Genome mining of *Leptographium wingfieldii*, an invasive species in Canadian forests, and related taxa in the order Ophiostomatales for the characterization of secondary metabolites and ribozymes

by

Abdullah Zubaer

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

DOCTOR OF PHILOSOPHY

Department of Microbiology

University of Manitoba

Winnipeg, Manitoba, Canada

ABSTRACT

Leptographium wingfieldii is an invasive fungal species in Canadian forests which was originally isolated in Europe (France and Greece). *Leptographium wingfieldii* and other fungi in the order Ophiostomatales (Ascomycota) are vectored by arthropods, and they can be either pathogenic to tree species or cause blue stain on sapwood in conifer (and hardwood) species. These fungi are ecologically and economically significant due to their impact on forest ecosystems and lumber industry. Whole genome sequences were obtained from *Leptographium wingfieldii* and related fungi (including *Leptographium procerum*, *Leptographium terebrantis*, *Grosmannia aureum*, *Ophiostoma minus*, and *Ophiostoma piliferum*). The mitochondrial genomes of these fungi were assembled and found to contain autocatalytic group I and group II introns, intron-encoded homing endonucleases along with intron-encoded reverse transcriptase enzymes that have applications in genome editing. These elements contribute toward the genetic diversity observed among the mitochondrial genomes studied. The study provided information to generate a mitochondrial intron landscape, identified complex intron arrangements, and demonstrated the correlation of mitogenome expansion with the number of introns. The whole genome sequence data were also analyzed with regards to the presence of nuclear genome encoded biosynthetic gene clusters (BGCs). This effort identified the presence of 205 BGCs categorized into PKS I, PKS III, NRPS, RiPPS, Terpenes, and hybrid types and these could be sources for potential antimicrobials and industrially important chemical compounds. The study provides a platform for downstream biochemical characterization and heterologous expression of the identified genetic elements, facilitating their functional annotation and explore their potential for industrial applications.

RÉSUMÉ

Leptographium wingfieldii est une espèce fongique envahissante présente dans les forêts canadiennes et initialement isolée en Europe (France et Grèce). *Leptographium wingfieldii* et d'autres champignons de l'ordre des Ophiostomatales (Ascomycota) sont transmis par des arthropodes et peuvent être pathogènes pour les espèces d'arbres ou provoquer le bleuissement de l'aubier chez les conifères (et les feuillus). Ces champignons sont d'importance écologique et économique en raison de leur impact sur les écosystèmes forestiers et l'industrie du bois d'œuvre. Des séquences génomiques complètes ont été obtenues à partir de *Leptographium wingfieldii* et de champignons apparentés (notamment *Leptographium procerum*, *Leptographium terebrantis*, *Grosmannia aureum*, *Ophiostoma minus* et *Ophiostoma piliferum*). Les génomes mitochondriaux de ces champignons ont été assemblés et contiennent des introns autocatalytiques des groupes I et II, des endonucléases de localisation codées par des introns ainsi que des enzymes de transcriptase inverse codées par des introns, qui ont des applications en édition génomique. Ces éléments contribuent à la diversité génétique observée parmi les génomes mitochondriaux étudiés. L'étude a fourni des informations permettant de générer un paysage intronique mitochondrial, identifié des arrangements complexes d'introns et démontré la corrélation entre l'expansion du mitogénome et le nombre d'introns. Les données de séquence du génome entier ont également été analysées concernant la présence de groupes de gènes biosynthétiques (GGB) codés par le génome nucléaire. Cet effort a identifié la présence de 205 GGB classés en PKS I, PKS III, NRPS, RiPPS, terpènes et types hybrides. Ceux-ci pourraient être des sources potentielles d'antimicrobiens et de composés chimiques d'importance industrielle. L'étude fournit une plateforme pour la caractérisation biochimique en aval et l'expression hétérologue des éléments génétiques identifiés, facilitant leur annotation fonctionnelle et explorant leur potentiel d'applications industrielles.

ACKNOWLEDGEMENTS

First of all, I want to express my deepest gratitude to my supervisor Dr. Georg Hausner. His guidance and training made me successfully complete this work. I am indebted to him for his compassion and care. He provided me with enough flexibility and freedom to think independently and grow as a scientist. I always found him supportive while facing difficulties in research and life. Words are not enough to thank him properly.

Second, I would like to thank my co-supervisor Dr. Olivier Tremblay-Savard. His teaching and guidance on the mathematical basis of comparative genomics, HMM, co-variance models were very helpful to navigate the genomics studies. I felt confident by his teaching and guidance to go deep into bioinformatics algorithms, and mathematical and statistical concepts used in genomics. I found him as a great supervisor and a great person I can always rely on.

Third, I am thankful to my committee members Dr. Deborah Court and Dr. Mike Domaratzki for their support and guidance. They gave me valuable suggestions in my research and provided their insight and wisdom to guide me to the right direction.

Finally, I want to thank the members of Hausner lab Aamer, Amal, Jigeesha, and Sawsan to be always friendly and helpful to me. I want to especially thank Alvan and Talal for their extensive support and compassion towards me. I found the faculties, students, and the office of the Department of Microbiology were helpful and supportive.

TABLE OF CONTENTS

PERMISSIONS/ AUTHOR CONTRIBUTIONS

Figures, tables and passages in chapter 2 and 3 of this thesis have been adapted from published works in academic journals that I have authored.

Chapter 2:

Author Contributions: Abdullah Zubaer (AZ), Alvan Wai (AW), Nikita Patel (NP), and Jordan Perillo (JP) have been working under the supervision of Georg Hausner (GH) and obtained data and contributed toward the analysis. AZ and AW took the lead with regards to assembling the datasets and the final analysis of the data.

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 in the article/ Supplementary Material. The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.618649/full#supplementary-material

Chapter 3:

Author Contributions:

Conceptualization – AZ, AW, GH. Writing – original draft: AZ. Writing – review & editing: AZ, AW, GH.

Data Availability Statement: The authors confirm that all research data has been deposited in NCBI GenBank (see Table 1, cjm-2024-0179suppla, and cjm-2024-0179supplb).

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Elaboration
BGC	Biosynthetic Gene Cluster
CM	Covariance Model
EDTA	Ethylenediaminetetraacetic acid
HE	Homing Endonuclease
HEG	Homing Endonuclease Gene
HMM	Hidden Markov Model
IEP	Intron-encoded Protein
MEA	Malt Extract Agar
NRPS	Non-ribosomal Protein Synthase
PCR	Polymerase chain reaction
PKS	Polyketide Synthase
PYG	Peptone Yeast Glucose
RiPP	Ribosomally synthesized and post-translationally modified peptides
RT	Reverse Transcriptase
SM	Secondary Metabolites
YPD	Yeast Peptone Dextrose

GENERAL INTRODUCTION

Ophiostomatales is an order of fungi assigned to the Division Ascomycota. The Order Ophiostomatales has undergone many taxonomic revisions, currently it includes 14 accepted genera with a proposal of possible up to 24 genera (de Beer et al. 2022). *Leptographium* and *Ophiostoma* are two well studied genera within the Ophiostomatales. Some members of *Ophiostoma* and *Leptographium* are causal agents of plant diseases such as Dutch elm disease (*Ophiostoma ulmi*), or conifer root diseases (*Leptographium terebrantis*) etc. In addition, many members of these genera are blue stain causing fungi. These fungi frequently grow in bark beetle galleries under the bark of infected trees and the fungal growth on the sapwood creates black and blue staining on the sapwood due to melanin produced by the fungus. Although this stain has no deleterious effects it is undesirable when sold as lumber. Blue stain therefore has a significant economic impact on the forestry/lumber related industries. Either pathogenic or blue stain causing, most members of *Leptographium* and *Ophiostoma* depend on bark beetles to propagate in the forest environment. Bark beetles live under the bark of trees and carry fungal spores with them, thus inoculate other trees as they migrate to new trees to establish their broods inside bark beetle galleries. The fungus can provide nutrients for the beetles, and secondary metabolites released by the fungus can potentially neutralize the plant's immune response (Six 2012). This creates a suitable environment for the beetles to propagate.

Fungal whole genome sequences offer an opportunity to explore mitochondrial genomes and features of the nuclear genome. Mitochondrial genomes are a source of mobile introns and intron-encoded homing endonuclease and reverse transcriptase enzymes. Those enzymes can be

utilized as genome editing tools. The mobile introns (group I and group II introns) are potentially autocatalytic (i.e., self-splicing) and can be described as ribozymes (RNA molecules) that have catalytic activities. Mitochondrial genomes can also provide sequences for fungal identification and for studying fungal evolution and they can be applied to study generic and ordinal relationships among fungi. The nuclear genomes might offer some insights into the categories of secondary metabolites (SMs) these fungi can produce. The genes associated with SM production are usually arranged in biosynthetic gene clusters (BGCs) and there are bioinformatics tools available to predict the presence and position of these genes. Members of the Ophiostomatales appear to produce SMs in part to attract insect vectors and in some instances SMs might facilitate their insect vectors to have a better chance to invade trees (Six and Wingfield 2011). In addition, there is considerable interest in fungal SMs as they may offer a range of natural products that are of pharmaceutical value (Bérdy 2005; Keller 2019).

Leptographium wingfieldi and related taxa are the subjects for part of this study. *Leptographium wingfieldii* was originally isolated in Europe (France and Greece) and it is assumed the species was introduced into Canada (reviewed in Hausner et al. 2005). Exotic fungi introduced to a different forest environment can pose ecological risks and challenges to the native flora. Species of *Leptographium* can be challenging to be identified based on morphological features, therefore it is necessary to have reliable DNA sequence-based methods for their identification.

During this study the mitochondrial genomes of *Ophiostoma minus* and *Ophiostoma piliferum* were explored and annotated as these species are considered serious blue-strain fungi

and are found in forest ecosystems in the Northern Hemisphere. These genomes provided an opportunity to conducts a comparative analysis with other recently published mitochondrial genomes for members of the Ophiostomatales. From this analysis the intron dynamics were examined, and a mitochondrial intron landscape was established for this Order of fungi. This work also identified novel intron configurations, nested introns, and side-by-side intron, in *Ophiostoma ips*. Overall, this work on mitochondrial genomes of the Ophiostomatales, showed the conservation of gene synteny, identified intron homing/insertion sites, and confirmed the correlation of intron number with mitogenome size, as seen in other fungi.

Combined, whole genome sequencing and genomic mining allowed for the characterization of mitochondrial genomes and nuclear biosynthetic gene clusters. This provides a resource for those searching for homing endonuclease, ribozymes, nested intron arrangements, potential sequence markers for taxonomic studies and the genetics of secondary metabolites production for the Ophiostomatales.

CHAPTER 1: LITERATURE REVIEW

CHAPTER 1: LITERATURE REVIEW

1.1 *LEPTOGRAPHIUM WINGFIELDII* AND THE OPHIOSTOMATALES FUNGI - BIOLOGY, TAXONOMY, AND ECONOMIC IMPORTANCE

Leptographium wingfieldii is a fungal species belonging to the order Ophiostomatales in the class Sordariomycetes which is in the phylum Ascomycota. The fungal phylum Ascomycota contains the largest number of fungi with diverse life histories and morphological features ranging from unicellular yeast (Saccharomyces), filamentous members (molds such as *Penicillium* species etc.) and multicellular "mushroom" like truffles (*Tuber*), morels (*Morchella* sp.), cup fungus (*Peziza*) etc. This phylum falls under the subkingdom Dikarya along with Basidiomycota (contains the mushrooms). Most Ascomycota fungi are filamentous and the characteristic feature of them is the formation of ascospores. Ascospores are sexual (meiotic) spores generated for fungal reproduction and these spores are formed within an ascus (a sac-like structure). The name 'Ascomycota' is derived from the term ascus and this group is also referred to as the "sac fungi" (Lutzoni et al. 2004; Hibbett et al. 2007).

Ophiostomatales includes more than 300 species (de Beer et al. 2022), most of them are non-pathogenic, however, some are plant pathogen (e.g., Dutch elm disease) and a few are human pathogens (e.g., Sporotrichosis). The taxonomy of Ophiostomatales has undergone many revisions due to a lack of definitive characters and many instances of convergent evolution of morphological features shared among unrelated fungal groups (Hausner et al. 1993; Spatafora and Blackwell 1994; Zipfel et al. 2006). With the availability of more molecular markers [such as ITS regions, actin (ACT), beta-tubulin (TUB2), calmodulin (CAL) and the translation

elongation factor 1-alpha (TEF1-α) (Linnakoski et al. 2016; Yin et al. 2016; de Beer et al. 2016; Jankowiak et al. 2017; Strzalka et al. 2020)], the order and its various genera have gone through more revisions in recent years (de Beer and Wingfield 2013; de Beer et al. 2016; Jankowiak et al. 2017). According to Hyde et al. (2020), this order of fungi contains two families Ophiostomataceae (11 genera) and Kathistaceae (3 genera), comprising a total of 14 well-defined genera. However, more recently Ophiostomatales have been re-examined, specifically the Ophiostomataceae. With a combination of genomic data and various nuclear markers, the finding suggested that there is the possibility of up to 24 lineages in this order (de Beer et al. 2022). The genus *Leptographium* is one of the genera in the Ophiostomatales and it is well known for the association of its members with various tree diseases such as plant root disease and frequently these fungi can cause blue stain on timbers. These fungi are saprophytic and thrive to live on wood (sapwood) under the bark of the trees. *Leptographium* species are filamentous and grow apically. The filamentous structure is called hyphae, and the complex network of hyphae is called the mycelium. *Leptographium* species have septate hyphae which means their hyphae has membranes (septa with pores) between cells. They can reproduce sexually by the formation of ascospores (formed inside asci that are housed within perithecia-like structures) however, asexual reproduction is more frequently encountered among these fungi. Indeed, for most members of *Leptographium* one typically only observes the asexual structures (conidiophore generating conidia). Historically the genus *Leprographium* was used to accommodate asexual members (see Upadhyay 1982) but recently the taxonomy of the fungi has been revised to a naming system (Wingfield et al. 2012) that avoids two names for fungi based on “asexual” (anamorph) and “sexual” (teleomorph) reproductive features (reviewed in de Beer et al. 2022).

Leptographium species maintain an association with bark beetles. The fungi and the beetles share a common habitat and mutually benefit each other. The bark beetles and their brood feed on the fungi and the adult beetles carry fungal spores on their body, thus spreading these fungi from tree to tree in the forest ecosystem. Usually, fungi produce certain chemicals (such as nitrogen, sterols, and other nutrients) that feed the beetles and help them to survive under the bark by possibly suppressing the plant immunity system (Bleiker and Six 2007; Lieutier et al. 2009; Six and Wingfield 2011), however, research showed exceptions of those phenomena in mountain pine beetles (Fortier et al. 2024).

There are at least 68 species in the genus *Leptographium* (Harrington and Cobb 1988; Jacobs and Wingfield 2001), but only a few species have been studied so far with regards to genomics. Species of *Leptographium* that have been studied at the genomic level so far are as follows: *L. procerum* (Harrington and Cobb 1983; Jacobs and Wingfield 2001), *L. lundbergii* (Jacobs et al. 2005), *L. wageneri* (Wagener and Mielke 1961), and *L. terebrantis* (Lee et al. 2006; Klepzig and Wilkens 1997). In general, there were frequent reports on these fungi, such as studies on *L. abietinum* (McBeath et al. 2004), *L. sinoprocerum* (Lu et al. 2008), *L. taigense* (Linnakoski et al. 2012a), *L. olivaceapini* (Davidson 1971; Yin et al. 2019), *L. panxianense* (Pan et al. 2020), *L. longiclavatum* (Lee et al. 2005), *L. conjunctum* (Paciura et al. 2010) etc.

Leptographium wingfieldii is the species of interest of the current study. The classification of *Leptographium wingfieldii* is as follows:
Kingdom: Fungi, Phylum: Ascomycota, Class: Sordariomycetes, Order: Ophiostomatales, Family: Ophiostomataceae, Genus: *Leptographium*, Species: *wingfieldii*.

Leptographium wingfieldii was originally described from Europe (France and Greece, Morelet 1988) but it appears this fungus has been recently introduced into North America along with its bark beetle vector (Jacobs et al. 2004). The strains used in the current thesis were collected from Northwest Ontario, Canada (described in Hausner et al. 2005). *L. wingfieldii* is a potential pathogen on white pine and it is a blue stain causing fungus on various conifer species. Blue stain is undesirable as it reduces the value of the lumber. In addition, as an invasive species, it might have ecological impact by competing with other fungi and colonizing tree species that may have little or no tolerance to this exotic species (Jacobs et al. 2004; Hausner et al. 2005). This fungus was originally found in association with the bark beetle *Tomicus piniperda*. However, due to the change in environment this fungus might create associations with native beetle species (observed in Jacobs et al. 2004). Its long-term impact on the forest ecosystem is still unknown and distinguishing this species from native species based on morphological characters can be challenging. Thus, the study of the genomes of this fungus and related species might offer opportunities for finding suitable genetic markers that can eventually be applied in high throughput DNA markers assisted identification strategies (Bergeron et al. 2019; Capron et al. 2020; Trollip et al. 2021).

1.2 Fungal genomics, mitochondrial genome and mobile introns

Fungal genomes consist of nuclear DNA and mitochondrial DNA. Although the chromosome numbers, genome size, and number of genes in the genome are variable among species, previous report showed the genomes range from 16.05 Mb to 43.82 Mb for the

Ophiostomatales fungi (de Beer et al. 2022). *Neurospora crassa*, a model filamentous fungus in Ascomycota, has a 43 Mb sized genome organized in seven chromosomes and encodes for about 10,000 genes (Galagan et al. 2003). *Ophiostoma ulmi* and *O. novo-ulmi* – are members of the Ophiostomatales that are estimated to have six chromosomes containing a 30-35 Mb genome harboring 8,000 - 10,000 genes (Hintz and Peberdy 1989). *Grosmannia clavigera* - a closely related fungus to *L. wingfieldii* - has a genome of 30 Mb with seven chromosomes which contains about 9,000 genes (DiGuistini et al. 2009).

Mitochondrial genomes of fungi usually get sequenced as a part of whole genome sequencing (WGS) efforts, where it is denoted as mitochondrial chromosome or chrM. However, mitochondrial genomes are also independently sequenced when studies focus on fungal mitochondria and its components, so the challenge in some instances is to recover pure mtDNA from many fungi. Mitochondrial genomes in fungi usually code for 15 protein-coding genes, two rRNA genes, and a complete set of tRNAs (covering all amino acids). The protein coding genes are ATP synthase subunit genes (*atp6, atp8, atp9*), NADH dehydrogenase subunit genes (*nad1, nad2, nad3, nad4, nad4L, nad5 and nad6*), cytochrome oxidase subunit genes (*cox1, cox2, cox3*, *cob*), and ribosomal protein coding gene (*rps3*). Two rRNA genes in mitochondrial genome are small ribosomal subunit gene (*rns*) and large ribosomal subunit gene (*rnl*), and there are about 20 - 26 tRNA genes in a mitochondrial genome (reviewed in Hausner 2003; Lang 2018). However, there are exceptions, and in some fungi, there is an absence of *atp* and *nad* genes in the mitogenomes, possibly due to gene loss by oxidative damage or transfer of organellar genes to the nuclear genome (Buschges et al. 1994; Schikora-Tamarit et al. 2021).

Although having similar number of genes, the sizes of the mitochondrial genomes (or mitogenome) among fungi are quite variable ranging from 11 kb (plus inverted repeats; in *Hanseniaspora uvarum*, Pramateftaki et al. 2006) to 272 kb (in *Morchella importuna*, Liu et al. 2020). Size variation of the mitogenomes in the Ophiostomatales ranges from 23.7 kb to 150.9 kb and the variability is in part due to introns that are selfish elements frequently encountered in the fungal mitogenomes (Zhang et al. 2019; Fonseca et al. 2021; Wai and Hausner 2021; Zubaer et al. 2018; Zubaer et al. 2021; Mukhopadhyay et al. 2023). These elements are frequently referred to as mobile introns as evidence has accumulated that shows these introns can move from intron-containing alleles to cognate intron-less alleles (reviewed in Dujon 1989). There are two types of mobile introns found in fungal mitogenomes, group I introns and group II introns (Fonseca et al. 2021). They can be distinguished based on their splicing mechanisms, and they are potential ribozymes as they can self-splice from the primary transcript (reviewed in Hausner et al. 2014; Mukhopadhyay and Hausner 2021; McNeil et al. 2016). Ribozymes (catalytic RNA) were first discovered independently by Sidney Altman and Thomas R. Cech, and they were awarded the Nobel prize in 1989 for their efforts (Stark et al. 1978; Guerrier-Takada et al. 1983; Cech et al. 1981; Zaug et al. 1983). Cech's group reported the presence of a self-splicing group I intron in the nuclear ribosomal large subunit gene from the protozoan *Tetrahymena* (Cech et al. 1981; Zaug et al. 1983). Later research showed that group I introns are widespread in viruses (phages) and almost all life forms - archaea, bacteria, fungi, plants, and animals (Haugen et al. 2005; Lang et al. 2007; Simon et al. 2008; Huchon et al. 2015; Nawrocki et al. 2018; Jenkins et al. 2022; Kobayashi et al. 2022).

Group I introns can reside in the protein coding genes in fungal mitochondria and introns are highly variable in their primary sequences. However, there is a pattern in the group I intron in which the terminal base (3' end) of this intron sequence is usually a 'G' (referred to as the omega G), and the last base of the upstream exon is usually a 'T'. Despite their differences at the sequence level, group I introns maintain a similar secondary and tertiary structure (Michel et al. 1982; Michel and Westhof 1990). Typically, there are ten paired regions/helical regions in the secondary structure in group I introns, designated as P1 to P10 (P = paired). The paired (P1 to P10) regions of group I are facilitated by complementary base pairings, and this is the key for establishing a three-dimensional configuration that promotes self-splicing activity. Due to variations in the intron core sequences and peripheral structures such as additional helical regions or the absence of some helical components, group I introns are categorized into five groups from IA to IE and further subdivided with designations such as IA1, IC2 etc. (Michel and Westhof 1990; Zhou et al. 2008).

The splicing reaction of a group I intron consists of two transesterification reactions. The splicing starts with an external GTP that initiates a nucleophilic reaction. This external guanosine (G, referred to as the alpha-G) associates with the GTP-binding pocket in the P7 region and the 3'OH group of the external G will promote the transesterification reaction that releases the upstream exon. The second transesterification reaction involves the 3'OH from the released upstream exon. RNA folding of the intron enables all active components to be in close proximity. The upstream exon 3' OH group will attack the downstream intron/exon junction destabilizing the phosphodiester bond at this junction and thus releasing the intron and joining the two exons. As stated previously these reactions are mediated by the intron RNA folding into

a conformation that generates a catalytic RNA molecule. Splicing also requires the presence of divalent cations (e.g., Mg+2) and in some instances protein factors (maturases) also enhance the activity of the reaction (Stahley and Strobel 2006; Vicens et al. 2008). The intron/exon junctions are moved into close proximity by the generation of the P1 fold (that includes the internal guide sequence) and the P10 interaction that pairs a short segment of the downstream exon with a sequence in the unpaired regions within the P1 helix. The end results of these interactions ensure the proper splicing of the intron and accurate joining of the exons (Vicens and Cech 2006). Usually, group I introns encode for open reading frames (ORFs). In fungal mitochondrial genomes these ORFs can encode homing nuclease genes (HEGs). The products of these genes are referred to as homing endonucleases (HEs) and they can catalyze the mobility of these introns (Dujon 1989). It was shown early on that during mating HEGs (and their host introns) display super-mendelian inheritance as they are preferentially passed on due to their mobility that makes them invasive type elements, although in general they are sequence specific and invade cognate loci (that lack introns) (Dujon et al. 1986). There are numerous examples in the literature suggesting that these elements can move horizontally, therefore group I introns (and their associated HEGs) can maintain a life cycle independent of the organisms that encode them (Koufopanou et al. 2002; Megarioti and Kouvelis 2020).

Homing endonuclease enzymes usually cleave DNA molecules in a manner that generates staggered ends. These endonucleases recognize and bind to a very specific string of 14 - 40 bp (depending on the type) long DNA sequence and HEs can either generate a double stranded cut or a single stranded nick in the target DNA that in turn will activate the double strand DNA-repair mechanism involving homologous recombination. This mechanism uses the

cognate intron-plus allele as a template to repair the cut or nicked intron-minus allele (Stoddard 2005; Stoddard 2014). This mechanism is referred to as intron-homing, i.e., the intron is going home making HEGs and their associated introns cite specific mobile units. There are two types of HEGs found in fungal mitogenomes - LAGLIDADG and GIY-YIG type HEGs, categorized based on short conserved amino-acid sequence motifs (Stoddard 2005; Hafez and Hausner 2012). Due to the HEGs specificity for a particular sequence, introns and their HEG partners can usually be found in specific locations across many fungal species, and it is possible to generate intron-maps using a reference species/sequence (Ferandon et al. 2010; Hafez et al. 2013; Guha et al. 2017; Zubaer et al. 2019).

Group II introns are less frequently observed in fungal mitochondrial genomes in comparison to group I introns. Group II introns are found in all domains of life (in bacteria, archaea and organellar genomes of protists, plants, fungi and less frequently in metazoans). So far, they have not been recorded from bacteriophages, but group II introns have been characterized from conjugating transposons and the reverse transcriptases of group II introns have been co-opted in some instance as components of prokaryotic defense systems against invading genetic elements (phages) (Ferat and Michel 1993; Lambowitz and Zimmerly 2011; Zimmerly and Semper 2015; McNeil et al. 2016; González-Delgado et al. 2021). In contrast to group I introns, a group II intron has six domains in its structure designated as domain I (DI) to domain VI (DVI). The intron-encoded ORF can reside on the second (DII), third (DIII), or more frequently in the fourth domain (DIV) (Michel and Westhof 1990; Toor et al. 2001; Simon et al. 2008; Hafez and Hausner 2011; Toor and Zimmerly 2002). The first and largest domain DI of group II introns is referred to as the scaffold domain. It is critical in facilitating tertiary

interactions that promote the proper folding of the intron including promoting interactions between intron and exon sequences to define the intron/exon splice junctions. Typically, DI includes exon binding sequences (EBS1 and EBS2) that can interact via H-bonding with the intron binding sequences (IBS1 and IBS2) located in the upstream exon, these interactions assemble the group II intron into a splicing competent structure. The splicing involves a two-step transesterification reaction mediated by RNA folding and the presence of Mg+2 cations: here an internal bulged adenosine's 3'OH group from domain DVI reacts with the upstream intron-exon junction; then the liberated upstream exon's 3'OH reacts with the downstream intron-exon junction releasing the intron and joining the exons together. (Michel and Ferat 1995; Daniels et al. 1996; Toor et al. 2008; Toor et al. 2010). The branching pathway mediated by the bulged internal adenine (referred to as the branch point) as described above results in the release of a lariat shaped RNA intron molecule, this contrasts to the alternative reaction where the intron is released as a linear molecule due to a hydrolysis reaction (McNeil et al. 2016; Mukhopadhyay and Hausner 2021).

Group II introns sometimes harbor an ORF encoding a reverse transcriptase-like enzyme that can mediate retro-homing - a form of intron homing that involves an RNA intermediate. Reverse transcriptase enzymes as encoded in group II introns typically have four functional domains which are as follows: 1. Maturase (X) domain (promotes proper intron RNA folding); 2. DNA binding (D) domain; 3. Endonuclease (E) domain; 4. Reverse transcriptase (RT) domain. After transcription of a gene containing a group II intron the intron encoded ORF (RT) can be translated, and the released intron RNA lariat can combine with its reverse transcriptase counterpart forming a ribonucleoprotein particle (RNP). The RNA component of the RNP with

its 3'OH nicks the sense strand of the target DNA, and the Endonuclease (E) domain of the reverse transcriptase enzyme makes a nick in the antisense strand of the DNA at the intron insertion target site. The intron target site in part is guided by the EBS sequences located in DI of the intron RNA. The intron RNA is reverse spliced into that gap present in the sense strand of the DNA sequence. The reverse transcriptase (RT) domain uses the 3'OH end of the DNA antisense strand as a primer and generates a cDNA by using the intron RNA as a template, thus forming a DNA-RNA hybrid at the intron homing site. The RNA part of the DNA-RNA hybrid eventually gets replaced with DNA by the host cell's repair mechanism (Lambowitz and Zimmerly 2004; Edgell et al. 2011). This process is referred to as retrohoming, as it is site specific but involves an RNA intermediate that is eventually reversed transcribed by the intron encoded RT.

Due to their self-splicing ability and intron homing/mobility mechanisms, introns and their intron encoded proteins (IEPs) can propagate themselves in the genomes and maintain a life cycle independent of the host genome; it also appears to promote their ability for lateral transfer across species boundaries. It should be noted that splicing of group I and II introns in many instances does require intron and/or nuclear genome encoded factors, usually to ensure appropriate intron folding and/or splicing efficiency (reviewed in Mukhopadhyay and Hausner 2024). It has been observed that when comparing fungal mitochondrial genomes, the intron content is highly variable, and introns appear to be rapidly lost and gained. Goddard and Burt (1999) when comparing mitochondrial genomes of members of the Saccharomycetales observed that HEGs appear to be gained and lost along with their intron hosts in addition they noted many examples of "degenerated" HEGs due to the presence of premature stop codons. From this work they postulated an HEG/Intron life cycle. They argued that group I introns and their HEGs

follow a scenario that best fits neutral evolution with regards to intron frequency (gain and loss) and the intron mobility. First, a group I intron encoding a HEG invades an empty homing site, however due to lack of selection the HEG is prone to accumulate mutations. This leads to the erosion of the HEG and the loss of mobility for the host intron. Eventually the intron will get lost, and this would generate a "new" target site that can be invaded again by a group I intron. This establishes a cycle of intron invasion, loss of HEG, loss of intron, and eventually a reinvasion by the intron to occupy the site again. This model suggests that at the population level some active versions of the group I intron (with active HEG) can be maintained by chance (drift), but many members would be at various stages of degradation. For an intron/HEG to persist it must keep invading "new" or empty sites or it will become extinct. Alternatively, an intron/HEG could escape this cycle if it benefits the host and is subject to selection.

The Goddard and Burt model has been widely accepted but the persistence of some introns (biased gain and losses) and the rapid streamlining (rapid loss of introns/HEGs) of mitochondrial genomes observed in some fungal lineages suggest that the Goddard and Burt model may not apply to all mobile introns (Pogoda et al. 2019). Zubaer et al. (2018) argued that for some group I introns "drift" may not be the best explanation for their persistence in some fungal mitochondrial genomes. For some introns selection might promote their presence as they may offer a benefit to the host genome such as fine-tuning gene regulation/expression for certain genes as splicing can be a rate limiting step under the control of intron and nuclear encoded factors in addition some introns encode genes such as the mL2450 group I intron that encodes the ribosomal protein RPS3 (Wai et al. 2019). Zubaer et al. (2018) also postulated that there are situations where introns share intron encoded splicing factors thus complementing each other in

situations where one member lost a functional IEP, a form of mutualism among some introns that would counter the loss/extinction predicted by models based on neutral evolution.

Introns and intron-encoded proteins are considered as selfish elements and maintain a life cycle independent of the organism. However, reports showed that introns may have phenotypic effects and possibly play roles important to the biology of the organism. Studies in the 1980s and 1990s showed a possible connection of senescence and the accumulation of circular versions of excised group II intron derived plasmids in *Podospora anserina* (reviewed in Begel et al. 1999). Although, this might be a consequence of other factors associated with triggering senescence in *Podospora* (Osiewacz 2024). Introns were found to be essential for the normal function of mitochondria where the introns play a role of regulators of gene expression in yeast (Rudan et al. 2018). Although group I and II introns have self-splicing capabilities many do depend on co-factors in order to splice efficiently. The processing of transcripts does involve intron and nuclear encoded factors that ensure that the intron RNAs assume splicing competent RNA folds. The presence of introns could act as a rate limiting factor in gene expression and can affect the organism's overall metabolism (Rose 2019). The involvement of nuclear encoded splicing factors for group I and II introns are a means of mitonuclear interplay that could make mitochondrial gene expression respond to environmental and developmental factors (reviewed in Mukhopadhyay and Hausner 2024). Transmissible mitochondrial hypovirulence, a phenomenon where a pathogenic fungus becomes less virulent due to mitochondrial defects, is an example where a splicing defective intron has an effect on the biology of fungi. It was reported that a splicing defective group II intron in the rns gene could be the agent that causes hypovirulence in *Cryphonectria parasitica* (chestnut blight fungus) by reducing the number of assembled

ribosomes (Baidyaroy et al. 2011). A mutation was found in isolates recovered from the last remaining stands of American chestnut trees in Michigan (USA). The “hypervirulent” strains can still infect and survive on their host trees but do not cause the devastating symptoms of their wild type counterparts.

Introns and intron-encoded proteins have applications as gene editing reagents, and they may have medical applications in pathogenic members as intron splicing would be a therapeutic target (Liu and Pyle 2024; Parenteau and Elela 2019; Fedorova et al. 2023). Ribozyme type introns are not found in mammalian genomes therefore group I and II introns offer targets unique to pathogens that carry them. Homing endonucleases are very precise in recognizing a specific DNA pattern/sequence and this specificity is crucial for their application in genome editing (Stoddard 2005; Stoddard 2014; Hafez and Hausner 2012; Guha et al. 2017). HEs can be part of a process that can mediate gene replacements or HEs can be used to generate mutations as double-stranded cuts are usually repaired by non-homologous end joining, a process that is error prone (Stoddard 2014). Recent developments in the CRISPR/Cas9 system are offering more economic options compared to HEs. Although HEs are more precise in their recognition of DNA targets, the cost of engineering HEGs for different target sites cannot compete against the cost of generating guide RNAs for the RNA guided Cas9 or related systems (Guha et al. 2017).

Group I introns have been touted as agents that could be applied in “repairing” defected mRNAs (Fiskaa and Birgisdottir 2010; Gomes et al. 2024); here trans-splicing group I introns can be envisioned as allowing for defective exons being replaced by functional (non-mutated) exons. Group II introns have been genetically engineered as bacterial genome edition reagents

referred to as targetrons. Taking advantage of the EBS/IBS interactions that allow group II introns and their associated RTs (as cofactors) to retrohome into genomic target sites, group II introns can be engineered to insert into specific locations generating mutations (site specific insertional mutations) (Enyeart et al. 2014; Befort and Lambowitz 2019).

1.3 Computational technology for the identification of group I and group II introns

Group I and group II introns have complex arrangements, and it is challenging to identify their positions and type accurately. One of the challenges is the presence of IEPs in the introns that complicate their identification where the programs fail to find the intron-exon boundaries due to the overlap of IEP with intron-core and the presence of IEP makes it complicated for computer programs to find the RNA folding partner. Another issue is their sequence variability; Mobile introns are highly variable in their primary structure (nucleotide sequence) level but conserved in the secondary structure. Based on their secondary structures group I and group II introns can be assigned to various subcategories (Michel and Westhoff 1990; Hausner et al. 2014; Zimmerly and Semper 2015).

There are a few technologies available to experimentally determine the structure of RNA molecules such as X-ray crystallography, NMR spectroscopy, cryogenic electron microscopy (cryo-EM) etc. Those techniques in association with modeling software can potentially generate high-resolution RNA structures. However, due to higher cost, longer time, and difficulties in sample preparation, experimental approaches are still limited and are not feasible at this time for

high throughput or rapid molecular structure determination of complex RNA molecules. Thus, various computational technologies were developed to deduce RNA structures in silico. A lot of the pioneering work on predicting RNA folds for group I and group II introns was done manually by carefully working with curated data sets composed of fungal mitochondrial group I and II intron sequences by Michel and co-workers (examples: Michel et al. 1982; Michel and Dujon 1983; Michel and Westhof 1990; Michel and Ferat 1995).

Over the last three decades various programs have been developed for the prediction of the secondary structures of RNA sequences, such as Mfold (now referred to as UNAfold; Zuker and Stiegler 1981), RNAfold (Hofacker et al. 1994), RNAstructure (Mathews et al. 2004) etc. and similar software are commonly used for RNA structure prediction (Fallmann et al. 2017; Opuu et al. 2022; Sato and Hamada 2023). Those programs use dynamic programming to calculate minimal energy of a possible RNA secondary structure (Zuker and Stiegler 1981). Another recent development is RNAcentral - a database and a set of tools based on sequence similarity to deduce RNA structures (The RNAcentral Consortium; Petrov et al. 2017). RNAweasel was developed for the identification of RNA structures from a given nucleotide sequence, and this has been further developed and implemented in MFannot for annotating mitochondrial genomes which are reservoirs of self-splicing RNAs (group I and group II introns) (Lang et al. 2007). The MFannot program is dedicated to address and solve the challenges in identifying and annotating group I and group II introns in the genomes. It uses manually curated profile HMM models and ERPIN models, and recently is starting to incorporate covariance models (CMs) to better identify group I and group II introns by recognizing evolutionarily conserved signatures in the primary sequence and secondary structure (Lang et al. 2023).

Another program named “Infernal” (Nawrocki et al. 2009; Eddy and Durbin 1994) works on a covariance model (CM) that creates probabilistic profiles of an RNA family using multiple sequence alignment data. The covariance models are statistical models that are used for searching homologous RNA sequences by integrating the primary sequence data and structural information. CM was developed on stochastic context free grammar (SCFG)-based RNA analysis method which is a derivative of the context-free grammar (SGF) of the Chomsky hierarchy of the transformational grammar theory (Sakakibara et al. 1994; Eddy and Durbin 1994). SGF deals with the palindromic language which is appropriate to apply for RNA secondary structures that maintain a long-distance pairwise correlation. CM is sometimes compared with profile-HMM where both models calculate position specific data, however, the difference is that CM base-paired positions have dependency on each other. The use of CM in the Infernal program made it useful in searching for homologs in a particular RNA family. In Infernal v1.1, models have been generated for group I and group II introns from bacterial origin (Nawrocki et al. 2018). Thus, the “cmbuild” function in the Infernal program can be used to create a customized model specific for fungal group I and group II introns. The study by Nawrocki et al. (2018) demonstrated that group I introns are present in archaeal genomes, previously it was assumed members of the Archaea lack group I introns (Hausner et al. 2014).

1.4 Fungal secondary metabolites and its relation to biosynthetic gene clusters

Fungi are great sources of natural products such as antibiotics, mycotoxins, agrochemicals, pharmaceuticals, pigments etc. Those natural products are usually secondary metabolites (SMs), which are not essential for fungal growth, reproduction or energy production but advantageous for its survival in a competitive environment. The exploration of fungal secondary metabolites played a great role in history proving the first ever antibiotic Penicillin. Further many significant contributions were made by the fungal SMs in the field of medicine, agriculture and industries (Zhgun 2023; Keller 2015). Traditionally, the discovery of SMs was solely dependent on chemical analysis from fungal extracts, which was very limited to a few organisms. In addition to that, fungi produce certain SMs only in certain natural environments which can be easily missed under laboratory growth conditions. The modern approach using genomic data - also known as genome mining - can overcome those issues by exploring genomic sequences of fungi and unveiling the unexpressed genetic potential for SM production. Combined with high throughput next generation sequencing protocols this process can be more efficient compared to traditional methods (Atanasov 2021; Mosunova et al. 2021; Domingo-Fernández et al. 2024).

The use of fungal extracts as natural products goes back to the early twentieth century with the discovery of Penicillin during the second world war. Classical work on fungal secondary metabolites was done by Harold Raistrick during the 1930s, who discovered many compounds from *Aspergillus*, *Penicillium* and *Fusarium* (Schor and Cox 2018). This trend was

common until the end of 1970s, however, due to repetitive rediscovery of the same chemicals, commercial enterprises became more interested in artificial chemical synthesis rather than hunting for natural products. That led to the "discovery void" for natural products or secondary metabolites for a decade (Li and Vederas 2009). Interest in finding natural products grew again in 1990 with publications of the first discovery of secondary metabolites gene cluster (SMGC) in *Penicillium chrysogenum* (Diez et al. 1989; Diez et al. 1990; Smith et al. 1990) and this can be considered as a milestone for the genomic phase of SM discovery. The idea of the involvement of a cluster of genes in SM production first established from a gene cloning experiment to transfer penicillin producing genes from *Penicillium crysogenum* to *Neurospora crassa* and *Aspergillus niger* (Trail et al. 1995). Advancement of genetics and genomics played the main role for the shift to a genome-driven approach for determining SMs in fungi by analyzing biosynthetic gene clusters (BGCs) (van der Berg et al. 2007). The key idea of this method is that a set of genes (usually 2 to 20) is involved with a secondary metabolite production so that discovery of BGC can indicate the production of that compound (Yu et al. 2004; Clevenger et al. 2017). The genome-driven approach opened up new possibilities for discovering novel BGCs and SMs; as it scans through the genomes and predicts all possible (cryptic, silent etc.) genes for SM production. A recent study that involved a genus-wide genomic screen of 24 *Penicillium* species indicated that there are 1,317 BGCs present within these species that show great potential to produce novel compounds (Nielsen et al. 2017). The genus-wide screening also made the whole process of SM discovery much faster by high throughput scanning of genomes using new algorithms. Another study of high throughput detection of BGCs in 1000 fungal genomes and their phylogenetic distribution is a good example that created an atlas to interpret and categorize BGCs (Robey et al. 2021). This atlas reveals insights into the diversity, evolution, and functional

potential of fungal BGCs, which are crucial for the production of secondary metabolites like antibiotics and other bioactive compounds.

Fungal secondary metabolites are grouped based on their precursor and enzyme involved. Examples of common precursors of SM are Acetyl-CoA, amino acid, prenyl diphosphate etc. However, the classification of the secondary metabolites is better understood with the key enzyme involved that catalyze the precursor into small molecules. Those enzymes are also found encoded in the BSGCs as well. The enzymes involved in the metabolic pathways of SM production are:

- Polyketides synthase PKS - multidomain enzyme
- Non-ribosomal peptides synthase NRPS - multidomain enzyme
- Terpene cyclase TC - one/two conserved domains
- Dimethyl-allyl-tryptophan synthase DMATS - one/two conserved domains

PKS is the most frequent among the SM groups. Both PKS and NRPS are large enzymes which contain multiple domains for different functions. In fungi, there are two types of PKSs found: type I PKS and type III PKS. Type I PKS can be further categorized into non-reducing (nrPKS), partially reducing (prPKS), and reducing PKS (rPKS). Type I PKS enzymes catalyze acetyl-CoA and malonyl-CoA by elongating their polyketide backbone (Cox 2007; Cox et al. 2018; Herbst et al. 2018). Type III PKSs are comparatively smaller enzymes with a keto-synthase (KS) domain that catalyzes acetyl-CoA (Shimizu et al. 2017). NRPS enzymes are multi-modular enzymes where each module has three conserved domains (Zhang et al. 2016).

The presence of conserved domains in those enzymes facilitate the search for the BGCs in fungal genomes. Although there are different strategies for finding BGCs in a genome (Ziemert et al. 2016), the mainstream approach of identifying BGCs is to find the genes encoding enzymes involved in SM biosynthesis. Based on the enzymes involved, the SMs can be categorized into five major classes – polyketides (PKs), non-ribosomal peptides (NRPs), ribosomally-synthesized and post-translationally modified peptides (RiPPs), terpenoids, and indole alkaloids. The genes that encode those enzymes can be determined by using BLAST (Altschul et al. 1990) and HMMER (Eddy 1998; Finn et al. 2011). Further sophisticated tools have been developed such as BAGEL (de Jong et al. 2006), CLUSEAN (Weber et al. 2009), SMURF (Khaldi et al. 2010), BiG SLiCE (Navarro-Muñoz et al. 2020; Kautsar et al. 2021) etc., but the antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) program (Medema et al. 2011; Blin et al. 2021) has become the most popular. Another notable development has been done by Takeda et al. (2014), which works independent of the prior information of the types of gene clusters in filamentous fungi, rather it looks for the arrangement of genes in a contiguous fashion. Other approaches for BGC-identification can be based on comparative genomics, phylogenetics, or based on target genes such as antibiotic-producing genes, transcriptional or epigenetic regulators etc. Recent development of BGC and SM databases such as antiSMASH-DB (Blin et al. 2019), IMG-ABC (Hadjithomas et al. 2015), and MIBiG 2.0 (Kautsar et al. 2020) are great resources for modern research in BGCs and SMs.

Once the BGCs are identified at the genomic level with the aid of different computational tools and databases, the challenge remains to linking specific metabolites to their corresponding biosynthetic gene clusters (BGCs). Caesar et al. (2023) presented a comprehensive metabolomics

and genomics study of 110 fungi and established a method to link metabolite-gene cluster pairs precisely. This study provides valuable insights into the natural products of fungi and how they can be harnessed for biotechnological and pharmaceutical applications, highlighting key BGCs associated with unique bioactive compounds. Kjærbølling et al. (2019) reviewed different strategies ranging from genome mining to heterologous expression for associating fungal biosynthetic gene clusters (BGCs) with their corresponding secondary metabolites.

1.5 BGCs in Ophiostomatales fungi

The natural reserve of SM is extraordinary, and the number would range in millions (Keller 2019). Although fungi claim a small fraction of the SM produced by all life forms (e.g., plants, bacteria etc.), the amount and the significance of fungal SM is noteworthy. Fungi have complex biochemical pathways that allow them to produce a range of industrially important enzymes (such as cellulases, lipases, ligninolytic enzymes), mycotoxins, antimicrobials and other valuable compounds such as alkaloids, pigments etc. (Bills and Gloer 2016). Discovered fungal compounds have been listed and reviewed in different literature (Schueffler and Anke 2014; Cacho et al. 2015; Goyal et al. 2017; Keller 2019). Although fungi have high potential for SM production, the vast majority of the fungi have not yet been explored. Moreover, the species explored were mostly limited to model organisms (Frisvad et al. 2007; Brakhage 2013). Thus, a lot of potential groups of fungi including Ophiostomatales fungi were left mostly unexplored. The earliest discoveries of SM in Ophiostomatales fungi were made by Ayer et al. (1983) using traditional approaches. They studied *Leptographium wageneri*, and discovered phenolic

compounds such as orcinol, orcinol monomethyl ether, 1,3,6,8-tetrahydroxyanthraquinone, and the a-L-rhamnopyranosides of orcinol and orcinol methyl ether with antibacterial activity (Ayer et al. 1983). *Ceratocystis clavigera*, *C. ips*, *C. huntii*, and *C. minor* (species later assigned to *Ophiostoma* as *O. minus*, or in the case of *C. huntii*, *Grosmannia*) were examined and P-phenethyl alcohol, tryptophol, prolylleucyl anhydride, tyrosol, 3-phenylpropane-1,2-diol (S), 6,8-dihydroxy-3-hydroxymethylisocoumarin, p-hydroxybenzaldehyde, phenylacetic acid, p-hydroxyphenylacetic acid, phenyllactic acid, p-hydroxyphenyllactic acid, and 2,3-dihydroxybenzoic acid, ceratenolone, 6,8-dihydroxy-3-methylisocoumarin, 8-hydroxy-6-methoxy -3-methylisocoumarin, 3,4-dihydro-6,8-dihydrox-y3 -methylisocoumarin, 3,4-dihydro-3-methy I-3,6,8-trihydroxy-3-methylisocoumarin, 3,4-dihydro-3-methyl-3,4,6,8-tetrahydroxyisocoumar, and succinic acid were extracted (Ayer et al. 1985, 1986). Two new xanthones, vertixanthone and hydroxyvertixanthone, two new -pyrones: vertipyronol and vertipyronediol, 1-Hydroxy-8-methoxyanthraquinone, 1,8-dimethoxynaphthalene, ß-sitosterol, ß-sitosteryl palmitate, isoevernin aldehyde, 1,3,6,8-tetrahydroxyanthraquinone, 3,4-dihydro-3,4,8-trihydroxy-l(2H)-naphthalenone, and mycoxanthone were extracted from *Leptographium wageneri* var. *pseudotsugae* (Ayer et al. 1989).

