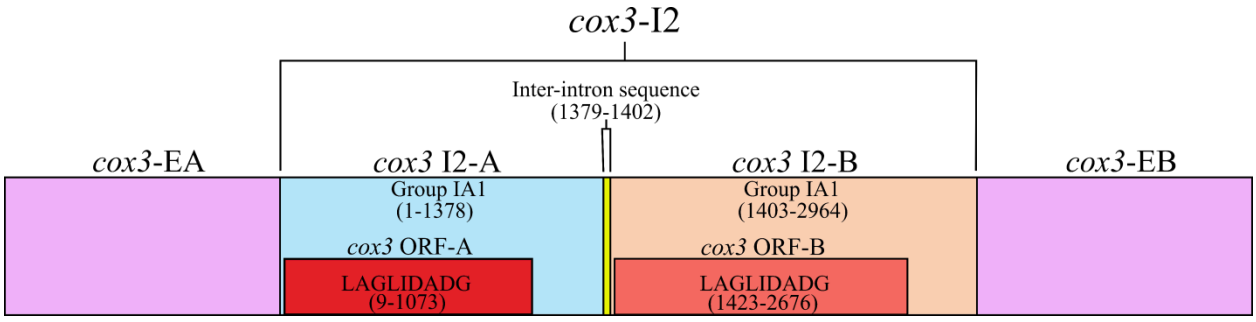
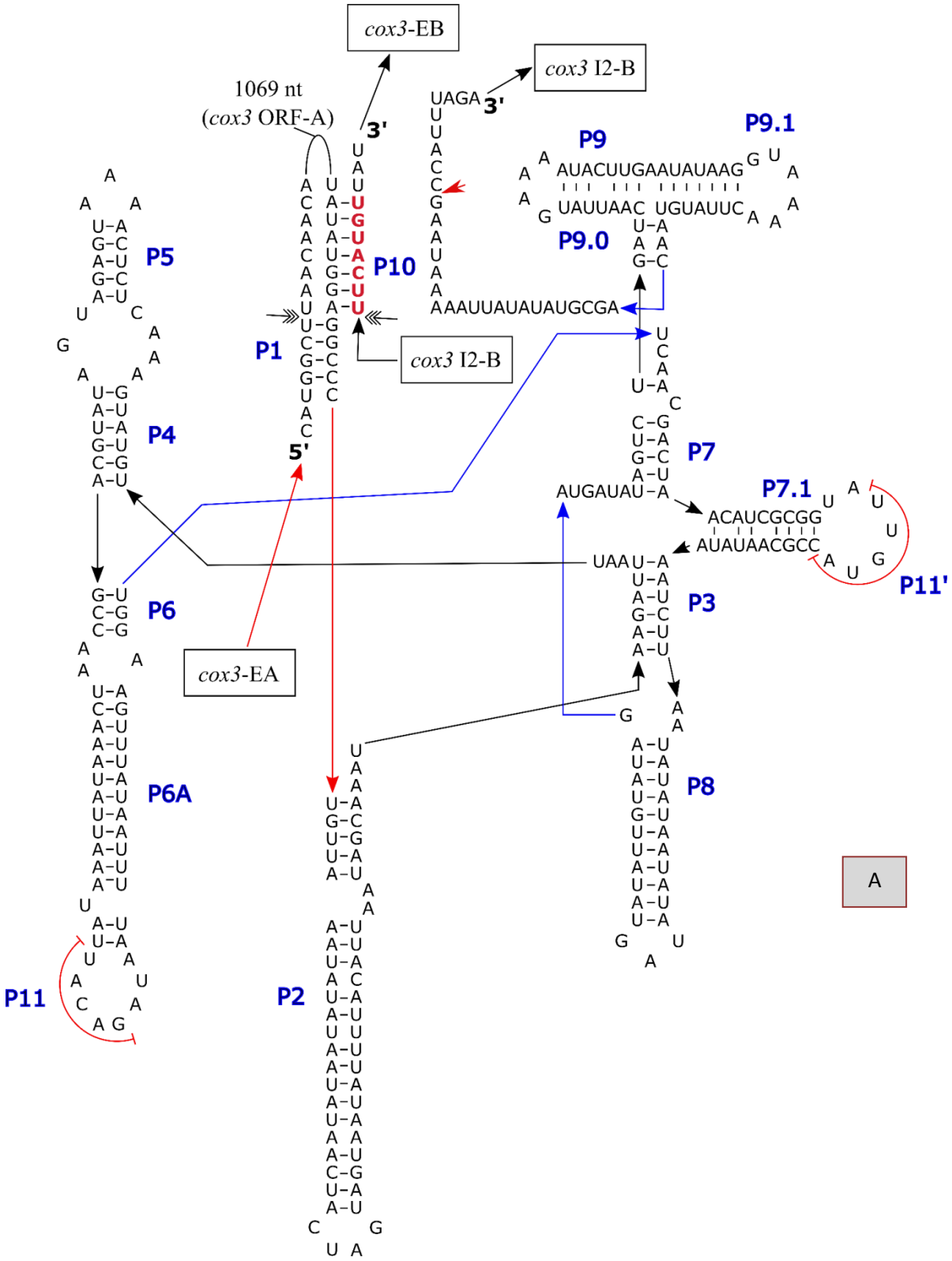


Figure 6: The *cox3* I2 640 intron schematic diagram. *cox3* I2 = the entire complex intron at *cox3* 640 position. *cox3* I2-A = *cox3* I2's upstream group IA intron, *cox3*-EA = upstream exon, *cox3* ORF-A = *cox3* I2-A's ORF, *cox3* I2- B = *cox3* I2's downstream group IA intron, *cox3* ORF-B = *cox3* I2-B's ORF, *cox3*-EB = downstream exon. Inter-intron sequence: sequence separating *cox3* I2-A and *cox3* I2-B. LAGLIDADG designation represents the type of homing endonuclease ORF encoded by group I introns. The numbers in brackets represent the position and length of each intron element relative to the start of *cox*3 I2.

						Aligned intron presence and type[a]	
Number	Organism	GenBank accession number or strain	Query Cover (%)[d]	Identity (%)[e]	Intron size (kb)[f]	*cox3* I2-A[b]	*cox3* I2-B[c]
1	*Ophiostoma ips*	NTMB01000349.1	---------	---------	2.964	IA	IA
2	*Grosmannia penicillata*	PCDK01000036.1	88	78.14	3.158	IA (5'par)	IA
3	*Endoconidiophora resinifera*	MK012641.1	92	76.21	3.260	IA	IA
4	*Endoconidiophora resinifera*	MH551223.1	92	76.21	3.260	IA	IA
5	*Endoconidiophora resinifera*	MK026449.1	92	76.21	3.259	IA	IA
6	*Endoconidiophora resinifera*	MK026450.1	92	76.21	3.259	IA	IA
7	*Nectria cinnabarina*	KT731105.1	87	68.38	3.246	IA	IA
8	*Hypomyces aurantius*	KU666552.1	95	71.97	3.056	IA (5'par)	IA
9	*Ophiostoma ips*	1487 (pending)	98	98.89	1.728	----------	IA
10	*Fusarium solani*	JN041209.1	92	73.40	1.711	----------	IA
11	*Aspergillus pseudoglaucus*	MK202802.1	91	73.60	1.742	----------	IA
12	*Esteya vermicola*	KY644696.1	91	79.14	1.717	----------	IA
13	*Fusarium culmorum*	KP827647.1	92	71.02	1.755	----------	IA (5'par)
14	*Parmotrem aultralucens*	MG807882.1	73	69.43	1.596	----------	IA
15	*Parmotrem aultralucens*	NC_040007.1	73	69.43	1.596	----------	IA
16	*Annulohypoxylon stygium*	MH620793.1	89	72.18	1.958	----------	IA
17	*Annulohypoxylon stygium*	MH620790.1	89	73.78	1.922	----------	IA

Table 6. *Ophiostoma ips* (GenBank accession number: NTMB01000349.1) mitochondrial *cox3* I2 annotation and comparison to similar introns used for comparative sequence analysis. [a]Refers to an intron being present and aligned/similar to the corresponding *O.ips* (GenBank accession number: NTMB01000349.1) *cox3* I2 intron component. [b]*cox3* I2 upstream intron component (*cox3* I2-A) in *O.ips* (GenBank accession number: NTMB01000349.1). [c]*cox3* I2 downstream intron component (*cox3* I2-B) in *O.ips* (GenBank accession number: NTMB01000349.1). [d]Refers to blastn (Altschul et al., 1990) query coverage (percent of query aligned to organism sequence). [e]Refers to blastn (Altschul et al., 1990) identity (percent of aligned query that is identical to organism sequence). [f]Intron size was determined by comparative sequence analysis with intron-less *cox3* gene sequences. IA represents group I intron type A, Intron type was determined using RNAweasel (Lang et al., 2007; Beck & Lang 2009), comparative sequence analysis, and GenBank database annotations.Query cover and identity were calculated using the blastn (Altschul et al., 1990) algorithm (with an expected threshold of $1x10^{-6}$) relative to *O.ips'* (GenBank accession number: NTMB01000349.1) *cox3* I2 (organisms in blue), and *cox3* I2-B only (organisms in red), as the queries. Relative to *S. cerevisiae* (GenBank accession number: KP263414.1), all organisms' introns were inserted at nucleotide position 640 in the mitochondrial *cox3* gene. The 5' par means only partial sequence (5') was predicted by RNAweasel.

The *cox3* I2 was annotated to exist as two intron modules in a tandem arrangement: an upstream group IA1 intron (*cox3* I2-A) corresponding to *cox3* I2 nucleotides 1-1378,and a downstream group IA1 intron (*cox3* I2-B) corresponding to *cox3* I2 nucleotides 1403-2964 (see *cox*3 intron schematic). This complex intron also appeared to contain a sequence separating *cox3* I2-A and *cox3* I2-B, referred to as the inter-intron sequence. The inter-intron sequence was annotated as corresponding to *cox3* I2 nucleotides 1379-1402, and therefore not part of either intron component. The presence of two introns, *cox3*I2-A's and *cox3*I2-B's core pairing sequences, and the intron classifications were initially determined by RNAweasel (Lang et al., 2007; Beck & Lang, 2009). Secondary structure model attributes based on models by Michel & Westhof (1990) and Deng et al., (2016), and alignment with similar intron sequences (Table S2) confirmed the intron boundaries and classifications. Furthermore, intron splicing requirements (Golden & Cech, 1996; Stahley & Strobel 2005; Stahley & Strobel, 2006) limited the number of possible *cox3* intron fold arrangements; only arrangements meeting all the criteria for self-splicing were considered. The inability to provide a plausible alternative model to this arrangement (such as a true twintron arrangement with *cox3* I2-A inserted inside *cox3* I2-B, or vice versa) also provided support for this annotation.


cox3-EB
cox3 I2-B
1069 nt
(cox3 ORF-A)
3'
UAGA
3'
P9
P9.1
AUACUUGAAUAUAAG
UAUUAAC UGUAUUC
P9.0
AAAUUAUAUAUGCGA
P10
P1
cox3 I2-B
5'
P5
P4
P7
P7.1
AUGAUAU-A
ACAUCGCGG
AUAUAACGCC
P11'
UAA U-A
P3
G-U
P6
C-G
C-G
cox3-EA
P6A
P8
P11
P2
A

