Mitochondrial genomes and concerted evolution in *Ceratocystis*

by

Kershney Naidoo

(MSc, University of Kwa-Zulu Natal)

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

in the Faculty of Natural and Agricultural Sciences, Department of Genetics and the Forestry and Agricultural Biotechnology Institute at the University of Pretoria

June 2013

Supervisor: Prof. Brenda D. Wingfield

Co-supervisors: Prof. Emma T. Steenkamp

Prof. Michael J. Wingfield

Declaration

I, the undersigned, hereby declare that the thesis submitted herewith for the degree *Philosophiae Doctor* to the University of Pretoria, contains my own independent work and has hitherto not been submitted for any degree at any other University.

.....

Ms Kershney Naidoo

June, 2013

"It always seems impossible until it's done"

Nelson Mandela

Table of Contents

Acknowledgements 1

Preface 4

Chapter One

Literature Review: Mitochondrial genomes in the Ascomycota 8

1.1 Introduction 9

1.2 General Characteristics of Mitochondrial Genomes 10

1.2.1 Mitochondrial Genomes of Fungi 11

1.2.1.1 Gene architecture and content 11
1.2.1.2 Genome size 12
1.2.1.3 Genetic code 13
1.2.1.4 Transport RNAs 14
1.2.1.5 Comparisons among the mitochondrial genomes of ascomycetes 15

1.3 Genome Sequencing 17
1.3.1 Mitochondrial Genomes 18
1.3.2 Mitochondrial Genome Assembly and Annotation 19
1.3.2.1 Assembly 19
1.3.2.2 Annotation 20
1.3.3 Mitochondrial genomes for phylogenetic inference and fungal systematics 20

1.4. Aims and rationale of this thesis 21

1.5. References 24

Figures 35

Chapter Two

Annotation and comparative analysis of three *Ceratocystis* mitochondrial genomes 47

Abstract 48

2.1. Introduction 49

2.2 Materials and methods 51
2.2.1 Fungal isolates and DNA isolation 51
2.2.2 Next generation DNA Sequencing 51
2.2.3 Assembly, analysis and annotation of sequence data 51

2.3. Results 53
2.3.1 Sequencing data 53
2.3.2 Annotation of each mitochondrial genome 53
2.3.2.1 *Ceratocystis fimbriata sensu stricto* 53
2.3.2.2 *Ceratocystis albifundus* 54
2.3.2.3 *Ceratocystis moniliformis* 55

2.4. Discussion 56

2.5. References 61

Tables 67
Figures 77
Supplementary Data – Figures 95

Chapter Three

Mitochondrial genome variation within *Ceratocystis albifundus* 96

Abstract 97

3.1. Introduction 98

3.2. Materials and methods 100
3.2.1 Fungal isolates and DNA isolation 100
3.2.2 Next generation DNA Sequencing 101
3.2.3 Sequence data assembly, analysis and annotation 101

3.3. Results 102
3.3.1 Sequencing data 102
3.3.2 Annotation of all *C. albifundus* mitochondrial genomes 102

3.4. Discussion 105

3.5. References 109

Tables 115
Figures 118

Chapter Four

Concerted evolution in the ribosomal RNA cistron 128

Abstract 129

4.1. Introduction 130

4.2. Materials and methods 132
4.2.1 Fungal Isolates 132
4.2.2 ITS PCR and sequencing 133
4.2.3 ITS PCR-RFLPs 134
4.2.4 Statistical Analysis 134

4.3. Results 135
4.3.1 Scoring of the ITS types 135
4.3.2 Statistical analysis of the ITS types 136

4.4. Discussion 137

4.5. References 140

Tables 145
Supplementary Table 148
Figures 149
Supplementary Data – Figures 155

Conclusions and Future Prospects 161

Summary 166

Supplementary Data CD

Acknowledgements

I would like to express my sincere appreciation to the following institutions and people who have been instrumental in assisting me with my thesis:

The provision of world-class research facilities from members of the Forestry and Agricultural Biotechnology Institution (FABI), Tree Protection Co-operative Programme (TPCP), the Centre for Tree Health Biotechnology (CTHB) and the Department of Genetics, University of Pretoria. I am very fortunate to be member of such an amazing research team. The Department of Science and Technology (DST) / National Research Foundation (NRF), the Centre of Excellence in Tree Health Biotechnology, University of Pretoria Mentorship award, as well as the Improving Academic Qualifications (IAQ) NRF Sabbatical grant is also acknowledged for their financial help.

My supervisors, Prof. Brenda Wingfield, Prof. Mike Wingfield and Prof. Emma Steenkamp for their dedication, mentorship, enthusiasm and critical review of my thesis. Their invaluable advice and encouragement has been truly inspirational to me and without their support, this thesis would not have been attainable. Prof. Brenda Wingfield, my primary supervisor, I cannot express how precious your guidance and mentorship has been to me. Your passion for science and the development of my academic career is truly appreciated. I hope to one day emulate you not only as the remarkable scientist that you are, but also possess your amazing leadership qualities that motivated me.

The Department of Genetics, both staff and students for their encouragement and support in the completion of my thesis. Prof. Paulette Bloomer, I thank you for your faith in me and inspiration that if I truly believe I can conquer any situation. To all my 6th floor laboratory colleagues, I have thoroughly enjoyed the wonderful environment we created. This made science truly fun.

Prof. Bernard Slippers, Dr. Martin Coetzee (co-author on the published work of Chapter Four), Dr. Albe van der Merwe and Mr. Simon Martin for their thought provoking conversations on concerted evolution as discussed in Chapter Four of this thesis. Special mention is made of Prof. Carl Roux and Prof. Jaco Greeff for their statistical assistance. Mrs. Glenda Brits is also thanked for her amazing help with the illustrations with this research chapter.

To the entire FABI family, I thank you for the many friendships that I have made; the memories will always be cherished. The moral support from every person from FABI, both staff and students, is truly precious to me. To Mrs. Jenny Hale, I will always remember how instrumental your help was at my initial move to Pretoria in 2006.

To my dear and special friends, I am so blessed to have such amazing people in my life. You have become like family to me. Mrs. Vivienne Clarence, you are my second mother, your support both moral and emotional is a treasure to me. To my best friend, Mr. Jaco Mare, you are an exceptionally dear friend, and I thank you for been an unwavering pillar of encouragement to me. Thank you for all the great moments that we shared may there be many more! To Dr. Pamela de Waal, I thank you for your mentorship and guidance with lecturing, but most importantly I am grateful for the wonderful friendship that we share. Working with you, has and always will be such a pleasure. Mr. Quentin Santana, thank you for your friendship, in times of need, I knew I could always count on you. Dr. Wubetu Bihon, it will always be a pleasure sharing a lab with you. Mrs. Hester Neethling and Ms. Ina Goosen, you both have been a wonderful support system to me, thank you for the many shared laughs and emotional guidance. To the many other dear friends, Mr. Van Zyl Spies, Ms. Louise le Roux (and all the Alex house mates), Mrs. Dominique Christian, Mr. Willie van Zyl, Ms. Phia van Collar, Mr. Nayan Joshi, Dr. Martin Coetzee and Dr. Jan Perold, I thank you all for the many fun times shared. My life is richer for knowing you all.

Finally to my family whose unconditionally love, encouragement and support has been my guiding light. To my mummy, you are my rock, your love and belief in me is truly amazing. I thank my daddy for always encouraging me to be cool, calm and collected. May I always make you both proud parents! To my sister, Ms. Shivani Naidoo, been your sister was God's greatest gift to me. You are the best friend and sister I could have ever wanted. Our bond transcends that of sisters, you are the one person that I know will always be there for me. Thank you.

To my Swami, Sri Satha Sai Baba, my faith in myself may have been tested at times but the strength from prayer to you always allowed me to climb any mountain. I live by your motto "Love all, Serve all, Help ever, Hurt never"

Preface

The advancement of next generation sequencing platforms has significantly improved the cost and efficiency of producing DNA sequence data. Complete characterizations of genomes including whole mitochondrial genomes are growing exponentially in many fields of biology. Mitochondria have organellar genomes which are small in comparison to their nuclear counterparts, making the study of their genomes potentially easier and providing a unique window into the life style of less studied organisms. Despite the increasing numbers of available genomes not many fully annotated and characterized Ascomycete fungal mitochondrial genomes are available. Within the order Microascales, no mitochondrial genomes were available at the time this study was initiated.

The **first chapter** of the thesis provides a broad literature overview of mitochondrial genomes within the Ascomycota. Aspects of mitochondrial biology are discussed, the genome size, gene architecture, structural signatures, genetic codon usage, transport RNA molecules, presence of introns and homing endonucleases and presence of plasmid elements. Comparisons of the available annotated Ascomycete mitochondrial genomes are also presented. This review also briefly discusses next generation sequencing platforms, with specific emphasis on how to manipulate DNA sequence generated from each technology in order to assemble and annotate fungal mitochondrial genomes. With the great variation observed in Ascomycete fungal mitochondrial genomes, the study of their genomes is a valuable tool for the identification of mitochondrial markers for species differentiation and the evaluation of their current evolutionary history.

The **second chapter** focuses on the complete characterization, annotation and comparative analysis of the mitochondrial genomes of three species within the genus *Ceratocystis*. This genus includes a number of species complexes and the differentiation of the species within these complexes are still controversial as the resolution provided by nuclear gene sequences has proved to be inconclusive. This genus includes a number of significant tree pathogens, thus understanding the classification of these species is an essential part of the disease management of these fungi. For this study two well-known pathogenic species namely *C. fimbriata* and *C. albifundus* were selected. The third species included was the saprophytic species *C. moniliformis*. The longer term aim of this research is to identify mitochondrial markers which could aid in the species delimitation, making use of the comparative analyses

of their mitochondrial genomes. Full characterization of these genomes is thus the first step in achieving this goal.

The most significant pathogen observed in *Acacia mearnsii* plantations is the wilt pathogen *C. albifundus*. This species is thought to be native to Southern Africa; however no concrete evidence exists as to the species' country of origin. As a consequence of the unidirectional inheritance of mitochondrial genomes it is possible to use them to investigate the pathways of pathogen invasion. **Chapter three** of this thesis investigates the variation in the mitochondrial genomes for six *C. albifundus* isolates from different geographical locations. The results of this study provides a platform for the identification of mitochondrial gene regions which can potentially allow for the development of markers to trace the movement of *C. albifundus* from native trees onto non-native plantation species and could also provide clues as to the centre of origin of this pathogen.

A chance discovery of the variation of the rRNA cistron allowed for a study on concerted evolution in C. *manginecans*. *Ceratocystis manginecans* was found to have two distinct sequences within the internally transcribed spacer (ITS) region of the ribosomal RNA cistron. As is the case with many *Ceratocystis* species *C. manginecans* it is homothallic. The variation within the rRNA cistron and the homothallism enabled us to investigate whether homogenization or a drift-like variation of the rRNA cistron occurred within its life stages by investigating the fluctuations in the two different ITS types in *C. manginecans*. The process of gene conversion and unequal crossing over as potential sources of concerted evolution were addressed in **chapter four**.

This thesis presents the first comprehensive study to date focusing on the mitochondrial genomes of a number of species within the genus *Ceratocystis*. Next generation sequencing platforms were employed in order to generate DNA sequence which could be assembled, annotated and characterized for selected mitochondrial genomes of *Ceratocystis* species. In addition variation in the rRNA cistron in a single species of *Ceratocystis* was investigated. The research chapters of the thesis are presented as stand-alone units [the last chapter has recently appeared in PLOS ONE] and some duplication between chapters has thus been

inevitable. The last section of the thesis provides the research highlights as well as some areas of future research.

Chapter One

Literature Review

Mitochondrial genomes in the Ascomycota

1.1 Introduction

Early in the 19th century, scientists' focus was drawn to the small rod-like organelle implicated in eukaryotic protein synthesis and energy metabolism. Officially, this organelle was described in the late 1800's by Carl Benda, a microbiologist, who defined it in terms of the Greek language, where *mitos* means "thread" and *khondrion* is "little granule" (Scheffler, 2001). Mitochondrial research has generally fascinated scientists globally, particularly so in the last century when it was shown that these organelles have their own DNA genome (Scheffler, 2001). These studies have also shown that the mitochondrion is a remnant of one or more endosymbiosis events between the ancestral eukaryotes and the bacterial class α-Proteobacteria (Bullerwell & Lang, 2005).

The mitochondrion serves as a powerhouse for energy production, fundamental to all biochemical and metabolic processes in the cell. This vital organelle is reserved throughout the cell cycle by constant replication, and is thus inherited by the daughter cells as intact copies upon cell division. Mitochondria contain the genes for all oxidative energy production processes (Yaffe, 1999), but are also involved in a surfeit of other cellular processes. These include ageing and apoptosis, the synthesis of iron-sulphur clusters, and the maintenance of calcium within the cell (Richardson *et al.*, 2010). All of these processes are crucial to the survival and functionality of the cell and thus, understanding the repertoire of genes contained within these organelles, their characteristics and mode of inheritance is significant.

Advances in the domain of biological information technology have enabled scientists to explore and answer questions once thought impossible. Furthermore, the ease with which DNA sequence data can be generated has meant that researchers are no longer restricted to studying only one or a few genes but that it is now possible to relatively easily generate data for whole genomes, including those of the mitochondria. Thus, more than 2000 sequenced mitochondrial genomes were available from the NCBI repository in 2012. Of these, 75 are fungal and 53 are of ascomycetes (Phylum Ascomycota, Kingdom Fungi), which are the focus of this review.

The purpose of this review is to critically address the literature relating to mitochondrial genomes across eukaryotes and with a particular focus on the Ascomycota. The various characteristics of mitochondrial genomes are discussed and compared in terms of gene architecture, gene content, size and genetic code. Furthermore, this review considers the impact of genome sequencing on our understanding of mitochondria, giving special attention

to genome assembly and annotation. Finally, the review briefly considers the potential uses of mitochondrial markers for phylogenetic inference and here the focus is on fungi in the genus *Ceratocystis*, which is the study organism of this thesis.

1.2. General Characteristics of Mitochondrial Genomes

Mitochondrial genomes are typically characterized as homoplasmic circular entities (Barr *et al.*, 2005), although there are some exceptions where these genomes are linear, for example, the slime mould *Physarum polycephalum* (Kawano *et al.*, 1982). These genomes are typically found in high copy number and are often conserved through alterations in the genetic code (Sammet *et al.*, 2010; Santos *et al.*, 2011; Santos *et al.*, 2004). However, the genome sizes of mitochondrial genomes differ among groups of eukaryotes; animal mitochondrial genomes are relatively small, while those of fungi are larger and the largest genomes are found in plants (Logan, 2003).

All mitochondrial genomes essentially contain the same suite of genes and these encode the subunits of proteins responsible for energy generation (Taylor, 1986). As such, mitochondria dictate a range of biological processes; the most fundamental being electron transport and ATP synthesis, translation and RNA processing, the synthesis of iron-sulphur clusters as well as calcium homeostasis (Bullerwell & Lang, 2005). Metabolism of lipids, nucleotides, amino acids and carbohydrates are also part of the functional roles of mitochondria in conjunction with DNA replication and repair (Logan, 2003). Certain mitochondrial genes are, however, encoded by nuclear genes and the products of these are then imported by translocase complex subunits into the mitochondrion (Adams *et al.*, 2000; Bullerwell & Lang, 2005; Taylor, 1986).

The mitochondrion is responsible for many biological processes. A number of these processes are encoded for by the nucleus [Figure 1] (Burger *et al.*, 2003). Involved in these processes are mitochondrial subunits common across all organisms; however there are specialized subunits specific to that of eukaryote organisms, often encoded by genes of a nuclear origin (Burger *et al.*, 2012). The various cytochrome *C* complexes (I-V) are directly involved in electron transportation and oxidative phosphorylation [Figure 1]. TIM translocases are implicated in protein import and insertion into the inner membrane, whilst

Sec-, Tat- and Oxa1-translocases are involved in protein export from the matrix into the inner membrane. The RNase P ribozyme is responsible for processing of the 5 prime ends of the tRNAs [Figure 1] (Burger *et al.*, 2003).

Eukaryotes, including fungi, have evolved mechanisms that allow them to maintain a homoplastic state in their mitochondrial genomes. Thus their zygotic progeny contains a single mitochondrial genotype (Basse, 2010). This allows the mitochondrial genome of an individual to be viewed as an intact copy of the genetic components of the maternal mitochondrion (Burger *et al.*, 2003). These generalizations are largely due to the apparent absence of genetic recombination among mitochondrial genomes, and also their uniparental mode of inheritance that allows for one of the parental mitochondria (in fungi this is typically the paternal one) to be eliminated in the zygote (Chen & Hebert, 1999). Furthermore, mitochondrial genomes are commonly viewed as small co-evolving genomes in which the relative evolutionary rates are independent of those for their nuclear counterparts (Ballard & Whitlock, 2004).

Mitochondrial genomics has emerged as a useful tool aiding in the study of evolutionary biology and systematics of eukaryotes. All of these issues have, however, received comprehensive attention in the scientific literature and will not be dealt with here. For example, Basse (2010), Lang (1999) and Barr (2005) reviewed the different modes of inheritance and maintenance of the genetic material of mitochondria. Takano (2010) detailed the current understanding of the evolution of mitochondria, while Atkins (2004) and Bruns (1991) reviewed the use of mitochondrial genetic information in the study of the evolutionary biology and systematics of eukaryotes (Atkins & Clark, 2004; Barr *et al.*, 2005; Basse, 2010; Bruns *et al.*, 1991; Lang *et al.*, 1999; Takano *et al.*, 2010).

1.2.1. Mitochondrial Genomes of Fungi

1.2.1.1 Gene architecture and content

Fungal mitochondrial genomes are diverse in their size and structure. They can be circular or linear in nature as was seen in early studies of the oomycete *Pythium oligandrum* and the ascomycete *Cephalosporium acremonium* (Bendich, 1993; Forget *et al.*, 2002; Rycovska *et al.*, 2004; Torriani *et al.*, 2008). In addition, high adenine and thymine content is prevalent in

these genomes along with an absence of methylation and conserved gene functionality (Campbell *et al.*, 1999; Duchene *et al.*, 2009). The increased AT content within the mitochondrial genome results in a bias in the genetic codon usage. These AT rich regions pose a difficulty when using computational software to assemble the sequence data as it increases the homoplasy (Shendure & Hanlee, 2008; Voelkerding *et al.*, 2009).

The mitochondrial genomes of ascomycete fungi typically include 14 protein coding gene regions, two ribosomal RNA (rRNA) genes and approximately 22 to 25 transport RNA (tRNA) genes (Pantou *et al.*, 2008). The genes involved in electron transport and phosphorylative processes producing ATP, include cytochrome *c* oxidase complex 1, 2 and 3; apocytochrome *b*, and ATPases. The standard repertoire of genes found in the mitochondrial genome include the NADH dehydrogenase complex 1, 2, 3, 4, 4L, 5, and 6; ATPase subunits 6, 8 and 9. In some cases, the ATPase 9 is found in both the nuclear and mitochondrial genomes of organisms (Hausner, 2003; Taylor, 1986), although it can also be absent from the mitochondrial genome, e.g., *Podospora anserina* (Kouvelis *et al.*, 2004).

1.2.1.2 Genome size

As mentioned above, fungal mitochondrial genomes are usually smaller than those found in plants, but larger than those in animals and also differ significantly from those of numerous protozoa (Burger *et al.*, 2003). Fungal mitochondrial genomes vary in size from approximately 24 kb to 100 kb (Hausner, 2003). In comparison the *Plasmodium* mitochondrial genome is only 6 kb in size, in contrast to for example, the rice mitochondrial genome that is approximately 80 times larger, at 490 kb (Notsu *et al.*, 2002). Some plant mitochondrial genomes are even larger, for example the cucurbit mitochondrial genome is approximately 3000 kb in size (Burger *et al.*, 2012). This size difference can primarily be attributed to the varying sizes and organization of intergenic regions (Scheffler, 2001) and the presence of mobile elements or introns that contain homing endonuclease genes. With current sequencing technologies constantly improving, contrasting mitochondrial architectures will continue to be undertaken and these studies will highlight exceptions to typical gene architectures and genome sizes currently known.

Many noteworthy literary reviews have addressed the great diversity and variation seen among mitochondrial genome sizes and composition across a range of eukaryotes, including

ascomycetes (Burger *et al.*, 2003; Gray *et al.*, 1999; Gray *et al.*, 1998; Lang *et al.*, 1999; Osiewacz, 2002; Taylor, 1986). No direct correlation has been established between the size of the mitochondrion and its gene content (Burger *et al.*, 2003). Rather it has been shown that the most significant differences can be attributed to variations described in the intergenic regions [non-coding regions], where the length and organization of repeat sequences often impacts on the architecture of a species mitochondrial genome [Figure 2] (Burger *et al.*, 2003; Hamari *et al.*, 2002). The presence of introns containing homing endonucleases which act as selfish mobile elements can substantially increase the size of the genome. Such mobile elements have also been implicated in recombinational events, introducing plasmids into the mitochondrial genomes (Bonen & Vogel, 2001; Hamari *et al.*, 2002; Hausner, 2003).

The number of shared common mitochondrial genes present in *Saccharomyces* is low [Figure 2] in comparison to the other mitochondrial genomes across the eukaryotes. Therefore, the role of mitochondrial-nuclear interactions is vital to cellular survival, as these species often transport necessary gene products from the nucleus into the energy driving organelles (Gray *et al.*, 1999). In some cases, it has been shown that organisms such as petite mutants of *S. cerevisiae* can exist without mitochondria, purely because their nuclei contain their mitochondrial genes (Contamine & Picard, 2000). It is worthy to mention that mitochondrial genomes may also contain unique ORFs not usually found within the genome, coding for genes specific to the organism (Gray *et al.*, 1999; Lane, 2009). However, the frequency for which this occurs is low in most organisms [Figure 2] (Burger *et al.*, 2003).

1.2.1.2 Genetic code

The genetic code determines how codons specify amino acids during translation or protein synthesis. Although there are many commonalities, the genetic code for mitochondria is typically different from those of nuclear genomes (Sammet *et al.*, 2010). In general, the genetic code is somewhat redundant in that multiple codons may code for the same amino acid. The wobble hypothesis (Sammet *et al.*, 2010; Santos *et al.*, 2011; Santos *et al.*, 2004; Silva *et al.*, 2004), suggests that codons can differ by a single nucleotide to code for the same or very similar amino acid, because the nucleotides forming the anticodon at position 34-36 within a tRNA molecule are subject to modification (Moura *et al.*, 2010). If the codon/anticodon recognition is compromised, it directly impacts on the post-translational editing of tRNA molecule (Duchene *et al.*, 2009). Watson and Crick (Crick, 1966)

determined that a minimum of 32 distinct tRNAs are all that is needed for complete recognition of all the codons in species using the standard genetic code (Lang *et al.*, 2012). However, mitochondria (and chloroplasts) have restrictions in terms of functionality and therefore do not always encode all 32 expected tRNAs.

As is true for most eukaryotes, fungal mitochondria use a genetic code that is slightly different from the so-called "universal" code (Santos *et al.*, 2011). The fungal mitochondrial code (i.e., Translation Table 4; Figure 3) is the same as that of Protozoa, Mycoplasma/Spiroplasma and Coelenterates (Lang *et al.*, 2012). The unusual nature of this particular mitochondrial code supports the fact that both the sense and non-sense codons can interchange in their identity. This is in contrast to nuclear genomes where only the stop codons appear to be able to change their identity and code for an amino acid (Bonitz *et al.*, 1980).

The mitochondrial exceptions in comparison to the universal code are shown where variants in the last codon position encode a different amino acid [Figure 3]. In many *Candida* species, the mitochondrial genome has CUG which codes for Serine instead of Leucine (Silva *et al.*, 2004), emphasizing the loss of stringency in codons which begin with cytosine. Changes in the genetic codes for the nuclear genomes of eukaryotic as well as the prokaryotic organisms can be linked to an alternative coding for their mitochondrial counterparts [Figure 3] (Silva *et al.*, 2004). Within the mitochondria, certain codons are more prone to changes in their identity, especially codons which begin with the adenine (A) or thymine (T) nucleotides whilst codons beginning with cytosine (C) are apparently less likely to change their identities (Santos *et al.*, 2011). Codons starting with guanine (G) are somehow resistant to changes in their identities. This implies that the strength of the codon-anticodon recognition system is most significant at the first codon position, thus important to the overall evolution of the non-standard genetic codes (Moura *et al.*, 2010).

1.2.1.3 Transport RNAs

A significant feature of mitochondrial genomes of ascomycete fungi is the variation in their tRNA composition and the number of tRNA genes they encode. Together with aminoacyl tRNA synthetases, tRNAs are essential for protein synthesis and a full complement of these units is thus required for the process of translation (Duchene *et al.*, 2009). Any deficiencies

in either aminoacyl tRNA synthetases or tRNAs species would thus result in erroneous translation, which in turn can have dire consequences such as an increased mutational rate (Hopper & Phizicky, 2003). Thus, it is essential that all tRNAs are encoded in the mitochondrial genome, or that these are "outsourced" from the nuclear genomes of an organism to ensure that all protein products are functionally translated.

Saccharomyces cerevisiae appears to have a sufficient number of tRNAs in the mitochondria to drive translation, which is unlike *Trypanosmona* or *Leishmania* species that lack tRNAs from their mitochondrial genomes completely (Hopper & Phizicky, 2003). However, in some *S. cerevisiae* isolates there are three tRNAs that are imported from the nucleus (Duchene *et al.*, 2009). These are tRNA lysine (CUU) and two tRNA glycines [Figure 3], although the full complement of 23 tRNAs is also synthesized directly within the mitochondrion [Figure 4]. It has been shown that the priori excessive nuclear $tRNA^{Lys}$ plays a vital role in the conditional environmental adaptation of mitochondrial protein synthesis when there was fluctuations in the temperature at which *S. cerevisiae* grow (Kamenski *et al.*, 2007). Imported tRNAs from the nucleus vary in both the number as well as the type of tRNA, depending largely on the particular organism, although importation of tRNAs from the nucleus is not an uncommon feature across all eukaryotes.

1.2.1.4 Comparisons among the mitochondrial genomes of ascomycetes

The gene content and organization of the mitochondrial genomes of ascomycete fungi is extremely variable [Figure 5] (Cummings *et al.*, 1990; Juhasz *et al.*, 2008; Kerscher *et al.*, 2001; Kouvelis *et al.*, 2004; Pantou *et al.*, 2006; Pantou *et al.*, 2008; Torriani *et al.*, 2008; Woo *et al.*, 2003; Wu *et al.*, 2009; Zivanovic *et al.*, 2005). Shown diagrammatically in Figure 5, the largest genome annotated is *P. anserina*, with a size of 100 314bp. It houses the standard cohort of 14 mitochondrial genes, with only *atp9* absent. Comparing this mitochondrial genome with those of other ascomycetes, suggests that much variation can be attributed to exceptionally large intragenic regions. For example, the *Kluyeromyces latis* mitochondrial genome with its eight genes and large intergenic regions is 40 219bp in size, which accounts for the fact that it is 40% smaller than that in *P. anserina*. Interestingly though, not all genomes are necessarily characterized by large intragenic regions. The mitochondrial genome of *Lecanicillium muscarium*, for example, is only 24 499bp in size, despite the fact that it contains all 14 of the essential genes. The standard cohort of essential

genes is mostly present across the spectrum of ascomycetes. The only exception currently known is in *K. latis*, which completely lacks any of the NADH genes. This is because these genes are encoded on the nuclear genome, and after translation, their protein products are imported into the mitochondrion (Zivanovic *et al.*, 2005). This is also the case for the missing *atp9* gene in the *P. anserina* mitochondrial genome (Cummings *et al.*, 1990).

The mitochondrial genes generally occur in the same order, although there is some variation with certain genes being more commonly linked. Such gene pairs include *nad4L* and *nad5*, *coxIII* and *nad6*, *atp8* and *atp6*, and *nad2* and *nad3*. The *nad4L* and *nad5* genes generally occur as a pair in *Aspergillus niger*, *Penicillium marneffei*¸ *Trichophyton rubrum*, *Lecanicillium muscarium* and *Podospora anserina* (Cummings *et al.*, 1990; Juhasz *et al.*, 2008; Kouvelis *et al.*, 2004; Woo *et al.*, 2003; Zivanovic *et al.*, 2005). These genes are also paired in *Mycosphaerella graminicola*, although they have been inverted relative to the other genes (Torriani *et al.*, 2008). The *coxIII* and *nad6* genes are linked in *A. niger*, *P. marneffei*, *T. rubrum*, *L. muscarium*, *V. dahlia*, *P. anserina* and *Fusarium oxysporium* (Cummings *et al.*, 1990; Cunnington, 2007; Juhasz *et al.*, 2008; Kouvelis *et al.*, 2004; Pantou *et al.*, 2006; Pantou *et al.*, 2008; Zivanovic *et al.*, 2005). The exceptions are in the mitochondrial genomes of *M. graminicola*, *Y. lipolytica* and *K. lactis* (Kerscher *et al.*, 2001; Torriani *et al.*, 2008). The *atp8* and *atp6* genes are linked in *A. niger*, *P. marneffei*¸ *M. graminicola*, *L. muscarium*, *Y. lipolica*, *V. dahlia*, *Y. lipolytica* and *F. oxysporium*, with the only exceptions *P. anserina* and *T. rubrum* (Cummings *et al.*, 1990; Cunnington, 2007; Kerscher *et al.*, 2001; Kouvelis *et al.*, 2004; Pantou *et al.*, 2006; Torriani *et al.*, 2008; Zivanovic *et al.*, 2005). The *nad2* and *nad3* genes are paired in *M. graminicola*, *L. muscarium*, *Y. lipolytica*, *V. dahlia*, *P. anserina* and *F. oxysporium*, while it is interrupted in *A. niger*, *P. marneffei*¸ *T. rubrum* and *K. lactis*. For all these exceptions, it is possible that inverted recombinational events could have disrupted the linkage between genes (Barr *et al.*, 2005). However, this could very well also be due to mis-assemblies and errors arising from sequencing assembly software analysis programmes (Metzker, 2010).

To some extent, mitochondrial genome size is dependent on the presence of introns. In plant and fungal mitochondrial genomes, introns as large as 5 kilo base pairs in size have been documented (Ferandon *et al.*, 2010). In ascomycetes [Figure 5], some species such as *Y. lypolytica* (47 916bp) and *P. anserina* have many introns, while those of *T. rubrum* (26 985bp) and *L. muscarium* have only one and those of *M. graminicola* (43 964bp) and *V. dahliae* (27 985bp) are completely void of introns (Cummings *et al.*, 1990; Joardar *et al.*,

2012; Kerscher *et al.*, 2001; Kouvelis *et al.*, 2004; Pantou *et al.*, 2006; Torriani *et al.*, 2008). Amongst all the described ascomycete mitochondrial genomes, the *coxI* gene of *P. anserina* has the largest repository of introns; it contains up to 16 introns, accounting for nearly 25% of the total genome size (Cummings *et al.*, 1990). Overall, it appears that the *cox* genes are most intron dense across ascomycetes, where they significantly impact on the size of the genome; e.g. the *coxI* gene itself ranges from 1550bp to as large as 24 000bp (Ferandon *et al.*, 2010). Thus it was established by Burger *et al*. (2003) that the number of genes and the size of the mtDNA have no significant correlation. Moreover it is now widely accepted that the high frequency and variation in size of introns, sometimes in combination with plasmids, nuclear and mitochondrial genomes contributes considerably to the size polymorphisms seen in fungal mitochondrial DNA's (Burger *et al.*, 2003).

The introns in mitochondrial genomes are classified as either Group I or Group II introns, and both forms can be present in the same species and sometimes even contained within the same gene (Hamari *et al.*, 2002; Hausner, 2003). Characteristic of Group I introns is the presence of a homing endonuclease gene that encodes a DNA endonuclease, which is responsible for the cleavage and transfer of the site-specific integration of the particular intron (Ferandon *et al.*, 2010). Group II introns are characterized as self-catalytic ribozymes, often found within genes on transposons or insertional elements as plasmids, as these locations aid in the introns mobility within the genome (Bonen & Vogel, 2001). In fungal mitochondrial genomes, Group I introns are found at a higher frequency than Group II introns (Hausner, 2003). The *P. anserina* the *coxI* gene houses 15 Group I introns and one Group II intron (Cummings *et al.*, 1990). The occurrence of both types of introns in the *cox1* gene has also been shown in the button mushroom, *Agarius bisporus*, which has not only the longest but also the largest mitochondrial gene in its *coxI* gene where it contains 18 Group I introns and one Group II intron (Ferandon *et al.*, 2010). It is for this reason that the *coxI* gene, which is popular as a DNA barcoding marker in animals (Boore & Brown, 1998; Likic *et al.*, 2010; Santamaria *et al.*, 2009), is too polymorphic to make it equivalently useful for fungal taxonomic classification purposes (Schoch *et al.*, 2009).

1.3. Genome Sequencing

Advancing technology has substantially pushed the boundaries in DNA sequence generation. Since the early 1960s when the first 10 bases of DNA sequence were published (Wu &

Kaiser, 1968), constant strides have been made to improve not only the technology but also the type of research questions that can be investigated (Ma & Fedorova, 2011; Shendure & Hanlee, 2008; Voelkerding *et al.*, 2009). Advances in sequencing technologies, from those considering single gene regions to complex complete genomes, have allowed an increasing number of new questions to be answered using comparative genomics.

Genomics is a growing field and data mining combined with powerful bioinfomatic platforms is the driving factor in the generation of annotated genomes for whole genomes as well as mitochondria (Haridass *et al.*, 2011). The emergence of improved sequencing technologies and assembly software programs, sequencing and annotation of genomes has become increasingly available to biologists. Bacterial genomes, due to their small size, have been the focus of much research in this field, but with recent developments in high throughput sequencing, whole fungal genomes are receiving increasing attention. The development of functional and comparative genomic studies on plant pathogenic fungi holds a vast potential to further enhance the pool of knowledge regarding both taxonomic questions and the various biochemical and molecular pathways that drive host-pathogen interactions.

1.3.1 Mitochondrial Genomes

Records show that nearly 600 complete mitochondrial genomes of metazoans are available on GenBank (Pantou *et al.*, 2008). A total of 97 fungal mitochondrial genomes could be found on the NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) in a search undertaken for the present review in January 2013. For the *Dikarya,* there are a total of 83 records of fully sequenced mitochondrial genomes and of these 75 are for Ascomycota and only 11 are for the Basidiomycota. Only 50 fully annotated mitochondrial genomes for fungi are accessible, most belonging to the Ascomycota. These mitochondrial genomes vary in size from as small as 18 kb to genomes larger than 100 kb and most contain the standard cohort of genes (Bullerwell & Lang, 2005).

Classical descriptions of sequenced mitochondrial genomes include the annotation of the gene products of the core proteins found within the inner mitochondria. These typically also detail their involvement in the electron transport system as well as the oxidative phosphorylation processes (Burger *et al.*, 2003). In addition to the standard genes, an additional subunit of the ATPase complex (*atp*9) as well as a ribosomal protein gene (*rps3*) is

found in fungal mitochondrial genomes (Bullerwell & Lang, 2005). Although these annotations typically include a minimalistic set of tRNAs, it is widely recognized that the absent tRNA genes are often found in the nuclear genome, from where their products are imported into mitochondria (Bullerwell & Lang, 2005).

Fungal genomes comprise an assortment of autonomously-self-replicating plasmids (Hausner, 2003). A prevalent area of research focuses on establishing evidence for plasmid integration into the mitochondrial genomes (Cahan & Kennell, 2005). To date, there is insufficient information implicating the presence of genetic material into the mitochondrial genome via horizontal transfer from these plasmids (Burger *et al.*, 2003).

The use of fully sequenced genomes aids in the identification of DNA recombination events, which can be linked to the appearance of the first eukaryotes (Schluter *et al.*, 2011). Taxonomic reclassification has received much attention, with phylogenomics used to define species positioning within the tree of life (Capella-Gutierrez *et al.*, 2012). Often, comparative phylogenomics is carried out to identify unique characteristics of species (Liu *et al.*, 2009). Such characteristics consider genes associated with RNA processing specifically proteins that enable the splicing of Group I or Group II introns. These maturases often feature the protein coding domains found in the intron itself and also include endonucleases that allow intron mobility (Chevalier & Stoddard, 2001). Gene rearrangements in mitochondrial genomes thus provide effective tools to study ancient phylogenetic relationships among species, because the rearrangements can address both evolutionary as well as mechanisms driving such rearrangements (Boore & Brown, 1998). There is also a trend to move towards whole small genome interrogation rather than homologous single gene comparisons in organisms to infer phylogenetic distances based on gene orders for particular species (Sankoff *et al.*, 1992). Mitochondrial genome research has also been exploited in terms of the variation and selective pressures acting on their GC content due to their elevated AT rich genomes (Clare *et al.*, 2008).

1.3.2 Mitochondrial Genome Assembly and Annotation

1.3.2.1 Assembly

DNA sequence assembly and annotation process take the form of an annotation pipeline incorporating many bioinfomatic tools (Ma & Fedorova, 2011) as is illustrated in Figure 6.

Once the sequence reads are obtained they are assembled into larger contigs, which in turn serve as the building scaffolds for the generation of supercontigs. The latter step is directly dependant on the availability of paired-end read libraries to ensure that orientation of the larger continuous fragment is correct [Figure 6]. The generation of a precise genome assembly is influenced by the accurate alignment of the reads in the correct genomic locality in combination with the alignment of these contigs on the correct scaffold. A common problem facing researchers in the assembly process is the misplacement of sequence reads, which results from repetitive sequences and sequence errors generated via the sequencing process itself (Metzker, 2010).

