Comparative analysis of mitochondrial genomes and complex intron arrangements among members of the Ophiostomatales

Jigeesha Mukhopadhyay (she/they)

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba, Canada

A thesis presented to the Faculty of Graduate Studies of the University of Manitoba in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

Members of the Ophiostomatales are of economic concern as many are blue-stain fungi including some plant pathogens. The taxonomy of these fungi has been challenging due to the convergent evolution of traits associated with insect dispersal. This project involved comparative sequence analysis of fungal mitochondrial genomes belonging to Ophiostomatales (phylum Ascomycota), which are also rich reservoirs of mobile genetic elements. The work is divided into two chapters. The first chapter is focused on three members of *Leptographium sensu lato* with inflated mitogenomes owing to abundant introns, including complex intron arrangements. Phylogeny based on mitogenome-derived protein sequences and nuclear markers helped in confirming the separation of *Leptographium* and *Grosmannia* species. The second chapter investigates the mitogenomes of nine strains of *Ophiostoma ips* with a varied global distribution which were sequenced and compared with other members of the Ophiostomatales. A mitogenome intron landscape demonstrated the distribution of the mobile genetic elements and provided insight into the evolutionary dynamics of introns among members of this group of fungi. Examples of complex or nested introns composed of two or three intron modules have been observed in some species, and RNA-Seq analysis suggests possible splicing pathways with regard to resolving complex introns in *Ophiostoma ips*. Mitochondrial DNA and RNA data for several members of the Ophiostomatales provide the basis for further studies relating to alternative splicing, evolutionary intron dynamics, and taxonomic studies.

Acknowledgements

Where there is hope, there is hardship -Sea, Bangtansonyeondan

My PhD journey and my love-hate relationship with Winnipeg are intertwined. A mid-September snow storm in my first semester back in 2019 did little to prepare me for how winter here settles with purpose, thawing out only to leave slippery glassy sidewalks and sticky ice on eye-lashes. But funny enough, Winnipeg -well, mostly my university and my lab- became a sanctuary of sorts that helped me to put together all the little pieces to myself on this road to independence, scientific endeavors, and tiny accomplishments to celebrate.

My first meeting with my supervisor, Dr. Georg Hausner, went something like this- "Do you know who else has reverse transcriptases other than viruses? Group II introns!" I'm thankful for having a supervisor who encourages scientific curiosity, teaches humility, emphasizes on healthy professional boundaries while always listening with empathy, and fills in the gaps with humor. As an extension of Dr. Hausner himself, I got acquainted with a lab full of proactive, driven, and helpful scientists from diverse cultural backgrounds who jumped in to assist me in my research and beyond. For this, I would like to thank Alvan, Amal, Zubaer, Aamer, and Sawsan, as well as the project and summer students who came and went. To Alvan, who became a wonderful teacher, friend, and co-author throughout my PhD research, I'm grateful for your patience with me, as much as for the lengthy conversations about life.

My appreciation goes to my committee members, Dr. Court, Dr. De Kievit,

Dr. Marcus, and Dr. McKenna, for all the advice and scientific inputs throughout my PhD program. To Dr. Court, thank you for letting me have a day of "ideal pancake like mycelia" while working on *Neurospora crassa* in your lab. To Teri, who has been a mentor beyond my scientific research, and has always encouraged and cheered me on my literary pursuits, I am thankful for the fun conference trip outings and conversations over coffee. Dr. Marcus, I appreciate you always reminding me of the importance of flexibility and sincerity when it comes to research deadlines. Dr. McKenna, I am thankful to you for bringing the fascinating "RNA World" closer to me through the RNA Salon. I also extend my gratitude to Dr. Franz Lang for his advice on my RNA-Seq data analysis.

The support from my department community has been crucial for a smooth sailing PhD and conference experience. Rizza, Jocelyn and the prep room folks, and the office coordinators, Stephanie, Jo, and Kerry have always been super helpful. I had an amazing experience TA-ing for several courses across the Faculty of Science, and I'm thankful to Dr. Chris Rathgeber for the long association and learning opportunity. Of the many smiling faces that greeted me in the hallways, I must mention Mike, who always stopped and asked me about my day.

My PhD project was supported by NSERC, SEGS, GETS, and UMGF funding. Additional travel bursaries also facilitated my conference trips where I got to present my work to a bigger scientific community.

Beyond my lab and department, the university community, especially the various student clubs that supported me in my advocacy efforts for the 2SLGBTQIA+ BIPOC community and opened up a whole new world of opportunities for me.

A special acknowledgement goes out to my parents, who are my constant cheerleaders and have always made an effort to support me and my career/life decisions. My dearest friends who I met along this journey, Sabrin and Dongbiao, I appreciate every moment of your presence in my life.

I would quickly like to add my favorite musicians, Kim Namjoon, Park Jimin,

and Min Yoongi, whose music, lyrics, and words have been healing, comforting, and encouraging throughout this challenging phase of my academic journey.

I am grateful for my time spent in the "Mobile Intron Group", a safe space brimming with opportunities, encouragement, and a sincere commitment to science. For everyone I have encountered on this journey, past and present, I hold deep appreciation for you in my heart.

Table of Contents

List of Figures

List of Tables

Chapter 1

Literature Review

The works presented in this chapter have been published.

a. Mukhopadhyay, J., and Hausner, G. (2021). Organellar Introns in Fungi, Algae, and Plants. *Cells*, 10(8), 2001. doi: 10.3390/cells10082001

MDPI permits the use of published content for the purposes of a thesis.

https://www.mdpi.com/authors/rights

b. Mukhopadhyay, J., and Hausner, G. (2024). Interconnected roles of fungal nuclear and intron encoded maturases: At the crossroads of mitochondrial intron splicing. *Biochemistry and Cell Biology*, doi: 10.1139/bcb-2024-0046

Canadian Science Publishing permits the use of published content for the purposes of a thesis.

https://cdnsciencepub.com/about/policies/publishing-policy

J.M. and G.H. conceived the idea for the reviews. J.M. wrote the first drafts of the manuscripts, and J.M. and G.H. reviewed, edited and approved the final manuscripts. All authors have read and agreed to the published versions of the manuscripts.

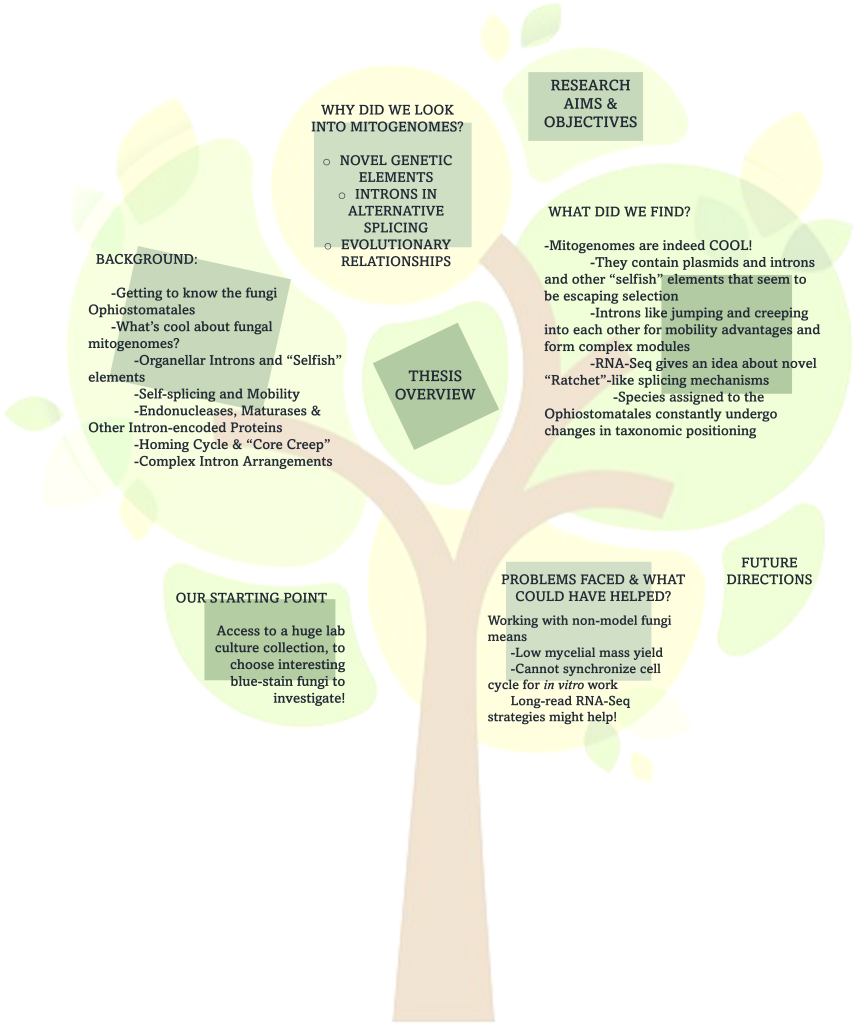

Figure 1.1: A tree summarizing my thesis research

1.1 The Fungi

Fungi are a diverse group of eukaryotic organisms, some of which have environmental and economic significance. They have a widespread global distribution. Fungi are heterotrophs (osmotrophs), and they obtain their nutrients by enzyme section and cell wall uptake processes (i.e., extracellular digestion). Their ecosystem-specific roles are diverse including decomposition, nutrient cycling, and nutrient transport (Dighton, 2007). Some are involved in symbiotic associations (lichens) while many are well-known pathogens on other plants and animals.

The role of fungi extends to applications in agriculture and food (processing, spoilage), biocontrol, and therapeutics (disease, pharmaceutics, poisons/toxins) (Guijarro et al., 2017; Katoch and Pull, 2017; Kernaghan et al., 2017; Macías-Rubalcava and Sánchez-Fernández, 2017; Narladkar et al., 2015; Oro et al., 2018; Østergaard and Olsen, 2011; Ribera et al., 2017).

The definition of fungi and organisms that belong to this group is not strict due to some unique features or traits (absence or challenging identification of synapomorphies; characteristics or traits derived from an ancestor) (reviewed by Richards et al., 2017). They have diverse ecological and morphological characteristics (can exist as haploid or diploid, asexual, sexual, or both or parasexual, and can be yeast-like, filamentous, mushroom-like, etc.). Environmental conditions also influence their morphology. Since traditional classification methods rely on phenotypic and functional observations, members of different fungal clades and lineages often give rise to taxonomic confusion. More recently, sequencing and phylogenetic approaches have been explored to study fungal taxonomy and biodiversity. A phylogenetic study using molecular data might contribute to the separation of several organisms that were once considered fungi, including the Oomycota (also known as water molds) and the slime molds (a polyphyletic group). For simplicity, the term fungi will be used to refer to organisms that belong to subkingdom, Dikarya, with an emphasis on fungi belonging in the phylum Ascomycota (unless otherwise stated), for the remainder

of this thesis.

The Order Ophiostomatales (Sordariomycetidae, Ascomycota), was described by Benny and Kimbrough (1980) accommodating the single family Ophiostomataceae. The Order includes species that are often associated with diseases affecting plants (Dutch Elm) and animal/human (sporotrichosis). They are mostly vectored by wood-infesting insects such as bark beetles (Wingfield et al., 1993; Seifert et al., 2013). De Beer et al. (2022) carried out comprehensive phylogenetic studies to unravel some of the taxonomic concerns within the Ophiostomatales. The Ophiostomatales as currently defined accommodates a single family, the Ophiostomataceae (De Beer et al., 2013a), which was initially described in 1932 (Nannfeldt, 1932). Currently 16 genera are accepted for the inclusion in the Ophiostamatles, however there could be up to 24 genera based on a recent analysis using four nuclear markers (De Beer et al., 2022).

1.2 Fungal Mitogenomes and their Variability

Mitochondria have been proposed to have originated from an ancient endosymbiosis of an ancestral free-living Alphaproteobacteria, which was integrated into the host cell (Carvalho et al., 2015; Munoz-Gomez et al., 2017). Sometimes referred to as the powerhouse of the cell, mitochondria are known for their essential role in energy metabolism, while also being crucial for carbohydrate, fatty acid, nucleic acid, and protein metabolism, apoptosis, disease, etc. Mitochondria house their own genome (mitochondrial DNA - mtDNA or mitogenome), with independent replication and inheritance, mainly encoding genes essential for protein synthesis and energy production. Throughout evolutionary history, a majority of early mt genes have been shuttled to the nucleus, and this has resulted in nuclear-encoded mt proteins that are translated by the cytosolic ribosomes and subsequently transferred to the mitochondria. Therefore, extant mtDNA mainly contains protein-coding genes that

are directly involved in oxidative phosphorylation, ATP production (ATPases) and translation (Verma et al., 2018). In most cases, a complete tRNA complement covering all twenty amino acids is also noted in the mitogenomes (Mukhopadhyay et al., 2022, 2024).

The mitogenome architecture is variable among eukaryotes, with circular mitogenomes assumed to be more common among fungi, however there are some linear mitogenomes (linear concatermers) that have been described in yeasts such as *Saccharomyces cerevisiae* (Ascomycota) (Williamson, 2002; Formaggioni et al., 2021). Typical mitogenomes in Ascomycota and Basidiomycota (Subkingdom Dikarya) fungi have a conserved core genome with 14 protein-coding genes, along with ribosomal and tRNA genes and a gene encoding for Ribosomal protein S3 (rps3) involved in protein synthesis. In some species of yeast, an exception to this conserved gene arrangement was observed in the form of an absence of the set of NADH dehydrogenase subunit genes (complex I) (reviewed by de Zamaroczy and Bernardi, 1986; Sankoff et al., 1992; Zivanovic et al., 2005). This absence seems to have been compensated for, by a nuclear counterpart, NADH:ubiquinone oxidoreductases (NADH dehydrogenases), which can act as alternatives for NAD(P)H oxidation carried out by complex I (De Vries et al., 1992; Overkamp et al., 2000; Videira and Duarte, 2002, Carneiro et al., 2004). This compensation also seems to be the trend for other genes missing as part of the "typical" composition, such as for genes encoding for ATP synthase subunit 9 (atp9) and rps3 (Lavín et al., 2008; Déquard-Chablat et al., 2011; Sellem et al., 2016). Sometimes, the mitochondrial rps3 (reviewed in Hausner, 2003) can be encoded within an *rnl* intron or be found as a free-standing gene or is nuclear encoded (Wai et al., 2019). Among the members of the Ophiostomatales, though gene order is highly conserved, some minor variations may be observed in terms of the tRNA gene set and the presence and absence of the *atp9* gene which might be compensated by the presence of a nuclear counterpart (Sellem et al., 2016; Franco et al., 2017; Zubaer et al., 2018). In addition, a few fungal mtDNAs appear to encode

a putative N-acetyltransferase (Duò et al., 2012; Zhang et al., 2017). Overall, gene content seems to be dynamic, which contributes to mtDNA size variability in fungi.

1.2.1 Organellar Introns and Other "Selfish" Elements

Introns, or intervening sequences, are segments that are transcribed but removed from a transcript before translation can proceed. There are several categories of introns, such as tRNA introns, nuclear spliceosomal introns, and group (Gr) I and Gr II introns (Parenteau and Elela, 2019). Introns are ubiquitous in eukaryotic genomes and have long been considered as 'junk RNA' but the huge energy expenditure in their transcription, removal, and degradation indicate that they may offer a significant evolutionary advantage (Frumkin et al., 2019; Zaffagni and Kadener, 2019). It has been speculated that Gr I and Gr II introns which are potential ribozymes, could have evolved in the RNA world (Koonin et al., 2006) whereas spliceosomal introns are potentially derived from Gr II introns that colonized the nuclear genome during eukaryogenesis (Martin and Koonin, 2006; Lang et al., 2007; Jo and Choi, 2014). In comparison to the impact of nuclear spliceosomal introns on gene regulation and expression, less is known about the contribution of organellar Gr I and II introns towards cellular processes.

Organellar genomes can be quite variable in size due to the absence and presence of introns; however, little is known about the mechanisms that promote intron gain and loss and why in some instances, the organellar genomes are devoid of introns, or in extreme examples where introns comprise most of the nucleotides that make up the genome (Dujon, 1989; Lambowitz and Zimmerly, 2004; Fedorova and Zingler, 2007). Organellar introns are potential mobile elements that can self-splice and, thus, minimize their impact on the host genes they have invaded. However, these introns can encode protein factors that catalyze their mobility and promote splicing. Additional genome encoded factors also appear to have been co-opted to enhance splicing and mobility. Although models have been developed that suggest

that organellar introns are neutral elements, their reliance on genome encoded factors or in some instances, *trans*-acting intron-encoded splicing factors would suggest that more complex evolutionary models may be needed to explain the biology of these elements (Lamb et al., 1983; Copertino and Hallick, 1991). The splicing of organellar introns could be a rate limiting step that can affect the expression of the genes in which they are embedded. Complex organellar introns composed of two or three intron modules have been characterized in some genomes and these might be platforms for alternative splicing (AS) or serve as molecular switches for modulating gene expression (Copertino and Hallick, 1991; Copertino and Hallick, 1993; Drager and Hallick, 1993). *Cis*- and *trans*-splicing combined with the possibility of these elements to transpose to new locations within genomes or between genomes provides a mechanism for generating genetic diversity. Gr I and Gr II intron are "building blocks" and these mobile or self-splicing modules can lead to the formation of complex-introns, variable gene architectures, and promote organellar genome evolution (Hafez and Hausner, 2015).

Organellar introns are mainly classified as self-splicing introns that are either Gr I and II type introns, based on differences in their splicing pathways and RNA secondary structural features. Sequence conservation among Gr I or Gr II introns is very low, but conservation among these elements can be found at the structural level (Anziano et al., 1982; van der Horst and Inoue, 1993; Adams et al., 2004; Michel et al., 2009). Gr I and II introns can have embedded open reading frames (ORFs) that encode for intron-encoded proteins (IEPs) (Jacquier and Dujon, 1985; Matsuura et al., 1997; Sellem and Belcour, 1997; Ho et al., 1997; Caprara and Waring, 2015; Belfort and Lambowitz, 2019).

Gr I and II introns at the RNA level assume RNA folds that generate the active sites needed for the splicing reactions to be initiated (fig. 1.2; Mukhopadhyay and Hausner, 2021).

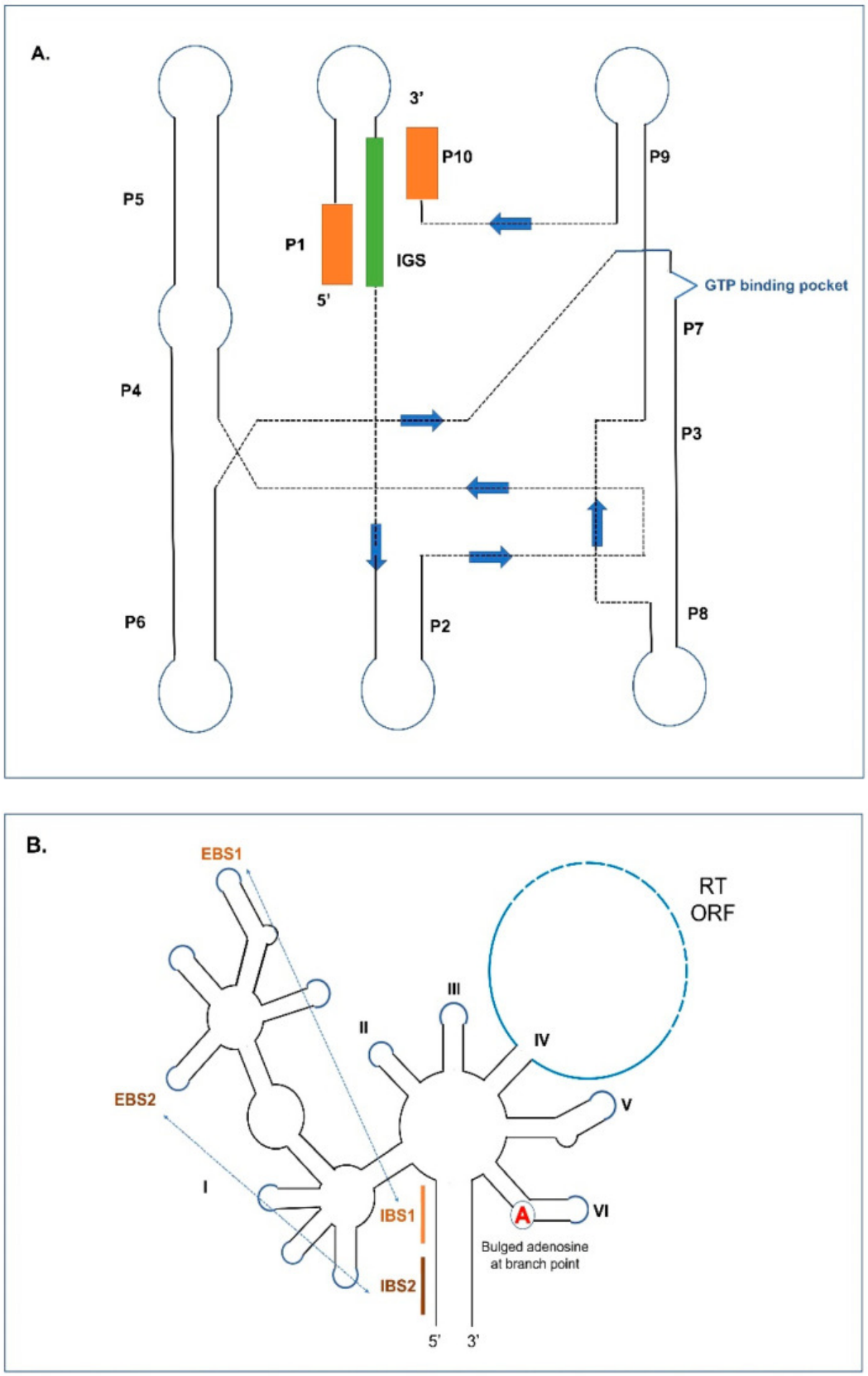

Figure 1.2: Secondary structure models for Group I and II introns

Gr I intron RNAs contain helical or paired regions referred to as P1 to P10 (some Gr I introns can have additional paired regions) with loop segments connecting the helical regions. The P1 helix includes a segment referred to as the internal guide sequence (IGS) which promotes the formation of the P10 helix, which ultimately coordinates the positions of the upstream and downstream exon sequences. The P7 helix includes a GTP binding pocket that recruits the GTP needed to initiate the splicing process. Gr I intron RNAs can self-splice by forming RNA folds where the P3 and P7 plus proximal P4, P5, P6, and P9 paired helical domains comprise the catalytic core components and the P1 and P10 helices generate the substrate domain. This RNA configuration brings the splice sites (intron/exon junctions) into proximity with each other. Gr I introns are classified into various subtypes based on variations within the primary, secondary, and tertiary structures (Lambowitz and Zimmerly, 2004; Haugen et al., 2005; Vicens and Cech, 2006; Burger et al., 2009; Glanz and Kück, 2009; Bonocora and Shub, 2009). These subtypes (IA to IE) can be further subdivided (e.g., IA1, IC3, etc.) (Michel and Westhof, 1990; Suh et al., 1999; Jackson et al., 2006).

Gr II intron RNA can be visualized as 6 stem-loop domains (referred to as domains DI to DVI) emerging from a central wheel. When ORFs are present they tend to be embedded within DIV. Domain I is the largest domain and is referred to as the scaffold domain that in part co-ordinates the folding and interactions of the other components of the Gr II intron RNA. So called exon binding sequences (EBSs) are embedded within DI that can interact via H-bonding with intron binding sequences (IBSs) that are located within the exons flanking the Gr II intron. The bulged adenine (A), referred to as the branch point, is located in DVI (Haack et al., 2019). DV is the most conserved segment of Gr II introns, and it is important in facilitating the interactions of DVI with the upstream intron–exon junction by coordinating the position of two Mg2+ ions that permits the 2′OH on the budged adenine to destabilize the phosphodiester bond (PDE) that links the upstream exon

with the intron (Jacquier and Rosbash, 1986; Peebles et al., 1986; Peebles et al., 1987; Jacquier and Michel, 1987; Daniels et al., 1996; Lambowitz and Zimmerly, 2004). Like Gr I introns, sequence conservation among Gr II introns is minimal except for DV and the intron boundaries, with GUGYG and AY (Y = pyrimidines) defining the 5′ and 3′ ends, respectively. Various subtypes of Gr II intron have been recognized based on structural features and the nature of EBS and IBS interactions (Costa et al., 2000; Pyle, 2010). Gr II intron RNAs found in organellar genomes can be classified into two major subgroups: IIA and IIB. These can be further divided into IIA1, IIA2, IIB1, and IIB2 (Michel and Ferat, 1995; Nagy et al., 2013; Smathers and Robart, 2019; Liu et al., 2020). Most mitochondrial Gr II introns can be assigned to subgroup IIA1, and many chloroplast Gr II introns can be assigned to subgroup IIB (Toor et al., 2001). Gr II-like introns lack the standard Gr II core structure, either due to loss of domains DI-IV or massive divergence from related Gr II introns (Robart and Zimmerly, 2005; Dabbagh and Preisfeld, 2016). Such degenerate Gr II introns were identified in mRNA genes of chloroplasts in euglenoid protists, having a conventional Gr II-type DVI domain with a bulged A residue, a streamlined version of DI, and absence of DII-DV (Copertino and Hallick, 1993).

Gr I introns are widespread and have been reported from nuclear rDNA (18S and 26S RNA genes) in fungi, protozoans, metazoans (Peebles et al., 1986; Peebles et al., 1987; Daniels et al., 1996; Beagley et al., 1996; Lasker et al., 1998; Müller et al., 2002; Pyle, 2010; Xu et al., 2013; Huchon et al., 2015; DePriest et al., 2016; Bernardino et al., 2017; Schuster et al., 2017; Corsaro et al., 2019; Johansen et al., 2021), bacterial genomes, phages and viruses, archaeal genomes, and they are encountered frequently in mitochondrial and plastid genomes (reviewed in Hausner, 2012). With regard to metazoans, Gr I introns have been noted in the mitochondrial genomes in members of Placozoa, Anthozoa, and Porifera (Huchon et al., 2015; Corsaro et al., 2019; Johansen et al., 2021). Noteworthy recent discoveries include putative Gr I introns in the nuclear rDNA internal transcribed spacers (ITS) in several fungi and closely

related unicellular organisms, such as *Mitosporidium* (Corsaro and Venditti, 2020), *Polychytrium* (Goryunov et al., 2021), *Amoeboaphelidium,* and *Nuclearia* (Goryunov et al., 2021). These ITS Gr I introns have been named spintrons (spacer introns) and examples of these introns have also been reported in sharks (Shivji et al., 1997).

Gr II introns also have a wide phylogenetic distribution and are found in bacteria and the organellar genomes of plants, fungi, protists, and animals (Bonen and Vogel, 2001; Dellaporta et al., 2006; Huchon et al., 2015). Though they are completely absent from nuclear eukaryotic genomes, they are hypothesized as the ancestors of spliceosomal introns during eukaryotic evolution (Palmer and Logsdon, 1991). There are rare occurrences of Gr II introns in some sponges (Huchon et al., 2015) and bilaterian animal genomes, an example being a *cytochrome oxidase* (*cox1*) gene-associated Gr II intron in the mitochondrial genome of the carnivorous polychaete *Nephtys* sp. (Vallès et al., 2008). In bacterial and archaeal genomes, Gr II introns have been identified which either entirely lack an ORF or encode novel reverse transcriptase (RT) ORFs and this indicates that certain lineages have adapted unusual RNA features, beyond being a retroelement (Simon et al., 2008). The vast majority of mitochondrial introns in angiosperms are of the Gr II type (Bonen, 2008).

It is speculated that Gr I introns are potential ancestors to the tRNA (bulge–helix–bulge) introns and Gr II introns are potential ancestors to the nuclear spliceosomal introns (Brown et al., 2014) (fig. 1.3 A, B; Mukhopadhyay and Hausner, 2021).

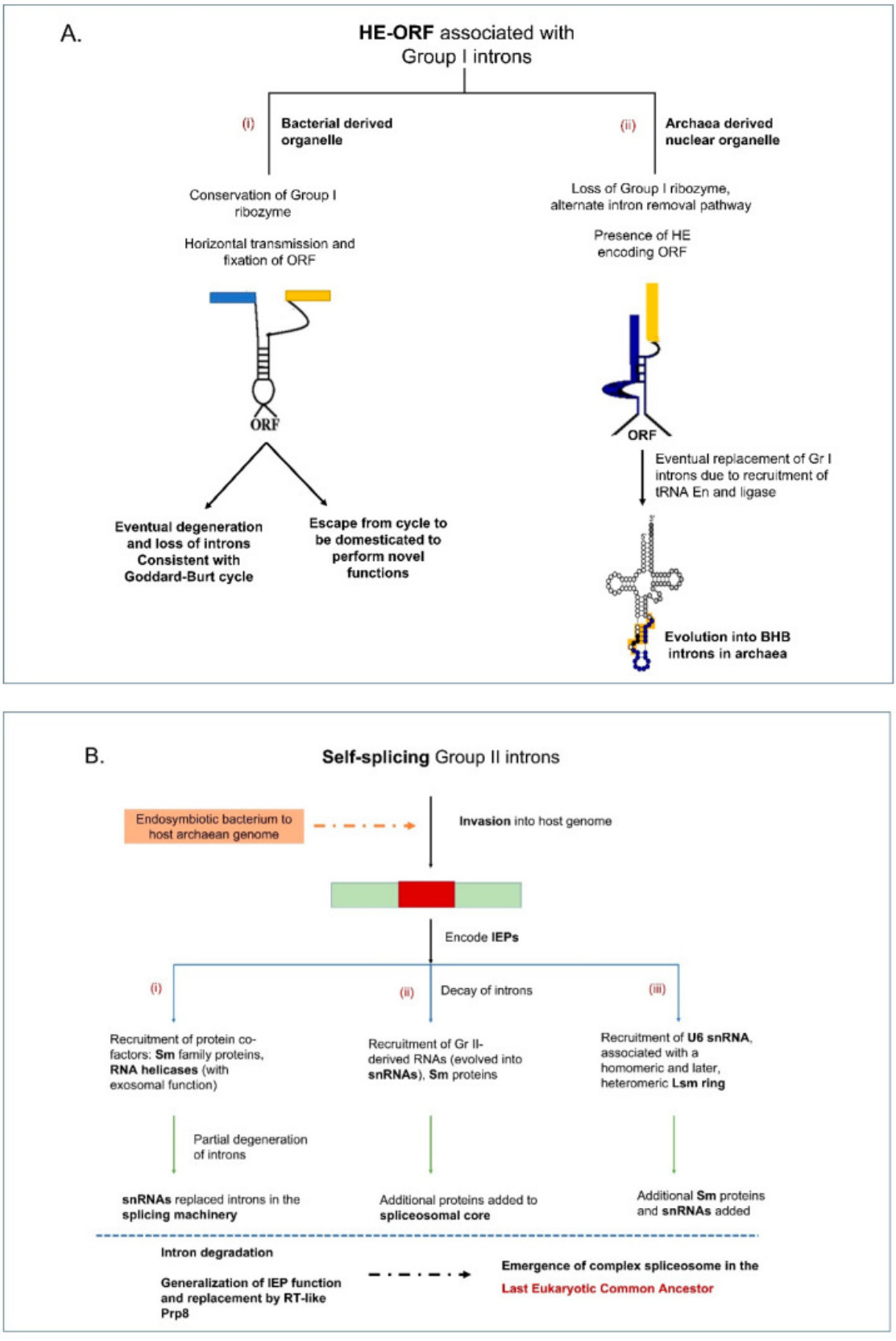

Figure 1.3: Probable evolutionary fates of self-splicing introns

Bulge–helix–bulge (BHB) introns are found within the tRNA genes (tDNAs) of the archaea and they may have evolved from Gr I intron ancestors but recruited tRNA endonucleases and ligases for intron removal (Schmidt and Matera, 2020). The canonical BHB structural motif is usually found at the exon-intron boundaries within the anticodon loop of pre-tRNA consisting of a 4 base-pair (bp) helix flanked on each 3′ side with 3 nt in the bulge (Nawrocki et al., 2018). There have been reports of possible tRNA introns in fungal mitochondrial genomes but they do not appear to fit into any known category of intron-type elements (Deng et al., 2021). Within chloroplast genomes, Gr II introns are frequently embedded within tRNA genes (Stern et al., 2010).

Among the fungi, Gr I introns are more frequently encountered compared to Gr II introns, however Gr I introns are not widespread among the Chlorophyta organellar genomes. Lichenization and maintenance of lichen symbiosis might have facilitated horizontal transfer of introns between symbionts, an example being chloroplast Gr I introns in the chlorophyte algae, Trebouxiophyceae (Del Hoyo et al., 2018). There is some speculation that plant organellar Gr I introns could have been horizontally transferred from fungal organellar genomes. It has also been speculated that introns located in the metazoan mitochondrial genomes may have been acquired via horizontal gene transfer (HGT) possibly mediated by viruses or by predation on fungi by some animals (Hausner, 2012; Johansen et al., 2021).

Gr II introns are more commonly distributed in both the mitochondria and chloroplast of the Chlorophyta. Organellar lineages of Gr II introns have been recognized that are the mitochondria-like lineage (ML) and the chloroplast-like lineage (CL) and this phylogenetic distribution is viewed to be indicative of their endosymbiotic origins of the chloroplasts and mitochondria and the co-evolution of their IEPs with the ribozyme intron RNA partners (Zimmerly et al., 2001; Lambowitz and Zimmerly, 2011). In plants, mitochondrial and chloroplast gene expression involves splicing of introns from the coding regions. In both chloroplast and mitochondrial

genomes of the angiosperms, Gr II introns tend to dominate, and these are usually found associated with the cytochrome (cyt) complexes and within ribosomal RNA and tRNA genes.

There have been reports demonstrating that organellar introns can promote genome size variation and genome rearrangement (Kroeger et al., 2009; Wu and Hao, 2014). For example, among fungal mitogenomes characterized for *Tramates* species (*Polyporales*, Basidiomycota), a huge variation was seen in intron numbers between different species due to intron gain and loss events, along with the discovery of novel introns, not detected before in Basidiomycota (Chen et al., 2021). A unique gene arrangement was also reported in the mitogenome of *Turbinellus floccosus* (Basidiomycota), with a specific bias for intron insertion downstream of a T base, preferably in the downstream site of a GT or AT, which might be a result of insertion or homing biased for specific conserved sites (Cheng et al., 2021).

Plasmids are genetic elements that can replicate autonomously from the main genome and they appear to persist by being cryptic (i.e., no phenotype/neutral) and in many instances by encoding at least one component required for their replication (Griffiths, 1995). These elements can potentially move between different species and in some instances these elements can insert by recombination into the mitogenome (Griffiths, 1995; Ferandon et al., 2008).

Plasmid-like elements (plMEs) that are derived from regions of the organellar genome have been found in a variety of organisms, with the best-studied examples found among the fungi (Hausner, 2003) such as *S. cereviseae* (Dujon and Belcour, 1989), *Aspergillus amstelodami* (Lazarus et al., 1980), *Podospora anserin*a (Begel et al., 1999; Albert and Sellem, 2002), *Ophiostoma novo-ulmi* (Abu-Amero et al., 1995), and *Cryphonectria parasitica* (Monteiro-Vitorello et al., 2009). Among the true plasmids, at least three broad categories can be recognized: (1) circular plasmids usually encoding a DNA polymerase (Griffiths, 1995); (2) linear plasmids with terminal inverted repeats encoding either a DNA or an RNA polymerase or both (Klassen

and Meinhardt, 2007); and (3) retroplasmids, which are linear or circular plasmids that usually encode an RT (Kennel and Cohen, 2004). Although true plasmids are mostly cryptic in nature, some plasmids have been associated with mitochondrial instabilities and senescence, the latter usually due to insertion of the plasmid into the mtDNA (Griffiths, 1992; Maheshwari and Navaraj, 2008; van Diepeningen et al., 2008; Nargang and Kennell, 2010), such as in *Neurospora* species, insertions are associated mutagenic effects and respiratory deficiencies and senescence (Court et al., 1992).

Miscellaneous features such as fragmented versions of both RNA and DNA polymerases have also been found in the mtDNAs of a few fungi, and these vestigial polymerase genes appear to be related to those found in true mitochondrial plasmids (Hermanns and Osiewacz, 1994). It is possible that these fragmented genes result from plasmid integration via recombination with the chromosomal mtDNA and that integrated plasmid sequences have degenerated over time in the absence of selection pressure. Linear plasmids with terminal inverted repeats might encode one or more ORFs such as phage-type single subunit DNA and/or RNA polymerases (*dpo* and *rpo*, respectively); in some instances, the *dpo* and *rpo* genes can be carried by two different plasmids (Kempken, 1995). In general, linear plasmids are cryptic (Formighieri et al., 2008; Himmelstrand et al., 2014). In a study on *Urnula criterium* (Pezizales) by Mukhopadhyay et al. (2022), four introns were identified along with a mitogenome segment derived from an integrated linear plasmid, however, no evidence of inverted repeats or duplications flanking this putative plasmid insert were detected. It has been postulated that linear plasmids can capture tRNA genes and possible non-coding ORFs (ncORFs) that might be beneficial to the host mitogenome thus ensuring the maintenance of the plasmid sequence, although it has been noted that over time the plasmid sequences (inverted repeats, *dpo* and/or *rpo* etc.) can erode due to mutations (Ferandon et al., 2008). In many fungi, mtDNA size variability may also be due to the presence of AT-rich intergenic spacers comprising

significant portions of that mtDNAs. Although these sequences might be strictly selfish DNAs, they might have regulatory functions or be DNA elements that are in a symbiotic relationship with their host's genome, for example, a non-coding sequence motif is the GC-rich PstI palindromes that are widely distributed throughout the *Neurospora* mitochondrial chromosome (Yin et al., 1981). Their structural conservation may be due to their potential for being involved in detrimental mtDNA recombination events, and for being sites of initiation of either DNA replication, transcription initiation or primary sites for the processing of transcripts.

1.2.2 Splicing Mechanisms of Group I and II introns

Self-splicing introns can be classified into different groups (I, II) based on RNA folding characteristics. The autocatalytic function of these introns depends on proper folding of the RNA molecule into a conserved three-dimensional structure involving several interactions including secondary and tertiary hydrogen bonding, base stacking, and coordination of monovalent and divalent cations (Kim and Cech, 1987; reviewed by Burke, 1988; Chastain and Tinoco, 1992; Copertino and Hallick, 1993; Adams et al., 2004; Stahley and Strobel, 2005). These introns are capable of metal-dependent self-splicing in vitro but may or may not require additional protein factors in vivo (Cech et al., 1981; Guo et al., 1995; Chien et al., 2009; Brown et al., 2014). It is interesting to note that although these introns possess highly conserved secondary and tertiary structures, the primary structures can be quite divergent, even within introns of the same subgroup (reviewed by Davies et al., 1982; reviewed by Saldanha et al., 1993).

The basic splicing reaction of Gr I introns (fig. 1.4; Mukhopadhyay and Hausner, 2021) is transesterification involving two nucleophilic attacks, first by a free guanosine (G) nucleoside co-factor, releasing the 3′OH on the upstream exon and the second attack (mediated by the 3′OH on the upstream exon), excising the intron, and joining the flanking exon sequences to form the spliced mature RNA.

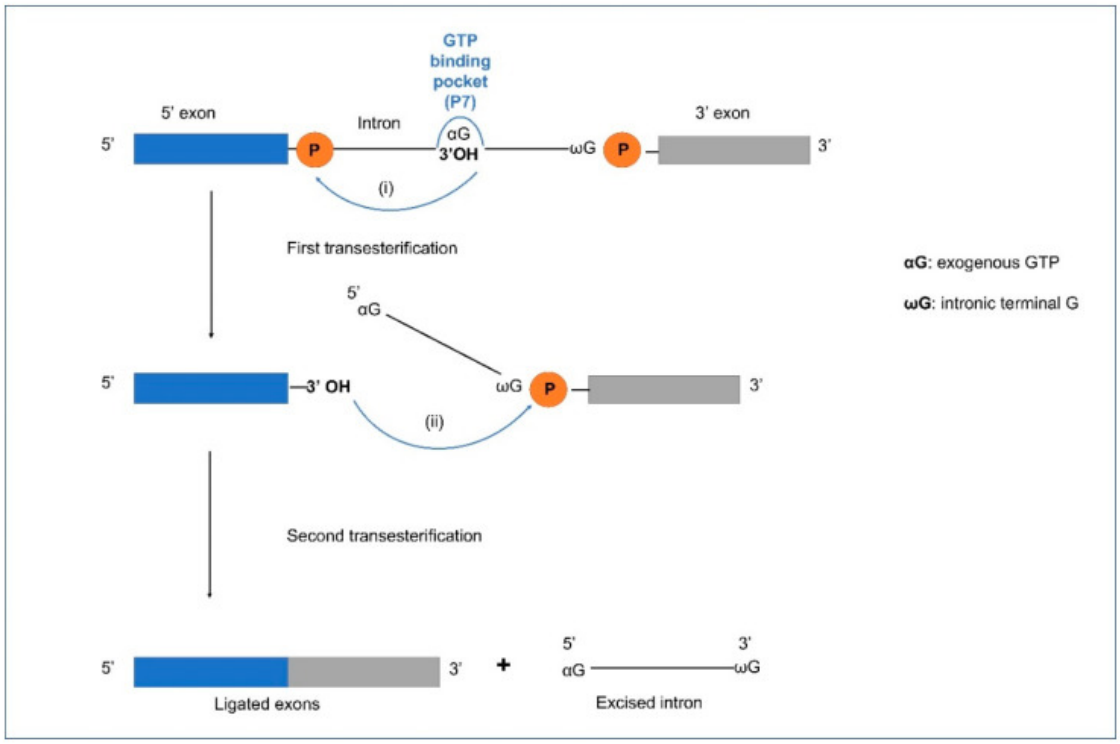

Figure 1.4: Schematic representation of splicing for Group I introns

Gr I introns may also transpose to new sites (ectopic integration) involving RNA intermediates through the process of reverse splicing. This involves a released Gr I intron RNA that can insert into a homologous or heterologous RNA facilitated by complementary base pairing between the intron IGS and complementary exon RNA sequences. Birgisdottir and Johansen (2005) investigated the site-specificity of reverse splicing of a Gr I intron located in ribosomal RNA. They used the twin-ribozyme intron (with two active sites derived from two Gr I intron modules) from the myxomycete *Didymium iridis* to show that Gr I introns can target rRNA molecules in a site-specific manner through reverse splicing (Birgisdottir and Johansen, 2005). This study showed that reverse splicing is a mechanism that provides Gr I introns with a mobility mechanism.

The Gr II intron splicing pathway (fig. 1.5; Mukhopadhyay and Hausner, 2021) also consists of two transesterification reactions.

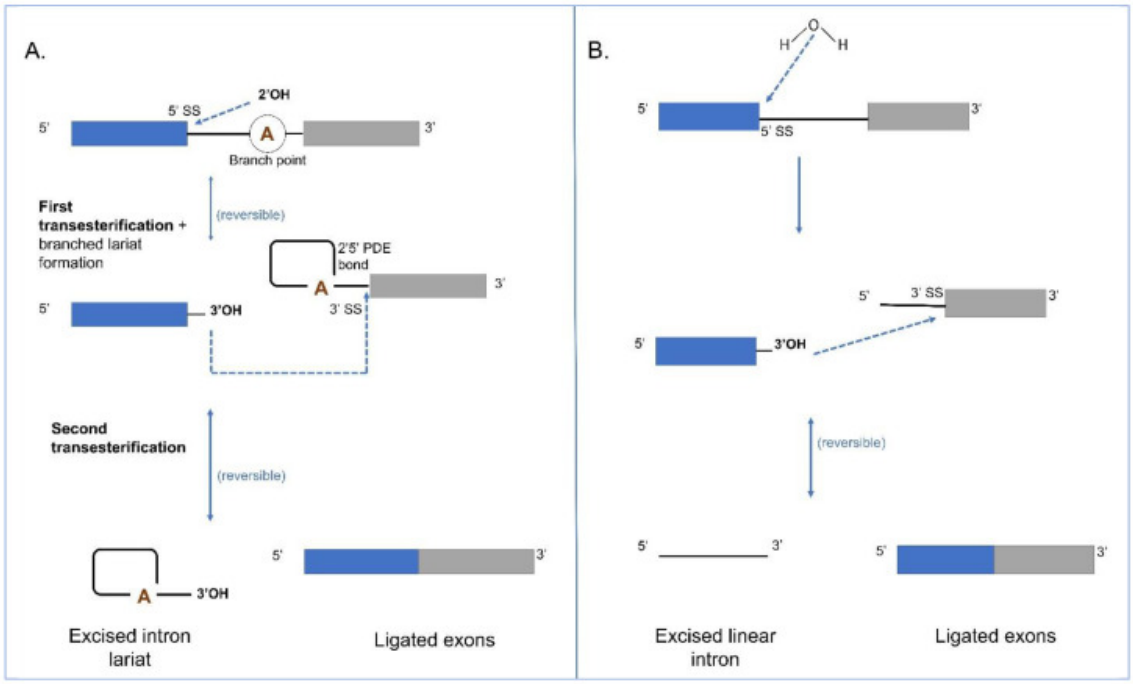

Figure 1.5: Schematic representation of splicing for Group II introns

In the first reaction, the 2′OH of an internal branch-point A, located near the 3′ end of the intron attacks the 5′ splice site (ss). This reaction, referred to as the "branching pathway", releases the 5′ exon and produces a branched lariat intermediate that is stabilized by a 5′-2′ linkage at the branch-point A (Haack et al., 2024). In the second reaction, the newly released 3′OH of the 5′ exon attacks the 3′ ss, resulting in the ligation of the 5′ and 3′ exons and excision of the intron lariat. The reversibility of the transesterification reactions allows the potential of reverse splicing of the excised intron into RNA or DNA molecules containing sequences that could base pair with the intron and exon binding sequences (Podar et al., 1998; Smathers and Robart, 2019).

For splicing to occur, the 5′ and 3′ exon sequences must be recognized by intron sequences in order to position the scissile phosphates in the active site (Lambowitz and Zimmerly, 2004). In the pre-catalytic structure, there is coordinated docking of the branch point and both splice sites in the active site before splicing (de Lencastre et al., 2005). This coordination requires two Mg2+ ions (Toor et al., 2008). Interestingly, in both Gr II introns and spliceosomal introns, the removal of the branched helix from the active site is required to allow the 3′' exon to enter and for the second step of splicing to proceed (Galej et al., 2016). In the post-catalytic lariat structure, the nucleobase component of the bulged A is disordered, and the lariat-PDE bond is located approximately 20 Å from the active site indicating that, after the first step catalysis, necessary reactants and products of the first step move out of the active site to make room for reactants of the second step. The 5′ and 3′ ends of the intron, in close proximity to the lariat bond, interact with each other through the first nucleotide, forming a non-canonical base pair with nucleotide A (Robart et al., 2014). Gr II intron splicing is facilitated by intron and host genome encoded factors.

Some Gr II introns encode a RT-like protein and these contain a maturase domain that can recognize its parent intron RNA by forming strong and highly specific

interactions with a specific RNA segment that is located within D4 of the intron. This protein-RNA interaction orients the IEP in a spatial position that facilitates splicing (Zhao and Pyle, 2017). In some instances, Gr II introns can be spliced out by hydrolysis where a water molecule or hydroxyl residue initiates the first step leading to the release of a linear intron RNA molecule (Robart and Zimmerly, 2005); the second step is the same as for the branching pathway (fig. 1.5).

In addition to canonical splicing, a peculiar backsplicing reaction observed in various intron types, sometimes generates a class of circular RNAs (circRNAs) in diverse eukaryotic species (Zhang et al., 2016; Mo et al., 2019). Two auto-catalytic complex Gr I introns that are harbored by the mitochondrial genomes of mushroom corals (Corallimorpharia) and one of these introns is processed by back-splicing (Chi et al., 2019). It has been suggested that in the latter example, backsplicing could generate transcripts more suitable for the expression of intron-encoded ORFs (Chi et al., 2019).

Another variation of "standard" intron splicing is *trans*-splicing (Glanz and Kück, 2009). *Trans*-splicing allows for fragmented genes (gene segments transcribed from different loci) to be assembled at the RNA level as intron components flanking the separated exons can assemble in *trans* and, therefore, facilitate the joining of the flanking exons generating a contiguous transcript (Nadimi et al., 2012; Dolan and Müller, 2014; Kamikawa et al., 2016). Plant mitochondria, in particular among the angiosperms, and to a lesser degree in the gymnosperms, have *trans*-split genes in which rearrangements have occurred due to various recombination events within Gr II introns (Mower, 2020; Guo et al., 2020) so that exons (and flanking half-introns) are dispersed within the genome, independently transcribed and the mRNA is generated through splicing in *trans* by assembling the intron components into the proper intron fold (Mitsuhashi et al., 2019). *Trans*-splicing of organellar introns can be associated with the requirement of RNA editing and multiple protein co-factors (Farré et al., 2012; Kück and Schmitt, 2021).

In a study of the model Gr II intron L1.LtrB from *Lactococcus lactis*, it was reported that the intron RNA can reverse splice into cellular RNAs by reverse splicing utilizing EBS–IBS interactions and the ectopically integrated intron RNAs can interact with upstream alternative circularization sites located on the intron-interrupted mRNA these promote intergenic *trans*-splicing products, such as novel Gr II intron splicing products with mRNA fragments inserted at the spice-junctions. These might provide new functions and mechanisms for generating genetic diversity (LaRoche-Johnston et al., 2018). Although this is not an organellar intron example it illustrates the versatility and potentially, adaptability of Gr II intron RNAs in promoting genetic diversity.

1.2.3 Intron-encoded Proteins: Homing Endonucleases, Reverse Transcriptases, and Maturases

Gr I and Gr II introns are commonly referred to as 'mobile introns' as they can move from an intron-containing allele to a cognate intron-less allele; this is sometimes referred to as "homing", or "retrohoming" for Gr II introns (fig. 1.6, 1.7; Mukhopadhyay and Hausner, 2021).

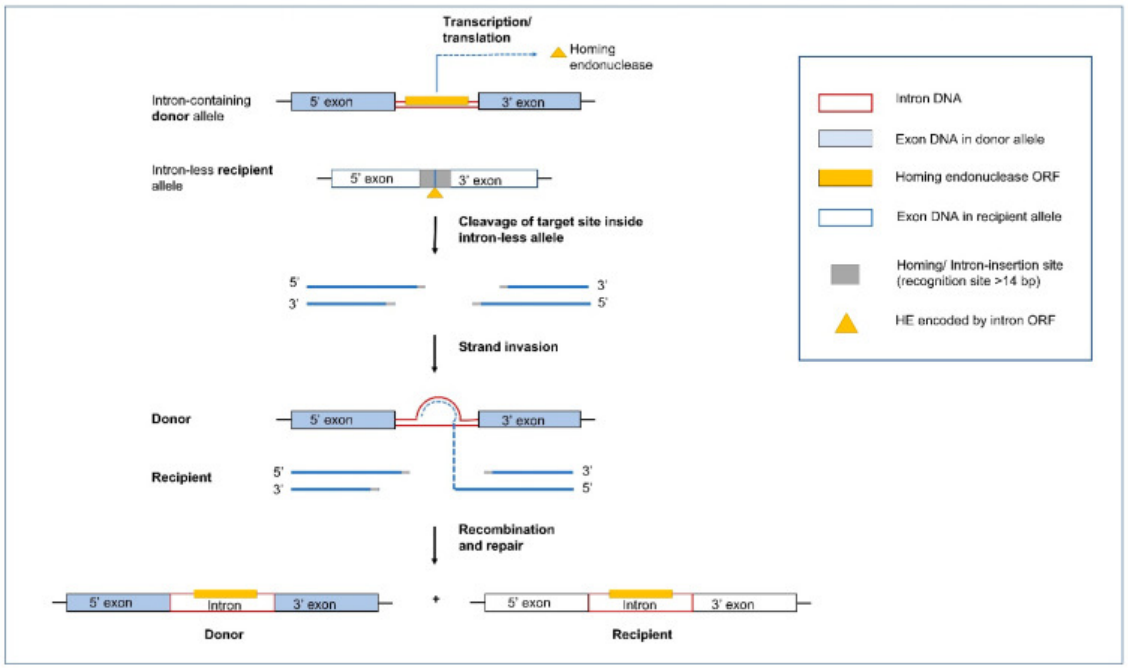

Figure 1.6: Mobility pathway of the Group I introns through homing

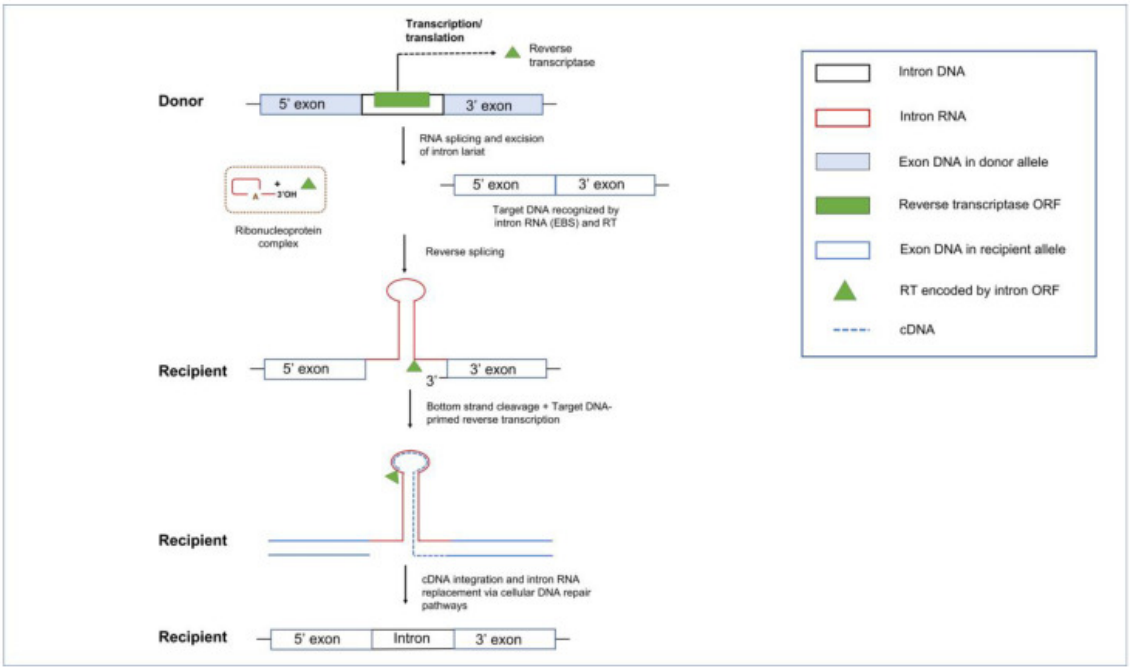

Figure 1.7: Mobility pathway of Group II introns through retrohoming

Mobility can be attributed to the presence of IEPs, such as homing endonucleases (HEs) for Gr I introns and RTs for Gr II introns. The IEPs for both types of introns have also been associated with maturase activities that facilitate (or enhance) the autocatalytic properties of these elements at the RNA level to meet the cellular demand for rapid transcript maturation and RNA turnover. Maturases are assumed to bind to intron RNAs and promote the formation or stability of splicing competent RNA folds (Zlotorynski, 2019; Moran et al., 1995; Gramespacher et al., 2019; Raghavan and Minnick, 2009).