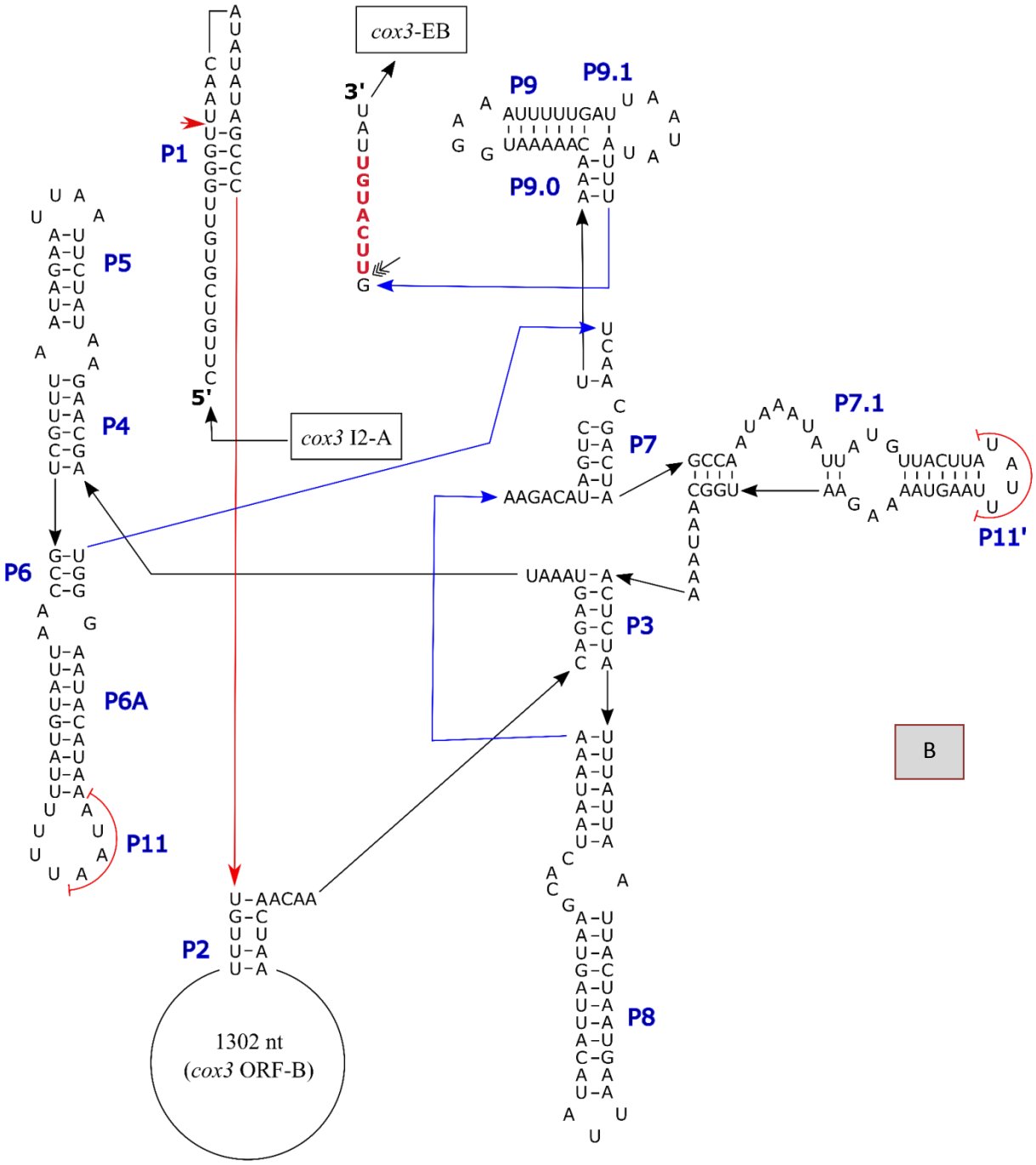

cox3-EB
3'
P1
5'
cox3 I2-A
P9
P9.1
P9.0
P5
P4
P6
P6A
P11
P7
P7.1
P11'
P3
P8
P2
1302 nt
(cox3 ORF-B)
B

Figure 7 (A, B). *cox3* I2 RNA secondary structure model composed of two introns. Pairing regions (P1-P11) labelled by blue text; tertiary P11 interactions shown by curved red lines. Hatched black arrows indicate intron-exon boundaries; small red arrows indicate intron - inter-intron sequence boundaries. Sequence in red is P10 forming sequence for *cox3* I2's upstream group IA1 intron (*cox3* I2-A) from downstream exon (*cox3*-EB). **(A)** *cox3*I2-A RNA secondary structure model. *cox3*-EA, upstream exon; *cox3* ORF-A, *cox3* I2-A's ORF; *cox3* I2-B, *cox3* I2's downstream group IA1 intron. **(B)** *cox3*I2-B RNA secondary structure model. *cox3*ORF-B, *cox3* I2-B's ORF.

With respect to splicing requirements, *cox3*I2-A's and *cox3* I2-B's catalytically required P1-P10 pairing regions (Michel & Westhof, 1990; Golden & Cech, 1996) could only form if the introns were arranged in tandem, since the nucleotides making up the pairing regions (with the exception of *cox3* I2-A's P10, which is discussed later) could only match (and be correctly positioned) if there were no unexpected large insertions (like an inserted intron) between them. In terms of alignment with similar intron sequences, some similar sequences aligned with approximately all of *cox3* I2 (organisms 1-8 in Supplementary Table S2), while others aligned to only a downstream segment (*cox3* I2-B) of *cox3* I2 (organisms 9-17 in Supplementary Table S2). This provided information regarding how many introns were present, and where the intron boundaries were. I also showed that *O. ips cox*3 I2 components have potential related sequences in other fungal mitochondrial genomes.

cox3 I2 was found to contain two ORFs. The *cox3* I2-A intron was annotated to encode a double motif LAGLIDADG homing endonuclease ORF (*cox3* ORF-A) in its P1 loop, corresponding to *cox3* I2 nucleotides 9-1073 as shown in Figure 6. The *cox3* I2-B ORF was annotated to encode a double motif LAGLIDADG homing endonuclease (*cox3* ORF-B) in its P2 loop, corresponding to *cox3* I2 nucleotides 1423-2676 as shown in *cox3* schematic.The proposed RNA secondary structure of *cox3* I2 (Figure 7) best accommodated a model that fits the tandem intron model and its introns' classifications (two group IA1 introns).

4.4.3. RT-PCR results confirm the splicing of the *cob*-490 and *cox*3-640 introns

The intron exon junctions were confirmed by *in silico* analysis as well as by the RT PCR and sequencing of the cDNA products. Both experiments confirmed the same intron insertion junction as shown below. The RT PCR showed that the "entire complex intron is removed". The splicing intermediates were not retrieved by RT PCR experiments but confirmation of absence of the whole intron in mature mRNA was obtained by RT PCR. Sequence analysis of recovered cDNA products confirmed the predicted intron/ exon junctions for both complex introns studied in *O. ips.*

The *cob*- 490 intron seems to be inserted at the following position;

[GTGCTATACCTTGAATTGGACAAGATATCGTTGAGT$^{\nabla}$TCATATGAGGAGGTTTCAGTGT TAATAATGC]

The *cox*3-640 intron seems to be inserted at the following position;

[TTTGGAACGGGGTTCCATGGCT$^{\nabla}$TTCATGTTATAATTGGAACTCTTTTTTT]

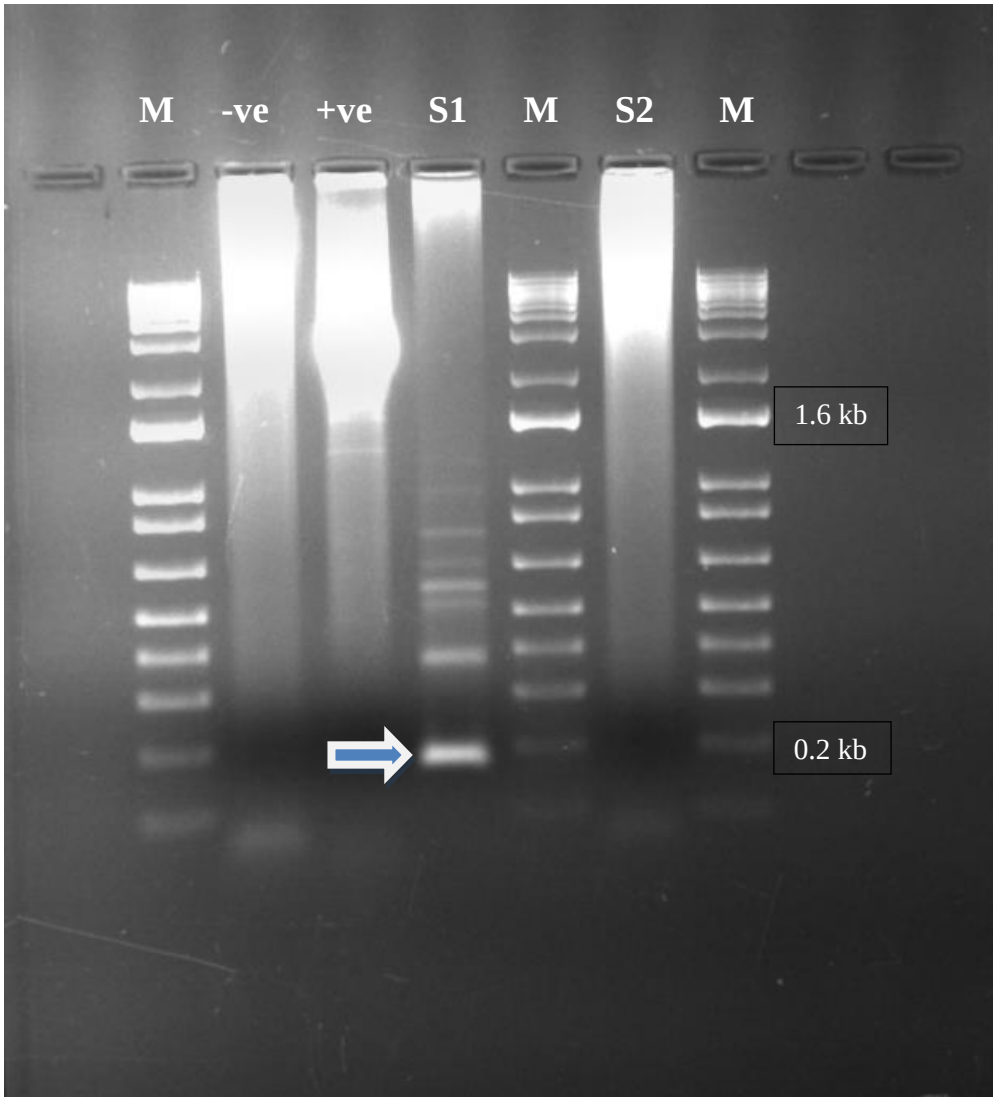


Figure 8. RT-PCR results of *cob*-490 intron showing spliced cDNA. Lane 1,5,7 = 1 kb plus DNA marker for band size prediction, Lane 2 = RNA for negative (-ve) , no RT step as a control, Lane 3 = positive (+ve) control–standard PCR (genomic DNA of WIN_M 1478 *O. ips*), Lane 4 = S1, the arrow targeted band shows RT PCR amplified cDNA by primers flanking the whole intron showing final intronless spliced version. Lane 6 = S2, RT results for primers based on amplying the sequences flanking the group II intron , but results are inconclusive.