1.3.2.2 Annotation

Initial annotation of the total DNA sequence obtained from a next generation sequencing platform, after *de-novo* assembly makes use of open reading frame predictive software. The NCBI offers users their freeware prediction tool using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to identify ORFs. For ascomycete fungi, such ORF predictions are based on the Mold mitochondrial genetic code. Genes encoding functional RNAs (e.g., rRNA) are identified using BLAST (Altschul *et al.*, 1997) or software that incorporates RNA secondary structure (e.g., RNase P using http://rna.tbi.univie.ac.at/bchek (Yusuf *et al.*, 2010)). The identification of tRNA genes is facilitated by tRNAscan-SE (Lowe & Eddy, 1997) in conjunction with available sequences from the NCBI database. All sequence alignments can be analysed using computational software programs such as the CLC genomics workbench (CLC Bio, Århus, Denmark).

1.3.3 Mitochondrial genomes for phylogenetic inference and fungal systematics

The consequence of substantial variability among mitochondrial genome sequences implies that there is a potential opportunity to use mitochondrial genes to infer phylogenetic relationships for fungi (Taylor, 1986). It is not only the sequence differences that make mitochondrial genomes useful, but also the lack of recombination and the fact that they are haploid. Analyses of gene rearrangements such as those observed in mitochondrial genomes, have also become a valuable tool for establishing evolutionary rates in many different species including fungi (Boore & Brown, 1998). Furthermore, mitochondrial DNA analyses allows

for the evolutionary history of a species to be interrogated with regard to their gene architecture, more specifically with rearrangements in gene architecture and the positioning of selfish intron mobile units (Lang *et al.*, 1999). The position of introns, transcription factors, origins of replication, as well as excision sites, which are all contained in a single unit, also become useful for phylogenetic inference, especially when these are viewed against a background lacking recombination (Sankoff *et al.*, 1992; Santamaria *et al.*, 2009; Simon & Weib, 2008).

Molecular identification of fungal species has largely been based on nuclear DNA markers (Schoch *et al.*, 2009; Schoch *et al.*, 2012). The most commonly used markers include the genes encoding the nuclear large and small rRNA subunits and the ribosomal internally transcribed spacer region, as well as those encoding beta-tubulin, elongation factor 1-alpha and RNA polymerase subunit II (Santamaria *et al.*, 2009). The intra and inter-species variation found in the mitochondrial genes makes these genes ideal candidate markers for species differentiation (Begerow *et al.*, 2012; Rubinoff, 2006). However, the presence of mobile introns is thought to hinder the PCR efficiency, although this is probably less of a problem among closely related species (Hamari *et al.*, 2002).

Mitochondrial markers provide the opportunity to develop differentiating barcodes due to their overall higher copy number (Seifert *et al.*, 2007). DNA barcoding essentially makes use of a fragment of DNA sequence which is standardized across a kingdom of species aiming at the ease of universally available primers to generate the sequence data. Preferably, the chosen barcode locus would be the same across kingdoms (Begerow *et al.*, 2012). Barcoding initiatives have commonly drawn on the use *coxI* for differentiation of species (Seifert *et al.*, 2007), which is seen as the essential gene barcode region for animals as well as a variety of species across the kingdoms (Lane, 2009). The mitochondrial *coxI* gene region, however, has largely been excluded as a potential marker due to its difficulty to be amplified in fungi as a result of a high intron presence (Schoch *et al.*, 2012). Thus the need to identify potential gene regions that can be universally used for phylogenetic inference becomes fundamental.

4. Aims and rationale of this thesis

The principal aim of this study, both inter- and intra-specifically, is to characterize the mitochondrial genomes of different isolates of *Ceratocystis* species. The genus *Ceratocystis*

resides in the phylum Ascomycota, within the order Microascales and the class Sordariomycetes. The Sordariomycete class is one of the largest monophyletic clades in the Ascomycota (Zhang *et al.*, 2006). Species of *Ceratocystis* gained its name from "ceratos" meaning horn and "cyst" symbolic of a sac-like pouch which are distinguishable characteristics of these fungi (Seifert & Okada, 1993). Species of *Ceratocystis* have a worldwide distribution and reports are increasing on a variety of hosts as well as vast geographical areas (Wingfield *et al.*, 2011; Wingfield, 1993).

These fungi are of great significance especially, economically important in that several species are pathogenic on crop plants and also associated with certain insect attacks on wood (Hunt, 1956). Some *Ceratocystis* species are responsible for staining forest products (van Wyk *et al.*, 2006; van Wyk *et al.*, 2010). The genus *Ceratocystis* includes species pathogenic to forest and shade tree diseases, agricultural crops and herbaceous plants (Alexopoulous, 1962). Overall, species within this genus infect a wide variety of commercially important food crops such as sugarcane, sweet potato, banana, pineapple, beans, mango, nut fruits, dates, and coffee. Agronomic crops such as cotton, rubber and tobacco as well as varieties of eucalypt, wattle and plane trees have also susceptible to these fungi (Grylls & Seifert, 1993). The reports of *C. fimbriata* from *Eucalyptus* spp. and *C. albifundus* from *Acacia* spp. have led to an increased concern with regard to their threat to the commercial industry especially in the tropics and the southern hemisphere (Roux *et al.*, 2004).

Gene genealogies and DNA based studies while initially allowing for the distinction of some species (Wingfield *et al.*, 2011; Witthuhn *et al.*, 1998) has proved to be less useful in resolving relationships between closely related species. The emerging rate at which new species are been discovered means that clear species differentiation tools are urgently needed in order to both identify and classify these *Ceratocystis* species. The taxonomic history of *Ceratocystis* currently includes a number of species that can be grouped into four main clades, namely the "*coerulescens*", "*fimbriata*", "*moniliformis*" and "*paradoxa*" clades (Wingfield *et al.*, 2011). Within each of these clades there are a number of cryptic species which have not been successfully delineated on the basis of morphology or multigene phylogenies (Wingfield *et al.*, 1996). Examples of such cryptic species complexes are *C. albifundus* within the "*fimbriata*" group, and *C. polonica*, *C. laricicola* within the "*coerulescens*" group (Montoya & Wingfield, 2006).

The rationale of this study was to investigate the mitochondrial genomes of selected species of *Ceratocystis* species that have been chosen to represent the *C. fimbriata* and *C. moniliformis* clade in recent phylogenetic analyses for this group of fungi (Montoya & Wingfield, 2006; Roux *et al.*, 2001; van Wyk *et al.*, 2006; Wingfield *et al.*, 2011). In addition six isolates of *C. albifundus* were chosen for an inter species characterisation of mitochondrial genomes. The larger vision is that the mitochondrial genomes might provide useful markers for species delineation within *Ceratocystis*. This study forms the first step in this process. The ability to define possible markers based on gene rearrangements within these mitochondrial genomes would then also greatly enhance the molecular classification of cryptic species complexes. At the outset, there was a hope that studying the mitochondria of *Ceratocystis* species might establish a framework that would lead to diagnostic tools for this group of fungi. This would then allow mitochondrial DNA gene complexes to be used in phylogenomics to differentiate between cryptic species in the various complexes for the larger group.

1.5. References

Adams K. L., Daley D. O., Qui Y. L., Whelan J., and Plamer J. D. (2000). Repeated, recent and diverse transfer of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408:** 354-357.

Al-Reedy R. M., Malireddy R., Dillman C. B., and Kennell J. C. (2012). Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genetics and Biology* **49:** 2-14.

Alexopoulous C. J. (1962). Introductory Mycology. *2nd Edition; John Wiley and Sons, Inc.***:** 217-240.

Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., and Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25:** 3389-3402.

Atkins S. D., and Clark I. M. (2004). Fungal molecular diagnostics: a mini review. *Journal of Applied Genetics* **45:** 3-15.

Ballard J. W. O., and Whitlock M. C. (2004). The incomplete natural history of mitochondria. *Journal of Molecular Ecology* **13:** 729–744.

Barr C. M., Neiman M., and Taylor D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168:** 39-50.

Basse C. W. (2010). Mitochondrial inheritance in fungi. *Current Opinion in Microbiology* **13:** 712-719.

Begerow D., Nilsson H., Unterseher M., and Maier W. (2012). Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Mycology and Biotechnology* **87:** 99-108.

Bendich A. J. (1993). Reaching for the ring: the study of mitochondrial genome structure. *Current Genetics* **24:** 279-290.

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

Bonitz S. G., Berlani R., Coruzzi G., Li M., Macino G., Noberga F. G., Noberga M. P., Thalenfeld B. E., and Tzagoloff A. (1980). Codon recognition rules in yeast mitochondria. *Proceedings of the National Academy of Science, USA* **77:** 3167-3170.

Boore J. L., and Brown W. M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics and Development* **8:** 668-674.

Bruns T. D., White T. J., and Taylor J. W. (1991). Fungal molecular systematics. *Annual Review of Ecological Systems* **22:** 525-264.

Bullerwell C. E., and Lang B. F. (2005). Fungal Evolution: the case of the vanishing mitochondrion. *Current Opinion in Microbiology* **8:** 362-369.

Burger G., Gray M. W., and Lang B. F. (2003). Mitochondrial genomes: anything goes. *Trends in Genetics* **19:** 709-716.

Burger G., Jackson C. J., and Waller R. F. (2012). Unusual Mitochondrial genomes and genes. *In* "Organelle Genetics" (C. E. Bullerwell, Ed.), pp. 41-77, Springer, Berlin, Heidelberg.

Cahan P., and Kennell J. C. (2005). Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi. *Molecular Genetics and Genomes* **273:** 462-473.

Campbell A., Mrazek J., and Karlin S. (1999). Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proceedings of the National Academy of Science, USA* **96:** 9184-9189.

Capella-Gutierrez S., MMarcet-Houben M., and Gabaldon T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BioMedical Central Biology* **10:** 1-13.

Chen J. Z., and Hebert P. D. N. (1999). Intra-individual sequence diversity and hierarchical approach to the study of mitochondrial DNA mutations. *Mutational Research* **434:** 205-217.

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

Clare E. L., Kerr K. C. R., von Konigslow T. E., Wilson J. J., and Hebert P. D. N. (2008). Diagonising Mitochondrial DNA diversity: Applications of a sentinel gene approach. *Journal of Molecular Evolution* **66:** 362-367.

Contamine V., and Picard M. (2000). Maintenance and Integrity of the mitochondrial genome: a Plethora of nuclear genes in the budding yeast. *Microbiology and Molecular Biology Reviews* **64:** 281-315.

Crick F. H. (1966). Codon-anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology* **19:** 548-555.

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

Cunnington J. H. (2007). Organization of the mitochondrial genome of *Fusarium oxysporum* (anamorphic Hypocreales). *Mycoscience* **48:** 403-406.

Duchene A.-M., Pujol C., and Marechal-Drouard L. (2009). Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Current Genetics* **55:** 1-18.

Ferandon C., Moukha S., Callac P., Benedetto J.-P., Castroviejo M., and Barroso G. (2010). The *Agarius bisporus cox1* gene: The longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *Public Library of Science ONE* **5:** 1-8.

Forget L., Ustinova J., Wang Z., Huss V. A. R., and Lang B. F. (2002). *Hyaloraphidium curvatum*: A linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. *Molecular Biology and Evolution* **19:** 310-319.

Gray M. W., Burger G., and Lang B. F. (1999). Mitochondrial Evolution. *Science* **283:** 1476-1481.

Gray M. W., Lang B. F., Cedergren R., Golding G. B., Lemieux C., Sankoff D., Turmel M., Brossard N., Delage E., Littlejohn T. G., Plante I., Rioux P., Saint-Louis D., Zhu D., and Burger G. (1998). Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Research* **26:** 865-887.

Grylls B. T., and Seifert K. A. (1993). A synoptic key to species of *Ophiostoma*, *Ceratocystis* and *Ceratocystiopsis*. In *Ceratocystis* and *Ophiostoma*: Taxonomy, Ecology and Pathogenicity. *(Wingfield, M.J., Seifert, K.A. and J.F. Webber, eds.) APS Press, St Paul, Minnesota***:** 261-268.

Hamari Z., Juhasz A., and Kevei F. (2002). Role of introns in mitochondrial genome diversity of fungi (a mini review). *Acta Microbiologica et Immunologica Hungarica* **49:** 331-335.

Haridass S., Breuill C., Bohlmann J., and Hsiang T. (2011). A biologist's guide to *de-novo* genome assembly using next-generation sequence data: A test with fungal genomes. *Journal of Microbiological Methods* **86:** 368-375.

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

Hopper A. K., and Phizicky E. M. (2003). tRNA transfers to the limelight. *Genes and Development* **17:** 162-180.

Hunt J. (1956). Taxonomy of the genus *Ceratocystis*. *Lloyadia* **19:** 1-58.

Joardar V., Abrams N. F., Hosteler J., Paulstelis P. J., Pakala S., Pakala B. B., Zafar N., O.O A., Payne G., Andrianopoulous A., Denning D. W., and Nierman W. C. (2012). Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genomes size variability. *BioMed Central Genomics* **13:** 1-12.

Juhasz A., Pfeiffer I., Keszthelyi A., Kucsera J., Vagvolgyi C., and Hamari Z. (2008). Comparative analysis of the complete mitochondrial genome of *Aspergillus niger* mtDNA type 1a and *Aspergillus tubingensis* mtDNA type 2b. *Federation of Microbiological Society, Microbiology Letters* **281:** 51-57.

Kamenski T., Kolensikova O., Jubenot V., Entelis N., Krasheninnikov I. A., Martin R. P., and Tarassov I. (2007). Evidence for an adaptation mechanism of mitochondrial translation via tRNA import from the cystol. *Molecular Cell* **26:** 625-637.

Kawano S., Suzuki T., and Kuroiwa T. (1982). Structural homogeneity of mitochondrial DNA in the mitochondrial nucleoid of *Physarum polycephalum*. *Biochimica et Biophysica Acta* **696:** 290-298.

Kerscher S., Durstewitz G., Casaregola S., Gaillardin C., and Brandt U. (2001). The complete mitochondrial genome of *Yarrowia lipolytica*. *Computational and Functional Genomics* **2:** 80-90.

Kouvelis V. N., Ghikas D. V., and Typas M. A. (2004). The analysis of the complete mitochondrial genome of *Lecanicillium muscarium* (synonym *Verticillium lecanni*) suggests a minimum common gene organization in mtDNAs of *Sordariomycetes*: phylogenetic implications. *Fungal Genetics and Biology* **41:** 930-940.

Lane N. (2009). On the origin of bar codes. *Nature* **462:** 272-274.

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

Lang B. F., Lavrov D., Beck N., and Steinberg S. V. (2012). Mitochondrial tRNA structure, identity, and evolution of the genetic code. *In* "Organelle Genetics" (C. E. Bullerwell, Ed.), pp. 431-474, Springer, Berlin, Heidelberg.

Likic V. A., McConville M. J., Lithgow T., and Bacic A. (2010). Systems Biology: The next frontier of Bioinformatics. *Advances in Bioinformatics* **2010:** 1-10.

Liu Y., Steenkamp E. T., Brinkmann H., Forget L., Philippe H., and Lang B. F. (2009). Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BioMedical Central Evolutionary Biology* **9:** 1-11.

Logan D. C. (2003). Mitochondrial dynamics. *New Phytologist* **160:** 463-478.

Lowe T., and Eddy S. R. (1997). TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25:** 955-964.

Ma L.-J., and Fedorova N. D. (2011). A practical guide to fungal genome projects: strategy, technology, cost and completion. *Mycology* **1:** 9-24.

Metzker M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics* **11:** 31-46.

Montoya M. M., and Wingfield M. J. (2006). A review of *Ceratocystis sensu stricto* with special reference to the species complexes *C. coerulescens* and *C. fimbriata*. *Revista Facultad Nacional de Agronomia, Medellin* **59:** 3045-3375.

Moura G., Paredes J. A., and Santos M. A. S. (2010). Development of the genetic code: Insights from a fungal codon reassignment. *Federation of Microbiological Society, Microbiology Letters* **584:** 334-341.

Notsu Y., Masood S., Nishikawa N., Kubo N., Akiduki G., Nakazono M., Hirai A., and Kadowaki K. (2002). The complete sequence of the rice (*Oryza sativa* L) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics* **268:** 434-445.

Osiewacz H. D. (2002). Genes, mitochondria and aging in filamentous fungi. *Aging Research Reviews* **1:** 425-442.

Pantou M. P., Kouvelis V. N., and Typas M. A. (2006). The complete mitochondrial genome of the vascular wilt fungus *Verticillium dahlia*: a novel gene order for *Verticillium* and a diagnostic tool for species identification. *Current Genetics* **50:** 125-136.

Pantou M. P., Kouvelis V. N., and Typas M. A. (2008). The complete mitochondrial genome of *Fusarium oxysporum*: Insights into fungal mitochondrial evolution. *Gene* **419:** 7-15.

Richardson D. R., Lane D. J., Becker E. M., Huang M. L., Whitnall M., Rahmanto Y. S., Sheftel A. D., and Ponka P. (2010). Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol. *Proceedings of the National Academy of Science, USA* **107:** 10775-10782.

Roux J., Harrington T. C., Steimel J., and Wingfield M. J. (2001). Genetic variation in the wattle wilt pathogen *Ceratocystis albifundus*. *Mycoscience* **42:** 327-332.

Roux J., Wyk M. v., Hatting H., and Wingfield M. J. (2004). *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. *Plant Pathology* **53:** 414-421.

Rubinoff D. (2006). Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology* **20:** 1026-1033.

Rycovska A., Valach M., Tomaska L., Bolotin-Fukuhara M., and Nosek J. (2004). Linear versus circular mitochondrial genomes: intraspecies variability of mitochondrial genome architecture in *Candida parapsilosis*. *Microbiology* **150:** 1571-1580.

Sammet S. G., Bastolla U., and Porto M. (2010). Comparison of translation loads for standard and alternative genetic codes. *BioMedical Central Evolutionary Biology* **10:** 1-14.

Sankoff D., Leduc G., Antoine N., Paquin B., and Lang B. F. (1992). Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome. *Proceedings of the National Academy of Science, USA* **89:** 6575-6579.

Santamaria M., Vicario S., Pappada G., Scioscia G., Scazzocchio C., and Saccone C. (2009). Towards barcode markers in Fungi: an Intron map of Ascomycota. *BioMedical Central Bioinformatics* **10:** 1471-1484.

Santos M. A. S., Gomes A. C., Santos M. C., Carreto L. C., and Moura G. R. (2011). The genetic code of the fungal CTG clade. *Comptes Rendus Biologies* **334:** 607-611.

Santos M. A. S., Moura G., Massey S., and Tuite M. F. (2004). Driving change: the evolution of alternative genetic codes. *Trends in Genetics* **20:** 95-102.

Scheffler I. E. (2001). A century of mitochondrial research: achievements and perspectives. *Mitochondrion* **1:** 3-31.

Schluter A., Ruiz-Trillo I., and Pujol A. (2011). Phylogenomic evidence for a Mycoccal contribution to the mitochondrial fatty acid beta-oxidation. *Public Library of Science ONE* **6:** e21989.

Schoch C. L., Seifert K. A., Huhndorf S., Robert V., Spouge J. L., Levesque C. A., Chen W., and Consortium F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi Proceedings of the National Academy of Science, USA* **PNAS** Early Edition**:** 1-6.

Schoch C. L., Sung G.-H., Lopez-Giraldez F., Townsend J. P., Miadlikowska J., Hofsetetter V., Robbertse B., Matheny B. P., Kauff F., Wang Z., Gueidan C., Andrie R. M., Trippe K., Ciufetti L. M., Wynns A., Fraker E., Hodkinson B. P., Bonito G., Groenewald J. Z., Arzanlou M., de Hoog S. G., Crous P. W., Hewitt D., Pfister D. H., Peterson K., Gryzenhout M., Wingfield M. J., Aptroot A., Suh S.-O., Blackwell M., Hillis D. M., Griffith G. W., Castlebury L. A., Rossman A. Y., Lumbsch T. H., Lucking R., Budel B., Rauhut A., Diederich P., Ertz D., Geiser D. M., Hosaka K., Inderbitzin P., Kohlymeyer J., Volkmann-Kohlymeyer B., Mostert L., O'Donnell K., Sipman H., Rogers J. D., Shoemaker R. A., Sugiyama J., Summerbell R. C., Untereiner W., Johnston P. R., Stenroos S., Zuccaro A., Dyer P. S., Crittenden P. D., Cole M. S., Hansen K., Trappe J. M., Yahr R., Lutzoni F., and Spatafora J. W. (2009). The Ascomycota tree of life: A phylum-wide phylogeny clarifies the origin and

evolution of fundamental reproductive and ecological traits. *Systematics Biology* **58:** 224-239.

Seifert K. A., and Okada G. (1993). *Graphium* pantographs of *Photon* species and similar pantographs of other ascomycetes. In: *Ceratocystis* and *Photon*: Taxonomy, Ecology and Pathogenicity. (*Wingfield, M.J., Seifert, K.A. and J.F. Webber, eds.) APS Press, St Paul, Minnesota.*

Seifert K. A., Samson R. A., Dewaard J. R., Houbraken J., Levesque C. A., Moncalvo J. M., Louis-Seize G., and Hebert P. D. (2007). Prospects for fungus identification using CO I DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Science, USA* **104:** 3901-3906.

Shendure J., and Hanlee J. (2008). Next-generation DNA sequencing. *Nature Biotechnology* **26:** 1135-1145.

Silva R. M., Miranda I., Moura G., and Santos M. A. S. (2004). Yeast as a model organism for studying the evolution of non-standard genetic codes. *Briefings in Functional Genomics and Proteomics* **3:** 35-46.

Simon U. K., and Weib M. (2008). Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Biology and Evolution* **25:** 2251-2254.

Takano H., Onoue K., and Kawano S. (2010). Mitochondrial fusion and inheritance of the mitochondrial genome. *Journal of Plant Research* **123:** 131-138.

Taylor J. W. (1986). Fungal evolutionary biology and mitochondrial DNA. *Experimental Mycology* **10:** 259-269.

Torriani S. F. F., Goodwin S. B., Kema G. H. J., Pangilinan J. L., and McDonald B. A. (2008). Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. *Fungal Genetics and Biology* **45:** 628-637.

van Wyk M., Roux J., Barnes I., Wingfield B. D., and Wingfield M. J. (2006). Molecular phylogeny of the *Ceratocystis moniliformis* complex and description of *C. tribiliformis* sp. nov. *Fungal Diversity* **21:** 181-201.

van Wyk M., Wingfield B. D., Marin M., and Wingfield M. J. (2010). New *Ceratocystis* species infecting coffee, cacao, citrus and native trees in Colombia. *Fungal Diversity* **40:** 103-117.

Voelkerding K. V., Dames S. A., and Durtschi J. D. (2009). Next-Generation Sequencing: From basic research to diagnostics. *Clinical Chemistry* **55:** 641-658.

Wingfield B. D., van Wyk M., Roos H., and Wingfield M. J. (2011). *Ceratocystis*: Emerging evidence for discrete generic boundaries. In: Ophiostomatiod fungi: expanding frontiers. *(Seifert, K.A and M.J Wingfield, eds.) CBS Biodiversity Series 12* **:** 57-64.

Wingfield M. J. (1993). Problems delineating the genus Ceratocystiopsis. In: Ceratocystis and Ophiostoma: Taxonomy, Ecology and Pathogenicity. *(Wingfield, M.J., Seifert, K.A. and J.F. Webber, eds.) APS Press, St Paul, Minnesota.***:** 21-25.

Wingfield M. J., De Beer C., Visser C., and Wingfield B. D. (1996). A new *Ceratocystis* species defined using morphological and ribosomal DNA sequence comparisons. *Systematics in Applied Microbiology* **19:** 191-202.

Witthuhn R. C., Wingfield B. D., Wolfaart M., and Harrington T. C. (1998). Monophyly of the conifer species in the *Ceratocystis coerulescens* complex based on DNA sequence data. *Mycologia* **90:** 96-101.

Woo P. C., Zhen H., Cai J. J., Yu J., Lau S. K., Wang J., Teng J. L., Wong S. S., Tse R. H., Chen R., Yang H., Lui B., and Yuen K. L. (2003). The mitochondrial genome of the thermal dimorphic fungus *Penicillium marneffei* is more closely related to those of molds than yeast. *Federation of Microbiological Society, Microbiology Letters* **555:** 469-477.

Wu R., and Kaiser A. D. (1968). Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology* **35:** 523-537.

Wu Y., Yang J., Yang F., Liu T., Leng W., Chu Y., and Jin Q. (2009). Recent dermatophyte divergence revealed by comparative and phylogenetic analysis of mitochondrial genomes. *BioMedical Central Genomics* **10:** 238.

Yaffe M. P. (1999). The machinery of mitochondrial inheritance and behaviour. *Science* **283:** 1493-1497.

Yusuf D., Marz M., Stadler P. F., and Hofacker I. L. (2010). Bcheck: a wrapper tool for detecting RNase P RNA genes. *BioMedical Central Genomics* **11:** 432-441.

Zhang N., Castlebury L. A., Miller A. N., Hunhndorf S. M., Schoch C. L., Seifert K. A., Rossman A. Y., Rogers J. D., Kohlymeyer J., Volkmann-Kohlymeyer B., and Sung G.-H. (2006). An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. *Mycologia* **98:** 1076-1087.

Zivanovic Y., Wincker P., Vacherie B., Bolotin-Fukuhara M., and Fukuhara H. (2005). Complete nucleotide sequence of the mitochondrial DNA from *Kluyveromyces latis*. *Federation of Microbiological Society*, *Yeast Research* **5:** 315-322.

Figure 1. The various biological pathways and processes occurring within the mitochondrion of ascomycetes (Burger *et al.*, 2003).

The various biological processes driven by the mitochondrion are depicted, where a number of mitochondrial components which are encoded for by the nucleus are shown in blue. The orange colourings show mitochondrial subunits, whilst the subunits in pink are encoded for by the mitochondria as is the case in some eukaroyotic organisms but not all. The RNA polymerase is generally encoded for within intronic regions of the core mitochondrial gene cob. Each complex represents an electron-transport system as well as. Complexes I-V (both *coxI*, *coxII*, *coxIII* as well as *nad1*, *nad2*, *nad3*, *nad4*) are responsible for electron-transport and oxidative phosphorylation pathways. Tat, Sec and Oxa1 are translocases that are involved with the protien export from the outter membrane into the inner membrane. RNAse P has a ribozmatic functionality to process the 5 prime end of tRNAs.

Figure 1.

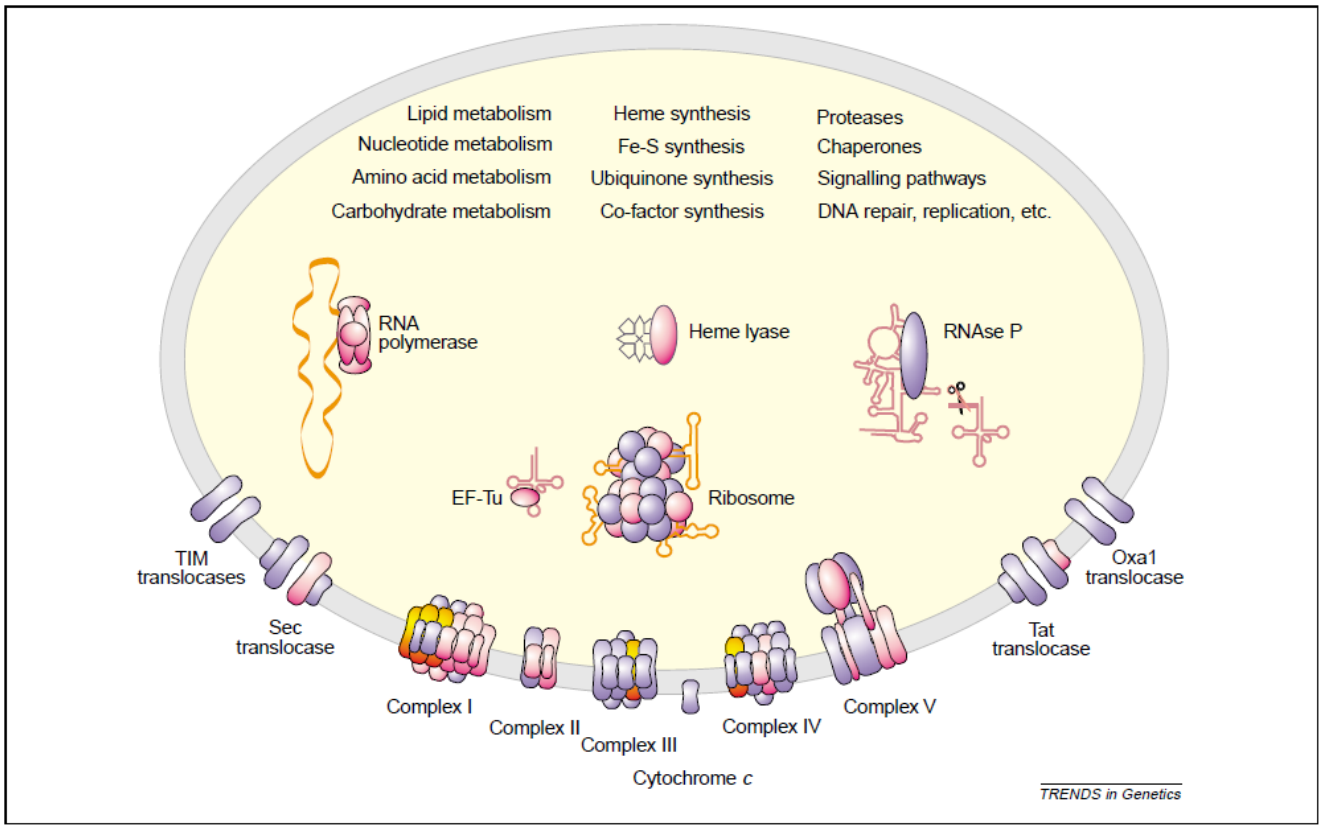


Figure 2. Mitochondrial genome size and gene components across eukaryotes (Burger *et al.*, 2003).

The diversity and the variation in the mitochondrial genomes size across a range of eukaryotes, including saccharomyces. The graph is illustrated to show the coding content of genuine mitochondrial genes most commonly found in mitochondria. The introns, intronic Open Reading Frames (ORFs), phage-like reverse transcriptases, and DNA polymerases are depicted in blue, whilst the intergenic regions are green.

Figure 2.

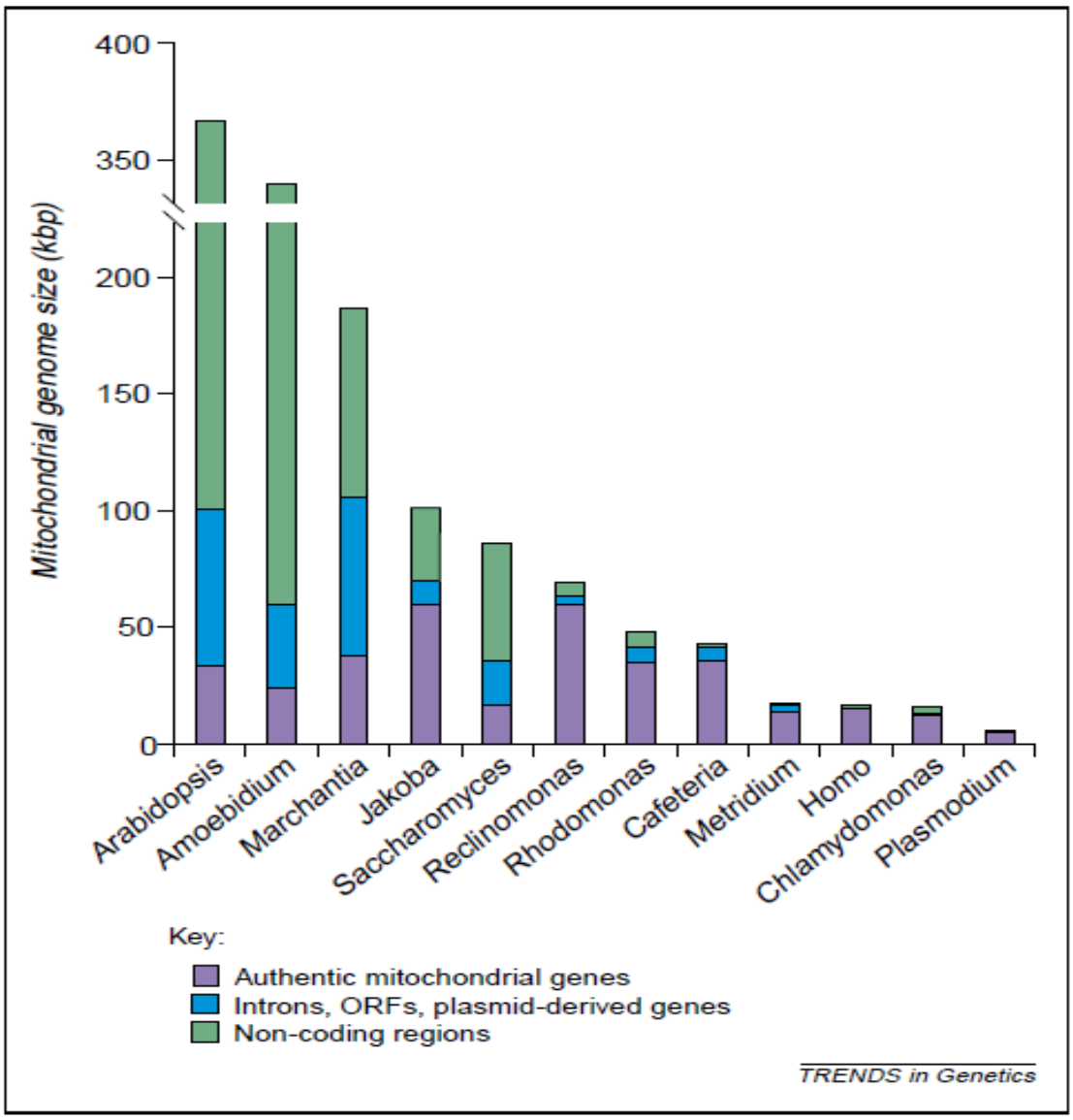


Figure 3. Diversity of the genetic code.

Shown are the three base nucleotide descriptions for the translation of the different amino acids present in mitochondrial and nuclear genomes for eukaryotic and prokaryotic systems. Adapted from (Silva *et al.*, 2004). In the mitochondrial genomes, both the sense and non-sense codons are inter-changeable. Only stop codons in the nuclear genomes are susceptible to change, with the only exception to this rule been Serine, shown in green for *Candida* species.

Figure 3.

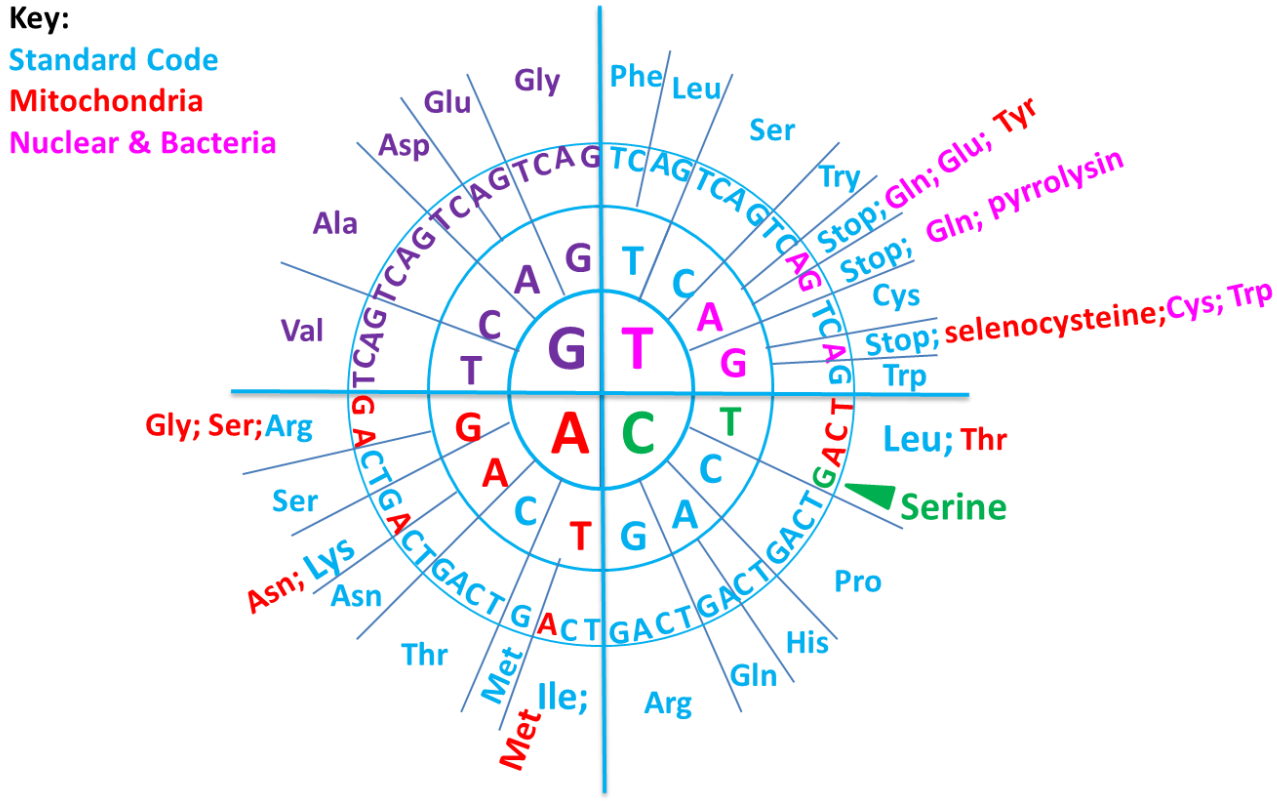


Figure 4. The mitochondria of *Saccharomyces cerevisiae* depicting the import of tRNA molecules from the nucleus into the mitochondrion.