Historically, the link between the mobility of Gr I introns and the intron-encoded HEs was determined by research on the "omega" intron (mitochondrial DNA *rnl* intron) in yeast (Dujon et al., 1976; Colleaux et al., 1988). From these pioneering studies till now, it has been shown that HEs are usually *cis*-acting and target-specific with some minor variability permitted in their homing or cleavage site an adaptation to prevent being eliminated due to neutral mutations at their target sites. Indeed, it has been recorded that the HE sequence-degeneracy usually corresponds to wobble positions within the reading frames of protein-coding host genes at the HE recognition site (Edgell et al., 2004; Scalley-Kim et al., 2007; Barzel et al., 2011).

Homing is initiated by HEs that are site-specific DNA endonucleases that recognize a specific target sequence in an intron-less allele, ranging from 14 to 44 bp in length (Chevalier and Stoddard, 2001). HEs are typically *cis*-acting by catalyzing the mobility of the genetic element that encode them. HEs introduce a double-stranded break (DSB), in the intron-less alleles triggering the host DSB-repair or synthesis-dependent strand annealing mechanism utilizing the intron/containing allele as a template to repair the break in the recipient intron-less allele. The end result is the non-reciprocal transfer of the intron-containing composite element (intron including the HE gene or HEG) into the intron-less allele (Jackson et al., 2006; Jacquier and Michel, 1987). In some instances, regions flanking the introns are also moved along to the new site along with the actual intron sequence due to gene

conversion.

There are currently six recognized classes of HEs, named after the conserved amino acid motifs: LAGLIDADG, H-N-H, His-Cys box, GIY-YIG, PD-(D/E)xK, and EDxHD (Stoddard, 2014). In fungal mitogenomes, the most abundant families are the LAGLIDADG and GIY-YIG HEs. In addition to these known classes, there are probably more classes yet to be characterized. For example, F-CphI encoded by ORF117 of cyanophage S-PM2 to represent a new HE family, defined by the motif DHHRN (Fang et al., 2018). Another novel example of a HE-derived element is the yeast mating-type switching endonuclease HO, which has become a domesticated member of an unorthodox homing genetic element family and it deviates from the classical definition of a HE in the sense that it does not propagate its own DNA sequence (does not 'home') and has rather been adapted for other normal cellular functions (Coughlan et al., 2020).

Some IEPs encoded by Gr I introns have dual maturase and HE activity and some HEs may have lost their DNA cutting activity and only act as maturases (Belfort, 2003). It has been reported that HEGs can move independently from their ribozyme partners (Mota and Collins, 1988), although recent studies suggest that intron encoded HEGs co-evolved with their ribozyme partners (Megarioti and Kouvelis, 2020).

Among the organellar genomes, the LAGLIDADG HE family is most frequently encountered. These enzymes can have one or two LAGLIDADG domains. The single motif LAGLIDADGs are active as homodimers and recognize sites that are palindromic in nature (18–22 bp with recognition flexibility). In contrast, double motif LAGLIDADGs are active as monomers and are not restricted to palindromic DNA target sequences allowing these elements to be more flexible in adapting to new target sites (Lamb et al., 1983; Gimble, 2000; Chevalier et al., 2005). Usually, LAGLIDADG HEs cut at their DNA binding site. The second most abundant HE family consists of the GIY endonucleases, characterized by a conserved amino acid

motif GIY-(X10–11)-YIG and these act as monomers that tend to cut upstream of their DNA binding site (Kowalski et al., 1999). These HEs, like the LADGLIDADGs, can be free-standing ORFs but may also be located within mobile Gr I introns in the fungal mtDNAs, in algae, and in the chloroplast DNA of plants (Stoddard, 2005).

LAGLIDADG HEGs can be components of Gr I and less frequently of Gr II introns (Toor et al., 2001; Belfort, 2003). HEGs can move along with their host introns or they can self-propagate within and in between genomes (Hausner et al., 2014). A GIY-YIG type endonuclease has also been found to be encoded within a Gr II intron (Copertino and Hallick, 1991; Lang et al., 2002). A LAGLIDADG HE encoded by a Gr II intron located within the small ribosomal subunit gene (*rns*) in the filamentous fungus *Leptographium truncatum* demonstrated to have the capacity of provide a means of mobility for its host introns but did not display any maturase activity (Mullineux et al., 2010). This study provided the first biochemical analysis of a Gr II intron that encodes a Gr I intron derived LAGLIDADG HE rather than a RT. Intron homing dependent on a group-II intron-encoded HE was also studied in the maize smut fungus, *Ustilago maydis* (Pfeifer et al., 2012). This study focused on a Gr II intron located within the *rnl* gene. The intron encoded HE (I-*Uma*I) represents a Mg2+ dependent endonuclease that requires both LAGLIDADG domains for activity and recognizes a minimum target site of 14 bp and can provide the Gr II intron with a DNA based homing mechanism.

HEGs are viewed to be diversity-generating elements and can propagate efficiently in organellar genomes by a homing mechanism that is based on homologous recombination. Their site-specific nature favors targeting conserved sequences and they persist in the presence of repetitive genes or genomes (multi-copy) and in organisms where there are opportunities for vertical and horizontal transmission of genetic elements (Hausner, 2012; Yan et al., 2018). For most cytoplasmic components (including mitochondria and chloroplasts), inheritance tends to be uniparental, ensuring uniparental inheritance of HEGs and their intron partners. However, so-

called "super-mendelian" inheritance has been noted in uni- or bi-parental inheritance systems in yeasts where introns can spread to the progeny efficiently due to the potential of mitochondria originating from the two mating types fusing and introns homing into intron-less alleles (Dujon, 2005). HEs catalyze intron mobility in organelles and there are indications that nuclear genes may impact the mobility and possible inheritance of mobile introns. A recent study on the impact of the nuclear gene *SXI1 a* in the *Cryptococcus neoformans* species complex on mitochondrial inheritance and intron/HEG transmission showed that disruption of this gene results in biparental mitochondrial inheritance and significant heteroplasmy and an increase in the number of introns in the progeny. Thus, this sex-determining nuclear gene *SXI1 a* appears to be critical for inhibition of the spread of HEGs in the mitochondrial genome and its absence leads to the over-transmission of HEG-containing introns (Yan et al., 2018). This study shows that there are mechanisms that can modulate the impact (or number) of organellar mobile elements.

Mobile Gr II introns spread within and between genomes by a mechanism of "retrohoming" in which the intron RNA inserts directly into a DNA site and is reverse-transcribed by an intron-encoded RT (Kaer and Speek, 2012). Gr II IEPs are multifunctional due to the presence of several protein domains: the N-terminal (NT) RT domain followed by the X (or maturase) domain. Some RTs also contain a DNA-binding domain (DBD) and an endonuclease (EN) domain at the C-terminus (CT) of the RT protein (Robart and Zimmerly, 2005).

In the retrohoming pathway, the basic steps include the release of the intron RNA lariat from the transcript followed by the intron RNA binding of the intron encoded RT protein to form a ribonucleoprotein (RNP) complex which can initiate retrohoming. The RNP complex can scan for the DNA target site utilizing the RT DBD and the EBS segments on the intron RNA. The EBS elements guide the RNP to a homing site via base pair complementarity and the 3′ end of the intron lariat can generate a single-stranded cut at the double-stranded DNA target site. This

single-stranded cut allows for the process of reverse splicing of the intron RNA into the DNA homing site. For those RTs that have an EN domain, the RT will cleave the complementary strand of the target site; this will generate a 3 prime end that can be utilized as a priming site for the RT to start the reverse transcription of the intron RNA into cDNA (Zimmerly et al., 1995a and b; Yang et al., 1996; Cousineau et al., 1998). The intron RNA will eventually be replaced by the DNA repair system. Retrohoming involves numerous host factors, such as components of the DNA repair and recombination machinery, and DNA replication components (Liu et al., 2020; Smith et al., 2005; Coros et al., 2009; Contreras et al., 2013; García-Rodríguez et al., 2019).

Gr II introns have diversified the genomes in all domains of life, and they are good candidates for being the progenitors of the eukaryotic spliceosomal introns. Evidence for their common ancestry is apparent from similarities between recently resolved crystallography-based structures of the eukaryotic spliceosome and structures of Gr II intron IEPs and selected Gr II intron RNAs. Shared features in RNA secondary structural elements, metal ion binding sites, and parallels in splicing mechanism have been discussed extensively in previous reviews (Qu et al., 2016; Novikova and Belfort, 2017; Chan et al., 2018; Haack and Toor, 2020; Gupta et al., 2014).

Compared to plant organellar intron splicing, fungal mitochondrial intron splicing is still poorly explored. The literature suggests that nuclear factors in synergy with intron-encoded maturases, belonging to ATP-dependent helicases and tRNA synthetases have direct impact on mitochondrial intron splicing through their association with intron RNA. In addition, several nuclear encoded accessory factors might drive the splicing impetus through translational activation, mitoribosome assembly, and phosphorylation-mediated RNA turnover. The expression of mitochondrial genes in fungi appears to be regulated primarily at the post-transcriptional level, thus protein-assisted splicing of introns by nuclear and mitochondrial encoded maturases are a means of mitonuclear interplay that could respond to environmental and

developmental factors promoting phenotypic adaptation and potentially speciation.

Maturases are defined as protein factors that can physically interact/contact the intron RNA in order to facilitate the intron RNA to fold into a splicing competent form. Splicing factors include all proteins/molecules that directly or indirectly promote splicing (includes maturases); this can also include components that enhance the expression of factors that promote the splicing of introns. Mitochondria require many nuclear-encoded factors for gene expression therefore organellar function is linked with nuclear gene expression. This establishes extensive inter-compartmental cross-talk that can integrate organellar gene expression into the cellular context as influenced by physiological, developmental and environmental cues (Cinget and Bélanger, 2020; Dujon, 2020; Hill, 2020; Nguyen et al., 2020).

Splicing of mitochondrial introns is a complex multi-step process that depends upon the cross-talk between intron-encoded maturases and nuclear-encoded splicing factors (Grivell, 1995). Multiple nuclear-encoded proteins appear to form a complex where the primary splicing factor directly interacts with the intron RNA, accessory proteins might aid in stabilizing these associations, optimizing steric, thermodynamic and kinetic conditions for the formation of splicing-competent folds, or through indirect roles such as transcriptional and/or translational activation. Nuclear encoded maturases can be assigned into two major categories, the DEAD box helicases/RNA chaperones and the tRNA synthetases. Among the DEAD-box helicases, CYT-19 from *N. crassa* and its yeast counterpart, Mss116p, have been studied extensively for their role in splicing. Among the tRNA synthetases, TyrRS (CYT-18) and LeuRS (NAM-2) have been extensively studied.

Aminoacyl-tRNA synthetases (aaRSs) are a conserved family of enzymes with an essential role in protein synthesis: ligating amino acids to cognate tRNA molecules for translation. In addition to their role in tRNA charging, aaRSs have acquired non-canonical functions, including post-transcriptional regulation of mRNA expression (Martinis et al., 1999a, b; Weiner and Maizels, 1999; Schimmel and Ribas de

Pouplana, 2000; Pang et al., 2014; Garin et al., 2021). Studies suggest that the aaRSs may use pre-existing protein motifs and modules, e.g. those that bind RNA, to gain new activities, but aaRSs can also acquire extra domains or insertions to confer or enhance secondary functions. Two mt tRNA synthetases, leucine [LeuRS or NAM2 (Labouesse et al., 1985; Herbert et al., 1988; Labouesse, 1990)] and tyrosine [TyrRS or CYT-18 (Akins and Lambowitz, 1987; Kamper et al., 1992)], have been shown to be required to facilitate Gr I intron splicing in *S. cerevisiae* and *N. crassa*, respectively (Lambowitz and Perlman, 1990; Dujardin and Herbert, 1997; Lambowitz et al., 1999).

The mt tyrosyl tRNA synthetase (TyrRS) from *N. crassa* (CYT-18 protein) is a bifunctional Gr I intron splicing cofactor that is capable of splicing multiple Gr I introns from a wide variety of sources by stabilizing the catalytically active intron structures (Geng and Paukstelis, 2014). These synthetases are encoded by nuclear genes and synthesized by cytosolic ribosomes; after translation, precursor mt proteins are maintained in an unfolded form though associated with chaperones and are translocated via an organelle-targeting signal to the mitochondria (Garin et al., 2020). They show dual localization, facilitated by mt import, to function in tRNA charging in both the cytosol and the mitochondria, and mt aaRS utilizes different tRNA recognition and charging elements compared to their cytosolic counterparts. Yeast mtTyrRS, which diverged from Pezizomycotina fungal mtTyrRSs prior to the evolution of splicing activity, binds Gr I intron and other RNAs non-specifically via its CT domain (CTD), but lacks further adaptations needed for Gr I intron splicing. Yeast mtTyrRS CTD can replace the CYT-18 protein to promote aminoacylation but not Gr I intron splicing (Lamech et al., 2014). Another tRNA synthetase, NAM2, which is a LeuRS from *S. cerevisiae* and *S. douglasii* is also involved in the splicing of the *cob*I4 intron (Herbert et al., 1988; Rho et al., 2002). These synthetases contain a connective polypeptide (CP1) domain (Starzyk et al., 1987; Hou et al., 1991) along with an intron-binding domain that has been shown to

facilitate self-splicing independent of full-length tRNA synthetase.

Both Gr I and Gr II introns encode proteins that can promote their mobility and it is assumed in many instances these proteins can act as maturases that encourage the formation of splicing competent intron RNA folds (Lazowska et al., 1980, 1989; Anziano et al., 1982; Weiss-Brummer et al., 1982; Goguel et al., 1992; Szczepanek and Lazowska, 1996; Gimble, 2000; Geese and Waring, 2001; Bolduc et al., 2003; Longo et al., 2005). Maturase activity has been demonstrated for a small set of intron-encoded proteins (reviewed in Belfort, 2003; Cui et al., 2004; Caprara and Waring, 2015; Edgell et al., 2011; McNeil et al., 2016; Zhao and Pyle, 2017; Mukhopadhyay and Hausner, 2021) including mitochondrial Gr II intron-encoded RTs (Kennell et al., 1993; Lazowska et al., 1994; Moran et al., 1995; Zimmerly et al., 1999; Robart and Zimmerly, 2005; Dai et al., 2008; Lambowitz and Zimmerly, 2011), and the consensus is that intron encoded proteins initially facilitated mobility and maturase activity evolved later. Maturase activity evolved in response to drift that introduced mutations within the intron sequence that might reduce splicing efficiency (Toor et al., 2001; Scalley-Kim et al., 2007, Edgell, 2009; Edgell et al., 2011; Megarioti and Kouvelis, 2020).

The shift from catalyzing intron mobility to intron RNA splicing required IEPs to gain the affinity to bind to intron RNA. In the yeast *cob* bI3 maturase, a mutation in the active site disrupted EN activity; this protein likely evolved from a DNA-binding to an RNA-binding protein (Bassi et al., 2002; Bassi and Weeks, 2003; Longo et al., 2005) and facilitated intron splicing at the expense of mobility. The authors proposed a model for the evolution of the bI3 maturase in which they suggest that initially the protein was an active DNA EN and invaded the target site and subsequent mutations resulted in the evolution of RNA binding properties. Genes encoding endonucleases spread rapidly in the genome and possibly the population (by horizontal transmission) and may also insert into ectopic sites at lower frequency because of flexibility in DNA target site recognition. Once all possible homing sites

have been occupied, there is little selective pressure on the EN activity and the ORF sequence can accumulate mutations that may fortuitously result in maturase activity. The protein is then effectively co-opted to promote efficient splicing of its own or related intron (Longo et al., 2005).

The interaction of intron-encoded and nuclear-encoded splicing factors are part of the mitonuclear interplay that allow mitochondrial gene expression to respond to nuclear cues in response to environmental and developmental signals (Visinoni and Delneri, 2022). These interactions may also be part of mitochondrial-nuclear co-adaptations that can result in phenotypic differences and impact the so-called "fitness landscape" with regards to the nuclear-mitochondrial combinations that can exist within a population (Dujon, 2020; Hill et al., 2020; Nguyen et al., 2020).

1.2.4 Homing Cycle and Other Evolutionary Speculations

The life cycle for the homing endonuclease genes (HEGs) commonly known as the 'homing cycle' was proposed (Goddard and Burt, 1999). According to this event, an empty site within a genome is invaded from another organelle or organism by a Gr I intron- or intein-associated HEG via horizontal transmission. Subsequently, the homing mechanism stably replicates the Gr I intron or intein gene and its associated ORF to identical loci in a recipient intron-less or intein-less cognate alleles. As these elements appear to be neutral, there is a lack of selection so inactivation and eventual elimination of the intron or the intein gene arises due to point mutations within the HEG leading eventually to the loss of the HEG and intron. Thus an empty site is regenerated and this step prepares the stage for the second invasion which continues the cycle of invasion and loss.

Koonin et al. (2006) proposed that Gr I and Gr II introns evolved in the pre-cellular RNA world. According to this speculation, the primordial pool of primitive genetic elements was conceptualized also as the source for the original lineages of viruses and related selfish elements. Moreover, it was also speculated that the mito-

chondrial endosymbionts that gave rise to the eukaryotic organelles probably carried with them mobile elements such as mobile introns and plasmids (Koonin et al., 2006; Martin and Koonin, 2006; Hausner, 2012).

The process of intron homing appears to be widespread. Based on reviews (Lambowitz et al., 1999), 30% of Gr I introns are estimated to contain internal ORFs and a significant number of them are assumed to be mobile. Gr I intron homing, so far, is the most wide-spread reported event compared to the homing exhibited by the Gr II intron. It is found in mitochondrial DNA of fungi, mitochondrial and chloroplast genomes of plants, algae and some protozoans as well as nuclear genomes of slime molds, ciliates, algae, fungi, soft corals and sponges (Gimble, 2000; Hafez and Hausner, 2012). In contrast, Gr I introns are less frequently encountered among bacteria (reviewed in Hausner et al., 2014). If present, they are predominately inserted within structural RNA genes such as tRNA (Paquin et al., 1997; Rudi et al., 2002) and rRNA genes (Salman et al., 2012), protein coding genes such as *nrdE* genes in some cyanobacteria (Meng et al., 1997; Fujisawa et al., 2010), *nrdE* and *recA* genes in various *Bacillus* species (Tourasse et al., 2006; Ko et al., 2002), flagellin gene in a thermophilic *Bacillus* species (Hayakawa and Ishizuka, 2009, 2012). Gr I introns are ancient, and have been in recently discovered in Archaea (Nawrocki et al., 2018). Currently, there are three speculations for such scarcity of Gr I introns among the prokaryotes. Homing is facilitated by the presence of multiple targets offered by repetitive DNAs or multi copy genomes such as chloroplast and mitochondrial DNAs in eukaryotes with lower mutation rates (Hausner, 2012). The absence of such multicopy targets in bacteria may be the first factor that explains why mobile introns such as Gr I introns are not so common amongst bacteria. Second, the extremely prevalent presence of primitive defense mechanisms, possibly, based on the RNA interference principle and the newly discovered CRISPR/Cas defense system in the bacterial genome might limit the spread of foreign DNA elements like mobile Gr I introns (Barrangou, 2013; Hausner et al., 2014; Silas et al., 2016). Unlike the

eukaryotic transcription and translation machineries which are compartmentalized, insertion of 11 Gr I introns into the protein-coding genes in bacteria, which exhibits coupled transcription and translation events may not be welcoming, the later is supposed to interfere by providing lesser time in proper folding of the Gr I introns to facilitate ribozyme formation and thus efficient splicing. This could be the third factor that would ultimately lead to the elimination of such mobile introns from the bacterial genomes (Ohman-Hedén et al., 1993; Edgell et al., 2000; Hausner et al., 2014).

The coevolution of fungal mitochondrial introns and their corresponding HEs (GIY-YIG and LAGLIDADG families) has recently been studied to propose a model called the "aenaon hypothesis" which describes the invasion of ancestral free-standing HEGs into introns through the evolutionary events of recombination, transposition, and HGT (Megarioti and Kouvelis, 2020). The coevolution of the HEGs associated with introns is crucial in understanding the variability in the organellar genomes. It has been observed that frequently intron encoded ORFs are fused to the upstream exons of their host genes. The inter-dependence of LAGLIDADG and GIY HEs with their upstream exons provides these ORFs access to the *cis* genetic elements that are required for their expression (Edgell et al., 2010; Guha et al., 2018). The "aenaon model" combines characteristics of the "intron late" and "intron early" theories (Koonin et al., 2006) proposing that there are ancestral introns and HEGs with conserved site recognition throughout fungal evolution. Mitochondrial introns have evolved bidirectionally, either, through migration to similar target sites but with different actual locations (ectopic, for example, orthologous introns associated with *cox1* and *cox2* genes) or through change in the intron RNA structure of their compact ancestral form (expansion through addition of new hairpins or the less common, reverse reduction). The expansion or reduction in the RNA structure of the introns themselves, is accompanied by the change in the overall architecture of the mitochondrial genomes, with the persistence of certain ancestral intron forms along

with the addition of new introns in the mitogenomes of higher fungi. Over time, free-standing HEGs have also been replaced with intron-associated HEs (Brankovics et al., 2018). Recent studies have shown that for LAGLIDADGs, minor amino acid changes within the HEs active sites can cause the HE to target different sites, thus allowing the HE plus their ribozyme host to transpose to new sites (Lamb et al., 1983). The "aenaon model" proposes a dynamic coevolution pattern of introns and HEGs, resulting in the great diversity of intron complements as observed among the fungal organellar genomes (Zubaer et al., 2018, 2021). These findings may also be relevant to non-fungal organellar genomes.

The genomes of fungal and plant mitochondria, chloroplast genomes of eubacteria, algae and plant encounter Gr II intron homing (Belfort et al., 1995). Gr II introns have been recorded in early branching metazoans (reviewed in Hausner et al., 2014; Huchon et al., 2015), however, they are rare in archaea (Rest and Mindell, 2003) and have not been found in the nuclear genomes of eukaryotes. It has been suggested that Gr II introns gave rise to spliceosomal introns and various types of retroelements which are highly abundant in eukaryotes (Xiong and Eickbush, 1990). It is interesting to note that spliceosomal introns are not known to be mobile (Lambowitz and Zimmerly, 2011). The structural similarities between Gr II introns and spliceosomal messenger RNA (mRNA) introns in eukaryotic genome suggest that they might be derived from once-mobile Gr II introns (Weiner, 1993; Sharp, 1994; Koonin et al., 2006). Engineered Gr II introns invading ectopic sites in the eukaryotic chromosome further support this theory (reviewed in Molina-Sánchez et al., 2015). Archaeal introns are present within tRNA and rRNA genes, although they have rare occurrence (Lykke-Andersen et al., 1997).

Two major events would have facilitated the transformation of Gr II introns (or a common ancestor) into spliceosomes. The first would be fragmentation of a Gr II intron RNA into pieces (such as the bipartite and tripartite Gr II introns observed in plants), ultimately resulting in functional RNAs that would operate on splice sites in

trans. In addition, the recruitment of splicing factors occurred to compensate for the loss of RNA structural features that optimized autocatalysis of the splicing reaction and to facilitate splicing reactions in *trans*-configurations. Finally, which is mechanistically dependent on the first, would be the ability of Gr II intron components to function as multiple-turnover enzymes that catalyze splicing at numerous sites. One interesting finding is the domestication of the Gr II intron-encoded RT into splicing factors such as the nuclear protein PrP8 which is an important component of the spliceosomal machinery involved in removing nuclear spliceosomal introns (Dlakic and Mushegian, 2011; Vosseberg and Snel, 2017). Other RT derived splicing factors are the nuclear encoded nMAT1, nMAT2, nMAT3, nMAT4 that facilitate splicing of introns in the mitochondrial genomes of the angiosperms (Mohr and Lambowitz, 2003; Malik et al., 2017), and matR (located in plant mitochondrial genomes) and matK (located in plant chloroplast genomes) (Hausner et al., 2006).

The ability of an intron to transpose to new sites might rely on the acquisition of novel target recognition sequences through accumulation of mutations in the IEPs. Reverse splicing of intron RNAs into RNA sequences is less efficient as there is a requirement for a reverse transcription step followed by recombination but the RNA interactions (wobble, etc.) might be more relaxed compared to protein mediated mobility. Homing endonuclease interactions with target sites is in part based on indirect readout (i.e., not necessarily based on specific nucleotide bases interacting with specific amino acids) so sometimes minor changes at the protein level can shift the enzymes specificity to a new target site (Lamb et al., 1983; Lambert et al., 2016). However, once a mobile intron has inserted into a new target site its persistence is based on drift (neutral evolution) and natural selection (Gimble, 2000; Gogarten and Hilario, 2016). If the element is neutral with regard to its impact on the host genome it can accumulate mutations leading to degeneration of the IEP coding and ribozymes sequences leading to the loss of the intron. Maintenance of an intron is dependent on the correlation between the energy burden of housing the intron and

possible advantages the intron may provide (Mullineux et al., 2011).

Goddard and Burt (1999) examined the distribution of HEGs and their host introns in members of the Saccharomycetales. They noticed that HEGs along with their intron partner move into a target site, but this is followed by the accumulation of mutations in the HE ORF sequence leading to the loss of the ORF and eventually the intron sequence is completely lost. This re-establishes the site for an orthologous HEG to reinvade this location. These observations were formulated into the HE lifecycle where mobile introns are neutral elements that, due to lack of selection, rapidly accumulate mutations, and degenerate. They can persist by outpacing drift at their sequence level by moving into empty target sites or invading new sites. The persistence within populations requires the opportunity for outcrossing, horizontal and vertical transfers, and cytoplasmic transmissions (such as heterokaryon formation). The pattern of rapid spread through a population followed by degeneration may be a common property of introns and other selfish genetic elements due to a lack of selection, however this balance can shift if introns provide benefits, such as acting as regulatory elements that can modulate gene expression, encoding essential proteins, or having been co-opted to provide essential functions. Intron conservation across some plant, metazoan, protozoan, and fungal organellar genomes would suggest that introns may not always be merely neutral elements. *Trans*-splicing introns might be a more permanent feature that allows fragmented genes to be assembled in trans at the RNA level. Introns encoding splicing factors (i.e., matK and matR) that can act in trans are also integral components of organellar gene expression.

The original concept of Goddard and Burt's (1999) HE lifecycle has also been applied to all types of mobile introns and has been modified as more organellar genomes have been characterized with regard to their intron contents. It is assumed that ribozyme-type introns originally existed as self-splicing introns that, in some instances, acquired ORFs. In the case of Gr I introns, HEs provide a means of more efficient mobility. In turn, HEs evolved maturase activity to promote efficient

splicing ensuring the composite element minimizes its impact to the host genome (Lambowitz et al., 1999). Composite introns as predicted by Goddard and Burt (1999) can be gained and lost and regained due to drift. However, there are other features of complex introns that are worth noting. Many intron sequences that encode proteins are fused to the upstream exons of their host genes. This has been referred to as 'core creep' where the intron ORF over time has incorporated upstream intronic sequences to fuse in-frame to the upstream exon (Edgell et al., 2011). This fusion would allow the intron-encoded protein to be more efficiently expressed, as it gains regulatory sequences of the host gene that optimize translation (Edgell et al., 2011; Guha et al., 2018).

The concept of core creep is also supported by the observation that there is a bias of mitochondrial introns towards being so-called phase 0 introns (i.e., they do not interrupt codons) (Zubaer et al., 2019, 2021). Intron insertion sites at phase 0 facilitate the fusion of intronic ORFs in-frame with the upstream exon. Composite introns can expand by gaining additional ORFs or intron modules (Guha et al., 2018), and there are instances of internal (non-fused) ORFs being expressed via AS that generates a product that fuses the intron ORF sequence to the upstream exon. This has been referred to splicing mediated core creep (Guha et al., 2018; Zubaer et al., 2021) and it demonstrates the complexity of the intron ORFs co-evolving with their intron host sequences and the host gene sequences for optimizing the expression of IEP and minimizing their impact on the host genomes. It has been suggested that the intron ORFs are essentially mutualistic towards their host introns and some IEPs may impact the persistence of other introns (reviewed in Zubaer et al., 2018) suggesting that drift is only one component that may explain the persistence and distribution of organellar introns.

Organellar genomes in plant and fungi are highly dynamic because of the loss and gain of mobile elements throughout evolution. Intron gain might occur due to homing (Gr I) or retrohoming (Gr II), and also due to HGT, leading to expanded

and inflated genomes. However, there are also instances of streamlining that can be due to drift or selection which might be lineage specific, where the cost of maintaining introns impact fitness. Among fungi, large-scale genome rearrangements are often accompanied by accumulation of intergenic and intronic sequences, resulting in the increase in mitogenome size (Mower et al., 2010; Ye et al., 2020). For example, the fungus *Endoconidiophora resinifera* boasts of an intron-rich mitochondrial genome due to large number of intron insertions in the *cox1* gene, partially justified by the probable benefit of introns as gene regulators contributing to their persistence within a population (Zubaer et al., 2018). Large-scale intron loss events were also recently reported among mitogenomes of ectomycorrhizal fungi belonging to the genus *Boletus*, which also helped explaining frequent gene position reversal throughout evolution (Li et al., 2021).

Over the course of evolution, several genetic elements have moved by horizontal transfer between chloroplast and mitochondrial genomes of heterokonts with recombinases likely promoting such gene exchange events. HGT between plants and fungal organelles was first demonstrated by the occurrence of a *cox1* Gr I intron in the vascular plant, *Peperomia polybotrya* (Vaughn et al., 1995). More recently, sequential HGT events have been noted for the ancestors of the orchid subfamily, Epidendroideae, which were obligate parasites on Basidiomycete fungi during early development. Unlike the previously documented HGT event between fungi and plants which involved transfer of a homing intron, this study documents the horizontal transfer of sequence comprising up to $\tilde{8}$ kbp involving at least a dozen fungal genes remnants, along with their intronic and intergenic regions. Overall, almost a third of the fungal genome was incorporated into the orchid genome, pointing at a genome-scale chimerism between a land plant mitogenome and a fungus (Sinn and Barrett, 2020). The dynamics of mitochondrial genome expansion and shrinkage are complex and can substantially vary among members of different genera or among species within the same genus. Extensive variation in content and size is seen among

plants and fungi, owing to dynamic gains and losses of repetitive noncoding DNA (intergenic spacers) and selfish genetic elements (introns and transposable elements) that have parasitized these genomes. Such gain and loss events are amplified in organisms with mutualistic lifestyles mainly guided by the need to eliminate potential redundancies and streamline the genome.

1.2.5 Complex Introns and Alternative Splicing Possibilities

The term 'twintron' was initially used to describe Gr II introns inserted within other Gr II or Gr III (derived from Gr II introns) introns; Gr III introns seem to have evolved by Gr II introns losing bits and pieces (Hong and Hallick, 1994). Later other variations were identified, such as Gr II in Gr I, Gr I within other Gr I, and lariat capping twin ribozyme introns (Hafez and Hausner, 2015; Rudan et al., 2018). The original definition implied that the internal components splice out first, allowing the "outer" components to be spliced together and achieve a splicing competent RNA fold. However, many different types of "nested" intron arrangements exist where splicing of the internal intron is not essential for the splicing of the external intron (Suzuki et al., 2013). It is still important for composite multiple intron-module arrangements to achieve splicing competent folds for its various components (Deng et al., 2016). In addition, intron and host genome encoded maturases or splicing factors must be involved in resolving transcripts with complex introns.

Twintrons were first described from *Euglena* chloroplast DNAs where the *psbF* gene which encodes the β-subunit of cytochrome b-559 contained a Gr II intron inserted within the structural domain V of another Gr II intron (Copertino and Hallick, 1991). A sampling of fungal mitochondrial *nad5* genes identified a complex Gr I intron with two sets of Gr I intron core sequences in the *Annulohypoxylon stygium*, while a second complex intron was observed in the *nad5* gene of *Cryphonectria parasitica,* where a Gr I intron core encodes a LAGLIDADG ORF that is interrupted by a Gr II intron module (Zubaer et al., 2019). A similar arrangement was previ-

ously characterized in the mitochondrial mS1247 intron in the *rns* gene (Hafez et al., 2013) and in the *cob*-506 intron (Guha et al., 2018). It has been speculated that the ORF-less Gr II intron, that in all cases was inserted in frame with regard to the LAGLIDADG ORF, could be a regulatory element that determines the expression of the IEP of the resident Gr I intron (Hepburn et al., 2021; Guha and Hausner, 2016). A variety of complex introns have been identified in fungal mitochondrial genomes (Guha et al., 2018; Zubaer et al., 2018 and 2019; Hafez et al., 2013; Guha et al., 2017). In some instances, complex introns composed of related introns may allow for various compatible RNA folding arrangements which can facilitate AS enhancing the expression of IEPs. A recent review showed that many different types of "nested" intron arrangements exist in both nuclear, prokaryotic, and organellar genomes, and there might be many mechanisms whereby they can be spliced out (Hafez and Hausner, 2015).

Zumkeller et al. (2020) reported five instances of twintrons in the mitogenome of lycophytes and hornworts, including an invasive "zombie" hypermobile Gr II intron (*cox1*i1149g2) in Lycopodiaceae that gave rise to two twintrons as an internal intron inserted into itself and into a newly identified succinate dehydrogenase *(sdh3)* intron (Takahara et al., 2002; Zumkeller et al., 2020). "Zombie" twintrons are composed of multiple intron modules but splice as one complete unit. A novel twintron configuration was reported in the mitochondrial genome in *Hypomyces aurantius*; in this fungus two Gr I introns were arranged side by side (tandem arrangement) within the *cox3* gene (Deng et al., 2016). With the rapid accumulation of organellar genome sequences for fungi more mobile introns along with complex intron configurations are bound to be encountered (Deng et al., 2018; Zubaer et al., 2021).

Twintron/complex introns offer new ribozyme scaffolds that could be engineered where the expression of intron-encoded ORFs can be regulated by the splicing of nested ribozyme type components (Hafez and Hausner, 2015). An example can be

found in the complex intron mS1247 from the thermophilic fungus, *Chaetomium thermophilum* var. *thermophilum*, in which the splicing of the internal Gr II intron reconstitutes a HE ORF of the external (resident) Gr I intron, facilitating its expression which, in turn, contributes to the mobility of the twintron to cognate intron-less *rns* genes (Guha and Hausner, 2014). It was also demonstrated that this complex intron could be expressed and spliced in *E. coli*; within *E. coli*, the expressions of the intron encoded protein could be manipulated by modulation the splicing of the internal Gr II intron module by manipulating Mg2+ concentrations in the media (Guha and Hausner, 2016). Essentially the internal intron module can be utilized as an on–off switch for the expression of intron-encoded heterologous proteins in *E. coli*.

The study of complex introns is interesting from an evolutionary point of view, with these configurations evolving independently multiple times involving different categories of introns, to generate complexity. These complex ribozymes my evolve into platform of alternative splicing or as regulatory elements that can modulate the expression of the host gene (Rudan et al., 2018; Chen et al., 2019). They may represent intermediate phases that favor the formation of new composite mobile elements and provide an avenue for introns that lack ORFs to achieve mobility by essentially receiving a free ride from the resident intron that encodes a functional HE.

1.3 Significance of Studying Fungal Mitogenomes, Introns and Encoded Proteins

Across the Mycota, mitogenomes show variability due to recombination events promoted by repeats and by the presence and activities of mobile elements such Gr I and Gr II introns and IEPs (Aguileta et al., 2014; Repar and Warnecke, 2017; Wu and Hao, 2019; Fonseca et al., 2021). Mitochondrial introns are gained and

lost and their impact on gene function and phenotypes is debatable (Goddard and Burt, 1999; Chatre and Ricchetti, 2014; Rudan et al., 2018). Organellar introns can be beneficial, as they encode proteins such as ribosomal proteins (rps3), aminotransferases, and N-acetyltransferases and *trans*-acting maturases (Hausner et al., 2006; Wai et al., 2019). In fungi, mitochondrial introns have been associated with fungicide resistance, a recent example being Gr I-D introns associated with the *cytochrome b (cob*) genes in the mitogenomes of 169 fungal species studied by Cinget and Bélanger (2020). Previously Grasso et al. (2006) have hypothesized that the presence of a *cob* Gr I mitochondrial intron blocked the mutation involved in the resistance against quinone outside inhibitors (QoI) fungicides (Grasso et al., 2006). Presence of Gr I introns in certain positions within the *cob* gene of fungi appeared to prevent mutations in the flanking exons in order to maintain sequences required for the P1 and P10 interactions that are required for intron splicing. This sets a constraint on mutations arising that are responsible for fungicide resistance. Overall, the correlation between this resistance phenotype and the Gr I intron is dependent on the presence of compatible homing sites in the mitogenomes and transient displacement of this intron (Cinget and Bélanger, 2020).

The presence of an ORF-less Gr II A1 intron inserted in the *rns* gene of the chestnut-blight fungus *Cryphonectria parasitica* appears to induce hypovirulence (Baidyaroy et al., 2011). This study showed that this intron spliced inefficiently resulting in a low production of mitochondrial ribosomes. In addition, it was demonstrated that the attenuated-virulence trait and the splicing-defective intron can be transferred asexually via hyphal contact from hypovirulent (intron) donor strains to virulent recipients. Hypovirulence results in the fungus infecting its host tree without causing detrimental consequences to the host. The attenuation of virulence is beneficial for the pathogen population as it prevents the extinction of the host species which, in turn, prevents the extinction of the pathogen.

Introns in general have been viewed as possible vehicles for regulating gene ex-

pression, as their removal can be a rate-limiting step for the expression of the genes that contain them (Rose, 2019). A study on yeast showed that the persistence of self-splicing mitochondrial introns is facilitated by an evolutionary lock-in, wherein the host genome has adapted to primordial invasion of introns in a way that subsequent intron loss could be deleterious (Rudan et al., 2018). The fitness of intron-less yeast strains was compared with the wild-type intron-rich strains and it was found that the strain without mitochondrial introns has altered mitochondrial morphology, gene expression, and metabolism impacting its growth and life span. Another example of introns serving as regulatory genetic elements, comes from the hexacoral mitochondrial Gr I introns associated with the *nad5* and *cox1* genes. Here, the introns appear to have been domesticated and gained novel host-specific functions beyond self-splicing (Johansen and Emblem, 2020).

Recently, Liu and Pyle (2024) have demonstrated the significance of Gr I introns that function as independent RNA processing units and splice with high efficiency in human pathogenic fungi such as *Candida albicans*. These introns have stably integrated within important housekeeping genes in the mitochondria, and contribute to the fitness of the systemic pathogen. Similarly, highly reactive Gr II intron splicing has been noted in human pathogenic fungi that function in mitochondrial regulatory network and host adaptation (Liu and Pyle, 2021). Such studies are paramount to the development of RNA-forward informatic tools for investigating human health and disease and exploring the potential of these robust self-splicing introns as anti-fungal drug targets.

It is interesting to note that if introns are indeed beneficial (Belfort, 2017; Rudan et al., 2018) for fine-tuning mitochondrial gene regulation, it is not based on specific introns, instead the mitogenome intron complement is composed of various introns (located at different sites) that are functionally redundant. The reliance on nuclear factors for organellar intron splicing impacts mitonuclear compatibilities (or incompatibilities) and potentially imposes reproductive barriers, thereby they could be

promoting speciation events (Dujon, 2020). There are still many speculations that need to be validated with regards to mitonuclear interactions and the associated "cross talk" that impact mitogenome sequence diversity and gene expression (Wu et al., 2022).

1.4 Research Aims and Objectives

The overall hypothesis for this research is that fungal mitogenome sizes correlate with the number of introns and large mitogenomes harbor many introns, including complex arrangements. Some complex introns can follow alternative splicing pathways depending on their structural RNA interactions and intron-encoded open reading frames.

The general aim of this work was to explore fungal mitochondrial genomes and transcriptomes to generate annotated genome maps and validate the annotations.

The three specific aims/objectives were:

1. To detect interesting genetic elements such as complex introns and plasmids, investigate intron insertion sites and gene synteny through comparative mitogenomics for the members of the Ophiostomatales.

2. To gain insights on the complex introns in the *cob* and *cox3* genes and their possible alternative splicing pathways in the mitochondrial transcriptome of two strains of *Ophiostoma ips*. Tto investigate the evolutionary position of fungi belonging to Ophiostomatales through phylogenetic studies.

To fulfill the above aims, the strategies were:

1. Generating mitogenome maps from selected fungal strains based on annotated Whole Generation Sequencing data

2. Aligning the RNA Sequencing reads of *O. ips* strains WIN(M) 1478 and 1480, to the annotated reference genomes to detect rare splice forms

3. Conducting phylogenetic analysis by using the concatenated dataset of amino acid sequences representing the core proteins

Chapter 2

General Materials and Methods

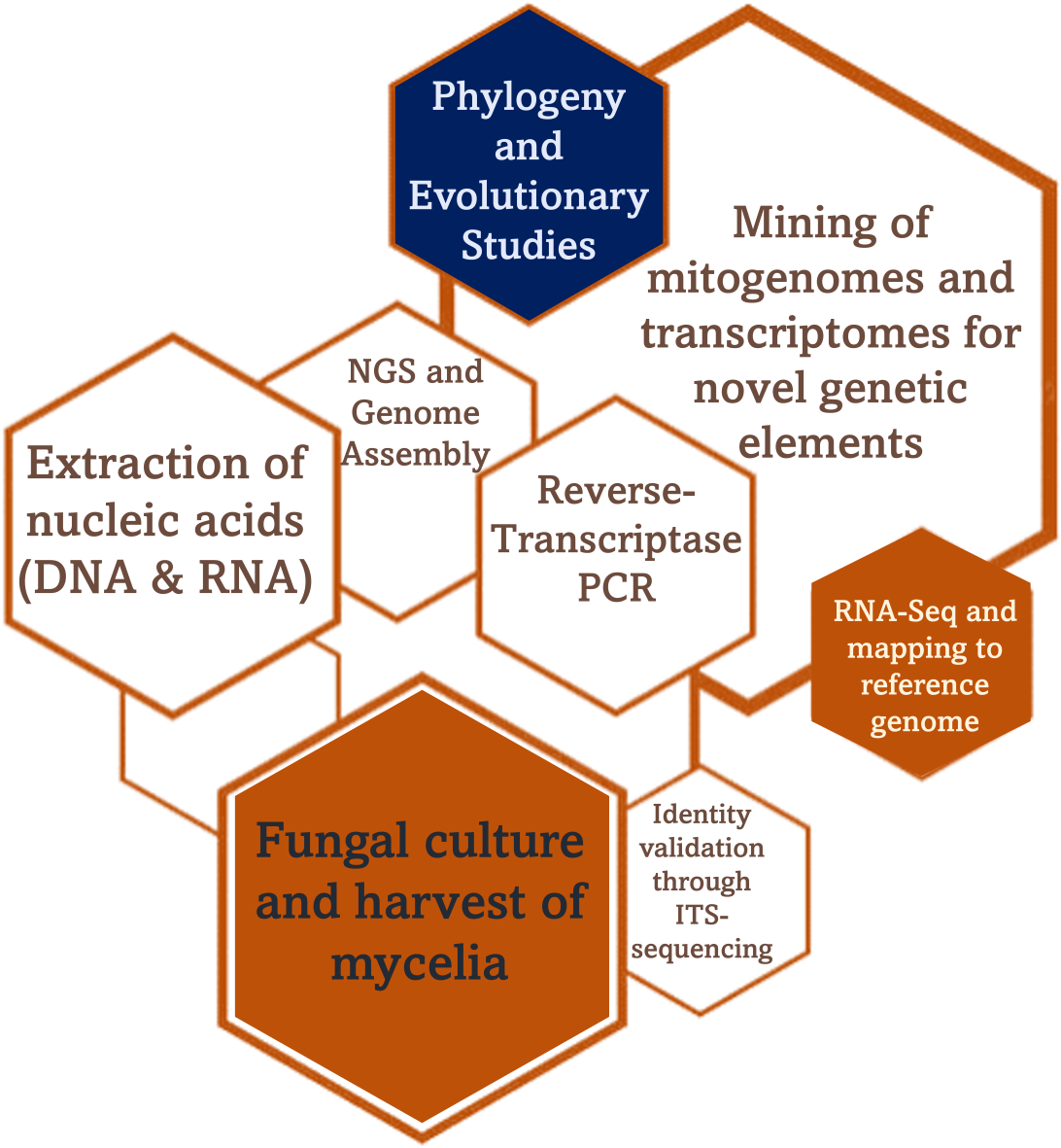

Figure 2.1: Flowchart of Methods

2.1 Nucleic Acid Extraction, Reagents, and Fungal Strains

All fungal strains were maintained in slants or media plates containing malt extract agar supplemented with yeast extract (for 1 L MEA Medium: Malt Extract 30 g, Yeast Extract 1 g, Agar 20 g).

Cultures were incubated in the dark at 20°C for up to 2 weeks. For the extraction of the nucleic acids, Erlenmeyer flasks containing 80 ml of MEA broth were each inoculated with agar blocks and incubated in the dark at 20°C for up to 10 days. Fungal mycelium was harvested from liquid media by vacuum filtration through a Whatman Grade 1 filter paper in a Büchner funnel. Mycelium was ground up in a pre-chilled mortar with a pestle and acid-washed sand and whole genome DNA for next generation sequencing was recovered and purified.

A brief protocol for harvesting of fungal mycelia and extraction of nucleic acids is given below:

2.1.1 Extraction of DNA

Fungi were grown on a solid medium and transferred to a liquid medium; mycelia were harvested by vacuum filtration and transferred to a Petri plate on ice temporary till grinding, and thereafter, to a mortar. The mycelia were mixed with acid-washed sand and isolation buffer (440 mM Sucrose, 10 mM Tris-HCl pH 8.0, 5 mM EDTA) in the ratio 1 g:1.5 g:2 ml. The mixture was ground with a pestle for 1 min or until the mixture became a slurry; an additional 5 ml isolation buffer was added, mixed, and transferred into a centrifuge tube.

The mixture was centrifuged (3000 rpm/15 min/4°C; Sorvall RC5B Plus SS-34 rotor); the supernatant was transferred into a new tube. After another centrifugation step (11,500 rpm/15 min/4°C; Sorvall RC5B Plus SS-34 rotor), the supernatant was discarded; an equal volume prewarmed (65°C) 2X CTAB buffer (2% CTAB, 100

mM Tris-HCl pH 8.0, 20 mM EDTA, 1.4 M NaCl) was added.

This incubation step could be for 1 h or over-night.

Incubation was carried out at 65°C with gentle mixing every 10-20 min (if incubating for 1 h). The tube was cooled at 15°C for 1 min; an equal volume of chloroform was added and mixed thoroughly. Centrifugation (3500 pm/10 min; Thermo IEC CENTRA CL-2 42609326) was carried out and 5 µl RNase A (100 mg/ml) was added before incubation at 65°C for 1 h (chloroform extraction was repeated until the aqueous layer was clear). 2.5-3 volumes 95% ethanol were added to the aqueous solution and mixed by gently inverting the tube and placed at -20°C for 1 h or overnight. Centrifugation was carried out at 3500 pm for 30 min (Thermo IEC CENTRA CL-2 42609326); the supernatant was decanted and 1 ml 70% ethanol is added. The final centrifugation step (3 500 rpm/10 min; Thermo IEC CENTRA CL-2 42609326) was carried out. Ethanol was completely removed by allowing drying and evaporation; the pellet was resuspended in 200 µl ddH20. The solution was transferred into a 1.5 ml microcentrifuge tube and stored at -20°C.

2.1.2 Extraction of RNA

This protocol is based on the RNeasy plant mini kit (Qiagen, Valencia, CA, US; catalog no. 74903 and 74904). Two strains of *Ophiostoma ips* WIN(M)1478 and 1480 were selected for RNA extraction and RNA Sequencing analysis (see Results and Discussion Chapter 4).

Cultures were grown in YPD (for 1 L YPD Medium: Yeast Extract 1 g, Peptone 1 g, Dextrose 3 g) broth for 7 to 10 days and mycelia were extracted using vacuum filtration. The harvested mycelia were put in Petri plates, and stored at -80°C till use.

A mortar was cooled by filling with liquid nitrogen; once cooled, the mycelia (from -80°C storage) were immediately added with additional liquid nitrogen (added twice in small amounts). Once dried, the mycelia was ground to powder. Once a

fine flour-like powder was obtained, the ground mycelia were scraped using a plastic inoculating loop and transferred using a steel spatula to an Eppendorf tube; Lysis mix (from the commerical kit) was added and vortexed to generate a lysate. The whole lysate was transferred to QIA shredder column and centrifuged for 2 min at 13,000 pm (Eppendorf centrifuge was used for all centrifugation steps in this protocol). The flow through was transferred (without touching cell pellet) to a new micro centrifuge tube. 96% chilled ethanol was added to the supernatant (at half volume of the supernatant) and immediately inverted and mixed vigorously. The sample in the microcentrifuge tube was added to the RNeasy spin column and centrifuged at 10,000 rpm for 15 secs (and put on hold for 30 sec). 700 µl of RW1 buffer was added to the RNeasy spin column and centrifuged for 15 s at 10,000 pm to wash the membrane; the flow through was discarded. 500 µl of RPE buffer was added to the RNeasy spin column and centrifuged for 15 s at 10,000 pm; the flow through was discarded. 500 µl of RPE buffer was added to the RNeasy spin column and centrifuged for 2 minutes at 10,000 rpm (to wash the membrane); the flow through was discarded. The column was put in a new microcentrifuge tube and centrifuged for 1 minute, drying the membrane. 50 µl RNase free water was added and centrifuged for 1 minute. The RNA was kept in 2 or more aliquots and frozen at -80°C .

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). The 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 Reverse-transcriptase PCR

RT-PCR protocol

SuperScript IV One-Step RT-PCR System protocol (Invitrogen, Carlsbad, CA, USA) was used for carrying out RT-PCR for *Ophiostoma ips* strains WIN(M) 1478 and 1480.

RT-PCR reaction mix

The components (Table 2.1) were combined in a 0.2 ml nuclease-free, PCR tube on ice and mixed gently to ensure that all the components are at the bottom of the amplification tube (centrifuged briefly when needed using an Eppendorf centrifuge).

*All components except the SuperScript IV RT Mix were used for no-RT control reactions

Reaction mixes were transferred to the preheated thermal cycler to start the RT-PCR program (Table 2.2).

Notes:

- All components, reaction mixes, and samples, were on ice
- The optimal conditions for RT-PCR depended on primer and target sequences
- The PCR extension time varied with the size of amplicon (appx. 30s/1 kb of amplicon)

  Reverse-transcriptase PCR was used as a preliminary strategy for detecting splicing products for the complex introns in *cob* and *cox3* genes in the mitogenomes of WIN(M) 1478 and 1480.

Table 2.1: 50 µl RT-PCR Mix

Component	**Volume**
2X Platinum SuperFi RT-PCR Master Mix	25 µl
Forward primer 10 µM	2.5 µl
Reverse primer 10 µM	2.5 µl
SuperScript IV RT Mix*	0.5 µl
Template RNA	2 µl
Nuclease-free water	17.5 µl

Table 2.2: Thermal Cycler Program for samples WIN(M) 1478 and 1480; *both RT steps

Step	**Temp**	**Time**	**Cycles**
RT	50°C	10 min	1*
RT inactivation/denaturation	98°C	2min	
Amplification	98°C/55°C/72°C	10sec/10sec/3min	35
Final extension	72°C	5min	1

Verification of cDNA was carried out using agarose gel electrophoresis on a 1% agarose gel.

Primers used for RT-PCR are tabulated in Table 2.3. Forward and Reverse primers were used in all possible combinations of matching forward and reverse primers to ensure consistency of results.

Table 2.3: *cob*490 and *cox3*-640 RT-PCR primers for WIN(M) 1478 and 1480

Primer name	**Orientation**	**Primer Sequence 5′-3′**
RTCOB-7F	Forward	GACATAGTCTAATTATATTTG
RTCOB-11Fgp2	Forward	GCGTAAGCGAAGATGTGG
RTCOB-6	Forward	CGCGTTATAGTAAATTATGC
RTCOB-8R	Reverse	CAAATATAATTAGACTATGTC
Ipscox3I2-F1	Forward	CAAGAAGTTGAATATGATC
Ipscox3I2-R1	Reverse	CAGTATGCAATAGCACCTTC
Ipscox3I2-F2	Forward	CCAGCAGTATGAGGAGGATTAG
Ipscox3I2-R2	Reverse	CCAGCAGTATGAGGAGGATTAG

Enzymatic Purification of PCR Products

PCR products were purified using ExoSAP-IT™ PCR Product Cleanup Reagent; Catalog number: 78200.200.UL (ThermoFisher Scientific). The ExoSAP-IT method utilizes two hydrolytic enzymes, Exonuclease I and Shrimp Alkaline Phosphatase, together in a specially formulated buffer, to remove unwanted dNTPs and primers from PCR products. Exonuclease I removes residual single-stranded primers and any extraneous single-stranded DNA produced in the PCR. Shrimp Alkaline Phosphatase removes the remaining dNTPs from the PCR mixture which would interfere with the subsequent reactions.

ExoSAP-IT cleanup reagent was added directly to the PCR product and incubated at 37°C for 15 minutes. After PCR treatment, ExoSAP-IT reagent was inactivated simply by heating to 80°C for 15 minutes.

2.2.1 Additional notes on PCR for fungal identity verification

Preliminary verification of the identity of fungal species was carried out using PCR and sequencing using primers against the internal transcribed spacer (ITS) region of the ribosomal RNA genes. This region with conserved sequences flanking a highly variable region has the highest probability of successful identification for the broadest range of fungi, with the most clearly defined barcode gap between inter- and

Table 2.4: Primers for amplification of ITS-region; SS/SSU= Small Subunit; LS/LSU= Large Subunit

Primer	**Orientation**	**Primer Sequence 5′-3′**
SSU Z	Forward	ATAACAGGTCTGTGATG
LSU 4	Reverse	TTGTGCGCTATCGGTCTC
SS3	Forward	GTCGTAACAAGGTCTCCG
LS2	Reverse	GATATGCTTAAGTTCAGCG

intraspecific variation (Schoch et al., 2012) . A list of primers used for ITS amplification is given in Table 2.4.

2.3 Next Generation Sequencing and Genome Assembly

The mtDNA preparations were sent to MicrobesNG (Units 1-2 First Floor, The Bio-Hub, Birmingham Research Park, 97 Vincent Drive, Birmingham, B15 2SQ, UK) for Illumina sequencing. Briefly, with regards to DNA preparation for MicrobesNG, quantification and quality assessment of the crude mtDNA extract were performed by Nanodrop 2000c UV-Vis spectrophotometer and agarose gel electrophoresis analyses. The extract was aliquoted into a 1.5 mL polypropylene conical microcentrifuge tube (labeled based on instructions provided by MicrobesNG) and diluted to 30 ng/µL DNA in a final volume of 100 µL with nuclease-free water.

MicrobesNG returned the sequencing reads, the adapters of which were trimmed using Trimmomatic version 0.30 (Bolger et al., 2014) using sliding window with a quality cutoff of 15. A draft assembly was also provided, assembled using SPAdes version 3.7 (Bankevich et al., 2012); and annotated contigs, annotations generated by Prokka version 1.14.3 (Seemann, 2014). In order to enhance the assembly and recover as many mitochondrial genome-derived reads, we re-assembled the data. The reads from MicrobesNG were assessed using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The as-

sembly from MicrobesNG was then evaluated by reassembling the reads using two approaches, SPAdes and A5-miseq (Tritt et al., 2012; Coil et al., 2015). The SPAdes assembly was generated using SPAdes version 3.14.0, setting the "–careful" option, which was used to obtain an assembly minimizing indels and mismatches. Using this option, a post-processing MismatchCorrector step is employed, that relies on the BWA mapper. The modified A5-miseq pipeline automates adapter trimming, error correction, quality filtering, scaffold generation, and the detection of misassemblies. BLASTn (Altschul et al., 1990; Johnson et al., 2008; Camacho et al., 2009) was used to search for contigs/scaffolds corresponding to the mtDNA in all assemblies.

