A study on the intron landscape of the cytochrome b gene, and mitochondrial-encoded N-acetyltransferase and ribosomal protein S3 genes of fungi in the subphylum Pezizomycotina

by

Alvan Wai

A Thesis to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Microbiology

University of Manitoba

Winnipeg, Manitoba, Canada

Abstract

The Master's program was based on mainly *in silico* work, employing several bioinformatics techniques/tools. The main focus of the research involved comparative sequence analysis of fungal mitochondrial genomes, which has previously been noted to be a rich source of mobile elements. The work is divided into two projects. The first was the study of intron distribution within the cytochrome b gene of filamentous fungi of the phylum Ascomycota. Cytochrome b genes (129) were sampled from the public database, GenBank, and compared via multiple sequence alignments. A total of 21 unique intron insertion sites were observed. Two introns displaying unique intron arrangements were explored in greater depths. One appears to be capable of alternative splicing, the other encodes an open reading frame interrupted by an intron. The results of the first project demonstrate the diversity of intron insertions within a single gene and its impact on genome evolution, and the potential to uncover novel introns. The second project was based on prior research that noted the presence/absence of N-acetyltransferase and ribosomal protein S3 genes within fungal mitochondrial genomes of the Ascomycota. In order to determine the uniqueness of this presence/absence, the protein sequences of the two genes were sampled from two different databases, GenBank and MycoCosm (Joint Genome Institute). Sequences were retrievable from two different compartments, nucleus and mitochondria. The two data sets were separately aligned, and each were subjected to phylogenetic analyses. The results suggest separate origins for the two genes. The N-acetyltransferase gene appears to have a nuclear origin, but some fungi have obtained a mitochondrial version(s). The ribosomal protein S3 gene appears to be of mitochondrial origin, but some fungi appear to encode two copies of separate origin (mitochondrial or nuclear) in the nuclear genome. The underlying mechanism of transfer of genetic elements between nucleus and mitochondria has yet been elucidated.

Acknowledgements

The Master's project was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Graduate Enhancement of Tri-Council Stipends (GETS) program of the Faculty of Graduate Studies (University of Manitoba; Winnipeg, Manitoba, Canada).

Much appreciation goes to Dr. Steven Zimmerly (University of Calgary; Calgary, Alberta, Canada) for all the advice he gave with regards to modeling of group II introns.

Much appreciation also goes to my committee members, Drs. Deborah Court and Sean McKenna for all the advice they have given me throughout the Master's program.

We would like to acknowledge Hay M, La D, Carta A, Chen T, and Corkery T, for all their contributions to the cytochrome b intron landscape project.

Valuable comments by two anonymous reviewers of the paper, that Chapter 2 was based on, were greatly appreciated.

I am very grateful for everyone, both past and present, that has accompanied me through my Master's project. Everyone has taught me a lot with regards to not only to materials related to the laboratory but also to various aspects of life.

A special acknowledgement goes my parents and siblings for their support and patience during the Master's program.

My deepest gratitude and appreciation goes to Dr. Georg Hausner. I would like to especially thank him for his patience, understanding, and vast insight of not only research but life as well.

Table of contents

List of tables

List of figures

List of abbreviations

aa	amino acid(s)
AAT	aminotransferase
AT	adenine, thymine
atp/ATP	adenosine triphosphate (gene)
blast(n/p)	Basic Local Alignment Search Tool (for nucleotides/proteins)
Bot	Botryosphaeriales
bp	base pair(s)
Cap	Capnodiales
cDNA	complementary DNA
CoA	coenzyme A
cox<number>	cytochrome c oxidase (gene), where <number> represents the subunit number
cpDNA	chloroplast genome
cytb	cytochrome b (gene)
d	derived or degenerated when referring to introns or proteins, respectively
Dia	Diaporthales
DNA	deoxyribonucleic acid
DSBR	double-strand break repair
EBS	exon binding site
Eur	Eurotiales
f	fused
G/GY	GIY-YIG homing endonuclease

GC	guanine, cytosine
Glo	Glomerellales
GNAT	Gcn5-related N-acetyltransferase
HE	homing endonuclease
HEG	homing endonuclease gene
Hel	Helotiales
His-Cys	histidine-cysteine
Hyp	Hypocreales
i	intron
IBS	intron binding site
IEP	intron-encoded protein
IGS	internal guide sequence
IS	insertion site
JGI	Joint Genome Institute
kb	kilobase pair(s)
L/LAG/LHE	LAGLIDADG homing endonuclease
Lec	Lecanorales
Mb	megabase pairs
Mic	Microascales
mL	millilitre(s)
mL	mitochondrial large ribosomal subunit gene
mRNA	mature RNA
mt	mitochondrial

mtDNA	mitochondrial genome
n	nuclear
N/NAT	N-acetyltransferase (gene)
NAD(P)H	nicotinamide adenine dinucleotide dehydrogenase subunit (phosphate)
nad<number>	nicotinamide adenine dinucleotide dehydrogenase subunit (gene), where <number> represents the subunit number
NCBI	National Center for Biotechnology Information
nt	nucleotide
OH	hydroxyl group
Ony	Onygenales
Oph	Ophiostomatales
P	pairing
PCR	polymerase chain reaction
Pel	Peltigerales
Pez	Pezizales
Ple	Pleosporales
pmol	picomole(s)
PYG	peptone, yeast extract, glucose
QoI	quinone outside inhibitor
rDNA	ribosomal DNA
RNA	ribonucleic acid
rnl	large ribosomal subunit (gene)
rns	small ribosomal subunit (gene)

rps3/RPS3	ribosomal protein S3 (gene/protein)
rRNA	ribosomal RNA
RT	reverse transcriptase
Sac	Saccharomycetales
senDNA	senescent DNA
Sor	Sordariales
trn<letter>	transfer RNA (gene), where **<letter>** represents the amino acid encoded by the tRNA; **A** = alanine; **C** = cysteine; **D** = aspartate; **E** = glutamate; **F** = phenylalanine; **G** = glycine; **H** = histidine; **I** = isoleucine; **K** = lysine; **L** = lysine; **M** = methionine; **N** = asparagine; **P** = proline; **Q** = glutamate; **R** = arginine; **S** = serine; **T** = threonine; **V** = valine; **W** = tryptophan; **Y** = tyrosine
tRNA	transfer RNA
Vsr	very short patch repair
WAG	Whelan and Goldman
Xyl	Xylariales

Chapter 1. Literature review

1.1. The Fungi

Fungi are a diverse group of eukaryotic organisms, some of which have environmental and/or economic importance. They are ubiquitous and highly versatile. Fungi are heterotrophs (osmotrophs), obtaining their nutrients by the secretion of enzymes and uptake through their cell wall. They play essential roles in ecosystems including in decomposition, nutrient cycling, and nutrient transport (Dighton, 2007). They can range from being pathogens to symbionts. Several fungi have applications in foods (agriculture, processing, spoilage, etc.), biocontrol, and medicine (disease, pharmaceutics, poisons/toxins) (Guijarro et al., 2017; Katoch and Pull, 2017; Kernaghan et al., 2017; Macías-Rubalcava and Sánchez-Fernández, 2017; Narladkar et al., 2015; Oro et al., 2018; Østergaard and Olsen, 2011; Ribera et al., 2017). Given their biodiversity and impact in the environment and society, much remains unknown of this group of organisms and thus warrants further study.

The definition of fungi and organisms that belong to this group, is currently not strictly defined; there does not seem to be features or traits that are unique to this group of organisms (i.e. absence of synapomorphy; a characteristic or trait derived from an ancestor) (reviewed by Richards et al., 2017). They are highly diverse in ecology and morphology (can exist as haploid or diploid, asexual, sexual, or both or parasexual, and can be yeast-like, filamentous, mushroom-like, etc.). Furthermore, environmental conditions can influence the morphology of fungi. This makes fungi difficult to classify based on traditional classification methods, which were based on phenotypic and functional observations, and were laborious and time-consuming. Modern

classification methods, based on DNA sequences, have revolutionized the study of fungal biodiversity. Phylogenetic study using molecular data eventually separated several organisms that were once considered fungi, including the Oomycota (also known as water molds) and the slime molds (a polyphyletic group). For simplicity, the term fungi will be used to refer to organisms that belong to subkingdom, Dikarya, with an emphasis on fungi belonging in the phylum Ascomycota (unless otherwise stated), for the remainder of this thesis.

1.2. Fungal mitochondrial genomes

The mitochondrion, sometimes referred to as the "powerhouse" of the cell, is best known for its essential role in energy metabolism, although it is also important in carbohydrate, fatty acid, nucleic acid, and protein metabolism, apoptosis, disease, etc. Mitochondria house their own genome, mainly encoding genes essential for protein synthesis and energy production. Fungal mitochondrial genomes (mtDNAs) are interesting with regards to size, which can vary by more than 10-fold; the smallest (18 844 bp), seen in the yeast, *Hanseniaspora uvarum* (Pramateftaki et al., 2006), and the largest reported so far (235 849 bp) in the filamentous Basidiomycota fungus, *Rhizoctonia solani* (Losada et al., 2014). With regards to filamentous fungi of the Ascomycota (the main focus of the thesis), the smallest mtDNA sequenced and publicly available thus far (July 2017) is from *Zasmidium cellare* at 23 743 bp (Goodwin et al., 2016) and the largest is from *Sclerotinia borealis* at 203 051 bp (Mardanov et al., 2014).

In the majority of currently sequenced fungi of the phylum Ascomycota, mtDNAs are typically composed of 14 protein-coding genes, which encode subunits involved in the electron transport chain (seven subunits of complex I: *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*; three subunits of complex III: *cox1*, *cox2*, and *cox3*; one subunit of complex IV: *cytb*), oxidative

phosphorylation (three subunits of ATP synthase: *atp6*, *atp8*, *atp9*), and one protein involved in protein synthesis (ribosomal protein: *rps3*), and non-coding RNA genes encoding typically at least 22 tRNA genes (two *trnL* and *trnR* genes) encoding for the 20 standard amino acids (*trnA* to *W*) and two rRNA genes encoding for the small (*rns*) and large (*rnl*) ribosomal subunits required for protein synthesis (summarized in Table 1.1). An exception to the "typical" composition of genes was observed in some species of yeast, where the set of NADH dehydrogenase subunit genes (complex I) may be absent (reviewed by de Zamaroczy and Bernardi, 1986; Sankoff et al., 1992; Zivanovic et al., 2005). Interestingly, these fungi possess nuclear-encoded alternative NADH:ubiquinone oxidoreductases (NADH dehydrogenases), which can act as alternatives for NAD(P)H oxidation carried out by complex I (De Vries et al., 1992; Overkamp et al., 2000; Videira and Duarte, 2002 Carneiro et al., 2004). Additionally, genes encoding for ATP synthase subunit 9 (*atp9*) and ribosomal protein S3 (*rps3*) appear to be present in some but not all mtDNAs. In cases where these genes are absent, the corresponding genes were predicted/shown to be nuclear-encoded (Lavín et al., 2008; Déquard-Chablat et al., 2011; Sellem et al., 2016). Furthermore, the *atp9* gene has been shown to be both nuclear- and mitochondrial-encoded in some species of fungi including *Aspergillus nidulans* (reviewed by Brown et al., 1985) and *Neurospora crassa* (van den Boogaart et al., 1982). One hypothesis is that such encoding pattern may be present in other fungi as well. The *rps3* gene was found to be absent from the mtDNAs of two fungi in the order Capnodiales (Torriani et al., 2008; Goodwin et al., 2016). In addition, a few fungal mtDNAs appear to encode a putative N-acetyltransferase (Duò et al., 2012, Zhang et al., 2017), a gene that has not been documented (to the author's knowledge) in fungal mtDNAs prior to work by Duò et al. (2012). Thus, gene content appears to be dynamic and, in part, gives rise to the mtDNA size variability observed in fungi.

Table 1.1. Summary of "typical" gene content in mitochondrial genomes of fungi in the phylum Ascomycota.

	Function	**Genes**
protein	electron transport chain	Complex I: *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6* Complex III: *cox1*, *cox2*, *cox3* Complex IV: *cytb*
	oxidative phosphorylation	ATP synthase: *atp6*, *atp8*, *atp9*
	protein synthesis	ribosomal protein: *rps3*
RNA	protein synthesis	tRNA: *A* to *W* rRNA: *rns*, *rnl*

1.2.1. Size variation of mitochondrial genomes

In general, mitochondrial genomes are highly variable in gene arrangement and size, which can range from 6 kb to greater than 3.9 Mb (Andersson et al., 2003; Burger et al., 2003). The large variation in sizes can be, in part, attributed to gene content (including gene or gene fragment duplications and inactivation), introns, and intergenic regions (Losada et al., 2014). With regards to the fungi, introns can significantly influence the size of mtDNA. Approximately 75% of the *Podospora anserina* mtDNA (Cummings et al., 1990) is composed of introns. Introns in *Agaricus bisporus* (Férandon et al., 2013) and *Sclerotinia borealis* (Mardanov et al., 2014), two of the larger fungal mtDNA sequenced thus far (sized at 135 005 bp and 203 051 bp, respectively), form 45.3% (61 092 bp) and 61.8% (125 394 bp) of their respective mtDNAs. Within a single gene, the number of introns can vary from 0 to as much as 19, in the case of the cytochrome c oxidase subunit I gene (*cox1*) of *Agaricus bisporus* (Férandon et al., 2010).

Mitochondrial introns can be composed of two elements. The first element is an intervening sequence (intron) that, when transcribed along with its host, is capable of self-removal from the site where it is situated; it is a self-splicing (autocatalytic) RNA molecule or ribozyme. The second element is an open reading frame (ORF) that potentially encodes an intron-encoded protein (IEP), which typically encodes a homing endonuclease or reverse transcriptase. The presence of these elements can present a challenge in the case of gene/genome annotations as the distribution of these elements appears to be sporadic/“patchy” (Yin et al., 2012; Guha et al., 2017b).

1.2.2. Self-splicing introns

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

1.2.2.1. Group I introns

Group I introns are the largest group of self-splicing introns, members of which can be found in both prokaryotes and eukaryotes (reviewed by Haugen et al., 2005; Vicens and Cech, 2006). They can be seen in all domains of life. Currently, they have been observed in bacteriophages, recently in archaea, bacteria, the nuclear genome (only in rRNA genes) of fungi and ciliates, and the organellar genome of protists, placozoans, corals, sea anemones, sponge, fungi, and plants (Belfort et al., 2002; Medina et al., 2006; Rot et al., 2006; Fukami et al., 2007; Signorovitch et al., 2007; Wang and Lavrov, 2008; reviewed by Hedberg and Johansen, 2013; Emblem et al., 2014; Nawrocki et al., 2018).

Group I introns are classified into subgroups (IA to IE) based on conserved primary and secondary structures (Michel and Westhof, 1990), IC1 and IE being more commonly found in nuclear genomes (rDNA) while IA, IB, IC2, IC3, and ID are more frequently seen in organellar genomes (Suh et al, 1999; Jackson et al., 2009). A group I intron RNA fold is typically composed of ten conserved helices, denoted as P1 to P10, although not all helices may be present in a given intron (reviewed by Lang et al., 2007). Helices P1 and P10 are formed by both exon and intron sequences (Davies et al., 1982; Michel and Westhof, 1990; Ritchings and Lewin, 1992). This underlies the importance of conservation of the exon sequences immediately flanking an intron; mutations to these sequences will affect important base-pairing involved in the formation of a splicing-competent intron. Helices P10-P1-P2 (substrate domain), P5-P4-P6 (scaffold domain), and P9-P7-P3-P8 (catalytic domain) form three larger, coaxially stacked, helical domains stabilized by base stacking interactions and tertiary hydrogen bonding (Adams et al., 2004b; Vicens and Cech, 2006; Hedberg and Johansen, 2013). This arrangement aligns the substrates (5′- and 3′-exons) close to each other creating a necessary scaffold for intron excision.

Splicing is carried out by a two-step transesterification reaction. The first transesterification step begins with the binding of an exogenous guanosine to the cleft formed by the catalytic domain, termed the guanosine binding site. The reaction proceeds with a nucleophilic attack made by the 3′ hydroxyl group (-OH) of the exogenous guanosine to the scissile phosphate at the 5′ splice junction. The exogenous guanosine forms a 3′-5′ phosphodiester bond with the 5′-end of the intron, leaving a free 3′-OH on the 5′-exon. The second transesterification step proceeds via a nucleophilic attack by the free 3′-OH of the 5′-exon to the scissile phosphate at the 3′ splice junction. This ultimately leads to the ligation of the two exon fragments and excision of a linear intron (Cech et al., 1981; Stahley and Strobel, 2005). The

splicing reaction can be enhanced by IEPs. An estimated 30% of group I introns contain an ORF which encodes an IEP (reviewed by Chevalier and Stoddard, 2001). The ORFs are typically confined to peripheral elements (ex. loops) of the intron, where they are less likely to interfere with formation of the catalytic core (reviewed by Lambowitz et al., 1999).

1.2.2.2. Group II introns

Group II introns are commonly encountered in plant chloroplast genomes and they are also found in archaeal and bacterial genomes and some organellar genomes of protists, fungi, plants, Placozoa, Porifera (demosponges), and the mtDNA of an annelid worm, *Nephtys*. Their occurrence in archaea is rare, and currently, there has been none found in nuclear genomes (Dellaporat et al., 2006; Vallès et al., 2008; Lambowitz and Zimmerly, 2011; Schuster et al., 2017).

Group II introns can be classified into subgroups IIA, IIB, and IIC based on the presence/absence of specific conserved sequences and structural features, including specific base-pairing patterns. The secondary structure of group II introns is organized into a wheel with six highly conserved stem-loop domains, denoted as domains I to VI (see Figure 1.1; reviewed by Michel et al., 2009). Short, conserved sequences (exon binding sites) within domain I form tertiary contacts to corresponding sequences (intron binding sites) of the 5′-exon, locking the exon in position. Domain I (which can be further divided into subdomains, denoted as "Ia" to "Id") interacts with domain V and is thought to function as a scaffold in the formation of the catalytic core (Koch et al., 1992; reviewed by Qin and Pyle, 1998). Domain II plays a pivotal role in recruiting domain III and partially direct folding of the catalytic core (Fedorova et al., 2003). Domain III interacts with the base of domain II and is involved in enhancing splicing

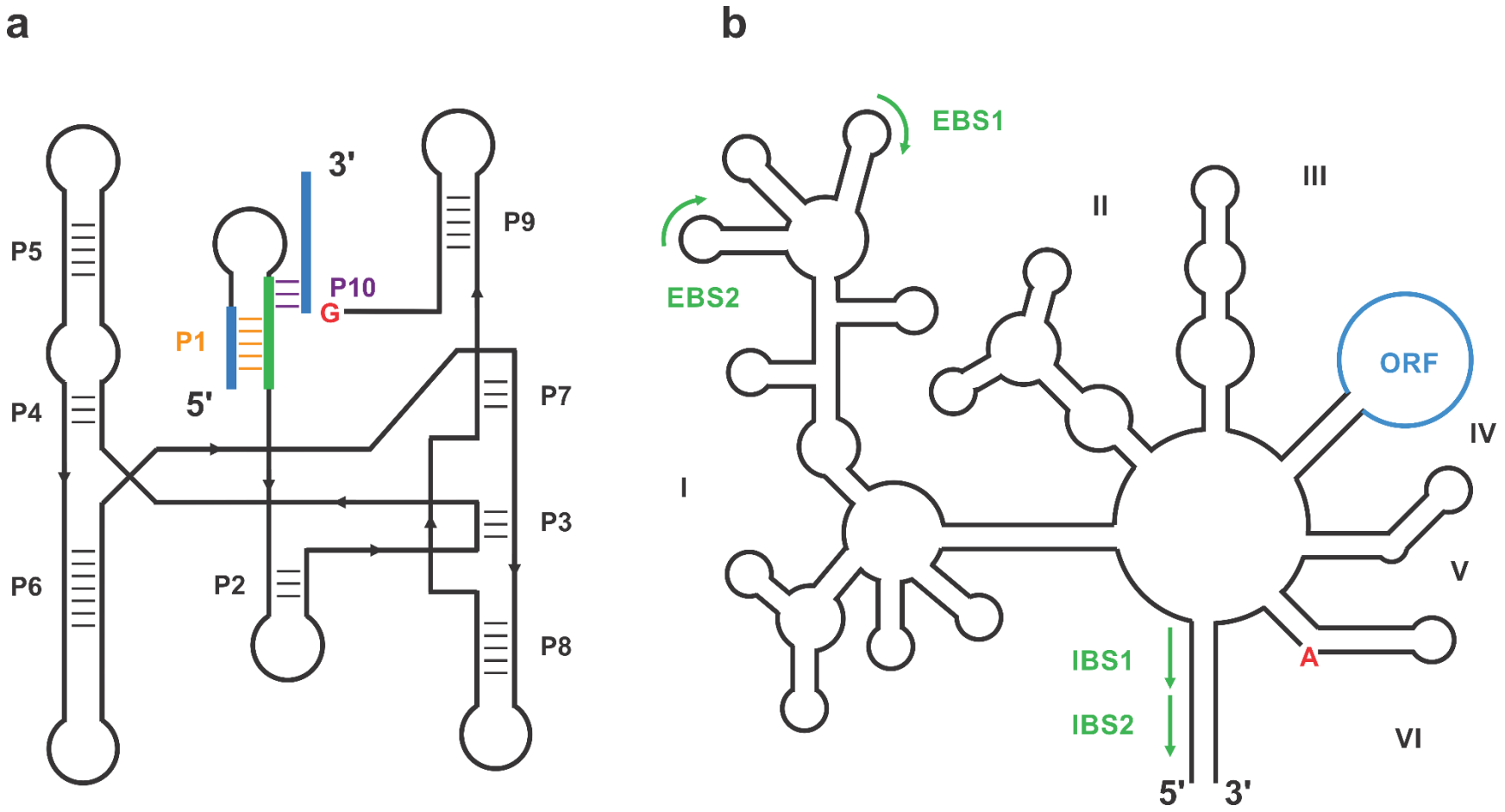

a
3'
P9
P5
P10
G
P1
5'
P4
P7
P3
P6
P2
P8
b
EBS1
II
III
EBS2
ORF
IV
I
V
IBS1
A
IBS2
VI
5'
3'

Figure 1.1. Schematic of conserved secondary structures adopted by group I and group II introns.

(a) Group I introns are characterized by ten major pairings (P1-P10) arranged into three helical stacks. **(b)** Group II introns are characterized by six major helical domains (DI-DVI) and are typically depicted in two dimensions as a structure resembling a “six-fingered hand”.

efficiency (Fedorova et al., 2003). The highly conserved, purine-rich linker region between domains II and III forms an important component of the catalytic core and seems to have multiple roles during splicing (Fedorova et al., 2003). Domain VI possess an adenosine bulge which initiates the first step of catalysis (reviewed by Michel and Ferat, 1995; reviewed by Toor et al., 2008; Marcia and Pyle, 2012).

Splicing of group II introns occurs in a similar manner as group I introns with some differences. One difference is the absence of an exogenous guanosine. Instead, group II introns contain an internal adenosine at the base of domain VI which bulges out of the helix. The adenosine bulge acts as the nucleophile which initiates the first of two reversible transesterification reactions. Another difference is, during splicing, the intron goes through a branched intermediate (the bulged adenosine forms a 2′-5′ phosphodiester bond with the 5′-end of the intron resulting in a looped intron attached to the 3′-exon) and is excised in the form of a lariat, rather than a linear intron typically seen in group I introns. Similar to group I introns, splicing of group II introns can be enhanced by IEPs, the ORF expressing a reverse transcriptase, are typically located in domain IV (reviewed by Lambowitz and Belfort, 2015).

1.2.2.3. Group III introns

Currently, the only known examples of group III introns are observed in the plastids of euglenoids (Doetsch et al., 1998). The secondary structure of group III introns resembles group II introns, although they lack domains II-IV (Michel and Ferat, 1995). The splicing mechanism is similar to group II introns, both of which proceed via a two-step transesterification reaction initiated by a bulged adenosine in domain VI (Copertino et al., 1994; Michel and Ferat, 1995; Doetsch et al., 1998). The requirement of IEPs to enhance splicing has yet to be elucidated.

Group III introns may contain an ORF analogous to those seen in group II introns, which encode an IEP that can assist in intron splicing, although evidence is mostly based on comparative sequence analysis (Mohr et al., 1993, Copertino et al., 1994).

Some group III introns have been reported to be arranged in twintron-like arrangements where the internal member must splice first in order for the external member to assume a splicing competent RNA fold (reviewed in Hafez and Hausner, 2015). In general, group III introns are viewed by some to be nothing more than a type of group II intron (reviewed by Robart and Zimmerly, 2005; reviewed by Zimmerly and Semper, 2015).

1.2.3. Mobile elements: Mobile introns and intron-encoded proteins

An interesting feature of self-splicing introns is their ability to insert into cognate intron-less alleles. This phenomenon was first observed through recombination studies of the large ribosomal gene (23S rRNA gene; *rnl*) in *Saccharomyces cerevisiae* (Michel et al., 1979). Subsequent studies (Jacquier and Dujon, 1985; Macreadie et al., 1985) revealed that the intron [physically mapped (Bos et al., 1978) and termed the ω allele at the time (Faye et al., 1979)] contained an IEP, which is required for insertion into intron-less alleles. Soon after, the IEP was shown to have endonuclease activity, with the ability to cut double-stranded deoxyribonucleic acid (DNA) molecules generating 4-nucleotide 3′ overhangs (Colleaux et al., 1986; Dujon et al., 1986). This ability to recognize and insert into its cognate intron-less allele was termed homing (Dujon et al., 1989).

Intron homing requires the expression of an intron-encoded protein referred to as a homing endonuclease (Gauthier et al., 1991). Homing endonuclease genes (HEGs) can be found

encoded within group I and group II introns and as part of inteins, or can be freestanding (reviewed by Stoddard, 2011).

The mechanism of intron homing involves expression of an intron-encoded homing endonuclease (HE), which makes a site-specific cleavage at a cognate intron-less allele. Cleavage generates a staggered cut which is repaired by the cell through the double-strand break repair (DSBR) pathway (Szostak et al., 1983). In the process, both the intron and the HEG are copied into the new site (gene conversion). Site-specific cleavage is attributed to the extensive recognition sequence required by these enzymes for DNA-binding. The recognition sequence can range anywhere from 14 to 40 bp (Li and Monnat, 2014). Interestingly, given the length of their recognition sequences, HEs can tolerate minor base pair substitution within the recognition site (Bryk et al., 1993; Lucas et al., 2001).

1.2.3.1. Homing endonucleases

HEs can be divided into at least six different families: LAGLIDADG, GIY-YIG, HNH, His-Cys, very short patch repair (Vsr), and PD-(D/E)xK HEs (Stoddard, 2011; Hafez and Hausner, 2012). The division is based on the presence of conserved sequence motifs. For example, the HNH HEs commonly seen in viruses possess a conserved asparagine (N) residue flanked at either end, with several amino acids in between, by a conserved histidine (H) residue (Mehta et al., 2004). The focus of this thesis will be on fungal HEs belonging to the LAGLIDADG and GIY-YIG families as these are most frequently encountered among fungal mtDNAs (reviewed by Stoddard, 2005).

1.2.3.1.1. LAGLIDADG homing endonucleases

The LAGLIDADG family of HEs (LHEs) can be found in archaea, bacteria, the mtDNAs of fungi, protozoa, Porifera (demosponges)**,** and the chloroplast genomes (cpDNAs) of algae and plants (reviewed by Chevalier et al., 2005; reviewed by Stoddard, 2005; Huchon et al., 2015; Schuster et al., 2017).

LHEs have two different structural forms, monomeric and dimeric. Dimeric LHEs (see Figure 1.2) are formed by dimerization of two separate LHEs, each containing a single LAGLIDADG motif. They have an αββαββααα topology (α = alpha helix, β = beta sheet; based on Heath et al., 1997). The LAGLIDADG motif is contained within C-terminal end of the first alpha helix. Two LAGLIDADG helices can interact to form a parallel two-helix bundle, a structure involved in protein dimerization. The two-helix bundle is positioned in the center of, and perpendicular to, 2 four-stranded antiparallel beta sheets (the four-stranded beta sheets are formed by 2 two-stranded beta sheets, connected by an alpha helix). Each of the four-stranded beta sheets form separate concave structures, called β-saddles (due to its resemblance to a horse saddle), which are stabilized by their three respective C-terminal helices. These three helices are positioned on the opposite side of the concave, stabilizing the beta sheets via hydrophobic interactions. Both beta sheets are involved in DNA recognition and binding. They form several important interactions within the two major grooves flanking the minor groove containing the cleavage site. These include hydrogen bonding, water-mediated contacts, and van der Waals interactions. The DNA recognition sequence of monomeric LHEs is palindromic or pseudo-palindromic due to the symmetrical nature of the homodimers (reviewed by Stoddard, 2011).

In contrast to the dimeric form, monomeric LHEs possess two LAGLIDADG motifs in tandem, separated by a linker region of approximately 90-120 aa. Their topology resembles a

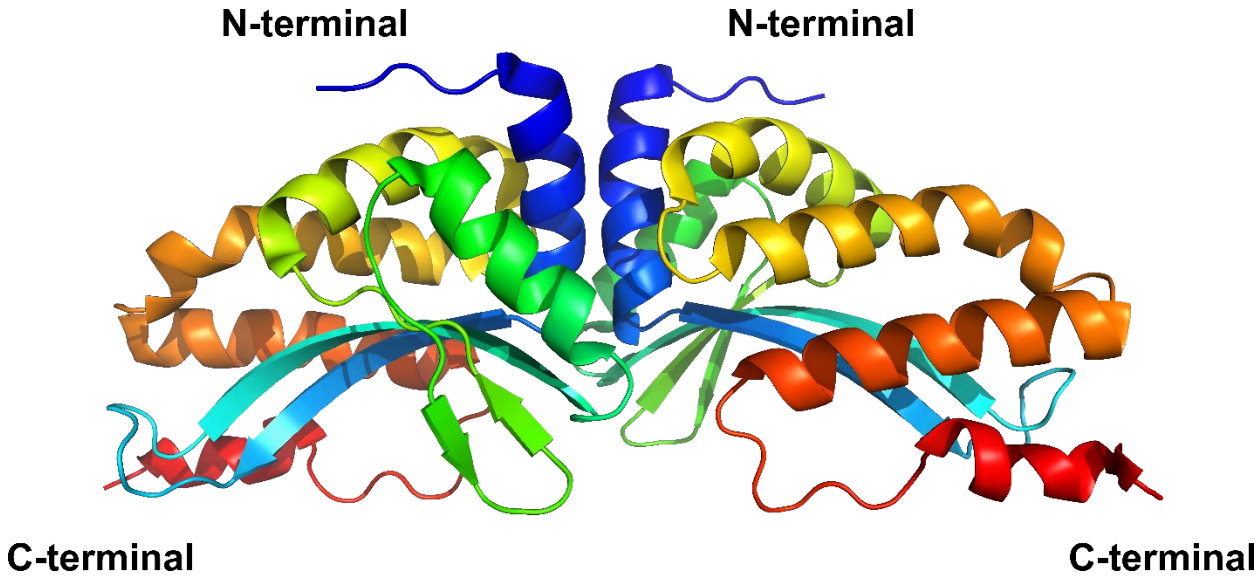

N-terminal
N-terminal
C-terminal
C-terminal

Figure 1.2. The structure of the LAGLIDADG homing endonuclease, I-*Cre*I.

I-*Cre*I is a homodimeric LAGLIDADG homing endonuclease. It is encoded within the intron of the chloroplast large ribosomal subunit gene of *Chlamydomonas reinhardtii*. The N- and C-terminals of each dimer are labeled. The image was generated from PDB ID: 1G9Z (Chevalier et al., 2001) using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

duplicated monomeric LHE, with two mixed α/β domains, typically containing two αββαββα motifs (reviewed by Stoddard, 2005).

Binding of a LHE results in a conformational change in the DNA, resulting in a bend at the minor groove. The bend distorts the DNA helix causing base twisting and unstacking. This exposes the bases of the minor groove to the catalytic domain and positions the scissile phosphates near conserved acidic residues of the two LAGLIDADG helices (Molina et al., 2012). These acidic residues are critical for catalytic activity, each coordinating a divalent cation (Schöttler et al., 2000). The divalent cation within each active site also coordinates a water molecule which will act as a nucleophile for strand cleavage. There is evidence suggesting that cleavage may occur sequentially or independently, favoring one strand over the other (Perrin et al., 1993; Ichiyanagi et al., 2000; Beylot and Spassky, 2001).

1.2.3.1.2. GIY-YIG homing endonucleases

The GIY-YIG family of HEs (GIYs) can be found in phage, archaea, bacteria, the mtDNAs of fungi and protozoa, and the cpDNAs and mtDNAs of algae (reviewed by Belfort and Bonocora, 2014; reviewed by Stoddard, 2005).

More research has been performed on fungal LHEs than GIYs. Thus, the following comparison of the two enzymes will contain information more relevant to phage and bacterial GIYs, which have been studied more extensively than their fungal counterparts.

In contrast to LHEs, GIYs are modular, possessing an N-terminal cleavage domain separate from the C-terminal DNA-binding domain. The two domains are joined by a flexible linker region (Van Roey et al., 2001). LHEs recognize DNA along the major groove whereas GIYs bind to the minor groove (Bryk et al., 1993). Although both LHEs and GIYs are site-

specific endonucleases, contacts along the DNA appears to be sequence-tolerant in both cases (Bryk et al., 1993; Lambert et al., 2016). GIYs recognize their targets via nuclease-associated modular DNA-binding domains (NUMODs), which are composed of alpha-helices in a helix-turn-turn conformation, whereas DNA recognition by LHEs are mediated by antiparallel beta-sheets (Sitbon and Pietrokovski, 2003). DNA cleavage by GIYs is mediated by tyrosine residue that acts as a general base whereas LHEs utilizes an aspartic acid residue, which acts as a general acid (Sokolowska et al., 2011). Both enzymes require divalent cations for activity.

1.2.4. Introns and fungal phenotypes

Mitochondrial introns may confer phenotypic variations among fungi. These include effects on virulence, antibiotic resistance, and growth and senescence. The following sections will give brief examples of potential phenotypic effects conferred by introns.