Due to the discovery void in natural compounds that followed in the 1990s studies, no reports could be found on the Ophiostomatales on SMs for decades. In recent years (since 2017), discovery of biosynthetic gene clusters in Ophiostomatales and related fungi were reported. The focus was on SMs that could be linked to plant pathogenicity or maintaining potential symbiotic relationships with insect vectors. The genome mining approach was conducted to study the genomes of *Ophiostoma ulmi* and *O. novo-ulmi* and a unique type of fujikurin-like gene cluster

(OpPKS8), siderophore compound (OpNRPS1), and pyrones and resorcylic acids (OpPKS10, a type III PKS) were found (Sbaraini et al. 2017). The fujikurin was previously reported as an antifungal compound collected from *Fusarium fujikuroi* (Brady and Clardy 2000). Its chemistry and BGC configuration were described by von Bargen (2015) showing four types based on structural variation. Orthologues of fujikurin were found in both pathogenic and non-pathogenic fungi and are expected to play a role in fungi interacting with plants as phytopathogens, endophytes, entomopathogenic, or opportunistic agents (Sbaraini et al. 2017). In addition to their ecological role, Sbaraini et al. (2017) mentioned their potential in biotechnological application and synthesis of similar compounds that could be scaled for medicinal use.

A study has been conducted on genomic analyses to examine the diversity and evolutionary patterns of polyketide biosynthesis gene clusters within the Ceratocystidaceae (in the order Microascales) fungal family – a group of fungi historically aligned with the Ophiostomatales due to similar life histories (Sayari et al. 2018). The study focuses on gene clusters responsible for polyketide synthesis and their final products (such as melanin) and explained their diversity across Ceratocystidaceae species and their evolutionary adaptation from distinct origins.

1.6 Rationale

The Ophiostomatales are a large group of economically important fungi, but their mitochondrial and nuclear genomes so far have been poorly studied. Exploring mitochondrial genomes might yield various novel intron configurations and nuclear genomes may offer insights into the potential of members of the Ophiostomatales (specifically *Leptographium*) for producing novel secondary metabolites. Mobile introns are composed of a ribozyme and in many instances a protein coding component. Both types of products have applications in biotechnology and therefore cataloging their locations ("intron landscapes") is a valuable resource for future workers. This may also provide insights into potential biases in the distribution of introns in mitochondrial genomes and offer some insights into their evolutionary dynamics. Exploring the nuclear genomes for biosynthetic gene clusters may be useful to assess the natural chemistry potential for these fungi, secondary metabolites might be important in supporting the lifestyles of these fungi such as their ability to colonize sapwood, be competitive towards other microbes and establish associations with various insects.

1.7 Hypothesis and Objectives

- Mobile introns contribute towards the genetic diversity and sizes of the mitochondrial genomes in members of the Ophiostomatales. Comparative mitochondrial genome analysis should potentially yield new insights on the evolution of mitochondrial mobile introns and the genetic diversity of mitochondrial genome for members of the Ophiostomatales.

- Members of the genus *Leptographium* encode biosynthetic gene clusters within their nuclear genomes, potentially encoding for the production of secondary metabolites that might be of pharmaceutical value. Applying bioinformatics tools should provide information on the types of BGC these fungi possess and their potential impact on the biology of these fungi.

CHAPTER 2: THE MITOGENOMES OF *OPHIOSTOMA MINUS* AND *OPHIOSTOMA PILIFERUM* AND COMPARISONS WITH OTHER MEMBERS OF THE OPHIOSTOMATALES

CHAPTER 2: THE MITOGENOMES OF *OPHIOSTOMA MINUS* AND *OPHIOSTOMA PILIFERUM* AND COMPARISONS WITH OTHER MEMBERS OF THE OPHIOSTOMATALES

2.1 ABSTRACT

Fungi assigned to the Ophiostomatales are of economic concern as many are blue-stain fungi and some are plant pathogens. The mitogenomes of two blue-stain fungi, *Ophiostoma minus* and *Ophiostoma piliferum*, were sequenced and compared with currently available mitogenomes for other members of the Ophiostomatales. Species representing various genera within the Ophiostomatales have been examined for gene content, gene order, phylogenetic relationships, and the distribution of mobile elements. Gene synteny is conserved among the Ophiostomatales but some members were missing the *atp9* gene. A genome wide intron landscape has been prepared to demonstrate the distribution of the mobile genetic elements (group I and II introns and homing endonucleases) and to provide insight into the evolutionary dynamics of introns among members of this group of fungi. Examples of complex introns or nested introns composed of two or three intron modules have been observed in some species. The size variation among the mitogenomes (from 23.7 kb to about 150 kb) is mostly due to the presence and absence of introns. Members of the genus *Sporothrix sensu stricto* appear to have the smallest mitogenomes due to loss of introns. The taxonomy of the Ophiostomatales has recently undergone considerable revisions; however, some lineages remain unresolved. The data showed that genera such as *Raffaelea* appear to be polyphyletic and the separation of *Sporothrix sensu stricto* from *Ophiostoma* is justified.

2.2 INTRODUCTION

Members of the Ophiostomatales are frequently associates of bark beetles that can serve as vectors for these fungi. Some members are referred to as ambrosia fungi as they exist in symbiotic relationships with wood boring ambrosia species (Vanderpool et al. 2018). Most species of the Ophiostomatales are either non-pathogenic or weak pathogens; some species can kill trees in combination with their beetle vectors or without any contribution by an arthropod vector (Wingfield et al. 2017). Many members of the Ophiostomatales cause blue-stain of sap wood in hard- and softwood species. Sap-staining fungi are responsible for considerable economic losses in the Forestry sector due to difficulties in exporting stained timbers/lumber products (Uzunovic and Byrne 2013).

Ophiostoma minus is an important agent of blue stain in various pine species (Klepzig 1998; Gorton and Webber 2000; Chang et al. 2019) and has been shown to be a potential pathogen of pine (Gorton et al. 2004; Ben Jamaa et al. 2007). *Ophiostoma piliferum* is a serious blue-stain agent on a variety of conifer species but it is not considered to be pathogenic on softwoods (Linnakoski et al. 2012b). Both *O. minus* and *O. piliferum* have been reported from many geographic regions and from a variety of hosts and they could represent species complexes (Chakravarty et al. 1994; Gorton and Webber 2000; Hafez and Hausner 2011; Jankowiak and Bilański 2013; Bilto and Hausner 2016).

Only a few mitochondrial genomes have been characterized so far for members of the Ophiostomatales (Abboud et al. 2018; Zhang et al. 2019). Fungal mitochondrial genomes encode

genes involved in translation, such as the small and large ribosomal subunit RNAs (*rns* and *rnl*) and a set of tRNAs, and protein components involved in electron transport chain and oxidative phosphorylation. This includes parts of Complex I (subunits of NADH dehydrogenase: *nad1* to *nad6*and *nad4L*; except for members of the Taphrinomycota and some members of the Saccharomycetales), components of Complex III (*cob*) and Complex IV (*cox1*, *cox2*, and *cox3*), plus members of Complex V (ATP synthase components: *atp6*, *atp8*, and usually *atp9*). Many fungi encode a ribosomal protein (*rps3*) (Hausner 2003; Freel et al. 2015; Wai et al. 2019) and the RNA (*rnpB* gene) component for RNaseP has also been recorded in some fungal mitochondrial genomes (Lang 2014). In addition, fungal mitogenomes can encode potential orphan genes (genes with unknown functions and a lack of detectable homologs) and in some members of the Ascomycota mitochondrial open reading frames (ORFs) have been detected that appear to encode putative N-acetyltransferases and amino-transferases (Wai et al. 2019).

Organellar introns in plants and fungi can be self-splicing (ribozymes). However, intron splicing is enhanced by intron- and/or host genome-encoded (nuclear or mitochondrial) factors (Lang et al. 2007; Hausner 2012; Schmitz-Linneweber et al. 2015). Based on intron RNA folds (secondary structure) and their splicing mechanisms fungal mitochondrial introns can be assigned to either group I or group II introns (Michel and Westhof 1990; Lambowitz et al. 1999). There are a few instances of complex introns where an intron has inserted into another intron, and these are sometimes referred to as twintrons or nested introns (Hafez and Hausner 2015; Deng et al. 2016; Deng et al. 2018; Guha et al. 2018; Zumkeller et al. 2020). Nested introns can be composed of group I intron modules or a combination of group I and group II

intron modules (Hafez et al. 2013; Guha and Hausner 2016; Guha et al. 2018). Group I and group II introns can encode intron-encoded proteins (IEPs) that can catalyze the movement of an intron from an intron-containing allele to cognate alleles that do not have introns (Dujon 1989), a process that is referred to as intron homing or retro-homing, if mediated by reverse transcriptase activity. Group I intron IEPs typically are homing endonucleases (HEs), which are DNA-cutting enzymes that facilitate intron homing or maturases that facilitate intron splicing. There are examples of intron IEPs that have maturase and HE activity (Belfort 2003; Caprara and Waring 2005). Two families of HEs, named after the presence of conserved amino-acid motifs, are found in fungal mitochondrial genomes: the LAGLIDADG and the GIY-YIG families of HEs (Stoddard 2014). HEs can be encoded by independent free-standing genes or their genes (HEGs) are embedded within intronic sequences. It has been reported that HEGs can move independently from their ribozyme partners (Mota and Collins 1988), although recent studies suggest that intron-encoded HEG co-evolve with their ribozyme partners (Megarioti and Kouvelis 2020). Finally, there are instances where group II introns encode HEGs; typically group II introns can be ORF-less or encode reverse transcriptases (Toor and Zimmerly 2002; Mullineux et al. 2010; Hafez and Hausner 2012; Zimmerly and Semper 2015).

Herein, we report the mitochondrial genomes for *O. minus* and *O. piliferum*. As more sequences for members of the Ophiostomatales become available mitochondrial DNA could provide a resource for developing markers that allow for distinguishing among various *Ophiostoma* species and allow for resolving some of the taxonomic issues that still need to be addressed with regards to circumscribing species complexes and lineages within *Ophiostoma sensu lato*.

2.3 Materials and Methods

2.3.1 Fungal Strains, Cultivation, and Preparation of DNA

A strain of *Ophiostoma minus* C262 [= WIN(M)495] [Northern Forest Research Centre, Edmonton, AB, Canada; isolated from Pinus contorta; WIN(M) = University of Manitoba]; and a strain of *Ophiostoma piliferum* UAMH 7459 [= NoF1929, = WIN(M)959; isolated near Nelsen, BC, Canada from Populus tremuloides; UAMH = UAMH Centre for Global Microfungal Biodiversity, Dalla Lana School of Public Health, University of Toronto] were grown at 20°C on Malt extract agar plates (per 1 L: 20 g Agar, 30 g Malt extract, and 1 g Yeast extract). After 7 days of growth, small agar plugs (20 plugs: ~ 2 mm × 2 mm) were transferred into 500 ml of YPD broth (per 1 L: 1 g Yeast extract, 1 g Peptone, 3 g Dextrose) medium. The cultures were incubated for 7 days at 20°C and the mycelium was harvested using vacuum filtration. Approximately four grams (wet weight) of mycelium were collected and DNA was extracted and quantified using the DNA extraction protocol described previously in Abboud et al. (2018).

2.3.2 Sequencing and Assembly of Mitochondrial Genomes

One hundred ng of DNA in 75 µl of H2O was supplied to McGill University and Génome Québec Innovation Centre (McGill University, QC, Canada) for shotgun Illumina sequencing using the MiSeq platform. The DNA preparations were part of a set of 20 DNA samples that were individually barcoded and thereafter pooled for sequencing (Abboud et al. 2018; Zubaer et al. 2018). Average size reads were 250 nt and average quality was 35 and quality offset was 33. The paired-end reads were trimmed to remove the barcodes/adaptor sequences and assembled de novo into contigs and scaffolds with the A5-miseq pipeline [(Coil et al. 2015); McGill University

and Génome Québec Innovation Centre; Canadian Center for Computational Genomics (C3G)]. For *O. minus*, a scaffold of 91,847 nt and for *O. piliferum* a scaffold of 69,966 nt could be recovered. These scaffolds, based on BLAST searches against the NCBI non-redundant database, contained mitochondrial sequences that showed matches with mitochondrial gene sequences previously deposited for *O. novo-ulmi* subsp. *novo-ulmi* (GenBank accession number: MG020143.1).

2.3.3 Annotation of Mitochondrial Genomes

The mitochondrial DNA sequences were annotated with the aid of the following programs: MFannot and RNAweasel (Gautheret and Lambert 2001; Lang et al. 2007). The MFannot program (setting genetic code 4; the mold, protozoan, and coelenterate mitochondrial code) predicts protein-encoding genes, ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), potential intron/exon junctions, and intron types. In addition, the online tRNAscan-SE2 (Chan and Lowe 2019) program was applied to the data set to verify the prediction of tRNA genes. For a precise annotation, gene sequences were individually verified starting with aligning gene sequences to similar sequences (acquired via BLASTn) in closely related fungi using the MAFFT (Katoh and Standley 2013) and AliView (Larsson 2014) programs. Intron/exon boundaries were predicted based on alignments with intron-less versions of cognate alleles from other fungal species. The RNAweasel program (Lang et al. 2007) was used to identify tRNAs, rRNAs, and intron types. All ORFs, introns, and intergenic spacers were identified using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) and nucleotide 6-frame translation-protein BLAST (BLASTx) searches against NCBI databases (Sayers et al. 2022). Feature tables were generated and, along with the mitochondrial genome sequences, these were analyzed further with

Artemis (Rutherford et al. 2000) to refine the annotation (GenBank accession numbers MW122509.1 and MW122508.1 for *O. minus* and *O. piliferum*, respectively). The output from Artemis was adjusted and applied to Circos program (Krzywinski et al. 2009) for visualization of the annotated mitogenomes showing the genes, tRNA, introns, and a feature-wise GC plot (calculating GC for every genetic feature such as exon, intron, intergenic region, etc.).

Intron nomenclature for *rns* and *rnl* introns was based on Johansen and Haugen (2001); for mtDNA protein coding sequence (CDS) we used the *Saccharomyces cerevisiae* CDS sequences to map introns (Guha et al. 2018), and for the *nad* genes we applied the *nad* sequences from *Neurospora crassa* (Zubaer et al. 2019).

2.3.4 Comparative Mitogenomic Analysis

The *O. minus* and *O. piliferum* mitogenomes were compared with other available sequences for members of the Ophiostomatales. In some instances, sequences were available in NCBI as whole genome sequence data sets; see Supplementary Table 2.1 for accession numbers (Van der Nest et al. 2014; Wingfield et al. 2015a; Wingfield et al. 2015b; Wingfield et al. 2016a; Wingfield et al. 2016b; Wingfield et al. 2017; Vanderpool et al. 2018). The whole genome assemblies were searched for mitochondrial sequences with BLASTn using the mitogenome of *O. novo-ulmi* subsp. *novo-ulmi* (Abboud et al. 2018) as the query. The recovered scaffolds were examined, annotated, and validated as described above utilizing MFannot, RNAweasel, and MAFFT (assemblies and annotations are presented in Supplementary Data File 1). A panintronic landscape was visualized with the aid of the Circos program.

2.3.5 RNA Folding of Selected Complex Introns

Group I and II intron classifications and their secondary core elements/folds were predicted by the RNAweasel program (Lang et al. 2007). For group I introns the P1, P2, P5, P6, P7.1 (and stem), and P9 helices were predicted by Mfold (Zuker 2003) and these helices were supported by comparative sequence analysis (including multiple sequence alignments) with related intron sequences. Group II introns were identified based on the highly conserved domain V sequence and the group II intron RNAs were drawn with the aid of Mfold to optimize the expected secondary and tertiary interactions known to stabilize group II introns RNAs in a splicing competent fold (Toor et al. 2001; Michel et al. 2009; Marcia et al. 2013). Secondary structure of introns and their features were based on existing models (Michel and Westhof 1990; Jaeger et al. 1991; Deng et al. 2016; Guha et al. 2018). Sequence alignments with related introns and their flanking boundaries sequences confirmed the intron boundaries and classifications. The final introns folds were drawn using CorelDRAW Graphics Suite X6 (Corel Corporation, Ottawa, ON, Canada).

2.3.6 Phylogenetic Analysis

A phylogenetic tree was generated to infer the phylogenetic position of *O. minus* and *O. piliferum* among other members of the Ophiostomatales. Twenty-three mitogenomes were available for the Ophiostomatales from the NCBI genome and GenBank databases. In addition, sequences utilized in Abboud et al. (2018) and Zubaer et al. (2018) were also included to evaluate the phylogenetic distribution of members the Ophiostomatales. The analysis was based on concatenated amino-acid sequences of 13 mitochondrial proteins encoded by the following genes: *atp6*, *8*, *cob*, *cox1–3*, *nad1–6*, and *nad4L*. The scaffold representing the mitogenome

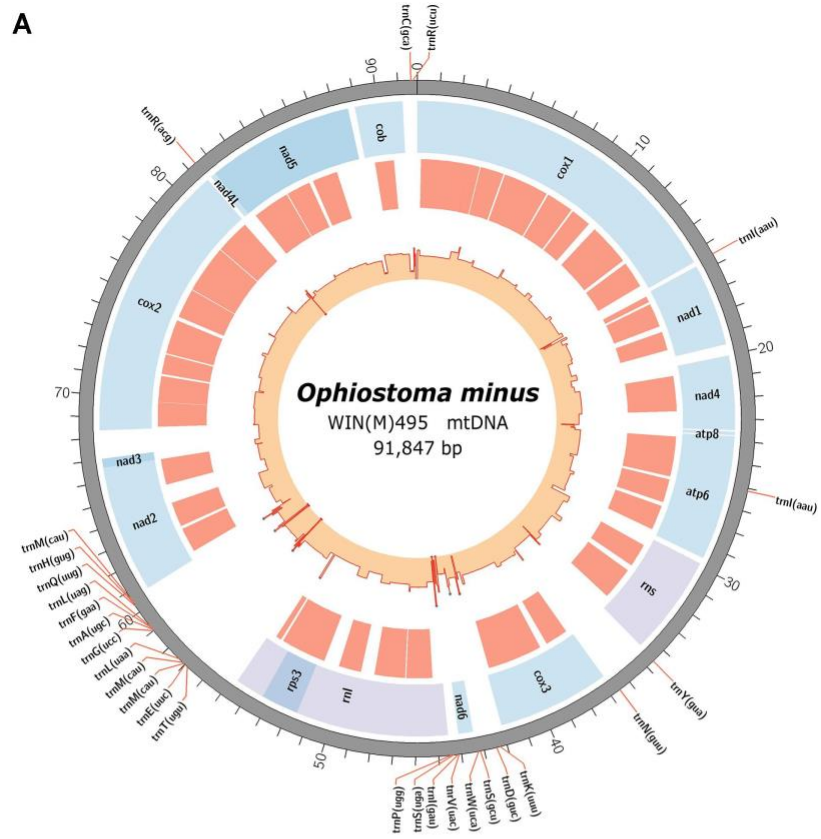
of *Ophiostoma ips* (NTMB01000349.1; scaffold_143) appears to be missing a small segment that includes *atp8*, therefore a separate analysis based on 12 concatenated protein sequences was conducted to confirm the position of *O. ips*. Forty-eight mitogenomes represented by concatenated amino-acid sequences were aligned with the MAFFT program using its iterative refinement method (FFT-NS-i). The aligned dataset was used for tree construction with the MrBayes program (Ronquist et al. 2012) applying the mixed model setting, which finally determined and used the best fit model as cpREV (Adachi et al. 2000). The analysis was performed by running 1 million generations with sample frequency set at 1000. For the sampled trees, the burn-in value was 25% to construct the majority rule consensus tree and for assessing posterior probability values. Two Eurotiales mitogenomes (*Aspergillus fumigatus* and *Penicillium digitatum*) were used as outgroup of the tree and re-rooted accordingly.

2.4 Results

2.4.1 Organization and Features of the Mitochondrial Genomes

The newly obtained mitogenomes of *O. minus* and *O. piliferum* can be represented as circular molecules of 91,847 and 69,966 nt, respectively (Figures 2.1A, 2.1B). The genomes encode the following 14 protein coding genes: *atp6*, *atp8*, *cob*, *cox1–cox3*, *nad1–nad6*, *nad4L*, and *rps3*. In addition, the genomes encode the following RNA structural genes: 27 tRNAs in *O. minus* and 25 tRNAs in *O. piliferum*, plus the small and large ribosomal RNA subunit genes

(*rns* and *rnl*, respectively). The ribosomal protein RPS3 is encoded by a group IA type intron inserted within the *rnl* gene. 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 (Aguileta et al. 2014; Abboud et al. 2018). The mitogenomes of *O. minus* and *O. piliferum* do not appear to encode the *atp9* and *rnpB* genes. For both mitogenomes all genes are encoded by the same strand and the gene order is as follows: *cox1*, *nad1*, *nad4*, *atp8*, *atp6*, *rns*, *cox3*, *nad6, rnl* (including *rps3*), *nad2*, *nad3*, *cox2*, *nad4L*, *nad5*, and *cob*. Gene order is conserved across the 25 examined members of the Ophiostomatales and in 22 species we noted the presence of the *atp9* gene, located between *nad3* and *cox2*.


A
Ophiostoma minus
WIN(M)495 mtDNA
91,847 bp
cox1
cox2
cob
nad1
nad2
nad3
nad4
nad4L
nad5
nad6
atp6
atp8
rnl
rns
rps3
trnI(aau)
trnY(gua)
trnN(guu)
trnK(uuu)
trnD(guc)
trnS(gcu)
trnW(uca)
trnV(uac)
trnP(ugg)
trnR(acg)
trnM(cau)
trnH(gug)
trnQ(uug)
trnL(uag)
trnF(gaa)
trnA(ugc)
trnG(ucc)
trnL(uaa)
trnM(cau)
trnM(cau)
trnE(uuc)
trnT(ugu)
trnC(gca)
trnR(ucg)

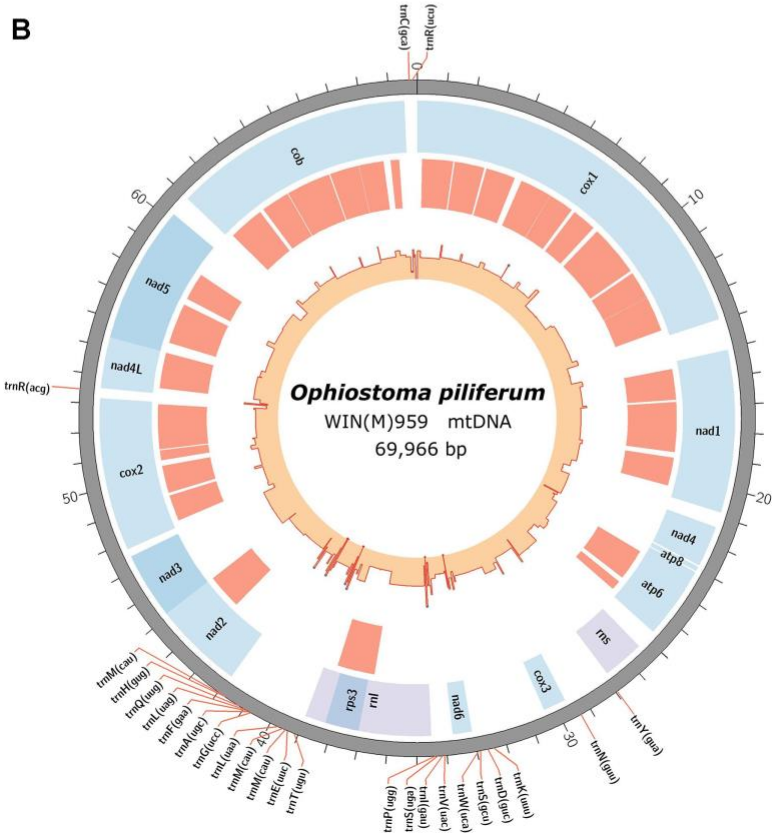

B
Ophiostoma piliferum
WIN(M)959 mtDNA
69,966 bp
cox1
cox2
cob
nad1
nad2
nad3
nad4
nad4L
nad5
nad6
atp6
atp8
rnl
rns
rps3
trnI(aau)
trnY(gua)
trnN(guu)
trnK(uuu)
trnD(guc)
trnS(gcu)
trnW(uca)
trnV(uac)
trnP(ugg)
trnR(acg)
trnM(cau)
trnH(gug)
trnQ(uug)
trnL(uag)
trnF(gaa)
trnA(ugc)
trnG(ucc)
trnL(uaa)
trnM(cau)
trnM(cau)
trnE(uuc)
trnT(ugu)
trnC(gca)
trnR(ucg)

Figure 2.1: Circular representation of the mitochondrial genomes showing the tRNA (pointed outward on the scale), protein coding genes (blue track), introns (red track), and genetic-feature-wise GC graph (GC% was calculated for annotated features instead of fixed window, showed in the innermost track). **(A)** The annotated mitochondrial genome of *Ophiostoma minus*; total size of the circular genome is 91,847 bp (GenBank accession: MW122509.1). **(B)** The annotated mitochondrial genome of *Ophiostoma piliferum*; total size of the circular genome is 69,966 bp (GenBank accession: MW122508.1).

C. brevicomis							YN	KDSW	VISP	TEMMLGAFLQHMV			+		R		CR
C. minuta							YN	KDSW	VISP	TEMMLGAFLQHMM			+		R		CR
E. vermicola							YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
F. purpurea							YN	KDSW	VISSP	TEMMLGAFLQHMI			+		R		CR
G. fragrans		N					YN	KDSW	VISP	TEMMLGAFLQHMM			+		R		CR
Gr. penicillata	R						YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
H. lignivorus							YN	KDSW	VISP	TEMMLGAFLQHMM			+		R		CR
L. lundbergii							YN	KDSW	VISP	TEMMLGAFLQHML			+		R		CR
O. ips							YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
O. minus		I			I		YN	KDSW	VISP	TEMMLGAFLQHM			-		R		CR
O. novo-ulmi							YN	KDSW	VISP	TEMMLGAFLQHM			-		R		CR
O. piliferum							YN	KDSW	VISP	TEMMLGAFLQHM			-		R		CR
R. albimanens		KK					YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
R. ambrosiae		N					YN	KNDSW	VISP	TEMMLGAAFLQHM			+		R		CR
R. arxii							YN	KDSW	VISP	TEMMLGAAFLQHM			+		R		CR
R. lauricola		F					YN	KDSW	VISP	TEMMLGAFLQHMX		K	+		R		CR
R. quercivora							YN	KDSW	VISP	TEMMLGAAFLQHML			+		RI		CR
R. quercus-mongolicae							YN	KDSW	VISP	TEMMLGAFLQHMI			+		R		CR
R. sp.							YN	KDSW	VISSP	TEMMLGAFLQHMM			+		R		CR
R. sulphurea						L	YN	KDSW	VISP	TEMMLGAFLQHML		R	+		R		KCR
S. brasiliensis		X					YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
S. globosa		X					YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
S. insectorum							YN	KDSW	VISSP	TEMMLGAFLQHMI			+		R		CR
S. pallida		F					YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR
S. schenckii		X					YN	KDSW	VISP	TEMMLGAFLQHM			+		R		CR

Figure 2.2: Schematic representation of gene order and position of tRNA gene clusters for members of the Ophiostomatales. Gene order is conserved across the 25 sampled Ophiostomatales with minor variations in tRNA composition and the presence or absence of the *atp9* gene. Genes encoding for tRNAs is represented using their respective single-letter amino acid codes. Intron-encoded tRNAs are represented by placing them under the gene that encodes them. Plus (+) and minus (–) signs represent presence and absence of a gene, respectively, and only applies to the *atp9* gene. See Supplementary Table 2.1 for GenBank NCBI accession numbers.

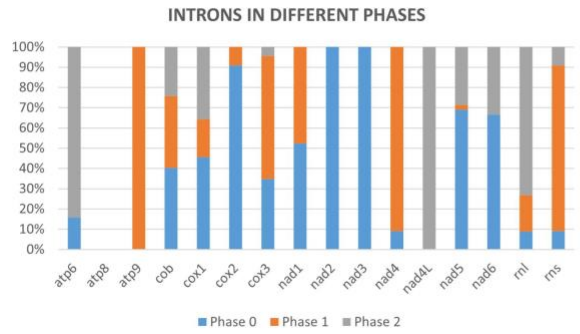
The synteny with regards to the tRNA genes are conserved among the examined members of the Ophiostomatales (see Figure 2.2). Most tRNAs genes are arranged in clusters located between the *cox3* and *nad6* genes (a cluster of 4–5 tRNA genes), the *nad6* and *rnl* genes (a cluster of 4–5 tRNA genes), and the largest grouping of tRNA genes was detected between the *rnl* and *nad2* genes (a cluster of 12–14 tRNA genes). In *O. minus*, one putative tRNA gene appears to be encoded within an intron in the *atp6* gene (*atp6*-i1 or *atp6*–173). Between the *cox2* and *nad4L*gene, all examined members of this order contain the tRNA gene for Arg (R); between the *rns* and *cox3* gene, the tRNA genes for Tyr (Y) and Asn (N), and the tRNA genes for Cys (C) and Arg (R) are positioned between the *cob* and *cox1*gene. The intergenic region separating the *cox1* and *nad1* genes appears to be quite diverse with regards to tRNA genes with 16 members showing no indication for the presence of tRNA genes and others showing the presence of tRNA genes for Asn (N), Arg (R) (intron-encoded, *cox1*–1281), Ile (I), Lys (K), Asn (N), Phe (F), or X (a highly derived tRNA gene predicted to bind to phenylalanine; Lang et al. 2012).

2.4.2 Ophiostomatales and Their Mitogenome Intron Complement

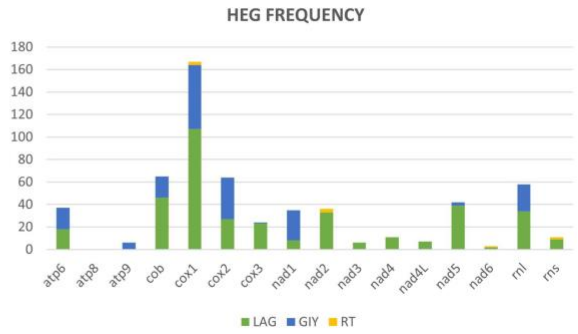
For the examined members of the Ophiostomatales the mitogenomes range in size from 23,830 bp (*Raffaelea* sp. RL272) to >150 kb (*Raffaelea quercivora*). The smallest mitogenomes belong to members of *Sporothrix sensu stricto* (*Sporothrix schenckii*: 26,095 bp; *Sporothrix globosa*: 26,671 bp), and the available sequences for *Graphilbum fragrans*: 25,567 bp and *Hawksworthiomyces lignivorus*: 27,092 bp. These smaller genomes are devoid of introns except for the RPS3 encoding group IA intron located in the *rnl* gene. In addition, the mitogenomes of *H. lignivorus* and *G. fragrans* contain one additional intron in the *cox1* gene.

Mitochondrial intron numbers range from 1 to 64 introns per genome across the examined Ophiostomatales mitogenomes (Supplementary Table 2.1 and Figure 2.3). Combined, 594 putative introns (complex/nested introns were treated as one item) were recorded based on structural features, 573 could be assigned to be group I introns, 15 introns are group II type introns, and six introns could not be assigned to any category. A total of 118 intron insertion sites were identified across the various mtDNA genes. Among these, 94 insertion sites were noted to be in protein coding genes with 55 sites in phase 0 (intron does not disrupt a codon), 19 in phase 1 (intron position after the first nucleotide of the codon), and 20 sites occupied a phase 2 position (intron insertion after the second nucleotide in the codon) (Figure 2.3A). With regards to the observed 594 introns, 505 were inserted in protein coding genes and among those 270 where in phase 0, 109 in phase 1, and 126 introns in phase 2 (Figure 2.3A). The rRNA genes had 24 intron insertion sites (18 within the *rnl* gene and six within the *rns* gene); these sites accounted for 89 introns. Twenty intron insertion sites among the 118 intron insertion sites had 10 or more introns present accounting for 246 introns.

A


INTRONS IN DIFFERENT PHASES
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
atp6 atp8 atp9 cob cox1 cox2 cox3 nad1 nad2 nad3 nad4 nad4L nad5 nad6 rnl rns
Phase 0 Phase 1 Phase 2

B


HEG FREQUENCY
180
160
140
120
100
80
60
40
20
0
atp6 atp8 atp9 cob cox1 cox2 cox3 nad1 nad2 nad3 nad4 nad4L nad5 nad6 rnl rns
LAG GIY RT

C

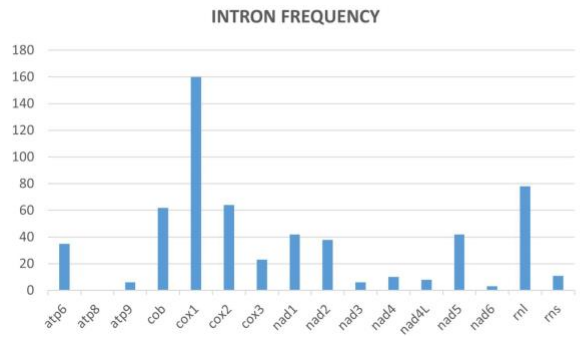

INTRON FREQUENCY
180
160
140
120
100
80
60
40
20
0
atp6 atp8 atp9 cob cox1 cox2 cox3 nad1 nad2 nad3 nad4 nad4L nad5 nad6 rnl rns

Figure 2.3: (A) Mitochondrial introns of the Ophiostomatales categorized according to intron phasing on a gene-by-gene basis. **(B)** Representation of the relative frequencies of different types of introns encoded open reading frames based on gene-by-gene basis (LAG, LAGLIDADG type homing endonucleases/maturases; GIY, GIY-YIG type homing endonucleases; RT, reverse transcriptases). **(C)** Relative distribution and number of mitochondrial introns recorded on a gene-by-gene basis among the examined members of the Ophiostomatales.

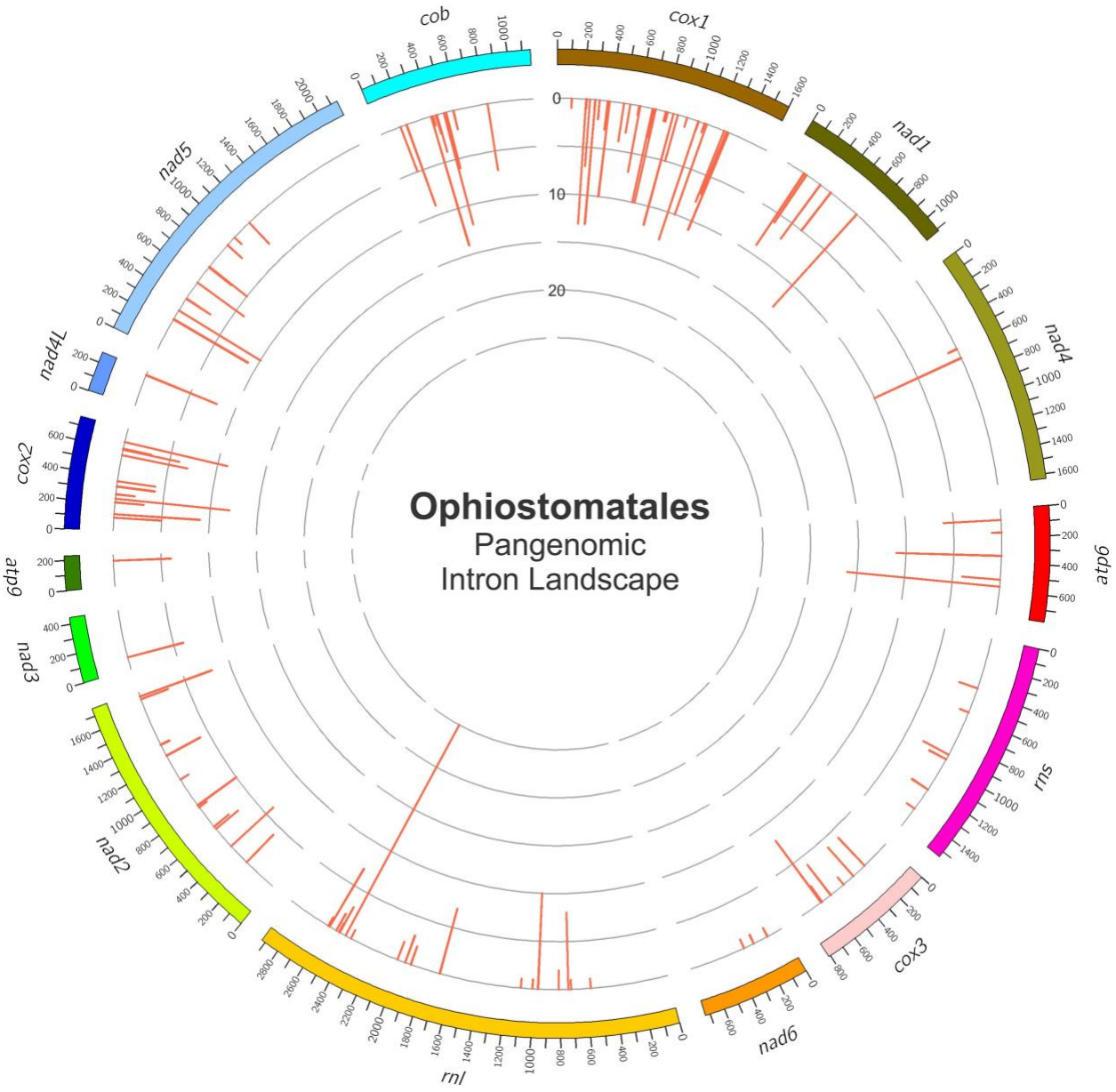

Ophiostomatales
Pangenomic
Intron Landscape
cob
cox1
nad1
nad4
atp6
nad3
nad2
rnl
nad6
cox3
rns
nad5
nad4L
cox2
atp9
0
10
20

Figure 2.4: The panintronic landscape for the studied members of the Ophiostomatales. The landscape was generated by Circos and shows all intron insertions sites and their frequencies. More detailed intron landscapes showing intron types, intron-encoded protein types, and introns phasing are shown in Supplementary Figures 2.1A–D.

Group I introns were either ORF-less or encoded LAGLIDADG or GIY-YIG type ORFs. Among the 15 group II introns, three were ORF-less, 9 encoded reverse transcriptase-like ORFs, and three encoded LAGLIDADG type ORFs (see Supplementary Table 2.2 and Figure 2.3B). Among the Ophiostomatales, the mS722 and mL952 group II introns encoded LAGLIDADG type ORFs. The mL2450 group IA intron encodes the RPS3 protein and in a few instances the rps3-coding sequence was fused in-frame to a LAGLIDADG HE-coding sequence (Gibb and Hausner, 2005). Among the 25 mitogenomes examined, the intron-rich genes were as follows (with total intron/insertion numbers listed in brackets):

cox1 (161), *rnl* (78), *cox2*(65), *cob* (62), *nad1* (42), *nad5* (42), *nad2* (38), and *atp6* (38).

The *cox1* gene was observed to have the most intron insertion sites at 29 (one element at *cox1*–264 could not be classified) with the *rnl* gene having 18 insertion sites (Figure 2.3C). The panintronic landscape for the studied members of the Ophiostomatales is illustrated in Figure 2.4 and more detailed intron landscapes showing intron types, IEPs, and introns phasing are shown in Supplementary Figures 2.1A–D.

2.4.3 Phylogenetic Groupings Observed With Mitogenome Analysis

The phylogenetic analysis of 48 concatenated mitochondrial protein sequences including 25 species that belong to the Ophiostomatales yielded a topology showing the following monophyletic groupings: Microascales, Hypocreales, Glomerellales, Sordariales, and Ophiostomatales (Figure 2.5). Within the Ophiostomatales, several lineages could be identified representing the following genera: *Ceratocystiopsis*, *Graphilbum, Hawksworthiomyces*, *Raffaelea sensu stricto*, *Ophiostoma sensu stricto*, and *Sporothrix* (Figure 2.5). However, the mitochondrial sequences failed to show monophyly for species assigned to *Sporothrix* and *Raffaelea*.

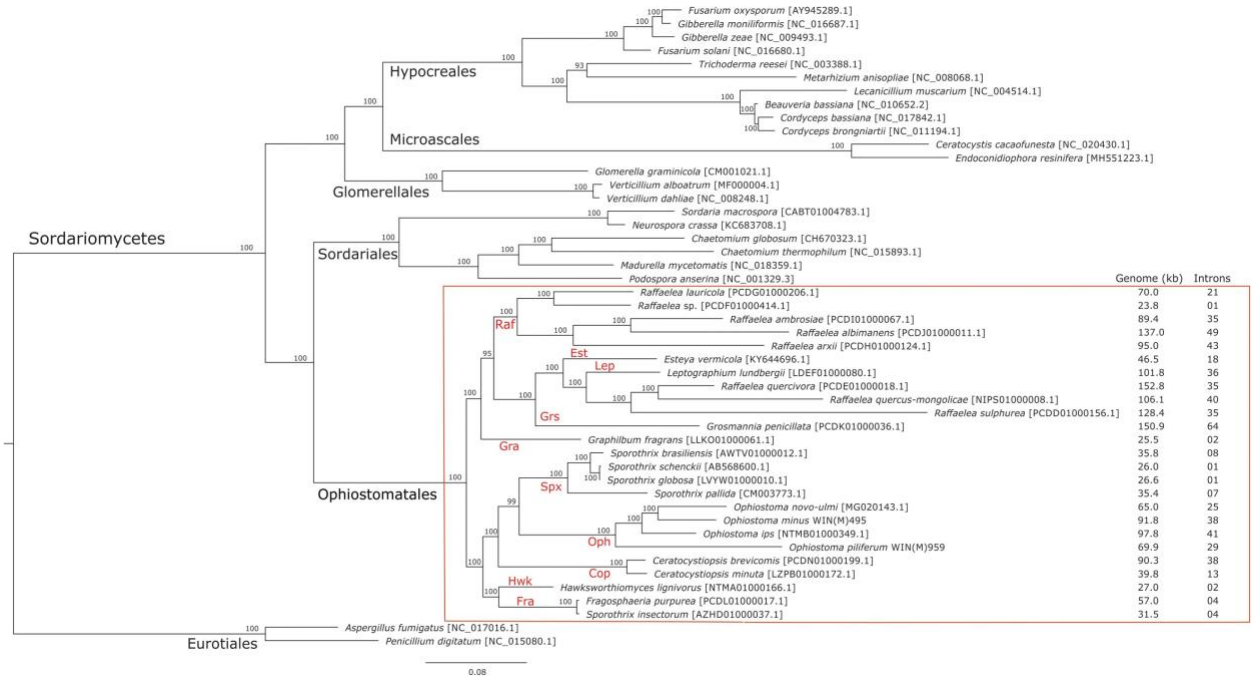

Fusarium oxysporum [AY945289.1]
Gibberella moniliformis [NC_016687.1]
Gibberella zeae [NC_009493.1]
Fusarium solani [NC_016680.1]
Trichoderma reesei [NC_003388.1]
Metarhizium anisopliae [NC_008068.1]
Lecanicillium muscarium [NC_004514.1]
Beauveria bassiana [NC_010652.2]
Cordyceps bassiana [NC_017842.1]
Cordyceps brongniartii [NC_011194.1]
Ceratocystis cacaofunesta [NC_020430.1]
Endoconidiophora resinifera [MH551223.1]
Hypocreales
Microascales
Glomerellales
Glomerella graminicola [CM001021.1]
Verticillium alboatrum [MF000004.1]
Verticillium dahliae [NC_008248.1]
Sordariomycetes
Sordaria macrospora [CABT01004783.1]
Neurospora crassa [KC683708.1]
Sordariales
Chaetomium globosum [CH670323.1]
Chaetomium thermophilum [NC_015893.1]
Madurella mycetomatis [NC_018359.1]
Podospora anserina [NC_001329.3]
Genome (kb) Introns
Raffaelea lauricola [PCDG01000206.1] 70.0 21
Raffaelea sp. [PCDF01000414.1] 23.8 01
Raf
Raffaelea ambrosiae [PCDI01000067.1] 89.4 35
Raffaelea albimanens [PCDJ01000011.1] 137.0 49
Raffaelea arxii [PCDH01000124.1] 95.0 43
Est
Esteya vermicola [KY644696.1] 46.5 18
Lep
Leptographium lundbergii [LDEF01000080.1] 101.8 36
Raffaelea quercivora [PCDE01000018.1] 152.8 35
Raffaelea quercus-mongolicae [NIPS01000008.1] 106.1 40
Grs
Raffaelea sulphurea [PCDD01000156.1] 128.4 35
Grosmannia penicillata [PCDK01000036.1] 150.9 64
Graphilbum fragrans [LLKO01000061.1] 25.5 02
Gra
Sporothrix brasiliensis [AWTV01000012.1] 35.8 08
Sporothrix schenckii [AB568600.1] 26.0 01
Sporothrix globosa [LVYW01000010.1] 26.6 01
Spx
Sporothrix pallida [CM003773.1] 35.4 07
Ophiostomatales
Ophiostoma novo-ulmi [MG020143.1] 65.0 25
Ophiostoma minus WIN(M)495 91.8 38
Ophiostoma ips [NTMB01000349.1] 97.8 41
Oph
Ophiostoma piliferum WIN(M)959 69.9 29
Ceratocystiopsis brevicomis [PCDN01000199.1] 90.3 38
Cop
Ceratocystiopsis minuta [LZPB01000172.1] 39.8 13
Hwk
Hawksworthiomyces lignivorus [NTMA01000166.1] 27.0 02
Fra
Fragosphaeria purpurea [PCDL01000017.1] 57.0 04
Sporothrix insectorum [AZHD01000037.1] 31.5 04
Aspergillus fumigatus [NC_017016.1]
Penicillium digitatum [NC_015080.1]
Eurotiales
0.08

Figure 2.5: Phylogenetic tree of mitogenomes showing the position of *Ophiostoma minus and Ophiostoma piliferum* among members of the Ophiostomatales. The tree topology (50% majority-rule consensus tree) was generated by the MrBayes program and involved 48 concatenated amino acid sequences. The tree is drawn to scale with branch length measured in the number of substitutions per site. Posterior probability values are indicated at the nodes. NCBI and GenBank accession numbers (except for *Verticillium alboatrum*, which refers to MitoFun database) are indicated in square brackets. For the members of the Ophiostomatales, mitogenome sizes, and total numbers of introns are listed for each genome. Taxonomic designations (Orders, and Genera for the Ophiostomatales) are indicated on the relevant branches. Raf, *Raffaelea*; Gra, *Graphilbum*; Est, *Esteya*; Grs, *Grosmannia*; Lep, *Leptographium*; Spx, *Sporothrix sensu stricto*; Cop, *Ceratocystiopsis*; Oph, *Ophiostoma sensu stricto*; Hwk, *Hawksworthiomyces*; Fra, *Fragosphaeria*.

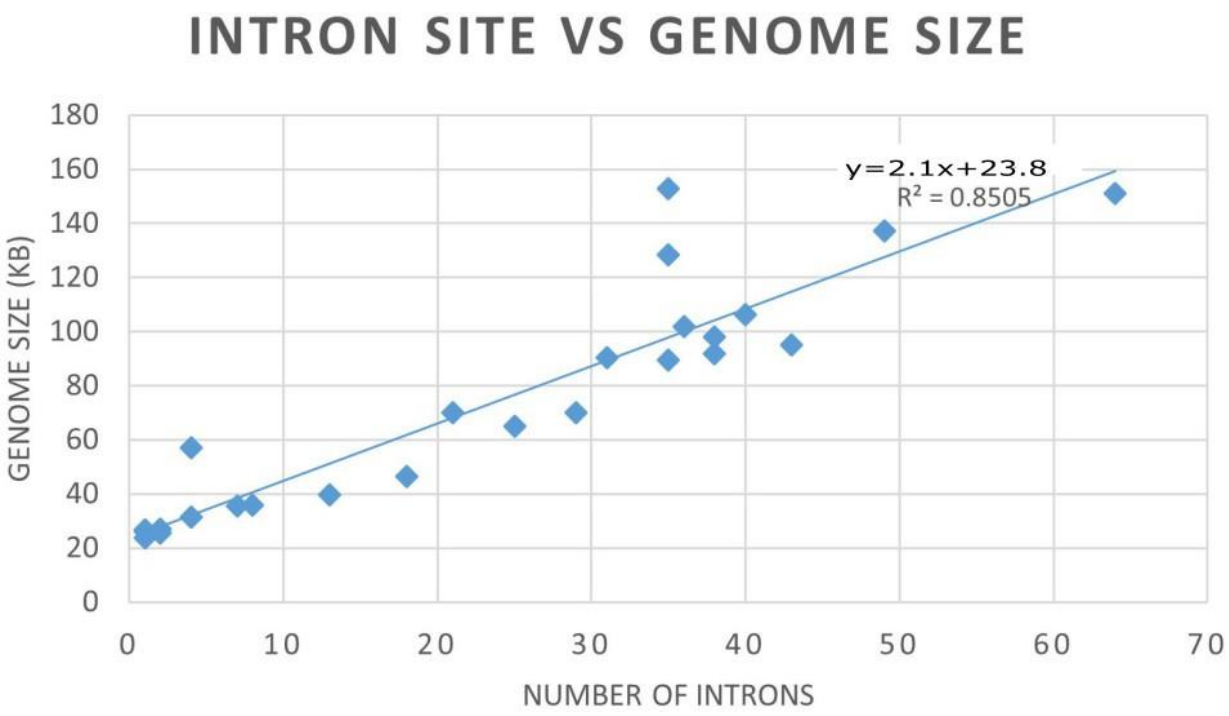

INTRON SITE VS GENOME SIZE
GENOME SIZE (KB)
NUMBER OF INTRONS
y=2.1x+23.8
R² = 0.8505
0
20
40
60
80
100
120
140
160
180
0
10
20
30
40
50
60
70

Figure 2.6: Graph depicting the relationship between mitogenome sizes and intron numbers per mitogenome for the examined members of the Ophiostomatales.