In some cases, the program Bandage (Wick et al., 2015) was used to examine the assembly graph files generated from SPAdes to aid in the recovery of potential mitogenome contigs. In order to more efficiently recover mtDNA derived reads and mitochondrial genome assemblies the program GetOrganelle (Jin et al., 2020) v1.7.5, with the organelle type set to fungus mitogenome (i.e., -F fungus_mt) was applied to the Illumina sequencing reads. In some instances, once a contig or a set of contigs were recovered, gaps and/or regions with low coverage were completed/validated by Sanger sequencing PCR products based on primers designed to bind to the 5 prime and 3 prime ends of the recovered linear contigs (or regions of low coverage).

The mtDNA contigs were annotated and refined using the MFannot program (Lang et al., 2023; setting "Genetic Code" to (4 - mold). Sequence alignments were generated for all protein-coding genes with *Tolypocladium inflatum* serving as a reference genome for naming introns according to the proposed nomenclature by Zhang and Zhang, 2019. The rDNA sequences were compared with those of *Escherichia coli* for intron annotations (naming) according to Johansen and Haugen (2001). Gene annotations were refined with Artemis (Rutherford et al., 2000) and the mtDNAs were visualized using Circos (Krzywinski et al., 2009). Circos was set up with the appropriate coordinates to highlight exon/intron configurations for

conserved protein-coding genes, noncoding genes, and to plot GC content. The GC plot was generated using a window size of 100 bp and a step size of 20 bp.

Predictions of tRNAs were also performed with tRNAscan-SE 2.0 (Chan and Lowe, 2019). The RNAweasel program (Eddy and Durbin, 1994; Gautheret and Lambert, 2001; Lang et al., 2007; Smith et al., 1994) was used to predict intron types/categories and intron core structural elements. Secondary structural elements of the Gr I introns including the P3, P4, P6, P7, and P8 helices, were predicted using RNAweasel. Other base-pairing interactions, namely the P1, P2, P5, P9 and P10 helices, were predicted in part through Mfold (Zuker, 2003) and UNAFold (Markham and Zuker, 2008), and manually by comparing the intron sequences against reference structures and alignments generated by Michel and Westhof (1990) as well as those provided within the Group I intron Sequence and Structure Database (GISSD) (Zhou et al., 2008). The final intron folds were manually drawn following the conventions proposed by Cech et al. (1994) using InkScape (https://inkscape.org/). Similarly, Gr II intron sequences were analyzed with RNAweasel and Mfold (using constraints based on previously published models) and RNA folds were generated based on multiple alignment comparative analysis and reference folds provided by Michel and Ferat (1995), Toor et al. (2001) and the group II intron database (`http://webapps2.ucalgary.ca/~groupii/`; Candales et al., 2012).

2.4 RNA Sequencing for WIN(M) 1478 and 1480

RNA was prepared according to specifications in the User Guide for Illumina Sequencing Technologies by Genome Quebec and submitted for preparing total RNA libraries without the following enrichment: poly(A) selection, depletion of rRNA and tRNA and size selection. One additional block was requested for deep sequencing on Illumina Novaseq6000 (PE100 25M reads/lane + additional block of reading).

For quality control, obtained paired end reads were run through FastQC for

quality control and Trimmomatic (Bolger, 2014) to remove low quality bases and adapters. Additionally Rcorrector (Song and Florea, 2015; courtesy of F. Lang, University of Montreal) was used to error-correct Illumina reads (i.e., correct for RT artifacts and random sequencing errors).

The reference genomes for WIN(M) 1478 and 1480 obtained through the annotation of NGS data were utilized for reference-based mapping, alongside *de novo* assembly using rnaSPAdes (Bushmanova et al., 2019). The reads were aligned to the reference genome using (RNA)STAR (Dobin et al., 2013). The output from STAR is a BAM file which was visualized on the Integrative Genomics Viewer, IGV (Robinson et al., 2011). IGV also provided confidence in the prediction of exonic and intronic regions (i.e., gene structure prediction from MFannot). On IGV, soft-clipped bases (bases in the $5'$ and $3'$ ends of the read that are not part of the alignment with the reference sequence) were viewed to facilitate detection of structural variations that might arise due to splicing (such as alternative splicing events).

2.5 Phylogenetic Analysis

2.5.1 Collection of Data

The fungal strains of species assigned to the Ophiostomatales and Pezizales used in this research have been tabulated in Tables 2.6 and 2.7.

The mitochondrial protein sequences were obtained from NCBI (`https://www.ncbi.nlm.nih.gov`) and MitoFun (`http://mitofun.biol.uoa.gr/`) sites, selecting 'organelle' and specifying 'fungi' and 'mitochondria' type in the search criteria. Thirteen different mitochondrial protein coding (amino acid) sequences were chosen and were concatenated. Additionally, nuclear marker sequences (ITS and beta-tubulin) were collected from NCBI (GenBank) and from whole genome assemblies for the ITS region (GenBank: OR146620 to OR146622) and partial beta-tubulin gene sequences (GenBank: OR146617-OR146619) to generate a potential species tree for

Table 2.5: Fungal Species and Strains. *Notes*:WIN(M) 1478(=80) have identical genomes; **O. adjuncti* ATCC34942 deposited by Davidson RW (1978) Mycologia 70: 35-40; PRJNA841745 refers to the Bioproject "Genome sequencing and assembly of dutch elm disease fungi and related species " submitted by Laval University; TB= Thunder Bay; WA= Washington; LA= Louisiana; BC= British Columbia; ON= Ontario

Identifier	**Genus, Species**	**Size (bp)**	**Introns**	**Source**
WIN(M) 1478	*Ophiostoma ips*	113671	46	TB, Canada
WIN(M) 1481	*O. ips*	111826	45	TB, Canada
WIN(M) 1486	*O. ips*	80045	32	Canada
WIN(M) 1487	*O. ips*	85284	33	Oregon, USA
WIN(M) 1488	*O. ips*	106900	45	WA, USA
WIN(M) 1001	*O. ips*	104698	46	New Zealand
WIN(M) 1515	ips-like/ complex	39957	12	New Zealand
WIN(M) 502	*O. adjuncti*	112641	45	ATCC34942*
SRR19396180	*O. ips*	96039	39	PRJNA841745*
SRR19396179	*O. montium*	76883	39	PRJNA841745*
NTMB01000349.1	*O. ips*	97849	42	LA, USA
WIN(M) 1376	*Leptographium* sp.	81823	27	New Zealand
WIN(M) 809	*L. aureum*	104547	37	BC, Canada
WIN(M) 1600	*Grosmannia fruticetum*	63821	25	ON, Canada
LJH211	*Urnula craterium*	43967	4	ON, Canada

Table 2.6: Additional Fungal Members of the Ophiostomatales for Phylogenetic Studies and Pearson Correlation.

Identifier	**Genus, Species**	**Size (bp)**	**Introns**
WIN(M) 495	*Ophiostoma minus*	91847	38
CBS 374.67	*Ophiostoma himal-ulmi*	111712	51
SRR10139932	*Ophiostoma novo-ulmi*subspecies *americana*	56451	24
MG020143.1	*Ophiostoma novo-ulmi* subspecies *novo-ulmi*	64848	27
SRR10139930	*Ophiostoma ulmi sensu lato*	59003	24
SRR10139873	*Ophiostoma ulmi*	77596	36
MW122509.1	*Ophiostoma minus*	91847	38
SRR869560	*Ophiostoma piceae*	33599	6
MW122508.1	*Ophiostoma piliferum*	69966	29
PCDN01000199.1	*Ceratocystiopsis brevicomis*	90376	38
LZPB01000172.1	*Ceratocystiopsis minuta*	39800	13
WIN(M) 51	*Ceratocystiopsis palidobrunnea*	29022	4
LDEF01000080.1	*Leptographium lundbergii*	101879	36
PCDK01000036.1	*Grosmannia penicillata*	150891	64
NC020430	*Ceratocystis cacaofunesta*	103147	37
KY644696.1	*Esteya vermicola*	46507	18
PCDD01000156.1	*Raffaelea sulphurea*	128408	36
PCDH01000124.1	*Raffaelea arxii*	95042	43
PCDJ01000011.1	*Raffaelea albimanens*	137049	49
PCDI01000067.1	*Raffaelea ambrosiae*	89461	35
PCDE01000018.1	*Raffaelea quercivora*	152890	35
NIPS01000008.1	*Raffaelea quercus-mongolicae*	106194	40
PCDF01000414.1	*Raffaelea* sp.	23830	1
CM003773.1	*Sporothrix pallida*	35458	7
AWTV01000012.1	*Sporothrix brasiliensis*	35826	9
LVYW01000010.1	*Sporothrix globosa*	26671	1
AB568600.1	*Sporothrix schenckii*	26095	1
PCDL01000017.1	*Fragosphaeria purpurea*	57056	5
NTMA01000166.1	*Hawksworthiomyces lignivorus*	27092	2
LLKO01000061.1	*Graphilbum fragrans*	25567	2

members of the *Leptographium sensu lato*.

2.5.2 Alignment of Sequences

In general, multiple concatenated protein sequence or nuclear marker sequence datasets were aligned using MAFFT (Multiple Alignment of Fast Fourier Transform) (Katoh and Standley, 2013) or MUSCLE program and the alignments were observed for features and conservation in AliView program (Larsson, 2014).

2.5.3 Phylogenetic Analysis of Mitochondrial Protein Coding Regions

For sequences of the fungal members assigned to the Ophiostomatales, the manually adjusted alignments containing 56 mitogenomes (for chapter 3) and 65 mitogenomes (for chapter 4) were analyzed with MrBayes 3.2.7a (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012) for inferring phylogenetic trees. A fixed-rate amino acid substitution model was estimated using the mixed model function implemented in MrBayes (MB). Rate variation among sites was modeled with a combination of the invariable sites model and gamma model (i.e., rates from a gamma distribution). Other parameters were left at their default values (i.e., a uniform distribution between 0.0 and 1.0 and an exponential distribution with a mean value of 1.0, respectively). The analysis was performed with 2,000,000 generations with a sampling frequency of 1,000. The cpREV+F+I was estimated with the highest probability. The first set of 25% of sampled trees was discarded (burn-in) and the remaining trees were used to construct the 50% majority rule consensus tree.

The aligned data set was also analyzed with programs contained within MEGA XI (Tamura et al., 2021): Maximum Likelihood (ML). For ML, the LG model (plus I and F) was applied, and 1,000 bootstrap replicas were analyzed to assess branch support values. The output of the phylogenetic analyses (i.e.,tree files) was visual-

ized with the FigTREE program 1.4.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and members of the Eurotiales were selected as the outgroup for the analysis.

2.5.4 Phylogenetic Analysis of Nuclear Markers: ITS and Beta-tubulin

For members of the *Leptographium sensu lato* (chapter 3), aligned sequences were used for generating phylogenetic trees with MEGA XI (ML, model T92 + G) and MrBayes (MB, setting mixed model converging with GTR+I+G, 1,000,000 generations, sample frequency at 100, and burn-in value at 30 percent).

Chapter 3

Exploring the Genus *Leptographium sensu lato*

3.1 The mitogenomes of *Leptographium aureum*, *Leptographium* sp., and *Grosmannia fruticeta*: expansion by introns

The work presented in this chapter has been published.

Mukhopadhyay, J., Wai, A., and Hausner, G. (2023) The mitogenomes of *Leptographium aureum*, *Leptographium* sp., and *Grosmannia fruticeta*: expansion by introns. *Frontiers in Microbiology*. 14: 1240407. doi: 10.3389/fmicb.2023.1240407.

JM, AW, and GH contributed towards the design of the project, and worked on the manuscript. JM and AW took the lead with regards to assembling the datasets and the final analysis of the data. JM prepared the first version of the manuscript. GH assembled the final version of the manuscript. All authors contributed to the article and approved the submitted version.

3.1.1 Abstract

Many members of the Ophiostomatales are of economic importance as they are bark-beetle associates and causative agents for blue stain on timber and in some instances contribute towards tree mortality. The taxonomy of these fungi has been challenging due to the convergent evolution of many traits associated with insect dispersal and a limited number of morphological characters that happen to be highly pleomorphic. This study examines the mitochondrial genomes for three members of *Leptographium sensu lato* [*Leptographium aureum* (also known as *Grosmannia aurea*), *Grosmannia fruticeta* (also known as *Leptographium fruticetum*), and *Leptographium* sp. WIN(M) 1376)] through Illumina sequencing combined with gene and intron annotations and phylogenetic analysis. Sequence analysis showed that gene content and gene synteny are conserved but mitochondrial genome sizes were variable: *G. fruticeta* at 63,821 bp, *Leptographium* sp. WIN(M) 1376 at 81,823 bp and *L. aureum* at 104,547 bp. The variation in size is due to the number of introns and intron-associated open reading frames. Phylogenetic analysis of currently available mitochondrial genomes for members of the Ophiostomatales supports currently accepted generic arrangements within this order and specifically supports the separation of members with Leptographium-like conidiophores into two genera, with *L. aureum* grouping with *Leptographium* and *G. fruticeta* aligning with *Grosmannia*. Overall, mitochondrial genomes are promising sequences for resolving evolutionary relationships within the Ophiostomatales.

3.1.2 Introduction

Members of the Ophiostomatales are usually characterized by producing ascocarps with asci that are randomly produced at the base of the ascocarp, and these asci deliquesce with the ascospores being released as a sticky droplet at the tip of the perithecial necks. In addition, many members produce slimy/sticky conidia on long-stalked conidiophores. These are morphological features that facilitate the dispersal

of their spores by arthropods (Six and Wingfield, 2011). Historically, the genera *Leptographium* and *Grosmannia* have been assigned to the Ophiostomatales with *Leptographium* accommodating anamorphic (mitotic) members and *Grosmannia* housing those species that have a *Leptographium* anamorph plus a sexual state (reviewed in Jacobs and Wingfield, 2001, 2013; Zipfel et al., 2006). With the introduction of the single name nomenclature as proposed by Hawksworth (2011), *Leptographium*, as the older generic name, should have priority. However, the position of some lineages of fungi with Leptographium-like anamorphs has not been resolved within the Ophiostomatales (Zipfel et al., 2006; De Beer and Wingfield, 2013; Jacobs and Wingfield, 2013; Jankowiak et al., 2018; De Beer et al., 2022). Recent studies showed that there are at least two clades (with the corresponding type specimens: *Leptographium lundbergii* and *Grosmannia penicillata*) that accommodate members of the Ophiostomatales with Leptographium-like conidial states, validating the genera *Leptographium* and *Grosmannia* (De Beer and Wingfield, 2013; De Beer et al., 2022).

Leptographium aureum (Rob.-Jeffr. & R.W. Davidson) M.J. Wingf. was isolated from bark beetle-infested, blue-stained pine (Robinson-Jeffrey and Davidson, 1968). In recent surveys, they have been isolated from the aggressive *Dendroctonus ponderosae* (Mountain pine beetle, MPB in western North America; McAllister et al., 2018) and less aggressive *D. murrayanae* (lodgepole pine beetle) beetle in British Columbia Canada (Six et al., 2011). This species was formerly designated as *Europhium aureum*, *Ophiostoma aureum*, and *Grosmannia aureum*, and more recently placed within *Leptographium* (reviewed in Zipfel et al., 2006, and De Beer et al., 2022), demonstrating the complexity and volatility of the taxonomy of this group of fungi. Most members of *Leptographium*/*Grosmannia* are potential agents of blue stain on timber and therefore of economic concerns to the forestry industry (Uzunovic and Byrne, 2013). Some *Leptographium*/*Grosmannia* species are associated with tree diseases (Harrington, 1988; Eckhardt et al., 2004; Jacobs et al., 2004;

Hausner et al., 2005) and there is always the potential of moving blue stain fungi into new regions due to the export of infected lumber/timber products and/or the movement/migration of their bark beetle vectors (Jacobs et al., 2004; Hausner et al., 2005; Humble and Allen, 2006). *Grosmannia fruticeta* (Alamouti, J.J. Kim & C. Breuil) M.L. Yin, Z.W. De Beer & M.J. Wingfield has been isolated from *Ips perturbatus* (northern spruce engraver bark beetle) in Northern British Columbia and the Yukon (Massoumi Alamouti et al., 2006, 2007). Frequently exotic fungi are cryptic in the early stages of their invasion and due to morphologically similar appearing native species they are hard to be identified and thus fail to be detected (Santini and Migliorini, 2022). Introduced fungal pathogens can have a severe ecological and economic impact. Therefore, it is important to monitor the movement of native and exotic fungi (Trollip et al., 2021). Biosecurity of forestry resources requires the development of accurate identification strategies. There is considerable interest to generate molecular data that can be used to identify potential plant pathogens or fungi of economic concerns using genomic approaches (Aylward et al., 2017; Trollip et al., 2021, 2022). Kulik et al. (2020, 2021) argued that fungal mitogenomes contain variable regions that could provide a source for molecular markers suitable for fungal identification.

Mitogenomes for the filamentous members of the Ascomycota show great variation in size although encoding a similar set of core genes: *rnl* and *rns* (large and small subunit RNAs involved in protein translation), *cob* and *cox1–3* (coding for components of the respiratory chain complexes), *atp6*, *atp8*, *atp9*, *nad1–6* and *nad4*L (coding for NADH dehydrogenase subunits) and a set of tRNA genes. Sometimes, the ribosomal protein RPS3 (*rps3*; Hausner, 2003; Wai et al., 2019) can be encoded within an *rnl* intron or be found as a free-standing gene. Fungal mitogenome architecture is variable as the results of various recombination events promoted by repeats and by the presence and activities of mobile elements such group I and group II introns and intron-encoded proteins (IEPs) such as homing

endonucleases (HEs) and reverse transcriptases (RTs) that facilitate intron mobility and mitochondrial DNA (mtDNA) architecture (Hausner, 2012; Aguileta et al., 2014; Wu and Hao, 2014, 2019; Franco et al., 2017; Repar and Warnecke, 2017; Kulik et al., 2020; Tan et al., 2022). Beyond phylogenetic applications, fungal mitochondrial genomes harbor genetic elements (ribozymes, complex introns (i.e., potentially co-operating ribozymes), and intron-encoded proteins such as endonucleases with novel cutting sites, reverse transcriptases) that have applications in biotechnology as genome editing tools and/or regulatory switches to control gene expression (Takeuchi et al., 2011; Guha et al., 2017; Belfort and Lambowitz, 2019). There is also considerable interest in developing antifungal compounds that target group I introns (Zhang et al., 2000; Malbert et al., 2023).

Currently, only two mitogenomes (mtDNA) have been described for the group *Leptographium/Grosmannia*, these are for *Leptographium lundbergii* and *Grosmannia penicillata* (Zubaer et al., 2021). In this study, we characterize the mitogenomes for the following strains: *L. aureum* (WIN(M) 809), *Leptographium* sp. (WIN(M) 1376), and *Grosmannia fruticeta* (WIN(M) 1600). This work is part of our ongoing effort to study the evolution of mitogenomes for members of the Ophiostomatales (Abboud et al., 2018; Zubaer et al., 2018, 2021; Wai and Hausner, 2021, 2022) with the potential of gaining more insight into the evolution and systematics of these fungi.

3.1.3 Materials and Methods

Source of culture, culturing methods, and extraction of nucleic acids

Cultures of *Leptographium aureum* CBS 438.69, ex-Type, (CBS-KNAW culture collection, Uppsalalaan 8, 3,584 CT, Utrecht, Netherlands); = WIN(M) 809; = UAMH 12546 (UAMH = Centre for Global Microfungal Biodiversity, University of Toronto, 223 College St., Toronto ON, Canada M5T 1R4), *Leptographium* sp. WIN(M) 1376 = J.R. 88-194A; = UAMH 12547, and *Grosmannia fruticeta* WIN(M) 1600 (=

UAMH 12545) were maintained on malt extract agar (MEA; supplemented with yeast extract, 30 g/L malt extract, 1 g/L yeast extract, 20 g/L agar) slants and agar plates (containing approximately 40 mL MEA) and incubated in the dark at 20°C for up to 2 weeks. For the extraction of nucleic acids, three-250 mL Erlenmeyer flasks containing 80 mL of PYG + ME broth (1 g/L peptone, 1 g/L yeast extract, 2 g/L D-glucose, and 3 g/L malt extract) were each inoculated with ten agar blocks (2 mm × 2 mm × 1 mm) and incubated in the dark at 20°C for up to 10 days. Fungal mycelium was harvested from liquid media by vacuum filtration through a Whatman® Grade 1 filter paper in a Büchner funnel. Mycelium was ground up in a pre-chilled mortar with a pestle and acid-washed sand and the DNA for next generation sequencing was recovered and purified as previously described (Wai and Hausner, 2021). DNA samples (30 ng/µL DNA in a final volume of 100 µL) were sent to MicrobesNG (Units 1–2 First Floor, The BioHub, Birmingham Research Park, 97 Vincent Drive, Birmingham, B15 2SQ, United Kingdom) for Illumina sequencing.

Assembly, analyses, and annotation of next generation sequencing data

Initial analyses of the Illumina reads were performed on the online server GALAXY (Afgan et al., 2018), specifically, the data was uploaded onto Galaxy Europe (usegalaxy.eu). The reads from MicrobesNG were initially assessed using FastQC v0.11.9.1 Assemblies were generated by using two programs: SPAdes vs. 3.14.0 (setting the "–careful" option and assembly graph option) and the A5-miseq pipeline (with "-end" option set to "5"; Tritt et al., 2012; Coil et al., 2015). NCBI BLAST+ blastn (Camacho et al., 2009) was used to search for sequences of interest (rDNA internal transcribed spacer regions, ITS; and beta-tubulin sequences) including contigs/scaffolds corresponding to fungal mtDNAs in all assemblies. In addition to GALAXY, the program Bandage (Wick et al., 2015) was used to examine the assembly graph files generated from SPAdes to aid in the recovery of potential mitogenome contigs.

To more efficiently recover mtDNA derived reads and mitochondrial genome assemblies, the program GetOrganelle (Jin et al., 2020) v1.7.5, with the organelle type set to fungus mitogenome (i.e., -F fungus_mt) was applied to the Illumina sequencing reads. In some instances, for *L. aureum*, a set of contigs were recovered, and gaps and/or regions with low coverage were completed/validated by Sanger sequencing PCR products based on primers designed manually (custom designed primers to accommodate for AT-richness of sequences) to bind to the 5′ and 3′ ends of the recovered linear contigs (or regions of low coverage; Supplementary Table 1).

The mtDNA contigs were annotated using the MFannot program (Lang et al., 2023; setting "Genetic Code" to 4)2 and RNAweasel3; setting to predict tRNAs, and group I (Gr I) and group II (Gr II) introns. Predictions of tRNAs were also performed with tRNAscan-SE 2.0 (Chan and Lowe, 2019). Annotations of protein-coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1–3*, *nad1–6*, *nad4L*, *rps3*) and nonstructural genes (i.e., *rnl* and *rns*, and the tRNAs), were verified by comparative sequence analysis from data obtained from GenBank (Benson et al., 2013). Sequence alignments were generated for all protein coding genes with *Tolypocladium inflatum* serving as a reference genome for naming introns according to the proposed nomenclature by Zhang and Zhang (2019). The rDNA sequences were compared with those of *E. coli* regarding intron annotations (naming) according to Johansen and Haugen (2001). Gene annotations were refined with Artemis (Rutherford et al., 2000) and the mtDNAs were visualized using Circos (Krzywinski et al., 2009). Circos was set up with the appropriate coordinates to highlight exon/intron configurations for conserved protein-coding genes, nonstructural genes, and GC content. The GC plot was generated using a window size of 100 bp and a step size of 20 bp.

The final annotated versions of the mtDNAs characterized in this study have been deposited in GenBank (GenBank accession numbers as follows: for strain WIN(M) 809: OQ851464; for strain WIN(M) 1600: OQ851465; and for strain WIN(M) 1376: OQ851466).

Phylogenetic analysis of mitochondrial protein coding regions

For the mitochondrial genomes a dataset was generated composed of 13 concatenated amino acid sequences, derived from the following protein-coding genes: *atp6*, *atp8*, *cob*, *cox1–3*, *nad1–6*, *nad4L*, and data were aligned using MAFFT version 7 (Katoh and Standley, 2013). Some members of the Ophiostomatales do not encode *atp9* within their mitogenomes and so it was not included in the dataset (Zubaer et al., 2021). Additional mitogenomes were obtained from GenBank and the Sequence Read Archive (SRA; Shumway et al., 2010; Leinonen et al., 2011). For more information on the collection, processing, and analyses of the additional mitogenomes, see Zubaer et al. (2021) and Wai and Hausner (2021). In total, 56 mitogenomes were included in the analysis and the tree was visualized and edited using FigTree version 1.4.4.4 The root was placed on the branch that splits the Eurotiales and Sordariomycetes and the outgroup consisted of two Eurotiales, *Aspergillus fumigatus* and *Penicillium digitatum*.

The alignment was manually adjusted with AliView version 1.25 (Larsson, 2014) and analyzed with MrBayes 3.2.7a (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012) for inferring phylogenetic trees. A fixed-rate amino acid substitution model was estimated using the model jumping (i.e., mixed model) function implemented in MrBayes (MB). Rate variation among sites was modeled with a combination of the invariable sites model and gamma model (i.e., rates from a gamma distribution). Other parameters were left at their default values (i.e., a uniform distribution between 0.0 and 1.0 and an exponential distribution with a mean value of 1.0, respectively). The analysis was performed with 2,000,000 generations with a sampling frequency of 1,000. The cpREV model was estimated with the highest probability. The first set of 25% of sampled trees was discarded (burn-in) and the remaining trees were used to construct the 50% majority rule consensus tree. The aligned data set was also analyzed with programs contained within MEGA XI (Tamura et al., 2021): Maximum Likelihood (ML). For ML, the LG model (plus I and F) was

applied, and 1,000 bootstrap replicas were analyzed to assess branch support values.

Phylogenetic analysis of nuclear markers: ITS and Beta-tubulin

Sequences were collected from NCBI (GenBank) and from whole genome assemblies for the internal transcribed spacer (ITS) region (GenBank: OR146620 to OR146622) and partial beta-tubulin (βT) gene sequences (GenBank: OR146617-OR146619) to generate a potential species tree for members of *Leptographium*. The data were aligned with MAFFT, and phylogenetic trees were generated with MEGA XI (ML, model T92 + G) and MrBayes (MB, setting mixed model converging with GTR + I + G, 1,000,000 generations, sample frequency at 100, and burnin value at 30%).

3.1.4 Results

The mitogenomes: size variation and gene content

The mitogenomes of *L. aureum* WIN(M) 809, *Leptographium* sp. WIN(M) 1376, and *Grosmannia fruticeta* WIN(M) 1600, are represented as circular molecules and are illustrated in figs. 3.1–3.3, respectively.

These vary in size with *L. aureum* at 104,547 bp, *L.* sp. WIN(M) 1376 at 81,823 bp, and *G. fruticeta* at 63,821 bp. Based on the program GetOrganelle v. 1.7.5 the coverage for the three mitogenomes were as follows: 2,491 fold for *L. aureum*, 1,555 fold for WIN(M) 1376, and 1,409 fold for *G. fruticeta*. All three genomes encode the ribosomal RNA genes *rns* and *rnl*, the mitochondrial core set of protein-coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1–3*, *nad1–6*, *nad4L*, and *rps3*). Like other members of the Ophiostomatales (and members of the Sordariomycetes) the ribosomal protein RPS3 is encoded within a Gr I intron embedded within the *rnl* gene (reviewed in Wai et al., 2019). The presence of *atp9* was noted for all three genomes and these also encode a set of tRNAs that cover all 20 standard amino acids. The mitogenomes for *L. aureum* and *L.* sp. WIN(M) 1376 encode 26 tRNA genes and the *G. fruticeta* genome encodes 25 tRNAs genes. The gene order and orientation

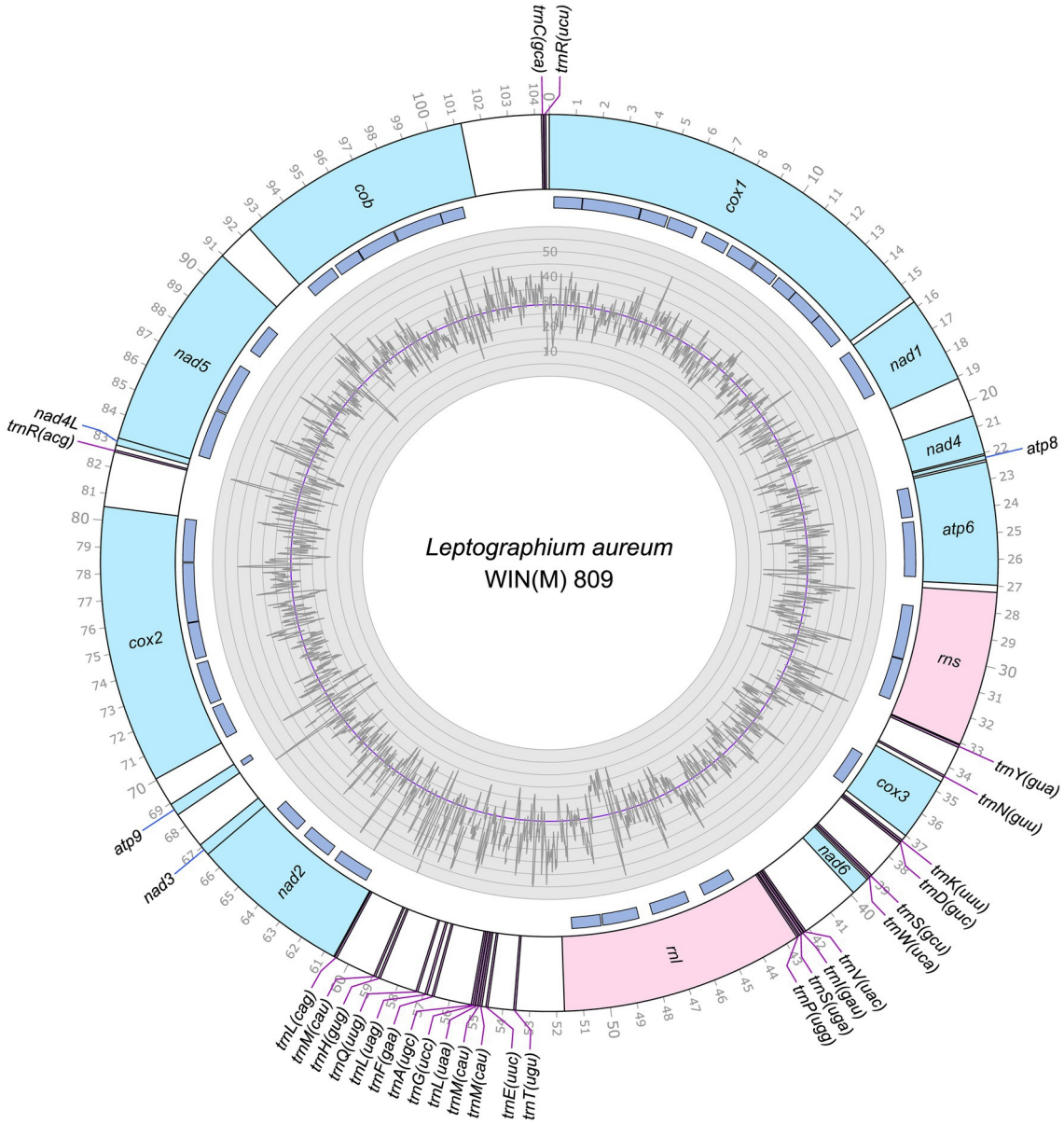

Figure 3.1: Circular representation of the mitochondrial genome WIN(M) 809. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

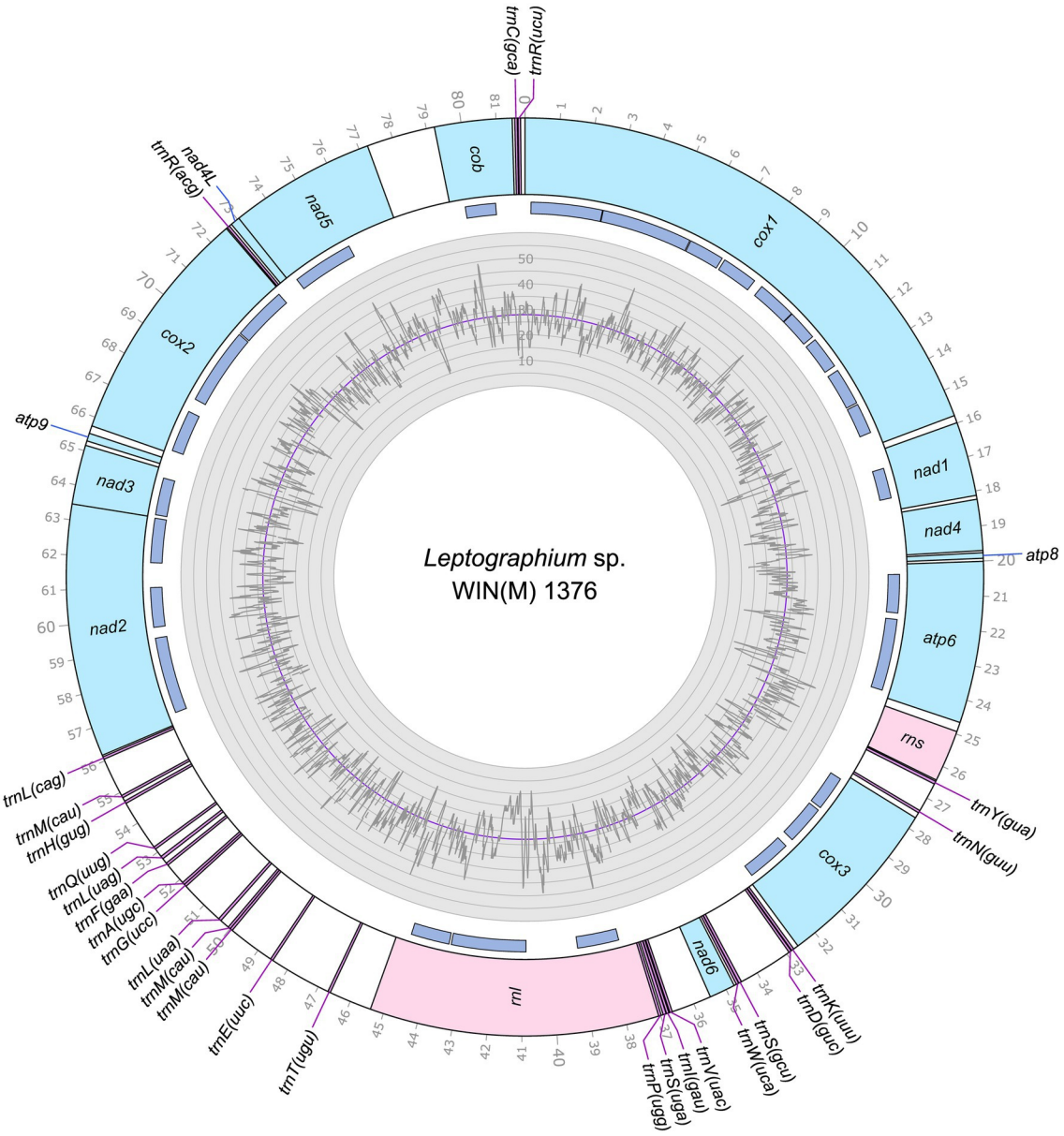

Figure 3.2: Circular representation of the mitochondrial genome of WIN(M) 1376. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

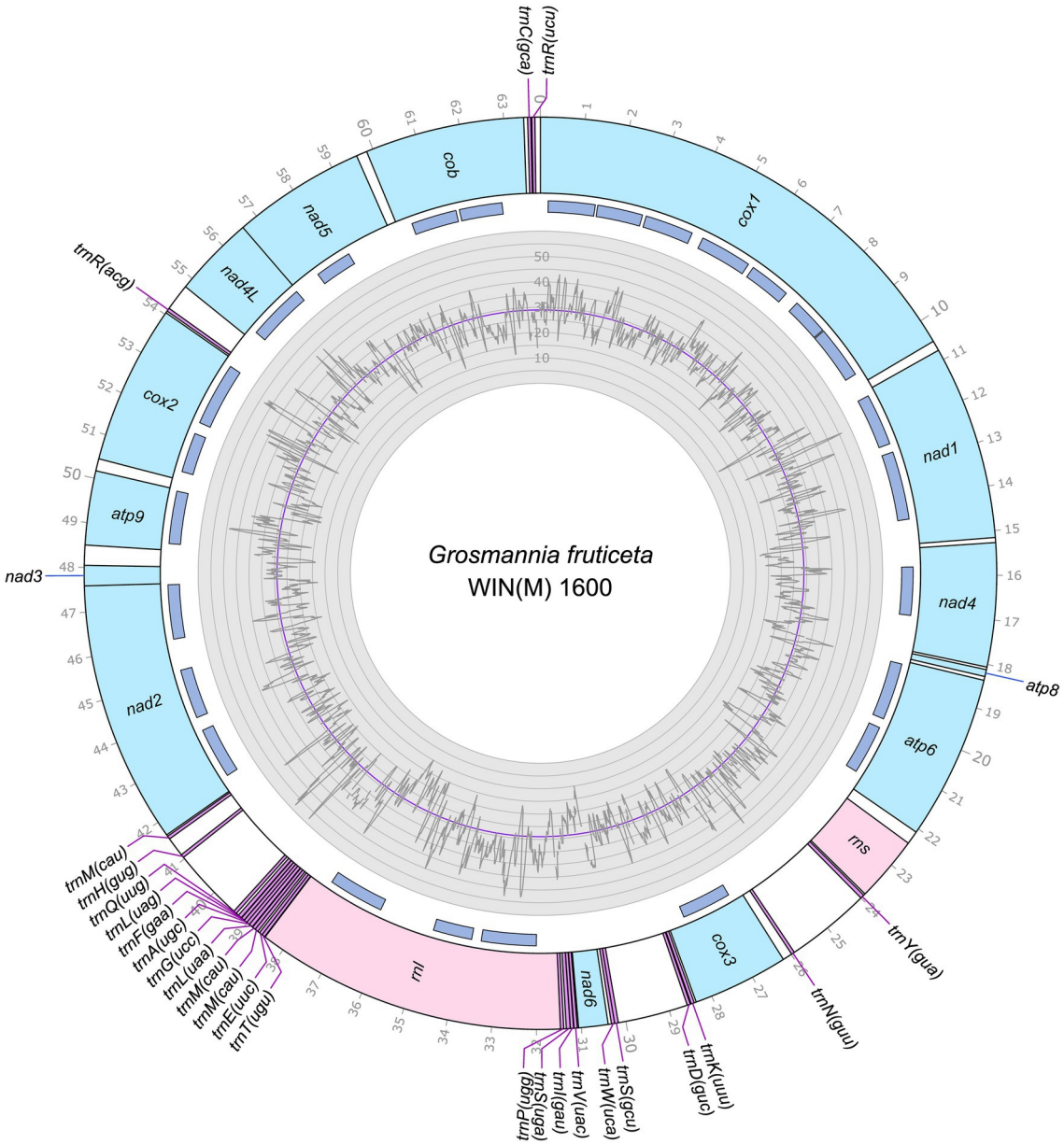

Figure 3.3: Circular representation of the mitochondrial genome of WIN(M) 1600. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

for all three genomes is identical and the same as previously reported for members of the Ophiostomatales (Wai and Hausner, 2021; Zubaer et al., 2021): *cox1*, *nad1*, *nad4*, *atp8*, *atp6*, *rns*, *cox3*, *nad6*, *rnl*, *nad2*, *nad3*, *atp9*, *cox2*, *nad4L*, *nad5*, *cob* (fig. 3.4). For *L. aureum* and *L.* sp. WIN(M) 1376 the *nad4L* and *nad5* genes are separated by one nucleotide and the *nad2* and *nad3* genes are "fused" although they are separated by a stop (TAA) codon. In *G. fruticeta* the *nad4L* and *nad5* genes overlap by one nucleotide and the *nad2* and *nad3* genes are fused as described for *L. aureum* and *L.* sp. WIN(M) 1376.

The tRNA genes are arranged into a few interspersed clusters with the majority of tRNAs arranged upstream and downstream of the *rnl* gene and a grouping of four tRNA genes between the *cox3* and *nad6* genes. *L. aureum*, *L.* sp. WIN(M) 1376, and *G. fruticeta* have the same tRNA gene configuration except for the larger tRNA cluster downstream of the *rnl* gene, here the *G. fruticeta* has 12 instead of 13 tRNA genes and it has a *trnL* as its last member for this cluster compared to *trnM* as seen in the other two species. This is the same as was observed previously for *G. penicillata* which has the *trnL* as its terminal member for this tRNA gene cluster whereas *L. lundbergii* has the *trnM* gene at this location (fig. 3.4; Zubaer et al., 2021). Most mitochondrial tRNA genes are single copies but there are exceptions. All three species have three copies of *trnM* (CAU), *L. aureum* and *L.* sp. WIN(M) 1376 have alternate versions for *trnL* (CAG, UAG and UAA), *trnR* (ACG and UCU) and *trnS* (GCU, UGA) and *G. fruticeta* has the same alternate versions for *trnL*, *trnR* and *trnS* except it lacks *trnL* (CAG).

Intron content of the studied mitogenomes

The mitochondrial genomes of *L. aureum*, *L.* sp. WIN(M) 1376 and *G. fruticeta* contain 37, 27 and 25 introns, respectively. In *L. aureum*, 36 are Gr I introns, encoding double-motif LAGLIDADG (LAG(2)) and GIY-YIG (GIY) type open reading frame (ORF) and one Gr II intron that encodes a reverse transcriptase-like (RT)

protein (see fig. 3.4). In *L.* sp. WIN(M) 1376 and *G. fruticeta*, all the introns can be assigned to Gr I, encoding LAG(2) and GIY ORFs. In some fungal mitogenomes, the *rnl* group IA (*rnl*-2,450; nomenclature based on Johansen and Haugen, 2001) intron encodes the RPS3 protein (Bullerwell et al., 2000). The *rnl*-2,450 intron in *L. aureum* encodes a LAG(2) (ORF434) located downstream of the rps3 ORF, whereas in *L.* sp. WIN(M) 1376 the *rnl*-2,450 encoded ORF is a fusion of the rps3 coding sequence with a LAG(2) coding sequence. In *G. fruticeta* WIN(M) 1600 the *rnl*-2,450 intron only codes for RPS3.

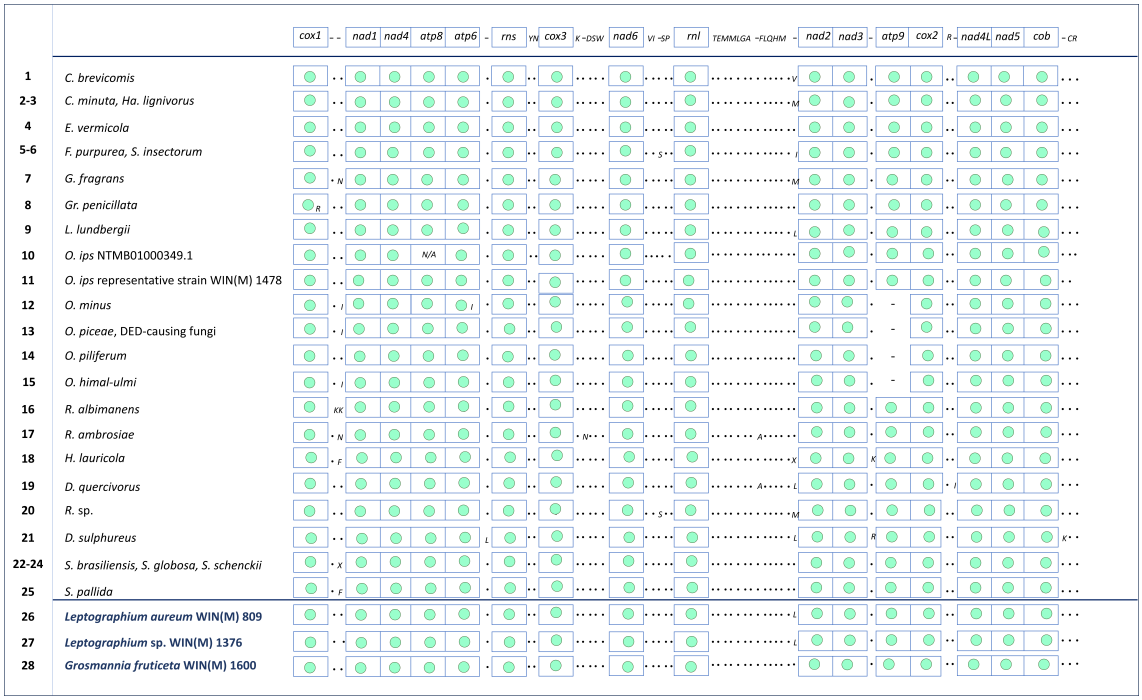
cox1
nad1
nad4
atp8
atp6
rns
cox3
nad6
rnl
nad2
nad3
atp9
cox2
nad4L
nad5
cob
1 C. brevicomis
2-3 C. minuta, Ha. lignivorus
4 E. vermicola
5-6 F. purpurea, S. insectorum
7 G. fragrans
8 Gr. penicillata
9 L. lundbergii
10 O. ips NTMB01000349.1
11 O. ips representative strain WIN(M) 1478
12 O. minus
13 O. piceae, DED-causing fungi
14 O. piliferum
15 O. himal-ulmi
16 R. albimanens
17 R. ambrosiae
18 H. lauricola
19 D. quercivorus
20 R. sp.
21 D. sulphureus
22-24 S. brasiliensis, S. globosa, S. schenckii
25 S. pallida
26 Leptographium aureum WIN(M) 809
27 Leptographium sp. WIN(M) 1376
28 Grosmannia fruticeta WIN(M) 1600

Figure 3.4: Gene synteny for 26 members of the Ophiostomatales. Amino acids are represented with the single-letter code. C.,*Ceratocystiopsis*; E., *Esteya*; F., *Fragosphaeria*; G., *Graphilbum*; Gr., *Grosmannia*; Ha., *Hawksworthiomyces*; L., *Leptographium*; O., *Ophiostoma*; R., *Raffaelea*; S., *Sporothrix* ; "N/A," not applicable; "–", absence of gene; "·", presence of tRNA genes and "green circle," presence of ribosomal and protein-coding genes (as in first line). The presence of *atp9* is observed for the three species reported in this study. Among the reported species of *Ophiostoma*, only *Ophiostoma ips* is known to encode atp9 in the mitogenome. Variation in terms of presence/absence is observed mostly for tRNA genes among the Ophiostomatales.

Fig. 3.5 summarizes the location of all introns, their classification, and the intron-encoded proteins (IEPs). The *atp9* group IA intron in *L. aureum* and *rnl* Gr IC1 intron in *G. fruticeta* show no evidence for an ORF (introns *atp9*-181 and *rnl*-965, respectively). The *nad1* Gr IA intron (*nad1*-145) appears to encode a degenerate GIY ORF. In *L. aureum*, *Leptographium* sp. WIN(M) 1376 and *G. fruticeta* the *cox1* gene has the most intron insertions (10, 9 and 7, respectively) followed by the *cox2* (5) and *cob* (5) genes in WIN(M) 809 which is the most intron-rich genome in this (fig. 3.5). Introns and intergenic sequences accounted for most of the genome size (greater than 80%), with 58% intron content in *L. aureum* and 55% intron content in *L.* sp. WIN(M) 1376 and *G. fruticeta* (figs. 3.6A–C). A complex intron can be recognized in the *cox1* gene at position 212 (based on Gr I intron naming nomenclature) in *L.* sp. WIN(M) 1376. A related version (same insertion site) of this intron was also recorded for *L. aureum* but here the intron is a Gr IB intron encoding a GIY ORF that is fused (in frame) to the upstream *cox1* exon. In *L.* sp. WIN(M) 1376 the intron is composed of a Gr IB module encoding a GIY YIG ORF and a partial Gr IB intron module that appears to be located within the P1 region of the host Gr IB intron (see fig. 3.7). Based on sequence comparison it appears that the original version, as represented by the *cox1*-212 intron in *L. aureum*, was invaded by a LAG(2)-type homing endonuclease gene (HEG) that moved along with it a partial Gr I B intron sequence. The LAG(2) HEG inserted into the N-terminal region of the resident GIY YIG ORF, fused in frame with the GIY component and thereby is in frame with the upstream *cox1* exon. The *cox1*-212 intron in *G. fruticeta* is composed of a Gr IB intron encoding a degenerated (fragmented) version of a GIY ORF due to the presence of premature stop codons.

Gene		*Leptographium aureum* WIN(M)809			*Leptographium* sp. WIN(M)1376			*Grosmannia fruticeta* WIN(M)1600		
		Insertion site	Intron subgroup	IEP	Insertion site	Intron subgroup	IEP	Insertion Site	Intron subgroup	IEP
cox1	1	212	IB	GIY	212[a]	IB	LAG(2), GIY	212	IB	GIY
	2	240	IB	LAG(2)	240	IB	LAG(2)	281	IB	LAG(2)
	3	281	IB	LAG(2)	281	IB	LAG(2)	386	IB	LAG(2)
	4	386	IB	LAG(2)	386	IB	LAG(2)	709	ID	LAG(2)
	5	867	IB	LAG(2)	709	ID	LAG(2)	867	IB	LAG(2)
	6	1057	IB	GIY	731	IB	LAG(2)	1262	IB	GIY
	7	1125	IB	LAG(2)	867	IB	LAG(2)	1296	IB	GIY
	8	1262	IB	GIY	1057	IB	GIY			
	9	1281	IB	GIY	1125	IB	LAG(2)			
	10	1296	IB	GIY						
nad1	1	145	IA	GIY	636	IB	GIY	145	IA	GIY[b]
	2							388	IA	GIY
nad4	1							505	IC2	LAG(2)
atp6	1	344	IB	LAG(2)	344	IB	LAG(2)	81	IC2	LAG(2)
	2	572	IC2	GIY	572	IC2	GIY	344	IB	LAG(2)
rns	1	913	ID	LAG(2)						
	2	952	II	LAG(2)						
cox3	1	640	IA	LAG(2)	219	IB	LAG(2)	333	IC2	LAG(2)
	2				333	IC2	LAG(2)			
	3				640	IA	LAG(2)			
rnl	1	1093	IC2	LAG(2)	742	IB	LAG(2)	742	IB	GIY
	2	1968	IC2	GIY	2450	IA	*	965	IC1	-
	3	2450	IA	*	2529	IC2	LAG(2)	2450	IA	*
	4	2529	IC2	LAG(2)						
nad2	1	378	IC2	LAG(2)	378	IC2	LAG(2)	378	IC2	LAG(2)
	2	765	IC2	LAG(2)	765	IC2	LAG(2)	765	IC2	LAG(2)
	3	1191	IC2	LAG(2)	1635	IA	LAG(2)	1635	IA	LAG(2)
nad3	1				90	IC2	LAG(2)			
atp9	1	181	IA	-				181	IA	GIY
cox2	1	108	IC2	LAG(2)	108	IC2	LAG(2)	228	IB	GIY
	2	318	IA	LAG(2)	592	IC2	GIY	552	IC2	GIY
	3	552	IC2	GIY	651	IC1	GIY			
	4	592	IC2	GIY						
	5	651	IC1	GIY						
nad5	1	248	ID	LAG(2)	248	ID	LAG(2)	717	IB	LAG(2)
	2	324	IC2	LAG(2)						
	3	1152	IC2	LAG(2)						
cob	1	201	IB	LAG(2)	506	IB	LAG(2)	393	ID	GIY
	2	393	ID	GIY				490	IA	LAG(2)
	3	437	IB	LAG(2)						
	4	490	IA	LAG(2)						
	5	506	IB	LAG(2)						

Figure 3.5: Summary of introns and intron-encoded proteins observed within the mitochondrial genomes of *Leptographium aureum* WIN(M) 809, *Leptographium* sp. WIN(M) 1376 and *Grosmannia fruticeta* WIN(M) 1600. Intron insertion sites for protein coding genes were based on *Tolypocladium inflatum* (NC 036382.1; Zhang and Zhang, 2019). No introns have been identified for *nad6* and *atp8* genes. Intron subgrouping applies only to group I introns; group II introns are annotated as "II"; Unidentifiable=No identifiable group I or II intron based on MFannot/RNAweasel results and manual inspection. Intron-encoded protein (IEP) associated with intron insertion site based on BLASTx results; LAG(2), double-motif LAGLIDADG homing endonuclease; RT, reverse transcriptase; GIY, GIY-YIG homing endonuclease; "–", no conserved motif detected based on BLASTx results. a. In WIN(M) 1,376, *cox1*-212 is a complex intron with one complete IB and a partial IB intron module. Both positioned at 212 with reference to *T. inflatum*. The second intron module is speculated to have invaded the resident intron module. b. Appears to encode a degenerate GIY. *Appears to encode a "rps3/HE-like fusion protein" (Bullerwell et al., 2000).