1.2.4.1. Introns and virulence

In 1985, Fulbright discovered a hypovirulent strain of *Cryphonectria parasitica* (synonym *Endothia parasitica*) different from hypovirulent strains documented earlier (Day et al., 1977; Jaynes and Elliston, 1982). The hypovirulent strain that was uncovered did not harbor dsRNA, which was later determined to be of viral origin (reviewed by Hillman and Suzuki, 2004). In 1993, Mahanti et al. found that hypovirulence of dsRNA-deficient strains may be connected to mitochondria and “mitochondrial dysfunction”. The relationship between hypovirulence and mitochondria was eventually explored in greater detail by Baidyaroy et al. (2011), which ultimately linked hypovirulence in *C. parasitica* with the presence of a group II

intron in the mitochondrial small ribosomal subunit gene. This intron was referred to as lnC9 or mS62, the latter nomenclature is based on intron position with respect to the 16S rRNA gene of *Escherichia coli* (Johansen and Haugen, 2001). The intron was noted to splice slowly and cells possess lower concentrations of mitochondrial ribosomes than virulent strains. Lower ribosome levels may the result of inefficient splicing of the group II intron from the ribosomal transcript. This slower splicing may be attributed to the lack of an ORF within the intron, which has been shown to enhance splicing (Carignani et al., 1983; Moran et al., 1994; Saldanha et al., 1999).

1.2.4.2. Introns and antibiotic resistance

Grasso et al. (2006) found that resistance to fungicides referred to as quinone outside inhibitors (QoI), which act on the cytochrome b protein, was strongly correlated with the presence or absence of introns, within the cytochrome b (*cob* or *cytb*) gene. Although resistance to antibiotics by mutation of the *cytb* gene has been observed previously (reviewed by Esposti et al., 1993), the association of resistance with occurrences of introns have not been documented (to the author's knowledge). The most common mutation displaying the strongest resistance effect is a glycine to alanine mutation at position 143 of the cytochrome b protein (G143A). Grasso et al. (2006) observed that many fungal plant pathogens which are resistant to QoIs appear to lack the group I intron that may be encoded immediately after codon 143. Grasso et al. (2006) suggested that, when present, the intron suppresses mutation of the codon from G**G**T (glycine) into G**C**T (alanine). This is due to base-pairing constraints imparted by the intron, whereby the last few nucleotides upstream of the intron is involved with forming the first pairing interaction (P1; see Figure 1) characteristic of group I introns. Thus, absence of a group I intron after codon 143

appears to “enable” the G143A mutation, which is, at least in part (Fernández-Ortuño et al., 2008), associated with QoI fungicide resistance.

As an interesting aside, a study conducted by Avila-Adame (2014) on the Ascomycota fungus, *Magnaporthe grisea*, suggests that the G143A phenotype can also be asexually transmitted, one alternative acquisition mechanism aside from random mutations in the genome. Asexual transmittance can occur via a process referred to as anastomosis where compatible [and potentially incompatible (Ishikawa et al., 2012)] vegetative hyphae fuse (reviewed by Leslie, 1993; reviewed by Saupe, 2000; Hickey et al., 2002; reviewed by Glass et al., 2004; Roca et al., 2005; Forgan et al., 2007; Craven et al., 2008). This fusion allows for exchange of cytoplasmic content. In this case, mitochondria can fuse and the mtDNAs can recombine, resulting in “sharing” of the mutated *cytb* gene.

Recent research in *Cryptococcus gattii* and *Cryptococcus neoformans* also suggests a role of group I introns in virulence and antibiotic resistance, in particular, intron absence is correlated with increased virulence and antibiotic resistance (Gomes et al., 2018). The research appears to be still in its early phase and thus is only briefly mentioned here.

1.2.4.3. Intron and growth and senescence

Introns appear to play a role in growth and senescence. In studies of the Ascomycota fungus, *Podospora anserina*, senescence was found to be linked with the presence of small circular DNAs, referred to as senDNAs (Jamet-Vierny et al., 1980). Multiple different senDNAs have been observed in the mitochondria and appear to be of mitochondrial origin (Belcour et al., 1981). One of the more commonly observed senDNAs, senDNA α, is identical in sequence to the first intron of the *cox1* gene (Osíewacz and Esser, 1984). The intervening sequence encodes for a

group II intron and reverse transcriptase/maturase (Sellem et al., 1990; Sainsard-Chanet et al., 1993). Earlier studies postulate senDNAs to be a contributing factor in senescence, whereby deletion (at the DNA level) of introns, potentially by the protein (reverse transcriptase) encoded by senDNA α, is associated with senescence (Sainsard-Chanet et al., 1993).

As an interesting side note, Schmidt et al. (1990) demonstrated (for first time with a group IIA intron) that the greater than 2.5 kb (including ORF) senDNA α is capable of self-splicing *in vitro*, at high salt concentrations. Thus, at least with regards to senDNA α, an active reverse transcriptase (or maturase) is not required for splicing of the intron (at the RNA level). In addition, based on its capacity to self-splice, the intron is presumed to be capable of folding into its active conformation even though it possesses a large sequence encoding for an ORF (ORF-less group II introns are approximately 0.8 kb) interrupting the intron sequence.

With regards to *Podospora* and senescence, later, Begel et al. (1999) demonstrated that *P. anserina* mutant strains missing the senDNA α sequence still undergoes senescence. This effectively eliminated the potential role of senDNA α as a causal agent of senescence. It appears that, rather than causing senescence, senDNA α contributes to it by introducing mtDNA instabilities, potentially by intron transposition into non-allelic sites (Sellem et al., 1993), inaccurate deletions of intron DNA (Belcour and Vierny, 1986), and/or non-homologous recombination (Sainsard-Chanet et al., 1998).

Furthermore, research also suggests a role of functional cytochrome c oxidase in senescence. In particular, the absence of a functional cytochrome c oxidase increases fungus longevity (Begel et al., 1999; Dufour et al., 2000). The lack of cytochrome c oxidase is not lethal as it appears to be compensated by alternative oxidase and using cyanide-resistant respiration (Dufour et al., 2000). The rationale for this phenomenon (decreased cytochrome c oxidase

increased longevity) was that reduced cytochrome c oxidase concentrations leads to reduced metabolism, that in turn leads to reduced production of harmful metabolic by-products (for example, reactive oxygen species), which can damage DNA (Perez-Campo et al., 1998). In other words, the reduction in metabolic rates would reduce the amount of sustained damage at any given time, reduce the rate of damage accumulation, and increase the amount of time for recovery (Dufour et al., 2000). The resulting overall effects would lead to an increase in longevity.

Interestingly, a decrease in longevity (amongst other phenotypic effects) of *Saccharomyces cerevisiae* was observed when all thirteen mitochondrial introns (one in *rnl*, seven in *cox1*, and five in the *cytb* gene) were removed (Rudan et al., 2018). Other effects of intron removal include increased *cox1* and *cytb* protein products and increased mitochondrial activity. Increasing *cox1* and *cytb* protein products and mitochondrial activity did not correspond with increasing concentrations of reactive oxygen species (ROS), as it appears that there is a corresponding increase in expression of the ROS scavenger, *sod2*. The authors hypothesize that some of the negative effects conferred by intron removal may be the result of increased mature *cox1* and *cytb* transcripts, leading to disequilibrium in protein products concentrations. This may affect the correct formation and functioning of complex III and complex IV of the electron transport chain (Rudan et al., 2018). Furthermore, the host gene may have adapted to the insertion of introns by increasing its own expression. Thus, intron removal would disrupt this adaptation.

Based on previous studies on *P. anserina* and complete intron removal in *S. cerevisiae* (Rudan et al., 2018), introns appear to play an important role in fungal longevity although the exact mechanism has yet been elucidated. Furthermore, additional studies are still required to

gain insight into the various functions of introns. One potential focus may be on the roles and effects of introns that appear to be highly conserved across divergent species of fungi (Guha et al., 2017b).

1.2.5. Applications of introns and intron-encoded proteins

One of the most prominent applications of introns in biotechnology was with engineered group II introns embedded into plasmid vectors referred to as "targetrons" (Guo et al., 2000; Karberg et al., 2001; Frazier et al., 2003; Zhong et al., 2003; Perutka et al., 2004; Yao and Lambowitz, 2007; reviewed by Liu et al., 2015). Targetrons are currently employed as a gene disruption system and is commercially sold by Sigma-Aldrich as the "TargeTron® Gene Knockout System" (https://www.sigmaaldrich.com/life-science/functional-genomics-and-rnai/targetron.html). The system is based on retrohoming of a group II intron into the target site, ultimately resulting in gene disruption. The specificity of the system involves base-pairing interactions between sequences within the group II intron (referred to as exon-binding sequences) with sequences (intron-binding sequences) directly upstream of the intron insertion site, as well as reverse transcriptase recognition. Possessing two different forms of target site recognition makes the targetron system highly specific. Furthermore, the discovery of heat-tolerant reverse transcriptases enables greater potentials in applications in biotechnology (Vellore et al., 2004; Ng et al., 2007; Mohr et al., 2013a).

The most well-known attempt in the application of homing endonucleases is its use as a gene drive (technique to propagate a gene(s) throughout a population), which can be used in controlling insect populations, for example those vectoring human disease-causing organisms (ex. malaria) (Chan et al., 2011; Windbichler et al., 2011). One strategy involves insertion of

LAGLIDADG or His-Cys homing endonuclease (ex. I-*Sce*I) gene (HEG) into the target gene (Galizi et al., 2014). HEG expression leads to cleavage of the target site, which is repaired mainly by homologous recombination (although non-homologous end joining, which is error prone and can introduce mutations, is also possible) using the uncut chromosome (or plasmid) as a template. In the process, the HEG is "copied" into the HEG-less allele (a process referred to as gene conversion), resulting in gene disruption.

The previous two examples demonstrate some of the ways in which mobile elements can be applied although not exhaustive [additional examples have been described in the following: Marcaida et al., 2010 (review); Takeuchi et al., 2011; Belfort and Bonocora, 2014 (review); Guha et al., 2017a (mini review)]. Thus, the study of homing endonuclease and HEGs may provide an important resource for their application in biotechnology.

1.3. Research objectives and rationale

There were two main objectives of the project: 1) study intron distribution in the *cytb* gene of fungi within the phylum Ascomycota, with an emphasis on filamentous fungi of the subphylum Pezizomycotina, and 2) study the gain and loss of two protein-coding (N-acetyltransferase and ribosomal protein S3) genes within the mtDNA.

By studying intron distribution, one can gain insight into the potential insertion sites where the mobile elements can be found and potentially speculate on their origins. It can also aid in gene/genome annotation by providing "coordinates" to where introns can be found. In the absence of an effective annotation/bioinformatics pipeline, taking a single gene approach allows for a more in-depth analysis of the components/elements of a gene, decreasing the likelihood of missing novel features (ex. N-acetyltransferase encoded within in cytochrome b intron of a few

fungi) or arrangements (ex. intron-within-intron/nested intron/twintron). Inadvertently, the single gene approach can limit genome-wide observations, which can also provide interesting observations (presence/absence of conserved and non-conserved genes, for example, ribosomal protein S3 and N-acetyltransferase).

Chapter 2: The intron landscape of the mtDNA *cytb* gene among the Ascomycota: introns and intron-encoded open reading frames

Chapter 2 has been published (Guha et al., 2017b) and was reformatted to be consistent with the format used in the thesis. Text, tables, and figures were modified (added or removed) where the author felt was appropriate.

Guha and Mullineux contributed some original data (sequencing of *cytb* gene for members of the Ophiostomatales). Dr. Hausner contributed towards the design of the experiment and helped with regards to editing the manuscript. Wai completed some the sequences needed for this work and recovered *cytb* gene sequences from NCBI. Wai also analysed and compiled all the data and wrote the first draft of the manuscript.

Guha TK*, Wai A*, Mullineux ST, Hausner G. 2017b. [*equal contribution]. The intron landscape of the mtDNA *cytb* gene among the Ascomycota: introns and intron-encoded open reading frames. Mitochondrial DNA A DNA Mapp Seq Anal. 20:1-10. doi: 10.1080/24701394.2017.1404042.

Taylor & Francis permits the use of published content for the purposes of a thesis.

2.1. Abstract

Fungal mitochondrial genes are frequently noted for the presence of introns. These introns are self-splicing and can be assigned to either group I or II introns and they can encode open reading frames (ORFs). This study examines the introns present within the cytochrome b gene of ascomycetes fungi. Cytochrome b gene sequences were sampled from GenBank and supplemented with our own data for species of *Leptographium* and *Ophiostoma*. Group I introns were encountered most frequently, many encoding either LAGLIDADG or GIY-YIG homing endonucleases. Numerous examples of different intron/ORF arrangements were observed including nested ORFs, multiple ORFs within a single intron and intron ORFs at various stages of erosion due to the accumulation of mutations. In addition, an example of a nested intron and one complex group II intron that could potentially allow for alternative splicing was noted. Documenting the distribution of introns within the same gene across a range of species allows for a better understanding of the evolution of introns and intronic ORFs. Intron landscapes also are a resource that can help in annotating genes and in bioprospecting for potentially active homing endonucleases, which are rare-cutting DNA endonucleases with applications in biotechnology.

2.2. Introduction

Among the fungi, mitochondrial genomes (mtDNAs) are highly variable in size and gene arrangement, ranging from 18 kb to more than 230 kb (Bullerwell et al., 2003; Losada et al., 2014; Aguileta et al., 2014; Wu and Hao, 2014; Freel et al., 2015). This variability is in part due to gene content, intergenic spacers, and group I and group II introns and intron encoded open reading frames (ORFs) (Procházka et al., 2010; Solieri, 2010). The autocatalytic function of

group I and II introns depends on folding of their RNAs into complex tertiary structures (Michel and Westhof, 1990; Chastain and Tinoco, 1992; Adams et al., 2004a; Zimmerly and Semper, 2015; Lambowitz and Belfort, 2015).

Group I and group II introns can encode homing endonucleases (HE) and reverse transcriptases (RT), respectively (Stoddard, 2005; Lambowitz and Zimmerly, 2011; Hausner, 2012; Wu et al., 2015). These intron-encoded proteins (IEPs) promote the mobility of their host introns and in some instances these proteins can act as maturases, i.e. they can assist in the splicing of the introns that encode them (Szczepanek and Lazowska, 1996; Bolduc et al., 2003; Belfort et al., 2002; Belfort, 2003; Lang et al., 2007).

Homing endonuclease genes (HEGs) generally are located within terminal loops of group I introns or in a few instances in the D3 or D4 segment of group II introns in a manner that does not interfere with the folding and splicing of the host intron (Guo et al., 1995; Toor and Zimmerly, 2002; Schäfer, 2003). Homing endonucleases (HEs) are DNA cleaving enzymes and they are assigned to various endonucleases protein families based on conserved amino acid motifs (Dujon, 1989; Stoddard, 2005; Hafez and Hausner, 2012). Among the fungi, the most frequently noted mtDNA encoded HEs are the LAGLIDADG and GIY-YIG families of HEs. HEs are very specific for their respective DNA recognition sequences that can range from 14 base pairs (bp) to 40 bp (Stoddard, 2005). Group I and II introns have applications as RNA reagents (Phylactou et al., 1998; Sullenger and Gilboa, 2002; Fiskaa and Birfisdottir, 2010) and can be coopted as regulatory elements (Guha and Hausner, 2014, 2016; McNeil et al., 2016). Homing endonucleases are rare-cutting DNA endonucleases that have been applied in genome editing and as agents for targeted mutagenesis (Stoddard, 2011, 2014; Takeuchi et al., 2011; Hafez and Hausner, 2012). HEs also have applications in synthetic biology (Liu et al., 2014) and

as components of cloning vectors and cloning strategies (Li et al., 2014). Group II intron RNA in combination with a reverse transcriptase can be engineered into so-called targetrons™ and these can be used in targeted mutagenesis (Enyeart et al., 2014). In addition, mtDNA introns have been associated with mtDNA defects in fungi (Sethuraman et al., 2008; Baidyaroy et al., 2011; Hausner, 2012), mtDNA rearrangements (Wu et al., 2015) and generating genetic diversity of sequences flanking introns (Repar and Warnecke, 2017), QoI fungicide resistance in fungal plant pathogens (Grasso et al., 2006), and in the attenuation of virulence in the plant pathogen *Cryphonectria parasitica* (Bertrand. 2000; Baidyaroy et al., 2011).

The cytochrome b (*cytb*) gene encodes for the cytochrome b subunit, a component of the cytochrome bc1 complex in the electron transport chain (Trumpower, 1990). An initial survey of the Ophiostomatales *cytb* gene revealed that it contains multiple insertions. The study was expanded to examine the *cytb* gene in other members of the Ascomycota. In this study, an intron landscape of the *cytb* gene for members of the Ascomycota was generated. Although, the long-term goal is to establish an intron landscape for the entire fungal mtDNA, computationally it is more feasible at this stage to focus on one gene and analyze in detail the types of introns and intron open reading frames (ORFs) that have invaded the gene. As *cytb* was previously noted (Yin et al., 2012) to be intron-rich, a detailed approach was taken to study the gene in regard to the diversity of introns and intron-encoded ORFs in order to propose scenarios for the evolution of intronic ORFs.

2.3. Materials and methods

2.3.1. Nucleic acid extraction, *cytb* gene amplification, and DNA sequencing

Strains of *Ophiostoma* and *Leptographium* were cultured on malt extract agar (Johnston and Booth, 1983). The strains were further cultured at room temperature in 50 mL PYG liquid media (peptone: 1 g/L, yeast extract: 1 g/L, glucose 3 g/L) for generating biomass suitable for nucleic acid extraction. Total nucleic acids were prepared according to methods described by Hausner et al. (1992).

Due to the length of the *cytb* gene (> 5 kb in some instances), it was amplified by the polymerase chain reaction (PCR) in three overlapping segments referred to as "front", "middle", and "back" regions. Primer sequences were designed for each region. Front primers: forward = 5′-TACTTACATATGGCAAGAGG-3′, reverse = 5′-ATAAAAGACATTTGACCATAAGG-3′. Middle primers: forward = 5′-ATGTCTTTATGAGGTGCTACAG-3′, reverse = 5′-ATAAAAGACATTTGACCATAAGG-3′. Back primers: forward = 5′-TAAGTGCTATACCTTGAATTGG-3′, reverse = 5′-TTAGGTATAGATCTTAATATTGC-3′. PCR reactions were set up according to the protocol provided by New England Biolabs for OneTaq Hot Start DNA Polymerase (catalog number: M0481). Typical thermal cycling conditions included: an initial denaturation at 94 °C for 30 s, followed by 30 cycles of denaturation at 94 °C for 30 s, primer annealing at 48 °C for 1 min, and primer extension at 68 °C for 3 min., and a final extension at 68 °C for 5 min. The conditions were optimized for different primer combinations.

PCR products were purified with the Wizard SV Gel and PCR Clean-Up system (Promega, Madison, USA) according to the manufacturer's protocol. Approximately 100 ng of PCR product was mixed with 1.3 pmol of each primer and was sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions at the DNA Technologies Unit NRC Saskatoon (110 Gymnasium

Place, Saskatoon, SK S7N 0W9, Canada). Initially, primers used for the PCR amplification were also used for sequencing. Thereafter, primers were designed as needed based on newly generated sequences in order to "primer walk" and sequence across the entire *cytb*-based amplicons. Sequencing results were analyzed using GeneDoc (Nicholas et al., 1997) and sequence contigs were generated using CAP3 (Huang and Madan, 1999).

2.3.2. Data mining for *cytb* sequences

Using the *cytb* of *Ophiostoma novo-ulmi* subspecies *novo-ulmi* (CM001753.1) as a query sequence, additional *cytb* sequences were extracted from the National Center for Biotechnology Information (NCBI) using the Basic Local Alignment Search Tool for nucleotides (blastn) (Altschul et al., 1990). Genome sequences were reanalyzed (annotated) with the program MFannot and RNAweasel (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl and http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl, respectively; Gautheret and Lambert, 2001; Lang et al., 2007) when required. The ORF Finder program (https://www.ncbi.nlm.nih.gov/orffinder/; Genetic code setting for molds #4; Sayers et al., 2010) was used to verify *cytb* exons and identify intron-encoded ORFs.

2.3.3. Sequence analysis

A total of 109 *cytb* sequences extracted from NCBI (GenBank; Benson et al., 2015) plus 20 from our own efforts were aligned using Kalign (Lassmann and Sonnhammer, 2005) using the program's default settings. The FASTA-formatted alignment was further analyzed using GeneDoc, ClustalX (Larkin et al., 2007), MAFFT

(http://mafft.cbrc.jp/alignment/server/index.html; alignment strategy setting: E-INS-I; Katoh and Standley, 2013), and the alignments were manually modified as necessary using GeneDoc. The aligned *cytb* nucleotide data set was also utilized to generate a phylogenetic tree using Maximum Likelihood criteria as implemented in W-IQ-TREE (Trifinopoulos et al., 2016) using default settings. Only exon sequences were maintained and the bootstrap option (1000 iterations) was implemented to estimate node support values.

The RNAweasel program (Eddy and Durbin, 1994; Gautheret and Lambert, 2001; Lang et al., 2007; Smith et al., 1994) was used to predict intron types/categories and intron core structural elements. Introns were named according to their insertion sites within the *cytb* sequence. Intron insertion site designations were based on comparative sequence analysis with the coding sequence of the *Saccharomyces cerevisiae* (KP263414.1) *cytb* gene. For example: the intron inserted after nucleotide position 506 would be abbreviated as *cytb*-506, where "*cytb*" denotes the cytochrome b gene and the number denotes the nucleotide position where the intron is located relative to the *S. cerevisiae cytb* coding sequence.

2.3.4. Intron folding

Secondary structural elements of the group I introns within the *cytb* genes, including the P3, P4, P6, P7, and P8 helices, were predicted using RNAweasel. Other base-pairing interactions, namely the P1, P2, P5, P9 and P10 helices, were predicted in part through Mfold (Zuker, 2003) and manually by comparing the intron sequences against reference structures and alignments generated by Michel and Westhof (1990) as well as those provided within the Group I intron Sequence and Structure Database (GISSD) (Zhou et al., 2008). The final intron folds were manually drawn following the conventions proposed by Cech et al. (1994) using

CorelDRAW Graphics Suite X6 (Corel Corporation, Ottawa). Similarly, group II intron sequences were analyzed with RNAweasel and Mfold (using constraints based on previously published models) and RNA folds were generated based on multiple alignment comparative analysis and reference folds provided by Michel and Ferat (1995), Toor et al. (2001) and the group II intron data base (http://webapps2.ucalgary.ca/~groupii/; Candales et al., 2012).

2.4. Results

2.4.1. The *cytb* gene architecture among the Ascomycota: the *cytb* intron landscape

Based on comparative sequence analysis and RNAweasel predictions, a total of 362 introns were observed in the *cytb* data set containing our data from 12 members of *Leptographium* and 8 strains representing members of the *Ophiostoma ulmi* and *novo-ulmi* species complex plus 109 sequences extracted from GenBank. It consisted of 350 group I introns, 10 group II introns, and two undefined introns (see Table 2.1, summarized in Table 2.2 and visualized in Figure 2.1), and were distributed in a total of 21 different intron insertion sites within the *cytb* gene (see Figure 2.1, simplified in Figure 2.2). The largest *cytb* genes were observed in *Bipolaris maydis* (KB733493.1)/*Cochliobolus heterostrophus* (KB445594.1), *Guignardia bidwellii* (JF785545.1), *Monilia yunnanensis* (HQ908793.1), *Monilinia fructicola* (GQ304941.1), *Monilinia fructigena* (HM149254.1), *Monilinia laxa* (GU952817.1), and *Pyrenophora tritici-repentis* (DS231662.1), each containing seven introns, although not every intron was shared amongst these species. In general, it was noted that some introns are wide spread across different organisms whereas others were more unique; for example: *cytb*-393 can be found in 94 out of the 129 fungi examined, *cytb*-490 was found in 68 fungal species, and *cytb*-506 was observed in 39, whereas *cytb*-159 had only one representative. In addition, related fungi

Table 2.1. Summary of introns inserted in the *cytb* gene sampled from 129 Pezizomycotina fungi from the National Center for Biotechnology Information database.

		Intron Insertion Site																					
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
Saccharomycetales (39)																						
***Barnettozyma californica* (KC993183.1)**	3		IB -								ID fL								IB fL			
***Candida albicans* (JQ864233.1)**	2							ID fG			ID -											
***Candida bohiensis* (KF214631.1)**	1																		? -			
***Candida chauliodes* (KF017574.1)**	1																		IB (ei) fL			
***Candida corydali* (KC993198.1)**	4							ID fG			ID fL		IA -						IB dL			
***Candida jiufengensis* (GU136397.1)**	1							ID fG														
***Candida labiduridarum* (KC993196.1)**	3							ID -					IA -						IB fL			
***Candida maltosa* (EU267175.1)**	2							ID G					I (d, A) -									
***Candida neerlandica* (EU334437.1)**	3							ID -					IA -						IB fL			
***Candida norvegica* (KF017573.1)**	1							ID fG														
***Candida orthopsilosis* (AY962590.1)**	1										ID fL											
***Candida oxycetoniae* (KC993187.1)**	2			II fRT														II fRT				
***Candida parapsilosis* (JQ062883.1)**	2							ID fG			ID fL											

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Candida pseudojiufengensis* (KC993179.1)**	4							ID fG			ID fL		I (d, A) -						IB -			
***Candida sake* (KC993194.1)**	1							ID fG														
***Candida santjacobensis* (KC993178.1)**	1							IB hp														
***Candida saraburiensis* (KC993174.1)**	1												IA -									
***Candida theae* (KC993195.1)**	1										ID fL											
***Candida viswanathii* (EF536359.1)**	1																	II fRT				
***Cyberlindnera jadinii* (KC993189.1)**	4							ID fG			ID fL		I (d, A) fL						IB fL			
***Cyberlindnera suaveolens* (KC993193.1)**	4							ID fG			ID fL								IB fL		IA fhp	
***Debaryomyces hansenii* (DQ508940.1)**	3							ID dG			ID L		IA fL									
***Galactomyces candidum* (HG530139.1)**	1							ID fG														
***Lachancea* sp. (HE983613.1)**	1							ID dG														
***Magnusiomyces capitatus* (KJ459952.1)**	2												IA dL						IB dL			
***Magnusiomyces ingens* (KJ459950.1)**	1												IA fL									
***Magnusiomyces tetrasperma* (KJ459951.1)**	2							ID fG			ID fL											

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Meyerozyma guilliermondii* (KC993176.1)**	2										ID fL		IA -									
***Millerozyma farinosa* (FN356025.1)**	5							ID G			ID fL		I (d, A) -	IB fL					IB fL			
***Ogataea philodendri* (KC993191.1)**	1												I (d, A) -									
***Ogataea thermophila* (KC993173.1)**	2							ID fG					IA fL									
***Pichia pastoris* (FR839632.1)**	3							ID fG			ID fL								IB (ei) fL			
***Saccharomyces cerevisiae* (CP006539.1)**	0																					
***Saccharomyces cerevisiae* (KP263414.1)**	5									II -	ID fL			IB (3′, p) fL					IB fL		IA -	
***Torulaspora franciscae* (KM595069.1)**	2							ID dG											IB dL			
***Torulaspora globosa* (KM595074.1)**	3							ID fG						IB fL					IB fL			
***Torulaspora pretoriensis* (KM595067.1)**	1							ID fG														
***Wickerhamomyces mucosus* (KC993197.1)**	4										ID fL		IA fL						IB fL		IA fhp	
***Yarrowia lipolytica* (CM004385.1)**	4										ID fL			IB fL					IB fL		IA fhp	
Pezizales (1)																						

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Pyronema omphalodes* (KU707476.1)**	3							ID hp	IC2 dL			IB -										
Botryosphaeriales (2)																						
***Guignardia bidwellii* (JF785545.1)**	7						ID fL	ID fG			ID dL, L		IA dL, dG	IB fL				II fRT		(cut off)		
***Guignardia citricarpa* (KC788404.1)**	2							ID dG, L	IC2 -		ID dL, dL											
Capnodiales (1)																						
***Mycosphaerella fijiensis* (AF343069.1)**	1													IB dL								
Pleosporales (6)																						
***Bipolaris maydis* (KB733493.1)**	7		IB L				ID dL, fG	ID fG			ID 2x N	IB L	IA fL									IB L
***Cochliobolus heterostrophus* (KB445594.1)**	7		IB L				ID dL, fG	ID fG			ID 2x N	IB L	IA fL									IB L
***Didymella pinodes* (KT946597.1)**	3		IB -					ID fG					IA fL									
***Phaeosphaeria nodorum* (EU053989)**	1													IB fL								
***Pyrenophora tritici-repentis* (DS231662.1)**	7		IB dL, L				ID fL	ID fG, L				IB 2x dL	IA fL	IB dL								IB -
***Venturia inaequalis* (AF004559.1)**	6	II fRT			IC1 dL				IC2 -					IB (3′ p) fL						IC1 hp		IB dL
Eurotiales (6)																						

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Aspergillus clavatus* (JQ354999.1)**	1							ID fG														
***Aspergillus flavus* (KP725058.1)**	1							ID fG														
***Aspergillus nidulans* (JQ435097.1)**	1													IB fL								
***Talaromyces marneffei* (AY347307.1)**	1							ID fG														
***Talaromyces stipitatus* (JQ354994.1)**	2							ID fG					IA fL									
***Xeromyces bisporus* (HG983520.1)**	4							ID dG			ID -		IA dL, L									IB dL, G
Onygenales (3)																						
***Ajellomyces capsulatus* (GG663449.1)**	1							ID fG														
***Epidermophyton floccosum* (AY916130.1)**	1							ID fG														
***Paracoccidioides brasiliensis* (AY955840.1)**	2		I (d, B1) -															II fRT				
Lecanorales (1)																						
***Lecanora strobilina* (KU308740.1)**	1												IA fL									
Peltigerales (2)																						
***Peltigera malacea* (JN088164.1)**	3							ID dG					IA dL									IB dL

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Peltigera membranacea* (JN088165.1)**	3							ID fG					IA dL									IB dL
Helotiales (8)																						
***Botryotinia fuckeliana* (FJ217742.1)**	3		IB L					ID fG					IA fL									
***Botryotinia fuckeliana* (FJ217744.1)**	4		IB L					ID fG			ID fL		IA fL									
***Monilia mumecola* (JN204425.1)**	6		IB dL, dG					ID fG			ID dL, L		I (d, A) fL	IB fL						I (d, A) hp		
***Monilia yunnanensis* (HQ908793.1)**	7		IB L					ID fG			ID dL, L		I (d, A) fL	IB dL						I (d, A) hp		IB dL, G
***Monilinia fructicola* (GQ304941.1)**	7		IB L					ID fG			ID fL		I (d, A) fL	IB fL						I (d, A) hp	I (d, A) fhp	
***Monilinia fructigena* (HM149254.1)**	7		IB L		IB dL	II -		ID fG					IA dL, dG	IB dL						I (d, A) hp		
***Monilinia laxa* (GU952817.1)**	7		IB dL, dG					ID fG			ID dL, dL		IA dL	IB dL						I (d, A) hp		IB dL, dG
***Sclerotinia borealis* (KJ434027.1)**	6				IC1 dL, dL			ID fG				IB dL, dG					? hp			IC1 dL	IA fhp	
Hypocreales (26)																						
***Beauveria bassiana* (EU371503.2)**	1							ID fG														
***Beauveria caledonica* (KT201150.1)**	2							ID fG						IB fL								
***Beauveria malawiensis* (KT201147.1)**	1							ID fG														

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Cordyceps bassiana* (EU100742.1)**	1							ID G														
***Cordyceps brongniartii* (EU100743.1)**	1							ID G														
***Cordyceps militaris* (KF432176.1)**	1							ID fG														
***Fusarium avenaceum* (JPYM01000083.1)**	1							ID fG														
***Fusarium circinatum* (JX910419.1)**	4		IB dL					ID dG					IA fL									IB dL
***Fusarium culmorum* (KP827647.1)**	6		IB L		IC1 fL			ID fG, hp					IA fL	IB fL						IC1 hp		
***Fusarium fujikuroi* (JX910420.1)**	1												IA dL									
***Fusarium gerlachii* (KM486533.1)**	5		IB fL		IC1 fL			ID fG, hp					IA fL	IB fL						IC1 hp		
***Fusarium graminearum* (DQ364632.1)**	5				IC1 fL			ID fG, hp					IA dL, L	IB fL						IC1 hp		
***Fusarium graminearum* (KP966561.1)**	5				IC1 fL			ID fG, hp					IA fL	IB fL						IC1 hp		
***Fusarium graminearum* (KR011238.1)**	6		IB L		IC1 fL			ID fG, hp					IA dL, L	IB fL						IC1 hp		
***Fusarium oxysporum* (EU035603.1)**	1							ID fG														
***Fusarium oxysporum* (EU035604.1)**	2							ID fG					IA fL									
***Fusarium oxysporum* (KR952337.1)**	1							ID fG														

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Fusarium temperatum* (KP742837.1)**	1							ID G, hp														
***Gibberella zeae* (DQ364632.1)**	5				IC1 fL			ID fG, hp					IA dL, L	IB fL						IC1 hp		
***Hirsutella minnesotensis* (KR139916.1)**	2							ID fG					IA fL									
***Hirsutella rhossiliensis* (KU203675.1)**	3							ID fG					IA fL									IB L
***Hypocrea jecorina* (AF447590.1)**	2							ID fG						IB fL								
***Hypomyces aurantius* (KU666552.1)**	3		IB (3′p) L					ID fG						IB (3′p) fL								
***Nectria cinnabarina* (KT731105.1)**	4		IB L, dL					ID fG					IA fL									IB L
***Neotyphodium lolii* (KF906135.1)**	4							ID fG					I (d, A) fL	IB -								IB L
***Ustilaginoidea virens* (JN204426.1)**	6		I (d, B1) dL					ID fG					IA fL	IB hp	IB L							IB L
Microascales (1)																						
***Ceratocystis cacaofunesta* (JX185564.1)**	2		IB L										IA fL									
Diaporthales (4)																						
***Chrysoporthe austroafricana* (KT380883.1)**	5		I (d, B1) fL			II L		ID fG					IA fL		IB L							
***Chrysoporthe cubensis* (KT380885.1)**	3		I (d, B1) fL					ID fG					IA fL									

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Chrysoporthe deuterocubensis* (KT380884.1)**	4		I (d, B1) fL					ID fG					IA fL									IB G, dL, G
***Cryphonectria parasitica* (KT428651.1)**	3		I (d, B1) fL										IA dL, dG									IB dG, L, G
Ophiostomatales (21)																						
***Leptographium lundbergii* (KY082967.1)**	0																					
***Leptographium lundbergii* (KY082962.1)**	2												IA fL	IB fL								
***Leptographium procerum* (KY082964.1)**	1							ID fG														
***Leptographium procerum* (KY082965.1)**	1							ID fG														
***Leptographium procerum* (KY082966.1)**	1							ID fG														
***Leptographium* sp. (KY082968.1)**	1													IB fL								
***Leptographium truncatum* (KY082969.1)**	1													IB fL								
***Leptographium truncatum* (KY082970.1)**	1													IB fL								
***Leptographium truncatum* (KY082971.1)**	1													IB fL								
***Leptographium truncatum* (KY082963.1)**	2												IA fL	IB fL								

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Leptographium wingfieldii* (KY082972.1)**	1													IB fL								
***Leptographium wingfieldii* (KY082973.1)**	1													IB fL								
***Ophiostoma novo-ulmi subsp. americana* (KY084297.1)**	4		IB L					ID fG				IB L	IA fL									
***Ophiostoma novo-ulmi* subsp. *americana* (KY084295.1)**	3							ID fG				IB L	IA fL									
***Ophiostoma novo-ulmi* subsp. *americana* (KY084296.1)**	4							ID fG				IB L	IA fL	IB fL								
***Ophiostoma novo-ulmi* subsp. *novo-ulmi* (CM001753.1)**	3							ID fG				IB dL	IA fL									
***Ophiostoma novo-ulmi* subsp. *novo-ulmi* (KY084294.1)**	3							ID fG				IB L	IA fL									
***Ophiostoma ulmi* (KY084298.1)**	5		IB L					ID fG				IB L	IA fL	IB fL								
***Ophiostoma ulmi* (KY084299.1)**	5		IB L					ID fG				IB L	IA fL	IB fL								
***Ophiostoma ulmi* (KY084300.1)**	5		IB L					ID fG				IB L	IA fL	IB fL								
***Sporothrix pallida* CM003773.1**	2							ID fG					IA fL									
Glomerellales (1)																						
***Colletotrichum lindemuthianum* (KF953885.1)**	1												IA fL									
Sordariales (6)																						

		Intron Insertion Site																				
		159	**201**	**212**	**278**	**289**	**378**	**393**	**405**	**415**	**429**	**437**	**490**	**506**	**562**	**616**	**677**	**684**	**756**	**776**	**806**	**820**
Phase		**0**	**0**	**2**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**2**	**1**	**1**	**2**	**0**	**0**	**2**	**2**	**1**
Total number of introns																						
***Chaetomium thermophilum* (JN007486.1)**	5							ID dG			ID fL	IB L	IA fL	IB, II dL								
***Neurospora crassa* (M37324.1)**	2							ID dG					IA fL									
***Neurospora crassa* (KC683708.1)**	2							ID fG					IA fL									
***Podospora anserina* (X51480.1)**	2		I (d, B1) L, G					ID fG														
***Podospora anserina* (X55026.1)**	3		I (d, B1) L, G					ID fG														IB L
***Sordaria macrospora* (NW_003542061.1)**	4		I (d, B1) L					ID fG					IA dL									IB dL
Xylariales (1)																						
***Annulohypoxylon stygium* (KF545917.1)**	6		IB (5′, p) 2x L					ID fG			ID N	IB L	IA L			ID fL						

Note: d = degenerated open reading frame (ORF); f = fused ORF; G = GIY-YIG homing endonuclease; hp = hypothetical protein; L = LAGLIDADG homing endonuclease; N = N-acetyltransferase; RT = reverse transcriptase; ei = extra insertion (only applies to introns).