2.4.4 Phylogeny vs. Mitogenome Size and Intron Numbers

Mitogenome sizes and intron content are quite variable among the examined members of the Ophiostomatales and do not necessarily correspond do the phylogenetic position of the species examined. Genome sizes do correspond to the number of introns they contain with smaller genomes containing few introns and larger mitogenomes being intron-rich (Figures 2.5, 2.6). Members of the genus *Sporothrix sensu stricto* appear to have smaller mitogenomes ranging from 26.1 to 35.9 kb, whereas members of *Raffaelea sensu stricto* have mitogenomes ranging from 23.8 to 137 kb. Mitogenome sizes for the clade that includes *Esteya*, *Leptographium*, members of *Raffaelea sensu lato*, and *Grosmannia* range from 46.5 to >150 kb. This clade also includes the two largest mitogenomes, *Grosmannia penicillata* and *Raffaelea quercivora*, both >150 kb. Members of *Ophiostoma sensu stricto*, which include *O. minus* and *O. piliferum*, range from 65 to 97.8 kb. Only two members of *Ceratocystiopsis* were available and their mitogenome sizes range from 39.8 to 90.3 kb, but these two examples demonstrate that sharing a recent common ancestor does not imply similar genome sizes or intron content. Intron numbers can be quite variable between or within the various clades that comprise the Ophiostomatales (Figure 2.5), with the possible exception for species belonging to *Sporothrix sensu stricto*. Plotting the intron number against the genome size for each genome shows a linear relationship with a strong (86%) correlation between intron numbers and genome sizes (Figure 2.6). Gene synteny and gene content is conserved among the Ophiostomatales, so intron content is a significant factor with regard to mitogenome size and variability (Figures 2.5, 2.6).

2.4.5 Complex and Novel Introns

In this study we found five potential novel intron arrangements that have not yet been reported for members of the Ophiostomatales (Gibb and Hausner 2005; Sethuraman et al. 2009; Mullineux et al. 2010; Rudski and Hausner 2012; Hafez et al. 2013). In *O. minus*, the first intron in the *atp6* gene (*atp6* i1 at position 81) encodes a double-motif LAGLIDADG ORF followed by a tRNA for I and this intron contains a segment at its 5′ end that is a partial duplication of the downstream exon. The *cox1*–281 intron in *Leptographium lundbergii* (LDEF01000080.1) and in *Raffaelea albimanens* (PCDJ01000011.1) is composed of a group IB type intron that encodes a double-motif LAGLIDADG is interrupted by a group IC2 intron module encoding a double-motif LAGLIDADG. The *cox2*–657 intron in *Raffaelea quercus-mongolicae* (NIPS01000008.1) appears to be composed of two group IC1 intron modules. Based on comparative analysis the original "resident" IC1 intron encodes a GIY-YIG type ORF in the P9 loop and near the N-terminal coding region of this ORF an IC1 intron has been inserted. More complex or nested introns were observed in the *cob* (cytochrome b; *cob* i4) and *cox3* (cytochrome 3; *cox*3 i2) genes in *Ophiostoma ips* (NTMB01000349.1) and these are described below in more detail. Plausible RNA folds for the *O. ips* complex introns are presented in Figures 2.7, 2.8.

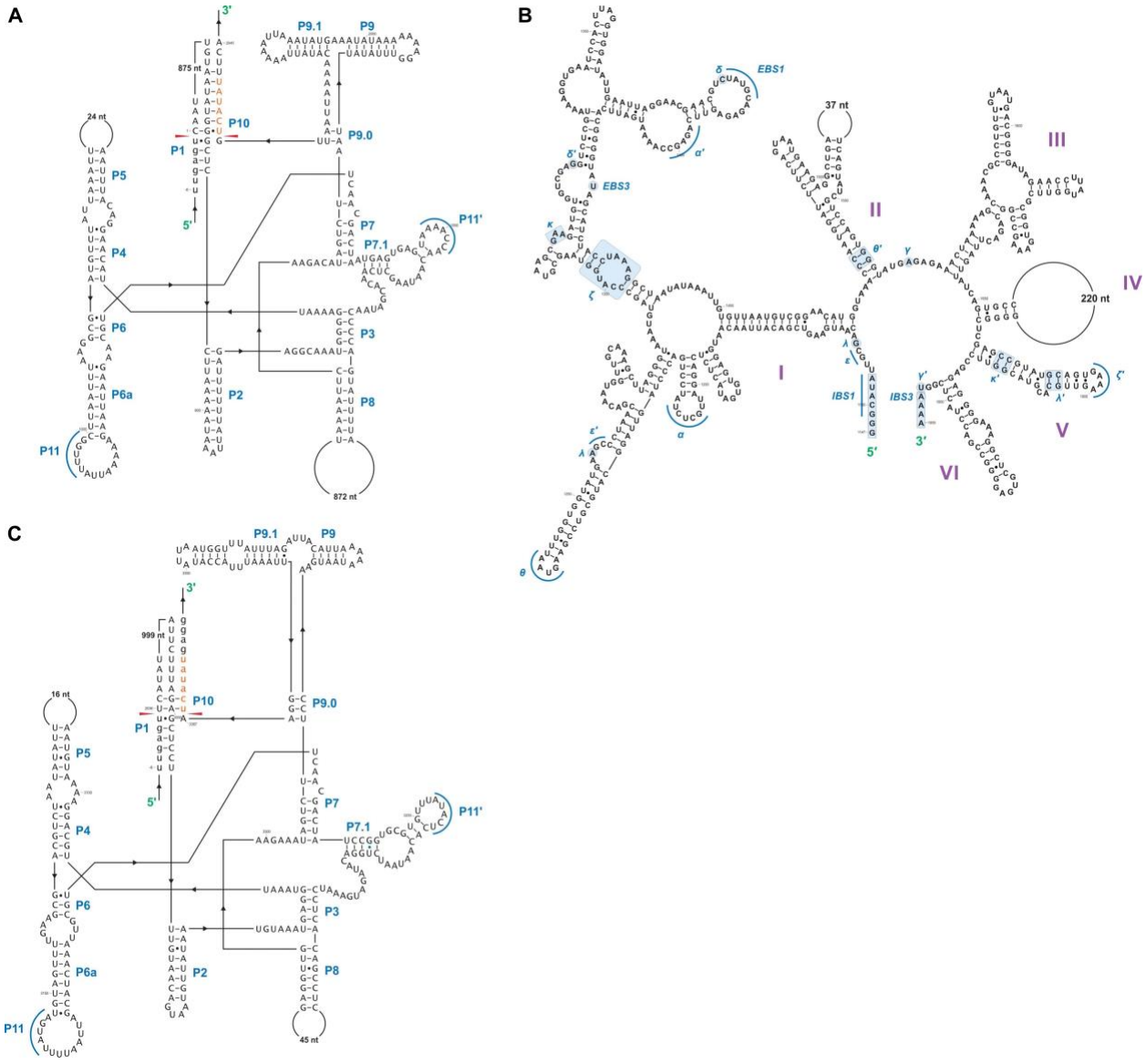

A
3'
P9.1
P9
875 nt
P10
P1
P9.0
24 nt
P5
P7
P11'
P4
P7.1
P6
P3
P6a
P2
P8
P11
872 nt
5'
B
δ
EBS1
δ'
EBS3
37 nt
II
III
κ
θ'
γ
IV
220 nt
ζ
I
λ
ε
IBS1
IBS3
V
α
5'
3'
VI
ε'
λ'
κ'
ζ'
γ'
θ
C
P9.1
P9
3'
999 nt
P10
P1
P9.0
16 nt
P5
P7
P11'
P4
P7.1
P6
P3
P6a
P2
P8
P11
45 nt
5'

Figure 2.7: Predicted *cob* I4 (*cob*-490) RNA fold composed of three intron modules **(A–C)**. Pairing regions [for group I introns: P1–P11; and domains (D) I–VI for the group II intron] are labeled by purple text; tertiary interactions are shown by blue lines. Exon and intron sequences are represented by lowercase and uppercase letters, respectively. Red arrows indicate intron-exon/pseudoexon boundaries. Orange subsequence in uppercase letters is exon-mimicking (pseudoexon) sequence, which is annotated as within *cob* I4's downstream group IA1 intron component (*cob* I4-C). Orange subsequence in lowercase letters is annotated as downstream exon sequence (*cob*-EB). **(A)** *cob*I4's upstream group IA1 intron (*cob* I4-A) RNA secondary structure model. **(B)** *cob* I4-B RNA secondary structure model. IBS, intron binding sequence; EBS, exon binding sequence. Helical domains I–VI branching from a central linker sequence ("six fingered hand") shown. Potential tertiary interactions (Greek letters) are indicated. **(C)** *cob* I4's downstream group IA1 intron (cob I4-C) RNA secondary structure model.

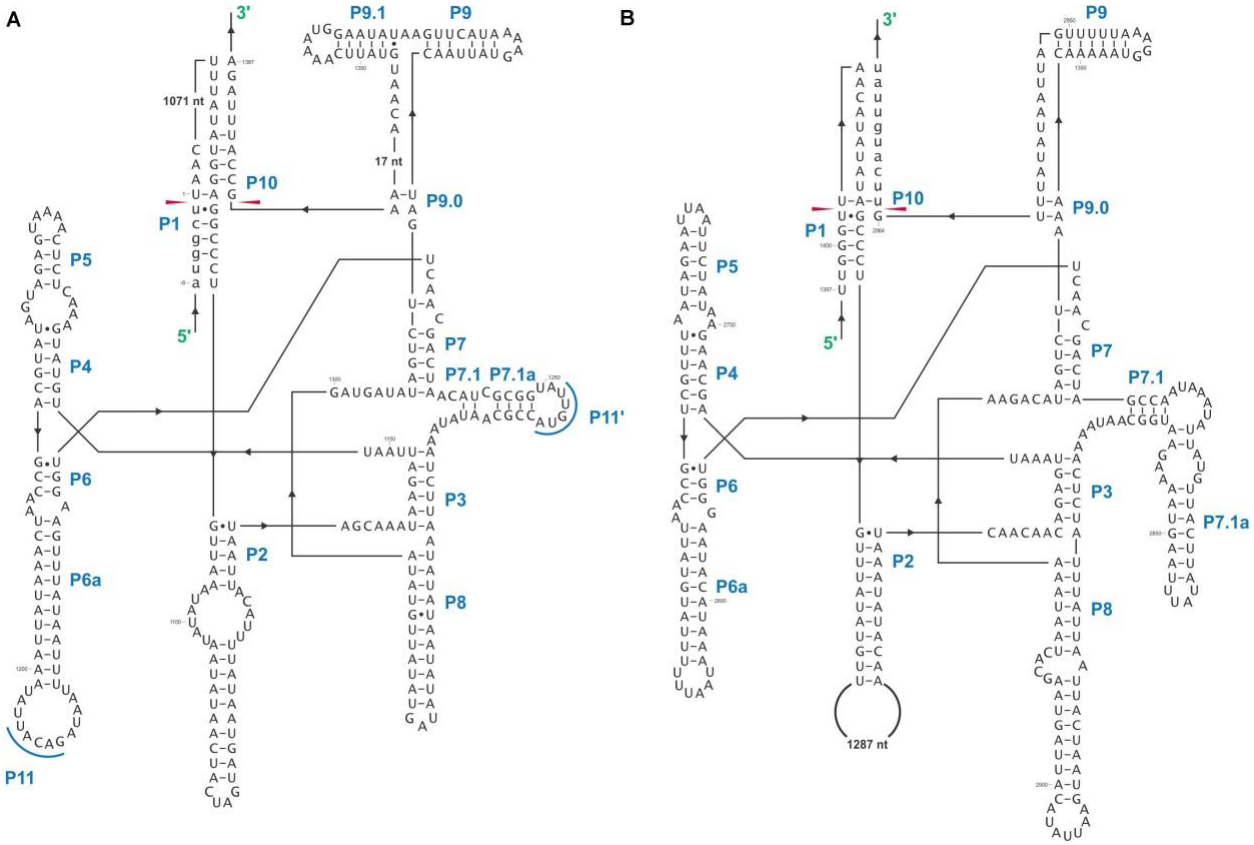

A
B
3'
5'
P1
P2
P3
P4
P5
P6
P6a
P7
P7.1
P7.1a
P8
P9
P9.1
P9.0
P10
P11
P11'
1071 nt
17 nt
1287 nt

Figure 2.8: Predicted *cox3* I2 (*cox3*–640) RNA secondary structure model composed of two intron modules. Pairing regions (P1–P11) are labeled by blue text; tertiary interactions are shown by blue lines. Exon and intron sequences are represented by lowercase and uppercase letters, respectively. Red arrows indicate intron-exon/pseudoexon boundaries. **(A)** *cox3* I2-A RNA secondary structure model. **(B)** *cox3* I2-B RNA secondary structure model.

A schematic overview of the *O. ips cob* i4 intron inserted at position 490 (relative to the *S. cerevisiae cob* coding sequence; GenBank accession number: KP263414.1) is shown in Figure 2.9. This complex intron consists of three distinct modules that contain all the necessary components for splicing. The three modules are a group I intron that is interrupted by a group II intron module and this composite element is inserted within the P1 loop of a group I intron module (presumable the resident intron). The group II intron appears to be ORF-less and is located within the P8 loop of the host group I intron module. The group I intron components contain ORFs that encode double motif LAGLIDADG type homing endonucleases. There is a short sequence separating the two group I intron modules. This so called "inter-intron module sequence" could be used as a "pseudoexon" by the internal group I intron component for the formation of the P10 helix or for the resident intron module for its P1 formation (Figures 2.9, 2.10). "Pseudoexon" is a term to describe intronic sequences that might be utilized during splicing by serving as "temporary exon" sequences; ultimately "pseudoexon" sequences are assumed to be removed when all intron components have been spliced out. The P1 and P10 helices are essential in aligning sequences that are to be spliced out or spliced together. The resident intron module can also form a P10 interaction with the downstream exon; this would allow the entire complex intron to splice out as one unit. The two group I intron modules belong to the same subtype (IA) and therefore the possibility exists that at the RNA level the two intron module components (P1–P9) can interact with each other in various combinations that may allow for various splicing pathways.

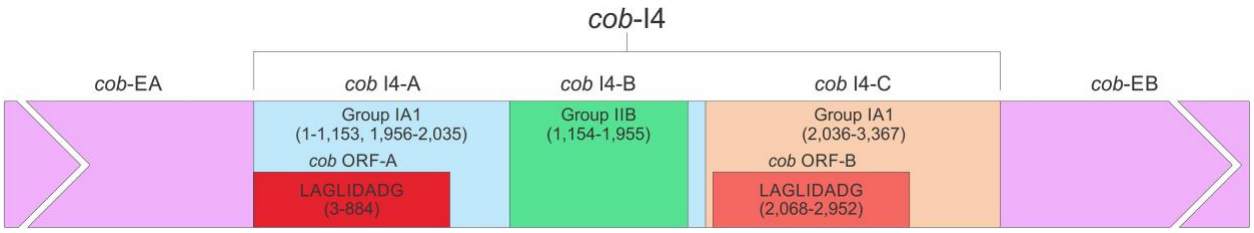

cob-I4
cob-EA
cob I4-A
cob I4-B
cob I4-C
cob-EB
Group IA1
(1-1,153, 1,956-2,035)
cob ORF-A
LAGLIDADG
(3-884)
Group IIB
(1,154-1,955)
Group IA1
(2,036-3,367)
cob ORF-B
LAGLIDADG
(2,068-2,952)

Figure 2.9: *cob*- 490 intron schematic diagram. *cob* i4, the entire complex intron at *cob* 490 position. *cob* i4-A, cob i4's upstream group IA intron; *cob*-EA, upstream exon; *cob* ORF-A, *cob*i4-A's ORF; *cob* i4-B, *cob*i4's middle group IIB intron; *cob* i4-C, *cob* i4's downstream group IA intron; *cob* ORF B, *cob* i4-C's ORF; *cob*-EB, downstream exon. LAGLIDADG represents type of homing endonuclease ORF encoded by group I intron. The numbers in brackets represent the position and length of each intron element relative to the start of *cob*i4. As the two group I introns are of the same subtype, their interactions can be interchangeable; components of the internal members can interact with components of the external member to form paired regions.

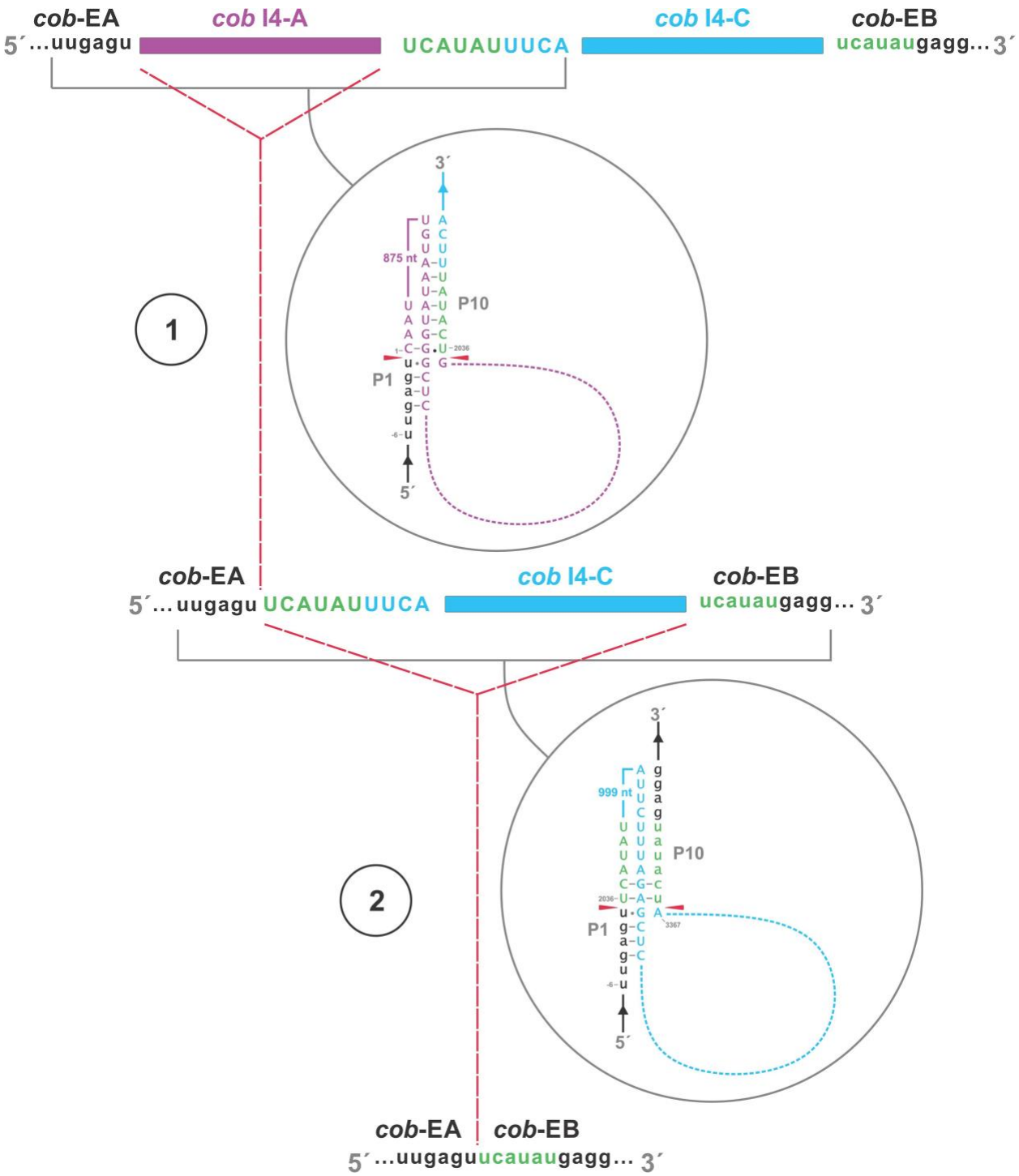

cob-EA
cob I4-A
cob I4-C
cob-EB
5´...uugagu
UCAUAUUUCA
ucauaugagg...3´
3´
875 nt
P10
P1
5´
1
cob-EA
cob I4-C
cob-EB
5´...uugagu UCAUAUUUCA
ucauaugagg...3´
3´
999 nt
P10
P1
5´
2
cob-EA
cob-EB
5´...uugaguucauaugagg...3´

Figure 2.10: Proposed RNA "ratchet-like" splicing model for *cob* I4 (*cob*-490). As the group I intron modules are both 1A types they have similar sequence elements that allow for the formation of the helical regions between the two modules (P1–P9), thus functionally they can act like tandem introns, i.e., side-by-side introns. Splicing occurs via a two-step process: (1) The "upstream" intron initially splices out using a sequence, referred to as a potential pseudoexon, located between the "upstream" and "downstream" introns, and identical to the first six nucleotides of the downstream exon. (2) Subsequent splicing of "downstream" intron results in joining of the upstream and downstream exons. *cob*-EA and *cob*-EB refer to upstream and downstream exons, respectively. The pseudoexon and downstream exon are represented in green in uppercase and lowercase, respectively. *cob* I4-A (purple) and *cob* I4-C (blue) refer to the "upstream" and "downstream" introns, respectively. Proposed P1 and P10 interactions are shown in gray circles.

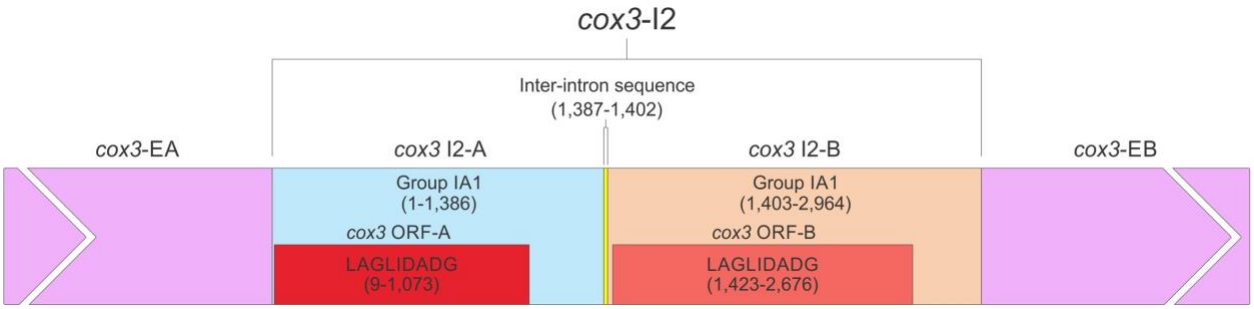

cox3-I2
Inter-intron sequence
(1,387-1,402)
cox3-EA
cox3 I2-A
cox3 I2-B
cox3-EB
Group IA1
(1-1,386)
cox3 ORF-A
LAGLIDADG
(9-1,073)
Group IA1
(1,403-2,964)
cox3 ORF-B
LAGLIDADG
(1,423-2,676)

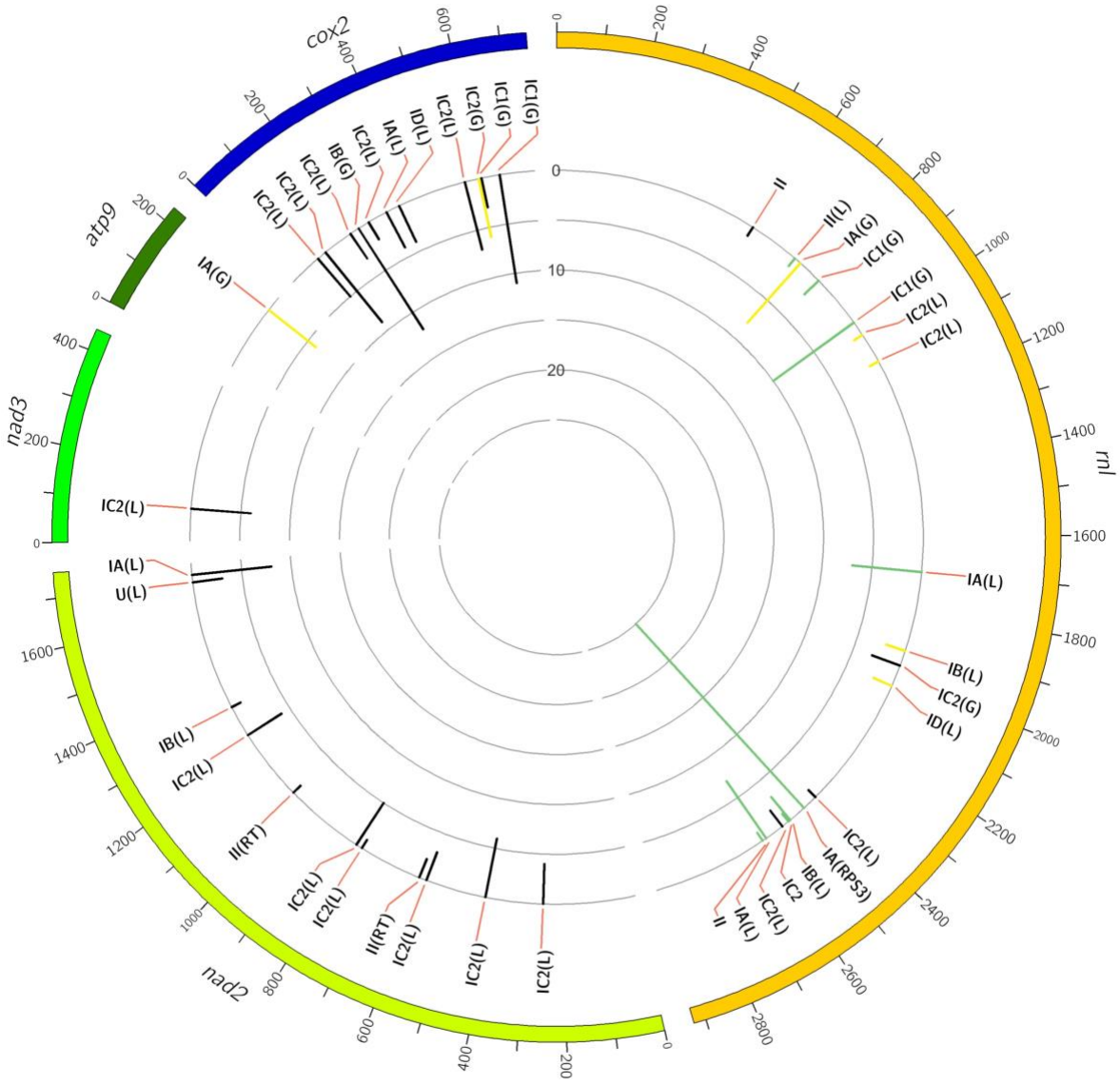
Figure 2.11: The *cox3*–640 intron schematic diagram. *cox3* i2, the entire complex intron at *cox3*–640 position. *cox3* i2-A, *cox3* i2's upstream group IA intron; *cox3*-EA, upstream exon; *cox3* ORF-A, *cox3* i2-A's ORF; *cox3* i2-B, *cox3*i2's downstream group IA intron; *cox3* ORF-B, *cox3* i2-B's ORF; *cox3*-EB, downstream exon. Inter-intron sequence: sequence separating *cox3* i2-A and *cox3* i2-B. LAGLIDADG designation represents the type of homing endonuclease ORF encoded by group I introns. The numbers in brackets represent the position and length of each intron element relative to the start of *cox3* i2.

The *O. ips cox3* i2 was confirmed as being *cox3*–640 relative to the *S. cerevisiae cox3* sequence (GenBank accession number: KP263414.1) with the total length of the intron being 2964 nt. The *cox3* i2 based on MFannot and RNAweasel analysis combined with BLASTn analysis was noted to be composed of two group I intron modules in a tandem arrangement (Deng et al. 2016): an upstream group IA1 intron module (*cox3* i2-A) corresponding to *cox3* i2 nucleotides 1–1386, and a downstream group IA1 intron module (*cox3* I2-B) corresponding to *cox3* i2 nucleotides 1403–2964 (see *cox*3 intron schematic) (see Figure 2.11). This complex intron also appeared to contain a sequence separating *cox3* i2-A and *cox3* i2-B intron modules, referred to as the inter-intron module sequence. The inter-intron module sequence was annotated as corresponding to *cox3* i2 nucleotides 1387–1402, and BLASTn did not show any related or similar sequences in GenBank. It does not appear to be part of either intron module; in addition, careful examination failed to reveal any sequences that could be utilized to form a suitable P10 interaction for the upstream intron module. That implies that the upstream intron module utilizes the downstream exon for the formation of the P10 interaction, which results in the splicing of the entire composite intron element. Similar tandem intron arrangements for the *cox3*–640 intron were observed in *Ceratocystiopsis brevicomis* (PCDN01000199.1) and *Grosmannia penicillata* (PCDK01000036.1).

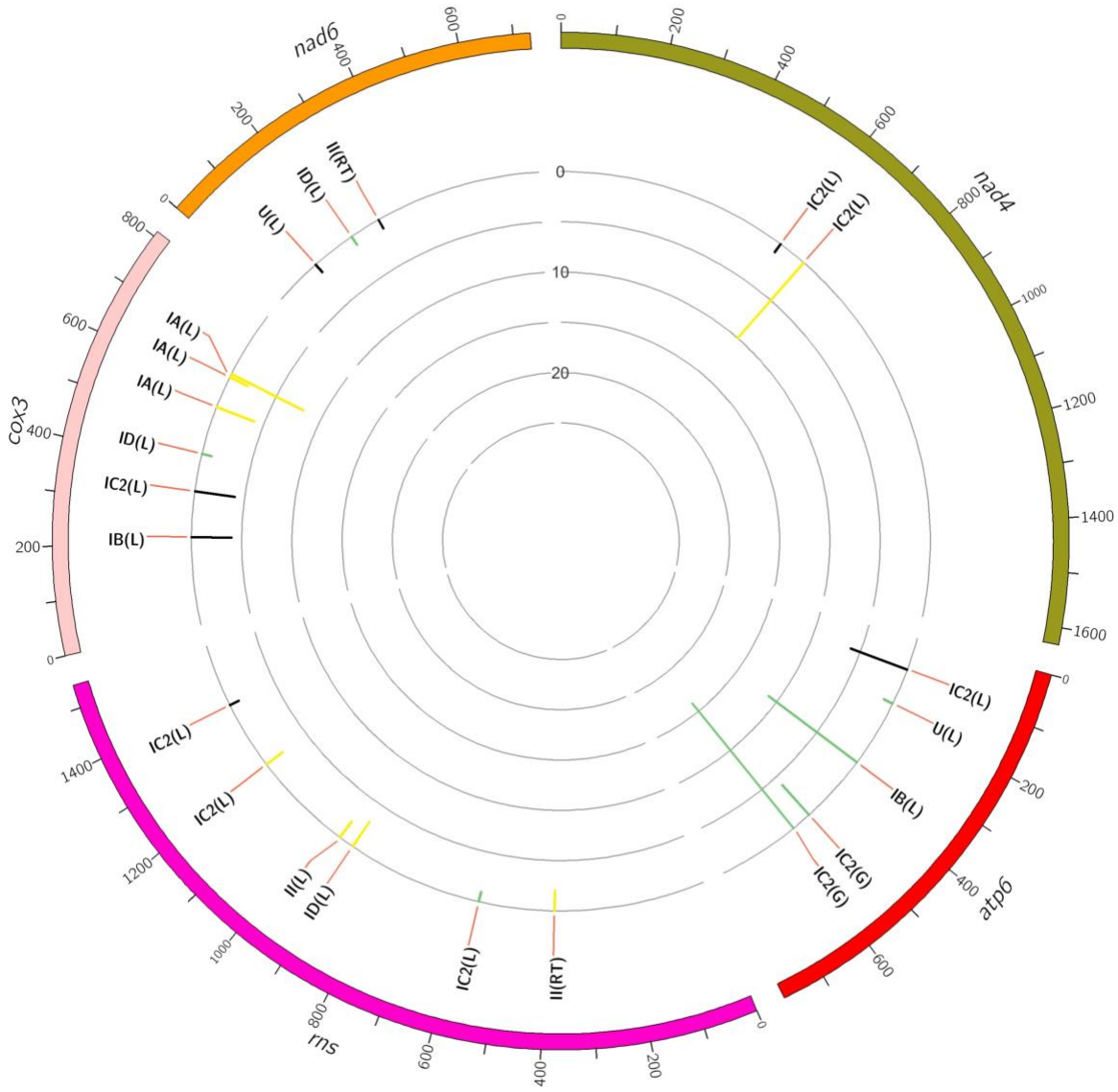
(A)


nad4L
cob
nad5
IC1(L)
IB(L)
ID(L)
IC2(L)
ID(L)
IB(L)
ID(L)
IB(L)
IB(G)
IC2(L)
IA(L)
IB(L)
IA(L)
IB(L)
ID(L)
IA
ID(G)
IB(L)
IB(L)
0
10
20

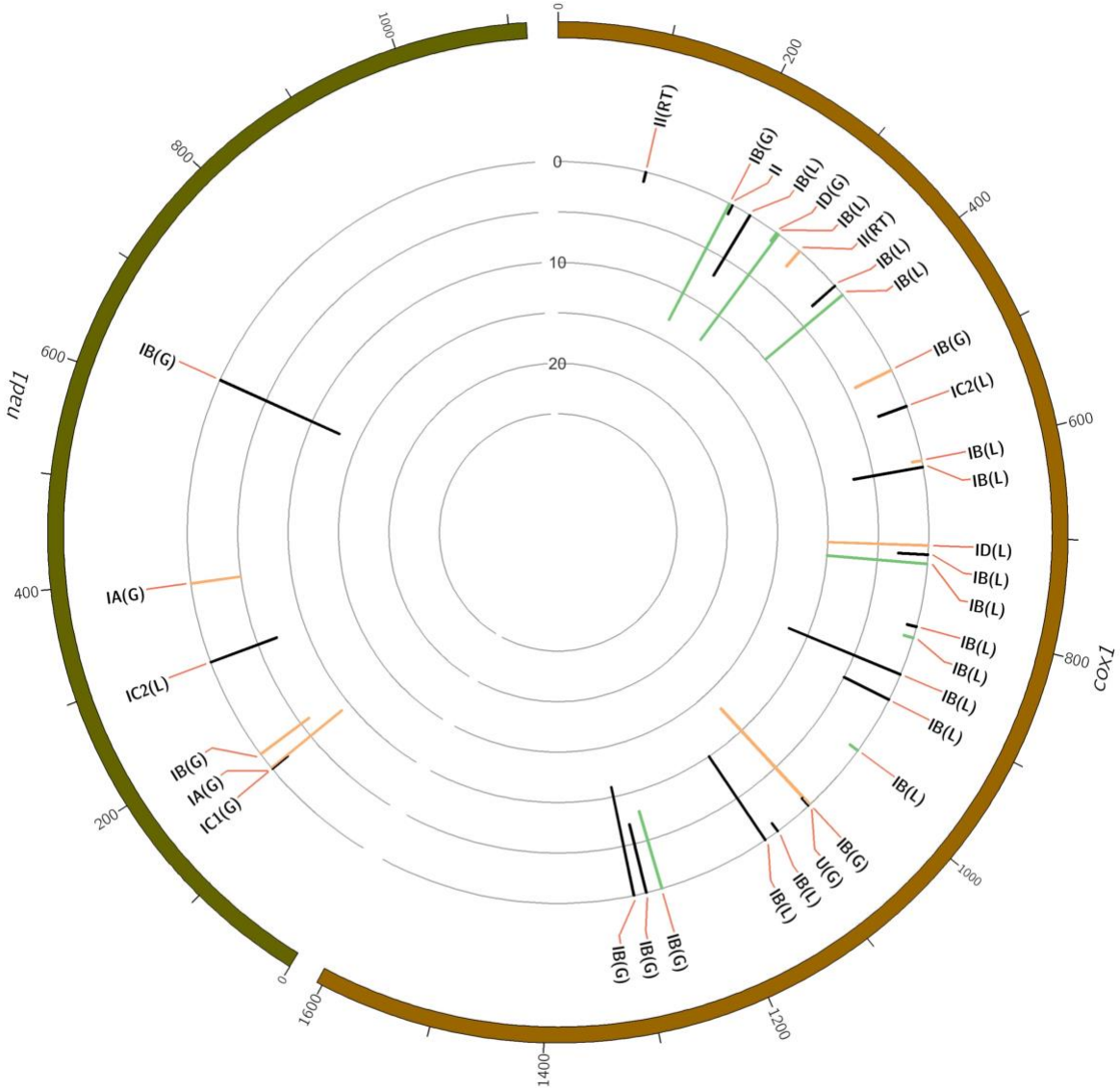
(B)


cox2
atp9
nad3
nad2
rnl
0
10
20
IC1(G)
IC2(G)
ID(L)
IC2(L)
IA(L)
IB(G)
IA(G)
II
II(L)
IA(G)
IC1(G)
IC1(G)
IC2(L)
IC2(L)
IC2(L)
IA(L)
U(L)
IA(L)
IB(L)
IC2(G)
ID(L)
IB(L)
IC2(L)
II(RT)
IC2(L)
IC2(L)
II(RT)
IC2(L)
IC2(L)
IC2(L)
IC2(L)
IA(RPS3)
IB(L)
IC2
IC2(L)
IA(L)
II

(C)


nad6
nad4
atp6
rns
COX3
0
10
20
II(RT)
ID(L)
U(L)
IC2(L)
IA(L)
IA(L)
IA(L)
ID(L)
IC2(L)
IB(L)
IC2(L)
U(L)
IB(L)
IC2(G)
IC2(G)
IC2(L)
IC2(L)
II(L)
ID(L)
IC2(L)
II(RT)

(D)


0
200
400
600
800
1000
1200
1400
1600
1800
COX1
tRNA-Leu
0
10
20
II(RT)
IB(G)
II
IB(L)
ID(G)
IB(L)
II(RT)
IB(L)
IB(L)
IB(G)
IC2(L)
IB(L)
IB(L)
ID(L)
IB(L)
IB(L)
IB(L)
IB(L)
IB(L)
IB(L)
IB(G)
U(G)
IB(L)
IB(L)
IB(G)
IB(G)
IB(G)
IB(G)
IA(G)
IC1(G)
IC2(L)
IA(G)
IB(G)

Supplementary Figure 2.1: Intron landscapes visualized with Circos generated diagrams. An overview of the intron landscape is illustrated in Figure 2.4 in **(A–D)** more detailed maps are presented maintaining the gene order as shown in Figure 2.4. Each panel shows one segment of the mitogenome: (A) cox1, nad1; (B) nad4, atp6, rns, cox3; (C) rnl, nad2, nad3, atp9, cox2; (D) nad4L, nad5, cob. Details were added with regards to intron types (subtypes), types of intron-encoded proteins (L, LAGLIDADG type ORFs; G, GIY-YIG type ORFs), and intron phasing [phase 0 = black lines, phase 1 = orange lines, and phase 2 = green lines, no phasing (for RNA gene) = purple line].

Supplementary Table 2.1: Comparison of the mitochondrial genomes and their intron complement for the studied members of the Ophiostomatales.

Name of the Organism	NCBI Accession	Genome Size	Number of Introns in Genes																Total Introns by Species
			atp6	*atp8*	*atp9*	*cob*	*cox1*	*cox2*	*cox3*	*nad1*	*nad2*	*nad3*	*nad4*	*nad4L*	*nad5*	*nad6*	*rnl*	*rns*	
Ceratocystiopsis brevicomis	PCDN01000199.1	90,376	3	0	1	5	10	5	1	3	3	0	1	1	2	0	3	0	38
Ceratocystiopsis minuta	LZPB01000172.1	39,800	1	0	1	1	6	0	0	1	0	0	0	1	1	0	1	0	13
Esteya vermicola	KY644696.1	46,507	1	0	0	3	4	2	3	1	0	0	1	0	1	0	2	0	18
Fragosphaeria purpurea	PCDL01000017.1	57,056	0	0	0	1	3	0	0	0	0	0	0	0	0	0	1	0	5
Graphilbum fragrans	LLKO01000061.1	25,567	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	2
Grosmannia penicillata	PCDK01000036.1	150,891	4	0	1	7	15	8	4	4	7	0	1	1	5	0	6	1	64
Hawksworthiomyces lignivorus	NTMA01000166.1	27,092	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	2
Leptographium lundbergii	LDEF01000080.1	101,879	2	0	0	1	13	5	2	3	2	1	0	0	2	0	3	2	36
Ophiostoma ips	NTMB01000349.1	97,849	3	N/A[a]	0	5	13	4	2	4	3	1	1	0	1	0	4	1	42
Ophiostoma minus [WIN(M)495]	MW122509.1	91,847	3	0	N/A[b]	1	9	7	2	3	3	0	0	0	3	0	5	2	38
Ophiostoma novo-ulmi	MG020143.1	65,095	2	0	N/A[b]	3	8	5	0	1	1	0	0	0	1	0	4	0	25
Ophiostoma piliferum [WIN(M)959]	MW122508.1	69,966	2	0	N/A[b]	6	9	4	0	3	1	0	0	1	2	0	1	0	29
Raffaelea albimanens	PCDJ01000011.1	137,049	2	0	1	6	8	4	3	2	5	0	1	1	9	1	6	0	49
Raffaelea ambrosiae	PCDI01000067.1	89,461	3	0	0	5	8	3	1	3	1	0	1	0	3	0	6	1	35
Raffaelea arxii	PCDH01000124.1	95,042	1	0	1	4	15	3	2	2	3	1	2	0	4	0	5	0	43
Raffaelea lauricola	PCDG01000206.1	70,090	2	0	0	2	6	3	0	3	1	0	0	1	0	0	3	0	21
Raffaelea quercivora	PCDE01000018.1	152,890	3	0	0	4	6	3	0	2	3	0	0	1	5	0	7	1	35
Raffaelea quercus-mongolicae	NIPS01000008.1	106,194	3	0	0	2	12	5	2	3	4	1	1	1	3	0	3	0	40
Raffaelea sp.	PCDF01000414.1	23,830	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Raffaelea sulphurea	PCDD01000156.1	128,408	2	0	1	3	6	4	1	3	1	0	1	0	0	2	9	3	36
Sporothrix brasiliensis	AWTV01000012.1	35,826	1	0	0	0	4	0	0	1	0	1	0	0	0	0	2	0	9
Sporothrix globosa	LVYW01000010.1	26,671	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Sporothrix insectorum	MK482392.1	31,500	0	0	0	0	3	0	0	0	0	0	0	0	0	0	1	0	4
Sporothrix pallida	CM003773.1	35,458	0	0	0	2	2	1	0	0	0	0	1	0	0	0	1	0	7
Sporothrix schenckii	AB568600.1	26,095	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Total Introns by Genes			38	0	6	62	161	65	23	42	38	6	11	8	42	3	78	11	594

[a] Mitochondrial scaffold incomplete and segment with *atp8* is missing.

[b] *atp9* is absent (as previously observed in *Ophiostoma novo-ulmi* (Abboud et al., 2018).

Supplementary Table 2.2: Intron subtypes and intron open reading frames (ORF) recorded within the mitogenomes of the examined members of the Ophiostomatales.

Gene	Position[a]	Phase	Subgroup[b]	Count[c]	ORF[d]	ORF count[e]
atp6	521	2	IC2	4	GIY	
	572	2	IC2	16	GIY	
atp8	N/A	N/A	N/A	N/A	N/A	N/A
atp9	187	1	IA	6	GIY	
cob	155	2	IB	5	LAG	
	201	0	IB	9	LAG	
	393	0	ID	14	GIY	
	406	1	IA	1	-	
	429	0	ID	2	LAG	
	437	2	IB	4	LAG	
	490	1	IA	12	LAG	
	506	2	IB	6	LAG	
	562	1	IA	2	LAG	
	820	1	IB	7	LAG	
cox1	105	0	II	1	RT	1
	212	2	IB	13	GIY	
	216	0	II	1	-	0

240	0	IB	7	LAG	
278	2	GIY	1	GIY	
281	2	IB	13	LAG	
313	1	II	2	RT	2
372	0	IB	3	LAG	
386	2	IB	10	LAG	
493	1	IB	4	GIY	
540	0	IC2	3	LAG	
607	1	IB	1	LAG	
615	0	IB	7	LAG	
709	1	ID	10	LAG	
720	0	IB	3	LAG	
731	2	IB	10	LAG	
807	0	IB	1	LAG	
821	2	IB	1	LAG	
867	0	IB	12	LAG	
900	0	IB	5	LAG	
971	2	IB	1	LAG	
1057	1	IB	13	GIY	
1059	0	U	1	GIY	
1107	0	IB	1	LAG	
1125	0	IB	10	LAG	

	1262	2	IB	8	GIY
	1281	0	IB	7	GIY
	1296	0	IB	11	GIY
cox2	93	0	IC2	5	LAG
	120	0	IC2	9	LAG
	207	0	IC2	3	LAG
	234	0	IB	12	GIY
	267	0	IC2	2	LAG
	324	0	IA	4	LAG
	363	0	ID	4	LAG
	558	0	IC2	7	GIY
	598	1	IC2	6	GIY
	606	0	IC1	3	GIY
	657	0	IC1	11	GIY
cox3	219	0	IB	4	LAG
	333	0	IC2	4	LAG
	428	2	ID	1	LAG
	550	1	IA	4	LAG
	631	1	IA	2	LAG
	640	1	IA	8	LAG
nad1	144	0	IC1	2	GIY
	145	1	IA	9	GIY

	166	1	IB	6	GIY	
	291	0	IC2	7	LAG	
	388	1	IA	5	GIY	
	636	0	IB	13	GIY	
nad2	258	0	IC2	4	LAG	
	420	0	IC2	6	LAG	
	591	0	IC2	3	LAG	
	612	0	II	2	RT	2
	792	0	IC2	1	LAG	
	810	0	IC2	5	LAG	
	1038	0	II	1	RT	1
	1242	0	IC2	4	LAG	
	1332	0	IB	1	LAG	
	1698	0	U	3	LAG	
	1719	0	IA	8	LAG	
nad3	90	0	IC2	6	LAG	
nad4	585	0	IC2	1	LAG	
	658	1	IC2	10	LAG	
nad4L	239	2	IC1	8	LAG	
nad5	248	2	ID	9	LAG	
	324	0	IC2	10	LAG	
	426	0	ID	3	LAG	

	570	0	IB	6	LAG	
	710	2	ID	3	LAG	
	717	0	IB	5	LAG	
	924	0	IB	2	LAG	
	1000	1	IB	1	GIY	
	1152	0	IC2	3	LAG	
nad6	120	0	U	1	LAG	
	233	2	ID	1	LAG	
	312	0	II	1	RT	1
rnl	576	N/A	II	1	-	0
	722	N/A	II	1	LAG	1
	742	N/A	IA/IB	8	GIY	
	812	N/A	IC1	2	GIY	
	965	N/A	IC1	10	GIY	
	1006	N/A	IC2	1	LAG	
	1096	N/A	IC2	1	LAG	
	1700	N/A	IA	7	LAG	
	1924	N/A	IB	2	LAG	
	1968	N/A	IC2	3	GIY	
	2029	N/A	ID	2	LAG	
	2406	N/A	IC2	1	LAG	
	2450	N/A	IA	25	RPS3	

	2501	N/A	IB	3	LAG	
	2507	N/A	IC2	1	-	
	2529	N/A	IC2	2	LAG	
	2585	N/A	IA	7	LAG	
	2597	N/A	II	1	-	0
rns	379	N/A	II	2	RT	2
	569	N/A	IC2	1	LAG	
	913	N/A	ID	3	LAG	
	952	N/A	II	2	LAG	2
	1210	N/A	IC2	2	LAG	
	1383	N/A	IC2	1	LAG	
Total	119	95	N/A	594	N/A	12

[a] Position of intron insertion (insertion site) with respect to a reference sequence without introns; see Materials and Methods for more details

[b] Refers to most observed intron subgroup within the insertion site and was based mainly on predictions by MFannot/RNAweasel and manually identified for those missed by the programs; U = undefined (i.e. no intron subgroup predicted by MFannot/RNAweasel and manually identified); more than one intron subgroup can be associated with one insertion site

[c] Refers to the number of fungi containing the insertion; some insertion sites may encode for more than one intron, where "intron" refers to (at least) the presence of elements composing the intron core

[d] Refers to the most observed Open Reading Frame (ORF) encoded within the intron insertion site; LAG = LAGLIDADG homing endonuclease, GIY = GIY-YIG homing endonuclease, RT = reverse transcriptase, - = no ORF containing identifiable (based on BLASTx) motif

[e] ORF count was restricted to only ORFs encoded within group II introns where ORFs could be unambiguously identified; several cases where a group I intron encoded multiple ORFs in various configurations (ex. one followed by another, one within another)

N/A = not applicable

2.5 Discussion

2.5.1 Mitogenomes of the Ophiostomatales

Fungal mitogenomes are usually represented as circular molecules and reported to range in size from 12.055 to >500 kb (James et al. 2013; Zubaer et al. 2018; Liu et al. 2020a; Liu et al. 2020b). Linear versions have also been observed, and some circular version could actually exist as linear concatemers generated by a rolling circle type DNA replication mechanism (Bendich 1993; Hausner 2003; Hausner 2012; Bullerwell and Lang 2005; Baidyaroy et al. 2011; Valach et al. 2011; Chen and Clark-Walker 2018). Mitochondrial genome architecture and size are highly variable among the fungi due to 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) (Aguileta et al. 2014; Wu and Hao 2014; Wu and Hao 2019; Franco et al. 2017; Repar and Warnecke 2017; Deng et al. 2018; Stone et al. 2018; Zubaer et al. 2018; Kolesnikova et al. 2019; Kulik et al. 2020; Liu et al. 2020b). Like other fungi, most of the mitogenome variation observed among the examined members of the Ophiostomatales is due to the absence and presence of introns.