Transport RNA molecule Lysine (CUU) and two tRNA Glycines are imported into the mitochondrion. All 23 tRNAs are synthesized in the mitochondrion. Adapted from (Duchene *et al.*, 2009).

Figure 4.

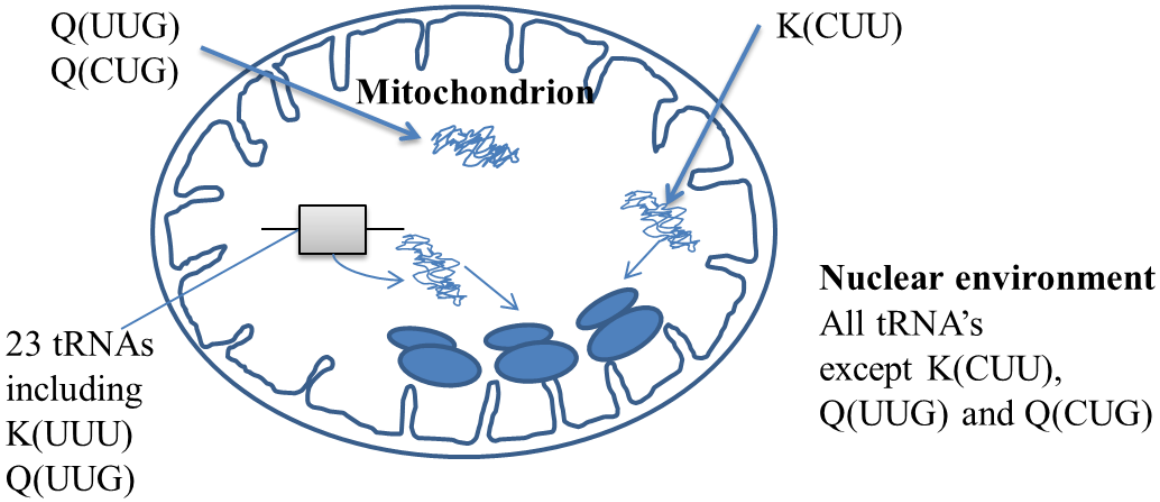


Figure 5. Comparison of different mitochondrial genomes within the Ascomycota.

Complete genome and gene sizes are shown proportionally. Introns are indicated with diagonal black solid lines within the gene, with the pink arrow showing the direction of transcription. Adapted from (Kouvelis *et al.*, 2004) with additional reference to (Cummings *et al.*, 1990; Juhasz *et al.*, 2008; Kerscher *et al.*, 2001; Pantou *et al.*, 2006; Pantou *et al.*, 2008; Torriani *et al.*, 2008; Woo *et al.*, 2003; Wu & Kaiser, 1968; Zivanovic *et al.*, 2005).

Figure 5.

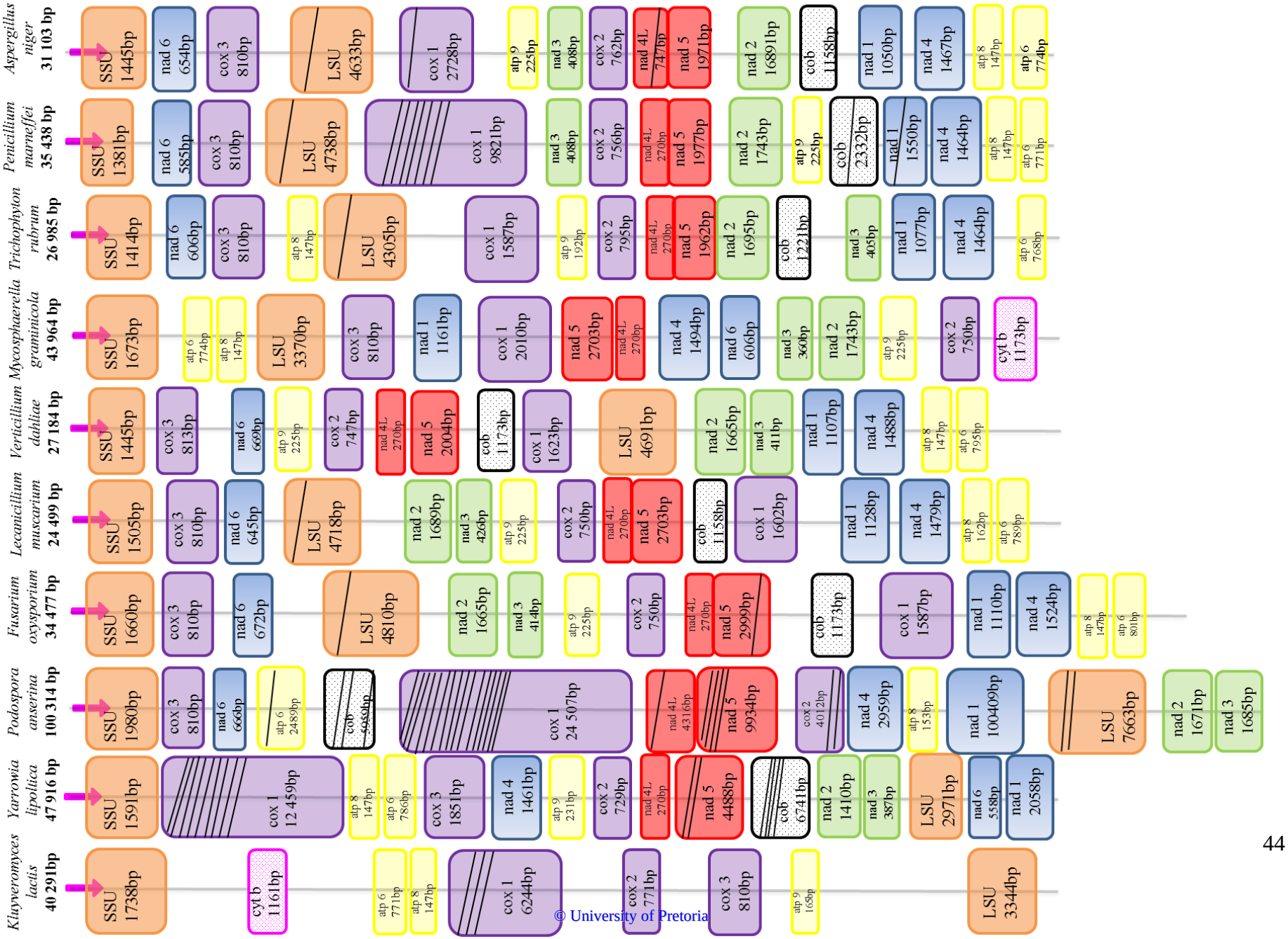
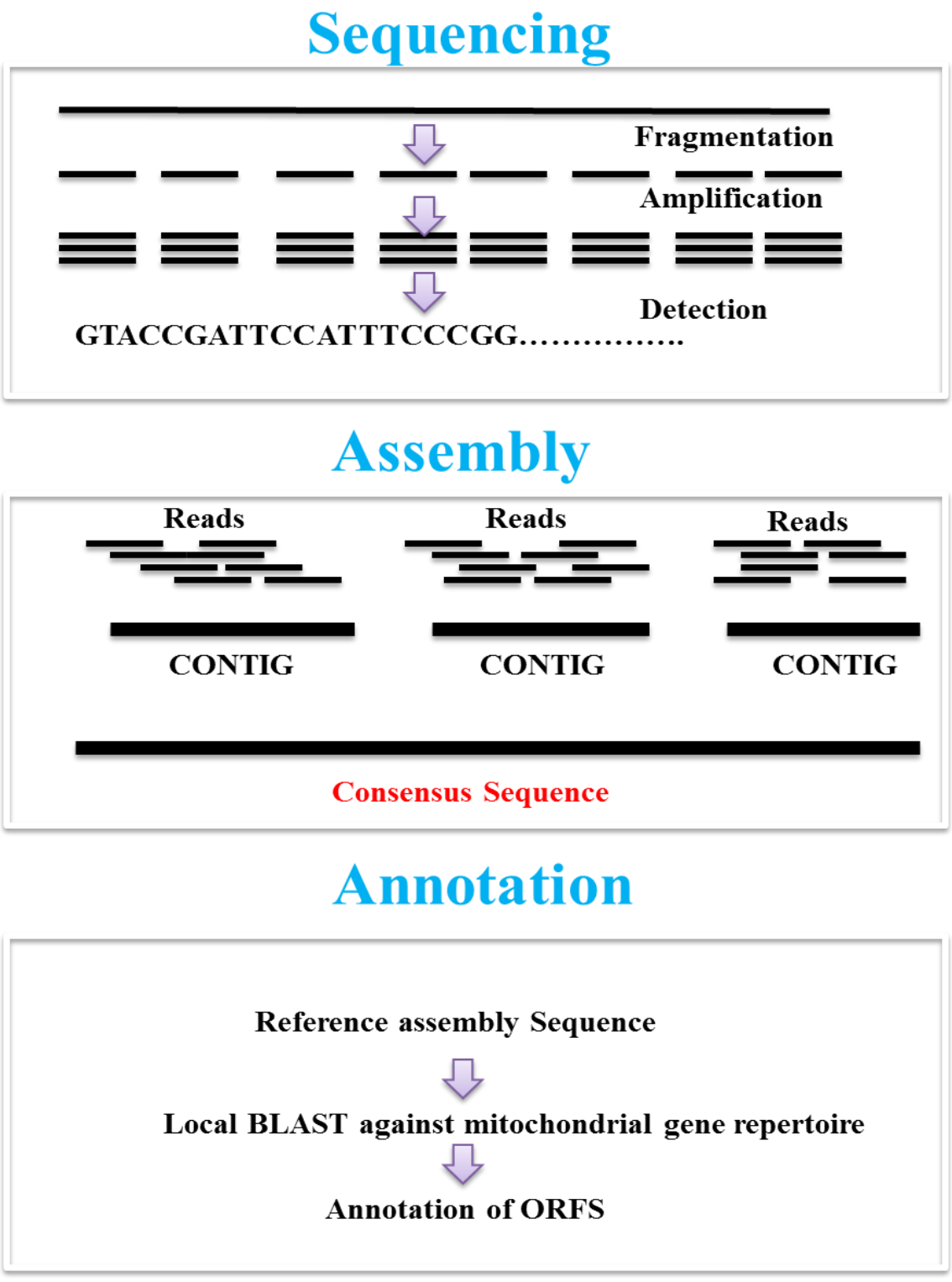


Figure 6. Summary describing genome sequencing, assembly and annotation pipelines as defined by Ma and Fedorova in 2010.

The blocks illustrate the steps undertaken when sequencing, assembling and annotating a mitochondrial eukaryotic genome. Different mitochondrial genes are depicted as horizontal black lines, and are aligned against known mitochondrial genes in order to complete the annotation process. Adapted from (Ma & Fedorova, 2011).

Figure 6.



Chapter Two

Annotation and comparative analysis of three *Ceratocystis* mitochondrial genomes

Abstract

Mitochondria are highly conserved organelles that evolve at an accelerated rate independently of their nuclear genome counterparts. Their genomes are, therefore, potentially valuable in the phylogenetic differentiation of cryptic species complexes. Very little is known regarding the mitochondria of *Ceratocystis* spp., neither of their relatives in the Microascales. In this study the aim was to compare the mitochondrial genomes of two pathogenic and one non-pathogenic *Ceratocystis* spp. These genomes were sequenced, annotated and compared. The mitochondrial genome of the pathogenic *Ceratocystis fimbriata sensu stricto* was 135 104 base pairs (bp) in size. The other pathogenic species, *Ceratocystis albifundus* was assembled into a 126 447 bp genome while the non-pathogenic species *Ceratocystis moniliformis* had a smaller (110 987 bp) mitochondrial genome. All three genomes contained 15 archetypal mitochondrial genes. Of these, ATP synthesis (ATP synthase subunits 6, 8 and 9) and the cytochrome oxidative phosphorylation suite of genes (cytochrome subunits I, II and III) were present. Genes responsible for the reduction of nicotinamide adenine dinucleotide ubiquinone oxidase subunits (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*) were also present. Other genes included apocytochrome b (*cob*), the large and small mitochondrial ribosomal RNA genes as well as the ribosomal protein S3 (*rps3*) encoded within the large ribosomal RNA subunit gene. A complement of 31 putative tRNA genes that recognized all the amino acids and accounted for the codon usage of the genes was found for *C. fimbriata sensu stricto*. In *Ceratocystis albifundus* and *C. moniliformis*, their mitochondrial genomes contained 33 and 24 putative tRNA genes respectively and these tRNAs also accounted for all the amino acid codons observed in these genomes. The tRNAs formed clusters between the core genes for all isolates. The major differences between the genomes could be ascribed to introns that were found in some genes. A variety of group I type introns were present across all the genomes, with cytochrome oxidase containing the highest number of introns. The genomes of *C. fimbriata* and *C. albifundus* were most similar to each other and very different to that of *C. moniliformis*. This is consistent with the fact that these two groups of *Ceratocystis* are very different and should reside in separate genera. Overall, the results of this study provide a framework for further genomic comparisons in *Ceratocystis* as well as other fungi in the Microascales.

2.1 Introduction

The genomes of mitochondria are typically small and encode for products that dictate a variety biological processes (Burger *et al.*, 2003; Gray *et al.*, 1999). The majority of these are proteins that are responsible for electron transport and ATP synthesis, translation and RNA processing, the synthesis of iron-sulphur clusters as well as calcium homeostasis (Bullerwell *et al.*, 2000). Metabolism of lipids, nucleotides, amino acids and carbohydrates are important functions of mitochondria, in conjunction with DNA replication and repair (Gray *et al.*, 1999).

Due to the fundamental role of the mitochondrion in the biology of eukaryotes, the genomes of these organelles have been widely studied (Ballard & Whitlock, 2004; Lang *et al.*, 1999). These genomes and/or the genes they encode are also commonly utilized in evolutionary and systematics studies (Capella-Gutierrez *et al.*, 2012; Liu *et al.*, 2009; Meusinier *et al.*, 2008), most of which have shown that mitochondrial genomes are characterized by evolutionary rates that are independent of that of their nuclear counterparts (Ballard & Whitlock, 2004). More than 600 mitochondrial genomes have been characterized and are available from public domain databases such as NCBI (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and the Gold Genomes database (Pantou *et al.*, 2008). These genomes vary considerably in size ranging from 18kb to genomes larger than 100kb and most contain the standard suite of genes (Al-Reedy *et al.*, 2012). Although many mitochondrial genomes exist as circular molecules, exceptions are found in for example the myxomycete *Physarum ploycephalum*, which is linear (Kawano *et al.*, 1982).

Compared to other eukaryotes, fungi are not well represented in the mitochondrial genome databases. Less than 20% of the publicly available sequences are fungal and even fewer (<15%) belong to the Ascomycota. Fungal mitochondrial genomes are characteristically larger in comparison to their animal counterparts (Hausner, 2003), but a magnitude smaller than those of plant mitochondria (Adams *et al.*, 2000) In general, the coding regions are highly conserved among fungal mitochondria, while polymorphism is mostly restricted to intergenic regions and introns, and in some cases plasmid sequences (Kouvelis *et al.*, 2004; Scheffler, 2001). It has been suggested that this high level of polymorphism, along with the high copy numbers and conservation through alterations in the genetic code, provides the potential for species differentiation (Sankoff *et al.*, 1992).

Various studies have used mitochondrial gene sequences to address taxonomic questions in fungi (Boore & Brown, 1998; Capella-Gutierrez *et al.*, 2012; Liu *et al.*, 2009). Most of these studies have focused on the use of multiple mitochondrial genes such as the cytochrome *c* oxidase complex, the apocytochrome *b* gene, the NADH dehydrogenase complex and the ATPase subunits or using single genes such as *atp6* that encodes the ATP synthase subunit 6 and *rns* that encodes the small ribosomal RNA subunit, to resolve relationships among species and genera (Capella-Gutierrez *et al.*, 2012; Feau *et al.*, 2011; van de Sande, 2012; White *et al.*, 1996). Some studies have also employed whole mitochondrial genome sequences to resolve relationships at deeper systematic levels. The study of Lui et al (Liu *et al.*, 2009), for example utilized mitochondrial gene sequences to show monophyly of the so-called Holomycota (i.e., the clade that includes Fungi and it's sistergroup Nucleariida) and paraphyly of in the "Zygomycota".

The genus *Ceratocystis* (order Microascales, class Sordariomycetes) includes many plant pathogens (Wingfield *et al.*, 2011). These pathogenic species are responsible for cankers, wilting and ultimately tree mortality (Roux *et al.*, 1999; Roux *et al.*, 2007; Roux & Wingfield, 2001; Roux *et al.*, 2004; van Wyk *et al.*, 2010; Wingfield *et al.*, 2011). The overall objective of this study was to characterize and fully annotate the mitochondrial genomes of *Ceratocystis* spp. To achieve this goal, the mitochondrial genomes of three species of *Ceratocystis* were considered. These included two well-known known plant pathogenic species in the *Ceratocystis fimbriata* complex, i.e., *C. fimbriata sensu stricto* that infects sweet potato (Halsted & Fairchild, 1891) and the tree pathogen *C. albifundus* (Morris *et al.*, 1993; Wingfield *et al.*, 1996). The saprophytic *C. moniliformis*, a member of the *C. moniliformis* complex (van Wyk *et al.*, 2006) was included for comparative purposes. Comparisons of total DNA sequence for each species was made based on the amino acid alignments from the annotated genomes of *Fusarium oxysporum*, *Podospora anserina* and *Verticillium dahliae* (Kouvelis *et al.*, 2004; Pantou *et al.*, 2008), which provided a framework for comparison.

2.2. Materials and methods

2.2.1 Fungal isolates and DNA isolation

Isolates of *C. fimbriata s.s.* (CMW 15049), *C. albifundus* (CMW 17274) and *C. moniliformis* (CMW 10134) were used in this study. They were grown at 25°C on 2% malt extract agar (MEA: 20% w/v; Biolab, Midrand, South Africa) supplemented with 100µg/L thymine. These isolates are maintained in the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa as well as in the culture collection of the Centraalbureau voor Schimmelcultures, Utrecht, Netherlands.

A single hyphal tip from each isolate was individually plated onto fresh MEA medium and allowed to grow for about two weeks prior to DNA isolation. To obtain high quality total genomic DNA, phenol-chloroform extractions were carried out as previously described (Roux *et al.*, 2004). DNA concentrations were quantified using a Qubit Flourometer (Invitrogen, USA). Approximately 20 DNA isolations from individual Petri dishes were pooled to obtain approximately 5µg of DNA for each isolate for the subsequent sequencing.

2.2.2 Next generation DNA Sequencing

DNA for the three isolates was subjected to either 454Pyrosequencing, Illumina or SOLiD next generation sequencing platforms (Metzker, 2010). 454 Pyrosequencing using the titanium technology was carried out at Inqaba Biotechnical Industries (PTY) LTD., South Africa. Illumina (MiSeq Sequencer) sequencing was carried out by the sequencing facility of the University of California, Davis as well as at the sequencing platform at the Agricultural Research Council facility, South Africa. The SOLiD sequencing using the Ion Torrent (Invitrogen, USA) was performed at SEQOMICS (Hungary, (PTY) LTD).

2.2.3 Assembly, analysis and annotation of sequence data

CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark) was used to assemble and annotate the mitochondrial DNA sequences. A *de-novo* assembly was generated using the long reads [>450bp] generated from the 454 Pyrosequencing data libraries for *C. fimbriata s.s.* A single contig of ~135 kb was produced that had BLAST

similarity to mitochondrial genes (Supplementary Data CD, C_fimbriata_Genebank_file). All the reads that mapped to the consensus sequence were then extracted and used for a second *de-novo* assembly. A re-assembly was then made using all the single reads obtained from the two sequencing platforms. This generated a consensus sequence for *C. fimbriata s.s.*, which was then used as a reference mitochondrial genome for the other two species used in this comparative study. The total reads for both *C. albifundus* and *C. moniliformis* were then individually mapped against the *C. fimbriata s.s.* reference sequence. This generated consensus sequences for each species.

The assembled mitochondrial genome sequences for *C. albifundus* and *C. moniliformis* were subjected to an open reading frame (ORF) prediction algorithm (CLC BIO-Genomics workbench) that identified the ORFs on the basis of the translational mitochondrial genetic code 4 (Fox, 1987). Once putative ORFs had been determined, these were subjected to a BLASTN similarity comparison (Altschul *et al.*, 1997). The tRNA genes were identified using $_{T}RNA_{SCAN}$-SE version 1.21 software (Lowe & Eddy, 1997). RNA_{WEASEL} (http://megasun.bch.umontreal.ca/RNAweasel) aided in the identification and curation of the intron and rRNA units present (Schattner *et al.*, 2005). Intronic ORFs that contained endonucleases as per the mitochondrial genetic code 4 (Fox, 1987) were also identified using BLAST searches (Altschul *et al.*, 1997). Manual annotation of all ORFs and tRNAs was made using CLCBioGenomics workbench using complete gene sequences as well as amino acid alignments from the annotated genomes of *Fusarium oxysporum*, *Podospora anserina* and *Verticillium dahliae* as a basis for comparison. The codon usage was confirmed using the freeware prediction programme Gene Infinity (http//www.geneinfinity.org) (Smith *et al.*, 2011). The amino acid alignments for each of the mitochondrial genes for the *Ceratocystis* species can be accessed on the Supplementary Data CD provided with this thesis. The introns present in some of the genes were further characterized and their homing endonuclease genes predicted using InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan) (Zdobnov & Apweiler, 2000).

2.3. Results

2.3.1 Sequencing data

The total sequenced DNA obtained for each species and the data used to produce each of the mitochondrial genomes is summarized in Table 1. The largest sequence data set [~ 14Gb] was obtained for *C. fimbriata s.s.* and this included sequence reads generated with 454 pyrosequencing and Illumina next generation sequencing technologies [Tables 1 and 2]. *Ceratocystis albifundus* sequences from 454 pyrosequencing, Illumina and SOLiD technologies amounted to a total of ~ 21 Gb of data [Tables 1 and 2]. A total of 16,4 Gb of sequence data were generated for *C. moniliformis* from the FLX 454 and Illumina sequencing runs [Tables 1 and 2]. The coverage of mitochondrial reads mapped to each genome was approximately 10-20%. The coverage of the mitochondrial contigs was approximately 3-4 times the coverage obtained for the single copy genes.

2.3.2 Annotation of mitochondrial genomes

2.3.2.1 *Ceratocystis fimbriata sensu stricto*

The mitochondrial genome of *C. fimbriata s.s* assembled into a circular fragment of 135 104 bp [Figure 1]. The coverage obtained from the *de novo* assembly was ~21 000 times, and comprised of nearly 6 million single and ~2.5 million paired reads [Table 1]. The mitochondrial genome contained 73.2 % of adenine and thymine nucleotides [Table 2]. It contained 14 putative mitochondrial protein coding genes together with the gene for the large (*rnnl*) and small (*rns*) ribosomal RNA subunits [Figure 1 and Table 3]. The protein coding genes included those for three ATP synthase subunits (*atp6*, *atp8* and *atp9*); three cytochrome oxidase subunits (*coxI*, *coxII* and *coxIII*); seven nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits (*nad1*-6 and *nad4L*); apocytochrome B (*cob*) and the *rps3* gene, which is encoded within the single Group I type intron harboured by the *rnnl* gene. All genes found were on the positive strand.

A total of 31 tRNA genes corresponding to the 20 standard amino acids were predicted [Table 4 and Figure 2]. In addition, the *C. fimbriata s.s* mitochondrial genome was predicted to have one possible suppressor tRNA present [Figure 2]. Suppressor tRNA molecules were

also predicted to include mutations within their anticodon loop. The *C. fimbriata s.s* mitochondrial genome also had an over-representation of the tRNA for the amino acid methionine, present five times in the genome [Figure 2]. The tRNA prediction software programme used the genetic code 4, taking into account the wobble hypothesis for mitochondrial genome characterization, which indicated that there was a strong bias towards codons ending with an adenine [Table 4].

The mitochondrial genes *nad5*, *coxI*, *coxIII* and *cob* all contained Group I introns [summarized in Table 5 and Figure 3]. The *nad5* gene was disrupted by two Group ID introns. Within these introns the predicted homing endonuclease genes (HEGs) all contained the motif LAGLIDADG. The gene encoding apocytochrome oxidase B contained two Group ID type introns. Both introns had ORFs predicting HEGs with LAGLIDADG and GIY-YIG motifs. The second intron [4 174bp] in the *cob* gene also had a 2 058 bp ORF of which the inferred amino acid sequence showed similarity to a reverse transcriptase polymerase. The ~19kb *coxI* gene contained four introns. These introns were comprised of Group IB and IC type introns. The HEGs for both of these encoded LAGLIDADG and GIY-YIG motifs. The *coxIII* gene had only Group IB type intron with an in frame HEG ORF with a LAGLIDADG motif. Only one intergenic ORF was detected that could be classified as possibly encoding a LAGLIDADG homing endonuclease [Figure 1; Figure 3].

Two of the *C. fimbriata s.s* ORFs showed high similarity to genes in *P. anserina* that are typically encoded by mitochondrial plasmids [Figure 1]. Three ORFs for which no similarity function could be determined were thus designated as hypothetical proteins. The hypothetical proteins predicted by CLC Genomics were in the vicinity of the mitochondrial plasmid genes which were predicted by software programme RNA_{WEASEL}.

2.3.2.2 *Ceratocystis albifundus*

The single and paired reads of *C. albifundus* [CMW 17274] were mapped against *C. fimbriata s.s* reference mitochondrial genome [Supplementary Figure 1]. All the gene regions had a high coverage [approximately 2500 times], comprising of ~27000 single and ~28000 paired reads. The AT content for *C. albifundus* was 72.9% [Table 1]. The *C. albifundus* genome was circular and 126 447 bp in size [Table 1 and Figure 4]. It harboured 14 core protein coding genes, *rns* and *rnnl*, as well as *rps3* within the single *rnnl* intron

[Table 3 and Figure 4]. The protein coding genes included *atp6*, *atp8*, and *atp9*, *coxI*, *coxII* and *coxIII*, *cob*, *nad1*-6 and *nad4L*); and apocytochrome B (*cob*). All genes were transcribed from the positive stand and the genome harboured one predicted LAGLIDADG-encoding HEG between the *atp6* and *cob* genes.

Ceratocystis albifundus had a total of 33 tRNA corresponding to 20 amino acids [Table 4 and Figure 5]. The tRNA gene that codes for the amino acid methionine was also found five times in the mitochondrial genome of *C. albifundus* [Figure 5]. There were also three tRNA genes for Serine (Ser) and four tRNAs for Threonine (Thr) [Figure 5]. There was a strong bias towards codons ending with an Adenine [Table 4].

Group I type introns were predicted in the *nad5*, *coxI*, *coxIII* and *cob* genes [Table 6; Figure 6]. There were two introns found within the *nad5* gene, both of which were predicted as Group ID types and contained putative HEGs encoding LAGLIDADG motif. The *cob* gene had three introns, two of which were Group IA and IB types with HEG ORFs all containing LAGLIDAD motifis, while the third was also IB but lacked a detectible ORF. Three introns of varying sizes were found in the *coxI* gene and all HEGs were either LAGLIDADG or GIY-YIG protein motifs. Two introns were found in the *coxIII* gene, both containing HEGs with a LAGLIDADG domain, where the second intron was biorfic.

2.3.2.3 *Ceratocystis moniliformis*

The total sequence data [Table 2] obtained for *C. moniliformis* [CMW 10234] was mapped against *C. fimbriata s.s* reference mitochondrial genome. The coverage obtained was 4458 times and generated a consensus circular mitochondrial genome of 110 987 bp in size [Table 2; Figure 7] made up of ~26500 single and ~15500 paired reads [Table 2]. The *C. moniliformis* sequence had an AT content of 70.3% and encoded the 14 core protein coding genes, *rns* and *rnnl*, together with *rps3* within the *rnnl* intron [Table 3 and Figure 7]. The *C. moniliformis* genome also contained twenty four tRNA genes [Table 4 and Figure 8]. These corresponded to the 18 amino acids, because the tRNA genes for the amino acids Valine and Isoleucine were not detected. The codon usage showed a strong bias for codons ending in Adenine [Table 4]. All the mitochondrial genes were transcribed from the positive stand.

Ceratocystis moniliformis contained only Group I type introns within its mitochondrial genes *nad5*, *coxI*, *coxIII* and *cob* [Table 5; Figure 9]. There was only one intron of Group IB type

predicted for the *nad5* gene, which corresponded to a HEG of the LAGLIDADG protein domain. The *cob* gene contained three introns, where the first one was predicted as a Group ID type with an in-frame ORF coding for a GIY-YIG domain. The second intron was Group I (derived A) harbouring two ORFs that both encoded HEGs with LAGLIDADG domains, while the third intron did not contain any ORFs. The *coxI* gene [~14000 bp] contained three introns. These introns comprised of Group IB and I (derived B2) type introns. The HEGs proteins had either LAGLIDADG or GIY-YIG motifs. The second intron of *coxI* did not have any ORFs and hence no HEGs. The *coxIII* gene had only a Group IC2 type intron with an in frame HEG ORF which had a LAGLIDADG domain.

2.4. Discussion

The mitochondrial genomes of the three *Ceratocystis* species were 135 104 bp, for *C. fimbriata s.s.*, 126 447 bp for *C. albifundus* and 110 987 bp for *C. moniliformis*. All three genomes were large but not outside the expected range seen in Ascomycete mitochondrial genomes, which are can generally range from 50kb to 120kp in size (Burger *et al.*, 2003). This size variation among the three genomes was not unusual it is well-recognised that this may be due to the length of the intergenic regions and the presence and size of introns (Lang *et al.*, 1999). In fact, mitochondrial introns can range from 0.15 to 5kb in size and influence genome size across the different genera (Ambrosio *et al.*, 2013). Another factor contributing to the size variation observed is the predicted plasmid gene elements that were predicted as known mitochondrial plasmid elements found in fungi such as *Neurospora crassa* and *Podospora anserina* (Hausner, 2003).

All three genomes considered in this study contained the 14 essential mitochondrial genes that have been reported previously for the Ascomycota (Hausner, 2003). These included the genes involved in electron transport and phosphorylative processes producing ATP include cytochrome *c* oxidase subunits 1, 2 and 3; apocytochrome *b*, and ATPases. The standard repertoire of genes found in the mitochondrial genome includes the NADH dehydrogenase subunits 1, 2, 3, 4, 4L, 5, and 6; ATPase subunits 6, 8 and 9 (Hausner, 2003). The *rrnl* along with the encoded *rps3* gene and *rns* genes were also present. This complete set of mitochondrial genes enhances our confidence in the assembly and annotation processes undertaken in this study for each of the mitochondrial genomes.

The genomes of *C. fimbriata*, *C. albifundus* and *C. moniliformis* all included a minimum set of tRNA genes similar to those found in other eukaryotes (Andersson *et al.*, 2003; Burger *et al.*, 2003; Hausner, 2003). *Ceratocystis fimbriata s.s.* and *C. albifundus* encoded 31 and 33 tRNA genes respectively, while *C. moniliformis* harboured only 24 tRNAs. The products of these genes are essential for protein synthesis, which requires a complete compliment of these units (Duchene *et al.*, 2009). Significant differences have, however, been reported among mitochondrial genomes of Ascomycetes (Lang *et al.*, 2012) and the number of tRNA genes per mitochondrial genome can range across the eukaryotes from 0 to 27. If the minimum tRNA complement is not present within the mitochondrial genome, then it must be imported from the nuclear genome (Adams & Palmer, 2003; Duchene *et al.*, 2009; Lang *et al.*, 2012). In *S. cerevisiae*, the minimum number of tRNAs required to interpret a degenerate amino acid code has been reported to be 24 (Bonitz *et al.*, 1980). In terms of genomic organization, the *Ceratocystis* tRNA genes formed a cluster between the core genes, although the clusters observed for *C. fimbriata s.s.* and *C. albifundus* were different from those of *C. moniliformis*.

The minimum set of tRNA genes encoded on a mitochondrial genome usually allow for the production of functional tRNA molecules associated with the standard set of 20 amino acids (Hausner, 2003). Although this was the case for *C. fimbriata s.s.* and *C. albifundus*, it was not true for *C. moniliformis* as its predicted tRNAs corresponds to only 18 cognate amino acids. For this fungus, tRNA genes for isoleucine and valine were not detected, although none of its genes harboured codons for these amino acids. Therefore, all the codons present across all the genomes specified for each *Ceratocystis* exon were recognized by the encoded tRNAs. Had the latter not been the case, mitochondrial translation in *C. moniliformis* would have been dependent on tRNAs that are nucleus-encoded and transported into the mitochondrion (Duchene *et al.*, 2009; Lang *et al.*, 2012). There are a many organisms that undergo a similar import system, however, the identity and number of tRNAs that are imported then differs based on the organisms specificity needs (Rubio & Alfonzo, 2012).

Transport RNA molecule redundancy is a feature of many mitochondrial genomes that do not import their tRNA molecules from the nuclear genome (Rubio & Alfonzo, 2012). This can in turn be linked to the possible suppressor tRNAs seen in the three species of *Ceratocystis* considered in this study These suppressor tRNA molecules are mutated tRNAs that allow for the insertion of a needed amino acid site within a protein encoding gene (Rubio & Alfonzo,

2012). The mutations allowing for this are typically located in the anticodon region, thus changing the tRNA's specificity for specific codons. The tRNA software used to analyse both *C. fimbriata s.s* and *C. albifundus* predicted the same suppressor tRNA or methionine which also contained an intron. Methionine and threonine tRNAs were present in the genome at a higher frequency than any of the other tRNAs for all three isolates. Suppressor tRNAs thus play a crucial role in ensuring that codon specificities are always met, as the mitochondrion of an organism is under constant demand for the production of energy that ultimately drives the biochemical and metabolic pathways of an organism (Rubio & Alfonzo, 2012). The occurrence of these suppressor tRNAs in the *Ceratocystis* mitochondrial genomes is, therefore, similar to suppressor tRNAs found within the mitochondria of *S. cerevisiae* (Bonitz *et al.*, 1980).

A number of the so-called signatures known for the mitochondrial genomes of Ascomycota (Scheffler, 2001) were observed in the three *Ceratocystis* genomes considered here. These included the fact that the *nad4L* and *nad5* genes are always found next to each other, and the consecutive succession of various gene pairs, for example, *nad2* and *nad3*, *nad1* and *nad4* as well as *atp8* and *atp6* (Kouvelis *et al.*, 2004). This commonality among the genomes of Ascomycota was also evident in the preferred start (ATG) and termination (TAA) codons (Bonitz *et al.*, 1980). However, *Ceratocystis* also employed an alternative start (ATA) and stop (AAT) codons at low frequency. The high adenine and thymine content prevalent in mitochondrial genomes along with an absence of methylation and conserved gene functionality are typical characteristics of mitochondria (Campbell *et al.*, 1999). The ribosomal protein S3 (*rps3*) was embedded within a specific *rnnl* Group I type intron for all species as is a trend among the Ascomycota (Sethuraman *et al.*, 2009b). The housing of the *rps3* gene within an intron ensures that these mobile elements are successfully proliferated within the mitochondrial genome, despite the lack of intron functionality significant to the organism (Sethuraman *et al.*, 2009a).

A large number of Group I introns were identified in the three *Ceratocystis* genomes analysed here. Mitochondrial genomes usually harbour Group I and Group II introns, and both types can be present in the same species and sometimes even contained within the same gene (Bonen & Vogel, 2001). Group II introns, are ribozymic, and are thus self-splicing, and like

Group I introns commonly occur in the rRNA, tRNA and mRNA of organelles of fungi (Bonen & Vogel, 2001). However, the fact that Group II type introns were not detected was not unexpected as Group I introns are found at a higher frequency in fungal mitochondrial genomes than Group II introns (Bonen & Vogel, 2001). A distinctive feature of Group I introns is the presence HEGs that encode DNA endonucleases, which are responsible for the cleavage and transfer of the site-specific integration of the particular intron (Hafez *et al.*, 2012).

Various types of Group I introns (Bonen & Vogel, 2001) were observed in the genes encoding cytochrome oxidase, nicotinamide adenine dinucleotide ubiquinone oxidoreductase complexes and the apocytochrome oxidase for all three *Ceratocystis* species investigated. However, not all mitochondrial genes shared the same type of introns for a particular gene. The largest number of introns was found in the *coxI* gene across the three different mitochondrial genomes of *Ceratocystis,* which is consistent with what has been seen before (Cummings *et al.*, 1989; Ferandon *et al.*, 2010). Although the *coxI* gene is a popular choice in plant and animal barcoding (Lane, 2009), the presence of these introns makes this gene problematic for use in fungal barcoding (Schoch *et al.*, 2012). This was abundantly evident from the *Ceratocystis* sequences where the introns differed considerably in size among the different *Ceratocystis* species, thus contributing gene size differences observed.

The Group I introns of the three *Ceratocystis* species were all predicted to harbour ORFs encoding LAGLIDADG or GIY-YIG type homing endonucleases. *Ceratocystis fimbriata s.s* and *C. albifundus* mitochondrial genomes both had an ORF prediction with a LAGLIDADG homing endonuclease similarity that was associated with a Group I type intron. This ORF is within intergenic region between the apocytochrome and ATPase 6 genes. It has highest similarity to a LAGLIDADG homing endonuclease as predicted by both BLAST analysis as well as the InterProScan protein domain prediction database. It is not uncommon for all introns to contain a coding domain sequence in frame with the upstream exon although this usually happens at low frequency (Bonen & Vogel, 2001).