Table 2.2. Summary of the number of introns and the intron/open reading frame configurations observed in the *cytb* gene, sampled from 129 Pezizomycotina fungi from the National Center for Biotechnology Information database.

IS	Total In	Total i	IA	IB	IC1	IC2	ID	II	UD	Total ORFs	fL	L	dL	fG	G	dG	fRT	RT	dRT	N	fhp	hp	ORF-minus
159	1	1						1		1							1						
201	33	33		33						37	5	22	6		2	2							3
212	1	1						1		1							1						
278	9	9		1	8					10	6		4										
289	2	2						2		1		1											1
378	4	4					4			6	2		2	2									
393	94	94		1			93			101		2		76	5	9						9	2
405	3	3				3				1			1										2
415	1	1						1		0													1
429	29	29					29			34	18	4	7							5			2
437	15	15		15						16		11	4			1							1
490	68	68	68							66	45	5	13			3							9
506	39	40		39				1		38	31		6									1	1
562	2	2		2						2		2											
616	1	1					1			1	1												
677	1	1							1	1												1	
684	4	4						4		4							4						
756	17	17		16					1	15	12		3										2
776	13	13	5		8					13			1									12	
806	6	6	6							5											5		1
820	18	18		18						24		8	9		5	2							1
Total	361	362	79	125	16	3	127	10	2	377	120	55	56	78	12	17	6	0	0	5	5	23	26

Note: IS = insertion site [numbering is with reference to homologous site in the *Saccharomyces cerevisiae* (KP263414.1) coding sequence], In = number of organisms with the insertion; IS = insertion site; i = number of intron (group I or group II) observed in the given insertion site; UD = undefined intron type; ORF = open reading frame; d = degenerated ORF; f = fused ORF; G = GIY-YIG homing endonuclease; hp = hypothetical protein; L = LAGLIDADG homing endonuclease; N = N-acetyltransferase; RT = reverse transcriptase.

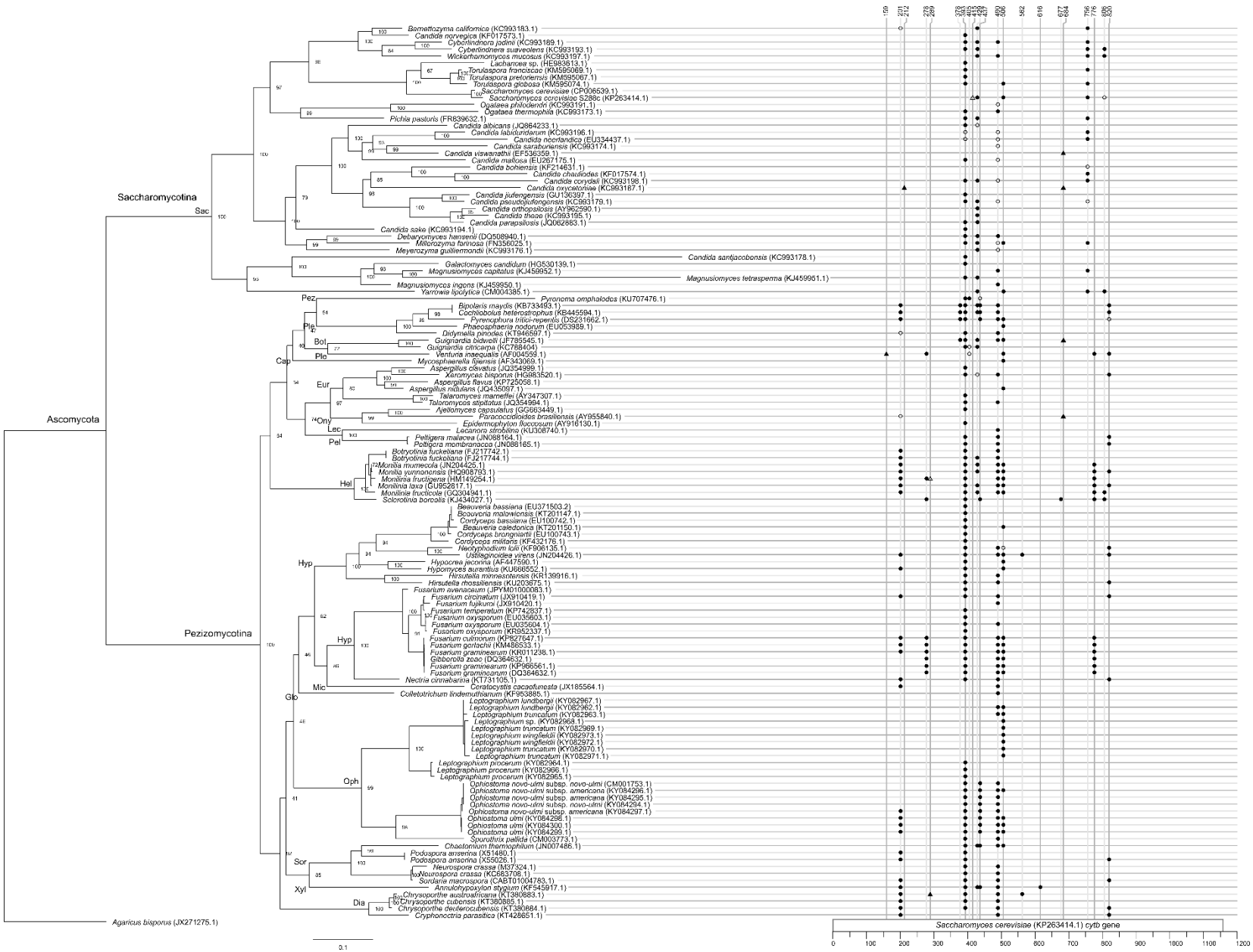

Ascomycota
Saccharomycotina
Sac
Pezizomycotina
Pez
Ple
Bot
Pto
Cap
Eur
Ony
Lec
Pel
Hel
Hyp
Hyp
Mic
Glo
Oph
Sor
Xyl
Dia
Agaricus bisporus (JX271275.1)
Saccharomyces cerevisiae (KP263414.1) cytb gene
0.1

Figure 2.1. Phylogenetic tree of *cytb* genes from fungi of the phylum Ascomycota.

A total of 129 cytochrome b (*cytb*) genes were sampled from GenBank (Benson et al., 2013). The phylogenetic tree was generated using W-IQ-TREE (Trifinopoulos et al., 2016) using default settings. Bootstrap values, based on 1 000 iterations, are indicated on the nodes. The scale on the lower right of the figure indicates the number of base pairs along the *Saccharomyces cerevisiae* (KP263414.1) *cytb* gene (1 158 bp). Intron insertion site numbered above vertical lines. Taxonomic rank (Order) are abbreviated and place by or at the node of the tree. Bot = Botryosphaeriales; Cap = Capnodiales; Dia = Diaporthales; Eur = Eurotiales; Glo = Glomerellales; Hel = Helotiales; Hyp = Hypocreales; Lec = Lecanorales; Mic = Microascales; Ony = Onygenales; Oph = Ophiostomatales; Pel = Peltigerales; Pez = Pezizales; Ple = Pleosporales; Sac = Saccharomycetales; Sor = Sordariales; Xyl = Xylariales. Circles = group I intron; Triangle = group II intron; Filled circle/triangle = ORF-plus; Open circle-triangle = ORF-minus.

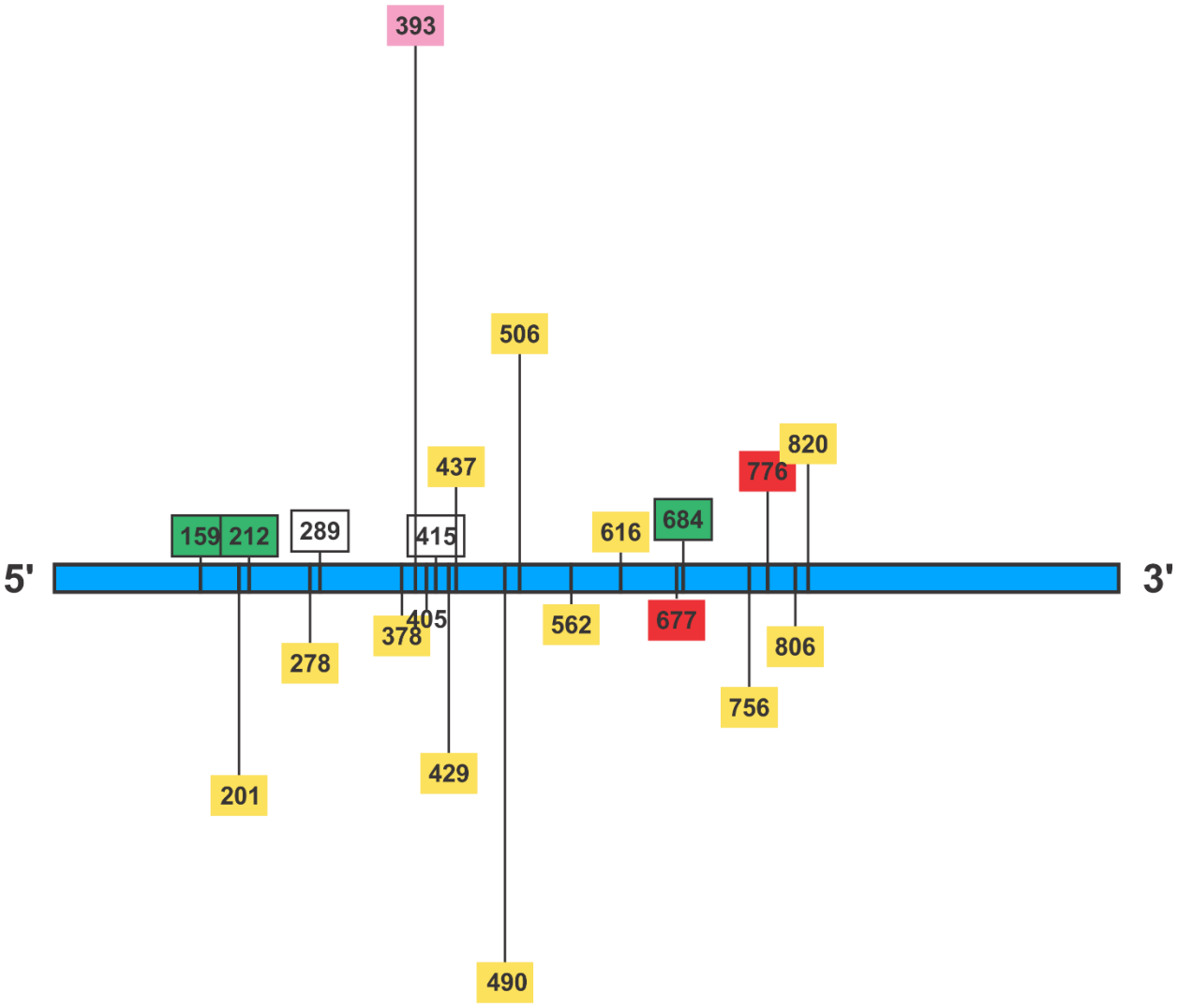

393
506
437
820
776
159
212
289
415
616
684
5'
3'
562
677
378
405
278
806
756
429
201
490

Figure 2.2. Intron landscape of the cytochrome b gene.

A total of 21 different intron insertion sites were observed among 129 fungal species of the phylum Ascomycota, sampled from the National Center for Biotechnology Information database. Intron insertion sites are numbered with respect to the homologous nucleotide (nt) position that is upstream to the intron insertion site with respect to the nt numbering in *Saccharomyces cerevisiae* (KP263414.1) coding sequence. Numbers within boxes = intron insertion site, vertical lines = relative number of representatives with intron, yellow box = introns typically encode for at least one LAGLIDADG homing endonuclease, pink box = introns typically encode for at least one GIY-YIG homing endonuclease, green box = introns typically encode for a reverse transcriptase, red box = introns typically encode for at least one hypothetical protein of unknown function, non-coloured box = unidentified intron, box without border = group I introns typically found at insertion site, boxes with border = group II introns typically found at insertion site.

can have different intron composition. These include species of *Leptographium*, which can have zero to two introns, some of which are shared among different species while others are unique to a particular strain. Variability with regards to intron distribution was observed with fungal strains of the same species, such as in *Botryotinia fuckeliana*, *Fusarium oxysporum*, and *Podospora anserina* (see Table 2.1, visualized in Figure 2.1).

Typically, it was recorded that at a specific *cytb* intron insertion site one would encounter the same type or subtype of intron encoding the same type of ORF (see Table 2.1, visualized in Figure 2.1). However, two exceptions were noted: *cytb*-393 can be a group ID or IB [in the case of *Candida santjacobensis* (KC993178.1)], and *cytb*-776 can either have a group IA (derived) or IC1 intron.

2.4.2. Group II introns and novel intron arrangements

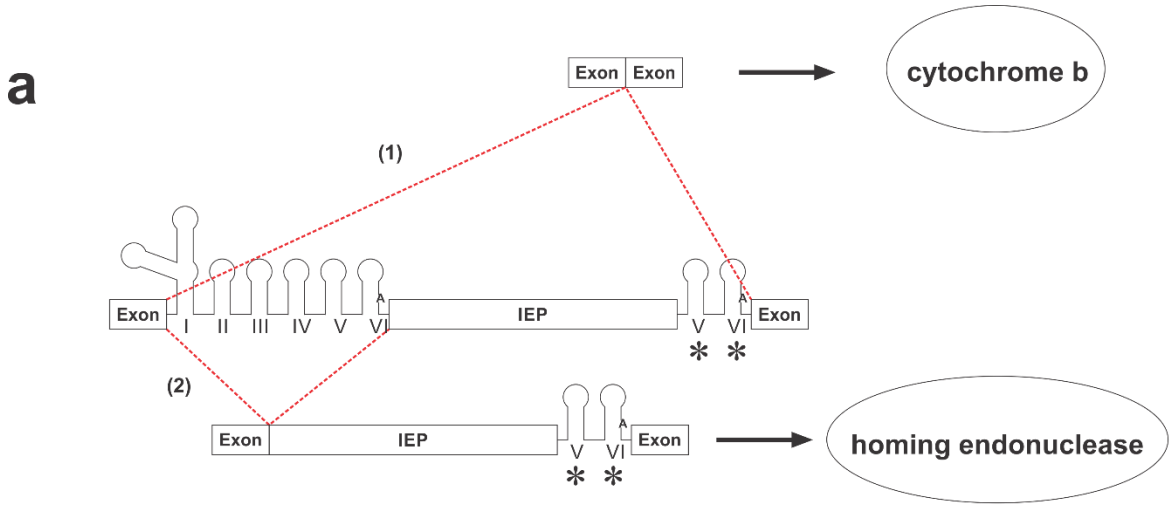
A total of five group II intron insertion sites (excluding *cytb*-506 nested group II intron) were found during this study, namely *cytb*-159, *cytb*-212, *cytb*-289, *cytb*-415, and *cytb*-684. The group II introns correspond to *cytb*-159 of *Venturia inaequalis* (AF004559.1), *cytb*-212 of *Candida oxycetoniae* (KC993187.1), *cytb*-289 of *Chrysoporthe austroafricana* (KT380883.1) and *Monilinia fructigena*, *cytb*-415 of *S. cerevisiae*, *cytb*-684 of *Candida oxycetoniae*, *Candida viswanathii* (EF536359.1), *Guignardia bidwellii* (JF785545.1), and *Paracoccidioides brasiliensis* (AY955840.1) (see Table 2.1). Further analysis through ORF Finder failed to identify any ORFs in the group II introns of *M. fructigena* and *S. cerevisiae*. The group II intron of *V. inaequalis* contains an ORF that has been identified to encode for a reverse transcriptase-like protein (Zheng and Köller, 1997).

A unique intron configuration was uncovered in the *cytb*-289 intron of *Chrysoporthe austroafricana* (KT380883.1). *Monilinia fructigena* has a group IIB intron (*cytb*-289) encoding a degenerated LAGLIDADG type ORF. However, an intron located at a homologous site in *Chrysoporthe austroafricana* (Ch.aus.) is more complex (see Figures 2.3a, b). Embedded within the Ch.aus *cytb*-289 group IIB intron is a partial IIB intron consisting of domains IV (segment), V and VI. This nested arrangement has the potential to allow for alternative splicing. Utilizing the internal V and VI domains could upon splicing generate a transcript whereby the ORF contained within the terminal part of this complex intron is be fused to the upstream exon and thus allowing for the efficient translation of the intron-encoded LAGLIDADG type ORF (see Figure 2.3a). Alternatively, the intron could also arrange in a way that connects the terminal segments (domains V and VI) of the intron with the anterior scaffold domain (DI) and the entire intron would be removed, and a functional mRNA could be generated for the host gene.

Another novel intron arrangement was noted for the *cytb*-506 of *Chaetomium thermophilum* (JN007486.1), is a nested intron composed of a group IIA1 intron inserted within the LAGLIDADG type ORF of the native group IB intron (see Figures 2.4a, b, c, d). The group II intron's exon binding sequence (EBS) appears to interact with IBS elements that are located upstream of the intron insertion site within the LAGLIDADG type ORF.

2.4.3. Group I introns and intron-encoded ORFs

Group I introns represent the most common type of intron found among the *cytb* genes, comprising 350 out of the 362 (~ 97%) introns examined. Group I introns were observed in sixteen different sites, namely at *cytb*-201, 278, 378, 393, 405, 415, 429, 437, 490, 506, 562, 616, 756, 776, 806, and 820 (see Table 2.1, summarized in Table 2.2). ORFs were frequently found


a
Exon
Exon
cytochrome b
(1)
Exon
I
II
III
IV
V
VI
IEP
V
VI
Exon
(2)
Exon
IEP
V
VI
Exon
homing endonuclease

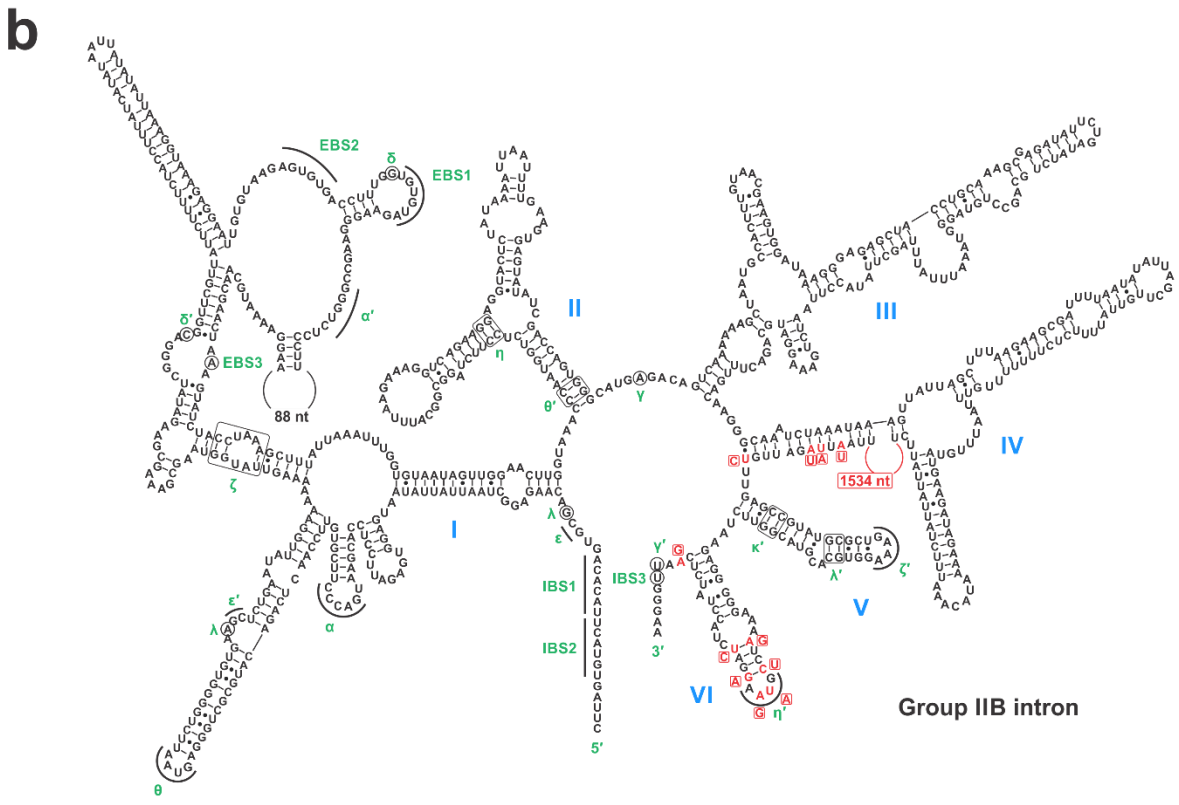

b
EBS2
δ
EBS1
α′
δ′
EBS3
88 nt
ζ
α
ε′
λ
θ
II
η
θ′
γ
I
III
IV
1534 nt
λ
ε
IBS1
IBS2
IBS3
5′
3′
γ′
κ′
λ′
ζ′
V
VI
η′
Group IIB intron

Figure 2.3. The *cytb*-289 intron of *Chrysoporthe austroafricana* (KT380883.1).

(a) Schematic overview of potential splicing pathways employed by the group IIB intron. The intron sequence contains a partially duplicated 3′ stem of domain IV, and duplicated domain V and VI flanking an intron-encoded protein (IEP), which encodes for a double-motif LAGLIDADG homing endonuclease. The duplicated sequences may allow for the intron to undergo alternative splicing. (1) Splicing using the 3′ duplicated segments results in the formation of a full-length cytochrome b mRNA. (2) Splicing using the upstream duplicated segments results in fusion of the IEP with the 5′ exon and potentially allowing for IEP expression. Differences are highlighted between the two duplicated sequences by red boxes and by denoting the nucleotide differences with *. **(b)** Proposed secondary structures of the *cytb*-289 group IIB intron. Note: 88 out of the 1534 nt in domain IV is part of the duplicated (upstream) group II intron sequence. EBS = exon binding site, IBS = intron binding site. Group II intron tertiary interactions are shown according to Michel and Ferat (1999).

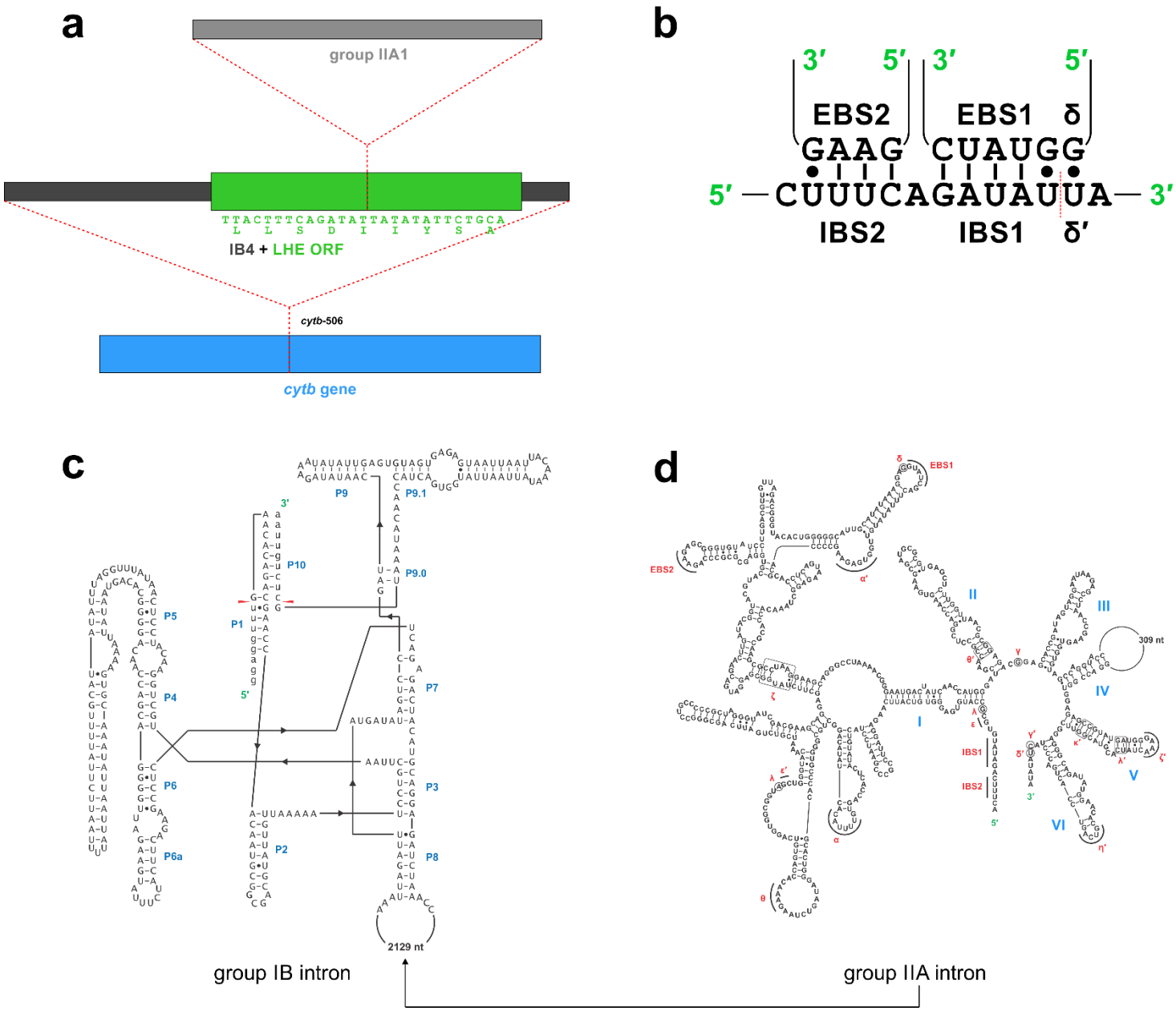

a
group IIA1
TTACTTTCAGATATTATATATTCTGCA
L L S D I I Y S A
IB4 + LHE ORF
cytb-506
cytb gene
b
3′ 5′ 3′ 5′
EBS2 EBS1 δ
GAAG CUAUGG
5′ — CUUUCAGAUAUUA — 3′
IBS2 IBS1 δ′
c
P9
P9.1
P10
P9.0
P1
P5
P4
P7
P6
P3
P6a
P2
P8
2129 nt
group IB intron
d
EBS1
EBS2
IBS1
IBS2
309 nt
I
II
III
IV
V
VI
group IIA intron

Figure 2.4. The *cytb*-506 intron of *Chaetomium thermophilum* (JN007486.1).

(a) Configuration of the *cytb*-506 intron. The intron is composed of a group IB4 encoding a double-motif LAGLIDADG homing endonuclease interrupted by a group IIA1 intron. Intron insertions are shown with dashed lines. The nucleotide sequence flanking the group IIA1 insertion is shown, with the corresponding amino acid sequence shown below the nucleotide sequence. The group IIA1 intron is inserted after the second nucleotide of a codon that encodes for isoleucine (I). **(b)** Intron/exon interactions of the group IIA1 intron with its host (double-motif LAGLIDADG) exon sequence. **(c)** Proposed structure of the group IB4 intron. **(d)** Proposed structure of group IIA1 intron. Note: EBS = exon binding site, IBS = intron binding site. Group II intron tertiary interactions are shown according to Michel and Ferat (1999).

within the group I introns examined and typically encode LAGLIDADG or GIY-YIG type HEs (see Table 2.1). Some GIY-YIG type ORFs were noted to be associated with or located within LAGLIDADG type ORFs and these were found in *cytb*-201 of *Monilia mumecola* (JN204425.1), *Monilinia laxa*, *Podospora anserina* (X51480.1), *P. anserina* (X55026.1), *cytb*-490 of *Monilinia fructigena*, and *cytb*-820 of *Monilia yunnanensis* and *Monilinia laxa* (see Table 2.1). In summary, among the 350 group I introns noted, 371 intron-encoded ORFs were predicted; 107 encoded GIY-YIG type ORFs, 231 encoded LAGLIDADG type ORFs, 5 encoded N-acetyltransferases, and 28 encoded hypothetical ORFs.

Many examples of degenerating or eroding ORFs (73/377) were detected due to premature stop codons or other mutations leading to the fragmentation of the resident ORFs along with the insertion of GIY-YIG (13) or LAGLIDADG (12) type HEGs into the resident ORFs (see Table 2.1).

Many examples were noted where the intron-encoded ORF is fused (209/377) to the upstream exon of the host gene. However, freestanding ORFs within introns or "bi-orfic" introns where the posterior or both ORFs were not fused to upstream exons were noted (see Table 2.2). Among the 350 *cytb* group I introns noted in this study, 40 showed evidence of containing two or more ORFs. Finally, it was noted that in some versions of the *cytb*-429 intron, a group ID intron, ORFs could be identified that encode for a putative N-acetyltransferase. The *cytb*-429 intron usually encodes LAGLIDADG type ORFs.