Gene order is conserved among the Ophiostomatales, and some minor variation was observed with regards to the tRNA gene set and the absence or presence of the *atp9* gene. It has been suggested that the loss of the mtDNA-encoded *atp9*gene can be compensated for by the presence of a nuclear-encoded version of this gene (Kanzi et al. 2016; Sellem et al. 2016; Franco et al. 2017; Zubaer et al. 2018).

Among the Ophiostomatales the number of mitochondrial introns and intron insertion sites is variable, except for mL2450. The mL2450 group IA intron encodes *rps3*, a configuration observed among many filamentous members of the Ascomycota. However, in some members of the Ascomycota *rps3* is a free-standing gene or missing from the mitochondrial genome (Korovesi et al. 2018; Wai et al. 2019). For the latter, such as in some members of the Capnodiales, (Wai et al. 2019) identified a nuclear-encoded ortholog of the mitochondrial-encoded version of *rps3*. Among the metazoans and the fungi, the nuclear and mitochondrial versions of *rps3* have been shown to be required for initiating protein translation and appear to have "moonlighting activities" by being involved in DNA repair, cell signaling, apoptosis, and potentially in gene regulation (Neu et al. 1998; Kim et al. 2013; Wang et al. 2019; Seshadri et al. 2020). Maintaining the *rps3* gene within the *rnl* group I intron might be a fortuitous association that provides the *rps3* gene a locus embedded within an essential gene (*rnl*) presumably transcribed at a high rate, and this may "protect" the intron from being eroded by drift.

2.5.2 Phylogenetic Analysis of Mitogenomes and Taxonomic Implications

The taxonomy of the Ophiostomatales has undergone considerable revisions in recent years; currently the Order Ophiostomatales includes two Families, Kathistaceae and the Ophiostomataceae. The latter includes the following genera: *Aureovirgo*, *Ceratocystiopsis*, *Fragosphaeria*, *Graphilbum, Hawksworthiomyces*, *Raffaelea sensu stricto*, *Ophiostoma sensu stricto*, and *Sporothrix sensu stricto* (Hyde et al., 2020). In addition, there are groupings for which monophyly is not certain such as *Leptographium sensu lato* (includes *Grosmannia* species, the *Raffaelea sulphurea* species complex, and *Esteya vermicola*) and *Ophiostoma sensu lato* (de Beer and Wingfield 2013; de Beer et al. 2013a; de

Beer et al. 2013b; de Beer et al. 2016a; de Beer et al. 2016b); the latter includes several smaller lineages with uncertain taxonomic positions (referred to as lineages A, B, C, and D; de Beer et al. 2016b).

The concatenated mitochondrial protein sequence data set confirmed the monophyly of the Ophiostomatales, but it shows unresolved relationships among members that belong to *Leptographium*, *Raffaelea*, *Sporothrix*, and *Grosmannia*, suggesting that future work should include more members of *Leptographium, Grosmannia, Graphilbum, Hawksworthiomyces, Fragosphaeria*, and *Aureovirgo* in order to resolve phylogenetic relationships among the Ophiostomataceae. The original concept of *Sporothrix* was viewed to be a polyphyletic grouping, and this lead to the designation of the Genus *Hawksworthiomyces* with *Hawksworthiomyces lignivorus* (formerly *Sporothrix lignivora*) as the type species (De Beer et al. 2016a). *Sporothrix insectorum* was shown (Zhang et al. 2019) not to group with species of *Sporothrix sensu stricto*, and the current analysis aligns this species with *Hawksworthiomyces lignivorus*.

The analysis confirmed that the circumscription of *Raffaelea sensu lato* does not define a natural grouping (De Beer and Wingfield 2013) with *Raffaelea quercivora*, *R. quercus-mongolicae*, and *R. sulphurea* possibly awaiting the designation of a new genus that can accommodate these ambrosia fungi. The placement of *Fragosphaeria purpurea* is unexpected as it groups with *Sporothrix insectorum* and *Hawksworthiomyces lignivorus*. This finding needs to be confirmed as previous studies based on nuclear markers placing *Fragosphaeria purpurea* very distantly from species of *Hawksworthiomyces* (de Beer et al. 2016b). The internal transcribed rDNA spacer sequences (ITS) recovered from the same genomic data set from which

the *Fragosphaeria purpurea* mitogenome was derived matches (at 100% identity) ITS data available in GenBank for *F. purpurea* (AB278192.1 -mitogenome; and MN511357.1 for ITS). The ITS data for the *Sporothrix insectorum* (MK482392.1) strain included in this mitogenome survey shares 99.47% identity with those available for *F. purpurea*; however, a strain of *Sporothrix insectorum* (CBS 756.73; MH860798.1 – ITS) sequenced by Vu et al. (2019) deviates from the above (MK482392.1), suggesting the species concept of *Sporothrix insectorum*needs to be re-evaluated.

2.5.3 The Mitogenome Intron Complement

Like a previous study (Zubaer et al. 2019), we observed that phase 0 introns outnumber introns positioned in phase 1 or 2. We speculated that this was due to core creep where the IEP coding region eventually extends to include the 5′ terminal intron sequence to fuse (in frame) to the upstream exon (Edgell et al. 2011). This would enhance the expression of the IEP as it would benefit from the host genes transcription and translation signals. This would entwine the intron and the HEG that may have started out as independent elements but now coevolve to maintain splicing and homing/mobility activities (Guha et al. 2018; Megarioti and Kouvelis 2020).

The sizes of the mitochondrial genomes among members of the Ophiostomatales appear to be linked to the number of introns they contain. Similar observations have been made for other fungal groups (Kanzi et al. 2016; Liang et al. 2017; Zubaer et al. 2018). Introns can be gained by events that allow for cytoplasm to be exchanged between members of the same species or different species that would allow for the fusion of mitochondrial organelles enabling intron homing events and/or recombination between the different mtDNAs. The maintenance of introns

has been assumed to be a matter of drift (neutral evolution) as a lack of selection would lead to an accumulation of mutations that could be deleterious to the intron and/or its protein coding components, leading to the eventual loss of the composite element (Goddard and Burt 1999). Survival of these elements depends on inserting into new sites or reinvading sites where the intron was lost, a continuous cycle of gain, degeneration, loss, and reinvasion. We noted that members of the genus *Sporothrix sensu stricto* that are known to be pathogens on mammals have the smallest mitogenomes. Proliferation of introns may not be compatible with the life histories of these fungi.

Some introns could be beneficial such as those that encode proteins like RPS3 or in some instances N-acetyltransferases or aminotransferases (Wai et al. 2019). The latter two examples may provide metabolic flexibility for certain fungi, providing adaptive advantages (Duò et al. 2012). Mitochondrial introns have been associated with resistance to fungicides (Cinget and Bélanger 2020) and hypovirulence (Baidyaroy et al. 2011). Introns are also a vehicle for modulating gene expression as their removal can be a rate limiting step for the expression of the genes (Rose 2019), for example, in *Saccharomyces cerevisiae* mitochondrial functioning is linked to mitochondrial intron splicing (Rudan et al. 2018). Introns could be environmental sensors (Belfort 2017), whereby splicing activity is influenced by environmental conditions and allows for gene expression to be fine-tuned to certain environmental cues.

2.5.4 Complex Introns

Complex introns are composed of several intron modules possible, the result of one mobile intron invading another intron. We characterized the *cox3* i2 and *cob*i4 introns of *O. ips* (GenBank accession number: NTMB01000349.1) using computational strategies. These complex introns were composed of two and three intron modules, respectively, and could provide a platform for alternative splicing that may optimize intron-encoded protein expression and/or modulate host gene expression. For *cox3* i2, a tandem group I intron, splicing was predicted to occur as a composite unit, and the downstream intron was presumed to be the native intron.

The *O. ips cob* i4 intron was more complex; here the resident group I intron within its P1 component houses another group I intron that has been invaded by a group II intron module. A homing endonuclease ORF was detected within each of the two group I intron modules. For the three intron modules, many possible splicing patterns could be envisioned. The group II intron could modulate the splicing and expression of the internal group I intron component of this complex intron (see Hafez et al. 2013; Guha and Hausner 2016; Guha et al. 2018) possibly having been co-opted to regulate the expression of the IEP. Plausible RNA interactions can be recognized that would allow for the group I modules to splice separately or as one composite intron. Detailed deep RNA sequencing analysis combined with RT-PCR based experiments are required to investigate if this insertion is a "zombie" intron (splices as one unit; Zumkeller et al. 2020) or a trintron where each module can splice individually or in various arrangements (isoforms) during RNA processing of the *cob* transcript. As the group I intron modules have similar sequence elements that allow for the formation of the helical regions (P1–P9)

functionally they can act like tandem introns, i.e., side-by-side introns. A plausible pathway is shown in Figure 2.10; a ratchet-like (see Hafez and Hausner 2015) mechanism could operate that removes the first intron module thereby 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 we refer 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. Similar splicing patterns demonstrating the "plasticity" of intron RNA folds have been previously observed (Sellem and Belcour 1994; Turk et al. 2013). Tandem type complex introns, such as *O. ips cox3* i2, have been observed and described in the literature (Deng et al. 2016; Zubaer et al. 2018). These configurations need closer examinations in future studies to understand their splicing pathways and impact on mitochondrial gene expression.

2.6 Conclusion

Comparing the mitogenomes of *O. minus* and *O. piliferum* with other available sequences for members of the Ophiostomatales showed that gene synteny is conserved and variability is mostly due to introns. The mitochondrial sequences show potential for resolving taxonomic questions among members of the Ophiostomatales, and as more genomes become available mitochondrial data will complement phylogenetic data based on nuclear markers. These insect-vectored fungi are potentially invasive and are a concern with regards to the biosecurity of forests; mitogenomics could provide a valuable tool in the identification and tracking of species belonging to the Ophiostomatales. Exploring fungal mitogenomes is important as some introns have the potential to serve as agents that can module gene expression and impact the phenotypes of the fungi that accommodate them (Cinget and Bélanger 2020; Medina et al. 2020). These types of introns could be considered the result of constructive neutral evolution whereby complex systems evolve by non-adaptive mechanisms (such as drift) (Stoltzfus 1999; Gray et al. 2010; Lukeš et al. 2011). In addition, ribozymes, complex introns (i.e., potentially co-operating ribozymes), and intron-encoded proteins 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).

CHAPTER 3: COMPARATIVE MITOGENOMICS OF *LEPTOGRAPHIUM WINGFIELDII* – AN INVASIVE FUNGAL SPECIES IN CANADIAN FORESTS

CHAPTER 3: COMPARATIVE MITOGENOMICS OF *LEPTOGRAPHIUM WINGFIELDII* – AN INVASIVE FUNGAL SPECIES IN CANADIAN FORESTS

3.1 ABSTRACT

Leptographium wingfieldii is a fungal associate of *Tomicus piniperda* (the pine shoot beetle) and pathogen of pines and this species is an agent of blue stain in sapwood on infected trees. This fungus was first reported from Europe and has been recently introduced to Canadian forests. Ten new mitogenomes have been sequenced and characterized, including seven strains of *L. wingfieldii*, two strains of *L. procerum* and one strain of *L. terebrantis*. The data were combined with other members of the Ophiostomatales collected from NCBI to gain more insight into the genetic diversity, evolution, and systematics of these fungi. The size of the studied mitogenomes of *Leptographium* species ranged from 41 kb to 126 kb with the number of potential mobile introns embedded within these mitogenomes ranging from 13 to 45. These data show that introns generate genetic diversity and confirms the contribution of mobile introns in genome expansion in Ophiostomatales fungi. This study also uncovered complex intron arrangements (twintrons) suggesting the potential of mobile introns generating complex ribozymes that may have implications in gene regulation.

3.2 Introduction

The Ophiostomatales (Ascomycota) currently includes 14 accepted genera, however there could be up to 24 genera based on a recent analysis using four nuclear markers (de Beer et al. 2022). Some of the more recognized genera are *Ophiostoma*, *Leptographium*, *Grosmannia*, *Sporothrix*, and *Ceratocystiopsis* (de Beer et al. 2022). Many fungi that belong to this order are of economic importance as they can impact forestry including urban forests (Roy et al. 2014). Some members are tree pathogens that cause wilt disease by colonizing the sapwood and invading tissues needed for water and nutrient transportation. Other members are non-pathogenic but can cause blue stains on sap wood on various hard and softwood species (Seifert 1993). Blue-stain is a concern to the lumber industry as stained wood is less desirable in the market and can be challenging to export (Uzunovic and Byrne 2013).

Many members of the Ophiostomatales are vectored by arthropods such as bark beetles (Six 2012). There appears to be a symbiotic relationship between the insect vectors and many of their fungal associates. In some instances, fungi have been implicated in facilitating the bark beetle's ability to invade and develop within host trees (DiGuistini et al. 2011; Shi et al. 2012; Liu et al. 2022; Zaman et al. 2023; Fortier et al. 2024). A genomic study on *Grosmannia clavigera* revealed the presence of genes that could be involved in the detoxification of plant-defense chemicals (DiGuistini et al. 2011). The relationship between fungi and the beetles were historically assumed to be very specific to each other, but it has been observed that there is flexibility in bark-beetles/fungus association (Six and Wingfield 2011). Fungi benefit in these systems as the insects will provide a means of dispersal, however, in some instance insects may

feed on the mycelium and essentially "farm" these fungi (reviewed in Vanderpool et al. 2018). The association of fungi and beetles is therefore mutually beneficial and an important component in pathogenesis and/or the spread of blue stains in forest ecosystems (Burgess et al. 2017). *Leptographium wingfieldii* was first described in Europe as a fungal associate of *Tomicus piniperda* (Morelet 1988), however, it has been recently isolated from introduced *T. piniperda* specimens in North America (Jacobs et al. 2004; Hausner et al. 2005). Although this fungus is primarily associated with the pine shoot beetle *Tomicus piniperda*, which is native to Europe and Asia, *L. wingfieldii* has now also been isolated from native North American beetles (Jacobs et al. 2004).

Many *Leptographium* species are associates of arthropods causing blue stain on sapwood, and some are associated with tree diseases, and therefore, these are of economic concerns to the forestry industry (Wingfield et al. 1993; Uzunovic et al. 1999; Uzunovic and Byrne 2013; Harrington 1988; Eckhardt et al. 2004; Jacobs et al. 2004; Hausner et al. 2005). The movement (export) of infected lumber/timer products combined with the migration of their bark beetle vectors has facilitated some blue stain fungi to move into new ecosystems (Jacobs et al. 2004; Hausner et al. 2005; Humble and Allen 2006). Introduced fungi tend to go undetected in the early stages of their invasion and due to morphologically similar appearing native species (Burgess et al. 2017; Santini and Migliorini 2022). To maintain the biosecurity of forests, it would be beneficial to apply genomic data for the rapid generation of molecular markers to identify fungi of economic concerns (Aylward et al. 2017; Trollip et al. 2021; Trollip et al. 2022). It has been argued that fungal mitogenomes may provide a source for molecular markers suitable for fungal identification (Kulik et al. 2020; Kulik et al. 2021).

Fungal mitochondrial DNAs (mtDNA) are usually illustrated as single circular molecules, however, there is evidence to suggest that they can also be linear concatemers composed of tandemly arranged mtDNA units (Valach et al. 2011; Lang 2018). The size variation of fungal mitogenomes ranges from 12.055 to >500 kb (James et al. 2013; Zubaer et al. 2018; Liu et al. 2020). Previous studies showed that the mitogenomes of the Ophiostomatales can range from 23.7 kb to about 150 kb (Zhang et al. 1999; Zubaer et al. 2021; Wai and Hausner 2021; Wai and Hausner 2022). The size variation is in large part due to the absence/presence of potential mobile introns and intron-encoded open reading frames (ORFs) or intron-encoded proteins (IEPs), along with intergenic spacers and the presence of non-conserved (or sometimes referred to as unknown) ORFs (Hausner 2012; Zubaer et al. 2021). Mitochondrial introns can be either group I or group II introns, differentiated by their splicing mechanisms and their secondary and tertiary folds at the RNA level. The IEPs can catalyze intron mobility (referred to as intron-homing) to cognate alleles and/or these IEPs enhance intron splicing efficiency by acting as maturases (Prince et al. 2022).

Mitochondrial introns are considered as selfish elements which maintain their own life-cycle independent of the organism. Propagation of introns among different species probably involves cytoplasmic exchange for the opportunity of mitochondrial fusions and of intron-homing. Organellar introns and intron-encoded genes appear to follow neutral evolution accumulating mutations leading to degeneration and eventual loss of the element (Goddard and Burt 1999). Although mobile introns are assumed to be neutral and/or selfish elements, it has been suggested that introns can have an impact in the organism's molecular biology and gene expression (Rudan et al. 2018). The presence of introns could act as rate limiting factor in gene

expression and can affect the organism's overall metabolism (Rose 2019). Introns in some instances could be part of the adaptation ("fine tuning" gene expression) of an organism to a particular environment and lifestyle (Rudan et al. 2018). Mitochondrial introns have been implicated to mechanisms related to hypovirulence and resistance to fungicides (Baidyaroy et al. 2011; Cinget and Belanger 2020). Mobile introns have also been suggested to have been co-opeted to serve as biosensors, and organellar introns could be targets for biocontrol agents and targets of antifungal drugs (Duo et al. 2012; Belfort 2017; Liu and Pyle 2021; Liu and Pyle 2024; Fedorova et al. 2018; Albert et al. 2023). In addition, components of group I and II introns such as the ribozymes, homing endonucleases, and reverse transcriptases have been developed into gene/genome editing reagents and other biotechnology tools (reviewed in Hafez and Hausner 2012; Enyeart et al. 2014; Stoddard 2014; Guha et al. 2017; Belfort and Lambowitz 2019).

This work is part of an ongoing effort to study the evolution of mitogenomes for members of the Ophiostomatales (Abboud et al. 2018; Zubaer et al. 2019; Zubaer et al. 2021; Wai and Hausner 2021; Wai and Hausner 2022; Mukhopadhyay et al. 2023). Herein, we focus on *L. wingfieldii*, an invasive species in Canadian forests, to gain new insights into their mitogenome diversity, such as intron composition, genomic architecture, and on the evolution and systematics of the Ophiostomatales.

3.3 Materials and Methods

3.3.1 Fungal cultures and DNA purification

Seven strains of *Leptographium wingfieldii*, two strains of *Leptographium procerum* and one strain of *Leptographium terebrantis* were sequenced for this study (Table 3.1). Some of these fungi were originally isolated from the pine shoot beetles (*Tomicus piniperda*) collected in northwestern Ontario, Canada (Hausner et al. 2005). Fungal strains were grown on Malt-Extract Agar (MEA; 30 g/L malt extract, 20 g/L agar, and supplemented with 1g/L yeast extract) media (slants and plates) at room temperature (23˚ C) for 8 to 10 days. For long term preservation, fungi were maintained on agar (MEA) slants at 4˚ C. For nucleic acid preparation fungi were first cultured on agar plates and from there inoculums (10 to 12 small agar blocks ~ 2 mm) were transferred to 250 ml Erlenmeyer flasks containing 100 ml of PYG broth (1 g/L peptone, 1 g/L yeast extract, 3 g/L D-glucose). The liquid cultures were incubated at room temperature for up to 10 days. Fungal mycelium was harvested from liquid cultures by vacuum filtration through a Whatman Grade 1 filter paper in a Buchner funnel.

Mycelium was transferred to a pre chilled mortar and ground up with a pestle and acid washed sand as previously described (Wai and Hausner 2021). The recovered nucleic acids were treated with RNase A (final concentration 100 µg/mL) as recommended by the manufacturer (ThermoFisher). The DNA preparation was resuspended in 100 µL of H_2O at 30 ng/µL and sent to MicrobesNG (Unit 1-2 First Floor, The BioHub, Birmingham Research Park, 97 Vincent Drive, Birmingham B15 2SQ, UK) for Illumina sequencing. DNA preparations for strains WIN(M)661 and WIN(M)662 were sequenced by Génome Québec using the MiSeq platform (Innovation Centre, McGill University). Here for each sample 75 µL of DNA (~ 1 µg) was

supplied within an Eppendorf 96-well twin.tec® PCR plate (Cat. No. 951020401) sealed with VWR® aluminum foil (Cat. No. 60941-074). The DNAs from different fungal samples were barcoded and combined into a single MiSeq run as described in Aboud et al. (2018).

3.3.2 Genome sequencing, assembly, and annotation

Genome sequence reads generated by Illumina platforms were initially analyzed with programs available through the online Galaxy platform (Galaxy (usegalaxy.eu); Afgan et al. 2018). The reads from MicrobesNG were assessed using FastQC v0.11.9 and assemblies (contigs) were generated by using SPAdes v3.15.4 (setting the "--careful" option and assembly graph option; Bankevich et al. 2012). The Galaxy NCBI BLAST+ blastn tool (Camacho et al. 2009) was used to search through the generated contigs for sequences of interest (rDNA internal transcribed spacer regions, ITS; and beta-tubulin sequences, βT) including contigs/scaffolds corresponding to fungal mtDNAs. To recover reads for the mitogenomes from the whole genome data set the program Bandage (Wick et al. 2015) was used to examine the assembly graph files generated from SPAdes, and the program GetOrganelle (Jin et al., 2020) v1.7.5, with the organelle type set to fungus mitogenome (i.e., -F fungus_mt) was also applied to the Illumina generated sequencing reads. In a few instances, the program MIRA (Chevreux et al. 1999) was used to extract mitochondrial-derived reads. The mitochondrial scaffolds were annotated with the online MFannot program (Lang et al. 2007; Prince et al. 2022; RNAweasel and MFannot (https://megasun.bch.umontreal.ca/RNAweasel/)). tRNAs, rRNAs, and intron types were predicted by MFannot and tRNAs were also screened by using the tRNAscan-SE 2.0 program (Chan and Lowe 2019). tRNA structures were generated using R2DT program (Sweeney et al.

2021), and sequence alignment was performed in MAFFT program (Katoh et al. 2019) to investigate tRNA duplication. The annotations generated by the MFannot program were further refined with the Artemis program (Rutherford et al. 2000). The genomes and gene synteny were visualized in the Circos (Krzywinski et al. 2009) program. Additional mitogenomes were obtained from GenBank and from the Sequence Read Archive (SRA; Leinonen et al. 2011). Data obtained from SRA were assembled and annotated as described above and as previously described (Wai and Hausner 2021; Zubaer et al. 2021).

The raw sequence data can be accessed via NCBI BioProject PRJNA1136001 and the Sequence Read Archive (SRA) accession numbers are reported in Table 3.1. The summary data for the various genomes sequenced in this study are provided in Supplementary Table 3.1.

3.3.3 Intron landscape

Building on previous data and analysis for mapping organellar introns (Zubaer et al. 2021), all newly obtained *L. wingfieldii* and *L. procerum* sequences along with three *L. wageneri* sequences from the NCBI database were examined for the position and types of introns. Initially, group I and group II introns were identified by the MFannot program. The introns found by MFannot were validated by multiple sequence alignment (MSA) with intron-less versions of genes using the MAFFT program (setting: E-INS-i; Katoh et al. 2019). Sequence alignments were generated for all protein coding genes from the listed fungi (Table 3.1) with the genes of *Tolypocladium inflatum* (NCBI accession NC_036382.1). *T. inflatum* has no introns in their protein coding genes and can be used as the reference genome for intron nomenclature in protein

coding genes (Zhang and Zhang 2019). The *rns* and *rnl* sequences were compared with corresponding sequences of *Escherichia coli* (NCBI accession AB035922.1) with regards to intron annotations and naming as described in Johansen and Haugen (2001).

3.3.4 Phylogenetic analysis

The current *Leptographium* dataset was incorporated into the reference dataset from our previous phylogenetic dataset for the Ophiostomatales (Zubaer et al. 2021). The final dataset consisted of 82 amino acid sequences prepared by extracting and concatenating 13 mitochondrial protein amino acid sequences from each mitogenome. The mitochondrial protein sequences were concatenated in alphabetical order as follows: *atp6*, *atp8*, *cob*, *cox1–3*, *nad1–4*, *nad4L*, *nad5–6*. The ATP9 amino acid sequence was not included as some members of the Ophiostomatales have lost this gene from their mitochondrial genomes (Zubaer et al. 2021; Wai et al. 2021; Mukhopadhyay et al. 2023). The sequence data were aligned in MAFFT (Katoh and Standley 2013), and the output used to generate the phylogenetic tree in MrBayes program (Huelsenbeck and Ronquist 2001; Ronquist et al. 2012) using the same models and parameters previously described (Zubaer et al. 2021). The concatenated amino acid sequences of *Aspergillus fumigatus* and *Penicillium digitatum* (Eurotiales) were used as outgroups of the tree and the tree was visualized in FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

3.4 Results

3.4.1 Organisation and features of the mitochondrial genomes for Leptographium wingfieldii and related species.

The mitochondrial genomes when assembled can be visualized as circular molecule and three examples representing one strain for each of *L. wingfieldii*, *L. terebrantis* and *L. procerum* are shown in Figure 3.1. For the *Leptographium* species examined, the expected set of fungal mitochondrial genes were observed (Hausner 2003; Lang 2018; Zubaer et al. 2021), including 15 protein coding genes: ATP synthase genes (*atp6*, *atp8* and *atp9*), cytochrome oxidase genes (*cob*, *cox1*, *cox2* and *cox3*), NADH dehydrogenase genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*), ribosomal protein S3 gene (*rps3*), and two rRNA genes: small subunit ribosomal RNA gene (*rns*) and large subunit ribosomal RNA gene (*rnl*) (Figure 3.1). The *rps3* gene is located inside a group IA intron in the *rnl* gene, which is a characteristic feature of mitogenomes for all members of the Ophiostomatales and many filamentous members of the Ascomycota (Sethuraman et al. 2009; Wai et al. 2021). All genes are found on the sense (+) strand of the DNA, and the arrangement of genes is the same in all the mitogenomes of *Leptographium* spp.

In the current dataset, all the genomes showed a continuous gene arrangement of *nad2* and *nad3* gene, where start codon of *nad3* is situated just after the stop codon of *nad2* gene. For the *nad4L* and *nad5* genes, only *L. procerum* genomes show the single nucleotide overlap between them, where the last base of the stop codon in *nad4L* is used as the first base of the start codon in *nad5*. Other *Leptographium* species have one intergenic base separating the *nad4L* and *nad5* genes.

Table 3.1: List of fungi examined for mitochondrial genomes in this study, including the GenBank/SRA accession numbers, genome sizes, and intron numbers per gene.

Name of the Organism and Strain	Accession (GenBank, SRA)	Genome Size (kb)	Number of Introns in Genes															Total Introns in Species
			atp6	*atp8*	*atp9*	*cob*	*cox1*	*cox2*	*cox3*	*nad1*	*nad2*	*nad3*	*nad4*	*nad4L*	*nad5*	*nad6*	*rnl*	*rns*	
L. wingfieldii WIN(M)1181	OP973817, SRR31369142	73.295	0	0	1	4	6	4	1	1	2	0	0	0	1	0	3	0	23
L. wingfieldii WIN(M)1192	OP973811, SRR29915607	74.809	1	0	1	4	6	4	1	1	2	0	0	0	1	0	3	0	24
L. wingfieldii WIN(M)1205	OP973812, SRR30606355	73.291	0	0	1	4	6	4	1	1	2	0	0	0	1	0	3	0	23
L. wingfieldii WIN(M)1238	OP973813, SRR30606179	73.291	0	0	1	4	6	4	1	1	2	0	0	0	1	0	3	0	23
L. wingfieldii WIN(M)1240	OP973814, SRR30334524	73.291	0	0	1	4	6	4	1	1	2	0	0	0	1	0	3	0	23
L. wingfieldii WIN(M)1281	OP973815, SRR30040522	73.291	0	0	1	4	6	4	1	1	2	0	0	0	1	0	3	0	23
L. wingfieldii WIN(M)1330	OP973816, SRR30675533	76.976	0	0	1	4	7	4	1	1	2	0	0	0	2	0	2	1	25
L. procerum WIN(M)661	OP936079, SRR31509245	41.446	1	0	0	1	4	0	0	1	1	0	1	1	1	0	2	0	13
L. procerum WIN(M)1211	OP963998, SRR31364221	41.443	1	0	0	1	4	0	0	1	1	0	1	1	1	0	2	0	13
L. terebrantis WIN(M)662	OP973818, SRR31509650	69.476	0	0	1	3	7	4	1	0	2	0	0	0	1	0	2	0	21
L. wageneri var. *pseudotsugae*	SRR14246565*	124.497	2	0	1	4	11	4	0	4	3	1	0	1	3	0	8	2	44
L. wageneri var. *wageneri*	SRR14246566*	118.26	1	0	1	3	11	4	0	4	4	1	0	1	3	0	7	2	42
L. wageneri var. *pseudotsugae*	SRR14246613*	126.48	2	0	1	4	11	5	0	4	3	1	0	1	3	0	8	2	45
Total Introns in Genes			8	0	11	44	91	45	8	21	28	3	2	5	20	0	49	7	342

*Third party sequences (read files from SRA database) were collected from NCBI. GenBank accession unavailable for those organisms. Mitochondrial genome sequence and annotation can be accessed in the supplementary data file.

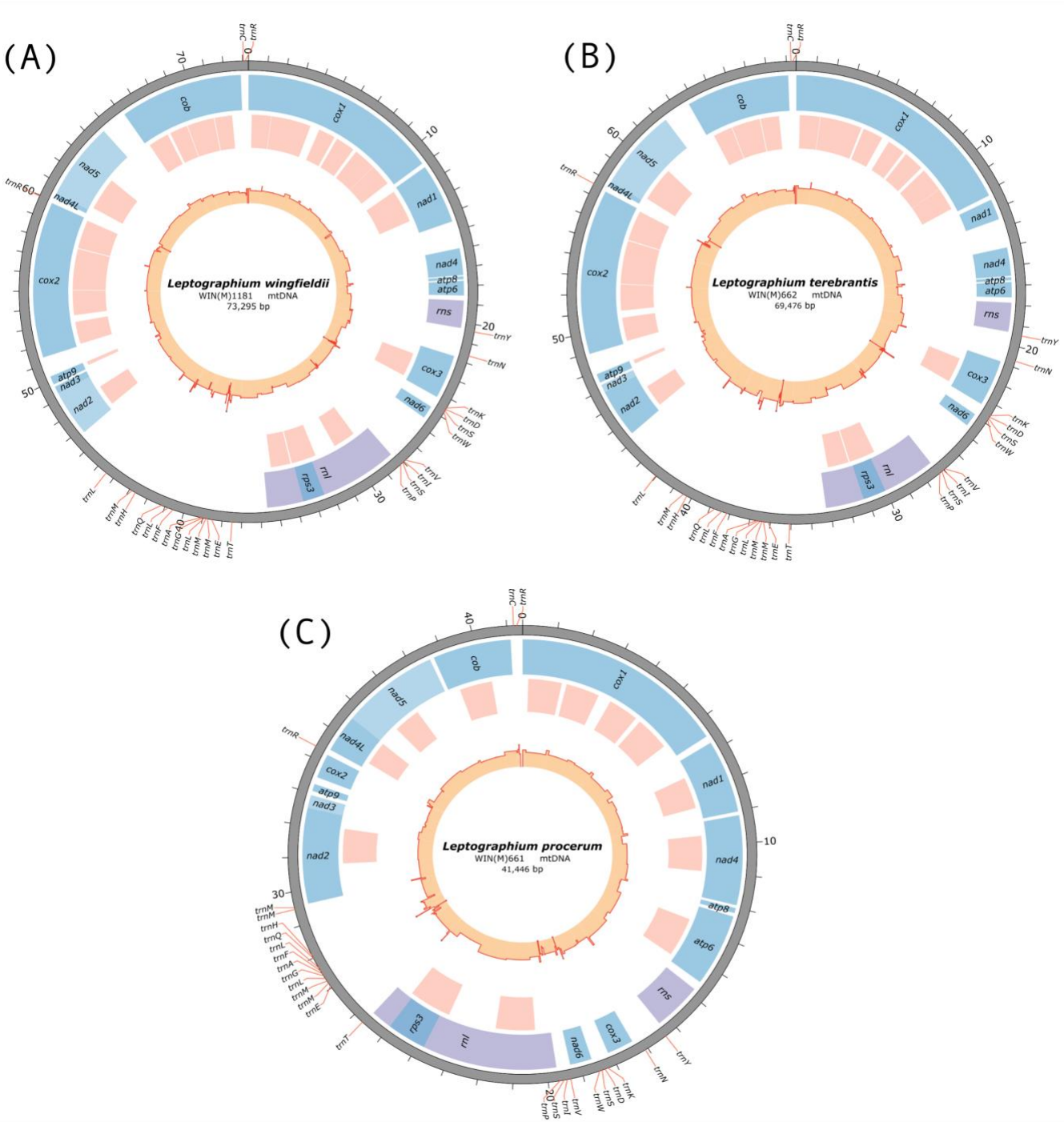

(A)
Leptographium wingfieldii
WIN(M)1181 mtDNA
73,295 bp
(B)
Leptographium terebrantis
WIN(M)662 mtDNA
69,476 bp
(C)
Leptographium procerum
WIN(M)661 mtDNA
41,446 bp
cob
cox1
nad1
nad4
atp8
atp6
rns
cox3
nad6
rnl
rps3
nad2
nad3
atp9
cox2
nad4L
nad5
trnR
trnC
trnY
trnN
trnK
trnD
trnS
trnW
trnV
trnI
trnS
trnP
trnL
trnM
trnH
trnQ
trnL
trnF
trnA
trnG
trnL
trnM
trnM
trnE
trnT

Figure 3.1: The mitochondrial genomes of (A) *Leptographium wingfieldii*, (B) *L. terebrantis*, and (C) *L. procerum*. The genomes are presented on circular scales according to their genome size on a scale of kilobases (i.e., one tick mark is 1 kb). Circular representation of the mitochondrial genomes generated by the program Circo showing the tRNA (pointed outward on the scale), protein coding genes (blue track), introns (red track), and genetic-feature-wise GC graph (GC% was calculated for annotated features instead of fixed window, showed in the innermost track).

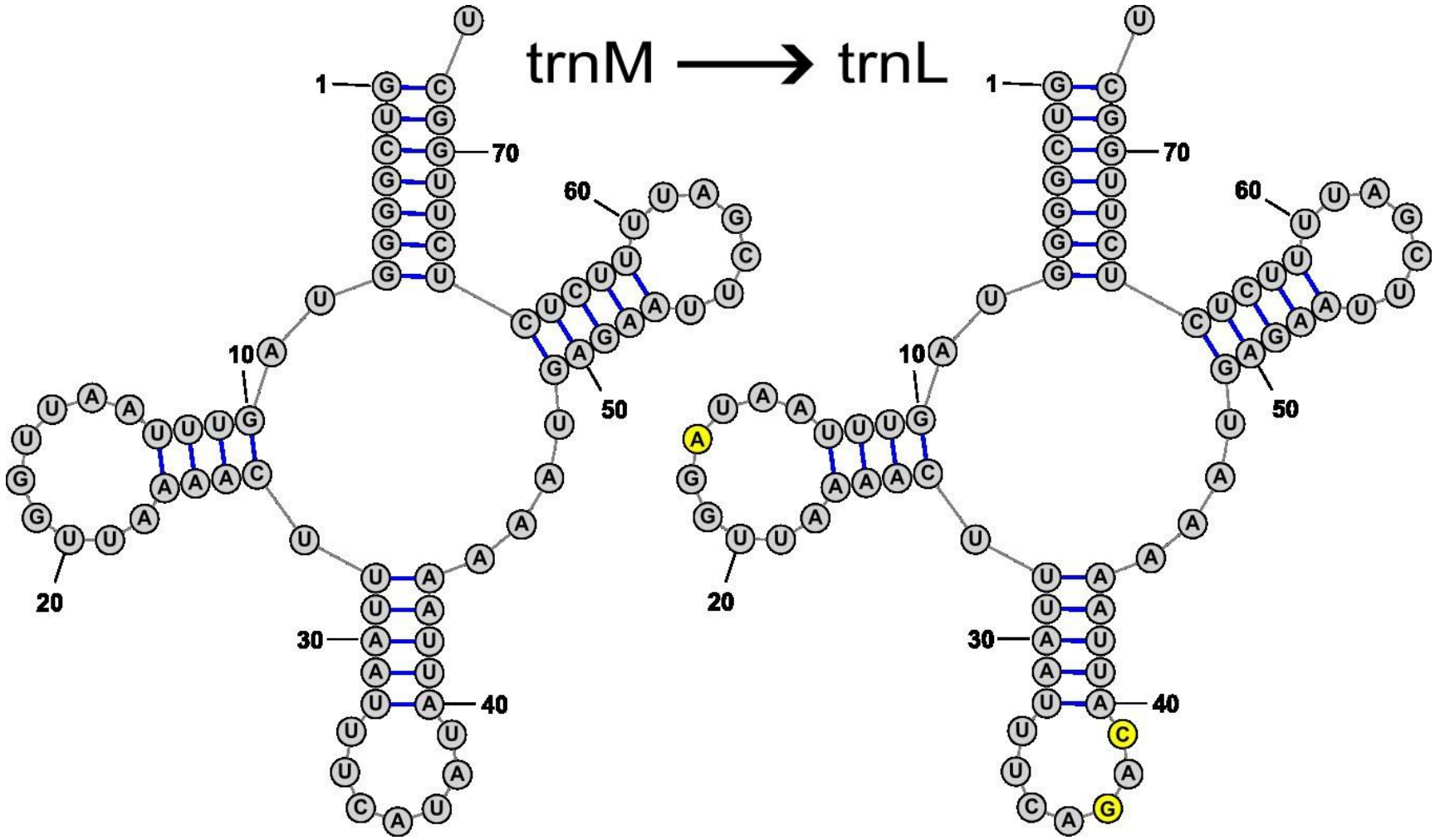

trnM ⟶ trnL
1
70
60
50
10
20
30
40

Figure 3.2: Secondary structure comparison of the *trnM* gene (upstream to *nad2* gene in *L. procerum*, left) the duplicated *trnL* (as seen in upstream to *nad2* gene in *L. wingfieldii and L. terebrantis*, right). The comparison shows three substitutions (highlighted in yellow) including one substitution in the anti-codon loop.

The mitochondrial genomes of *Leptographium* code for 26 tRNAs that contain the complete set for coding for all 20 amino acids. Multiple copies were recorded for *trnS* (2 copies), *trnR* (2 copies), *trnL* (2 copies) and *trnM* (3 copies). An additional copy of *trnM* is found in *L. procerum* upstream the *nad2* gene. At the same location, the *trnL* genes was observed (instead of *trnM*) in the mitogenome of *L. wingfieldii* and *L. terebrantis*. Sequence alignment and secondary structure comparison of these versions (located upstream of *nad2*) of *trnM* and *trnL* gene showed that there are three substitutions including a mutation in the anti-codon that appears to potentially convert *trnM* to *trnL* (Figure 3.2, Supplementary Figure 3.1). The tRNA genes found to be concentrated in certain regions of the mitogenomes. For example, thirteen tRNA genes are situated downstream of the *rnl* gene and 4 tRNA genes are upstream of the *rnl* gene. The pattern of tRNA arrangement is conserved among the *Leptographium* species (Figure 3.1, Supplementary Table 3.2).

3.4.2 Mitogenome expansion, mobile introns, and complex introns

The genome sizes of *L. terebrantis* and *L. wingfieldii* strains range from 69.476 kb to 76.976 kb, respectively, while the mitogenomes for the two examined *L. procerum* strains ranged from 41.443 to 41.446 kb. The *L. wageneri* var. *wageneri* and *L. wageneri* var. *pseudotsugae* mitogenomes are larger, comprising 118.260 to 126.480 kb (Table 3.1). The genome size differences among the *Leptographium* species examined are related to the number of introns. The plot of intron numbers (intron insertion sites) and genome size shows a linear trend (Figure 3.3), which supports our previous finding of introns being the major factor in mitochondrial genome expansion among members of the Ophiostomatales (Zubaer et al. 2021).

Introns were found in all conserved mitochondrial protein coding genes (except *rps3*) and rRNA genes except for *atp8* and *nad6* genes. The presence of introns in a gene was mapped with regards to the reference *Tolypocladium inflatum* for protein coding gene positions, and *E. coli* for rRNA gene (*rns* and *rnl*) positions. A pan-genomic intron landscape is presented for all intron homing sites observed among the examined *Leptographium* species in Figure 3.4. In total, 55 intron insertion sites were identified in the dataset, and a total of 342 introns were found distributed in those intron sites across the studied 13 mitogenomes (Table 3.1, Supplementary Table 3.3). The 2450 position in *rnl* gene (mL2450 in Johansen and Haugen 2001 notation) was occupied in all members by a group IA intron that encodes the *rps3* gene. The *cox1* gene was found to have the highest number of intron insertion sites (91 introns), with the *rnl* and *cob* gene having 49 and 45 intron insertion sites. Among the 342 recorded introns, there were two group II introns were found: one is cox1-216 found in the strains of *L. wageneri*, and another one is rns-952 (mS952 in Johansen and Haugen 2001 notation). The mS952 group II intron encodes IEPs that are LAGLIDADG type homing endonucleases (HEs) in *L. wingfieldii* WIN(M)1330 and strains of *L. wageneri*. This group II intron was previously described by Mullineux et al. (2011). The remaining introns found in this study are group I introns mostly harbouring homing endonuclease genes (HEGs) that encode IEPs that can be assigned to the LAGLIDADG or GIY-YIG HE families and some introns lacked ORFs.

Group I intron RNAs can fold due to complementary sequences hydrogen bonding and thus forming paired (P) regions (helices) that are referred to as P1 to P10 with single-stranded loop sequences connecting the helices (Michel and Westhof 1990). Based on variations within the primary, secondary, and tertiary structures, group I introns have been arranged into

categories: classes IA to IE and these can be further subdivided, e.g., IA1, IC3 etc. (Michel and Westhof 1990; Hausner et al. 2014). For more details on the various types of introns, such as group I intron categories and potential ORFs, see Supplementary Table 3.3.

Among the introns found in this investigation, there is a complex intron noted in position 281 in *cox1* gene in *L. wageneri* strains. Complex introns are designated as intron arrangements where an intron is either nested or fused to another intron. Here, a group IB intron appears to have been invaded by a IC2 type group I intron. The resident IB intron encodes a double-motif LAGLIDADG ORF possibly in P1 and the ORF is interrupted by the IC2 module that encodes a double-motif LAGLIDADG ORF in P9.

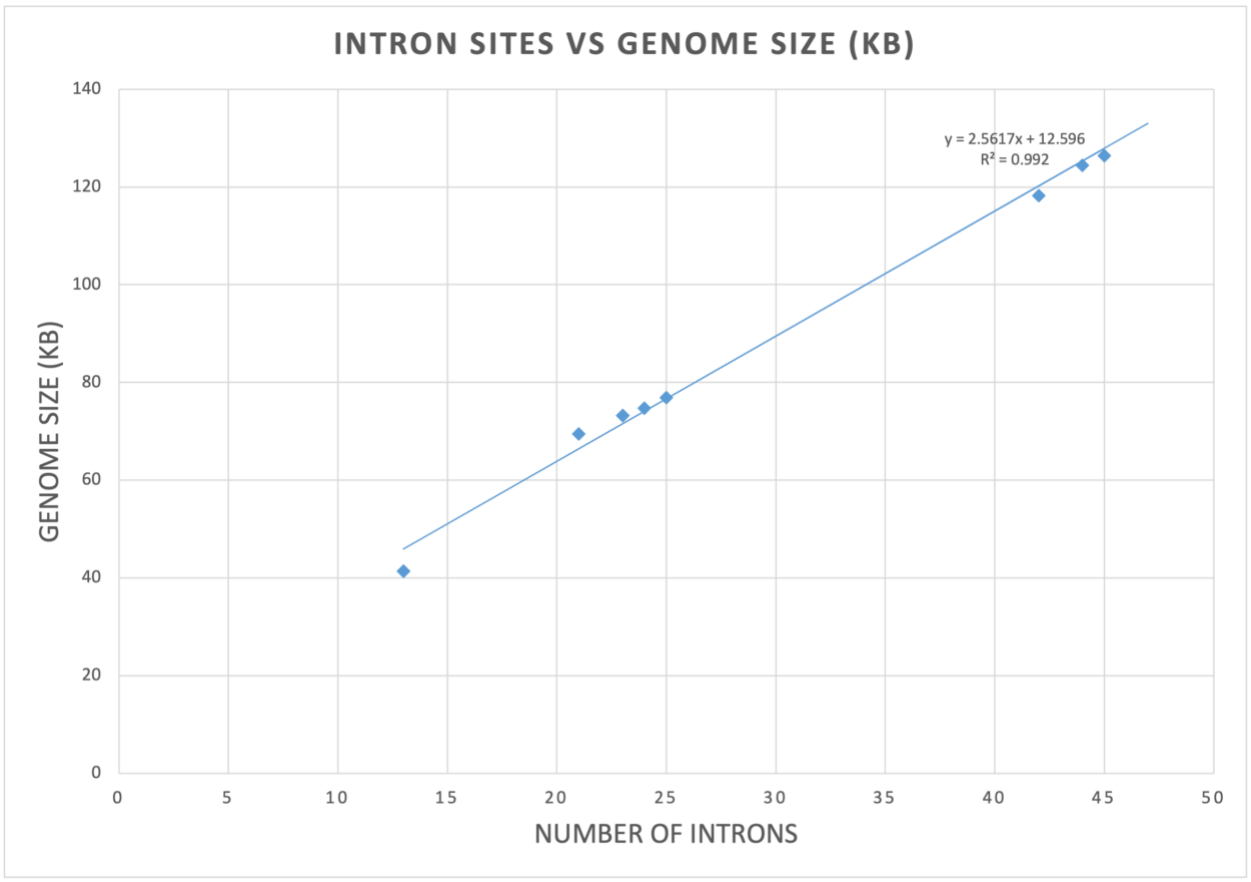

INTRON SITES VS GENOME SIZE (KB)
GENOME SIZE (KB)
NUMBER OF INTRONS
y = 2.5617x + 12.596
R² = 0.992
0
20
40
60
80
100
120
140
0
5
10
15
20
25
30
35
40
45
50

Figure 3.3: Correlation graph of genome size vs intron insertion sites for the 13 *Leptographium* genomes examined in this study. Intron numbers are on the x-axis and genome sizes are on the y-axis. The graph shows a linear trend with strong correlation.

