UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Distinguishing between cryptic species in the *Ceratocystis fimbriata sensu lato* species complex

by

Arista Fourie

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural & Agricultural Sciences,
Department of Genetics and
Forestry and Agricultural and Biotechnology Institute
at the University of Pretoria, Pretoria, South Africa

June 2014

Supervisor: Dr. Irene Barnes
Co-supervisors: Prof. Michael J. Wingfield
Prof. Brenda D. Wingfield

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Faculty of Natural and Agricultural Sciences

Declaration

I, the undersigned, **Arista Fourie**, declare that the dissertation, which I hereby submit for the degree **Magister Scientiae** at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

_____

Arista Fourie

June 2014

Table of contents

Acknowledgements

I want to show my deepest gratitude towards my supervisors, Dr. Irene Barnes, Prof. Brenda Wingfield and Prof. Mike Wingfield for their guidance and support throughout this study. Your input in the planning and execution of the research as well as in the writing of this dissertation made this immense task achievable. Thank you for your time, effort and involvement. A special thanks to Irene for the example you set, for your positive attitude and motivation, you truly are an inspiration.

Thank you to my lab partners at the FABI facility for their friendship and interest in my studies. Thank you for all your help on the technical aspects of my research as well as for personal motivation. A special thanks to Katrin Fitza, who has also become a dear friend beyond academic environments.

For the friends that accompanied me throughout my studies, thank you Elrea, Velushka and Brigitte for your friendship and interest in my well-being. I appreciate the impact each of you have made in my life over the past three years. I am also grateful for every friend I have been blessed with, thank you for ensuring I had enjoyable times even among the academic obligations.

A special thanks to the sequencing facility, specifically Mrs. Renate Zipfel and Mrs. Gladys Shabangu for their assistance and supportive service with regard to the generation of the sequence data. Thank you to the culture collection team at FABI for your friendly assistance and help during the course of this study. Thanks also to Dr. Tuan Dong and Dr. Wilhelm de Beer for their assistance in the planning and design of some of the primers used in this study.

I am very grateful for the financial support I obtained during the course of my studies. I want to acknowledge the Forestry and Agricultural Biotechnology Institute (FABI) and, specifically, the Tree Protection Co-operative Program (TPCP) for financial support to execute this research and for providing a stimulating research environment. This research is also supported by the National Research Foundation (NRF) of South Africa and the University of Pretoria. I also want to acknowledge the Thuthuka Foundation for the provision of a student bursary grant for the studies performed.

I mention the following acknowledgements lastly, simply because it formed the foundation for my studies throughout the course of my degree. I am extremely blessed with a family that believes in me and motivates me in everything I do. Thank you for supporting me, firstly on a financial basis, but mostly on a spiritual and emotional basis. Thank you mams, dad and

Alice for being with me every step of the way. Most importantly, I know ‘I can do all things through Christ that strengthens me’ (Phil. 4:13). I am grateful for a Father that gave me the opportunity and the talents to perform this research and for being the source of my joy and strength. None of this would have been possible without the presence of God in every part of my life.

Preface

Species in the *Ceratocystis fimbriata sensu lato* species complex are primarily pathogens that cause disease on a diversity of host plants in almost every continent. These pathogens have had a significant economic impact in the fruit crop and plantation industries in countries where their commercial products are a major source of income. The species in this complex are phylogenetically closely related and contain negligible morphological differences between them. Events such as host jumps or geographical expansions of a specific species can have detrimental effects in countries where the disease is newly introduced. It is thus essential to accurately identify and distinguish between all the species in this complex for effective disease management.

The classification of *Ceratocystis* species has been strongly reliant on DNA sequence data and phylogenetic analyses in the past ten years due to the limitations of other species recognition criteria. In the *C. fimbriata s.l*. species complex some of the distinct species were primarily classified by the phylogeny of the rRNA Internal Transcribed Spacer region (ITS) with little support from the, generally used, β-tubulin 1 (βT 1) and Translation Elongation Factor 1-α (EF 1-α) gene regions. Although support from a single gene region is acceptable in fungal taxonomy, ITS is less reliable in *Ceratocystis* due to the occurrence of intragenomic ITS sequence variation that can classify a single isolate as two different species.

The first chapter of this dissertation provides an overview of the different methods that have been used over the last two centuries to identify, classify and describe species in *Ceratocystis.* These methods have ranged from morphological to biochemical comparisons, mating studies and phylogenetic analyses. In this chapter the effectiveness and limitations of these methods for species delineation are discussed. The significant impact DNA sequence data have had on species delineation is highlighted and, subsequently, the challenges and pitfalls of applying this method are also presented. Moreover, additional approaches that can be used for the identification of *Ceratocystis* species are discussed.

Some species in the *C. fimbriata s.l.* complex are distinguished only by the ITS region with a lack of resolution in other gene regions. The aim of the second chapter of this dissertation was to consider two alternative approaches for the reliable identification of species in the complex. Firstly, the potential of other phylogenetic gene regions to be used for resolving cryptic species boundaries were compared to the three gene regions currently applied in the complex. The second approach was to investigate the potential of single nucleotide polymorphisms (SNPs) to be used as diagnostic markers for species identification. Here nine SNP markers were developed and applied to 15 known species to determine the resolution they provide between cryptic species.

In the third chapter a greater focus is placed on the highly pathogenic species, *C. acaciivora* and *C. manginecans,* and the challenges encountered with their taxonomy. Recent research has suggested these species might be conspecific, due to the occurrence of multiple ITS haplotypes in individual isolates. This could influence previous distribution and population studies performed on either of the species. The aim of this chapter was, firstly, to determine the species identity of *Ceratocystis* species causing disease on *Acacia* trees in Vietnam, based on DNA sequence data. A clonal population of *C. manginecans* occurs in Oman and Pakistan, which indicates the pathogen has been introduced into these countries, but no studies have been performed in South East Asian countries where the species also occur. An additional aim was, thus, to evaluate the genetic diversity and relatedness of different populations of *C. manginecans* in South East Asian countries. A new set of microsatellite markers was designed for this study and these will prove useful for further studies on this economically important group of fungi. This study expands the knowledge on the current distribution of the *C. manginecans* species as well as the genetic diversity of its populations in South East Asia.

The research in this dissertation highlights the importance of accurate identification of cryptic fungal species for scientific studies. Accurate species identification is essential for the implementation of quarantine and trade regulations in plantation industries to prevent spread of the pathogen to a new area. The alternative markers investigated in this study for the reliable identification of closely related species should provide a foundation for similar investigations in other fungal genera.

Chapter 1

Taxonomy of species in the fungal genus *Ceratocystis*

1. Introduction

Classifying organisms into species is essential for communication among scientists as well as to ensure accuracy and consistency in scientific research. For experimental studies to be reproducible, a reliable identification system is required to ensure that the same organism is used in repeated studies. Species delineation in the Fungal Kingdom is of great importance, as many of the fungi are important human and plant pathogens. Without accurate distinction between pathogenic fungal species, it would be impossible to monitor the geographic distribution of a pathogen and implement quarantine measures (Harrington and Rizzo, 1999). Likewise, the development of biological control agents or plant resistance breeding would be inaccurate (Shivas and Cai, 2012).

There is a general agreement that species can be defined as "segments of evolutionary lineages" (Giraud et al., 2008) that "share an evolutionary history with each other more than they do with organisms of other species" (Hey, 2006). Species concepts are the subject of much controversy and this has led to the establishment of approximately 24 alternative species concepts (De Queiroz, 2007; Mayden, 1997). These concepts provide either a theoretical definition of what a species is or they suggest criteria that might be used to recognise species boundaries (Giraud et al., 2008). The main controversy is not with regards to the most accurate species definition (theoretical species concepts) but rather the best criteria to be used for species delineation (operational species concepts).

The three most widely used species recognition criteria for the fungi underpin the morphological, biological and phylogenetic species concepts (Harrington and Rizzo, 1999; Taylor et al., 2000). The focus of this review is on the development and application of methods that are based on these three criteria and the impact that they have had on the identification of species in the fungal kingdom. Here special emphasis is placed on the genus *Ceratocystis* due to the fact that this group of fungi are the main topic of the presented dissertation.

Ceratocystis species are found on a diversity of trees and other crop plants and have been reported in almost every continent (Baker et al., 2003; Roux and Wingfield, 2009; Van Wyk et al., 2006). Lifestyles of the species range from highly pathogenic to non-pathogenic species with some displaying saprophytic and others biotrophic lifestyles on their hosts. Pathogenic *Ceratocystis* species have caused significant economic loss in some countries (Al Adawi et al., 2006; Engelbrecht and Harrington, 2005). It is essential, therefore, to identify these species accurately in order to be aware of their distribution and the economic impact they might have. Furthermore, accurate identification is necessary where effective quarantine measures are required to prevent introductions into new areas.

The genus *Ceratocystis* has a long history of confused classification, emerging due to the limitations of the methods applied for species delineation in earlier years (Bakshi, 1951; Hunt, 1956; Upadhyay, 1981). Recent studies on species in this genus have included

morphological, biological and phylogenetic species recognition. This establishes *Ceratocystis* as a good model organism to investigate the impact of these different criteria on the classification of fungal species in general.