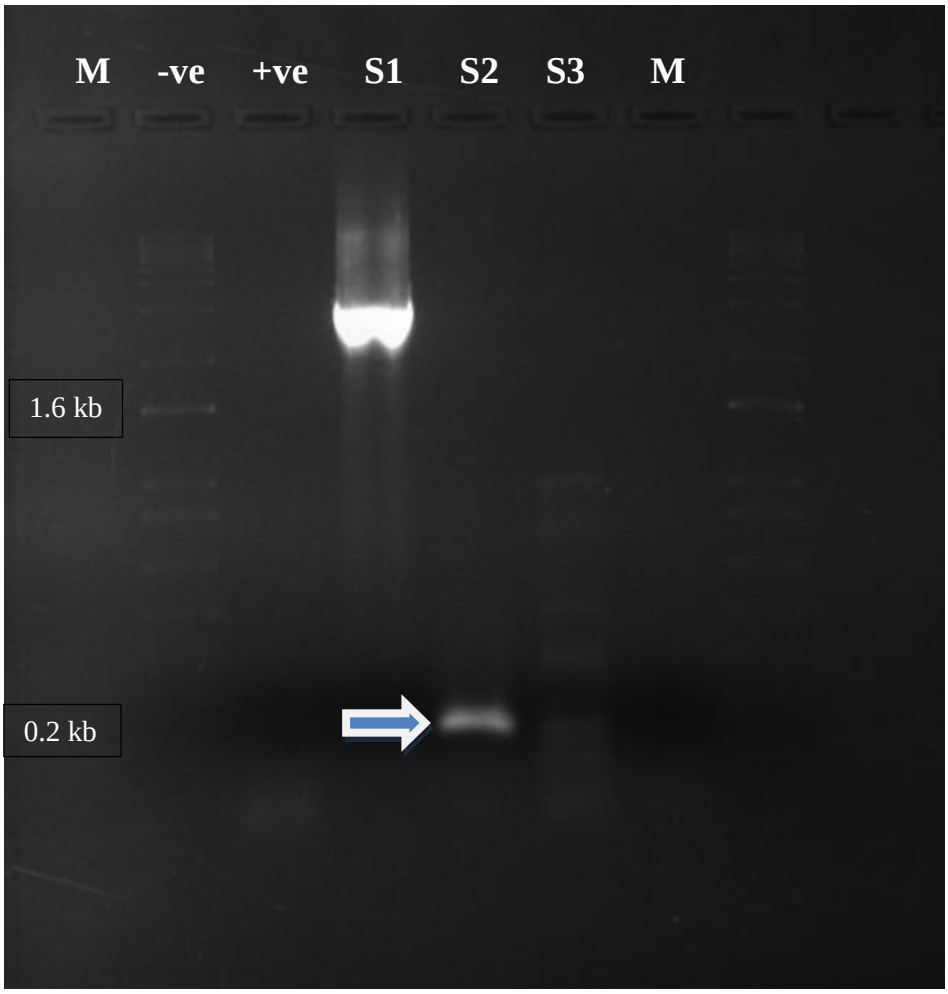
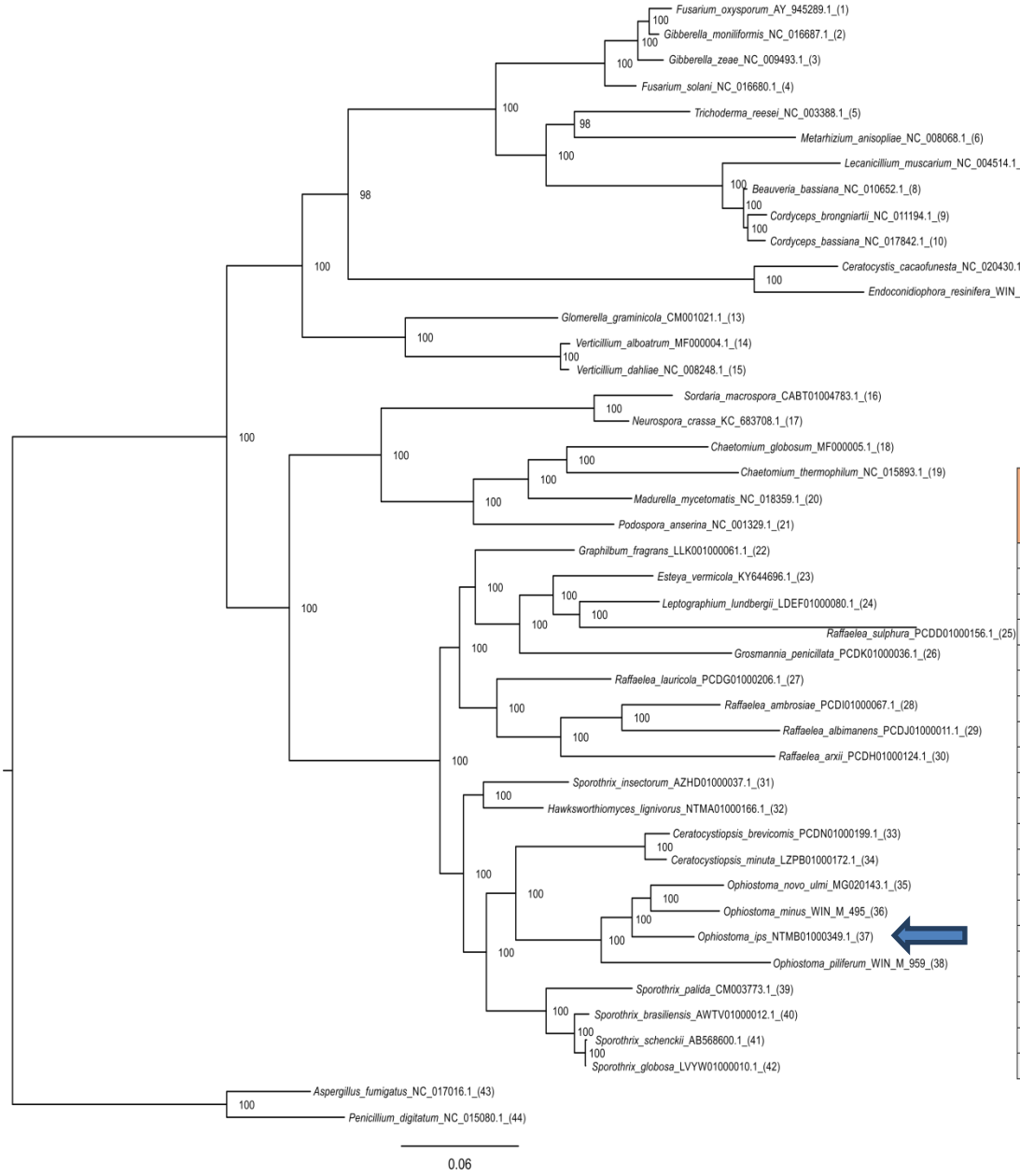


Figure 9. RT-PCR results of *cox*3-640 intron showing cDNA. Lane 1,7 = 1 kb plus DNA marker (M) for band size prediction, Lane 2 = RNA for negative (-ve) control, Lane 3 = positive (+ve) control (genomic DNA of WIN_M 1478 *O. ips*), Lane 4 = S1, the arrow indicates RT PCR amplified cDNA by primers flanking the whole intron showing final intronless spliced version. S2 = results for primers flanking the downstream intron of *cox*3-640, S3 = results for the primers flanking the upstream intron of *cox*3-640.



	cob-490 intron presence	*cox*3-640 intron presence
22	-	-
23	+ (1183 bp)	+ (1538 bp)
24	-	+ (1566 bp)
25	-	-
26	+ (2109 bp)	+ (2979 bp)
27	+ (2184 bp)	-
28	+ (1177 bp)	-
29	+ (2655 bp)	+ (1616 bp)
30	+ (1191 bp)	-
31	-	-
32	-	-
33	+ (1176 bp)	+ (3110 bp)
34	-	-
35	+ (1189 bp)	-
36	-	+ (3112bp)
37	+ (3367 bp)	+ (2964 bp)
38	+ (1269 bp)	-
39	+ (1164 bp)	-
40	-	-
41	-	-
42	-	-

Figure 10. A tree generated using concatenated amino acid sequences for 13 mitochondrial encoded proteins for the Ophiostomatales and related fungal taxa with the aim to show the presence or absence of the *cob*-490 and *cox*3-640 introns. The sizes noted for related members of the Ophiostomatales. The tree was generated by the Mr. Bayes program. Node support values based on posterior probability values were obtained from a consensus tree. The *cytb*-490 "trintron" is unique to *O. ips* but the *cox3* twintron is present in other *Ophiostoma* species.

4.5. Discussion

The complexities of introns that are made up of two or three separate individual intron modules are of great interest for variety of reasons. Complex introns provide insight into novel features of self-splicing introns, and the consequences that those features may have with regards to "orderly splicing". These are ultimately complex ribozymes that evolved by ribozymes aggregating in one larger unit forming complex ribozymes. The complex introns studied herein, aside from being composed of multiple intron components, possess unique features, such as the pseudo-exon sequence in *cob* I4, and the inter-intron sequence (probably also a pseudo-exon type sequence) in *cox3* I2, and the missing P10 pairing region in *cox3* I2-B. The latter suggests that for this intron combination the downstream component may not be able to splice independently of the downstream intron module.These distinct attributes are reflected in the intron secondary structure modeling; and these may features also have implications regarding complex intron splicing, mobility and formation.

In terms of *cob* I4 splicing, being composed of three intron components and the presence of a pseudo-exon sequence complicates and increases the number of ways in which the introns may splice out. Due to the presence of the pseudo-exon sequence, both *cob* I4-A (regardless of when or if *cob* I4-B splices out of *cob* I4-A) and *cob* I4-C may splice separately (ie. independently of the downstream intron module).

If *cob* I4-C were to splice out before *cob* I4-A, then *cob* I4-A would be forming its P10 with the pseudo-exon sequence ligated to the downstream exon. In comparison, if *cob* I4-A were to splice out before *cob* I4-C, then *cob* I4-A would still be using the pseudo-exon sequence to form its P10; however, the pseudo-exon sequence would still be linked to *cob* I4-C. Thus, since *cob* I4-A can and *cob* I4-C must use the pseudo-exon sequence for splicing, *cob* I4-A and *cob* I4-C may splice

independently of each other, and the order may be affected by competing affinities of *cob* I4-A and *cob* I4-C to the pseudo-exon sequence.

The sequential splicing of *cob* I4-A and *cob* I4-C (in either order) is possible. Guha & Hausner (2016) demonstrated that for some group II introns the IBS2-EBS2 interaction is not essential (or even required) for splicing. Therefore, it is still possible for *cob* I4-B to splice out of *cob* I4 independently of *cob* I4-A, since *cob* I4-B still possesses IBS1-EBS1 and IBS3-EBS3 tertiary interactions to bring the flanking sequences (*cob* I4-A sequence) into close proximity.

The splicing pattern may depend on host gene regulatory factors, evolutionary considerations, and intron mobility factors. Splicing out as one large composite unit likely simplifies and speeds up the removal process, allowing for processing and thus faster translation of the *cob* mRNA, whereas splicing of the various intron modules independently from each other may have the opposite effect, allowing for greater regulation or fine tuning of gene expression. These intron modules could be rate limiting steps that allow for modulating the expression of the host gene. In terms of evolution, splicing as one large unit could be deemed energetically favorable, as the splicing mechanism with the fewest steps requires the least amount of energy, which could provide an evolutionary or selective advantage to the host fungus. However, splicing independently may also aid with intron mobility and persistence; independent splicing allows for either *cob* I4-A or *cob* I4-C (depending on which splices out first) to remain attached to the exon sequences for a longer period of time (avoiding degradation), which could increase their ORF expression levels, contributing to increased intron invasion of new sites and ultimately in the intron/IEP persistence. These combinations of factors make it unlikely for splicing topredominantly to occur in only one order; it is possible that each of the proposed splicing orders occur at different frequencies during the production of the many *cob* mRNAs.

Splicing of *cox3* I2 does not appear to be as complex as the splicing of *cob* I4; based on the proposed model, *cox3* I2-A and *cox3* I2-B likely splice out together, rather than independently as separate components (Figure 6). This is a direct consequence of the absence of the P10 pairing region in *cox3* I2-B and the presence of an inter-intron sequence. As discussed earlier, the P10 region is required for intron self-splicing and ligation of the sequences (usually exon) flanking the intron (Michel & Westhof, 1990; Golden & Cech, 1996); for *cox3* I2, the only way a P10 region composed of downstream exon can form is if it is also composed of *cox3* I2-A's internal guide sequence. This would result in both the upstream and downstream exons taking part in *cox3* I2-A's P1 and P10 regions respectively, causing *cox3* I2 to splice out as a single unit. In addition to the lack of *cox3* I2-B's P10, the presence of an inter-intron sequence also suggests that the intron likely splices out as a single unit. If both *cox3* I2-A and *cox3* I2-B spliced out independently, neither splicing event would remove the inter-intron sequence, since it is not within the splicing boundaries of either intron component.

Although *cox3* I2 likely splices out as a single unit, it is possible that this was not always the case. *cox3* I2 was noted to be able to form an additional P10 with inter-intron sequence rather than downstream exon. It should be noted that the upstream intron in the tandem intron model proposed by Deng et al., (2016) could also form a P10 with the downstream exon (allowing the tandem intron to splice as a unit). This suggests that the previously published tandem intron by Deng et al. shares many features with the *cox3* I2 of *O. ips*. These types of arrangements may allow for alternative splicing whereby the removal of the upstream intron generates an intermediate whereby the downstream intron is fused to the upstream exon with the downstream ORF in frame with the upstream exon. This is viewed to be advantageous allowing for more efficient translation of the intron encoded ORF (Edgell et al., 2011).

Intron formation and mobility are related processes for complex introns like *cob* I4 and *cox*3 I2; complex intron formation depends on intron mobility. Typically when discussing complex introns, one of the intron components is referred to as the presumptive native intron, while the other or others are termed, "invaders". However, determining which intron is the native intron is not always possible.

However, there is evidence to suggest that *cob* I4-B is not the native intron. As stated in the results, *cob* I4-B was annotated as being inserted in the P8 loop of *cob* I4-A, therefore, *cob* I4-B could not have been present in *O. ips* (GenBank accession number: NTMB01000349.1) *cob*-490 position before *cob* I4-A. In addition, similar *cob*-490 introns only possessed a group II intron (corresponding to *cob* I4-B) if an intron similar to *cob* I4-A was also present, suggesting that the presence of *cob* I4-B is associated with the presence of *cob* I4-A.