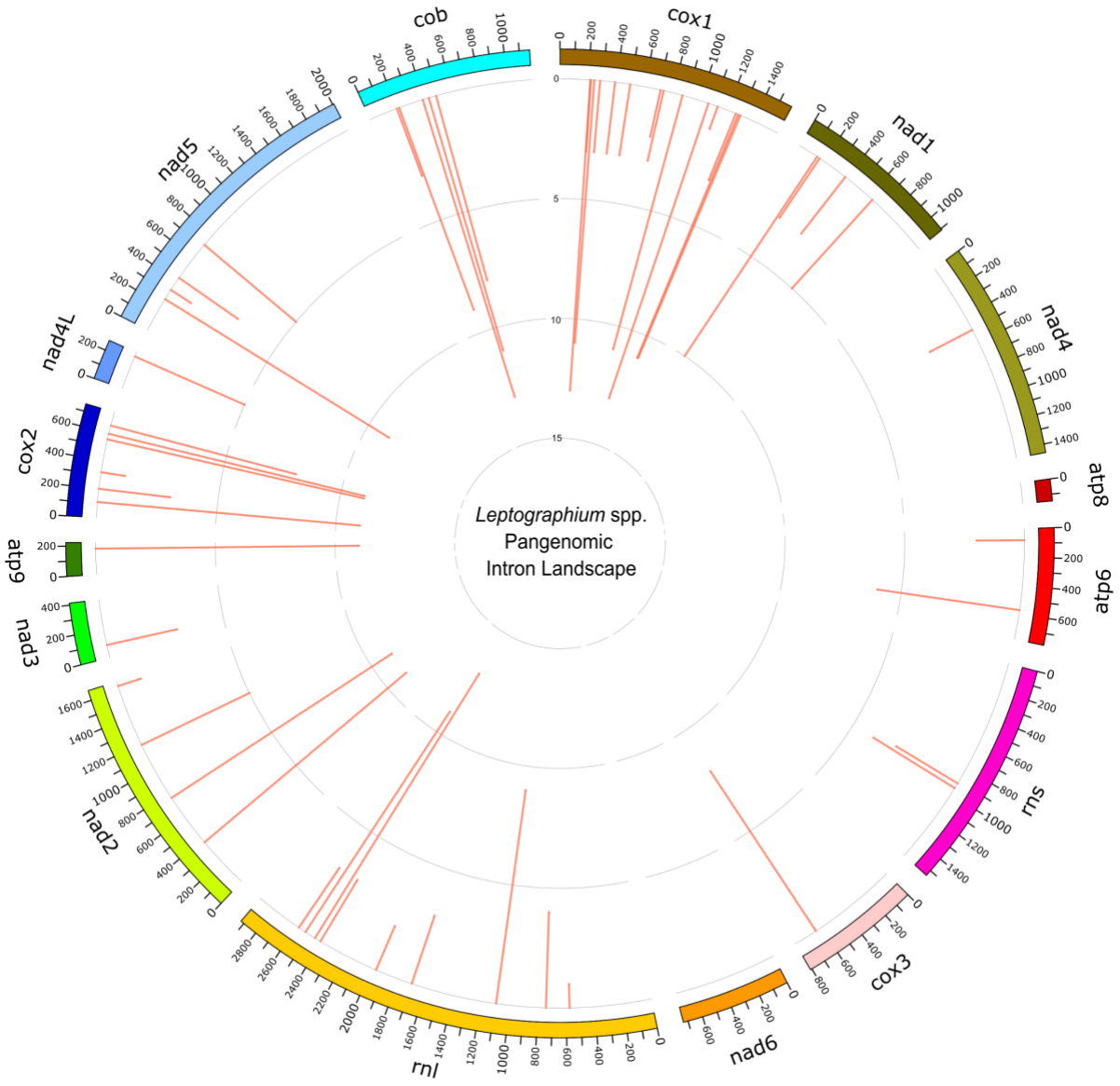

Leptographium spp.
Pangenomic
Intron Landscape
cob
cox1
nad1
nad4
atp8
atp6
rns
cox3
nad6
rnl
nad2
nad3
atp9
cox2
nad4L
nad5

Figure 3.4: The pan-intronic landscape for the 13 *Leptographium* mitogenomes examined in this study. The landscape shows all the intron-homing sites and intron frequencies on those positions. The graph was prepared by using a vertical scale ranging from 0 to 15. The intron-homing sites are positioned on the reference genes represented in the outer track using a scale of number of base pairs.

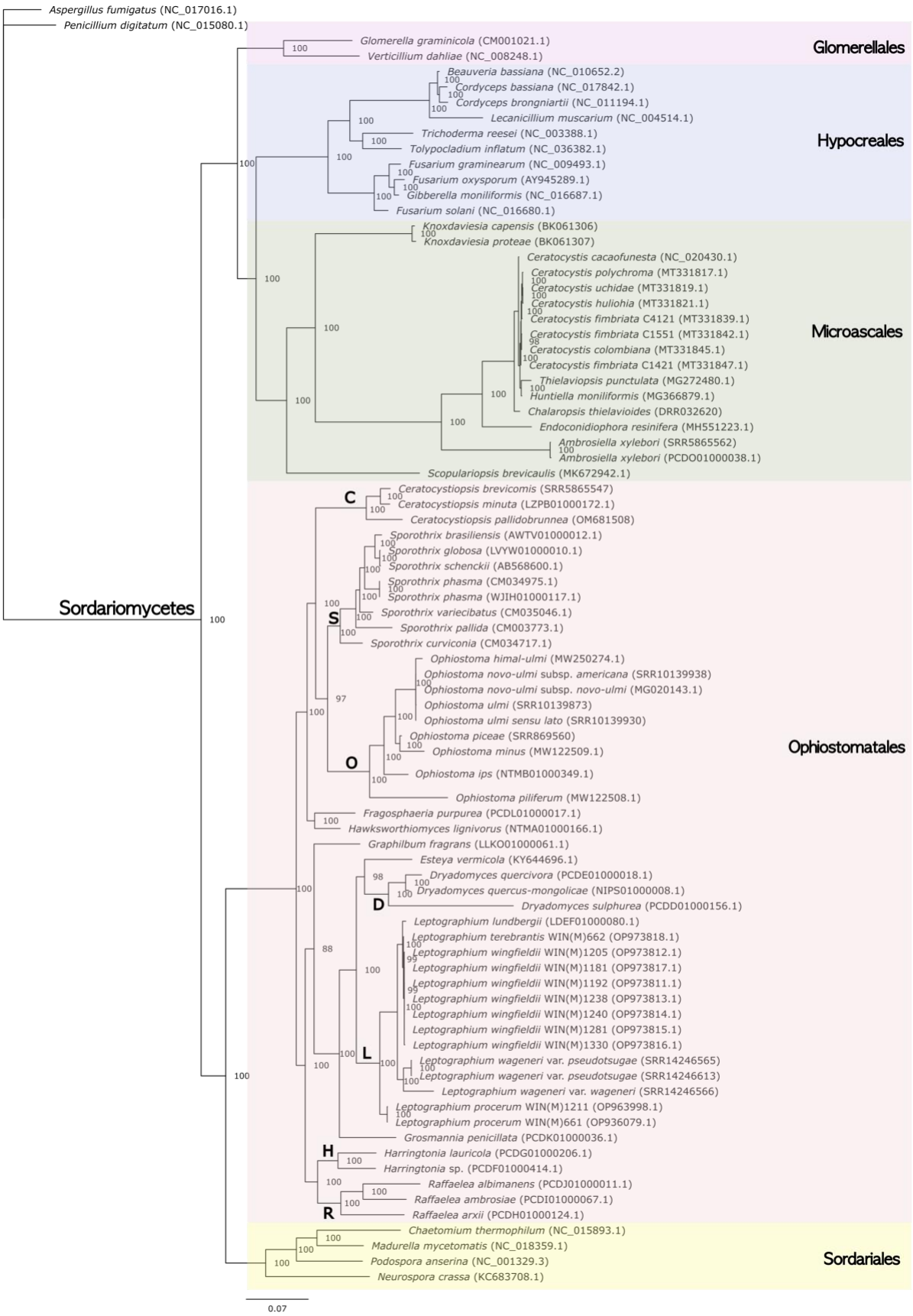

Aspergillus fumigatus (NC_017016.1)
Penicillium digitatum (NC_015080.1)
Glomerella graminicola (CM001021.1)
Verticillium dahliae (NC_008248.1)
Glomerellales
Beauveria bassiana (NC_010652.2)
Cordyceps bassiana (NC_017842.1)
Cordyceps brongniartii (NC_011194.1)
Lecanicillium muscarium (NC_004514.1)
Trichoderma reesei (NC_003388.1)
Tolypocladium inflatum (NC_036382.1)
Fusarium graminearum (NC_009493.1)
Fusarium oxysporum (AY945289.1)
Gibberella moniliformis (NC_016687.1)
Fusarium solani (NC_016680.1)
Hypocreales
Knoxdaviesia capensis (BK061306)
Knoxdaviesia proteae (BK061307)
Ceratocystis cacaofunesta (NC_020430.1)
Ceratocystis polychroma (MT331817.1)
Ceratocystis uchidae (MT331819.1)
Ceratocystis huliohia (MT331821.1)
Ceratocystis fimbriata C4121 (MT331839.1)
Ceratocystis fimbriata C1551 (MT331842.1)
Ceratocystis colombiana (MT331845.1)
Ceratocystis fimbriata C1421 (MT331847.1)
Thielaviopsis punctulata (MG272480.1)
Huntiella moniliformis (MG366879.1)
Chalaropsis thielavioides (DRR032620)
Endoconidiophora resinifera (MH551223.1)
Ambrosiella xylebori (SRR5865562)
Ambrosiella xylebori (PCDO01000038.1)
Microascales
Scopulariopsis brevicaulis (MK672942.1)
Sordariomycetes
C
Ceratocystiopsis brevicomis (SRR5865547)
Ceratocystiopsis minuta (LZPB01000172.1)
Ceratocystiopsis pallidobrunnea (OM681508)
Sporothrix brasiliensis (AWTV01000012.1)
Sporothrix globosa (LVYW01000010.1)
Sporothrix schenckii (AB568600.1)
Sporothrix phasma (CM034975.1)
Sporothrix phasma (WJIH01000117.1)
S
Sporothrix variecibatus (CM035046.1)
Sporothrix pallida (CM003773.1)
Sporothrix curviconia (CM034717.1)
Ophiostoma himal-ulmi (MW250274.1)
Ophiostoma novo-ulmi subsp. americana (SRR10139938)
Ophiostoma novo-ulmi subsp. novo-ulmi (MG020143.1)
Ophiostoma ulmi (SRR10139873)
Ophiostoma ulmi sensu lato (SRR10139930)
Ophiostoma piceae (SRR869560)
Ophiostoma minus (MW122509.1)
O
Ophiostoma ips (NTMB01000349.1)
Ophiostoma piliferum (MW122508.1)
Fragosphaeria purpurea (PCDL01000017.1)
Hawksworthiomyces lignivorus (NTMA01000166.1)
Graphilbum fragrans (LLKO01000061.1)
Esteya vermicola (KY644696.1)
Dryadomyces quercivora (PCDE01000018.1)
Dryadomyces quercus-mongolicae (NIPS01000008.1)
D
Dryadomyces sulphurea (PCDD01000156.1)
Leptographium lundbergii (LDEF01000080.1)
Leptographium terebrantis WIN(M)662 (OP973818.1)
Leptographium wingfieldii WIN(M)1205 (OP973812.1)
Leptographium wingfieldii WIN(M)1181 (OP973817.1)
Leptographium wingfieldii WIN(M)1192 (OP973811.1)
Leptographium wingfieldii WIN(M)1238 (OP973813.1)
Leptographium wingfieldii WIN(M)1240 (OP973814.1)
Leptographium wingfieldii WIN(M)1281 (OP973815.1)
Leptographium wingfieldii WIN(M)1330 (OP973816.1)
L
Leptographium wageneri var. pseudotsugae (SRR14246565)
Leptographium wageneri var. pseudotsugae (SRR14246613)
Leptographium wageneri var. wageneri (SRR14246566)
Leptographium procerum WIN(M)1211 (OP963998.1)
Leptographium procerum WIN(M)661 (OP936079.1)
Grosmannia penicillata (PCDK01000036.1)
H
Harringtonia lauricola (PCDG01000206.1)
Harringtonia sp. (PCDF01000414.1)
Raffaelea albimanens (PCDJ01000011.1)
Raffaelea ambrosiae (PCDI01000067.1)
R
Raffaelea arxii (PCDH01000124.1)
Ophiostomatales
Chaetomium thermophilum (NC_015893.1)
Madurella mycetomatis (NC_018359.1)
Podospora anserina (NC_001329.3)
Neurospora crassa (KC683708.1)
Sordariales
0.07

Figure 3.5: Phylogenetic tree for members of the Ophiostomatales along with sequences from other Orders in the Sordariomycetes to verify the position of *Leptographium wingfieldii* and other *Leptographium* species. For the Ophiostomatales, nodes that define different genera are designated with a single letter as follows C: *Ceratocystiopsis*, S: *Sporothrix*, O: *Ophiostoma*, D: *Dryadomyces*, L: *Leptographium*, H: *Harringtonia*, R: *Raffaelea*. The tree was generated by MrBayes using the concatenated and aligned mitochondrial amino acid sequences of thirteen conserved protein coding genes from 82 fungal samples. *Aspergillus fumigatus* and *Penicillium digitatum* were used as outgroups. The branches are drawn to scale, where the length corresponds to the number of substitutions per site. Posterior probability values are indicated at the nodes.

3.4.3 Phylogeny of selected members of Leptographium

The phylogenetic tree was constructed based on 82 mitochondrial genomes including 10 *Leptographium* genomes sequenced during this study (Figure 3.5). In addition, we extracted mitogenomes from GenBank and from data available at the Sequence Read Archive (SRA) for strains representing *L. lundgergii*, *L. wageneri* var. *wageneri* and *L. wageneri* var. *pseudotsugae* (Shumway et al. 2010; Leinonen et al. 2011). We included 80 sequences representing various Orders that belong to the Class Sordariomycetes, and two sequences from the Order Eurotiales serving as the outgroup. The phylogenetic tree provides support for the monophyly of sequences sampled from the Glomerellales, Hypocreales, Microascales, Ophiostomatales, and Sordariales. The Ophiostomatales form a single clade containing 47 species and these supported the monophyly of the following genera: *Ceratocystiopsis*, *Sporothrix*, *Ophiostoma*, *Fragosphaeria, Hawksworthiomyces, Graphilbum*, *Esteya*, *Dryadomyces*, *Leptographium sensu stricto*, *Grosmannia*, *Harringtonia*, and *Raffaelea*. With regards to all seven studied sequences for the *Leptographium wingfieldii* strains (including the WIN(M)1181, the ex-paratype), they clustered together with sequences from *L. terebrantis*, and *L. lundbergii*. The three sequences that represent *L. wageneri* var. *wageneri* and *L. wageneri* var. *pseudotsugae* were positioned on a slightly deeper node and sequences obtained for strains of *L. procerum* were the branch from the node that supports the monophyly for the sampled members of *Leptographium sensu stricto*. Overall, the phylogenetic tree agrees with the generic concepts recently proposed for the Ophiostomatales by de Beer et al. (2022). The limited mitogenome-derived data on *L. wageneri var. wageneri* and *L. wageneri* var. *pseudotsugae* does support the latest proposal that these varieties should be recognized as distinct species, *Leptographium wageneri* and *Leptographium pseudotsugae*, respectively (Choi et al. 2023).

3.5 Discussion

3.5.1 Genomic variations in Leptographium

The mitogenomes of *Leptographium* species sampled in this study encode the typical genes observed in fungal mitogenomes, which include 15 protein coding genes, 2 rRNA genes, and a set of tRNAs for all 20 amino acids (Lang et al. 2018). Gene synteny among the examined *Leptographium* species was found to be conserved. In general, mitogenome synteny is conserved for all members of the Ophiostomatales that have been characterized so far (Abboud et al. 2018; Zubaer et al. 2021; Wai and Hausner 2021; Mukhopadhyay et al. 2023). *Leptographium* mitogenomes show the presence of the *atp9* gene, while some other members of the Ophiostomatales appear to lack *atp9* (see Wai et al. 2021). Previous studies also noted the one nucleotide overlap, and continuous/fused gene arrangements found between *nad4L* and *nad5* genes and *nad2* and *nad3* genes, respectively (Zubaer et al. 2018; Wai et al. 2021).

It has been reported that tRNAs in mitogenomes can be subject to duplication and gene loss, explained by the tandem duplication/random loss (TDRL) model (Jiang et al. 2016). However, duplication may also give rise to a tRNA with different amino acid preference by accumulating mutation in the anti-codon bases – a process called tRNA remolding (Jiang et al. 2016). tRNA remolding was previously reported for *Ceratocystiopsis pallidobrunnea* (Wai and Hausner 2022) and in other fungi (Song et al. 2024); in *C. pallidobrunnea* a duplicated *trnM* appeared to change to *trn*. Similar situation was found in the current study where one mutation changed the anticodon of *trnM* to *trnL* as found in *L. wingfieldii* and *L. terebrantis*. However, the final impact of the mutation in the anticodon is not clear as the amino acid acceptor stem remains unchanged.

A characteristic feature of most fungal mitogenomes is the distribution of tRNAs with a bias to being located upstream and downstream of the *rnl* gene (Zubaer et al. 2018; Wai et al. 2019). The examined *Leptographium* species showed a similar trend of tRNAs clustering around the *rnl* gene in the mitogenome and the remaining tRNAs usually being located between the following genes: *rns*/*cox3*, *cox3*/*nad6*, *cox2*/*nad4L*, and *cob*/*cox1*. This might generate transcripts that ensures that the *rnl* RNA is co-expressed at a similar rate as the flanking tRNAs. In addition, it has been assumed that in mitochondria the location of tRNAs might serve as signals for RNA processing of the primary (assumed to be polycistronic) transcripts, referred to as the punctuation model (Ojala et al. 1981; D'Souza and Minczuk 2018; Golik 2024; Zhang et al. 2023). There is support for this model in fungal mitochondria for some polycistronic transcripts (reviewed in Varassas and Kouvelis 2022), but in the Ophistomatales there are genes that are not flanked by tRNAs (see supplementary Table 3.2) so other, not yet identified processing motifs may be important in resolving those transcripts.

3.5.2 The intron complements of the Leptographium mitogenomes

Fungal mitochondrial genomes frequently contain group I and group II introns. These elements are potential ribozymes with self-splicing activity, but under cellular conditions splicing is most likely assisted by either intron- or nuclear-encoded maturases/splicing factors that promote intron RNA folding into splicing competent configurations (Lang et al. 2007; Mukhopadhyay and Hausner 2024). The intron landscape map (Figure 3.4) shows that intron insertion sites or homing sites are conserved among different species. Intron bias for being

"intron-rich" in the *cox1* gene is found to be a common phenomenon in members of the Ophiostomatales (Zubaer et al. 2021; Wai and Hausner 2021).

A complex intron was detected in this study at position *cox1*-281, and this illustrates the dynamic nature of these elements that are under continuous pressure to insert into new locations for their survival to avoid extinction (Goddard and Burt 1999). Nested intron arrangements are of interest as they may have the potential to evolve new complex ribozymes or generate regulatory elements that can be resolved by splicing pathways that can enhance the expression of the individual's intron module's open reading frames (Wallweber et al. 1997; Guha et al. 2018; Zubaer et al. 2021). The resolution of complex introns may also impose a rate limiting step in the expression of the host gene. The possible splicing pathways for the *cox1*-281 intron will be explored in future studies as this may impact RNA processing of primary transcripts and gene expression of the components that are present on polycistronic mRNAs along with the genes that host these types of complex/nested introns.

The dependence of organellar introns on nuclear- and intron-encoded factors for efficient splicing links nuclear responses to environmental conditions to mitochondrial gene expression (Prince et al. 2022; Mukhopadhyay and Hausner 2024). It is therefore possible that the organism's life history along with drift and selection might be impacted by the presence of introns. Within the Ophiostomatales, members that are animal pathogens typically have small mitochondrial genomes with few introns (Teixeira et al. 2014; Abboud et al. 2018). Conversely, those members causing blue-stain tend to have larger mitogenomes with higher number of introns (Zubaer et al. 2018; Mukhopadhyay et al. 2023). The current results showed that the

number of introns and genome sizes are similar in *L. wingfieldii* and *L. terebrantis*, while varieties of *L. wageneri* show larger mitogenomes with higher number of introns. Overall findings in the current study confirm previous reports on the Ophiostomatales where introns are a major factor in mitogenome expansion (Zubaer et al. 2021; Mukhopadhyay et al. 2023).

3.5.3 Phylogenetic analysis of mtDNAs from various members of the Ophiostomatales

The current study utilized mitogenome-derived data for addressing the taxonomic position of *L. wingfieldii* within the Ophiostomatales. The data showed *L. wingfieldii*, *L. terebrantis*, and *L. lundbergii* are closely related to each other, whereas *L. procerum* was found to be a more distantly related to *L. wingfieldii*. The current mitogenome data also suggests that genera *Esteya* and *Dryadomyces*, share a recent common ancestor with *Leptographium* and along with *Grosmannia* all four genera are monophyletic. *Leptographium* was traditionally considered to be the asexual (anamorph) state of *Grosmannia.* This is still a complex issue as some members of *Grosmannia* do indeed form *Leptographium*-like asexual states and some members of *Leptographium* produce perithecia (reviewed in Zipfel et al. 2006; de Beer et al. 2022).

With the introduction of the single name nomenclature for fungi (Hawksworth et al. 2011), *Leptographium* has priority as the older generic name. However, the position of some lineages of fungi with *Leptographium*-like anamorphs within the Ophiostomatales has not been resolved (Zipfel et al. 2006; De Beer and Wingfield 2013; Jankowiak et al. 2018). Moreover, studies have shown that there are at least two clades (with the corresponding type specimens: *Leptographium lundbergii* and *Grosmannia penicillata*) that group members of the

Ophiostomatales with *Leptographium*-like asexual states. This validates the genera *Leptographium* and *Grosmannia* (De Beer and Wingfield 2013; Nel et al. 2021; de Beer et al. 2022).

The mitogenome-based data also supports the division of the genus *Raffaelea sensu lato* into three genera, which have recently been erected: *Dryadomyces, Harringtonia,* and *Raffaelea* (de Beer et al. 2022). This study suggests that mitogenomes should be further explored as a taxonomic marker for species identification and verifying and delimitating generic boundaries for members of the Ophiostomatales.

3.6 Conclusions

Leptographium wingfieldii, L. procerum, and *L. terebrantis* are saprotrophs and they are sometimes associated with tree mortality (Harrington 1988; Hausner et al. 2005; Jacobs and Wingfield 2001, 2013; Mensah et al. 2021). Varieties of *L. wageneri*, are known causal agents of black stain root disease (Bennett et al. 2021). *Leptographium wingfieldii*, a potential tree pathogen, has been recently introduced into Canadian forests. It comes with challenges of adaptation and risks to the native forest ecology. This fungus invades sapwood of certain conifers and maintains symbiotic/mutualistic relationships with bark beetles, and thus pose a risk of an outbreak of a plant disease. Mitogenomes offer a suite of potential markers that can aid in the identification of invasive fungi and in resolving taxonomic issues within the

Ophiostomatales. Exploration of the mitogenome and mobile elements in *L. wingfieldii* and related species is potentially beneficial to understand the biology and taxonomy of this species. Characterizing mitogenomes and cataloging introns and other novel intron derivatives is a resource for locating ribozymes and intron-encoded proteins (homing endonucleases and reverse trancriptases) that have applications to biotechnology.

Supplementary Figures.

Supplementary data are available for this Chapter at https:

https://cdnsciencepub.com/doi/10.1139/cjm-2024-0179

CHAPTER 4: BIOSYNTHETIC GENE CLUSTERS IN MEMBERS OF *LEPTOGRAPHIUM WINGFIELDII* AND RELATED TAXA

CHAPTER 4: BIOSYNTHETIC GENE CLUSTERS IN MEMBERS OF *LEPTOGRAPHIUM WINGFIELDII* AND RELATED TAXA

4.1 ABSTRACT

Leptographium wingfieldii belongs to the order Ophiostomatales. It is a pathogen of pines and an agent of blue stain in sapwood on infected trees. This fungus is vectored by bark beetles such as *Tomicus piniperda* (the pine shoot beetle). The movement of bark beetle and their fungal associates can have severe ecological and economic impacts. Based on previous surveys conducted on the *Leptographium* collection housed as part of the WIN collection (University of Manitoba), several fungal candidates that showed antimicrobial activity while being challenged with multidrug resistant bacterial strains of *Staphylococcus aureus.* Those fungi were examined in more detail and studied at the genomic level to investigate the genetic basis of their antimicrobial activities. The whole genome of seven strains of *Leptographium wingfieldii*, one strain of *L. procerum*, one strain of *L. aureum*, and one strains of *Penicillium citrinum* were obtained using the Illumina sequencing platform. The taxonomic identity of the *Leptographium* strains was confirmed by applying five different nuclear marker genes (ITS2, ACT, CAL, TEF, B-TUB). The sequence data (reads) were assembled in SPAdes and the contigs were submitted to a fungal specific antiSMASH program to identify potential biosynthetic gene clusters. A comparative analysis using the BiG-SCAPE program showed the relationships among biosynthetic gene clusters (BGCs), and their clustering with known BGC families. The analysis revealed 205 BGCs which are related to 97 BGC families. The gene clusters of the studied

Leptographium strains were further annotated using BLAST to provide a reference for other members of the Ophiostomatales.

4.2 Introduction

Leptographium wingfieldii is a fungus that belongs to the order Ophiostomatales in the phylum Ascomycota. *L. wingfieldii* was originally described in Europe (Morlet 1988) and later was recognized as an invasive species in Canada (Jacobs et al. 2014; Hausner et al. 2005). Being a newly introduced species, this fungus poses a risk to the Canadian forest ecology. Although *L. wingfieldii* was not found to be a significant pathogen to conifers in Canada, it can cause blue stain on sapwood which is an undesirable characteristic for the lumber industry (Uzunovic et al. 1999). Stained lumber is less desirable for export and therefore limits the marketing of affected lots of wood products. This fungus is primarily associated with the pine shoot beetle *Tomicus piniperda* and maintains a symbiotic relation by helping beetles to live under the bark of host trees by presumably producing secondary metabolites that appear to stabilize the effect of the hosts immune system (Jacobs et al. 2004; de Beer and Wingfield 2013; Zaman et al. 2023). In general, there are examples where bark beetles or similar arthropods feed on blue stain fungi and keep farming them for their survival (Vanderpool et al. 2018). Fungi get the advantage by being vectored to other trees in the forest with the movement of beetles that carry their spores. The association of the fungi and beetles were initially thought to be species specific, however, it has been observed that members of the Ophiostomatales can cooperate with other beetles (and vice versa) (Six and Wingfield 2011). To adapt to a new environment, fungi can enter into a

symbiotic relationship with native bark beetle species, thus spreading more efficiently in a new location, colonizing the forest, eventually posing a risk in the forest environment (Six and Winfield 2011; Linnakoski et al. 2012a; Linnakoski et al. 2012b).

The filamentous fungi in Ascomycota can be an abundant source of natural products such as antibiotics, toxins, pigments etc. (Keller 2015). For example, Penicillin (*Penicillium* sp.; Fleming 1929), aflotoxin (*Aspergillus* sp.; Sweeney and Dobson 1998), Lovastatin (*Aspergillus terreus*; Alberts et al. 1980), Fumonisins (*Fusarium verticillioides*; Gelderblom et al. 1988), Cyclosporin (*Tolypocladium inflatum*; Borel et al. 1976), Eleniol (*Ophiostoma quercus*; Oh et al. 2010), Terbinafine (*Sporothrix schenckii*; Córdoba et al. 2018), Ophiocordin (*Ophiostoma* sp.; Takahashi et al. 1998) etc. Thus, *L. wingfieldii* can be considered a potential producer of valuable secondary metabolites with industrial and medical importance. Secondary metabolites (SM) - a derivative of primary metabolites - are not involved in the primary function of an organism such as growth and reproduction. But they can provide the organisms with an advantage to survive while living with stress in a competitive environment. So called biosynthetic gene clusters (BGCs) are associated with the production of secondary metabolites (SMs). Secondary metabolites are usually derivates of Acetyl-CoA, however, they can also be derivatives of aromatic amino acids (Nielsen and Nielsen 2017). They are categorized based on the major enzyme that is related to their production. The major classes of SMs are Polyketides and Non-ribosomal peptides, and these are categorized based on the enzymes involved in their synthesis: Polyketide synthase (PKS) and Non-ribosomal peptide synthase (NRPS). Besides those two major types of SMs there are terpenoids, alkaloids, and post-translationally modified peptides. PKSs are structurally and functionally diverse. They are categorized in three types –

PKS I, PKS II, and PKS III based on their structure and enzymatic function (Chan et al. 2009; Cox and Simpson 2009). In fungi, pathways involving PKS are most frequently encountered (87% of all BGCs), but only two types of PKS enzymes are usually encountered in fungi (PKS I and PKS III); PKS II so far has only been reported from bacteria (Jenke-Kodama et al. 2005). PKS-I is the best studied enzyme involved in SM production, and it encodes for a large multidomain protein involved in the production of many important SMs, PKS I is comprised of the beta-ketoacyl synthase (KS), acyl carrier protein (ACP) and acyl transferase (AT) domains. Some versions of PKS I may also contain domains for starter acyltransferase (SAT), thioesterase (TE), C-methyltransferase (MET), and product template domain (PT) etc. (Kroken et al. 2003; Cox 2007).

Discoveries of SMs from fungi traditionally involved chemical analysis from fungal extracts. Although traditional strategies resulted in the discovery of important SMs, the re-discovery of the same compounds in other fungi plus the advancement of chemical synthesis discouraged scientists and industries from working on screening fungi for SMs. Instead, the focus shifted on synthetic compound development rather than exploring natural sources for molecules. After a few decades of discovery void of new antibiotics, with the progress of genomics technologies, the screening organisms for SMs has become again become more fashionable. Screening for the potential production of different SMs combined with exploring genomic data. Dereplication strategies have made genomic mining an attractive alternative to traditional methods. By genome mining, it is possible to predict the possible production of valuable chemicals from a fungus, including revealing their cryptic genotype and potential to produce novel secondary metabolites. Many BGCs remain latent, their expression dependent on

specific conditions. Consequently, when fungi are cultivated in laboratory settings, they often fail to produce SMs. Therefore, genome mining can be a preliminary screening strategy for finding fungal strains that warrant further experimental and chemical analysis.

As stated previously, SM producing is dependent on multiple genes that can be arranged in clusters. These clusters of genes that are responsible for producing SMs, are referred to as biosynthetic gene clusters. Usually there are 2-20 genes that can make up a BGC. The genes of a BGC are grouped together and are in the same genomic region and they can be expressed together (Shwab and Keller 2008; Osbourn 2010). It has been proposed that the coordinated gene expression of BGCs provides a selective advantage postulated from the observation that genes encoding enzymes and regulatory factors involved in the production of SMs are often physically clustered together in fungal genomes. This organization enables their coordinated regulation, allowing fungi to rapidly and efficiently produce SMs in response to environmental stimuli (Khaldi and Wolfe 2011; Rokas et al. 2020). A typical BGC contains core genes that encode enzymes responsible for the primary steps in synthesizing the secondary metabolite. These core genes are often flanked by additional genes whose products play roles in supporting the biosynthetic process, such as regulating BGC gene expression, post-translational modifications of core enzymes, modification of the secondary metabolite, and facilitating cellular transport (Brown et al. 2012; Osbourn 2010).

The early studies on members of the Ophiostomatales producing secondary metabolites were performed from 1983 to 1989 by Ayer and coworkers. They extracted natural products from *Leptographium wageneri*, and *Leptographium* (=*Verticicladiella*) sp. (Ayer et al. 1989).

Recently, a study has been conducted on the Dutch-elm disease causing agent *Ophiostoma ulmi* and *O. novo-ulmi* and it was found that the fujikurin-like gene cluster (OpPKS8) is present in their genomes (Sbaraini et al. 2017). Apart from the search for industrially or medically important compounds, a few studies were conducted to elucidate the disease-causing genes and toxic compounds produced by members of the Ophiostomatales (DiGuistini et al. 2011; Ibarra Caballero et al. 2020; Lah et al. 2017; Sayari 2018). Previous study in search of novel antimicrobial compounds from *Leptographium* spp. screened over 100 strains against multi-drug-resistant *Staphylococcus aureus* (Ibrahim 2020). That study found 13 *Leptographium* strains showed antimicrobial activity against resistant bacteria. The current study is exploring the genomes of strains of *Leptographium wingfieldii* and related fungi that were previously found to be potential producers of medically important small molecules.

4.3 Material and Methods

4.3.1 Fungal strains and their culture

Leptographium wingfieldii and related fungal species collected from forests in Canada and were stored at 4℃ on malt extract agar slants (MEA, 30 g malt extract (VWR life science, Ohio, USA), 1 g yeast extract (Bacto™, Dickinson & Co., Maryland, USA) and 20 g bacteriological agar (Bacto™ BD) per liter) in the WIN collection (Department of Microbiology, University of Manitoba, Winnipeg, MB, Canada). Ten fungi including five strains of *Leptographium wingfieldii*, two *Leptographium* sp. (species not determined), one strain of *Leptographium procerum*, one strain of *Grosmannia* (=*Leptographium*) *aureum*, and one strain of *Penicillium citrinum* were selected for the current study based on previous work suggesting the production of secondary metabolites with antimicrobial activity. The fungal strains were inoculated and grown on MEA plates at room temperature (23ºC) for 7 to 10 days. The mycelia were transferred to 100 mL ME broth (30 g malt extract and 1 g of yeast extract (Bacto™ BD) per liter. The fungi were harvested from the broth after 7 to 10 days of growth at room temperature. These cultures were used for DNA extraction. For long term preservation, the fungi were cultured on 10 mL MEA slants and submitted to a culture collection facility (Table 4.1).

4.3.2 DNA extraction and genome sequencing

Fungal mycelia were harvested by vacuum filtration from the ME broth. DNA were extracted from those mycelia by using Cetyltrimethylammonium bromide (CTAB) buffer [2%

CTAB (Fisher Scientific), 100 mM Tris-HCl, pH 8 (Fisher Scientific), 20 mM Ethylenediaminetetraacetic acid (EDTA) (Fisher Scientific), 1.4 M NaCl] followed by chloroform-extraction and ethanol precipitation (as previously described in Hausner et al. 1992). Extracted nucleic acids were dissolved in sterile water and treated with RNase and the DNA solutions were finally transferred into 1.5 mL microcentrifuge tubes. DNA concentrations were quantified with a NanoDrop and 100 µl (30 ng/µl) of DNA were shipped to MicrobesNG at the Birmingham University (UK) sequencing facility (Illumina MiSeq platform). The resulting sequence reads were assembled (using the tools provided in GALAXY) and genome completeness was assessed using the BUSCO program (Manni et al. 2015).

4.3.3 Taxonomic identification

Five different nuclear markers (ITS2, ACT, CAL, TEF, B-TUB) were used to identify and confirm the taxonomic identity of the studied fungi (Strzałka et al. 2020). Sequences of these marker genes were searched for by using *Leptographium wingfieldii* WIN1181 sequence as queries with blastn in NCBI GenBank database. Similar sequences were downloaded and sorted. The genome sequences (contigs) were searched using the NCBI blast+ program as implemented in the Galaxy platform and copies of the five different marker genes were extracted from all genomes used in this study. Each partial gene set was aligned with the MAFFT program (Katoh and Standley 2013), and the alignments were analyzed with MrBayes (Ronquist et al. 2012) program (using GTR substitution model and 500 sample frequency running for a million generations with 25% burnin of samples) for establishing phylogenetic relationships. The phylogenetic trees for five different marker genes were used for the identification of the

unknown strains of *Leptographium* and to confirm the taxonomic identity of the *Leptographium wingfieldii*.

4.3.4 Identification and characterization of BGS from the draft genomes

Illumina sequencing generated paired end-reads that were assembled into contigs by the SPAdes program (Bankevich et al. 2012). The genomic reads were submitted into the NCBI SRA database (Table 4.1). The assembled contigs were uploaded onto the antiSMASH program (Blin et al. 2023) with strict and relaxed settings for each of the genomes to find all strong matches and more distant matches. The previously known BGCs from databases with strict search were identified and their potential products' structures were obtained from ChemSpider (https://www.chemspider.com). The resulting unknown "hits"/matched genes of a BGC were subjected to the BLAST program to find their potential identity and putative gene function. The BiG-SCAPE program (Navarro-Muñoz et al. 2020) was used to perform a comparative analysis to categorize the BGCs into clusters using all the antiSMASH results as the input data.

4.4 Results

4.4.1 Genome sequencing and completeness

Genome sequences of seven strains of *Leptographium wingfiedii*, a strain of *L. procerum*, a strain of *L. aureum*, and a strain of *Penicillium citrinum* were sequenced and the reads were further assembled to generate contigs. From the contigs, the genome size of *L. wingfieldii* and related fungi were determined (Table 4.1). The genomic contigs were assessed for completeness by estimating the presence of complete genomic information using the BUSCO program. The analysis was based on reference genomes from the corresponding taxonomic lineage (Sordariomycetes for *Leptographium*, and Eurotiales for *Penicillium*). All genomes yielded BUSCO scores ranging from 95% to 97.5%, indicating that the genomes are 95% to 97.5% complete with respect to the expected set of single-copy orthologous genes (Figure 4.1, Table 4.1).

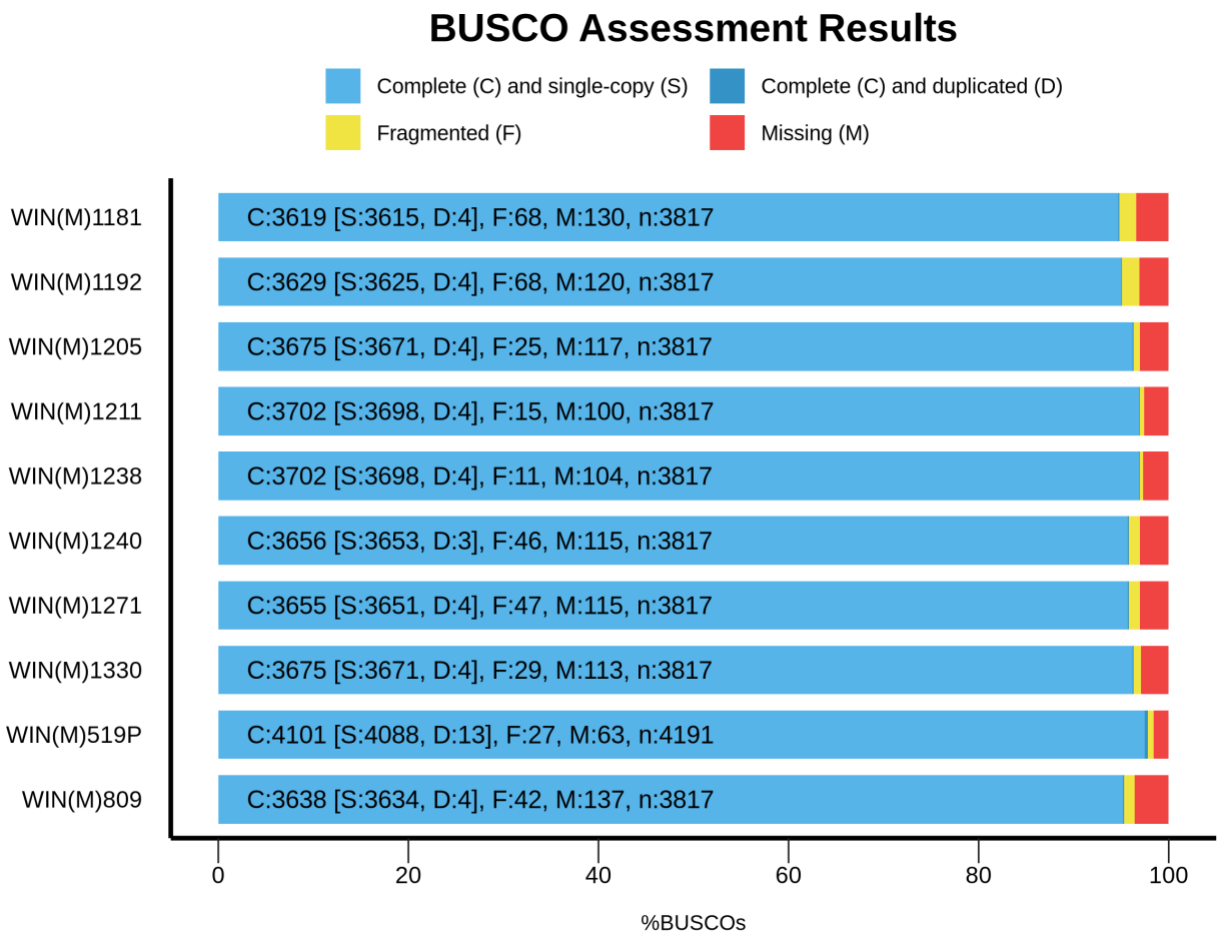

BUSCO Assessment Results
Complete (C) and single-copy (S)
Complete (C) and duplicated (D)
Fragmented (F)
Missing (M)
WIN(M)1181
C:3619 [S:3615, D:4], F:68, M:130, n:3817
WIN(M)1192
C:3629 [S:3625, D:4], F:68, M:120, n:3817
WIN(M)1205
C:3675 [S:3671, D:4], F:25, M:117, n:3817
WIN(M)1211
C:3702 [S:3698, D:4], F:15, M:100, n:3817
WIN(M)1238
C:3702 [S:3698, D:4], F:11, M:104, n:3817
WIN(M)1240
C:3656 [S:3653, D:3], F:46, M:115, n:3817
WIN(M)1271
C:3655 [S:3651, D:4], F:47, M:115, n:3817
WIN(M)1330
C:3675 [S:3671, D:4], F:29, M:113, n:3817
WIN(M)519P
C:4101 [S:4088, D:13], F:27, M:63, n:4191
WIN(M)809
C:3638 [S:3634, D:4], F:42, M:137, n:3817
0
20
40
60
80
100
%BUSCOs

Figure 4.1: Genome completeness results obtained from the BUSCO program showing the number of single copy orthologs that are complete (C), present as a single copy (S), fragmented (F), duplicated (D) and missing (M). The number of duplicated orthologs was very low suggesting the was little or no contamination of these genomes with other DNAs.

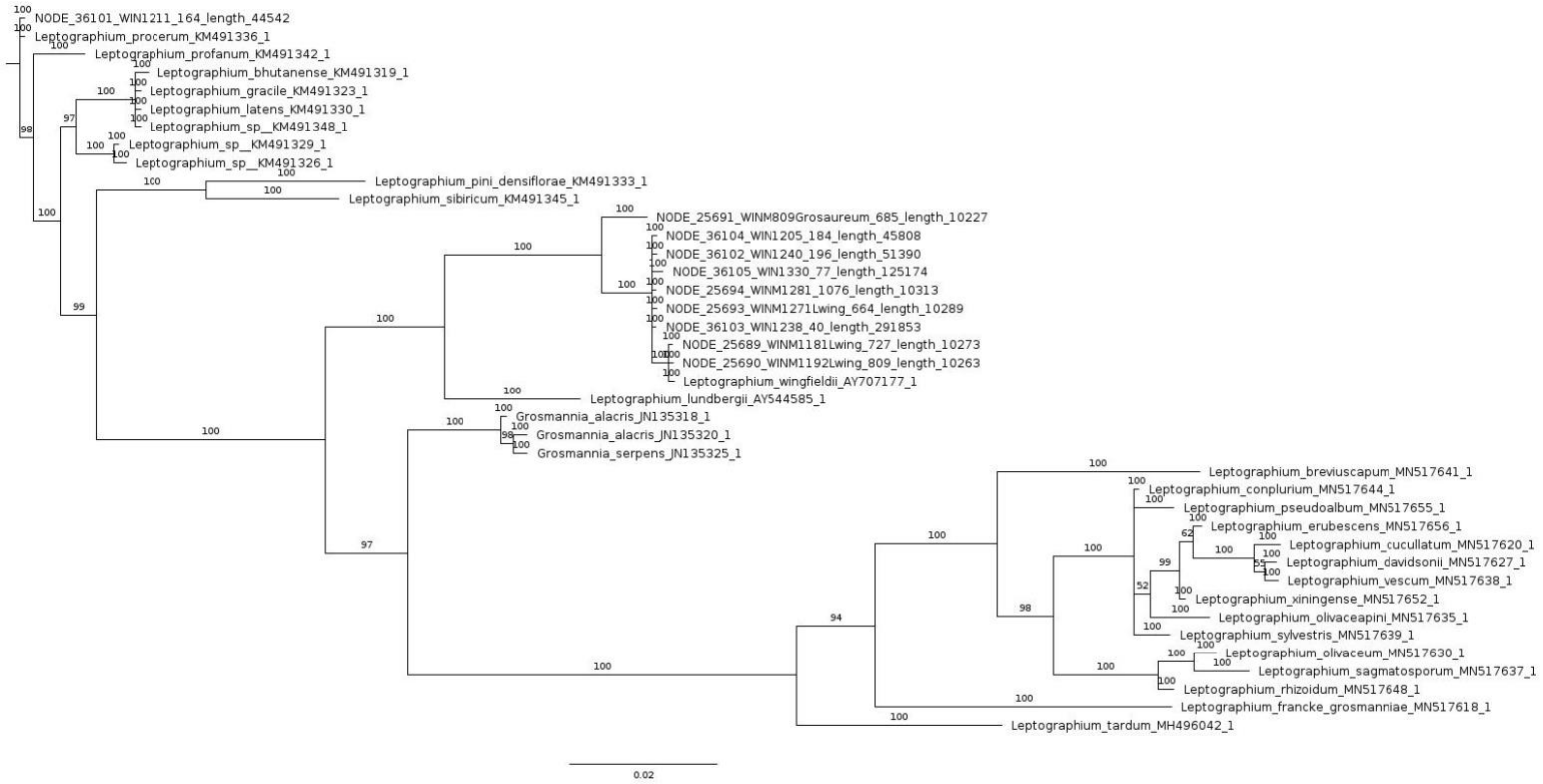
Table 4.1: List of fungi sequenced, their NCBI SRA accession, genome size, and their genome completeness scores (based on single copy orthologs) from BUSCO analysis.