The first criterion applied to species delineation in *Ceratocystis* was based on the morphological variation observed in fungal cultures as well as variation in asexual and sexual reproductive structures (Upadhyay, 1981). All *Ceratocystis* species have well-developed sexual structures (ascomata) with globose bases and long necks that release sticky drops of ascospore masses (Bakshi, 1950; Malloch and Blackwell, 1993; Spatafora and Blackwell, 1994; Upadhyay, 1981). This morphological adaptation is optimal for spore dispersal by insects and the structures are thus highly conserved in the species (Malloch and Blackwell, 1993). Since morphological variation between *Ceratocystis* species is relatively limited, biological species recognition was introduced to define taxa. Intersterility tests enabled the further identification of several cryptic species that appeared similar based only on morphological differences (Engelbrecht and Harrington, 2005). Other methods to differentiate between these fungi included isozyme variation and analysis of restriction fragment length polymorphisms (RFLPs) (Johnson et al., 2005; Witthuhn et al., 1999). The aforementioned methods revealed a limited amount of variation between species and more than 70% of the currently defined *Ceratocystis* species have only been identified and described since the incorporation of the phylogenetic species concept, based on DNA sequence data.

Presently, DNA sequence comparisons are most commonly used for species classification in *Ceratocystis.* These data provided the primary support for the grouping of the genus into three major species complexes, *C. fimbriata sensu lato (s.l.), C. coerulescens s.l*. and *C. moniliformis s.l.* (Wingfield et al., 2013; Witthuhn et al., 1999). Sequence data also allowed for the description of numerous cryptic species in each of the species complexes (Van Wyk et al., 2006; Van Wyk et al., 2012; Van Wyk et al., 2011a). The three gene regions most widely used are the Internal Transcribed Spacer regions 1 and 2 (ITS1 and 2) of the ribosomal DNA, β-tubulin 1 (βT 1) and Translation Elongation Factor 1-α (EF 1-α) (Mbenoun et al., 2013; Van Wyk et al., 2011b; Wingfield et al., 2013). However, recent phylogenetic studies analysing these regions separately have shown limited resolution between some species in the βT 1 and EF 1-α genealogies resulting in the ITS region providing the primary support for distinct species (Van Wyk et al., 2011a; Van Wyk et al., 2012).

The ITS region has been selected as the DNA barcode region for fungal species identification (Schoch et al., 2012). The problem with using it for taxonomic purposes in *Ceratocystis* species is that multiple ITS types have been reported in single isolates (Al Adawi et al., 2013; Naidoo et al., 2013). This has also been reported for *Fusarium* and a few other fungal genera (Lindner et al., 2013; Naidoo et al., 2013; O'Donnell and Cigelnik, 1997). The ITS region should, thus, be used with caution for fungal species delineation in these genera

and requires the incorporation of additional sequence data to support the definition of species boundaries.

This review provides an overview on the species classified in the three main species complexes of the *Ceratocystis* genus and the various methods applied for morphological, biological and phylogenetic recognition of the species. The impact that each of the currently applied methods has had on *Ceratocystis* taxonomy and the limitations of these methods are discussed. Potential alternative approaches that could reliably support the existence of distinct species in the genus, as suggested by ITS, are also investigated.

2. Taxonomy of *Ceratocystis* species

The genus *Ceratocystis* is classified in the Microascales and it forms part of the Sordariomycetes (Lutzoni et al., 2004; Zhang et al., 2006). The type species of the genus, *Ceratocystis fimbriata senu stricto* (*s.s.*), was described in 1890 by Ellis and Halsted as the causal agent of black rot on sweet potato (*Ipomoea batatas*) (Halsted, 1890). Subsequently, multiple new species have been described in this genus and sequence data for the ribosomal SSU were used to classify the species into three main species complexes exemplified by *C. fimbriata sensu lato* (*s.l.*), *C. coerulescens s.l.* and *C. moniliformis s.l.* (Fig. 1) (Wingfield et al., 2013; Witthuhn et al., 1999). The three species complexes are currently under intensive taxonomic investigation and they will likely be treated at the generic level in the near future (Z. W. De Beer, Personal communication).

A potential fourth complex in *Ceratocystis* includes *Thielaviopsis* species. These are the entirely anamorphic species residing in *Ceratocystis s.l.* and include *T. australis*, *T. basicola*, *T. neocaledoniae*, *T. ovoidea*, *T. populi* and *T. thielavioides* (Fig. 1) but the monophyly of all six species in the phylogeny has not been fully resolved when ITS and LSU data is considered (Wingfield et al., 2013). Four additional species that do not group with any of the named complexes represent two potentially smaller, distinct complexes. The one includes *C. fagacearum* and *C. adiposa* and the other includes *C. paradoxa* and *C. radicicola* (Harrington, 2008; Wingfield et al., 2013). The discovery of additional, novel species such as *C. norvegica* (Reid et al., 2010) grouping closely with *C. fagacearum* and *C. adiposa,* might confirm their status as distinct complexes.

The *C. coerulescens* complex was the first species complex to be described in *Ceratocystis* (Harrington et al., 1996; Harrington and Wingfield, 1998). A characteristic morphological trait of species in this complex is the production of oblong-cylindrical ascospores (Fig. 3A) (Bakshi, 1951; Reid et al., 2010) whereas species in the other two complexes produce ascospores with hat-shaped sheaths (Fig. 3B) (Van Wyk et al., 2011b). A broad spectrum of species recognition criteria have been applied for species delimitation in this complex , including morphological comparisons, intersterility tests, isozyme analyses and

phylogenetic analyses (Table 1) (Harrington and McNew, 1997; Harrington et al., 1996; Marin et al., 2005; Stipes, 1970). This has enabled the identification of a total of 12 species in the complex (Table 1).

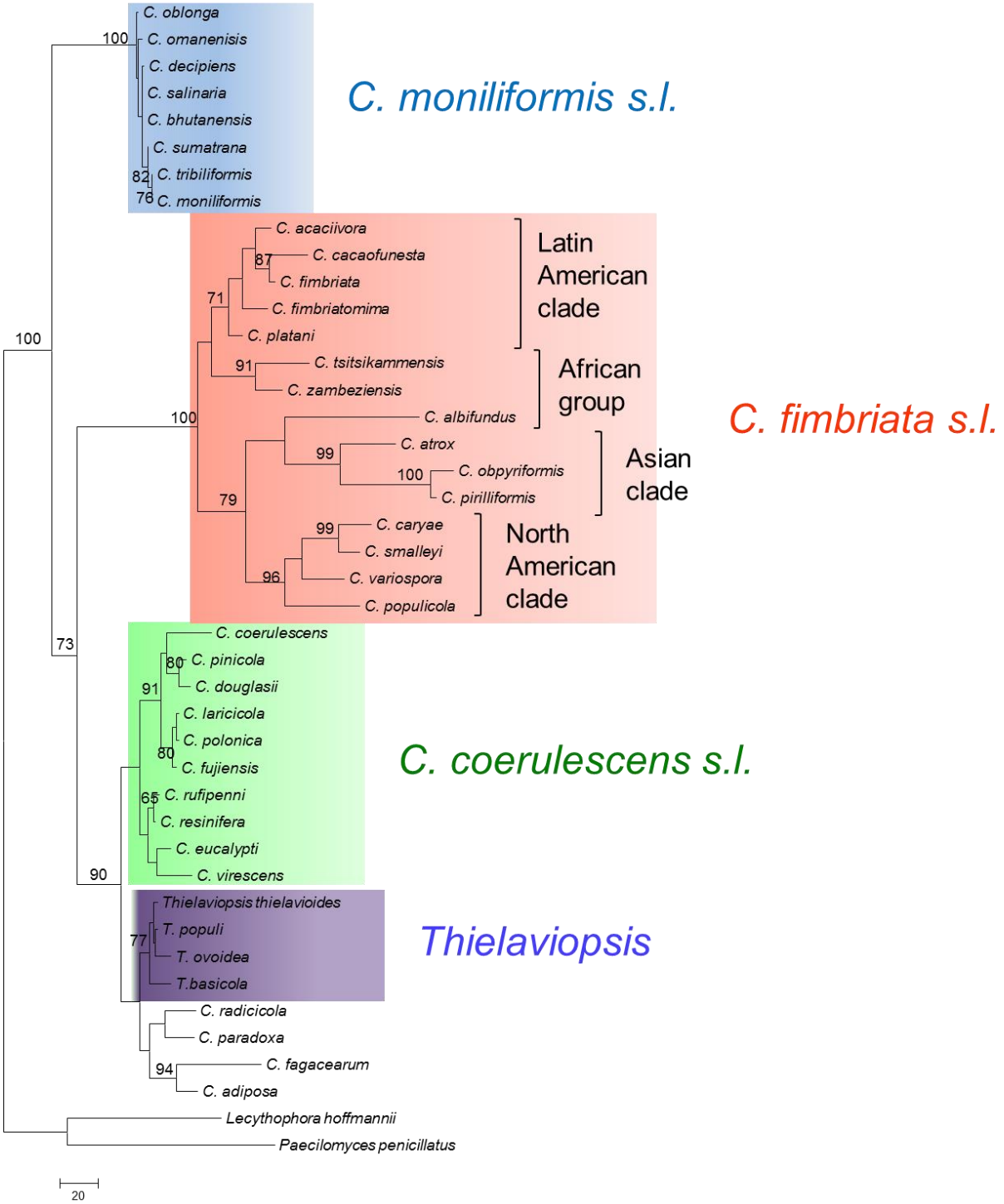


Fig. 1. Maximum parsimony tree of a representation of the *Ceratocystis* species complexes, based on ITS sequence data. Bootstrap values are indicated above the branches. This indicates the separation of species into the main species complexes. The *C. moniliformis*, *C. fimbriata* and *C. coerulescens* complexes are well supported and a potential fourth complex consists of *Thielaviopsis* species. The *C. fimbriata s.l.* species complex can further be subdivided in four groups and these are also indicated in the figure. Four additional species

group outside all complexes. All sequences were obtained from NCBI GenBank database and the MP tree was personally constructed for the purpose of this literature review.