2.5. Discussion

2.5.1. *cytb* intron landscape

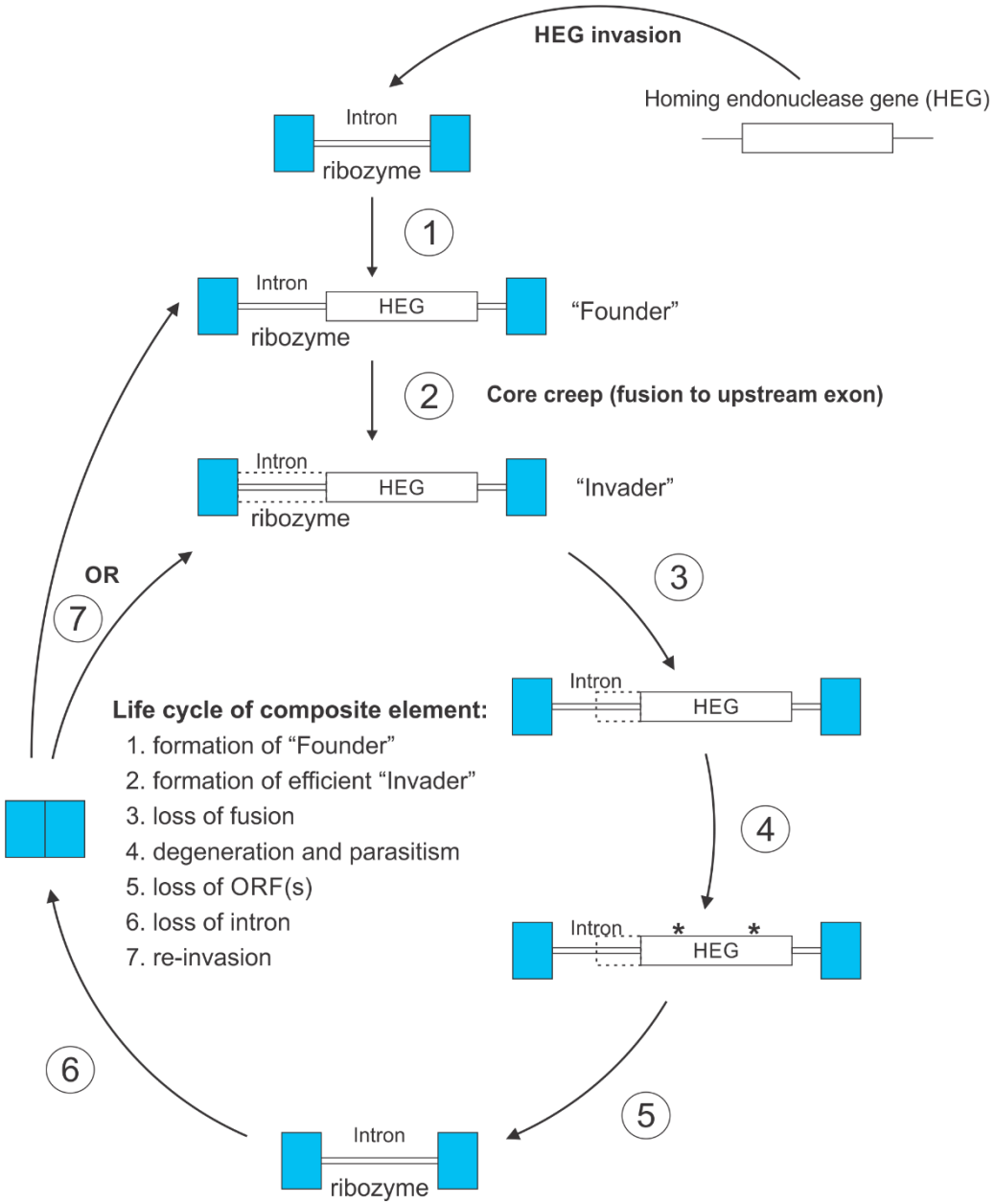
A recent study performed by Yin et al. (2012) took a more general approach in examining the *cytb* gene, with various aspects being similar to this study. In their work, they examined 69 fungal intron-containing *cytb* genes and noted 172 introns (but the insertion site of only 171 introns were determined). They observed a "patchy" (sporadic) intron distribution and found 35 different intron insertion sites among members of the Mycota within *cytb*. The focus of their study was to document the gain and loss of introns and how transposition and horizontal transfer contribute towards intron distribution. Similar findings were observed in this study, but the focus was the Ascomycota. Here, 21 intron insertion sites were noted (see Figure 2.2, detailed in Figure 2.1). The current study focused on intron types and intron ORFs and intron folds for selected introns that appear to be unique compared to standard group I or group II introns (see Figures 2.3 and 2.4). For one, as expected for a particular *cytb* insertion site, typically the same category of intron is present; a reflection that intron mobility is usually facilitated by intron encoded homing endonuclease that target (i.e., cleave) specific homing sites (Dujon, 1989; Belfort et al., 2002). The 3′-end of the gene (~ 350 nt) appears to be devoid of introns (see Figure 2.2, detailed in Figure 2.1). The latter has been observed for other genes and intron content and it is usually speculated that intron loss might be due to replacing genes or parts of genes with reverse transcribed version of the mRNA (Sverdlov et al., 2004). As reverse transcription starts at the 3′ end of the transcript and usually cDNAs are not completed for the entire mRNA due to low processivity of the RT, the 3′ regions are more likely subjected to intron loss. Some introns were more frequently encountered, in particular *cytb*-393 (94 representatives), 490 (68), 506 (39), 201 (33), and 429 (29) (see Table 2.1, summarized in Table 2.2). These locations may be highly conserved and thus preferred templates for HE activity. Alternatively, maintaining a

potentially biased intron distribution might be indicative that some mobile elements can benefit the host gene (Novikova et al., 2016).

2.5.2. Homing endonuclease genes: From intron invasion, to intron/HEG mutualism, to drift, and decay

Collectively, the *cytb* introns and their associated ORFs may offer further insight into the so called: "intron and homing endonuclease life cycle" (see Figure 2.5). Emblem et al. (2014), based on the analysis of sea anemone *cox1* mtDNAs introns, distinguished five stages in the cycle of invasion of mitochondrial group I introns: (1) initially, what they referred to as a "young" intron, have the HEG-encoding component fused in-frame with the upstream exon; (2) in more "mature" introns, the intron-encoded HE ORF connection with the upstream exon is lost and; (3) eventually, the introns show evidence of ORF degeneration; (4) leading to the loss of the ORF; and finally (5) the introns are lost. This model is essentially an expansion of the model proposed by Goddard and Burt (1999) which was based on the yeast omega system (*rnl* group IA intron and its LAGLIDADG type ORF), which appears to go through a cycle of invasion of the mL2449 site [insertion site nomenclature proposed by Johansen and Haugen (2001)], followed by deterioration (decay) of the ORF and eventual loss of the intron. Essentially, the lack of selection causes the ORF sequences to drift and accumulate mutations. Thus, survival is based on the composite intron to invade new sites (in cognate alleles or ectopic sites) in order to outpace its decay. For both models described above, a site that lost the intron is a site that can be reinvaded by an intron.

Emblem's model is compatible with the concept of core creep proposed by Edgell et al. (2011); here, the argument is presented that intron ORFs, over time, expand and the ORF


HEG invasion
Homing endonuclease gene (HEG)
Intron
ribozyme
1
Intron
HEG
ribozyme
"Founder"
2
Core creep (fusion to upstream exon)
Intron
HEG
ribozyme
"Invader"
3
OR
7
Intron
HEG
Life cycle of composite element:
1. formation of "Founder"
2. formation of efficient "Invader"
3. loss of fusion
4. degeneration and parasitism
5. loss of ORF(s)
6. loss of intron
7. re-invasion
4
Intron
HEG
6
5
Intron
ribozyme

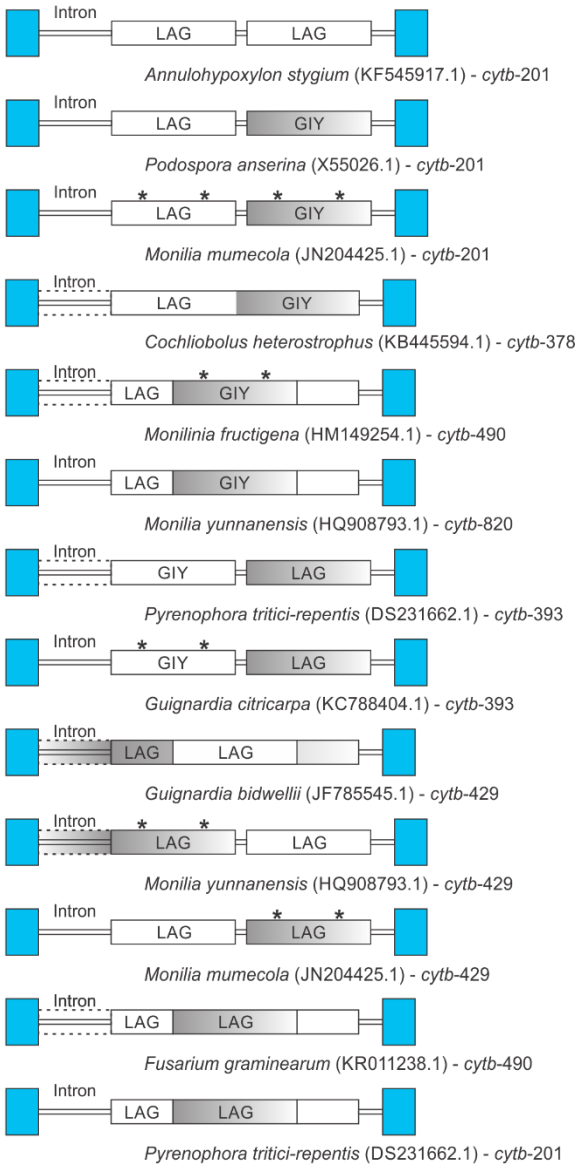

Intron
LAG
LAG
Annulohypoxylon stygium (KF545917.1) - cytb-201
Intron
LAG
GIY
Podospora anserina (X55026.1) - cytb-201
Intron
LAG
GIY
Monilia mumecola (JN204425.1) - cytb-201
Intron
LAG
GIY
Cochliobolus heterostrophus (KB445594.1) - cytb-378
Intron
LAG
GIY
Monilinia fructigena (HM149254.1) - cytb-490
Intron
LAG
GIY
Monilia yunnanensis (HQ908793.1) - cytb-820
Intron
GIY
LAG
Pyrenophora tritici-repentis (DS231662.1) - cytb-393
Intron
GIY
LAG
Guignardia citricarpa (KC788404.1) - cytb-393
Intron
LAG
LAG
Guignardia bidwellii (JF785545.1) - cytb-429
Intron
LAG
LAG
Monilia yunnanensis (HQ908793.1) - cytb-429
Intron
LAG
LAG
Monilia mumecola (JN204425.1) - cytb-429
Intron
LAG
LAG
Fusarium graminearum (KR011238.1) - cytb-490
Intron
LAG
LAG
Pyrenophora tritici-repentis (DS231662.1) - cytb-201

Figure 2.5. Proposed intron and homing endonuclease life cycle.

Overview of the various intron open reading frame (ORF) configurations observed among the surveyed cytochrome b (*cytb*) introns and a possible model depicting the possible evolutionary history of composite introns invading the *cytb* gene (see text for details). Illustrated is a possible route for the formation of a "founder" composite element evolving via core creep to a more efficient "invader" element that once fixed within a certain location will due to drift eventually decay. Examples of *cytb* intron ORF arrangements illustrate intronic ORFs extending toward the upstream exon (core creep) or mutating by losing the connection to the upstream exon, in addition some introns gain multiple ORFs (either LAGLIDADG or GIY-YIG type ORFs or both) and some ORFs are starting to erode by accumulating mutations (*). Instances where intronic ORFs appear to occur in tandem arrangements or have inserted into other intronic ORFs are also illustrated. Intronic ORFs illustrated in gray are assumed to be secondary invaders, based on comparative analysis of introns located at homologous sites. GIY = GIY-YIG type ORFs; LAG = LAGLIDADG type ORFs.

includes intron core sequences and “creeps” towards the upstream exon. The result is fusion of the intron ORF to the upstream exon that can now be potentially expressed as a fusion protein, which can be resolved post translation. This would allow the intron ORF to be expressed more efficiently as it gains the regulatory sequences of the host gene that optimize translation.

One would have to assume that the initial event leading to the formation of the composite mobile intron involved a HEG invading an intron (Loizos et al., 1994; Derbyshire and Belfort, 1998; Bonocora and Shub, 2009; Zeng et al., 2009; Barzel et al., 2011) and, here most likely, the HEG would reside within the intron probably in a loop region that would not interfere with ribozyme activity (Schäfer, 2003). We refer to this element as the “founder” mobile intron. Once this combination has invaded a site, core creep could increase the efficiency of the intron ORF expression (translation). Some intron ORFs have been shown to have maturase activity (Bolduc et al., 2003; Belfort, 2003; Caprara and Waring, 2005). Thus, efficient expression of the intron-encoded protein would increase the splicing efficiency of the intron favouring the expression of the host gene. This phase could be defined the “invader” stage, as now the intron/HE combination has established a configuration that is mutualistic, (i.e., minimizes the impact on the host gene and favours the expression of the IEP, which would increase the probability of intron mobility and efficient intron splicing). This would be expected to be the most common form of an active composite element. As suggested by Goddard and Burt (1999), once a mobile intron has exhausted its target site (fixation), drift (i.e., a lack of selection) would result in the accumulation of mutations. The first sign of this may be the loss of the intron ORF being fused to the upstream exon. Eventually, this loss of fusion can either be compensated for by alternative splicing or an intron-encoded cryptic ribosome binding site permitting the expression of the IEP. This phase would correspond to Emblem’s “mature intron stage”.

What complicates defining homing endonucleases and their host intron "life cycles" is the promiscuous nature of HEGs; as noted (see Table 2.1), numerous examples of introns with two ORFs or introns with ORFs that were disrupted by HEGs. In addition, examples of LAGLIDADG type ORFs invading either LAGLIDADG type ORFs or positioning themselves into a new and different location within a previously colonized intron, or GIY-YIG type ORFs invading (disrupting) LAGLIDADG type ORFs or inserting into a previously colonized intron (examples *cytb*-201 and *cytb*-820) were noted. At this stage, it is difficult to evaluate if the extra IEP contributes towards mobility or provides maturase activity for the host intron. The invasion of a GIY-YIG HEG into a pre-existing LAGLIDADG HEG has been observed previously in the *cytb*-i2 of *Podospora curvicolla* (Saquez et al., 2000). The HEG inserted after the first motif of a double-motif LAGLIDADG type ORF in this case (Saquez et al., 2000). GIY-YIG HEGs may benefit from being fused to the upstream LAGLIDADG coding component as it gains access to the *cis*-acting regulatory sequences required for expression. These GIY-YIG type ORFs are therefore "parasitic" and may eventually replace the original resident ORF. Saquez et al. (2000) proposed that invading pre-existing HE ORFs might be a strategy to keep to the total numbers of active HEs the same, thus minimizing the impact of HE activity on the host genome.

In addition, it was also noted that many introns encode ORFs that are starting to degrade/degenerate. With bi-orfic examples, various combinations were noted where either one or both HE ORFs were at various stages of degeneration due to the accumulation of mutations (see Figure 2.5, summarized in Table 2.1 and Table 2.2). Overall, a phase of "parasitism" by other HEGs along with ORF degeneration appears to ultimately lead to ORFs being lost. One would assume that the lack of an active HE along with splicing deficiencies will eventually cause the introns to be lost, as predicted by the Goddard and Burt model (1999).

2.5.3. Evolutionary strategies for optimizing or regulating expression of intron ORFs

It is worth noting that in some instances of "bi-orfic introns" (*Podospora anserina nad1*-i4 and *cox1*-i7) or single freestanding intron ORFs (*S. cerevisiae cox1*-i1 referred to as aI5β), alternative splicing can generate transcripts where the internal ORF sequence gets fused to the upstream exon sequence (Sellem and Belcour, 1994, 1997; Sellem et al., 1996; Turk et al., 2013; Sulo et al., 2017). Thus, core creep is not the only mechanism that may enhance the expression of the intron-encoded protein. During this study, an instance where coopting components of an intron could allow for alternative splicing was also noted. Alternative splicing may enhance or facilitate the long-term survival of the HEG. The *Chrysoporthe austroafricana cytb*-289 group IIB intron contains a partially duplicated group IIB intron sequence making up domains V and VI, which could provide a platform for alternative splicing analogous to a bacterial group IIC introns described recently by McNeil et al. (2014).

Coopting components of introns, such as the partial duplication of group II intron core elements seen in the *Chrysoporthe austroafricana cytb*-289 group IIB intron or the in-frame insertion of a group IIA1 intron in the *Chaetomium thermophilum cytb*-506 intron-encoded LAGLIDADG type ORF, would allow for improved expression and or regulation of intron ORFs (Guha and Hausner, 2014, 2016). It has been suggested that in some instances mobile introns could be environmental sensors (Belfort, 2017) these types of intron arrangements that includes potential platforms for alternative splicing or regulation of intron encoded ORFs with introns could be such sensors that deserves future investigations (Guha et al., 2017a).

2.6. Conclusions

The large compilation of introns and intron ORFs presented in this study provides a broader picture as to the evolutionary dynamics of introns and their encoded ORFs. Although many intron ORFs are prone to degeneration, introns and their ORFs also display evolutionary versatility. These include core creep of intron ORFs to generate efficient mobile units, or components of introns serving as regulatory elements to facilitate ORF expression. Some HEGs appear to be parasitizing HEGs to gain a host intron or invade new locations without increasing the intron or HEG load on the host genome. The evolutionary versatility of mobile introns has made them successful in spreading among the fungi and in comprising a significant component of the mitochondrial genome in many fungi.

Chapter 3: Intron-encoded ribosomal proteins and N-acetyltransferases within the mitochondrial genomes of fungi: Here today, gone tomorrow?

Chapter 3 is part of a manuscript that is (at the time of writing of the thesis) in progress and was reformatted to be consistent with the format used in the thesis. Text, tables, and figures were modified (added or removed) where the author felt was appropriate.

Authors and contributions

Author list: Alvan Wai, Chen Shen, Andrell Carta, Alexandra Dansen, Pedro Crous, Georg Hausner

Shen, Carta, and Dansen contributed either some original data or helped with regards to data mining. Dr. Crous allowed access to their genomic data posted within MycoCosm. Dr. Hausner helped with regards to the design of the experiment and editing of the manuscript. Wai did all the data compilation and data analysis. All tables and figures were prepared by Wai along with the first draft of the manuscript.

3.1. Abstract

In the mitochondrial genomes of the filamentous Ascomycota, aside from the usual "core" set of genes, one can sometimes encounter genes encoding for ribosomal protein S3 (*rps3*), N-acetyltransferase, and in a few instances aminotransferases. Based on a survey using sequence data from various databases, it was observed that these genes can be located within introns or exist as freestanding genes in intergenic regions. Furthermore, they can also be absent from fungal mitochondrial genomes. Overall, the *rps3* gene is highly conserved among fungal mitochondrial genomes. However, the N-acetyltransferase gene was less frequently encountered and may be a more recent import from the nuclear genome. Both genes serve as examples of genetic elements that appear to be capable of "cycling" or mobilizing between introns and intergenic regions. This "cycling" mechanism is currently not understood but may involve recombination events and/or movement via RNA intermediates.

3.2. Introduction

Genes encoded by fungal mitochondrial genomes (mtDNAs) can be classified as: (1) RNA-encoding genes involved in translation, which include the small and large ribosomal subunits (*rns* and *rnl*, respectively), and the 20 essential tRNAs (*trnA*-*W*); (2) genes encoding for components of the respiratory chain, which include the cytochrome oxidase subunits (*cox1*, *cox2*, and *cox3*), the cytochrome b (*cytb*), subunits of NADH dehydrogenase (*nad1* to *nad6*, including *nad4L*; not encoded in the mtDNAs of members of the Taphrinomycotina and some members of the Saccharomycetales), and components of ATP synthase (*atp6*, *atp8*, and *atp9*); and (3) in some instances, a gene for a ribosomal protein, referred to as *var1*, *S5*, or *rps3* (Wolf and Del Giudice, 1988; Bullerwell et al., 2003; Kennell and Cohen, 2004; Jung et al., 2009, 2012;

Procházka et al., 2010; Solieri, 2010; Eldarov et al., 2011; Mardanov et al., 2014; Freel et al., 2015). With regards to *var1* and *S5*, which are gene designations used in the literature for fungal mtDNAs, Bullerwell et al. (2003) showed that these genes are all homologs of the *rps3* gene.

Mitochondrial genomes can also contain introns and intron-encoded open reading frames (ORFs), along with unknown (or undefined) reading frames (uORFs) encoding for hypothetical proteins. Mitochondrial introns can be assigned to either group I or group II, based on secondary structures and splicing mechanism. Fungal mitochondrial introns are potential mobile elements, composed of a ribozyme component capable of facilitating RNA splicing and in some instances a protein-coding component. These intron-encoded proteins (IEPs) can catalyze intron mobility and, in some cases, can act as maturases to aid splicing (Hausner, 2003, 2012; Lang et al., 2007). Group I introns typically encode homing endonuclease genes (HEGs) whereas group II introns tend to encode reverse transcriptases (RTs). With regards to the latter, cases of group II introns encoding HEGs, probably derived from group I introns, has been observed (Belfort et al., 2002; Toor and Zimmerly, 2002; Sethuraman et al., 2013; Lambowitz and Belfort, 2015). Unusual group I intron arrangements have been noted within the mtDNA of metazoans. For example, within the phylum Cnidaria, the *nad5* group I intron can encode essential mitochondrial protein-coding genes such as *nad1*, *nad3*, etc. (Beagley et al., 1996, 1998; Emblem et al., 2011, 2014; Arrigoni et al., 2016; Chi and Johansen, 2017). Among the fungi, these types of complex introns have so far only been observed in the white-rot fungus, *Flammulina velutipes* (Agaricales) where the *nad5* group I intron (subgroup IB) encodes the *cox3* gene (Yoon et al., 2012).

N-acetyltransferases (NATs) are a large superfamily of enzymes that are universal in their distribution. These use acyl-CoAs as acetyl group donors in order to acetylate their cognate substrates (reviewed by Polevoda and Sherman, 2002, 2003; Vetting et al., 2005; Polevoda et al.,

2009; Arnesen et al., 2009; Salah Ud-Din et al., 2016). NATs (of the RimJ/RimL family) have been associated with a variety of functions such as translation, ribosome structure and biogenesis, post-translational modification, protein turnover, histone modification, antibiotic resistance in bacteria, and as chaperones (reviewed by Vetting et al., 2005; Marchler-Bauer et al., 2015; Salah Ud-Din et al., 2016). Acetylation is one of the most common protein modifications among Eukaryotes but there has also been reports of NAT activity in Bacteria (Tanaka et al., 1989; Vetting et al., 2008). In a study conducted by Duò et al. (2012), an ORF encoding a putative NAT was noted in the mtDNA of *Phialocephala subalpina*. In order to evaluate how prevalent *NAT*-encoding genes are among fungal mtDNAs, a survey was conducted on the currently available fungal mtDNA sequences.

The fungal mtDNA *rps3* gene poses an interesting enigma. Among many filamentous fungi of the Ascomycota, it appears to be frequently encoded within a group IA intron, referred to as mL2449, located in the U11 region of the *rnl* gene (Burke and RajBhandary, 1982; Bullerwell et al., 2003). It can also be found in intergenic regions as a freestanding gene or absent (reviewed by Sethuraman et al., 2009a, b; Jung et al., 2012; Freel et al., 2014; Mardanov et al., 2014; Goodwin et al., 2016). In addition, the nuclear-encoded (cytoplasmic) version of RPS3 has been shown to have multiple functions (i.e. it is a moonlighting protein), including DNA repair and host defense (Wool, 1996). In *Schizosaccharomyces pombe*, the mitochondrial-encoded version of this protein is associated with a mutator phenotype presumably implicating its role in DNA repair (Zimmer et al., 1991). Across the fungi, RPS3 is extremely variable with regards to arrangement and size, ranging from 227 to 1453 amino acids (Bullerwell et al., 2000; Sethuraman et al., 2009b). With regards to *Saccharomyces cerevisiae*, due to its high variability and AT-rich nature, the *var1* (*rps3*) gene has been speculated to be an example of a gene that

evolved by endogenous mechanisms whereby non-coding AT-rich spacer sequences eventually evolved into the *var1* ORF (reviewed by Bernardi, 2005). The objective was to survey the mtDNAs of fungi that belong to the Pezizomycotina (which includes the filamentous ascomycetes) for the presence of novel intron-encoded ORFs in order to gain a better understanding of factors that may influence the mtDNA architecture.

3.3. Material and Methods

3.3.1. Sequence analysis

Fungal mtDNAs contained within the NCBI organellar resource database (http://www.ncbi.nlm.nih.gov/genome/organelle/) and Joint Genome Institute (JGI) Genome Portal (MycoCosm; http://genome.jgi.doe.gov/; Nordberg et al., 2014) were examined. In order to distinguish between sequences from two the two databases, the source will be indicated within parenthesis. When referring to data from NCBI, GenBank accession numbers (specified in: https://www.ncbi.nlm.nih.gov/Sequin/acc.html) will be used. For data from JGI, the source will be explicitly stated along with an identifier in order to distinguish between different strains. In some instances, mtDNAs were reanalyzed (annotated) with the programs, MFannot and RNAweasel (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl and http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl, respectively; Gautheret and Lambert, 2001; Lang et al., 2007). Intron/exon junctions were confirmed by blastn analysis; comparing intron-containing alleles with cognate alleles that lack introns/insertions. The ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html; Genetic code setting for molds #4) was used to search for potential ORFs within introns and intergenic regions. The online resource Basic Local Alignment Search Tool (BLAST;

http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul et al., 1990) was used to retrieve amino acid from GenBank, which shared similarities to the mtDNA NAT sequence of *Annulohypoxylon stygium* (AHB33527.1) or RPS3 sequence of *Ophiostoma novo-ulmi* (CM001753.1). In addition, the NAT and RPS3 sequences were also collected from the JGI Genome Portal by blastn using the NAT nucleotide sequence from *Stemphylium lycopersici* (KX453765.1) as the query. The recovered NAT or RPS3 sequences were translated using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) and compiled within GeneDoc (Nicholas et al., 1997). Both datasets were aligned with the online program MAFFT (http://mafft.cbrc.jp/alignment/server/index.html; alignment strategy setting: E-INS-I; Katoh and Standley, 2013). The alignments were adjusted in GeneDoc, where necessary.

3.3.2. Intron folding

The fifth intron of the *nad5* gene [*nad5*-i5 or *nad5*-710; latter nomenclature based on the convention proposed by Johansen and Haugen (2001) for ribosomal genes] of *Phyllosticta citriasiana* (JGI) was predicted by the RNAweasel program (Lang et al., 2007) to be a group ID intron. This was confirmed and folded based on a personal dataset of group ID introns. The intron and flanking exon sequences (~ 10 nt) were aligned by the Multiple Alignment using Fast Fourier Transform program, MAFFT (alignment strategy setting: E-INS-I). P1, P2, P5, and P9 helices, were predicted by Mfold (Zuker, 2003) and supported by comparative sequence analysis. Introns were drawn using CorelDRAW Graphics Suite X6 (Corel Corporation, Ottawa).

3.3.3. Phylogenetic analysis of NAT and RPS3 amino acid sequences

The phylogenetic tree for the NAT and RPS3 sequences were inferred based on Bayesian analyses (MrBayes, version 3.2.6; Ronquist and Huelsenbeck, 2003; Ronquist, 2004; Ronquist et al., 2012). The best substitution models to use for the analyses were determined in MEGA7 (Kumar et al., 2016). The Whelan and Goldman (WAG; Whelan and Goldman, 2001) model was selected for the NAT dataset and the cpREV (Adachi et al., 2000) model was applied for the RPS3 dataset in Bayesian analyses. In MrBayes, the following parameters were selected: four chains were run simultaneously for 5 000 000 generations and trees were sampled every 100 generations. The first 40% of the trees generated were discarded ("burn-in"). Remaining trees were used to compute the majority rule consensus tree. A tree (see Figure 3.1) was also generated based on the RPS3 dataset using FastTree (Price et al., 2010) with default settings.

3.4. Results

3.4.1. Among the Pezizomycotina, the *rps3* gene can be intron-encoded, freestanding, or missing from the mitochondrial genome

A schematic phylogenetic tree was generated based on currently available phylogenies for the Ascomycota constructed on the basis of multigene analyses (see Figure 3.1; Hibbett et al., 2007; Schoch et al., 2009). The absence/presence and position of the *rps3* gene was indicated within the context of the phylogenetic position among various members of the Ascomycota (see Figure 3.1, additional information in Table 3.1).

3.4.2. Encoding pattern of the *rps3* gene within the Pezizomycetes

The Pezizales are currently viewed to be the earliest branching members among the

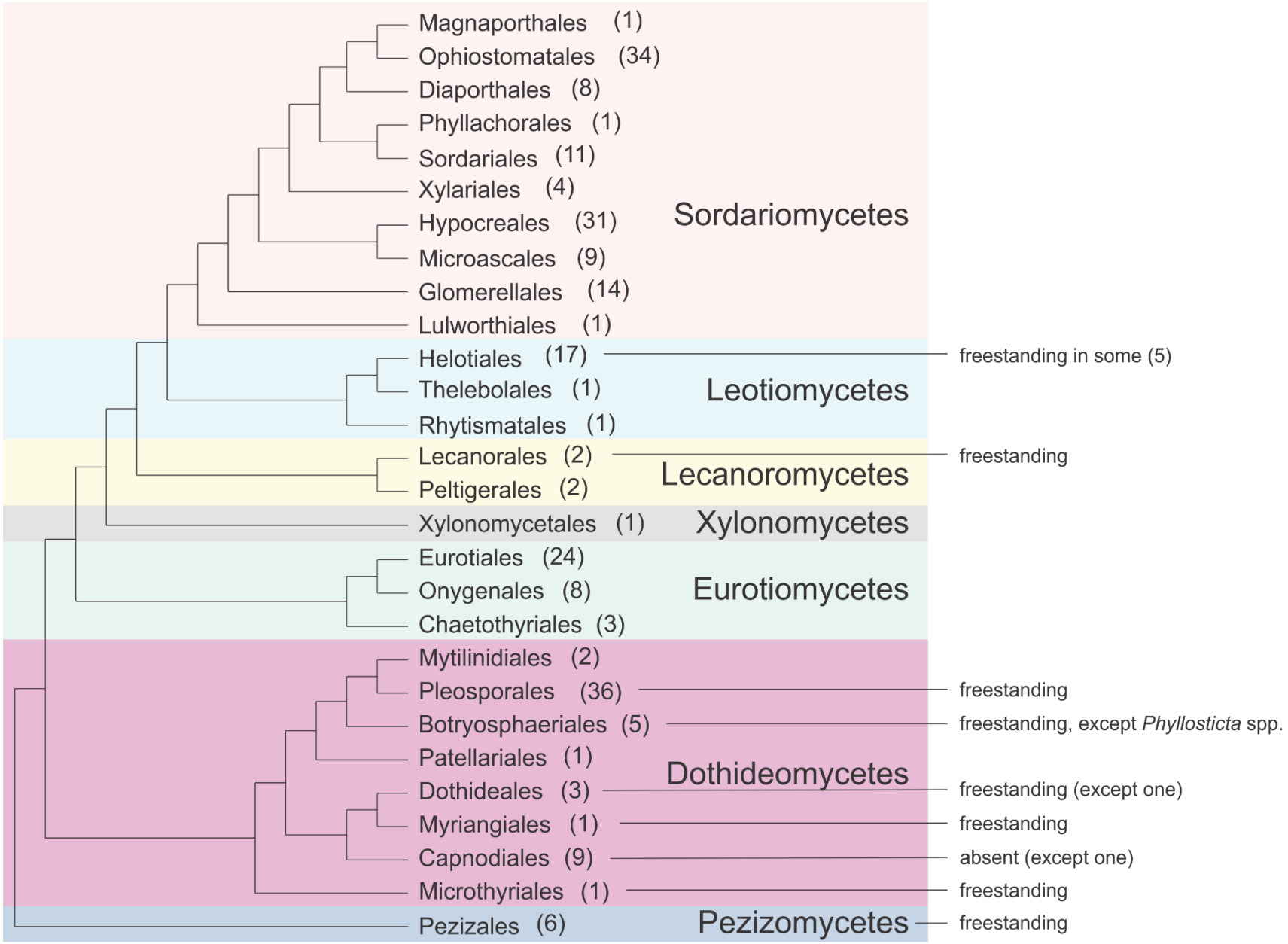

Magnaporthales (1)
Ophiostomatales (34)
Diaporthales (8)
Phyllachorales (1)
Sordariales (11)
Xylariales (4)
Hypocreales (31)
Microascales (9)
Glomerellales (14)
Lulworthiales (1)
Sordariomycetes
Helotiales (17)
Thelebolales (1)
Rhytismatales (1)
Leotiomycetes
freestanding in some (5)
Lecanorales (2)
Peltigerales (2)
Lecanoromycetes
freestanding
Xylonomycetales (1)
Xylonomycetes
Eurotiales (24)
Onygenales (8)
Chaetothyriales (3)
Eurotiomycetes
Mytilinidiales (2)
Pleosporales (36)
Botryosphaeriales (5)
Patellariales (1)
Dothideales (3)
Myriangiales (1)
Capnodiales (9)
Microthyriales (1)
Dothideomycetes
freestanding
freestanding, except *Phyllosticta* spp.
freestanding (except one)
freestanding
absent (except one)
freestanding
Pezizales (6)
Pezizomycetes
freestanding

Figure 3.1. Schematic phylogenetic tree based on current phylogeny of the Ascomycota. Orders are labeled on the leaves of the tree. Numbers within parenthesis beside the order names represent the number of fungi sampled from the respective order. Orders with members encoding freestanding *rps3* genes or do not encode a mitochondrial *rps3* gene (i.e. absent) are indicated. Orders not indicated as freestanding or absent have *rps3* genes encoded within the large ribosomal subunit-encoded group IA intron located at position mL2449 (nomenclature based on Johansen and Haugen, 2001). Fungi (seven in total) that have not been resolved at the order level are not included in the tree. See Table 3.1 for more details. In brief, this includes three species from the Dothideomycetes, two from the Leotiomycetes, two from the Sordariomycetes.

Table 3.1. The encoding pattern of the *rps3* gene among Pezizomycotina fungi surveyed from the Joint Genome Institute MycoCosm web portal and the National Center for Biotechnology Information.