Two ORFs in the *C. fimbriata s.s.* mitochondrial genome encoded genes for mitochondrial plasmid genes that showed high similarity to the sequences of plasmids in *P. anserina*. These ORFs had distinct sizes and the smaller gene was found between the *coxII* and *nad4L* genes, whilst the larger gene disrupted a tRNA cluster. While no EST evidence is available for

these ORFs, the codon usage of these genes can be accommodated by the existing tRNAs in produced by the mitochondrion, suggesting that these genes may still be transcribed or at least originated from plasmids in which the genes could be transcribed in the mitochondrion. This can partially be accounted for by the insertion of a plasmid into this genome as has been seen in literature (Campbell *et al.*, 1999; Hausner, 2003). Fungal mitochondrial plasmids can be found as either autonomously replicating circular or linear-double stranded extra-chromosomal DNA molecules and in certain instances can be linked to mitochondrial instabilities within the genome (Hausner, 2003). The results of a previous study has shown that traces of plasmid-like elements, similar to those observed in *C. fimbriata* and *C. albifundus*, also occur in *Neurospora crassa* and *Podospora anserina* (Hausner, 2003). However in these cases, the plasmid elements are found within mitochondrial genes (Hausner, 2003).

This comparative mitochondrial genomic study has shown that for the three species of *Ceratocystis* that were analysed, both *C. fimbriata* and *C. albifundus* had mitochondrial genomes very similar to that of *C. cacaofunesta* (Ambrosio *et al.*, 2013) which was recently published. This increases our confidence in the characterization of these mitochondrial genomes as the species of *C. fimbriata*, *C. albifundus* and *C. cacaofunesta* all belong to the larger *C. fimbriata* sensu lato complex (Wingfield *et al.*, 2011). The mitochondrial genome of the saprophytic *C. moniliformis* is very different to the aforementioned species, thus also confirming that it is phylogenetically very different to members of the *C. fimbriata* complex, most appropriately residing in a discrete genus (Wingfield *et al.*, 2011).

This is the first report where a comparative study of the mitochondrial genomes was specifically characterized for members of the Microascalean order. Like the mitochondrial genomes of *C. fimbriata s.s.*, *C. albifundus* and *C. moniliformis*, that of *C. cacaofunesta* are also larger than the average for Ascomycota. The latter species also encoded the same suite of genes as the other three species (Ambrosio *et al.*, 2013; Andersson *et al.*, 2003; Campbell *et al.*, 1999; Lang *et al.*, 2012; Pantou *et al.*, 2006; Pantou *et al.*, 2008; van de Sande, 2012; Zivanovic *et al.*, 2005). The results of this study also provide a foundation for the exploitation of these *Ceratocystis* mitochondrial genomes in either complete mitogenomic comparative investigations or by using a combination of different mitochondrial genes to aid in future fungal systematic studies.

2.5. References

Adams K. L., Daley D. O., Qui Y. L., Whelan J., and Plamer J. D. (2000). Repeated, recent and diverse transfer of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408:** 354-357.

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

Al-Reedy R. M., Malireddy R., Dillman C. B., and Kennell J. C. (2012). Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genetics and Biology* **49:** 2-14.

Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., and Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25:** 3389-3402.

Ambrosio A. B., do Nascimento L. C., Oliveira B. V., Teixeria P. J., Tiburcio R. A., Thomazella D. P., Leme A. F., Carazolle M. F., Vidal R. O., Mieczkowski P., Meinhardt L. W., Pereira G. A., and Cabrera O. G. (2013). Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome. *BioMedical Central Genomics* **14:** 1-38.

Andersson S. G. E., Karlberg O., Canback B., and Kurland C. G. (2003). On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society B* **358:** 165-179.

Ballard J. W. O., and Whitlock M. C. (2004). The incomplete natural history of mitochondria. *Journal of Molecular Ecology* **13:** 729–744.

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

Bonitz S. G., Berlani R., Coruzzi G., Li M., Macino G., Noberga F. G., Noberga M. P.,

Thalenfeld B. E., and Tzagoloff A. (1980). Codon recognition rules in yeast mitochondria. *Proceedings of the National Academy of Science, USA* **77:** 3167-3170.

Boore J. L., and Brown W. M. (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinion in Genetics and Development* **8:** 668-674.

Bullerwell C. E., Burger G., and Lang B. F. (2000). "A novel motif for identifying rpS3 homologs in fungal mitochondrial genomes," Elsevier, Amsterdam.

Burger G., Gray M. W., and Lang B. F. (2003). Mitochondrial genomes: anything goes. *Trends in Genetics* **19:** 709-716.

Campbell A., Mrazek J., and Karlin S. (1999). Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proceedings of the National Academy of Science, USA* **96:** 9184-9189.

Capella-Gutierrez S., MMarcet-Houben M., and Gabaldon T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BioMedical Central Biology* **10:** 1-13.

Cummings D. J., Michel F., and McNally K. L. (1989). DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I mitochondrial gene from *Podospora anserina*: a gene with sixteen introns. *Current Genetics* **16:** 381-406.

Duchene A.-M., Pujol C., and Marechal-Drouard L. (2009). Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Current Genetics* **55:** 1-18.

Feau N., Decourcelle T., Husson C., MDesprez-Loustau M.-L., and Dutech C. (2011). Finding single copy genes out of sequenced geneomes for multilocus phylogenetics in non-model fungi. *Public Library of Science ONE* **6:** e18803.

Ferandon C., Moukha S., Callac P., Benedetto J.-P., Castroviejo M., and Barroso G. (2010). The *Agarius bisporus cox1* gene: The longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *Public Library of Science ONE* **5:** 1-8.

Fox T. D. (1987). Natural variation in the genetic code. *Annual Reviews of Genetics* **21:** 67-91.

Gray M. W., Burger G., and Lang B. F. (1999). Mitochondrial Evolution. *Science* **238:** 1476-1481.

Hafez M., Iranpour M., Mullineux S.-T., Sethuraman J., Wosnitza K. M., Lehn P., Kroeker J., Loewen P. C., Reid J., and Hausner G. (2012). Identification of group I introns within the SSU rDNA gene in species of *Ceratocystiopsis* and related taxa. *Fungal Biology* **116:** 98-111.

Halsted B. D., and Fairchild D. G. (1891). Sweet-potato black rot. *Journal of Mycology* **VII:** 1-11.

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

Kawano S., Suzuki T., and Kuroiwa T. (1982). Structural homogeneity of mitochondrial DNA in the mitochondrial nucleoid of *Physarum polycephalum*. *Biochimica et Biophysica Acta* **696:** 290-298.

Kouvelis V. N., Ghikas D. V., and Typas M. A. (2004). The analysis of the complete mitochondrial genome of *Lecanicillium muscarium* (synonym *Verticillium lecanni*) suggests a minimum common gene organization in mtDNAs of *Sordariomycetes*: phylogenetic implications. *Fungal Genetics and Biology* **41:** 930-940.

Lane N. (2009). On the origin of bar codes. *Nature* **462:** 272-274.

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

Lang B. F., Lavrov D., Beck N., and Steinberg S. V. (2012). Mitochondrial tRNA structure, identity, and evolution of the genetic code. *In* "Organelle Genetics" (C. E. Bullerwell, Ed.), pp. 431-474, Springer, Berlin, Heidelberg.

Liu Y., Steenkamp E. T., Brinkmann H., Forget L., Philippe H., and Lang B. F. (2009). Phylogenomic analyses predict sistergroup relationship of nucleariids and Fungi and paraphyly of zygomycetes with significant support. *BioMedical Central Evolutionary Biology* **9:** 1-11.

Lowe T. M., and Eddy S. R. (1997). TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25:** 955-964.

Metzker M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics* **11:** 31-46.

Meusinier I., Singer G. A. C., Landry J.-F., Hickey D. A., Hebert P. D. N., and Hajibabaei (2008). A universal DNA mini-barcode for biodiversity analysis. *BioMedical Central Genomics* **9:** 1-4.

Morris M. J., Wingfield M. J., and De Beer C. (1993). Gummosis and wilt of *Acacia mearnsii* in South Africa caused by *Ceratocystis fimbriata*. *Plant Pathology* **42:** 814-817.

Pantou M. P., Kouvelis V. N., and Typas M. A. (2006). The complete mitochondrial genome of the vascular wilt fungus *Verticillium dahlia*: a novel gene order for *Verticillium* and a diagnostic tool for species identification. *Current Genetics* **50:** 125-136.

Pantou M. P., Kouvelis V. N., and Typas M. A. (2008). The complete mitochondrial genome of *Fusarium oxysporum*: Insights into fungal mitochondrial evolution. *Gene* **419:** 7-15.

Roux J., Dunlop R., and Wingfield M. J. (1999). Susceptibility of elite *Acacia mearnsii* families to Ceratocystis wilt in South Africa. *Journal of Forestry Research* **4:** 187-190.

Roux J., Heath R. N., Labuschange L., Nkuenam G. K., and Wingfield M. J. (2007). Occurrence of the wattle wilt pathogen, *Ceratocystis albifundus* on native South African trees. *Forest Pathology* **37:** 292-302.

Roux J., and Wingfield M. J. (2001). First report of Ceratocystis wilt of *Acacia mearnsii* in

Uganda. *Plant Disease* **85:** 1029.

Roux J., Wyk M. v., Hatting H., and Wingfield M. J. (2004). *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. *Plant Pathology* **53:** 414-421.

Rubio M. A. T., and Alfonzo J. D. (2012). tRNA Modification, editing, and importing in Mitochondria. *In* "Organelle Genetics" (C. E. Bullerwell, Ed.), Springer, Berlin, Heidelberg.

Sankoff D., Leduc G., Antoine N., Paquin B., and Lang B. F. (1992). Gene order comparisons for phylogenetic inference: Evolution of the mitochondrial genome. *Proceedings of the National Academy of Science, USA* **89:** 6575-6579.

Schattner P., Brooks A. N., and Lowe T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33:** W686-W689.

Scheffler I. E. (2001). A century of mitochondrial research: achievements and perspectives. *Mitochondrion* **1:** 3-31.

Schoch C. L., Seifert K. A., Huhndorf S., Robert V., Spouge J. L., Levesque C. A., Chen W., and Consortium F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 6241-6246.

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

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

Smith A. C., Blackshaw J. A., and Roobinson A. J. (2011). MitoMiner: A data warehouse for

mitochondrial proteomics data. *Nucleic Acids Research* **40:** DOI: 10.1093.

van de Sande W. W. J. (2012). Phylogenetic analysis of the complete Mitochondrial genome of *Madurella mycetomatis* confirms if taxonomic position within the order Sordiales. *Public Library of Science ONE* **7:** e38654.

van Wyk M., Roux J., Barnes I., Wingfield B. D., and Wingfield M. J. (2006). Molecular phylogengy of the *Ceratocystis moniliformis* complex and description of *C. tribiliformis* sp. nov. *Fungal Diversity* **21:** 181-201.

van Wyk M., Wingfield B. D., and Wingfield M. J. (2010). Four new *Ceratocystis* spp. associated with wounds on *Eucalyptus, Schizolobium* and *Terminalia* trees in Ecuador. *Fungal Diversity* **46:** 111-131.

White E. E., Foord B. M., and Kinloch B. B. (1996). Genetics of *Croonartium rubicon*. II. Variation in the ribosomal gene cluster. *Canadian Journal of Botany* **74:** 461-468.

Wingfield B. D., van Wyk M., Roos H., and Wingfield M. J. (2011). *Ceratocystis*: Emerging evidence for discrete generic boundaries. In: Ophiostomatiod fungi: expanding frontiers. *(Seifert, K.A and M.J Wingfield, eds.) CBS Biodiversity Series 12* **:** 57-64.

Wingfield M. J., Visser C., Wingfield B. D., and de Beer C. (1996). A new *Ceratocystis* species defined using morphological and ribosomal DNA sequence comparisons. *Systematic and Applied Microbiology* **19:** 191-202.

Zdobnov E. M., and Apweiler R. (2000). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics Application Note* **17:** 847-848.

Zivanovic Y., Wincker P., Vacherie B., Bolotin-Fukuhara M., and Fukuhara H. (2005). Complete nucleotide sequence of the mitochondrial DNA from *Kluyveromyces latis*. *Federation of European Microbiological Yeast Research* **5:** 315-322.

Table 1. Sequence data, coverage and mitochondrial genome size for the three species

Mitochondrial Assembly character	***Ceratocystis fimbriata s.s* CMW 15049**	***Ceratocystis albifundus* CMW 17274**	***Ceratocystis moniliformis* CMW 10134**
Total DNA sequence data obtained	20.98 Gb	13.98 Gb	16.41 Gb
mt DNA (bp) genome	135 104 bp	126 447 bp	110 987 bp
mt DNA coverage	21 772x	2573x	4458x
Single reads used	6 252 466	279 896	265 934
Paired reads used	2 527 432	28 6142	154 902
AT content	72.7%	72.9%	70.3%

Table 2. Sequencing technology employed for each species.

Ceratocystis species	Next generation Sequencing platform employed	Total data generated
Ceratocystis fimbriata s.s	454 Pyrosequencing	1.78 Gb
CMW 15049	Illumina	19.2 Gb
Ceratocystis albifundus	454 Pyrosequencing	0.5 Gb
CMW 17274	Illumina	12.58 Gb
	SOLiD	0.9 Gb
Ceratocystis moniliformis	454 Pyrosequencing	0.78 Gb
CMW 10134	Illumina	15.63 Gb

Table 3. Mitochondrial gene suites depicting the coding domain sequences (base pairs) and the amino acids present across the three *Ceratocystis* isolates.

Mitochondrial genes	***Ceratocystis fimbriata s.s* CMW 15049**		***Ceratocystis albifundus* CMW 17274**		***Ceratocystis moniliformis* CMW 10134**	
ATP synthase subunits	Exon size	Amino Acids	Exon size	Amino Acids	Exon size	Amino Acids
atp6	528 bp	176	528 bp	176	528 bp	176
atp8	198 bp	66	189 bp	63	198 bp	63
atp9	225 bp	75	225 bp	75	225 bp	75
Nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits	Exon size	Amino Acids	Exon size	Amino Acids	Exon size	Amino Acids
nad1	1 116 bp	372	1 116 bp	372	1 116 bp	372
nad2	1 890 bp	630	1 890 bp	630	1 890bp	630
nad3	1 086 bp	362	1 086 bp	362	1 086 bp	362
nad4	1 329 bp	443	1 329 bp	443	1 329 bp	443
nad4L	1 218bp	406	1 218 bp	406	1 218 bp	406
nad5	1 989 bp	663	1 989 bp	663	1 989 bp	663
nad6	819 bp	273	819 bp	273	819 bp	273
Cytochrome oxidase subunits	Exon size	Amino Acids	Exon size	Amino Acids	Exon size	Amino Acids
coxI	1 485 bp	495	1 485 bp	495	1 485 bp	495
coxII	807 bp	269	807 bp	269	807 bp	269
coxIII	810 bp	210	810 bp	210	810 bp	210
Apocytochrome gene	Exon size	Amino Acids	Exon size	Amino Acids	Exon size	Amino Acids
cob	1 173 bp	391	1 173 bp	391	1 173 bp	391
Ribosomal protein	Exon size	Amino Acids	Exon size	Amino Acids	Exon size	Amino Acids
rps3	621	244	621	244	621	244

Table 4. tRNA codon usage for the different amino acids predicted for all three *Ceratocystis* species.

tRNA amino acid	*Ceratocystis fimbriata* CMW 15049	*Ceratocystis albifundus* CMW 17274	*Ceratocystis moniliformis* CMW 10134
Alanine (Ala)	GCA	GCA	GCA
Arginine (Arg)	CGT	CGT	CGT; AGA
Asparagine (Asn)	AAC	AAC	AAC
Aspartic acid (Asp)	GAC	GAC	GAC
Cysteine (Cys)	TGC	TGC	TGC
Glutamic Acid (Glu)	GAA	GAA	GAA
Glutamine (Gln)	CAA	CAA	CAA
Glycine (Gly)	GGA	GGA	GGA
Histidine (His)	CAC	CAC	CAC
Isoleucine (Ile)	ATC (x2)	ATC (x2)	Absent
Leucine (Leu)	TTA	TTA	TTA
Lysine (Lys)	AAA	AAA (x2)	AAA
Methionine (Met)	ATG (x4); ATA	ATG (X4); ATA	ATG (x3)
Phenylalanine (Phe)	TTC	TTC	TTC
Proline (Pro)	CCA	CCA	CCA
Serine (Ser)	TCA (x2); AGC	TCA (x2); AGC	AGC; TCG
Threonine (Thr)	CTA (x2); ACT; CTT	CTA (x2); ACA; CTT	ACA; CTA
Tryptophan (Trp)	TGA (x2)	TGA (x2)	TGA (x2)
Tyrosine (Tyr)	TAC	TAC	TAC
Valine (Val)	GTA	GTA	Absent

Table 5. Summary of the introns present in *Ceratocystis fimbriata s.s* CMW 15049[a].

Gene[b]	Intron Number[c]	*Ceratocystis fimbriata s.s* CMW 15049			
nad5 (7 531 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e] size	Homing endonuclease gene (HEG) prediction[f]
	1	Group ID	1 891 bp	ORF 1. 1 200 bp	LAGLIDADG
	2	Group ID	3 651 bp	ORF 1. 1 356 bp ORF 2. 864 bp	LAGLIDADG LAGLIDADG
coxI (19 415 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e] size	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	3 780 bp	ORF 1. 975 bp ORF 2. 1 038 bp ORF 3. 525 bp	GIY-YIG LAGLIDADG LAGLIDADG
	2	Group IC	3 169 bp	ORF 1. 1 356 bp	LAGLIDADG
	3	Group IB	7 426 bp	ORF 1. 927 bp ORF 2. 603 bp ORF 3. 1 902 bp ORF 4. 999 bp ORF 5. 1 113 bp	LAGLIDADG LAGLIDADG LAGLIDADG LAGLIDADG
	4	Group IB	3 555 bp	ORF 1. 1 239 bp ORF 2. 777 bp ORF 3. 1 059 bp	GIY-YIG LAGLIDADG LAGLIDADG
coxIII (5 170 bp)		Intron Type[d]	Size	Opening Reading Frame (ORF)[e] size	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	4 360 bp	ORF 1. 669 bp ORF 2. 951 bp ORF 3. 1 293 bp	LAGLIDADG LAGLIDADG LAGLIDADG
cob (14 813 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e] size	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	4 174 bp	ORF 1. 1 356 bp ORF 2. 525 bp ORF 3. 2 055 bp	LAGLIDADG LAGLIDADG Reverse Transcriptase polymerase
	2	Group IB	4 516 bp	ORF 1. 560 bp ORF 2. 1 083 bp	LAGLIDADG GIY-YIG

Footnote:

[a] Represented is the group type of the intron/s found in that mitochondrial gene along with the sizes of the intron.

[b] The complete gene size is given in brackets for the respective mitochondrial gene.

[c] The intron number as determined clockwise on the positive stand within each mitochondrial gene.

[d] The intron type as predicted using RNA_{WEASEL}(Schattner *et al.*, 2005).

[e] Open reading frame as determined using CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark) ORF finder and the mitochondrial genetic code 4 (Fox, 1987).

[f] The homing endonuclease genes (HEGs) protein domains as characterized using InterProScan.

Table 6. Summary of the introns present in *Ceratocystis albifundus* CMW 17274[a].

Gene[b]	Intron Number[c]	*Ceratocystis albifundus* CMW 17274			
nad5 (7 843 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e] size	Homing endonuclease gene (HEG) prediction[f]
	1	Group ID	2 837 bp	ORF 1. 1 199 bp ORF 2. 918 bp	LAGLIDADG LAGLIDADG
	2	Group ID	3 017 bp	ORF 1. 543 bp ORF 2. 1 233 bp	LAGLIDADG LAGLIDADG
coxI (16 455 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e] size	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	3 884 bp	ORF 1. 996 bp ORF 2. 1 038 bp	GIY-YIG LAGLIDADG
	2	Group IB	7 531 bp	ORF 1. 567 bp ORF 2. 603 bp ORF 3. 1 221 bp ORF 4. 1 113 bp	LAGLIDADG LAGLIDADG LAGLIDADG LAGLIDADG
	3	Group IB	3 555 bp	ORF 1. 1 161 bp	GIY-YIG
coxIII (5 921 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e]	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	2 015 bp	ORF 1. 1 101 bp	LAGLIDADG
	2	Group I (derived IA)	2 815 bp	ORF 1. 558 bp ORF 2. 1 281 bp	LAGLIDADG LAGLIDADG
cob (7 746 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e]	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	3 844 bp	ORF 1. 333 bp	LAGLIDADG
	2	Group IA	1 652 bp	ORF 1. 1 047 bp	LAGLIDADG
	3	Group IB	1 077 bp	No ORFs predicted	No HEGs predicted

Footnote:

[a] Represented is the group type of the intron/s found in that mitochondrial gene along with the sizes of the intron.

[b] The complete gene size is given in brackets for the respective mitochondrial gene.

[c] The intron number as determined clockwise on the positive stand within each mitochondrial gene.

[d] The intron type as predicted using RNA_{WEASEL}(Schattner *et al.*, 2005).

[e] Open reading frame as determined using CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark) ORF finder and the mitochondrial genetic code 4 (Fox, 1987).

[f] The homing endonuclease genes (HEGs) protein domains as characterized using InterProScan.

Table 7. Summary of the introns present in *Ceratocystis moniliformis* CMW 10134[a].

Gene[b]	**Intron Number[c]**	***Ceratocystis moniliformis* CMW 10134**			
nad5 (5 513 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e]	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	3 524 bp	ORF 1. 1 233 bp	LAGLIDADG
coxI (14 748 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e]	Homing endonuclease gene (HEG) prediction[f]
	1	Group IB	7 423 bp	ORF 1. 567 bp ORF 2. 603 bp ORF 3. 1 221 bp ORF 4. 1 113 bp	LAGLIDADG LAGLIDADG LAGLIDADG LAGLIDADG
	2	Group IB	1 369 bp	No ORFs predicted	No HEGs predicted
	3	Group I (derived B2)	4 471 bp	ORF 1. 717 bp ORF 2. 807 bp ORF 3. 768 bp	LAGLIDADG GIY-YIG LAGLIDADG
coxIII (3 703 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e]	Homing endonuclease gene (HEG) prediction[f]
	1	Group IC2	2 893 bp	ORF 1. 1 185 bp	LAGLIDADG
cob (8 708 bp)		Intron type[d]	Size	Opening Reading Frame (ORF)[e]	Homing endonuclease gene (HEG) prediction[f]
	1	Group ID	3 022 bp	ORF 1. 309 bp	GIY-YIG
	2	Group I (derived A)	2 718 bp	ORF 1. 666 bp ORF 2. 879 bp	LAGLIDADG LAGLIDADG
	3	IB	1 795 bp	No ORFs predicted	No HEGs predicted

Footnote:

[a] Represented is the group type of the intron/s found in that mitochondrial gene along with the sizes of the intron.

[b] The complete gene size is given in brackets for the respective mitochondrial gene.

[c] The intron number as determined clockwise on the positive stand within each mitochondrial gene.

[d] The intron type as predicted using RNA_{WEASEL}(Schattner *et al.*, 2005).

[e] Open reading frame as determined using CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark) ORF finder and the mitochondrial genetic code 4 (Fox, 1987).

[f] The homing endonuclease genes (HEGs) protein domains as characterized using InterProScan.

Figure 1. Schematic representation of the mitochondrion of *Ceratocystis fimbriata s.s* CMW 15049.

The gene arrangement of the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment (135 104 bp). All genes were transcribed on the positive stand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; pal 2-1 plasmid genes are purple; intergenic LAGLIDADG homing endonuclease (HE) is bright blue; the *rns* is brown and the *rnnl* is orange. Hypothetical proteins are shown in pink.

Figure 1.

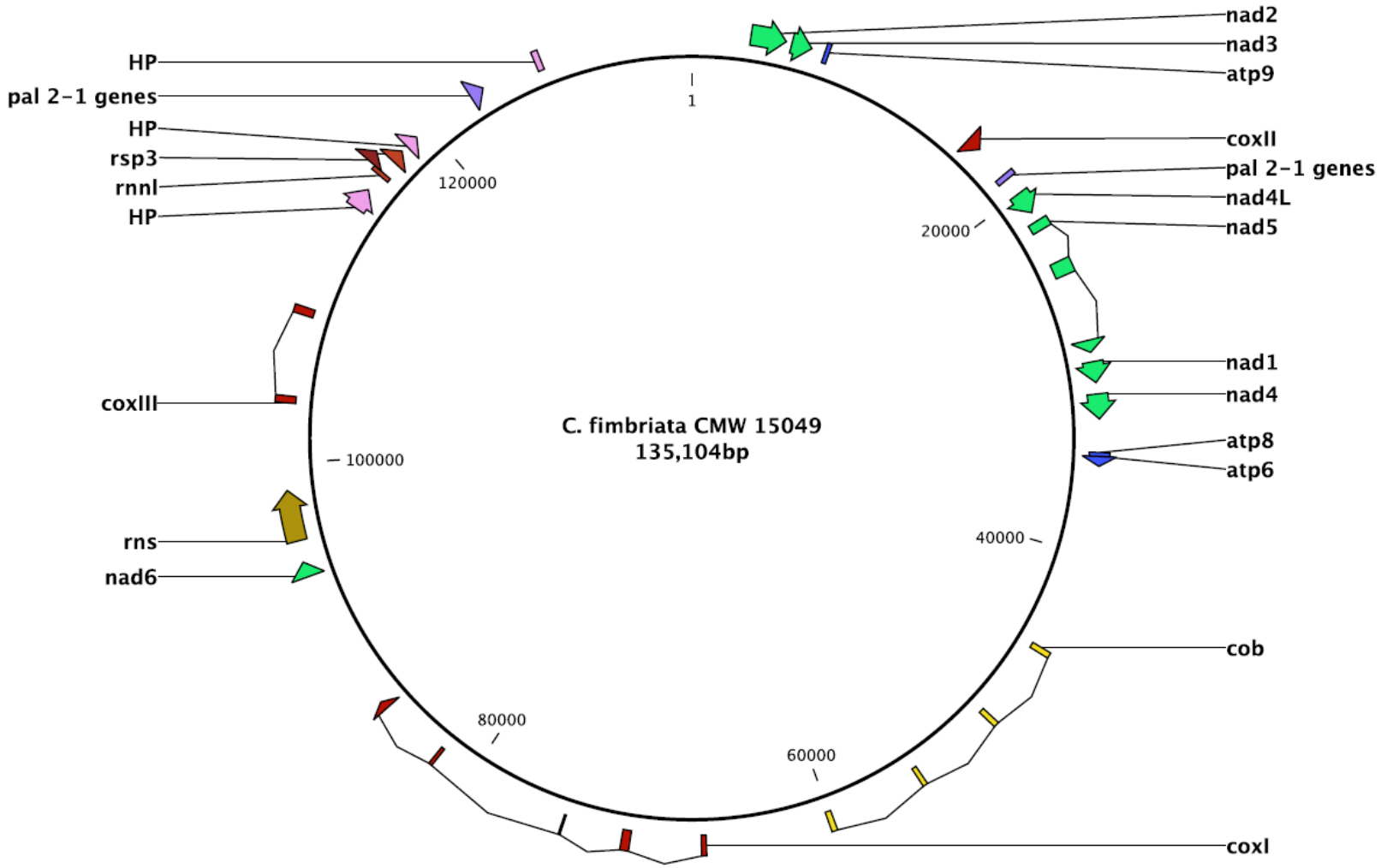


Figure 2. The 31 tRNA for *Ceratocystis fimbriata s.s* CMW 15049.

The tRNA clusters were identified using $_{T}RNA_{SCAN}$-SE (Lowe & Eddy, 1997). The 15 core mitochondrial genes, *rnnl* and *rns* are included in the figure to provide context.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; pal 2-1 plasmid genes are purple; intergenic LAGLIDADG homing endonuclease is bright blue; the *rns* is brown and the *rnnl* is orange.

Figure 2.

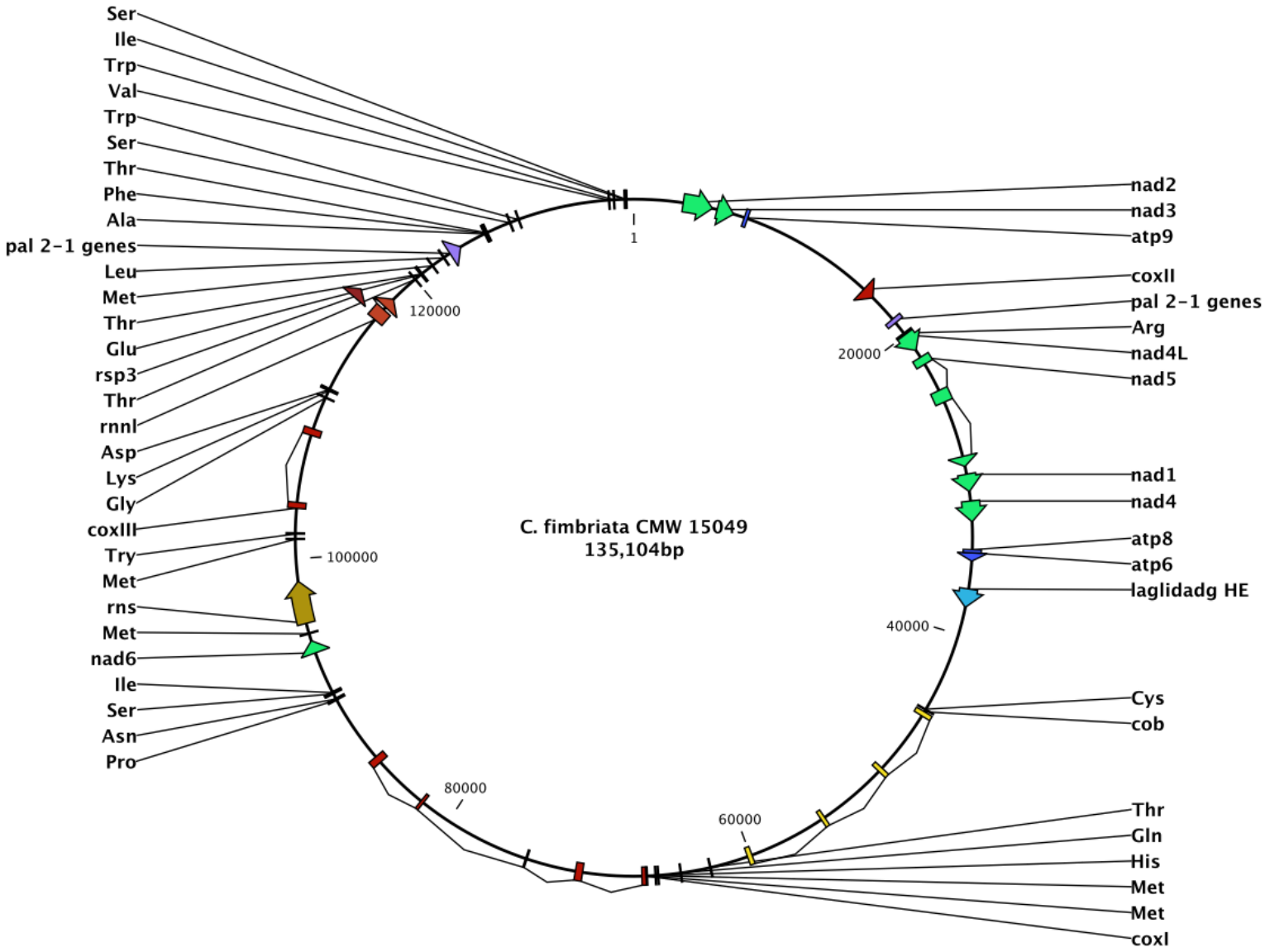


Figure 3. Schematic representation of mitochondrial genes and introns of *Ceratocystis fimbriata s.s* CMW 15049.

The gene arrangement and intron presence within the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment (135 104 bp). All genes were transcribed on the positive stand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; intergenic LAGLIDADG homing endonuclease (HE) is bright blue; the *rns* is brown and the *rnnl* is orange. Introns were predicted as Group I types where Group IB is light purple; Group ID is aqua blue and Group IC2 is pink. Homing endonuclease genes (HEGs) were predicted for all open reading frames found within the intron. The LAGLIDADG HEGs are shown as grey whilst the GIY-YIG HEGs are brown.

Figure 3.

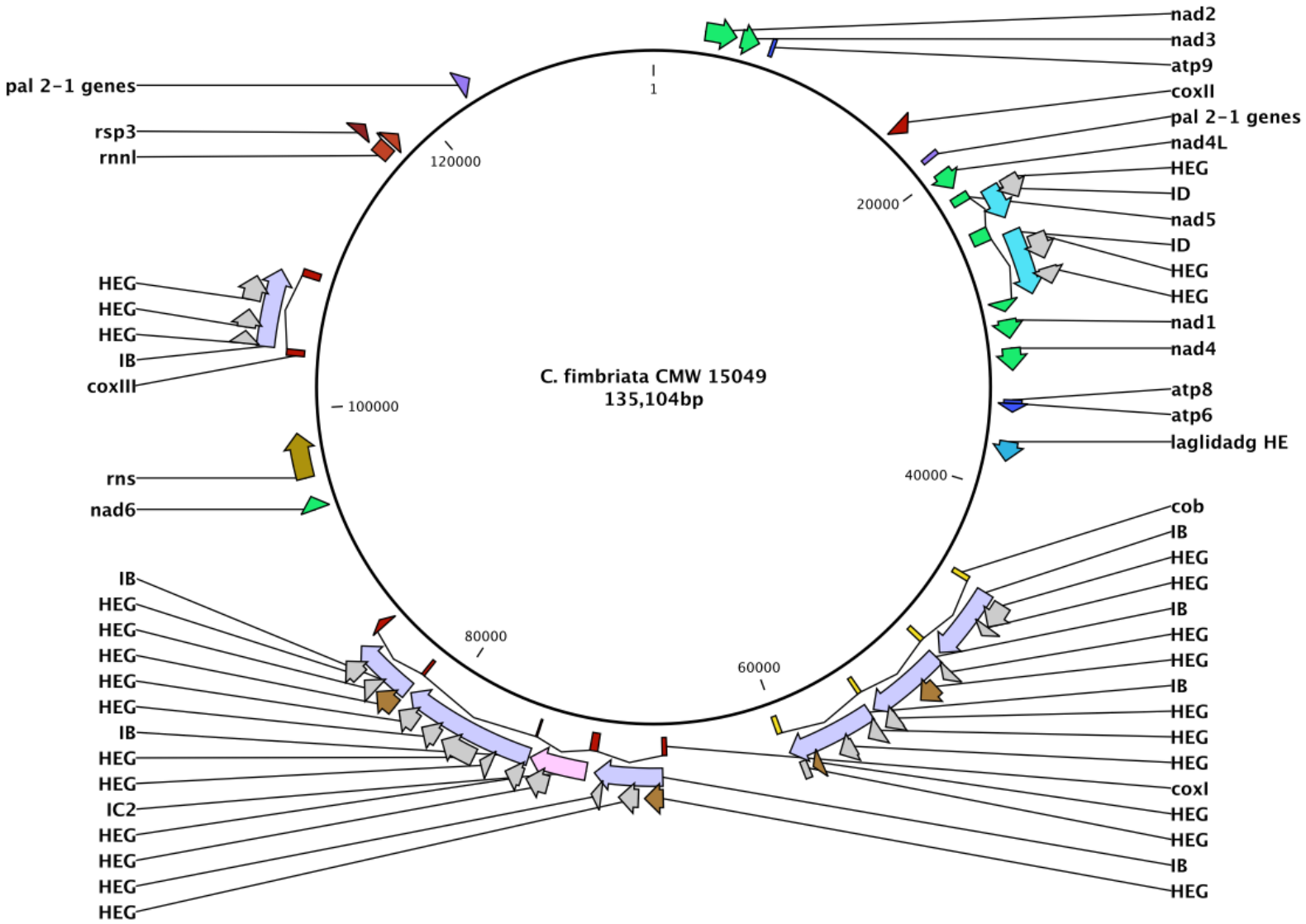


Figure 4. Schematic representation of the mitochondrion of *Ceratocystis albifundus* CMW 17274.

The gene arrangement of the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment (12 6447 bp). All genes were transcribed on the positive stand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; intergenic LAGLIDADG homing endonuclease (HE) is bright blue; the *rns* is brown and the *rnnl* is orange.

Figure 4.