All species in the *C. coerulescens* complex result in staining of sapwood but some are saprophytic and others pathogenic (Wingfield et al., 1997). Species in this complex occur both on gymnosperms and angiosperms. Species found on angiosperms (such as *Eucalyptus* spp., *Quercus* spp. and *Acer saccharum*) include *C. erinaceus* (Bohar, 1996), *C. virescens* (Harrington et al., 1996) and *C. eucalypti* (Kile et al., 1996). However, the majority of species are found on gymnosperms, specifically conifers such as spruce (*Picea* spp.), larch (*Larix* spp.), pine (*Pinus* spp.) and fir (*Pseudotsuga menziesii*) (Harrington and Wingfield, 1998; Wingfield et al., 1997). Three pathogenic species in this complex, *C. laricicola, C. fujiensis* and *C. polonica,* have specific associations with bark beetles and can result in tree death following beetle infestations of the host (Harrington and Wingfield, 1998; Marin et al., 2005; Witthuhn et al., 2000). Additional species that are also pathogenic to their hosts include *C. virescens, C. eucalypti* and *C. rufipenni* (Kile, 1993; Kile et al., 1996; Wingfield et al., 1997).

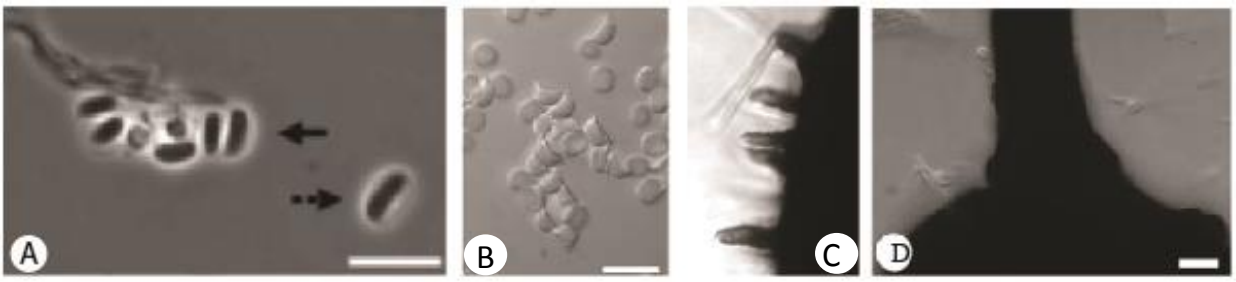


Fig. 3. Distinct micro morphological structures, unique to each *Ceratocystis* species complex. A) oblong-cylindrical shaped ascospores unique to *C. coerulescens* species (Marin et al., 2005), B) characteristic hat-shaped ascospores, as observed in the majority of *C. fimbriata* and *C. moniliformis* species, C) conical spines covering the ascomatal base, unique to the *C. moniliformis* complex (Kamgan et al., 2012) and D) disc-shaped base of the ascomatal neck, also observed only in *C. moniiformis* species (Van Wyk et al, 2011).

Ceratocystis moniliformis was first considered to represent a species complex in 2002 when a new species, *C. moniliformopsis*, was identified and classified as closely related to *C. moniliformis* (Van Wyk et al., 2004; Yuan and Mohammed, 2002). Since then, 13 additional species have been identified in this complex based on morphological and phylogenetic analyses (Table 1). All species of the *C. moniliformis* complex can be recognised by two characteristic traits of their ascomata; the ascomatal bases are covered with conical spines (Fig. 3C) (Upadhyay, 1981) and the base of the ascomatal necks end in a disc-shaped structure that easily breaks off from the base of the ascomata (Fig. 3D) (Hunt, 1956; Kamgan et al., 2012; Van Wyk et al., 2011b).

The majority of species in the *C. moniliformis* complex are non-pathogenic and mainly saprophytes. However, their growth in woody tissue often results in black, grey or blue staining of wood that reduces the value of the timber (Van Wyk et al., 2006; Yuan and

Mohammed, 2002). Species in this complex have been reported from a broad range of hosts, including *Eucalyptus* spp, *Acacia* spp, pine and mango (Al-Subhi et al., 2006; Nkuekam et al., 2013; Tarigan et al., 2010). Of the 16 currently recognised species in the complex, only *C. omanensis* on mangoes, and *C. chinaeucensis* on *Eucalyptus* spp., have been associated with mild symptoms of pathogenicity on the host trees (Al-Subhi et al., 2006; Al Adawi et al., 2006; Chen et al., 2013). With regard to insect association, *Ceratocystis bhutanensis* is the only *Ceratocystis* species that is associated with conifer-infesting bark beetles (Van Wyk et al., 2004).

Ceratocystis fimbriata s.l. represents the largest species complex in the genus and includes 32 described species (Fig. 2). Since the identification of species, additional to *C. fimbriata s.s.* in this complex, the term *sensu stricto* was reserved for isolates from sweet potato classified along with other species in the *C. fimbriata s.l.* species complex (Engelbrecht and Harrington, 2005). Species in the *C. fimbriata s.l.* complex can be recognised by a combination of two morphological traits. These include the absence of spines on their ascomatal bases, unlike species in the *C. moniliformis* complex, and ascospores with hat-shaped sheaths, unlike species in the *C. coerulescens* complex (Fig. 3B) (Engelbrecht and Harrington, 2005). The complex has been further subdivided into four groups, based primarily on phylogenies generated from the ITS, βT 1 and EF 1-α gene regions, but also on geographical distribution (Harrington, 2000; Johnson et al., 2005; Mbenoun et al., 2013). The groups consist of a Latin-American/South American clade, an African geographical group, a North-American and an Asian/Indo-pacific clade (Fig. 1). The naming of the groups was based on the countries and hosts where the initial species occurred but as additional species have been described, the names of some clades do not accurately reflect the geographic distribution of the species. The African group, on the other hand, is not currently considered a monophyletic clade but corresponds to the geographic occurrence of the species.

Species in the Latin-American clade are located in regions throughout Central and South America, the Caribbean and some in North America (Johnson et al., 2005) with a total of 15 species ascribed to this clade (Table 1). However, species in this clade do not occur only in Latin American countries but can be found on hosts native to Latin America. For example *C. fimbriata s.s.* occurs on sweet potato in countries in America as well as in Oceania in countries such as Papua New Guinea (Harrington, 2000). The African group consists of five species from South Africa and Tanzania (Mbenoun et al., 2013) and the Asian clade includes eight species from Japan, Africa, Australia and Indonesia (Table 1). Few species have been described in the North American clade and none of these have been reported outside North American countries. The clade consists of four host-specific species *C. harringtonii, C. variospora, C. caryae* and *C. smalleyi* (Johnson et al., 2005).