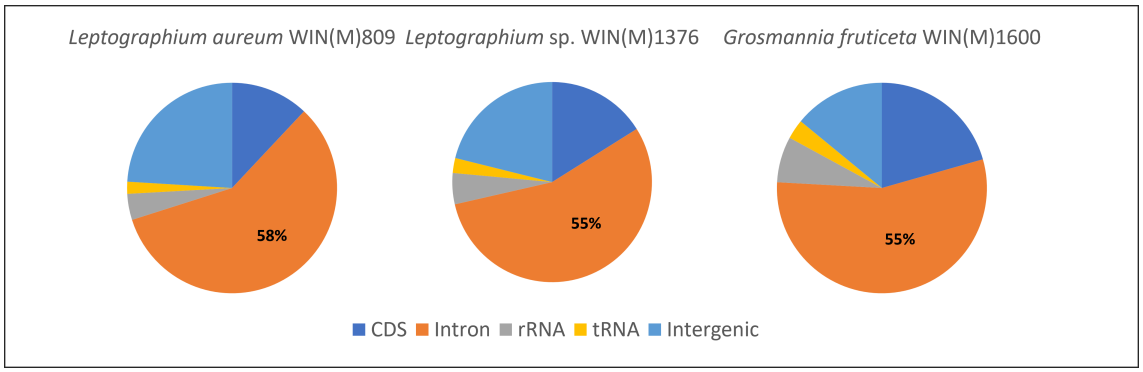

Figure 3.6: Composition of the mitochondrial genomes of WIN(M) 809, WIN(M) 1376 and showing the proportion of introns (including intron-encoded ORFs)

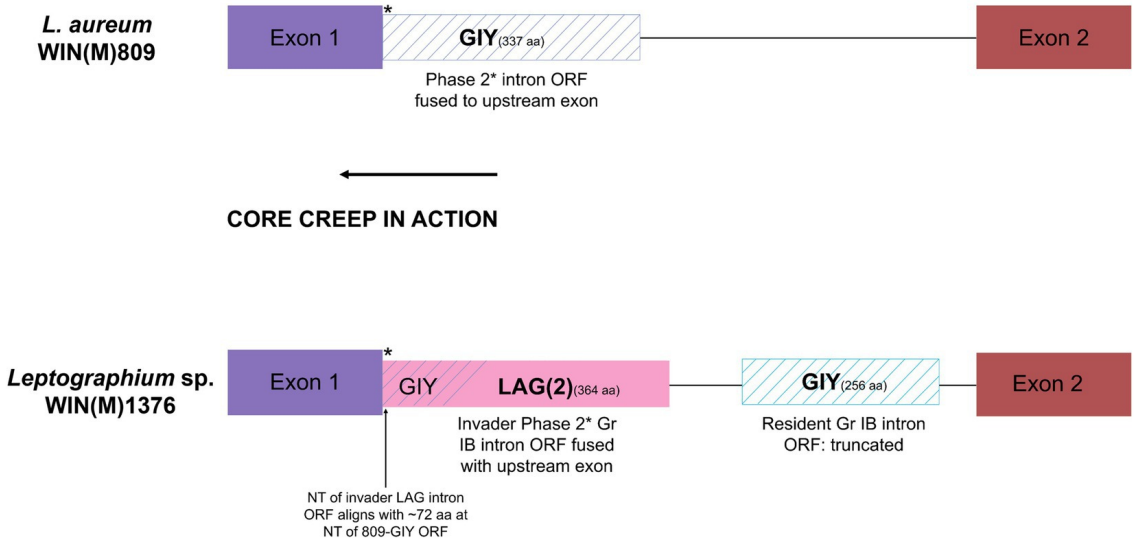

Figure 3.7: Schematic representation of the *cox1*-212 intron of *L. aureum* WIN(M) 809 and *Leptographium* sp. WIN(M) 1376. The *cox1*-212 intron in *L. aureum* belongs to the subgroup Group I and is a phase 2 intron (interrupts the second position of the codon) as the intron ORF (GIY) is fused to the upstream exon (see Figure 3.5). Similarly, the first intron in the *cox1* gene in WIN(M) 1376 is also a phase 2 intron (subgroup IB) where the LAG(2) ORF is fused with the upstream exon. It is speculated that this is the result of 'core creep' where the intron ORF over time, has incorporated upstream intronic sequences to fuse in-frame to the upstream exon (Edgell et al., 2011; Mukhopadhyay and Hausner, 2021). This fusion would allow the IEP to be more efficiently expressed, as it gains regulatory sequences of the host gene that optimize translation. The N-terminus of the LAG(2) ORF in *L.* sp. WIN(M) 1376 coincides with the N-terminus of the GIY intron in L. aureum, indicating that the remaining part of the resident GIY ORF has been displaced downstream due to the invasion of the LAG(2) intron in *L.* sp. WIN(M) 1376. There also is a partial IB intron component in *cox1*-212 intron of *L.* sp. WIN(M) 1376.

Phylogenetic analysis of the mitochondrial genomes

The mitogenome based phylogeny is based on the dataset comprised of 56 sequences that include 34 sequences representing 12 of the currently accepted 14 genera of the Ophiostomatales (De Beer et al., 2022). All members of the Ophiostomatales can be derived from one branch with high levels of confidence (100%) based on ML and MB analysis (fig. 3.8). The inferred phylogeny yielded a topology that supported the monophyly of the following groupings: Microascales, Hypocreales, Glomerellales, Sordariales, and the Ophiostomatales. Among the Ophiostomatales, we were able to sample from 12 genera: *Ophiostoma*, *Sporothrix*, *Ceratocystiopsis*, *Fragosphaeria*, *Hawksworthiomyces*, *Raffaelea*, *Harringtonia* (previously referred to as the *Raffaelea lauricola* complex; De Beer and Wingfield, 2013, De Beer et al., 2022), *Graphilbum*, *Grosmannia*, *Esteya*, *Leptographium*, and *Dryadomyces* (now including the *Raffaelea sulphurea* complex; De Beer and Wingfield, 2013; De Beer et al., 2022).

The mitogenome-based tree agrees with phylogenies published based on nuclear markers (reviewed in De Beer et al., 2022). The *L. aureum* and *L.* sp. WIN(M) 1376 sequences grouped with *L. lundbergii* and *G. fruticeta* grouped with *G. penicillata*. With regards to *L. aureum*, *Leptographium* sp. WIN(M) 1376, and *G. fruticeta* the mitogenomes confirm phylogenetic placements as obtained from the analysis of nuclear rDNA ITS regions and partial beta-tubulin sequences (Supplementary figs. 1A,B; image 1 and 2 respectively). The mitogenome data does confirm the separation of species with Leptographium-like conidial states into the two genera *Leptographium* and *Grosmannia*.

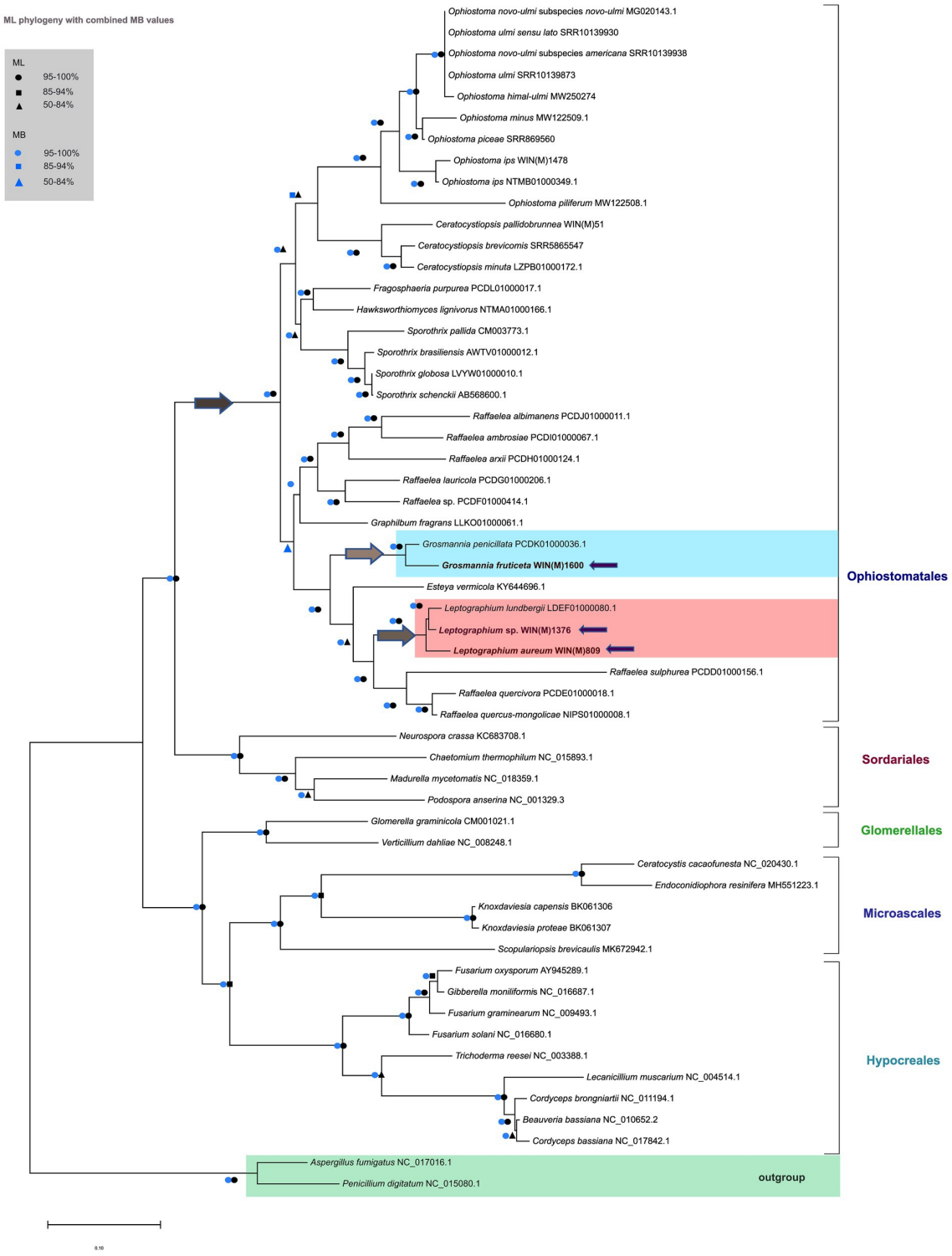
ML phylogeny with combined MB values
ML
95-100%
85-94%
50-84%
MB
95-100%
85-94%
50-84%
Ophiostoma novo-ulmi subspecies novo-ulmi MG020143.1
Ophiostoma ulmi sensu lato SRR10139930
Ophiostoma novo-ulmi subspecies americana SRR10139938
Ophiostoma ulmi SRR10139873
Ophiostoma himal-ulmi MW250274
Ophiostoma minus MW122509.1
Ophiostoma piceae SRR869560
Ophiostoma ips WIN(M)1478
Ophiostoma ips NTMB01000349.1
Ophiostoma piliferum MW122508.1
Ceratocystiopsis pallidobrunnea WIN(M)51
Ceratocystiopsis brevicomis SRR5865547
Ceratocystiopsis minuta LZPB01000172.1
Fragosphaeria purpurea PCDL01000017.1
Hawksworthiomyces lignivorus NTMA01000166.1
Sporothrix pallida CM003773.1
Sporothrix brasiliensis AWTV01000012.1
Sporothrix globosa LVYW01000010.1
Sporothrix schenckii AB568600.1
Raffaelea albimanens PCDJ01000011.1
Raffaelea ambrosiae PCDI01000067.1
Raffaelea arxii PCDH01000124.1
Raffaelea lauricola PCDG01000206.1
Raffaelea sp. PCDF01000414.1
Graphilbum fragrans LLKO01000061.1
Grosmannia penicillata PCDK01000036.1
Grosmannia fruticeta WIN(M)1600
Esteya vermicola KY644696.1
Leptographium lundbergii LDEF01000080.1
Leptographium sp. WIN(M)1376
Leptographium aureum WIN(M)809
Raffaelea sulphurea PCDD01000156.1
Raffaelea quercivora PCDE01000018.1
Raffaelea quercus-mongolicae NIPS01000008.1
Neurospora crassa KC683708.1
Chaetomium thermophilum NC_015893.1
Madurella mycetomatis NC_018359.1
Podospora anserina NC_001329.3
Glomerella graminicola CM001021.1
Verticillium dahliae NC_008248.1
Ceratocystis cacaofunesta NC_020430.1
Endoconidiophora resinifera MH551223.1
Knoxdaviesia capensis BK061306
Knoxdaviesia proteae BK061307
Scopulariopsis brevicaulis MK672942.1
Fusarium oxysporum AY945289.1
Gibberella moniliformis NC_016687.1
Fusarium graminearum NC_009493.1
Fusarium solani NC_016680.1
Trichoderma reesei NC_003388.1
Lecanicillium muscarium NC_004514.1
Cordyceps brongniartii NC_011194.1
Beauveria bassiana NC_010652.2
Cordyceps bassiana NC_017842.1
Aspergillus fumigatus NC_017016.1
Penicillium digitatum NC_015080.1
outgroup
Ophiostomatales
Sordariales
Glomerellales
Microascales
Hypocreales
0.10

Figure 3.8: Phylogenetic relationships of 56 fungal species belonging to the Ascomycota are presented, based on concatenated amino acid sequences, composed of *atp6*, *atp8*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*. *Aspergillus fumigatus* and *Penicillium digitatum* are selected as the outgroups. Maximum Likelihood (LG+I+F) as implemented in MEGA XI was used to generate the phylogenetic tree using the bootstrap option (1,000 pseudoreplicates) to estimate node support values. A second tree was constructed based on Bayesian inference and the posterior probability (PP) support values from the 50 percent majority Bayesian (MB) consensus tree are represented in combination with the Maximum Likelihood node support values (bootstrap support) on this tree. Key nodes of interest are indicated by arrows; dark grey arrow shows the node branching into the Ophiostomatales; light grey arrows indicate the nodes of branching into *Grosmannia* and *Leptographium* groups. The three fungal species in this study are also indicated by purple arrows. Black circles, squares and triangles represent bootstrap support (95–100 percent), (85–94 percent), and (50–84 percent) respectively for the ML analysis while blue circles, squares and triangles correspond to the PP support (95–100 percent), (85–94 percent), and (50–84 percent) respectively for the MB analysis, at the branches. Branch lengths are proportional to the number of substitutions per site (see scale bar).

3.1.5 Discussion

Mitochondrial Genome Organization

Mitochondria contain their own genomes and they have been reported to be derived from alpha-proteobacteria through endosymbiosis (Muñoz-Gómez et al., 2017). Compared with the more conservative genome sizes in animals, fungal mitogenomes show a wide variation in size and differ from their animal counterparts in terms of lower substitution rates, presence of introns and associated mobile elements, higher noncoding DNA, and recombination-associated repair mechanisms (Hausner, 2012; Freel et al., 2015; Deng et al., 2018; Lang, 2018; Sandor et al., 2018; Fonseca et al., 2021). Fungal mitochondrial genomes can be presented as a single circular molecule or be linear concatemers composed of mtDNA units in tandem arrangements (Valach et al., 2011; Lang, 2018). Although fungi share a common set of mitochondrial core genes, there is high variability in terms of gene order among the fungi, both between and within the major phyla (i.e., basidiomycetes, ascomycetes, and early diverging fungi; Aguileta et al., 2014).

Among the Ophiostomatales, mitogenomes can vary greatly in size ranging from 23.7 kb to about 150 kb (Zhang et al., 2019; Wai and Hausner, 2021; Zubaer et al., 2021). Size variation among the Ophiostomatales mitogenomes is in part due to the absence/presence of potential mobile introns, intron-encoded open reading frames, intergenic spacers, duplications, presence of non-conserved (or sometimes referred to as unknown) open reading frames (non-conserved or ncORFs and unidentified or uORFs; Wai and Hausner, 2021; Zubaer et al., 2021). Organellar introns can be either Gr I or Gr II introns, differentiated by their splicing mechanisms and their secondary and tertiary folds at the RNA level. The introns can encode open reading frames (ORFs) that express proteins that may catalyze their mobility to cognate alleles and/or enhance splicing efficiency by acting as maturases (reviewed by Prince et al., 2022).

Although mitochondrial gene synteny is highly conserved among the Ophiostomatales, the main variation arises with respect to the tRNA genes and the presence or absence of *atp9*. For the three fungal species studied, shared tRNA synteny was observed for *L. aureum* WIN(M) 809 and *L.* sp. WIN(M) 1376 with *L. lundbergii* whereas *G. fruticeta* WIN(M) 1600 shared tRNA synteny with *Esteya vermicola* (fig. 3.4) and for all these mitogenomes, the presence of *atp9* was noted. The absence of the *atp9* gene has been observed in other fungal taxa and presumable nuclear mitochondrial-derived versions of *atp9* have been identified Déquard-Chablat et al. (2011), Franco et al. (2017), Zubaer et al. (2018), Wai and Hausner (2021), and Zaccaron et al. (2021) suggesting that within the fungi a copy of the mitochondrial *atp9* gene was transferred to the nuclear genome. A study by Déquard-Chablat et al. (2011) demonstrated that many fungi have one or more copies of nuclear-encoded version(s) of *atp9* that encode the ATP9 peptide component for the mitochondrial ATP synthase. Scanning the nuclear genome contigs for *L. aureum*, *G. fruticeta*, and *L.* sp. WIN(M) identified nuclear-encoded *atp9* genes (GenBank accession numbers: OR271575-OR271577) therefore it can be assumed that members of the Ophiostom-

atales have nuclear-encoded versions of *atp9* and in some instances the mitochondrial version of *atp9* has been lost (see Zubaer et al., 2019; Wai and Hausner, 2021).

In fungi, most genes related to mitochondrial function are found in the nuclear genome (Bolender et al., 2008). It has been speculated that the escape of the mtDNA involves transfer to the cytosol and, consequently, to the nucleus where these genetic fragments are integrated with the aid of mobile elements that use the non-homologous (NHEJ) machinery (Berg and Kurland, 2000; Tsuji et al., 2012; Korovesi et al., 2018). The selective advantage of mtDNA transfer to the nucleus has been attributed to escaping the build-up of deleterious mutations, protection of DNA from mitochondrial mutagens (oxygen radicals) and thus the fixing of beneficial mutations (Allen and Raven, 1996; Blanchard and Lynch, 2000; Saccone et al., 2000). However, a core set of protein coding genes, rRNA genes and tRNA genes are retained in the mitochondrial genome for efficient local control of energy metabolism (Lang et al., 2012; Allen, 2015).

Other common gene arrangements among fungi include the fusion of *nad2* and *nad3* genes and the overlap between the ORFs of the *nad4L* and *nad5* genes. *L. aureum* and *L.* sp. WIN(M) 1376 are closely related to *L. lundbergii* and *G. fruticeta* is closely related to *G. penicillata*, and these related members of the Ophiostomatales show the fusion of *nad2-nad3* genes. As for the *nad4L-nad5* ORF arrangements, there is one nucleotide (T) separation in *L. lundbergii,* and one nucleotide overlap in *G. penicillata* (as observed in *G. fruticeta*). In the latter, the overlap of the stop and initiation codons between these genes in these pairs is the cause for their contiguity.

The tRNA gene clusters noted in the studied fungi (fig. 3.4) have also been observed in other members of the Ophiostomatales such as the tRNA genes reported upstream and downstream of the *rnl* gene (Wai and Hausner, 2021). In some metazoan mitogenomes, it has been proposed that tRNAs are positioned in a pattern that promotes the resolution of polycistronic mRNAs, referred to as the punctuation model (Salinas-Giegé et al., 2015). In the fungi examined in this work,

not all genes are separated by tRNAs so the punctuation model may not apply to the Ophiostomatales. The tRNA clusters near the rDNA genes (*rns* and *rnl*) may ensure all RNA molecules needed for translation are expressed at appropriate amounts.

Genome expansion due to intron proliferation

The three members of *Leptographium sensu lato* in this study were reported to have intron-rich genomes with introns and intergenic regions covering more than 80% of the mitogenome size, reinforcing their role in generating mitochondrial genome diversity. Related members such as *G. penicillata* at 150.9 kb (64 introns) that groups with *G. fruticeta* groups and *L. lundbergii* at 101.8 kb (36 introns) that groups with *L. aureum* and *L.* sp. WIN(M) 1376 also boast of intron-rich genomes correlated to their large genome sizes. This is consistent across different fungal phyla with variable genome sizes, sometimes associated with large-scale gene rearrangements, and complex intron dynamics (Zubaer et al., 2018, 2021; Li et al., 2020, 2021, 2023), indicating that phylogenetic positions do not necessarily correlate with genome size or intron content. One example of significant mitogenomic collinearity and consistent mitochondrial gene arrangement with vastly different mitogenome sizes has been reported for two members of the Pleosporales, with *Exserohilium rostratum* genome at 64,620 bp housing 17 introns (and 17 intronic ORFs) while the closely related *Exserohilium turcicum* at 264,948 bp contained 70 introns (and 126 intronic ORFs). This great difference in intron number suggests that gain/loss events of introns frequently occur in the mitogenome evolution in these fungi (Ma et al., 2022).

The variation in intron numbers among fungal mitochondrial genomes is puzzling with examples of streamlining (intron-loss) and genome expansions driven by intron-gain among the various members of the Ophiostomatales (Abboud et al., 2018; Zubaer et al., 2021). Introns and intergenic DNA in mitogenomes were originally considered a genetic liability, as they are targets for deleterious and potentially

lethal mutations (Lynch et al., 2006). The mutational burden hypothesis (MBH) was postulated to explain the origin of organellar genome size (Lynch et al., 2006) which stated that introns and intergenic DNAs tend to accumulate when natural selection is less efficient at purging hazardous non-coding DNA (Xiao et al., 2017).

Elevated sequence evolution near mobile introns and increased density of single nucleotide polymorphisms (SNP) in exon regions approaching intron boundaries have been noted in yeast mitogenomes (Repar and Warnecke, 2017), which correlate with homing endonuclease recognition sites. The rapid turnover of mobile introns can significantly impact genome size, but there are only a limited number of available intron insertion sites due to the requirement of conserved sequences for homing endonuclease. Thus, the expansion and contraction of mitogenomes (due to the gain and loss of introns) may cause only subtle change per event, but they take place persistently within the space limit (reviewed in Hao, 2022).

The presence and influence of non-coding elements in mitogenomes have been explored across several fungal groups (Fonseca et al., 2020; Araújo et al., 2021; Kulik et al., 2021). Models have been proposed that argue that mobile introns and their mobility-promoting IEPs are examples of "neutral evolution" (Goddard and Burt, 1999; reviewed in Hao, 2022). The model is supported by examples of introns and associated ORFs that invade a site followed by degeneration of the ORF, intron, and eventual loss of the element, which generates an "empty site" that can be reinvaded. Lack of selection permits a cycle of invasion, degeneration, loss, and reinvasion. Homing endonuclease genes (which can be free-standing or embedded within introns) are classified as selfish or parasitic elements that can transpose by breaking the DNA at specific sites, which consequently leads to gene conversion (Barzel et al., 2011; Stoddard, 2014). Mobile introns can move horizontally and invade new genomes, facilitated by IEPs (HEs for Gr I introns and reverse transcriptases for Gr II introns; reviewed by Hausner, 2012, Mukhopadhyay and Hausner, 2021). Intron and HEG origins and their coevolution have been speculated (Loizos et al., 1994;

Megarioti and Kouvelis, 2020) and it has been postulated that introns serve as a "safe haven" for these HEGs to get established. Edgell (2009) states that both HEGs and introns are selfish, independent elements, but they often benefit from being in close genomic proximity (e.g., HEGs are encoded by introns, and introns can use the encoded endonuclease to spread more easily). The association of HEGs with introns ensures their mutual persistence in the genome, evading the negative selective pressure to eliminate selfish elements from the genome and thus outpacing degeneration with mobility to unoccupied (cognate or new) sites by outcrossing and other means (Stoddard, 2014). Transient hyphal fusion might allow the exchange of cytoplasm between different fungal strains that could lead to fusion of mitochondria to facilitate mtDNA recombination and homing or ectopic transposition of introns and HEGs. Intron loss can be mitigated by intron components (ribozyme or ORF) gaining essential functions such as acting as trans-acting maturases or the intron encoding essential proteins (such as RPS3).

Although initially viewed as autocatalytic, organellar introns typically require co-factors to achieve splicing-competent configurations, these can be intron- or nuclear-encoded (Lang et al., 2007; Prince et al., 2022). The requirement of additional components can turn these introns into regulatory elements whose removal becomes a rate-limiting step in the expression of their host genes. Rudan et al. (2018) suggested that mitochondrial intron splicing is an essential component for normal mitochondrial function in *Saccharomyces cerevisiae*. If this is true for other fungi, the splicing of mitochondrial introns and their reliance on *cis*- and *trans*-acting factors encoded by the mitochondrial and nuclear genomes may be a fortuitous mechanism that evolved to allow for finetuning mitochondrial function to specific environments and life histories. For members of the Ophiostomatales, mtDNA intron landscapes have been generated and biases with regards to intron insertion sites and genes that are more likely to be intron-rich, have been noted, but there are no conserved introns found in all members of this group of fungi (except for *rnl*-2450 that encodes for

RPS3; Zubaer et al., 2021; Wai and Hausner, 2022). This might imply that fine-tuning mitochondrial gene regulation is not based on specific introns, instead based on an assemblage (or intron complement) composed of various introns (at various sites) that are "functionally" redundant. Finally, the reliance on nuclear factors for organellar intron splicing impacts mitonuclear compatibilities (or incompatibilities) and potentially imposes reproductive barriers, thereby promoting speciation events (reviewed in Dujon, 2020).

Complex introns are composed of multiple intron modules, which may be the result of introns invading other introns. Twintron-like arrangements have been identified in fungal mitochondrial genes, which splice out in consequent inter-dependent steps or as zombies (Zumkeller et al., 2020) where the composite intron splices as one unit. High intron load may promote more recombination events, but it may also place constraints on those exon sequences that are required for proper intron splicing (P1 and P10 interactions). Horizontal transfer among closely related (or distant) species of introns in addition to co-conversion of flanking markers promotes variability among members of the population. Thus, complex introns could serve as platforms for alternative splicing pathways that can be explored in terms of their potential in differential expression patterns for their associated ORFs. In this study, a complex intron was predicted in the *cox1* gene of *L.* sp. WIN(M) 1376 where a double-motif LAGLIDADG HEG invaded a resident Gr I intron ORF and brought along a partial Gr IB intron segment, fused in-frame with the GIY ORF, and the upstream *cox1* intron. This fusion is an example of "core creep" (Edgell et al., 2011; Mukhopadhyay and Hausner, 2021) whereby the intronic ORF extended ("creeped") towards the upstream exon and has fused in frame with the exon (Edgell et al., 2011). This pseudo-"exonization" event (intron sequences becoming coding sequences, but eventually spliced out) requires that the intron RNA folding capacity is not altered due to changes at the nucleotide level that result in the intron ORF being fused to the upstream exon. This might facilitate the intron-encoded ORFs that are present

in protein-coding genes to be more efficiently translated. In the same instance "core creep" is achieved by alternate splicing; here transcripts are generated that fuse the intron-encoded reading frame with the upstream exon segment (Turk et al., 2013; Sellem et al., 2016; Guha et al., 2018; Zubaer et al., 2018, 2021; Deng et al., 2020).

Phylogenetic analysis of mitogenomes representing various genera of the Ophiostomatales

Comparative mitogenomics has been applied to elucidate the population structure and genomic organization of fungal species, and study factors related to fungal pathogenicity, production of substances and enzymes of commercial interest, and other aspects of fungal evolution and biology including their coevolution with their associated vectors (Stajich, 2017; Jankowiak et al., 2018). The taxonomy of the Ophiostomatales has undergone considerable revisions in recent years; currently, the order Ophiostomatales includes two families, Kathistaceae and the Ophiostomataceae and monophyly is often not certain for the groupings *Leptographium sensu lato* (includes *Grosmannia* species) and *Ophiostoma sensu lato* (De Beer and Wingfield, 2013; De Beer et al., 2016). In this study, phylogenetic analysis relying on both DNA-based markers (ITS, beta-tubulin) and mitogenome-derived amino acid sequences helped in confirming the separation of *Leptographium* and *Grosmannia* species, with *L. aureum* and *L.* sp. strain WIN(M) 1376 grouping with *L. lundbergii* and *G. fruticeta* grouping with *G. penicillata.* In addition, the mitogenome-derived data supports the division of *Raffaelea sensu lato* (De Beer and Wingfield, 2013) into three genera: *Raffaelea*, *Harringtonia*, and *Dryadomyces* (De Beer et al., 2022).

3.1.6 Conclusions and applications

This study showed that the structural variation and variabilities in size and composition in fungal mitogenomes could mainly be explained by the presence of accessory elements (introns and their encoded products). This study also identified an ex-

ample of a complex intron and a phenomenon linked to it, referred to as "core creep" where intron encoded ORFs have fused to the upstream exons to enhance the expression of intron encoded proteins. The use of concatenated mitogenome derived amino acid sequence-based phylogenies has the potential in resolving taxonomic groupings within the Ophiostomatales. As more mitogenome sequences and multi-loci sequence data for reference specimens become available for members of the Ophiostomatales, these can serve as a resource for DNA-based markers that can be utilized for large-scale phylogenetic genomic-based diagnostics and species delimitation. This could eventually lead to the development of a platform that can be used to recognize and explore the taxonomic diversity within the Ophiostomatales and assist in developing high-throughput strategies in biosecurity services to detect pathogens and invasive fungi (Trollip et al., 2021, 2022).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at:

https://www.ncbi.nlm.nih.gov/genbank/, OQ851464

https://www.ncbi.nlm.nih.gov/genbank/, OQ851465

https://www.ncbi.nlm.nih.gov/genbank/, OQ851466

https://www.ncbi.nlm.nih.gov/genbank/, OR146620

https://www.ncbi.nlm.nih.gov/genbank/, OR146621

https://www.ncbi.nlm.nih.gov/genbank/, OR146622

https://www.ncbi.nlm.nih.gov/genbank/, OR146617

https://www.ncbi.nlm.nih.gov/genbank/, OR146618

https://www.ncbi.nlm.nih.gov/genbank/, OR146619

https://www.ncbi.nlm.nih.gov/genbank/, OR271575

https://www.ncbi.nlm.nih.gov/genbank/, OR271576

https://www.ncbi.nlm.nih.gov/genbank/, OR271577

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1240407/full#supplementary-material

Chapter 4

Exploring the Genus *Ophiostoma sensu stricto*

4.1 Characterization of the mitochondrial genomes for *Ophiostoma ips* from various geographic origins and related species: large intron rich genomes and complex intron arrangements

The work presented in this chapter is in preparation for submission.

Mukhopadhyay, J., Wai, A., and Hausner, G. (2024) Characterization of the mitochondrial genomes for *Ophiostoma ips* from various geographic origins and related species: large intron-rich genomes and complex intron arrangements.

JM, AW, and GH contributed towards the design of the project, and worked on the manuscript. JM and AW took the lead with regards to assembling the datasets and the final analysis of the data. JM prepared the first draft of the manuscript. JM and GH assembled the final version of the manuscript. All authors contributed

to the article.

4.1.1 Abstract

Members of the Ophiostomatales are of economic concern as many are blue-stain fungi and some are plant pathogens and their taxonomic groupings have been revised several times in the past. The mitogenomes of members assigned to this order exhibit size polymorphism despite having highly conserved gene order, owing to the variable number of introns and intron insertion sites. Complex introns consisting of nested or tandem arrangements of group I and II introns have been previously noted in several species of this order. In this work, eleven blue-stain fungi including nine strains of *Ophiostoma ips* with a varied distribution across North America and New Zealand were sequenced and compared with other members of the Ophiostomatales. A pan-mitogenome intron landscape has been prepared to demonstrate the distribution of the mobile genetic elements (group I and II introns and intron-encoded homing endonucleases) and to provide insight into the evolutionary dynamics of introns among members of this group of fungi. The size variation among these mitogenomes (from about 23.8 kb to 152 kb) showed high correlation to the presence and absence of introns. Examples of complex or nested introns composed of two or three intron modules have been observed in some *Ophiostoma ips* strains; RNA-Seq analysis suggested possible alternative splicing pathways with regards to resolving these complex arrangements in *O. ips*. Species representing other genera within the Ophiostomatales have also been included to generate a diverse comparative dataset to examine gene content, gene order, phylogenetic relationships, and the distribution of mobile elements. Mitochondrial DNA and RNA data for *O. ips* provides the basis for further studies relating to gene annotation, alternative splicing, evolutionary intron dynamics and as a resource for taxonomic studies.

4.1.2 Introduction

Species assigned to the Ophiostomatales are usually characterized by producing ascocarps with asci that are randomly produced at the base of the ascocarp; the asci are short-lived and deliquesce with the ascospores being released as a sticky droplet at the tip of the perithecial necks. Many of these features along with tall conidiophores are usually associated with adaptations towards insect dispersal. The taxonomy of the Ophiostomatales has undergone many revisions (Zipfel et al., 2006; De Beer et al., 2013a; De Beer et al., 2022) the Order accommodated fungi that are currently placed into two families, *Ceratocystidaceae* and *Ophiostomataceae*. DNA sequence analysis revealed that convergent evolution of morphological characters generated conflicting taxonomic schemes for this group of fungi (Hausner et al., 1993a,b and c; Wingfield et al., 1993; Réblová et al., 2011), and now Ceratocystidaceae has been assigned to the Microascales and the Ophiostomataceae to the Ophiostomatles (De Beer et al., 2013a, 2016). Molecular data also helped in resolving generic boundaries for members of these families (Zipfel et al., 2006; De Beer et al., 2013a, b; De Beer et al., 2022). Currently, 16 genera are accepted for inclusion in the Ophiostamatales; however there could be up to 24 genera based on a recent analysis using four nuclear markers (De Beer et al., 2022). These genera include economically important tree pathogens and blue-stain fungi (Olchowecki and Reid, 1974, Wingfield et al., 1993, Hausner et al., 2005, Fisher et al., 2012). *Ophiostoma ips* (Rumbold) Nannfeldt is an economically important bark-beetle associated blue-stain (sapstain) fungus (Seifert, 1993) that has a global distribution and appears to be vectored by a broad range of insects, including *Ips* and *Dendroctonus* species (Rumbold, 1931; Rane et al., 1987; Uzunovic et al., 1999; Benade et al., 1995; Zhou et al., 2004a and b; Thwaites et al., 2005; Suh et al., 2013; Wingfield et al., 2017a, b). This species produces pleomorphic asexual states and morphologically is very similar to *Ophiostoma montium* and *Ophiostoma adjuncti* (Updadhyay, 1981), and hence its taxonomy, based on morphological characteristics, can be challenging at times (Kim

et al., 2003).

The mitochondrion is an important organelle present almost universally among eukaryotes, and being semi-autonomous, it contains its own genetic material (mitochondrial DNA or mtDNA), with nuclear genes supporting mitochondrial processes in *trans* (Sandor et al., 2018). Mitogenomes for filamentous members of the Ascomycota show great variation in size despite encoding a similar set of core genes: *rnl* and *rns* (large and small subunit RNAs involved in protein translation), *cob* and *cox1–3* (coding for components of the respiratory chain complexes), *atp6*, *atp8*, *atp9*, *nad1–6* and *nad4L* (coding for NADH dehydrogenase subunits) and a set of tRNA genes. Sometimes, the ribosomal protein RPS3 (*rps3*; Hausner, 2003; Wai et al., 2019) can be encoded within an *rnl* intron or be found as a free-standing gene.

Fungal mitogenomes harbor mobile genetic elements such as group I (Gr I) and group II (Gr II) introns. These elements essentially are composed of a ribozyme component and in some instances an open reading frame (ORF) that encodes intron encoded proteins (IEPs) that catalyze intron mobility and in many instances promote intron splicing (Belfort et al., 2002; Lambowitz and Zimmerly, 2011). The mode of splicing and the corresponding RNA folds distinguish between the two types of mitochondrial introns (Mukhopadhyay and Hausner, 2021). For Gr I introns IEPs tend to be site specific homing endonucleases (HEs) and for Gr II introns, reverse transcriptases (RTs) and these IEPs have applications in biotechnology such as genome editing reagents and/or possible regulatory switches to control gene expression (Takeuchi et al., 2011; Hafez and Hausner, 2012; Stoddard, 2014; Guha et al., 2017; Belfort and Lambowitz, 2019). In fungal mitogenomes two families of HEs can be recognized based on the presence of conserved amino-acid motifs: the LAGLIDADG or GIY-YIG families of HEs (Stoddard 2014). Homing endonuclease genes (HEGs) can be free-standing or embedded within Gr I intron sequences. HEGs can move independently from their ribozyme partners (Mota and Collins, 1988), but it has been suggested that intron-encoded HEG tend to co-evolve with their ribozyme

partners (Megarioti and Kouvelis, 2020; reviewed in Mukhopadhyay and Hausner, 2021). Finally, there are instances where Gr II introns encode HEGs; typically Gr II introns can be ORF-less or encode reverse transcriptases (Toor and Zimmerly, 2002; Mullineux et al., 2010; Zimmerly and Semper, 2015).

In some instances mobile introns can invade an existing intron and form a nested or complex intron arrangement. Originally described as twintrons, an intron-within-intron configuration and the position of the internal intron module within the pre-existing (resident) intron can determine whether or not its splicing is essential for the splicing of the resident intron (i.e., potentially co-operating ribozymes; Copertino and Hallick, 1994). It was assumed that the internal intron module splices first before the resident intron module can generate a splicing competent RNA fold (reviewed in Hafez and Hausner, 2015). The name 'zombie' twintrons (half-dead, half-alive) was suggested by Zumkeller et al. (2020) for cases where splicing of external introns does not depend on prior splicing of internal introns (that may have degenerated), there the nested intron arrangement can splice out as one unit.

Complex intron modules have been identified in the mitogenomes of various fungi including members of the Ophiostomatales (Hafez et al., 2013; Guha and Hausner, 2014; Guha et al., 2018; Zubaer et al., 2021; Mukhopadhyay et al., 2023). In some examples the splicing of the internal intron module will reconstitute the ORF that it interrupted, thereby allowing for its expression (Guha and Hausner, 2014). Other complex introns have the potential to generate splicing intermediates whereby the internally located intron ORF gets fused to the upstream exon (Guha et al., 2018; Zubaer et al., 2021). The latter might be an example of splicing mediated 'core-creep' (Zubaer et al., 2021). It has been reported by Edgell et al. (2011) that many intron encoded ORF sequences are fused, in frame, with the upstream exon of their host gene and this might enhance the expression of the IEP.

This study examines the mitogenomes from nine strains of *O. ips* isolated from North America and New Zealand. The sequences were compared to other avail-

able sequences from members of the Ophiostomatales, including the closely related species *Ophiostoma montium* and *Ophiostoma adjuncti*, plus other members of the Ascomycota. The *O. ips* mitogenomes were observed to be large (greater than 100 kb) and rich in introns including novel intron arrangements such as twintron (two intron modules) in the *cox3*-640 and a 'trintron' (composed of three intron modules) in *cob*-490. RNAseq data was generated to complement the genomic data for two strains of *O. ips* and this validated the gene annotations and provided preliminary data to support a proposed ratchet model (see Hafez and Hausner 2015) for the processing of the *cob*-490 (Zubaer et al., 2021).

This study is part of the long term project to generate and apply mitogenomes for members of the Ophiostomatales as a resource for developing markers for diagnostics and taxonomic applications plus to gain a better understanding of the evolutionary dynamics of introns and mitogenomes diversity.

4.1.3 Materials and Methods

Source of culture, culturing methods, and extraction of nucleic acids

Seven strains of *Ophiostoma ips*: WIN(M) 1478 [= TB 3.11; = UAMH12572; UAMH = UAMH Centre for global microfungal biodiversity, Toronto, ON, Canada], WIN(M) 1480 [=TB 3.10; = UAMH12573] and WIN(M) 1481 [=TB 2.5; =UAMH12574] [isolated TB = Thunder Bay, Canada, Hausner et al. (2005); WIN(M) = University of Manitoba]; WIN(1486) [=ATCC24285, ATCC = American Type Culture Collection; isolated Canada; = UAMH12575]; WIN(1487) [=CBS137.36, CBS =Westerdijk Fungal Biodiversity Institute; AUT-strain; Oregon, isolated USA; = UAMH12579]; WIN(M) 1488 [=SYPT1-USA-B Colette Breuil; =UAMH12576] [UBC; isolated West Coast- Washington State]; WIN(M) 1001 [=JR 88-130A';JR = James Reid; from New Zealand; =UAMH12577] and strains belonging to two related species *Ophiostoma* sp. WIN(M) 1515 [=JR 82-83a; from New Zealand; = UAMH12578], and *O. adjuncti* WIN(M) 502 [=ATCC34942; ex-T; iso-

lated New Mexico, USA; ; =UAMH12571] were maintained on malt extract agar (MEA; supplemented with yeast extract, 30 g/L malt extract, 1 g/L yeast extract, 20 g/L agar) slants and agar plates (containing approximately 40 mL MEA). Cultures were incubated in the dark at 20°C for up to 2 weeks. For the extraction of nucleic acids, three 250 mL Erlenmeyer flasks containing 80 mL of PYG + ME broth (1 g/L peptone, 1 g/L yeast extract, 2 g/L D-glucose, and 3 g/L malt extract) were each inoculated with ten agar blocks (2 mm × 2 mm × 1 mm) and incubated in the dark at 20°C for up to 10 days. Fungal mycelium was harvested from liquid media by vacuum filtration through a Whatman® Grade 1 filter paper in a Büchner funnel. Mycelium was ground up in a pre-chilled mortar with a pestle and acid-washed sand and whole genome DNA for next generation sequencing was recovered and purified as previously described (Wai and Hausner 2021). DNA samples (30 ng/µL DNA in a final volume of 100 µL) were sent to MicrobesNG (Units 1–2 First Floor, The BioHub, Birmingham Research Park, 97 Vincent Drive, Birmingham, B15 2SQ, United Kingdom) for whole genome Illumina sequencing.

For two strains of *O. ips* [WIN(M) 1478 and 1480], RNA was prepared from 7-10 day old cultures (˜50 mg wet weight of fungal mycelium) using the RNAeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) following the protocol outlined by the manufacturer. Contaminating DNA was removed from the nucleic acid samples using the TURBO-DNase kit (Applied Biosystems, ABI, Foster City, CA, USA). RNA was prepared according to specifications in the the User Guide for Illumina Sequencing Technologies by Genome Quebec and submitted for preparing total RNA libraries without poly(A) selection, depletion of rRNA and tRNA and size selection. RNA samples for RNAseq analysis were prepared in duplicates for two strains of *O. ips* and the entire RNA seq analysis was carried out twice with RNA samples prepared in 2021 and 2022.

Assembly, analyses, and annotation of next generation sequencing mitogenome data

Initial analyses of the Illumina reads were performed on the online platform GALAXY (https://usegalaxy.eu/; Afgan et al., 2018). The reads from MicrobesNG were initially assessed using FastQC v0.11.9.1 Assemblies were generated by using two programs: SPAdes vs. 3.14.0 (setting the "–careful" option and assembly graph option) and the independent (not part of Galaxy) A5-miseq pipeline (with "-end" option set to "5"; Tritt et al., 2012; Coil et al., 2015). The program Bandage (Wick et al., 2015) was used to examine the assembly graph files generated from SPAdes to aid in the recovery of potential mitogenome contigs. To more efficiently recover mtDNA derived reads and for mitochondrial genome assemblies, the program GetOrganelle (Jin et al., 2020) v1.7.5, with the organelle type set to fungus mitogenome (i.e., -F fungus_mt) was applied to the Illumina sequencing reads. Additional mitogenomes, including sequences for *Ophiostoma montium* (SRR19396179), *Ophiostoma ips* (SRR19396180) and other members of the Ophiostomatales were obtained from GenBank and the Sequence Read Archive (SRA; Shumway et al., 2010; Leinonen et al., 2011). The collection, processing, and analyses of SRA data for mitogenomes was previously described in Wai and Hausner (2021) and Mukhopadhyay et al. (2023). The mtDNA contigs were annotated using the MFannot program (Lang et al., 2023; setting "Genetic Code" to (4 - mold) and RNAweasel (Prince et al., 2022); setting to predict tRNAs, and Gr I and Gr II introns. Predictions of tRNAs were also performed with tRNAscan-SE 2.0 (Chan and Lowe, 2019). Annotations of protein-coding genes (*atp6*, *atp8*, *atp9*, *cob*, *cox1–3*, *nad1–6*, *nad4L*) and noncoding genes (i.e., *rnl* and *rns*, and the tRNAs), were verified by comparative sequence analysis from data obtained from GenBank (Benson et al., 2013). Sequence alignments were generated for all protein coding genes with corresponding genes from *Tolypocladium inflatum* serving as a reference genome for naming introns (based on the position of the insertion) according to the proposed nomenclature by

Zhang and Zhang (2019). The rRNA coding genes were compared with those (16S and 21S RNA) of *E. coli* with regards to intron annotations/naming as described by Johansen and Haugen (2001). Gene annotations were refined with Artemis (Rutherford et al., 2000) and the mtDNAs were visualized using Circos (Krzywinski et al., 2009). Circos was set up with the appropriate coordinates to highlight exon/intron configurations for conserved protein-coding genes, noncoding genes, and to plot GC content. The GC plot was generated using a window size of 100 bp and a step size of 20 bp.

Reverse Transcriptase PCR and RNA Sequencing

For one-step RT-PCR, the SuperScript IV One-Step RT-PCR kit (Invitrogen, Carlsbad, CA, USA) was used. *Ophiostoma ips cob*-490 and *cox3*-640 (relative to the *S. cerevisiae cob* coding sequence; GenBank accession number: KP263414.1) RT-PCR primers were used (see Table 2.3) for amplification, followed by verification of cDNA using agarose gel electrophoresis on 1% gel.

Sequencing of RNA libraries (i.e., RNAseq) was performed on an Illumina Novaseq6000 platform (PE100 25M reads/lane) along with additional blocks of reads to ensure deep sequencing (in total, greater than 100M reads were generated per sample). For quality control, obtained paired end reads were run through FastQC for quality control and Trimmomatic (Bolger, 2014) to remove low quality bases and adapters. Additionally Rcorrector (courtesy of F. Lang, University of Montreal) was used to error-correct Illumina reads (i.e., correct for RT artifacts and random sequencing errors).

RNA-Seq reads mapping and visualization

The reference genomes for WIN(M) 1478 and 1480 obtained through the annotation of NGS data were utilized for reference-based mapping, alongside *de novo* assembly using rnaSPAdes (https://github.com/ablab/spades; Bushmanova et al.,

2019). Since the mitogenome sequences were identical for WIN(M) 1478 and 1480, the two independent RNA-Seq runs ensured both technical and biological replicates. The reads were aligned to the reference genome using (RNA)STAR (https://github.com/alexdobin/STAR; Dobin et al., 2013). The output from STAR is a BAM file which was visualized on the Integrative Genomics Viewer, IGV (http://www.broadinstitute.org/igv; Robinson et al., 2011). IGV also provided confidence in the prediction of exonic and intronic regions (i.e., gene structure prediction from MFannot). On IGV, soft-clipped bases (bases in the 5′ and 3′ ends of the read that are not part of the alignment with the reference sequence) were viewed to facilitate detection of structural variations that might arise due to splicing (such as alternative splicing events).

Phylogenetic analysis of mitochondrial protein-coding regions

For the mitochondrial genomes a dataset was generated composed of 13 concatenated amino acid sequences, derived from the following protein-coding genes: *atp6*, *atp8*, *cob*, *cox1–3*, *nad1–6*, *nad4L*, and data were aligned using MAFFT version 7 (Katoh and Standley, 2013). Some members of the Ophiostomatales do not encode *atp9* within their mitogenomes and so it was not included in the dataset (Zubaer et al., 2021; Wai et al., 2021). The alignment containing 65 mitogenomes was manually adjusted with AliView version 1.25 (Larsson, 2014) and analyzed with MrBayes 3.2.7a (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012) for inferring phylogenetic trees. A fixed-rate amino acid substitution model was estimated using the mixed model function implemented in MrBayes (MB). Rate variation among sites was modeled with a combination of the invariable sites model and gamma model (i.e., rates from a gamma distribution). Other parameters were left at their default values (i.e., a uniform distribution between 0.0 and 1.0 and an exponential distribution with a mean value of 1.0, respectively). The analysis was performed with 2,000,000 generations with a sampling frequency of 1,000. The CPrev+F+I was estimated

with the highest probability. The first set of 25% of sampled trees was discarded (burn-in) and the remaining trees were used to construct the 50% majority rule consensus tree. The output of the phylogenetic analyses (i.e., tree files) was visualized with the FigTREE program 1.4.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and members of the Eurotiales were selected as the outgroup for the analysis. The aligned data set was also analyzed with programs contained within MEGA XI (Tamura et al., 2021): Maximum Likelihood (ML). For ML, the LG model (plus I and F) was applied, and 1,000 bootstrap replicas were analyzed to assess branch support values.

4.1.4 Results

The mitogenomes: size variation and gene content

Previously an *O. ips* mitogenome for one strain was assembled (see Zubaer et al., 2021) based on whole genome data (NTMB01000349.1) from strain CBS138721 (isolated in the United States; CBS = Westerdijk Fungal Biodiversity Institute) published by Wingfield et al. (2017a). However, the mitogenome was not complete (*atp8* region was missing). Therefore, in this study additional strains were examined from two different geographic regions to get a better representation of the mitogenome for this economically important species. The pangenomic intron landscape for the mitogenomes of *O. ips* strains WIN(M) 1478, WIN(M) 1480, WIN(M) 1481, WIN(M) 1486, WIN(M) 1487, WIN(M) 1488, WIN(M) 1001, WIN(M) 1515, and *O. adjuncti* WIN(M) 502 is given in fig. 4.1. Each mitogenome has been represented as a circular molecules [suppl. fig. 4.11-4.18], respectively. These vary in size with WIN(M) 1478 and 1480 being the largest at 113,671 bp, and WIN(M) 1515 being the smallest at 39,957 bp. The strain WIN(M) 1515 was originally identified as *O. ips* but the size of the ascospores were reported to be above the range typically observed for *O. ips* (James Reid notes from 1982). The mitogenome sizes of the rest are as follows: WIN(M) 1481 at 111,826 bp, WIN(M) 1486 at 80,045 bp, WIN(M) 1487 at 85,284 bp, WIN(M) 1488 at 106,900 bp, *O. adjuncti* WIN(M) 502 at 112,641 bp, and

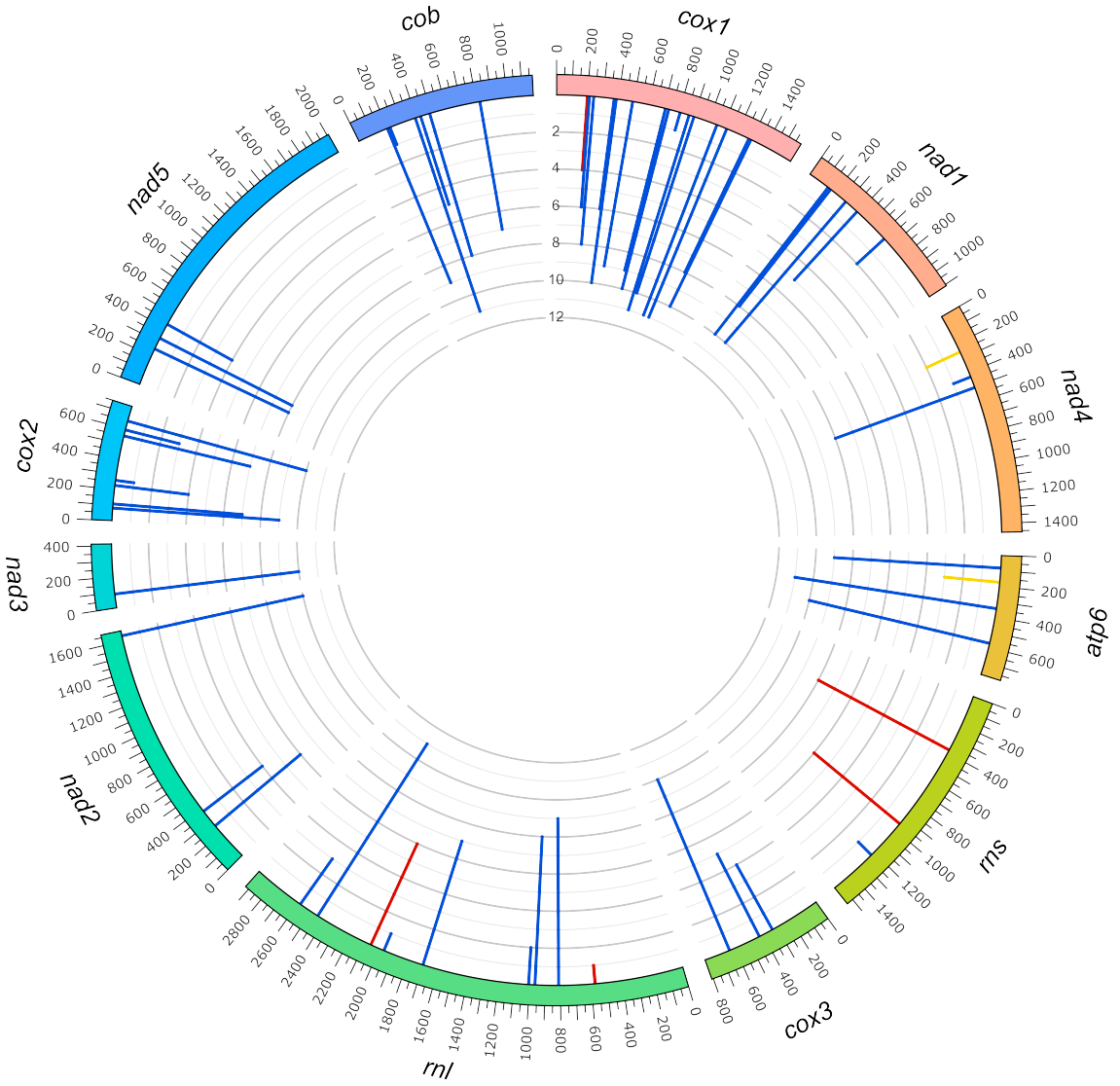

Figure 4.1: The pangenomic intron landscape for the studied members of the Ophiostomatales. The landscape was generated by Circos and shows all intron insertions sites and their frequencies.

WIN(M) 1001 at 104,698 bp. Additionally, we included *O. montium* SRR19396179 (76,883 bp), and *O. ips* NTMB01000349.1 (97,849 bp). All mitogenomes studied encoded the ribosomal RNA genes *rns* and *rnl*, the mitochondrial core set of 14 protein-coding genes including *rps3* (*atp6*, *atp8*, *atp9*, *cob*, *cox1–3*, *nad1–6*, *nad4L*, and *rps3*). As observed in other members of the Ophiostomatales (and members of the Sordariomycetes), the ribosomal protein RPS3 is encoded within a Gr IA-type intron embedded within the *rnl* gene (*rnl* position mL2450 reviewed in Wai et al., 2019). The presence of *atp9* was recorded for all studied *O. ips* strains, *O. montium* and *O. adjuncti* mitogenomes. Previously it was observed that some members of the Genus *Ophiostoma* (*O. minus*, *O. piliferum* (Zubaer et al., 2021) and *O. piceae*, *O. ulmi* and *O. novo-ulmi* (Wai and Hausner 2021) lost the mtDNA version of the *atp9* gene.

All nine mitogenomes encode a set of 25 tRNAs that cover all 20 standard amino acids. The tRNA genes are arranged into a few interspersed clusters with the majority of tRNA genes arranged upstream and downstream of the *rnl* (largest cluster between *rnl* and *nad2*) gene and a smaller grouping of tRNA genes between the *cox3*/*nad6* and *nad6*/*rnl* genes. The tRNA configuration is the same for all the studied *O. ips* and related mitogenomes with *trnM* as the last member of the largest cluster downstream of the *rnl* gene. Each of the studied genomes have three copies of *trnM* (CAU), two copies with alternate versions for *trnL* (UAG and UAA), and two copies with alternative versions of *trnR* (ACG and UCU), and two copies with alternative versions of *trnS* (GCU, UGA).

The gene order and orientation were all identical and as previously observed for members of the Ophiostomatales (Mukhopadhyay et al., 2023; Wai and Hausner, 2021; Zubaer et al.,, 2021): *cox1, nad1, nad4, atp8, atp6, rns, cox3, nad6, rnl, nad2, nad3, atp9, cox2, nad4L, nad5, cob* (fig. 4.2). Mauve progressive alignment of the mitogenomes showed that, despite the size polymorphism among the strains of the same species, the genomes are collinear, and showed no evidence of rearrangements

(fig. 4.3). Other features noted are the fusion of the *nad2* and *nad3* genes and the overlap between the *nad4L* and *nad5* ORFs by one nucleotide, i.e., the last nucleotide of the *nad4L* stop codon serves as the first nucleotide of the *nad5* ORF. These gene arrangements have been previously observed in other members of the Ophiostomatales and Ascomycota (Mukhopadhyay et al., 2023; Aguileta et al., 2014; Abboud et al., 2018).

Figure 4.2: Gene synteny for 37 members of the Ophiostomatales. Amino acids are represented with the single-letter code. Genus names: C., *Ceratocystiopsis*; D., *Dryadomyces*; E., *Esteya*; F., *Fragosphaeria*; G., *Graphilbum*; Gr., *Grosmannia*; H., *Harringtonia*; Ha., *Hawksworthiomyces*; L., *Leptographium*; O., *Ophiostoma*; R., *Raffaelea*; S., *Sporothrix*; N/A, not applicable; –, absence of gene; ·, presence of genes.* Indicates that the *O. ips* strains WIN(M) 1478, 1480, 1481, 1486, 1487, 1488, 1001, and the *ips*-like *Ophiostoma* sp. WIN(M) 1515 are represented by *O. ips* WIN(M) 1478 and they shared identical gene synteny. Related species *O. montium* and *O. adjuncti* also share the gene order. The *O. ips* strain NTMB01000349.1 seemed to be missing the *atp8* gene containing segment in the scaffold 143 (Zubaer et al. 2021).