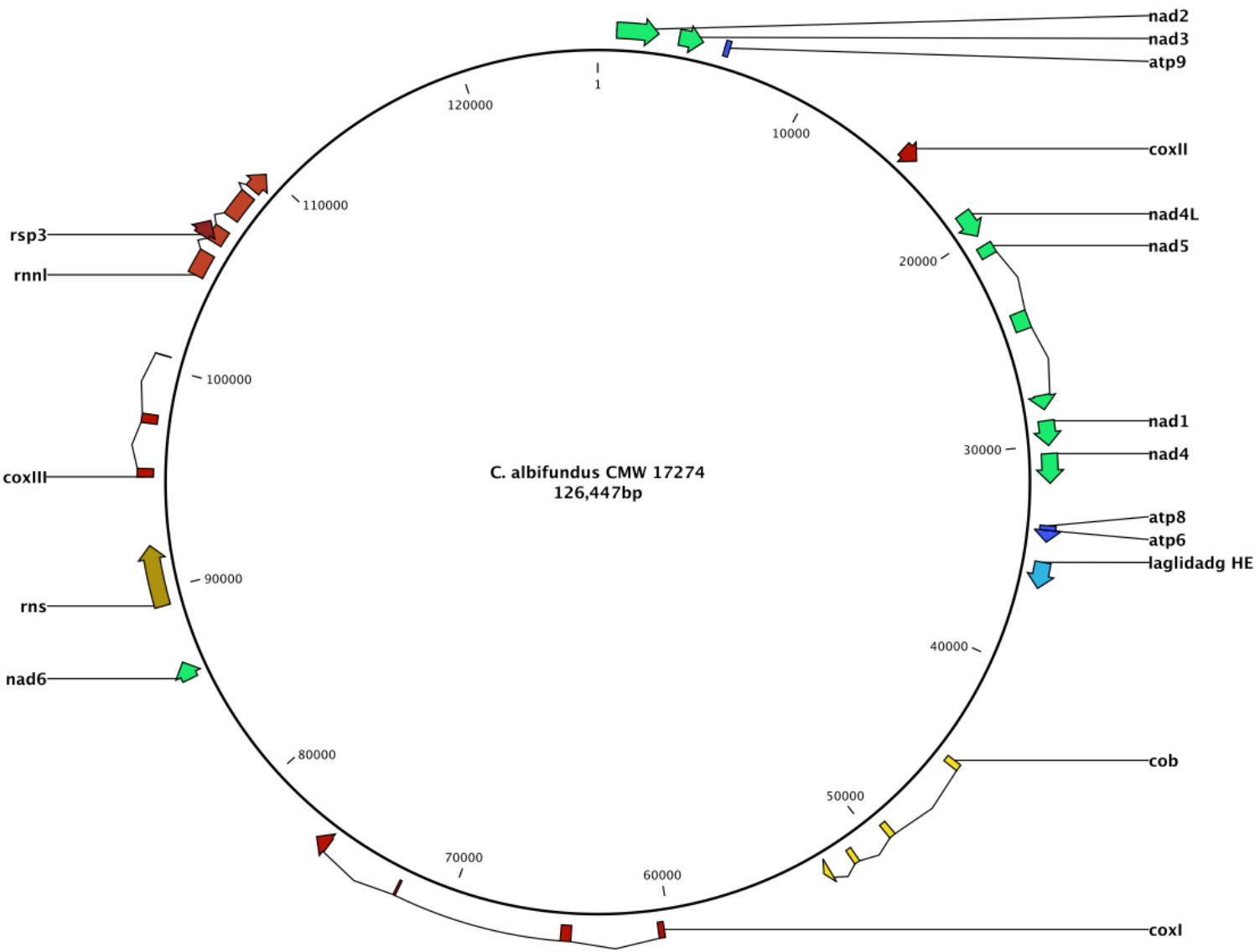


Figure 5. The 33 tRNA for *Ceratocystis albifundus* CMW 17274.

The tRNA clusters were identified using $_{T}RNA_{SCAN}$-SE (Lowe & Eddy, 1997). The 15 core mitochondrial genes, *rnnl* and *rns* are included in the figure to provide context. The five methionine (Met) tRNA's are highlighted as this amino acid is over represented. There were also three tRNAs for Serine (Ser) and four tRNAs for Threonine (Thr).

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* gene are red; intergenic LAGLIDADG homing endonuclease (HE) is bright blue; the *rns* is brown and the *rnnl* is orange.

Figure 5.

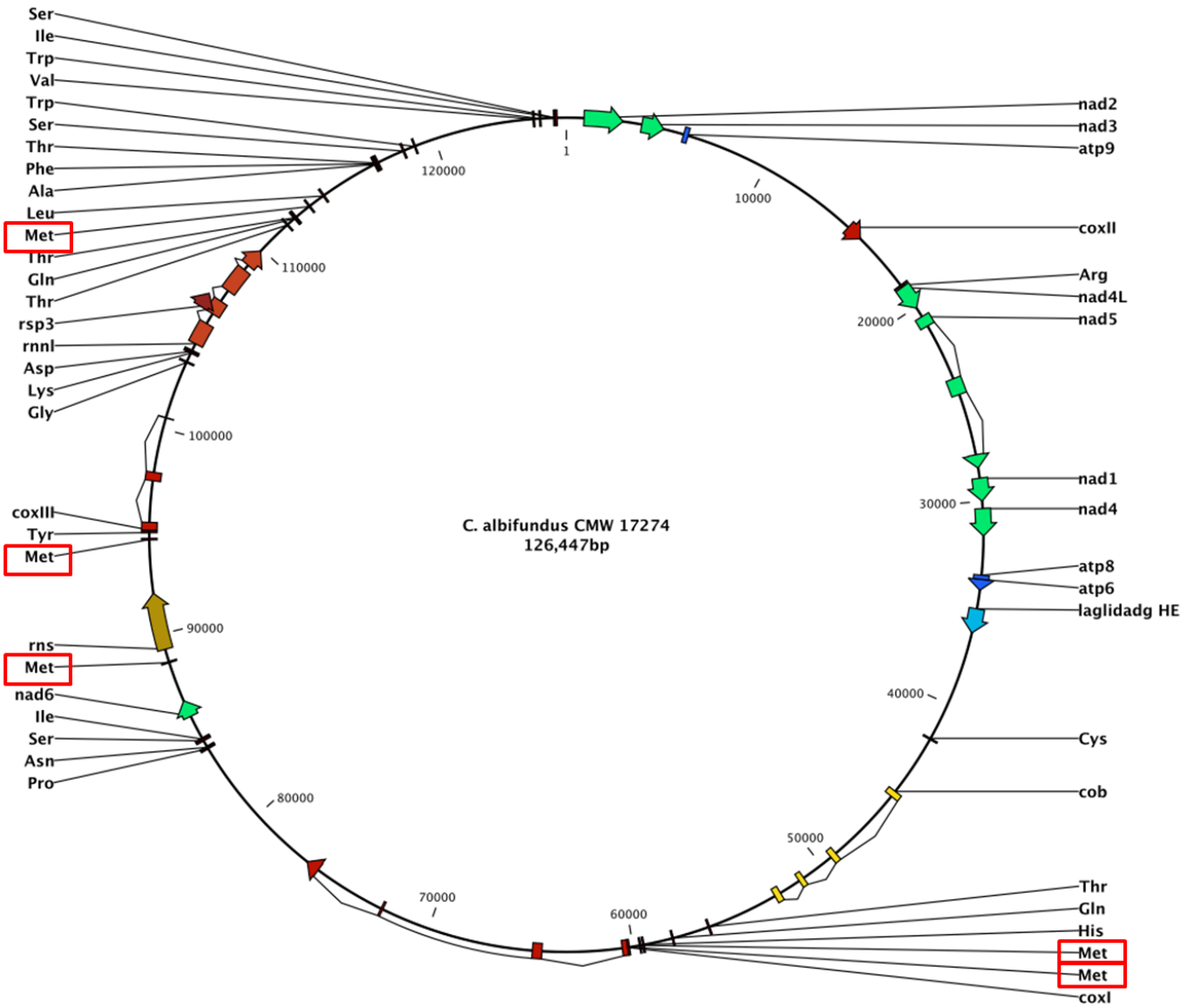


Figure 6. Schematic representation of mitochondrial genes and introns of *Ceratocystis albifundus* CMW 17274.

The gene arrangement and intron presence within the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment (126 447 bp). All genes were transcribed on the positive stand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; intergenic LAGLIDADG homing endonuclease is bright blue; the *rns* is brown and the *rnnl* is orange. Introns were predicted as Group I types where Group IA is bright green; Group I (derived A) is bright blue; Group IB is navy and Group ID is aqua blue. Homing endonuclease genes (HEGs) were predicted for all open reading frames found within the intron. The LAGLIDADG HEGs are shown as dark purple whilst the GIY-YIG HEGs are dark brown.

Figure 6.

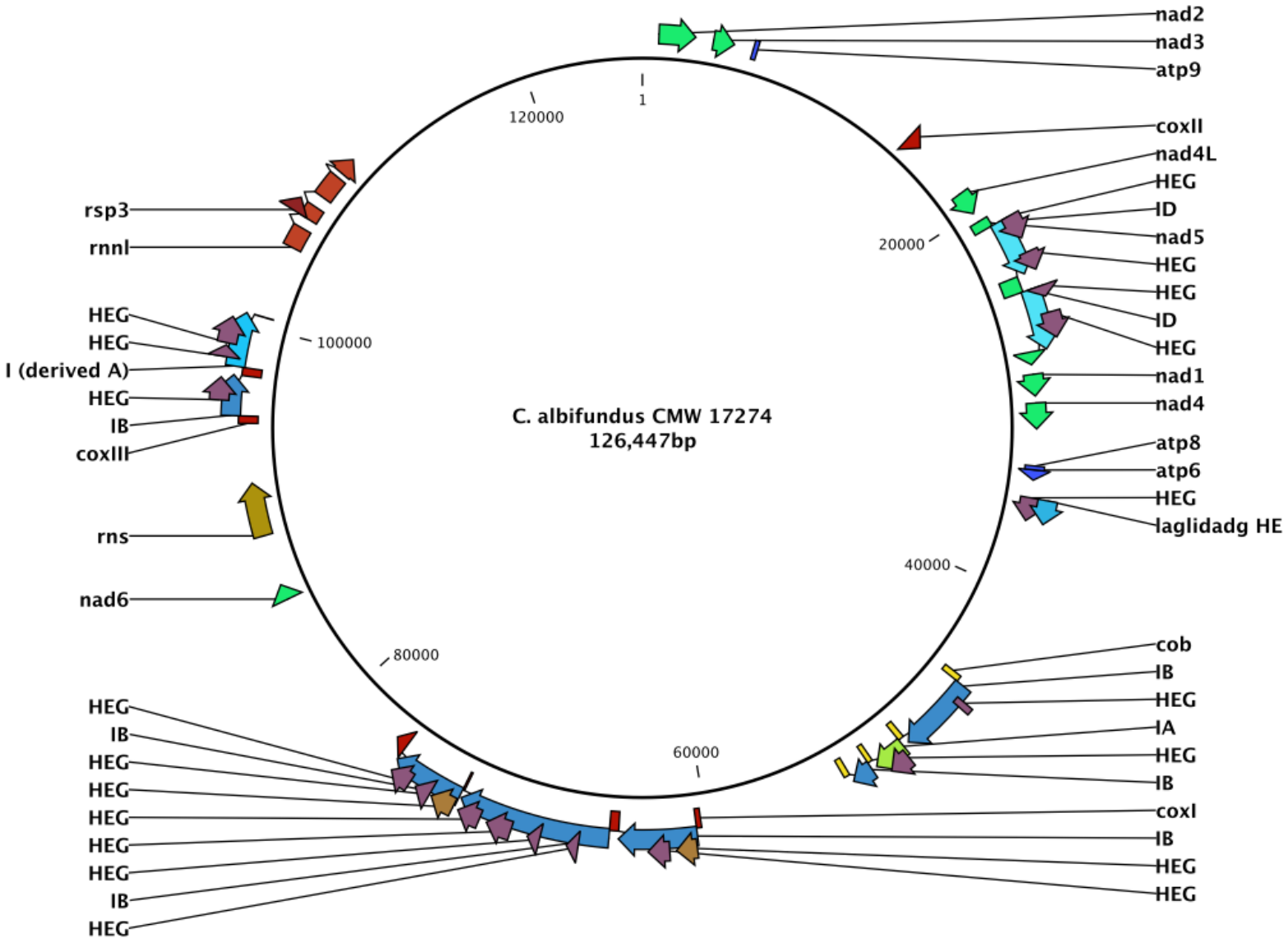


Figure 7. Schematic representation of the mitochondrion of *Ceratocystis moniliformis* CMW 10134.

The gene arrangement of the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment (110 987 bp). All genes were transcribed on the positive stand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; the *rns* is brown and the *rnnl* is orange.

Figure 7.

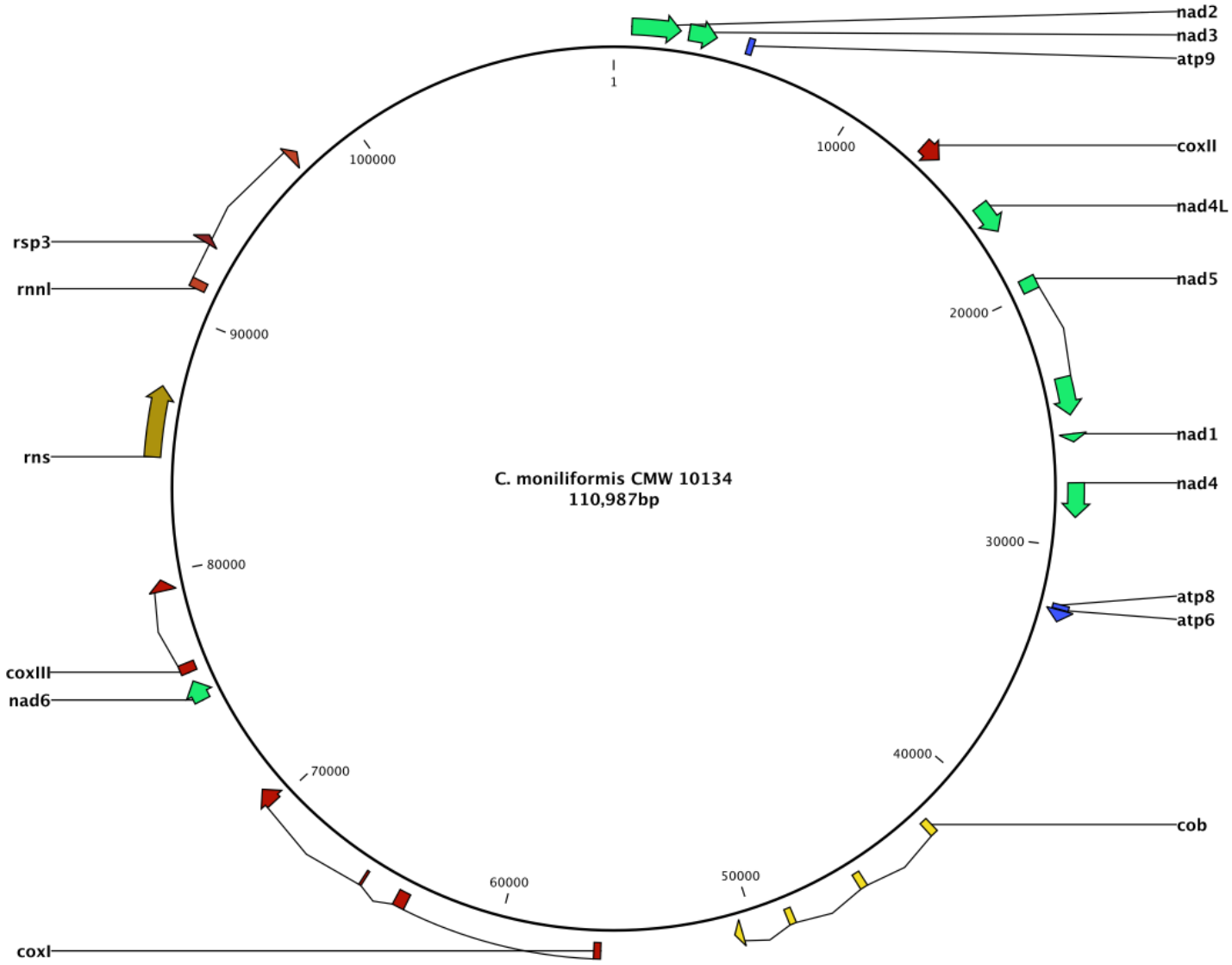


Figure 8. The 24 tRNA for *Ceratocystis moniliformis* CMW 10134.

The tRNA clusters were identified using $_{T}RNA_{SCAN}$-SE (Lowe & Eddy, 1997). The 15 core mitochondrial genes, *rnnl* and *rns* are included in the figure to provide context. The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; the *rns* is brown and the *rnnl* is orange.

Figure 8.

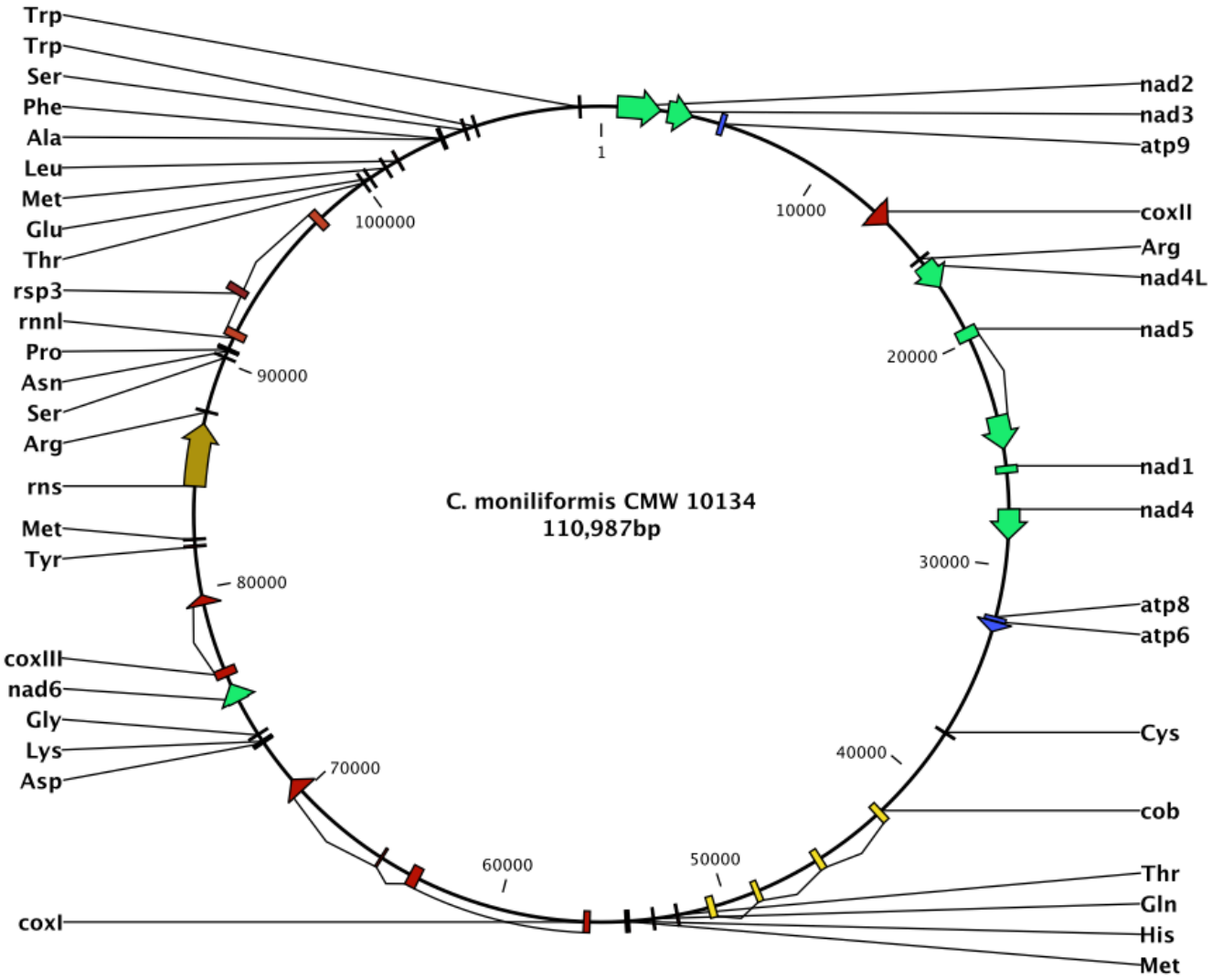
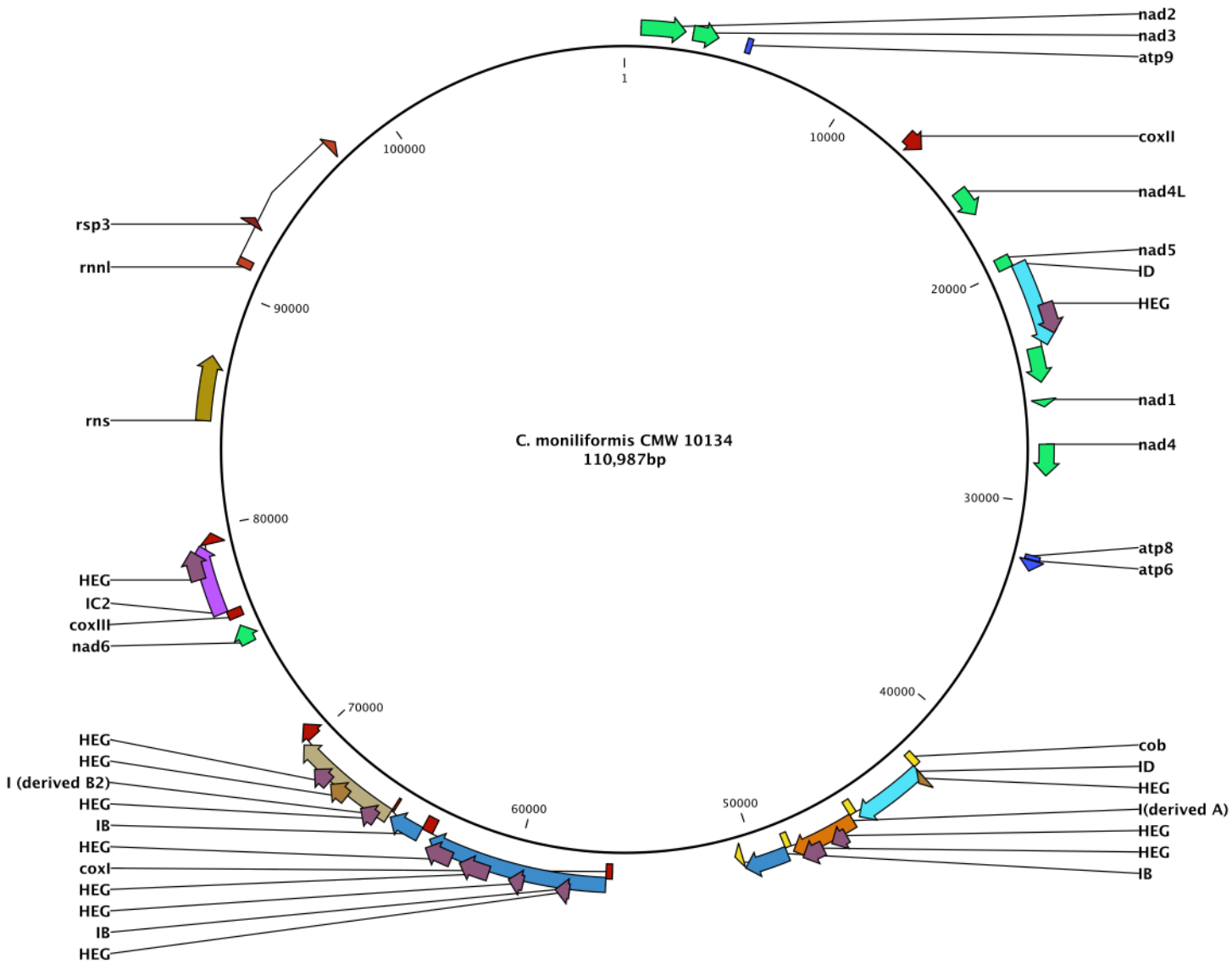


Figure 9. Schematic representation of mitochondrial genes and introns of *Ceratocystis moniliformis* CMW 10134.

The gene arrangement and intron presence within the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment (110 987 bp). All genes were transcribed on the positive stand.

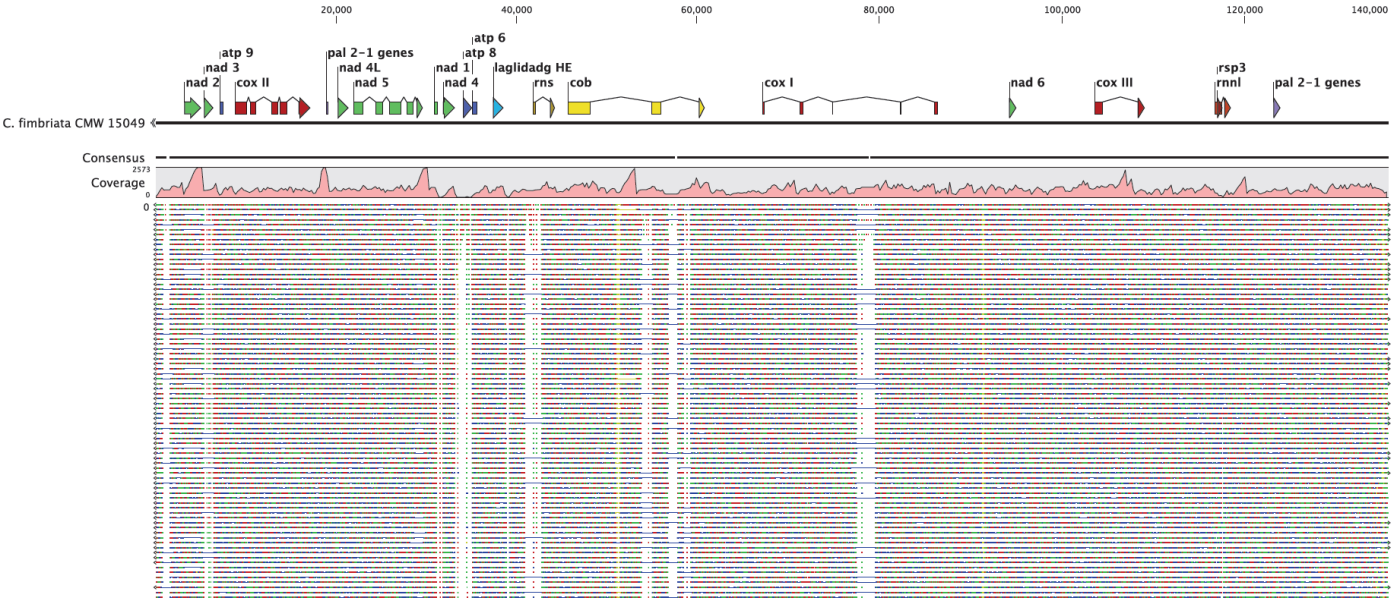
The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; the *rns* is brown and the *rnnl* is orange. Introns were predicted as Group I types where Group I (derived A) is bright orange; Group IB is navy; Group I (derived B2) is charcoal; Group ID is aqua blue and Group IC2 is light purple. Homing endonuclease genes (HEGs) were predicted for all open reading frames found within the intron. The LAGLIDADG HEGs are shown as dark purple whilst the GIY-YIG HEGs are dark brown.

Figure 9.



Supplementary Data

Figure S1. Mapping of *C. albifundus* CMW 17274 reads against the *C. fimbriata s.s* CMW 15049 reference contig.



Chapter Three

Mitochondrial genome variation within *Ceratocystis albifundus*

Abstract

Ceratocystis species include a number of disease causing plant pathogens, affecting the commercial forestry industry which cause great economic losses. *Ceratocystis albifundus* is a tree pathogen of particular interest as a consequence of its broad host range. This species is thought to be native to Southern Africa but its country of origin is unclear. Understanding the genetic diversity and origin of this species is an important focus for Acacia plantation management. Mitochondrial genomes offer potential markers to investigate pathogen spread and origin. The aim of this study was to characterize the mitochondrial genomes of six *C. albifundus* isolates in order to determine the variation present in the mitochondrial genomes of a number of geographically disparate isolates. Three isolates were selected from South Africa, namely KwaZulu Natal, Gauteng and Limpopo provinces and three from elsewhere in Africa, Kenya, Tanzania and Zambia. The mitochondrial genomes for all six *C. albifundus* isolates showed high conservation and synteny in their gene architecture. All the *C. albifundus* mitochondrial genomes were observed to be circular but varied in their absolute mitochondrial gene sizes. The *C. albifundus* isolates from Gauteng South Africa, the Kenyan and Tanzanian isolates were the largest at 126 447 bp in size. The KwaZulu Natal isolate was ten bases smaller at 126 437 bp whilst the Limpopo and Zambian isolates of *C. albifundus* were the smallest at 126 353 bp. Variation in the tRNA genes varied from 30 to 33, the differences observed were in the number of Methionine, Serine and Threonine tRNA genes present. Only Group I type introns containing homing endonuclease genes were observed. The Tanzanian and Kenyan isolates were identical in sequence. The exon regions for all isolates compared had a greater similarity than the combination of the intronic and exonic regions. The intergenic comparison revealed that the KwaZulu Natal isolate was most dissimilar to all other isolates investigated. This report represents the first intraspecies mitochondrial genomic analysis for *Ceratocystis* and provides the platform for the identification of mitochondrial gene regions which can potentially allow for the development of markers to trace the movement of *C. albifundus* from native trees onto non-native plantation species.

3.1. Introduction

A number of wilt and canker pathogens are found within the genus *Ceratocystis* (Wingfield *et al.*, 1993). The members of this genus have a diverse host range affecting both native and non-native plantation forests globally (Roux *et al.*, 2007). Many of these species have been characterized as pathogens affecting the commercial forestry industry (Montoya & Wingfield, 2006) in South Africa and globally (Barnes *et al.*, 2005; Chen *et al.*, 2013; Kamgan *et al.*, 2008; Kamgan *et al.*, 2012; Roux *et al.*, 2004). Morphological and DNA-based characters have shown that this genus is separated into three large clades, represented by the *Ceratocystis fimbriata sensu lato*, *Ceratocystis moniliformis* sensu lato and *Ceratocystis coerulescens* sensu lato (Wingfield *et al.*, 2011). The *C. fimbriata sensu lato* clade is of considerable economic importance. It includes pathogens of commercial crops such as sugarcane, (Lewton-Brain, 1907), pineapple (De Seynes, 1886) sweet potato (Halsted & Fairchild, 1891), mango (van Wyk *et al.*, 2007), coffee, cocoa and citrus plantations (Grylls & Seifert, 1993; van Wyk *et al.*, 2010). The forestry species affected include varieties of eucalypt (Roux *et al.*, 2004), wattle and plane plantations which have fallen susceptible to *Ceratocystis* diseases (Barnes *et al.*, 2005; Chen *et al.*, 2013; Kamgan *et al.*, 2012; Roux *et al.*, 2004; Wingfield *et al.*, 2011).

Ceratocystis albifundus was first identified as the disease causing agent of *Acacia mearnsii* in South African plantations in 1996 (Roux *et al.*, 1999; Roux *et al.*, 2001; Wingfield *et al.*, 1996). South African savannah ecosystems have also fallen susceptible to *C. albifundus* diseases (Heath *et al.*, 2009). *Ceratocystis albifundus* has also been reported to occur on *Protea* species native to South Africa (Groter, 1977; Roux *et al.*, 2001; Wingfield *et al.*, 1996). From a global perspective, plantation species affected by *C. albifundus* include the silver and green wattle species found in Australia [*A. dealbata* and *A. decurrens* respectively] (Barnes *et al.*, 2005). Characteristic symptoms associated with this pathogen are the formation of cankers and bark lesions. Other symptoms include serve decolourization of the sapwood, dieback and ultimately tree death (Barnes *et al.*, 2005).

C. albifundus is believed to be native to southern Africa (Roux *et al.*, 2001). This notion is supported by restriction length polymorphism analysis of mitochondrial genes and nuclear gene profiling (Roux *et al.*, 2001; Witthuhn *et al.*, 1999). The nuclear gene profiling was achieved using a multi-locus microsatellite marker $(CAT)_5$ (DeScenzo & Harrington, 1994) and resulted in unique nuclear fingerprints for the different *C. albifundus* isolates investigated

in that study (Roux *et al.*, 2001). In addition, the comparison of populations of isolates from South Africa and Uganda also suggested that *C. albifundus* is native to southern Africa (Barnes *et al.*, 2005). None of the population studies carried out thus far has shown conclusive evidence as to which region in southern Africa is the centre of origin of this fungus. This is mainly due to the low resolving power of the markers employed, because the previously used nuclear gene sequences and the microsatellites have limited variation.

Mitochondrial markers probably represent viable alternatives to the molecular markers employed in previous studies on the population genetics of *C. albifundus*. This is due to mitochondrial genomes typically exhibiting high degrees of variation, even among closely related individuals and members of the same species as was shown in *Fusarium* species (Al-Reedy *et al.*, 2012). It can be assumed that the mitochondrion of *C. albifundus* is an autonomously replicating entity and is inherited in a uni-parental fashion (Barr *et al.*, 2005) thereby making allowing it to be a potentially powerful tool in investigating the movement of pathogens. This is particularly important for *C. albifundus* because the pathogen is believed to have made a significant host jumps from native African tree species to a non-native species of *Acacia*. The reasoning for such an assumption lies in the fact that not all mitochondria are inherited uni-parentally, and forms of bi-parental inheritance have been observed in fungi, however in low frequencies (Barr *et al.*, 2005).

Fungal mitochondrial genomes can vary greatly in size, ranging from as small as 24 kb to as large as 100kb (Burger *et al.*, 2003). The mitochondrial genomes of Ascomycota characteristically have 15 coding gene regions, two ribosomal RNA gene units and approximately 25 transport RNA [tRNA] gene units (Lang *et al.*, 1999). However, significant variation also has been observed in gene content, genome architecture and intergenic regions in their composition within the mitochondrial genomes for both model and non-model Ascomycota (Juhasz *et al.*, 2008; Kerscher *et al.*, 2001; Kouvelis *et al.*, 2004; Pantou *et al.*, 2006; Torriani *et al.*, 2008; Woo *et al.*, 2003). Mitochondria are directly involved in electron transportation and oxidative phosphorylation pathways requiring ATPases as well as the apocytochrome B and the cytochrome C oxidase complex genes, all of which ultimately produce ATP for cellular functionality (Lang *et al.*, 1999). Seven nicotinamide adenine dinucleotide ubiquinone oxidoreductase subunits as well as the ribosomal protein S3 are also present in the repertoire of genes found within the mitochondrial genome (Lang *et al.*, 1999).

The mitochondrial genomes of four *Ceratocystis* species have been sequenced, i.e., *C. cacaofunesta*, *C. fimbriata*, *C. albifundus* and *C. moniliformis* (Ambrosio *et al.*, 2013; Naidoo *et al.*, 2013). These studies however investigated only a single isolate of the various species of *Ceratocystis*, thus establishment of the value of mitochondrial genes in populations studies of different isolates of *C. albifundus* required that each isolates' mitochondrial genome be determined. The aim of this study was therefore to sequence the mitochondrial genomes for additional strains of this fungus and to compare these genomes with the ultimate goal of aiding the identification of specific regions that could be exploited for population genetic studies. To achieve these objectives, the mitochondrial genomes of six *C. albifundus* isolates obtained from different hosts and geographic origins were sequenced, after which the genomes were fully characterized in order to determine the degree of variation associated with the various regions.

3.2. Materials and Methods

3.2.1 Fungal Isolates and DNA isolation

Six isolates of *Ceratocystis albifundus* [CMW 17274, CMW 4068, CMW 17620, CMW 248685, CMW 13980 and CMW 24860] that vary in terms of host and geographic origin were used in this study [Table 1]. *Ceratocystis albifundus* CMW 17274 has been used previously and its mitochondrial genome fully annotated and characterized (Naidoo *et al.*, 2013). These isolates were grown at 25°C on 2% malt extract agar (MEA: 20% w/v; Biolab, Midrand, South Africa) supplemented with thymine. Isolates are maintained in the culture collection of the Tree Protection Cooperative Programme, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa.

A single hyphal tip from each isolate was individually plated onto fresh MEA medium and allowed to grow for two weeks prior to DNA isolation. High quality total genomic DNA was extracted using the Phenol-chloroform method as previously described by Roux and colleagues (Roux *et al.*, 2004). The DNA extracts from approximately 50 individual petri plates were pooled for each isolate in order to produce approximately 5µg of DNA for sequencing. DNA concentrations were quantified using a Qubit Flourometer (Invitrogen, USA).

3.2.2 Next generation DNA Sequencing

The DNA from *C. albifundus* isolates was sequenced using an Illumina (MiSeq Sequencer) next generation sequencing platform at the UC Davis Sequencing Facility in California. Mate-paired libraries of 300 and 500bp fragments were prepared and sequenced [Table 2].

3.2.3 Sequence data assembly, analysis and annotation

The CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark) was used to assemble and annotate the mitochondrial DNA sequences. A *de novo* assembly was generated using the 454 Pyrosequencing long reads [>450bp] data libraries for *C. albifundus*. The resulting consensus sequence for *C. albifundus* CMW 17274 was then used as a reference mitochondrial genome for the other five species. The total reads for each isolate was individually mapped against the *C. albifundus* CMW 17274 reference sequence in order to obtain a single consensus contig. This resulted in a single consensus mitochondrial sequence contig been generated for each of the different isolates.

Each assembled mitochondrial genome for all the *C. albifundus* isolates was then subjected to an open reading frame prediction algorithm (CLC-BIO Genomics Workbench) which identified the ORFs on the basis of the translational mitochondrial genetic code 4 (Fox, 1987). Individual ORFs were then subjected to BLASTn (Altschul *et al.*, 1997) similarity comparisons against the nucleotide database of the National Center for Biotechnology Information [NCBI; www.ncbi.nlm.nih.gov]. The tRNA genes were identified using $_{T}RNA_{SCAN}$-SE version 1.21 software (Lowe & Eddy, 1997). RNA_{WEASEL} (http://megasun.bch.umontreal.ca/RNAweasel) aided in the identification and curation of the intron and rRNA gene units present (Schattner *et al.*, 2005). Intronic ORFs that contained endonuclease genes as per the mitochondrial genetic code 4 (Fox, 1987) were identified by BLAST searches (Altschul *et al.*, 1997). Manual annotation of all the ORFs and tRNAs were done using CLCBioGenomics workbench using complete gene sequences as well as amino acid alignments from the annotated genomes of *Fusarium oxysporium*, *Podospora anserina* and *Verticillium dahlia* as a base line for comparison. The codon usage was confirmed using a freeware prediction programme [http//www.geneinfinity.org]. The exon regions for the 14 core mitochondrial genes for each isolate were aligned. These alignments

were then subjected to an individual as well as a combined pairwise comparison carried out using CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark). The intergenic regions were also compared by aligning all the isolates of *C. albifundus*. The same approach was undertaken for the combined exon and introns for each of the 14 core mitochondrial genes to determine the overall sequence similarity percentage as well as the number of differences in the alignment position where the two sequences differed.