Current sequence information suggests that *cox3* I2's native intron was *cox3* I2-B. When only one intron module is present in other fungi at this site it is similar to I2-B. Therefore, it is likely that *cox3* I2-B is the "resident" intron at the *cox3*-640 position, prior to adjacent invasion by *cox3* I2-A to form the proposed tandem intron.

Regarding the inter-intron sequence in the cox2-I2 complex intron, its origin is unknown but as stated above it is potentially critical in allowing for alternative splicing. One could speculate that it was at one point a component of *cox3* I2-A or *cox3* I2-B, but selection towards whole *cox3* I2 splicing altered the location of *cox3* I2-A's P10 or *cox3* I2-B's P1 pairing regions.

As mentioned previously, the IEPs encoded within fungal mitochondrial introns can facilitate host intron homing (site specific mobility) or transposition to new sites (Hafez et al., 2013). Intron encoded homing endonucleases, like the double motif LAGLIDADG homing endonucleases

encoded in *cox3* I2 and *cob* I4, not only require translation by the fungal host, but also require a specific DNA nucleotide sequence 14-40 base pairs long to catalyze double-stranded breaks and subsequent intron homing or transposition (Chevalier & Stoddard, 2001).

The high similarity between *cob* I4's ORFs, the tandem arrangement of *cob* I4-A and *cob* I4-C, and the presence of the pseudo-exon sequence suggests that the two homing endonucleases encoded by *cob* I4-A and *cob* I4-C could possible recognize similar sequences for catalyzing doubled-stranded breaks and intron homing. Thus, the presence of a pseudo-exon sequence may increase the occurrence of complex introns composed of introns encoding highly similar ORFs. In contrast in the *cox3* I2 the ORFs in the group I intron *cox3* I2 modules were dissimilar; they probably recognize different sequences for catalysis. But here the inter intron sequence again may serve the function of alternative splicing but it may not benefit the intron ORFs (compared to the cox3 I2 intron ORFs) but it may provide a regulatory switch that controls the expression of the host gene. The nature of this inter-intron sequence that essentially can act like a pseudo exon sequence (as it could participate in the formation of P1 and P10 helices) needs further investigation.

Future work involving these introns will include RNA sequencing experiments (ie. RNAseq) to confirm or refute proposed intron boundaries and splicing mechanisms. Deep RNA sequences may uncover alternative splicing events and rare transcripts that may have been in missed in our RT PCR work. RNA would have to be harvested from cultures at different stages (ages) as some transcripts may not be present in older cultures.

Regardless of which ORF is expressed in complex introns, IEP ORFs may accumulate mutations quickly, presumably due to their host introns being evolutionarily neutral in regard to host gene function (Goddard & Burt, 1999); thus it has been observed that introns and their ORFs quickly

erode and get lost. However, to avoid extinction, these introns and their ORFs must continuously "home" to intron-less alleles, spread horizontally into new genomes, and/or transpose into sites (Goddard & Burt, 1999; Guha et al., 2018). The ability to transpose into novel sites (potentially forming complex introns) may be facilitated by the same mutations in introns/ORFs that can contribute to intron extinction, since some mutations in ORF sequences could lead to the creation of novel homing endonucleases with new/altered recognition sequencespecificity. For example, McMurrough et al., (2018) showed that it takes very few changes in the DNA binding component of LAGLIDADG type endonuclease in order for different sequences to be invaded. This is partially due to the interactions between these types of endonucleases with their DNA target sites. It relies in part on "indirect" contact or "indirect readout" – that means the protein recognizes the DNA topology at the cleavage site not necessarily the exact DNA nucleotide sequence.

This study explores the first plausible three-component self-splicing intron and a unique tandem intron model. The unique information provided by such rare arrangements can give insight into the structure and migration of mobile introns and their IEPs present within fungal mitochondrial genomes. The immediate relevancy of these complex, ribozyme type introns and their encoded proteins is apparent by their potential uses in genome editing (Hafez & Hausner, 2012; Stoddard, 2014; Abboud et al., 2018) or introns being used as regulatory RNAs in controlling the expression of heterologous proteins (Hafez & Hausner, 2015; and see Guha et al., 2016). Furthermore, mobile, self-splicing mitochondrial introns have been associated with mitochondrial diseases in fungi, mitochondrial DNA instability in fungi, and hypo-virulence in fungal plant pathogens, such as the Chestnut blight fungus *Cryphonectria parasitica* (Baidyaroy et al., 2011; Abboud et al., 2018). The specific findings from this study can be utilized for structural and comparative sequence analysis

with other complex intron arrangements in fungal mitochondrial genomes. In addition, this study provides a basis from which other group IA and group IIB introns may be annotated and modeled.

CHAPTER 5: GENERAL CONCLUSIONS

The first project demonstrated that mtDNAs can yield a phylogenetic tree for fungi with good node support values. The tree represents most of the sequences available in NCBI and MitoFun. A total of 205 sequences were taken into account. The future of the first project can be a large review of features of fungal mitochondrial DNA. This study also suggests that with the advent of more rapid and affordable NGS approaches such as nanopore (MinION) based techniques, mtDNA sequences could be more rapidly collected for many fungi. For pathogens and invasive fungi this could be very useful in terms of monitoring their spread over geographic regions. Members of the Ophiostomatales which are vectored by insect can spread quite rapidly via their vectors or by importing/exporting lumber. Therefore, the monitoring of exotic fungi requires accurate molecular markers. These fungi are a concern with regards to the biosecurity of Canadian forests and new arrivals pose serious threats and need to be identified accurately and rapidly.

The second project demonstrated the presence of a novel three-component intron in *cob*-490 position. The study also confirmed the presence of a *cox*3-640 twintron in the mtDNA of *O. ips*.The complex introns observed in *O. ips* probably are a product of intron mobility, and these complex intron configurations might offer regulatory platforms for alternative splicing or modulating host gene expression. Multistep splicing pathways offer more opportunity for regulation; it is known that splicing of group I and II introns is assisted by so called maturases that can be encoded by introns or by the host genomes. Ribozymes (such as group I and group II introns) and their encoded proteins have applications in biotechnology; plus these mobile introns might “shape” fungal mtDNAs and thus are important in understanding the evolution of fungal mtDNAs. Information obtained from complex introns can provide insight into the evolution of fungal mtDNAs, introns and ribozymes. Results from this study can serve as a template for modelling other complex group I and group II introns that have yet to be discovered. Future work includes reverse transcriptase PCR and RNA

sequencing (RNAseq) in order to confirm the proposed intron boundaries, the splicing mechanisms, and splicing pathways. Complex introns are mechanisms whereby new types of mobile elements could evolve. In addition complex introns could be developed into regulatory sequences applied in biotechnology were modulating gene expression is important. Recently nested group II introns were applied for controlling the expression of heterologous proteins in *E. coli*. The study of Guha et al. (2017) demonstrated that in *E. coli* a mtDNA fungal homing endonuclease (HE) could be turn on or off by controlling the splicing of the group II intron that was inserted into the HE ORF. This was a proof of principle that ribozyme type introns would be useful molecular switches for controlling the expression of genes.

CHAPTER 6. APPENDICES

Table S1. A list of the strains used to generate the phylogenetic tree as shown in Figure 1

Phylum	No.	Species name	Accession number	mtDNA size (bp)	%GC content	Class	Order
Asco-mycota	1	*Opegrapha vulgata*	NC_035825.1	38937	30.8	Arthoniomycetes	Arthoniales
	2	*Cladophialophora bantiana*	NC_030600.1	26821	24.5	Eurotiomycetes	Chaetothyriales
	3	*Cladophialophora carrionii*	CM004524.1	26795	25.5	Eurotiomycetes	Chaetothyriales
	4	*Chrysoporthe austroafricana*	NC_030522.1	190834	30.3	Sordariomycetes	Diaporthales
	5	*Chrysoporthe cubensis*	NC_030524.1	89084	29.7	Sordariomycetes	Diaporthales
	6	*Chrysoporthe deuterocubensis*	NC_030523.1	124412	30	Sordariomycetes	Diaporthales
	7	*Cryphonectria parasitica*	KT428651.1	158413	44.1	Sordariomycetes	Diaporthales
	8	*Diaporthe longicolla*	NC_027509.1	53439	34.4	Sordariomycetes	Diaporthales
	9	*Aspergillus cristatus*	CM004508.1	77649	28.22	Eurotiomycetes	Eurotiales
	10	*Aspergillus kawachii*	AP012272.1	31222	26.4	Eurotiomycetes	Eurotiales
	11	*Aspergillus ustus*	NC_025570.1	33007	25.2	Eurotiomycetes	Eurotiales
	12	*Penicillium nordicum*	KR952336.1	28520	25.4	Eurotiomycetes	Eurotiales
	13	*Penicillium polonicum*	NC_030172.1	28192	25.6	Eurotiomycetes	Eurotiales
	14	*Penicillium roqueforti*	KR952335.1	29908	25.4	Eurotiomycetes	Eurotiales
	15	*Talaromyces marneffei*	NC_005256.1	35438	24.6	Eurotiomycetes	Eurotiales
	16	*Talaromyces pinophilus*	CP017352.1	31729	N/A	Eurotiomycetes	Eurotiales
	17	*Xeromyces bisporus*	HG983520.1	69886	29.4	Eurotiomycetes	Eurotiales

	18	*Colletotrichum acutatum*	NC_027280.1	30892	30.5	Sordariomycetes	Glomerellales
	19	*Colletotrichum fioriniae*	NC_030052.1	30020	30	Sordariomycetes	Glomerellales
	20	*Colletotrichum graminicola*	NW_007361658.1	39649	49.1	Sordariomycetes	Glomerellales
	21	*Colletotrichum lindemuthianum*	NC_023540.1	36957	30.9	Sordariomycetes	Glomerellales
	22	*Colletotrichum lupini*	NC_029213.1	36554	29.9	Sordariomycetes	Glomerellales
	23	*Colletotrichum salicis*	NC_035496.1	33950	30.4	Sordariomycetes	Glomerellales
	24	*Colletotrichum tamarilloi*	NC_029706.1	30824	30.5	Sordariomycetes	Glomerellales
	25	*Verticillium dahliae*	NC_008248.1	27184	27.3	Sordariomycetes	Glomerellales
	26	*Verticillium nonalfalfae*	NC_029238.1	26139	26.9	Sordariomycetes	Glomerellales
	27	*Cairneyella variabilis*	NC_029759.1	27186	26.3	Leotiomycetes	Helotiales
	28	*Glarea lozoyensis*	KF169905.1	45038	29.7	Leotiomycetes	Helotiales
	29	*Marssonina brunnea*	NC_015991.1	70379	29.3	Leotiomycetes	Helotiales
	30	*Rhynchosporium agropyri*	NC_023125.1	68904	29.4	Leotiomycetes	Helotiales
	31	*Rhynchosporium orthosporum*	NC_023127.1	49539	28.8	Leotiomycetes	Helotiales
	32	*Rhynchosporium secalis*	NC_023128.1	68729	29.3	Leotiomycetes	Helotiales
	33	*Acremonium chrysogenum*	NC_023268.1	27266	26.5	Sordariomycetes	Hypocreales
	34	*Acremonium fuci*	NC_029851.1	24565	28.8	Sordariomycetes	Hypocreales
	35	*Beauveria caledonica*	NC_030636.1	38316	26.3	Sordariomycetes	Hypocreales
	36	*Beauveria malawiensis*	NC_030635.1	44135	26.7	Sordariomycetes	Hypocreales