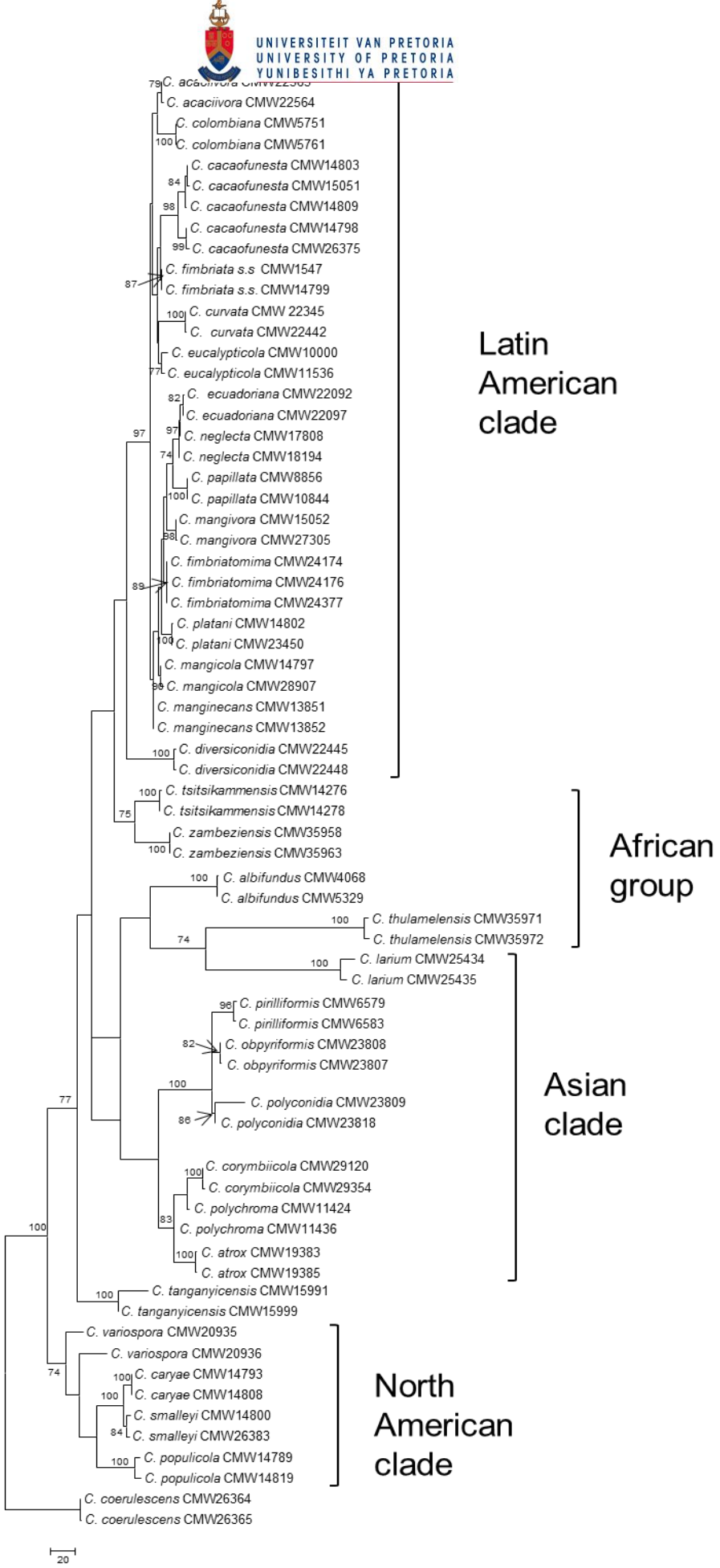


Fig. 2. One of the most parsimonious trees representing all 32 species in the *C. fimbriata s.l.* species complex. The species complex has been subdivided into four main groups, which is indicated in the tree. *Ceratocystis coerulescens* was selected as outgroup. The bootstrap

support values for the branches are indicated above each branch. Sequence data were obtained from the NCBI GenBank database and some were generated in Chapter 2. The MP tree was personally constructed for the purpose of this literature review.

The majority of species in the *C. fimbriata s. l.* species complex are pathogenic to a wide range of fruit and forest trees. Symptoms of infection can range from root rot to vascular staining, cankers and wilting that can result in significant loss of agricultural crops or tree plantations (Al Adawi et al., 2013; Engelbrecht and Harrington, 2005; Kajitani and Masuya, 2011; Van Wyk et al., 2007a; Van Wyk et al., 2009b). Two examples of highly pathogenic *C. fimbriata s.l.* species that have had a significant economic impact in their respective habitats include *C. platani* and *C. manginecans* (Ocasio-Morales et al., 2007; Van Wyk et al., 2007a). In USA and certain European countries infection of oriental plane (*Platanus orientalis*) and American plane trees (*P. occidentalis*) by *C. platani,* has resulted in tree death of fully grown plane trees within the first few years of infection (Engelbrecht and Harrington, 2005; Ocasio-Morales et al., 2007). The pathogen has rapidly spread throughout numerous European countries where it is an exotic species, causing extensive disease (Chapin and Arcangioli, 2007; Hadden and Hugh-Jones, 2011; Ocasio-Morales et al., 2007). In Oman and Pakistan, *C. manginecans* causes severe mango wilt that can result in tree death within six months after infection (Al Adawi et al., 2006; Al Adawi et al., 2003). Infection by the pathogen has affected up to 60% of the mango crop in some regions and has resulted in severe yield and economic losses in both countries (Al Adawi et al., 2013; Al Adawi et al., 2006; FAOSTAT, 2011).

Table 1 Summary of the classification methods applied for species delineation in the *Ceratocystis* species complexes[a]

	Morphological criteria				Biological criteria	Phylogenetic criteria				
Species	Macro morphology [b]	Micro morphology [b]	Isozyme/ allozyme [d]	RFLP+ SSRs [e]	Intersterility [c]	MAT-2 [f]	ITS [f]	BT [f]	EF [f]	Reference
***C. coerulescens* s.l. complex**										
1 *C. coerulescens*	0	0	-	X	-	0	X	-	-	Bakshi, 1950
2 *C. douglasii*	X	X	X	X	-	X	X	-	-	Wingfield et al., 1997
3 *C. erinaceus*	Partially	X	-	-	-	-	-	-	-	Bohar, 1996
4 *C. eucalypti*	X	X	-	-	-	X	X	-	-	Kile et al., 1996
5 *C. fujiensis*	X	0	-	-	X	X	X	X	-	Marin et al., 2005
6 *C. laricicola*	X	0	Partially	0	X	X	X	X	-	Harrington and Wingfield, 1998
7 *C. norvegica*	X	X	-	-	-	-	X	-	-	Reid et al., 2010
8 *C. pinicola*	Partially	Partially	-	X	-	X	X	-	-	Harrington and Wingfield, 1998
9 *C. polonica*	X	0	Partially	0	X	X	X	X	-	Harrington and Wingfield, 1998
10 *C. resinifera*	Partially	0	-	X	X	X	X	-	-	Harrington and Wingfield, 1998
11 *C. rufipenni*	Partially	0	X	X	X	X	X	-	-	Harrington et al., 1996
12 *C. virescens*	0	Partially	X	X	-	X	X	-	-	Harrington et al., 1996
***C. moniliformis* s.l. complex**										
1 *C. moniliformis*	X	-	-	-	-	-	-	-	-	Moreau, 1952
2 *C. moniliformopsis*	X	X	-	-	-	-	Partially	X	X	Yuan and Mohammed, 2002
3 *C. bhutanensis*	X	Partially	-	-	-	-	Partially	X	X	Van Wyk et al., 2004
4 *C. omanensis*	X	Partially	-	-	-	-	X	X	X	Al-Subhi et al., 2006
5 *C. tribiliformis*	X	Partially	-	-	-	-	Partially	Partially	Partially	Van Wyk et al., 2006
6 *C. savannae*	X	Partially	-	-	-	-	X	X	X	Kamgan et al., 2008
7 *C. oblonga*	X	Partially	-	-	-	-	X	X	X	Heath et al. 2009
8 *C. inquinans*	0	0	-	-	-	-	X	X	X	Tarigan et al., 2009
9 *C. microbasis*	0	Partially	-	-	-	-	X	X	X	Tarigan et al., 2009
10 *C. sumatrana*	0	0	-	-	-	-	X	X	X	Tarigan et al., 2009
11 *C. sublaevis*	0	Partially	-	-	-	-	0	0	X	Van Wyk et al., 2011b

Table 1 (continue)

		Morphological criteria				Biological criteria	Phylogenetic criteria				
	Species	Macro morphology [b]	Micro morphology [b]	Isozyme/ allozyme [d]	RFLP+ SSRs [e]	Intersterility [c]	MAT-2 [f]	ITS [f]	BT [f]	EF [f]	Reference
12	*C. tyalla*	0	Partially	-	-	-	-	Partially	Partially	X	Kamgan et al., 2011
13	*C. salinaria*	0	0	-	-	-	-	0	X	Partially	Kamgan et al., 2012
14	*C. decipiens*	0	0	-	-	-	-	0	X	X	Kamgan et al., 2012
15	*C. chinaeucensis*	0	0	-	-	-	-	X	X	X	Chen et al., 2013
16	*C. cryptoformis*	0	0	-	-	-	-	0	X	0	Mbenoun et al., 2013
	***C. fimbriata* s.l. complex**										
	Latin-American clade										
1	*C. acaciivora*	0	Partially	-	-	-	-	X	0	Partially	Tarigan et al., 2009
2	*C. cacaofunesta*	0	Partially	-	X	X	-	X	-	-	Engelbrecht and Harrington, 2005
3	*C. colombiana*	X	X	-	-	-	-	X	Partially	Partially	Van Wyk et al., 2010
4	*C. curvata*	Partially	Partially	-	-	-	-	X	Partially	Partially	Van Wyk et al., 2011b
5	*C. diversiconidia*	Partially	Partially	-	-	-	-	X	Partially	Partially	Van Wyk et al., 2011b
6	*C. ecuadoriana*	Partially	X	-	-	-	-	X	Partially	Partially	Van Wyk et al., 2011b
7	*C. eucalypticola*	0	X	-	-	-	-	X	0	0	Van Wyk et al., 2012
8	*C. fimbriata* s.s.	X	Partially	-	X	X	-	X	-	-	Engelbrecht and Harrington, 2005
9	*C. fimbriatomima*	0	Partially	-	-	-	-	X	Partially	X	Van Wyk et al., 2009b
10	*C. manginecans*	0	X	-	-	-	-	X	0	Partially	Van Wyk et al., 2007a
11	*C. mangicola*	0	0	-	-	-	-	X	Partially	-	Van Wyk et al., 2011a
12	*C. mangivora*	0	0	-	-	-	-	X	Partially	0	Van Wyk et al., 2011a
13	*C. neglecta*	0	X	-	-	-	-	X	-	-	Van Wyk et al., 2008
14	*C. papillata*	X	X	-	-	-	-	X	Partially	Partially	Van Wyk et al., 2010
15	*C. platani*	X	Partially	-	X	X	-	X	-	-	Engelbrecht and Harrington, 2005
	Asian clade										
16	*C. atrox*	X	X	-	-	-	-	X	X	X	Van Wyk et al., 2007b
17	*C. corymbiicola*	0	X	-	-	-	-	Partially	Partially	Partially	Kamgan et al., 2011
18	*C. ficicola*	Partially	X	-	-	-	-	X	-	-	Kajitani and Masuya, 2011
19	*C. larium*	0	X	-	-	-	-	X	X	X	Van Wyk et al., 2009a
20	*C. obpyriformis*	Partially	X	-	-	-	-	X	Partially	Partially	Heath et al. 2009
21	*C. pirilliformis*	0	X	-	-	-	-	X	-	-	Barnes et al., 2009
22	*C. polychroma*	X	X	-	-	-	-	X	X	X	Van Wyk et al., 2004
23	*C. polyconidia*	0	0	-	-	-	-	X	Partially	Partially	Heath et al. 2009
	North-American clade										
24	*C. caryae*	X	Partially	X	-	X	-	X	X	X	Johnson et al., 2005
25	*C. harringtonii*	0	X	X	-	X	-	X	-	-	Johnson et al., 2005
26	*C. smalleyi*	X	X	X	-	X	-	X	X	X	Johnson et al., 2005
27	*C. variospora*	0	Partially	X	-	X	-	X	-	-	Johnson et al., 2005
	African clade										
28	*C. albifundus*	X	X	-	X	-	-	X	-	-	Wingfield et al., 1996
29	*C. tanganyicensis*	Partially	X	-	-	-	-	X	Partially	Partially	Heath et al. 2009
30	*C. tsitsikammensis*	Partially	X	-	-	-	-	X	Partially	Partially	Kamgan et al., 2008
31	*C. thulamelensis*	Partially	0	-	-	-	-	X	X	X	Mbenoun et al., 2013
32	*C. zambeziensis*	0	0	-	-	-	-	X	X	X	Mbenoun et al., 2013