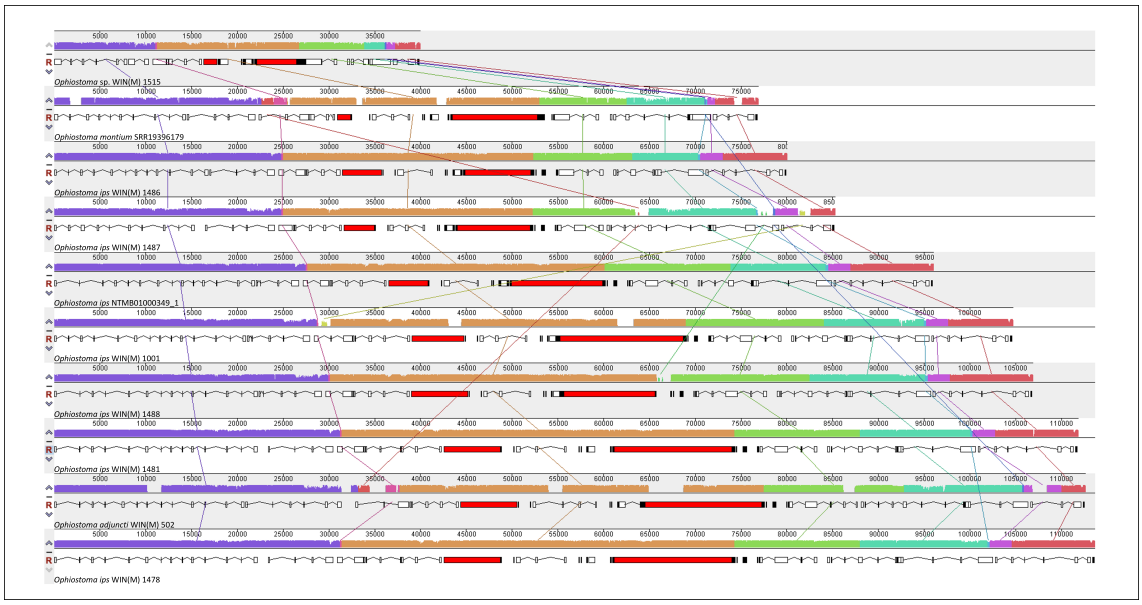

Figure 4.3: Mauve progressive alignment of the mitochondrial genomes for the ten members of the *O. ips* complex (*O. ips* strains, *O. ips*-like and related species). The mitogenomes are co-linear and therefore show no evidence of genome rearrangement. Lines connecting Locally Collinear Blocks (LCBs) are shown that represent conserved (homologous) segments of sequences across the genomes. WIN(M) 1515 is selected as the reference genome due to the smallest size and least number of introns among the studied species. The other species with correspondingly larger genomes and huge number of introns can be viewed as evidence of intron gain/loss events with respect to WIN(M) 1515.

The mitogenome intron content

The pangenomic intron landscape for the studied members of the Ophiostomatales is illustrated in fig. 4.1. The *cox1* gene was recorded with the most intron insertion sites [13 in WIN(M) 1478, 1480, 1481, 1488, 1001; 11 in WIN(M) 1486; 9 in WIN(M) 1487; 5 in WIN(M) 1515], followed by the *rnl* [6 in WIN(M) 1478, 1480, 1481, 1001; 4 in WIN(M) 1488; 3 in WIN(M) 1486 and 1487; 1 in WIN(M) 1515] and the *cob* gene [5 in WIN(M) 1478, 1480, 1481, 1488; 4 in WIN(M) 1001, WIN(M) 1486; 2 in WIN(M) 1487]. For the related species, *O. adjuncti* WIN(M) 502, 12 intron insertion sites were noted in *cox1*, 6 in *rnl*, and 2 in *cob*.

The mitogenomes of WIN(M) 1478, 1480, and 1001 contain 46 introns, while mitogenomes of WIN(M) 1481 and 1488 contain 45 introns. *O. ips* SRR19396180 with a smaller genome at 96,039 bp contains 39 introns, whereas WIN(M) 1487 at 85,284 bp contains 33 introns and WIN(M) 1486 at 80,045 bp contains 32 introns. Among the related species, *O. adjuncti* WIN(M) 502 contains 45 introns, whereas *O. montium* SRR19396179 contains 30 introns. The smallest of the studied mitogenomes, The *O. ips*-like WIN(M) 1515 strain contains only 12 introns. A related species, *O. adjuncti* WIN(M) 502 has a mitogenome size of 112,641 bp with 45 introns. Fig. 4.4 summarizes the location of all introns, their classification and intron-encoded ORFs. Gr I introns encoded either GIY-YIG or LAGLIDADG type ORFs with the *nad1*-145 Gr I introns in WIN(M) 1488 and 1001 showing no evidence of an ORF. For *atp6*-173 in WIN(M) 1478, 1480, and 1481, it was difficult to identify the intron type due to a partial imperfect duplication of the downstream exon at the beginning of the intron.

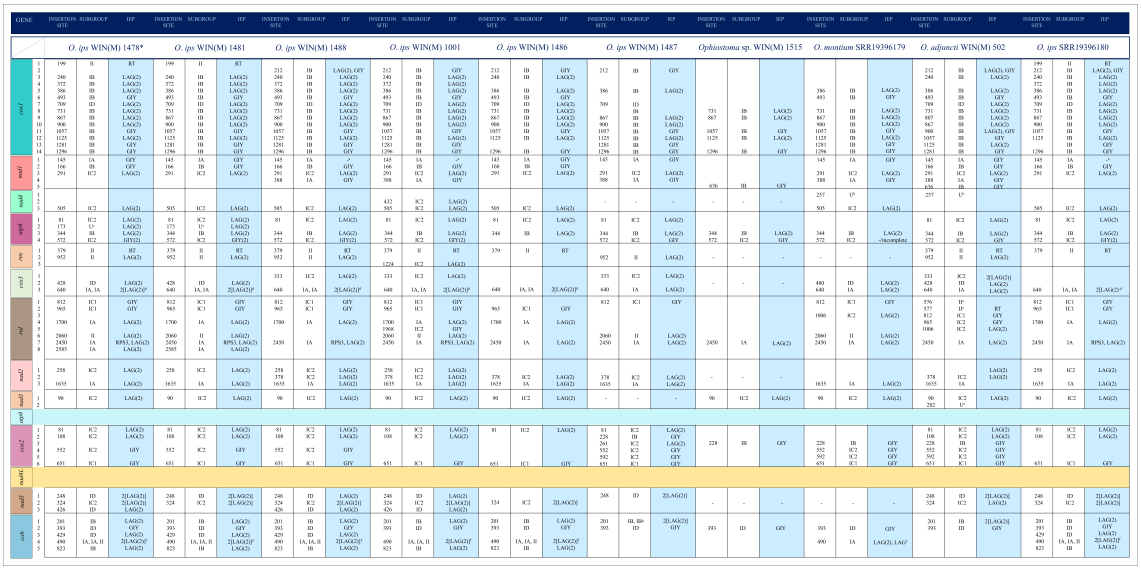

Figure 4.4: Summary of introns and intron-encoded proteins observed within the mitochondrial genomes of the studied members of the Ophiostomatales. Intron insertion sites for protein coding genes were based on *Tolypocladium inflatum* (NC_036382.1; Zhang and Zhang, 2019). No introns have been identified for *nad6* and *atp8* genes. Intron subgrouping applies only to group I introns; group II introns are annotated as "II"; Unidentifiable = No identifiable group I or II intron based on MFannot/RNAweasel results and manual inspection. Intron-encoded protein (IEP) associated with intron insertion site based on BLASTx results; LAG(2), double-motif LAGLIDADG homing endonuclease; RT, reverse transcriptase; GIY, GIY-YIG homing endonuclease; "–", no conserved motif detected based on BLASTx results. a. Does not appear to encode ORF (GIY); b. The beginning of the intron resembles ending of *nad4* gene; for WIN(M) 502, encodes additional nucleotides in comparison to *Ophiostoma montium* (SRR19396179) with *trnR(ccu)* accounting for $\tilde{7}1$ bp of the "extra" $\tilde{1}781$ bp [LAG (could be single/double-motif; note: this is in addition to LAG(2) shared by both]; c. For *atp6*-173, and *nad3*-282, there is partial imperfect duplication of the downstream exon at the beginning of the intron; it is difficult to identify the intron type; *atp6*-173 encodes trnI (aau) in WIN(M) 1478, 1480, and 1481; d. Complex twintron *cox3*-640 consisting of two group IA intron modules in tandem; e. Two group II introns at positions *rnl*-576 and *rnl*-577 are separated by a mini-exon (A) in an arrangement based on domain V; one RT-ORF encoded by *rnl*-577; f. Complex trintron consisting of a group II intron module between two gr IA modules; for *O. montium* (SRR19396179), *cob*-490 (IA) encodes LAG(2) that appears to be interrupted by potentially single-motif LAG; g. *cob*-201 complex intron with two modules of group IB introns (according to MFannot) and encoding two double-motif LAGLIDADG endonucleases.

Most of the observed introns belong to Gr I, while Gr II introns are mostly found in the ribosomal genes (*rnl*, *rns*), and in the *cox1* and *cob* genes of some members. The *atp6*-572 Gr I intron encoded two GIY-ORFs in all members, except WIN(M) 1515 and WIN(M) 1487 which encoded only one GIY-ORF. In WIN(M) 1486, an intron was not detected at *atp6*-572. The *cox1*-212 intron in addition to the GIY-ORF encoded by WIN(M) 1001, 1486, and 1487, encodes a double motif-LAGLIDADG ORF in WIN(M) 1488. Similarly, *nad5*-248 in WIN(M) 1478, 1480, and 1481 encodes two double motif-LAGLIDADG ORFs whereas in WIN(M) 1001, 1487 and 1488, only one double motif-LAGLIDADG ORF is encoded. The *rps3* gene is encoded within an *rnl* intron (mL2450) in WIN(M) 1478, 1480, 1481, and 1001, and here the *rps3* sequence is part of a gene fusion where the C-terminus rps3 coding segment is fused in frame with a double-motif LAGLIDADG coding region (Sethuraman et al., 2009).

Among the eleven studied mitogenomes, including *O. ips* and related strains, we identified the following Gr II introns: four *cox1*-199 introns, two *rnl*-576, 577 introns (two modules as part of a neighboring arrangement), and eight *rns*-379 introns encode for an RT, whereas the remaining encode LAGLIDADG HE ORFs. The *rns*-952 Gr II introns in *O. ips* WIN(M) 1478, 1480, 1481, 1487, 1488, and *O. adjuncti* WIN(M) 502, and the *rnl*-2060 Gr II introns in *O. ips* WIN(M) 1001, 1478, 1480, 1481, 1487 and *O. montium* SRR19396179 encode double-motif LAGLIDADG HE ORFs. Gr II introns are observed in the complex intron (trintron) arrangement in *cob*-490 in WIN(M) 1001, 1478, 1480, 1481, 1488, and *O. ips* SRR19396180. Upon initial inspection in WIN(M) 502, two Gr II introns appeared to form a tandem twintron-like arrangement at the *rnl*-576 position however a one nucleotide mini-exon could be identified that denotes Gr II introns at rnl-576 and rnl-577 (see next section for more details).

Novel intron arrangements: Focus on complex introns

In this study, complex/nested intron arrangements were observed in the *cob* (cytochrome b; *cob*-490) and *cox*3 (cytochrome 3; *cox3*-640) genes in the following *O. ips* strains: WIN(M) 1478, 1480, 1481, 1488, and 1001. Of these, the complex *cox3*-640 was also noted in WIN(M) 1486, but a single Gr IA module was observed at *cob*-490. Both complex arrangements have been reported in *O. ips* CBS138721 (NTMB01000349.1) and their RNA folds have been described in Zubaer et al. (2021, see figs. 7 and 8). A schematic overview of these complex introns, *cob*-490 and *cox3*-640 can be found in Zubaer et al. 2021 (figs. 9 and 11).

i. The *cox3*-640 twintron

The *cox3*-640 complex intron appears to be composed of two Gr IA1 intron modules in a tandem arrangement with the upstream and downstream Gr IA1 intron modules also separated by an inter-intron module sequence. BLASTn did not show any related and similar sequences to the inter-intron module in GenBank and alternative or intermediate spliced forms for this complex intron were not observed by RNA-Seq analysis. Careful examination of the "inter-intron" module sequence failed to reveal any sequences that could be utilized to form a suitable P10 interaction for the upstream intron module (similar to previous notes from Zubaer et al., 2021).

ii. The *cob*-490 trintron

The *cob*-490 complex intron consists of three distinct modules that contain all the necessary components for splicing. The three modules are a Gr IA1 intron that is interrupted by a Gr IIB intron module and this composite element is inserted within the P1 loop of a Gr IA1 intron module (presumably the resident intron). The Gr II intron appears to be ORF-less and is located within the P8 loop of the host Gr I intron module. Both Gr I intron components contain ORFs that encode double motif LAGLIDADG type homing endonucleases. There is a short sequence separating the two Gr I intron modules. This so-called "inter-intron module sequence" could be used as a "pseudoexon" by the internal Gr I intron component for the formation of

the P10 helix or for the resident intron module for its P1 formation (fig. 10; Zubaer et al., 2021). This is supported by the RNA-Seq data that supports the splicing of the "upstream" intron prior to the "downstream" intron.

RNA-Seq reads were mapped to the reference annotated genomes of WIN(M) 1478 and 1480 and visualized on IGV. This visualization showed evidence for transcript reads in which the pseudoexon sequence 'TCATAT' was fused to the upstream exon sequence (the last 8 nt on the 3' end being 'CGTTGAGT') which represents the "intermediate splicing step" according to the ratchet model where the upstream intron has been spliced in the first step leading to this fusion (fig. 4.5). Two independent runs of RNA sequencing yielded reproducible reads of this sequence, indicating that they are not likely sequencing artifacts, supporting the proposed model that this complex intron would use ratchet splicing as mode of RNA processing. Preliminary RT-PCR results showed only the end product which is the spliced out composite trintron unit for *cob*-490 and alternative or intermediate splice products could not be detected.

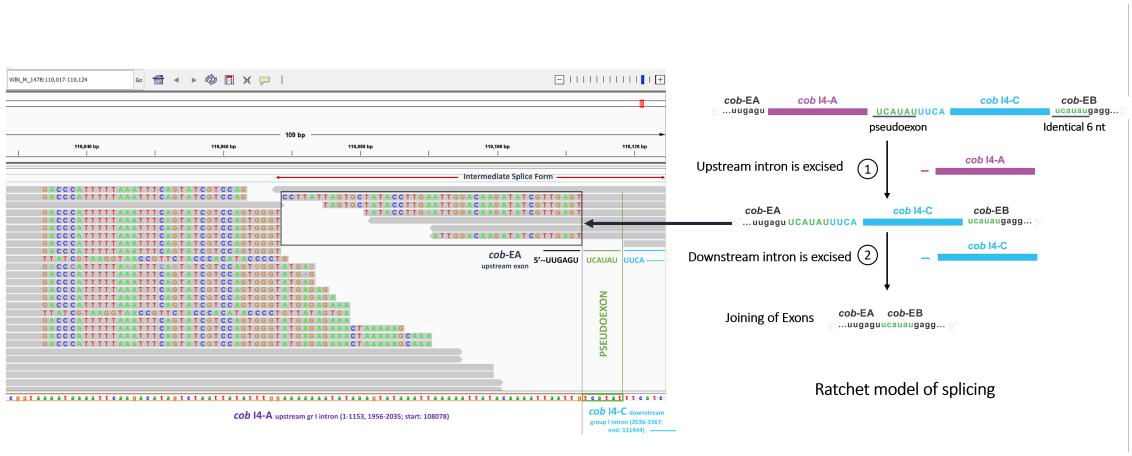

Figure 4.5: RNA-Seq reads mapped to the annotated mitogenome of WIN(M) 1478, showing the "pseudoexon" sequence. The IGV snapshot shows the mapped RNA-Seq reads, with the following highlighted: the upstream intron sequence of *cob*-490 (relative to *T. inflatum*) in purple (*cob* I4-A; group I intron with start position 108078 relative to the mitogenome sequence; spanning from 1-1153 and 1956-2035 relative to the ungapped I4 intron sequence); the downstream intron sequence of *cob*-490 in blue(*cob* I4-C; group I intron with end position 111444 relative to the mitogenome sequence; spanning from 2036-3367 relative to the ungapped I4 intron sequence) containing the pseudoexon sequence (TCATAT) in green (spanning from 2036-2041 relative to the ungapped I4 intron sequence). The numbers in brackets represent the positional range of each intron element taking the start of *cob* I4 as position 1. RNASeq reads show the "intermediate splice form" as hypothesized in the "ratchet" splicing model from Zubaer et al. (2021) has been shown where the *cob*-EA upstream exon sequence is fused to the pseudoexon sequence (part of *cob* I4-C) indicating that for these mapped reads, the upstream intron *cob* I4-A has been spliced out in a preliminary splicing step. This map validates the ratchet splicing model of the complex intron where intron modules utilize the pseudoexon sequence for splicing in multiple steps, through intermediate splice forms.The right panel shows a simplified schematic adapted from the ratchet model (Figure 10, Zubaer et al. 2021) that shows the intermediate splicing step after the upstream intron I4-A has spliced out leading to the fusion of the *cob*-EA with the "pseudoexon" sequence of the downstream intron I4-C.

iii. The *cob*-201 intron

In a strain listed as *Ophiostoma ips* WIN(M) 1487, the *cob*-201 intron appears to be a complex intron composed of two Gr IB intron modules, according to MFannot. These modules encode LAGLIDADG ORFs the introns are arranged in tandem (schematically represented in fig. 4.6). However, the second intron module at the 5′ end contains a short 18 nt sequence that could be a duplication of the downstream exon; this could lead to an alternate interpretation whereby the two Gr IB introns are separated by an 18 nt exon. The "intron" 18 nt component potentially encodes the amino acid sequence that is expected for the downstream exon. What further complicates the interpretation is that the intron version of the 18 nt nucleotides resembles the exonic version of this sequence in *O. ips* WIN(M) 1478 and related strains (see next section) and the WIN(M) 1487 downstream exonic 18 nt sequence resembles the corresponding sequence as found in *O. adjuncti* WIN(M) 502.

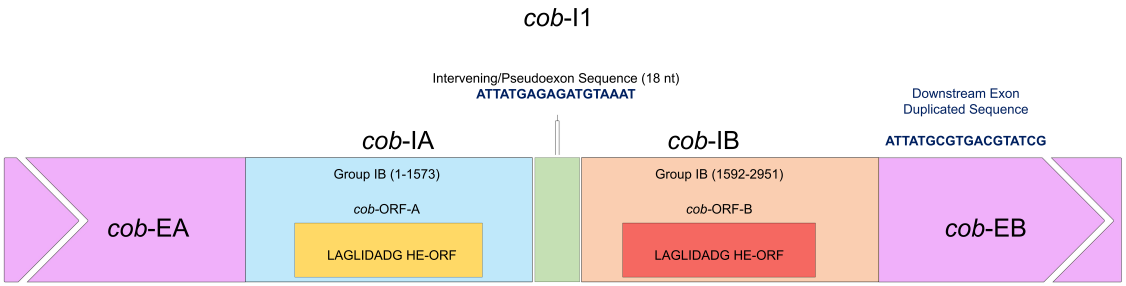

Figure 4.6: A schematic diagram of the *cob*-201 intron in *O. ips* WIN(M) 1487. *cob*-I1, the entire complex twintron at *cob*-201 position (relative to *T. inflatum*). *cob*-IA is the upstream group IB intron; *cob*-EA, upstream exon; *cob*-ORF-A, encoded by *cob*-IA; *cob*-IB is the downstream group IB intron; *cob*-ORF-B, encoded by *cob*-IB; *cob*-EB, downstream exon. LAGLIDADG represents the type of homing endonuclease ORF encoded by the two group IB intron modules. The numbers in brackets represent the positional range of each intron element taking the start of *cob*-I1 as position 1. The two group I introns are separated by an intervening sequence of length 18 nt which can be interpreted as a "pseudoexon" which seems to be duplicated (partially) as a downstream exon sequence. This tandemly arranged twintron can also be interpreted as two independent group IB intron modules at positions 201 and 219, respectively (relative to *T. inflatum*).

iv. The *rnl*-576 and 577 co-operating neighbors

A novel intron arrangement has also been identified in *Ophiostoma adjuncti* WIN(M) 502 (schematically represented in fig. 4.7). Here two Gr II introns (based on domain V identified through RNAweasel) have been predicted at positions *rnl*-576 and *rnl*-577. Possible interactions between intron and exon binding sequences (IBS1, 2, 3 and EBS1, 2 and 3, respectively) at the folded secondary structure levels for the two introns are shown in the fig. 4.8 (for detailed view of intron folds, see suppl. figs. 4.19-4.20). Based on alignment with *Escherichia coli*, the intron at *rnl*-576 does not start or end with typical, conserved nucleotides (begins with 'UUGCG' and ends with 'GGU' instead of the conserved 'GUGYG'). Though not as common, 'UUGCG' has been observed previously (Michel et al., 1989). The non-canonical ending is favored by the possible IBS3/EBS3 interactions highlighted in fig. 4.8. The *rnl*-577 Gr II intron appears to encode an RT-ORF (in domain IV) while the *rnl*-576 intron is ORF-less. These two intron modules appear to be separated by a mini-exon ('A' nucleotide). Both introns based on structural features are Gr IIB types. The EBS3/IBS3 interactions favour the existence of the one nucleotide exon at this position; a similar configuration was previously reported in Mayer et al. (2021) for members of the genus *Ceratocystis*.

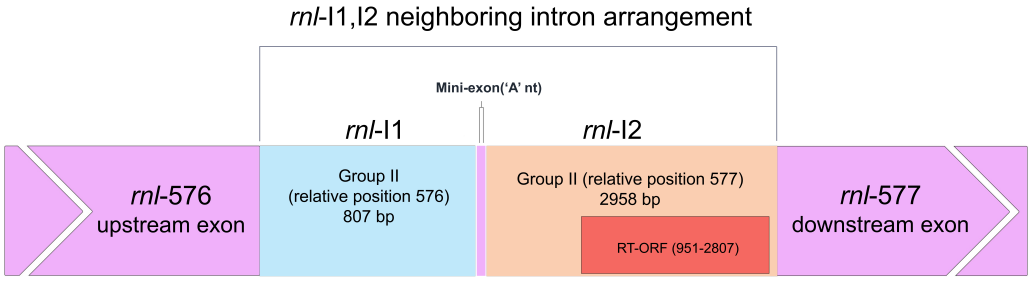

Figure 4.7: A schematic diagram of the neighboring arrangement of the *rnl*-576 and 577 introns in *O. adjuncti* WIN(M) 502.*cob*-I1, the entire intron arrangement starting at *rnl*-576 position (relative to *E. coli*). *rnl*-I1 is the first group II intron at position 576; *cob*-ORF-A, encoded by *cob*-IA; *rnl*-I2 is the second group II intron at position 577; RT-ORF, encoded by *rnl*-I2 and the range is given taking the first nucleotide of the intron as position 1; the upstream and downstream exons are also shown. The two group II introns are separated by a single nucleotide 'A' which can be called a "miniexon".

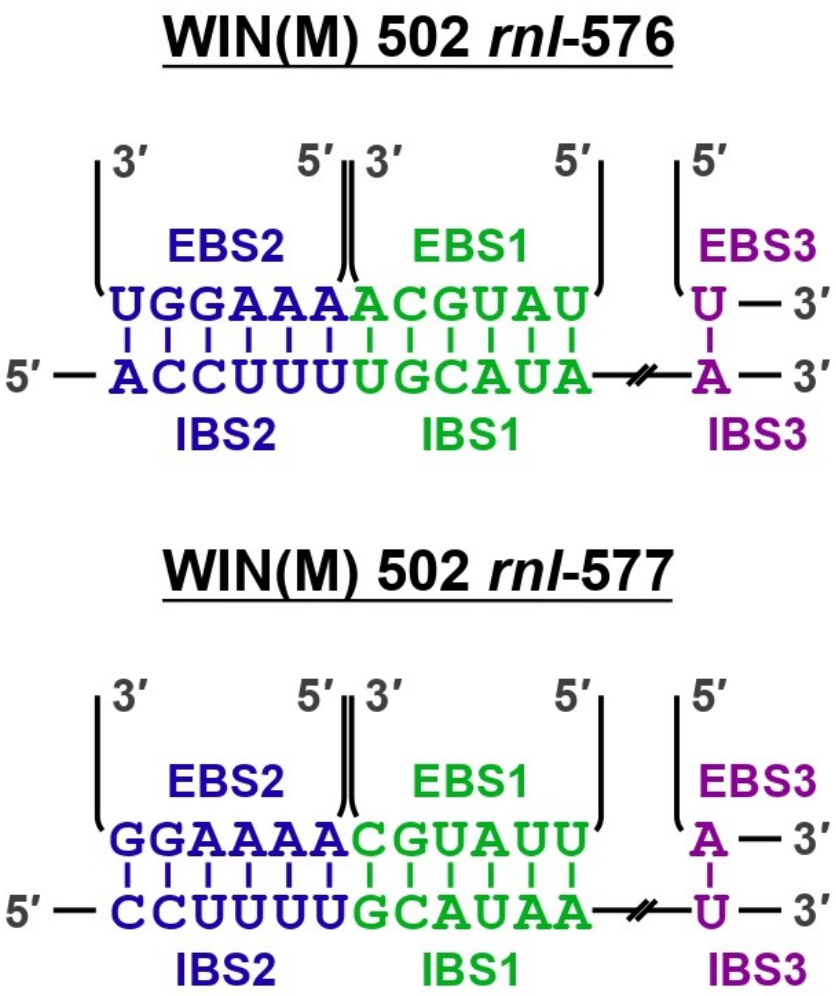

Figure 4.8: IBS-EBS interactions of WIN(M) 502 *rnl*-576 and *rnl*-577. Possible interactions are shown at the folded secondary structure levels, based on domain V identified through RNAweasel.

Phylogenetic groupings observed with mitogenome analysis

The phylogenetic analysis of 65 concatenated mitochondrial protein sequences including 43 species that belong to the Ophiostomatales yielded a topology showing monophyletic groupings for the Microscales, Hypocreales, Glomerellales, Sordariales, and Ophiostomatales (fig. 4.9; suppl. fig. 4.21). Within the Ophiostomatales, several lineages could be identified and labeled representing the following genera: *Ceratocystiopsis*, *Raffaelea*, *Harringtonia, Dryadomyces, Leptographium*, *Grosmannia*, *Esteya*, *Ophiostoma sensu stricto*, *Fragosphaeria*, *Hawksworthiomyces*, *Graphilbum*, and *Sporothrix* (De Beer et al., 2013, 2016, 2022). All members of the Ophiostomatales can be derived from one branch with high levels of confidence (100%) based on ML (MEGA) and MB analyses. Mitochondrial sequences show that species historically assigned to the genus *Raffaelea* resolved into three distinct clades. This supports the currently proposed genera *Harringtonia* (with *Harringtonia lauricola* (*Raffaelea lauricola*) and *D. quercus-mongolicae*), and *Dryadomyces* (inlcuding *D. quercivorus*, and *D. sulphureus*). This is consistent with the findings by De Beer et al. (2022) and Araújo et al. (2022). Otherwise, all remaining genera and species included in the analysis resided in clades consistent with current generic concepts applied to the Ophiostomatales (De Beer et al., 2022).

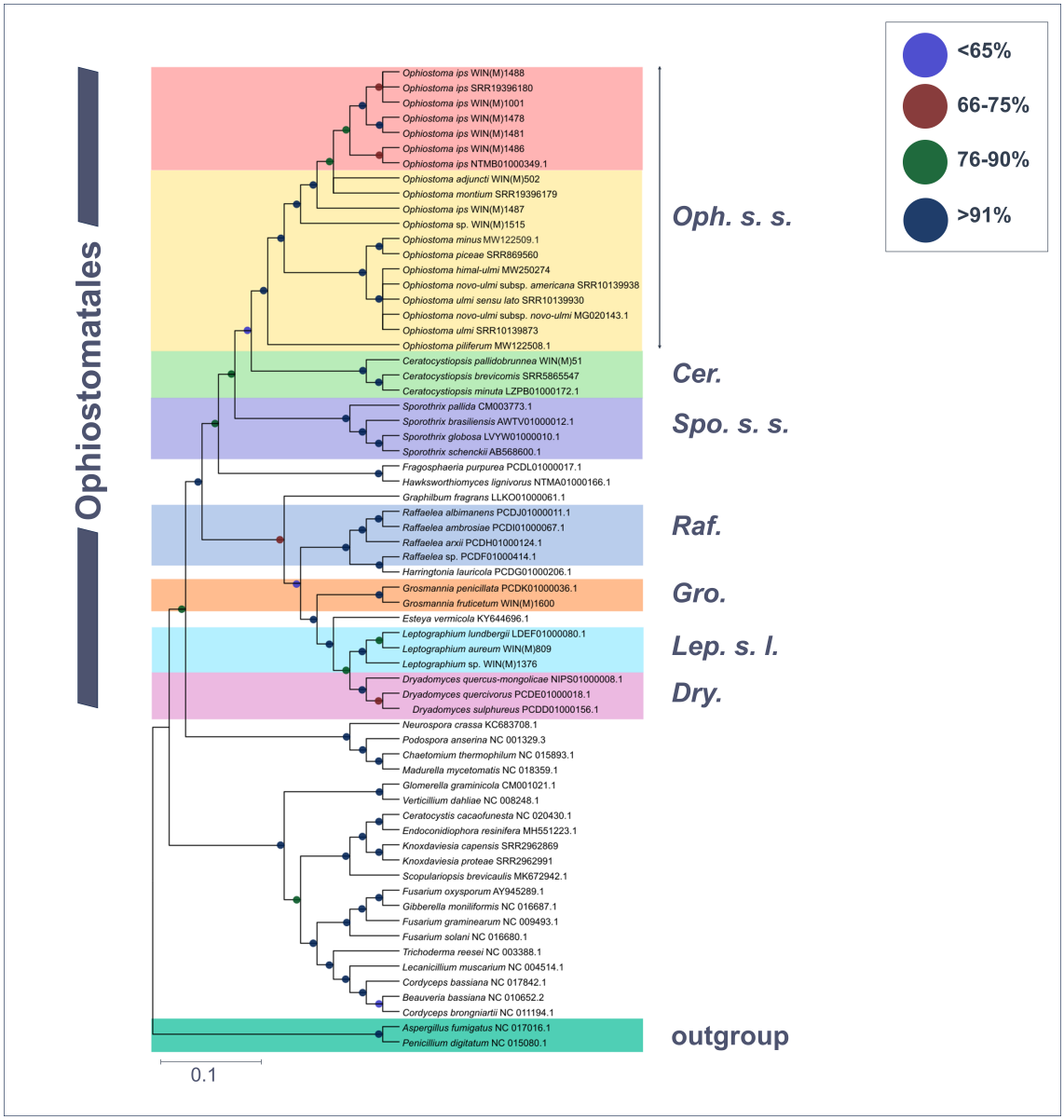
Ophiostomatales
Ophiostoma ips WIN(M)1488
Ophiostoma ips SRR19396180
Ophiostoma ips WIN(M)1001
Ophiostoma ips WIN(M)1478
Ophiostoma ips WIN(M)1481
Ophiostoma ips WIN(M)1486
Ophiostoma ips NTMB01000349.1
Ophiostoma adjuncti WIN(M)502
Ophiostoma montium SRR19396179
Ophiostoma ips WIN(M)1487
Ophiostoma sp. WIN(M)1515
Ophiostoma minus MW122509.1
Ophiostoma piceae SRR869560
Ophiostoma himal-ulmi MW250274
Ophiostoma novo-ulmi subsp. americana SRR10139938
Ophiostoma ulmi sensu lato SRR10139930
Ophiostoma novo-ulmi subsp. novo-ulmi MG020143.1
Ophiostoma ulmi SRR10139873
Ophiostoma piliferum MW122508.1
Ceratocystiopsis pallidobrunnea WIN(M)51
Ceratocystiopsis brevicomis SRR5865547
Ceratocystiopsis minuta LZPB01000172.1
Sporothrix pallida CM003773.1
Sporothrix brasiliensis AWTV01000012.1
Sporothrix globosa LVYW01000010.1
Sporothrix schenckii AB568600.1
Fragosphaeria purpurea PCDL01000017.1
Hawksworthiomyces lignivorus NTMA01000166.1
Graphilbum fragrans LLKO01000061.1
Raffaelea albimanens PCDJ01000011.1
Raffaelea ambrosiae PCDI01000067.1
Raffaelea arxii PCDH01000124.1
Raffaelea sp. PCDF01000414.1
Harringtonia lauricola PCDG01000206.1
Grosmannia penicillata PCDK01000036.1
Grosmannia fruticetum WIN(M)1600
Esteya vermicola KY644696.1
Leptographium lundbergii LDEF01000080.1
Leptographium aureum WIN(M)809
Leptographium sp. WIN(M)1376
Dryadomyces quercus-mongolicae NIPS01000008.1
Dryadomyces quercivorus PCDE01000018.1
Dryadomyces sulphureus PCDD01000156.1
Neurospora crassa KC683708.1
Podospora anserina NC 001329.3
Chaetomium thermophilum NC 015893.1
Madurella mycetomatis NC 018359.1
Glomerella graminicola CM001021.1
Verticillium dahliae NC 008248.1
Ceratocystis cacaofunesta NC 020430.1
Endoconidiophora resinifera MH551223.1
Knoxdaviesia capensis SRR2962869
Knoxdaviesia proteae SRR2962991
Scopulariopsis brevicaulis MK672942.1
Fusarium oxysporum AY945289.1
Gibberella moniliformis NC 016687.1
Fusarium graminearum NC 009493.1
Fusarium solani NC 016680.1
Trichoderma reesei NC 003388.1
Lecanicillium muscarium NC 004514.1
Cordyceps bassiana NC 017842.1
Beauveria bassiana NC 010652.2
Cordyceps brongniartii NC 011194.1
Aspergillus fumigatus NC 017016.1
Penicillium digitatum NC 015080.1
0.1
Oph. s. s.
Cer.
Spo. s. s.
Raf.
Gro.
Lep. s. l.
Dry.
outgroup
<65%
66-75%
76-90%
>91%

Figure 4.9: Phylogenetic relationships of 65 fungal species belonging to the Ascomycota, including 43 members of the Ophiostomatales are presented, based on concatenated amino acid sequences, composed of *atp6*, *atp8*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*. *Aspergillus fumigatus* and *Penicillium digitatum* are selected as the outgroups. Maximum Likelihood (LG + I + F) as implemented in MEGA XI was used to generate the phylogenetic tree using the bootstrap option (1,000 pseudoreplicates) to estimate node support values. A second tree was constructed based on Bayesian inference and the posterior probability (PP) support values from the 50% majority Bayesian (MB) consensus tree (Cprev+F+I model) are represented in combination with the Maximum Likelihood node support values (bootstrap support) on this tree. Branch lengths are proportional to the number of substitutions per site (see scale bar). NCBI/MitoFun accession numbers for each sequence are indicated, and where not available, a strain identifier is used. "*Cer.*" = "*Ceratocystiopsis*", "*Drya*" = *Dryadomyces* (=*R. sulphurea complex*)", "*Est.*" = "*Esteya*", "*Fra.*" = "*Fragosphaeria*", "*Gra.*" = "*Graphilbum*", "*Gro.*" = "*Grosmannia*", "*Har.*" = "*Harringtonia gen. nov.* (=*R. lauricola complex*), "*Haw.*" = "*Hawksworthiomyces*", "*Lep. s. l.*" = "*Leptographium sensu lato*", 698 "*Oph. s. s.*" = "*Ophiostoma sensu stricto*", "*Raf.*" = "*Raffaelea*".

Based on the analysis of the mtDNA derived data we can infer one grouping of *O. ips* strains that includes WIN(M) 1478 (=1480), WIN(M) 1481, WIN(M) 1486, WIN(M) 1488, WIN(M) 1001, *O. ips* NTMB01000349.1, and *O. ips* SRR19396180. However, *O. ips* WIN(M) 1487 appears to branch at a more basal position, prior to a branch that groups *O. adjuncti* WIN(M) 502 with a sequence reported for *O. montium* SRR19396179. A strain listed as *O. ips* WIN(M) 1515 was the most basal member for the "*ips*" complex examined in this study ("*ips*" complex as defined by De Beer et al. (2013 and 2022).

Differences between the *O. ips* strains centered around WIN(M) 1478 groups and WIN(M) 1487 and WIN(M) 1515 are the overall mtDNA intron content and the complex *cob* and *cox3* introns are absent in WIN(M) 1487 and WIN(M) 1515. In addition the strain WIN(M) 1515 has a much smaller genome at 39,957 bp compared to the other *O. ips* strains. As was previously reported, mitogenome sizes do appear to correspond to intron numbers; this is exemplified by WIN(M) 1515 with 12 introns and *Ophiostoma piceae* at 33,599 bp with 6 introns compared with WIN(M) 1478 (=80) at 113,671 bp and 46 introns.

4.1.5 Discussion

Mitochondrial genome organization

The mitogenomes from isolates of *O. ips* and related taxa from various locations (N. America: USA and Canada and New Zealand) show size polymorphism that can be attributed to the difference in their content of introns and intronic ORFs. Similar observations have been noted for other fungal groups where mtDNA variations among members of the same genus are due to introns (Zubaer et al., 2018, 2021; Wai and Hausner, 2021; Mukhopadhyay et al., 2023). Gene order is highly conserved among the Ophiostomatales, with some minor variations observed in terms of the tRNA gene set and the presence/absence of the *atp9* gene (for eg. in the "*ips*" complex, *Ophiostoma s. s.*) which might be compensated by the presence of a nuclear counterpart (Déquard-Chablat et al., 2011; Sellem et al., 2016; Franco et al., 2017; Zubaer et al., 2018; Fonseca et al., 2020, Wai and Hausner, 2021).

Genome size and intron proliferation

Among fungi, large-scale genome rearrangements were often accompanied by accumulation of intergenic and intronic sequences, resulting in the increase in mitogenome size (Mower et al., 2010; Zubaer et al., 2018; Ye et al., 2020). However, there have also been reports of large-scale intron loss events, for example in mitogenomes of Lichen fungi (Pogoda et al., 2019) and ectomycorrhizal fungi belonging to the genus *Boletus* (Li et al., 2021) and this leads us to speculate that the dynamics of mitochondrial genome expansion and shrinkage are complex and can vary substantially among members of different genera or among species within the same genus. In the current study, there seems to be a strong positive correlation between the number of introns and mitogenome size based on a Pearson correlation coefficient (86%; fig. 4.10). This would support that mobile introns and their ORFs are a major factor for generation of mitogenome size variation in fungal mitogenomes.

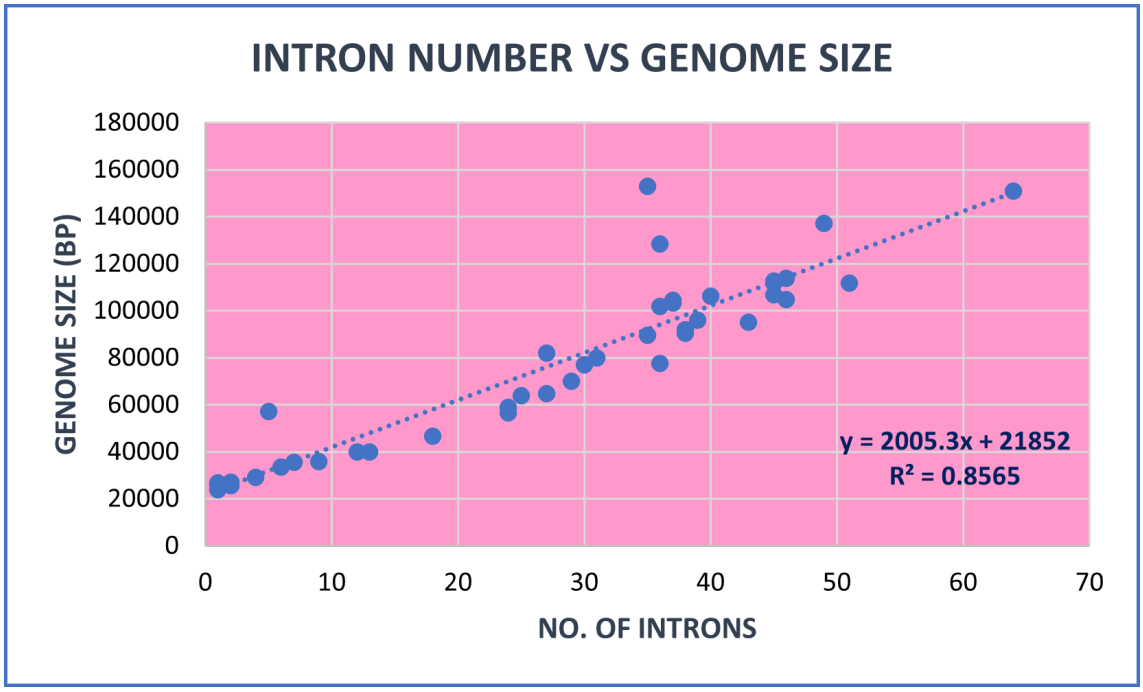

Figure 4.10: Graph depicting the relationship between mitogenome sizes and intron numbers per mitogenome for 43 examined members of the Ophiostomatales. This graph shows a linear relationship with a strong correlation of ∼86% (Pearson Correlation Coefficient of Determination=0.8565) between intron numbers and genome sizes.

Introns can also potentially integrate into new locations ectopically by invading non-cognate intron-less alleles. Once a mobile intron has been inserted into a new target site, its persistence usually relies on drift (neutral evolution) (Goddart and Burt, 1999; Gimble, 2000; Gogarten and Hilario, 2006). Whether an intron is maintained in a population depends on drift vs selection based on the correlation between the energy burden of housing the intron and possible advantages provided by introns such as fine tuning gene regulation (Rudan et al., 2018), functioning as environmental sensors (Belfort, 2017), hypovirulence (Baidyaroy et al., 2011), and interference with the effects of fungicides (Cinget and Bélanger, 2020).

The intron complement for the *O. ips* lineage

In fungi, an abundance of Gr I introns is noted in comparison to Gr IIs; this is in contrast to plant mitogenomes where Gr II introns have a wide distribution (Huchon et al., 2015). In this study, we have identified a total of 416 introns across 63 insertion sites in the eleven mitogenomes, out of which 32 are Gr II introns (among these 6 are part of a complex intron arrangement). Group I introns are observed across 55 insertion sites and group II introns across 6 sites There appears to be a bias towards Gr I introns in the examined mitogenomes. This leads to the speculation that the process of homing might be more successful in fungal mitogenomes in comparison to retrohoming, and Gr II introns might favour invading ribosomal genes (*rns*, *rnl*) or the N-terminal region of the *cox1* gene, that shows high sequence conservation. Some Gr II introns insert into Gr I introns for their persistence as part of complex intron. Encoding HEs over RTs (such as by gaining and using LAGLIDADGs such as *rns*-952 and *rnl*-2060) might be a way in which some Gr II introns can adopt a DNA based homing strategy (Pfeifer et al., 2012).

Novel and complex intron arrangements

In the context of a complex intron, the word 'twintron' was initially used to describe Gr II introns inserted within other Gr II or Gr III (derived from Gr II introns) introns; later other variations were identified, such as Gr II in Gr I, or Gr I within other Gr I, and lariat capping twin ribozyme introns (Hafez and Hausner 2015). Composite introns might expand by gaining additional ORFs (such as HEGs) or intron modules (Guha et al., 2018), and there are instances of internal (non-fused) ORFs being expressed via alternative splicing (Turk et al., 2013; Wai et al., 2019) that could lead to splicing mediated 'core-creep' (Edgell et al., 2011; Guha et al., 2018) where a transcript is generated that fuses intron ORF to the upstream exon (Guha et al., 2018; Wai et al., 2019; Mukhopadhyay and Hausner, 2021). In some instances, "secondary invaders" can insert in-frame to the resident intron ORF sequences that are already fused to the upstream exon. A possible example of this core creep was observed in a *cox1* complex intron in *Leptographium* sp. WIN(M) 1376, where a LAGLIDADG ORF fused in-frame with a resident GIY-YIG ORF and the upstream exon, probably taking advantage of this "pseudo-exonization" for more efficient translation of the ORFs (Mukhopadhyay et al., 2023). Some notable novel or complex intron arrangements in this study are described in the following sections.

i. The *cox3*-640 'zombie' twintron

The *cox3*-640 twintron consists of two Gr I intron modules in tandem that appear to splice as a composite unit based on RT-PCR results and RNA-Seq, and the downstream intron module was presumed to be the original resident (native) intron (Zubaer et al., 2021). As alternative splice forms could not be detected through RNA-Seq mapped reads, it is proposed that the complex intron splices out as a single composite unit or "zombie" [as suggested by Zumkeller et al. (2020)] and this might be an energetically favorable cellular process. Tandem type complex introns, such as *O. ips cox3*-640, have been observed and described in *Ceratocystiopsis brevicomis*

(PCDN01000199.1) and *Grosmannia penicillata* (PCDK01000036.1) (see Zubaer et al., 2021 supplementary data).

ii. The *cob*-490 trintron utilizes a 'ratchet' pathway

An alternative splicing route was modeled as the "ratchet" pathway by Zubaer et al. (2021) for the *cob*-490 complex intron. This model supports both the "tandem" (side-by-side) and "nested" (upstream intron fully contained within the downstream/resident intron) arrangements since: (i) complete intron pairings exist in the "tandem" arrangement and all pairings of the "upstream" introns can be modeled to to be contained (nested) within the P1 loop of the resident "downstream" intron; (ii) both "upstream" and "downstream" introns have their own P1' elements, but they can share the same P1 interaction.

In the proposed "ratchet" pathway for *cob*-490, the first intron module is removed in a primary step, generating an intermediate RNA molecule that regenerates suitable sequences for the second intron module to assume a splicing competent RNA fold including P1 and P10 interactions that allow for its removal and joining of the flanking exons. Splicing of the upstream intron component could generate a transcript whereby the downstream located ORF is fused in frame with the upstream exon, optimizing the expression of the downstream intron-encoded LAGLIDADG protein, a scenario referred to as splicing-mediated core creep (Guha et al., 2018) where transcripts are generated that fuse the downstream located ORF sequence with the upstream exon.

The ratchet model is based on the assumption that the resident Gr I intron utilizes the pseudoexon (intronic sequences utilized as "temporary exons" for splicing) for its P1-P10 interactions. By using this sequence which is identical to the first six nt of the downstream exon (see fig. 10; Zubaer et al., 2021), the upstream and downstream introns splice out sequentially. The P1 and P10 helices are essential for the intron module in aligning intron/exon splice junction sequences so that the introns can be spliced out (and by default exons are spliced together). The two Gr I

intron modules of the complex intron belong to the same subtype IA and the helices of these two introns could potentially interact in multiple combinations. Alternately, the resident Gr I intron can directly interact with the downstream exon for P10 formation which would lead to the entire complex *cob*-490 intron to be spliced out as a single unit (similar to the observation for the *cox3*-640 twintron).

However, it has to be noted that the splicing process is assumed to be spontaneous and unless the splicing intermediates can be isolated *in vitro* by synchronizing fungal cultures, it might be difficult to detect physical evidence of the rare splice pathways and intermediates. Filamentous fungal cultures are difficult to synchronize and control time range-specific experiments are difficult to conduct for a non-model fungal species such as *O. ips.* Based on the IGV visualization of mapped RNA-Seq reads and observation of clipped bases, we have found reproducible evidence for transcripts mapping to the pseudoexon sequence 'TCATAT' associated the upsteam exonic sequences (the last 8 nt on the 3' end being 'CGTTGAGT'). This arrangement represents the "intermediate splicing step" from the ratchet model where the upstream intron has been spliced out leading to the fusion of the upstream exon with the pseudoexon sequence which results in a splice competent fold leading to the excision of the downstream intron. Since these sequences were reported for two independent RNA-Seq runs using biological replicates WIN(M) 1478 and 1480, we concluded that these were not simply sequencing artifacts. Overall, the data suggests that it is difficult to determine which splice route is followed ("zombie" like in *cox3*-640 or "ratchet" like in some instances for *cob*-490) since the end products of the "zombie" and "ratchet" splicing pathways are the same. RNA enrichment by mitochondria isolation combined with Nanopore long read sequencing might have been able to capture the full mitochondrial transcriptome in the model fungi, as was done for example in *S. cerevisiae* (Putzeys et al., 2022, 2024), and allowed the analysis and visualization of splicing variants with single base pair resolution, and the quantification of the abundance of RNA species at different processing stages. How-

ever, its reproducibility in a non-model filamentous fungal system remains highly debatable. For now, we can rely on our evidence of alternative splice forms and the existing knowledge of similar splicing patterns demonstrating the "plasticity" of intron RNA folds that have been previously observed (Sellem and Belcour, 1994; Turk et al., 2013).

iii. The *cob*-201 complex intron with two group IB modules

Two Gr IB modules were identified at the *cob*-201 position in WIN(M) 1487. According to the MFannot annotation, this complex arrangement consists of two Gr IB introns which could also be interpreted as two independent intron modules at the positions 201 and 219 (relative to the *cob* sequence in the annotation reference, *T. inflatum*). The 18 nt "intervening sequence" separating the two intron modules could potentially represent a pseudoexon that is required for the splicing of the two intron modules and eventually it is removed during the processing of the *cob* transcript. The duplication of this 18 nt sequence could have been introduced by the invading intron module as homing endonuclease mediated mobility can also modify flanking sequences due to gene conversion.

iv. The *rnl*-576 and *rnl*-577 intron cooperating "neighbors" separated by a single-nucleotide exon

Two Gr II intron modules were identified at *rnl*-576 and *rnl*-577 positions in WIN(M) 502. In this neighboring-introns arrangement, a single 'A' nucleotide serves as the exon to separate the two introns and an RT is encoded by the ORF housed in the second (downstream) intron (*rnl*-577). Based on the conservation of the introns, one potential scenario could be that *rnl*-577 inserted prior to the insertion of the ORF-less *rnl*-576 that might have been inserted into the IBS of *rnl*-577 by reverse splicing. One can speculate that, later in time, the insertion of *rnl*-576 interfered with the splicing of *rnl*-577, by which the downstream intron has now become dependent on the splicing of *rnl*-576, prior to splicing of itself. This scenario is reflected by secondary structure modeling (fig. 4.8) which shows that the second

intron has still maintained its "native" exon binding sequences (EBS1 and EBS2, which form approximately the last 12 nucleotides of the 5′-exon); the corresponding intron binding sequences (IBS1 and IBS2) for the second intron is restored after splicing of the first intron. The splicing reaction resembles a ratchet-like arrangement as seen in *cob*-490 as one intron module splices, that event "generates" the splice site for the other intron module. However, here there is no exon separating those intron modules. In *rnl*-576 and *rnl*-577 the EBS1 for the *rnl*-577 intron seems to interact with the 'A' exon plus the upstream exon, and thus, the second intron seems to have its IBS sequences shifted by one nucleotide compared to the first intron. This would support the existence of the 'A' exon.

Previously an example of two *rnl* Gr II introns separated by an "A" single nucleotide exon was described in two *Ceratocystis* species, *C. lukuohia* and *C. huliohia*. The two elements referred to as SPAM 3 and SPAM 7 (Mayers et al., 2021) are possible orthologs of the *rnl*-576 and *rnl*-577 introns observed in *O. adjuncti* WIN(M) 502. The authors proposed a tandem splicing mechanism that is essentially the same as what we propose for the *rnl*-576 and *rnl*-577 introns. Related introns have been observed in *Ophiocordyceps sinensis* KP835313.1, *Graphilbum pis-grandicollis* sp. *nov* JADHKH000000000.1 and *Graphilbum rectangulosporium* JADHKI010000081.1, but only the downstream intron *rnl*-577 is present, whereas in *Grosmannia penicillata* MLJV01000071.1, only the *rnl*-576 can be identified (see suppl. alignment). Both of these introns compete for the same/similar *rnl* insertion sites based on compatibility of IBS/EBS sequences.

Examples of micro-exons in mitogenomes are scarce; short 3 bp exons were previously reported in fungi (Xavier et al., 2012), a single-bp microexon has been noted in the mtDNA cox1 gene for a placozoa metazoan (Osigus et al., 2017). Micro- and mini-exons have been extensively studied with regards to spliceosomal introns in plants (Guo and Liu, 2015; Li et al., 2015); however, few examples have been reported in fungal mitochondrial genomes. This could potentially be attributed to

misannotation, since standard annotators adopting a statistical model such as Hidden Markov Model (HMM) that relies on a training dataset and a minimum of three nucleotide based codons, do not favor microexon prediction (Yu et al., 2022). Lang et al. (2023) emphasized on the shortcomings of MFannot and HMM-specific approach in detecting mini-exons and the need for improved and robust algorithms for correct identification of short N-terminal exons and mini-exons. Further investigation of mitogenomes among various genera of fungi could facilitate discovery of novel micro-exons and their role in splicing of complex introns or intron arrangements.

Phylogenetic analysis of mitogenomes

In this study, we examined mitogenomes for species that can be assigned to the various genera of the Ophiostomatales. Historically due to overlapping and sometimes conflicting morphological characters this group of fungi has been challenging with regards to generic assignment and species identification (Upadhyay, 1981; Zipfel et al., 2006; Plattner et al., 2009). The data set applied in this study consisted of 65 sequences and the tree topology confirmed the currently proposed generic lineages proposed for members of the Ophiostomatales (De Beer et al., 2022). The data set also showed the monophyly of strains belonging to *O. ips* and taxa related to *O. ips*, such as *O. adjuncti* and *O. montium* (Upadhyay 1981). The data also suggests that *O. ips* complex/lineage shares a common ancestor with other members of Ophiostoma *sensu stricto* such as *O. piceae*, members of the *O. ulmi* species complex and *O. piliferum*.

Ophistoma ips strains from North America and New Zealand all grouped together with two exceptions, strains WIN(M) 1487 and WIN(M) 1515. With regards to WIN(M) 1515 it was previously noted by J. Reid (personal communication) that the ascospores were larger than that described for *O. ips,* although the conidial states for WIN(M) 1515 appeared to match descriptions for *O. ips*. The mitogenome phylogenetic analysis along with its significantly smaller mitogenome appears to

suggest it should be referred to as "*ips*-like" *Ophiostoma* sp. and more studies are needed to assign it to a species. The position of WIN(M) 1487 is surprising. This isolate deposited as CBS 137.36 by Rumbold (original authority for *Ophiostoma ips*) has been designed by CBS (Westerdijk Fungal Biodiversity Institute) as the authoritative (AUT) strain (the original Type or ex-Type for *O. ips* are lost). A previous study showed that based on nuclear markers and ability to grow at 35°C the strain CBS 137.36 belongs to *O. ips* and can be distinguished from *O. montium* (Kim et al., 2003). In most fungi, mitochondria are uniparentally inherited but on rare occasions segments of mitochondrial genomes can move horizontally between different fungal species (see Mayers et al., 2021). The similarity of the *cob*-201 intron and the 18 nt segment in the downstream exon in the CBS 137.36 with *O. montium* may hint at a possible hybridization/introgression event between these two species (or other members of the "*ips*" lineage). Hybridization and introgression events have been recorded for members of *Ophiostoma ulmi* and the subspecies of *Ophiostoma novo-ulmi* (Stukenbrock 2016; Hessenauer et al., 2020). In future studies more strains of *O. ips* and *O. montium* need to be examined at both the nuclear and mitogenomic level to provide evidence for this speculation.