	Location of *rps3*^	*rnl* intron(s)*	Order
Pezizomycetes			
***Caloscypha fulgens* (ATCC 42695; JGI) - v1.0**	*trnR-trnQ*	> 3	Pezizales
***Choiromyces venosus* (120613-1; JGI) - v1.0**	*trnW-trnR*	> 3	Pezizales
***Gyromitra esculenta* (CBS 101906; JGI) - v1.0**	*trnR-trnQ*	> 3	Pezizales
***Morchella importuna* (SCYDJ1-A1; JGI) - v1.0**	*trnR-trnQ*	> 3	Pezizales
***Pyronema omphalodes* (AMO66548.1)**	*trnL-Rnrt*	3	Pezizales
***Tuber borchii* (Tbo3840; JGI) - v1.0**	*trnW-trnR*	> 3	Pezizales
Eurotiomycetes			
***Cladophialophora bantiana* (ANK79138.1)**	*rnl*	I?(ORF?); IA(*rps3*)	Chaetothyriales
***Coccodinium bartschii* (CBS 121709; JGI) - v1.0**	*rnl*	I?(-); IB(GIY); IA(*rps3*)	Chaetothyriales
***Exophiala dermatitidis* (EHY51767.1)**	*rnl*		Chaetothyriales
***Aspergillus clavatus* (AFD96021.1)**	*rnl*		Eurotiales
***Aspergillus fumigatus* (AFE02838.1)**	*rnl*		Eurotiales
***Aspergillus glaucus* (JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Aspergillus luchuensis* (CBS 106.47; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Aspergillus nidulans* (AFC69019.1)**	*rnl*		Eurotiales
***Aspergillus oryzae* (AFM82517.1)**	*rnl*		Eurotiales
***Aspergillus ruber* (EYE89930.1)**	*rnl*		Eurotiales
***Aspergillus terreus* (AFD96057.1)**	*rnl*		Eurotiales
***Neosartorya fischeri* (AFD95940.1)**	*rnl*		Eurotiales
***Penicillium bilaiae* (ATCC 20851; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium brevicompactum* (JGI) - v2.0**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium chrysogenum* (BAA02977.1)**	*rnl*		Eurotiales
***Penicillium digitatum* (ADU57311.1)**	*rnl*		Eurotiales
***Penicillium glabrum* (DAOM 239074; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium janthinellum* (ATCC 10455; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium lanosocoeruleum* (ATCC 48919; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium raistrickii* (ATCC 10490; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium roqueforti* (KR952335.1)**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium solitum* (JN696111.1)**	*rnl*	IA(*rps3*)	Eurotiales
***Penicillium urticae* (BAA03421.1)**	*rnl*		Eurotiales
***Talaromyces aculeatus* (ATCC 10409; JGI) - v1.0**	*rnl*	IA(*rps3*)	Eurotiales
***Talaromyces marneffei* (AAQ54923.1)**	*rnl*		Eurotiales

	Location of *rps3*^	*rnl* intron(s)*	Order
***Talaromyces stipitatus* (AFD95916.1)**	*rnl*		Eurotiales
***Xeromyces bisporus* (HG983520.1)**	*rnl*		Eurotiales
***Arthroderma obtusum* (ACR19608.1)**	*rnl*		Onygenales
***Arthroderma otae* (ACR19622.1)**	*rnl*	IA(*rps3*)	Onygenales
***Arthroderma uncinatum* (ACR19605.1)**	*rnl*		Onygenales
***Epidermophyton floccosum* (AAW78226.1)**	*rnl*		Onygenales
***Histoplasma capsulatum* (EER36345.1)**	*rnl*		Onygenales
***Paracoccidioides brasiliensis* (AAY30338.1)**	*rnl*		Onygenales
***Trichophyton mentagrophytes* (ACR19584.1)**	*rnl*		Onygenales
***Trichophyton rubrum* (ACR19560.1)**	*rnl*		Onygenales
Dothideomycetes			
***Dothidotthia symphoricarpi* (JGI) - v1.0**	*Rnrt-trnC*	?	Botryosphaeriales
***Phyllosticta capitalensis* (CBS 128856; JGI) - v1.0**	*nad5*		Botryosphaeriales
***Phyllosticta citriasiana* (JGI) - v1.0**	*nad5*	> 3	Botryosphaeriales
***Phyllosticta citribraziliensis* (CBS 100098; JGI) - v1.0**	*nad5*	> 3	Botryosphaeriales
***Saccharata proteae* (CBS 121410; JGI) - v1.0**	*nad2-atp6*	intronless	Botryosphaeriales
***Baudoinia panamericana* (EMC96256.1)**	nuclear	?	Capnodiales
***Cercospora zeae-maydis* (JGI) - v1.0**	nuclear	IA(-?)	Capnodiales
***Dissoconium aciculare* (JGI) - v1.0**	nuclear	intronless	Capnodiales
***Dothistroma septosporum* (EME43844.1)**	nuclear		Capnodiales
***Mycosphaerella fijiensis* (JGI) - v2.0**	nuclear	intronless	Capnodiales
***Pseudocercospora fijiensis* (EME79251.1)**	nuclear		Capnodiales
***Pseudovirgaria hyperparasitica* (CBS 121739; JGI) - v1.0**	*rnl*	IB(?); IA(*rps3*); LAG?	Capnodiales
***Septoria musiva* (SO2202; JGI) - v1.0**	nuclear	IA(LAG1)	Capnodiales
***Sphaerulina musiva* (EMF13745.1)**	nuclear		Capnodiales
***Delphinella strobiligena* (CBS 735.71; JGI) - v1.0**	*nad6-cox3*	> 3	Dothideales
***Lineolata rhizophorae* (ATCC 16933; JGI) - v1.0**	*rnl*	IA(*rps3*)	Dothideales
***Setomelanomma holmii* (CBS 110217; JGI) - v1.0**	*nad3-1dan*	intronless	Dothideales
***Tothia fuscella* (CBS 130266; JGI) - v1.0**	*rns-trnK*	intronless	Microthyriales
***Elsinoë ampelina* (CECT 20119; JGI) - v1.0**	*trnR-rns*	IA(-?)	Myriangiales
***Lophium mytilinum* (CBS 269.34; JGI) - v1.0**	*rnl*	IA(*rps3*)	Mytilinidiales
***Mytilinidion resinicola* (CBS 304.34; JGI) - v1.0**	*rnl*	IA(*rps3*)	Mytilinidiales
***Patellaria atrata* (JGI) - v1.0**	*rnl*	IA(*rps3*)	Patellariales

	Location of *rps3*^	*rnl* intron(s)*	Order
***Aaosphaeria arxii* (CBS 175.79; JGI) - v1.0**	*nad4-nad1*	intronless	Pleosporales
***Alternaria alternata* (SRC1lrK2f; JGI) - v1.0**	*trnI-1dan*	intronless?	Pleosporales
***Alternaria brassicicola* (JGI)**	*trnV-1dan*		Pleosporales
***Amniculicola lignicola* (CBS 123094; JGI) - v1.0**	*trnG-cox3*	IC1(-); IC1(GIY)	Pleosporales
***Bimuria novae-zelandiae* (CBS 107.79; JGI) - v1.0**	*trnF-atp6*?	intronless	Pleosporales
***Bipolaris cookei* (MF784482.1)**	*cox3-1dan*	IC1(GIY)	Pleosporales
***Bipolaris maydis* (ENH99013.1)**	*cox3-1dan*	intronless	Pleosporales
***Bipolaris oryzae* (KI964241.1)**	*cox3-1dan*		Pleosporales
***Byssothecium circinans* (CBS 675.92; JGI) - v1.0**	*nad4L-cytb*	intronless	Pleosporales
***Clathrospora elynae* (CBS 161.51; JGI) - v1.0**	*cox3-1dan*	intronless	Pleosporales
***Clohesyomyces aquaticus* (JGI) - v1.0**	*trnD-nad4L*	IC1(GIY)	Pleosporales
***Cochliobolus lunatus* (m118; JGI) - v2.0**	*cox3-1dan*	intronless	Pleosporales
***Corynespora cassiicola* (CCP; JGI) - v1.0**	*trnC-trnI*	intronless	Pleosporales
***Delitschia confertaspora* (ATCC 74209; JGI) - v1.0**	*cox2-cox3*	intronless	Pleosporales
***Didymella pinodes* (KT946597.1)**	*cytb-nad1*	intronless	Pleosporales
***Karstenula rhodostoma* (CBS 690.94; JGI) - v1.0**	*nad5-atp6*	intronless	Pleosporales
***Leptosphaeria biglobosa* (FO905608.1)**	*Wnrt-trnL*	intronless	Pleosporales
***Leptosphaeria maculans* (FP929115.1)**	*nad2-1dan*	> 3	Pleosporales
***Lindgomyces ingoldianus* (ATCC 200398; JGI) - v1.0**	*cytb-5dan*	IC1(GIY)	Pleosporales
***Lophiostoma macrostomum* (JGI) - v1.0**	*cox2-nad4*	intronless	Pleosporales
***Lophiotrema nucula* (CBS 627.86; JGI) - v1.0**	*trnC-trnC*	intronless	Pleosporales
***Melanomma pulvis-pyrius* (JGI) - v1.0**	*trnC-nad4*	intronless	Pleosporales
***Ophiobolus disseminans* (CBS 113818; JGI) - v1.0**	*cytb-nad4L*	intronless	Pleosporales
***Paraconiothyrium sporulosum* (AP3s5-JAC2a; JGI) - v1.0**	*nad5-atp6*	intronless	Pleosporales
***Parastagonospora nodorum* (ABU49441.1)**	*cox2-nad4*	intronless	Pleosporales
***Phaeosphaeriaceae* sp. (PMI 808; JGI) - v1.0**	*trnV-nad1*	IC1(-)	Pleosporales
***Polyplosphaeria fusca* (CBS 125425; JGI) - v1.0**	*trnC-cox3*	intronless	Pleosporales
***Pyrenochaeta* sp. (DS3sAY3a; JGI) - v1.0**	*cox2-trnV*	intronless	Pleosporales
***Setosphaeria turcica* (NY001; JGI) - v1.0**	*atp6-nad1*	> 3	Pleosporales
***Shiraia bambusicola* (AJI44518.1)**	*trnL-cox3*	intronless	Pleosporales
***Sporormia fimetaria* (JGI) - v1.0**	*nad5-nad1*	IB(-)	Pleosporales
***Stagonospora* sp. (SRC1lsM3a; JGI) - v1.0**	*cytb-1dan*	intronless	Pleosporales
***Trematosphaeria pertusa* (CBS 122368; JGI) - v1.0**	*nad2-atp6*	intronless	Pleosporales

	Location of *rps3*^	*rnl* intron(s)*	Order
***Trichodelitschia bisporula* (CBS 262.69; JGI) - v1.0**	*rnpB-trn*?	intronless	Pleosporales
***Verruculina enalia* (CBS 304.66; JGI) - v1.0**	*cox3-nad4*	intronless	Pleosporales
***Westerdykella ornate* (CBS 379.55; JGI) - v1.0**	*trnV-cytb*	intronless	Pleosporales
***Lepidopterella palustris* (JGI) - v1.0**	*rnl*	IA(*rps3*)	*incertae sedis*
***Lizonia empirigonia* (CBS 542.76; JGI) - v1.0**	*cytb-nad4L*	intronless	*incertae sedis*
***Macroventuria anomochaeta* (CBS 525.71; JGI) - v1.0**	*trnC-nad1*	intronless	*incertae sedis*
Lecanoromycetes			
***Lecanora strobilina* (AMZ84236.1)**	*nad4-cytb*	intronless	Lecanorales
***Usnea florida* (ATCC 18376; JGI) - v1.0**	*nad5-trnV*	> 3	Lecanorales
***Peltigera malacea* (AEK48294.1)**	*rnl*	IC1(-);IC1(GIY);IA(*rps3*)	Peltigerales
***Peltigera membranacea* (AEK48323.1)**	*rnl*	> 3	Peltigerales
Leotiomycetes			
***Acephala macrosclerotiorum* (EW76-UTF0540; JGI) - v1.0**	*cox2-nad4L*	intronless	Helotiales
***Botrytis cinerea* (AGN49003.1)**	*rnl*		Helotiales
***Bulgaria inquinans* (CBS 118.31; JGI) - v1.0**	*rnl*	> 3	Helotiales
***Cadophora* sp. (DSE1049; JGI) - v1.0**	*rnl*	IA(*rps3*)	Helotiales
***Cairneyella variabilis* (KU168424.1)**	*rnl*	IA(*rps3*)	Helotiales
***Glarea lozoyensis* (AGN74485.1)**	*rnl*		Helotiales
***Leptodontium* sp. (PMI 412; JGI) - v1.0**	*rnl*	IA(*rps3*)	Helotiales
***Loramyces macrosporus* (CBS 235.53; JGI) - v1.0**	*cox2-nad4L*	intronless	Helotiales
***Marssonina brunnea* (JN204424.1)**	*cox2-nad4L*	IC1(GIY)	Helotiales
***Phialocephala scopiformis* (CBS 120377; JGI) - v1.0**	*cox2-nad4L*	intronless	Helotiales
***Phialocephala subalpina* (AEI52989.1)**	*cox2-nad4L*	intronless	Helotiales
***Rhynchosporium agropyri* (AHC02296.1)**	*rnl*		Helotiales
***Rhynchosporium orthosporum* (AHC02389.1)**	*rnl*		Helotiales
***Rhynchosporium secalis* (AHC02417.1)**	*rnl*		Helotiales
***Rutstroemia firma* (CBS 115.86; JGI) - v1.0**	*rnl*	> 3	Helotiales
***Sarcotrochila macrospora* (ACZ97563.1)**	*rnl*		Helotiales
***Sclerotinia borealis* (AHX82983.1)**	*rnl*	> 3	Helotiales
***Coccomyces strobi* (CBS 202.91; JGI) - v1.0**	*rnl*	> 3	Rhytismatales
***Thelebolus microsporus* (ATCC 90970; JGI) - v1.0**	*rnl*	IA(*rps3*)	Thelebolales
***Pseudogymnoascus destructans* (AEFC01003580.1)**	*rnl*		*incertae sedis*
***Pseudogymnoascus pannorum* (KR055655.1)**	*rnl*	IA(*rps3*)	*incertae sedis*

	Location of *rps3*^	*rnl* intron(s)*	Order
Sordariomycetes			
***Chrysoporthe austroafricana* (AMX22137.1)**	*rnl*	> 3	Diaporthales
***Chrysoporthe cubensis* (AMX22244.1)**	*rnl*	I?(LAG2?); IA(*rps3*)	Diaporthales
***Chrysoporthe deuterocubensis* (AMX22200.1)**	*rnl*	> 3	Diaporthales
***Coniella* sp. (B22-T-1; JGI) - v1.0**	*rnl*	IA(*rps3*); IA(LAG)	Diaporthales
***Cryphonectria parasitica* (KT428651.1)**	*rnl*	> 3	Diaporthales
***Cryphonectria parasitica* (EP155; JGI) - v2.0**	*rnl*	> 3	Diaporthales
***Diaporthe longicolla* (KP137411.1)**	*rnl*	IA(*rps3*)	Diaporthales
***Lollipopaia minuta* P26 (CBS 116597; JGI) - v1.0**	*rnl*	IC1(-); IA(*rps3*)	Diaporthales
***Colletotrichum acutatum* (AKJ86848.1)**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum caudatum* (CBS 131602; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum cereale* (CBS 129662; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum eremochloae* (CBS 129661; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum falcatum* (MAFF306170; JGI)**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum godetiae* (CBS 193.32; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum graminicola* (EFQ24852.1)**	*rnl*		Glomerellales
***Colletotrichum lindemuthianum* (AHI96224.1)**	*rnl*		Glomerellales
***Colletotrichum lupini* (CBS 109225; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum phormii* (CBS 102054; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum somersetensis* (CBS 131599; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Colletotrichum zoysiae* (MAFF235873; JGI) - v1.0**	*rnl*	IA(*rps3*)	Glomerellales
***Glomerella graminicola* (CM001021.1)**	*rnl*	IA(*rps3*)	Glomerellales
***Verticillium dahliae* (ABC60419.1)**	*rnl*		Glomerellales
***Acremonium alcalophilum* (JGI) - v2.0**	*rnl*	IA(*rps3*)	Hypocreales
***Acremonium chrysogenum* (KF757229.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Acremonium implicatum* (AJI44519.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Beauveria bassiana* (ABY61762.1)**	*rnl*		Hypocreales
***Beauveria pseudobassiana* (AGW46921.1)**	*rnl*		Hypocreales
***Cordyceps brongniartii* (ABU50166.1)**	*rnl*		Hypocreales
***Cordyceps militaris* (AGY61382.1)**	*rnl*		Hypocreales
***Claviceps purpurea* (CCE35431.1)**	*rnl*		Hypocreales
***Clonostachys rosea* (CBS 125111; JGI) - v1.0**	*rnl*	IA(*rps3*)	Hypocreales
***Fusarium acuminatum* (CDL73441.1)**	*rnl*		Hypocreales

	Location of *rps3*^	*rnl* intron(s)*	Order
***Fusarium avenaceum* (JQGE01000002.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Fusarium circinatum* (AFX93769.1)**	*rnl*		Hypocreales
***Fusarium culmorum* (AKE07416.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Fusarium fujikuroi* (AFX93798.1)**	*rnl*		Hypocreales
***Fusarium graminearum* (ABC86618.1)**	*rnl*		Hypocreales
***Fusarium oxysporum* (AAW67496.1)**	*rnl*		Hypocreales
***Fusarium oxysporum* (AAX21823.1)**	*rnl*		Hypocreales
***Fusarium oxysporum* f. sp. *lycopersici* (4287; JGI) - v2**	*rnl*	IA(*rps3*)	Hypocreales
***Fusarium solani* (AEJ72888.1)**	*rnl*		Hypocreales
***Hypomyces aurantius* (ANC62706.1)**	*rnl*	> 3	Hypocreales
***Lecanicillium muscarium* (AAO14657.1)**	*rnl*		Hypocreales
***Metarhizium anisopliae* (AAW58815.1)**	*rnl*		Hypocreales
***Metarhizium robertsii* (EXU94425.1)**	*rnl*		Hypocreales
***Nectria cinnabarina* (AND76626.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Neotyphodium lolii* (AHI96208.1)**	*rnl*	3	Hypocreales
***Pochonia chlamydosporia* (AGY95315.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Purpureocillium* sp. (UdeA0106; JGI) - v1.0**	*rnl*	IA(*rps3*)	Hypocreales
***Trichoderma asperellum* (AKK32420.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Trichoderma gamsii* (ANC73538.1)**	*rnl*	IA(*rps3*)	Hypocreales
***Trichoderma harzianum* (KR952346.1)**	*rnl*		Hypocreales
***Trichoderma reesei* (AAL74166.1)**	*rnl*		Hypocreales
***Lindra thalassiae* (JK4322; JGI) - v1.0**	*rnl*	IA(*rps3*)	Lulworthiales
***Magnaporthiopsis poae* (ATCC 64411; JGI)**	*rnl*	> 3	Magnaporthales
***Ceratocystis autographa* (UAMH 4934; our own sequence)**	*rnl*		Microascales
***Ceratocystis cacaofunesta* (AFO38088.1)**	*rnl*		Microascales
***Ceratocystis coerulescens* (ACZ97560.1)**	*rnl*		Microascales
***Ceratocystis curvicollis* (ACV41157.1)**	*rnl*		Microascales
***Ceratocystis fimbriata* (ACZ97568.1)**	*rnl*		Microascales
***Gondwanamyces proteae* (ACZ97562.1)**	*rnl*		Microascales
***Kernia pachypleura* (ACZ97564.1)**	*rnl*		Microascales
***Sphaeronaemella fimicola* (ACZ97566.1)**	*rnl*		Microascales
***Sphaeronaemella fimicola* (ACZ97567.1)**	*rnl*		Microascales
***Ceratocystiopsis brevicomis* (ACV41155.1)**	*rnl*		Ophiostomatales

	Location of *rps3*^	*rnl* intron(s)*	Order
***Ceratocystiopsis minuta-bicolor* (ACV41170.1)**	*rnl*		Ophiostomatales
***Ceratocystiopsis parva* (ACV41169.1)**	*rnl*		Ophiostomatales
***Ceratocystiopsis rollhanseniana* (ACV41168.1)**	*rnl*		Ophiostomatales
***Cornuvesica falcata* (ACZ97561.1)**	*rnl*		Ophiostomatales
***Graphilbum curvicolle* (ACV41157.1)**	*rnl*		Ophiostomatales
***Grosmannia aurea* (ACV41162.1)**	*rnl*		Ophiostomatales
***Grosmannia europhioides* (ACV41151.1)**	*rnl*		Ophiostomatales
***Grosmannia europhioides* (ACV41156.1)**	*rnl*		Ophiostomatales
***Grosmannia laricis* (ACV41166.1)**	*rnl*		Ophiostomatales
***Grosmannia penicillata* (ACV41146.1)**	*rnl*		Ophiostomatales
***Grosmannia penicillata* (ACV41148.1)**	*rnl*		Ophiostomatales
***Grosmannia piceaperda* (ACV41152.1)**	*rnl*		Ophiostomatales
***Grosmannia pseudoeurophioides* (ACV41163.1)**	*rnl*		Ophiostomatales
***Leptographium pityophilum* (ACV41147.1)**	*rnl*		Ophiostomatales
***Leptographium procerum* (ACV41165.1)**	*rnl*		Ophiostomatales
***Leptographium truncatum* (ACV41150.1)**	*rnl*		Ophiostomatales
***Leptographium truncatum* (ACV41164.1)**	*rnl*		Ophiostomatales
***Leptographium truncatum* (ACV41167.1)**	*rnl*		Ophiostomatales
***Ophiostoma distortum* (ACV41160.1)**	*rnl*		Ophiostomatales
***Ophiostoma himal-ulmi* (ACV41177.1)**	*rnl*		Ophiostomatales
***Ophiostoma ips* (ACV41172.1)**	*rnl*		Ophiostomatales
***Ophiostoma ips* (ACV41173.1)**	*rnl*		Ophiostomatales
***Ophiostoma megalobrunneum* (ACV41171.1)**	*rnl*		Ophiostomatales
***Ophiostoma minus* (ACV41174.1)**	*rnl*		Ophiostomatales
***Ophiostoma minus* (ACV41175.1)**	*rnl*		Ophiostomatales
***Ophiostoma nigrum* (ACV41161.1)**	*rnl*		Ophiostomatales
***Ophiostoma novo-ulmi* subsp. *americana* (AAY59060.1)**	*rnl*		Ophiostomatales
***Ophiostoma novo-ulmi* subsp. *americana* (AAY59061.1)**	*rnl*		Ophiostomatales
***Ophiostoma tetropii* (ACV41158.1)**	*rnl*		Ophiostomatales
***Ophiostoma tetropii* (ACV41159.1)**	*rnl*		Ophiostomatales
***Ophiostoma ulmi* (ACV41153.1)**	*rnl*		Ophiostomatales
***Sporothrix schenckii* (AB568599.1)**	*rnl*		Ophiostomatales
***Sporothrix* sp. (ACV41149.1)**	*rnl*		Ophiostomatales

	Location of *rps3*^	*rnl* intron(s)*	Order
***Plectosphaerella cucumerina* (DS2psM2a2; JGI) - v1.0**	*rnl*	IA(*rps3*)	Phyllachorales
***Chaetomium thermophilum* (AEI26005.1)**	*rnl*		Sordariales
***Gelasinospora tetrasperma* (ACZ97559.1)**	*rnl*		Sordariales
***Madurella mycetomatis* (AEY94393.1)**	*rnl*		Sordariales
***Neurospora crassa* (AGG15994.1)**	*rnl*		Sordariales
***Neurospora tetrasperma* (FGSC 2509; JGI) mat_a v2.0**	*rnl*	IA(*rps3*)	Sordariales
***Podospora anserina* (CAA32860.1)**	*rnl*		Sordariales
***Podospora curvicolla* (TEP21a; JGI) - v1.0**	*rnl*	IA(*rps3*)	Sordariales
***Sordaria fimicola* (FJ717863.1)**	*rnl*		Sordariales
***Thielavia antarctica* (CBS 123565; JGI) - v1.0**	*rnl*	IA(*rps3*)	Sordariales
***Thielavia appendiculata* (CBS 731.68; JGI) - v1.0**	*rnl*	IA(*rps3*)	Sordariales
***Thielavia terrestris* (JGI) - v2.0**	*rnl*	IA(*rps3*)	Sordariales
***Annulohypoxylon stygium* (AHB33506.1)**	*rnl*	> 3	Xylariales
***Anthostoma avocetta* (NRRL 3190; JGI) - v1.0**	*rnl*	IA(*rps3*)	Xylariales
***Microdochium bolleyi* (J235TASD1; JGI) - v1.0**	*rnl*	IA(*rps3*)	Xylariales
***Xylaria hypoxylon* (OSC100004; JGI) - v1.0**	*rnl*	IA(*rps3*)	Xylariales
***Acidothrix acidophila* (CBS 136259; JGI) - v1.0**	*rnl*	IA(*rps3*)	*incertae sedis*
***Sodiomyces alkalinus* (JGI) - v1.0**	*rnl*	IA(*rps3*)	*incertae sedis*
Xylonomycetes			
***Xylona heveae* (TC161; JGI) - v1.0**	*rnl*	IA(*rps3*)	Xylonomycetales

^ some genes are written reverse indicating they are encoded on the opposite DNA strand relative to the *rps3* gene

* empty cells represent *rnl* genes that were not examined in greater beyond noting that it at least encodes for the mL2449 intron containing the *rps3* gene

Abbreviations used for genes and introns: *cox* = cytochrome oxidase subunit, *cytb* = cytochrome b; *nad* = NADH dehydrogenase subunit; *rnl* = large ribosomal subunit; *rps3* = ribosomal protein S3; *trn* = transfer ribonucleic acid (tRNA); letter beside "*trn*" = amino acid carried by the tRNA; IA, IB, IC1 = group I intron subtype; The open reading frame (ORF) encoded by an intron is denoted within parenthesis. GIY = GIY-YIG type homing endonuclease; LAG = LAGLIDADG type homing endonuclease; v = version; ? = status of component could not be determined; - = absence of ORF within intron; > 3 = more than three introns (ORF or ORF-minus)

filamentous Ascomycota (Hibbett et al., 2007; Lücking et al., 2009). Six sequences from the Pezizomycetes were surveyed. In all cases, the *rps3* gene is freestanding and the *rnl* gene appears to have an intron inserted at mL2449 encoding for a double-motif LAGLIDADG type ORF. A similar arrangement is seen in some members of the Saccharomycetales. In these fungi, *var1* (*rps3*), when present, is a freestanding gene and the *rnl* gene may contains an intron (mL2449) that encodes for a LAGLIDADG homing endonuclease. A well-known example is the omega (ω) system of *Saccharomyces cerevisiae* (reviewed by Dujon, 1989; Goddard and Burt, 1999).

The next few sections (**3.4.3** to **3.4.6**) will be covering the various arrangements/encoding patterns of the *rps3* gene. The reader is directed to Figure 3.1 and Table 3.1 for additional information.

3.4.3. Encoding pattern of the *rps3* gene within the Dothideomycetes

Among members of the Dothideomycetes, the *rps3* gene was noted to have a variety of arrangements. For the only member of the Patellariales [*Patellaria atrata* (JGI)] and the two Mytilinidiales surveyed [*Lophium mytilinum* (CBS 269.34; JGI) and *Mytilinidion resinicola* (CBS 304.34; JGI)], the *rps3* gene is encoded within the mL2449 intron, an arrangement that is frequently observed among most members of the Pezizomycotina. For members of the Dothideales, two examples of freestanding *rps3* genes [*Delphinella strobiligena* (CBS 735.71; JGI) and *Setomelanomma holmii* (CBS 110217; JGI)] and one example of a *rnl* intron-encoded (mL2449) version [*Lineolata rhizophorae* (ATCC 16933; JGI)] was noted. Freestanding *rps3* genes were found among all surveyed members (36) of the Pleosporales. With regards to the Microthyriales and Myriangiales, only one representative mtDNA sequence (where the location

of the *rps3* gene could be determined) was available for each [*Tothia fuscella* (CBS 130266; JGI) and *Elsinoë ampelina* (CECT 20119; JGI), respectively]. In the both instances, the *rps3* gene appears to be freestanding. Three fungi that were not phylogenetically resolved at the order level were noted. The *rps3* gene is encoded in the mL2449 intron for *Lepidopterella palustris* (JGI), and freestanding in *Lizonia empirigonia* (CBS 542.76; JGI) and *Lizonia empirigonia* (CBS 542.76; JGI). A novel arrangement was observed in *Phyllosticta citriasiana* where the *rps3* gene was located within the fifth intron of *nad5* (*nad5*-i5 or *nad5*-710). The intron was a large (6372 nt) group ID intron that appears to encode two identifiable ORFs: *rps3* (1191 or 1716 nt, depending on interpretation) and a large ORF (2389 or 2574 nt, depending on interpretation) with unknown function. The group ID intron appears to still adopt the conserved secondary structure of group I introns even though it encodes for a large insertion. The two ORFs were located within the terminal loop segment of P2.1b of the group ID intron (see Figure 3.2).

3.4.3.1. The *rps3* gene is absent from mitochondrial genomes of the Capnodiales (Dothideomycetes)

The *rps3* gene appears to be absent from the mtDNA of eight (out of the nine; note: *Baudoinia panamericana* (JGI) mtDNA could not be found and thus should probably be removed from the analysis) sampled members of the Capnodiales but present *in Pseudovirgaria hyperparasitica* (CBS 121739; JGI). In the latter case, it is encoded within the mL2449 intron.

In the nuclear genome of members of the Capnodiales (where nuclear data was available), two forms of the *rps3* gene was noted; one that appears to be bacterial-like (similar to the mitochondrial-encoded version) and the other was a cytoplasmic version (typically encoded

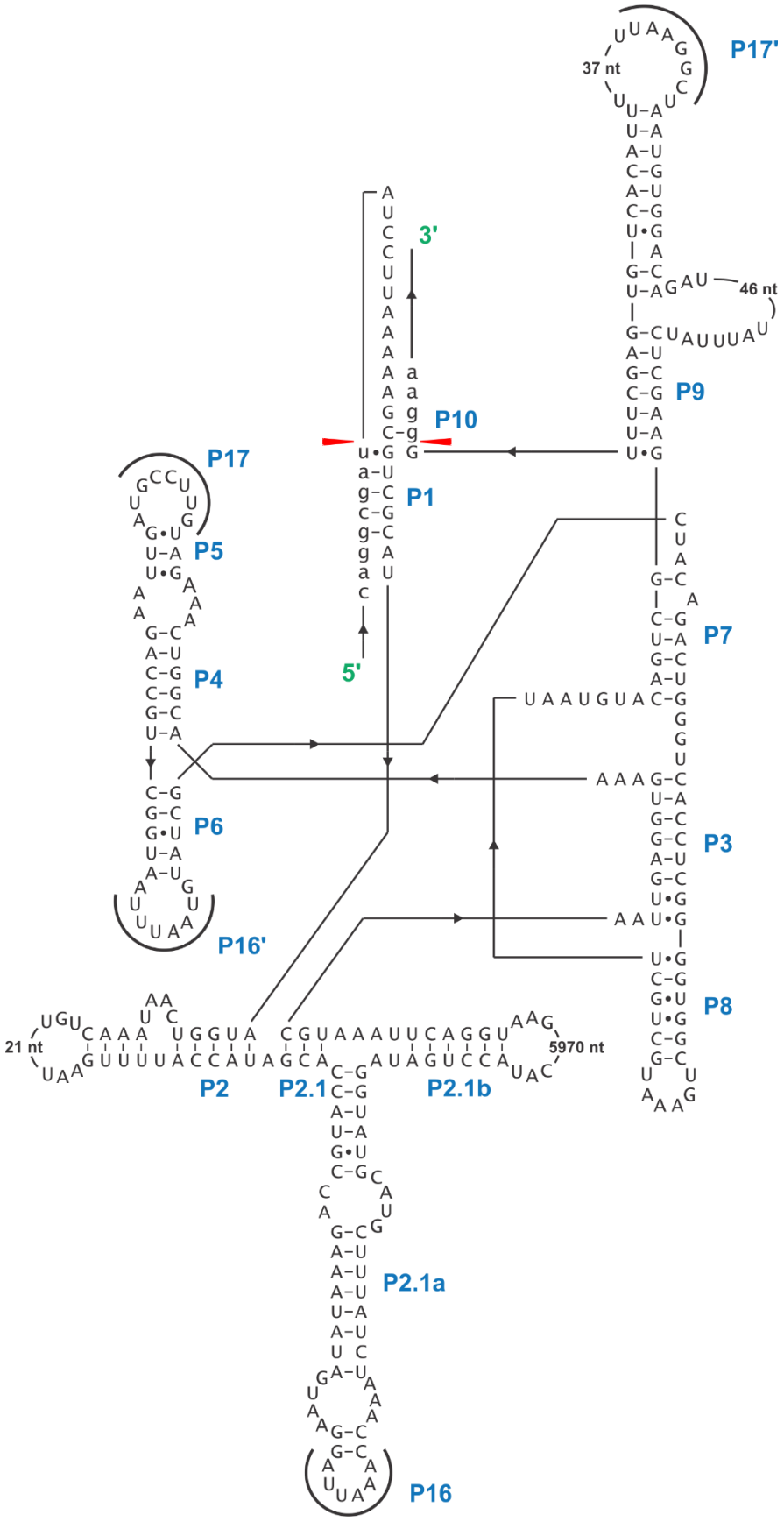

P17'
37 nt
46 nt
3'
5'
P10
P1
P9
P17
P5
P4
P7
P6
P3
P16'
P8
21 nt
5970 nt
P2
P2.1
P2.1b
P2.1a
P16

Figure 3.2. The proposed secondary structure of the *nad5*-710 intron of *Phyllosticta citriasiana* (JGI).

The 5′- and 3′-ends, and pairing interactions (P1-P10, plus P16/P16′ and P17/P17′) are indicated in green and blue, respectively. A large insertion is found within the terminal loop of P2.1b. The large insertion encodes for a *rps3* gene and an open reading frame of unknown function. Exon sequences are represented as lowercase letters. Intron sequences are represented as uppercase letters. Red arrows indicate 5′ and 3′ splice sites. Structure was manually drawn using CorelDRAW Graphics Suite X6 (Corel Corporation, Ottawa).

within the nuclear genome). Based on phylogenetic analysis, the bacterial-like *rps3* gene grouped with mitochondrial-encoded RPS3 sequences and were separate from the cytoplasmic versions (see Figure 3.3). This seems to suggest that nuclear-encoded bacterial-like *rps3* genes are more similar to mitochondrial-encoded *rps3* genes than to nuclear-encoded cytoplasmic *rps3* genes.

3.4.4. Encoding pattern of the *rps3* gene within the Leotiomycetes

Within the Leotiomycetes, freestanding *rps3* genes were noted for some members (5/17) of the Helotiales, including *Marssonina brunnea* (JN204424.1)*, Phialocephala subalpina* (Duò et al., 2012; AEI52989.1), *Acephala macrosclerotiorum* (JGI), *Loramyces macrosporus* (JGI), and *Phialocephala scopiformis* (JGI). In all cases, the *rnl* genes lacked the mL2449 intron. Other members (12/17) of this class possessed a *rnl* intron-encoded *rps3* gene.

3.4.5. Encoding pattern of the *rps3* gene within the Lecanoromycetes

Within the Lecanoromycetes, all members of the Lecanorales examined were noted to contain a freestanding version of the *rps3* gene, but for the Peltigerales, only *rnl* intron-encoded versions were noted.