a An X indicates that the method was effective in confirming the species as distinct, if only partially informative this was indicated accordingly and 0 indicates the method was not informative. Methods not considered for species delineation were indicated with a dash.

b Macro morphology refers to culture appearance and growth characteristics whereas micromorphology considers dimensions of the reproductive structures including the length of the ascomatal necks, size and shape of ascomatal bases and the variation in conidial shapes.

c Intersterility tests confirmed distinct biological species during mating studies.

d Isozyme/alozyme investigations compared the electrophoretic mobility of various enzymes to identify distinct species.

e Restriction fragment length polymorphism (RFLP) and microsatellites (SSRs) phenotypically investigates DNA sequence variation by digesting DNA with enzymes (RFLP) or screening DNA samples for length polymorphisms (SSRs). These methods have not been used to delineate species but can indicate potential species boundaries.

f The four gene regions were applied in phylogenetic analysis to distinguish between species.

3. Methods used to identify *Ceratocystis* species

Species delineation in *Ceratocystis* was exclusively based on morphological variation during the 1900's (Hunt, 1956). Since the late 1990's, alternative identification methods where

considered, such as protein biochemical characteristics (Harrington et al., 1996), intersterility (Harrington and McNew, 1998) and DNA sequence data (Witthuhn et al., 1999). The following section of this review considers the impact that these methods have had on the classification of the 61 currently defined *Ceratocystis* species.

3.1. Morphology

The approach for morphological species recognition is aimed at identifying a characteristic that differs between two organisms in its physical appearance (Ruse, 1969). Comparisons can be made based on both the macro and micro morphological characters between fungi as well as differences in biochemical properties (Bessey, 1950; Harrington and Rizzo, 1999). Macro morphological characteristics considered in *Ceratocystis* species include the optimal growth temperature and culture colour. The optimal growth temperature for *Ceratocystis* species range from 20 °C to 25 °C (Kajitani and Masuya, 2011; Kamgan et al., 2012; Kile et al., 1996) but some species have more extreme temperature optima for growth. *Ceratocystis albifundus*, for example, is native to Africa and grows optimally at 30 °C (Wingfield et al., 1996) whereas some species in the *C. coerulescens* complex, found primarily in North America and Europe, grow optimally below 20 °C (Harrington et al., 1996). The species *C. bhutanensis* occurs at high altitude, in the Himalayas, and can survive temperatures as low as 4°C (Van Wyk et al., 2004). Colony colour can be used to help identify the general complex of the species. A dark grey to olive-green colour is unique in *C. coerulescens* species (Fig. 4A) (Bakshi, 1951; Harrington et al., 1996), an olive-green colony colour is common in *C. fimbriata s.l.* species (Fig. 4B) and white to dark brown mycelia is mostly found in *C. moniliformis* species (Fig. 4C) (Kamgan et al., 2012; Van Wyk et al., 2011b).

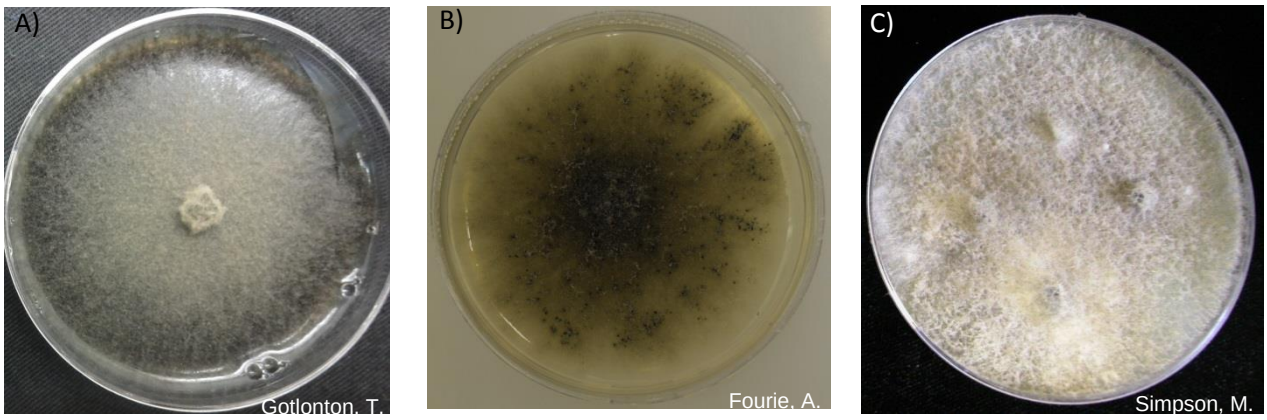


Fig. 4. Variation in culture morphology with characteristics that are unique to each of the three *Ceratocystis* species complexes. A) *C. coerulescens s.l.* species have a grey to green colour B) *C. fimbriata s.l.* species have an olive green culture colour and C) *C. moniliformis s.l.* species have a white to brown colour.

Micromorphological characteristics considered for *Ceratocystis* species descriptions include both the sexual fruiting structures (ascomata and ascospores) (Fig. 5A), and asexual reproductive structures that include conidiophores (Fig. 5C), that produce conidia (Fig. 5D), and chlamydospores (Fig. 5E). Morphological variation in the ascomata include the shape, dimensions and colour of the ascomatal bases, the length of the ascomatal necks, the presence of either convergent or divergent ostiolar hyphae at the tips of the necks (Fig. 5B) and the shape of the ascospores produced by the ascomata (Figs. 3A and 3B) (Upadhyay, 1981). Variation in the shape of the conidia, produced by the conidiophore, can be characteristic to a species. Shapes can range for cylindrical to bacilliform to oblong or barrel-shaped conidia (Fig. 5D) (Engelbrecht and Harrington, 2005; Kamgan et al., 2012). The absence of chlamydospore production can also be used to distinguish between species (Van Wyk et al., 2007b).

Specific examples where morphological traits have provided the primary support for distinct species include *C. eucalypti* producing sheathed, elongated ascospores (Kile et al., 1996), *C. albifundus* having light coloured ascomatal bases (Wingfield et al., 1996) and *C. ficicola* that have significantly larger ascomata than any other *Ceratocystis* species (Kajitani and Masuya, 2011) (Fig. 6). However, the majority of species in *Ceratocystis* do not possess such distinct morphological traits, due to strong evolutionary selection in the overall structure of the ascomata (Malloch and Blackwell, 1993; Upadhyay, 1981), resulting in limited variable characteristics. Another limitation in using morphology as a criterion is that the number of morphological differences required to distinguish two species is subjective and reliant on the opinion of the researcher involved (Ruse, 1969).

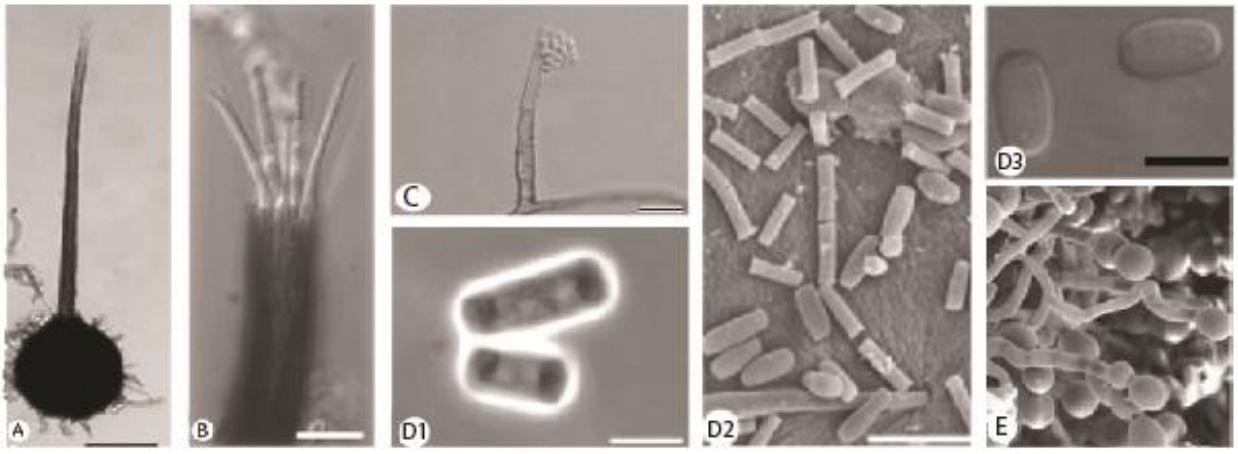


Fig. 5. An illustration of typical micro morphological characteristics of *Ceratocystis* species: A) ascomata with a long ascomatal neck (Van Wyk et al., 2010), B) divergent ostiolar hyphae at the tip of the ascomatal neck (Kamgan et al., 2012), C) conidiophore structures (Van Wyk et al, 2004) that produce D) conidia with varying shapes such as cylindrical (D1) (Marin et al., 2005), oblong, bacilliform (D2) (Kamgan et al., 2012) and barrel-shaped (D3) (Van Wyk et al, 2011). E) Chlamydospore survival structures (Kamgan et al., 2012).