4.1.6 Concluding Remarks

Comparative mitogenome analysis of fungal members of the Ophiostomatales is an ongoing effort to understand the size variation, phylogenetic groupings, and novel and complex intron arrangements. Mitochondrial introns might also be important therapeutic targets against fungi; for example, taking advantage of auto-splicing Gr I introns linked to susceptibility to aminoglycosides like pentamidine (reviewed in Gomes et al., 2024). Thus mapping of introns is a resource for keeping track of intron positions. A comprehensive look into the splicing pathways of these complex introns, and how that affects the translation of their encoded ORFs might require greater depth of sequencing through long-read technologies, combined with *in vitro*

splicing assays. Putzeys et al. (2022, 2024) recognized the challenges with short-read RNA-Seq technology that fail to capture key transcriptional events in phages and presented a Nanopore-based long-read strategy that could generate a full-length transcriptional map and help to uncover complex transcriptomes (Putzeys et al., 2022). This study showed the potential of using this long-read strategy to identify Gr I intron splicing events in phages, which could be extended to other prokaryotic systems, but their success in resolving RNA processing events in filamentous fungal systems is yet to be explored. Another setback of RNA-Seq for mitochondrial transcriptomes is the fact that until now, poly(A) tails are assumed to be absent from mtDNA transcripts and enrichment of libraries through poly(A) selection could not be performed for the fungal strains in this study. Koster et al. (2023) used Nanopore long-read sequencing for better data resolution when looking into splicing events in *S. cerevisiae*, combined with enzymatically added poly(A) tails to the natively non-polyadenylated mitochondrial RNA samples. If this strategy could be successfully applied to filamentous fungi, it could significantly reduce the "noise" from rRNAs and tRNAs. Finally, the phylogenetic groupings of the fungal members of the Ophiostomatales can be refined by combining analyses using nuclear and mitogenomic sequences and a constantly expanding repertoire of available fungal whole genome sequences and high quality assemblies.

Data availability statement

The datasets presented in this study have been submitted to online repositories as following:

GenBank accessions for submitted mitogenome sequences:

O. ips WIN(M) 1478: PQ043840

O. ips WIN(M) 1488: PQ043838

O. ips WIN(M) 1487: PQ043837

O. ips WIN(M) 1486: PQ043836

O. ips WIN(M) 1481: PQ043835

O. ips WIN(M) 1001: PQ043834

O. ips WIN(M) 1480: PQ043832

O. adjuncti WIN(M) 502: PQ043833

Ophiostoma sp. WIN(M) 1515: PQ043839

RNA Sequencing Reads can be accessed at:

Bioproject PRJNA1143047 (includes biosample SAMN42992279 and SRA SRX25593433)

https://www.ncbi.nlm.nih.gov/sra/PRJNA1143047

Supplementary material

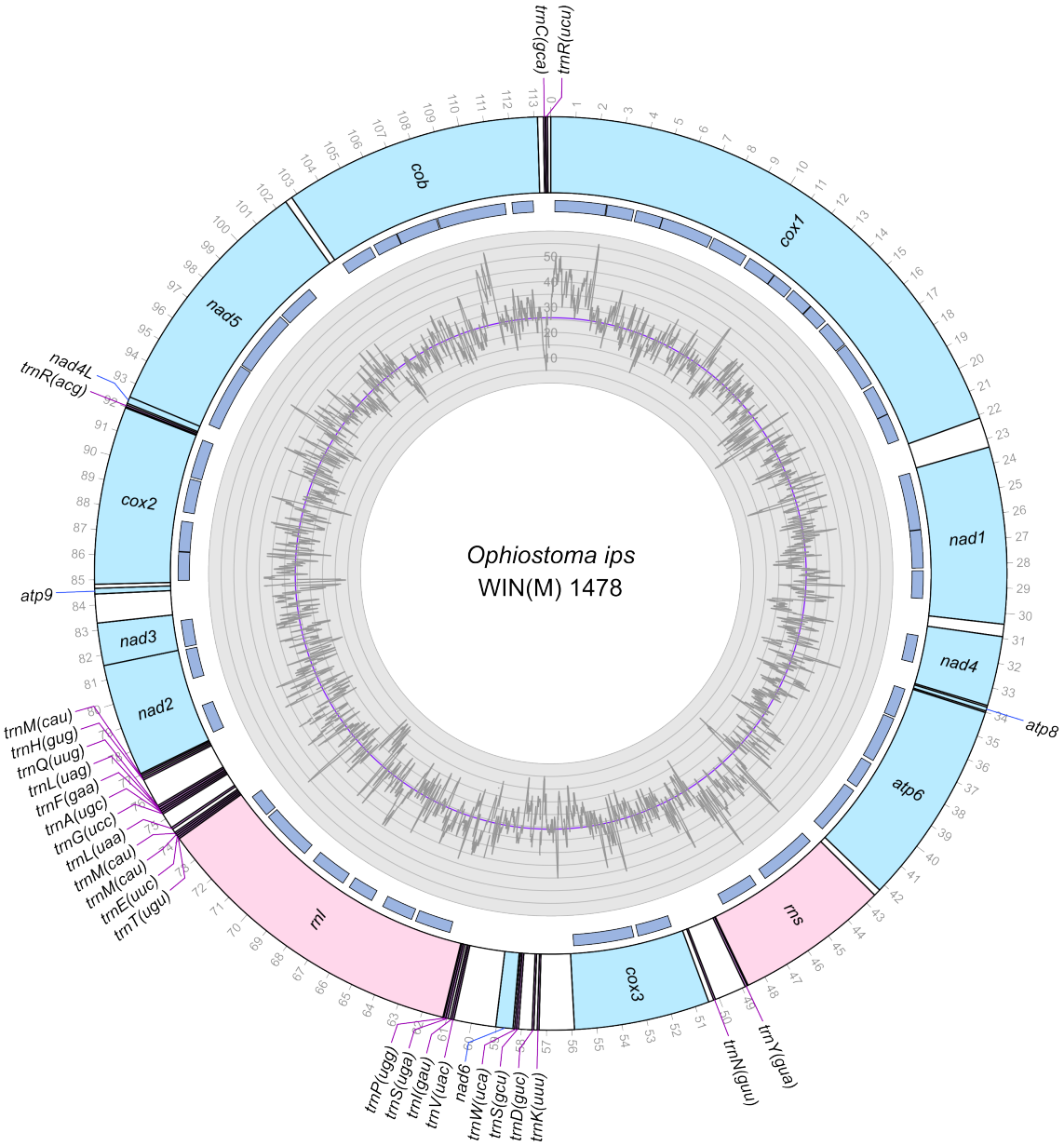

Figure 4.11: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1478. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

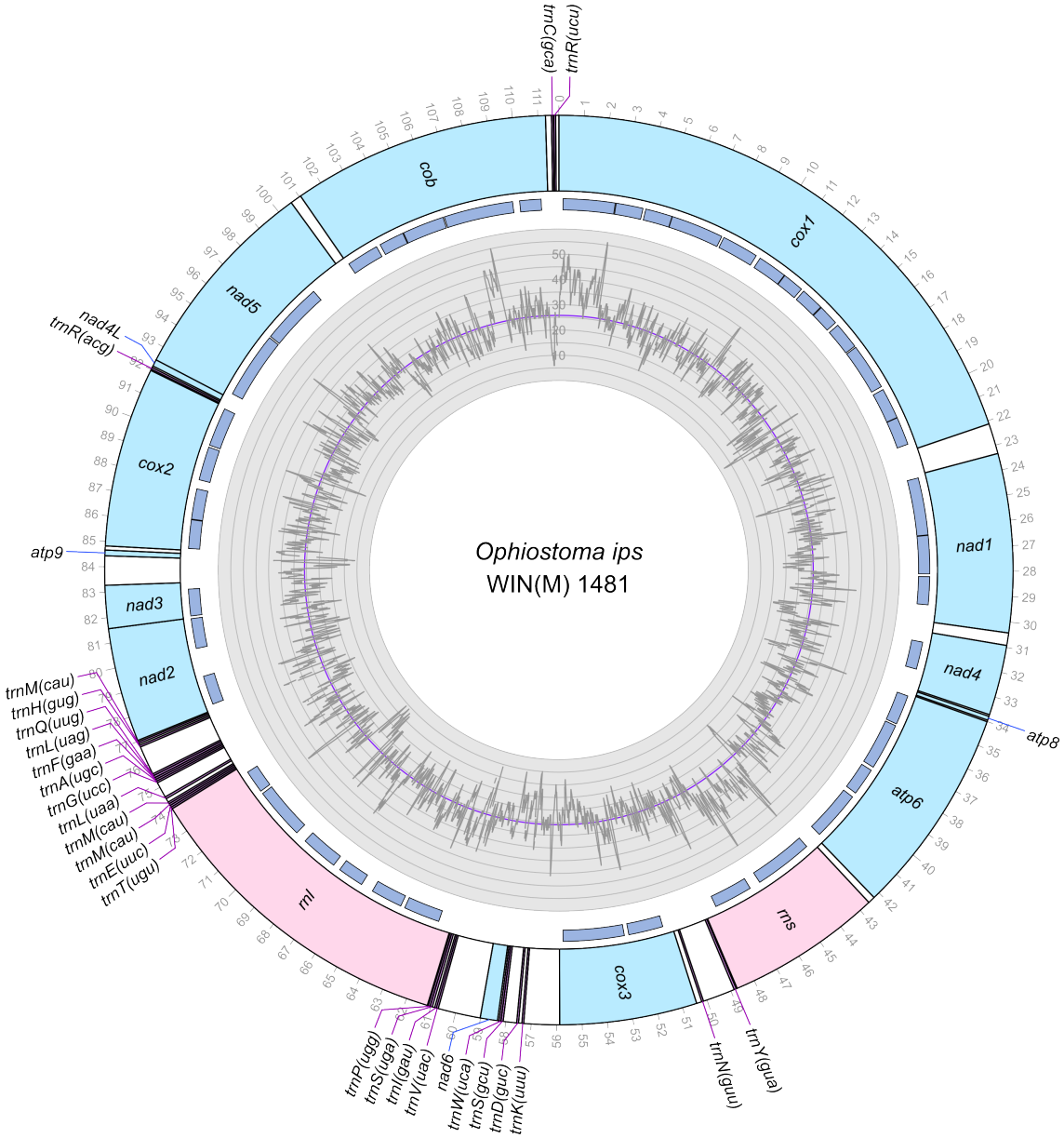

Figure 4.12: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1481. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

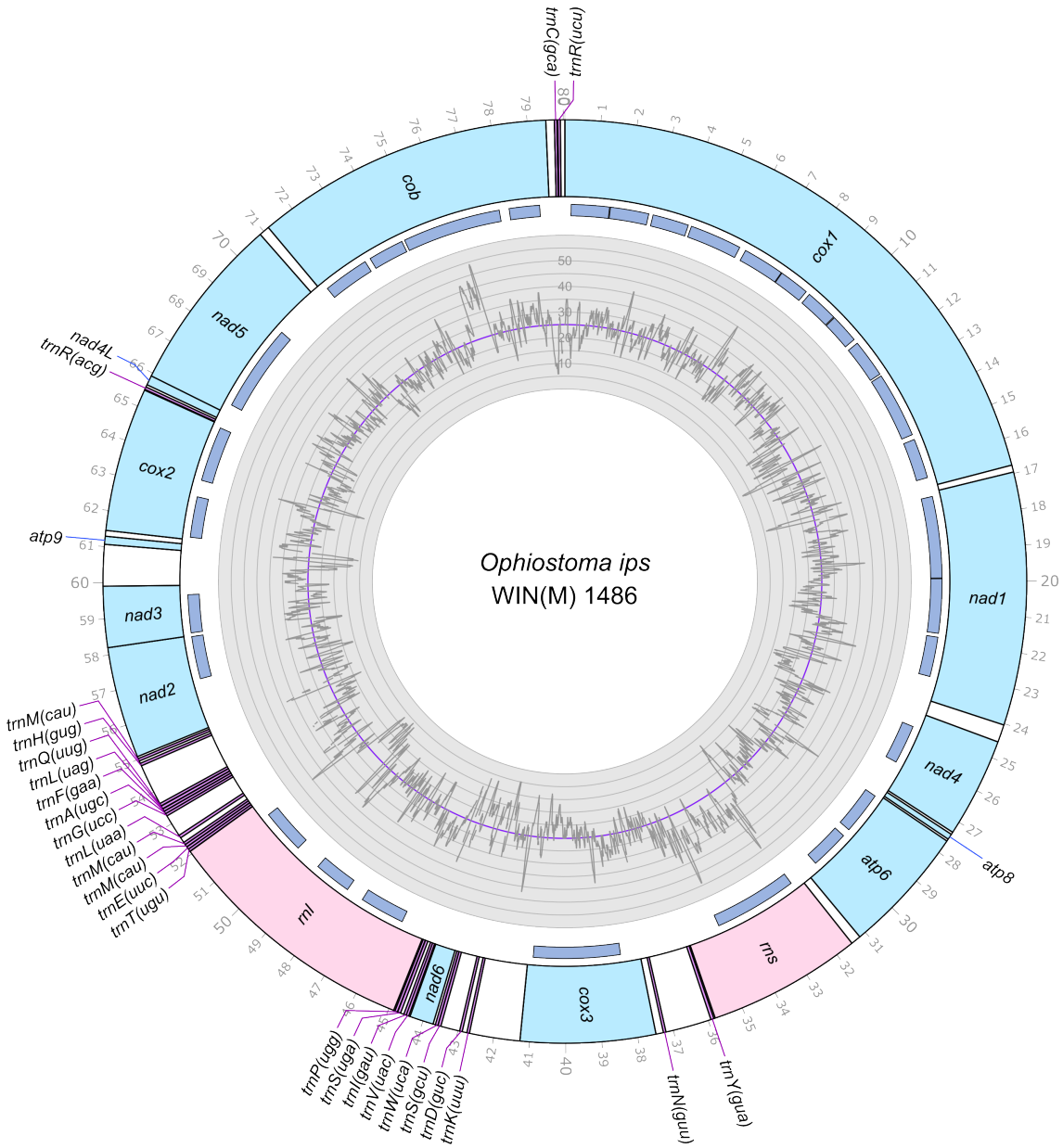

Figure 4.13: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1486. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

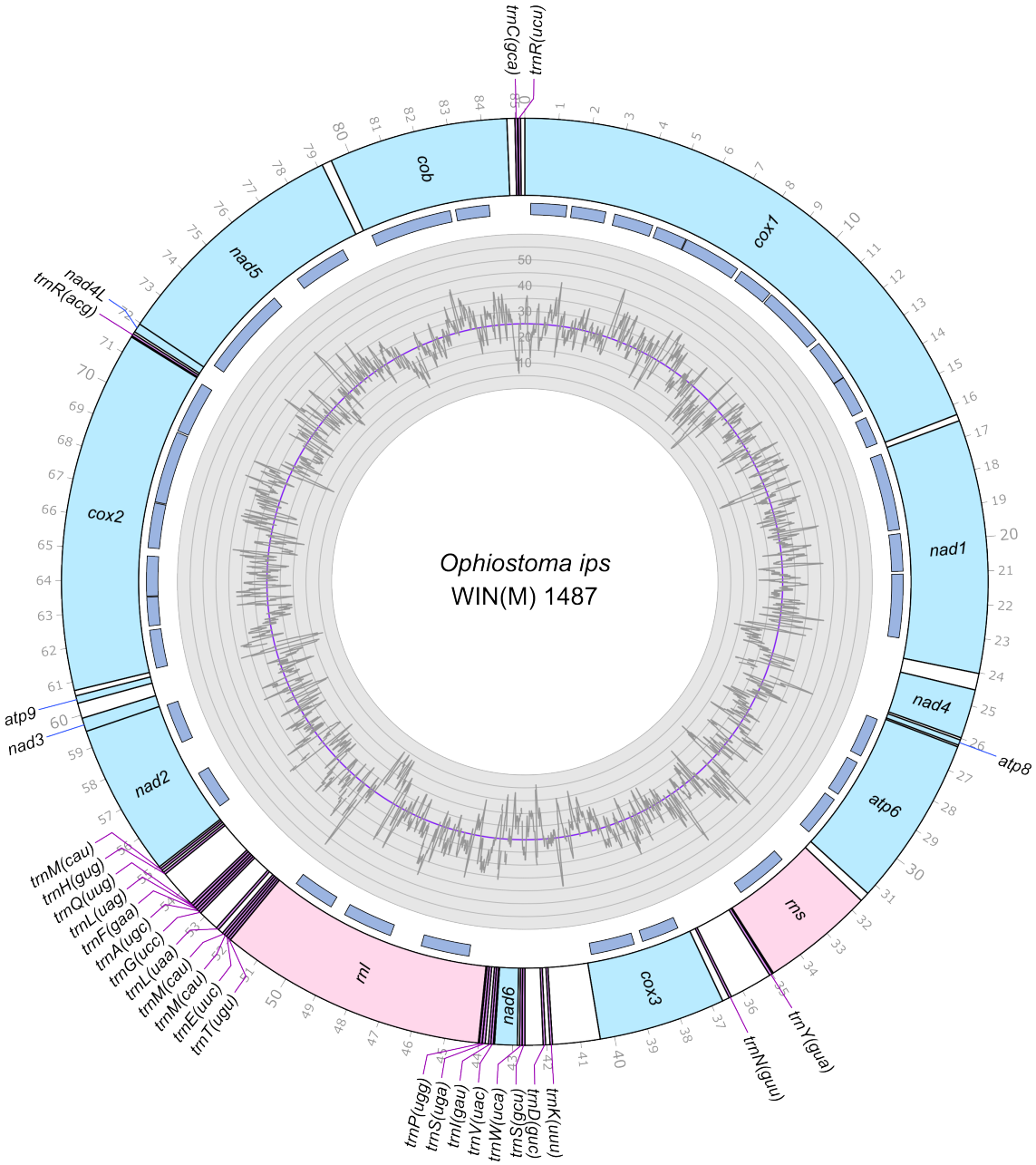

Figure 4.14: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1487. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

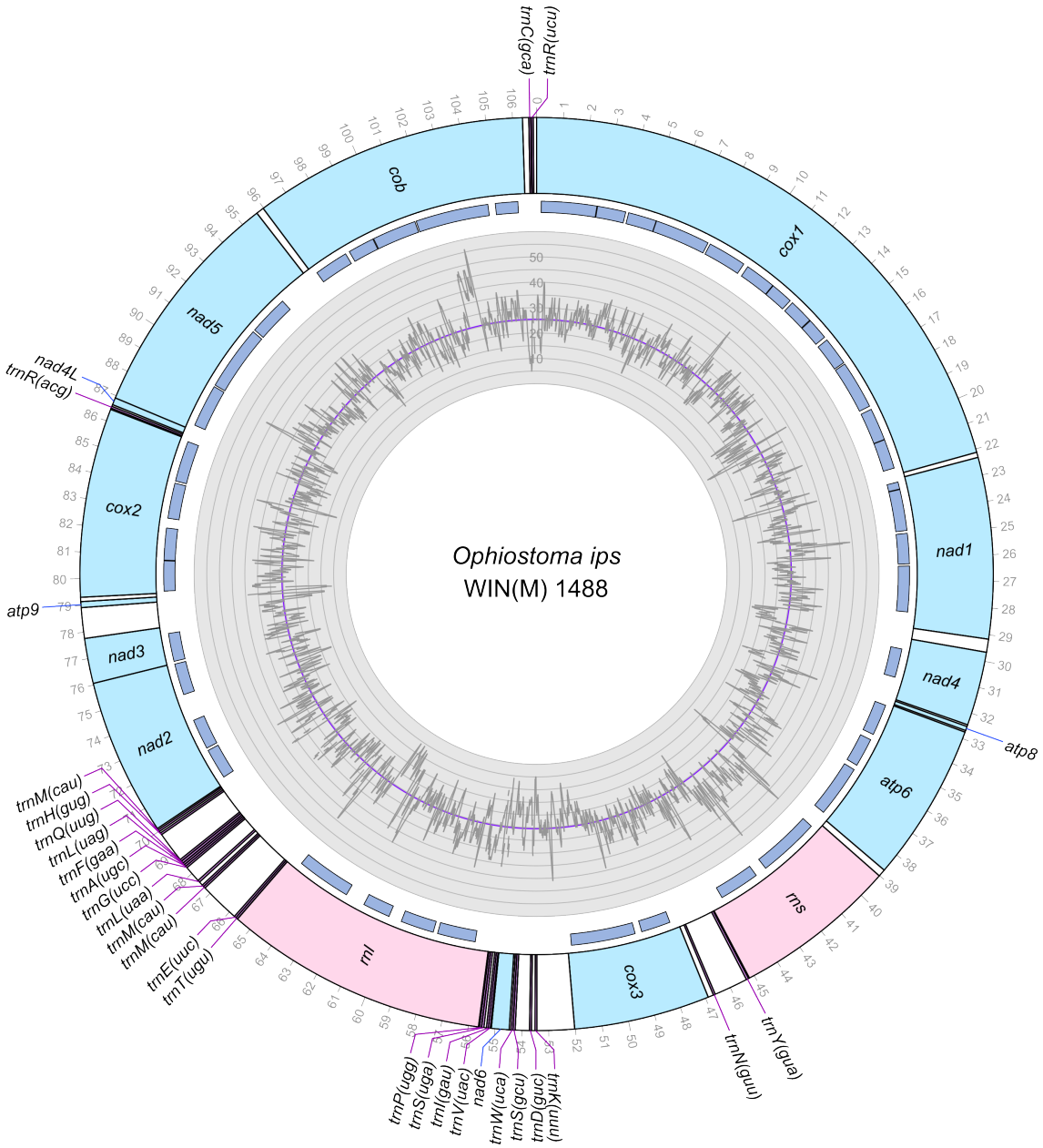

Figure 4.15: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1488. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

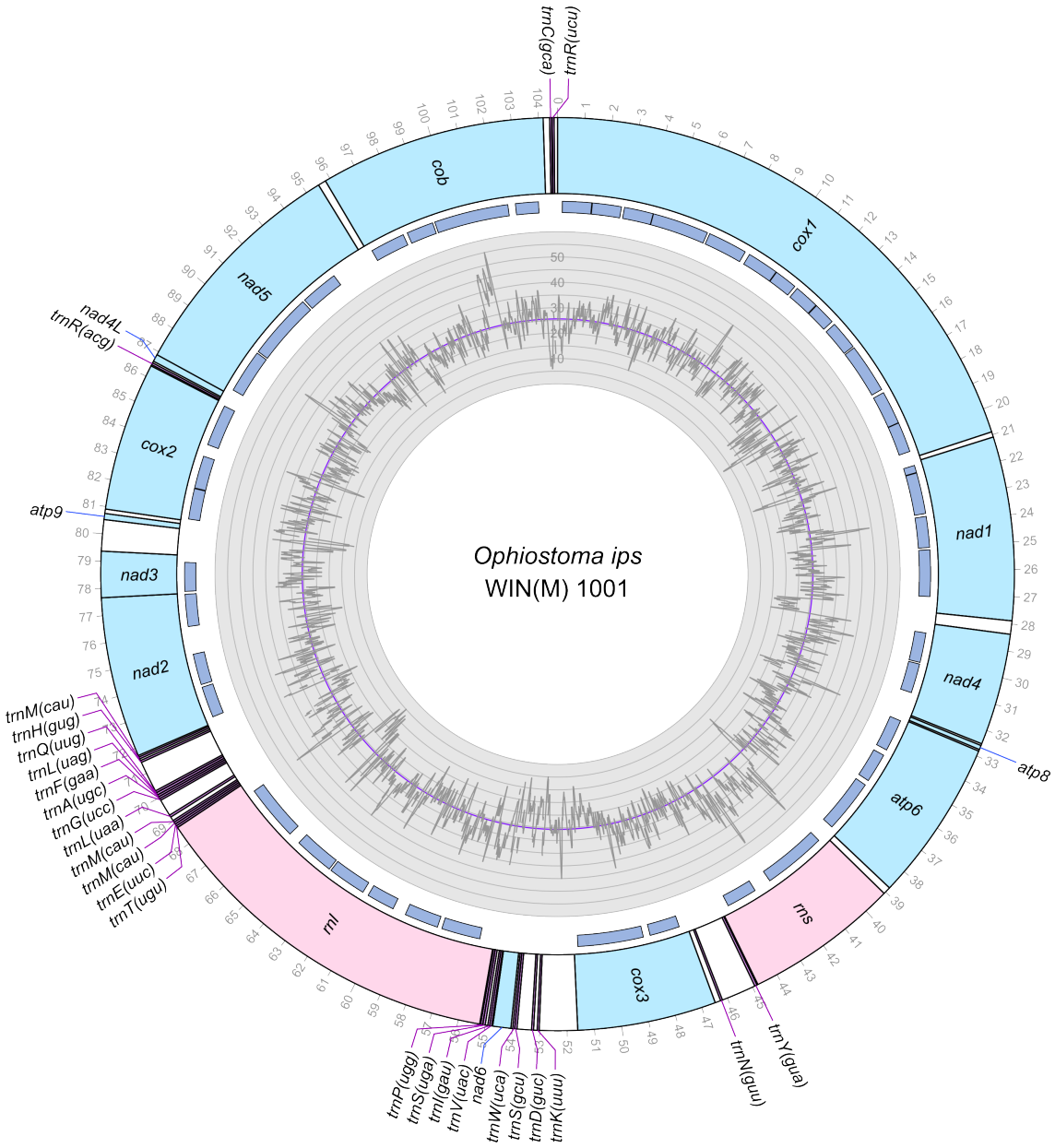

Figure 4.16: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1001. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

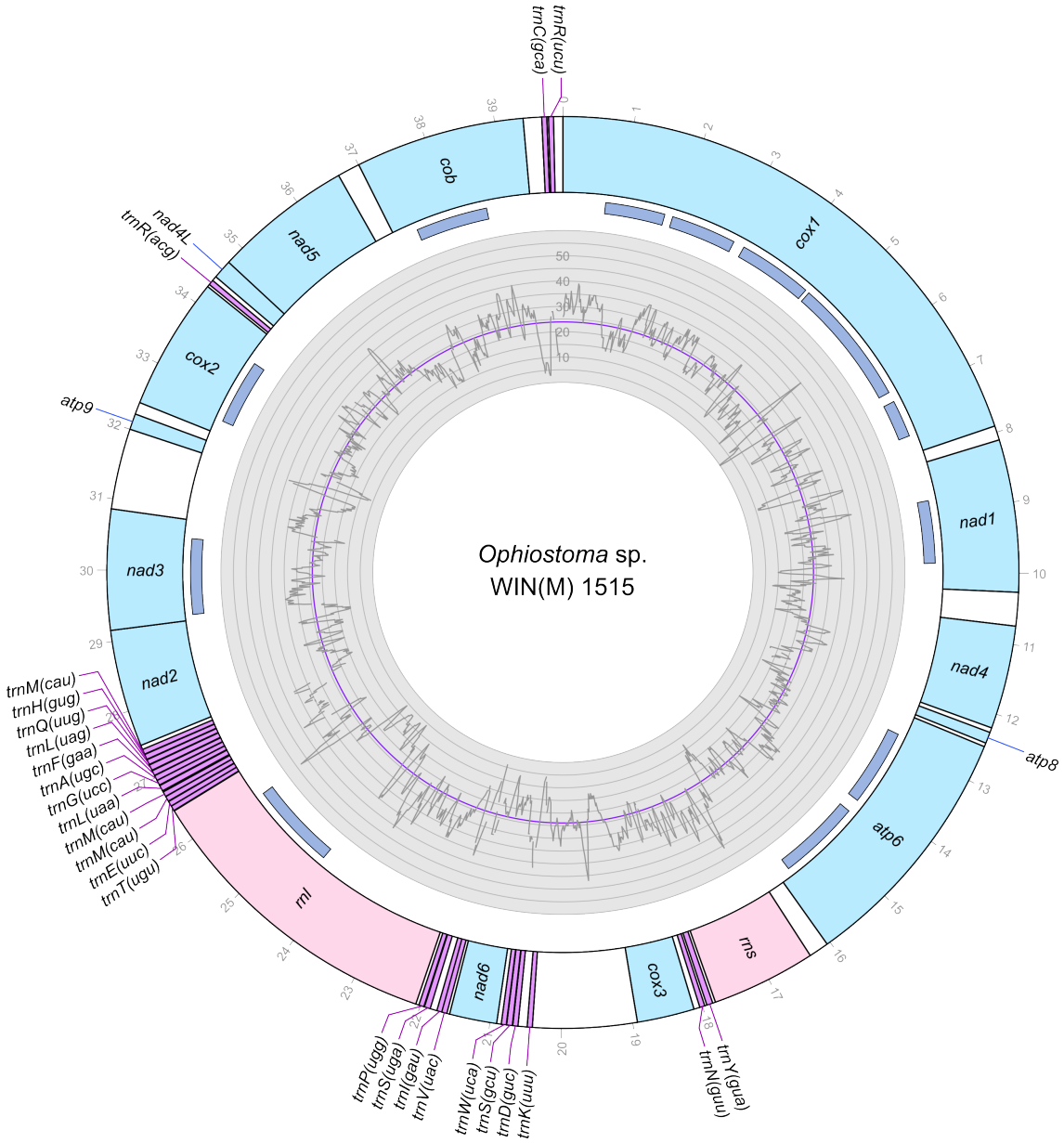

Figure 4.17: Circular representation of the mitochondrial genomes of *O. ips* WIN(M) 1515. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

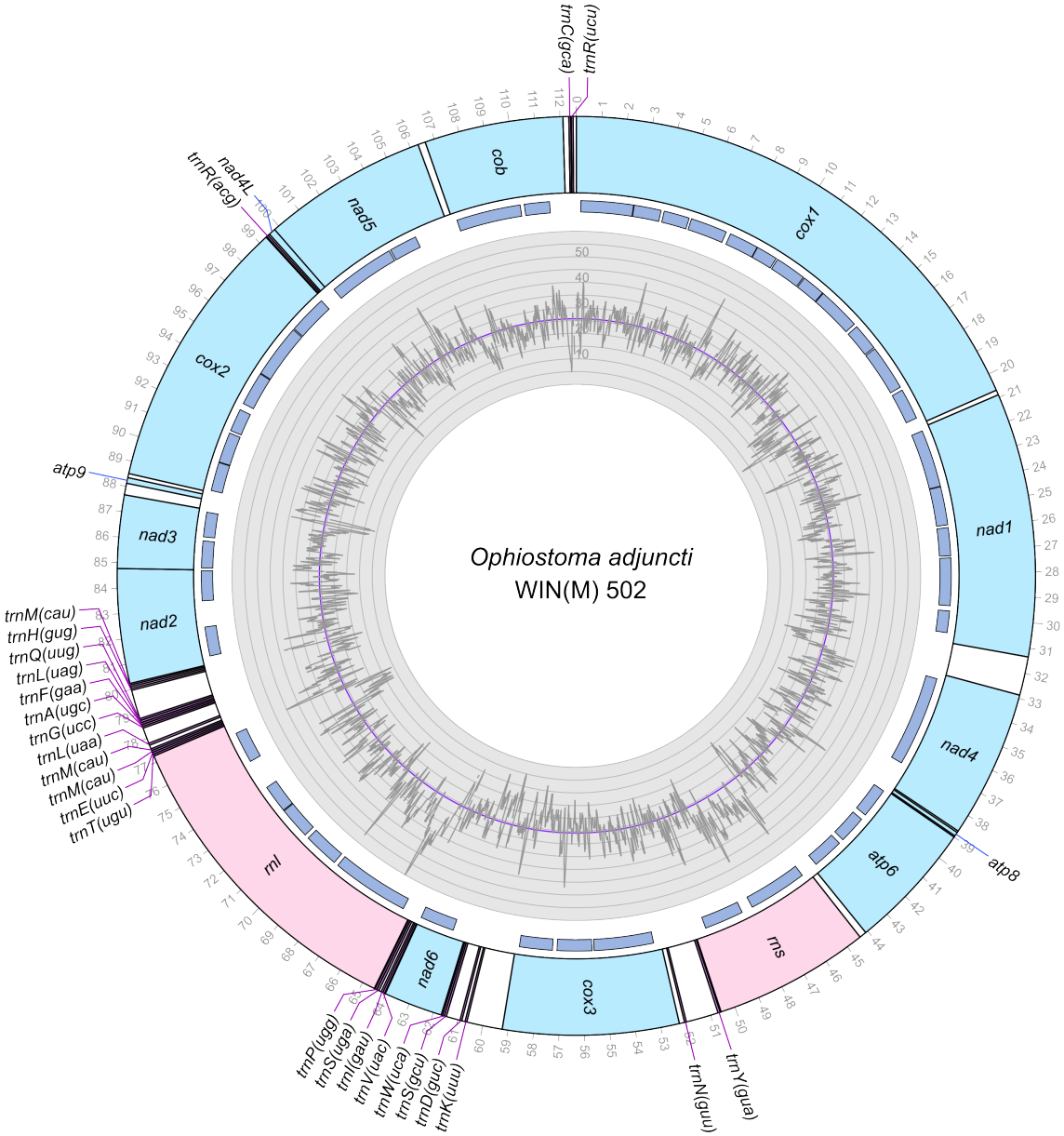

Figure 4.18: Circular representation of the mitochondrial genomes of *O. adjuncti* WIN(M) 502. Genes, introns, and GC plot are shown on the outer, middle, and inner tracks, respectively. The purple line of the GC plot corresponds to the average GC content of the mitochondrial genomes. The tick marks on the outer track label every 1,000$^{\text{th}}$ nucleotide, starting from the beginning of the *cox1* gene. All labeled genes are encoded on the same strand.

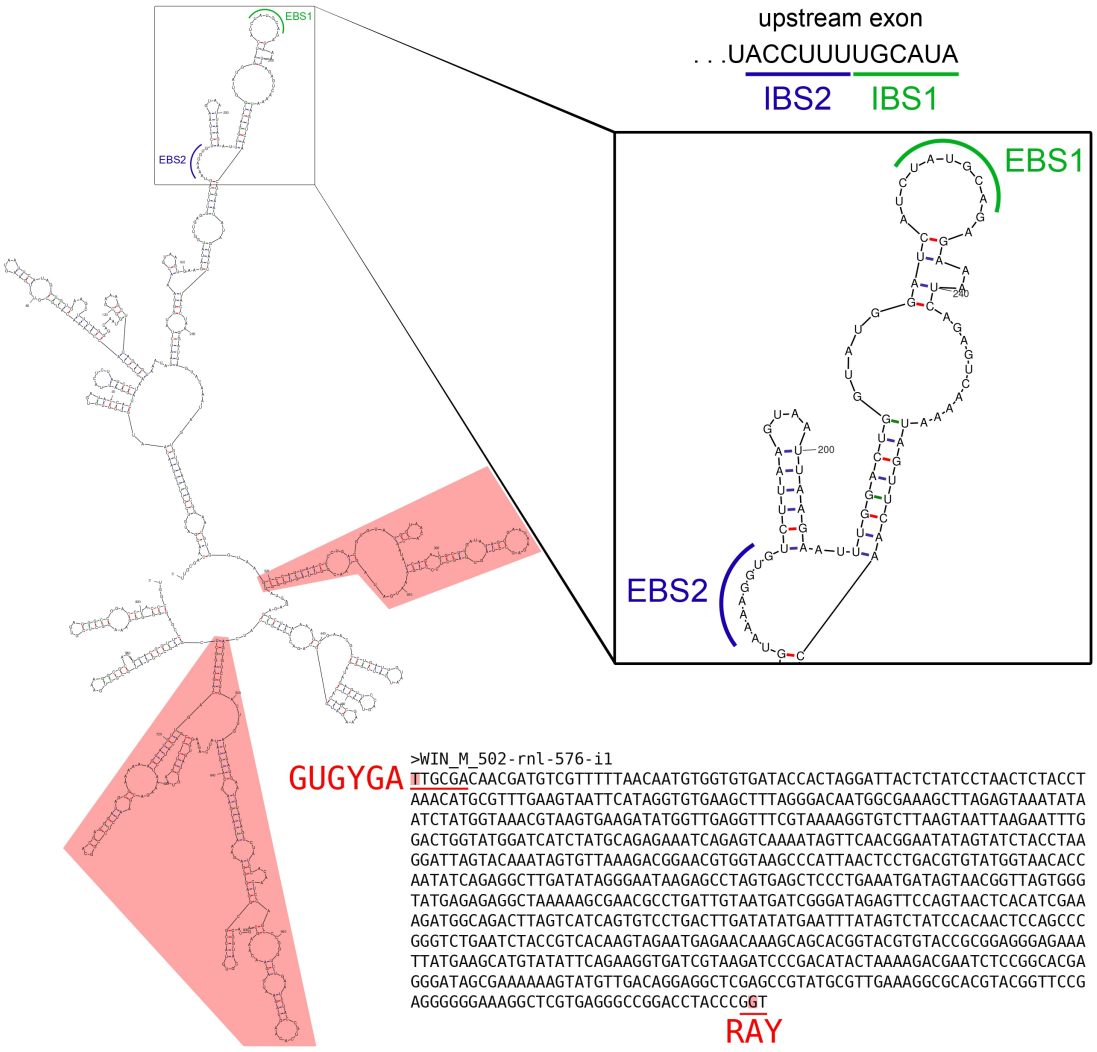

Figure 4.19: mfold model of *rnl*-576 in WIN(M) 502. Domains II and IV are highligted in red. It is indicated that *rnl*-576 does not start or end with typical, conserved nucleotides

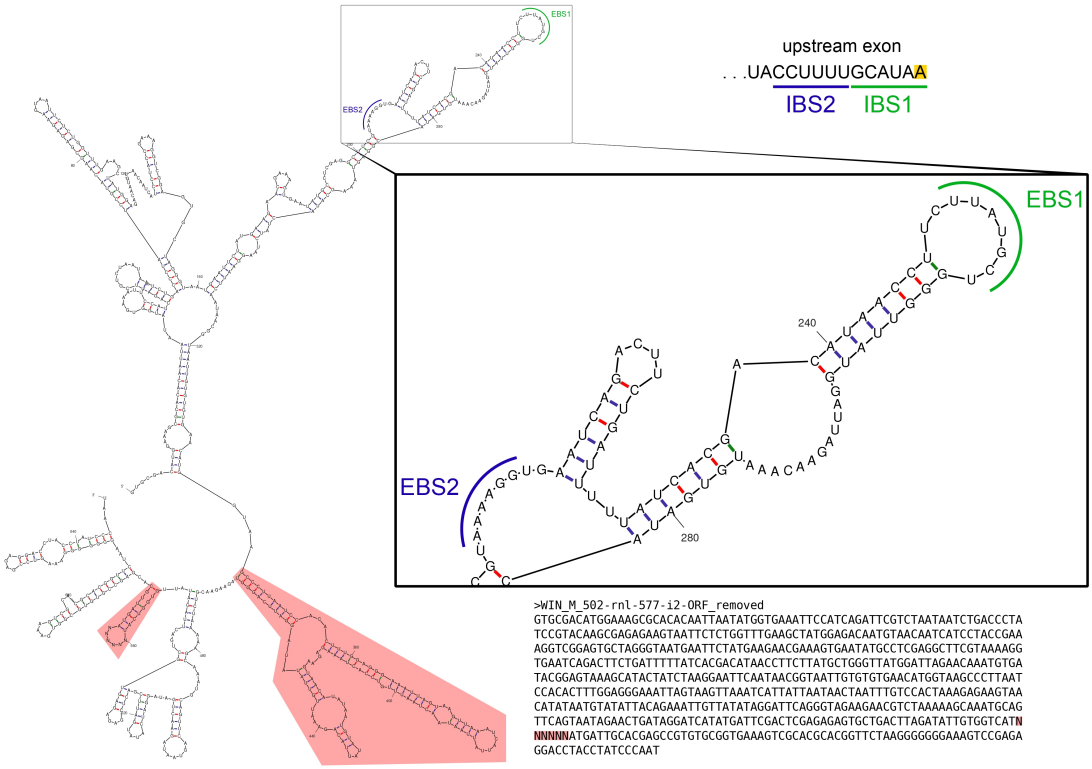

Figure 4.20: mfold model of *rnl*-577 in WIN(M) 502. Domains II and IV are highligted in red. It is indicated that for domain IV, a region coding for a RT was removed and replaced by six Ns.

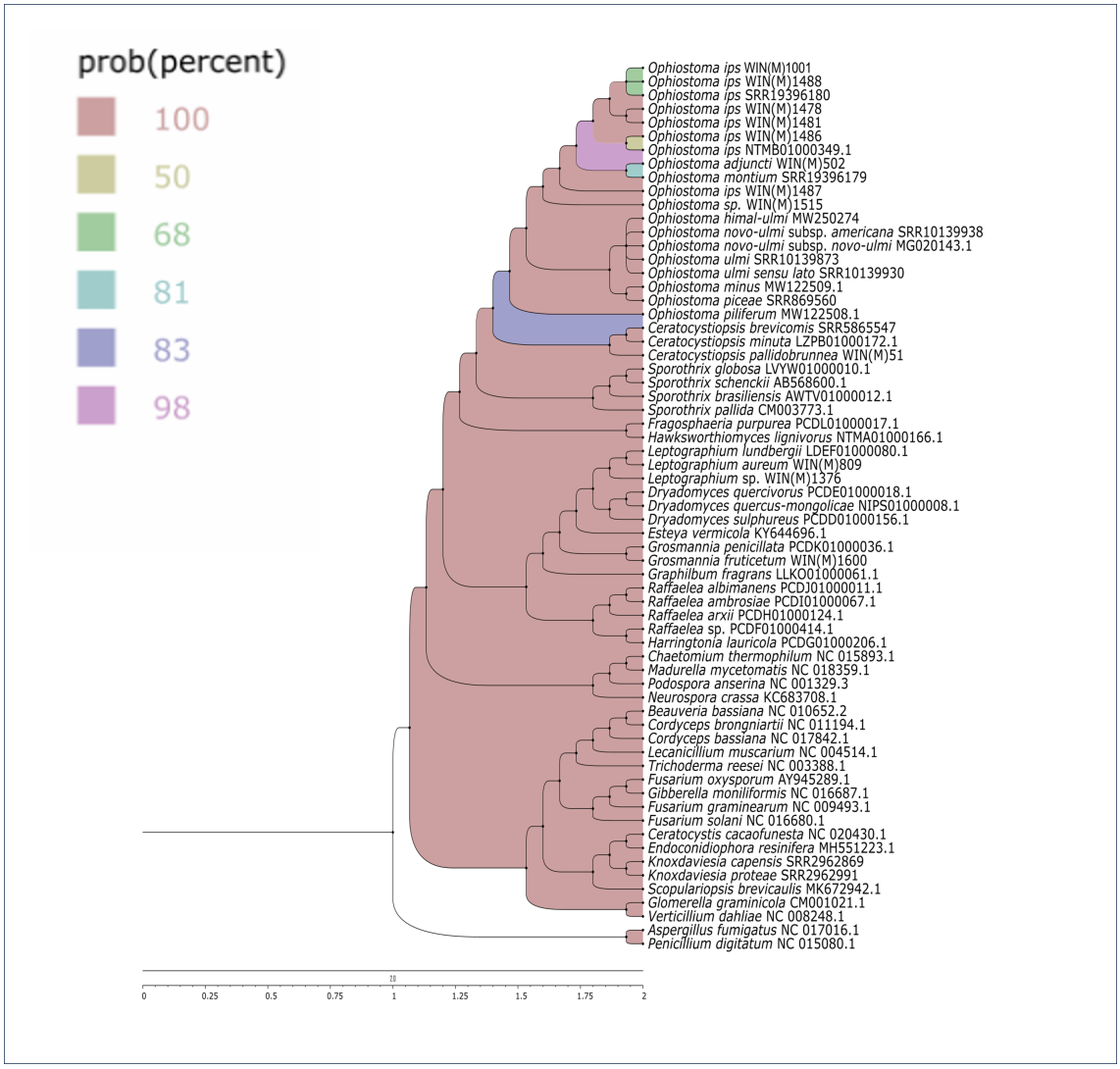

Figure 4.21: Phylogenetic relationships of 65 fungal species belonging to the Ascomycota, including 43 members of the Ophiostomatales are presented, based on concatenated amino acid sequences. *Aspergillus fumigatus* and *Penicillium digitatum* are selected as the outgroups. The tree was constructed based on Bayesian inference and the posterior probability (PP) support values from the 50% majority Bayesian (MB) consensus tree (Cprev+F+I model) are represented. PP support values are indicated at the nodes of the tree. Branch lengths are proportional to the number of substitutions per site (see scale bar).

Chapter 5

Conclusions and Future Directions

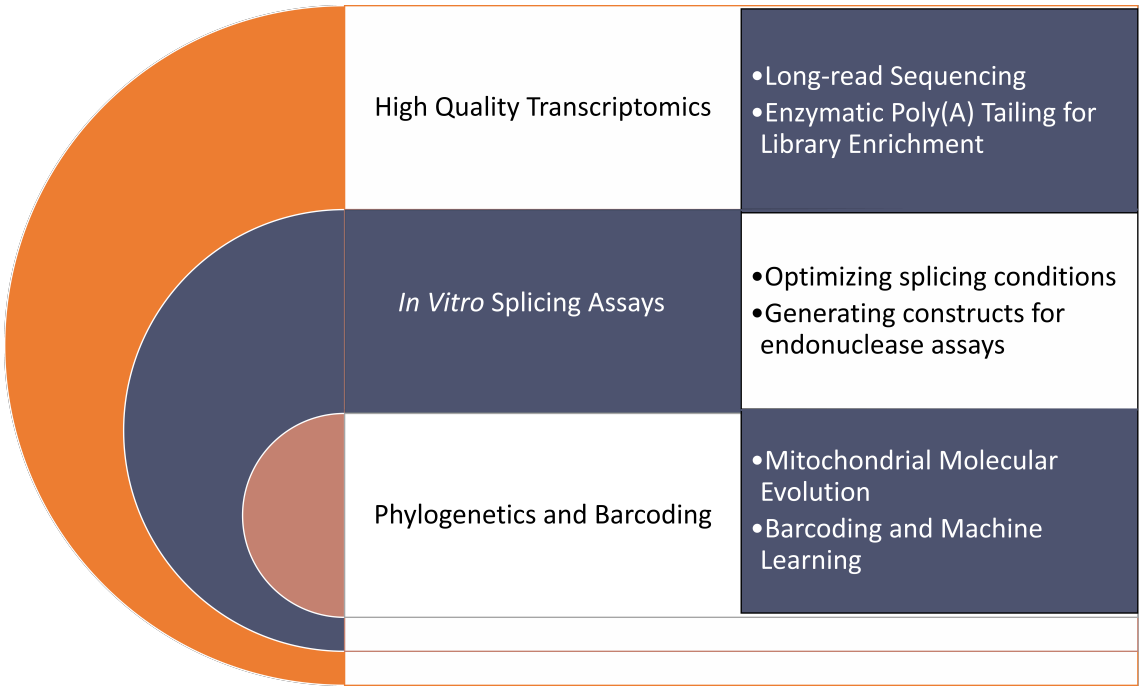

Figure 5.1: Potential Future Directions

Fungal mitogenomes are of great interest since they contain diverse elements that contribute to their size dynamicity. In some fungal lineages, mitogenome expansions seemed to coincide with intron proliferation, which led to the speculation that introns might be important regulatory elements that are retained due to a "survival" advantage, and not just based on neutral evolution or drift (Megarioti and Kouvelis, 2020; Mukhopadhyay and Hausner, 2021).

Comparative mitogenome analysis of fungal members of the Ophiostomatales is an ongoing effort to comprehensively understand the size variation, phylogenetic groupings, and novel and complex intron arrangements. The mitogenomes explored in this study presented examples of introns modules in nested/tandem and neighboring arrangements, and indicated scenarios that favored the integration of HEGs such as LAGLIDADG, as well as RTs into the introns. While some HEGs have been reported to move independently of their ribozyme partners (Mota and Collins, 1988), more frequently, they seem to have co-evolved with their encoding introns (Megarioti and Kouvelis, 2020; reviewed in Mukhopadhyay and Hausner, 2021). The splicing of mitochondrial introns is facilitated by nuclear and intron encoded factors (such as maturases etc.); therefore the processing of mtDNA transcripts are part of the mitonuclear interplay that allow mitochondrial gene expression to respond to nuclear cues in response to environmental and developmental signals (Cinget and Bélanger, 2020; Visinoni and Delneri, 2022). These interactions may also be part of mitochondrial-nuclear co-adaptations that can result in phenotypic differences and impact the so called "fitness landscape" with regards to the nuclear-mitochondrial combinations that can exist within a population (Dujon, 2020; Hill et al., 2020; Nguyen et al., 2020).

In Chapter 3, we see the example of invasion by a LAGLIDADG HEG into a resident *cox1* intron in *Leptographium* sp. WIN(M) 1376 where a "core creep" scenario seemed to favor the in-frame fusion of the intron-encoded HE-ORFs with the upstream exon and facilitate more efficient translation. In another example of *cob*-490 in *O. ips* WIN(M) 1478 and 1480 (Chapter 4), RNA-Seq results gave a preliminary overview of alternative splice isoforms, which could potentially facilitate the differential expression of the encoded LAGLIDADG HEGs. It could be speculated that in the neighboring arrangement of *rnl*-576, 577 as seen in *O. adjuncti* WIN(M) 502, there seems to be competition for two group II introns to insert into these *rnl* sites and generate a similar potential splicing scenario analogous to the *cob*-490

ratchet pathway. In *cob*-201 in *O. ips* WIN(M) 1487, a twintron could accommodate LAGLIDADG HE-ORFs in each intron module of the complex arrangement. These complex introns, along with their HEGs could provide an evolutionary advantage for their mutual persistence and mobility in the mitogenome (see fig. 7, Mukhopadhyay and Hausner, 2021), and contribute to the size dynamicity. A comprehensive look into the splicing pathways of these complex introns, and how that affects the translation of their encoded ORFs might require greater depth of sequencing through long-read technologies, combined with *in vitro* splicing assays.

In addition, comparative mitogenome analysis facilitates the resolution of issues concerning fungal taxonomy and population genetics. This is relevant in terms of forest pathology as several members of the Ophiostomatales are plant-pathogenic and invasive (Kulik et al., 2021, 2023; Trollip et al., 2022). In Chapter 3, a paper on members of the *Leptographium sensu lato* is presented, with the phyloegentic analysis combining approaches using mitogenomic sequences as well as DNA-based markers (ITS, beta-tubulin). This helped us validate the support for the division of *Raffaelea sensu lato* (De Beer and Wingfield, 2013) into three genera: *Raffaelea*, *Harringtonia*, and *Dryadomyces* (De Beer et al., 2022). In Chapter 4, we focused on members of the *Ophiostoma sensu stricto* comprised of *O. ips* and related strains that could be assigned to the "*O. ips* complex/lineage". It is interesting to note that for the blue-stain causing *O. ips*, there is a substantial difference in mitogenome sizes and intron components among strains from various geographic locations across the world (Canada, USA, New Zealand). The *ips*-like *Ophiostoma* sp. WIN(M) 1515 and the *O. ips* WIN(M) 1487 both seemed to group together with related members such as *O. montium* and *O. adjuncti.* Overlapping and often conflicting morphological features pose concerns for taxonomic classification, which could benefit from more comprehensive studies with bigger datasets and combining the results of nuclear and mitogenomic- sequence based phylogenies.

In this section, we will explore the different directions that work on fungal mi-

tochondrial sequences could take and their potential and limitations.

Long-read Technologies for In-depth Sequencing

RNA-Seq has proved to be a powerful method of studying the transcriptome and has not only enabled the improvement of existing fungal genome annotations but has also facilitated research into a wide variety of biological processes in fungi by revealing their underlying genetic regulation. Recently, emphasis has been put on improving annotations of pathogenic fungi and understanding the molecular mechanisms of pathogenicity, including the gene regulatory networks that underpin them. This might enable researchers to theorize about potential new novel therapies, networks, and targets of drugs in their efforts to develop effective treatments. Although mitogenome sequences of a growing number of fungi are reported, only a few papers regarding fungal mitogenomes investigated expression of mitochondrial genes. Expression quantification of mitochondrial genes was only performed for several yeasts, including *Candida albicans* (Kolondra et al., 2015), *S. cerevisiae* (Turk et al., 2013), and *Schizosaccharomyces pombe* (Shang et al., 2018), and filamentous fungi, including *Hirsutella thompsonii* (Wang et al., 2018), *Ophiocordyceps sinensis* (Li et al., 2015), *Pestalotiopsis fci* (Zhang et al., 2017a), and *Tolypocladium infatum* (Zhang et al., 2017b).

Putzeys et al. (2022, 2024) recognized the challenges with short-read RNA-Seq technology that fail to capture key transcriptional events in phages such as initiation and termination, since this strategy does not differentiate between primary and processed transcripts and cannot successfully delineate transcriptional boundaries. They presented a Nanopore-based long-read strategy that could generate a full-length transcriptional map and help to uncover complex transcriptomes (Putzeys et al., 2022) and utilized this strategy to successfully profile transcriptomes from virulent phages across major taxonomic clades (Putzeys et al., 2024). This study showed the potential of using this long-read strategy to identify group I intron splic-

ing events in phages, which could be extended to other prokaryotic systems, but the success in resolving RNA processing events in filamentous fungal systems is yet to be explored through similar strategies. In non-model filamentous fungi such as *O. ips*, it is difficult to either synchronize or yield abundant mycelial mass. Another setback of RNA-Seq on Illumina NovaSeq6000 platform is the fact that till now, poly(A) tails are assumed to be absent from mtDNA transcripts and enrichment of libraries through poly(A) selection could not be performed for the fungal strains in this study. This might be remedied if long poly(A) tails could be enzymatically added to the natively non-polyadenylated mitochondrial RNA samples, ensuring successful adapter ligation, which has been attempted in the model fungal system, *Saccharomyces cerevisiae* (Koster et al., 2023). This might reduce the "noise" from rRNAs and tRNAs. Koster et al. also suggested the use of Nanopore long read sequencing for better data resolution when looking into splicing events, which can be used as a potential future strategy for our RNA-Seq studies. Finally, the phylogenetic groupings of the fungal members of the Ophiostomatales can be refined by combining analyses using nuclear and mitogenomic sequences and a constantly expanding repertoire of available fungal whole genome sequences and high-quality assemblies.