3.3. Results

3.3.1 Sequencing data

The *C. albifundus* CMW 17274 mitochondrial genome was established to be 126 447 bp in size and was generated from a total of 14GB of DNA sequence data [Table 2]. The total DNA sequence data obtained for the other five *C. albifundus* isolates of approximately 6 to 8GB was mapped to the reference *C. albifundus* CMW 17274 isolate [Table 2]. The observed single reads which mapped against the *C. albifundus* CMW 17274 mitochondrial genome ranged between approximately 640 000 and 1250 000, whilst the paired reads ranged between 700 000 and 3435 000 respectively [Table 2]. The overall coverage of the consensus mitochondrial sequence for each isolate was exceptionally high, ranging from approximately 117 000 to 135 000 times for the different isolates [Table 2], however the overall coverage of the total mitochondrial genome was between 1 to 5 times. The exon regions had the highest paired coverage as these as are conserved sequences across the mitochondrial genomes, whilst the intergenic sequences had a lower coverage for the different isolates [Table 2].

3.3.2 Annotation of all *C. albifundus* mitochondrial genomes

All the mitochondrial gene regions had a high coverage, thus ensuring confidence in the consensus mitochondrial genome was sufficient for the isolates CMW 4068, CMW 17620, CMW 24860, CMW 24685 and CMW 13980. There was a strong AT bias for all the *C. albifundus* isolates with a 73% AT content in the complete genome content [Table 2]. The *C. albifundus* mitochondrial genomes were all circular and with CMW 24685 and CMW 24860 being the largest at 126 447 bp in size, the same as CMW 17274, CMW 4068 was ten bases

smaller at 126 437 bp whilst CMW 17620 and CMW 13980 were the smallest at 126 353 bp [Table 1].

All mitochondrial genomes contained the same 14 core protein coding genes, as well as the *rns* and *rnnl*, and the *rps3* gene within the single *rnnl* intron [Figure 1]. The protein coding genes included *atp6*, *atp8*, and *atp9*, *coxI*, *coxII* and *coxIII*, *cob*, *nad1-6* and *nad4L*); and apocytochrome B. All genes presented in the same gene order and were transcribed from the same strand. In each of the mitochondrial genomes for the different *C. albifundus* isolates one predicted intergenic homing endonuclease gene with a LAGLIDADG motif was observed between the *atp6* and *cob* genes [Figure 2].

The total tRNA genes present for each *C. albifundus* isolate ranged between 31 and 33 [Table 3]. All the tRNA genes corresponded to the minimum 20 amino acids needed to code for all the mitochondrial genes within the genomes. There was an over-representation of the tRNA gene which contains the anticodon for the amino acid methionine across each of the *C. albifundus* mitochondrial genomes. The tRNA gene for Methionine was present five times in the mitochondrial genomes of *C. albifundus* isolate CMW 17274, CMW 4068 and CMW 17620 and present four times in the mitochondrial genomes of *C. albifundus* CMW 24585, CMW 24860 and CMW 13980 [Table 3]. There were also differences in the number of times the tRNA genes for Serine (Ser) and Threonine (Thr) were present in all isolates of *C. albifundus* mitochondrial genomes [Table 3]. Overall there was a strong bias towards codons ending with an A for all *C. albifundus* mitochondrial genomes [Table 3].

For all the intronic regions, only Group I type introns were predicted across all the mitochondrial genomes of the six *C. albifundus* isolates. These introns occurred in the *nad5*, *coxI*, *coxIII* and *cob* genes [Figure 2]. Two introns were found within the *nad5* gene, both of which were predicted as Group ID types and contained putative homing endonuclease genes (HEGs) encoding LAGLIDADG protein motif [Figure 2]. The *cob* gene had three introns, two of which were Group IA and IB types with HEG ORFs all containing LAGLIDADG protein motifs, while the third was also Group type IB, however it lacked an in-frame HEG ORF. The introns across all *C. albifundus* isolates were the same size and with the same intron Group type predicted. The HEGs were either LAGLIDADG or GIY-YIG protein motifs [Figure 2]. Two introns were found in the *coxIII* gene, both containing HEGs with a

LAGLIDADG domain, where the second intron was biorfic, containing two ORFs within the Group I type intron [Figure 2].

The overall sequence identity between all isolates was determined using pairwise evaluation in the CLC Bio Genomics workbench software [version 5.9] (CLC Bio, Århus, Denmark) [Figure 3]. When only coding regions [exons] were compared a very high similarity was seen between all genomes. *Ceratocystis albifundus* CMW 17274 from Gauteng was most similar to the *C. albifundus* CMW 4068 isolate from KwaZulu Natal [Figure 3]. The sequence identity for *C. albifundus* CMW 24860 from Tanzania had the highest overall similarity to *C. albifundus* CMW 24685 (100%) in comparison to all other isolates [Figure 3].

When comparing the combined intron and exon regions the percentage identity was exactly the same for *C. albifundus* CMW 24860 and CMW 24685, been the Kenyan and Tanzanian isolates respectively. An averaged nucleotide difference for each isolate was pairwise compared and the percentage similarities as well as difference among these isolates showed the major difference occurred in the total mitochondrial genome size. A ten base difference in the size of the Gauteng *C. albifundus* CMW 17374, CMW 24865 (Kenya) and CMW 24860 (Tanzania) (which are the largest mitochondrial genomes sizes across all the isolates) compared to the *C. albifundus* CMW 4068 (KwaZulu Natal) isolate was observed in the intergenic region between the *atp9* and *coxII* mitochondrial genes.

The total sequence coding domains [exons] in conjunction with the introns for the mitochondrial genes *nad5*, *cob*, *coxI* and *coxIII* showed a significant percentage identity [Figure 4]. This correlated with the observed similarity seen in when only comparing the exon sequences for all the different *C. albifundus* isolates. The percentage overlapping sequence alignment for any two sequences showed similarity values greater than 97% [Figure 4]. The *C. albifundus* CMW 24865 and CMW 24860 isolates from Kenya and Tanzania respectively had 100% similarity for all the intronic regions as well as the homing endonuclease genes contained within. The nucleotide differences in terms of the number of alignment positions where each sequence differed was highest for *C. albifundus* CMW 4068 in comparison to the other African isolates (*C. albifundus* CMW 24685; CMW 24860 and CMW 13980 from Kenya, Tanzania and Zambia respectively) [Figure 4].

The greatest differences were observed in the intergenic regions across all *C. albifundus* isolates. These differences may be attributed to the bias created when using a predetermined reference sequence, namely *C. albifundus* CMW 17274, or that in actual fact each isolates mitochondrial genome is different. The *C. albifundus* CMW 4068 displayed the highest overall dissimilarity to all other isolates in the number of nucleotides differences [Figure 5]. It also had a significantly lower overall percentage sequence identity when compared to the other isolates from South Africa. The isolates from Kenya and Tanzania shared a hundred percentage identity and no differences in the number of nucleotides that deviated when aligned [Figure 5].

3.4. Discussion

The reference mitochondrial genome of *C. albifundus* CMW 17274 allowed for the mitochondrial genomes of five other isolates of *C. albifundus* (CMW 17620; CMW 4068; CMW 24685; CMW 24860 and CMW 13980) to be fully annotated and characterized. All the mitochondrial genomes were similar in terms of synteny, genome size, gene orientation, coding domain sequences as well as their intronic regions. The genomes differed in size by only ten bases in size when mapped to the reference *C. albifundus* CMW 17274 mitochondrial genome, with all isolates having a greater than 98% overall similarity, although those for the Kenya and Tanzanian isolates were identical. All the mitochondrial genome signatures in terms of particular gene successions (*nad2* and *nad3*; *nad1* and *nad4*; *atp8* and *atp6*) as described in literature were observed (Ambrosio *et al.*, 2013; Lang *et al.*, 1999; Naidoo *et al.*, 2013). The exceptionally high coverage obtained for each of the consensus sequence derived from the reference mapping gives us confidence in both the gene conservation and level of sequence synteny observed for all the *C. albifundus* isolates [Table 2]. The overall bias in AT composition and skewed GC content is in line with what has been observed as a common feature of many characterized mitochondrial genomes (Ambrosio *et al.*, 2013; Juhasz *et al.*, 2008; Kerscher *et al.*, 2001; Kouvelis *et al.*, 2004; Notsu *et al.*, 2002; Pantou *et al.*, 2006).

All the *C. albifundus* mitochondrial genomes had near identical tRNA clusters dispersed between the core mitochondrial genes, which are consistent with what has been seen before for the mitochondrial genomes of Ascomycota (Hopper & Phizicky, 2003; Lang *et al.*, 2012). The largest tRNA cluster containing 14 tRNA genes was observed in the sequence

downstream from the large subunit and consists of duplicate genes with the same sequence and anticodon for Serine and Tryptophan genes. The variation amongst the *C. albifundus* isolates was seen in the presence or absence of these two tRNA genes along with a reduction in the number of tRNA Methionine present [highlighted in Table 3]. This tRNA was over-represented in *C. albifundus* CMW 17274 only (seen a total of five times) [Table 3], whilst in the other isolates it was observed four times in the mitochondrial genome. Nevertheless, the coding bias prediction software revealed that all the amino acids anticodons were recognized for the core mitochondrial genes, specifically the exon regions present in all *C. albifundus* isolates, thus there is no need to import tRNAs from the nuclear genome for this species (Hopper & Phizicky, 2003; Kamenski *et al.*, 2007).

The tRNA genes have the tendency to be congregated near the protein coding gene regions as this has been shown to ensure their functionality and expression (Lang *et al.*, 2012). This is in conjunction with the presence of RNA polymerase III enzyme, imported from the nuclear genome, which ultimately transcribes the tRNA genes. Interestingly, the mitochondrial RNA polymerase II has been shown to be related to virus polymerases (Hirt *et al.*, 1999). This is the case of the first tRNA cluster, observed between the *cob* and *coxI* genes, in which the tRNAs coding for Threonine, Glutamine, Histidine and Methionine occur. In the second tRNA cluster comprising of tRNAs specific for Isoleucine, Serine, Asparagine and Proline is also observed to be between the coding regions of the *coxI* and *nad6* genes mitochondrial genes.

The highest number of intronic regions observed in the cytochrome oxidase genes (*coxI* and *coxIII*), was consistent with previously characterized mitochondrial genomes such as *C. fimbriata*, *C. albifundus*, *C. moniliformis* as well as *C. cocoafunesta* (Ambrosio *et al.*, 2013; Naidoo *et al.*, 2013) was prevalent in all the *C. albifundus* isolates annotated. This particular mitochondrial gene is known to be exceptionally large (up to 24kb in total mitochondrial gene size) (Ferandon *et al.*, 2010) and although all the *C. albifundus* isolates contained an approximately 19kb *coxI* gene, in which a high synteny in the intron type and position as well as the predicted homing endonuclease was observed for the *coxI* gene. There was no difference in the intron type as well as the homing endonuclease predicted across the *C. albifundus* isolates. The protein motifs found within the ORFs in these introns corresponded to either LAGLIDADG or GIY-YIG types. Only the *cob* gene presented an intronic ORF with no HEG prediction. This was consistent for all the *C. albifundus* isolates [Figure 2], and

although it is not uncommon for introns to contain ORFs with no HEG, this occurrence is rare.

The mitochondrial genomes of the South African isolates (*C. albifundus* CMW 17274; CMW 17620 and CMW 4068) showed the greatest similarity to each other. The other three isolates from elsewhere on the African continent (*C. albifundus* CMW 24865; CMW 24860 and CMW 13980) also displayed a higher sequence percentage identity. This apparent lack of variation between certain isolates of a species is consistent with a comparative study carried out on *Aspergillus* and *Penicillium* species, where they showed only six synonymous single nucleotide polymorphisms (SNPs) that occurred in the coding regions (Joardar *et al.*, 2012). There were no synonymous SNPs observed in the mitochondrial gene regions for any of the *C. albifundus* isolates. These SNP's will be further investigated in order to determine if they are in fact real and not artefacts of mis-assemblies.

The actual differences between the genomes in terms of their size can be attributed to deletions within the intragenic region between the *atp9* and *coxII* mitochondrial genes. This difference is localized to this region only for the *C. albifundus* CMW 4068 isolate from KwaZulu Natal, South Africa. The difference in the mitochondrial genome size for the isolates CMW 17260 and CMW 13980 from Limpopo and Zambia was recognized as transitional changes where purines were substituted with pyrimidines, thus converting adenine or guanines to cytosines or thymines. The lack of transversions suggests that the mitochondrial genomes of the different *C. albifundus* isolates preferred interchanges in the ring structure of the purines from a two-ring conformation to a single ring pyrimidine. We can thus speculate that the presence of point mutations is was higher in the *C. albifundus* CMW 17620 and CMW 13890 isolates respectively. Similar transitional changes have been observed in comparative studies done on yeasts and human mitochondrial DNA (Blanchard & Schmidt, 1996).

The comparative pairwise comparison of the nucleotide differences in terms of their alignment position for each isolate was higher when comparing the total intronic and exonic regions for a particular gene than just the exon regions. The intergenic comparison showed that the KwaZulu Natal isolate, *C. albifundus* CMW 4068 was the most dissimilar to both the South African and African isolates. The fact that both the Tanzanian and Kenyain isolates of *C. albifundus* share a hundred percent similarity is suggestive that these isolates share a common parent and are the result of a recent introduction of this pathogen into the region

(Roux *et al.*, 2005). The fact that the South African isolates were not identical correlates to previous studies which showed that the origin of *Ceratocystis* must lie in Southern Africa (Barnes *et al.*, 2005; Chen *et al.*, 2013; Roux *et al.*, 2004; Wingfield *et al.*, 2011).

This represents the first intraspecies study comparing different geographically distinct *C. albifundus* isolates within the order Microascales. We were able to fully annotate each isolates' mitochondrial genome and investigate their sequence similarity. From the exon alignment variation seen across the different isolates of *C. albifundus* the mitochondrial genes *atp8*, *atp9*, *nad5*, *coxIII* and *cob* were the most informative. Individual alignments for each mitochondrial gene region showed variation and are thus presented as potential genes to be further exploited in extensive phylogenetic analysis especially for intra and interspecies differentiation. This should allow further studies on the spread of the fungus particularly from native to plantation areas.

3.5. References:

Al-Reedy R. M., Malireddy R., Dillman C. B., and Kennell J. C. (2012). Comparative analysis of *Fusarium* mitochondrial genomes reveals a highly variable region that encodes an exceptionally large open reading frame. *Fungal Genetics and Biology* **49:** 2-14.

Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W., and Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25:** 3389-3402.

Ambrosio A. B., do Nascimento L. C., Oliveira B. V., Teixeria P. J., Tiburcio R. A., Thomazella D. P., Leme A. F., Carazolle M. F., Vidal R. O., Mieczkowski P., Meinhardt L. W., Pereira G. A., and Cabrera O. G. (2013). Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome. *BioMedical Central Genomics* **14:** 1-38.

Barnes I., Nakabonge G., Roux J., Wingfield B. D., and Wingfield M. J. (2005). Comparison of populations of the wilt pathogen *Ceratocystis albifundus* in South Africa and Uganda. *Plant Pathology* **54:** 189-195.

Barr C. M., Neiman M., and Taylor D. R. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168:** 39-50.

Blanchard J. L., and Schmidt G. W. (1996). Mitochondrial DNA migration events in yeast and humans: Integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Molecular Biology and Evolution* **13:** 537-548.

Burger G., Gray M. W., and Lang B. F. (2003). Mitochondrial genomes: anything goes. *Trends in Genetics* **19:** 709-716.

Chen S., van Wyk M., Roux J., Wingfield M. J., Xie Y., and Zhou X. (2013). Taxomony and pathogenicity of *Ceratocystis* species on *Eucalyptus* trees in South China, including *C. chinaeucensis* sp. nov. *Fungal Diversity* **58:** 267-279.

De Seynes J. (1886). Recherches Veg. Infer. III. *In* "Wismer, C.A (1961) Pineapple disease. Sugarcane disease of the world" (J. P. Martin, E. V. Abbett, and C. G. Huges, Eds.), Volume 1.

DeScenzo R. A., and Harrington T. C. (1994). Use of $(CAT)_5$ as a DNA fingerprinting probe for fungi *Phytopathology* **84:** 534-540.

Ferandon C., Moukha S., Callac P., Benedetto J.-P., Castroviejo M., and Barroso G. (2010). The *Agarius bisporus cox1* gene: The longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *Public Library of Science ONE* **5:** 1-8.

Fox T. D. (1987). Natural variation in the genetic code. *Annual Reviews of Genetics* **21:** 67-91.

Groter G. J. M. A. (1977). Index of plant pathogens and the diseases they cause in cultivated plants in South Africa. *In* "Science Bulletin No. 392", Pretoria, South Africa: Plant Protection Research Institution, Department of Agricultural Technical Services.

Grylls B. T., and Seifert K. A. (1993). A synoptic key to species of *Ophiostoma*, *Ceratocystis* and *Ceratocystiopsis*. In *Ceratocystis* and *Ophiostoma*: Taxonomy, Ecology and Pathogenicity. *(Wingfield, M.J., Seifert, K.A. and J.F. Webber, eds.) APS Press, St Paul, Minnesota***:** 261-268.

Halsted B. D., and Fairchild D. G. (1891). Sweet-potato black rot. *Journal of Mycology* **VII:** 1-11.

Heath R. N., van Wyk M., Wingfield M. J., and Roux J. (2009). Insect associations of *Ceratocystis albifundus* in South Africa and patterns of association in a native savannae ecosystem *Environmental Entomology* **38:** 356-364.

Hirt R. P., Logsdon J. M., Healy B., Dorey M. W., Doolittle W. F., and Embley T. M. (1999). Microsporidia are related to fungi: Evidence from the largest subunit of RAN polymerase II and other proteins. *Proceedings of the National Academy of Science* **96:** 580-585.

Hopper A. K., and Phizicky E. M. (2003). tRNA transfers to the limelight. *Genes and Development* **17:** 162-180.

Joardar V., Abrams N. F., Hosteler J., Paulstelis P. J., Pakala S., Pakala B. B., Zafar N., O.O A., Payne G., Andrianopoulous A., Denning D. W., and Nierman W. C. (2012). Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genomes size variability. *BioMedical Central Genomics* **13:** 1-12.

Juhasz A., Pfeiffer I., Keszthelyi A., Kucsera J., Vagvolgyi C., and Hamari Z. (2008). Comparative analysis of the complete mitochondrial genome of *Aspergillus niger* mtDNA type 1a and *Aspergillus tubingensis* mtDNA type 2b. *Federation of Microbiological Society, Microbiology Letters* **281:** 51-57.

Kamenski T., Kolensikova O., Jubenot V., Entelis N., Krasheninnikov I. A., Martin R. P., and Tarassov I. (2007). Evidence for an adaptation mechanism of mitochondrial translation via tRNA import from the cystol. *Molecular Cell* **26:** 625-637.

Kamgan N. G., Jacobs K., de Beer Z. W., Wingfield M. J., and Roux J. (2008). *Ceratocystis* and *Ophiiostoma* species including three new taxa, associated with wounds on native South African trees. *Fungal Diversity* **29:** 37-59.

Kamgan N. G., Wingfield M. J., Mohammed C., Carnegie A. J., Pegg G. S., and Roux J. (2012). *Ceratocystis* species, including two new species associated with nitidulid beetles, on eucalypts in Australia. *Antonie van Leeuwenhoek* **101:** 217-241.

Kerscher S., Durstewitz G., Casaregola S., Gaillardin C., and Brandt U. (2001). The complete mitochondrial genome of *Yarrowia lipolytica*. *Computational and Functional Genomics* **2:** 80-90.

Kouvelis V. N., Ghikas D. V., and Typas M. A. (2004). The analysis of the complete mitochondrial genome of *Lecanicillium muscarium* (synonym *Verticillium lecanni*) suggests a minimum common gene organization in mtDNAs of *Sordariomycetes*: phylogenetic implications. *Fungal Genetics and Biology* **41:** 930-940.

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

Lang B. F., Lavrov D., Beck N., and Steinberg S. V. (2012). Mitochondrial tRNA structure, identity, and evolution of the genetic code. *In* "Organelle Genetics" (C. E. Bullerwell, Ed.), pp. 431-474, Springer, Berlin, Heidelberg.

Lewton-Brain L. (1907). Rind disease of the sugar cane. *In* "Wismer, C.A (1961) Pineapple disease. Sugarcane disease of the world" (J. P. Martin, E. V. Abbett, and C. G. Huges, Eds.), Volume 1

Lowe T. M., and Eddy S. R. (1997). TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* **25:** 955-964.

Montoya M. M., and Wingfield M. J. (2006). A review of *Ceratocystis sensu stricto* with special reference to the species complexes *C. coerulescens* and *C. fimbriata*. *Revista Facultad Nacional de Agronomia, Medellin* **59:** 3045-3375.

Morris M. J., Wingfield M. J., and De Beer C. (1993). Gummosis and wilt of *Acacia mearnsii* in South Africa caused by *Ceratocystis fimbriata*. *Plant Pathology* **42:** 814-817.

Naidoo K., Steenkamp E. T., Wingfield M. J., and Wingfield B. D. (2013). Annotation and comparative analysis of three *Ceratocystis* mitochondrial genomes. *In* "Genetics", Ph.D Thesis: University of Pretoria, Pretoria, South Africa.

Notsu Y., Masood S., Nishikawa N., Kubo N., Akiduki G., Nakazono M., Hirai A., and Kadowaki K. (2002). The complete sequence of the rice (*Oryza sativa* L) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics* **268**.

Pantou M. P., Kouvelis V. N., and Typas M. A. (2006). The complete mitochondrial genome of the vascular wilt fungus *Verticillium dahlia*: a novel gene order for *Verticillium* and a diagnostic tool for species identification *Current Genetics* **50:** 125-136.

Pantou M. P., Kouvelis V. N., and Typas M. A. (2008). The complete mitochondrial genome of *Fusarium oxysporum*: Insights into fungal mitochondrial evolution. *Gene* **419:** 7-15.

Roux J., Dunlop R., and Wingfield M. J. (1999). Susceptibility of elite *Acacia mearnsii* families to Ceratocystis wilt in South Africa. *Journal of Forestry Research* **4:** 187-190.

Roux J., Harrington T. C., Steimel J., and Wingfield M. J. (2001). Genetic variation in the wattle wilt pathogen *Ceratocystis albifundus*. *Mycoscience* **42:** 327-332.

Roux J., Heath R. N., Labuschange L., Nkuenam G. K., and Wingfield M. J. (2007). Occurrence of the wattle wilt pathogen, *Ceratocystis albifundus* on native South African trees. *Forest Pathology* **37:** 292-302.

Roux J., Meke G., Kanyi B., Mwangi L., Mabaga A., Hunter G. C., Nakabonge G., Heath R. N., and Wingfield M. J. (2005). Diseases of plantation forestry trees species in Eastern and Southern Africa. *South African Journal of Science* **101:** 409-413.

Roux J., and Wingfield M. J. (1997). Survey and virulence of fungi occurring on diseased *Acacia mearnsii* in South Africa. *Forest Ecology and Management* **99:** 327-336.

Roux J., Wyk M. v., Hatting H., and Wingfield M. J. (2004). *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. *Plant Pathology* **53:** 414-421.

Schattner P., Brooks A. N., and Lowe T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* **33:** W686-W686.

Torriani S. F. F., Goodwin S. B., Kema G. H. J., Pangilinan J. L., and McDonald B. A. (2008). Intraspecific comparison and annotation of two complete mitochondrial genome sequences from the plant pathogenic fungus *Mycosphaerella graminicola*. *Fungal Genetics and Biology* **45:** 628-637.

van Wyk M., Al-Dawi A. O., Khan I. A., Deadman M. L., Al-Jahwari A., Wingfield B. D., Ploetz R., and Wingfield M. J. (2007). *Ceratocystis manginecans* sp. nov., causal agent of a destructive mango wilt disease in Oman and Pakistan. *Fungal Diversity* **27**.

van Wyk M., Wingfield B. D., Marin M., and Wingfield M. J. (2010). New *Ceratocystis* species infecting coffee, cacao, citrus and native trees in Colombia *Fungal Diversity* **40:** 103-117.

Wingfield B. D., van Wyk M., Roos H., and Wingfield M. J. (2011). *Ceratocystis*: Emerging evidence for discrete generic boundaries. In: Ophiostomatiod fungi: expanding frontiers. *(Seifert, K.A and M.J Wingfield, eds.) CBS Biodiversity Series 12* **:** 57-64.

Wingfield M. J., De Beer C., Visser C., and Wingfield B. D. (1996). A new *Ceratocystis* species defined using morphological and ribosomal DNA sequence comparisons. *Systematics in Applied Microbiology* **19:** 191-202.

Wingfield M. J., Seifert K. A., and Webber J. A. (1993). *Ceratocystis* and *Ophiostoma*: Taxonomy, ecology and pathogenicity *American Phytopathological Society, St Paul, Minnesota*.

Witthuhn R. C., Wingfield B. D., Wingfield M. J., and Harrington T. C. (1999). PCR based identification and phylogeny of species of *Ceratocystis sensu stricto*. *Mycological Research* **103:** 743-749.

Woo P. C., Zhen H., Cai J. J., Yu J., Lau S. K., Wang J., Teng J. L., Wong S. S., Tse R. H., Chen R., Yang H., Lui B., and Yuen K. L. (2003). The mitochondrial genome of the thermal dimorphic fungus *Penicillium marneffei* is more closely related to those of molds than yeast. *Federation of Microbiological Society, Microbiology Letters* **555:** 469-477.

Zivanovic Y., Wincker P., Vacherie B., Bolotin-Fukuhara M., and Fukuhara H. (2005). Complete nucleotide sequence of the mitochondrial DNA from *Kluyveromyces latis*. Federation of Microbiological Society, Yeast Research **5:** 315-322.

Table 1. Geographic origin and hosts of the *Ceratocystis albifundus* isolates used in this study.

***Ceratocystis albifundus* isolate**	**Location isolated from**	**Host**
CMW 17274[a]	Leeuwfontein Nature Reserve, Gauteng, South Africa	*Faurea saligna*
CMW 4068[a]	Bloemdal Farm, Pietermaritzburg, KwaZulu Natal, South Africa	*Acacia mearnsii*
CMW 17620[a]	Kruger National Park, Limpopo, South Africa	*Terminalia serecia*
CMW 24685[b]	Kenya	*Acacia mearnsii*
CMW 13980[b]	Kaloko Trust, Zambia	*Parinari curatifolia*
CMW 24860[b]	Tanganika Wattle Company, Njombe, Tanzania	*Acacia mearnsii*

Footnote:

[a] Reference: (Morris *et al.*, 1993; Roux *et al.*, 1999; Roux & Wingfield, 1997)

[b] Reference: (Roux *et al.*, 2005)

Table 2. Sequence data, mitochondrial genome size and coverage as well as the percentage AT content for the six different *Ceratocystis albifundus* isolates used in this study.

Mitochondrial Assembly character	***Ceratocystis albifundus* CMW 17274 (Gauteng)**	***Ceratocystis albifundus* CMW 4068 (KwaZulu Natal)**	***Ceratocystis albifundus* CMW 17620 (Limpopo)**	***Ceratocystis albifundus* CMW 24685 (Kenya)**	***Ceratocystis albifundus* CMW 24860 (Tanzania)**	***Ceratocystis albifundus* CMW 13980 (Zambia)**
Total DNA sequence data	13.98 GB	6.81GB	7.54GB	7.54GB	8.07GB	7.32GB
mt DNA (bp) genome	126 447 bp	126 437 bp	126 353 bp	126 447 bp	126 447 bp	126 353 bp
mt DNA coverage[a]	2573x[b]	117 941x[c]	130 271x[c]	135 171x[c]	129 452x[c]	117 287x[c]
Single reads[d]	279 896	966 578	640 299	780 183	855 181	1249 997
Paired reads[d]	28 6142	1293 890	703 614	1152 7708	1613 710	3434 898
AT content	72.9%	72.9%	72.9%	72.9%	72.9%	72.9%

Footnote:

[a]Mitochondrial genome coverage based on mapping against single and paired Illumina reads for each *Ceratocystis albifundus* isolate.

[b]Mitochondrial genome coverage based on mapping against *Ceratocystis fimbriata* CMW 15049 reference mitochondrial sequence.

[c]Mitochondrial genome coverage based on mapping against *Ceratocystis albifundus* CMW 17274 reference mitochondrial sequence.

[d]Single and paired reads which generated the consensus mitochondrial genome sequence for each of the *C. albifundus* isolates respectively.

Table 3. tRNA codon usage for the different amino acids predicted for all six *Ceratocystis* species

tRNA amino acid	*Ceratocystis albifundus* CMW 17274	*Ceratocystis albifundus* CMW 17620	*Ceratocystis albifundus* CMW 4068	*Ceratocystis albifundus* CMW 24685	*Ceratocystis albifundus* CMW 24860	*Ceratocystis albifundus* CMW 13980
Alanine (Ala)	GCA	GCA	GCA	GCA	GCA	GCA
Arginine (Arg)	CGT	CGT	CGT	CGT	CGT	CGT
Asparagine (Asn)	AAC	AAC	AAC	AAC	AAC	AAC
Aspartic acid (Asp)	GAC	GAC	GAC	GAC	GAC	GAC
Cysteine (Cys)	TGC	TGC	TGC	TGC	TGC	TGC
Glutamic Acid (Glu)	GAA	GAA	GAA	GAA	GAA	GAA
Glutamine (Gln)	CAA	CAA	CAA	CAA	CAA	CAA
Glycine (Gly)	GGA	GGA	GGA	GGA	GGA	GGA
Histidine (His)	CAC	CAC	CAC	CAC	CAC	CAC
Isoleucine **(Lle)**	ATC (x2)	ATC (x2)	ATC (x2)	ATC (x2)	ATC (x2)	ATC (x2)
Leucine (Leu)	TTA	TTA	TTA	TTA	TTA	TTA
Lysine (Lys)	AAA	AAA (x2)	AAA (x2)	AAA (x2)	AAA (x2)	AAA (x2)
Methionine (Met)	ATG (x4); ATA	ATG (x4); ATA	ATG (x4); ATA	ATG (x4)	ATG (x4)	ATG (x4)
Phenylalanine (Phe)	TTC	TTC	TTC	TTC	TTC	TTC
Proline (Pro)	CCA	CCA	CCA	CCA	CCA	CCA
Serine (Ser)	TCA (x2); AGC	TCA (x2); AGC	TCA (x2); AGC	TCA; AGC	TCA; AGC	TCA (x2); AGC
Threonine (Thr)	CTA (x2); ACT; CTT	CTA; ACA; CTT	CTA; ACA; CTT	CTA (x2); ACA	CTA (x2); ACA	CTA (x2); ACA
Tryptophan (Trp)	TGA (x2)	TGA (x2)	TGA (x2)	TGA (x2)	TGA (x2)	TGA (x2)
Tyrosine (Tyr)	TAC	TAC	TAC	TAC	TAC	TAC
Valine (Val)	GTA	GTA	GTA	GTA	GTA	GTA
Total tRNAs present	33	33	33	31	31	31

Footnote: The tRNAs highlighted in yellow represent the differences seen in the presence or absence of particular tRNA gene codons for Methionine, Serine and Threonine for the different *C. albifundus* isolates as predicted $_{T}RNA_{SCAN}$-SE version 1.21 software (Lowe & Eddy, 1997).

Figure 1. Schematic representation of the mitochondrial genome for all *Ceratocystis albifundus* isolates.

The same gene organization of the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* and *rns*) are shown as a circular fragment for all *Ceratocystis albifundus* -CMW 17274; -CMW 4068; -CMW 17620; -CMW 24685; -CMW 24860 and –CMW 13980 isolates. All the mitochondrial genes are apparently transcribed from the positive strand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; the *cob* gene is yellow; the *cox* genes and *rpS3* gene are red; the intergenic LAGLIDADG homing endonuclease (HE) gene is bright blue; the *rns* is brown and the *rnnl* is orange.

Figure 1.

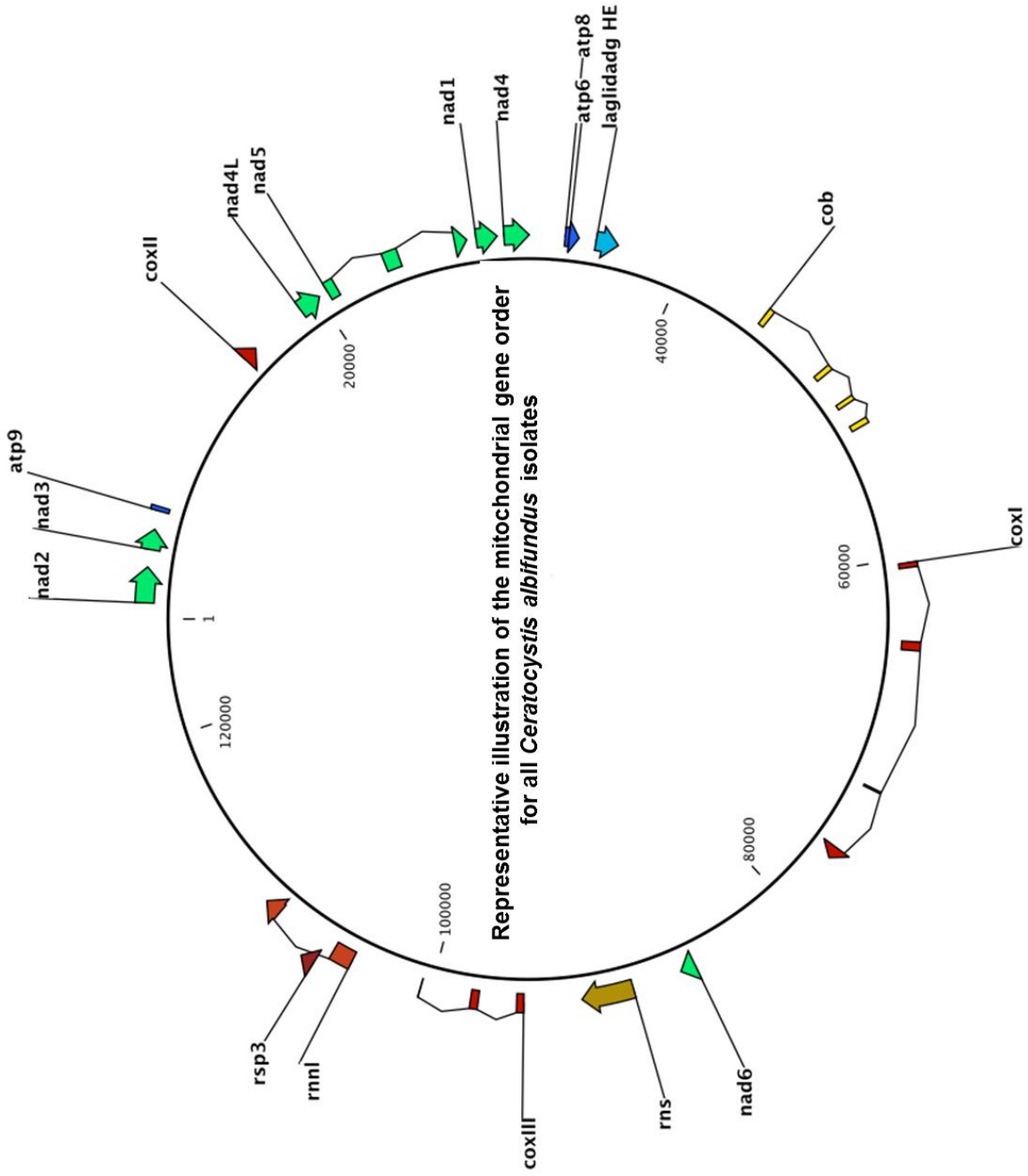


Figure 2. Schematic representative of the mitochondrial genes, introns and homing endonuclease genes found across all *Ceratocystis albifundus* isolates.

The gene arrangement and intron presence within the 15 core mitochondrial genes as well as the large and small ribosomal subunits (*rnnl* & *rns*) are represented as a circular fragment all *Ceratocystis albifundus* -CMW 17274; -CMW 4068; -CMW 17620; -CMW 24685; -CMW 24860 and –CMW 13980 isolates. All genes were transcribed on the positive stand.

The genes are coded for as follows: *nad* genes are green; *atp* genes are dark blue; *cob* gene is yellow; *cox* and *rps3* genes are red; intergenic LAGLIDADG homing endonuclease is bright blue; the *rns* is brown and the *rnnl* is orange. Introns were predicted as Group I types where Group IA is bright green; Group I (derived A) is bright blue; Group IB is navy and Group ID is aqua blue. Homing endonuclease genes (HEGs) were predicted for all open reading frames found within the intron. The LAGLIDADG HEGs are shown as dark purple whilst the GIY-YIG HEGs are dark brown.

Figure 2.