3.2. Protein biochemical analysis

In the 1970's, analyses of the biochemical properties of proteins were applied to *Ceratocystis* species for taxonomic purposes (Sprecher and Hanssen, 1983; Stipes, 1970). One of the first studies to investigate the utility of biochemical variation compared the banding patterns of total soluble mycelial proteins of three *Ceratocystis* species (Stipes, 1970). Intraspecific conservation of proteins was evident in the study as ±10 proteins were consistently identified within a species. The different protein phenotypes could confirm the species *C. coerulescens*, *C. fagacearum* and *C. variospora* to be distinct.

Volatile aromatic metabolites were also considered for species delimitation in five *Ceratocystis* species (Sprecher and Hanssen, 1983). Methyl heptenyl compounds were present in all *C. coerulescens s.l.* species investigated but higher levels were found in *C. virescens* than in *C. coerulescens s.s.* In contrast, the compound 2-phenylacatate was detected in only a few *C. coerulescens s.l.* isolates. These data could be used to distinguish the *C. coerulescens* complex from other species complexes and potentially distinguished between some species within the complex. Variation was also found between *C. fimbriata s.l.* species based on their ability to produce monoterpenes, for example *C. variospora* produced higher levels of monoterpenes than the other species (Sprecher and Hanssen, 1983). The study could not identify volatiles conserved within a species but some of the strains considered as a single species in the 1980's were likely distinct cryptic species. These compounds have not been investigated subsequently and a re-evaluation in currently defined species would be informative. However, the analyses of aromatic metabolites do not appear to have great value in the definition of closely related species of *Ceratocystis* as the amount of variation that can be observed is still limiting. Variation has also been found primarily between distantly related species and not cryptic species.

Since the 1990's, biochemical comparisons between species were predominantly based on enzyme variation with which 12 species in *Ceratocystis* have been distinguished (Fig. 6) (Harrington et al., 1996). Variation in the biochemical properties of isozymes can be identified based on the electrophoretic ability of an enzyme (Harrington et al., 1996). In the *C. coerulescens* complex 10 isozymes were considered (Harrington et al., 1996) and the data confirmed seven distinct species in the complex, corresponding to the variation found in their morphology. *Ceratocystis polonica* and *C. laricicola* are morphologically indistinguishable but variation in a single isozyme was the first indication that they were distinct species. This was later confirmed by DNA sequence data (Marin et al., 2005). In the *C. fimbriata s.l.* species complex, variation in 12 allozymes resulted in distinct electrophoretic phenotypes that supported the classification of the species *C. variospora, C. harringtonii, C. caryae* and *C. smalleyi* that reside in the North American clade (Johnson et al., 2005).

3.3. Intersterility tests

The basis of a biological species is that is has to consist of “interbreeding populations that reproduce naturally and produce fertile offspring” (Voigt and Kirk, 2011). Biological species recognition is based on mating experiments where two individuals, representative of the two populations, are brought in close proximity, allowed to mate and the intersterility between them evaluated (Harrington and Rizzo, 1999). This approach has been applied for species recognition in *Ceratocystis* since the 1990’s (Harrington and McNew, 1998). The majority of *Ceratocystis* species investigated to date are homothallic and thus mainly reproduce sexually through self-fertilization which can complicate mating experiments (Engelbrecht and Harrington, 2005; Ferreira et al., 2010; Harrington and McNew, 1997). However, the mechanism of unidirectional mating-type switching in *Ceratocystis* (Harrington and McNew, 1997) results in the production of self-sterile offspring that can be used for mating studies in this genus. This mechanism has been studied extensively only for some species in the *C. fimbriata s.l.* complex.

All self-fertile *Ceratocystis* species (termed a MAT-2 self-fertile strain) contain two mating type loci (MAT-1 and MAT-2) and can produce both self-fertile and self-sterile offspring (Harrington and McNew, 1997, 1998). A self-sterile isolate (MAT-1 self-sterile strain) is the result of the mating type switching mechanism, where the MAT-2 gene is deleted from the isolate’s genome. MAT-2 self-sterile isolates can be induced under laboratory conditions and contain both mating type loci but are self-sterile. These two self-sterile isolates of opposite mating types (MAT-1 and MAT-2) can be used for mating experiments (Engelbrecht and Harrington, 2005). Results from a mating study are regarded as positive when the two isolates on a single plate can mate and produce viable ascospores (Engelbrecht and Harrington, 2005). If sexual structures are malformed or absent, the isolates are considered intersterile.

Mating studies have supported the delineation of 13 species in *Ceratocystis* (Fig. 6). An example is the seven species that could be distinguished within the *C. coerulescens* species complex (Harrington and McNew, 1998). Intersterility or very poor levels of interfertility between isolates supported the monophyly of *C. pinicola, C. coerulescens, C. resinifera, C. rufipenni, C. laricicola, C. polonica* and *C. douglasii* as suggested by morphology, isozymes and ITS sequence data (Marin et al., 2005). Mating studies have also provided the primary support for species boundaries of three host specific species in the *C. fimbriata s.l.* complex, namely *C. fimbriata s.s.* on sweet potato, *C. platani* on sycamore trees and *C. cacaofunesta* on cacao trees (Engelbrecht and Harrington, 2005). Complications of mating studies include the challenges to cultivate certain species and the fact that species can show partial or complete interfertility in a controlled laboratory environment, due to pre-zygotic mating recognition, that would not have occurred in nature (Taylor et al., 2000).

3.4. Molecular data

The application of molecular markers and DNA sequencing has revolutionised fungal classification methods by exponentially increasing the number of variable characters that can be compared between species (Balajee et al., 2009; Lutzoni et al., 2004). DNA sequence variation has been analysed both at a phenotypic level as well as by direct sequencing of a gene region. Phenotypic molecular markers applied for species delineation include DNA fingerprinting techniques such as restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP) (Coetzee et al., 2000; Harrington and Rizzo, 1999; Witthuhn et al., 1999). Delineation based on DNA sequences and phylogenetic analyses have enabled the identification of numerous cryptic species that cannot be distinguished by either morphological or biological species recognition criteria (Cai et al., 2011; Krüger et al., 2012; Mbenoun et al., 2013; Taylor et al., 2000; Wingfield et al., 2013).

3.4.1. Phenotypic molecular analysis

RFLP analyses do not require knowledge of a specific DNA sequence but the banding pattern created by restriction enzyme digestion of the DNA fragment can indicate, at a phenotypic level, the genetic variation present between species. One of the first RFLP analyses performed on *Ceratocystis* species was based on a 1.6 kb fragment of rDNA digested with 18 restriction enzymes (Witthuhn et al., 1999). Variation in RFLP patterns were sufficient to distinguish between all 11 species investigated (Table 1; Fig. 6).

In the *C. coerulescens* complex, RFLP analysis of rDNA and mitochondrial DNA supported the suggestion that the species *C. major* and *C. adiposa* are the same due to their identical restriction profiles (Hausner et al., 1993). These data also corrected the classification of the species *C. coerulescens, C. paradoxa* and *C. radicicola* in *Ceratocystis*, previously classified as *Endoconidiophora* based on morphological data (Hausner et al., 1993). Subsequent RFLP analysis of the β-tubulin gene region in seven *C. coerulescens* species confirmed the utility of this phenotypic analysis method for species delineation in the complex (Loppnau and Breuil, 2003). In the *C. fimbriata s.l.* species complex, RFLP analysis of the ITS region could separate isolates on coffee trees in Colombia into two clades (Marin et al., 2005). These were later confirmed to be distinct species, *C. colombiana* and *C. papillata,* based on DNA sequence data (Van Wyk et al., 2010).

3.4.2. DNA sequencing for fungal species identification

The primary focus in current fungal classification studies is to reliably distinguish between cryptic species based on DNA sequence data (Bickford et al., 2007; Blackwell et al., 2006; Shivas and Cai, 2012). This shift in the methods applied for fungal species identification has

become evident in numerous studies in the last 10 years (Begerow et al., 2010; Rintoul et al., 2012; Roe et al., 2010; Seifert, 2009). One of the collaborative projects that clearly illustrates the importance of DNA sequence data is the "Fungal Barcode of Life" project (FBoL, http://www.fungalbarcoding.org), aimed at identifying a universal DNA barcode that could be applied to the identification of all fungal species (Seifert, 2009). Of the six gene regions considered with the greatest potential as a barcode marker the ITS region best complied to all the criteria (Schoch et al., 2012; Schoch et al., 2011).