In Vitro Splicing Assays

For the biochemical characterization of alternative splice isoforms preliminarily detected through RNA-Seq analyses (Chapter 4), one way could be utilizing *in vitro* splicing assays by manipulation splicing conditions such as temperature or magnesium ions. Highly reactive self-splicing Gr II introns within mitogenomes of human pathogenic fungi have been assayed using designed precursor molecules and their dependence for magnesium ion for self-splicing was evaluated (Liu and Pyle, 2021). These introns were found to readily self-splice and could fold into catalytically active conformations under near physiological conditions. Similarly, Gr I introns from

phytopathogenic fungi were tested *in vitro* for their self-splicing activity, considering variables such as buffer composition, arbitrary 5′ and 3′ exon lengths, incubation time, and annealing conditions (Malbert et al., 2023). This work facilitated the validation of a *trans*-ribozyme designed from Gr I introns and paved the way for discovery of new anti-fungal molecules. To apply similar strategies to complex introns identified through our study, one would require to optimize an *in vitro* transcription reaction mixture, as well as determine the splicing competency of the composite module, as well as individual intron modules in complex arrangements by modifying buffer salts. A similar approach was previously tested by Guha and Hausner (2016) where splicing events for the nested twintron S1247 were detected by PCR-amplifying cDNAs generated from *in vitro* transcribed RNA. Further experiments for determining the activity of intron-encoded endonucleases were carried out using a suitable expression vector constructs to express the HE protein in *E. coli*. To monitor the cleavage reactions and functionality of the HE-ORF, a substrate plasmid containing a DNA segment with the putative target site (additionally, target site-free control plasmid) was generated for cleavage assay and mapping (akin to restriction digestion mapping). One potential challenge to replicate such experiments for the fungal complex trintron is the size of the construct for plasmid construction for *in vitro* transcription. A possible way to reduce construct sizes could be through the removal of the intron ORF sequences and only keeping the relevant intron core sequences needed for splicing. Intron RNA splicing tends to required magnesium (Gregan et al., 2001; Rangan and Woodson, 2003). Experimental standardization (titration) to determine magnesium concentration that could slow down the splicing sufficiently to detect intermediate splicing steps and isoforms could also be tedious but might be useful to capture intermediate if the intron modules are highly reactive. However, there is immense potential in designing these experiments to see the effects of magnesium ions on the alternative splicing of complex intron modules, and the expression of their encoded LAGLIDADG endonucleases.

Phylogenetics and Taxonomic Surveys

Fungi assigned to the Ophiostomatales are noted for their association and co-evolution with their vector insects, such as bark beetles (Hausner et al., 2005; Wingfield et al., 2017; Jankowiak et al., 2022). Due to similar adaptive features for the vector associations and insect-mediated dispersals, these fungi frequently have overlapping morphological features, making their identification and classification difficult. Ophiostomatoid genera of the Ophiostomatales that are most commonly associated with beetles include members of the following genera: *Affroraffaelea*, *Aureovirgo*, *Ceratocystiopsis*, *Fragosphaeria*, *Graphilbum*, *Leptographium*, *Ophiostoma*, *Raffaelea*, and *Sporothrix* (Hyde et al., 2020). The bark-beetle associated Ophiostomatales fungi could pose bio-surveillance concerns such as invasion as exotic species due to their wide-spread global distribution through insect-vectored dispersal or by anthropogenic causes (movement of contaminated timber). Trollip et al. (2021) surveyed ophiostomatoid fungi in Australia using whole genome sequencing (WGS) and phylogenetics for improved molecular diagnostics for forest biosecurity. They used WGS assemblies to extract commonly used barcoding loci (ITS, beta-tubulin, calmodulin, elongation factor) and generated locus-specific binned reads for multi-locus phylogenetic analysis. As such, our current work on the *Leptographium sensu lato* (Chapter 3) using nuclear markers such as ITS and beta-tubulin could be expanded to include multiple DNA loci, and correlated with morphological studies. This could also be extended or integrated into the modular multi-barcode amplicon sequencing pipeline established by Trollip et al. (2022). For their study, they utilized Ophiostomatales as a model since this group of fungi exhibit several concerns for fungal species identification such as extensive taxonomic revisions due to convergent evolution, presence of cryptic species (species that cannot be distinguished based on morphological criteria), and data availability limitations (De Beer et al., 2021; Jankowiak et al., 2022).

Another direction for the phylogenetic work could be the integrating of machine

learning for biosurveillance and using phylogenetic signals to test phytopathogenic traits of Ascomycota fungi (Dort et al., 2023). Their work utilized a supervised machine learning approach that combined genomic and phylogenetic data and predicted fungal phytopathogenic lifestyles and biosurveillance characteristics. Predictive genomics could be crucial for the advancement of biosurveillance pipelines of plant-pathogenic or economically important ophiostomatoid fungi (Dort et al., 2023). Overall, phylogenetic approaches can be used to answer different questions, about evolutionary and taxonomic relationships in fungi, as long as substantial datasets and genome assemblies are available in public databases, to explore.

References

Abboud, T.G., Zubaer, A., Wai, A., and Hausner, G. (2018). The complete mitochondrial genome of the Dutch elm disease fungus *Ophiostoma novo-ulmi* subsp. *novo-ulmi. Can. J. Microbiol.*, 64, 339–348. doi: 10.1139/cjm-2017-0605

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

Adams, P.L., Stahley, M.R., Gill, M.L., Kosek, A.B., Wang, J., and Strobel, S.A. (2004). Crystal structure of a group I intron splicing intermediate. *RNA*, 10, 1867–1887. doi: 10.1261/rna.7140504

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible, and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.*, 46, W537–W544. doi: 10.1093/nar/gky379

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., and Gabaldón, T. (2014). High variability of mitochondrial gene order among fungi. *Genome Biology and Evolution,* 6(2), 451-465. https://doi.org/10.1093/gbe/evu028

Akins, R.A., and Lambowitz, A.M. (1987). A protein required for splicing group I in-

trons in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. *Cell*, 50, 331-345.

Albert, B., and Sellem, C.H. (2002). Dynamics of the mitochondrial genome during *Podospora anserina* aging. *Curr Genet*, 40, 365–373

Allen, J.F. (2015). Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, 112, 10231–10238. doi: 10.1073/pnas.1500012112

Allen, J.F., and Raven, J.A. (1996). Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J. Mol. Evol.*, 42, 482–492. doi: 10.1007/BF02352278

Anziano, P.Q., Hanson, D.K., Mahler, H.R., and Perlman, P.S. (1982). Functional domains in introns: Trans-acting and *cis*-acting regions of intron 4 of the *cob* gene. *Cell*, 30, 925–932. doi: 10.1016/0092-8674(82)90297-5

Araújo, D.S., de-Paula, R.B., Tomé, L.M.R., Quintanilha-Peixoto, G., Salvador-Montoya, C.A., del-Bem, L.E., et al. (2021). Comparative mitogenomics of Agaricomycetes: diversity, abundance, impact and coding potential of putative open-reading frames. *Mitochondrion,* 58, 1–13. doi: 10.1016/j.mito.2021.02.002

Araújo, J.P.M, Lebert, B.M., Vermeulen, S., Brachmann, A., Ohm, R.A., Evans, H.C., and de Bekker, C. (2022). Masters of the manipulator: Two new hypocrealean general, Niveomyces (Cordycipitaceae) and Torrubiellomyces (Ophiocordycipitaceae), parasitic on the zombie ant fungus *Ophiocordyceps camponoti-floridani*. *Persoonia*, 49(1), 171–194. https://doi.org/10.3767/persoonia.2022.49.05

Aylward, J., Steenkamp, E.T., Dreyer, L. L., Roets, F., Wingfield, B.D., and Wingfield, M.J. (2017). A plant pathology perspective of fungal genome sequencing. *IMA Fungus*, 8, 1–15. doi: 10.5598/imafungus.2017.08.01.01

Baidyaroy, D., Hausner, G., Hafez, M., Michel, F., Fulbright, D.W., and Bertrand,

H. (2011). A 971-bp insertion in the *rns* gene is associated with mitochondrial hypovirulence in a strain of *Cryphonectria parasitica* isolated from nature. *Fungal Genet. Biol*, 48, 775–783. doi: 10.1016/j.fgb.2011.05.006

Bankevich, A., Nurk, S., Antipov, D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.*, 19(5), 455-77. doi: 10.1089/cmb.2012.0021.

Barrangou, R. (2013). CRISPR-Cas systems and RNA-guided interference. *Wiley Interdiscip. Rev. RNA*, 4(3), 267-278

Barzel, A., Naor, A., Privman, E., Kupiec, M., and Gophna, U. (2011). Homing endonucleases residing within inteins: Evolutionary puzzles awaiting genetic solutions. *Biochem. Soc. Trans.*, 39, 169–173. doi: 10.1042/BST0390169

Bassi, G.S., and Weeks, K.M. (2003). Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst. *Biochemistry*, 42(33), 9980-9988. https://doi.org/10.1021/bi0346906

Bassi, G.S., de Oliveira, D.M., White, M.F., and Weeks, K.M. (2002). Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc. Natl. Acad. Sci. USA*, 99(1), 128-133. doi: 10.1073/pnas.012579299

Beagley, C.T., Okada, N.A., and Wolstenholme, D.R. (1996). Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: One intron contains genes for subunits 1 and 3 of NADH dehydrogenase. *Proc. Natl. Acad. Sci. USA.*, 93, 5619–5623. doi: 10.1073/pnas.93.11.5619

Begel, O., Boulay, J., Albert, B., Dufour, E., and Sainsard-Chanet, A. (1999). Mitochondrial group II introns, cytochrome c oxidase, and senescence in *Podospora anserina*. *Mol Cell Biol.*, 19, 4093–4100

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

Belfort, M. (2017). Mobile self-splicing introns and inteins as environmental sensors. *Curr. Opin. Microbiol.*, 38, 51–58. doi: 10.1016/j.mib.2017.04.003

Belfort, M., and Lambowitz, A.M. (2019). Group II Intron RNPs and reverse transcriptases: From retroelements to research tools. *Cold Spring Harb Perspect Biol.*, 11, a032375. doi: 10.1101/cshperspect.a032375

Belfort, M., and Perlman, P.S. (1995). Mechanisms of intron mobility. *J Biol Chem.*, 270(51), 30237-40. doi: 10.1074/jbc.270.51.30237

Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B., and Lambowitz, A.M. (2002). Mobile introns: Pathways and proteins. In Mobile DNA II (eds Craig NL, Craigie R, Gellert M, Lambowitz AM), pp. 761–783 ASM Press, Washington D.C

Benade, E., Wingfield, M.J., and Van Wyk, P.S. (1997). Conidium development in Sporothrix anamorphs of Ophiostoma. *Mycological Research*, 101, 1108–1112

Benny, G.L., and Kimbrough, J.W. (1980). A synopsis of the orders and families of Plectomycetes with keys to genera. *Mycotaxon*, 12, 1–91

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.*, 41, D36–D42. doi: 10.1093/nar/gks1195

Berg, O.G., and Kurland, C.G. (2000). Why mitochondrial genes are most often found in nuclei. *Mol. Biol. Evol.*, 17, 951–961. doi: 10.1093/oxfordjournals.molbev.a026376

Bernardino, A.F., Li, Y., Smith, C.R., and Halanych, K.M. (2017). Multiple introns in a deep-sea Annelid (*Decemunciger: Ampharetidae*) mitochondrial genome. *Sci. Rep.*, 7, 4295. doi: 10.1038/s41598-017-04094-w

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

Birgisdottir, A.B., and Johansen, S. (2005). Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. *Nucleic Acids Res.*, 33, 2042–2051. doi: 10.1093/nar/gki341

Blanchard, J.L., and Lynch, M. (2000). Organellar genes: why do they end up in the nucleus? *Trends Genet.*, 16, 315–320. doi: 10.1016/s0168-9525(00)02053-9

Bolduc, J.M., Spiegel, P.C., Chatterjee, P., Brady, K.L., Downing, M.E., Caprara, M.G., et al. (2003). Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. *Genes & Development*, 17(23), 2875-2888. doi: 10.1101/gad.1109003

Bolender, N., Sickmann, A., Wagner, R., Meisinger, C., and Pfanner, N. (2008). Multiple pathways for sorting mitochondrial precursor proteins. *EMBO Rep.*, 9, 42–49. doi: 10.1038/sj.embor.7401126

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-20. doi: 10.1093/bioinformatics/btu170.

Bonen, L. (2008). *Cis*- and *trans*-splicing of group II introns in plant mitochondria. *Mitochondrion.*, 8, 26–34. doi: 10.1016/j.mito.2007.09.005

Bonen, L., and Vogel, J. (2001). The ins and outs of group II introns. *Trends Genet.*, 17, 322–331. doi: 10.1016/S0168-9525(01)02324-1

Bonocora, R.P., and Shub, D.A. (2009). A likely pathway for formation of mobile group I introns. *Curr. Biol.*, 19, 223–228. doi: 10.1016/j.cub.2009.01.033

Brankovics, B., Kulik, T., Sawicki, J., Bilska, K., Zhang, H., de Hoog, G.S., van der Lee, T.A., Waalwijk, C., and van Diepeningen, A.D. (2018). First steps towards mitochondrial pan-genomics: Detailed analysis of *Fusarium graminearum* mitogenomes. *PeerJ.*, 6, e5963. doi: 10.7717/peerj.5963

Brown, G.G., Colas des Francs-Small, C., and Ostersetzer-Biran, O. (2014). Group II intron splicing factors in plant mitochondria. *Front. Plant Sci.*, 5, 35. doi: 10.3389/fpls.2014.00035

Bullerwell, C. E., Burger, G., and Lang, B. F. (2000). A novel motif for identifying rps3 homologs in fungal mitochondrial genomes. *Trends Biochem. Sci.*, 25, 363–365. doi: 10.1016/s0968-0004(00)01612-1

Burger, G., Yan, Y., Javadi, P., and Lang, B.F. (2009). Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends Genet.*, 25, 381–386. doi: 10.1016/j.tig.2009.07.003

Burke, J.M. (1988). Molecular genetics of group I introns: RNA structures and protein factors required for splicing–a review. *Gene*, 73(2), 273-294

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A.D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience*, 8(9), giz100. doi: 10.1093/gigascience/giz100

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421. doi: 10.1186/1471-2105-10-421

Caprara, M.G., and Waring, R.B. Group I introns and their maturases: Uninvited, but welcome guests. In: Belfort M., Wood D.W., Stoddard B.L., Derbyshire V., editors. *Homing Endonucleases and Inteins.* Springer; Berlin/Heidelberg, Germany: 2015

Carneiro, P., Duarte, M., and Videira, A. (2004). The main external alternative NAD(P)H dehydrogenase of *Neurospora crassa* mitochondria. *Biochim Biophys Acta.*, 1608(1), 45-52

Carvalho, D.S., Andrade, R.F.S., Pinho, S.T.R., Góes-Neto, A., Lobão, T.C.P., Bomfim, G.C., et al. (2015). What are the evolutionary origins of mitochondria? A complex network approach. *PLoS One*, 10, e0134988. 10.1371/journal.pone.0134988

Cech, T.R., Zaug, A.J., and Grabowski, P.J. (1981). In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell*, 27(3 Pt 2), 487-496

Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol.*, 1962, 1–14. doi: 10.1007/978-1-4939-9173-0_1

Chan, R.T., Peters, J.K., Robart, A.R., Wiryaman, T., Rajashankar, K.R., and Toor, N. (2018). Structural basis for the second step of group II intron splicing. *Nat. Commun.*, 9, 4676. doi: 10.1038/s41467-018-0667

Chang, R., Duong, T.A., Taerum, S.J., et al. (2017). Ophiostomatoid fungi associated with conifer-infesting beetles and their phoretic mites in Yunnan, China. *MycoKeys*, 28, 19–64

Chastain, M, and Tinoco, I. Jr. (1992). A triple-base structural domain in RNA. *Biochemistry*, 31(51), 12733-12741

Chatre, L., and Ricchetti, M. (2014). Are mitochondria the Achilles' heel of the Kingdom Fungi? *Curr. Opin. Microbiol.*, 20, 49–54. doi: 10.1016/j.mib.2014.05.001

Chen, C., Li, Q., Fu, R., Wang, J., Deng, G., Chen, X., and Lu, D. (2021). Comparative mitochondrial genome analysis reveals intron dynamics and gene rearrangements in two *Trametes* species. *Sci. Rep.*, 11, 2569. doi: 10.1038/s41598-021-82040-7

Chen, S.Y., Li, C., Jia, X., and Lai, S.J. (2019). Sequence and evolutionary features for the alternatively spliced exons of eukaryotic genes. *Int. J. Mol. Sci.*, 20, 3834. doi: 10.3390/ijms20153834

Cheng, J., Luo, Q., Ren, Y., Luo, Z., Liao, W., Wang, X., and Li, Q. (2021). Panorama of intron dynamics and gene rearrangements in the phylum Basidiomycota as revealed by the complete mitochondrial genome of *Turbinellus floccosus*. *Appl Microbiol. Biotechnol.*, 105, 2017–2032. doi: 10.1007/s00253-021-11153-w

Chevalier, B., Monnat, R.J., and Stoddard, B.L. The laglidadg homing endonuclease family. In: Belfort M., Wood D.W., Stoddard B.L., Derbyshire V., editors.

Homing Endonucleases and Inteins. Volume 16. Springer; Berlin/Heidelberg, Germany: 2005

Chevalier, B.S., and Stoddard, B.L. (2001). Homing endonucleases: Structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, 29, 3757–3774. doi: 10.1093/nar/29.18.3757

Chi, S.I., Dahl, M., Emblem, Å., and Johansen, S.D. (2019). Giant group I intron in a mitochondrial genome is removed by RNA back-splicing. *BMC Mol. Biol.*, 20,16. doi: 10.1186/s12867-019-0134-y

Chien, M.F., Tosa, S., Huang, C.C., and Endo, G. (2009). Splicing of a bacterial group II intron from *Bacillus megaterium* is independent of intron-encoded protein. *Microbes Environ.*, 24(1), 28-32

Cinget, B., and Bélanger, R.R. (2020). Discovery of new group I-D introns leads to creation of subtypes and link to an adaptive response of the mitochondrial genome in fungi. *RNA Biology,* 17(9), 1252–1260. https://doi.org/10.1080/15476286.2020.1763024

Coil, D., Jospin, G., and Darling, A.E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, 31, 587–589. doi: 10.1093/bioinformatics/btu661

Colleaux, L., d'Auriol, L., Galibert, F., and Dujon, B. (1988). Recognition and cleavage site of the intron encoded omega transposase. *Proc. Natl. Acad. Sci. USA.*, 85, 6022–6026. doi: 10.1073/pnas.85.16.6022

Contreras, L.M., Huang, T., Piazza, C.L., Smith, D., Qu, G., Gelderman, G., Potratz, J.P., Russell, R., and Belfort, M. (2013). Group II intron-ribosome association protects intron RNA from degradation. *RNA.*, 19, 1497–1509. doi: 10.1261/rna.039073.113

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

Copertino, D.W., and Hallick, R.B. (1993). Group II and group III introns of twintrons: Potential relationships with nuclear pre-mRNA introns. *Trends Biochem. Sci.*, 18, 467–471. doi: 10.1016/0968-0004(93)90008-B.

Copertino, D.W., Hall, E.T., Van Hook, F.W., Jenkins, K.P., and Hallick, R.B. (1994). A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Res*, 22(6), 1029-36. doi: 10.1093/nar/22.6.1029

Coros, C.J., Piazza, C.L., Chalamcharla, V.R., Smith, D., and Belfort, M. (2009). Global regulators orchestrate group II intron retromobility. *Mol. Cell.*, 34, 250–256. doi: 10.1016/j.molcel.2009.03.014

Corsaro, D., and Venditti, D. (2020). Putative Group I introns in the eukaryote nuclear internal transcribed spacers. *Curr. Genet.*, 66, 373–384. doi: 10.1007/s00294-019-01027-0

Corsaro, D., Köhsler, M., Venditti, D., Rott, M.B., and Walochnik, J. (2019). Recovery of an acanthamoeba strain with two Group I introns in the nuclear 18S rRNA Gene. *Eur. J. Protistol.*, 68, 88098. doi: 10.1016/j.ejop.2019.01.007

Costa, M., Michel, F., and Westhof, E. (2000). A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.*,19, 5007–5018. doi: 10.1093/emboj/19.18.5007

Coughlan, A.Y., Lombardi, L., Braun-Galleani, S., Martos, A.A., Galeote, V., Bigey, F., Dequin, S., Byrne, K.P., and Wolfe, K.H. (2020). The yeast mating-type switching endonuclease HO is a domesticated member of an unorthodox homing genetic element family. *eLife.*, 9, e55336. doi: 10.7554/eLife.55336

Court, D.A., and Bertrand, H. (1992). Genetic organization and structural features of maranhar, a senescence-inducing linear mitochondrial plasmid of *Neurospora crassa*. *Curr Genet*, 22, 385–397

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J.E., Yang, J., Mills, D., Manias, D., Dunny, G., Lambowitz, A.M., and Belfort, M. (1998). Retro-

homing of a bacterial group II intron: Mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell.*, 94, 451–462. doi: 10.1016/S0092-8674(00)81586-X

Cui, X., Matsuura, M., Wang, Q., Ma, H., and Lambowitz, A.M. (2004). A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. *Journal of Molecular Biology*, 340(2), 211-231

Dabbagh, N., and Preisfeld, A. (2016). The chloroplast genome of *Euglena mutabilis* —Cluster arrangement, intron analysis, and intrageneric trends. *J. Eukaryot. MicroBiol.*, 64, 31–44. doi: 10.1111/jeu.12334

Dai, L., Chai, D., Gu, S. Q., Gabel, J., Noskov, S.Y., Blocker, F.J., et al. (2008). A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase. *Molecular Cell*, 30(4), 472-485.

Daniels, D.L., Michel, W.J., Jr., and Pyle, A.M. (1996). Two competing pathways for self-splicing by group II introns: A quantitative analysis of in vitro reaction rates and products. *J. Mol. Biol.*, 256, 31–49. doi: 10.1006/jmbi.1996.0066

Davies, R.W., Waring, R.B., Ray, J.A., Brown, T.A., and Scazzocchio, C. (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature*, 300(5894), 719-724

De Beer, Z.W., and Wingfield, M.J. (2013). Emerging lineages in the Ophiostomatales. In *Ophiostomatoid fungi: Expanding frontiers.* Edited by K. A. Seifert, Z. W. Beerde, and M. J. Wingfield. CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands. pp. 21–46.

De Beer, Z.W., Procter, M., Wingfield, M.J., Marincowitz, S., and Duong, T.A. (2022). Generic boundaries in the Ophiostomatales reconsidered and revised. *Stud. Mycol.*, 101, 57–120. doi: 10.3114/sim.2022.101.02

De Beer, Z.W., and Wingfield, M.J. (2013b). Emerging lineages in the Ophiostom-

atales. In: The Ophiostomatoid Fungi: Expanding Frontiers, (Seifert KA, De Beer ZW, Wingfield MJ, ed.): 21–46. CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands

De Beer, Z.W., Duong, T.A., and Wingfield, M.J. (2016). The divorce of *Sporothrix* and *Ophiostoma*: solution to a problematic relationship. *Studies in Mycology*, 83, 165–191

De Beer, Z.W., Seifert, K, and Wingfield, M.J. (2013a). The ophiostomatoid fungi: their dual position in the Sordariomycetes. In: The Ophiostomatoid Fungi: Expanding Frontiers, (Seifert KA, De Beer ZW, Wingfield MJ, ed.): 1–19. CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands

De Beer, Z.W., Seifert, K.A., and Wingfield, M.J. (2013b). A nomenclator for ophiostomatoid genera and species in the *Ophiostomatales* and *Microascales*. In: *The Ophiostomatoid Fungi: Expanding Frontiers*, (Seifert KA, De Beer ZW, Wingfield MJ, ed.): 245–322. CBS-KNAW Fungal Biodiversity Centre, Utrecht, The Netherlands.

De Errasti, A., De Beer, Z.W., Coetzee, M.P.A, et al. (2016). Three new species of Ophiostomatales from Nothofagus in Patagonia. *Mycological Progress*, 15, 17

De Lencastre, A., Hamill, S., and Pyle, A.M. (2005). A single active-site region for a group II intron. *Nat. Struct. Mol. Biol.*, 12, 626–627. doi: 10.1038/nsmb957

De Silva, D., Poliquin, S., Zeng, R., Zamudio-Ochoa, A., Marrero, N., Perez-Martinez, X., et al. (2017). The DEAD-box helicase Mss116 plays distinct roles in mitochondrial ribogenesis and mRNA-specific translation. *Nucleic Acids Res.*, 45, 6628–6643. doi: 10.1093/nar/gkx426

De Vries, S., Van Witzenburg, R., Grivell, L.A., and Marres, C.A. (1992). Primary structure and import pathway of the rotenone-insensitive NADH-ubiquinone oxidoreductase of mitochondria from *Saccharomyces cerevisiae*. *Eur J Biochem.*, 203(3), 587-592

De Zamaroczy, M., and Bernardi, G. (1986). The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae*–a review. *Gene*, 47(2-3), 155-177

Del Hoyo, A., Álvarez, R., Gasulla, F., Casano, L.M., and Del Campo, E.M. (2018). Origin and evolution of chloroplast Group I introns in lichen algae. *J. Phycol.*, 54, 66–78. doi: 10.1111/jpy.12600

Dellaporta, S.L., Xu, A., Sagasser, S., Jakob, W., Moreno, M.A., Buss, L.W., and Schierwater, B. (2006). Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proc. Nat. Acad. Sci. USA.*, 103, 8751–8756

Deng, Y., Hsiang, T., Li, S., Lin, L., Wang, Q., Chen, Q., et al. (2018). Comparison of the mitochondrial genome sequences of six *Annulohypoxylon stygium* isolates suggests short fragment insertions as a potential factor leading to larger genomic size. *Front. Microbiol.* 9:2079. doi: 10.3389/fmicb.2018.02079

Deng, Y., Wu, X., Wen, D., Huang, H., Chen, Y., Mukhtar, I., Yue, L., Wang, L., and Wen, Z. (2021). Intraspecific mitochondrial DNA comparison of mycopathogen *Mycogone perniciosa* provides insight into mitochondrial transfer RNA introns. *Phytopathology*, 111, 639–648. doi: 10.1094/PHYTO-07-20-0281-R

Deng, Y., Zhang, Q., Ming, R., Lin, L., Lin, X., Lin, Y., Li, X., Xie, B., and Wen, Z. (2016). Analysis of the mitochondrial genome in *Hypomyces aurantius* reveals a novel twintron complex in fungi. *Int. J. Mol. Sci.*,17, 1049. doi: 10.3390/ijms17071049

Deng, Y., Zhang, X., Xie, B., Lin, L., Hsiang, T., Lin, X., et al. (2020). Intra-specific comparison of mitochondrial genomes reveals host gene fragment exchange via intron mobility in *Tremella fuciformis*. *BMC Genomics*, 21, 426. doi: 10.1186/s12864-020-06846-x

DePriest, P.T., and Been, M.D. (1992). Numerous Group I introns with variable

distributions in the ribosomal DNA of a lichen fungus. *J. Mol. Biol.*, 228, 315–321. doi: 10.1016/0022-2836(92)90819-6

Déquard-Chablat, M., Sellem, C.H., Golik, P., Bidard, F., Martos, A., Bietenhader, M., et al. (2011). Two nuclear life cycle-regulated genes encode interchangeable subunits c of mitochondrial ATP synthase in *Podospora anserina*. *Mol. Biol. Evol.*, 28, 2063–2075. doi: 10.1093/molbev/msr025

Dighton, J. (2007). Nutrient cycling by saprotrophic fungi in terrestrial habitats. In: Environmental and Microbial Relationships. Volume 4. Springer Berlin Heidelberg. pp. 287-300. https://doi.org/10.1007/978-3-540-71840-6_16

Dlakić, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA*, 17, 799–808. doi: 10.1261/rna.2396011

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. doi: 10.1093/bioinformatics/bts635.

Dolan, G.F., and Müller, U.F. (2014). *Trans*-splicing with the group I intron ribozyme from *Azoarcus*. *RNA.*, 20, 202–213. doi: 10.1261/rna.041012.113

Dort, E.N., Layne, E., Feau, N., Butyaev, A., Henrissat, B., Martin, F.M., Haridas, S., Salamov, A., Grigoriev, I.V., Blanchette, M., and Hamelin, R.C. (2023). Large-scale genomic analyses with machine learning uncover predictive patterns associated with fungal phytopathogenic lifestyles and traits. *Sci Rep.*, 13(1), 17203. doi: 10.1038/s41598-023-44005-w.

Drager, R.G., and Hallick, R.B. (1993). A complex twintron is excised as four individual introns. *Nucleic Acids Res.*, 21, 2389–2394. doi: 10.1093/nar/21.10.2389

Dujardin, G., and Herbert, C.J. (1997). Aminoacyl tRNA synthetases involved in group I intron splicing. In R. F. Gesteland & R. Schroeder (Eds.), *Ribosomal RNA and Group I Introns* (pp. 179-198). Landes Bioscience.

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

Dujon, B. Homing Endonucleases and the Yeast Mitochondrial ω Locus—A Historical Perspective. In: Belfort M., Wood D.W., Stoddard B.L., Derbyshire V., editors. *Homing Endonucleases and Inteins.* Volume 16. Springer; Berlin/Heidelberg, Germany: 2005.

Dujon, B. (2020). Mitochondrial genetics revisited. *Yeast (Chichester, England),* 37(2), 191–205. https://doi.org/10.1002/yea.3445

Dujon, B., and Belcour, L. (1989). Mitochondrial DNA instabilities and rearrangements in yeasts and fungi. In: Berg DE, Howe MM (eds) Mobile DNA. AMS Press, Washington DC, pp 861–878

Dujon, B., Bolotin-Fukuhara, M., Coen, D., Deutsch, J., Netter, P., Slonimski, P.P., and Weill, L. (1976). Mitochondrial genetics. XI. Mutations at the mitochondrial locus omega affecting the recombination of mitochondrial genes in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.*, 143, 131–165. doi: 10.1007/BF00266918

Duò, A., Bruggmann, R., Zoller, S., Bernt, M., and Grünig, C.R. (2012). Mitochondrial genome evolution in species belonging to the *Phialocephala fortinii* s.l. - *Acephala applanata* species complex. *BMC Genomics*, 13, 166. doi: 10.1186/1471-2164-13-166

Eckhardt, L.G., Jones, J.P., and Klepzig, K.D. (2004). Pathogenicity of *Leptographium* species associated with loblolly pine decline. *Plant Dis.*, 88, 1174–1178. doi: 10.1094/PDIS.2004.88.11.1174

Edgell, D.R. (2009). Selfish DNA: homing endonucleases find a home. *Curr. Biol.*, 19, R115–R117. doi: 10.1016/j.cub.2008.12.019

Edgell, D.R., Chalamcharla, V.R., and Belfort, M. (2011). Learning to live together: mutualism between self-splicing introns and their hosts. *BMC Biol.*, 9:22. doi: 10.1186/1741-7007-9-22

Edgell, D.R., Belfort, M., and Shub, D.A. (2000). Barriers to intron promiscuity in bacteria. *J. Bacteriol.*, 182, 5281-5289

Edgell, D.R., Derbyshire, V., Van Roey, P., LaBonne, S., Stanger, M.J., Li, Z., Boyd, T.M., Shub, D.A., and Belfort, M. (2004). Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nat. Struct. Mol. Biol.*, 11, 936–944

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

Farré, J.C., Aknin, C., Araya, A., and Castandet, B. (2012). RNA editing in mitochondrial trans-introns is required for splicing. *PLoS ONE*, 7, e52644. doi: 10.1371/journal.pone.0052644

Fedorova, O., and Zingler, N. (2007). Group II introns: Structure, folding and splicing mechanism. *Biol. Chem.*, 388, 665–678. doi: 10.1515/BC.2007.090

Ferandon, C., Chatel Sel, K., Castandet, B., Castroviejo, M., and Barroso, G. (2008). The *Agrocybe aegerita* mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid integration site. *Fungal Genet Biol*, 45, 292–301

Ferrareze, P.A.G, et al. (2017). Transcriptional Analysis Allows Genome Reannotation and Reveals that *Cryptococcus gattii* VGII Undergoes Nutrient Restriction during Infection. *Microorganisms*, 5, 49. doi: 10.3390/microorganisms5030049

Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., and Gurr, S.J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484, 186-194

Fonseca, P.L.C., Badotti, F., De-Paula, R.B., Araújo, D.S., Bortolini, D.E., Del-Bem, L.E., et al. (2020). Exploring the relationship among divergence time and coding and non-coding elements in the shaping of fungal mitochondrial

genomes. *Front. Microbiol.* 11:765. doi: 10.3389/fmicb.2020.00765

Fonseca, P., De-Paula, R.B., Araújo, D.S., Tomé, L., Mendes-Pereira, T., Rodrigues, W., et al. (2021). Global characterization of fungal mitogenomes: new insights on genomic diversity and dynamism of coding genes and accessory elements. *Front. Microbiol.* 12, 787283. doi: 10.3389/fmicb.2021.787283

Formaggioni, A., Luchetti, A., and Plazzi, F. (2021). Mitochondrial genomic landscape: a portrait of the mitochondrial genome 40 years after the first complete sequence. *Life*, 11, 663. 10.3390/life11070663

Formighieri, E.F., Tiburcio, R.A., Armas, E.D., et al. (2008). The mitochondrial genome of the phytopathogenic basidiomycete *Moniliophthora perniciosa* is 109 kb in size and contains a stable integrated plasmid. *Mycol Res.*, 112, 1136–1152. doi: 10.1016/j.mycres.2008.04.014

Franco, M.E.E., López, S.M.Y., Medina, R., Lucentini, C.G., Troncozo, M.I., Pastorino, G.N., et al. (2017). The mitochondrial genome of the plant-pathogenic fungus *Stemphylium lycopersici* uncovers a dynamic structure due to repetitive and mobile elements. *PLoS One*, 12:e0185545. doi: 10.1371/journal.pone.0185545

Freel, K.C., Friedrich, A., and Schacherer, J. (2015). Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.*, 15:fov023. doi: 10.1093/femsyr/fov023

Frumkin, I., Yofe, I., Bar-Ziv, R., Gurvich, Y., Lu, Y.Y., Voichek, Y., Towers, R., Schirman, D., Krebber, H., and Pilpel, Y. (2019). Evolution of intron splicing towards optimized gene expression is based on various *Cis*- and *Trans*-molecular mechanisms. *PLoS Bio.*, e3000423. doi: 10.1371/journal.pbio.3000423

Fujisawa, T., Narikawa, R., Okamoto, S., Ehira, S., Yoshimura, H., Suzuki, I., Masuda, T., Mochimaru, M., Takaichi, S., Awai, K., Sekine, M., Horikawa, H., Yashiro, I., Omata, S., Takarada, H., Katano, Y., Kosugi, H., Tanikawa, S.,

Ohmori, K., Sato, N., Ikeuchi, M., Fujita, N., and Ohmori, M. (2010). Genomic structure of an economically important cyanobacterium, *Arthrospira* (*Spirulina*) *platensis* NIES39. *DNA Res.*, 17, 85-103

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. *Nature*, 537, 197–201. doi: 10.1038/nature19316

García-Rodríguez, F.M., Neira, J.L., Marcia, M., Molina-Sánchez, M.D., and Toro, N. (2019). A group II intron-encoded protein interacts with the cellular replicative machinery through the β-sliding clamp. *Nucleic Acids Res.*, 47, 7605–7617. doi: 10.1093/nar/gkz468

Garin, S., Levi, O., Cohen, B., Golani-Armon, A., and Arava, Y. S. (2020). Localization and RNA Binding of Mitochondrial Aminoacyl tRNA Synthetases. *Genes (Basel),* 11(10), 1185. https://doi.org/10.3390/genes11101185

Garin, S., Levi, O., Forrest, M. E., Antonellis, A., and Arava, Y. S. (2021). Comprehensive characterization of mRNAs associated with yeast cytosolic aminoacyl-tRNA synthetases. *RNA Biology*, 18(12), 2605-2616. doi: 10.1080/15476286.2021.1935116

Geese, W.J., and Waring, R.B. (2001). A comprehensive characterization of a group IB intron and its encoded maturase reveals that protein-assisted splicing requires an almost intact intron RNA. *Journal of Molecular Biology*, 308(4), 609-622.

Geng, C., and Paukstelis, P.J. (2014). An in vitro peptide complementation assay for CYT-18-dependent group I intron splicing reveals a new role for the N-terminus. *Biochemistry*, 53(8), 1311-1319. doi: 10.1021/bi401614h

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

Glanz S., and Kück U. (2009). Trans-splicing of organelle introns-a detour to con-

tinuous RNAs. *Bioessays*, 31, 921–934. doi: 10.1002/bies.200900036

Goddard, M. R., and Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. U. S. A.*, 96, 13880–13885. doi: 10.1073/pnas.96.24.13880

Gogarten, J.P., and Hilario, E. (2006). Inteins, introns, and homing endonucleases: Recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.*, 6, 94. doi: 10.1186/1471-2148-6-94

Goguel, V., Delahodde, A., and Jacq, C. (1992). Connections between RNA splicing and DNA intron mobility in yeast mitochondria: RNA maturase and DNA endonuclease switching experiments. *Molecular and Cellular Biology*, 12(2), 696-705.

Golik, P. (2023). RNA processing and degradation mechanisms shaping the mitochondrial transcriptome of budding yeasts. *IUBMB Life*, 2779. doi: 10.1002/iub.2779

Gomes, R.M.O.D.S., Silva, K.J.G.D., and Theodore, R.C. (2024). Group I introns: Structure, splicing and their applications in medical mycology. *Genet Mol Biol.*, 47Suppl 1(Suppl 1), e20230228. doi: 10.1590/1678-4685-GMB-2023-0228

Goryunov, D.V., Sotnikova, E.A., Goryunova, S.V., Kuznetsova, O.I., Logacheva, M.D., Milyutina, I.A., Fedorova, A.V., Fedosov, V.E., and Troitsky, A.V. (2021). The mitochondrial genome of nematodontous moss *Polytrichum commune* and Analysis of intergenic repeats distribution among bryophyta. *Diversity*, 13, 54. doi: 10.3390/d13020054

Gramespacher, J.A., Burton, A.J., Guerra, L.F., and Muir, T.W. (2019). Proximity induced splicing utilizing caged split inteins. *J. Am. Chem. Soc.*, 141,13708–13712. doi: 10.1021/jacs.9b05721

Grasso, V., Palermo, S., Sierotzki, H., Garibaldi, A., and Gisi, U. (2006). Cytochrome b gene structure and consequences tor resistance to Qo inhibitor

fungicides in plant pathogens. *Pest. Manag. Sci.*, 62, 465–472. doi: 10.1002/ps.1236

Gregan, J., Kolisek, M., and Schweyen, R.J. (2001). Mitochondrial Mg(2+) homeostasis is critical for group II intron splicing in vivo. *Genes. Dev.*, 15(17), 2229-37. doi: 10.1101/gad.201301

Griffiths, A.J.F (1995). Natural plasmids of filamentous fungi. *Microbiol Rev*, 59, 673–685

Griffiths, A.J.F. (1992). Fungal senescence. *Annu Rev Genet*, 26, 351–372

Grivell, L.A. (1995). Nucleo-mitochondrial interactions in mitochondrial gene expression. Crit *Rev Biochem Mol Biol.*, 30(2), 121-64. doi: 10.3109/10409239509085141

Guha, T. K., Wai, A., and Hausner, G. (2017). Programmable genome editing tools and their regulation for efficient genome engineering. *Comput. Struct. Biotechnol. J.*, 15, 146–160. doi: 10.1016/j.csbj.2016.12.006

Guha, T. K., Wai, A., Mullineux, S., and Hausner, G. (2018). The intron landscape of the mtDNA cytb gene among the Ascomycota: introns and intron-encoded open reading frames. *Mitochondrial DNA A: DNA Mapp. Seq. Anal.*, 29, 1015–1024. doi: 10.1080/24701394.2017.1404042

Guha, T.K., and Hausner, G. (2014). A homing endonuclease with a switch: Characterization of a twintron encoded homing endonuclease. *Fungal Genet. Biol.*, 65, 57–68. doi: 10.1016/j.fgb.2014.01.004

Guha, T.K., and Hausner, G. (2016). Using group II introns for attenuating the in vitro and in vivo expression of a homing endonuclease. *PLoS ONE*, 11, e0150097. doi: 10.1371/journal.pone.015009

Guijarro, B., Larena, I., Melgarejo, P., and De Cal, A. (2017). Adaptive conditions and safety of the application of Penicillium frequentans as a biocontrol agent on stone fruit. *Int J Food Microbiol.*, 254, 25-35. doi: 10.1016/j.ijfoodmicro.2017.05.004

Guo, L., and Liu, C.M. (2015). A single-nucleotide exon found in Arabidopsis. *Sci. Rep.*, 5, 18087.

Guo, W., Zhu, A., Fan, W., Adams, R.P., and Mower, J.P. (2020). Extensive Shifts from Cis- to Trans-splicing of Gymnosperm Mitochondrial Introns. *Mol. Biol. Evol.*, 37, 1615–1620. doi: 10.1093/molbev/msaa029

Guo, W.W., Moran, J.V., Hoffman, P.W., Henke, R.M., Butow, R.A., and Perlman, P.S. (1995). The mobile group I intron 3 alpha of the yeast mitochondrial COXI gene encodes a 35-kDa processed protein that is an endonuclease but not a maturase. *J Biol Chem.*, 270(26), 15563-15570

Gupta, K., Contreras, L.M., Smith, D., Qu, G., Huang, T., Spruce, L.A., Seeholzer, S.H., Belfort, M., and Van Duyne, G.D. (2014). Quaternary arrangement of an active, native group II intron ribonucleoprotein complex revealed by small-angle X-ray scattering. *Nucleic Acids Res.*, 42, 5347–5360. doi: 10.1093/nar/gku140

Haack, D.B., Rudolfs, B., Zhang, C., Lyumkis, D., and Toor, N. (2024). Structural basis of branching during RNA splicing. *Nat Struct Mol Biol.* 31(1):179-189. doi: 10.1038/s41594-023-01150-0.

Haack, D.B., and Toor, N. (2020). Retroelement origins of pre-mRNA splicing. *Wiley Interdiscip Rev. RNA.*, 11, e1589. doi: 10.1002/wrna.1589

Haack, D.B., Yan, X., Zhang, C., Hingey, J., Lyumkis, D., Baker, T.S., and Toor, N. (2019). Cryo-EM Structures of a Group II Intron Reverse Splicing into DNA. *Cell.* 178(3):612-623.e12. doi: 10.1016/j.cell.2019.06.035.

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

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

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

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

Hao, W. (2022). From genome variation to molecular mechanisms: what we have learned from yeast mitochondrial genomes? *Front. Microbiol.*, 13:806575. doi: 10.3389/fmicb.2022.806575

Harrington, T. C. (1988). "*Leptographium* species, their distributions, hosts and insect vectors" in *Leptographium Root Diseases of Conifers.* eds. T. C. Harrington and F. W. Cobbs Jr. (St. Paul, MN: APS Press), 1–40.

Haugen, P., Simon, D.M., and Bhattacharya, D. (2005). The natural history of group I introns. *Trends Genet.*, 21, 111–119. doi: 10.1016/j.tig.2004.12.007

Hausner, G. (2003). Fungal Mitochondrial Genomes. *Fungal Genomics*, 3, 101

Hausner, G. (2012). "Introns, mobile elements and plasmids" in *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression.* ed. C. E. Bullerwell (Berlin: Springer Verlag), 329–358.

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

Hausner, G., Iranpour, M., Kim, J.-J., Breuil, C., Davis, C. N., Gibb, E. A., et al. (2005). Fungi vectored by the introduced bark beetle *Tomicus piniperda* in Ontario, Canada and comments on the taxonomy of *Leptographium lundbergii*, *Leptographium terebrantis*, *Leptographium truncatum*, and *Leptographium wingfieldii*. *Can. J. Bot.* 83, 1222–1237. doi: 10.1139/b05-095

Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E.R., Karol, K.G., McCourt, R.M., and Zimmerly, S. (2006). Origin and evolution of the chloro-

plast trnK (matK) intron: A model for evolution of group II intron RNA structures. *Mol. Biol. Evol.*, 23, 380–391. doi: 10.1093/molbev/msj047

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

Hausner, G., Reid, J., and Klassen, G.R. (1993c). On the subdivision of Ceratocystis s.l., based on partial ribosomal DNA sequences. *Canadian Journal of Botany*, 71, 52–63.

Hausner, G., Reid, J., and Klassen, GR. (1993a). Ceratocystiopsis: a reappraisal based on molecular criteria. *Mycological Research*, 97, 625–633.

Hawksworth, D. L. (2011). A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. *IMA Fungus* 2, 155–162. doi: 10.5598/imafungus.2011.02.02.06

Hayakawa, J., and Ishizuka, M. (2009). A group I self-splicing intron in the flagellin gene of the thermophilic bacterium *Geobacillus stearothermophilus*. *Biosci. Biotechnol. Biochem.*, 73, 2758- 2761.

Hayakawa, J., and Ishizuka, M. (2012). Temperature-dependent self-splicing group I introns in the flagellin genes of the thermophilic *Bacillus* species. *Biosci. Biotechnol. Biochem.*, 76, 410-413

Hepburn, N.J., Schmidt, D.W., and Mower, J.P. (2012). Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Mol. Biol. Evol.*, 29, 3111–3120. doi: 10.1093/molbev/mss130

Herbert, C.J., Labouesse, M., Dujardin, G., and Slonimski, P.P. (1988). The NAM2 proteins from S. cerevisiae and S. douglasii are mitochondrial leucyl-tRNA synthetases and are involved in mRNA splicing. *EMBO Journal*, 7, 473-483.

Hermanns, J., Asseburg, A., and Osiewacz, H.D. (1994). Evidence for a life span-prolonging effect of a linear plasmid in a longevity mutant of *Podospora anserina*. *Mol Gen Genet.*, 243(3), 297-307. doi: 10.1007/BF00301065.

Hill, G.E. (2020). Genetic hitchhiking, mitonuclear coadaptation, and the origins of mt DNA barcode gaps. *Ecology and Evolution,* 10(17), 9048–9059. https://doi.org/10.1002/ece3.6640

Himmelstrand, K., Olson, A., Brandström Durling, M., Karlsson, M., and Stenlid, J. (2014). Intronic and plasmid-derived regions contribute to the large mitochondrial genome sizes of Agaricomycetes. *Curr Genet.*, 60(4), 303-13. doi: 10.1007/s00294-014-0436-z. Epub 2014 Jul 11

Hong, L., and Hallick, R.B. (1994). A group III intron is formed from domains of two individual group II introns. *Genes Dev.*, 8, 1589–99

Ho, Y., Kim, S.J., and Waring, R.B. (1997). A protein encoded by a group I intron in *Aspergillus nidulans* directly assists RNA splicing and is a DNA endonuclease. *Proc. Natl. Acad. Sci. USA*, 94, 8994–8999. doi: 10.1073/pnas.94.17.8994

Hou, Y. M., Shiba, K., Mottes, C., and Schimmel, P. (1991). Sequence determination and modeling of structural motifs for the smallest monomeric aminoacyl-tRNA synthetase. *Proc. Natl. Acad. Sci. USA*, 88, 976-980.

Huchon, D., Szitenberg, A., Shefer, S., Ilan, M., and Feldstein, T. (2015). Mitochondrial group I and group II introns in the sponge orders Agelasida and Axinellida. *BMC Evol. Biol.*, 15, 278. doi: 10.1186/s12862-015-0556-1

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Humble, L.M., and Allen, E.A. (2006). Forest biosecurity: alien invasive species and vectored organisms. *Can. J. Plant Pathol.*, 28, S256–S269. doi: 10.1080/07060660609507383

Hyde, K.D., Norphanphoun, C., Maharachchikumbura, S., Bhat, D., Jones, E.,

Bundhun, D., et al. (2020). Refined families of sordariomycetes. *Mycosphere*, 21, 7019. doi: 10.5943/mycosphere/11/1/7

Jackson, S.A., Koduvayur, S., and Woodson, S.A. (2006). Self-splicing of a group I intron reveals partitioning of native and misfolded RNA populations in yeast. *RNA*, 12, 2149–2159. doi: 10.1261/rna.184206

Jacobs, K., and Wingfield, M.J. (2001). *Leptographium* species: tree pathogens, insect associates, and agents of blue-stain. *Am. Phytopathol. Soc.*, 55121, 207.

Jacobs, K., and Wingfield, M.J. (2013). An overview of *Leptographium* and *Grosmannia*. In *Ophiostomatoid fungi: Expanding frontiers*. Edited by K. A. Seifert, Z. W. Beerde, and M. J. Wingfield. CBS-KNAW Fungal Biodiversity Centre, Utrecht. pp.:47–56.

Jacobs, K., Bergdahl, D. R., Wingfield, M. J., Halik, S., Seifert, K. A., Bright, D. E., et al. (2004). *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. *Mycol. Res.*, 108, 411–418. doi: 10.1017/s0953756204009748

Jacquier, A., and Dujon, B. (1985). An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, 41, 383–394. doi: 10.1016/S0092-8674(85)80011-8

Jacquier, A., and Michel, F. (1987). Multiple exon-binding sites in class II self-splicing introns. *Cell*, 50, 17–29. doi: 10.1016/0092-8674(87)90658-1

Jacquier, A., and Rosbash, M. (1986). Efficient trans-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5′ exon-intron interaction. *Science*, 234, 1099–1104. doi: 10.1126/science.2430332

Jankowiak, R., Ostafińska, A., Aas, T., Solheim, A., Bilański, P., Linnakoski, R., Mukhopadhyay, J., and Hausner, G. (2018). Three new *Leptographium* spp. (Ophiostomatales) infecting hardwood trees in Norway and Poland. *Antonie Van Leeuwenhoek*, 111, 2323–2347. doi: 10.1007/s10482-018-1123-8

Jankowiak, R., Strzałka, B, Bilański. P, et al. (2017). Two new *Leptographium* spp. reveal an emerging complex of hardwood-infecting species in the Ophiostomatales. *Antonie van Leeuwenhoek*, 110, 1537–1553

Jin,J. J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.*, 21, 241. doi: 10.1186/s13059-020-02154-5

Jo, B., and Choi, S.S. (2014). Introns: The functional benefits of introns in genomes. *Genom. Inform.*, 13, 112–118. doi: 10.5808/GI.2015.13.4.112

Johansen, S., and Emblem, Å. Mitochondrial Group I introns in hexacorals are regulatory genetic elements. In: Soto L.A., editor. *Advances in the Study of Benthic Zone.* IntechOpen; London, UK: 2020

Johansen, S., and Haugen, P. (2001). A new nomenclature of group I introns in ribosomal DNA. *RNA* 7, 935–936. doi: 10.1017/s1355838201010500

Johansen, S.D., Chi, S.I., Dubin, A., and Jørgensen, T.E. (2021). The mitochondrial genome of the sea anemone *Stichodactyla haddoni* reveals catalytic introns, insertion-like element, and unexpected phylogeny. *Life*, 11, 402. doi: 10.3390/life11050402

Kaer, K., and Speek, M. (2012). Intronic retroelements. *Mobile Gen. El.*, 2, 154–157. doi: 10.4161/mge.20774

Kamikawa, R., Shiratori, T., Ishida, K., Miyashita, H., and Roger, A.J. (2016). Group II Intron-Mediated Trans-Splicing in the Gene-Rich Mitochondrial Genome of an Enigmatic Eukaryote, *Diphylleia rotans*. *Genome Biol. Evol.*, 8, 458–466. doi: 10.1093/gbe/evw011

Kamper, U., Kuck, U., Cherniack, A.D., and Lambowitz, A.M. (1992). The mitochondrial tyrosyl-tRNA synthetase of *Podospora anserina* is a bifunctional enzyme active in protein synthesis and RNA splicing. *Molecular and Cellular Biology*, 12, 499-511.

Katoch, M., and Pull, S. (2017). Endophytic fungi associated with *Monarda citri-*

odora, an aromatic and medicinal plant and their biocontrol potential. *Pharm Biol.*, 55(1), 1528-1535. doi: 10.1080/13880209.2017.1309054

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30, 772–780. doi: 10.1093/molbev/mst010

Kempken, F. (1995). Horizontal transfer of a mitochondrial plasmid. *Mol Gen Genet.*, 248(1), 89-94. doi: 10.1007/BF02456617

Kennel, J.C., and Cohen, S.M. (2004). Fungal mitochondria: genomes, genetic elements and gene expression. In: Arora DK (ed) The handbook of fungal biotechnology, 2nd edn. Marcel Dekker Inc, New York, pp 131–143

Kennell, J.C., Moran, J.V., Perlman, P.S., Butow, R.A., and Lambowitz, A.M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell*, 73(1), 133-146.

Kernaghan, G., Mayerhofer, M., and Griffin, A. (2017). Fungal endophytes of wild and hybrid Vitis leaves and their potential for vineyard biocontrol. *Can J Microbiol.* 63(7), 583-595. doi: 10.1139/cjm-2016-0740

Kim, J.J., Kim, S.H., Lee, S., and Breuil, C. (2003). Distinguishing *Ophiostoma ips* and *Ophiostoma montium*, two bark beetle-associated sapstain fungi. *FEMS Microbiol Lett.*, 222(2), 187-92. doi: 10.1016/S0378-1097(03)00304-5

Kim, S.H., and Cech, T.R. (1987). Three-dimensional model of the active site of the self-splicing rRNA precursor of Tetrahymena. *Proc. Natl. Acad. Sci. USA.*, 84(24), 8788-8792

Klassen, R., and Meinhardt, F. (2007). Linear protein primed replicating plasmids in eukaryotic microbes. In: Meinhardt F, Klassen R (eds) Microbial linear plasmids. Springer, Berlin, Germany, pp 188–226

Ko, M., Choi, H., and Park, C. (2002). Group I self-splicing intron in the recA gene of Bacillus anthracis. *J Bacteriol.*, 184, 3917-3922

Kolondra, A., Labedzka-Dmoch, K., Wenda, J.M., Drzewicka, K., and Golik, P.

(2015). The transcriptome of *Candida albicans* mitochondria and the evolution of organellar transcription units in yeasts. *BMC Genomics*, 16, 827. doi: 10.1186/s12864-015-2078-z

Koonin, E.V., Senkevich, T.G., and Dolja, V.V. (2006). The ancient virus world and evolution of cells. *Biol. Direct.*, 1, 29. doi: 10.1186/1745-6150-1-29

Korovesi, A. G., Ntertilis, M., and Kouvelis, V. N. (2018). Mt-rps3 is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. *Mol. Phylogenet. Evol.*, 127, 74–86. doi: 10.1016/j.ympev.2018.04.037

Koster, C.C., Kleefeldt, A.A., van den Broek, M., Luttik, M., Daran, J.M., and Daran-Lapujade, P. (2023). Long-read direct RNA sequencing of the mitochondrial transcriptome of *Saccharomyces cerevisiae* reveals condition-dependent intron abundance. *Yeast*, doi: 10.1002/yea.3893

Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Pietrokovski, S., Baxter, S.M., and Derbyshire, V. (1999). Configuration of the catalytic GIY–YIG domain of intron endonucleases I-*Tev* I: Coincidence of computational and molecular findings. *Nucleic Acids Res.*, 27, 2115–2125. doi: 10.1093/nar/27.10.2115

Kroeger, T.S., Watkins, K.P., Friso, G., van Wijk, K.J., and Barkan, A. (2009). A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proc. Natl. Acad. Sci. USA*, 106, 4537–4542. doi: 10.1073/pnas.0812503106

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.*, 19, 1639–1645. doi: 10.1101/gr.092759.109

Kück, U., and Schmitt, O. (2021). The chloroplast *Trans*-Splicing RNA-Protein supercomplex from the green alga *Chlamydomonas reinhardtii*. *Cells*, 10, 290. doi: 10.3390/cells10020290

Kulik, T., Bilska, K., and Zelechowski, M. (2020). Promising perspectives for

detection, identification, and quantification of plant pathogenic fungi and oomycetes through targeting mitochondrial DNA. *Int. J. Mol. Sci.*, 21, 2645. doi: 10.3390/ijms21072645

Kulik, T., Van Diepeningen, A. D., and Hausner, G. (2021). Editorial: the significance of mitogenomics in mycology. *Front. Microbiol.*, 11, 628579. doi: 10.3389/fmicb.2020.628579

Labouesse, M. (1990). The yeast mitochondrial leucyl-tRNA synthetase is a splicing factor for the excision of several group I introns. *Molecular Genetics and Genomics*, 224, 209-221.