Species	**Strain**	**NCBI SRA Accession**	**Genome Size (Mb)**	**BUSCO Genome completeness score (%)**			
				Complete and single copy	**Complete and duplicated**	**Fragmented**	**Missing**
L. wingfieldii	WIN(M)1181	SRR31369142	32.2	94.7	0.1	1.8	3.4
L. wingfieldii	WIN(M)1192	SRR29915607	31.9	95.0	0.1	1.8	3.1
L. wingfieldii	WIN(M)1271	SRR30014474	30.0	95.7	0.1	1.2	3.0
L. wingfieldii	WIN(M)1240	SRR30334524	32.1	95.7	0.1	1.2	3.0
L. wingfieldii	WIN(M)1238	SRR30606179	32.1	96.9	0.1	0.3	2.7
L. wingfieldii	WIN(M)1205	SRR30606355	32.1	96.2	0.1	0.7	3.1
L. wingfieldii	WIN(M)1330	SRR30675533	32.4	96.2	0.1	0.8	3.0
L. procerum	WIN(M)1211	SRR31364221	29.5	96.9	0.1	0.4	2.6
L. aureum	WIN(M)809	SRR31365496	31.1	95.2	0.1	1.1	3.6
P. citrinum	WIN(M)519P	SRR31383536	34.5	97.5	0.3	0.6	1.5

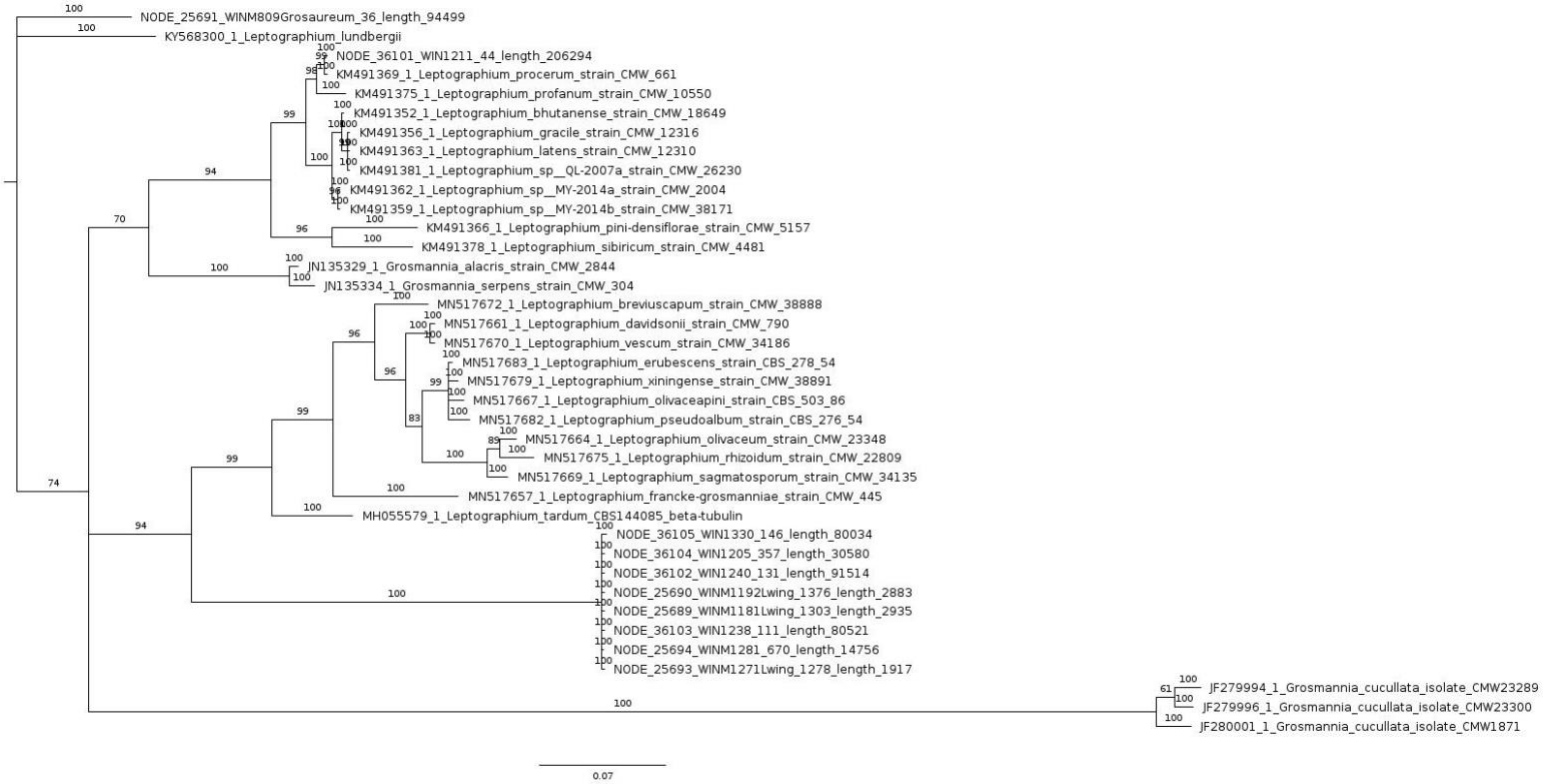
4.4.2 Taxonomic identification of Leptographium sp.

Two strains (WIN(M)1205 and WIN(M)1330) of *Leptographium* were unknown with regards to species identity. It was also necessary to confirm the taxonomic identity of the other strains used in this study. For Ophiostomatales species designations can be challenging due to lack of morphological characters and convergent evolution (de Beer et al. 2022). The phylogenetic trees of the five marker genes confirmed the identities of WIN(M)1205 and WIN(M)1330, as being *Leptographium wingfieldii*. These strains of *L. sp.* grouped among the other strains of *L. wingfieldii* from our collection and *L. wingfieldii* sequences collected from the NCBI GenBank database. Five different trees based on different marker gene sequences show similar results, confirming the identities of the species used in this study (Figure 4.2).

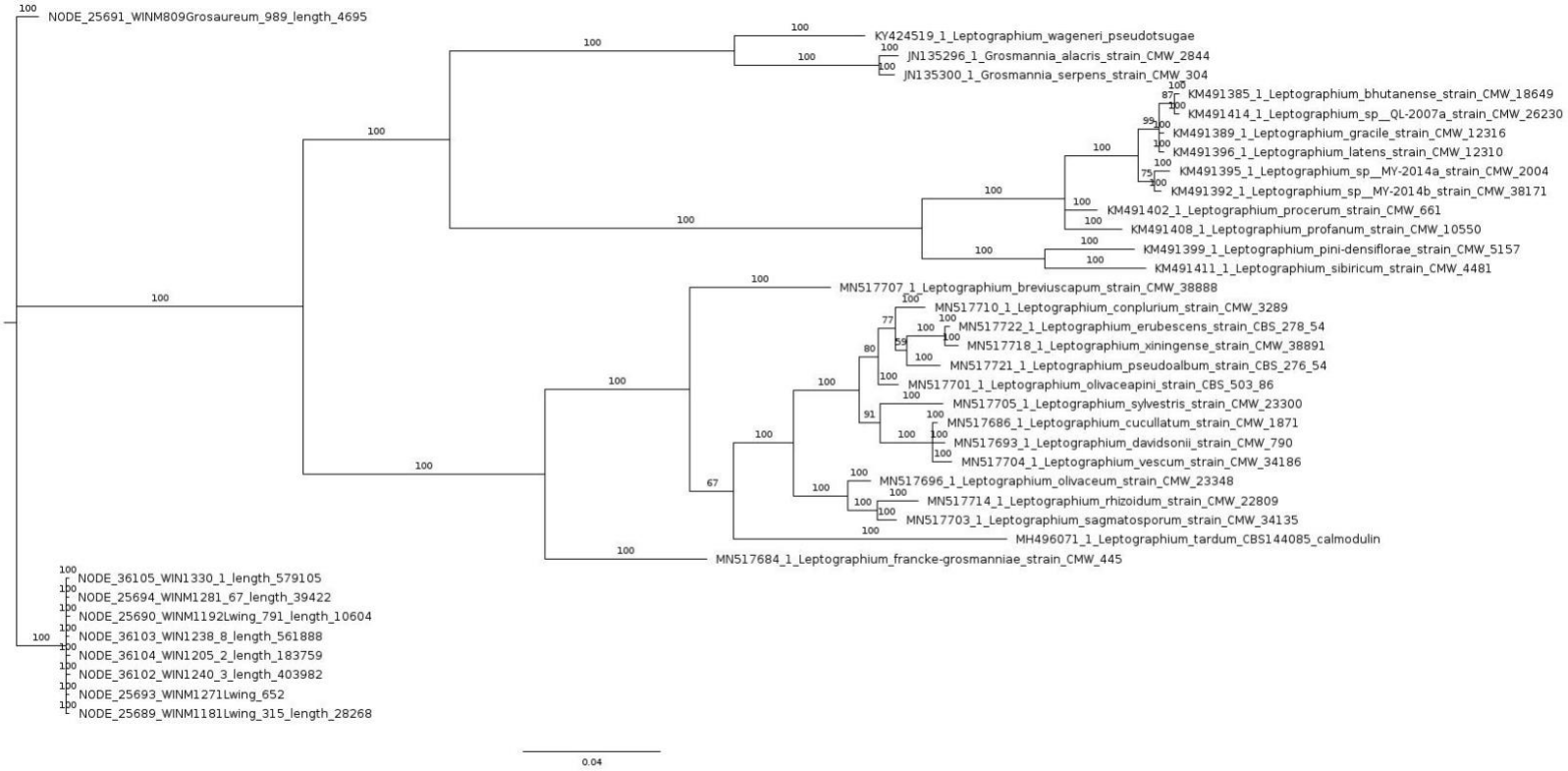
(A)


NODE_36101_WIN1211_164_length_44542
Leptographium_procerum_KM491336_1
Leptographium_profanum_KM491342_1
Leptographium_bhutanense_KM491319_1
Leptographium_gracile_KM491323_1
Leptographium_latens_KM491330_1
Leptographium_sp__KM491348_1
Leptographium_sp__KM491329_1
Leptographium_sp__KM491326_1
Leptographium_pini_densiflorae_KM491333_1
Leptographium_sibiricum_KM491345_1
NODE_25691_WINM809Grosaureum_685_length_10227
NODE_36104_WIN1205_184_length_45808
NODE_36102_WIN1240_196_length_51390
NODE_36105_WIN1330_77_length_125174
NODE_25694_WINM1281_1076_length_10313
NODE_25693_WINM1271Lwing_664_length_10289
NODE_36103_WIN1238_40_length_291853
NODE_25689_WINM1181Lwing_727_length_10273
NODE_25690_WINM1192Lwing_809_length_10263
Leptographium_wingfieldii_AY707177_1
Leptographium_lundbergii_AY544585_1
Grosmannia_alacris_JN135318_1
Grosmannia_alacris_JN135320_1
Grosmannia_serpens_JN135325_1
Leptographium_breviuscapum_MN517641_1
Leptographium_conplurium_MN517644_1
Leptographium_pseudoalbum_MN517655_1
Leptographium_erubescens_MN517656_1
Leptographium_cucullatum_MN517620_1
Leptographium_davidsonii_MN517627_1
Leptographium_vescum_MN517638_1
Leptographium_xiningense_MN517652_1
Leptographium_olivaceapini_MN517635_1
Leptographium_sylvestris_MN517639_1
Leptographium_olivaceum_MN517630_1
Leptographium_sagmatosporum_MN517637_1
Leptographium_rhizoidum_MN517648_1
Leptographium_francke_grosmanniae_MN517618_1
Leptographium_tardum_MH496042_1
0.02

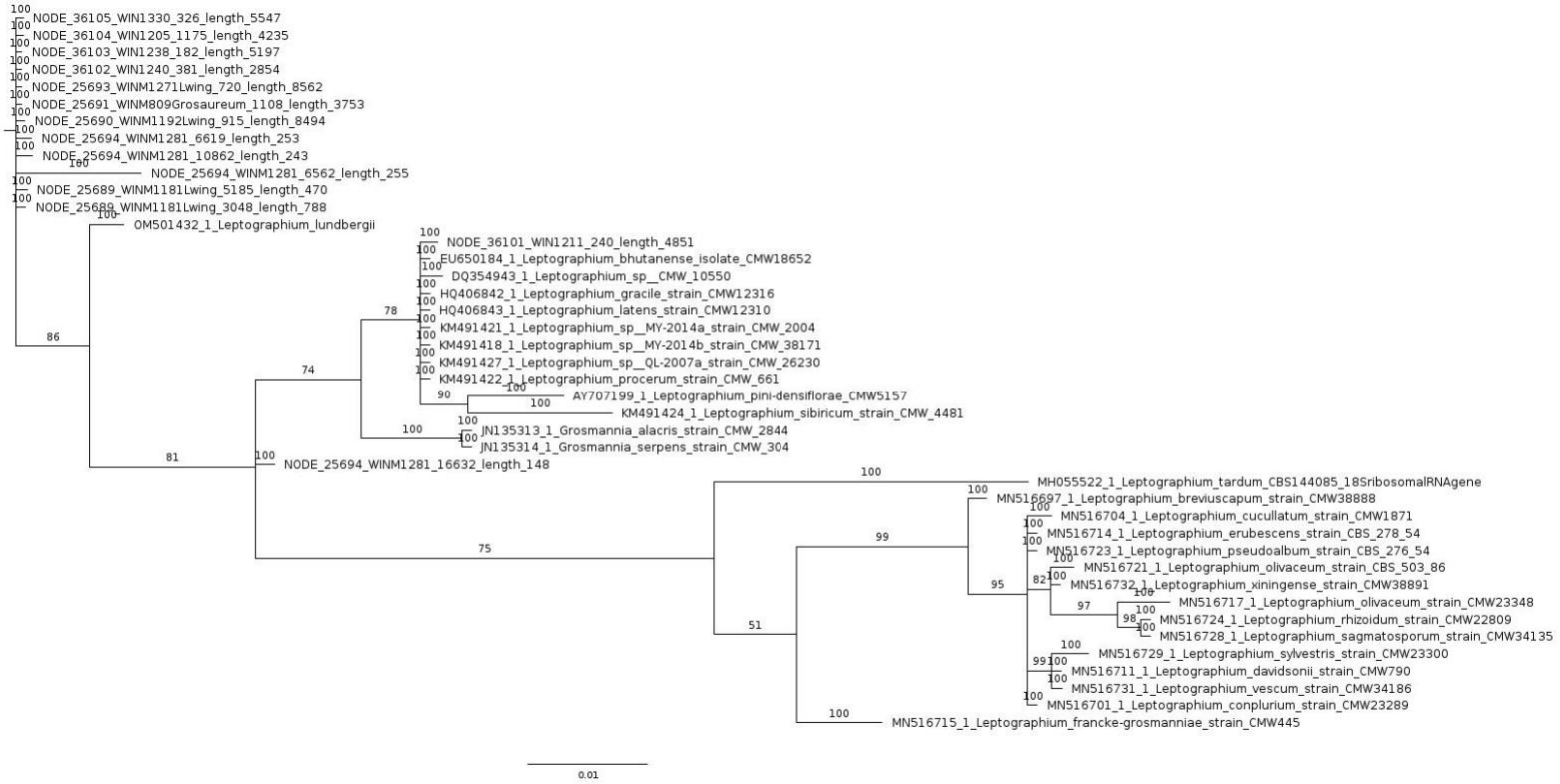
(B)


NODE_25691_WINM809Grosaureum_36_length_94499
KY568300_1_Leptographium_lundbergii
NODE_36101_WIN1211_44_length_206294
KM491369_1_Leptographium_procerum_strain_CMW_661
KM491375_1_Leptographium_profanum_strain_CMW_10550
KM491352_1_Leptographium_bhutanense_strain_CMW_18649
KM491356_1_Leptographium_gracile_strain_CMW_12316
KM491363_1_Leptographium_latens_strain_CMW_12310
KM491381_1_Leptographium_sp__QL-2007a_strain_CMW_26230
KM491362_1_Leptographium_sp__MY-2014a_strain_CMW_2004
KM491359_1_Leptographium_sp__MY-2014b_strain_CMW_38171
KM491366_1_Leptographium_pini-densiflorae_strain_CMW_5157
KM491378_1_Leptographium_sibiricum_strain_CMW_4481
JN135329_1_Grosmannia_alacris_strain_CMW_2844
JN135334_1_Grosmannia_serpens_strain_CMW_304
MN517672_1_Leptographium_breviuscapum_strain_CMW_38888
MN517661_1_Leptographium_davidsonii_strain_CMW_790
MN517670_1_Leptographium_vescum_strain_CMW_34186
MN517683_1_Leptographium_erubescens_strain_CBS_278_54
MN517679_1_Leptographium_xiningense_strain_CMW_38891
MN517667_1_Leptographium_olivaceapini_strain_CBS_503_86
MN517682_1_Leptographium_pseudoalbum_strain_CBS_276_54
MN517664_1_Leptographium_olivaceum_strain_CMW_23348
MN517675_1_Leptographium_rhizoidum_strain_CMW_22809
MN517669_1_Leptographium_sagmatosporum_strain_CMW_34135
MN517657_1_Leptographium_francke-grosmanniae_strain_CMW_445
MH055579_1_Leptographium_tardum_CBS144085_beta-tubulin
NODE_36105_WIN1330_146_length_80034
NODE_36104_WIN1205_357_length_30580
NODE_36102_WIN1240_131_length_91514
NODE_25690_WINM1192Lwing_1376_length_2883
NODE_25689_WINM1181Lwing_1303_length_2935
NODE_36103_WIN1238_111_length_80521
NODE_25694_WINM1281_670_length_14756
NODE_25693_WINM1271Lwing_1278_length_1917
JF279994_1_Grosmannia_cucullata_isolate_CMW23289
JF279996_1_Grosmannia_cucullata_isolate_CMW23300
JF280001_1_Grosmannia_cucullata_isolate_CMW1871
0.07

(C)


NODE_25691_WINM809Grosaureum_989_length_4695
KY424519_1_Leptographium_wageneri_pseudotsugae
JN135296_1_Grosmannia_alacris_strain_CMW_2844
JN135300_1_Grosmannia_serpens_strain_CMW_304
KM491385_1_Leptographium_bhutanense_strain_CMW_18649
KM491414_1_Leptographium_sp__QL-2007a_strain_CMW_26230
KM491389_1_Leptographium_gracile_strain_CMW_12316
KM491396_1_Leptographium_latens_strain_CMW_12310
KM491395_1_Leptographium_sp__MY-2014a_strain_CMW_2004
KM491392_1_Leptographium_sp__MY-2014b_strain_CMW_38171
KM491402_1_Leptographium_procerum_strain_CMW_661
KM491408_1_Leptographium_profanum_strain_CMW_10550
KM491399_1_Leptographium_pini-densiflorae_strain_CMW_5157
KM491411_1_Leptographium_sibiricum_strain_CMW_4481
MN517707_1_Leptographium_breviuscapum_strain_CMW_38888
MN517710_1_Leptographium_conplurium_strain_CMW_3289
MN517722_1_Leptographium_erubescens_strain_CBS_278_54
MN517718_1_Leptographium_xiningense_strain_CMW_38891
MN517721_1_Leptographium_pseudoalbum_strain_CBS_276_54
MN517701_1_Leptographium_olivaceapini_strain_CBS_503_86
MN517705_1_Leptographium_sylvestris_strain_CMW_23300
MN517686_1_Leptographium_cucullatum_strain_CMW_1871
MN517693_1_Leptographium_davidsonii_strain_CMW_790
MN517704_1_Leptographium_vescum_strain_CMW_34186
MN517696_1_Leptographium_olivaceum_strain_CMW_23348
MN517714_1_Leptographium_rhizoidum_strain_CMW_22809
MN517703_1_Leptographium_sagmatosporum_strain_CMW_34135
MH496071_1_Leptographium_tardum_CBS144085_calmodulin
MN517684_1_Leptographium_francke-grosmanniae_strain_CMW_445
NODE_36105_WIN1330_1_length_579105
NODE_25694_WINM1281_67_length_39422
NODE_25690_WINM1192Lwing_791_length_10604
NODE_36103_WIN1238_8_length_561888
NODE_36104_WIN1205_2_length_183759
NODE_36102_WIN1240_3_length_403982
NODE_25693_WINM1271Lwing_652
NODE_25689_WINM1181Lwing_315_length_28268
0.04

(D)


NODE_36105_WIN1330_326_length_5547
NODE_36104_WIN1205_1175_length_4235
NODE_36103_WIN1238_182_length_5197
NODE_36102_WIN1240_381_length_2854
NODE_25693_WINM1271Lwing_720_length_8562
NODE_25691_WINM809Grosaureum_1108_length_3753
NODE_25690_WINM1192Lwing_915_length_8494
NODE_25694_WINM1281_6619_length_253
NODE_25694_WINM1281_10862_length_243
NODE_25694_WINM1281_6562_length_255
NODE_25689_WINM1181Lwing_5185_length_470
NODE_25689_WINM1181Lwing_3048_length_788
OM501432_1_Leptographium_lundbergii
NODE_36101_WIN1211_240_length_4851
EU650184_1_Leptographium_bhutanense_isolate_CMW18652
DQ354943_1_Leptographium_sp__CMW_10550
HQ406842_1_Leptographium_gracile_strain_CMW12316
HQ406843_1_Leptographium_latens_strain_CMW12310
KM491421_1_Leptographium_sp__MY-2014a_strain_CMW_2004
KM491418_1_Leptographium_sp__MY-2014b_strain_CMW_38171
KM491427_1_Leptographium_sp__QL-2007a_strain_CMW_26230
KM491422_1_Leptographium_procerum_strain_CMW_661
AY707199_1_Leptographium_pini-densiflorae_CMW5157
KM491424_1_Leptographium_sibiricum_strain_CMW_4481
JN135313_1_Grosmannia_alacris_strain_CMW_2844
JN135314_1_Grosmannia_serpens_strain_CMW_304
NODE_25694_WINM1281_16632_length_148
MH055522_1_Leptographium_tardum_CBS144085_18SribosomalRNAgene
MN516697_1_Leptographium_breviuscapum_strain_CMW38888
MN516704_1_Leptographium_cucullatum_strain_CMW1871
MN516714_1_Leptographium_erubescens_strain_CBS_278_54
MN516723_1_Leptographium_pseudoalbum_strain_CBS_276_54
MN516721_1_Leptographium_olivaceum_strain_CBS_503_86
MN516732_1_Leptographium_xiningense_strain_CMW38891
MN516717_1_Leptographium_olivaceum_strain_CMW23348
MN516724_1_Leptographium_rhizoidum_strain_CMW22809
MN516728_1_Leptographium_sagmatosporum_strain_CMW34135
MN516729_1_Leptographium_sylvestris_strain_CMW23300
MN516711_1_Leptographium_davidsonii_strain_CMW790
MN516731_1_Leptographium_vescum_strain_CMW34186
MN516701_1_Leptographium_conplurium_strain_CMW23289
MN516715_1_Leptographium_francke-grosmanniae_strain_CMW445
0.01

(E)

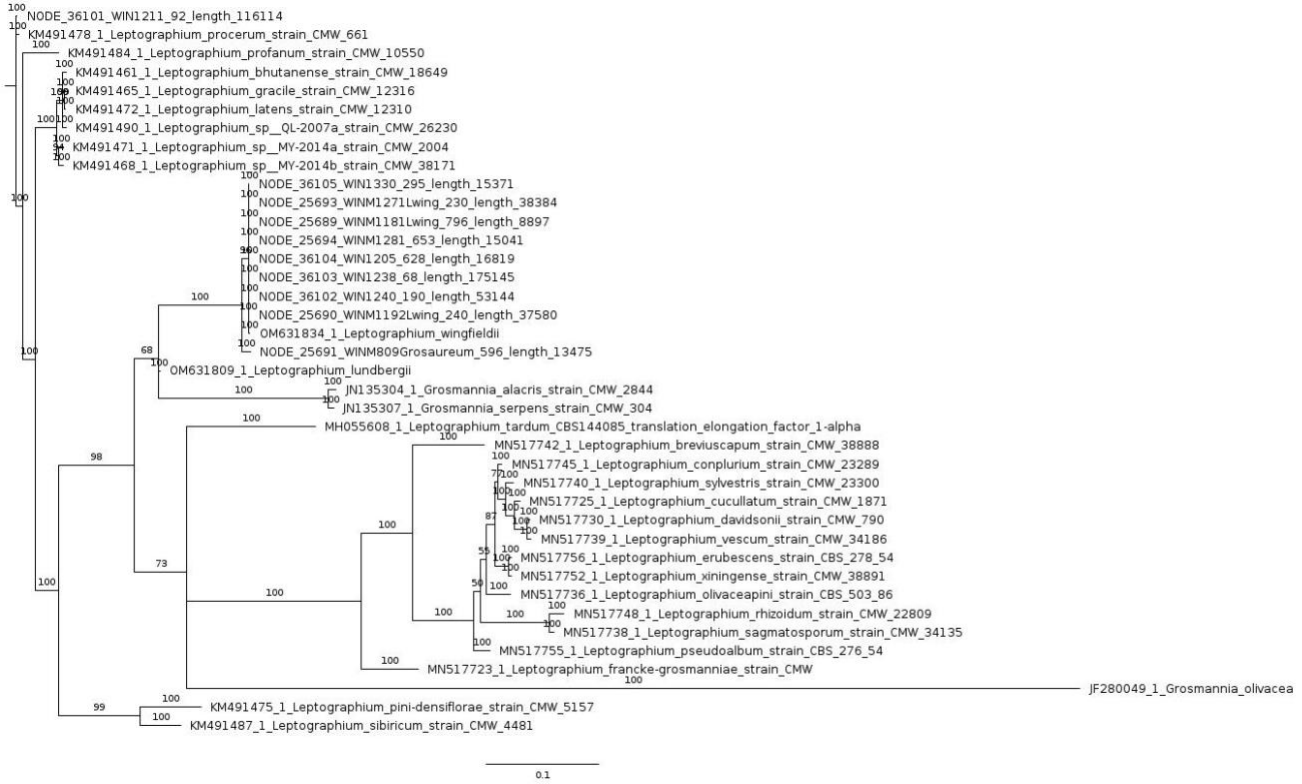

NODE_36101_WIN1211_92_length_116114
KM491478_1_Leptographium_procerum_strain_CMW_661
KM491484_1_Leptographium_profanum_strain_CMW_10550
KM491461_1_Leptographium_bhutanense_strain_CMW_18649
KM491465_1_Leptographium_gracile_strain_CMW_12316
KM491472_1_Leptographium_latens_strain_CMW_12310
KM491490_1_Leptographium_sp__QL-2007a_strain_CMW_26230
KM491471_1_Leptographium_sp__MY-2014a_strain_CMW_2004
KM491468_1_Leptographium_sp__MY-2014b_strain_CMW_38171
NODE_36105_WIN1330_295_length_15371
NODE_25693_WINM1271Lwing_230_length_38384
NODE_25689_WINM1181Lwing_796_length_8897
NODE_25694_WINM1281_653_length_15041
NODE_36104_WIN1205_628_length_16819
NODE_36103_WIN1238_68_length_175145
NODE_36102_WIN1240_190_length_53144
NODE_25690_WINM1192Lwing_240_length_37580
OM631834_1_Leptographium_wingfieldii
NODE_25691_WINM809Grosaureum_596_length_13475
OM631809_1_Leptographium_lundbergii
JN135304_1_Grosmannia_alacris_strain_CMW_2844
JN135307_1_Grosmannia_serpens_strain_CMW_304
MH055608_1_Leptographium_tardum_CBS144085_translation_elongation_factor_1-alpha
MN517742_1_Leptographium_breviuscapum_strain_CMW_38888
MN517745_1_Leptographium_conplurium_strain_CMW_23289
MN517740_1_Leptographium_sylvestris_strain_CMW_23300
MN517725_1_Leptographium_cucullatum_strain_CMW_1871
MN517730_1_Leptographium_davidsonii_strain_CMW_790
MN517739_1_Leptographium_vescum_strain_CMW_34186
MN517756_1_Leptographium_erubescens_strain_CBS_278_54
MN517752_1_Leptographium_xiningense_strain_CMW_38891
MN517736_1_Leptographium_olivaceapini_strain_CBS_503_86
MN517748_1_Leptographium_rhizoidum_strain_CMW_22809
MN517738_1_Leptographium_sagmatosporum_strain_CMW_34135
MN517755_1_Leptographium_pseudoalbum_strain_CBS_276_54
MN517723_1_Leptographium_francke-grosmanniae_strain_CMW
JF280049_1_Grosmannia_olivacea
KM491475_1_Leptographium_pini-densiflorae_strain_CMW_5157
KM491487_1_Leptographium_sibiricum_strain_CMW_4481
0.1

Figure 4.2: Taxonomic position of the *Leptographium wingfieldii* and identification of the unknown strains inferred from the phylogenetic trees based on the applied marker genes/sequences (A) ACT: actin, (B) TUB2: beta-tubulin, (C) CAL: calmodulin (D) ITS: internal transcribed spacer (E)TEF1-α: translation elongation factor 1-alpha extracted from the sequenced genomes, and sequences collected from NCBI Genbank database. Trees show the position of *Leptographium wingfieldii* sequences and confirm the identities of the *Leptographium* spp. (WIN(M)1205 and WIN(M)1330). The tree topology (50% majority-rule consensus tree) was generated by the MrBayes program and visualized in the FigTree program (Rambaut 2014).

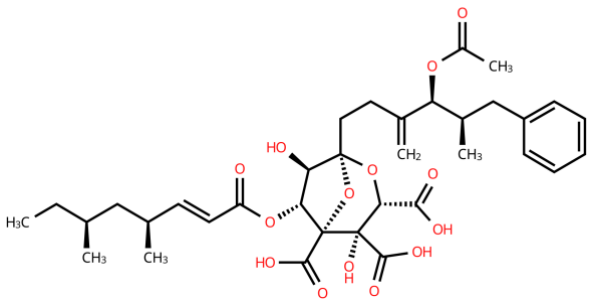
4.4.3 BGCs in Leptographium genomes

The BGCs were identified from the analysis with antiSMASH for every genome in our dataset, and all BGCs were further annotated and tabulated in the Supplementary Table 4.1. The antiSMASH search with strict setting for the genomes showed 13 to 16 BGCs. From each genome of *L. wingfieldii*, there were three strong matches with previously annotated BGCs by the antiSMASH search. The compounds are squalestatin S1 (Terpene), neosartorin (PKS), and naphthalene (PKS) which are present in *L. wingfieldii* strain WIN(M)1181 and WIN(M)1205. The other *L. wingfieldii* strains (WIN(M)1192, WIN(M)1271, WIN(M)1240, WIN(M)1238, WIN(M)1330) contain those three BGCs in addition there is an extra copy of neosartorin. *L. procerum* WIN(M)1211 has five known BGCs which are squalestatin (Terpene), clavaric acid (Terpene), shanorellin (PKS), bikaverin (PKS), and phyllostictine A (hybrid). *Leptographium* (=*Grosmannia*) *aureum* WIN(M)809 has eight known BGCs that are squalestatin (Terpene), clavaric acid (Terpene), dichlorodiaporthin (PKS), rosamicin (PKS), and scytalone (PKS). *Penicillium citrinum* WIN(M)519P has seven known BGCs, ucs1025a (hybrid), squalestatin (Terpene), equisetin (PKS), sorbicillin (PKS), naphthopyrone (PKS), and trypacidin (PKS). All those known compounds' structures are depicted in Figure 4.3. The gene arrangement of squalestatin and naphthalene in strain WIN(M)1181 and WIN(M)1205 were found to be identical. However, the gene arrangements of neosartorin BGCs do not share any similarities (Figure 4.3). All the strains have the core biosynthetic gene, additional biosynthetic genes, transport related genes, and other genes in their neosartorin BGCs, but the arrangement of those genes showed no pattern. Five strains of *L. wingfieldii* have a second copy of neosartorin, but the two copies from the same genome are completely different in gene arrangement. Only exception

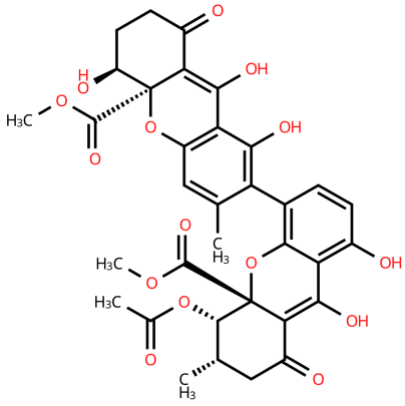
of this high variability in gene synteny is the WIN(M)1192 and WIN(M)1271 which have identical gene arrangement (Figure 4.3).

With the relaxed search setting in antiSMASH, members of *Leptographium* showed 19 to 25 BGCs in their genomes. The BGCs were further categorized and summarized based on their types such as PKS I, PKS III, NRPS, terpenes etc. A total of 205 BGCs were found among all the examined genomes of the study and the majority (86 BGCs) of the total BGCs are PKS I (Table 4.2). Those BGCs were further analyzed by the BiG-SCAPE program to establish their relationships with known BGC families from Pfam and MiBIG database. The 205 BGCs can be assigned into 97 BGC families based on BiG-SCAPE analysis and the program assigned similar BGCs into the same BGC families (Table 4.2). The comparative analysis also revealed the gene arrangement within the BGCs and represented BGC-clusters in cladograms, reflecting overall sequence similarities. In addition to the clustering, singletons were identified, which did not group with any clusters, indicating that these BGCs are unique to particular genomes. The BGCs clustered into 23 families, and 74 BGCs remained singletons. Clustered BGCs in each family showed similar gene synteny with minor variations (Table 4.3).

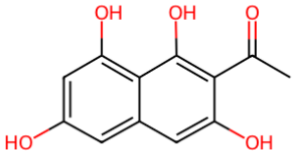
(A)



Squalestatin



Neosartorin



Naphthalene

(B)

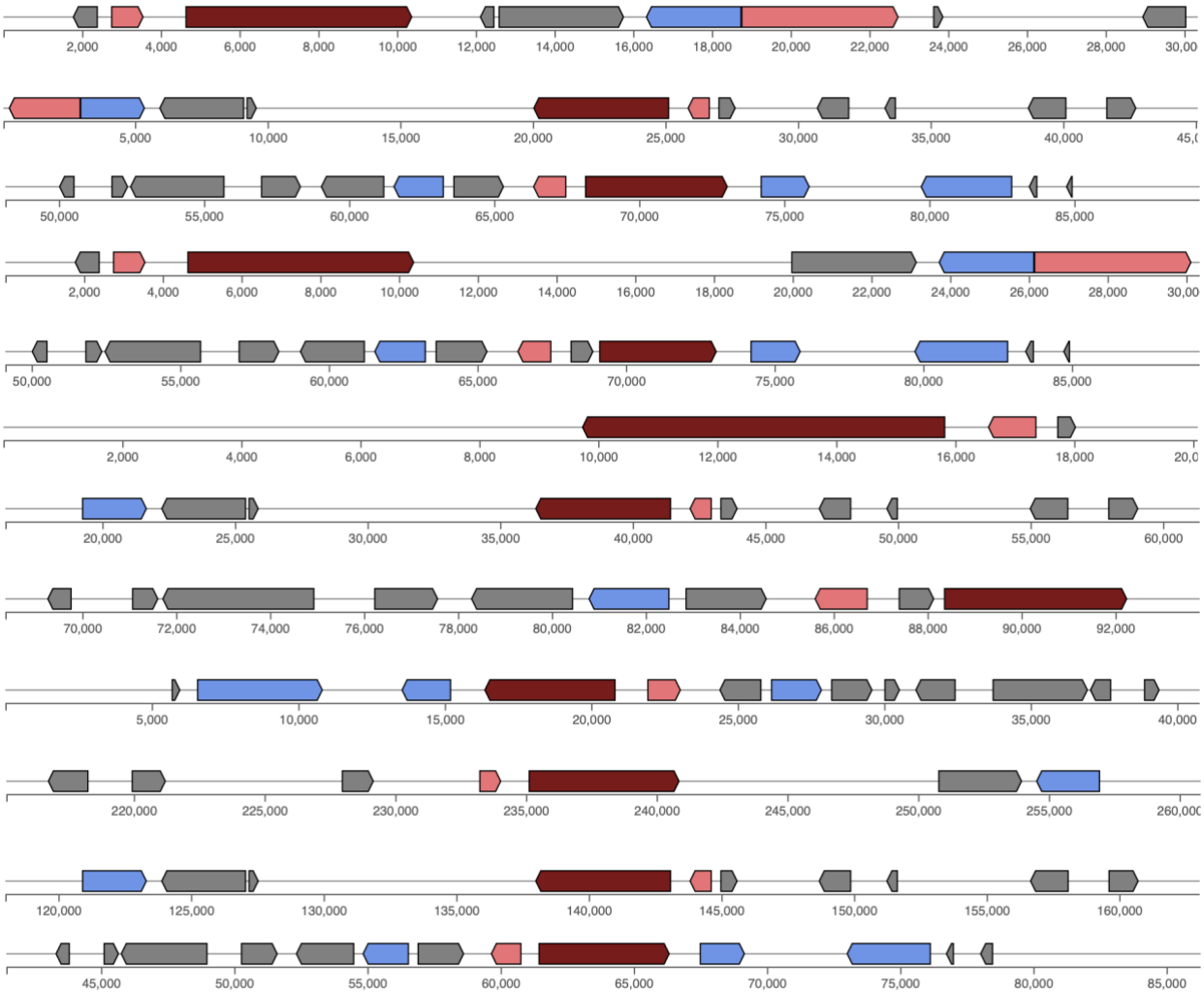
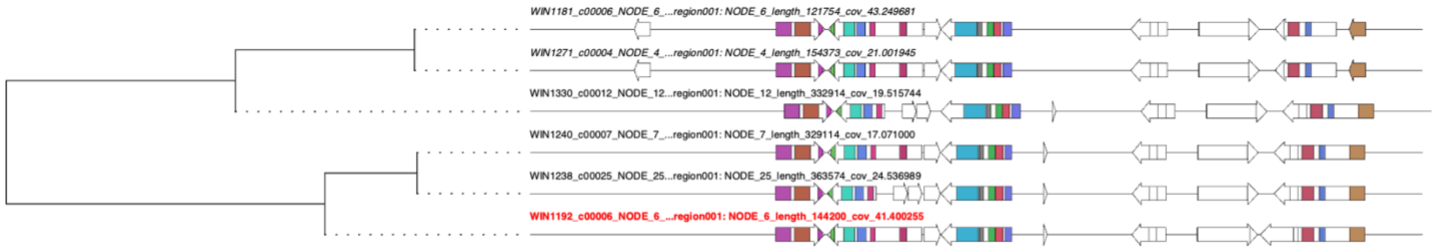


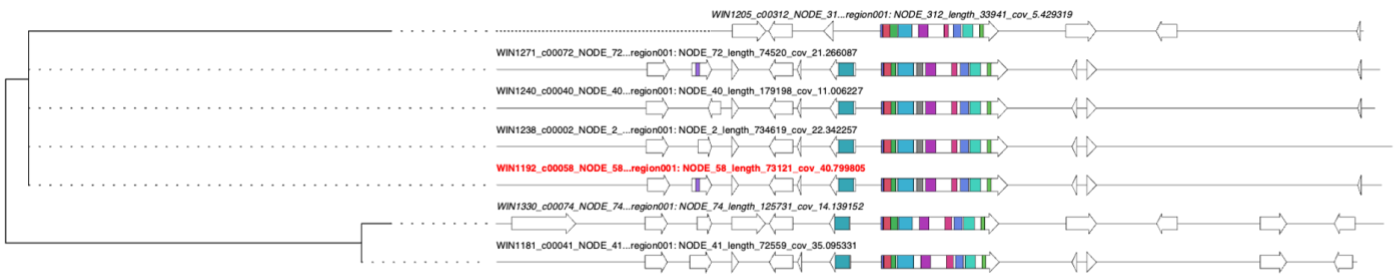
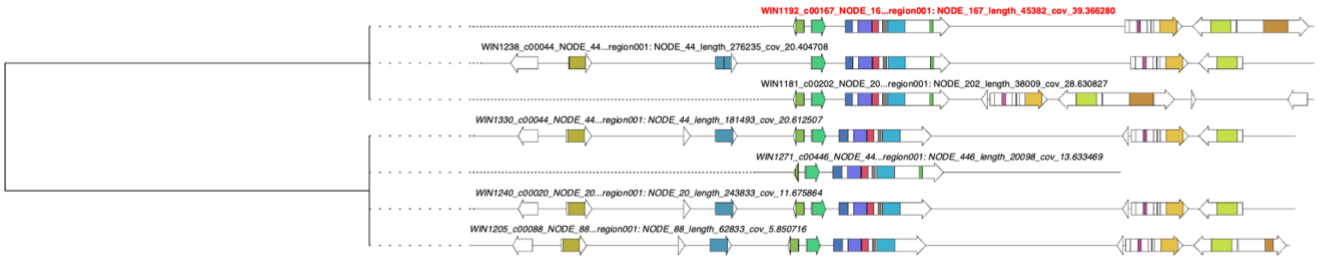
Figure 4.3: (A) The structures of small molecules potentially produced by previously known BGCs based on the antiSMASH database. (B) The neosartorin producing BGCs from Leptographium wingfieldii strains. The gene synteny in BGCs, the dark red color represents the core biosynthetic gene, light red represents additional biosynthetic gene, blue color represents transport-related gene, and gray color represents other genes related to the BGC.

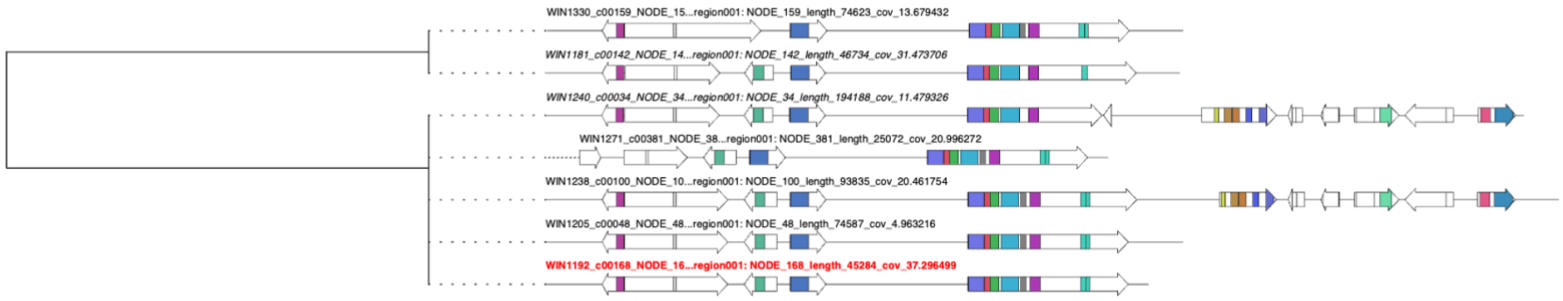
Table 4.2: List of BGCs in *Leptographium* species and related fungi found from antiSMASH analysis in strict and relaxed settings. The BGCs are also tabulated in categories such as PKS, NRPS, terpenes etc.

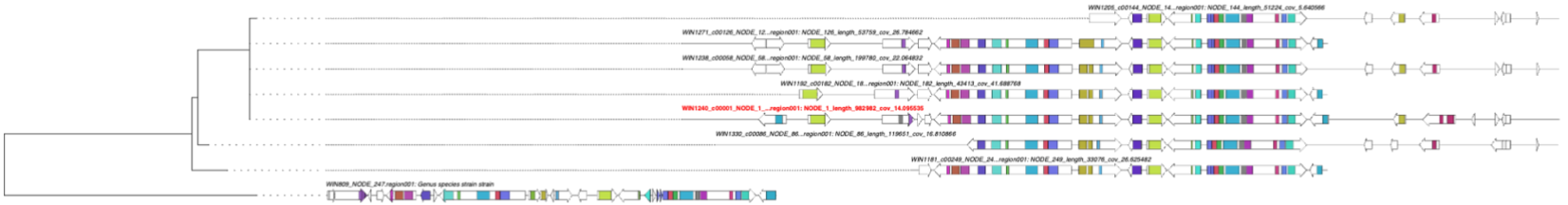
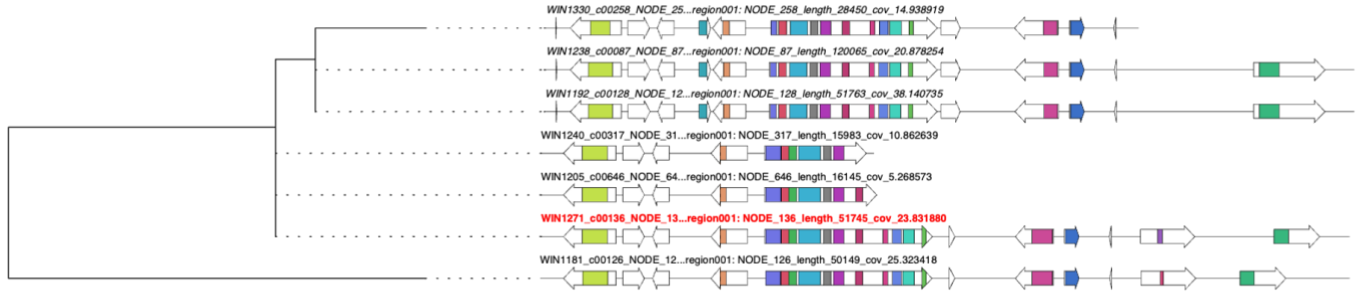
Species	**Strain**	**Number of BGCs (strict)**	**Number of BGCs (relaxed)**	**BGC numbers in categories**					
				PKS I	**PKS III**	**NRPS/ NRPS-like**	**Terpene**	**RiPP/RiPP-like**	**Hybrids**
L. wingfieldii	WIN(M)1181	14	19	9	1	7	2	0	0
L. wingfieldii	WIN(M)1192	16	20	10	1	6	2	0	1
L. wingfieldii	WIN(M)1271	15	20	10	1	7	2	0	0
L. wingfieldii	WIN(M)1240	16	20	10	1	6	2	0	1
L. wingfieldii	WIN(M)1238	15	19	9	1	6	2	0	1
L. wingfieldii	WIN(M)1205	14	18	9	1	6	2	0	0
L. wingfieldii	WIN(M)1330	15	20	10	1	7	2	0	0
L. procerum	WIN(M)1211	13	19	4	1	7	4	0	3
L. aureum	WIN(M)809	13	25	7	1	5	3	9	0
P. citrinum	WIN(M)519P	16	25	8	0	12	1	0	4
Total number of BGCs		**147**	**205**	**86**	**9**	**69**	**22**	**9**	**10**
Number of BGC Family matches			**97**	**28**	**9**	**33**	**9**	**11**	**7**

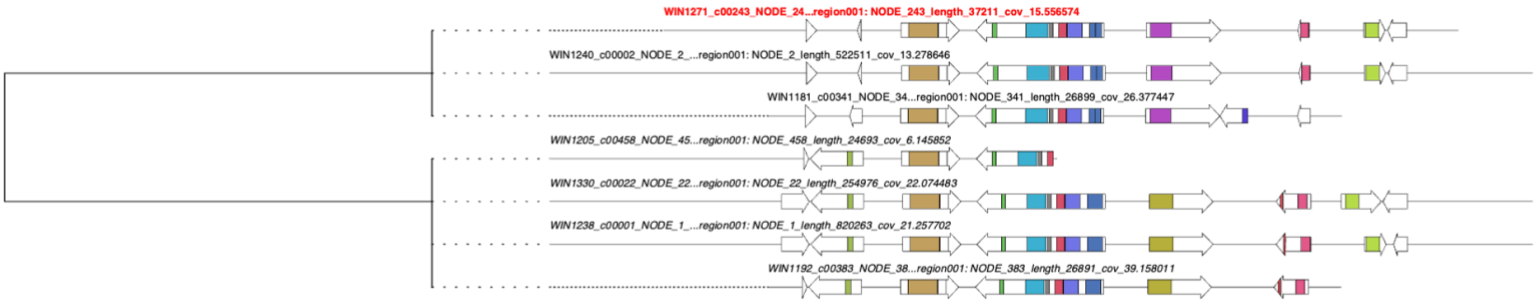
Table 4.3: The list of BGC clusters formed in BiG-SCAPE analysis for *Leptographium* species. BGC clusters are categorized according to the corresponding BGC family. Number of clusters and singletons belong to BGC types are tabulated followed by BGC families and clustered BGCs that are presented in cladograms to visualize the similarities and variations in gene arrangements in BGCs in different *Leptographium* genomes.