	37	*Beauveria pseudobassiana*	NC_022708.1	28006	27.5	Sordariomycetes	Hypocreales
	38	*Clonostachys rosea*	NC_036667.1	40921	27.9	Sordariomycetes	Hypocreales
	39	*Cordyceps militaris*	NC_022834.1	33277	26.8	Sordariomycetes	Hypocreales
	40	*Epichloe festucae*	NC_032064.1	88744	27.5	Sordariomycetes	Hypocreales
	41	*Epichloe typhina*	NC_032063.1	84630	27	Sordariomycetes	Hypocreales
	42	*Fusarium avenaceum*	JQGD01000004.1	49402	33.1	Sordariomycetes	Hypocreales
	43	*Fusarium circinatum*	NC_022681.1	67109	31.4	Sordariomycetes	Hypocreales
	44	*Fusarium commune*	NC_036106.1	47526	32.4	Sordariomycetes	Hypocreales
	45	*Fusarium culmorum*	NC_026993.1	103844	31.7	Sordariomycetes	Hypocreales
	46	*Fusarium graminearum*	HG970331.1	95638	31.8	Sordariomycetes	Hypocreales
	47	*Fusarium mangiferae*	NC_029194.1	30629	31.3	Sordariomycetes	Hypocreales
	48	*Hirsutella rhossiliensis*	NC_030164.1	62483	28.2	Sordariomycetes	Hypocreales
	49	*Ilyonectria destructans*	NC_030340.1	42895	28.2	Sordariomycetes	Hypocreales
	50	*Parengyodontium album*	NC_032302.1	28081	25.9	Sordariomycetes	Hypocreales
	51	*Tolypocladium inflatum*	NC_036382.1	25328	27.8	Sordariomycetes	Hypocreales
	52	*Tolypocladium ophioglossoides*	NC_031384.1	35159	27.5	Sordariomycetes	Hypocreales
	53	*Trichoderma asperellum*	KR952346.1	29999	27.8	Sordariomycetes	Hypocreales
	54	*Trichoderma gamsii*	NC_030218.1	29303	28.3	Sordariomycetes	Hypocreales
	55	*Trichoderma hamatum*	NC_036144.1	32763	27.7	Sordariomycetes	Hypocreales
	56	*Bryoria tenuis*	NC_034786.1	84295	30.2	Lecanoromycetes	Lecanorales
	57	*Cladonia rangiferina*	NC_036309.1	59116	29.6	Lecanoromycetes	Lecanorales
	58	*Hypogymnia vittata*	NC_035730.1	38888	30.7	Lecanoromycetes	Lecanorales
	59	*Imshaugia aleurites*	NC_035550.1	32029	30.5	Lecanoromycetes	Lecanorales

	60	*Lecanora strobilina*	NC_030051.1	39842	29.2	Lecanoromycetes	Lecanorales
	61	*Phyllopsora corallina*	NC_034779.1	39591	28.4	Lecanoromycetes	Lecanorales
	62	*Usnea ceratina*	NC_035940.1	65539	38.9	Lecanoromycetes	Lecanorales
	63	*Ceratocystis cacaofunesta*	NC_020430.1	103147	26.8	Sordariomycetes	Microascales
	64	*Lomentospora prolificans*	CM008229.1	23987	26.9	Sordariomycetes	Microascales
	65	*Nannizzia nana (Arthroderma obtusum)*	NC_012830.1	24105	24.5	Eurotiomycetes	Onygenales
	66	*Esteya vermicola*	NC_035254.1	46507	24.9	Sordariomycetes	Ophiostomatales
	67	*Ophiostoma novo-ulmi*	CM001753.1	66357	24.49	Sordariomycetes	Ophiostomatales
	68	*Sporothrix pallida*	CM003773.1	35458	N/A	Sordariomycetes	Ophiostomatales
	69	*Gomphillus americanus*	NC_034790.1	28370	28.4	Lecanoromycetes	Ostropales
	70	*Graphis lineola*	NC_035824.1	24945	29.5	Lecanoromycetes	Ostropales
	71	*Coccocarpia palmicola*	NC_034332.1	73992	25.3	Lecanoromycetes	Peltigerales
	72	*Leptogium hirsutum*	NC_034928.1	120920	24.8	Lecanoromycetes	Peltigerales
	73	*Peltigera dolichorrhiza*	NC_031804.1	51156	26.8	Lecanoromycetes	Peltigerales
	74	*Ricasolia amplissima*	NC_035826.1	82333	29.8	Lecanoromycetes	Peltigerales
	75	*Pyronema omphalodes*	NC_029745.1	191189	43	Pezizomycetes	Pezizales
	76	*Didymella pinodes*	NC_029396.1	55973	29.5	Dothideomycetes	Pleosporales
	77	*Phaeosphaeria nodorum*	NC_009746.1	49761	29.4	Dothideomycetes	Pleosporales
	78	*Pithomyces chartarum*	NC_035636.1	68926	28.6	Dothideomycetes	Pleosporales
	79	*Shiraia bambusicola*	NC_026869.1	39030	25.2	Dothideomycetes	Pleosporales
	80	*Pneumocystis*	NC_020331.1	35626	25.5	Pneumocystidomycetes	Pneumocystis

		jirovecii					
	81	*Pneumocystis murina*	NC_020332.1	24608	29.8	Pneumocystidomycetes	Pneumocystis
	82	*Pseudogymnoascus pannorum*	NC_027422.1	26918	28.1	Leotiomycetes	Pseudeurotiaceae
	83	*Annulohypoxylon stygium*	NC_023117.1	133782	29.9	Sordariomycetes	Xylariales
	84	*Arthrinium arundinis*	NC_035508.1	48975	29.4	Sordariomycetes	Xylariales
	85	*Pestalotiopsis fici*	NC_031828.1	69529	28.4	Sordariomycetes	Xylariales
	86	*Zasmidium cellare*	NC_030334.1	23743	27.84	Dothideomycetes	Capnodiales
	87	*Zymoseptoria tritici*	NC_010222.1	43964	31.93	Dothideomycetes	Pleosporales
	88	*Exophiala dermatitidis*	N/A	26004	25.88	Eurotiomycetes	Chaetothyriales
	89	*Aspergillus clavatus*	JQ_354999	35056	24.96	Eurotiomycetes	Eurotiales
	90	*Aspergillus flavus*	JQ_355000	29205	26.16	Eurotiomycetes	Eurotiales
	91	*Aspergillus fumigatus*	NC_017016	30696	25.48	Eurotiomycetes	Eurotiales
	92	*Aspergillus nidulans*	NC_017896	33227	24.94	Eurotiomycetes	Eurotiales
	93	*Aspergillus niger*	NC_007445	31103	26.90	Eurotiomycetes	Eurotiales
	94	*Aspergillus oryzae*	JQ_354998	29202	26.15	Eurotiomycetes	Eurotiales
	95	*Aspergillus terreus*	JQ_355001	24658	27.09	Eurotiomycetes	Eurotiales
	96	*Aspergillus tubingensis*	NC_007597	33656	26.78	Eurotiomycetes	Eurotiales
	97	*Neosartorya fischeri*	JQ_354995	34373	25.43	Eurotiomycetes	Eurotiales
	98	*Penicillium chrysogenum*	AM_920464	31790	24.93	Eurotiomycetes	Eurotiales
	99	*Penicillium digitatum*	NC_015080	28978	25.34	Eurotiomycetes	Eurotiales
	100	*Penicillium marneffei*	NC_005256	35438	24.63	Eurotiomycetes	Eurotiales
	101	*Penicillium solitum*	JN_696111	28601	25.47	Eurotiomycetes	Eurotiales
	102	*Ajellomyces capsulatus*	GG_663449	40201	23.38	Eurotiomycetes	Onygenales
	103	*Arthroderma*	NC_012830	24105	24.47	Eurotiomycetes	Onygenales

		obtusum					
	104	*Arthroderma otae*	NC_012832	23943	24.25	Eurotiomycetes	Onygenales
.	105	*Arthroderma uncinatum*	NC_012828	28530	23.45	Eurotiomycetes	Onygenales
	106	*Epidermophyton floccosum*	NC_007394	30910	23.43	Eurotiomycetes	Onygenales
	107	*Microsporum gypseum*	DS_989840	23539	24.72	Eurotiomycetes	Onygenales
	108	*Paracoccidioides brasiliensis*	NC_007935	71335	21.00	Eurotiomycetes	Onygenales
	109	*Trichophyton mentagrophytes*	NC_012826	24297	24.03	Eurotiomycetes	Onygenales
	110	*Trichophyton rubrum*	NC_012824	26985	23.51	Eurotiomycetes	Onygenales
	111	*Peltigera malacea*	NC_016955	63363	27.24	Lecanoromycetes	Peltigerales
	112	*Peltigera membranacea*	NC_016957	62785	27.10	Lecanoromycetes	Peltigerales
	113	*Botryotinia fuckeliana*	KC_832409	82212	29.89	Leotiomycetes	Helotiales
	114	*Phialocephala subalpina*	NC_015789	43742	27.95	Leotiomycetes	Helotiales
	115	*Rhynchosporium commune*	NC_023126	69581	29.39	Leotiomycetes	Helotiales
	116	*Sclerotinia borealis*	NC_025200	203051	32.52	Leotiomycetes	Helotiales
	117	*Sclerotinia sclerotiorum*	KT_283062	128852	30.90	Leotiomycetes	Helotiales
	118	*Pseudogymnoascus destructans*	NC_033907	32181	28.52	Leotiomycetes	Incertae Sedis
	119	*Pneumocystis carinii*	NC_013660	26119	29.81	Pneumocystidomycetes	Pneumocystidales
	120	*Candida sake*	KC_993194	26205	25.73	Saccharomycetes	Saccharomycetales
	121	*Wickerhammomyces canadiensis*	NC_001762	27694	18.14	Saccharomycetes	Saccharomycetales
	122	*Chrysoporthe cubensis*	NC_030524	89084	29.71	Sordariomycetes	Diaporthales

	123	*Colletotrichum gloeosporioides*	KX_885104	55169	34.55	Sordariomycetes	Glomerellales
	124	*Glomerella graminicola*	CM_001021	39649	29.89	Sordariomycetes	Glomerellales
.	125	*Verticillium dahliae*	NC_008248	27184	27.32	Sordariomycetes	Glomerellales
	126	*Beauveria bassiana*	NC_010652	29961	27.25	Sordariomycetes	Hypocreales
	127	*Cordyceps bassiana*	NC_010652	32263	26.99	Sordariomycetes	Hypocreales
	128	*Cordyceps brongniartii*	NC_011194	33926	27.33	Sordariomycetes	Hypocreales
	129	*Fusarium gerlachii*	NC_025928	93428	31.91	Sordariomycetes	Hypocreales
	130	*Fusarium oxysporum*	AY_945289	34477	30.98	Sordariomycetes	Hypocreales
	131	*Fusarium solani*	NC_016680	62978	28.88	Sordariomycetes	Hypocreales
	132	*Gibberella moniliformis*	NC_016687	53753	32.61	Sordariomycetes	Hypocreales
	133	*Gibberella zeae*	NC_009493	95676	31.83	Sordariomycetes	Hypocreales
	134	*Hirsutella minnesotensis*	KR_139916	52245	28.41	Sordariomycetes	Hypocreales
	135	*Hirsutella vermicola*	KY_465721	53793	25.27	Sordariomycetes	Hypocreales
	136	*Hypomyces aurantius*	NC_030206	71638	28.31	Sordariomycetes	Hypocreales
	137	*Lecanicillium muscarium*	NC_004514	24499	27.15	Sordariomycetes	Hypocreales
	138	*Lecanicillium saksenae*	N/A	25919	26.53	Sordariomycetes	Hypocreales
	139	*Metarhizium anisopliae*	NC_008068	24673	28.40	Sordariomycetes	Hypocreales
	140	*Nectria cinnabarina*	NC_030252	69895	28.71	Sordariomycetes	Hypocreales
	141	*Ophiocordyceps sinensis*	NC_034659	157539	30.21	Sordariomycetes	Hypocreales
	142	*Pochonia chlamydosporia*	NC_022835	25615	28.28	Sordariomycetes	Hypocreales
	143	*Trichoderma reesei*	NC_003388	42130	27.24	Sordariomycetes	Hypocreales
	144	*Sporothrix schenckii*	NC_015923	27125	25.10	Sordariomycetes	Ophiostomatales
.	145	*Chaetomium*	JN_007486	127206	31.43	Sordariomycetes	Sordariales

		thermophilum					
	146	*Madurella mycetomatis*	NC_018359	45590	26.87	Sordariomycetes	Sordariales
	147	*Neurospora crassa*	KC_683708	64840	36.13	Sordariomycetes	Sordariales
	148	*Podospora anserina*	NC_001329	100314	30.01	Sordariomycetes	Sordariales
	149	*Sordaria macrospora*	CABT01004783	88423	33.64	Sordariomycetes	Sordariales
Basidio-mycota	150	*Lentinula edodes*	NC_018365	121394	30.70	Agaricomycetes	Agaricales
	151	*Moniliophthora perniciosa*	NC_005927	109103	31.89	Agaricomycetes	Agaricales
	152	*Moniliophthora roreri*	NC_015400	93722	27.63	Agaricomycetes	Agaricales
	153	*Pleurotus ostreatus*	NC_009905	73242	26.35	Agaricomycetes	Agaricales
	154	*Schizophyllum commune*	NC_003049	49704	21.86	Agaricomycetes	Agaricales
	155	*Cantharellus cibarius*	NC_020368	58656	26.79	Agaricomycetes	Cantharellales
	156	*Rhizoctonia solani*	NC_021436	235849	35.92	Agaricomycetes	Cantharellales
	157	*Microbotryum lychnidis-dioicae*	NC_020353	107808	33.76	Microbotryomycetes	Microbotryales
	158	*Phlebia radiata*	NC_020148	156348	31.24	Agaricomycetes	Polyporales
	159	*Trametes cingulata*	NC_013933	91500	24.46	Agaricomycetes	Polyporales
	160	*Phakopsora meibomiae*	NC_014352	32520	34.87	Pucciniomycetes	Pucciniales
	161	*Phakopsora pachyrhizi*	NC_014344	31825	34.55	Pucciniomycetes	Pucciniales
	162	*Rhodotorula taiwanenesis*	HF_558455	40392	40.87	Urediniomycetes	Sporidiales
	163	*Tilletia indica*	NC_009880	65147	28.86	Exobasidiomycetes	Tilletiales
	164	*Tilletia walkeri*	NC_010651	59352	28.79	Exobasidiomycetes	Tilletiales
	165	*Cryptococcus neoformans*	NC_004336	24874	34.98	Tremellomycetes	Tremellales