3.4.6. Encoding pattern of the *rps3* gene within the Eurotiomycetes, Sordariomycetes, and Xylonomycetes

For all surveyed sequences for members of the Eurotiomycetes, Sordariomycetes, and Xylonomycetes, the *rps3* gene was noted to be encoded within mL2449 group I intron. Although it has been reported that *Aspergillus nidulans* and *Penicillium solitum* lack the *rps3* gene (Goodwin et al., 2016), our analysis agrees with Joardar et al. (2012); it appears that the *rnl* gene

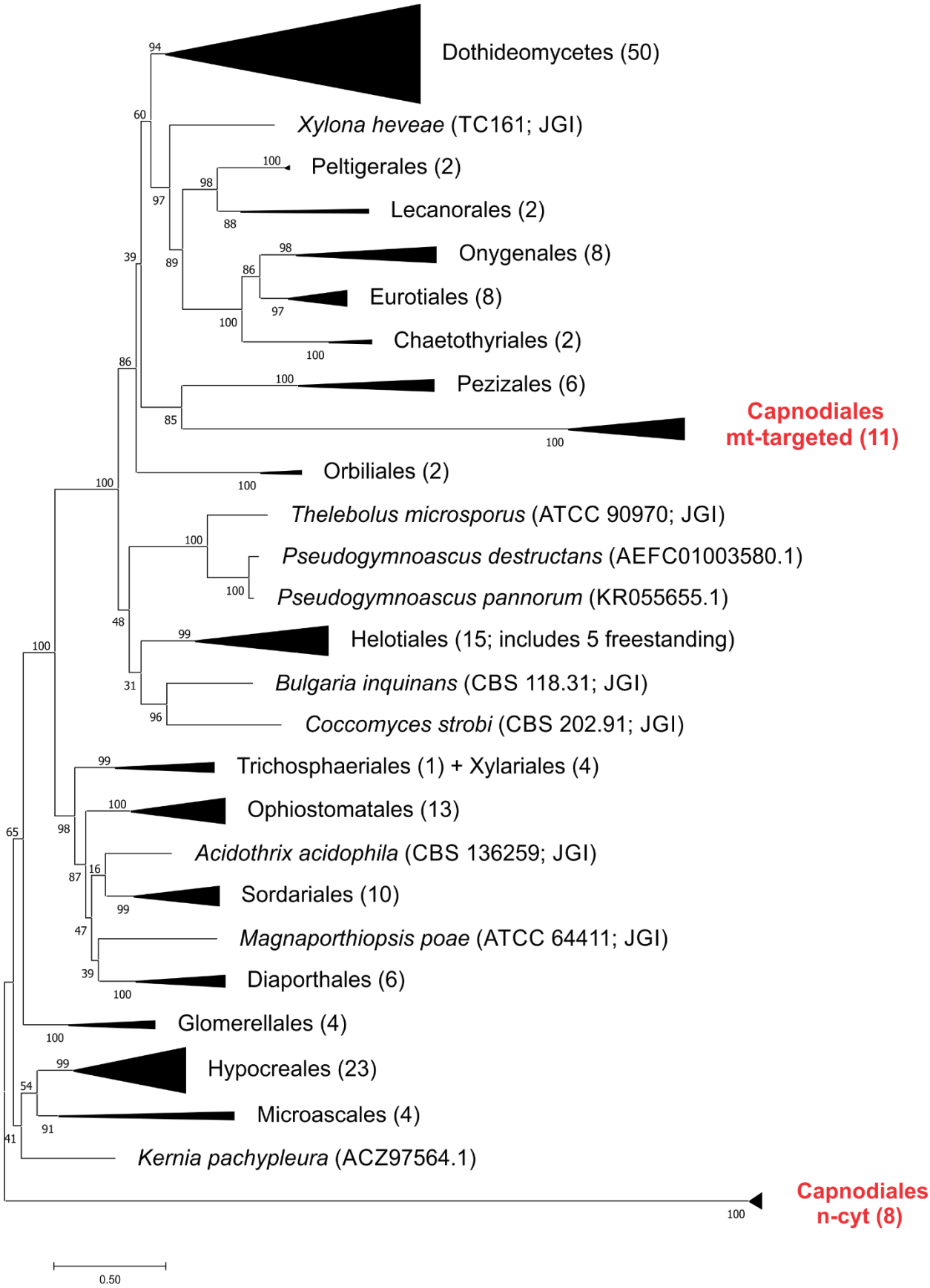

Dothideomycetes (50)
Xylona heveae (TC161; JGI)
Peltigerales (2)
Lecanorales (2)
Onygenales (8)
Eurotiales (8)
Chaetothyriales (2)
Pezizales (6)
Capnodiales mt-targeted (11)
Orbiliales (2)
Thelebolus microsporus (ATCC 90970; JGI)
Pseudogymnoascus destructans (AEFC01003580.1)
Pseudogymnoascus pannorum (KR055655.1)
Helotiales (15; includes 5 freestanding)
Bulgaria inquinans (CBS 118.31; JGI)
Coccomyces strobi (CBS 202.91; JGI)
Trichosphaeriales (1) + Xylariales (4)
Ophiostomatales (13)
Acidothrix acidophila (CBS 136259; JGI)
Sordariales (10)
Magnaporthiopsis poae (ATCC 64411; JGI)
Diaporthales (6)
Glomerellales (4)
Hypocreales (23)
Microascales (4)
Kernia pachypleura (ACZ97564.1)
Capnodiales n-cyt (8)
0.50

Figure 3.3. Phylogenetic analysis of mitochondrial- and nuclear-encoded RPS3 amino acid sequences from fungi of the order Capnodiales.

The phylogenetic tree was generated using 188 ribosomal protein S3 (RPS3) amino acid sequences by FastTree (Price et al., 2010). The order Capnodiales, is indicated in red. Nodes were collapsed based on class or order, where possible. The class or order name is indicated at the leaves of the tree. The number of representatives of each class or order are indicated within parenthesis. Fungi could not be grouped into a single collapsed node are indicated with their genus species name. The strain information, when available, is indicated within parenthesis. JGI = Joint Genome Institute (MycoCosm; source of mitochondrial sequence); mt-targeted = nuclear-encoded mitochondrial RPS3; n-cyt = nuclear cytoplasmic RPS3; Node support values = bootstrap values based on 1 000 pseudoreplicates. Scale bar = 0.5 amino acid substitutions per site.

in both cases has been misannotated. In *Aspergillus nidulans* (AFC69019.1), the *rps3* gene is intron-encoded. For *Podospora solitum* (JN696111.1), the mL2449 intron is also present but the *rps3* gene appears to be degenerated as the ORF can only be assembled by correcting frameshift mutations.

3.4.7. "Hitchhiking" *rps3* genes

Several versions of the mL2449 encoded *rps3* gene fused to double-motif LAGLIDADG type homing endonuclease sequences have been observed previously in *Cryphonectria parasitica* and in members of the Ophiostomatales (Hausner et al., 1999; Gibb and Hausner, 2005; Sethuraman et al., 2009a; Kanzi et al., 2016). In this survey, an example in *Monacrosporium haptotylum* of a double-motif LAGLIDADG coding region fused to the C-terminus coding region of a freestanding version of the *rps3* gene was noted. Freestanding versions of the *rps3* gene were also noted in *Gyromitra esculenta* (JGI) and *Morchella importuna* (JGI). With regards to the latter, the *rps3* gene appears to exist as a fusion ORF. The N-terminal component encodes a GIY-YIG type homing endonuclease gene and the C-terminal component encodes the *rps3* gene.

3.4.8. In fungal mitochondrial genomes, NAT open reading frames can be intron-encoded or freestanding

While investigating the fungal mtDNA *cytb* gene for the presence of introns and homing endonucleases (Guha et al., 2017b), a few instances (see Table 3.2) of some unusual ORFs were noted in an intron that is located at position 429 (between nucleotide 429 and 430; the numbering is with respect to the homologous position in the *Saccharomyces cerevisiae cytb* nucleotide

sequence; CP006539.1). In the mtDNA of *Annulohypoxylon stygium* (KF545917.1), two NAT ORFs were annotated by the authors; one located within the *cytb*-429 intron (AHB33527.1) and the other was freestanding (AHB33515.1) in the following gene arrangement: *cox2-NAT-tRNA-Arg-nad4L* genes. Previously, it was reported for the mtDNA of *Phialocephala subalpina* (JN031566.1; Duò et al., 2012), where a *NAT* gene was part of a somewhat similar gene arrangement: *cox2-rps3-NAT-tRNAsup-nad4L*. In this work, while searching GenBank, additional examples of mtDNA-encoded NAT sequences were noted. Mitochondrial sequences from *Cochliobolus heterostrophus* and related species were recovered that contained two NAT ORFs within the *cytb*-429 intron. Only one *cytb*-429 intron-encoded NAT ORF was observed in *Pyrenophora teres* (EFQ96341.1). A non-intron-encoded version of a NAT ORF was recovered from *Fusarium avenaceum* (KIL83566.1). The NAT ORF was situated between the *nad5* and *cytb* genes but encoded on the opposite reading frame relative to the flanking genes. In summary, various mtDNA NAT arrangements among the fungi were observed including freestanding, two copies with one encoded by the *cytb*-429 intron and the other being freestanding, and within the *cytb*-429 intron with either one or two copies of the NAT ORF. All *cytb* intron-versions NAT ORFs observed were encoded in the opposite DNA strand relative to the host *cytb* gene. Based on NCBI (blastp), the mtDNA NAT amino acid sequences contain an acetyltransferase (GNAT) domain that catalyses N-acetyltransferase reactions. They can be assigned to pfam13301 or COG1670 (RimL) protein domains (Marchler-Bauer et al., 2015).

3.4.9. Phylogenetic analysis of mitochondrial- and nuclear-encoded NATs

A phylogenetic analysis of NAT amino acid revealed that all mitochondrial-encoded NAT [with the exception a nuclear-encoded NAT from *Pestalotiopsis fici* (ETS85844.1)] formed

Table 3.2. The encoding pattern of the *NAT* gene among Pezizomycotina fungi surveyed from the National Center for Biotechnology Information.

	Location	Insertion^*	Order
***Acremonium chrysogenum* (KFH46433.1)**	n		Hypocreales
***Alternaria brassicicola* (PHFN01000012.1)**	mt	*1xoc-trnR*	Pleosporales
***Alternaria solani* (DQ209284.1)**	mt	*cytb*-429	Pleosporales
***Annulohypoxylon stygium* (AHB33515.1)**	mt	*cox2-nad4L*	Xylariales
***Annulohypoxylon stygium* (AHB33527.1)**	mt	*cytb*-429	Xylariales
***Aschersonia aleyrodis* (KZZ90792.1)**	n		Hypocreales
***Aspergillus terreus* (EAU30554.1)**	n		Eurotiales
***Beauveria bassiana* (EJP70772.1)**	n		Hypocreales
***Ceratocystis fimbriata* (PHH55288.1)**	n		Microascales
***Cladophialophora bantiana* (KIW95554.1)**	n		Chaetothyriales
***Cladophialophora carrionii* (ETI22638.1)**	n		Chaetothyriales
***Cochliobolus carbonum* (EUC26919.1)**	mt	*cytb*-429	Pleosporales
***Cochliobolus carbonum* (EUC26921.1)**	mt	*cytb*-429	Pleosporales
***Cochliobolus heterostrophus* (EMD85037.1)**	mt	*cytb*-429	Pleosporales
***Cochliobolus heterostrophus* (EMD85039.1)**	mt	*cytb*-429	Pleosporales
***Cochliobolus miyabeanus* (EUC39512.1)** [#]	mt	*cytb*-429?	Pleosporales
***Cochliobolus miyabeanus* (EUC39514.1)** [#]	mt	*cytb*-429?	Pleosporales
***Cochliobolus victoriae* (EUN20374.1)** [#]	mt	*cytb*-429?	Pleosporales
***Colletotrichum gloeosporioides* (EQB58383.1)**	n		Glomerellales
***Colletotrichum graminicola* (EFQ25412.1)**	n		Glomerellales
***Colletotrichum higginsianum* (CCF36591.1)**	n		Glomerellales
***Colletotrichum sublineola* (KDN64323.1)**	n		Glomerellales
***Coniochaeta ligniaria* (OIW31834.1)**	n		Sordariales
***Cordyceps militaris* (EGX96378.1)**	n		Hypocreales
***Daldinia* sp. (OTB17272.1)**	n		Xylariales
***Diaporthe helianthi* (POS76385.1)**	n		Diaporthales
***Drechmeria coniospora* (ODA81274.1)**	n		Hypocreales
***Emmonsia crescens* (KKZ60448.1)**	n		Onygenales
***Eutypa lata* (EMR66796.1)**	n		Xylariales
***Fusarium avenaceum* (KIL83566.1)**	mt	*btyc-5dan*	Hypocreales
***Fusarium fujikuroi* (KLO86718.1)**	n		Hypocreales
***Fusarium oxysporum* (EWY89876.1)**	n		Hypocreales
***Fusarium oxysporum* (EWY89877.1)**	n		Hypocreales

	Location	Insertion^*	Order
***Fusarium verticillioides* (EWG50349.1)**	n		Hypocreales
***Magnaporthe oryzae* (EHA49915.1)**	n		Magnaporthales
***Metarhizium acridum* (EFY86333.1)**	n		Hypocreales
***Metarhizium album* (KHN96892.1)**	n		Hypocreales
***Metarhizium anisopliae* (KFG82641.1)**	n		Hypocreales
***Metarhizium brunneum* (KID75149.1)**	n		Hypocreales
***Metarhizium majus* (KID99091.1)**	n		Hypocreales
***Microdochium bolleyi* (KXJ86251.1)**	n		Xylariales
***Nectria haematococca* (EEU44146.1)**	n		Hypocreales
***Neonectria ditissima* (KPM43351.1)**	n		Hypocreales
***Neurospora crassa* (KHE88515.1)**	n		Sordariales
***Neurospora tetrasperma* (EGZ77264.1)**	n		Sordariales
***Ophiostoma piceae* (EPE10432.1)**	n		Ophiostomatales
***Penicillium digitatum* (EKV07681.1)**	n		Eurotiales
***Penicillium rubens* (CAP94545.1)**	n		Eurotiales
***Pestalotiopsis fici* (ETS85844.1)**	n		Xylariales
***Pestalotiopsis fici* (ETS81215.1)**	n		Xylariales
***Pestalotiopsis fici* (AOW71168.1)**	mt	*Cnrt-Cnrt*	Xylariales
***Phialocephala subalpina* (AEI52990.1)**	mt	*cox2-nad4L*	Helotiales
***Podospora anserina* (CDP22703.1)**	n		Sordariales
***Pseudogymnoascus pannorum* (KFY46591.1)**	n		Leotiomycetes
***Purpureocillium lilacinum* (OAQ83156.1)**	n		Hypocreales
***Pyrenophora teres* (EFQ96341.1)**	mt	*cytb*-429	Pleosporales
***Rasamsonia emersonii* (KKA25341.1)**	n		Eurotiales
***Rosellinia necatrix* (GAP86717.2)**	n		Xylariales
***Sporothrix brasiliensis* (KIH90051.1)**	n		Ophiostomatales
***Sporothrix schenckii* (ERS97658.1)**	n		Ophiostomatales
***Stachybotrys chartarum* (KFA74112.1)**	n		Hypocreales
***Stemphylium lycopersici* (AOS52859.1)**	mt	*btyc-2xoc*	Pleosporales
***Thielaviopsis punctulata* (KKA26070.1)**	n		Microascales
***Togninia minima* (EOO00824.1)**	n		Calosphaeriales
***Torrubiella hemipterigena* (CEJ88750.1)**	n		Hypocreales
***Torrubiella hemipterigena* (CEJ91063.1)**	n		Hypocreales

	Location	Insertion^*	Order
***Trichoderma atroviride* (EHK39837.1)**	n		Hypocreales
***Trichoderma reesei* (ETS00603.1)**	n		Hypocreales
***Trichoderma virens* (EHK22913.1)**	n		Hypocreales
***Ustilaginoidea virens* (KDB16562.1)**	n		Hypocreales
***Verticillium dahliae* (EGY17399.1)**	n		Glomerellales

^ some genes are written reverse indicating they are encoded on the opposite DNA strand relative to the *rps3* gene

***** only applies to mitochondrial-encoded N-acetyltransferases

[#] only intron sequences were found but were highly similar to related sequences in other *Cochliobolus* spp.

Abbreviations used for genes: mt = mitochondrial-encoded; n = nuclear-encoded; *cox* = cytochrome oxidase subunit, *cytb* = cytochrome b; *nad* = NADH dehydrogenase subunit; *trn* = transfer ribonucleic acid (tRNA); letter beside "*trn*" = amino acid carried by the tRNA; ? = status of component could not be determined.

a single clade (see Figure 3.4, Table 3.2). The tree topology suggests that the mitochondrial versions are derived from a nuclear-encoded version. Furthermore, the *NAT* gene was transferred from the nuclear genome to the mtDNA on at least one occasion. The phylogenetic tree also suggests that in instances where two NAT ORFs were located within one intron (*cytb*-429), the two NAT copies are paralogs; NAT sequences encoded towards the 5′-end and 3′-end form their own respective clades but are connected to a common node suggesting they share a common ancestor. The freestanding and the intron-encoded versions of the *Annulohypoxylon stygium* NAT amino acid sequences failed to group together, suggesting that they may have diverged significantly since the ancestral gene sequence entered the mtDNA. In summary, intergenic versions of the *NAT* gene were recorded for members of the Helotiales (1), Hypocreales (1), Pleosporales (2), and Xylariales (2). The remaining versions of the *NAT* gene appear to be found within members of the Pleosporales and Hypocreales. These appear to be encoded within an intron of the *cytb* gene (see Table 3.2).

3.4.10. More additions to mitochondrial genomes: members of the aminotransferase superfamily

Recent reports and mtDNA sequence data have revealed another novel intron-encoded ORF that appears to be a potential nuclear transfer or horizontal transfer into the mtDNA by some yet unknown mechanism. In *Juglanconis oblonga*, a member of the Diaporthales, the *cox1*-i12 intron (group IB intron) encodes a putative aminotransferase (AAT) (KY575054.1). Another example of a freestanding version of the *AAT* gene was noted in a member of the Xylariales, *Pestalotiopsis fici* (KX870077.1; Zhang et al., 2017). Here, the putative *AAT* is located upstream of the *nad4L* gene.

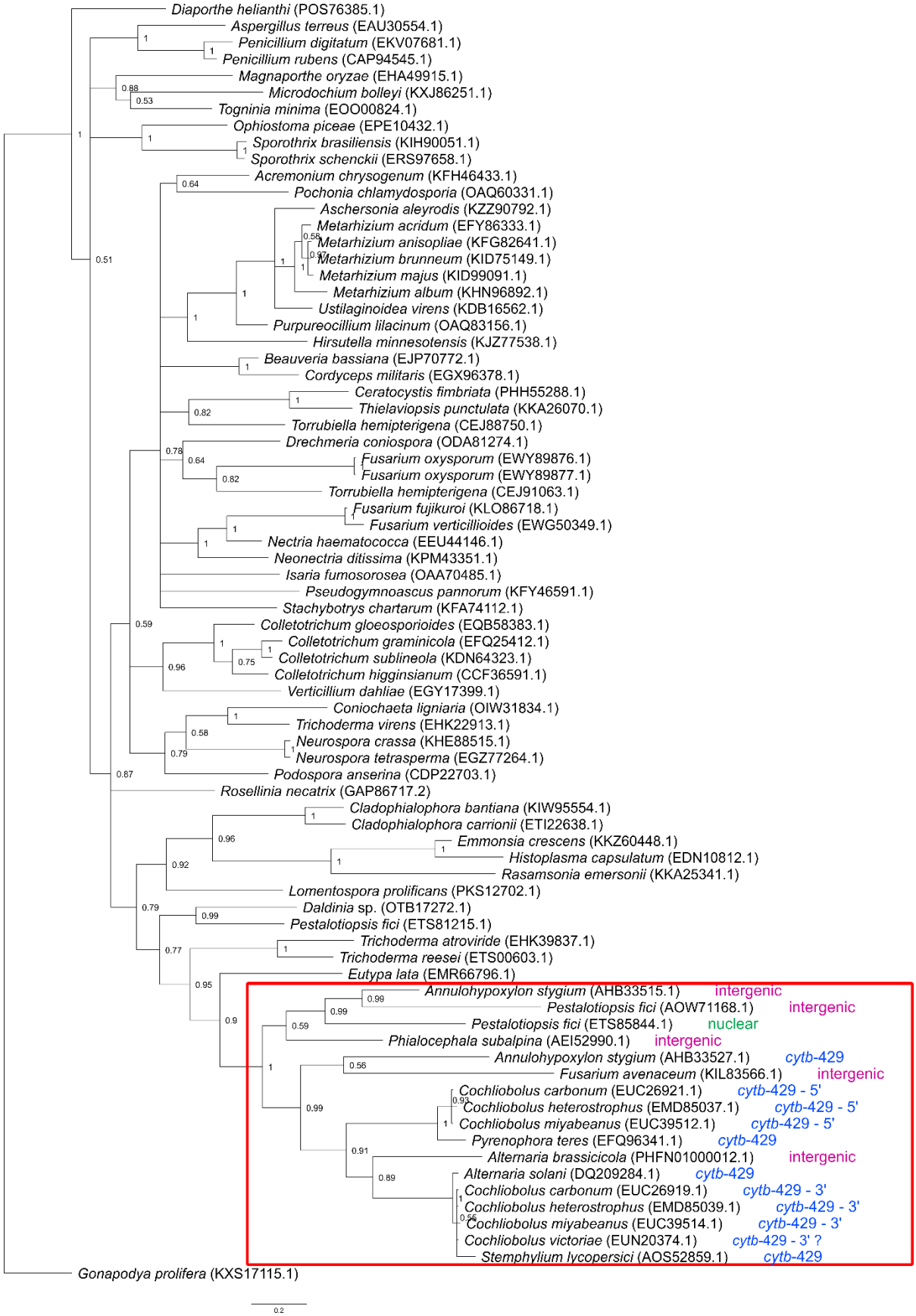

Diaporthe helianthi (POS76385.1)
Aspergillus terreus (EAU30554.1)
Penicillium digitatum (EKV07681.1)
Penicillium rubens (CAP94545.1)
Magnaporthe oryzae (EHA49915.1)
Microdochium bolleyi (KXJ86251.1)
Togninia minima (EOO00824.1)
Ophiostoma piceae (EPE10432.1)
Sporothrix brasiliensis (KIH90051.1)
Sporothrix schenckii (ERS97658.1)
Acremonium chrysogenum (KFH46433.1)
Pochonia chlamydosporia (OAQ60331.1)
Aschersonia aleyrodis (KZZ90792.1)
Metarhizium acridum (EFY86333.1)
Metarhizium anisopliae (KFG82641.1)
Metarhizium brunneum (KID75149.1)
Metarhizium majus (KID99091.1)
Metarhizium album (KHN96892.1)
Ustilaginoidea virens (KDB16562.1)
Purpureocillium lilacinum (OAQ83156.1)
Hirsutella minnesotensis (KJZ77538.1)
Beauveria bassiana (EJP70772.1)
Cordyceps militaris (EGX96378.1)
Ceratocystis fimbriata (PHH55288.1)
Thielaviopsis punctulata (KKA26070.1)
Torrubiella hemipterigena (CEJ88750.1)
Drechmeria coniospora (ODA81274.1)
Fusarium oxysporum (EWY89876.1)
Fusarium oxysporum (EWY89877.1)
Torrubiella hemipterigena (CEJ91063.1)
Fusarium fujikuroi (KLO86718.1)
Fusarium verticillioides (EWG50349.1)
Nectria haematococca (EEU44146.1)
Neonectria ditissima (KPM43351.1)
Isaria fumosorosea (OAA70485.1)
Pseudogymnoascus pannorum (KFY46591.1)
Stachybotrys chartarum (KFA74112.1)
Colletotrichum gloeosporioides (EQB58383.1)
Colletotrichum graminicola (EFQ25412.1)
Colletotrichum sublineola (KDN64323.1)
Colletotrichum higginsianum (CCF36591.1)
Verticillium dahliae (EGY17399.1)
Coniochaeta ligniaria (OIW31834.1)
Trichoderma virens (EHK22913.1)
Neurospora crassa (KHE88515.1)
Neurospora tetrasperma (EGZ77264.1)
Podospora anserina (CDP22703.1)
Rosellinia necatrix (GAP86717.2)
Cladophialophora bantiana (KIW95554.1)
Cladophialophora carrionii (ETI22638.1)
Emmonsia crescens (KKZ60448.1)
Histoplasma capsulatum (EDN10812.1)
Rasamsonia emersonii (KKA25341.1)
Lomentospora prolificans (PKS12702.1)
Daldinia sp. (OTB17272.1)
Pestalotiopsis fici (ETS81215.1)
Trichoderma atroviride (EHK39837.1)
Trichoderma reesei (ETS00603.1)
Eutypa lata (EMR66796.1)
Annulohypoxylon stygium (AHB33515.1) intergenic
Pestalotiopsis fici (AOW71168.1) intergenic
Pestalotiopsis fici (ETS85844.1) nuclear
Phialocephala subalpina (AEI52990.1) intergenic
Annulohypoxylon stygium (AHB33527.1) cytb-429
Fusarium avenaceum (KIL83566.1) intergenic
Cochliobolus carbonum (EUC26921.1) cytb-429 - 5'
Cochliobolus heterostrophus (EMD85037.1) cytb-429 - 5'
Cochliobolus miyabeanus (EUC39512.1) cytb-429 - 5'
Pyrenophora teres (EFQ96341.1) cytb-429
Alternaria brassicicola (PHFN01000012.1) intergenic
Alternaria solani (DQ209284.1) cytb-429
Cochliobolus carbonum (EUC26919.1) cytb-429 - 3'
Cochliobolus heterostrophus (EMD85039.1) cytb-429 - 3'
Cochliobolus miyabeanus (EUC39514.1) cytb-429 - 3'
Cochliobolus victoriae (EUN20374.1) cytb-429 - 3' ?
Stemphylium lycopersici (AOS52859.1) cytb-429
Gonapodya prolifera (KXS17115.1)
0.2

Figure 3.4. Phylogenetic analysis of mitochondrial-encoded NAT amino acid sequences from fungi of the subphylum Pezizomycotina.

The phylogenetic tree was generated using 77 N-acetyltransferase (NAT) amino acid sequences by Bayesian inference (MrBayes 3.2.6; Ronquist et al., 2012). The clade formed by mitochondrial-encoded NATs [excluding *Pestalotiopsis fici* (ETS85844.1)] is outlined with a red box. The location where the NAT open reading frame is encoded is indicated after the organism name (and after the identifier). intergenic = *NAT* gene is encoded within an intergenic region; nuclear = *NAT* gene is encoded within the nuclear genome; *cytb*-429 = *NAT* gene is encoded within a group ID intron of the cytochrome b gene (*cytb*) at position 429 (nomenclature based on Johansen and Haugen, 2001); 5′ and 3′ = (only applies to *cytb*-429 introns that encode two *NAT* genes) indicates the relative position of the NAT ORF (i.e. closer to the 5′- or 3′-end of the *cytb* gene); node support values = posterior probability based on 5 000 000 generations; scale bar = 0.2 amino acid substitutions per site.

3.5. Discussion

3.5.1. The position of the *rps3* gene among the Pezizomycotina

Based on previous studies (Sethuraman et al., 2009b), the following conclusion could be drawn. The *rps3* gene is freestanding among members of the Saccharomycotina and Taphrinomycotina, but among most members of the Pezizomycotina, (filamentous ascomycetes) the *rps3* gene tends to be encoded within a group IA intron located at position mL2449 in the large ribosomal subunit (*rnl*) gene (Bullerwell et al., 2003). Previously, a freestanding mtDNA-encoded *rps3* gene was noted in *Phaeosphaeria nodorum*, suggesting a rare event whereby the *rps3* gene was relocated into an intergenic region (Sethuraman et al., 2009b). In this study, all members of the Pleosporales that were surveyed also have freestanding *rps3* genes. It was also noted that the Pezizales, which are considered to be early branching filamentous Ascomycota fungi, also encode for freestanding versions.

In general, the greatest diversity of *rps3* arrangements was observed in members of the Dothideomycetes (see Figure 3.1). It can be intron-encoded (within intron mL2449), freestanding (i.e. encoded within intergenic regions), or absent from the mtDNA. Furthermore, among the the Botryosphaeriales, two different *rps3* gene encoding patterns were observed. The first was in a group IA intron of the mitochondrial *rnl* gene (i.e. within intron mL2449). The other encoding pattern was within a group ID intron of the *nad5* gene (*nad5*-710). The latter configuration was observed in the *Phyllosticta* spp. surveyed (see Table 3.1). The mechanism of how a group ID intron in a *nad5* gene could have acquired a *rps3* gene is still unclear.

This study also uncovered additional examples that suggest that the *rps3* gene relocated from the mL2449 intron to become freestanding in some members of the Helotiales and

Lecanorales. In some instances, the relocation of the *rps3* gene appears to be followed by the loss of the mL2449 intron that houses this gene in other members of the Ascomycota (see Table 3.1).

It was previously shown that the intron-encoded RPS3 is an essential protein in *Neurospora crassa* and it is incorporated in to the small subunit of the ribosome (LaPolla and Lambowitz, 1981). It is noteworthy that the *Neurospora crassa* mL2449 intron has been extensively characterized by various groups. It does not appear to readily splice and requires various nuclear-encoded factors in order to be efficiently removed from the rRNA precursor (Akins and Lambowitz, 1987; Guo et al., 1991; reviewed by Hausner, 2012). This suggests that maintaining the mL2449 intron comes at a cost to the host cell. Thus, loss of the *rps3* gene would not warrant the maintenance of the intron, which itself will be lost due to lack of selection.

3.5.2. The case of the lost mtDNA *rps3* gene in some members of the Capnodiales

Fungi of the Capnodiales do not encode for a *rps3* gene within the mtDNA. Instead, they appear to encode two different copies within the nuclear genome; one that is bacterial like (and similar to the mitochondrial-encoded version) and a cytoplasmic version (typically nuclear-encoded). At this stage, it can only be speculated that these bacterial-like *rps3* genes are nuclear counterparts that could substitute for "lost" mtDNA *rps3* genes. There are several possible scenarios of how the bacterial-like *rps3* genes may have arisen: 1) these genes represent the mtDNA *rps3* gene that was transferred from the organellar genome to the nuclear genome or 2) the mitochondrial version of *rps3* was lost but a nuclear-encoded *rps3*-like gene was coopted to express the mitochondrial RPS3 protein.

3.5.3. Speculations on the formation of freestanding *rps3* genes

The question remains: How did freestanding versions of the *rps3* gene get established? Usually, to explain relocation of genes or rearrangements within a genome, one invokes homologous or non-homologous recombination along with gene conversion events, usually stimulated by the presence of repeats or regions prone to single- or double-stranded breaks (Gross et al., 1984; Zinn et al., 1988; Aguileta et al., 2014; Wu and Hao, 2014; Wu et al., 2015). It is known that fungal mtDNAs are AT-rich (Wolf and Del Giudice, 1988). It was noted that both *NAT* and *rps3* gene sequences are flanked by AT-rich sequences. These may have the same effect as repeats increasing the probability of illegitimate recombination events. Reverse transcriptase (RT) activity may be another explanation for a mechanism that could mediate the relocation of the *rps3* gene. Sources for mitochondrial RT activity could be group II introns or retroplasmids (Hausner, 2012). However, this would require ectopic retrohoming, which one would assume is a rare event. Retrohoming is a phenomenon where a RT binds to a non-cognate transcript and generates a cDNA that is inserted into a new location.

It is also worth noting that RTs can transfer from one transcript to another during cDNA synthesis, a phenomenon referred to as template switching (Chiang et al., 1994; Mohr et al., 2013b). Template switching can lead to the fusion of different transcripts or transcript segments, forming chimeric RNAs (fusion transcripts). RT activity and template switching may explain the freestanding *Phaeosphaeria nodorum rps3* gene that is fused with a short *cox1* segment (Sethuraman et al., 2009b). In this case, one could propose a scenario during reverse transcription of a *rps3* transcript, where the RT switched onto a *cox1* transcript. The result was the formation of a *cox1-rps3* chimeric cDNA, which could then insert into a new location.

Another possibility is based on the observation that in some instances, an intron-encoded RPS3 ORF is fused with a LAGLIDADG type homing endonuclease (Hausner et al., 1999;

Sethuraman et al., 2009b). A scenario that could generate such fusion may be that the RPS3 ORF "hitchhiked" during a HEG ectopic integration event. Following integration into a new site, the mL2449 intron was lost and the HEG that was attached to the *rps3* gene would degrade over time due to lack of selective pressures (Goddard and Burt, 1999).

3.5.4. The *NAT* gene is a recent arrival from the nuclear genome

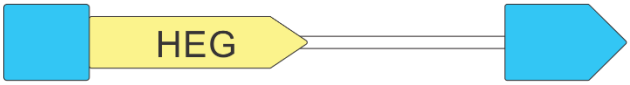
Among the filamentous fungi (Pezizomycotina), the presence of *NAT* genes in the mtDNA was only noted in a few instances (twelve, excluding fungi encoding two copies in the mtDNA). Furthermore, the distribution appears to be sporadic (i.e. it can be found in members of the Pleosporales, Helotiales, Hypocreales, and Xylariales). The uncommon occurrence and sporadic distribution of *NAT* genes among the Pezizomycotina suggests that they transferred into fungal mtDNAs potentially several times among various lineages.

Transfer of DNA from the mitochondrial into the nuclear genome is a process that has been assumed to have occurred since the beginning of the origin of mitochondria (Lang et al., 1999; Gray, 2015). Among plants, there is evidence of DNA transfer between mitochondria and chloroplast, and potentially even nuclear DNA (Thorsness and Fox, 1990; Thorsness and Weber, 1996; Hazkani-Covo et al., 2010). However, with regards to functional gene transfer, evidence points toward a unidirectional transfer [i.e. from organelle to nucleus (Thorsness and Fox, 1990; Palmer et al., 2000) or from endosymbiont to organelles (Bilewitch and Degan, 2011)]. Recent evidence suggests that mitochondria have the ability to take up DNA and RNA from the cytoplasm but the translocation pathways through the mitochondrial outer and inner membranes are still elusive (Koulintchenko et al., 2003; Niazi et al., 2013; Weber-Lofi et al., 2015). It has been hypothesized that genes may be translocated from one compartment into another via RNA

intermediates. These would provide an avenue of removing introns and promoter sequences that may not be operational in the organellar genomic context (Palmer et al., 2000; Niazi et al., 2013). This pathway would require a RT step and recombination into the genome. Moreover, if the gene was to gain a function, it would need to acquire compatible codon usage, promoter, ribosome binding site, etc. Group II intron-encoded RTs could be potential agents that can promote the incorporation of foreign DNA into the mtDNA. It is currently not known if the mitochondrial-encoded NAT proteins are functional but the open reading frames appear to be preserved based on homology to the nuclear-encoded counterparts. With regards to expression, as mtDNAs tend to encode polycistronic messages (Kennell and Lambowitz, 1989), the freestanding *NAT* genes could be transcribed as part of a larger transcript unit. Intron-encoded *NAT* genes, on the other hand, would have to be transcribed from a different promoter than the host gene because they are encoded on opposite strands.