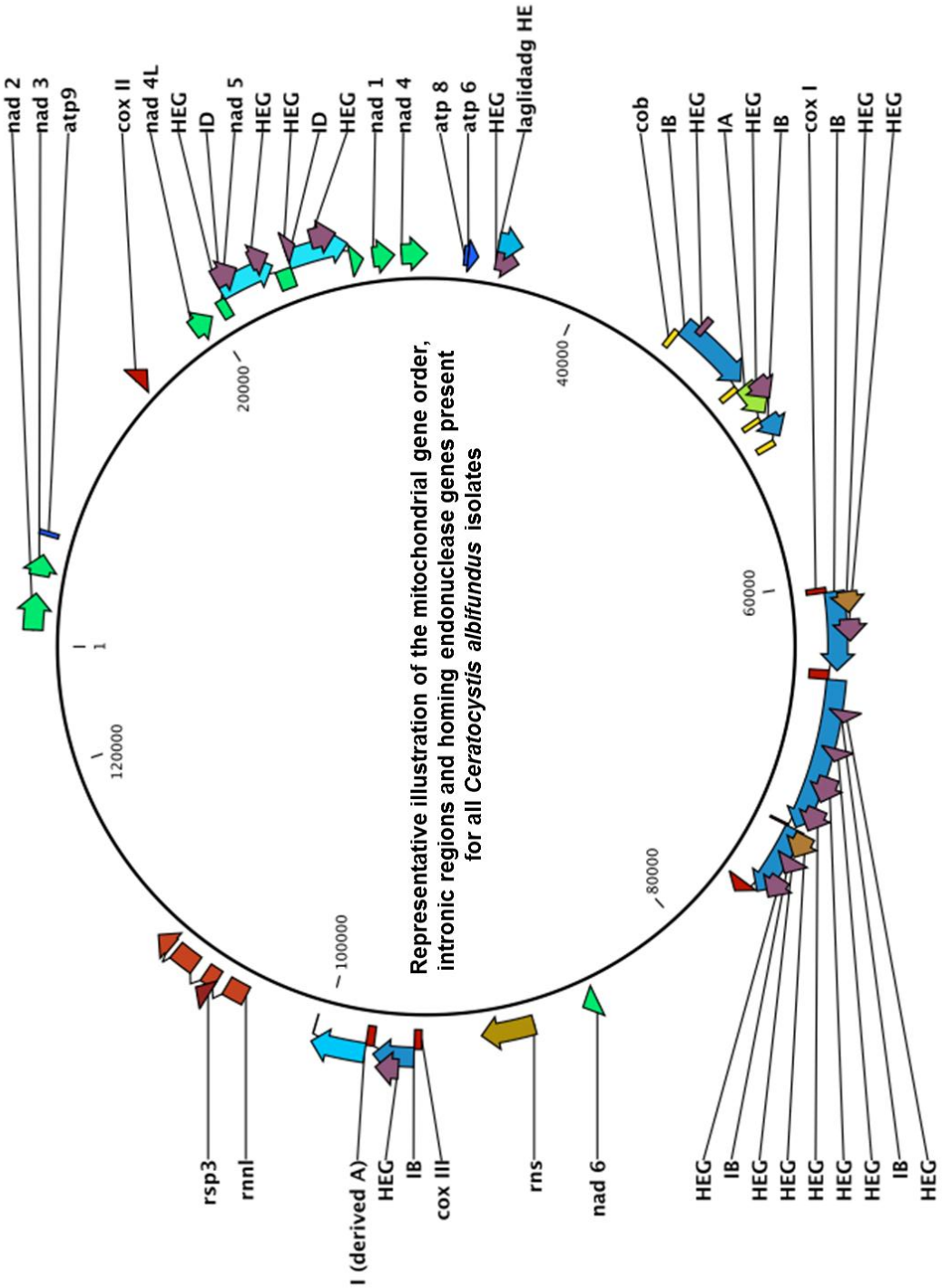


Figure 3. Diagrammatic illustration tabling the percentage sequence identity and the differences in the nucleotide sequence alignment position for the coding sequence domains [exons] for all 14 mitochondrial genes present across the different *Ceratocystis albifundus* isolates.

The different *C. albifundus* isolates are represented as follows:

1. *C. albifundus* CMW 17274 [Gauteng, South Africa]
2. *C. albifundus* CMW 17620 [Limpopo, South Africa]
3. *C. albifundus* CMW 4068 [KwaZulu Natal, South Africa]
4. *C. albifundus* CMW 24685 [Kenya, Africa]
5. *C. albifundus* CMW 24860 [Tanzania, Africa]
6. *C. albifundus* CMW 13980 [Zambia, Africa]

Figure 3.

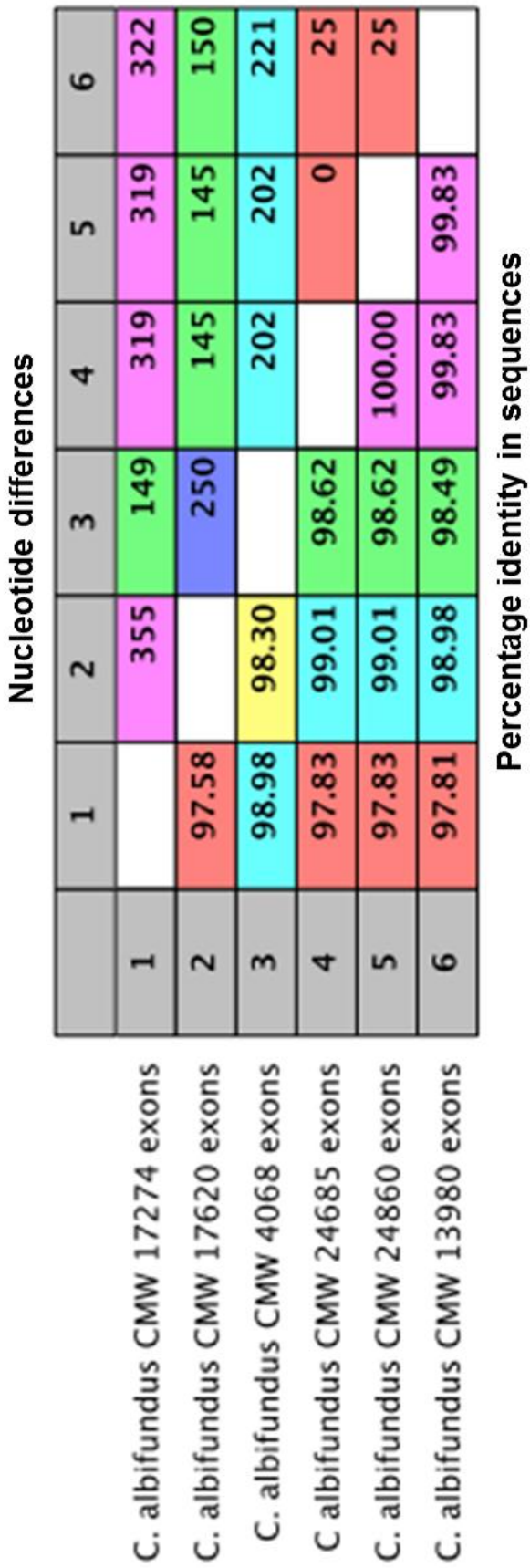


Figure 4. Diagrammatic illustration tabling the percentage sequence identity and the differences in the nucleotide sequence alignment position for the coding sequence domains [exons] as well as the introns for the *nad5*, *cob*, *coxI* and *coxIII* mitochondrial genes present across the different *Ceratocystis albifundus* isolates.

The different *C. albifundus* isolates are represented as follows:

1. *C. albifundus* CMW 17274 [Gauteng, South Africa]
2. *C. albifundus* CMW 17620 [Limpopo, South Africa]
3. *C. albifundus* CMW 4068 [KwaZulu Natal, South Africa]
4. *C. albifundus* CMW 24685 [Kenya, Africa]
5. *C. albifundus* CMW 24860 [Tanzania, Africa]
6. *C. albifundus* CMW 13980 [Zambia, Africa]

Figure 4.

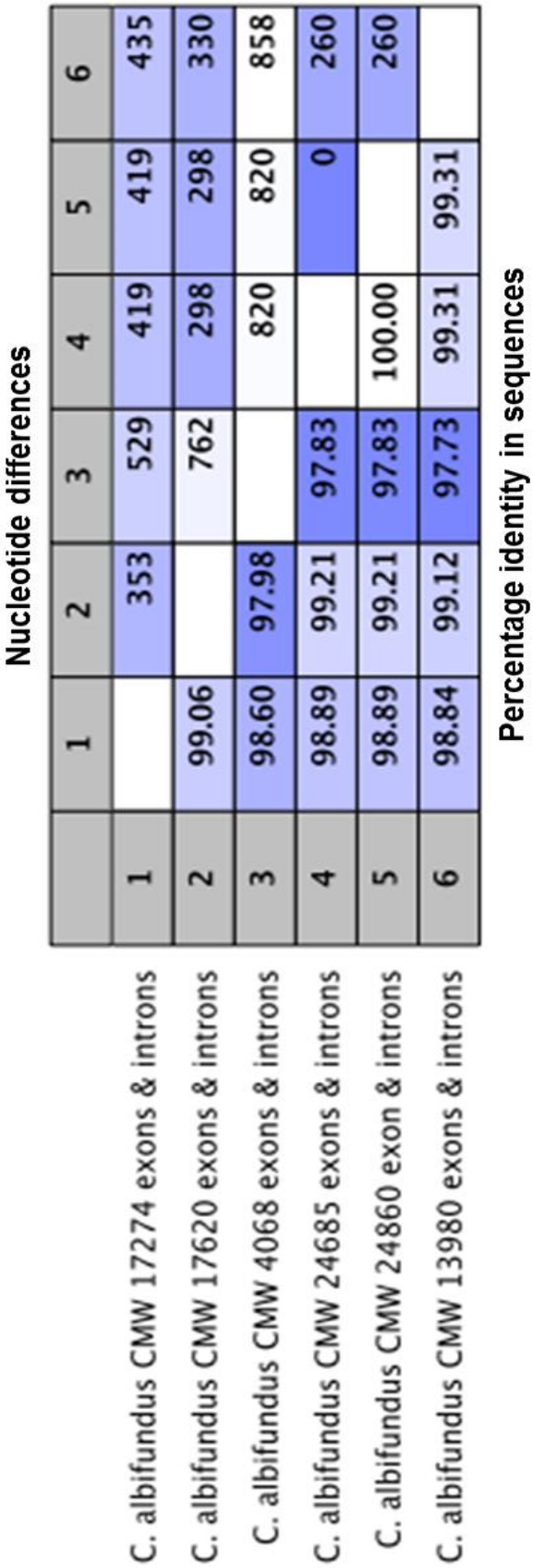
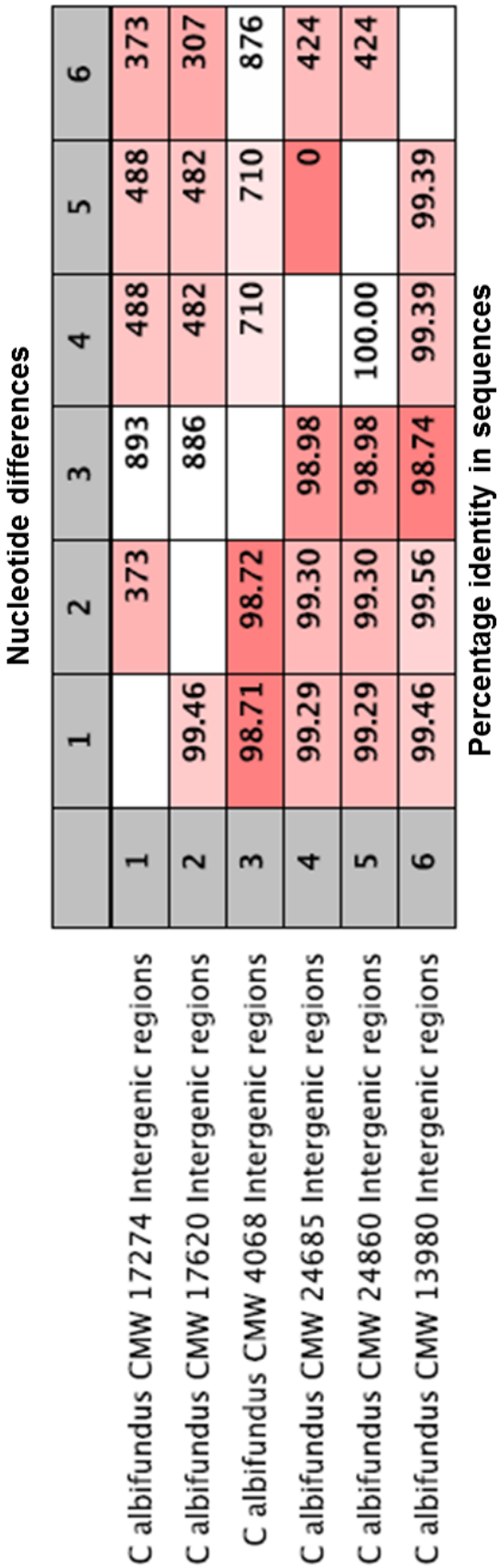


Figure 5. Diagrammatic illustration tabling the percentage sequence identity and the differences in the nucleotide sequence alignment position for the intergenic sequence regions in the mitochondrial genomes present across the different *Ceratocystis albifundus* isolates.

The different *C. albifundus* isolates are represented as follows:

1. *C. albifundus* CMW 17274 [Gauteng, South Africa]
2. *C. albifundus* CMW 17620 [Limpopo, South Africa]
3. *C. albifundus* CMW 4068 [KwaZulu Natal, South Africa]
4. *C. albifundus* CMW 24685 [Kenya, Africa]
5. *C. albifundus* CMW 24860 [Tanzania, Africa]
6. *C. albifundus* CMW 13980 [Zambia, Africa]

Figure 5.



Chapter Four

Concerted evolution in the ribosomal RNA cistron

This chapter of the thesis has been published as:

Naidoo K, Steenkamp ET, Coetzee MPA, Wingfield MJ, Wingfield BD (2013) Concerted Evolution in the Ribosomal RNA Cistron. **PLoS ONE 8(3): e59355**.

doi:10.1371/journal.pone.0059355

Abstract

Gene conversion is the mechanism proposed to be responsible for the homogenization of multigene families such as the nuclear ribosomal gene clusters. This concerted evolutionary process prevents individual genes in gene clusters from accumulating mutations. The mechanism responsible for concerted evolution is not well understood but recombination during meiosis has been hypothesized to play a significant role in this homogenization. In this study we tested the hypothesis of unequal crossing over playing a significant role in gene conversion events within the ribosomal RNA cistron during meiosis, mitosis or both life stages in the fungal tree pathogen *Ceratocystis manginecans*. *Ceratocystis manginecans*, a haploid ascomycete, reproduces homothallically and was found to have two distinct sequences within the internally transcribed spacer (ITS) region of the ribosomal RNA cistron. The different ITS types were scored using PCR-RFLP assays and chi-square analyses to determine the level of significance of the changes in the ratios of the ITS types. The relative ratios of the two ITS sequence types changed when the fungal isolates were cultured vegetatively or allowed to produced sexual structures and spores. These active changes were shown to occur more frequently during meiosis than mitosis. The evidence presented provides concrete support for homogenization in the rRNA gene clusters found in this fungus and that the most reasonable explanation for this process is unequal crossing over.

4.1. Introduction

Current understanding of the structure and evolutionary history of genes encoding the RNA subunits comprising of ribosomes is based on research conducted over more than 40 years (Eickbush & Eickbush, 2007; Liao, 2000; Liao, 2003). The small ribosomal subunit (18S) together with the large subunit RNAs (5.8S and 28S) are all processed from a single precursor RNA after its transcription from the rRNA cistron. The ribosomal 5S gene is transcribed independently of the rRNA cistron and in many cases, is also located in the intergenic region of the rRNA cistron (Rooney & Ward, 2005). Within the cistron, the genes encoding the respective RNA subunits are separated by the internal transcribed spacer regions (ITS1 and ITS2), while the entire region is flanked by the 3- and 5-prime intergenic spacer regions (White *et al.*, 1990). Due to the high level of cellular demand for ribosomes, the rRNA cistrons occur in large head-to-tail tandem arrays at one or a few chromosomal loci (Nei & Rooney, 2005). However, through evolutionary time, the same DNA sequence is maintained in each cistron of these arrays, even though differences between the cistrons of different species are allowed to accumulate (Long & Dawid, 1980). This is not consistent with classical evolutionary expectation, where all the members of a gene family would evolve independently (Nei & Rooney, 2005). Consequently, the term "concerted evolution" has been introduced to describe the form of evolution in which DNA or gene repeats evolve as single units in concert (Eickbush & Eickbush, 2007; Liao, 2003; Nei *et al.*, 1997; Nei *et al.*, 2000; Nei & Rooney, 2005).

In their model of concerted evolution, Brown and colleagues (Brown *et al.*, 1972) proposed that the homogenization process in DNA or gene repeats requires the individual repeats to evolve in a manner which is dependent on one another (Nei & Rooney, 2005). In this way the transmission and accumulation of mutations occurring in the repeat region becomes homogenized (i.e., mutations spread throughout the rRNA repeat array to all the member cistrons). Although the exact mechanisms determining concerted evolution remains unclear (Liao, 2003), two fundamental processes are thought to drive the homogenization process; unequal crossing over and gene conversion (Holliday, 1964; Lindegren, 1953). The essential difference between these two evolutionary forces is that gene conversion maintains the copy number of a gene at a constant size, while unequal crossing over has the potential to cause fluctuations in the gene copy number from one generation to the next (Dover, 1982; Pinhal *et al.*, 2011). Unequal crossing over represents a form of homologous recombination between repeats or cistrons, located at dissimilar positions within a locus on two chromosomes or

between different cistrons on the same chromosome (Eickbush & Eickbush, 2007; Ganley & Kobayashi, 2011). Gene conversion results in unidirectional or non-reciprocal DNA transfer between DNA duplexes (Liao, 2000) (at the same or different loci) due to homologous recombination that was initiated by DNA double stranded breaks (Chen *et al.*, 2007; Lange *et al.*, 2011). The process of unequal crossing over can arise during mitosis within or between sister chromatids or during meiosis within or between homologous or non-homologous chromosome pairs (Eickbush & Eickbush, 2007). The effects of gene conversion have also been observed during both meiosis and mitosis (Chen *et al.*, 2007; Eickbush & Eickbush, 2007).

Most evidence for concerted evolution of the rRNA cistron is derived from research using metazoan model organisms (e.g., *Drosophila* and *Xenopus*) and microorganisms such as *Saccharomyces cerevisiae* (Brown *et al.*, 1972; Eickbush & Eickbush, 2007; Nei & Rooney, 2005). Evidence for this process can also be inferred from DNA sequence data available from a variety of plants and animals (Innan, 2011; Lichten, 2001). The growing number of genomes accessible for study has provided additional support for the existence of concerted evolution (Liao, 2000). Empirical evidence for the involvement of unequal crossing over in the homogenization of rRNA cistrons is also available from a range of eukaryotes, particularly *S. cerevisiae* (Eickbush & Eickbush, 2007). However, experimental evidence for the contribution of gene conversion in the concerted evolution of DNA or gene repeats is mostly restricted to the homogenization of protein-encoding gene families (Chen *et al.*, 2007; Liao, 1999).

Although various studies have considered concerted evolution in non-model plants (Cronn *et al.*, 1996) and animals (Elder & Turner, 1995), few studies have focused on the homogenization of multicopy genes and rRNA cistrons in non-model fungi (Elder & Turner, 1995; Lumbsch & Leavitt, 2011). For fungi, the rRNA cistron is of primary importance because various regions of the rRNA cistron are frequently targeted for DNA-based identification (Liao, 2000). For example, the ITS region is utilized as the standard DNA barcoding region for fungal identification (Schoch *et al.*, 2012). The potentially far-reaching consequences of concerted evolution on fungal taxonomy and diagnostics (i.e., by influencing the sequence and evolutionary trajectories as well as the resulting phylogenies of the rRNA cistron), requires a more detailed understanding of this process in fungi.

Ceratocystis manginecans is a homothallic ascomycete and an important plant pathogen (van Wyk *et al.*, 2007). In the laboratory environment, this fungus is maintained in culture by repeatedly transferring mycelial strands from agar cultures. Its homothallic nature results in the production of sexual structures without the requirement of outcrossing (van Wyk *et al.*, 2007) [Figure 1]. Therefore, in a single isolate, both sexual and asexual reproduction occurs, without a change to the genetic structure of the organism during either mitosis or meiosis.

In recent studies, some isolates of *C. manginecans* were identified that contain two distinct rRNA ITS variants, a phenomenon that has also been observed in other fungi (Aanen *et al.*, 2001; O'Donnell & Cigelnik, 1997) as well as some higher eukaryotes (Liao, 2000). In addition, we have seen some variation in the ratios of these ITS variants while maintaining these fungi in culture. These observations and the reproductive strategy of *C. manginecans* in the laboratory provided an opportunity to investigate concerted evolutionary processes in this fungus. Our specific aims were to determine whether homogenization or a drift-like variation of the rRNA cistron occurred during meiosis, mitosis or both life stages and whether the effects of this process could be seen as fluctuations of the different ITS types. We were able to provide statistical evidence of significant changes occurring within the rRNA cistron during both meiosis as well as mitosis. These changes could be linked to concerted evolution and ultimately to the occurrence of unequal recombinational crossing and potentially gene conversion of rRNA cistrons.

4.2. Materials and Methods

4.2.1 Fungal Isolates

Two sets of four *C. manginecans* isolates were used in this study. The first set of isolates (CMW 13581, 13584, 23641 and 23643) apparently harbour a single ITS type [Figure S1], while those in the second set harbour two types of ITS (CMW 13852, CMW 17568, CMW 17570 and CMW 23635) (van Wyk *et al.*, 2007). All isolates were maintained on 2% malt extract agar (MEA: 20% w/v; Biolab, Midrand, South Africa) at 25°C.

The second isolate set was used to test whether homogenization of rRNA cistrons occurred during mitosis. For this purpose five sequential sub-cultures for each of the four isolates were prepared, which was done by transferring a single hyphal tip of each isolate to fresh MEA medium and allowing it to grow for approximately two weeks. The entire process was

repeated an additional four times, each time using hyphal tips from the new sub-culture [Figure 2]. All of these isolate sets, together with the original isolates, have been deposited in and can be obtained from the culture collection (CMW) of the Tree Protection Cooperative Programme, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Pretoria, South Africa.

The sexual spores produced by each of the four isolates in the set harbouring two ITS types were used to test whether homogenization of the rRNA cistrons occurred during meiosis. This was accomplished by allowing the four isolates to produce sexual structures (perithecia) by incubating them on MEA at 25°C for one week. From the mature sexual structures (perithecia) that developed, a single ascospore (meiospore) mass was collected for each isolate after which individual ascospores were used to provide the DNA template for the amplification of the ITS region [Figure 2].

4.2.2 ITS PCR and sequencing

DNA was isolated from all cultures and subcultures and used as templates in PCRs with primers ITS1 and ITS4 (White *et al.*, 1990) using protocols described previously (Sokal & Rohlf, 1995; van Wyk *et al.*, 2006). DNA was not isolated from ascospores but the ITS region was amplified directly from single ascospores using PCR. To control for the possible impact of PCR artifacts, a number of thermostable DNA polymerases were tested. These included Expand Taq [Roche Diagnostics, Mannheim, Germany], FastStart Taq [Roche Diagnostics, Mannheim, Germany], and SuperTherm Taq [Fermentas, Inqaba Biotechnical Industries (PTY) LTD., South Africa].

All amplicons were cloned using the pGEM-T Vector system II [Promega, Anatech Instruments, South Africa]. In each case, cloned inserts were amplified directly from 20 randomly selected recombinants using vector-specific primers and PCR as described previously (Witthuhn *et al.*, 2000). After purification with Sephadex® G-50 columns (SIGMA), the products were sequenced in both directions using the original vector-specific primers, the Big Dye terminator version 3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) and an ABI PRISM ™ 3100 Autosequencer (Applied BioSystems, Foster City, California, USA). All sequences were then manually aligned using BioEdit (Hall, 1999) to visualize the two ITS types or variants [referred to as Y and Z] [Figure 2].

4.2.3 ITS PCR-RFLPs

In order to analyse the large number of cloned inserts that were generated, a PCR-RFLP (Restriction Fragment Length Polymorphism) technique was developed to differentiate between the two ITS sequence types. The ITS region of the one sequence type [type Y, GenBank accession number KC261853], contained the restriction site recognised by the enzyme *TscAI* [Fermentas, Inqaba Biotechnical Industries (PTY) LTD., South Africa] in two positions. This site is present only once in the other ITS sequence type [type Z, GenBank accession number KC261852]. The restriction enzyme, TscAI was thus used to differentiate between the two ITS sequence types. Briefly, the PCR-RFLP technique employed involved direct amplification of 40 randomly selected recombinants using vector-specific primers and PCR as described above. The PCR products were then digested with TscAI according to the supplier's specifications. The RFLP fragments were stained with Gel Red [Anatech Instruments, South Africa], separated using agarose (3%, w/v, Supplier) gel electrophoresis and visualized under UV trans-illumination.

2.4 Statistical Analysis

Chi-square analyses (Sokal & Rohlf, 1995) were conducted to establish whether the ratios of the different ITS sequence types were statistically different in the sub-cultures that had been produced, as well as in the single ascospores that were subjected to PCR. A 99.999% level of confidence was applied. The number of degrees of freedom (df=4) was established based on the number of transfers made (i.e., n=5, thus n-1=4). The null hypothesis was that no significant change would be observed between the ratios after meiosis or mitosis.

The combined chi-square statistic tested the same null hypothesis under the conditions of mitosis and meiosis. Each sub-cultured isolate thus served as a replica test allowing for all the individual chi-square analyses to be combined. The probability values for the combined statistics were calculated using the initial chi-square value gained from the original replica tests for the respective isolate and then correlating this to the associated lowest possible probability in the R-statistical programming software (R_Development_Core_Team, 2007). This software allows for the lowest mathematical probability to be determined as an inverse of the chi-square derivative. The logarithm of this probability is then cumulatively added for all the replicas to generate a total probability value, which can then be checked against the

number of degrees of freedom. In this case, the number of degrees of freedom is established by taking twice the number of separate tests and probabilities carried out initially (thus 2k = 2(4) = 8 degrees of freedom). A 99.999% level of confidence was also applied.

4.3. Results

4.3.1 Scoring of the ITS types

Use of the universal ITS primers (White *et al.*, 1990) yielded PCR products of 646 base pairs (bp) in length. Sequence analysis of *C. manginecans* isolates CMW 13581, 13584, 23641 and 23643 confirmed that they all harboured a single ITS type [Figure S1], while isolates CMW 13582, 17568, 17570 and 23635 contained a combination of both ITS types. After cloning the latter ITS fragments into the pGEM®-T Easy vector, and sequencing the insert, no differences could be ascribed to the polymerase used for amplification. In all the various sequence sets (i.e., sequences generated for the 20 cloned inserts of the five vegetative culture sets for each of the four original *C. manginecans* isolates, as well as the five single-ascospores from the four isolates) examined, we observed two different ITS types (Y and Z) [Figure S2].

PCR-RFLP analysis of the 40 randomly selected *Escherichia coli* clones generated from a particular isolate and its sub-cultured replicates or single ascospores showed the expected RFLP profiles. The size of the undigested PCR fragment from the *E. coli* clone was 920bp. Digestion of the PCR clone products of ITS sequence type Z yielded a doublet of 460bp. The ITS sequence type Y produced three fragments [525bp, 335bp and 60pb], although the smallest fragment was too small for regular detection using agarose gel electrophoresis.

3.2 Statistical analysis of the ITS types

For each of the isolates, CMW 13582, 17570, 17568 and 23635, a set of 600 cloned ITS fragments were analysed (i.e., 60 for each of five vegetative cultures and 60 representing the amplicons generated from five single ascospores). The observed Z:Y ratios for the ITS types in each set of 60 cloned amplicons after meiosis and mitosis are summarized in supplementary information [Table S1]. To determine the chi-square statistic, the ITS type frequency for each original isolate (i.e., the expected frequency) was compared with the

observed frequency after meiosis and mitosis [Table S1]. While it is possible that there could be more than two ITS types, these would need to exist at a frequency of less than one percent, not to have been observed in our analyses. It was anticipated that the copy number of the rRNA cistron is around 120 to 130 per cell (Ide *et al.*, 2010), which makes the possibility of missing additional variants negligible.

In each isolate, there was a significant oscillation in the relative frequencies of the different ITS variants [Table S1]. For example, the ascospore-derived data for CMW 17570.1 showed 36 ITS Z types and 24 ITS Y types, while those for CMW 17570.2 had 7 ITS Z and 53 ITS Y. These dramatic changes were observed less often in the mitotically-derived data. For example, in isolates CMW 17570.a and CMW 23635.a, the proportion of the ITS type Z variant remained lower than the ITS type Y variant [Table S1]. At a 99.999% level of confidence, the chi-square analysis showed that the frequency of the two ITS types changed significantly during both meiosis and mitosis. These analyses also showed that the ITS type ratios changed more often as a consequence of meiosis [Table 1].

Contrast to what was observed for isolates CMW 17570, 17568 and 23635, the results for isolate CMW 13582 showed a significant change in frequency of the ITS types in both the meiotic and mitotic sets of observations [Table 1]. For each of the meiotic and mitotic sets of observations, a combined chi-square value was also obtained (Sokal & Rohlf, 1995) [Table 2, Table 3]. This value considered each isolate as a separate replicate of the experiment and probability values were obtained using R programming software (R_Development_Core_Team, 2007). This approach employed the prediction of probability statistics by making use of the initial chi-square values. Overall, the values obtained for the meiotic observations for the different isolates produced a highly significant deviation at a 99.999% level of confidence, indicating that these ITS type frequencies were not due to a chance sampling event.

4.4. Discussion

Concerted evolution has been hypothesized to be the result of gene conversion or unequal cross over during recombination. In this study, we have provided direct evidence of unequal recombination and the potential of gene conversion occurring within the rRNA cistron of the fungal pathogen *C. manginecans*. The observed changeability in the ratios of the Z:Y ITS types is a direct consequence of unequal cross over during recombination, while the lack of a

second ITS type likely emerged because a non-reciprocal recombination between cistrons (Eickbush & Eickbush, 2007). Our results showed that the processes of concerted evolution are highly dynamic as reflected by the dramatic changes in the ratios of the two ITS types analysed. This is similar to what was observed by Ganley and Kobayashi (Ganley & Kobayashi, 2011) who used a tagged *S. cerevisiae* rDNA unit to monitor the frequency of duplication and deletion events. In *C. manginecans*, however, we were able to show the dynamic flux of different ITS types in both mitosis and meiosis against an identical genetic background without the need to tag the ITS region. Concerted evolutionary processes could, therefore, be directly linked to unequal crossing over events occurring during both the meiotic and mitotic lifecycles within the rRNA cistron of *C. manginecans*.

The fluctuations observed in the frequencies at which both the ITS Z and Y variants occurred across the four *C. manginecans* isolates examined, suggest an associated fluctuation in the overall copy number of the rRNA cistron in these isolates. Similar evidence of fluctuations in the rRNA copy number has been recorded in other eukaryotes such as *Drosophila melanogaster* (Arnheim *et al.*, 1980; Dover, 1982; Ritossa, 1968; Rooney & Ward, 2005). However, the need for large amounts of rRNA to be produced in a cell probably precludes the copy numbers of these genes ever becoming inordinately low (Pinhal *et al.*, 2011). Ide et al. (Ide *et al.*, 2010) had demonstrated in their study of *S. cerevisiae* that an increase or decrease in the copy number of a highly transcribed gene led to increased sensitivity to DNA damage and increased cell toxicity, respectively. Various authors have suggested that a specific mechanism acts to maintain the multicopy nature of highly transcribed genes (Tartof, 1988). This previous work on *S. cerevisiae* thus suggests that a similar multicopy maintenance mechanism probably functions to preserve the 120 to 130 copies (authors unpublished) of the rRNA cistron in *C. manginecans*.

In this study, the effects of concerted evolution were more pronounced after meiosis than mitosis. Statistical analyses of the frequencies of the two ITS types examined, indicated that the ratios changed more often and sometimes more dramatically as a consequence of meiosis than mitosis. In CMW 13582, for example, the ratio of the ITS types Z and Y changed from 55:5 in one generation to 9:51 in the subsequent meiotic generation. This change in ratio was found in a single ascospore and was thus the result of a single meiotic event. In order to generate an ascospore having a 9:51 ratio of the two ITS variants, a reduction of the rRNA cistron copy number to around 12 copies would be required, if this were the result of a single cell division. Ascospore development in fungi such as *C. manginecans* involves three cell

divisions, the first two are those normally observed during meiosis and then there is an additional cell division to produce eight ascospores. But, whichever way the problem is examined, at least 4 divisions would be needed to change from a ratio of 55:5 to 9:51 [Figure S3], if the copy number of the cistron remained the same or was in some way constrained [Figure S3A]. However, if there is no restriction placed on the cistron copy number then the reduction in the ITS sequence types can occur in just one division [Figure S3B]. This, therefore, suggests that significant reduction in the copy number of the cistrons can occur as a consequence of unequal recombination.

Although the occurrence of multiple ITS types in *C. manginecans* suggests, at first glance, a relaxation in the mechanisms driving concerted evolution of the rRNA cistron, there are other more plausible explanations for their occurrence. For example, in their seminal work on the rRNA cistrons of humans and other primates, Arnheim et al. (Arnheim *et al.*, 1980) showed that some regions of this unit remain polymorphic and that these polymorphisms are maintained through evolutionary time by natural selection. But unlike in the primate situation, all individuals of *C. manginecans* do not harbour both types of cistron [Figure S1]. This is probably because the *C. manginecans* polymorphisms examined in this study do not represent alleles that were conserved during evolution. An alternative hypothesis would be that ITS types Y and Z represent polymorphisms that were united into the same genome following an interspecies hybridization event. In fact, intraspecific polymorphisms in the rRNA cistron are commonly thought to be a property of hybrid species (de Sousa Queiro *et al.*, 2011; O'Donnell & Cigelnik, 1997; Odorico & Miller, 1997). Following this hypothesis, the dynamic processes of concerted evolution led to the loss of the second ITS type from some isolates of this fungus (CMW 13581, 13584, 23641 and 23643).

Unequal crossing over, though considered to be rare during mitosis, has previously been reported to occur during this life stage (LaFave & Sekelsky, 2009). It was demonstrated that *S. cerevisiae* tagged mutants underwent mitotic recombination during the interphase process of the cycle replication (Szostak & Wu, 1980). These sister chromatids were found to be recombining and resulted in the production of diploid cells (Tartof, 1974). The mitotic chromosomes thus allow for the formation of homolog pairs, which greatly enhance the ease of recombination due to the formation of synaptonemal complexes (Zickler & Kleckner, 1999). Literature also verifies that crossing over events are driven by various genes acting in unity to achieve the variation (Zickler & Kleckner, 1999). Studies have, however, not been able to show empirical evidence for the unequal crossing over event actively occurring during

the cellular divisions of meiosis and mitosis within the same biological organism (Zickler & Kleckner, 1999). The results of this study, therefore, provide evidence that recombination occurs in a non-reciprocal manner and that the overall concerted evolution of the rRNA cistron is mediated in a random and not directed fashion.

A model is proposed to explain our results using unequal crossing over during recombination, [Figure 3] in a hypothetical situation where there are 20 copies of the same repeat unit of the haploid fungus *C. manginecans* that reproduces homothallically. If the repeat units recombine only in a reciprocal process [Figure 3A] then the ratios of the different ITS types would have remained static. In order to explain the differing ratios observed in this study, it is necessary to propose that non-reciprocal crossing over occurs [Figure 3B]. It is thus possible that the copy number of these multigene families can be dramatically reduced or expanded. However, there would be selection against any individual cell in which the copy number drops below a minimum level to sustain cell function, while there is selection against cells containing more than a certain maximum number of copies. Via this dynamic process, the copy number of this multicopy element is thus maintained at some ideal number and the gene sequences continuously homogenised to maintain identical (or near identical) sequences in each unit. Ultimately, these unequal crossing over events have the potential to result in gene conversion [Figure 3, Gene conversion Event and thus loss of one or the other ITS type.

Previous models addressing concerted evolutionary processes have commonly been based on hybrid organisms containing multiple copies of the ITS repeat units which homogenize over the generations through meiotic and mitotic cellular divisions (de Sousa Queiro *et al.*, 2011). The present study also considered an organism in which isolates had two different ITS sequences. The fact that this fungus exists in a haploid state, a situation common to many fungi but unique to most other eukaryotes, and the fact that it is able to undergo sexual reproduction without outcrossing (homothallic) provided us with a unique opportunity to observe experimentally changes in the ratios of the different rRNA types during meiosis and mitosis. Other than the ITS ratios for isolates CMW 17570, 17568 and 23635, which did not change significantly in their mitotic lifecycles, the results suggest that the repeat units of the rRNA cistron can undergo significant size changes during relatively few cell divisions as was the case for isolate CMW 13582. Thus our presented model summarises the overall observations and illustrates how concerted evolution is a consequence of unequal recombination which ultimately over time leads to gene conversion.

4.5 References

Aanen D. K., Kuyper T. W., and Hoekstra R. F. (2001). A widely distributed ITS polymorphism within a biological species of the ectomycorrhizal fungus *Hebeloma velutipes*. *Mycological Research* **105:** 284 - 290.

Arnheim N., Krystal M., Schmickel R., Wilson G., Ryder O., and Zimmer E. (1980). Molecular evidence for genetic exchange among ribosomal genes on nonhomologous chromosomes in man and apes. *Proceedings of the National Academy of Sciences of the United States of America* **77:** 7323-7327.

Brown D. D., Wensink P. C., and Jordan E. (1972). Comparison of Ribosomal DNAs of *Xenopus-Laevis* and *Xenopus-Mulleri*: Evolution of Tandem Genes. *Journal of Molecular Biology* **63:** 57-73.