	166	*Trichosporon asahii*	CM_001777	32568	28.90	Tremellomycetes	Tremellales
	167	*Ustilago maydis*	NC_008368	56814	31.20	Ustilaginomycetes	Ustilaginales
	168	*Flammulina velutipes*	NC_021373.1	88508	16.5	Agaricomycetes	Agaricales
	169	*Tricholoma matsutake*	NC_028135.1	76037	20.7	Agaricomycetes	Agaricales
	170	*Malassezia sympodialis*	LT671829.1	38623	32.08	Malasseziomycetes	Malasseziales
	171	*Jaminaea angkorensis*	NC_023248.1	29999	32.2	Exobasidiomycetes	Microstromatales
	172	*Fomitopsis palustris*	NC_034349.1	63479	24	Agaricomycetes	Polyporales
	173	*Ganoderma lucidum*	NC_021750.1	60635	26.7	Agaricomycetes	Polyporales
	174	*Ganoderma meredithae*	NC_026782.1	78447	26.1	Agaricomycetes	Polyporales
	175	*Ganoderma sinense*	NC_022933.1	86451	26.8	Agaricomycetes	Polyporales
	176	*Ganoderma applanatum*	NC_027188.1	119803	26.7	Agaricomycetes	Polyporales
	177	*Hericium coralloides*	NC_033903.1	72961	18	Agaricomycetes	Russulales
	178	*Heterobasidion irregulare*	NC_024555.1	114193	22.78	Agaricomycetes	Russulales
	179	*Serendipita indica*	FQ859090.1	63682	26.3	Agaricomycetes	Sebacinales
	180	*Rhodotorula mucilaginosa*	NC_036340.1	47023	40.43	Microbotryomycetes	Sporidiobolales
	181	*Tremella fuciformis*	NC_036422.1	35058	0	Tremellomycetes	Tremellales
	182	*Sporisorium scitamineum*	CP010939.1	88018	32.4	Ustilaginomycetes	Ustilaginaceae
	183	*Ustilago bromivora*	LT558140.1	177540	36	Ustilaginomycetes	Ustilaginales
Other	184	*Gigaspora margarita*	JQ041882.1	96998	45.02	Glomeromycetes	Gigaspora
	185	*Gigaspora rosea*	NC_016985.1	97350	44.84	Glomeromycetes	Gigaspora
	186	*Glomus cerebriforme*	NC_022144.1	59633	46.74	Glomeromycetes	Glomerales
	187	*Lichtheimia*	NC_024200.1	31830	34.09	Mucoromycotina	Mucorales

		hongkongensis					
	188	*Phycomyces blakesleeanus*	NC_027730.1	62082	37.9	Mucoromycotina	Phycomyces
	189	*Rhizophagus fasciculatus*	NC_029185.1	72251	37.1	Glomeromycotina	Glomerales
	190	*Rhizophagus intraradices(Glomu intraradices)*	NC_012056.1	70606	37.2	Glomeromycotina	Glomerales
	191	*Spizellomyces punctatus*	NC_003052.1	58830	31.8	Chytridiomycetes	Spizellomycetales
	192	*Parasitella parasitica*	NC_024944.1	83361	30.9	Mucoromycotina	Mucorales
Chytridio-mycota	193	*Allomyces macrogynus*	NC_001715	57473	39.49	Blastocladiomycetes	Blastocladiales
	194	*Blastocladiella emersonii*	NC_011360	36503	35.10	Blastocladiomycetes	Blastocladiales
	195	*Harpochytrium sp.*	NC_004760	24169	36.21	Monoblepharidomycetes	Monoblepharidales
	196	*Hyaloraphidium curvatum*	NC_003048	29593	43.18	Monoblepharidomycetes	Monoblepharidales
	197	*Monoblepharella sp.*	NC_004624	60432	39.32	Monoblepharidomycetes	Monoblepharidales
	198	*Rhizophydium sp.*	NC_003053	68834	22.97	Chytridiomycetes	Rhizophydiales
	199	*Spizellomyces punctatus*	NC_003052	58830	31.78	Chytridiomycetes	Spizellomycetales
Zygo-mycota	200	*Funneliformis mosseae*	KT_371477	134925	42.87	Glomeromycetes	Glomerales
	201	*Rhizophagus irregularis*	NC_014489	70820	37.23	Glomeromycetes	Glomerales
	202	*Zancudomyces culisetae*	NC_006837	58654	18.49	Incertae Sedis	Harpellales
	203	*Mortierella verticillata*	NC_006838	58745	27.88	Incertae Sedis	Mortierellales
	204	*Absidia glauca*	KU_196782	63080	28.18	Zygomycetes	Mucorales

	205	*Rhizopus oryzae*	NC_006836	54178	26.17	Incertae Sedis	Mucorales

Table S2. Intron-less *cob* and *cox3* sequences used for comparative sequence

Number	Organism	GenBank accession number
1	*Aspergillus flavus*	KP725058.1
2	*Aspergillus fumigatu*	JQ346807.1
3	*Aspergillus ustus*	KM245566.1
4	*Colletotrichum fructicola*	KM885304.1
5	*Cordyceps militaris*	KP722500.2
6	*Cordyceps militaris*	KP722499.1
7	*Fusarium mangiferae*	KP742838.1
8	*Fusarium proliferatum*	LT841261.1
9	*Fusarium solani*	JN041209.1
10	*Gibberella moniliformis*	JX910421.1
11	*Lecanicillium saksenae*	KT585676.1
12	*Magnaporthe grisea*	AY245427.1
13	*Magnaporthe grisea*	AY245424.1
14	*Magnaporthe grisea*	AY245425.1
15	*Magnaporthe grisea*	AY245426.1
16	*Microsporum canis*	FJ385030.1
17	*Monilia mumecola*	JN572107.1
18	*Phialocephala subalpina*	JN031566.1
19	*Pseudogymnoascus pannorum*	KR055655.1
20	*Saccharomyces cerevisiae*	CP006539.1
21	*Sordariam acrospora*	XM_024656032.1
22	*Thermothelomyces thermophila*	XM_003658940.1
23	*Verticillium dahliae*	DQ351941.1
24	*Sporothrix schenckii*	AB568599.1
25	*Sporothrix schenckii*	AB568600.1
26	*Acremonium chrysogenum*	KF757229.1
27	*Beauveria pseudobassiana*	KF297618.1
28	*Lecanicilliummu scarium*	AF487277.1
29	*Tolypocladiumin flatum*	NC_036382.1
30	*Beauveria bassiana*	KT201148.1
31	*Beauveria caledonica*	KT201150.1
32	*Beauveria malawiensis*	KT201147.1

33	*Chrysoporthe cubensis*	KT380885.1
34	*Chrysoporthe deuterocubensis*	KT380884.1
35	*Cladophialophora bantiana*	KX257489.1
36	*Colletotrichum lindemuthianum*	KF953885.1
37	*Cordyceps militaris*	KP722513.2
38	*Diaporthe longicolla*	KP137411.1
39	*Hirsutella minnesotensis*	KR139916.1
40	*Hirsutella thompsonii*	NC_040165.1
41	*Hirsutella vermicola*	KY465721.1
42	*Hypocrea jecorina*	AF447590.1
43	*Madurella mycetomatis*	JQ015302.1
44	*Mycosphaerella graminicola*	EU090238.1
45	*Neurospora crassa*	KY498478.1
46	*Ophiostoma novo-ulmi*	MG020143.1
47	*Pestalotiopsis fici*	KX870077.1
48	*Podospora anserina*	X55026.1
49	*Pseudocercospora mori*	NC_037198.1
50	*Pyricularia oryzae*	D88389.1
51	*Rhynchosporium agropyri*	KF650572.1
52	*Rhynchosporium secalis*	KF650575.1
53	*Saccharomyces cerevisiae*	KP263414.1
54	*Sordariam acrospora*	XM_003342320.1
55	*Sporothrix schenckii*	XM_016737278.1
56	*Tolypocladium ophioglossoides*	KX455872.1

Table S2. Intron-less mitochondrial *cob* and *cox3* sequences used for comparative sequence analysis.Organisms in green correspond to sequences used for the determination of intron-exon boundaries for *cob* I4. Organisms in orange correspond to sequences used for the determination of intron-exon boundaries for both *cob* I4 and *cox3* I2. Organisms in pink correspond to sequences used for the determination of intron-exon boundaries for *cox3* I2. Sequences were collected from GenBank using blastn with exon sequences [determined by preliminary analysis with MFannot flanking *cob* I4 and *cox3* I2 as the queries].


Blastocladiella emersonii (NC_011360)
Chytridiomycota
Zancudomyces culisetae (NC_006837)
Zygomycota
Hyaloraphidium curvatum (NC_003048)
Harpochytrium sp. (NC_004623)
Monoblepharella sp. (NC_004624)
Chytridiomycota
Spizellomyces punctatus (NC_003052)
Rhizophydium sp. (NC_003053)
Rhizopus oryzae (NC_006836)
Zygomycota
Lichtheimia hongkongensis (NC_024200)
Parasitella parasitica (NC_024944)
Phycomyces blakesleeanus (NC_027730)
Absidia glauca (KU_196782)
Mucoromycota
Mortierella verticillata (NC_006838)
Zygomycota
Rhizophagus irregularis (JQ_514225)
Zygomycota
Rhizophagus intraradices (NC_012056)
Rhizophagus fasciculatus (NC_029185)
Mucoromycota
Glomus cerebriforme (NC_022144)
Funneliformis mosseae (KT_371477)
Zygomycota
Gigaspora margarita (JQ_041882)
Mucoromycota
Gigaspora rosea (NC_016985)
Mucoromycota
Other group
Trichosporon asahii (JH_925097)
Cryptococcus neoformans (NC_004336)
Tremella fuciformis (NC_036422)
Ustilago maydis (NC_008368)
Ustilago bromivora (LT_558140)
Jaminaea angkorensis (NC_023248)
Tilletia indica (NC_009880)
Tilletia walkeri (NC_010651)
Malassezia sympodialis (LT671829)
Microbotryum lychnidis (NC_020353)
Rhodotorula taiwanensis (HF_558455)
Rhodotorula_mucilaginosa (NC_036340)
Phakopsora pachyrhizi (NC_014344)
Phakopsora meibomiae (NC_014352)
Schizophyllum commune (NC_003049)
Moniliophthora perniciosa (NC_005927)
Moniliophthora roreri (NC_015400)
Lentinula edodes (NC_018365)
Flammulina velutipes (NC_021373)
Pleurotus ostreatus (NC_009905)
Tricholoma matsutake (NC_028135)
Trametes cingulata (NC_013933)
Ganoderma lucidum (NC_021750)
Ganoderma meredithae (NC_026782)
Ganoderma sinense (NC_022933)
Ganoderma applanatum (NC_027188)
Fomitopsis palustris (NC_034349)
Phlebia radiata (NC_020148)
Hericium coralloides (NC_033903)
Heterobasidion irregulare (NC_024555)
Serendipita indica (FQ859090)
Cantharellus cibarius (NC_020368)
Rhizoctonia solani (NC_021436)
Basidiomycota
Pneumocystis carinii (NC_013660)
Pneumocystis murina (NC_020332)
Pneumocystis jirovecii (NC_020331)
Pneumocystis
Candida sake (KC_993194)
Saccharomycetales
88.8/85
100/100
100/100
100/100
100/100
100/100
100/100
77.7/97
100/100
100/100
88.5/100
100/100
98.9/100
100/100
100/100
100/100
100/100
98.4/79
100/100
100/99
74.1/64
100/100
99.9/90
100/100
99.2/100
96.1/55
100/100
93.1/99
100/100
100/100
88.7/98
31.9/49
100/100
100/100
100/83
100/100
100/100
59/93
100/100
100/100
100/100
99.4/100
100/100
100/100
99.9/98
90.2/100
99.9/98
98.1/100
99.3/100
99.5/88
100/100
100/88
49.7/59
100/100
100/100
35.7/80
100/100