The growing application of sequence data for fungal taxonomy, specifically of the ITS region, is evident in the submissions of ITS sequences to GenBank, at the National Center for Biotechnology Information (NCBI) (Benson et al., 2012). A total of 630 900 fungal ITS sequences are available on this public sequence database and 83% of these were submitted in the last five years. A recent collaborative effort, UNITE, was established to re-evaluate the bulk of these sequences and to create a reliable fungal ITS sequence database (Kõljalg et al., 2013). A total of 352 050 ITS sequences were confirmed as reliable, describing 52 481 potential species. This is more than double the number of species listed with ITS sequences in the latest edition of the Dictionary of the Fungi (Kirk et al., 2008), illustrating the rapid increase in sequence data generation and associated species descriptions.

For species descriptions based on phylogeny, a species can be considered phylogenetically distinct only if it contains "a derived character that arose monophyletically and is shared by the members of the species" (Voigt and Kirk, 2011). Phylogenetic trees are constructed by comparing DNA sequence variation between isolates and calculating the genetic distance between them (Avise and Wollenberg, 1997; Holder and Lewis, 2003; Yang and Rannala, 2012). Since 1996, the taxonomy of closely related *Ceratocystis* species was based on the ITS gene region and one of the first cryptic species identified was *C. albifundus* (Wingfield et al., 1996). Subsequent phylogenetic analyses, based only on the ITS region identified 17 novel cryptic species in the *Ceratocystis* genus in the last 17 years (Table 1; Fig. 6). Examples are the three species, *C. virescens, C. douglasii* and *C. rufipenni,* that were initially identified as *C. coerulescens* and supported to be distinct species based on ITS data (Witthuhn et al., 1998). Within the *C. fimbriata s.l.* species complex, *C. pirilliformis* was also primarily distinguished from *C. fimbriata s.s.* based on ITS sequence data (Barnes et al., 2003). In the North American clade of this species complex, this gene region confirmed the variation observed in the morphology and allozymes of *C. caryae, C. smalleyi, C. harringtonii* and *C. variospora* (Johnson et al., 2005).

Species delineation based on a single gene genealogy can be biased, since the threshold for the percentage of sequence similarity (97% or 99% for example) required to classify isolates as a single species, will vary between researchers (Taylor et al., 2000). This subjectivity prompted the development and application of the genealogical concordance

phylogenetic species recognition (GCPSR) concept where distinct species have to be confirmed by the concordance of multiple gene genealogies (Taylor et al., 2000). Due to its informativeness and reliability, the GCPSR is currently the most widely used approach for species delineation in the fungal Kingdom and it has led to the identification of numerous cryptic species (Amalfi et al., 2012; Blackwell, 2010; Duong et al., 2012; Hibbett and Taylor, 2013; Lutzoni et al., 2004; Marin et al., 2005; Rintoul et al., 2012)

Since the incorporation of ITS sequence data for *Ceratocystis* species delineation in the late 1990's, and in concordance with the GCPSR, it has become mandatory to include at least two gene region for species identification. In numerous studies, the additional gene regions consisted of either MAT-2, βT 1 or EF 1-α in addition to ITS data (Marin et al., 2005; Van Wyk et al., 2007a; Van Wyk et al., 2004; Van Wyk et al., 2009a). An example of the efficacy of this approach was the combination of ITS, βT 1 and MAT-2 that could distinguish between *C. polonica* and *C. laricicola* (Marin et al., 2005) where morphological data, isozymes and initial ITS sequence data did not provide any distinction (Harrington et al., 1996). In the past 10 years, the application of the GCPSR approach for the identification of species enabled the delineation of 36 of the 61 currently described *Ceratocystis* species (Table 1; Fig. 6).

The three most widely used gene regions in *Ceratocystis* classification have shown limitations in the past three years (Van Wyk et al., 2012; Van Wyk et al., 2011b). Even though the combined phylogenetic trees of ITS, βT 1 and EF 1-α have identified many cryptic species, the individual βT 1 and EF 1-α gene genealogies contributed very little to the phylogenetic signal for some species. An example is found in the species *C. mangicola* and *C. mangivora*, isolated from mango trees in Brazil (Van Wyk et al., 2011a), where ITS was the only region that contained significant variation between the two species. The ITS gene region also provided the primary support for the species status of *C. colombiana, C. corymbiicola, C. curvata, C. ecuadoriana* and *C. eucalypticola* (Kamgan et al., 2012; Van Wyk et al., 2012; Van Wyk et al., 2010; Van Wyk et al., 2011b) in recent studies.

The application of the ITS region in the GCPSR approach for *Ceratocystis* species classification has introduced further complications in the robust identification of species in this genus. Firstly, the region contains multiple indels in some *Ceratocystis* species causing significant length variations. These length variations complicate accurate alignments of the sequences for phylogenetic analyses (Schmitt et al., 2009). Additionally, intragenomic variation of the ITS region has been reported in various fungal species, including *Fusarium, Laetiporus* and recently *Ceratocystis* (Aanen et al., 2001; Al Adawi et al., 2013; Lindner et al., 2013; O'Donnell and Cigelnik, 1997). Such variation in sequences could result in incorrect and unreliable classification of the species.

In the *C. fimbriata s.l.* species complex, two different copies or "haplotypes" of the ITS sequence repeat, differing by seven nucleotides, have been identified in some *C.*

manginecans isolates (Al Adawi et al., 2013; Naidoo et al., 2013). Depending on the dominant ITS haplotype, a single isolate was identified as either *C. manginecans* or *C. acaciivora* (Al Adawi et al., 2013). The primary concern here is that βT 1 and EF 1-α provided no support for the differentiation between these two species (Al Adawi et al., 2013). An in depth analysis of five species in the Latin-American clade of *C. fimbriata s.l.* identified between two to three different ITS haplotypes in an individual isolate (Harrington et al., 2014). If all these haplotypes from a single isolate were included in a species phylogeny, they would be distributed in at least 3 clades, representing different species. Intraspecific ITS variation was also observed among some of the isolates from a clonal population. This confirms that the ITS region is highly variable within isolates and within species. The fact that multiple ITS types are most likely also present in other *C. fimbriata s.l.* species, that have been defined primarily based on ITS sequences, raises questions relating to their validity.

The low phylogenetic resolution obtained from the βT1 and EF 1-α gene regions for some species in the *C. fimbriata s.l.* complex might indicate that the regions evolve more slowly than the ITS region (Rintoul et al., 2012). This would make the regions uninformative for delineation of closely related species and alternative regions would have to be considered. Alternatively, the species distinguished only by ITS might contain variable copies of ITS, suggesting new species, but could be almost identical in the rest of their genomes. Additional sequence data must thus be considered to determine how accurate the species classification, based on the ITS region, is in the *C. fimbriata s.l.* complex.

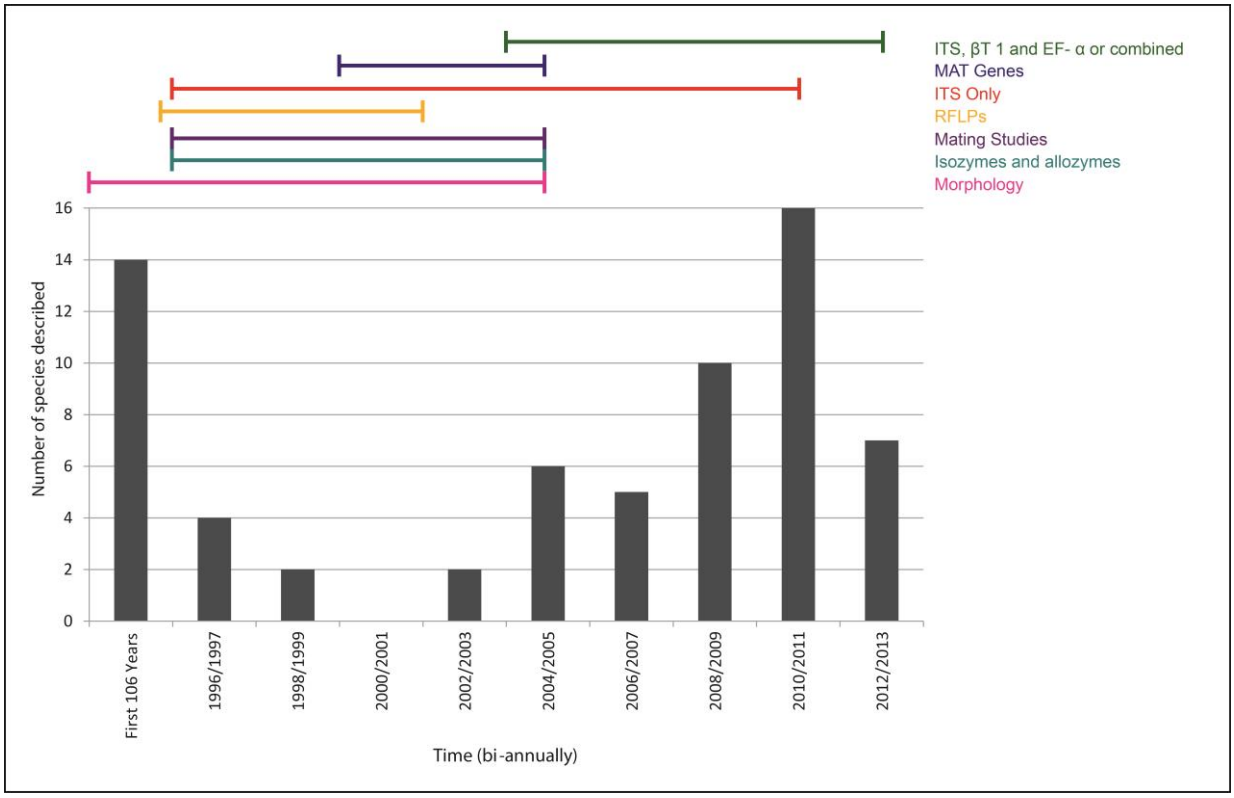


Fig. 6. An illustration of the most widely used methods applied for species delineation in the *Ceratocystis* genus from its first description in 1890. The x-axis indicates the time period during which the method was applied, starting with the first 106 years before DNA sequence data was used, followed by bi-annual increments. The y-axis indicates the number of species described during the different time periods. Morphology includes both macro and micro morphological traits; protein analysis refers to isozyme and allozyme variation; RFLPs refer to restriction fragment length polymorphisms. The ITS and MAT gene regions were used before a genealogical concordance approach was considered and are thus indicated separately from the combined gene analysis.