Chen J.-M., Cooper D. N., Chuzhanova N., Ferec C., and Patrinos G. P. (2007). Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics* **8:** 762-775.

Cronn R. C., Zhao X., Paterson A. H., and Wendel J. F. (1996). Polymorphism and Concerted Evolution in a Tandemly Repeated Gene Family: 5S Ribosomal DNA in Diploid and Alloploid Cottons. *Journal of Molecular Biology* **42:** 685-705.

de Sousa Queiro C., de Carvalho Batista F. R., and de Oliveria L. O. (2011). Evolution of the 5.8S nrDNA gene and internal transcribed spacers in *Carapichea ipecacuanha* (Rubiaceae) within a phylogeographic context. *Molecular Phylogenetics and Evolution* **59:** 293-302.

Dover G. A. (1982). Molecular drive: a cohesive mode of species evolution. *Nature (London)* **299:** 111-117.

Eickbush T. H., and Eickbush D. G. (2007). Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics* **175:** 477-485.

Elder J. F., and Turner B. J. (1995). Concerted Evolution of Repetitive DNA Sequences in Eukaryotes. *Quartly Review of Biology* **70:** 297-320.

Ganley A. R. D., and Kobayashi T. (2011). Monitoring the Rate and Dynamics of Concerted Evolution in the Ribosomal DNA repeats of *Saccharomyces cerevisiae* using Experimental Evolution. *Molecular Biology and Evolution* **28:** 2883-2891.

Hall T. A. (1999). BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95-98.

Holliday R. (1964). A mechanism for gene conversion in fungi. *Genetical Research* **5:** 282-304.

Ide S., Miyazaki T., Maki H., and Kobayashi T. (2010). Abundance of Ribosomal RNA gene copies maintains genome integrity. *Science* **327:** 693-696.

Innan H. (2011). Gene Conversion in Duplicated Genes. *Genes* **2:** 394-396.

LaFave M., and Sekelsky J. (2009). Mitotic Recombination: Why? When? How? Where? *Public Library of Science Genetics* **5:** 1-2.

Lange C., Zerulla K., Breuert S., and Soppa J. (2011). Gene conversion results in the equalization of genome copies in the polypoid haloarchaeon *Haloferax volcanii*. *Molecular Microbiology* **83:** 666-677.

Liao D. (1999). Concerted Evolution: Molecular Mechanism and Biological Implications. *American Journal of Human Genetics* **64:** 24-30.

Liao D. (2000). Gene conversion drives within genic sequences: Concerted evolution of ribosomal RNA genes in bacteria and archaea. *Journal of Molecular Evolution* **51:** 305-317

Liao D. (2003). Concerted Evolution. *In* "Nature Encyclopaedia of the Human Genome" (D. N. Cooper, Ed.), pp. 938-942, Nature Publishing Group, London, New York and Tokyo.

Lichten M. (2001). Meiotic recombination: Breaking the genome to save it. *Current Biology* **11:** 253-256.

Lindegren C. C. (1953). Gene conversion n *Saccharomyces*. *Journal of Genetics* **51:** 625-637.

Long E. O., and Dawid I. B. (1980). Repeated Genes in Eukaryotes. *Annual Review of Biochemistry* **49:** 727-764.

Lumbsch H. T., and Leavitt S. D. (2011). Goodbye morphology? A paradigm shift in the delimitation of species in lichenized fungi. *Fungal Diversity* **50:** 59-72.

Nei M., Gu X., and Sitnikova T. (1997). Evolution by the Birth-and-Death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 7799-7806.

Nei M., Rogozin I. B., and Piontkivska H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proceedings of the National Academy of Sciences of the United States of America* **97:** 10866-10871.

Nei M., and Rooney A. P. (2005). Concerted and Birth-and-Death Evolution of Multigene Families. *Annual Review of Genetics* **39:** 121-152.

O'Donnell K., and Cigelnik E. (1997). Two Divergent Intragenomic rDNA ITS2 Types within a Monophyletic Lineage of the Fungus Fusarium Are Non-orthologous. *Molecular Phylogenetics and Evolution* **7:** 103-116.

Odorico D. M., and Miller D. J. (1997). Variation in the ribosomal internal transcribed spacers and 5.8S rDNA among five species of *Acropora* (Cnidaria; Scleractinia): Patterns of Variation consistent with Recticulate Evolution. *Molecular Biology and Evolution* **14:** 465-473.

Pinhal D., Yoshimura T. S., Araki C. S., and Martins C. (2011). The 5S rDNA family evoles through concerted and birth-and-death evolution in fish genomes: an example from fresheater stingrays. *BMC Evolutionary Biology* **11:** 151.

R_Development_Core_Team (2007). A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* **R version 2.6.1 (2007-11-26)**.

Ritossa F. M. (1968). Unstable redundancy of genes for ribosomal RNA. *Proceedings of the National Academy of Sciences of the United States of America* **60:** 509-516.

Rooney A. P., and Ward T. J. (2005). Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 5084-5089.

Schoch C. L., Seifert K. A., Huhndorf S., Robert V., Spouge J. L., Levesque C. A., Chen W., and Consortium F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 6241-6246.

Sokal R. R., and Rohlf F. J. (1995). "Biometry: the principles and practices of statistics in biological research," W.H. Freeman and Co.:New York.

Szostak J. W., and Wu R. (1980). Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature (London)* **284:** 426-430.

Tartof K. D. (1974). Unequal mitotic sister chromatid exchange as the mechanism of ribosomal RNA gene magnification. *Proceedings of the National Academy of Sciences of the United States of America* **71:** 1272-1276.

Tartof K. D. (1988). Unequal crossing over then and now. *Genetics* **120:** 1-6.

van Wyk M., Al Adawi A. O., Khan I. A., Deadman M. L., Al Jahwari A. A., Wingfield B. D., Ploetz R., and Wingfield M. J. (2007). *Ceratocystis manginecans* sp. nov., causal agent of a destructive mango wilt disease in Oman and Pakistan. *Fungal Diversity* **27:** 213-230.

van Wyk M., Roux J., Barnes I., Wingfield B. D., and Wingfield M. J. (2006). Molecular phylogeny of the *Ceratocystis moniliformis* complex and description of *C. tribiliformis sp. nov*. *Fungal Diversity* **21:** 181 - 201.

White T. J., Bruns T., Lee S., and Taylor J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *In* "PCR protocols: A sequencing guide to methods and applications" (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, Eds.), pp. 315-322 Academic Press, Inc, San Diego.

Witthuhn R. C., Harrington T. C., Wingfield B. D., Steimel J. P., and Wingfield M. J. (2000). Deletion of the *MAT*-2 mating type gene during uni-directional mating-type switching in *Ceratocystis*. *Current Genetics* **38:** 48-52.

Zickler D., and Kleckner N. (1999). Meiotic chromosomes: Integrating structure and function. *Annual Review of Genetics* **33:** 603-754.

Table 1. Summary of Chi-Square values across the replica tests for the respective isolates[a].

Replica Isolate	Meiosis χ2value[b]	Mitosis χ2value[c]
$\chi^2_{0.01\,[4*]} = 13.28$[d]		
CMW 13582	**18.34**	**19.32**
CMW 17568	**24.92**	11.99
CMW 17570	**26.11**	3.04
CMW 23635	**24.11**	2.21

Footnote:

[a]For each isolate the corresponding degrees of freedom(df)* was calculated as (n-1), where n is the number of transfers per isolate.

[b,c]Combined replica data for respective meiosis and mitosis chi-square values as determined in supplementary information [**Table S1**]. These data are based on five sequential rounds of either mitotic transfers or sexual crosses for each isolate.

[d]The corresponding chi-square value when a 99.999% level of confidence is applied at 4 degrees of freedom. The level of significance was established on the basis of the χ^2prediction been greater or less than the respective replica isolate χ^2value. Thus, values in bold denote significant levels of change, whilst the clear blocks denote non-significant change.

Table 2. Combined Chi-Square value statistics across all replicas for meiosis[a].

Replica Isolate	**Meiosis χ^2 value[b]**	**Probability Value[c]**	**ln (P)[d]**; (Sokal & Rohlf, 1995)
CMW 13582	18.34	0.010	-4.605
CMW 17568	24.92	5.221×10^{-5}	-9.860
CMW 17570	26.11	3.007×10^{-5}	-10.412
CMW 23635	24.11	7.592×10^{-5}	-9.486
•At 2*k*, where *k* = the number of separate tests and probabilities, thus degrees of freedom = 2(4) =8			**$-2 \sum \ln(P) = 68.726$***
$\chi^2_{0.001\,[8]}$ = **68.726>1.214 x 10 $^{-12}$, thus highly significant** (R_Development_Core_Team, 2007)			

Footnote:

[a]The different fungal isolate serves as a replica of the same statistical test for meiosis.

[b]Chi-square value as determined from the original replica isolates as shown in supplementary information [**Table S1**] for that specific isolate.

[c]Lowest possible probability score associated to that chi-square value determined by R-statistical software algorithmic programming (R_Development_Core_Team, 2007).

[d]The logarithmic value of the probability[c]. *The formula used to derive the sum of all the separate tests for each replica.

•The number of degrees of freedom is established by taking twice the number of separate tests and probabilities carried out initially (thus 2*k* = 2(4) = 8 degrees of freedom). A 99.999% level of confidence was applied.

Table 3. Combined Chi-Square value statistics across all replicas for mitosis[a].

Replica Isolate	**Mitosis χ^2 value[b]**	**Probability Value[c]**	**ln (P)[d]**; (Sokal & Rohlf, 1995)
CMW 13582	19.32	0.0007	-7.264
CMW 17568	11.98	0.0175	-4.046
CMW 17570	3.04	0.551	-0.596
CMW 23635	2.21	0.7155	-0.335
•At 2*k* where *k* = the number of separate tests and probabilities, thus degrees of freedom = 2(4) =8			**-2 ∑ ln(P) = 24.482***
$\chi^2_{0.001\ [8]}$ = **24.482 > 0.00186, thus highly significant** (R_Development_Core_Team, 2007)			

Footnote:

[a]The different fungal isolate serves as a replica of the same statistical test for mitosis.

[b]Chi-square value as determined from the original replica isolates as shown in supplementary information [**Table S1]** for that specific isolate.

[c]Lowest possible probability score associated to that chi-square value determined by R programming (R_Development_Core_Team, 2007).

[d]The logarithmic value of the probability[c] .

[*]The formula used to derive the sum of all the separate tests for each replica.

[•]The number of degrees of freedom is established by taking twice the number of separate tests and probabilities carried out initially (thus 2*k* = 2(4) = 8 degrees of freedom). A 99.999% level of confidence was applied.

Table S1. Summary of the observed ITS sequence types from 60 cloned amplicons across all test replicas for both meiosis and mitosis based on five sequential rounds of either mitotic transfers or sexual crosses for each isolate◄.

Spore Isolate	Meiosis ITS Type		Hyphal Tip Isolate	Mitosis ITS Type		Spore Isolate	Meiosis ITS Type		Hyphal Tip Isolate	Mitosis ITS Type	
	Z	Y		Z	Y		Z	Y		Z	Y
CMW 13582.1	17	43	CMW 13582.a	56	4	CMW 17570.1	36	24	CMW 17570.a	3	57
CMW 13582.2	55	5	CMW 13582.b	12	48	CMW 17570.2	7	53	CMW 17570.b	13	47
CMW 13582.3	9	51	CMW 13582.c	18	42	CMW 17570.3	43	17	CMW 17570.c	6	54
CMW 13582.4	34	26	CMW 13582.d	32	28	CMW 17570.4	9	51	CMW 17570.d	2	58
CMW 13582.5	17	43	CMW 13582.e	11	49	CMW 17570.5	58	2	CMW 17570.e	4	56
Expected value	26.4	33.6	Expected value	25.8	34.2	Expected value	30.6	29.4	Expected value	5.6	54.4
χ^2	10.27	8.07	χ^2	11.01	8.30	χ^2	12.79	13,31	χ^2	2.76	0.28
Combined χ^2	**18.34**		**Combined χ^2**	**19.32**		**Combined χ^2**	**26.11**		**Combined χ^2**	**3.04**	
Spore Isolate	**Meiosis ITS Type**		**Hyphal Tip Isolate**	**Mitosis ITS Type**		**Spore Isolate**	**Meiosis ITS Type**		**Hyphal Tip Isolate**	**Mitosis ITS Type**	
	Z	Y		Z	Y		Z	Y		Z	Y
CMW 17568.1	12	48	CMW 17568.a	37	23	CMW 23635.1	21	39	CMW 23635.a	14	46
CMW 17568.2	54	6	CMW 17568.b	4	56	CMW 23635.2	57	3	CMW 23635.b	13	47
CMW 17568.3	10	50	CMW 17568.c	15	45	CMW 23635.3	11	49	CMW 23635.c	9	51
CMW 17568.4	29	31	CMW 17568.d	6	54	CMW 23635.4	57	3	CMW 23635.d	2	58
CMW 17568.5	2	58	CMW 17568.e	23	37	CMW 23635.5	33	27	CMW 23635.e	11	48
Expected value	21.4*	38.6	Expected value	17	43	Expected value	35.8	24.2	Expected value	9.8	50.2
χ^2	16.03**	8.89	χ^2	8.59	3.40	χ^2	9.73	14.39	χ^2	1.85	0.36
Combined χ^2	**24.92[a]**		**Combined χ^2**	**11.99**		**Combined χ^2**	**24.12**		**Combined χ^2**	**2.21**	

*Expected value calc. = $\frac{\text{total \# clones x total \# of ITS type Z for that isolate}}{\text{total \# of clones for the experiment}}$

i.e Expected value = $\frac{60 \times 107}{300}$

= 21.4

**$\chi^2 = \frac{\Sigma(12-21.4)^2 + (54-21.4)^2 + (10-21.4)^2 + (29-21.4)^2 + (2-21.4)^2}{(5 \times 21.4)}$

χ^2 = 16.03 for ITS Z types

[a] Combined χ^2 value = ITS type Z χ^2 + ITS type Y χ^2
i.e. Combined χ^2 value = 16.03 + 8.89
= 24.92

◄Each new generation of meiotic and mitotic growth is shown as a numerical and alphabetical succession, respectively. Blocks shaded in grey show a statistically significant chi-square value whilst clear blocks maintain a non-significant chi-square value. The calculated expected frequency for the particular ITS type variant is denoted by *. This frequency was calculated for each individual isolate. ** Denotes the calculated chi-square value for the ITS sequence type Z in the meiotic life cycle of the isolate CMW 17568. [a]The combined chi-square analysis was determined using the specific chi-square values of the ITS sequence type Z and Y frequencies for each isolate.

Figure 1. Life cycle of a *Ceratocystis* species.

Shown here for this typical ascomycete includes both sexual (meiotic) and asexual (mitotic) cycles. In this case both the meiotic and mitotic states occur in a single haploid culture, a condition known as homothallism in fungi.

Figure 1.

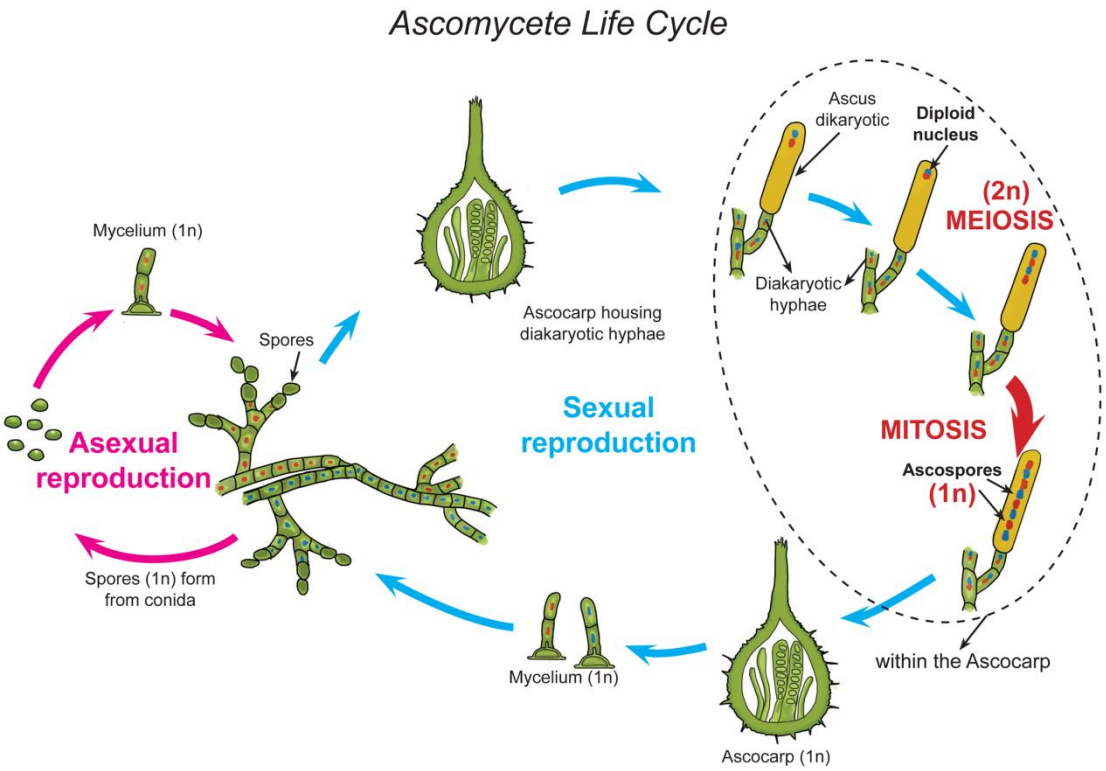


Figure 2. Diagrammatic representation of the methodology employed in the experimental design.

The fungus *Ceratocystis manginecans*, isolate CMW 17568 was derived from single meiospores (ascospores) to generate the meiotic progeny, and the four sequential single hyphal tip isolations generated the mitotic generations of the fungus.

Figure 2.

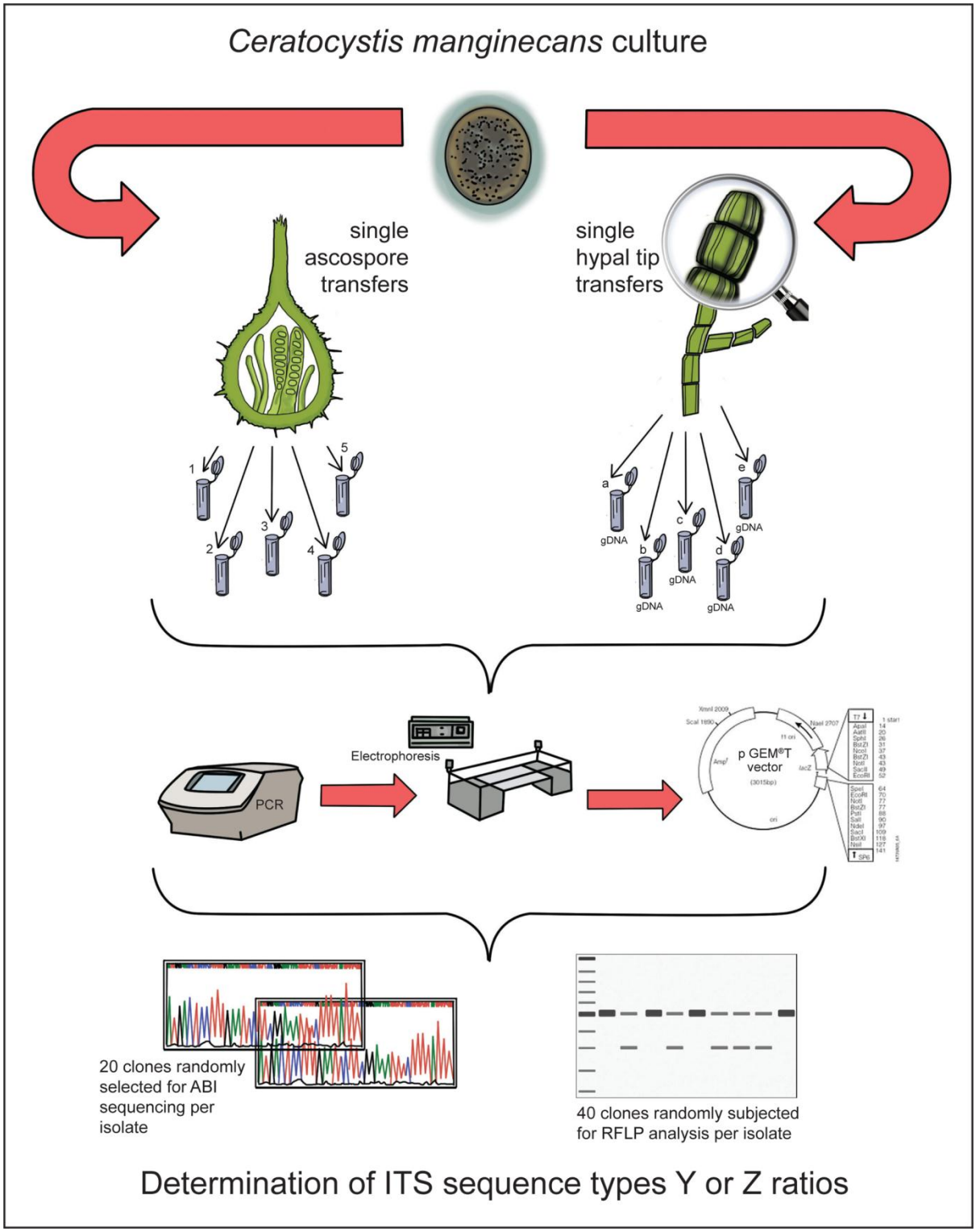
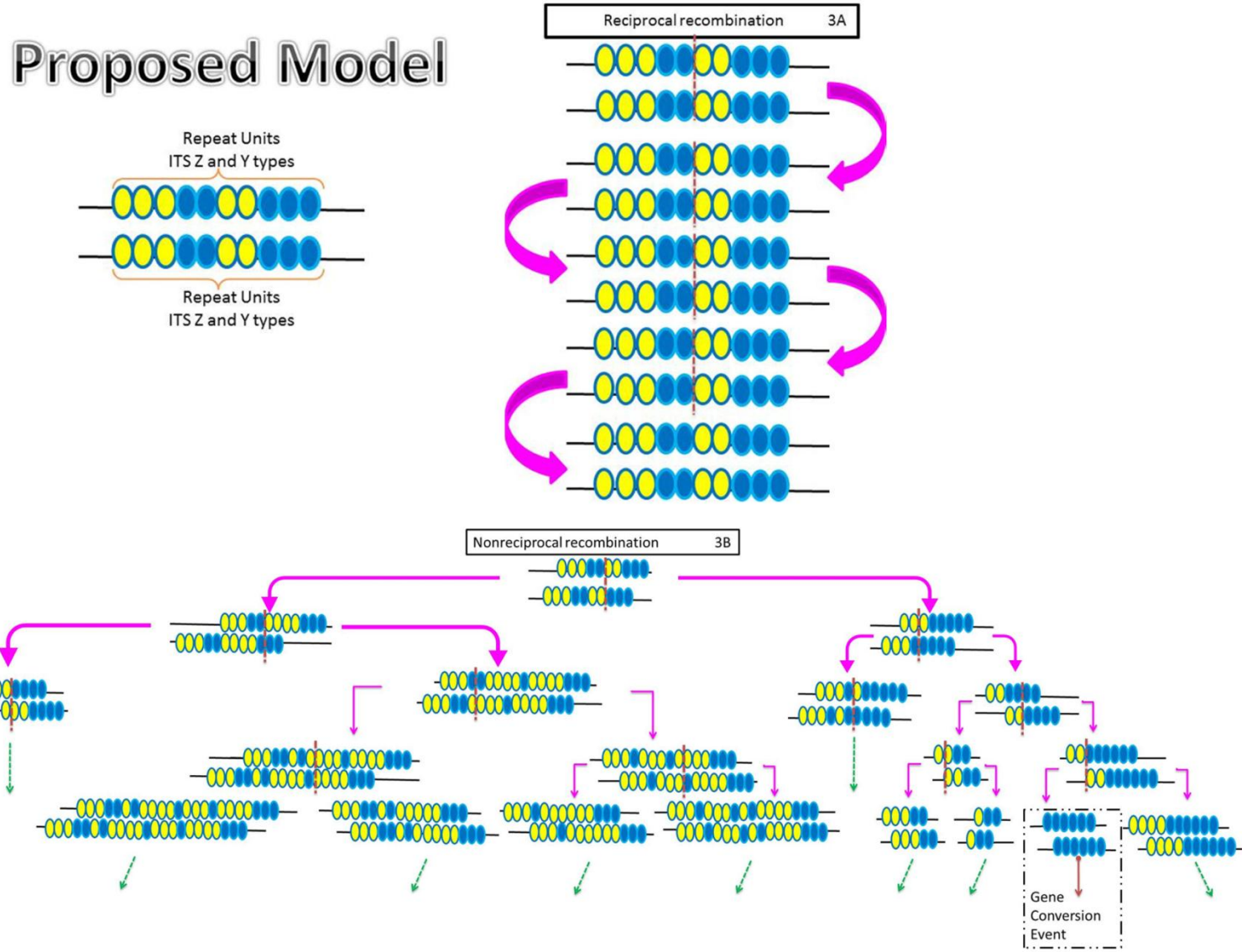


Figure 3. Diagram illustrating a proposed model of concerted evolution in *Ceratocystis manginecans*.

Concerted evolution is seen as a result of both gene conversion and unequal crossing over occurring during both the meiotic and mitotic life cycles of *Ceratocystis manginecans* on a hypothetical chromosome which contain both ITS type Z and Y sequences. Yellow circles represent ITS type Z sequence repeat units whilst blue circles represent ITS type Y sequence repeat units. Crossing over (broken red lines) is shown hypothetically as reciprocal recombination [3A] or nonreciprocal recombination [3B] events, in each case the repeat units are subjected to gain and loss events which can result in Gene conversion (solid red line). The broken green arrows indicate multiple unequal crossing events.

Figure 3.



Supplementary Data

Figure S1. Aligned DNA sequence.

Screen print showing the aligned DNA sequences of only ITS sequence type Y in *Ceratocystis manginecans* isolates CMW 13581, 13584, 23641 and 23643.

Figure S1.

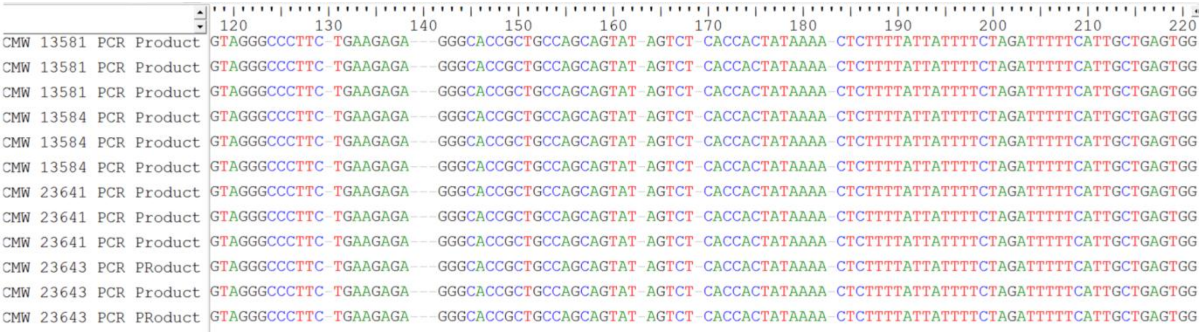


Figure S2. Aligned DNA sequence.

Screen print of aligned DNA sequences showing the differences in the sequences of the two ITS sequence types (Z and Y) in *Ceratocystis manginecans* CMW 17568.

Figure S2.

Figure S3. Hypothetical rRNA cistron showing meiotic and mitotic divisions.

Illustration of a hypothetical situation for *Ceratocystis manginecans* where we assume that its rRNA cistron has 600 copies undergoing meiotic and mitotic divisions. In this example, under meiotic conditions each division represents either an increase or a decrease in the ITS sequence types. Figure S3A depicts a restricted cistron size scenario whilst Figure S3B has no size restriction.

Figure S3.

Conclusion and future prospects

Mitochondria are organelles that are found in eukaryotic cells, where they primarily function as an electron transport system in correlation with oxidative phosphorylation mechanisms to generate energy for an organism. In recent years the study of the genomes of these organelles has gained momentum across all research fields. This is due to the advances in sequencing platforms accelerating the amount of mitochondrial DNA data available for whole mitochondrial genome characterization in combination with proteomic and transcriptomics analyses. Such research allows researchers the potential to address questions pertaining to various aspects of an organisms' genetic background, taking into account the identification of putative mitochondrial proteins which may have downstream applications implicating them in a plethora of hosts, pathogenicity and species differentiation scenarios.

Fungal mitochondrial genetics focuses primarily on the variation observed in their genome sizes, gene architecture and the presence or absence of mobile elements. Within the Ascomycota phylum exists examples of fully characterized mitochondrial genomes. It was observed that most distinguishable feature of these mitochondrial genomes was the difference in the genome size, ranging on average from 24 kb to 100 kb. This considerable genome size variation is largely due to the presence or absence of mobile elements such as introns, the possible insertion of plasmids into the mitochondrial genomes at various points during a fungus' evolutionary history. The studies carried out in this thesis provide an initial start in the description of fungi found in the genus *Ceratocystis*. I have with the thesis focused on the characterisation of these mitochondrial genomes. At the inception of this study, there were no mitochondrial genomes available from the order Microascales. The fact that the first *Ceratocystis* mitochondrial genome, that of *Ceratocystis cacaofunesta* was published only this year serves to highlight the scarcity of data on mitochondrial genomes in the Microascales. I have with the thesis characterized the mitochondrial genomes of the species *Ceratocystis fimbriata*, *Ceratocystis albifundus* and *Ceratocystis moniliformis* and thus added to the data now available for the mitochondrial genomes of *Ceratocystis*. The results of the study are largely congruent with the published data on *C. cacaofunesta* with the exception of a single large inversion.

Mitochondrial genomes contain a very AT rich content and many repetitive DNA sequence regions. This combination makes it difficult to assemble these genomes from short DNA

sequences. As total genomic DNA was isolated for the different next generation sequencing, it was imperative that only mitochondrial DNA sequence reads was used for the *de-novo* assemblies. It is thus not improbable that partial mitochondrial genes were thus annotated in this sifting process. While most of the sequence data generated is short DNA sequences, there were longer DNA sequences (> 500 bp) for both C. *fimbriata* and *C. albifundus* mitochondrial genomes generated using the FLX 454 technology. Thus the use of these longer DNA sequence fragments to *de-novo* assemble these genomes increased the confidence that the mitochondrial genomes for *C. fimbriata, C. albifundus and C. moniliformis* were correct for this species. While *C. cacaofunesta* might in fact have this inversion the gene order of all the *Ceratocystis* mitochondrial genomes should be further confirmed using PCR to validate the true gene order.

The resulting characterized genomes of *C. fimbriata*, *C. albifundus* and *C. moniliformis* allowed for a comparative mitochondrial genome analysis for these species. The substantially large mitochondrial genomes for each of these species was not anticipated, as mitochondrial genomes previously described, though varying in their genome size, were not in the magnitude that of *Ceratocystis*. With the recent publication of a completely annotated mitochondrial genome of *C. cocoafunesta*, our results with regard to gene, intron and tRNA annotation are very much comparable to published work, increasing our confidence in our findings. All mitochondrial genes presented in *C. fimbriata*, *C. albifundus* and *C. moniliformis* correlated to that described in the mitochondrial genome of *C. cocoafunesta*. This further validates our conclusions for the comparative analyses carried out.

Within the genus *Ceratocystis* three broadly classified phylogenetically different clades exists. The *C. fimbriata sensu lato* clade is however shrouded in cryptic species complexes. One such potential species complex found within this clade is the different isolates of *C. albifundus*. Thus the study of the mitochondrial genomes of different isolates of *C. albifundus* can be investigated in order to establish if there is a comparable variation among the isolates of *C. albifundus* spp. The intraspecies characterization of the mitochondrial genomes of six isolates of *C. albifundus* showed very little variation in their gene content and organization. The sequencing and assembly of the mitochondrial genomes of isolates from different parts of the African continent shed light on the uncertainty regarding the country of

origin of this species which was previously thought to be from South Africa. Our resulted showed high levels of similarity among the isolates with the only variation seen in the intergenic regions of their mitochondrial genomes. This was an unexpected result, as we anticipated that there would be a significant variation in the isolates especially as the isolates were chosen on their different geographical location. We did observe that the isolates from Tanzania and Kenya shared a hundred percentage sequence similarity, suggestive that the origin of this pathogen may very well be in the Southern parts of Africa, and not South Africa itself. The mitochondrial genes *atp8*, *atp9*, *nad5*, *coxIII* and *cob* all presented as potential genes to be further investigated as these gene regions showed a discriminatory power between the different isolates. Therefore, the potential of these mitochondrial genes to allow for further studies detailing the spread of the fungus particularly from native to plantation areas can also be addressed. This intraspecies mitogenomic analysis of isolates from the genus *C. albifundus* and whilst there are studied done of a similar nature, very few focus on the use mitochondrial genomes of the Ascomycota. This sets the tone for using a similar approach for an interspecies study where a number of isolates for the different *Ceratocystis* species can be characterized and potentially delimit the species complexes where current phylogenetic classification is problematic.

Species characterization has relied on morphological as well as phylogenetic analyses. For fungi the latter has largely depended on the regions of ribosomal RNA cistron, most commonly the intergenic spacer region (ITS). This is particularly the case for the genus *Ceratocystis*. The ITS region is in fact the region of preference for barcoding for all but a few fungal taxa. The use of the ITS gene region in some fungal groups is controversial as a consequence of there being more than one possible sequence type. The ribosomal RNA cistron has been the focus of much attention from geneticists as a consequence of the fact that despite it being composed of tandem gene repeat units, these are all apparently identical. The manner in which this similarity is maintained has been a subject for great debate over the years. As a chance observation during this research we discovered that *Ceratocystis manginecans* contained two different ITS sequences and that the ratio of these sequences was not constant during culturing. Allowing this fungus to undergo both mitotic and meiotic cell division, we provided empirical evidence that unequal crossing of these different ITS types ultimately leads to gene conversion events, specifically that of concerted evolution of the ribosomal RNA cistron.

We have been able propose a model which shows concerted evolution in action within *C. manginecans*. While *C. manginecans* is perhaps not the most tangible model organism to study such a fundamental process in eukaryotic cells, the combination between it being homothallic and having the two ITS types made it uniquely suited to this study. Our study is further supported by the observations seen in *Saccharomyces cerevisiae* where recombination during mitosis was previously demonstrated. In another study also focusing on *S. cerevisiae*, it was shown that the copy number of the rRNA cistron is probably maintained as a consequence of selection, where isolates where the copy number goes below a certain minimum threshold, these isolates grow more slowly. It is likely that cells which accumulate too high a copy number would be similarly less fit. Thus it seems that this multi gene family retains its identity through unequal recombination events, constantly losing single mutations in the population. The possibility of a mutation becoming fixed via this recombinational process can also hold true as we have in possession isolates of *C. manginecans* which contain only one ITS sequence type. It can be speculated that the isolates containing the two different ITS DNA sequence types could very well be as a result of a hybridization event; however this needs further empirical evidence investigating the actual copy number of the rRNA cistron. This discovery of two ITS types lays the foundation for subsequent findings in other species of *Ceratocystis* which may also contain more than one ITS type which at present makes their phylogenetic species classification difficult. Once again this comes back to the fact that even though the ITS gene is a popular barcode it remains problematic for the clear differentiation of certain fungal taxa.

The future prospects of the studies undertaken in this thesis have far reaching potential not only for mitogenomic analyses but also for species identification and population studies. In this study we have examined the mitochondrial genomes of only 3 species. We will shortly have access to the whole genome sequence of another 14 *Ceratocystis* species. Interrogation of the mitochondrial genomes for each of these species in combination with an interspecies analysis provides us with the tools for unlocking many of the species which form cryptic species complexes. The DNA sequencing data for the different *Ceratocystis* species can also be exploited via the use of a combination of diverse mitochondrial genes to aid in future systematic studies. Questions pertaining to the mode of mitochondrial inheritance and hybridization events among these species can also be addressed by investigating molecular clocks tracing ancestral history specifically for the genus *Ceratocystis*.

Summary

The objective of this study was to characterize the mitochondrial genomes of the species within the genus *Ceratocystis* and investigate the evolutionary process of the ribosomal RNA cistron found within these fungi. *Ceratocystis* incorporates a number of pathogenic species affecting a variety of hosts, making the study of these fungi economically significant. The fortuitous identification of a *Ceratocystis* species, *C. manginecans*, which contained two different internally transcribed spacer sequence types within the ribosomal rRNA cistron, enabled a study of concerted evolution in this fungus. Using this non-model organism we were able to show empirical evidence for unequal crossing over and gene conversion as the ultimate forces acting on this gene region dictating a concerted evolutionary effect. We suggest that this process is true for all eukaryotes. Using the knowledge drawn from previously characterized and annotated mitochondrial genomes of other eukaryotes, the genomes of three *Ceratocystis* species, namely *Ceratocystis fimbriata*, *Ceratocystis albifundus* and *Ceratocystis moniliformis* were fully assembled and annotated for comparative analysis. This comparative study addressed the genome size, gene content, tRNA presence as well as intron types and their homing endonucleases found among these three mitochondrial genomes. An interspecies characterization was then undertaken using the mitochondrial genomes of six different *C. albifundus* isolates from different geographical locations in Africa. Genetic variation and similarities among these isolates supports the previous hypothesis that the origin of this fungus is Southern Africa. It is hoped that the research presented in this thesis will contribute to the improved understanding of the mitochondrial genomes in *Ceratocystis* species.