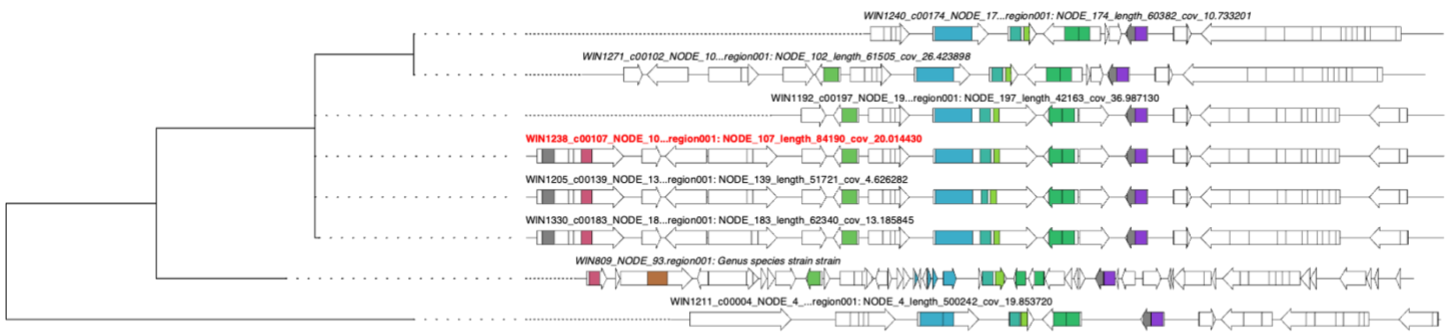
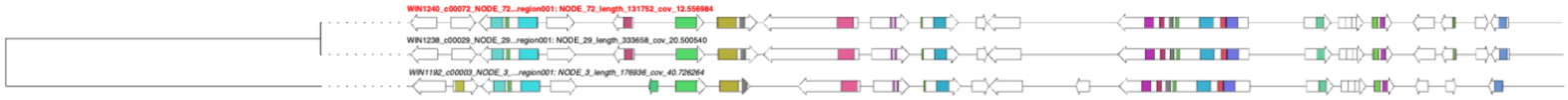
BGC Family	Type	Singleton	Clusters		
	PKS I	18	10		
FAM_01940					
FAM_01946					

	
FAM_01950	
FAM_01951	

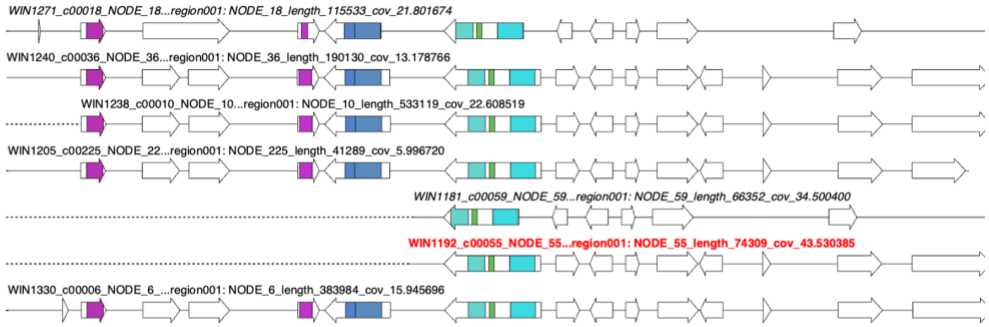
	
FAM_01981	
FAM_02009	

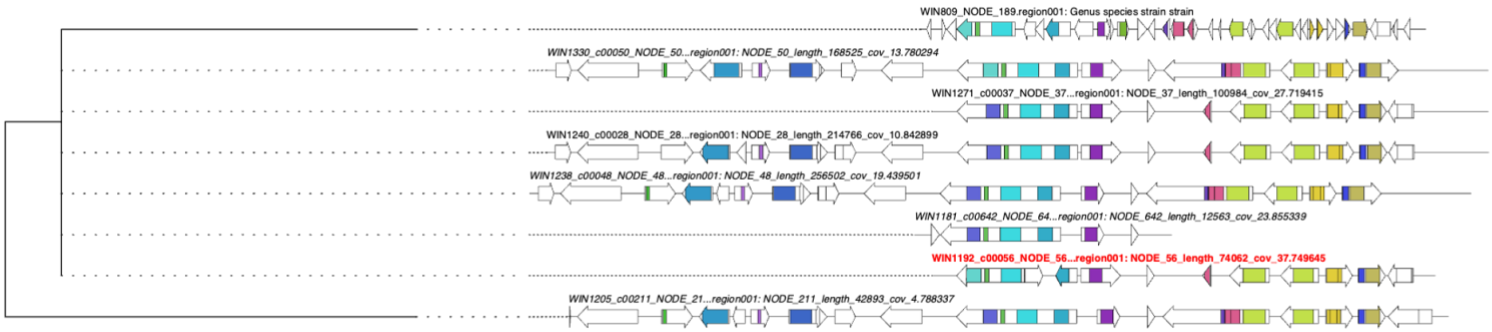
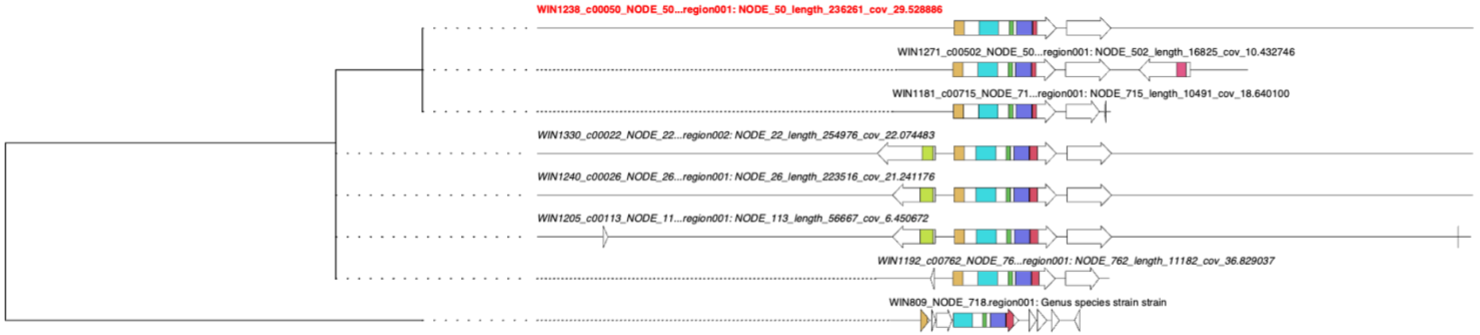
	
FAM_02015	
FAM_02045	

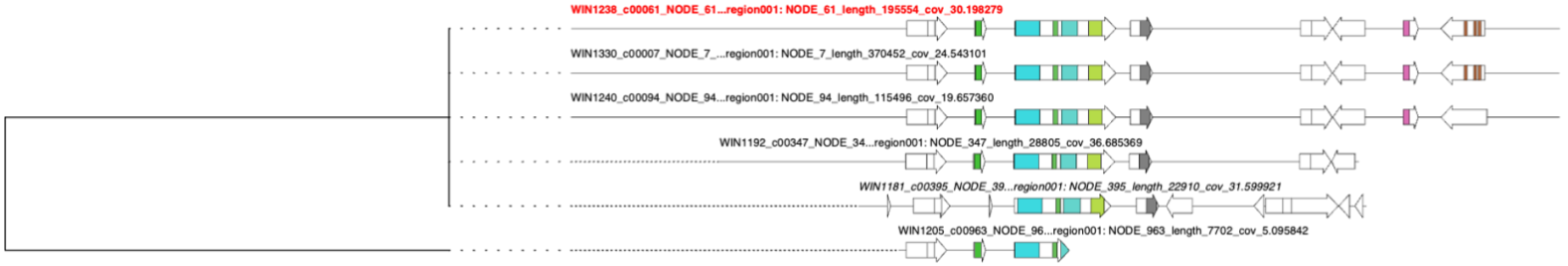
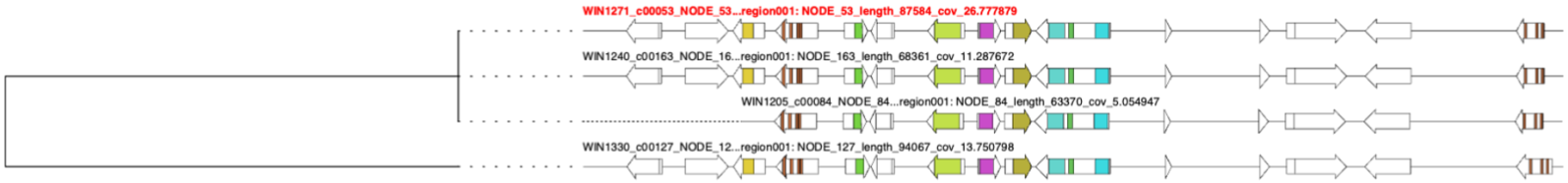
<table>
<tr><td>FAM_02046</td><td colspan="5"> 
</td></tr>
<tr><td>FAM_02047</td><td colspan="5"> 
</td></tr>
<tr><td></td><td>PKS III</td><td>7</td><td>2</td><td></td><td></td></tr>
<tr><td>FAM_02014</td><td colspan="5"></td></tr>
</table>

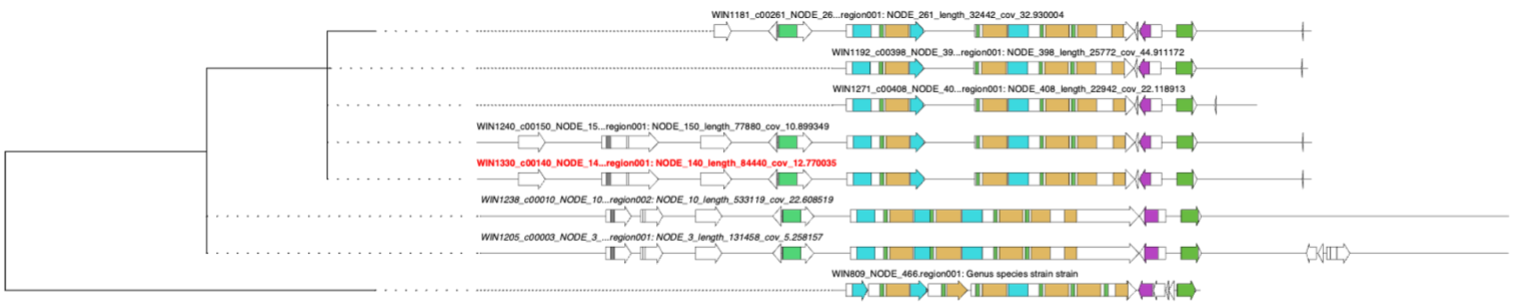
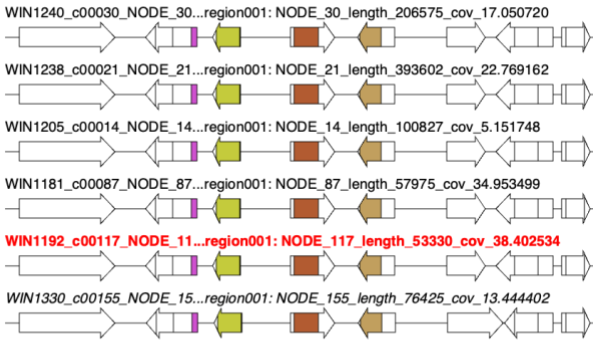
<table>
<tr><td></td><td colspan="5"> 
</td></tr>
<tr><td>FAM_02026</td><td colspan="5"> 
</td></tr>
<tr><td></td><td>NRPS</td><td>26</td><td>7</td><td></td><td></td></tr>
<tr><td>FAM_01939

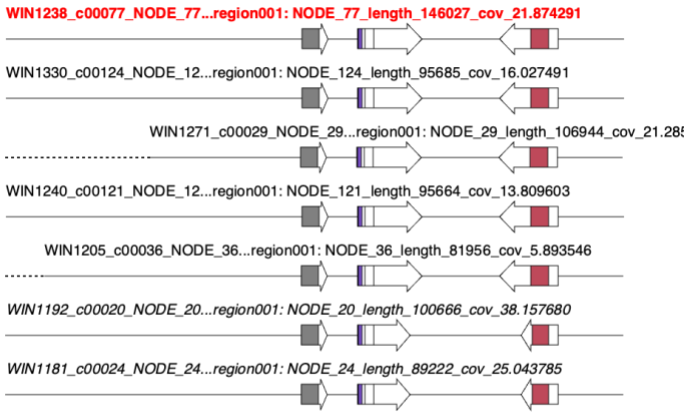
FAM_02043</td><td colspan="5"></td></tr>
</table>

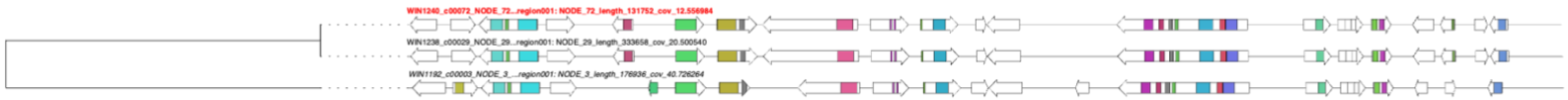
	
FAM_01944	
FAM_01945	

	
FAM_02006	
FAM_02008	

	
FAM_02040	
FAM_02068	

					
	Terpene	6	3		
FAM_01948					
FAM_02011					

<table>
<tr><td></td><td colspan="5"> 
</td></tr>
<tr><td>FAM_02080</td><td colspan="5"> 
</td></tr>
<tr><td></td><td>Hybrid</td><td>6</td><td>1</td><td></td><td></td></tr>
<tr><td>FAM_02026</td><td colspan="5"></td></tr>
</table>

					
	Others (RiPPs)	11	0		

4.5 Discussion

The current study sequenced the whole genome for *Leptographium wingfieldii* for the first time, along with the sequences for *L. procerum*, *L. aureum*, and *P. citrinum*. The genomes were found to be nearly complete based on the BUSCO analysis. The genome size of *Leptographium* species ranges from 29.5 to 32.4 Mb that is in the range of other studies that investigate genomes for members of the Ascomycota (Mohanta and Bae 2015). Genome data is a source to extract the nuclear marker genes/sequences for taxonomic identification in eukaryotes. Sometimes use of a single marker gene fails to provide the correct taxonomic validation, this is a common issue for fungi (Lücking et al. 2021). The current dataset had two strains unresolved for species designation, which in this study could be confirmed as *L. wingfieldii* based on the phylogenetic trees built using marker genes/sequences. Multiple marker genes are also necessary because sometimes a gene is missing or not retrievable for a species. The dataset included WIN(M)1181 strain of *L. wingfieldii* which is an ex paratype (Zubaer et al. 2024) that increased the confidence of the fungal identification. In addition to the species confirmation, the *Grosmannia aureum* fungal strain was found to be clustered with *Leptographium* species thus it confirms its proposed placement in the genus *Leptographium* as *Leptographium aureum*, that was also reported previously based on various nuclear marker genes/sequences (de Beer et al. 2022).

The known BGCs in *L. wingfieldii* are squalestatin, neosartorin, and naphthalene. Squalestatin was previously extracted from *Aspergillus* sp. and *Phoma* sp. It is an inhibitor of squalene synthase enzyme which is responsible for converting farnesyl pyrophosphate to

squalene in cholesterol biosynthesis. Previous studies suggested that squalestatin can be used as a cholesterol lowering agent, this compound can also be used in certain carcinogenic tumors that depend on cholesterol pathways (Bergstrom et al. 1993; Bonsch et al. 2016). Gene arrangement of squalestatin BGCs in *L. wingfieldii* were found quite conserved suggesting their similar functionality among all the *L. wingfieldii* strains. Naphthalene is a simple aromatic hydrocarbon well-known for its toxic and repellent properties. Naphthalene can be produced by fungi for multiple functions such as chemical defense, quorum sensing, stress response etc. (Cerniglia et al. 1978). Neosartorin was previously studied in *Aspergillus fumigatus* and *Aspergillus novofumigatus* (Frisvad and Larsen 2016; Matsuda et al. 2018). Structurally, neosartorin belongs to the bis-naphthopyrone family of natural products which are part of fungal defense against predatory organisms (Xu et al. 2019). It showed antibacterial effect against broad-spectrum gram-positive bacteria (Ola et al. 2014). Unlike squalestatin or naphthalene BGCs, the BGCs for neosartorin is highly variable. Except for WIN(M)1181 and WIN(M)1205, the rest of the *L. wingfieldii* strains present a second copy of neosartorin BGC. The gene arrangement of the second copy is different from the first copy; thus, it can be assumed that it may not have originated from duplication or duplication happened early in the evolution of these fungi and the second "copy" via drift changed significantly from the original ancestral version.

While *L. wingfieldii* strains showed similar BGCs, other *Leptographium* species (*L. procerum* and *L. aureum*) presented different types of BGCs. *L. procerum* has clavaric acid, bikaverin, phyllostictine BGCs. Clavaric acid is an inhibitor of farnesyl transferase that can have implication in cancer therapy (Jayasuriya et al. 1998). Bikaverin is a red pigment and mycotoxin previously found in *Fusarium fujikuroi*, which is also assumed to contribute to plant disease

(Wiemann et al. 2013). Phyllostictine were reported to have herbicidal property and being necrotic to plant tissues (Evidente et al. 2008). *L. aureum* potentially produces rosamicin, dichlorodiaporthin, scytalone etc. Rosamicin is an antimicrobial compound effective for gram-negative and gram-positive bacteria. It was found effective to treat infection by *Mycoplasma* sp. and *Ureaplasma* sp. (Crowe and Sanders 1974; Baumueller et al. 1977). Dichlorodiaporthin is a mycotoxin (Zhang et al. 2022), and Scytalone is a key intermediate for the biosynthesis of melanin which is crucial for fungal pathogenicity and protection against environmental stress (Bell et al. 1976). Apart from *Leptographium* species, BGCs are in *P. citrinum* is different and they tend to produce different types of natural products. The latter is not surprising as *P. citrinum* belongs to a different Order of fungi with different life histories and therefore this species (and other members of the Eurotiales) maintain BGCs that potentially benefit them in their survival and fitness. This species was included to provide an outgroup that was expected to contrast significantly with the examined species of *Leptographium*.

Among the number of BGCs identified by antiSMASH program, only a few provided "hits" with previously known BGCs. The majority of the BGCs and their products are not known. One must keep in mind that antiSMASH predictions are based on what is in the curated database. In the future the addition of more fungal genomes, such as for members of the Ophiostomatales, should enable antiSMASH to make improved predictions for more fungi. The antiSMASH program categorizes BGCs with the currently applied database matching the key enzyme (or core enzymes), and majority of the BGCs found belong to the PKS I group. PKS I enzymes synthesize a wide range of structurally diverse polyketides. Further using BiG-SCAPE program, similar BGCs in different *Leptographium* genomes can be clustered together into 23

groups; they can be speculated as having evolutionary relatedness and belonging to the same BGC family. With minor variation, the clusters showed genetic resemblance in gene arrangement among the *Leptographium* species. The singleton BGCs, which were not in clusters with others, could be to a particular strain and should be further studied for their novel characteristics. In addition to the clustered BGC and some singleton showed matches with databases such as MIBiG or Pfam, there are some unknown BGCs that might produce novel natural products that should be explored in future studies.

4.6 Conclusion

Leptographium wingfieldii and related species were found to have the potential to produce natural products that can be used as antimicrobials. The genome mining approach revealed many biosynthetic gene clusters (BGCs) that are common among the examined *Leptographium* species but also some unique BGCs have been noted for some strains. The current study provides a valuable resource for further exploration for a specific BGC and its product(s). It also provides an understanding of *Leptographium* genomes and potential diversity of BGCs in the Ophiostomatales.

GENERAL CONCLUSION

Leptographium wingfieldii is a blue stain causing fungus that infests pine trees and maintains a symbiotic relation with the bark beetle *Tomicus piniperda*. This fungal species was originally reported in Europe, and was introduced to the forests of Northwestern Ontario, Canada. Introduced species can pose a risk on the forest ecology in Canada. The current thesis uncovers the genomic information and annotated genes involved in the production of BGCs and the mitochondrial genome along with its mobile intron components. This is the first study that examines the genomics of *Leptographium wingfieldii*, and other related Ophiostomatales fungi (*L. procerum*, *L. terebrantis*, *L. aureum*, *O. minus*, *O. piliferum*). The whole genome mining of the sequenced organisms along with sequences collected from the NCBI databases were performed to establish an understanding about Ophiostomatales fungi and their mitochondrial genomes, including their taxonomy, common genomic features, genome evolution, potential production of natural products, and evolutionary dynamics of mitochondrial selfish elements (group I & II introns, and IEPs).

The mitogenomes of *O. minus* and *O. piliferum* were compared with mitogenomes of other members of the Ophiostomatales to explore gene content, gene order, phylogenetic relationships, and the distribution of mobile genetic elements (group I and II introns and homing endonucleases). The study found that gene order is largely conserved among the Ophiostomatales, with minor variations in tRNA composition and the presence or absence of the *atp9* gene. It highlighted that mitogenome size differences among Ophiostomatales, ranging from 23.7 kb to 150 kb, are correlated with the number of introns in the genomes. The *Sporothrix*

sensu stricto species exhibiting smaller mitogenomes due to intron loss. This dynamic intron evolution is further highlighted by the discovery of complex or nested introns composed of multiple intron modules, indicating that these introns contribute significantly to mitogenomic plasticity. These complex introns suggest that several introns can sometimes target the same insertion site, or sometimes mobile introns insert into existing introns (resident introns). The resolution of these complex introns may involve several splicing events to ensure the proper removal of all components to ensure that the host gene is functional. These complex introns should be explored in more detail with regards to their splicing pathways and it is somewhat speculative but maybe the removal of these complex introns is a rate limiting step that may impact the expression of the "host" gene. Introns may provide an opportunity of fine-tuning gene regulation. The phylogenetic analysis in this study supported recent taxonomic revisions within the Ophiostomatales, providing evidence that some genera, such as *Raffaelea*, may be polyphyletic. It also validated the separation of *Sporothrix* sensu stricto from *Ophiostoma*, offering clarity on the taxonomy of these fungi. By exploring phylogenetic relationships, genome structure, and mobile genetic elements, this study provides a comprehensive understanding of the genetic diversity, genome evolution, and possible ecological adaptations of Ophiostomatales fungi as intron loss (or mitogenome streamlining) was observed in members of the Genus *Sporothrix* that are animal parasites.

The mitochondrial genomes of *Leptographium* species with a focus on the exotic *Leptographium wingfieldii* showed that the mitogenomes for this group ranged from 41 kb to 126 kb in size, again the size variation appears to primarily be driven by mobile introns, which contribute to genome expansion and complexity. The introns were found to be the main

impacting factor for the observed mitogenome expansion. Other key findings include the presence of conserved gene arrangements, a high density of introns in certain genes like *cox1* and *rnl*, and complex intron arrangements (e.g., twintrons in *cox1* gene in *L. wageneri,* where a IC2 intron is nested in IB intron), which may play roles in gene regulation and adaptation. Phylogenetic analysis confirmed that *L. wingfieldii* is closely related to *L. terebrantis* and *L. lundbergii*, while *L. procerum* is more distantly related. The study suggested that mitogenome sequences can be used as a marker for identifying invasive fungi and resolving taxonomic challenges.

Finally, this work also explored the nuclear genomes of *L. wingfieldii* and related species to identify the biosynthetic gene clusters. The impetus for this study was that some of the examined strains were found in a previous study to possess antibacterial activity against the multi-drug-resistant bacteria. The genomes showed 19 to 25 BGCs in *Leptographium* and these were further categorized and studied comparatively. BGCs in *Leptographium* can be assigned to following categories: PKS I, PKS III, terpenes, NRPS, RiPPs and hybrid types. These fungi potentially produce squalestatin, neosartorin, naphthalene, rosamicin, bikaverin, scytalone and other natural products that are either mycotoxins, pigments, antibacterials, or metabolic inhibitors. These compounds can potentially be applied in medical or other industries. However, this will require more detailed chemistry work to isolate, purify and identify the chemical structures of these compounds. In addition, work is needed on their toxicity to mammalian cells and mode of action. This study should be followed up with more detailed transcriptomic work to assess which gene clusters are actively expressed. Plus, detailed metabolomic studies are needed

on these fungi to potentially correlate the presence, expression of BGCs with the presence of certain metabolites (i.e., SMs).

The current thesis studied the nuclear and mitochondrial genomes of *L. wingfieldii* and related fungi. Future research on the mitogenomics of blue-stain fungi such as *Leptographium wingfieldii* and other Ophiostomatales species should focus on further elucidating the functional roles of mobile introns, complex intron arrangements, and their encoded proteins on the organisms' biology and utilize them as biotechnological tool. These elements, while traditionally viewed as selfish genetic elements, may have adaptive significance in fungal survival and possible pathogenesis. Fungal natural products have a significant role in human history and the filamentous fungi in Ophiostomatales also have the potential to produce some important novel compounds. Further research on those BGCs discovered in this study and devising a heterologous expression system in more suitable fungal model systems for more functional characterization can open a new paradigm in the genome mining research.

REFERENCES

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

Adachi J, Waddell PJ, Martin W, Hasegawa M. 2000. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. J Mol Evol. 50(4):348–358. https://doi.org/10.1007/s002399910038.

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, et al. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46(W1):W537–W544. https://doi.org/10.1093/nar/gky379.

Aguileta G, de Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, et al. 2014. High variability of mitochondrial gene order among fungi. Genome Biol Evol. 6:451–465. https://doi.org/10.1093/gbe/evu028.

Alberts AW, Chen J, Kuron G, Hunt V, Huff J, Hoffman C, Rothrock J, Lopez M, Joshua H, Harris E, et al. 1980. Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. Proc Natl Acad Sci USA. 77(7):3957–3961. doi: 10.1073/pnas.77.7.3957.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403-410.

Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. 2021. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 20(3):200–216. https://doi.org/10.1038/s41573-020-00114-z.

Ayer WA, Browne L, Hiratsuka Y. 1986. Metabolites of *Ceratocystis ips* and other blue-stain fungi associated with the mountain pine beetle. Can J Microbiol. 32(3):231-237. https://doi.org/10.1139/m86-043.

Ayer WA, Craw PA. 1989. Metabolites of *Leptographium wageneri* var. pseudotsugae, a pathogen of Douglas-fir. Journal of Natural Products. 52(1):154-159. https://doi.org/10.1021/np50061a024.

Ayer WA, Lee SP. 1983. Metabolites produced by *Leptographium wageneri*, the cause of black stain root disease of conifers. Can J Chem. 61(9):1830-1835. https://doi.org/10.1139/v83-317.

Ayer WA, Segal RP, Tsuneda A. 1985. Metabolites of *Ceratocystis clavigera*, a blue stain fungus. Can J Micro. 31(6):705-710. https://doi.org/10.1139/m85-131.

Baidyaroy D, Hausner G, Hafez M, Michel F, Fulbright DW, 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. Fun Gen Biol. 48(8):775-83.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021.

Baumueller A, Hoyme U, Madsen PO. 1977. Rosamicin--a new drug for the treatment of bacterial prostatitis. Antimicrob Agents Chemother. 12(2):240-2. doi: 10.1128/AAC.12.2.240.

Begel O, Boulay J, Albert B, Dufour E, Sainsard-Chanet A. 1999. Mitochondrial group II introns, cytochrome c oxidase, and senescence in *Podospora anserina*. Mol Cell Biol. 19(6):4093–4100. https://doi.org/10.1128/MCB.19.6.4093.

Belfort M, Lambowitz AM. 2019. Group II intron RNPs and reverse transcriptases: From retroelements to research tools. Cold Spring Harbor Perspect Biol. 11(4):a032375. doi:10.1101/cshperspect.a032375.

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

Bell AA, Puhalla JE, Tolmsoff WJ, Stipanovic RD. 1976. Use of mutants to establish (+)-scytalone as an intermediate in melanin biosynthesis by *Verticillium dahliae*. Can J Microbiol 22(6): 1015–1021.

Ben Jamaa ML, Lieutier F, Yart A, Jerraya A, Khouja ML. 2007. The virulence of phytopathogenic fungi associated with the bark beetles *Tomicus piniperda* and *Orthotomicus erosus* in Tunisia. For Pathol. 37:51–63. https://doi.org/10.1111/j.1439-0329.2007.00478.x.

Bendich AJ. 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr Genet. 24:279–290. https://doi.org/10.1007/BF00336777.

Bennett PI, Tabima JF, Leon AL, Browning J, Wingfield MJ, LeBoldus JM. 2021. Spatial genetic structure of the insect-vectored conifer pathogen *Leptographium wageneri* suggests long-distance gene flow among Douglas-fir plantations in western Oregon. Front For Glob Change. 4:695981. https://doi.org/10.3389/ffgc.2021.695981.

Bérdy J. 2005. Bioactive microbial metabolites. J Antibiot (Tokyo). 58(1):1–26.

Bergeron MJ, Feau N, Stewart D, Tanguay P, Hamelin RC. 2019. Genome-enhanced detection and identification of fungal pathogens responsible for pine and poplar rust diseases. PLoS One. 14(2):e0210952. doi:10.1371/journal.pone.0210952.

Bergstrom JD, Kurtz MM, Rew DJ, Amend AM, Karkas JD, Bostedor RG, Bansal VS, Dufresne C, VanMiddlesworth FL, Hensens OD, et al. 1993. Zaragozic acids: a family of fungal metabolites that are picomolar competitive inhibitors of squalene synthase. Proc Natl Acad Sci U S A. 90(1):80-4. doi: 10.1073/pnas.90.1.80.

Bills GF, Gloer JB. 2016. Biologically Active Secondary Metabolites from the Fungi. Microbiol Spectr. 4(6). doi: 10.1128/microbiolspec.FUNK-0009-2016. PMID: 27809954.

Bilto IM, Hausner G. 2016. The diversity of mtDNA rns introns among strains of *Ophiostoma piliferum*, *Ophiostoma pluriannulatum*, and related species. *Springerplus* 5:1408. doi:10.1186/s40064-016-3076-6.

Bleiker KP, Six DL. 2007. Dietary benefits of fungal associates to an eruptive herbivore: potential implications of multiple associates on host population dynamics. Environ Entomol. 36(6):1384-96.

Blin K, Pascal Andreu V, de Los Santos ELC, Del Carratore F, Lee SY, Medema MH, Weber T. 2019. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. 47(D1):D625-D630. doi: 10.1093/nar/gky1060.

Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, et al. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. 2023 Jul 5;51(W1):W46-W50. doi: 10.1093/nar/gkad344.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49(W1):W29-W35.

Bonsch B, Belt V, Bartel C, Duensing N, Koziol M, Lazarus CM, Bailey AM, Simpson TJ, Cox RJ. 2016. Identification of genes encoding squalestatin S1 biosynthesis and in vitro production of new squalestatin analogues. Chem Commun (Camb). 52(41):6777-80. doi: 10.1039/c6cc02130a.

Borel JF, Feurer C, Gubler HU, Stähelin H. 1976. Biological effects of cyclosporin A: a new antilymphocytic agent. Agents Actions. 6(4):468–475. doi: 10.1007/BF01973261.

Brady SF, Clardy J. 2000. Long-chain N-acyltyrosine synthases from environmental DNA. J Am Chem Soc. 122(26):12903-12904. doi:10.1021/ja003172e.

Brakhage AA. 2013. Regulation of fungal secondary metabolism. Nat Rev Microbiol. 11(1):21-32. doi: 10.1038/nrmicro2916.

Brown DW, Butchko RA, Busman M, Proctor RH. 2012. Identification of gene clusters associated with fusaric acid, fusarin, and perithecial pigment production in *Fusarium verticillioides*. Fungal Genet Biol. 49(7):521-32. doi: 10.1016/j.fgb.2012.05.010.

Bullerwell CE, Lang BF. 2005. Fungal evolution: the case of the vanishing mitochondrion. *Curr Opin Microbiol* 8:362–369. doi:10.1016/j.mib.2005.06.009.

Burgess TI, Crous CJ, Slippers B, Hantula J, Wingfield MJ. 2017. Tree invasions and biosecurity: eco-evolutionary dynamics of hitchhiking fungi. *AoB Plants* 8:plw076. doi:10.1093/aobpla/plw076.

Buschges R, Bahrenberg G, Zimmermann M, Wolf K. 1994. NADH: ubiquinone oxidoreductase in obligate aerobic yeasts. Yeast. 10:475–479. doi:10.1002/yea.320100406.

Cacho RA, Tang Y, Chooi YH. 2015. Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. Front Microbiol. 5:774.

Caesar LK, Butun FA, Robey MT, Ayon NJ, Gupta R, Dainko D, Bok JW, Nickles G, Stankey RJ, Johnson D, et al. 2023. Correlative metabologenomics of 110 fungi reveals metabolite-gene cluster pairs. Nat Chem Biol. 19(7):846-854. doi: 10.1038/s41589-023-01276-8.

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

Caprara MG, Waring RB. 2005. Group I introns and their maturases: uninvited, but welcome guests. In: Belfort M, Wood DW, Stoddard BL, Derbyshire V, editors. Homing Endonucleases and Inteins. Berlin (Germany): Springer; p. 103–119. https://doi.org/10.1007/3-540-29474-0_7

Capron A, Stewart D, Hrywkiw K, Allen K, Feau N, Bilodeau G, Tanguay P, Cusson M, Hamelin RC. 2020. In situ processing and efficient environmental detection (iSPEED) of tree pests and pathogens using point-of-use real-time PCR. PLoS One. 15(4). doi:10.1371/journal.pone.0226863.

Cech TR, Zaug AJ, Grabowski PJ. 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. doi:10.1016/0092-8674(81)90390-1.

Cerniglia CE, Hebert RL, Szaniszlo PJ, Gibson DT. 1978. Fungal transformation of naphthalene. Arch Microbiol. 117(2):135-43. doi: 10.1007/BF00402301.

Chakravarty P, Trifonov L, Hutchison LJ, Hiratsuka Y, Ayer WA. 1994. Role of *Sporormiella similis* as a potential bioprotectant of *Populus tremuloides* wood against the blue-stain fungus *Ophiostoma piliferum*. Can J For Res. 24:2235–2239. doi: 10.1139/x94-286.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Methods Mol Biol. 1962:1–14. doi: 10.1007/978-1-4939-9173-0_1.

Chan YA, Podevels AM, Kevany BM, Thomas MG. 2009. Biosynthesis of polyketide synthase extender units. Nat Prod Rep. 26(1):90–114. doi: 10.1039/B801658P.

Chang R, Duong TA, Taerum SJ, Wingfield MJ, Zhou X, Yin M, et al. 2019. Ophiostomatoid fungi associated with the spruce bark beetle *Ips typographus*, including 11 new species from China. Persoonia Mol Phylogeny Evol Fungi. 42:50–74. doi: 10.3767/persoonia.2019.42.03.

Chen XJ, Clark-Walker GD. 2018. Unveiling the mystery of mitochondrial DNA replication in yeasts. Mitochondrion. 38:17–22. doi: 10.1016/j.mito.2017.07.009.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. German Conference on Bioinformatics. 99(1):45–56.

Choi D, Harrington TC, Shaw DC, Stewart JE, Klopfenstein NB, Kroese DR, Kim MS. 2023. Phylogenetic analyses allow species-level recognition of *Leptographium wageneri* varieties that cause black stain root disease of conifers in western North America. Front Plant Sci. 14:1286157. doi: 10.3389/fpls.2023.1286157.

Cinget B, Bélanger RR. 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 Biol. 17(9):1252–1260. doi: 10.1080/15476286.2020.1763024.

Clevenger KD, Bok JW, Ye R, Miley GP, Verdan MH, Velk T, Chen C, Yang K, Robey MT, Gao P, et al. 2017. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat Chem Biol. 13(8):895–901.

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

Córdoba S, Isla G, Szusz W, Vivot W, Hevia A, Davel G, Canteros CE. 2018. Molecular identification and susceptibility profile of *Sporothrix schenckii* sensu lato isolated in Argentina. Mycoses. 61(7):441-448. doi: 10.1111/myc.12760.

Cox RJ, Simpson TJ. 2009. Fungal type I polyketide synthases. Methods Enzymol.459:49-78. doi: 10.1016/S0076-6879(09)04603-5.

Cox RJ, Skellam E, Williams K. 2018. Biosynthesis of fungal polyketides. In: Physiology and Genetics: Selected Basic and Applied Aspects. p. 385-412.

Cox RJ. 2007. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. Org Biomol Chem. 5(13):2010–2026.

Crowe CC, Sanders WE Jr. 1974. Rosamicin: evaluation in vitro and comparison with erythromycin and lincomycin. Antimicrob Agents Chemother. (3):272-5. doi: 10.1128/AAC.5.3.272.

D'Souza AR, Minczuk M. 2018. Mitochondrial transcription and translation: overview. Essays Biochem. 62(3):309–320. doi: 10.1042/EBC20170102.

Daniels DL, Michels WJ Jr, Pyle AM. 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(1):31–49.

Davidson RW. 1971. New species of *Ceratocystis*. Mycologia. 63:5–15. doi:10.1080/00275514.1971.12019076.

de Beer ZW, Duong TA, Wingfield MJ. 2016a. The divorce of *Sporothrix* and *Ophiostoma*: solution to a problematic relationship. *Stud Mycol*. 83:165–191. doi: 10.1016/j.simyco.2016.07.001.

de Beer ZW, Marincowitz S, Duong TA, Kim JJ, Rodrigues A, Wingfield MJ. 2016b. *Hawksworthiomyces* gen. nov. (*Ophiostomatales*), illustrates the urgency for a decision on how to name novel taxa known only from environmental nucleic acid sequences (ENAS). Fungal Biol. 120:1323–1340. doi: 10.1016/j.funbio.2016.07.004.

de Beer ZW, Procter M, Wingfield MJ, Marincowitz S, Duong TA. 2022. Generic boundaries in the Ophiostomatales reconsidered and revised. Stud Mycol. 101:57–120.

de Beer ZW, Seifert KA, Wingfield MJ. 2013a. A nomenclator for ophiostomatoid genera and species in the *Ophiostomatales* and *Microascales*. In: Seifert KA, de Beer ZW, Wingfield MJ, editors. Ophiostomatoid fungi: expanding frontiers. Utrecht (Netherlands): CBS-KNAW Fungal Biodiversity Centre; p. 245–322

de Beer ZW, Seifert KA, Wingfield MJ. 2013b. The Ophiostomatoid fungi: their dual position in the *Sordariomycetes*. In: Seifert KA, de Beer ZW, Wingfield MJ, editors. Ophiostomatoid fungi: expanding frontiers. Utrecht (Netherlands): CBS-KNAW Fungal Biodiversity Centre; p. 1–19. doi: 10.1016/s0953-7562(09)80327-4.

de Beer ZW, Wingfield MJ. 2013. Emerging lineages in the Ophiostomatales. In: Seifert KA, de Beer ZW, Wingfield MJ, editors. Ophiostomatoid fungi: expanding frontiers. Utrecht (Netherlands): CBS-KNAW Fungal Biodiversity Centre. p. 21–46.

de Jong A, van Hijum SA, Bijlsma JJ, Kok J, Kuipers OP. 2006. BAGEL: a web-based bacteriocin genome mining tool. Nucleic Acids Res. 34:W273-9.

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, Zhang Q, Ming R, Lin L, Lin X, Lin Y, et al. 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.

Diez B, Barredo JL, Alvarez E, Cantoral JM, van Solingen P, Groenen MAM, Veenstra AE, Martin JF. 1989. Two genes involved in penicillin biosynthesis are linked in a 5.1 Kb *SalI* fragment in the genome of *Penicillium chrysogenum*. Mol Gen Genet. 218:572–576.

Diez B, Gutierrez S, Barredo JL, van Solingen P, van der Voort LHM, Martin JF. 1990. Identification and characterization of the *pcbAB* gene and linkage to *pcbC* and *penDE* genes. J Biol Chem. 265:16358–16365.

DiGuistini S, Wang Y, Liao NY, Taylor G, Tanguay P, Feau N, Henrissat B, Chan SK, Hesse-Orce U, Alamouti SM, et al. 2011. Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. Proc Natl Acad Sci USA. 108(6):2504–2509. doi: 10.1073/pnas.1011289108.

Domingo-Fernández D, Gadiya Y, Preto AJ, Krettler CA, Mubeen S, Allen A, Healey D, Colluru V. 2024. Natural products have increased rates of clinical trial success throughout the drug development process. J Nat Prod. 87(7):1844–1851. doi: 10.1021/acs.jnatprod.4c00581.

Dujon B, Colleaux L, Jacquier A, Michel F, Monteilhet C. 1986. Mitochondrial introns as mobile genetic elements: the role of intron-encoded proteins. Basic Life Sci. 40:5–27. doi: 10.1007/978-1-4684-5251-8_2.

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

Duo A, Bruggmann R, Zoller S, Bernt M, Grunig CR. 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 LG, Jones JP, Klepzig KD. 2004. Pathogenicity of *Leptographium* species associated with loblolly pine decline. Plant Dis. 88(11):1174–1178. doi: 10.1094/PDIS.2004.88.11.1174.

Eddy SR, Durbin R. 1994. RNA sequence analysis using covariance models. Nucleic Acids Res. 22:2079–2088.

Eddy SR. 1998. Profile hidden Markov models. Bioinformatics. 14(9):755–763. doi: 10.1093/bioinformatics/14.9.755.

Edgell DR, Chalamcharla VR, 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.

Enyeart PJ, Mohr G, Ellington AD, Lambowitz AM. 2014. Biotechnological applications of mobile group II introns and their reverse transcriptases: gene targeting, RNA-seq, and non-coding RNA analysis. Mob DNA. 5(1):2. doi: 10.1186/1759-8753-5-2.

Evidente A, Cimmino A, Andolfi A, Vurro M, Zonno MC, Motta A. 2008. Phyllostoxin and phyllostin, bioactive metabolites produced by *Phyllosticta cirsii*, a potential mycoherbicide for *Cirsium arvense* biocontrol. J Agric Food Chem. 56(3):884-8. doi: 10.1021/jf0731301.

Fallmann J, Will S, Engelhardt J, Grüning B, Backofen R, Stadler PF. 2017. Recent advances in RNA folding. J Biotechnol. 261:97–104. doi: 10.1016/j.jbiotec.2017.07.007.

Fedorova O, Arhin G, Pyle AM, Frank AT. 2023. In silico discovery of group II intron RNA splicing inhibitors. ACS Chem Biol. 18(9):1968–1975. doi: 10.1021/acschembio.3c00160.

Fedorova O, Jagdmann GE Jr, Adams RL, Yuan L, Van Zandt MC, Pyle AM. 2018. Small molecules that target group II introns are potent antifungal agents. Nat Chem Biol. 14(12):1073–1078. doi: 10.1038/s41589-018-0142-0.

Férandon C, Moukha S, Callac P, Benedetto JP, Castroviejo M, Barroso G. 2010. The Agaricus bisporus cox1 gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group i introns. PLoS One. 5(11):e14048. doi:10.1371/journal.pone.0014048.

Ferat JL, Michel F. 1993. Group II self-splicing introns in bacteria. Nature. 364(6435):358-61. doi:10.1038/364358a0.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29-37.

Fiskaa T, Birgisdottir AB. 2010. RNA reprogramming and repair based on trans-splicing group I ribozymes. New Biotechnol. 27(3):194–203. doi: 10.1016/j.nbt.2010.02.013.

Fleming A. 1929. On the antibacterial action of cultures of a Penicillium, with special reference to their use in the isolation of B. influenzae. Br J Exp Pathol. 10(3):226–236.

Fonseca PLC, De-Paula RB, Araújo DS, Tomé LMR, Mendes-Pereira T, Rodrigues WFC, Del-Bem LE, Aguiar ERGR, Góes-Neto A. 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.

Fortier CE, Musso AE, Evenden ML, Zaharia LI, Cooke JEK. 2024. Evidence that Ophiostomatoid fungal symbionts of mountain pine beetle do not play a role in overcoming lodgepole pine defenses during mass attack. Mol Plant Microbe Interact. 37(5):445–458. doi: 10.1094/MPMI-06-23-0077-R.

Franco MEE, López SMY, Medina R, Lucentini CG, Troncozo MI, Pastorino GN, 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 KC, Friedrich A, Schacherer J. 2015. Mitochondrial genome evolution in yeasts: an all-encompassing view. FEMS Yeast Res. 15:fov023. doi: 10.1093/femsyr/fov023.

Frisvad JC, Andersen B, Thrane U. 2008. The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. Mycol Res. 112(Pt 2):231-40. doi:10.1016/j.mycres.2007.08.018.

Frisvad JC, Larsen TO. 2016. Extrolites of *Aspergillus fumigatus* and other pathogenic species in *Aspergillus* section *Fumigati*. Front Microbiol. 6:1485. doi: 10.3389/fmicb.2015.01485.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus Neurospora crassa. Nature. 422(6934):859-68. doi:10.1038/nature01554.

Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J Mol Biol. 313:1003–1011. doi: 10.1006/jmbi.2001.5102.

Gelderblom WCA, Jaskiewicz K, Marasas WFO, Thiel PG, Horak RM, Vleggaar R, Kriek NPJ. 1988. Fumonisins: novel mycotoxins with cancer-promoting activity produced by *Fusarium moniliforme*. Appl Environ Microbiol. 54(7):1806-11.

Gibb EA, Hausner G. 2005. Optional mitochondrial introns and evidence for a homing-endonuclease gene in the mtDNA *rnl* gene in *Ophiostoma ulmi* s. lat. Mycol Res. 109:1112–1126. doi: 10.1017/S095375620500376X.

Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci. 96(24): 13880–13885. doi: 10.1073/pnas.96.24.13880.

Golik P. 2024. RNA processing and degradation mechanisms shaping the mitochondrial transcriptome of budding yeasts. IUBMB Life. 76(1):38-52. doi: 10.1002/iub.2779.

Gomes RMOS, Silva KJG, Theodoro RC. 2024. Group I introns: Structure, splicing and their applications in medical mycology. Genet Mol Biol. 47(Suppl 1):e20230228. doi: 10.1590/1678-4685-GMB-2023-0228.

González-Delgado A, Mestre MR, Martínez-Abarca F, Toro N. 2021. Prokaryotic reverse transcriptases: from retroelements to specialized defense systems. FEMS Microbiol Rev. 45(6):fuab025. doi: 10.1093/femsre/fuab025.

Gorton C, Kim SH, Henricot B, Webber J, Breuil C. 2004. Phylogenetic analysis of the blue stain fungus *Ophiostoma minus* based on partial ITS rDNA and β-tubulin gene sequences. Mycol Res. 108:759–765. doi: 10.1017/S0953756204000012.

Gorton C, Webber JF. 2000. Reevaluation of the status of the bluestain fungus and bark beetle associate *Ophiostoma minus*. Mycologia. 92:1071–1079. doi: 10.2307/3761474.

Goyal S, Ramawat KG, Merillon JM. Different Shades of Fungal Metabolites: An Overview. In: Mérillon JM, Ramawat K, editors. Fungal Metabolites Reference Series in Phytochemistry. Switzerland: Springer Cham; 2017. p. 1–29.

Gray MW, Lukeš J, Archibald JM, Keeling PJ, Doolittle WF. 2010. Irremediable complexity? Science. 330:920–921. doi: 10.1126/science.1198594.

Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. 35:849–857. doi: 10.1016/0092-8674(83)90117-4.

Guha TK, 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.0150097.

Guha TK, Wai A, 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 TK, Wai A, Mullineux ST, Hausner G. 2018. The intron landscape of the mtDNA cytb gene among the Ascomycota: introns and intron-encoded open reading frames. Mitochondrial DNA A. 29(7):1015–1024. doi: 10.1080/24701394.2017.1404042.

Hadjithomas M, Chen IM, Chu K, Ratner A, Palaniappan K, Szeto E, Huang J, Reddy TB, Cimermančič P, Fischbach MA, et al. 2015. IMG-ABC: A knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. mBio. 6(4):e00932. doi: 10.1128/mBio.00932-15.

Hafez M, Hausner G. 2011. The highly variable mitochondrial small-subunit ribosomal RNA gene of Ophiostoma minus. Fungal Biol. 115(11):1122-37. doi:10.1016/j.funbio.2011.07.007.

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

Hafez M, Hausner G. 2015. Convergent evolution of twintron-like configurations: One is never enough. RNA Biol. 12(12):1275-88. doi: 10.1080/15476286.2015.1103427.

Hafez M, Majer A, Sethuraman J, Rudski SM, Michel F, Hausner G. 2013. The mtDNA *rns* gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. Fungal Genetics and Biology. 53:71-83.

Harrington TC, Cobb FW. 1983. Pathogenicity of Leptographium and Verticicladiella spp. Isolated from Roots of Western North American Conifers. Phytopathology. 73:596-599.

Harrington TC. 1988. Leptographium species, their distributions, hosts and insect vectors. In: *Leptographium Root Diseases of Conifers.* Harrington TC, Cobbs FW Jr, editors. St. Paul: APS Press. p. 1–40.

Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. Trends Genet. 21(2):111-9. doi: 10.1016/j.tig.2004.12.007.

Hausner G, Hafez M, Edgell DR. 2014. Bacterial group I introns: mobile RNA catalysts. Mobile DNA. 5(1):8. https://doi.org/10.1186/1759-8753-5-8

Hausner G, Iranpour M, Kim J-J, Breuil C, Davis CN, Gibb EA, Reid J, Loewen PC, Hopkin AA. 2005. Fungi vectored by the introduced bark beetle *Tomicus piniperda* in Ontario, Canada, and comments on the taxonomy of *Leptographium lundbergii, Leptographium terebrantis, Leptographium truncatum*, and *Leptographium wingfieldii*. Canadian Journal of Botany. 83(10): 1222-1237.

Hausner G, Reid J, Klassen GR. 1993. On the phylogeny of *Ophiostoma*, *Ceratocystis* s.s., and Microascus, and relationships within *Ophiostoma* based on partial ribosomal DNA sequences. Can J Bot. 71(9): 1249-1265.