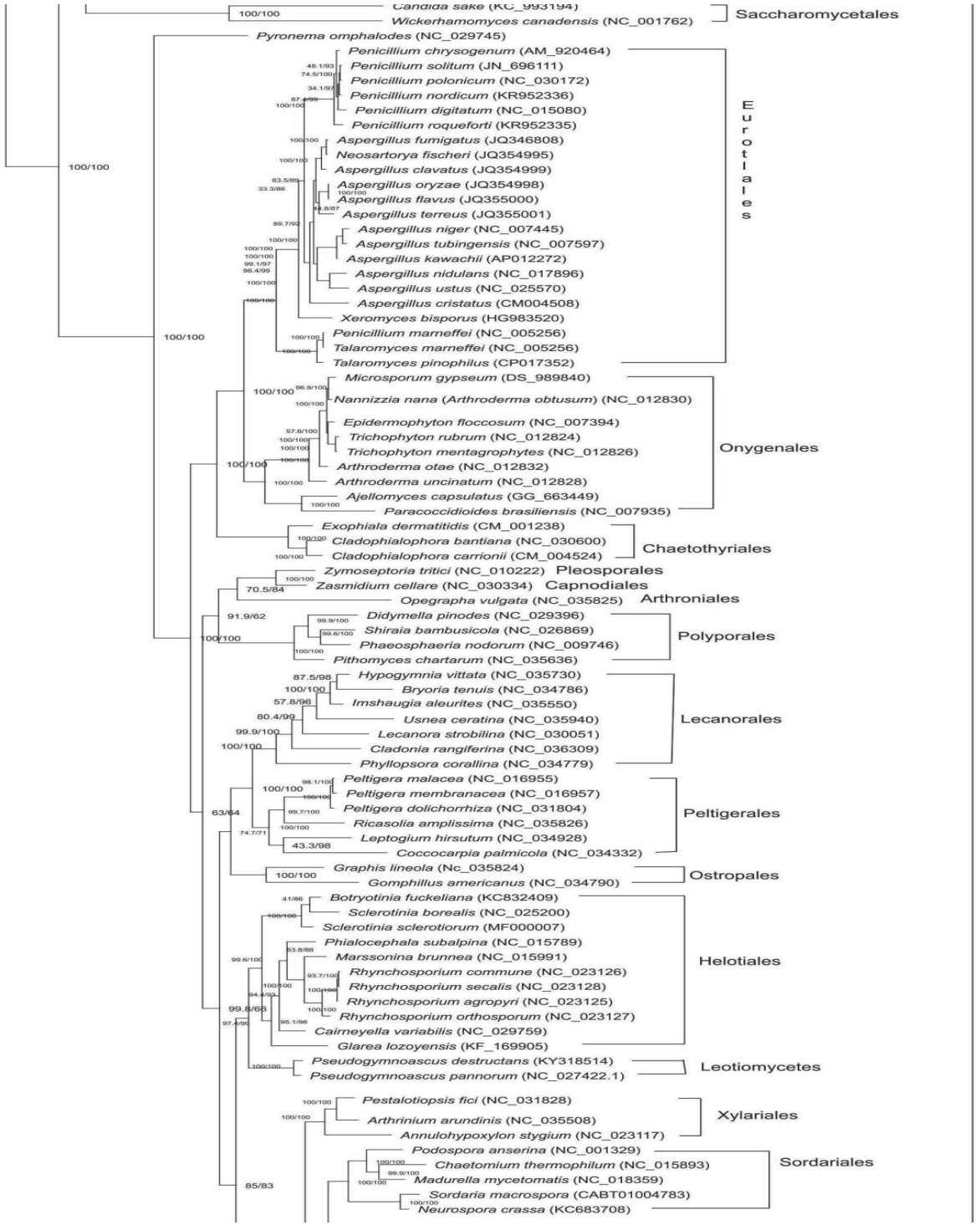

Candida sake (KC_993194)
Wickerhamomyces canadensis (NC_001762)
Saccharomycetales
Pyronema omphalodes (NC_029745)
Penicillium chrysogenum (AM_920464)
Penicillium solitum (JN_696111)
Penicillium polonicum (NC_030172)
Penicillium nordicum (KR952336)
Penicillium digitatum (NC_015080)
Penicillium roqueforti (KR952335)
Aspergillus fumigatus (JQ346808)
Neosartorya fischeri (JQ354995)
Aspergillus clavatus (JQ354999)
Aspergillus oryzae (JQ354998)
Aspergillus flavus (JQ355000)
Aspergillus terreus (JQ355001)
Aspergillus niger (NC_007445)
Aspergillus tubingensis (NC_007597)
Aspergillus kawachii (AP012272)
Aspergillus nidulans (NC_017896)
Aspergillus ustus (NC_025570)
Aspergillus cristatus (CM004508)
Xeromyces bisporus (HG983520)
Penicillium marneffei (NC_005256)
Talaromyces marneffei (NC_005256)
Talaromyces pinophilus (CP017352)
Eurotiales
Microsporum gypseum (DS_989840)
Nannizzia nana (Arthroderma obtusum) (NC_012830)
Epidermophyton floccosum (NC_007394)
Trichophyton rubrum (NC_012824)
Trichophyton mentagrophytes (NC_012826)
Arthroderma otae (NC_012832)
Arthroderma uncinatum (NC_012828)
Ajellomyces capsulatus (GG_663449)
Paracoccidioides brasiliensis (NC_007935)
Onygenales
Exophiala dermatitidis (CM_001238)
Cladophialophora bantiana (NC_030600)
Cladophialophora carrionii (CM_004524)
Chaetothyriales
Zymoseptoria tritici (NC_010222)
Pleosporales
Zasmidium cellare (NC_030334)
Capnodiales
Opegrapha vulgata (NC_035825)
Arthroniales
Didymella pinodes (NC_029396)
Shiraia bambusicola (NC_026869)
Phaeosphaeria nodorum (NC_009746)
Pithomyces chartarum (NC_035636)
Polyporales
Hypogymnia vittata (NC_035730)
Bryoria tenuis (NC_034786)
Imshaugia aleurites (NC_035550)
Usnea ceratina (NC_035940)
Lecanora strobilina (NC_030051)
Cladonia rangiferina (NC_036309)
Phyllopsora corallina (NC_034779)
Lecanorales
Peltigera malacea (NC_016955)
Peltigera membranacea (NC_016957)
Peltigera dolichorrhiza (NC_031804)
Ricasolia amplissima (NC_035826)
Leptogium hirsutum (NC_034928)
Coccocarpia palmicola (NC_034332)
Peltigerales
Graphis lineola (Nc_035824)
Gomphillus americanus (NC_034790)
Ostropales
Botryotinia fuckeliana (KC832409)
Sclerotinia borealis (NC_025200)
Sclerotinia sclerotiorum (MF000007)
Phialocephala subalpina (NC_015789)
Marssonina brunnea (NC_015991)
Rhynchosporium commune (NC_023126)
Rhynchosporium secalis (NC_023128)
Rhynchosporium agropyri (NC_023125)
Rhynchosporium orthosporum (NC_023127)
Cairneyella variabilis (NC_029759)
Glarea lozoyensis (KF_169905)
Helotiales
Pseudogymnoascus destructans (KY318514)
Pseudogymnoascus pannorum (NC_027422.1)
Leotiomycetes
Pestalotiopsis fici (NC_031828)
Arthrinium arundinis (NC_035508)
Annulohypoxylon stygium (NC_023117)
Xylariales
Podospora anserina (NC_001329)
Chaetomium thermophilum (NC_015893)
Madurella mycetomatis (NC_018359)
Sordaria macrospora (CABT01004783)
Neurospora crassa (KC683708)
Sordariales
Ascomycota

Figure S1. Phylogenetic tree generated by W-IQ-TREE program

REFERENCES

Abboud, T. G., Zubaer, A., Wai, A., & Hausner, G. (2018). The complete mitochondrial genome of the Dutch elm disease fungus *Ophiostoma novo-ulmi* subsp. *Novo-ulmi*. *Canadian Journal of Microbiology*, *64*(5), 339–348. https://doi.org/10.1139/cjm-2017-0605

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). The RNA World and the Origins of Life. *Molecular Biology of the Cell. 4th Edition*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK26876/

Basse, C. W. (2010). Mitochondrial inheritance in fungi. *Current Opinion in Microbiology*, *13*(6), 712–719. https://doi.org/10.1016/j.mib.2010.09.003

Belfort, M. (2003). Two for the price of one: A bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes & Development*, *17*(23), 2860–2863. https://doi.org/10.1101/gad.1162503

Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B., & Lambowitz, A. M. (2002). Mobile Introns: Pathways and Proteins. *Mobile DNA II*, 761–783. https://doi.org/10.1128/9781555817954.ch31

Belfort, M., & Lambowitz, A. M. (2019). Group II Intron RNPs and Reverse Transcriptases: From Retroelements to Research Tools. *Cold Spring Harbor Perspectives in Biology*, *11*(4). https://doi.org/10.1101/cshperspect.a032375

Bernier, L. (1993). Conventional and Molecular Genetic Approaches to the Study of Pathogenicity in Ophiostoma ulmi sensu lato. In M. B. Sticklen & J. L. Sherald (Eds.), *Dutch Elm Disease Research: Cellular and Molecular Approaches.* 293–307. https://doi.org/10.1007/978-1-4615-6872-8_22

Bilto, I. M., & Hausner, G. (2016). The diversity of mtDNA rns introns among strains of *Ophiostoma piliferum*, *Ophiostoma pluriannulatum* and related species. *SpringerPlus*, *5*(1), 1408. https://doi.org/10.1186/s40064-016-3076-6

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. https://doi.org/10.1016/s0968-0004(00)01612-1

Bullerwell, C. E., Forget, L., & Lang, B. F. (2003). Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences. *Nucleic Acids Research*, *31*(6), 1614–1623. https://doi.org/10.1093/nar/gkg264

Bullerwell, Charles E., & Gray, M. W. (2004). Evolution of the mitochondrial genome: Protist connections to animals, fungi and plants. *Current Opinion in Microbiology*, *7*(5), 528–534. https://doi.org/10.1016/j.mib.2004.08.008

Bullerwell, Charles E, & Lang, B. F. (2005). Fungal evolution: The case of the vanishing mitochondrion. *Current Opinion in Microbiology*, *8*(4), 362–369. https://doi.org/10.1016/j.mib.2005.06.009

Burger, G., Gray, M. W., & Franz Lang, B. (2003). Mitochondrial genomes: Anything goes. *Trends in Genetics*, *19*(12), 709–716. https://doi.org/10.1016/j.tig.2003.10.012

Calderone, R., Li, D., & Traven, A. (2015). System-level impact of mitochondria on fungal virulence: To metabolism and beyond. *FEMS Yeast Research*, *15*(4). https://doi.org/10.1093/femsyr/fov027

Candales, M. A., Duong, A., Hood, K. S., Li, T., Neufeld, R. A. E., Sun, R., … Zimmerly, S. (2012). Database for bacterial group II introns. *Nucleic Acids Research*, *40*(D1), D187–D190. https://doi.org/10.1093/nar/gkr1043

Chatre, L., & Ricchetti, M. (2014). Are mitochondria the Achilles' heel of the Kingdom Fungi? *Current Opinion in Microbiology*, *20*, 49–54. https://doi.org/10.1016/j.mib.2014.05.001

Copertino, D. W., & Hallick, R. B. (1991). Group II twintron: An intron within an intron in a chloroplast cytochrome b-559 gene. *The EMBO Journal*, *10*(2), 433–442. https://doi.org/10.1002/j.1460-2075.1991.tb07965.x

Davis, B. M., & Waldor, M. K. (2002). Mobile Genetic Elements and Bacterial Pathogenesis. *Mobile DNA II*, 1040–1059. https://doi.org/10.1128/9781555817954.ch45

Deininger, P. L., & Roy-Engel, A. M. (2002). Mobile Elements in Animal and Plant Genomes. *Mobile DNA II*, 1074–1092. https://doi.org/10.1128/9781555817954.ch47