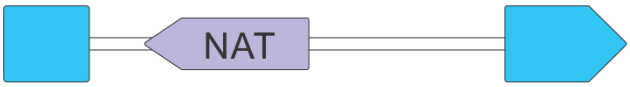
Group I intron-encoded *NAT* genes might benefit from being in this position as they are part of a potential mobile element that could spread vertically or horizontally within and between populations, and potentially across species boundaries (Hausner, 2012; Wu et al., 2015). If these *NAT* genes are functional, they may benefit the intron (host) as natural selection may now preserve them. This is in contrast to introns that carry no ORFs that may benefit the host (ex. promote splicing or dispersal of the intron) and thus may be more susceptible to loss. Usually, the *cytb*-429 group ID intron, when present, encodes a LAGLIDADG type ORF (Yin et al., 2012). See Figure 3.5 for intron configurations that have been noted within the Pezizomycotina. Maintenance of the *cytb*-429 intron is of some relevance with regards to plant pathogenic fungi. A mutation upstream of the potential intron insertion site for the *cytb*-429 intron that results in a glycine to alanine substitution at the amino acid position 143 (G143A) appears to be linked to

Xeromyces bisporus
(HG983520.1)

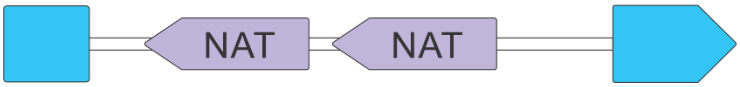


Chaetomium thermophilum
(JN007486.1)

Monilia yunnanensis
(HQ908793.1)



Annulohypoxylon stygium
(KF545917.1)



Cochliobolus heterostrophus
(KB445594.1)

Figure 3.5. Encoding patterns noted in the *cytb*-429 intron of fungi of the Pezizomycotina. Schematic representation of the encoding patterns of the intron inserted at position 429 of the cytochrome b gene (*cytb*-429; nomenclature based on Johansen and Haugen, 2001). When present, the intron encoded at *cytb*-429 is typically a group ID intron. The intron may encode a homing endonuclease gene (HEG), the product of which is typically a double-motif LAGLIDADG type homing endonuclease. In a few instances, the intron may encode for two HEGs. In this survey, additional configurations were observed, namely, an intron which encodes for either one or two copies of an N-acetyltransferase (*NAT*) open reading frame (ORF). When present, the NAT ORF is encoded on the opposite strand of the host gene. Genes are depicted as arrows to indicate the direction of the reading frame, from 5′ to 3′. A fungus with an example of each configuration is indicated next to their respective configuration.

resistance against QoI fungicides (Grasso et al., 2006). The presence of the group ID intron appears to prevent the occurrence of the mutation (and thus QoI fungicide resistance). It was suggested that the mutation interferes with proper splicing of the intron (Perea and Jacq, 1985).

3.6. Conclusions

The *NAT*, *rps3*, and the recently described mtDNA-encoded *AAT* genes may offer some insights into genes that are not necessarily part of the "core" set of fungal mtDNA genes. *NAT*, *rps3*, and *AAT* can be either freestanding, intron-encoded, or absent from the mtDNA. *NAT* and *AAT* are most likely "foreign" sequences that were imported from the nuclear genome. The *rps3* gene is most likely native to the mtDNA but can be lost, likely due to the presence of a nuclear-encoded analog or homolog that can replace the mitochondrial-encode *rps3* gene. Mechanisms that underlie the relocation of genes between introns and intergenic regions, or between different compartments are at this point elusive. They could involve RNA intermediates and the activity of reverse transcriptases, or recombination events driven by the AT-rich nature of the mtDNA (and thus, generating many small segments of similar, short AT-rich sequences). Most mtDNA annotations focus on the "core" set of genes, but genes such as *NAT*, *rps3*, and *AAT* may potentially have important roles, including contributions to the fitness of the fungus. Thus, these genes deserve to be carefully annotated and characterized.

Chapter 4: Conclusions and future directions

The mitochondrial genome (mtDNA) of fungi is highly variable in size. Currently (as of August 2018), the smallest mtDNA (excluding obligate parasitic fungi, ex. *Rozella allomycis*, which is 12 055 bp; GenBank accession number: KC702881.1; James et al., 2013) thus far belongs to the wine yeast, *Hanseniaspora uvarum* (GenBank accession number: DQ058142.1; Pramateftaki et al., 2006), sized at 18 844 bp, and the largest belongs to the potato pathogen, *Rhizoctonia solani* (GenBank accession number: KC352446.1; Losada et al., 2014), at 235 849 bp. One explanation for the cause of such size variation can be attributed to the presence of a large number of introns (mobile elements), which is observed in some of the largest fungal mtDNAs sequenced to date (Férandon et al., 2013; Mardanov et al., 2014; Salavirta et al., 2014; Li et al., 2015; Kanzi et al., 2016; Nadimi et al., 2016). Polymorphism of mtDNA size as a result of introns can be observed down to the species level (Al-Reedy et al., 2012; Joardar et al., 2012; Beaudet et al., 2013; Kanzi et al., 2016; Liang et al., 2017; Ruan et al., 2017; de Queiroz et al., 2018) and strain, as seen in *Saccharomyces cerevisiae* (Wolters et al., 2015). Another potential explanation is due to variability in gene content. In some species of yeast, the set of NADH dehydrogenase subunit genes (*nad1-6*, including *nad4L*) typically encoded within the mtDNA is absent (reviewed by de Zamaroczy and Bernardi, 1986; Sankoff et al., 1992; Zivanovic et al., 2005). Members of the Pleosporales appear to lack two ATP synthase subunit genes (namely, *atp8* and *atp9*) within their genome [*Didymella pinodes* (GenBank accession number: KT946597.1); *Parastagonospora nodorum* (GenBank accession number: EU053989.1; Hane et al., 2007); *Shiraia bambusicola* (GenBank accession number: KM382246.1; Shen et al., 2015)]. *In silico* analysis suggests that they may potentially be nuclear-encoded and imported into the

mitochondria (Franco et al., 2017). Similar observations were noted when only one ATP synthase subunit gene (*atp9*) was missing from the mtDNA (Lavín et al., 2008; Déquard-Chablat et al., 2011).

The diversity of mobile elements (intron composition varies even among closely related fungi) and the presence/absence of several, what are regarded as conserved genes, makes these two aspects of fungal mtDNAs interesting candidates to examine in order to study their overall effects on mtDNA variability and dynamics.

4.1. The intron landscape of the cytochrome b gene

In Chapter 2, potential sites of intron insertion within the cytochrome b (*cytb*) gene were examined. The result was an overview of intron distribution along the gene, which we referred to as an "intron landscape". Intron landscapes have been examined previously, for example with the mitochondrial-encoded small ribosomal subunit gene (Hafez et al., 2013), cytochrome c oxidase subunit 1 gene (not extensively; Férandon et al., 2010), and the *cytb* gene (Yin et al., 2012).

The *cytb* gene was further explored with an emphasis on members of the subphylum Pezizomycotina (phylum Ascomycota). The study by Yin et al. (2012) examined a total of 172 introns from 69 different fungal species spread across the subkingdom Dikarya (includes the phylums Ascomycota and Basidiomycota), whereas our study examined 362 introns from 129 fungi, all of which are members of the Pezizomycotina. Yin et al. (2012) noted 35 intron insertion sites whereas we noted 21.

Although there were differences among the two studies, both seem to suggest that intron insertion appears, in some aspects, to be sporadic or "patchy". Intron insertions at certain positions appear to be rare whereas a few were noted to be highly conserved. There were no

notable trends to suggest a potential cause of this observation. One may speculate that certain insertions may be beneficial or the mobile composite element may be highly efficient at insertion/invasion. Group I introns were most frequently noted (96.69% of all introns observed). Group II introns were much less frequently observed (2.76%). Some introns were unidentified in this study. Typically, the most encountered intron-encoded open reading frame (ORF) for group I introns were of the LAGLIDADG homing endonuclease family (66%). GIY-YIG homing endonucleases (encountered in 30.57% of group I introns) were typically noted at one insertion site (*cytb*-393) although they may appear in other insertion sites and in a configuration that suggests it may be interrupting the "resident" ORF. The term "resident" was based on a majority rules criterion, where the most common ORF/intron is presumed to have populated the insertion site initially. Group II introns typically encoded reverse transcriptases although instances of group II introns without an associated ORF were also noted.

Aside from the typical one intron, one ORF configuration, we also noted intron configurations from introns with no ORFs to introns showing complex arrangements (ex. the *cytb*-506 intron of *Chaetomium thermophilum*). Based on our compiled results, we generated a model that builds upon the preexisting model by Goddard and Burt (1999), mainly incorporating the concept of core creep (Edgell et al., 2011) and extension of the intron-encoded ORF such that it becomes in-phase with the upstream exon. The fusion can potentially allow for more efficient expression of the intron-encoded ORF as it can use the start codon (and promoter elements) of its host gene.

Two novel intron arrangements were noted in this study namely, the *cytb*-289 intron of *Chrysoporthe austroafricana* and the *cytb*-429 intron of *Annulohypoxylon stygium* (expanded upon in Chapter 3). The former was speculated to undergo alternative splicing to permit

expression of the host gene and the intron-encoded ORF. The latter was a group I intron that encodes an N-acetyltransferase, a gene that is typically nuclear-encoded. This anomaly, in part, led to some of the work performed in Chapter 3.

Overall, by taking a single gene approach (using currently available bioinformatics resources), we can gain insight into the evolutionary dynamics of introns and intron-encoded ORFs and discover novel intron configurations. Introns and their associated ORFs appear to be highly adaptable allowing them to spread and persist within the mtDNA. Their capacity to mobilize within and potentially between genomes may explain the high variability in mtDNA sizes observed in fungi. Intron landscapes can serve as a resource for genome annotation, which can be difficult, especially with regards to large/complex genomes. Annotation programs are available but still requires manual curation and can be fairly involved, requiring time and resources. See Haridas et al. (2018) for an example of the genome annotation process.

4.2. Mitochondrial-encoded N-acetyltransferase and ribosomal protein S3

In Chapter 3, both the encoding patterns of N-acetyltransferase and ribosomal protein S3 were examined. Initial studies (Duò et al., 2012, Zhang et al., 2017) found that fungal mtDNAs can encode for N-acetyltransferases (*NAT*), which are typically nuclear-encoded. Furthermore, the ribosomal protein S3 (*rps3*) gene, a highly conserved gene within fungal mtDNAs, appears to be missing in members of the Capnodiales (Torriani et al., 2008; Goodwin et al., 2016). Thus, the objective of Chapter 3 was to examine the presence/absence of these two genes within the mtDNAs of members of the Pezizomycotina in order to gain insight on their effects with regards to genome architecture.

The *NAT* gene is typically not mitochondrial-encoded which made its first documentations (Duò et al., 2012) unique and interesting. In order to determine its prevalence within fungal (subphylum Pezizomycotina) mtDNAs and its relationship with nuclear-encoded *NAT* genes, a comparative approach was taken, involving the use of data mining techniques and phylogenetic analysis. A search was conducted using BLAST against the non-redundant protein sequences database on National Center for Biotechnology Information (NCBI) in order to uncover additional examples of mitochondrial-encoded *NAT* genes. Only a few sequences (16) were found, of which, many (7) were noted in *Cochliobolus* spp. Interestingly, there appears to be two copies of the NAT gene encoded within a *cytb* gene intron in *Cochliobolus* spp. Additionally, all other instances (3) of intron-encoded NAT genes were encoded within the same *cytb* gene intron, namely *cytb*-429, on the opposite strand to the host gene. Furthermore, the intron insertion site (*cytb*-429) is commonly associated with antibiotic resistance (ex. Grasso et al., 2006; Banno et al., 2009; Luo et al., 2010). In other instances (6), the *NAT* gene was noted to be freestanding within the mtDNA. Phylogenetic analysis of mitochondrial- and nuclear-encoded NAT protein sequences suggests that the two genes are different from each other. The mitochondrial-encoded *NAT* gene was noted at a much lower frequency. Thus, one scenario may be that the mitochondrial-encoded NAT gene may have arisen from a transfer event from the nuclear genome and has evolved separate from nuclear-encoded *NAT* genes.

A survey of currently available mtDNAs of members of the Pezizomycotina suggests that the *rps3* gene is typically intron-encoded among the later-branching members (class Eurotiomycetes and later) but freestanding in early-branching members (includes the classes Pezizomycetes and Dothideomycetes). In the order Capnodiales (within the class Dothideomycetes), the *rps3* gene appeared to be absent from the mtDNA. Searching nuclear

scaffolds of members of the Capnodiales, two versions of the *rps3* gene could be identified. Phylogenetic analysis of the two versions, together with mitochondrial-encoded RPS3's suggest that one version may be the mitochondrial counterpart but was transferred to the nucleus for this order of fungi.

4.3. Mitochondrial genomes of fungi are diverse: contributions by introns and genes

Based on work performed in Chapters 2 and 3, introns and genes can be major contributors to mtDNA diversity. In particular presence/absence of introns can cause significant size variability of mtDNAs. This is observed to a lesser degree with regards to presence/absence of *NAT* and *rps3* gene. Introns can be versatile accommodating various intron configurations. Two novel intron configurations observed in the Master's work include an intron capable of alternative splicing (i.e. *cytb*-289 intron of *Chrysoporthe austroafricana*) and an intron that can encode for an *NAT* gene (i.e. *cytb*-429). Both *NAT* and *rps3* genes can be either mitochondrial- or nuclear-encoded. Mitochondrial-encoded *NAT* and *rps3* genes can be either freestanding or intron-encoded. Overall, it appears that the mtDNA is dynamic, capable of undergoing expansion and contraction from intron mobility and transfer of genes between itself and the nuclear genome.

4.4. Potential future work to consider

Much of the work presented was performed *in silico*. Thus, additional *in silico* analysis and experimental work is required to further support/confirm some of the observations noted. The following is a list of potential work/example questions to pursue in the future.

In Chapter 2, the *cytb*-289 intron of *Chrysoporthe austroafricana* was speculated to undergo alternative splicing. Can an alternative splicing assay be performed to confirm this? If the intron does undergo alternative splicing what are the products and at what ratio do they occur at?

In Chapter 2, the *cytb*-506 intron of *Chaetomium thermophilum* was modeled to be a nested intron, whereby a double-motif LAGLIDADG homing endonuclease ORF encoded within a group IB4 intron is interrupted by a group IIA1 intron. Is the removal of the group IIA1 intron required prior to splicing of the group IB4 intron (i.e. it is a twintron)? Is expression of the intron-encoded ORF (i.e. LAGLIDADG homing endonuclease) required for intron splicing?

In Chapter 3, a number of *NAT* genes were noted to be mitochondrial-encoded but whether they are expressed or active/functional has yet been elucidated. Thus, some potential follow-ups to this work is to examine if the genes are expressed. If they are, what are the levels of expression and are they active? If they are active, what are their target(s)? By examining these aspects, we may gain some insight as to the reason why some *NAT* genes are mitochondrial-encoded.

In Chapter 3, members of the Capnodiales are missing a mitochondrial-encoded *rps3* gene but were noted to possess two nuclear-encoded versions of the gene. Additional *in silico* analyses can be performed to examine whether the mitochondrial-related nuclear-encoded version have any features that suggest it may be targeted to the mitochondria, which one would speculate is the case given the essential role(s) of the RPS3 protein. This can be further supported by experimental work (potentially by adding a fluorescent tag to the protein and observing if it localizes to the mitochondria; although a potential problem with this approach is that the fluorescent tag may interfere with mitochondrial import).

In general, it is not well understood as to the expression levels of intron-encoded ORFs relative to each other. Transcriptomics/proteomics may be methods that can provide insight with regards to this aspect. Some follow-up questions that may be considered include: How do expression levels of intron-encoded ORFs inserted at the same insertion sites compare? Does the total number of introns-encoded ORFs affect their expression? In other words, if a mtDNA contains numerous intron-encoded ORFs, does this affect the expression of each ORF (as one would presume that expression of each ORF would be energy/resource demanding)? Transcriptomics and proteomics may also provide insight as to whether some intron-encoded ORFs are transcribed as fusion transcripts and whether these fusion transcripts are processed or are expressed as fusion proteins. This can help in supporting/refining the model of the lifecycle of composite elements presented in Chapter 2. One would speculate that at least some of the intron-encoded ORFs are transcribed as a fusion transcript. This is based on observations within many *cytb* introns where the intron-encoded ORF was in phase with the upstream exon.

References

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.

Adams PL, Stahley MR, Gill ML, Kosek AB, Wang J, Strobel SA. 2004a. Crystal structure of a group I intron splicing intermediate. RNA. 10(12):1867-1887.

Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. 2004b. Crystal structure of a self-splicing group I intron with both exons. Nature. 430(6995):45-50.

Aguileta G, De Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, Gabaldón T. 2014. High variability of mitochondrial gene order among fungi. Genome Biol Evol. 6(2):451-465. doi: 10.1093/gbe/evu028.

Akins RA, Lambowitz AM. 1987. A protein required for splicing group I introns in *Neurospora* mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell. 1987 Jul 31;50(3):331-345.

Al-Reedy RM, Malireddy R, Dillman CB, Kennell JC. 2012. Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. Fungal Genet Biol. 49(1):2-14. doi: 10.1016/j.fgb.2011.11.008.

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

Andersson SG, Karlberg O, Canbäck B, Kurland CG. 2003. On the origin of mitochondria: a genomics perspective. Philos Trans R Soc Lond B Biol Sci. 358(1429):165-177; discussion 177-179.

Arnesen T, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, Varhaug JE, Vandekerckhove J, Lillehaug JR, Sherman F, Gevaert K. 2009. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. Proc Natl Acad Sci U S A. 106(20):8157-8162. doi: 10.1073/pnas.0901931106.

Arrigoni R, Vacherie B, Benzoni F, Barbe V. 2016. The complete mitochondrial genome of *Acanthastrea maxima* (Cnidaria, Scleractinia, Lobophylliidae). Mitochondrial DNA 27(2):927-928. doi: 10.3109/19401736.2014.926489.

Avila-Adame C. 2014. Transmission of the G143A QoI-resistance point mutation through anastomosis in *Magnaporthe grisea*. Pest Manag Sci. 70(12):1918-1923. doi: 10.1002/ps.3758.

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. Fungal Genet Biol. 48(8):775-783. doi: 10.1016/j.fgb.2011.05.006.

Banno S, Yamashita K, Fukumori F, Okada K, Uekusa H, Takagaki M, Kimura M, Fujimura M. 2009. Characterization of QoI resistance in *Botrytis cinerea* and identification of two types of mitochondrial cytochrome b gene. Plant Pathology. 58(1):120-129.

Barzel A, Obolski U, Goquarten JP, Kupiec M, Hadany L. 2011. Home and away- the evolutionary dynamics of homing endonucleases. BMC Evol Biol. 11:324. doi: 10.1186/1471-2148-11-324.

Beagley CT, Okada NA, Wolstenholme DR. 1996. Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc Natl Acad Sci U S A. 93(11):5619-5623.

Beagley CT, Okimoto R, Wolstenholme DR. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics. 148(3):1091-1108.

Beaudet D, Nadimi M, Iffis B, Hijri M. 2013. Rapid mitochondrial genome evolution through invasion of mobile elements in two closely related species of arbuscular mycorrhizal fungi. PLoS One. 8(4):e60768. doi: 10.1371/journal.pone.0060768.

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.

Belcour L, Begel O, Mossé MO, Vierny C. 1981. Mitochondrial DNA amplification in senescent cultures of *Podospora anserina*: Variability between the retained, amplified sequences. Curr Genet. 3(1):13-21. doi: 10.1007/BF00419575.

Belcour L, Vierny C. 1986. Variable DNA splicing sites of a mitochondrial intron: relationship to the senescence process in *Podospora*. EMBO J. 5(3):609-614.

Belfort M, Bonocora RP. 2014. Homing endonucleases: from genetic anomalies to programmable genomic clippers. Methods Mol Biol. 1123:1-26. doi: 10.1007/978-1-62703-968-0_1.

Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz AM. 2002. Mobile Introns: Pathways and proteins. In: Mobile DNA II. Editors: Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington DC: ASM Press. pp. 761-783.

Belfort M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. Genes Dev. 17(23):2860-2863.

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.

Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2015. GenBank. Nucleic Acids Res. 43(Database issue):D30-D35. doi: 10.1093/nar/gku1216.

Bernardi G. 2005. Lessons from a small, dispensable genome: the mitochondrial genome of yeast. Gene. 354:189-200.

Bertrand H. 2000. Role of mitochondrial DNA in the senescence and hypovirulence of fungi and potential for plant disease control. Annu Rev Phytopathol. 38:397-422. doi: 10.1146/annurev.phyto.38.1.397.

Beylot B, Spassky A. 2001. Chemical probing shows that the intron-encoded endonuclease I-*Sce*I distorts DNA through binding in monomeric form to its homing site. J Biol Chem. 276(27):25243-25253.

Bilewitch JP, Degnan SM. 2011. A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. BMC Evol Biol. 11:228. doi: 10.1186/1471-2148-11-228.

Bolduc JM, Spiegel PC, Chatterjee P, Brady KL, Downing ME, Caprara MG, Waring RB, Stoddard BL. 2003. Structural and biochemical analyses of DNA and RNA binding by a bifunctional homing endonuclease and group I intron splicing factor. Genes Dev. 17(23):2875-2888.

Bonocora RP, Shub DA. 2009. A likely pathway for formation of mobile group I introns. Curr Biol. 19(3):223-228. doi: 10.1016/j.cub.2009.01.033.

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

Bryk M, Quirk SM, Mueller JE, Loizos N, Lawrence C, Belfort M. 1993. The td intron endonuclease I-*Tev*I makes extensive sequence-tolerant contacts across the minor groove of its DNA target. EMBO J. 12(5):2141-2149.

Bullerwell CE, Burger G, Lang BF. 2000. A novel motif for identifying rps3 homologs in fungal mitochondrial genomes. Trends Biochem Sci. 25(8):363-365.

Bullerwell CE, Leigh J, Seif E, Longcore JE, Lang BF. 2003. Evolution of the Fungi and their mitochondrial genomes. In: Applied mycology and biotechnology. Volume III: Fungal genomics. Editors: Arora DK, Khachatourians GG. New York: Elsevier Science. pp. 133-59.

Burke JM, RajBhandary UL. 1982. Intron within the large rRNA gene of *N. crassa* mitochondria: a long open reading frame and a consensus sequence possibly important in splicing. Cell. 31(3 Pt 2):509-520.

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

Candales MA, Duong A, Hood KS, Li T, Neufeld RA, Sun R, McNeil BA, Wu L, Jarding AM, Zimmerly S. 2012. Database for bacterial group II introns. Nucleic Acids Res. 40(Database issue):D187-D190. doi: 10.1093/nar/gkr1043.

Caprara MG, Waring RB. 2005. Group I introns and their maturases: uninvited, but welcome guests. In: Homing endonucleases and inteins. Volume 16. Editors: Belfort M, Stoddard BL, Wood DW, Derbyshire V. Springer-Verlag, New York (NY). pp. 103-119.

Carignani G, Groudinsky O, Frezza D, Schiavon E, Bergantino E, Slonimski PP. 1983. An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. Cell. 35(3 Pt 2):733-742.

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

Cech TR, Damberger SH, Gutell RR. 1994. Representation of the secondary and tertiary structure of group I introns. Nat Struct Biol. 1(5):273-280.

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.

Chan YS, Naujoks DA, Huen DS, Russell S. 2011. Insect population control by homing endonuclease-based gene drive: an evaluation in *Drosophila melanogaster*. Genetics. 188(1):33-44. doi: 10.1534/genetics.111.127506.

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

Chevalier BS, Monnat RJ Jr, Stoddard BL. 2001. The homing endonuclease I-*Cre*I uses three metals, one of which is shared between the two active sites. Nat Struct Biol. 8(4):312-316.

Chevalier BS, Stoddard BL. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. 29(18):3757-3774.

Chi SI, Johansen SD. 2017. Zoantharian mitochondrial genomes contain unique complex group I introns and highly conserved intergenic regions. Gene. 628:24-31. doi: 10.1016/j.gene.2017.07.023.

Chiang CC, Kennell JC, Wanner LA, Lambowitz AM. 1994. A mitochondrial retroplasmid integrates into mitochondrial DNA by a novel mechanism involving the synthesis of a hybrid cDNA and homologous recombination. Mol Cell Biol. 14(10):6419-6432.

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

Colleaux L, d'Auriol L, Betermier M, Cottarel G, Jacquier A, Galibert F, Dujon B. 1986. Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. Cell. 44(4):521-533.

Copertino DW, Hall ET, Van Hook FW, Jenkins KP, Hallick RB. 1994. A group III twintron encoding a maturase-like gene excises through lariat intermediates. Nucleic Acids Res. 22(6):1029-1036.

Copertino DW, Hallick RB. 1993. Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. Trends Biochem Sci. 18(12):467-471.

Craven KD, Vélëz H, Cho Y, Lawrence CB, Mitchell TK. 2008. Anastomosis is required for virulence of the fungal necrotroph *Alternaria brassicicola*. Eukaryot Cell. 7(4):675-683. doi: 10.1128/EC.00423-07.

Davies RW, Waring RB, Ray JA, Brown TA, Scazzocchio C. 1982. Making ends meet: a model for RNA splicing in fungal mitochondria. Nature. 300(5894):719-724.

Day PR, Dodds JA, Elliston JE, Jaynes RA, Anagnostakis SL. 1977. Double-stranded RNA in *Endothia parasitica* [from Europe and North America, biological control, chestnut blight]. Phytopathology (USA).

de Queiroz CB, Santana MF, Pereira Vidigal PM, de Queiroz MV. 2018. Comparative analysis of the mitochondrial genome of the fungus *Colletotrichum lindemuthianum*, the causal agent of anthracnose in common beans. Appl Microbiol Biotechnol. 102(6):2763-2778. doi: 10.1007/s00253-018-8812-0.

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

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

Dellaporta SL, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. 2006. Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. Proc Natl Acad Sci U S A. 103(23):8751-8756.

Déquard-Chablat M, Sellem CH, Golik P, Bidard F, Martos A, Bietenhader M, di Rago JP, Sainsard-Chanet A, Hermann-Le Denmat S, Contamine V. 2011. Two nuclear life cycle-regulated genes encode interchangeable subunits c of mitochondrial ATP synthase in *Podospora anserina*. Mol Biol Evol. 28(7):2063-2075. doi: 10.1093/molbev/msr025.

Derbyshire V, Belfort M. 1998. Lightning strikes twice: intron-intein coincidence. Proc Natl Acad Sci USA. 95(4):1356-1357.

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

Doetsch NA, Thompson MD, Hallick RB. 1998. A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg? Mol Biol Evol. 15(1):76-86.

Dufour E, Boulay J, Rincheval V, Sainsard-Chanet A. 2000. A causal link between respiration and senescence in *Podospora anserina*. Proc Natl Acad Sci U S A. 97(8):4138-4143.

Dujon B, Belfort M, Butow RA, Jacq C, Lemieux C, Perlman PS, Vogt VM. 1989. Mobile introns: definition of terms and recommended nomenclature. Gene. 82(1):115-118.

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.

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

Duò A, Bruggmann R, Zoller S, Bernt M, Grünig 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.

Eddy SR, Durbin R. 1994. RNA sequence analysis using covariance models. Nucleic Acids Res. 22(11):2079-2088.

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.

Eldarov MA, Mardanov AV, Beletsky AV, Ravin NV, Skryabin KG. 2011. Complete sequence and analysis of the mitochondrial genome of the methylotrophic yeast *Hansenula*

polymorpha DL-1. FEMS Yeast Res. 11(6):464-472. doi: 10.1111/j.1567-1364.2011.00736.x.

Emblem Å, Karlsen BO, Evertsen J, Johansen SD. 2011. Mitogenome rearrangement in the cold-water scleractinian coral *Lophelia pertusa* (Cnidaria, Anthozoa) involves a long-term evolving group I intron. Mol Phylogenet Evol. 61(2):495-503. doi: 10.1016/j.ympev.2011.07.012.

Emblem Å, Okkenhaug S, Weiss ES, Denver DR, Karlsen BO, Moum T, Johansen SD. 2014. Sea anemones possess dynamic mitogenome structures. Mol Phylogenet Evol. 75:184-193. doi: 10.1016/j.ympev.2014.02.016.

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.

Esposti MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A. 1993. Mitochondrial cytochrome b: evolution and structure of the protein. Biochim Biophys Acta. 1143(3):243-271.

Faye G, Dennebouy N, Kujawa C, Jacq C. 1979. Inserted sequence in the mitochondrial 23S ribosomal RNA gene of the yeast *Saccharomyces cerevisiae*. Mol Gen Genet. 168(1):101-109.

Fedorova O, Mitros T, Pyle AM. 2003. Domains 2 and 3 interact to form critical elements of the group II intron active site. J Mol Biol. 330(2):197-209.

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.

Férandon C, Xu J, Barroso G. 2013. The 135 kbp mitochondrial genome of *Agaricus bisporus* is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. Fungal Genet Biol. 55:85-91. doi: 10.1016/j.fgb.2013.01.009.

Fernández-Ortuño D, Torés JA, de Vicente A, Pérez-García A. 2008. Field resistance to QoI fungicides in *Podosphaera fusca* is not supported by typical mutations in the mitochondrial cytochrome b gene. Pest Manag Sci. 64(7):694-702. doi: 10.1002/ps.1544.

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

Forgan AH, Knogge W, Anderson PA. 2007. Asexual genetic exchange in the barley pathogen *Rhynchosporium secalis*. Phytopathology. 97(5):650-654. doi: 10.1094/PHYTO-97-5-0650.

Franco MEE, López SMY, Medina R, Lucentini CG, Troncozo MI, Pastorino GN, Saparrat MCN, Balatti PA. 2017. The mitochondrial genome of the plant-pathogenic fungus *Stemphylium lycopersici* uncovers a dynamic structure due to repetitive and mobile elements. PLoS One. 12(10):e0185545. doi: 10.1371/journal.pone.0185545.

Frazier CL, San Filippo J, Lambowitz AM, Mills DA. 2003. Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. Applied and environmental microbiology. 69(2):1121-1128.

Freel KC, Friedrich A, Hou J, Schacherer J. 2014. Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. Genome Biol Evol. 6(10):2586-2594. doi: 10.1093/gbe/evu203.

Freel KC, Friedrich A, Schacherer J. 2015. Mitochondrial genome evolution in yeasts: an all-encompassing view. FEMS Yeast Res. 15(4):fov023. doi: 10.1093/femsyr/fov023.

Fukami H, Chen CA, Chiou CY, Knowlton N. 2007. Novel group I introns encoding a putative homing endonuclease in the mitochondrial *cox1* gene of Scleractinian corals. J Mol Evol. 64(5):591-600.

Fulbright DW. 1985. A cytoplasmic hypovirulent strain of *Endothia parasitica*. Phytopathology. 75:1328 (abstract).

Galizi R, Doyle LA, Menichelli M, Bernardini F, Deredec A, Burt A, Stoddard BL, Windbichler N, Crisanti A. 2014. A synthetic sex ratio distortion system for the control of the human malaria mosquito. Nat Commun. 5:3977. doi: 10.1038/ncomms4977.

Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. J Mol Biol. 313(5):1003-1011.

Gauthier A, Turmel M, Lemieux C. 1991. A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron. Curr Genet. 19(1):43-47.

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(Pt 10):1112-1126.

Glass NL, Rasmussen C, Roca MG, Read ND. 2004. Hyphal homing, fusion and mycelial interconnectedness. Trends Microbiol. 12(3):135-141.

Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. Proc Natl Acad Sci USA. 96(24):13880-13885.

Gomes FEES, Arantes TD, Fernandes JAL, Ferreira LC, Romero H, Bosco SMG, Oliveira MTB, Del Negro GMB, Theodoro RC. 2018. Polymorphism in mitochondrial group I introns among *Cryptococcus neoformans* and *Cryptococcus gattii* genotypes and its association with drug susceptibility. Front Microbiol. 9:86. doi: 10.3389/fmicb.2018.00086.

Goodwin SB, McCorison CB, Cavaletto JR, Culley DE, LaButti K, Baker SE, Grigoriev IV. 2016. The mitochondrial genome of the ethanol-metabolizing, wine cellar mold *Zasmidium cellare* is the smallest for a filamentous ascomycete. Fungal Biol. 120(8):961-974. doi: 10.1016/j.funbio.2016.05.003.

Grasso V, Palermo S, Sierotzki H, Garibaldi A, Gisi U. 2006. Cytochrome b gene structure and consequences for resistance to Qo inhibitor fungicides in plant pathogens. Pest Manag Sci. 62(6):465-472.

Gray MW. 2015. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proc Natl Acad Sci U S A. 112(33):10133-10138. doi: 10.1073/pnas.1421379112.

Gross SR, Hsieh TS, Levine PH. 1984. Intramolecular recombination as a source of mitochondrial chromosome heteromorphism in *Neurospora*. Cell. 38(1):233-239.

Guha TK, Hausner G. 2014. A homing endonuclease with a switch: Characterization of a twintron encoded homing endonuclease. Fungal Genet Biol. 65:57-68. doi: 10.1016/j.fgb.2014.01.004.

Guha TK, Hausner G. 2016. Using group II introns for attenuating the *in vitro* and *in vivo* expression of a homing endonuclease. PLoS One. 11(2):e0150097. doi: 10.1371/journal.pone.0150097.

Guha TK, Wai A, Hausner G. 2017a. 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. 2017b. The intron landscape of the mtDNA *cytb* gene among the Ascomycota: introns and intron-encoded open reading frames. Mitochondrial DNA A DNA Mapp Seq Anal. 20:1-10. doi: 10.1080/24701394.2017.1404042.

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

Guo H, Karberg M, Long M, Jones JP 3rd, Sullenger B, Lambowitz AM. 2000. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. Science. 289(5478):452-457.

Guo QB, Akins RA, Garriga G, Lambowitz AM. 1991. Structural analysis of the *Neurospora* mitochondrial large rRNA intron and construction of a mini-intron that shows protein-dependent splicing. J Biol Chem. 266(3):1809-1819.

Guo WW, Moran JV, Hoffman PW, Henke RM, Butow RA, Perlman PS. 1995. The mobile group I intron 3 alpha of the yeast mitochondrial *COXI* gene encodes a 35-kDa processed protein that is an endonuclease but not a maturase. J Biol Chem. 270(26):15563-15570.

Hafez M, Hausner G. 2012. Homing endonucleases: DNA scissors on a mission. Genome. 55(8):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):1275-1288. 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 Genet Biol. 53:71-83. doi: 10.1016/j.fgb.2013.01.005.

Hane JK, Lowe RG, Solomon PS, Tan KC, Schoch CL, Spatafora JW, Crous PW, Kodira C, Birren BW, Galagan JE, Torriani SF, McDonald BA, Oliver RP. 2007. Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. Plant Cell. 19(11):3347-3368.

Haridas S, Salamov A, Grigoriev IV. 2018. Fungal genome annotation. Methods Mol Biol. 1775:171-184. doi: 10.1007/978-1-4939-7804-5_15.

Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. Trends Genet. 21(2):111-119.

Hausner G, Monteiro-Vitorello CB, Searles DB, Maland M, Fulbright DW, Bertrand H. 1999. A long open reading frame in the mitochondrial LSU rRNA group-I intron of *Cryphonectria parasitica* encodes a putative S5 ribosomal protein fused to a maturase. Curr Genet. 35(2):109-117.

Hausner G. 2003. Fungal Mitochondrial Genomes. Fungal Genomics. 3:101.

Hausner G. 2012. Introns, mobile elements, and plasmids. In: Organelle genetics - Evolution of organelle genomes and gene expression. Editor: Bullerwell CE. New York (NY): Springer-Verlag. pp. 329-357.

Hausner G. Reid J, Klassen GR. 1992. Do galeate-ascospore members of the Cepahloascaceae, Endomycetaceae and Ophiostomataceae share a common phylogeny? Mycologia. 84(6):870-881. doi: 10.2307/3760285.

Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6(2):e1000834. doi: 10.1371/journal.pgen.1000834.

Heath PJ, Stephens KM, Monnat RJ Jr, Stoddard BL. 1997. The structure of I-*Cre*1, a group I intron-encoded homing endonuclease. Nat Struct Biol. 4(6):468-476.

Hedberg A, Johansen SD. 2013. Nuclear group I introns in self-splicing and beyond. Mob DNA. 4(1):17. doi: 10.1186/1759-8753-4-17.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, Thorsten Lumbsch H, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalpers JA, Vilgalys R, Aime MC, Aptroot A, Bauer R, Begerow D, Benny GL, Castlebury LA, Crous PW, Dai YC, Gams W, Geiser DM, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G, Hosaka K, Humber RA, Hyde KD, Ironside JE, Kõljalg U, Kurtzman CP, Larsson KH, Lichtwardt R, Longcore J, Miadlikowska J, Miller A, Moncalvo JM, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, Rogers JD, Roux C, Ryvarden L, Sampaio JP, Schüssler A, Sugiyama J, Thorn RG, Tibell L, Untereiner WA, Walker C, Wang Z, Weir A, Weiss M, White MM, Winka K, Yao YJ, Zhang N. 2007. A higher-level phylogenetic classification of the Fungi. Mycol Res. 111(Pt 5):509-547.

Hickey PC1, Jacobson D, Read ND, Glass NL. 2002. Live-cell imaging of vegetative hyphal fusion in *Neurospora crassa*. Fungal Genet Biol. 37(1):109-119.

Hillman BI, Suzuki N. 2004. Viruses of the chestnut blight fungus, *Cryphonectria parasitica*. Adv Virus Res. 63:423-472.

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

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 Evol Biol. 15:278. doi: 10.1186/s12862-015-0556-1.

Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K. 2000. Crystal structure of an archaeal intein-encoded homing endonuclease PI-*Pfu*I. J Mol Biol. 300(4):889-901.

Ishikawa FH, Souza EA, Shoji JY, Connolly L, Freitag M, Read ND, Roca MG. 2012. Heterokaryon incompatibility is suppressed following conidial anastomosis tube fusion in a fungal plant pathogen. PLoS One. 7(2):e31175. doi: 10.1371/journal.pone.0031175.

Jackson CJ, Barton RC, Clark CG, Kelly SL. 2009. Molecular characterization of a subgroup IE intron with wide distribution in the large subunit rRNA genes of dermatophyte fungi. Med Mycol. 47(6):609-617. doi: 10.1080/13693780802385445.

Jacquier A, Dujon B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell. 41(2):383-394. Curr Biol. 23(16):1548-1553. doi: 10.1016/j.cub.2013.06.057.

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.

Jamet-Vierny C, Begel O, Belcour L. 1980. Senescence in *Podospora anserina*: amplification of a mitochondrial DNA sequence. Cell. 21(1):189-194.

Jaynes RA, Elliston JE. 1982. Hypovirulent isolates of *Endothia parasitica* associated with large American chestnut trees. Plant Diseases (USA). 66:769-772.

Joardar V, Abrams NF, Hostetler J, Paukstelis PJ, Pakala S, Pakala SB, Zafar N, Abolude OO, Payne G, Andrianopoulos A, Denning DW, Nierman WC. 2012. Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genome size variability. BMC Genomics. 13:698. doi: 10.1186/1471-2164-13-698.

Johansen S, Haugen P. 2001. A new nomenclature of group I introns in ribosomal DNA. RNA. 7(7):935-936.

Johnston A, Booth C. 1983. Plant pathologist's pocketbook. 2nd ed. Kew: Commonwealth Mycological Institute.

Jung PP, Friedrich A, Reisser C, Hou J, Schacherer J. 2012. Mitochondrial genome evolution in a single protoploid yeast species. G3 (Bethesda). 2(9):1103-1111. doi: 10.1534/g3.112.003152.

Jung PP, Schacherer J, Souciet JL, Potier S, Wincker P, de Montigny J. 2009. The complete mitochondrial genome of the yeast *Pichia sorbitophila*. FEMS Yeast Res. 9(6):903-910. doi: 10.1111/j.1567-1364.2009.00540.x.

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(6):e0156104. doi: 10.1371/journal.pone.0156104.

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol. 19(12):1162-1167.

Katoch M, Pull S. 2017. Endophytic fungi associated with *Monarda citriodora*, an aromatic and medicinal plant and their biocontrol potential. Pharm Biol. 55(1):1528-1535. doi: 10.1080/13880209.2017.1309054.

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.

Kennell JC, Cohen SM. 2004. Fungal mitochondria: genomes, genetic elements and gene expression. In: The handbook of fungal biotechnology, 2nd ed. Editor: Arora DK. New York: Marcel Dekker Inc. pp. 131-143.

Kennell JC, Lambowitz AM. 1989. Development of an in vitro transcription system for *Neurospora crassa* mitochondrial DNA and identification of transcription initiation sites. Mol Cell Biol. 9(9):3603-3613.

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

Kim SH, Cech TR. 1987. Three-dimensional model of the active site of the self-splicing rRNA precursor of *Tetrahymena*. Proc Natl Acad Sci U S A. 84(24):8788-8792.

Koch JL, Boulanger SC, Dib-Hajj SD, Hebbar SK, Perlman PS. 1992. Group II introns deleted for multiple substructures retain self-splicing activity. Mol Cell Biol. 12(5):1950-1958.

Koulintchenko M, Konstantinov Y, Dietrich A. 2003. Plant mitochondria actively import DNA via the permeability transition pore complex. EMBO J. 22(6):1245-1254.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 33(7):1870-1874. doi: 10.1093/molbev/msw054.

Lambert AR, Hallinan JP, Shen BW, Chik JK, Bolduc JM, Kulshina N, Robins LI, Kaiser BK, Jarjour J, Havens K, Scharenberg AM, Stoddard BL. 2016. Indirect DNA sequence recognition and its impact on nuclease cleavage activity. Structure. 24(6):862-873. doi: 10.1016/j.str.2016.03.024.

Lambowitz AM, Belfort M. 2015. Mobile bacterial group II introns at the crux of eukaryotic evolution. Microbiol Spectr. 3(1):MDNA3-0050-2014. doi: 10.1128/microbiolspec.MDNA3-0050-2014.

Lambowitz AM, Caprara MG, Zimmerly S, Perlman PS. 1998. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. The RNA world. 2:451-485.

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, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. Annu Rev Genet. 33:351-397.

Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23(3):119-125.

LaPolla RJ, Lambowitz AM. 1981. Mitochondrial ribosome assembly in *Neurospora crassa*. Purification of the mitochondrially synthesized ribosomal protein, S-5. J Biol Chem. 256(13):7064-7067.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23(21): 2947-2948.

Lassmann T, Sonnhammer EL. 2005. Kalign - an accurate and fast multiple sequence alignment program. BMC Bioinformatics. 6:298.

Lavín JL, Oguiza JA, Ramírez L, Pisabarro AG. 2008. Comparative genomics of the oxidative phosphorylation system in fungi. Fungal Genet Biol. 45(9):1248-1256.

Leslie JF. 1993. Fungal vegetative compatibility. Annu Rev Phytopathol. 31:127-150.

Li H, Monnat RJ Jr. 2014. Homing endonuclease target site specificity defined by sequential enrichment and next-generation sequencing of highly complex target site libraries. Methods Mol Biol. 1123:151-163. doi: 10.1007/978-1-62703-968-0_12.

Li MV, Shukla D, Rhodes BH, Lall A, Shu J, Moriarity BS, Largaespada DA. 2014. HomeRun Vector Assembly System: a flexible and standardized cloning system for assembly of multi-modular DNA constructs. PLoS One. 9(6):e100948. doi: 10.1371/journal.pone.0100948.

Li Y, Hu XD, Yang RH, Hsiang T, Wang K, Liang DQ, Liang F, Cao DM, Zhou F, Wen G, Yao YJ. 2015. Complete mitochondrial genome of the medicinal fungus *Ophiocordyceps sinensis*. Sci Rep. 5:13892. doi: 10.1038/srep13892.

Liang X, Tian X, Liu W, Wei T, Wang W, Dong Q, Wang B, Meng Y, Zhang R, Gleason ML, Sun G. 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.

Liu JK, Chen WH, Ren SX, Zhao GP, Wang J. 2014. iBrick: a new standard for iterative assembly of biological parts with homing endonucleases. PLoS One. 9(10):e110852.

Liu YJ, Zhang J, Cui GZ, Cui Q. 2015. Current progress of targetron technology: development, improvement and application in metabolic engineering. Biotechnol J. 10(6):855-865. doi: 10.1002/biot.201400716.

Loizos N, Tillier ER, Belfort M. 1994. Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. Proc Natl Acad Sci USA. 91(25):11983-11987.

Losada L, Pakala SB, Fedorova ND, Joardar V, Shabalina SA, Hostetler J, Pakala SM, Zafar N, Thomas E, Rodriguez-Carres M, Dean R, Vilgalys R, Nierman WC, Cubeta MA. 2014. Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. FEMS Microbiol Lett. 352(2):165-173. doi: 10.1111/1574-6968.12387.

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

Lücking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT. 2009. Fungi evolved right on track. Mycologia. 101(6):810-822.

Luo CX, Hu MJ, Jin X, Yin LF, Bryson PK, Schnabel G. 2010. An intron in the cytochrome b gene of *Monilinia fructicola* mitigates the risk of resistance development to QoI fungicides. Pest Manag Sci. 66(12):1308-1315. doi: 10.1002/ps.2016.

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

Macreadie IG, Scott RM, Zinn AR, Butow RA. 1985. Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. Cell. 41(2):395-402.

Mahanti N, Bertrand H, Monteiro-Vitorello C, Fulbright DW. 1993. Elevated mitochondrial alternative oxidase activity in dsRNA-free, hypovirulent isolates of *Cryphonectria parasitica*. Physiological and molecular plant pathology. 42(6):455-463.

Marcaida MJ, Muñoz IG, Blanco FJ, Prieto J, Montoya G. 2010. Homing endonucleases: from basics to therapeutic applications. Cell Mol Life Sci. 67(5):727-748. doi: 10.1007/s00018-009-0188-y.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43(Database issue):D222-D226. doi: 10.1093/nar/gku1221.

Marcia M, Pyle AM. 2012. Visualizing group II intron catalysis through the stages of splicing. Cell. 151(3):497-507. doi: 10.1016/j.cell.2012.09.033.

Mardanov AV, Beletsky AV, Kadnikov VV, Ignatov AN, Ravin NV. 2014. The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. PLoS One. 9(9):e107536. doi: 10.1371/journal.pone.0107536.

McNeil BA, Semper C, Zimmerly S. 2016. Group II introns: versatile ribozymes and retroelements. Wiley Interdiscip Rev RNA. 7(3):341-355. doi: 10.1002/wrna.1339.

McNeil BA, Zimmerly S. 2014. Novel RNA structural features of an alternatively splicing group I intron from *Clostridium tetani*. RNA. 20(6):855-866. doi: 10.1261/rna.042440.113.

Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL. 2006. Naked corals: skeleton loss in Scleractinia. Proc Natl Acad Sci U S A. 103(24):9096-9100.

Mehta P, Katta K, Krishnaswamy S. 2004. HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. Protein Sci. 13(1):295-300.

Michel F, Costa M, Westhof E. 2009. The ribozyme core of group II introns: a structure in want of partners. Trends Biochem Sci. 34(4):189-199. doi: 10.1016/j.tibs.2008.12.007.

Michel F, Ferat JL. 1995. Structure and activities of group II introns. Annu Rev Biochem. 64:435-461.

Michel F, Grandchamp C, Dujon B. 1979. Genetic and physical characterization of a segment of yeast mitochondrial DNA involved in the control of genetic recombination. Biochimie. 61(9):985-1010.

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.

Mohr G, Hong W, Zhang J, Cui GZ, Yang Y, Cui Q, Liu YJ, Lambowitz AM. 2013a. A targetron system for gene targeting in thermophiles and its application in *Clostridium thermocellum*. PLoS One. 8(7):e69032. doi: 10.1371/journal.pone.0069032.

Mohr G, Perlman PS, Lambowitz AM. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucleic Acids Res. 21(22):4991-4997.

Mohr S, Ghanem E, Smith W, Sheeter D, Qin Y, King O, Polioudakis D, Iyer VR, Hunicke-Smith S, Swamy S, Kuersten S, Lambowitz AM. 2013b. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. RNA. 19(7):958-970. doi: 10.1261/rna.039743.113.

Molina R, Redondo P, Stella S, Marenchino M, D'Abramo M, Gervasio FL, Epinat JC, Valton J, Grizot S, Duchateau P, Prieto J, Montoya G. 2012. Non-specific protein-DNA interactions control I-*Cre*I target binding and cleavage. Nucleic Acids Res. 40(14):6936-6945. doi: 10.1093/nar/gks320.

Moran JV, Mecklenburg KL, Sass P, Belcher SM, Mahnke D, Lewin A, Perlman P. 1994. Splicing defective mutants of the *COXI* gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. Nucleic Acids Res. 22(11):2057-2064.

Mueller GM and Bills GF. 2004. Biodiversity of fungi - Inventory and monitoring methods. Elsevier Academic Press, MA, USA. pp. 1-4.

Nadimi M, Stefani FO, Hijri M. 2016. The large (134.9 kb) mitochondrial genome of the glomeromycete *Funneliformis mosseae*. Mycorrhiza. 26(7):747-755. doi: 10.1007/s00572-016-0710-7.

Narladkar BW, Shivpuje PR, Harke PC. 2015. Fungal biological control agents for integrated management of *Culicoides* spp. (Diptera: Ceratopogonidae) of livestock. Vet World. 8(2):156-163. doi: 10.14202/vetworld.2015.156-163.

Nawrocki EP, Jones TA, Eddy SR. 2018. Group I introns are widespread in archaea. Nucleic Acids Res. doi: 10.1093/nar/gky414.

Ng B, Nayak S, Gibbs MD, Lee J, Bergquist PL. 2007. Reverse transcriptases: intron-encoded proteins found in thermophilic bacteria. Gene. 393(1-2):137-144.

Niazi AK, Mileshina D, Cosset A, Val R, Weber-Lotfi F, Dietrich A. 2013. Targeting nucleic acids into mitochondria: progress and prospects. Mitochondrion. 13(5):548-558. doi: 10.1016/j.mito.2012.05.004.

Nicholas KB, Nicholas HB, Deerfield II DW. 1997. GeneDoc: analysis and visualization of genetic variation. EMB News. 4:14.

Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV, Dubchak I. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. 42(Database issue):D26-D31. doi: 10.1093/nar/gkt1069.

Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, Belfort M. 2016. Intein clustering suggests functional importance in different domains of life. Mol Biol Evol. 33(3):783-799. doi: 10.1093/molbev/msv271.

Oro L, Feliziani E, Ciani M, Romanazzi G, Comitini F. 2018. Volatile organic compounds from *Wickerhamomyces anomalus, Metschnikowia pulcherrima* and *Saccharomyces cerevisiae* inhibit growth of decay causing fungi and control postharvest diseases of strawberries. Int J Food Microbiol. 265:18-22. doi: 10.1016/j.ijfoodmicro.2017.10.027.

Osíewacz HD, Esser K. 1984. The mitochondrial plasmid of *Podospora anserina*: A mobile intron of a mitochondrial gene. Curr Genet. 8(4):299-305. doi: 10.1007/BF00419728.

Østergaard LH, Olsen HS. 2010. Industrial applications of fungal enzymes. In: Industrial Applications. Springer Berlin Heidelberg. pp. 269-290.

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

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A. 97(13):6960-6966.

Perea J, Jacq C. 1985. Role of the 5' hairpin structure in the splicing accuracy of the fourth intron of the yeast cob-box gene. EMBO J. 4(12):3281-3288.

Perez-Campo R, López-Torres M, Cadenas S, Rojas C, Barja G. 1998. The rate of free radical production as a determinant of the rate of aging: evidence from the comparative approach. J Comp Physiol B. 168(3):149-158.

Perrin A, Buckle M, Dujon B. 1993. Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. EMBO J. 12(7):2939-2947.

Perutka J, Wang W, Goerlitz D, Lambowitz AM. 2004. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. J Mol Biol. 336(2):421-439.

Phylactou LA, Darrah C, Wood MJ. 1998. Ribozyme-mediated trans-splicing of a trinucleotide repeat. Nat Genet. 18(4):378-381.

Polevoda B, Arnesen T, Sherman F. 2009. A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. BMC Proc. 3 Suppl 6:S2. doi: 10.1186/1753-6561-3-S6-S2.

Polevoda B, Sherman F. 2002. The diversity of acetylated proteins. Genome Biol. 3(5):reviews0006.

Polevoda B, Sherman F. 2003. Composition and function of the eukaryotic N-terminal acetyltransferase subunits. Biochem Biophys Res Commun. 308(1):1-11.

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.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 5(3):e9490. doi: 10.1371/journal.pone.0009490.

Procházka E, Poláková S, Piskur J, Sulo P. 2010. Mitochondrial genome from the facultative anaerobe and petite-positive yeast *Dekkera bruxellensis* contains the NADH dehydrogenase subunit genes. FEMS Yeast Res. 10(5):545-557. doi: 10.1111/j.1567-1364.2010.00644.x.

Qin PZ, Pyle AM. 1998. The architectural organization and mechanistic function of group II intron structural elements. Curr Opin Struct Biol. 8(3):301-308.

Repar J, Warnecke T. 2017. Mobile introns shape the genetic diversity of their host genes. Genetics. 205(4):1641-1648. doi: 10.1534/genetics.116.199059.

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

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

Ritchings BW, Lewin AS. 1992. Mutational evidence for competition between the P1 and the P10 helices of a mitochondrial group I intron. Nucleic Acids Res. 20(9):2349-2353.

Robart AR, Zimmerly S. 2005. Group II intron retroelements: function and diversity. Cytogenet Genome Res. 110(1-4):589-597.

Roca MG, Arlt J, Jeffree CE, Read ND. 2005. Cell biology of conidial anastomosis tubes in *Neurospora crassa*. Eukaryot Cell. 4(5):911-919.

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

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.

Ronquist F. 2004. Bayesian inference of character evolution. Trends Ecol Evol. 19(9):475-481.

Rot C, Goldfarb I, Ilan M, Huchon D. 2006. Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. BMC Evol Biol. 6:71.

Ruan J, Cheng J, Zhang T, Jiang H. 2017. Mitochondrial genome evolution in the *Saccharomyces sensu stricto* complex. PLoS One. 12(8):e0183035. doi: 10.1371/journal.pone.0183035.

Rudan M, Bou Dib P, 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. pii: e35330. doi: 10.7554/eLife.35330.

Sainsard-Chanet A, Begel O, Belcour L. 1993. DNA deletion of mitochondrial introns is correlated with the process of senescence in *Podospora anserina*. J Mol Biol. 234(1):1-7.

Sainsard-Chanet A, Begel O, d'Aubenton-Carafa Y. 1998. Two co-existing mechanisms account for the large-scale deletions of mitochondrial DNA in *Podospora anserina* that involve the 5' border of a group-II intron. Curr Genet. 34(4):326-335.

Salah Ud-Din AI, Tikhomirova A, Roujeinikova A. 2016. Structure and functional diversity of GCN5-related N-Acetyltransferases (GNAT). Int J Mol Sci. 17(7). pii: E1018. doi: 10.3390/ijms17071018.

Salavirta H, Oksanen I, Kuuskeri J, Mäkelä M, Laine P, Paulin L, Lundell T. 2014. Mitochondrial genome of *Phlebia radiata* is the second largest (156 kbp) among fungi

and features signs of genome flexibility and recent recombination events. PLoS One. 9(5):e97141. doi: 10.1371/journal.pone.0097141.

Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM. 1999. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. Biochemistry. 38(28):9069-9083.

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

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

Saquez C, Lecellier G, Koll F. 2000. Intronic GIY-YIG endonuclease gene in the mitochondrial genome of *Podospora curvicolla*: evidence for mobility. Nucleic Acids Res. 28(6):1299-1306.

Saupe SJ. 2000. Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. Microbiol Mol Biol Rev. 64(3):489-502.

Schäfer B. 2003. Genetic conservation versus variability in mitochondria: the architecture of the mitochondrial genome in the petite-negative yeast *Schizosaccharomyces pombe*. Curr Genet. 43(5):311-326.

Schmidt U, Riederer B, Mörl M, Schmelzer C, Stahl U. 1990. Self-splicing of the mobile group II intron of the filamentous fungus *Podospora anserina* (COI I1) *in vitro*. EMBO J. 9(7):2289-2298.

Schoch CL1, Sung GH, López-Giráldez F, Townsend JP, Miadlikowska J, Hofstetter V, Robbertse B, Matheny PB, Kauff F, Wang Z, Gueidan C, Andrie RM, Trippe K, Ciufetti

LM, Wynns A, Fraker E, Hodkinson BP, Bonito G, Groenewald JZ, Arzanlou M, de Hoog GS, Crous PW, Hewitt D, Pfister DH, Peterson K, Gryzenhout M, Wingfield MJ, Aptroot A, Suh SO, Blackwell M, Hillis DM, Griffith GW, Castlebury LA, Rossman AY, Lumbsch HT, Lücking R, Büdel B, Rauhut A, Diederich P, Ertz D, Geiser DM, Hosaka K, Inderbitzin P, Kohlmeyer J, Volkmann-Kohlmeyer B, Mostert L, O'Donnell K, Sipman H, Rogers JD, Shoemaker RA, Sugiyama J, Summerbell RC, Untereiner W, Johnston PR, Stenroos S, Zuccaro A, Dyer PS, Crittenden PD, Cole MS, Hansen K, Trappe JM, Yahr R, Lutzoni F, Spatafora JW. 2009. The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. Syst Biol. 58(2):224-239. doi: 10.1093/sysbio/syp020.

Schöttler S, Wende W, Pingoud V, Pingoud A. 2000. Identification of Asp218 and Asp326 as the principal Mg^{2+} binding ligands of the homing endonuclease PI-*Sce*I. Biochemistry. 39(51):15895-15900.

Schuster A, Lopez JV, Becking LE, Kelly M, Pomponi SA, Wörheide G, Erpenbeck D, Cárdenas P. 2017. Evolution of group I introns in Porifera: new evidence for intron mobility and implications for DNA barcoding. BMC Evol Biol. 17(1):82. doi: 10.1186/s12862-017-0928-9.

Sellem CH, Belcour L. 1994. The *in vivo* use of alternate 3'-splice sites in group I introns. Nucleic Acids Res. 22(7):1135-1137.

Sellem CH, Belcour L. 1997. Intron open reading frames as mobile elements and evolution of a group I intron. Mol Biol Evol. 14(5):518-526.

Sellem CH, d'Aubenton-Carafa Y, Rossignol M, Belcour L. 1996. Mitochondrial intronic open reading frames in *Podospora*: mobility and consecutive exonic sequence variations. Genetics. 143(2):777-788.

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(7):e1006161. doi: 10.1371/journal.pgen.1006161.

Sellem CH, Lecellier G, Belcour L. 1993. Transposition of a group II intron. Nature. 366(6451):176-178.

Sellem CH, Sainsard-Chanet A, Belcour L. 1990. Detection of a protein encoded by a class II mitochondrial intron of *Podospora anserina*. Mol Gen Genet. 224(2):232-240.

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

Sethuraman J, Majer A, Iranpour M, Hausner G. 2009b. Molecular evolution of the mtDNA encoded *rps3* gene among filamentous ascomycetes fungi with an emphasis on the Ophiostomatoid fungi. J Mol Evol. 69(4):372-385. doi: 10.1007/s00239-009-9291-9.

Sethuraman J, Okoli CV, Majer A, Corkery TL, Hausner G. 2008. The sporadic occurrence of a group I intron-like element in the mtDNA *rnl* gene of *Ophiostoma novo-ulmi* subsp. *americana*. Mycol Res. 112(Pt 5):564-582. doi: 10.1016/j.mycres.2007.11.017.

Sethuraman J, Rudski SM, Wosnitza K, Hafez M, Guppy B, Hausner G. 2013. Evolutionary dynamics of introns and their open reading frames in the U7 region of the mitochondrial

rnl gene in species of *Ceratocystis*. Fungal Biol. 117(11-12):791-806. doi: 10.1016/j.funbio.2013.10.002.

Shen XY, Li T, Chen S, Fan L, Gao J, Hou CL. Characterization and phylogenetic analysis of the mitochondrial genome of *Shiraia bambusicola* reveals special features in the order of Pleosporales. PLoS One. 10(3):e0116466. doi: 10.1371/journal.pone.0116466.

Signorovitch AY, Buss LW, Dellaporta SL. 2007. Comparative genomics of large mitochondria in placozoans. PLoS Genet. 3(1):e13.

Sitbon E, Pietrokovski S. 2003. New types of conserved sequence domains in DNA-binding regions of homing endonucleases. Trends Biochem Sci. 28(9):473-477.

Smith SW, Overbeek R, Woese CR, Gilbert W, Gillevet PM. 1994. The genetic data environment an expandable GUI for multiple sequence analysis. Comput Appl Biosci. 10(6):671-675.

Sokolowska M, Czapinska H, Bochtler M. 2011. Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. Nucleic Acids Res. 39(4):1554-1564. doi: 10.1093/nar/gkq821.

Solieri L. 2010. Mitochondrial inheritance in budding yeasts: towards an integrated understanding. Trends Microbiol. 18(11):521-530. doi: 10.1016/j.tim.2010.08.001.

Spatafora JW, Aime MC, Grigoriev IV, Martin F, Stajich JE, Blackwell M. 2017. The fungal tree of life: from molecular systematics to genome-scale phylogenies. Microbiol Spectr. 5(5). doi: 10.1128/microbiolspec.FUNK-0053-2016.

Stahley MR, Strobel SA. 2005. Structural evidence for a two-metal-ion mechanism of group I intron splicing. Science. 309(5740):1587-1590.

Stoddard BL. 2005. Homing endonuclease structure and function. Q Rev Biophys. 38(1):49-95.

Stoddard BL. 2011. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. Structure. 19(1):7-15. doi: 10.1016/j.str.2010.12.003.

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.

Suh SO, Jones KG, Blackwell M. 1999. A Group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of Group I introns. J Mol Evol. 48(5):493-500.

Sullenger BA, Gilboa E. 2002. Emerging clinical applications of RNA. Nature. 418(6894):252-258.

Sulo P, Szabóová D, Bielik P, Poláková S, Šoltys K, Jatzová K, Szemes T. 2017. The evolutionary history of *Saccharomyces* species inferred from completed mitochondrial genomes and revision in the 'yeast mitochondrial genetic code'. DNA Res. doi: 10.1093/dnares/dsx026.

Sverdlov AV, Babenko VN, Rogozin IB, Koonin EV. 2004. Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism of intron insertion. Gene. 338(1):85-91.

Szczepanek T, Lazowska J. 1996. Replacement of two non-adjacent amino acids in the *S. cerevisiae* bi2 intron-encoded RNA maturase is sufficient to gain a homing-endonuclease activity. EMBO J. 15(14):3758-3767.

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. Cell. 33(1):25-35.

Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, Gloor GB, Scharenberg AM, Edgell DR, Stoddard BL. 2011. Tapping natural reservoirs of homing endonucleases for targeted

gene modification. Proc Natl Acad Sci U S A. 108(32):13077-13082. doi: 10.1073/pnas.1107719108.

Tanaka S, Matsushita Y, Yoshikawa A, Isono K. 1989. Cloning and molecular characterization of the gene *rimL* which encodes an enzyme acetylating ribosomal protein L12 of *Escherichia coli* K12. Mol Gen Genet. 217(2-3):289-293.

Thorsness PE, Fox TD. 1990. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. Nature. 346(6282):376-379.

Thorsness PE, Weber ER. 1996. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. Int Rev Cytol. 165:207-234.

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.

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.

Torriani SF, Goodwin SB, Kema GH, Pangilinan JL, McDonald BA. 2008. Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. Fungal Genet Biol. 45(5):628-637. doi: 10.1016/j.fgb.2007.12.005.

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 44(W1):W232-W235. doi: 10.1093/nar/gkw256.

Trumpower BL. 1990. Cytochrome bc1 complexes of microorganisms. Microbiol Rev. 54(2):101-129.

Turk EM, Das V, Seibert RD, Andrulis ED. 2013. The mitochondrial RNA landscape of *Saccharomyces cerevisiae*. PLoS One. 8(10):e78105. doi: 10.1371/journal.pone.0078105.

Vallès Y, Halanych KM, Boore JL. 2008. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS One. 3(1):e1488. doi: 10.1371/journal.pone.0001488.

van den Boogaart P, Samallo J, Agsteribbe E. 1982. Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa*. Nature. 298(5870):187-189.

Van Roey P, Waddling CA, Fox KM, Belfort M, Derbyshire V. 2001. Intertwined structure of the DNA-binding domain of intron endonuclease I-*Tev*I with its substrate. EMBO J. 20(14):3631-3637.

Vellore J, Moretz SE, Lampson BC. 2004. A group II intron-type open reading frame from the thermophile *Bacillus* (*Geobacillus*) *stearothermophilus* encodes a heat-stable reverse transcriptase. Appl Environ Microbiol. 70(12):7140-7147.

Vetting MW, Bareich DC, Yu M, Blanchard JS. 2008. Crystal structure of RimI from *Salmonella typhimurium* LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. Protein Sci. 17(10):1781-1790. doi: 10.1110/ps.035899.108.

Vetting MW, S de Carvalho LP, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS. 2005. Structure and functions of the GNAT superfamily of acetyltransferases. Arch Biochem Biophys. 433(1):212-226.

Vicens Q, Cech TR. 2006. Atomic level architecture of group I introns revealed. Trends Biochem Sci. 31(1):41-51.

Videira A, Duarte M. 2002. From NADH to ubiquinone in *Neurospora* mitochondria. Biochim Biophys Acta. 1555(1-3):187-191.

Wang X, Lavrov DV. 2008. Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae. PLoS One. 3(7):e2723. doi: 10.1371/journal.pone.0002723.

Weber-Lotfi F, Koulintchenko MV, Ibrahim N, Hammann P, Mileshina DV, Konstantinov YM, Dietrich A. 2015. Nucleic acid import into mitochondria: New insights into the translocation pathways. Biochim Biophys Acta. 1853(12):3165-3181. doi: 10.1016/j.bbamcr.2015.09.011.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 18(5):691-699.

Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT, Baker D, Monnat RJ Jr, Burt A, Crisanti A. 2011. A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature. 473(7346):212-215. doi: 10.1038/nature09937.

Wingfield MJ, Benade E, Van Wyk PS, Visser C. 1995. Conidium development in *Ceratocystis autographa*. Mycological Research. 99(11):1289-1294.

Wolf K, Del Giudice L. 1988. The variable mitochondrial genome of ascomycetes: organization, mutational alterations, and expression. Adv Genet. 25:185-308.

Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. BMC Genomics. 16:451. doi: 10.1186/s12864-015-1664-4.

Wool IG. 1996. Extraribosomal functions of ribosomal proteins. Trends Biochem Sci. 21(5):164-165.

Wu B, Buljic A, Hao W. 2015. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. Mol Biol Evol. 32(10):2559-2570. doi: 10.1093/molbev/msv127.

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

Yao J, Lambowitz AM. 2007. Gene targeting in gram-negative bacteria by use of a mobile group II intron ("Targetron") expressed from a broad-host-range vector. Appl Environ Microbiol. 73(8):2735-2743.

Yin LF, Hu MJ, Wang F, Kuang H, Zhang Y, Schnabel G, Li GQ, Luo CX. 2012. Frequent gain and loss of introns in fungal cytochrome b genes. PLoS One. 2012;7(11):e49096. doi: 10.1371/journal.pone.0049096.

Yoon H, You YH, Woo JR, Park YJ, Kong WS, Lee BM, Kim JG. 2012. The mitochondrial genome of the white-rot fungus *Flammulina velutipes*. J Gen Appl Microbiol. 58(4):331-337.

Zeng Q, Bonocora RP, Shub DA. 2009. A free-standing homing endonuclease targets an intron insertion site in the *psbA* gene of cyanophages. Curr Biol. 19(3):218-222. doi: 10.1016/j.cub.2008.11.069.

Zhang S, Wang XN, Zhang XL, Liu XZ, Zhang YJ. 2017. Complete mitochondrial genome of the endophytic fungus *Pestalotiopsis fici*: features and evolution. Appl Microbiol Biotechnol. 101(4):1593-1604. doi: 10.1007/s00253-017-8112-0.

Zheng D, Köller W. 1997. Characterization of the mitochondrial cytochrome b gene from *Venturia inaequalis*. Curr Genet. 32(5):361-366.

Zhong J, Karberg M, Lambowitz AM. 2003. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. Nucleic Acids Res. 31(6):1656-1664.

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(Database issue):D31-D37.

Zimmer M, Krabusch M, Wolf K. 1991. Characterization of a novel open reading frame, urf a, in the mitochondrial genome of fission yeast: correlation of urf a mutations with a mitochondrial mutator phenotype and a possible role of frameshifting in urf a expression. Curr Genet. 19(2):95-102.

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

Zinn AR, Pohlman JK, Perlman PS, Butow RA. 1988. *In vivo* double-strand breaks occur at recombinogenic G + C-rich sequences in the yeast mitochondrial genome. Proc Natl Acad Sci U S A. 85(8):2686-2690.

Zivanovic Y, Wincker P, Vacherie B, Bolotin-Fukuhara M, Fukuhara H. 2005. Complete nucleotide sequence of the mitochondrial DNA from *Kluyveromyces lactis*. FEMS Yeast Res. 5(4-5):315-322.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31(13):3406-3015.