Labouesse, M., Dujardin, G., and Slonimski, P.P. (1985). The yeast nuclear gene NAM2 is essential for mitochondrial DNA integrity and can cure a mitochondrial RNA-maturase deficiency. *Cell*, 41, 133-143.

Lamb, M.R., Anziano, P.Q., Glaus, K.R., Hanson, D.K., Klappe, H.J., Perlman, P.S., and Mahler, H.R. (1983). Functional domains in introns. RNA processing intermediates in cis- and trans-acting mutants in the penultimate intron of the mitochondrial gene for cytochrome b. *J. Biol. Chem.*, 258, 1991–1999. doi: 10.1016/S0021-9258(18)33086-2

Lambert, A.R., Hallinan, J.P., Shen, B.W., Chik, J.K., Bolduc, J.M., Kulshina, N., Robins, L.I., Kaiser, B.K., Jarjour, J., Havens, K., et al. (2016). Indirect DNA Sequence Recognition and Its Impact on Nuclease Cleavage Activity. *Structure*, 24, 862–873. doi: 10.1016/j.str.2016.03.024

Lambowitz, A.M., and Perlman, P.S. (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends in Biochemical Sciences*, 15, 440-444.

Lambowitz, A.M., and Zimmerly, S. (2004). Mobile group II introns. *Annu Rev. Genet.*, 38, 1–35. doi: 10.1146/annurev.genet.38.072902.091600

Lambowitz, A.M., and Zimmerly, S. (2011). Group II introns: Mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol.*, 3, a003616. doi:

10.1101/cshperspect.a003616

Lambowitz, A.M., Caprara, M.G., Zimmerly, S., and Perlman, P.S. (1999). Group I and group II ribozymes as RNPs: Clues to the past and guides to the future. In R. F. Gesteland, T. R. Cech, & J. F. Atkins (Eds.), *The RNA World* (2nd ed., pp. 451-485). Cold Spring Harbor Laboratory Press.

Lamech, L. T., Saoji, M., Paukstelis, P. J., & Lambowitz, A. M. (2016). Structural Divergence of the Group I Intron Binding Surface in Fungal Mitochondrial Tyrosyl-tRNA Synthetases That Function in RNA Splicing. *Journal of Biological Chemistry*, 291(22), 11911-11927. doi: 10.1074/jbc.M116.725390

Lang, B. F. (2018). "Mitochondrial genomes in fungi" in *Molecular Life Sciences*. eds. R. D. Wells, J. S. Bond, J. Klinman, and B. S. S. Masters (New York, NY: Springer), 722–728.

Lang, B.F., Beck, N., Prince, S., Sarrasin, M., Rioux, P., and Burger, G. (2023). Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. *Front. Plant Sci.*, 14, 1222186. doi: 10.3389/fpls.2023.1222186

Lang, B. F., Laforest, M. J., and Burger, G. (2007). Mitochondrial introns: a critical view. *Trends Genet.*, 23, 119–125. doi: 10.1016/j.tig.2007.01.006

Lang, B.F., Lavrov, D., Beck, N., and Steinberg, S.V. (2012). "Mitochondrial tRNA structure, identity, and evolution of the genetic code" in *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*. ed. C. E. Bullerwell (Berlin Heidelberg: Springer-Verlag), 431–474.

Lang, B.F., O'Kelly, C., Nerad, T., Gray, M.W., and Burger, G. (2002). The closest unicellular relatives of animals. *Curr. Biol.*, 12, 1773–1778. doi: 10.1016/S0960-9822(02)01187-9

LaRoche-Johnston, F., Monat, C., Coulombe, S., and Cousineau, B. (2018). Bacterial group II introns generate genetic diversity by circularization and *trans*-splicing from a population of intron-invaded mRNAs. *PLoS Gen.*, 14,

e1007792. doi: 10.1371/journal.pgen.1007792

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Lasker, B.A., Smith, G.W., Kobayashi, G.S., Whitney, A.M., and Mayer L.W. (1998). Characterization of a single group I intron in the 18S rRNA gene of the pathogenic fungus *Histoplasma capsulatum*. *Med Mycol.*, 36, 205–212. doi: 10.1080/02681219880000311

Lavín, J.L., Oguiza, J.A, Ramírez, L., and Pisabarro, A.G. (2008). Comparative genomics of the oxidative phosphorylation system in fungi. *Fungal Genet Biol.*, 45(9), 1248-1256

Lazarus, C.M., Earl, A.J., Turner, G., and Kuntzel, H. (1980). Amplification of a mitochondrial DNA sequence in the cytoplasmically inherited "ragged" mutant of *Aspergillus amstelodami*. *Eur. J. Biochem.*, 106, 633–641

Lazowska, J., Claisse, M., Gargouri, A., Kotylak, Z., and Spyridakis, A. (1989). Protein encoded by the third intron of cytochrome b gene in Saccharomyces cerevisiae is an mRNA maturase: Analysis of mitochondrial mutants, RNA transcripts proteins and evolutionary relationships. *Journal of Molecular Biology*, 205(2), 275-289. doi: 10.1016/0022-2836(89)90341-0

Lazowska, J., Jacq, C., and Slonimski, P. (1980). Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. *Cell*, 22, 333-348.

Lazowska, J., Meunier, B., and Macadre, C. (1994). Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. *EMBO Journal*, 13(20), 4963-4972.

Leinonen, R., Sugawara, H., and Shumway, M., International Nucleotide Sequence Database Collaboration. (2011). The sequence read archive. *Nucleic Acids Res.*, 39, D19–D21. doi: 10.1093/nar/gkq1019

Li, Q., Wu, P., Li, L., Feng, H., Tu, W., Bao, Z., Xiong, C., Gui, M., and Huang, W. (2021). The first eleven mitochondrial genomes from the ectomycorrhizal fungal genus (*Boletus*) reveal intron loss and gene rearrangement. *Int. J. Biol Macromol.*, 172, 560–572. doi: 10.1016/j.ijbiomac.2021.01.087

Li, Q., Xiao, W., Wu, P., Zhang, T., Xiang, P., Wu, Q., et al. (2023). The first two mitochondrial genomes from *Apiotrichum* reveal mitochondrial evolution and different taxonomic assignment of Trichosporonales. *IMA Fungus*, 14:7. doi: 10.1186/s43008-023-00112-x

Li, X., Li, L., Bao, Z., Tu, W., He, X., Zhang, B., et al. (2020). The 287,403 bp mitochondrial genome of ectomycorrhizal fungus *Tuber calosporum* reveals intron expansion, tRNA loss, and gene rearrangement. *Front. Microbiol.*, 11:591453. doi: 10.3389/fmicb.2020.591453

Li, Y., Hu, XD., Yang, RH. et al. (2015). Complete mitochondrial genome of the medicinal fungus *Ophiocordyceps sinensis*. *Sci Rep*, 5, 13892. https://doi.org/10.1038/srep13892

Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., et al. (2021). A genome-scale phylogeny of the kingdom Fungi. *Curr. Biol.*, 31, 1653–1665.e5. doi: 10.1016/j.cub.2021.01.074

Li, Y.I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.*, 25, 1–13.

Liu, N., Dong, X., Hu, C., Zeng, J., Wang, J., Wang, J., Wang, H.W., and Belfort, M. (2020). Exon and protein positioning in a pre-catalytic group II intron RNP primed for splicing. *Nucleic Acids Res.*, 48, 11185–11198. doi: 10.1093/nar/gkaa773

Liu, T., and Pyle, A.M. (2021). Discovery of highly reactive self-splicing group II introns within the mitochondrial genomes of human pathogenic fungi, *Nucleic Acids Res.*, 49, 21(2), 12422–12432

Liu, T., and Pyle, A.M. (2024). Highly Reactive Group I Introns Ubiquitous in Pathogenic Fungi. *J Mol Biol.*, 5, 436(8):168513. doi: 10.1016/j.jmb.2024.

Loizos, N., Tillier, E. R., and Belfort, M. (1994). Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc. Natl. Acad. Sci. U. S. A.*, 91, 11983–11987. doi: 10.1073/pnas.91.25.11983

Longo, A., Leonard, C.W., Bassi, G.S., Berndt, D., Krahn, J.M., Hall, T.M., and Weeks, K. M. (2005). Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nature Structural & Molecular Biology*, 12(9), 779-787.

Lykke-Andersen, J., Aagaard, C., Semionenkov, M., and Garrett, R.A. (1997). Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem Sci.*, 22(9), 326-31. doi: 10.1016/s0968-0004(97)01113-4

Lynch, M., Koskella, B., and Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. *Science*, 311, 1727–1730. doi: 10.1126/science.1118884

Ma, Q., Geng, Y., Li, Q., Cheng, C., Zang, R., Guo, Y., et al. (2022). Comparative mitochondrial genome analyses reveal conserved gene arrangement but massive expansion/contraction in two closely related *Exserohilum* pathogens. *Comput. Struct. Biotechnol. J.*, 20, 1456–1469. doi: 10.1016/j.csbj.2022.03.016

Macías-Rubalcava, M.L., and Sánchez-Fernández, R.E. (2017). Secondary metabolites of endophytic Xylaria species with potential applications in medicine and agriculture. *World J Microbiol Biotechnol.*, 33(1), 15

Maheshwari, R., and Navaraj, A. (2008). Senescence in fungi: the view from *Neurospora*. *FEMS Microbiol Lett.*, 280(2), 135-43. doi: 10.1111/j.1574-6968.2007.01027.x.

Malbert, B., Labaurie, V., Dorme, C., and Paget, E. (2023). Group I intron as a potential target for antifungal compounds: development of a trans-

splicing high-throughput screening strategy. *Molecules*, 28, 4460. doi: 10.3390/molecules28114460

Malik, S., Upadhyaya, K.C., and Khurana, S.P. (2017). Phylogenetic Analysis of Nuclear-Encoded RNA Maturases. *Evol. Bioinform. Online.*, 13, 1176934317710945. doi: 10.1177/1176934317710945

Martin, W., and Koonin, E.V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. *Nature*, 440, 41–45. doi: 10.1038/nature04531

Martinis, S.A., Plateau, P., Cavarelli, J., and Florentz, C. (1999a). Aminoacyl-tRNA synthetases: a family of expanding functions. *EMBO Journal*, 18, 4591-4596.

Martinis, S.A., Plateau, P., Cavarelli, J., and Florentz, C. (1999b). Aminoacyl-tRNA synthetases: a new image for a classical family. *Biochimie*, 81, 683-700.

Massoumi Alamouti, S., Kim, J. J., and Breuil, C. (2006). A new *Leptographium* species associated with the northern spruce engraver, *Ips perturbatus*, in Western Canada. *Mycologia*, 98, 149–160. doi: 10.3852/mycologia.98.1.149

Massoumi Alamouti, S., Kim, J. J., Humble, L. M., Uzunovic, A., and Breuil, C. (2007). Ophiostomatoid fungi associated with the northern spruce engraver, *Ips perturbatus*, in Western Canada. *Antonie Van Leeuwenhoek* ,91, 19–34. doi: 10.1007/s10482-006-9092-8

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., Cavanagh, S., Dunny, G.M., Belfort, M., and Lambowitz, A.M. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: Biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.*, 11, 2910–2924. doi: 10.1101/gad.11.21.2910

Mayers, C.G., Harrington, T.C., Wai, A., and Hausner, G. (2021). Recent and Ongoing Horizontal Transfer of Mitochondrial Introns Between Two Fungal Tree

Pathogens. *Front Microbiol.*, 12, 656609. doi: 10.3389/fmicb.2021.656609

McAllister, C. H., Fortier, C. E., St Onge, K. R., Sacchi, B. M., Nawrot, M. J., Locke, T., et al. (2018). A novel application of RNase H2-dependent quantitative PCR for detection and quantification of *Grosmannia clavigera*, a mountain pine beetle fungal symbiont, in environmental samples. *Tree Physiol.*, 38, 485–501. doi: 10.1093/treephys/tpx147

McNeil, B.A., Semper, C., and Zimmerly, S. (2016). Group II introns: Versatile ribozymes and retroelements. *Wiley Interdisciplinary Reviews: RNA*, 7(3), 341-355.

Megarioti, A.H., and Kouvelis, V.N. (2020). The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). *Genome Biol. Evol.*, 12, 1337–1354. doi: 10.1093/gbe/evaa126

Meng, Q., Zhang, Y., and Liu, X.Q. (2007). Rare group I intron with insertion sequence element in a bacterial ribonucleotide reductase gene. *J. Bacteriol.*, 189, 2150-2154

Michel, F., and Ferat, J.L. (1995). Structure and activities of group II introns. *Annu. Rev. Biochem.*, 64, 435–446. doi: 10.1146/annurev.bi.64.070195.002251

Michel, F., and Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, 216, 585–610. doi: 10.1016/0022-2836(90)90386-Z

Michel, F., Costa, M., and Westhof, E. (2009). The ribozyme core of group II introns: A structure in want of partners. *Trends Biochem. Sci.*, 34, 189–199. doi: 10.1016/j.tibs.2008.12.007

Mitsuhashi, H., Homma, S., Beermann, M.L., Ishimaru, S., Takeda, H., Yu, B.K., Liu, K., Duraiswamy, S., Boyce, F.M., and Miller, J.B. (2019). Efficient system for upstream mRNA trans-splicing to generate covalent, head-to-tail, protein multimers. *Sci. Rep.*, 9, 2274. doi: 10.1038/s41598-018-36684-7

Mo, D., Li, X., Raabe, C.A., Cui, D., Vollmar, J.F., Rozhdestvensky, T.S., Skryabin, B.V., and Brosius, J. (2019). A universal approach to investigate circRNA protein coding function. *Sci. Rep.*, 9, 11684. doi: 10.1038/s41598-019-48224-y

Mohr, G., and Lambowitz, A.M. (2003). Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res.*, 31, 647–652. doi: 10.1093/nar/gkg153

Molina-Sánchez, M.D., López-Contreras, J.A., Toro, N., and Fernández-López, M. (2015). Genomic characterization of Sinorhizobium meliloti AK21, a wild isolate from the Aral Sea Region. *Springerplus*, 4, 259. doi: 10.1186/s40064-015-1062-z

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

Moran, J.V., Zimmerly, S., Eskes, R., Kennell, J.C., Lambowitz, A.M., Butow, R.A., and Perlman, P.S. (1995). Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Molecular and Cellular Biology*, 15(5), 2828-2838.

Mota, E.M., and Collins, R.A. (1988). Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. *Nature*, 332, 654–656. doi: 10.1038/332654a0

Mower, J.P., Stefanović, S., Hao, W., Gummow, J.S., Jain, K., Ahmed, D., and Palmer, J.D. (2010). Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.*, 8, 150. doi:10.1186/1741-7007-8-150

Mukhopadhyay, J., and Hausner, G. (2021). Organellar Introns in Fungi, Algae, and Plants. *Cells*, 10(8), 2001. doi: 10.3390/cells10082001

Mukhopadhyay, J., Wai, A., and Hausner, G. (2023). The mitogenomes of *Leptographium aureum*, *Leptographium* sp., and *Grosmannia fruticeta*: expansion by introns. *Front Microbiol*, 14, 1240407. doi: 10.3389/fmicb.2023.1240407.

Mukhopadhyay, J., Wai, A., Hutchison, L.J., and Hausner, G. (2022). The mitogenome of *Urnula craterium*. *Can. J. Microbiol.*, 68(8), 561-568. doi: 10.1139/cjm-2022-0012.

Müller, W.E., Böhm, M., Grebenjuk, V.A., Skorokhod, A., Müller, I.M., and Gamulin, V. (2002). Conservation of the positions of metazoan introns from sponges to humans. *Gene*, 295, 299–309. doi: 10.1016/S0378-1119(02)00690-X

Mullineux, S.T., Costa, M., Bassi, G.S., Michel, F., and Hausner, G. (2010). A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. *RNA*, 16, 1818–1831. doi: 10.1261/rna.2184010

Mullineux, S.T., Willows, K., and Hausner, G. (2011). Evolutionary dynamics of the mS952 intron: A novel mitochondrial group II intron encoding a LAGLIDADG homing endonuclease gene. *J. Mol. Evol.*, 72, 433–449. doi: 10.1007/s00239-011-9442-7

Muñoz-Gómez, S.A., Wideman, J.G., Roger, A.J., and Slamovits, C.H. (2017). The origin of mitochondrial cristae from alphaproteobacteria. *Mol. Biol. Evol.*, 34, 943–956. 10.1093/molbev/msw298

Nadimi, M., Beaudet, D., Forget, L., Hijri, M., and Lang, B.F. (2012). Group I intron-mediated trans-splicing in mitochondria of *Gigaspora rosea* and a robust phylogenetic affiliation of arbuscular mycorrhizal fungi with Mortierellales. *Mol. Biol. Evol.*, 29, 2199–2210. doi: 10.1093/molbev/mss088

Nagy, V., Pirakitikulr, N., Zhou, K.I., Chillón, I., Luo, J., and Pyle, A.M. (2013). Predicted group II intron lineages E and F comprise catalytically active ri-

bozymes. *RNA*, 19, 1266–1278. doi: 10.1261/rna.039123.113

Nannfeldt, J.A. (1932). Studien uber die Morphologie und Systematik der nicht-lichenisierten inoperculaten Discomyceten. *Nova Acta Regiae Societatis Scientiarum Upsaliensis IV*, 8, 1–368

Nargang, F.E., and Kennell, J.C. (2010). Mitochondria and respiration. In: Borkovich KA, Ebbole DJ (eds) Cellular and molecular biology of filamentous fungi. ASM Press, Washington DC, pp 155–178

Narladkar, B.W., Shivpuje, P.R., and Harke, P.C. (2015). Fungal biological control agents for integrated management of Culicoides spp. (Diptera: Ceratopogonidae) of livestock. *Vet World.*, 8(2), 156-163. doi: 10.14202/vetworld.2015.156-163

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

Nguyen, T.H.M., Sondhi, S., Ziesel, A., Paliwal, S., and Fiumera, H.L. (2020). Mitochondrial-nuclear coadaptation revealed through mtDNA replacements in *Saccharomyces cerevisiae*. *BMC Evolutionary Biology,* 20(1), 128. https://doi.org/10.1186/s12862-020-01685-6

Novikova, O., and Belfort, M. (2017). Mobile Group II Introns as ancestral eukaryotic elements. *Trends Genet.*, 33, 773–783. doi: 10.1016/j.tig.2017.07.009

Ohman-Heden, M., Ahgren-Stalhandske, A., Hahne, S., and Sjoberg, B.M. (1993). Translation across the 5′-splice site interferes with autocatalytic splicing. *Mol Microbiol.*, 7, 975–982

Olchowecki, A., and Reid, J. (1974). Taxonomy of the genus Ceratocystis in Manitoba. *Can. J. Bot.*, 52,1675-1711

Oro, L., Feliziani, E., Ciani, M., Romanazzi, G., and Comitini, F. (2018). Volatile organic compounds from *Wickerhamomyces anomalus*, *Metschnikowia pulcherrima* and *Saccharomyces cerevisiae* inhibit growth of decay causing fungi and control postharvest diseases of strawberries. *Int J Food Microbiol.*, 265,

18-22. doi: 10.1016/j.ijfoodmicro.2017.10.027

Osigus, H.J., Eitel,M., and Schierwater, B. (2017). Deep RNA sequencing reveals the smallest known mitochondrial micro exon in animals: the placozoan cox1 single base pair exon. *PLoS One*, 12, e0177959.

Østergaard, L.H., and Olsen, H.S. (2010). Industrial applications of fungal enzymes. In: Industrial Applications. Springer Berlin Heidelberg. pp. 269-290

Overkamp, K.M., Bakker, B.M., Kötter, P., van Tuijl, A., de Vries, S., van Dijken, J.P., Pronk, J.T. (2000). In vivo analysis of the mechanisms for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria. *J Bacteriol.*, 182(10), 2823-2830

Palmer, J.D., Logsdon, J.M., Jr. (1991). The recent origins of introns. *Curr Opin Genet. Dev.*, 1, 470–477. doi: 10.1016/S0959-437X(05)80194-7

Pang, Y.L., Poruri, K., and Martinis, S.A. (2014). tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdisciplinary Reviews: RNA*, 5(4), 461-480. doi: 10.1002/wrna.1224

Paquin, B., Kathe, S.D., Nierzwicki-bauer, S.A., and Shub D.A. (1997). Origin and evolution of group I introns in cyanobacterial tRNA genes. Origin and Evolution of Group I Introns in Cyanobacterial tRNA Genes. *J. Bacteriol.*, 179, 6798–6806. 10.1128/jb.179.21.6798-6806.1997

Parenteau, J., and Elela, S.A. (2019). Introns: Good day junk is bad day treasure. *Trends Genet.*, 35, 923–934. doi: 10.1016/j.tig.2019.09.010

Peebles, C.L., Benatan, E.J., Jarrell, K.A., and Perlman, P.S. (1987). Group II intron self-splicing: Development of alternative reaction conditions and identification of a predicted intermediate. *Cold Spring Harb Symp Quant. Biol.*, 52, 223–232. doi: 10.1101/SQB.1987.052.01.027

Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrell, K.A., and Cheng, H.L. (1986). A self-splicing RNA excises an intron lariat. *Cell*, 44, 213–223. doi: 10.1016/0092-8674(86)90755-5

Pfeifer, A., Martin, B., Kämper, J., and Basse, C.W. (2012). The mitochondrial LSU rRNA group II intron of *Ustilago maydis* encodes an active homing endonuclease likely involved in intron mobility. *PLoS ONE*, 7, e49551. doi: 10.1371/journal.pone.0049551

Plattner, A., Kim, J.J., Reid, J., Hausner, G., Woon, Y.L., Yamaoka, Y. and Breuil, C. (2009). Resolving taxonomic and phylogenetic incongruence within the species *Ceratocystiopsis minuta*. *Mycologia*, 101, 878-887

Podar, M., Chu, V.T., Pyle, A.M., Perlman, P.S. (1998). Group II intron splicing in vivo by first-step hydrolysis. *Nature*, 391, 915–918. doi: 10.1038/36142

Pogoda, C.S., Keepers, K.G., Nadiadi, A.Y., Bailey, D.W., Lendemer, J.C., Tripp, E.A., and Kane, N. C. (2019). Genome streamlining via complete loss of introns has occurred multiple times in lichenized fungal mitochondria. *Ecology and evolution*, 9(7), 4245–4263. https://doi.org/10.1002/ece3.5056

Prince, S., Munoz, C., Filion-Bienvenue, F., Rioux, P., Sarrasin, M., and Lang, B.F. (2022). Refining mitochondrial intron classification with ERPIN: identification based on conservation of sequence plus secondary structure motifs. *Front. Microbiol.*, 13, 866187. doi: 10.3389/fmicb.2022.866187

Putzeys, L., Boon, M., Lammens, E.M., Kuznedelov, K., Severinov, K., and Lavigne, R. (2022). Development of ONT-cappable-seq to unravel the transcriptional landscape of Pseudomonas phages. *Comput Struct Biotechnol J.*, 20, 2624-2638. doi: 10.1016/j.csbj.2022.05.034.

Putzeys, L., Intizar, D., Lavigne, R., and Boon, M. (2024). Obtaining Detailed Phage Transcriptomes Using ONT-Cappable-Seq. *Methods Mol Biol.*, 2793, 207-235. doi: 10.1007/978-1-0716-3798-2_14.

Pyle, A.M. (2010). The tertiary structure of group II introns: Implications for biological function and evolution. *Crit Rev. Biochem Mol. Biol.*, 45, 215–232

Qu, G., Kaushal, P.S., Wang, J., Shigematsu, H., Piazza, C.L., Agrawal, R.K., Belfort, M., Wang, H.W. (2016). Structure of a group II intron in complex

with its reverse transcriptase. *Nat. Struct. Mol. Biol.*, 23, 549–557. doi: 10.1038/nsmb.3220

Raghavan, R., and Minnick, M.F. (2009). Group I introns and inteins: Disparate origins but convergent parasitic strategies. *J. Bacteriol.*, 191, 6193–6202. doi: 10.1128/JB.00675-09

Rane, K.K., and Tattar, T.A. (1987). Pathogenicity of blue-stain fungi associated with *Dendroctonusterebrans*. *Plant Dis,* 71, 8798–8883. 10.1094/PD-71-0879

Rangan, P., and Woodson, S.A. (2003). Structural requirement for Mg2+ binding in the group I intron core. *J Mol Biol.* 329(2), 229-38. doi: 10.1016/s0022-2836(03)00430-3

Réblová, M., Gams, W., and Seifert, KA. (2011). *Monilochaetes* and allied genera of the *Glomerellales*, and a reconsideration of families in the *Microascales*. *Studies in Mycology*, 68, 163–191

Repar, J., and Warnecke, T. (2017). Mobile introns shape the genetic diversity of their host genes. *Genetics*, 205, 1641–1648. doi: 10.1534/genetics.116.199059

Rest, J.S., and Mindell, D.P. (2003). Retroids in archaea: phylogeny and lateral origins. *Mol Biol Evol.*, 20(7), 1134-42. doi: 10.1093/molbev/msg135.

Rho, S.B., Lincecum Jr., T.L., and Martinis, S.A. (2002). An inserted region of leucyl-tRNA synthetase plays a critical role in group I intron splicing. *EMBO Journal*, 21, 6874-6881

Ribera, J., Fink, S., Bas, M.D., and Schwarze, F.W. (2017). Integrated control of wood destroying basidiomycetes combining Cu-based wood preservatives and Trichoderma spp. *PLoS One*, 12(4), e0174335. doi: 10.1371/journal.pone.0174335

Richards, T.A., Leonard, G., and Wideman, J.G. (2017). What defines the "Kingdom" Fungi? *Microbiol Spectr.*, 5(3). doi: 10.1128/microbiolspec.FUNK-0044-2017

Robart, A.R., and Zimmerly, S. (2005). Group II intron retroelements: Function and

diversity. *Cytogenet. Genome Res.*, 110, 589–597. doi: 10.1159/000084992

Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K.R., and Toor, N. (2014). Crystal structure of a eukaryotic group II intron lariat. *Nature*, 514, 193–197. doi: 10.1038/nature13790

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat Biotechnol.*, 29(1), 24-6. doi: 10.1038/nbt.1754

Robinson-Jeffrey, R.C., and Davidson, R.W. (1968). Three new *Europhium* species with *Verticicladiella* imperfect states on blue-stained pine. *Can. J. Bot.*, 46, 1523–1527. doi: 10.1139/b68-210

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, 61, 539–542. doi: 10.1093/sysbio/sys029

Rose, A.B. (2019). Introns as Gene Regulators: A brick on the accelerator. *Front. Genet.*, 9, 672. doi: 10.3389/fgene.2018.00672

Rudan, M., Bou Dib, P., Musa, M., Kanunnikau, M., Sobočanec, S., Rueda, D., et al. (2018). Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *elife* , 7:e35330. doi: 10.7554/eLife.35330

Rudi, K., Fossheim, T., and Jakobsen, K.S. (2002). Nested evolution of a tRNA(Leu)(UAA) group I intron by both horizontal intron transfer and recombination of the entire tRNA locus. *J Bacteriol.*, 184(3), 666-71. doi: 10.1128/JB.184.3.666-671.2002

Rumbold, C.T. (1931) Two blue-stain fungi associated with bark-beetle infestation of pines. *Journal of Agricultural Research*, 43, 847–873

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics*,

16, 944–945. doi: 10.1093/bioinformatics/16.10.944

Saccone, C., Gissi, C., Lanave, C., Larizza, A., Pesole, G., and Reyes, A. (2000). Evolution of the mitochondrial genetic system: an overview. *Gene*, 261, 153–159. doi: 10.1016/s0378-1119(00)00484-4

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

Salinas-Giegé, T., Giegé, R., and Giegé, P. (2015). tRNA biology in mitochondria. *Int. J. Mol. Sci.*, 16, 4518–4559. doi: 10.3390/ijms16034518

Salman, V., Amann, R., Shub, D.A., and Schulz-Vogt, H.N. (2012). Multiple self-splicing introns in the 16S rRNA genes of giant sulfur bacteria. *Proc. Natl. Acad. Sci. USA.*, 109, 4203-4208

Sandor, S., Zhang, Y., and Xu, J. (2018). Fungal mitochondrial genomes and genetic polymorphisms. *Appl. Microbiol. Biotechnol.*, 102, 9433–9448. doi: 10.1007/s00253-018-9350-5

Sankoff, D., Leduc, G., Antoine, N., Paquin, B., Lang, B.F., and Cedergren, R. (1992). Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci. USA.*, 89(14), 6575-6579

Santini, A., and Migliorini, D. (2022). Invasive alien plant pathogens: the need of new detection methods. *Methods Mol. Biol.*, 2536, 111–118. doi: 10.1007/978-1-0716-2517-0_7

Scalley-Kim, M., McConnell-Smith, A., Stoddard, B.L. (2007). Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.*, 372, 1305–1319. doi: 10.1016/j.jmb.2007.07.052

Schimmel, P., and Ribas de Pouplana, L. (2000). Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends in Biochemical Sciences*, 25, 207-209.

Schmidt, C.A., and Matera, A.G. (2020). tRNA introns: Presence, processing, and purpose. *Wiley Interdiscip Rev. RNA.*, 11, e1583. doi: 10.1002/wrna.1583

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A.,

and Chen, W. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci USA*. 109(16), 6241-6. doi: 10.1073/pnas.1117018109

Schuster, A., Lopez, J.V., Becking, L.E., Kelly, M., Pomponi, S.A., Wörheide, G., Erpenbeck, D., and Cárdenas, P. (2017). Evolution of Group I introns in Porifera: New evidence for intron mobility and implications for DNA barcoding. *BMC Evol. Biol.*, 17, 1–21. doi: 10.1186/s12862-017-0928-9

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-9. doi: 10.1093/bioinformatics/btu153.

Seifert, K.A., De Beer, Z.W., and Wingfield, M.J. (2013). The ophiostomatoid fungi: expanding frontiers. CBS Biodiversity Series 12. CBS-KNAW Biodiversity Centre, Utrecht, The Netherlands

Seifert, K.A., Okada, G. (1993). *Graphium* anamorphs of *Ophiostoma* species and similar anamorphs of other Ascomycetes. In: *Ceratocystis and Ophiostoma: Taxonomy, Ecology and Pathogenicity*, (Wingfield MJ, Seifert KA, Webber J, eds.), 27–41. American Phytopathological Society Press, St. Paul, Minnesota

Sellem, C. H., di Rago, J. P., Lasserre, J. P., Ackerman, S. H., and Sainsard-Chanet, A. (2016). Regulation of aerobic energy metabolism in *Podospora anserina* by two paralogous genes encoding structurally different c-subunits of ATP synthase. *PLoS Genet.*, 12:e1006161. doi: 10.1371/journal.pgen.1006161

Sellem, C.H., and Belcour, L. (1994). The in vivo use of alternate 3'-splice sites in group I introns. *Nucleic Acids Res.*, 22, 1135–1137. doi: 10.1093/nar/22.7.1135

Sellem, C.H., and Belcour, L. (1997). Intron open reading frames as mobile elements and evolution of a group I intron. *Mol. Biol. Evol.*, 14, 518–526

Sellem, C.H., di Rago, J.P., Lasserre, J.P., Ackerman, S.H., and Sainsard-Chanet, A. (2016). Regulation of aerobic energy metabolism in *Podospora anserina* by two paralogous genes encoding structurally different c-subunits of ATP

synthase. *PLoS Genet.*, 12(7), e1006161. doi: 10.1371/journal.pgen.1006161

Sethuraman, J., Majer, A., Iranpour, M., and Hausner, G. (2009). Molecular evolution of the mtDNA encoded rps3 gene among filamentous ascomycetes fungi with an emphasis on the Ophiostomatoid fungi. *J Mol Evol.*,69(4), 372-85. doi: 10.1007/s00239-009-9291-9.

Shang, J., Yang, Y., Wu, L., Zou, M., and Huang, Y. (2018). The *S. pombe* mitochondrial transcriptome. *RNA.*, 24(9), 1241-1254. doi: 10.1261/rna.064477.117. Epub 2018 Jun 28

Sharp, P.A. (1994). Split genes and RNA splicing. *Cell,* 77*,* 805-815

Shivji, M.S., Stanhope, M.J., and Rogers, S.O. Group I introns ("spintrons") are present in shark nuclear ribosomal DNA internal transcribed spacers; Proceedings of the American Elasmobranch Society, 13th Annual Meeting; Seattle, WA, USA. 26 June–2 July 1997

Shumway, M., Cochrane, G., and Sugawara, H. (2010). Archiving next generation sequencing data. *Nucleic Acids Res.*, 38, D870–D871. doi: 10.1093/nar/gkp1078

Silas, S., Mohr, G., Sidote, D.J., Markham, L.M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A.M., and Fire, A.Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science*, 351(6276), 4234

Simon, D.M., Clarke, N.A., McNeil, B.A., Johnson, I., Pantuso, D., Dai, L., Chai. D., and Zimmerly, S. (2008). Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA*, 14, 1704–1713. doi: 10.1261/rna.1056108

Sinn, B.T., Barrett, C.G. (2020). Ancient Mitochondrial Gene Transfer between Fungi and the Orchids. *Mol. Biol. Evol.*, 37, 44–57. doi: 10.1093/molbev/msz198

Six, D. L., and Wingfield, M. J. (2011). The role of phytopathogenicity in bark

beetle-fungus symbioses: a challenge to the classic paradigm. *Annu. Rev. Entomol.*, 56, 255–272. doi: 10.1146/annurev-ento-120709-144839

Six, D.L., De Beer, Z.W., Duong, T.A., Carroll, A.L., and Wingfield, M.J. (2011). Fungal associates of the lodgepole pine beetle, *Dendroctonus murrayanae*. *Antonie Van Leeuwenhoek*, 100, 231–244. doi: 10.1007/s10482-011-9582-1

Smathers, C.M., and Robart, A.R. (2019). The mechanism of splicing as told by group II introns: Ancestors of the spliceosome. *Biochim. Biophys. Acta (BBA) Gene Regul. Mech.*, 1862, 194390. doi: 10.1016/j.bbagrm.2019.06.001

Smith, D., Zhong, J., Matsuura, M., Lambowitz, A.M., and Belfort, M. (2005). Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev.*, 19, 2477–2487. doi: 10.1101/gad.1345105

Song, L., and Florea, L. (2015). Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience*, 4, 48. doi: 10.1186/s13742-015-0089-y

Stahley, M.R., and Strobel, S.A. (2005). Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science*, 309(5740), 1587-1590

Stajich, J. E. (2017). Fungal genomes and insights into the evolution of the kingdom. *Microbiol. Spectrum*, 5. doi: 10.1128/microbiolspec.FUNK-0055-2016

Starzyk, R.M., Webster, T.A., and Schimmel, P. (1987). Evidence for dispensable sequences inserted into a nucleotide fold. *Science*, 237, 1614-1618.

Stern, D.B., Goldschmidt-Clermont, M., and Hanson, M.R. (2010). Chloroplast RNA metabolism. *Annu Rev. Plant Biol.*, 61, 125–155. doi: 10.1146/annurev-arplant-042809-112242

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

Stoddard, B.L. (2005). Homing endonuclease structure and function. *Q Rev. Biophys.*, 38, 49–95. doi: 10.1017/S0033583505004063

Suh, D.Y., Hyun, M.W., Kim, J.J., Son, S.Y., and Kim, S.H. (2013). *Ophios-*

toma ips from Pinewood Nematode Vector, Japanese Pine Sawyer Beetle (*Monochamus alternatus*), in Korea. *Mycobiology*, 41(1), 59-62. doi: 10.5941/MYCO.2013.41.1.59

Suh, S.O., Jones, K.G., Blackwell, M. (1999). A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: Evidence for a new major class of Group I introns. *J. Mol. Evol.*, 48, 493–500. doi: 10.1007/PL00006493

Suzuki, H., Kameyama, T., Ohe, K., Tsukahara, T., and Mayeda, A. (2013). Nested introns in an intron: Evidence of multi-step splicing in a large intron of the human dystrophin pre-mRNA. *FEBS Lett.*, 587, 555–561. doi: 10.1016/j.febslet.2013.01.057

Szczepanek, T., and 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.

Takahara, K., Schwarze, U., Imamura, Y., Hoffman, G.G., Toriello, H., Smith, L.T., Byers, P.H., and Greenspan, D.S. (2002). Order of intron removal influences multiple splice outcomes, including a two-exon skip, in a *col5a1* acceptor-site mutation that results in abnormal pro-α1(v) n-propeptides and ehlers-danlos syndrome type I. *AJHJ.*, 71, 451–465. doi: 10.1086/342099

Takeuchi, R., Lambert, A. R., Mak, A. N. S., Jacoby, K., Dickson, R. J., Gloor, G. B., et al. (2011). Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U. S. A.*, 108, 13077–13082. doi: 10.1073/pnas.1107719108

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.*, 38, 3022–3027. doi: 10.1093/molbev/msab120

Tan, H., Yu, Y., Fu, Y., Liu, T., Wang, Y., Peng, W., et al. (2022). Comparative analyses of *Flammulina filiformis* mitochondrial genomes reveal high length

polymorphism in intergenic regions and multiple intron gain/loss in cox1. *Int. J. Biol. Macromol.*, 221, 1593–1605. doi: 10.1016/j.ijbiomac.2022.09.110

Thwaites, J.M., Farrell, R.L., Duncan, S.M., Reay, S.D., Blanchette, R.A., Hadar, E., Hadar, Y., Harrington, T.C., and McNew, D. (2005). Survey of potential sapstain fungi on Pinus radiata in New Zealand. *New Zealand Journal of Botany*, 43(3), 653-663

Tocchini-Valentini, G.D., Fruscoloni, P., and Tocchini-Valentini, G.P. (2011). Evolution of introns in the archaeal world. *Proc. Natl. Acad. Sci. USA.*, 108, 4782-4787

Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. *Science*.320(5872):77-82. doi: 10.1126/science.

Toor, N., and Zimmerly, S. (2002). Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA*, 8, 1373–1377. doi: 10.1017/s1355838202023087

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

Tourasse, N.J., Helgason, E., Økstad, O.A., Hegna, I.K., and Kolstø, A.B. (2006). The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J. Appl. Microbiol.*, 101, 579–593

Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., and Minh, B.Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.*, 44(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Tritt, A., Eisen, J. A., Facciotti, M. T., and Darling, A. E. (2012). An integrated pipeline for de novo assembly of microbial genomes. *PLoS One*, 7, e42304. doi: 10.1371/journal.pone.0042304

Trollip, C., Carnegie, A. J., Dinh, Q., Kaur, J., Smith, D., Mann, R., et al. (2021). Ophiostomatoid fungi associated with pine bark beetles and infested pines

in South-Eastern Australia, including *Graphilbum ipis-grandicollis* sp. nov. *IMA Fungus*, 12:24. doi: 10.1186/s43008-021-00076-w

Trollip, C., Carnegie, A. J., Rodoni, B., and Edwards, J. (2022). Draft genome sequences for three *Ophiostoma* species acquired during revisions of australian plant pathogen reference collections. *Microbiol. Resour. Announc.*, 11, e0017522. doi: 10.1128/mra.00175-22

Tsuji, J., Frith, M. C., Tomii, K., and Horton, P. (2012). Mammalian NUMT insertion is non-random. *Nucleic Acids Res.* 40, 9073–9088. doi: 10.1093/nar/gks424

Turk, E. M., Das, V., Seibert, R. D., and Andrulis, E. D. (2013). The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. *PLoS One*, 8, 78105. doi: 10.1371/journal.pone.0078105

Upadhyay, H.P. (1981). *A monograph of Ceratocystis and Ceratocystiopsis*. University of Georgia Press.

Uzunovic, A., and Byrne, T. (2013). "Wood market issues relating to blue-stain caused by Ophiostomatoid fungi in Canada" in *Ophiostomatoid Fungi: Expanding Frontiers*. eds. K. A. Seifert, Z. W. de Beer, and M. J. Wingfield (Utrecht: CBS-KNAW Fungal Biodiversity Centre), 201–212.

Uzunovic, A., Yang, D.Q., Gagné, P., Breuil, C., Bernier, L., Byrne, A., Gignac, M., Kim, S.H. (1999). Fungi that cause sapstain in Canadian softwoods. *Can. J. Microbiol.*, 45(11), 914-922

Valach, M., Farkas, Z., Fricova, D., Kovac, J., Brejova, B., Vinar, T., et al. (2011). Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Res.*, 39, 4202–4219. doi: 10.1093/nar/gkq1345

Vallès, Y., Halanych, K.M., and Boore, J.L. (2008). Group II introns break new boundaries: Presence in a bilaterian's genome. *PLoS ONE*, 3, e1488. doi: 10.1371/journal.pone.0001488

Van der Horst, G., and Inoue, T. (1993). Requirements of a Group I Intron

for Reactions at the 3′ Splice Site. *J. Mol. Bio.*, 229, 685–694. doi: 10.1006/jmbi.1993.1072

Van Diepeningen, A.D., Debets, A.J.M., Slakhorst, S.M., and Hoekstra, R.F. (2008). Mitochondrial pAL2-1 plasmid homologs are senescence factors in *Podospora anserina* independent of intrinsic senescence. *Biotechnol J.*, 3, 791–802

Vaughn, J.C., Mason, M.T., Sper-Witis, G.L., Kuhlman, P., and Palmer, J.D. (1995). Fungal Origin by Horizontal Transfer of a Plant Mitochondrial Group I Intron in the Chimeric *Coxl* Gene of *Peperomia*. *J. Mol. Evol.*, 41, 563–572. doi: 10.1007/BF00175814

Verma, S., Shakya, V.P.S., Idnurm, A. (2018). Exploring and exploiting the connection between mitochondria and the virulence of human pathogenic fungi. *Virulence*, 9, 426–446. 10.1080/21505594.2017.1414133

Vicens, Q., and Cech, T.R. (2006). Atomic level architecture of Group I introns revealed. *Trends Biochem. Sci.*, 31, 41–51. doi: 10.1016/j.tibs.2005.11.008

Videira, A., and Duarte, M. (2002). From NADH to ubiquinone in *Neurospora* mitochondria. *Biochim Biophys Acta.*, 1555(1-3), 187-191

Visinoni, F., and Delneri, D. (2022). Mitonuclear interplay in yeast: from speciation to phenotypic adaptation. *Curr. Opin. Genet. Dev.*, 76:101957. doi: 10.1016/j.gde.2022.101957

Vosseberg, J., and Snel, B. (2017). Domestication of self-splicing introns during eukaryogenesis: The rise of the complex spliceosomal machinery. *Biol. Direct.*, 12, 30. doi: 10.1186/s13062-017-0201-6

Wai, A., and Hausner, G. (2021). The mitochondrial genome of *Ophiostoma himal-ulmi* and comparison with other fungi causing Dutch elm disease. *Can. J. Microbiol.*, 67, 584–598. doi: 10.1139/cjm-2020-0589

Wai, A., and Hausner, G. (2022). The compact mitogenome of *Ceratocystiopsis pallidobrunnea*. *Can. J. Microbiol.* 68, 569–575. doi: 10.1139/cjm-2022-0038

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

Wang, L., Zhang, S., Li, J.H., and Zhang, Y.J. (2018). Mitochondrial genome, comparative analysis and evolutionary insights into the entomopathogenic fungus *Hirsutella thompsonii*. *Environ Microbiol.*, 20(9), 3393-3405. doi: 10.1111/1462-2920.14379. Epub 2018 Sep 24.

Wang, Z., Liu, Y., Wang, H., Meng, X., Liu, X., Decock, C., Zhang, X., and Lu, Q. (2020). Ophiostomatoid fungi associated with *Ips subelongatus*, including eight new species from northeastern China. *IMA Fungus*, 11, 3. doi: 10.1186/s43008-019-0025-3

Weiner, A.M. (1993). mRNA splicing and autocatalytic introns: Distant cousins or the products of chemical determination? *Cell, 72*, 161-164

Weiner, A.M., and Maizels, N. (1999). A deadly double life. *Science*, 284, 63-64.

Weiss-Brummer, B., Rödel, G., Schweyen, R.J., and Kaudewitz, F. (1982). Expression of the split gene cob in yeast: Evidence for a precursor of a "maturase" protein translated from intron 4 and preceding exons. *Cell*, 29, 527-536.

Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*, 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Williams, S.B., Baributsa, D., and Woloshuk, C. (2014). Assessing purdue improved crop storage (PICS) bags to mitigate fungal growth and aflatoxin contamination. *Journal of Stored Products Research*, 59, 190–196. 10.1016/j.jspr.2014.08.003

Williamson, D. (2002). The curious history of yeast mitochondrial DNA. *Nat. Rev. Genet.*, 3, 475–481. 10.1038/nrg814

Wingfield, B.D., Berger, D.K., Steenkamp, E.T., Lim, H.J., Duong, T.A., Bluhm, B.H., de Beer, Z.W., De Vos, L., Fourie., et al. (2017). IMA Genome-F 8: Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. *IMA Fungus*, 8(2), 385-396. doi: 10.5598/imafungus.2017.08.02.10

Wingfield, M.J., Barnes, I., De Beer, Z.W., Roux, J., Wingfield, B.D., and Taerum, S.J. (2017). Novel associations between ophiostomatoid fungi, insects and tree hosts: current status—future prospects. *Biol. Invasions*, 19, 3215–3228. doi: 10.1007/s10530-017-1468-3

Wingfield, M.J., Seifert, K.A., and Webber, J.F. (1993). *Ceratocystis* and *Ophiostoma*: taxonomy, ecology, and pathogenicity. American Phytopathological Society, St. Paul, Minnesota, USA

Wu, B., and Hao, W. (2014). Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. *G3 (Bethesda)*, 4(4), 605-612. https://doi.org/10.1534/g3.113.009910.

Wu, B., and Hao, W. (2014). *Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns.* G3 (Bethesda, Md.), 4, 605–612. doi: 10.1534/g3.113.009910

Wu, B., and Hao, W. (2019). Mitochondrial-encoded endonucleases drive recombination of protein-coding genes in yeast. *Environmental microbiology*, 21, 4233–4240. doi: 10.1111/1462-2920.14783

Wu, B., Hao, W., and Cox, M. P. (2022). Reconstruction of gene innovation associated with major evolutionary transitions in the kingdom Fungi. *BMC Biol.*, 20, 144. doi: 10.1186/s12915-022-01346-8

Xavier, B.B., Miao, V.P., Jónsson, Z.O., ans Andrésson, Ó.S. (2012). Mitochondrial genomes from the lichenized fungi Peltigera membranacea and Peltigera malacea: features and phylogeny. *Fungal Biol.*, 116, 802–814.

Xiao, S., Nguyen, D. T., Wu, B., and Hao, W. (2017). Genetic drift and indel

mutation in the evolution of yeast mitochondrial genome size. *Genome Biol. Evol.*, 9, 3088–3099. doi: 10.1093/gbe/evx232

Xiong, Y., and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, 9(10), 3353-62. doi: 10.1002/j.1460-2075.1990.tb07536.x

Xu, C., Wang, C., Sun, X., Zhang, R., Gleason, M.L., Eiji, T., and Sun, G. (2013). Multiple Group I introns in the small-subunit rDNA of *Botryosphaeria dothidea*: Implication for intraspecific genetic diversity. *PLoS ONE*, 8, e67808

Yan, Z., Li, Z., Yan, L., Yu, Y., Cheng, Y., Chen, J., Liu, Y., Gao, C., Zeng, L., Sun, X., et al. (2018). Deletion of the sex-determining gene *SXI1a* enhances the spread of mitochondrial introns in *Cryptococcus neoformans*. *Mobile DNA*, 9, 24. doi: 10.1186/s13100-018-0129-0

Yang, J., Zimmerly, S., Perlman, P.S., and Lambowitz, A.M. (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*, 381, 332–335. doi: 10.1038/381332a0

Ye, J., Cheng, J., Ren, Y., Liao, W., and Li, Q. (2020). The First Mitochondrial Genome for Geastrales (*Sphaerobolus stellatus*) Reveals Intron Dynamics and Large-Scale Gene Rearrangements of Basidiomycota. *Front. Microbiol.*, 11, 1970. doi: 10.3389/fmicb.2020.01970

Yin, M., Wingfield, M.J., Zhou ,X., et al. (2019). Taxonomy and phylogeny of the Leptographium olivaceum complex (Ophiostomatales, Ascomycota), including descriptions of six new species from China and Europe. *MycoKeys*, 60, 93–123

Yin, M.L., Wingfield, M.J., Zhou, X.D., et al. (2020). Phylogenetic re-evaluation of the Grosmania penicillata complex (Ascomycota, Ophiostomatales), with the description of five new species from China and USA. *Fungal Biology*, 124, 110–124

Yin, S., Heckman, J., and RajBhandary, U.L. (1981). Highly conserved GC-rich

palindromic DNA sequences flank tRNA genes in *Neurospora crassa* mitochondria. *Cell*, 26(3 Pt 1), 325-32. doi: 10.1016/0092-8674(81)90201-4

Yu, H., Li, M., Sandhu, J., Sun, G., Schnable, J.C., Walia, H., Xie, W., Yu, B., Mower, J.P., and Zhang, C. (2022). Pervasive misannotation of microexons that are evolutionarily conserved and crucial for gene function in plants. *Nat Commun.*, 13(1), 820. doi: 10.1038/s41467-022-28449-8.

Zaccaron, A. Z., De Souza, J. T., and Stergiopoulos, I. (2021). The mitochondrial genome of the grape powdery mildew pathogen *Erysiphe necator* is intron rich and exhibits a distinct gene organization. *Sci. Rep.*, 11, 13924. doi: 10.1038/s41598-021-93481-5

Zaffagni, M., and Kadener, S. (2019). Craving for introns. *Mol. Cell.*, 73, 1095–1096. doi: 10.1016/j.molcel.2019.03.008

Zhang, S., and Zhang, Y. J. (2019). Proposal of a new nomenclature for introns in protein-coding genes in fungal mitogenomes. *IMA Fungus*, 10, 15. doi: 10.1186/s43008-019-0015-5

Zhang, S., Wang, X.N., Zhang, X.L., Liu, X.Z., and Zhang, Y.J. (2017). Complete mitochondrial genome of the endophytic fungus *Pestalotiopsis fici*: features and evolution. *Appl Microbiol Biotechnol.*, 101(4), 1593-1604. doi: 10.1007/s00253-017-8112-0

Zhang, S., Zhang, Y. J., and Li, Z. L. (2019). Complete mitogenome of the entomopathogenic fungus *Sporothrix insectorum* RCEF 264 and comparative mitogenomics in Ophiostomatales. *Appl. Microbiol. Biotechnol.*, 103, 5797–5809. doi: 10.1007/s00253-019-09855-3

Zhang, X., Dong, R., Zhang, Y., Zhang, J.L., Luo, Z., Zhang, J., Chen, L.L., and Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.*, 26, 1277–1287. doi: 10.1101/gr.202895.115

Zhang, Y., Bell, A., Perlman, P. S., and Leibowitz, M. J. (2000). Pentamidine

inhibits mitochondrial intron splicing and translation in *Saccharomyces cerevisiae*. *RNA*, 6, 937–951. doi: 10.1017/s1355838200991726

Zhang, Y.J., Zhang, H.Y., Liu, X.Z., and Zhang, S. (2017). Mitochondrial genome of the nematode endoparasitic fungus *Hirsutella vermicola* reveals a high level of synteny in the family Ophiocordycipitaceae. *Appl Microbiol Biotechnol.*, 101(8), 3295-3304. doi: 10.1007/s00253-017-8257-x.

Zhao, C., and Pyle, A.M. (2017). The group II intron maturase: A reverse transcriptase and splicing factor go hand in hand. *Curr. Opin. Struct. Biol.*, 47, 30–39

Zhao, Z., Liu, H., Wang, C., and Xu, J.R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*, 14, 274

Zhou, X., De Beer, Z.W., Cibrian, D, et al. (2004a). Characterisation of *Ophiostoma* species associated with pine bark beetles from Mexico, including *O. pulvinisporum* sp. *nov. Mycological Research*, 108, 690–698

Zhou, X., De Beer, Z.W., Harrington, T.C., et al. (2004b). Epitypification of *Ophiostoma galeiforme* and phylogeny of species in the *O. galeiforme* complex. *Mycologia*, 96, 1306–1315

Zhuang, F., Mastroianni, M., White, T.B., and Lambowitz, A.M. (2009). Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. *Proc. Natl. Acad. Sci. USA.*, 106, 18189–18194. doi: 10.1073/pnas.0910277106

Zimmerly, S., and Semper, C. (2015). Evolution of group II introns. *Mob. DNA*, 6, 7. doi: 10.1186/s13100-015-0037-5

Zimmerly, S., Guo, H., Perlman, P.S., and Lambowitz, A.M. (1995). Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, 82, 545–554. doi: 10.1016/0092-8674(95)90027-6

Zimmerly, S., Hausner, G., and Wu, X. (2001). Phylogenetic relationships

among group II intron ORFs. *Nucleic Acids Res.*, 29, 1238–1250. doi: 10.1093/nar/29.5.1238

Zimmerly, S., Moran, J.V., Perlman, P.S., and Lambowitz, A.M. (1999). Group II intron reverse transcriptase in yeast mitochondria: Stabilization and regulation of reverse transcriptase activity by the intron RNA. *Journal of Molecular Biology*, 289(3), 473-490.

Zipfel, R. D., de Beer, Z. W., Jacobs, K., Wingfield, B. D., and Wingfield, M. J. (2006). Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. *Stud. Mycol.*, 55, 75–97. doi: 10.3114/sim.55.1.75

Zipfel, R.D., De Beer, Z.W., Jacobs, K., et al. (2006). Multi-gene phylogenies define Ceratocystiopsis and Grosmannia distinct from Ophiostoma. *Studies in Mycology*, 55, 75–97

Zivanovic, Y., Wincker, P., Vacherie, B., Bolotin-Fukuhara, M., and Fukuhara, H. (2005). Complete nucleotide sequence of the mitochondrial DNA from *Kluyveromyces lactis*. *FEMS Yeast Res.*, 5(4-5), 315-322

Zlotorynski, E. (2019). Intron definition, exon definition and back-splicing revisited. *Nat. Rev. Mol. Cell Biol.*, 20, 661. doi: 10.1038/s41580-019-0178-3

Zubaer, A., Wai A., Patel, N., Perillo, J., and Hausner, G. (2021). The Mitogenomes of *Ophiostoma minus* and *Ophiostoma piliferum* and Comparisons With Other Members of the Ophiostomatales. *Front. Microbiol.*, 12, 618649. doi: 10.3389/fmicb.2021.618649

Zubaer, A., Wai, A., and Hausner, G. (2018). The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. *Sci. Rep.*, 8,17591. doi: 10.1038/s41598-018-35926-y

Zubaer, A., Wai, A., and Hausner, G. (2019). The fungal mitochondrial *Nad5* pan-genic intron landscape. *Mitochondrial DNA Part A*, 30, 835–842. doi: 10.1080/24701394.2019.1687691

Zubaer, A., Wai, A., Patel, N., Perillo, J., and Hausner, G. (2021). The mi-

togenomes of *Ophiostoma minus* and *Ophiostoma piliferum* and comparisons with other members of the Ophiostomatales. *Front. Microbiol.*, 12:618649. doi: 10.3389/fmicb.2021.618649

Zumkeller, S., Gerke, P., and Knoop, V. (2020). A functional twintron, 'zombie' twintrons and a hypermobile group II intron invading itself in plant mitochondria. *Nucleic Acids Res.*, 48, 2661–2675. doi: 10.1093/nar/gkz1194