Deng, Y., Zhang, Q., Ming, R., Lin, L., Lin, X., Lin, Y., … Wen, Z. (2016). Analysis of the Mitochondrial Genome in Hypomyces aurantius Reveals a Novel Twintron Complex in Fungi. *International Journal of Molecular Sciences*, *17*(7), 1049. https://doi.org/10.3390/ijms17071049

Dujon, B. (1989). Group I introns as mobile genetic elements: Facts and mechanistic speculations — a review. *Gene*, *82*(1), 91–114. https://doi.org/10.1016/0378-1119(89)90034-6

Ebersberger, I., de Matos Simoes, R., Kupczok, A., Gube, M., Kothe, E., Voigt, K., & von Haeseler, A. (2012). A Consistent Phylogenetic Backbone for the Fungi. *Molecular Biology and Evolution*, *29*(5), 1319–1334. https://doi.org/10.1093/molbev/msr285

Enyeart, P. J., Mohr, G., Ellington, A. D., & Lambowitz, A. M. (2014). Biotechnological applications of mobile group II introns and their reverse transcriptases: Gene targeting, RNA-seq, and non-coding RNA analysis. *Mobile DNA*, *5*(1), 2. https://doi.org/10.1186/1759-8753-5-2

Fang, X., Jiang, Y., Li, K., & Zeng, Q. (2018). F-CphI represents a new homing endonuclease family using the Endo VII catalytic motif. *Mobile DNA*, *9*(1), 27. https://doi.org/10.1186/s13100-018-0132-5

Ferat, J.-L., & Michel, F. (1993). Group II self-splicing introns in bacteria. *Nature*, *364*(6435), 358. https://doi.org/10.1038/364358a0

Gimble, F. S. (2000). Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiology Letters*, *185*(2), 99–107. https://doi.org/10.1111/j.1574-6968.2000.tb09046.x

Gray, M. W., Burger, G., & Lang, B. F. (1999). Mitochondrial evolution. *Science (New York, N.Y.)*, *283*(5407), 1476–1481. https://doi.org/10.1126/science.283.5407.1476

Guha, Tuhin K., Wai, A., Mullineux, S.-T., & Hausner, G. (2018). The intron landscape of the mtDNA *cytb* gene among the Ascomycota: Introns and intron-encoded open reading frames. *Mitochondrial DNA Part A*, *29*(7), 1015–1024. https://doi.org/10.1080/24701394.2017.1404042

Guha, Tuhin Kumar, Wai, A., & Hausner, G. (2017). Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Computational and Structural Biotechnology Journal*, *15*, 146–160. https://doi.org/10.1016/j.csbj.2016.12.006

Hafez, M., & Hausner, G. (2012). Homing endonucleases: DNA scissors on a mission. *Genome*, *55*(8), 553–569. https://doi.org/10.1139/g2012-049

Hafez, M., & Hausner, G. (2015). Convergent evolution of twintron-like configurations: One is never enough. *RNA Biology*, *12*(12), 1275–1288. https://doi.org/10.1080/15476286.2015.1103427

Hafez, M., Majer, A., Sethuraman, J., Rudski, S. M., Michel, F., & Hausner, G. (2013). The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: Twintrons, introns, and intron-encoded proteins. *Fungal Genetics and Biology*, *53*, 71–83. https://doi.org/10.1016/j.fgb.2013.01.005

Hausner, G., Iranpour, M., Kim, J.-J., Breuil, C., Davis, C. N., Gibb, E. A., … Hopkin, A. A. (2005). Fungi vectored by the introduced bark beetle *Tomicus piniperda* in Ontario, Canada, and comments on the taxonomy of *Leptographium lundbergii* , *Leptographium terebrantis* , *Leptographium truncatum* , and *Leptographiumwingfieldii*. *Canadian Journal of Botany*, *83*(10), 1222–1237. https://doi.org/10.1139/b05-095

Hausner, Georg. (2012). Introns, Mobile Elements, and Plasmids. In Charles E. Bullerwell (Ed.), *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression* (pp. 329–357). https://doi.org/10.1007/978-3-642-22380-8_13

Hausner, Georg. (2003). Fungal Mitochondrial Genomes, Plasmids and Introns.*Applied mycology and biotechnology,*Vol. *3(Fungal Genomics),*101-131

Hausner, Georg, Hafez, M., & Edgell, D. R. (2014). Bacterial group I introns: Mobile RNA catalysts. *Mobile DNA*, *5*(1), 8. https://doi.org/10.1186/1759-8753-5-8

Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., … Zhang, N. (2007). A higher-level phylogenetic classification of the Fungi. *Mycological Research*, *111*(5), 509–547. https://doi.org/10.1016/j.mycres.2007.03.004

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., … Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, *443*(7113), 818–822. https://doi.org/10.1038/nature05110

James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., & Stajich, J. E. (2013). Shared Signatures of Parasitism and Phylogenomics Unite Cryptomycota and Microsporidia. *Current Biology*, *23*(16), 1548–1553. https://doi.org/10.1016/j.cub.2013.06.057

Kennell, J. C., Moran, J. V., Perlman, P. S., Butow, R. A., & Lambowitz, A. M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell*, *73*(1), 133–146. https://doi.org/10.1016/0092-8674(93)90166-N

Kim, J., Kim, S. H., Lee, S., Breuil, C., & Nannfeldt, O. I. (rumbold. (2003). Distinguishing *Ophiostoma ips* and *Ophiostoma montium,* two bark beetle-associated sapstain fungi. *FEMS*.

Knoop, V., & Brennicke, A. (1994). Evidence for a group II intron in *Escherichia coli* inserted into a highly conserved reading frame associated with mobile DNA sequences. *Nucleic Acids Research*, *22*(7), 1167–1171

Kolesnikova, A. I., Putintseva, Y. A., Simonov, E. P., Biriukov, V. V., Oreshkova, N. V., Pavlov, I. N., … Krutovsky, K. V. (2019a). Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related Armillaria species. *BMC Genomics*, *20*(1), 351. https://doi.org/10.1186/s12864-019-5732-z

Kolesnikova, A. I., Putintseva, Y. A., Simonov, E. P., Biriukov, V. V., Oreshkova, N. V., Pavlov, I. N., … Krutovsky, K. V. (2019b). Mobile genetic elements explain size variation in the mitochondrial genomes of four closely-related Armillaria species. *BMC Genomics*, *20*(1), 351. https://doi.org/10.1186/s12864-019-5732-z

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. 5.*Mol. Biol. Evol.*33(7):1870–1874

Lambowitz, A. M., & Belfort, M. (1993). Introns as mobile genetic elements. *Annual Review of Biochemistry*, *62*(1), 587–622. https://doi.org/10.1146/annurev.bi.62.070193.003103

Lambowitz, A. M., & Belfort, M. (2015). Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution. *Microbiology Spectrum*, *3*(1), MDNA3-0050–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0050-2014

Lambowitz, A. M., Caprara, M. G., Zimmerly, S., & Perlman, P. S. (1999). *as RNPs: Clues to the Past and Guides to the Future*. https://doi.org/10.1101/087969589.37.451

Lambowitz, A. M., & Zimmerly, S. (2011a). Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harbor Perspectives in Biology*, *3*(8). https://doi.org/10.1101/cshperspect.a003616

Lambowitz, A. M., & Zimmerly, S. (2011b). Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harbor Perspectives in Biology*, *3*(8). https://doi.org/10.1101/cshperspect.a003616

Lang, B. F., Laforest, M.-J., & Burger, G. (2007). Mitochondrial introns: A critical view. *Trends in Genetics*, *23*(3), 119–125. https://doi.org/10.1016/j.tig.2007.01.006

Lehmann, K., & Schmidt, U. (2003). Group II introns: Structure and catalytic versatility of large natural ribozymes. *Critical Reviews in Biochemistry and Molecular Biology*, *38*(3), 249–303. https://doi.org/10.1080/713609236

Linnakoski, R., de Beer, Z. W., Niemelä, P., & Wingfield, M. J. (2012). Associations of Conifer-Infesting Bark Beetles and Fungi in Fennoscandia. *Insects*, *3*(1), 200–227. https://doi.org/10.3390/insects3010200

Losada, L., Pakala, S. B., Fedorova, N. D., Joardar, V., Shabalina, S. A., Hostetler, J., … Cubeta, M. A. (2014). Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiology Letters*, *352*(2), 165–173. https://doi.org/10.1111/1574-6968.12387

Maleszka, R., Skelly, P. J., & Clark-Walker, G. D. (1991). Rolling circle replication of DNA in yeast mitochondria. *The EMBO Journal*, *10*(12), 3923–3929.

Marcet-Houben, M., & Gabaldón, T. (2009). The Tree versus the Forest: The Fungal Tree of Life and the Topological Diversity within the Yeast Phylome. *PLOS ONE*, *4*(2), e4357. https://doi.org/10.1371/journal.pone.0004357

Martin, W. F., Garg, S., & Zimorski, V. (2015). Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1678), 20140330. https://doi.org/10.1098/rstb.2014.0330

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. https://doi.org/10.1016/0022-2836(90)90386-Z

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. https://doi.org/10.1261/rna.2184010

Muszewska, A., Steczkiewicz, K., Stepniewska-Dziubinska, M., & Ginalski, K. (2019). Transposable elements contribute to fungal genes and impact fungal lifestyle. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-40965-0

Nawrocki, E. P., Jones, T. A., & Eddy, S. R. (2018). Group I introns are widespread in archaea. *Nucleic Acids Research*, *46*(15), 7970–7976. https://doi.org/10.1093/nar/gky414

Paquin, B., & Shub, D. A. (2001). Introns: Group I Structure and Function. https://doi.org/10.1038/npg.els.0000884

Saldanha, R., Mohr, G., Belfort, M., & Lambowitz, A. M. (1993). Group I and group II introns. *The FASEB Journal*, *7*(1), 15–24. https://doi.org/10.1096/fasebj.7.1.8422962

Sambrook, J., & Russell, D. W. (2001). *Molecular cloning: A laboratory manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press

Sandor, S., Zhang, Y., & Xu, J. (2018). Fungal mitochondrial genomes and genetic polymorphisms. *Applied Microbiology and Biotechnology*, *102*(22), 9433–9448. https://doi.org/10.1007/s00253-018-9350-5

Spatafora, J. W., Aime, M. C., Grigoriev, I. V., Martin, F., Stajich, J. E., & Blackwell, M. (2017). The Fungal Tree of Life: From Molecular Systematics to Genome-Scale Phylogenies. *Microbiology Spectrum*, *5*(5). https://doi.org/10.1128/microbiolspec.FUNK-0053-2016

Stoddard, B. L. (2006). Homing endonuclease structure and function. *Quarterly Reviews of Biophysics*, *38*(01), 49. https://doi.org/10.1017/S0033583505004063

Stoddard, B. L. (2011). Homing Endonucleases: From Microbial Genetic Invaders to Reagents for Targeted DNA Modification. *Structure*, *19*(1), 7–15. https://doi.org/10.1016/j.str.2010.12.003

Stoddard, B. L. (2014). Homing endonucleases from mobile group I introns: Discovery to genome engineering. *Mobile DNA*, *5*(1), 7. https://doi.org/10.1186/1759-8753-5-7

Szczepanek, T., & Lazowska, J. (1996). Replacement of two non-adjacent amino acids in the S.cerevisiae bi2 intron-encoded RNA maturase is sufficient to gain a homing-endonuclease activity. *The EMBO Journal*, *15*(14), 3758–3767. https://doi.org/10.1002/j.1460-2075.1996.tb00746.x

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

Wai, A., Shen, C., Carta, A., Dansen, A., Crous, P. W., & Hausner, G. (2019). Intron-encoded ribosomal proteins and N-acetyltransferases within the mitochondrial genomes of fungi: Here today, gone tomorrow? *Mitochondrial DNA Part A, 30*(3), 573–584. https://doi.org/10.1080/24701394.2019.1580272

Xu, J., & Wang, P. (2015). Mitochondrial inheritance in basidiomycete fungi. *Fungal Biology Reviews*, *29*(3), 209–219. https://doi.org/10.1016/j.fbr.2015.02.001

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P. S., & Lambowitz, A. M. (1995). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell*, *83*(4), 529–538. https://doi.org/10.1016/0092-8674(95)90092-6

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*, 75–97.

Zubaer, A., Wai, A., & Hausner, G. (2018). The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. *Scientific Reports*, *8*. https://doi.org/10.1038/s41598-018-35926-y

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595