Hausner G. 2003. Fungal mitochondrial genomes, introns and plasmids. In: Arora DK, Khachatourians GG, editors. Applied Mycology and Biotechnology. Volume III: Fungal Genomics. New York: Elsevier Science. p. 101–131.

Hausner G. 2012. Introns, mobile elements and plasmids. In: Bullerwell CE, editor. Organelle genetics: evolution of organelle genomes and gene expression. Berlin: Springer Verlag. p. 329–358.

Hawksworth DL, Crous PW, Redhead SA, Reynolds DR, Samson RA, Seifert KA, et al. 2011. The Amsterdam declaration on fungal nomenclature. IMA Fungus. 2(1):105–112. https://doi.org/10.5598/imafungus.2011.02.01.14

Herbst DA, Townsend CA, Maier T. 2018. The architectures of iterative type I PKS and FAS. Nature Chemical Biology. 14(5):474-479. doi:10.1038/s41589-018-0022-y.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, et al. 2007. A higher-level phylogenetic classification of the Fungi. Mycol Res. 111(Pt 5):509-47. doi: 10.1016/j.mycres.2007.03.004.

Hintz WE, Peberdy JF. 1989. Chromosome-length polymorphisms in Ophiostoma ulmi. Current Genetics. 15(3):179-186. doi:10.1007/BF00311026.

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. Monatsh Chem.125:167-88.

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

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17(8):754-755. doi:10.1093/bioinformatics/17.8.754.

Humble LM, Allen EA. 2006. Forest biosecurity: alien invasive species and vectored organisms. Canadian Journal of Plant Pathology. 28:S256-S269.

Hyde KD, Norphanphoun C, Maharachchikumbura SS, Bhat DJ, Jones EB, Bundhun D, Chen YJ, Bao DF, Boonmee S, Calabon MS, et al. 2020. Refined families of sordariomycetes. Mycosphere. 11(1):305–1059. doi:10.5943/mycosphere/11/1/7.

Ibarra Caballero JR, Jeon J, Lee YH, Fraedrich S, Klopfenstein NB, Kim MS, Stewart JE. 2019. Genomic comparisons of the laurel wilt pathogen, *Raffaelea lauricola*, and related tree pathogens highlight an arsenal of pathogenicity related genes. Fungal Genet Biol. 125:84-92. doi: 10.1016/j.fgb.2019.01.012.

Ibrahim A. 2020. Identification of secondary metabolites extracted from *Leptographium* spp. in search of novel biochemical compounds with probable antimicrobial/therapeutic potential [dissertation]. University of Manitoba.

Jacobs K, Bergdahl DR, Wingfield MJ, Halik S, Seifert KA, Bright DE, Wingfield BD. 2004. *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. Mycol. Res. 108(Pt 4):411–418. https://doi.org/10.1017/s0953756204009748.

Jacobs K, Solheim H, Wingfield BD, Wingfield MJ. 2005. Taxonomic re-evaluation of *Leptographium lundbergii* based on DNA sequence comparisons and morphology. Mycol Res. 109(Pt 10):1149-61. doi: 10.1017/s0953756205003618.

Jacobs K, Wingfield MJ, Wingfield BD. 2001. Phylogenetic relationships in *Leptographium* based on morphological and molecular characters. Can J Bot. 79(6): 719-732.

Jacobs K, Wingfield MJ. 2001. *Leptographium* species: Tree pathogens, insect associates, and agents of blue-stain. St. Paul (MN): American Phytopathological Society. 207 p.

Jacobs K, Wingfield MJ. 2013. An overview of *Leptographium* and *Grosmannia*. In: Seifert KA, de Beer ZW, Wingfield MJ, editors. Ophiostomatoid fungi: Expanding frontiers. Utrecht: CBS-KNAW Fungal Biodiversity Centre. p. 47-56.

Jaeger L, Westhof E, Michel F. 1991. Function of P11, a tertiary base pairing in self-splicing introns of subgroup IA. J. Mol. Biol. 221:1153–1164. https://doi.org/10.1016/0022-2836(91)90925-V.

James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, Stajich JE. 2013. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr. Biol. 23(16):1548–1553. https://doi.org/10.1016/j.cub.2013.06.057.

Jankowiak R, Bilański P. 2013. Ophiostomatoid fungi associated with root-feeding bark beetles on Scots pine in Poland. For. Pathol. 43:422–428. https://doi.org/10.1111/efp.12049.

Jankowiak R, Ostafińska A, Aas T, Solheim A, Bilański P, Linnakoski R, Hausner G. 2018. Three new *Leptographium* spp. (Ophiostomatales) infecting hardwood trees in Norway and Poland. Antonie van Leeuwenhoek. 111(12):2323–2347. https://doi.org/10.1007/s10482-018-1123-8.

Jankowiak R, Strzałka B, Bilański P, et al. 2017. Diversity of Ophiostomatales species associated with conifer-infesting beetles in the Western Carpathians. Eur. J. Forest. Res. 136:939–956. https://doi.org/10.1007/s10342-017-1081-0.

Jayasuriya H, Silverman KC, Zink DL, Jenkins RG, Sanchez M, Pelaez F, Vilella D, Lingham RB, Singh SB. 1998. Clavaric acid: a triterpenoid inhibitor of farnesyl-protein transferase from *Clavariadelphus truncatus*. J Nat Prod. 61(12):1568-70. doi: 10.1021/np980200c.

Jenke-Kodama H, Sandmann A, Müller R, Dittmann E. 2005. Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol. 22(10):2027–2039. https://doi.org/10.1093/molbev/msi193

Jenkins HL, Graham R, Porter JS, Vieira LM, de Almeida ACS, Hall A, O'Dea A, Coppard SE, Waeschenbach A. 2022. Unprecedented frequency of mitochondrial introns in colonial bilaterians. Sci. Rep. 12(1):10889. https://doi.org/10.1038/s41598-022-14477-3.

Jiang P, Li H, Song F, Cai Y, Wang J, Liu J, Cai W. 2016. Duplication and remolding of tRNA genes in the mitochondrial genome of *Reduvius tenebrosus* (Hemiptera: Reduviidae). Int. J. Mol. Sci. 17(6):951. https://doi.org/10.3390/ijms17060951.

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21(1):241.

Johansen S, Haugen P. 2001. A new nomenclature of group I introns in ribosomal DNA. RNA. 7:935–936. https://doi.org/10.1017/S1355838201010500.

Kanzi AM, Wingfield BD, Steenkamp ET, Naidoo S, van der Merwe NA. 2016. Intron derived size polymorphism in the mitochondrial genomes of closely related *Chrysoporthe* species. PLoS One. 11:e0156104. https://doi.org/10.1371/journal.pone.0156104.

Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20(4):1160–1166. https://doi.org/10.1093/bib/bbx108.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30(4): 772–780. doi: 10.1093/molbev/mst010.

Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Duran HGS, Andreu VP, et al. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 48(D1):D454-D458. doi:10.1093/nar/gkz882.

Kautsar SA, van der Hooft JJJ, de Ridder D, Medema MH. 2021. BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. Gigascience. 2021 Jan 13;10(1):giaa154. doi: 10.1093/gigascience/giaa154.

Keller NP. 2015. Translating biosynthetic gene clusters into fungal armor and weaponry. Nat Chem Biol. 11(9):671-7. doi: 10.1038/nchembio.1897.

Keller NP. 2019. Fungal secondary metabolism: regulation, function and drug discovery. Nat Rev Microbiol. 17(3):167-180.

Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND. 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 47(9):736-41.

Khaldi N, Wolfe KH. 2011. Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. Int J Evol Biol. 2011:423821. doi: 10.4061/2011/423821.

Kim YJ, Kim HD, Kim J. 2013. Cytoplasmic ribosomal protein S3 (rpS3) plays a pivotal role in mitochondrial DNA damage surveillance. Biochim. Biophys. Acta Mol. Cell Res. 1833:2943–2952. https://doi.org/10.1016/j.bbamcr.2013.07.015.

Klepzig KD, Wilkens RT. 1997. Competitive interactions among symbiotic fungi of the southern pine beetle. Applied and Environmental Microbiology. 63(2):621-627.

Klepzig KD. 1998. Competition between a biological control fungus, *Ophiostoma piliferum*, and symbionts of the southern pine beetle. Mycologia. 90:69–75. https://doi.org/10.1080/00275514.1998.12026880.

Kobayashi G, Itoh H, Kojima S. 2022. Mitogenome of a stink worm (Annelida: Travisiidae) includes degenerate group II intron that is also found in five congeneric species. Scientific Reports. 12(1):4449. https://doi.org/10.1038/s41598-022-08103-5.

Kolesnikova AI, Putintseva YA, Simonov EP, Biriukov VV, Oreshkova NV, Pavlov IN, et al. 2019. Mobile genetic elements explain size variation in the mitochondrial genomes of four closely related *Armillaria* species. BMC Genomics. 20:351. https://doi.org/10.1186/s12864-019-5732-z.

Korovesi AG, Ntertilis M, Kouvelis VN. 2018. Mt-*rps3* is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. Mol Phylogenet Evol. 127:74–86. https://doi.org/10.1016/j.ympev.2018.04.037.

Koufopanou V, Goddard MR, Burt A. 2002. Adaptation for horizontal transfer in a homing endonuclease. Molecular Biology and Evolution. 19(3):239–246. https://doi.org/10.1093/oxfordjournals.molbev.a004077.

Kroken S, Glass NL, Taylor JW, Yoder OC, Turgeon BG. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc Natl Acad Sci USA. 100(26):15670–15675. doi: 10.1073/pnas.2532165100.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19(9):1639–1645. https://doi.org/10.1101/gr.092759.109.

Kulik T, Bilska K, Żelechowski M. 2020. Promising perspectives for detection, identification, and quantification of plant pathogenic fungi and oomycetes through targeting mitochondrial DNA. Int J Mol Sci. 21(7):2645. https://doi.org/10.3390/ijms21072645.

Kulik T, Brankovics B, van Diepeningen AD, Bilska K, Żelechowski M, Myszczyński K, Molcan T, Stakheev A, Stenglein S, Beyer M, et al. 2020. Diversity of mobile genetic elements in the mitogenomes of closely related *Fusarium culmorum* and *F. graminearum* sensu stricto strains and its implication for diagnostic purposes. Front Microbiol. 11:1002. doi:10.3389/fmicb.2020.01002.

Kulik T, van Diepeningen AD, Hausner G. 2021. Editorial: "The Significance of Mitogenomics in Mycology". Front Microbiol. 11:628579.

Lah L, Löber U, Hsiang T, Hartmann S. 2017. A genomic comparison of putative pathogenicity-related gene families in five members of the Ophiostomatales with different lifestyles. Fungal Biol. 121(3):234-252. doi: 10.1016/j.funbio.2016.12.002.

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS. 1999. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: Gesteland RF, Cech TR, Atkins JF, editors. The RNA World. New York (NY): Cold Spring Harbor Laboratory Press. p. 451–485.

Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. Annu Rev Genet. 38:1-35. doi: 10.1146/annurev.genet.38.072902.091600. PMID: 15568970.

Lambowitz AM, Zimmerly S. 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol. 3(8):a003616. doi: 10.1101/cshperspect.a003616.

Lang BF, Beck N, Prince S, Sarrasin M, Rioux P, 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 BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23(3):119-25. doi: 10.1016/j.tig.2007.01.006.

Lang BF, Lavrov D, Beck N, Steinberg SV. 2012. Mitochondrial tRNA structure, identity, and evolution of the genetic code. In: Bullerwell CE, editor. Organelle genetics: evolution of organelle genomes and gene expression. Berlin: Springer-Verlag Berlin Heidelberg. p. 431–474. https://doi.org/10.1007/978-3-642-22380-8_17.

Lang BF. 2014. Mitochondrial genomes in fungi. In: Wells RD, Bond JS, Klinman J, Masters BSS, Bell E, editors. Molecular life sciences. New York (NY): Springer New York. p. 1–7. https://doi.org/10.1007/978-1-4614-6436-5_113-2.

Lang BF. 2018. Mitochondrial genomes in fungi. In: Wells RD, Bond JS, Klinman J, Masters BSS. Molecular life sciences. New York (NY): Springer. p. 722–728. https://doi.org/10.1007/978-1-4614-1531-2_113.

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

Lee S, Kim JJ, Breuil C. 2005. *Leptographium longiclavatum* sp. nov., a new species associated with the mountain pine beetle, *Dendroctonus ponderosae*. Mycol Res. 109(Pt 10):1162-70. doi: 10.1017/s0953756205003588.

Lee S, Kim JJ, Breuil C. 2006. Diversity of fungi associated with mountain pine beetle, *Dendroctonus ponderosae*, and infested lodgepole pines in British Columbia. Fungal Diversity. 22:91–105.

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

Li JW, Vederas JC. 2009. Drug discovery and natural products: end of an era or an endless frontier? Science. 325(5937):161-5. doi: 10.1126/science.1168243.

Liang X, Tian X, Liu W, Wei T, Wang W, Dong Q, Wang B, Meng Y, Zhang R, Gleason ML, et al. 2017. Comparative analysis of the mitochondrial genomes of Colletotrichum gloeosporioides sensu lato: insights into the evolution of a fungal species complex

interacting with diverse plants. BMC Genomics. 18(1):171. doi: 10.1186/s12864-016-3480-x.

Lieutier F, Yart A, Salle A. 2009. Stimulation of tree defenses by Ophiostomatoid fungi can explain attack success of bark beetles on conifers. Ann For Sci. 66:801. https://doi.org/10.1051/forest/2009066.

Linnakoski R, de Beer ZW, Duong TA, Niemelä P, Pappinen A, Wingfield MJ. 2012a. *Grosmannia* and *Leptographium* spp. associated with conifer-infesting bark beetles in Finland and Russia, including *Leptographium taigense* sp. nov. Antonie Van Leeuwenhoek. 102(2):375-99. doi: 10.1007/s10482-012-9747-6.

Linnakoski R, de Beer ZW, Niemelä P, Wingfield MJ. 2012b. Associations of conifer-infesting bark beetles and fungi in Fennoscandia. Insects. 3(1):200–227. doi: 10.3390/insects3010200.

Linnakoski R, Mahilainen S, Harrington A, Vanhanen H, Eriksson M, Mehtätalo L, Pappinen A, Wingfield MJ. 2016. Seasonal succession of fungi associated with *Ips typographus* beetles and their phoretic mites in an outbreak region of Finland. PLoS One. 11(5):e0155622. doi: 10.1371/journal.pone.0155622.

Liu F, Ye F, Cheng C, Kang Z, Kou H, Sun J. 2022. Symbiotic microbes aid host adaptation by metabolizing a deterrent host pine carbohydrate d-pinitol in a beetle-fungus invasive complex. Science Advances. 8(51):eadd5051. https://doi.org/10.1126/sciadv.add5051.

Liu T, Pyle AM. 2021. Discovery of highly reactive self-splicing group II introns within the mitochondrial genomes of human pathogenic fungi. Nucleic Acids Research. 49(21):12422–12432. https://doi.org/10.1093/nar/gkab1077.

Liu T, Pyle AM. 2024. Highly reactive group I introns ubiquitous in pathogenic fungi. J Mol Biol. 436(8):168513.

Liu W, Cai Y, Zhang Q, Chen L, Shu F, Ma X, Sun Y, Zhang S, Liang Y, Peng X, et al. 2020a. The mitochondrial genome of Morchella importuna (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames

and repetitive sequences. Int J Biol Macromol. 143:373–381. doi: 10.1016/j.ijbiomac.2019.12.056.

Liu W, Cai Y, Zhang Q, Shu F, Chen L, Ma X, Sun Y, Zhang S, Liang Y, Peng X, et al. 2020b. Subchromosome-scale nuclear and complete mitochondrial genome characteristics of Morchella crassipes. Int J Mol Sci. 21:483. doi: 10.3390/ijms21020483.

Lu Q, Decock C, Zhang XY, et al. 2008. *Leptographium sinoprocerum* sp. nov., an undescribed species associated with *Pinus tabuliformis – Dendroctonus valens* in northern China. Mycologia 100: 275–290.

Lücking R, Aime MC, Robbertse B, Miller AN, Aoki T, Ariyawansa HA, Cardinali G, Crous PW, Druzhinina IS, Geiser DM, et al. 2021. Fungal taxonomy and sequence-based nomenclature. Nat Microbiol. 6(5):540-548. doi: 10.1038/s41564-021-00888-x.

Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. 2011. How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life. 63(7):528–537. doi: 10.1002/iub.489.

Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, Dentinger B, Padamsee M, Hibbett D, James TY, Baloch E, et al. 2004. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. Am J Bot. 91(10):1446-80. doi: 10.3732/ajb.91.10.1446.

Malbert B, Labaurie V, Dorme C, Paget E. 2023. Group I Intron as a Potential Target for Antifungal Compounds: Development of a Trans-Splicing High-Throughput Screening Strategy. Molecules. 28(11):4460. doi: 10.3390/molecules28114460.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc. 1(12):e323. doi: 10.1002/cpz1.323.

Marcia M, Somarowthu S, Pyle AM. 2013. Now on display: a gallery of group II intron structures at different stages of catalysis. Mob DNA. 4:14. doi:10.1186/1759-8753-4-14.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A. 101(19):7287-92.

Matsuda Y, Gotfredsen CH, Larsen TO. 2018. Genetic Characterization of Neosartorin Biosynthesis Provides Insight into Heterodimeric Natural Product Generation. Org Lett. 20(22):7197-7200. doi: 10.1021/acs.orglett.8b03123.

McBeath JH, Cheng M. 2004. First report of *Leptographium abietinum* associated with blue stain on declining western Siberian larch in Alaska. Plant Health Progress. 5(1). doi: 10.1094/PHP-2004-0326-01-HN.

McNeil BA, Semper C, Zimmerly S. 2016. Group II introns: versatile ribozymes and retroelements. Wiley interdisciplinary reviews. RNA. 7(3):341–355. https://doi.org/10.1002/wrna.1339

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39:W339-46.

Medina R, Franco ME, Bartel LC, Martinez Alcántara V, Saparrat MC, Balatti PA. 2020. Fungal mitogenomes: relevant features to planning plant disease management. Front Microbiol. 11:978. doi: 10.3389/fmicb.2020.00978.

Megarioti AH, Kouvelis VN. 2020. The coevolution of fungal mitochondrial introns and their homing endonucleases (GIY-YIG and LAGLIDADG). Genome Biol Evol. 12(8):1337–1354. doi: 10.1093/gbe/evaa126.

Michel F, Costa M, Westhof E. 2009. The ribozyme core of group II introns: a structure in want of partners. Trends Biochem Sci. 34:189–199. 10.1016/j.tibs.2008.12.

Michel F, Dujon B. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. EMBO J. 2(1):33-38. doi:10.1002/j.1460-2075.1983.tb01376.x

Michel F, Ferat JL. 1995. Structure and activities of group II introns. Annu Rev Biochem. 64:435-61. doi: 10.1146/annurev.bi.64.070195.002251.

Michel F, Jacquier A, Dujon B.1982. Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. Biochimie. 64(10):867-881. doi:10.1016/s0300-9084(82)80349-0

Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol. 216(3):585-610. doi: 10.1016/0022-2836(90)90386-Z.

Mohanta TK, Bae H. 2015. The diversity of fungal genome. Biol Proced Online. 17:8. doi: 10.1186/s12575-015-0020-z

Morelet M. 1988. Observations sur trois deutéromycètes inféodés aux pins. Annales de la Société des Sciences Naturelles et d'Archéologie de Toulon et du Var. 40(1):41-45.

Mosunova O, Navarro-Muñoz JC, Collemare J. 2021. The biosynthesis of fungal secondary metabolites: from fundamentals to biotechnological applications. Encyclopedia of Mycology. 2:458–476.

Mota EM, Collins RA. 1988. Independent evolution of structural and coding regions in a Neurospora mitochondrial intron. Nature. 332:654–656. 10.1038/332654a0.

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

Mukhopadhyay J, Hausner G. 2024. Interconnected roles of fungal nuclear- and intron-encoded maturases: at the crossroads of mitochondrial intron splicing. Biochem Cell Biol. 102(5):351-372. doi: 10.1139/bcb-2024-0046.

Mukhopadhyay J, Wai A, 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.

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

Mullineux ST, Willows K, Hausner G. 2011. Evolutionary dynamics of the mS952 intron: a novel mitochondrial group II intron encoding a LAGLIDADG homing endonuclease gene. J Mol Evol. 72(5-6):433–449. doi: 10.1007/s00239-011-9442-7.

Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, et al. 2020. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol. 16(1):60–68. doi: 10.1038/s41589-019-0400-9.

Nawrocki EP, Jones TA, Eddy SR. 2018. Group I introns are widespread in archaea. Nucleic Acids Res. 46(15):7970–7976. doi: 10.1093/nar/gky414.

Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics. 25(10):1335–1337.

Nel WJ, Wingfield MJ, de Beer ZW, Duong TA. 2021. Ophiostomatales fungi associated with wood-boring beetles in South Africa including two new species. Antonie van Leeuwenhoek. 114(6):667–686. doi: 10.1007/s10482-021-01548-0.

Neu R, Goffart S, Wolf K, Schäfer B. 1998. Relocation of urf a from the mitochondrion to the nucleus cures the mitochondrial mutator phenotype in the fission yeast *Schizosaccharomyces pombe*. Mol Gen Genet. 258:389–396. doi: 10.1007/s004380050746.

Nielsen JC, Grijseels S, Prigent S, Ji B, Dainat J, Nielsen KF, Frisvad JC, Workman M, Nielsen J. 2017. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in Penicillium species. Nat Microbiol. 2:17044.

Oh H, Jensen PR, Murphy BT, Fiorilla C, Sullivan JF, Ramsey T, Fenical W. 2010. Cryptosphaerolide, a cytotoxic Mcl-1 inhibitor from a marine-derived ascomycete related to the genus Cryptosphaeria. J Nat Prod. 73(5):998-1001. doi: 10.1021/np1000889.

Ojala D, Montoya J, Attardi G. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature. 290:470–474. doi: 10.1038/290470a0.

Ola ARB, Debbab A, Aly AH, Mandi A, Zerfass I, Hamacher A, Kassack MU, Brötz-Oesterhelt H, Kurtan T, Proksch P. 2013. Absolute configuration and antibiotic activity of neosartorin from the endophytic fungus *Aspergillus fumigatiaffinis*. Bioorg Med Chem. 21(19):5790-5796. doi:10.1016/j.bmc.2013.08.008.

Opuu V, Merleau NSC, Messow V, Smerlak M. 2022. RAFFT: Efficient prediction of RNA folding pathways using the fast Fourier transform. PLoS Comput Biol. 18(8):e1010448. doi: 10.1371/journal.pcbi.1010448.

Osbourn A. 2010. Gene clusters for secondary metabolic pathways: an emerging theme in plant biology. Plant Physiol. 154(2):531-5. doi: 10.1104/pp.110.161315.

Osiewacz HD. 2024. Impact of mitochondrial architecture, function, redox homeostasis, and quality control on organismic aging: Lessons from a fungal model system. Antioxid Redox Signal. 40(16-18):948-967. doi:10.1089/ars.2023.0487.

Paciura D, de Beer ZW, Jacobs K, Zhou XD, Ye H, Wingfield MJ. 2010. Eight new *Leptographium* species associated with tree-infesting bark beetles in China. Persoonia. 25:94-108. doi: 10.3767/003158510X551097.

Pan Y, Lu J, Zhou XD, Yu ZF, Chen P, Wang J, Ye H. 2020. Two new species of *Leptographium* associated with *Tomicus* spp. infesting *Pinus* spp. in Southwestern China. Int J Syst Evol Microbiol. 70(8):4798–4807. doi: 10.1099/ijsem.0.004349.

Parenteau J, Abou Elela S. 2019. Introns: Good day junk is bad day treasure. Trends Genet. 35(12):923–934. doi: 10.1016/j.tig.2019.09.010.

Pogoda CS, Keepers KG, Nadiadi AY, Bailey DW, Lendemer JC, Tripp EA, Kane NC. 2019. Genome streamlining via complete loss of introns has occurred multiple times in lichenized fungal mitochondria. Ecol Evol. 9(7):4245–4263. doi: 10.1002/ece3.5056.

Pramateftaki PV, Kouvelis VN, Lanaridis P, Typas MA. 2006. The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. FEMS Yeast Res. 6(1):77–90. doi: 10.1111/j.1567-1364.2005.00018.x.

Prince S, Munoz C, Filion-Bienvenue F, Rioux P, Sarrasin M, Lang BF. 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.

Rambaut A. 2014. FigTree: A graphical viewer of phylogenetic trees. Available from: http://tree.bio.ed.ac.uk/software/figtree/

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

Robey MT, Caesar LK, Drott MT, Keller NP, Kelleher NL. 2021. An interpreted atlas of biosynthetic gene clusters from 1,000 fungal genomes. Proc Natl Acad Sci USA. 118(19):e2020230118.

Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. 2020. Biosynthetic gene clusters and the evolution of fungal chemodiversity. Nat Prod Rep. 37(7):868-878. doi: 10.1039/c9np00045c.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61(3):539–542. doi: 10.1093/sysbio/sys029.

Rose AB. 2019. Introns as gene regulators: A brick on the accelerator. Front Genet. 9:672. doi: 10.3389/fgene.2018.00672.

Roy BA, Alexander HM, Davidson J, et al. 2014. Increasing forest loss worldwide from invasive pests requires new trade regulations. Front Ecol Environ. 12:457–465.

Rudan M, Dib BP, Musa M, Kanunnikau M, Sobočanec S, Rueda D, Warnecke T, Kriško A. 2018. Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. eLife. 7:e35330. doi: 10.7554/eLife.35330.

Rudski SM, Hausner G. 2012. The mtDNA *rps3* locus has been invaded by a group I intron in some species of *Grosmannia*. Mycoscience. 53:471–475. doi: 10.1007/s10267-012-0183-2.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics. 16(10):944–945. doi: 10.1093/bioinformatics/16.10.944.

Sakakibara Y, Brown M, Hughey R, Mian IS, Sjölander K, Underwood RC, Haussler D. 1994. Stochastic context-free grammars for tRNA modeling. Nucleic Acids Res. 22(23):5112–5120. doi: 10.1093/nar/22.23.5112.

Santini A, Migliorini D. 2022. Invasive alien plant pathogens: The need of new detection methods. Methods Mol Biol. 2536:111–118.

Sato K, Hamada M. 2023. Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery. Brief Bioinform. 24(4):bbad186. doi: 10.1093/bib/bbad186.

Sayari M, Steenkamp ET, van der Nest MA, Wingfield BD. 2018. Diversity and evolution of polyketide biosynthesis gene clusters in the Ceratocystidaceae. Fungal Biol. 122(9):856–866. doi: 10.1016/j.funbio.2018.04.011.

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res. 50(D1):D20-D26. doi: 10.1093/nar/gkab1112.

Sbaraini N, Da Silva NM, Gonçalves AR, et al. 2017. Genome-wide analysis of secondary metabolite gene clusters in Ophiostoma ulmi and Ophiostoma novo-ulmi reveals a fujikurin-like gene cluster with a putative role in infection. Front Microbiol. 8:1063. doi: 10.3389/fmicb.2017.01063.

Schikora-Tamarit MÀ, Marcet-Houben M, Nosek J, Gabaldón T. 2021. Shared evolutionary footprints suggest mitochondrial oxidative damage underlies multiple complex I losses in fungi. Open Biol. 11(4):200362. doi: 10.1098/rsob.200362.

Schmitz-Linneweber C, Lampe MK, Sultan LD, Ostersetzer-Biran O. 2015. Organellar maturases: a window into the evolution of the spliceosome. Biochim Biophys Acta Bioenerg. 1847:798–808. doi: 10.1016/j.bbabio.2015.01.009.

Schor R, Cox R. 2018. Classic fungal natural products in the genomic age: the molecular legacy of Harold Raistrick. Nat Prod Rep. 35(3):230–256.

Schueffler A, Anke T. 2014. Fungal natural products in research and development. Nat Prod Rep. 31(10):1425–1448.

Seifert KA. 1993. Sapstain of commercial lumber by species of Ophiostoma and Ceratocystis. In: Wingfield MJ, Seifert KA, Webber JF, editors. Ceratocystis and Ophiostoma: taxonomy, ecology and pathology. St. Paul: APS Press. p. 141–151.

Sellem CH, 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 CH, di Rago JP, Lasserre JP, Ackerman SH, 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.

Seshadri SR, Banarjee C, Barros MH, Fontanesi F. 2020. The translational activator Sov1 coordinates mitochondrial gene expression with mitoribosome biogenesis. Nucleic Acids Res. 48:6759–6774. doi: 10.1093/nar/gkaa424.

Sethuraman J, Majer A, Friedrich NC, Edgell DR, Hausner G. 2009. Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. Mol Biol Evol. 26(10):2299–2315. doi: 10.1093/molbev/msp145.

Shi Z, Wang B, Clarke SR, Sun J. 2012. Effect of associated fungi on the immunocompetence of red turpentine beetle larvae, Dendroctonus valens (Coleoptera: Curculionidae: Scolytinae). Insect Sci. 19(5):579–584. doi: 10.1111/j.1744-7917.2011.01484.x.

Shimizu Y, Ogata H, Goto S. 2017. Type III polyketide synthases: functional classification and phylogenomics. Chembiochem. 18(1):50–65. doi: 10.1002/cbic.201600522.

Shwab EK, Keller NP. 2008. Regulation of secondary metabolite production in filamentous ascomycetes. Mycol Res. 112(Pt 2):225-30.

Simon DM, Clarke NA, McNeil BA, Johnson I, Pantuso D, Dai L, Chai D, Zimmerly S. 2008. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. RNA. 14(9):1704–1713. doi: 10.1261/rna.1056108.

Six DL, Wingfield MJ. 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 DL. 2012. Ecological and evolutionary determinants of bark beetle-fungus symbioses. Insects. 3:339–366.

Smith DJ, Burnham MK, Edwards J, Earl AJ, Turner G. 1990. Cloning and heterologous expression of the penicillin biosynthetic gene cluster from *Penicillum chrysogenum*. Biotechnology. 8(1):39-41.

Song X, Geng Y, Xu C, Li J, Guo Y, Shi Y, Ma Q, Li Q, Zhang M. 2024. The complete mitochondrial genomes of five critical phytopathogenic Bipolaris species: features, evolution, and phylogeny. IMA Fungus. 15(1):15. https://doi.org/10.1186/s43008-024-00149-6.

Spatafora JW, Blackwell M. 1994. The polyphyletic origins of Ophiostomatoid fungi. Mycol Res. 98(1):1–9.

Stahley MR, Strobel SA. 2006. RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. Curr Opin Struct Biol. 16(3):319–326. https://doi.org/10.1016/j.sbi.2006.04.005.

Stark BC, Kole R, Bowman EJ, Altman S. 1978. Ribonuclease P: an enzyme with an essential RNA component. Proc Natl Acad Sci. 75:3717–3721. doi: 10.1073/pnas.75.8.3717.

Stoddard BL. 2005. Homing endonuclease structure and function. Q Rev Biophys. 38(1):49–95. doi: 10.1017/S0033583505004063.

Stoddard BL. 2014. Homing endonucleases from mobile group I introns: discovery to genome engineering. Mob DNA. 5(1):7. doi: 10.1186/1759-8753-5-7.

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. J Mol Evol. 49:169–181. doi: 10.1007/PL00006540.

Stone CL, Frederick RD, Tooley PW, Luster DG, Campos B, Winegar RA, et al. 2018. Annotation and analysis of the mitochondrial genome of *Coniothyrium glycines*, causal agent of red leaf blotch of soybean, reveals an abundance of homing endonucleases. PLoS One. 13:e0207062. doi: 10.1371/journal.pone.0207062.

Strzałka B, Jankowiak R, Bilański P, Patel N, Hausner G, Linnakoski R, Solheim H. 2020. Two new species of Ophiostomatales (Sordariomycetes) associated with the bark beetle *Dryocoetes alni* from Poland. MycoKeys. 68:23–48. doi: 10.3897/mycokeys.68.50035.

Sweeney BA, Hoksza D, Nawrocki EP, Ribas CE, Madeira F, Cannone JJ, Gutell R, Maddala A, Meade CD, Williams LD, et al. 2021. R2DT is a framework for predicting and visualising RNA secondary structure using templates. Nat Commun. 12(1):3494. doi: 10.1038/s41467-021-23555-5.

Sweeney MJ, Dobson ADW. 1998. Mycotoxin production by *Aspergillus*, *Fusarium* and *Penicillium* species. *Int J Food Microbiol.* 43(3):141–158. doi: 10.1016/S0168-1605(98)00112-3.

Takahashi S, Nakagawa K, Tanaka A, Ogura K, Seto S. 1998. Isolation and structural elucidation of ophiocordin, a novel antifungal compound produced by *Ophiostoma piliferum.* J Antibiot (Tokyo). 51(6):523–528. doi: 10.7164/antibiotics.51.523.

Takeda I, Umemura M, Koike H, Asai K, Machida M. 2014. Motif-independent prediction of a secondary metabolism gene cluster using comparative genomics: application to sequenced genomes of Aspergillus and ten other filamentous fungal species. DNA Res. 21(4):447–457.

Takeuchi R, Lambert AR, Mak ANS, Jacoby K, Dickson RJ, Gloor GB, et al. 2011. Tapping natural reservoirs of homing endonucleases for targeted gene modification. Proc Natl Acad Sci USA. 108:13077–13082. doi: 10.1073/pnas.1107719108.

Teixeira MM, de Almeida LG, Kubitschek-Barreira P, Alves FL, Kioshima ES, Abadio AK, Fernandes L, Derengowski LS, Ferreira KS, Souza RC, et al. 2014. Comparative genomics of the major fungal agents of human and animal Sporotrichosis: *Sporothrix schenckii* and *Sporothrix brasiliensis*. BMC Genomics. 15:943. doi: 10.1186/1471-2164-15-943.

The RNAcentral Consortium, Petrov AI, Kay SJE, Kalvari I, Howe KL, Gray KA, Bruford EA, Kersey PJ, Cochrane G, Finn RD, Bateman A. 2017. RNAcentral: a comprehensive database of non-coding RNA sequences. Nucleic Acids Res. 45(D1):D128–D134. doi: 10.1093/nar/gkw1008.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA. 7(8):1142–1152. doi:10.1017/s1355838201010251.

Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM. 2010. Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. RNA. 16(1):57–69.

Toor N, Keating KS, Taylor SD, Pyle AM. 2008. Crystal structure of a self-spliced group II intron. Science. 320(5872):77–82. doi: 10.1126/science.1153803.

Toor N, Zimmerly S. 2002. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA. 8(11):1373–1377. doi:10.1017/s1355838202023087.

Trail F, Mahanti N, Rarick M, Mehigh R, Liang SH, Zhou R, Linz JE. 1995. Physical and transcriptional map of an aflatoxin gene cluster in *Aspergillus parasiticus* and functional disruption of a gene involved early in the aflatoxin pathway. Appl Environ Microbiol. 61(7):2665-73.

Trollip C, Carnegie AJ, Dinh Q, Kaur J, Smith D, Mann R, Rodoni B, Edwards J. 2021. Ophiostomatoid fungi associated with pine bark beetles and infested pines in south-eastern Australia, including *Graphilbum ipis-grandicollis* sp. nov. IMA Fungus. 12(1):24.

Trollip C, Carnegie AJ, Rodoni B, Edwards J. 2022. Draft genome sequences for three Ophiostoma species acquired during revisions of Australian plant pathogen reference collections. Microbiol Res Announc. 11(6):e0017522. https://doi.org/10.1128/mra.00175-22.

Turk EM, Das V, Seibert RD, Andrulis ED. 2013. The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. PLoS One. 8:e78105. doi: 10.1371/journal.pone.0078105.

Upadhyay HP. 1981. A monograph of *Ceratocystis* and *Ceratocystiopsis*. Athens: University of Georgia Press. 176 pp.

Uzunovic A, Byrne T. 2013. Wood market issues relating to blue-stain caused by Ophiostomatoid fungi in Canada. In: Seifert KA, de Beer ZW, Wingfield MJ, editors. Ophiostomatoid Fungi: Expanding Frontiers. Utrecht: CBS-KNAW Fungal Biodiversity Centre. p. 201–212.

Uzunovic A, Yang DQ, Gagné P, Breuil C, Bernier L, Byrne A, Gignac M, Kim SH. 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, Pfeiffer I, Kucsera J, Tomaska L, Lang BF, et al. 2011. Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. Nucleic Acids Res. 39(10):4202–4219. doi: 10.1093/nar/gkq1345.

van den Berg MA, Westerlaken I, Leeflang C, Kerkman R, Bovenberg R. 2007. Functional characterization of the penicillin biosynthetic gene cluster of *Penicillium chrysogenum* Wisconsin54-1255. Funct Gen Biol. 44(9):830–844.

Van der Nest MA, Beirn LA, Crouch JA, Demers JE, Wilhelm de Beer Z, De Vos L, et al. 2014. Draft genomes of *Amanita jacksonii, Ceratocystis albifundus, Fusarium circinatum, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana, and Sclerotinia echinophila*. IMA Fungus. 5:473–486. doi: 10.5598/imafungus.2014.05.02.11.

Vanderpool D, Bracewell RR, McCutcheon JP. 2018. Know your farmer: Ancient origins and multiple independent domestications of ambrosia beetle fungal cultivars. Mol Ecol. 27(8):2077–2094. doi: 10.1111/mec.14394.

Varassas SP, Kouvelis VN. 2022. Mitochondrial transcription of entomopathogenic fungi reveals evolutionary aspects of mitogenomes. Front Microbiol. 13:821638. https://doi.org/10.3389/fmicb.2022.821638.

Vicens Q, Cech TR. 2006. Atomic level architecture of group I introns revealed. Trends Biochem Sci. 31(1):41–51. https://doi.org/10.1016/j.tibs.2005.11.008.

Vicens Q, Paukstelis PJ, Westhof E, Lambowitz AM, Cech TR. 2008. Toward predicting self-splicing and protein-facilitated splicing of group I introns. RNA. 14(10):2013–2029. https://doi.org/10.1261/rna.1027208.

von Bargen KW, Niehaus EM, Krug I, Bergander K, Würthwein EU, Tudzynski B, Humpf HU. 2015. Isolation and structure elucidation of Fujikurins A-D: Products of the PKS19 gene cluster in *Fusarium fujikuroi*. J Nat Prod. 78(8):1809–1815. doi: 10.1021/np5008137.

Vu D, Groenewald M, de Vries M, Gehrmann T, Stielow B, Eberhardt U, et al. 2019. Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. Stud Mycol. 92:135–154. doi: 10.1016/j.simyco.2018.05.001.

Wagener WW, Mielke JL. 1961. A staining fungus root disease in ponderosa, Jeffrey, and pinyon pines. Plant Dis Rep. 45(10):831–835.

Wai A, Hausner G. 2021. The mitochondrial genome of *Ophiostoma himal-ulmi* and comparison with other fungi causing Dutch elm disease. Can J Micro. 67(8): 584-598.

Wai A, Hausner G. 2022. The compact mitogenome of *Ceratocystiopsis pallidobrunnea*. Can J Micro. 68(9): 569–575. https://doi.org/10.1139/cjm-2022-0038.

Wai A, Shen C, Carta A, Dansen A, Crous PW, 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(3):573–584. doi: 10.1080/24701394.2019.1580272.

Wallweber GJ, Mohr S, Rennard R, Caprara MG, Lambowitz AM. 1997. Characterization of Neurospora mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3' splice site for an intron ORF. RNA. 3(2):114–131.

Wang G, Lin J, Shi Y, Chang X, Wang Y, Guo L, et al. 2019. Mitochondrial genome in *Hypsizygus marmoreus* and its evolution in Dikarya. BMC Genomics. 20:765. doi: 10.1186/s12864-019-6133-z.

Wang R, Dong L, Chen Y, Qu L, Li E, Wang Q, Zhang Y. 2017. The complete mitochondrial genome of nematophagous fungus Esteya vermicola. Mitochondrial DNA B Resour. 2(1):196-197. doi: 10.1080/23802359.2017.1307700.

Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, Wohlleben W. 2009. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. J Biotechnol. 140(1-2):13-7.

Wick RR, Schult MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 31(20):3350–3352. https://doi.org/10.1093/bioinformatics/btv383.

Wiemann P, Sieber CM, von Bargen KW, Studt L, Niehaus EM, Espino JJ, Huß K, Michielse CB, Albermann S, Wagner D, et al. 2013. Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. PLoS Pathog. (6):e1003475. doi: 10.1371/journal.ppat.1003475.

Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W, Crouch JA, de Beer ZW, De Vos L, Duong TA, Fields CJ, et al. 2015a. IMA Genome-F 4: Draft genome sequences of *Chrysoporthe austroafricana*, *Diplodia scrobiculata*, *Fusarium nygamai*, *Leptographium lundbergii*, *Limonomyces culmigenus*, *Stagonosporopsis tanaceti*, and *Thielaviopsis punctulata*. IMA Fungus. 6(1):233–248. doi:10.5598/imafungus.2015.06.01.15.

Wingfield BD, Ambler JM, Coetzee MP, de Beer ZW, Duong TA, Joubert F, Hammerbacher A, McTaggart AR, Naidoo K, Nguyen HD, et al. 2016a. IMA Genome-F 6: Draft genome sequences of *Armillaria fuscipes*, *Ceratocystiopsis minuta*, *Ceratocystis adiposa*, *Endoconidiophora laricicola*, *E. polonica*, and *Penicillium freii* DAOMC 242723. IMA Fungus. 7(1):217–227. doi: 10.5598/imafungus.2016.07.01.11.

Wingfield BD, Barnes I, de Beer ZW, De Vos L, Duong TA, Kanzi AM, Naidoo K, Nguyen HD, Santana QC, Sayari M, et al. 2015b. IMA Genome-F 5: Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysoporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum*, and *Thielaviopsis musarum*. IMA Fungus. 6(2):493–506. doi: 10.5598/imafungus.2015.06.02.13.

Wingfield BD, Berger DK, Steenkamp ET, Lim HJ, Duong TA, Bluhm BH, de Beer ZW, De Vos L, Fourie G, Naidoo K, et al. 2017a. Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. IMA Fungus. 8(2):385–396. https://doi.org/10.5598/imafungus.2017.08.02.10.

Wingfield BD, Duong TA, Hammerbacher A, van der Nest MA, Wilson A, Chang R, de Beer ZW, Steenkamp ET, Wilken PM, Naidoo K, et al. 2016b. Draft genome sequences for *Ceratocystis fagacearum*, *C. harringtonii*, *Grosmannia penicillata*, and *Huntiella bhutanensis*. IMA Fungus. 7(2):317–323. https://doi.org/10.5598/imafungus.2016.07.02.11.

Wingfield MJ, Barnes I, de Beer ZW, Roux J, Wingfield BD, Taerum SJ. 2017b. Novel associations between ophiostomatoid fungi, insects, and tree hosts: current status—

future prospects. Biological Invasions. 19:3215–3228. https://doi.org/10.1007/s10530-017-1468-3.

Wingfield MJ, de Beer ZW, Slippers B, Wingfield BD, Groenewald JZ, Lombard L, Crous PW. 2012. One fungus, one name promotes progressive plant pathology. Molecular Plant Pathology. 13(6):604–613. https://doi.org/10.1111/j.1364-3703.2011.00768.x.

Wingfield MJ, Seifert KA, Webber JF. 1993. *Ceratocystis* and *Ophiostoma*: taxonomy, ecology, and pathogenicity. St. Paul: American Phytopathological Society. 293 pp.

Wu B, Hao W. 2014. Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. G3 Genes Genomes Genet. 4:605–612. doi:10.1534/g3.113.009910.

Wu B, Hao W. 2019. Mitochondrial-encoded endonucleases drive recombination of protein-coding genes in yeast. Environ Microbiol. 21:4233–4240. doi:10.1111/1462-2920.14783.

Xu Y, Vinas M, Alsarrag A, Su L, Pfohl K, Rohlfs M, Schäfer W, Chen W, Karlovsky P. 2019. Bis-naphthopyrone pigments protect filamentous ascomycetes from a wide range of predators. Nat Commun. 10(1):3579. doi: 10.1038/s41467-019-11377-5.

Yin M, Wingfield MJ, Zhou X, de Beer ZW. 2016. Multigene phylogenies and morphological characterization of five new *Ophiostoma* spp. associated with spruce-infesting bark beetles in China. Fungal Biol. 120(4):454–470.

Yin M, Wingfield MJ, Zhou X, Linnakoski R, de Beer ZW. 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. doi:10.3897/mycokeys.60.39069.

Yu J, Chang PK, Ehrlich KC, Cary JW, Bhatnagar D, Cleveland TE, Payne GA, Linz JE, Woloshuk CP, Bennett JW. 2004. Clustered pathway genes in aflatoxin biosynthesis. Appl Environ Microbiol. 70(3):1253–1262.

Zaman R, May C, Ullah A, Erbilgin N. 2023. Bark beetles utilize Ophiostomatoid fungi to circumvent host tree defenses. Metabolites. 13(2):239. doi:10.3390/metabo13020239.

Zaug AJ, Grabowski PJ, Cech TR. 1983. Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. Nature. 301(5901):578–583. doi:10.1038/301578a0.

Zhang D, Li S, Fan M, Zhao C. 2022. The novel compounds with biological activity derived from soil fungi in the past decade. Drug Des Devel Ther. 16:3493-3555. doi: 10.2147/DDDT.S377921.

Zhang J, Liu N, Cacho RA, Gong Z, Liu Z, Qin W, Tang C, Tang Y, Zhou J. 2016. Structural basis of nonribosomal peptide macrocyclization in fungi. Nat Chem Biol. 12(12):1001–1003. doi:10.1038/nchembio.2202.

Zhang S, Zhang YJ, Li ZL. Complete mitogenome of the entomopathogenic fungus *Sporothrix insectorum* RCEF 264 and comparative mitogenomics in Ophiostomatales. Appl Microbiol Biotechnol. 2019 Jul;103(14):5797-5809. doi: 10.1007/s00253-019-09855-3.

Zhang S, Zhang YJ. 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 YJ, Fan XP, Li JN, Zhang S. 2023. Mitochondrial genome of *Cordyceps blackwelliae*: organization, transcription, and evolutionary insights into *Cordyceps*. IMA Fungus. 14(1):13. doi:10.1186/s43008-023-00118-5.

Zhgun AA. 2023. Fungal BGCs for production of secondary metabolites: main types, central roles in strain improvement, and regulation according to the Piano Principle. Int J Mol Sci. 24(13):11184. doi:10.3390/ijms241311184.

Zhou Y, Lu C, Wu QJ, Wang Y, Sun ZT, Deng JC, Zhang Y. 2008. GISSD: group I intron sequence and structure database. Nucleic Acids Res. 36(suppl_1):D31–D37.

Ziemert N, Alanjary M, Weber T. 2016. The evolution of genome mining in microbes – a review. Nat Prod Rep. 33:988–1005.

Zimmerly S, Semper C. 2015. Evolution of group II introns. Mobile DNA. 6:7. doi:10.1186/s13100-015-0037-5.

Zipfel RD, de Beer ZW, Jacobs K, Wingfield BD, Wingfield MJ. 2006. Multi-gene phylogenies define *Ceratocystiopsis* and *Grosmannia* distinct from *Ophiostoma*. Stud Mycol. 55:75–97. doi:10.3114/sim.55.1.75.

Zubaer A, Wai A, Hausner G. 2018. The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. Sci Rep. 8(1):17591. doi:10.1038/s41598-018-35926-y.

Zubaer A, Wai A, Hausner G. 2019. The fungal mitochondrial Nad5 pan-genic intron landscape. Mitochondrial DNA A DNA Mapp Seq Anal. 30(8):835–842. doi:10.1080/24701394.2019.1687691.

Zubaer A, Wai A, Hausner G. 2024. Comparative mitogenomics of *Leptographium procerum, Leptographium terebrantis*, and *Leptographium wingfieldii*, an invasive fungal species in Canadian forests. Can J Microbiol. doi: 10.1139/cjm-2024-0179.

Zubaer A, Wai A, Patel N, Perillo J, 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.

Zuker M, Stiegler P. 1981. Optimal computer folding of large RNA sequences using thermodynamic and auxiliary information. Nucleic Acids Res. 9(1):133–148.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415. doi:10.1093/nar/gkg595.

Zumkeller S, Gerke P, 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.