4. Alternative approaches for species delimitation

4.1. Alternative gene regions

The *C. fimbriata s.l.* species complex includes ±13 species that cannot reliably be differentiated with current morphological and phylogenetic approaches, besides the ITS phylogeny. This necessitates the addition of extra gene regions needed for a GCPSR approach to differentiate between the species. Alternative genes to consider are the protein coding genes found to be informative in previous fungal phylogenetic studies. Two such collaborative projects include FBoL and “Assembling the Fungal Tree of Life” (AFTOL, http://aftol.org) projects (Lutzoni et al., 2004; Schoch et al., 2012). The aim of the AFTOL project has been to reassess the higher level phylogenetic relationships between all genera in the fungal kingdom, primarily based on DNA sequence data. Three of the regions used in these projects, the ribosomal large subunit (LSU), ribosomal small subunit (SSU) and EF 1-α regions, have already been applied to *Ceratocystis* (Van Wyk et al., 2009b; Wingfield et al., 1994). Future studies could focus on utilizing additional regions such as the 1st and 2nd largest subunits of the RNA polymerase II (RPBI and RPBII), mitochondrial SSU, Mcm7 and ATP6 for phylogenetic application.

The rapid increase in publicly available genome sequence data (Kuramae et al., 2006) has enabled large scale genome-comparison studies as recently performed on 21 fungal genomes (Aguileta et al., 2008; Marthey et al., 2008). The study identified 246 single copy protein coding genes as potential phylogenetic markers, now available on the FUNYBASE public database (Marthey et al., 2008). The phylogenetic utility of the gene regions were evaluated and two of the regions, Mcm7 and Tsr1, showed potential as additional phylogenetic markers for Ascomycetes (Aguileta et al., 2008). The regions have been applied in various phylogenetic studies (including FBoL) to resolve both genus level and species level relationships (Raja et al., 2011; Schmitt et al., 2009). Two additional gene regions recently identified from FUNYBASE, MS204 and FG1093 (Walker et al., 2012b), were highly informative for delineation of closely related species and were suggested as potential phylogenetic markers for the Sordariomycete class (Walker et al., 2012a).

Alternative gene regions that might be considered as high resolution phylogenetic markers in *Ceratocystis* species are mitochondrial genes. The mitochondrial cytochrome c

oxidase I (CoxI) gene is the official barcode for species identification in the animal Kingdom and was thus considered as a marker for fungal classification (Schoch et al., 2012). However, it has been excluded from phylogenetic studies in fungi due to multiple introns, the presence of homologous copies and the low phylogenetic signal it provides in many genera (Chase and Fay, 2009; Gilmore et al., 2009; Seifert, 2009). Nonetheless, the CoxI region was more informative than ITS in *Penicillium* species delineation and contained limited introns (Seifert et al., 2007). Thus, it should not be disregarded until it has been evaluated for *Ceratocystis* species.

Even though the phylogenetic application of mitochondrial genes in the Fungal Kingdom is discouraged (Schoch et al., 2012; Seifert, 2009), recent studies have identified some regions with low intron density which increases their potential as phylogenetic markers. For example, a bioinformatic approach was used to identify the abundance of introns in the mitochondrial sequences of all ascomycota submitted on Genbank (Santamaria et al., 2009). The study confirmed the high density of introns throughout most mitochondrial gene regions and the only gene without introns was NADH dehydrogenase subunit 6 (ND6). Three additional regions, ND3 and the 3' end of ND4 and ND5, had very low levels of mobile introns. These four mitochondrial genes could serve as phylogenetic markers but their utility still needs to be confirmed as very few fungal phylogenetic studies have applied them to date.

An alternative group of gene regions that can be investigated for species delimitation in phytopathogenic fungi are pathogenic factors such as effector molecules. Effectors are involved in the secondary line of defence during plant-pathogen interactions by assisting in host infection processes and suppressing host defence responses (Chakrabarti et al., 2011; Stergiopoulos and De Wit, 2009). These molecules can be related to the host on which the fungus occurs (Hogenhout et al., 2009) and have the potential to show variation between related species. The SIX6 effector gene was compared in four *formae speciales* of *Fusarium oxysporum* and contained a significant amount of amino acid sequence differences to distinguish between the *formae speciales* (Chakrabarti et al., 2011). Such markers might not be selectively neutral and cannot be used to deduce the evolutionary history of the species (Rintoul et al., 2012), but they could contain sequence variation between species that might be useful for diagnostic marker development.

One of the pathogenicity factors investigated in multiple *Ceratocystis* species is the pathogen associated molecular pattern (PAMP) protein; cerato-platanin (CP) (Comparini et al., 2009). The CP protein is present in the mycelia and plays a role in fungal growth and development during host infection (Baccelli et al., 2012). This single copy protein was sequenced and aligned for the host-specific species *C. platani, C. variospora, C. harringtonii, C. cacaofunesta* and other *C. fimbriata s.l.* species from different hosts (Comparini et al., 2009). Multiple amino acid changes were identified between *C. platani, C. variospora* and *C.*

harringtonii with fewer differences between closely related *C. fimbriata s.l.* species. However, at a nucleotide level there might be a significant level of variation between closely related species. Further analysis in the larger species complex is required but it could serve as a diagnostic marker to distinguish between cryptic species.

4.2. Polymorphic markers

The GCPSR approach for species delineation is based on sequence data from multiple gene regions but the amount of information obtained from these markers remains an under representation of the overall variation present in an organism's genome (Fitzpatrick et al., 2006). The increase in available genome sequence data can be utilised for the discovery of genome-wide polymorphisms that can be compared between different species. Two types of polymorphism that have potential for species recognition include microsatellites and single nucleotide polymorphisms (SNPs). Once polymorphic sites have been identified, rapid diagnostic tools can be developed to screen for the polymorphisms in numerous individuals, making large scale comparative studies feasible (Garvin et al., 2010; Guichoux et al., 2011).

4.2.1. Microsatellites

Microsatellites are known to occur in prokaryotes and eukaryotes and tend to be highly polymorphic (Dettman and Taylor, 2004). Variation in the number of microsatellite repeats has been used extensively for population studies but also has the potential to recognise distinct species (Burgess et al., 2001; De Wet et al., 2003; Pavlic, 2009). Advantages of these markers are their high mutation rate that can detect recent evolutionary events (Bhargava and Fuentes, 2010), they are co-dominant markers that allows for the detection of both alleles in an individual, results obtained are easy to score and the data generated are reproducible (Dettman and Taylor, 2004).

A collection of 58 microsatellite makers have been designed for species in the *C. fimbriata s.l.* complex to investigate the genetic diversity and structure of populations (Barnes et al., 2001; Rizatto et al., 2010; Steimel et al., 2004). Some of these markers have also shown potential to detect distinct *Ceratocystis* species. Microsatellite variation in host specific isolates of *C. fimbriata s.l.* could group the isolates from different areas and different hosts into six distinct clades (Barnes et al., 2001). Some of the clades have since been described as separate species, including *C. fimbriata s.s*. from sweet potato, *C. platani* from plane trees in Europe (Engelbrecht and Harrington, 2005), *C. colombiana* from coffee and citrus trees (Van Wyk et al., 2010) and *C. harringtonii* from *Poplus* sp. in North America (Johnson et al., 2005). A set of 10 additional microsatellite markers, specific to species in the *C. fimbriata s.l.* complex, were recently developed (Simpson et al., 2013) and could distinguish between 12 species in the complex.

The collection of 58 microsatellite markers (Barnes et al., 2001; Steimel et al., 2004) have been investigated in only a few of the *C. fimbriata s.l.* species. These markers might also distinguish many of the newly identified, cryptic species and can be investigated in future studies. The 10 new markers (Simpson et al., 2013) have conclusively confirmed the use of microsatellite markers for *C. fimbriata s.l.* species identification but have not been evaluated on some of the problematic species, such as *C. mangicola, C. mangivora* and *C. eucalypticola.* The variation between these species can be investigated in the 10 microsatellite markers.

4.2.2. Single nucleotide polymorphisms

A single nucleotide polymorphism (SNP) is the result of a single basepair mutation when a nucleotide, commonly found in the population, is replaced by an alternative nucleotide (Schork et al., 2000). When this mutation is transferred from one generation to the next and is eventually fixed in at least 1% of a population, the single nucleotide variation is termed a SNP (Human Genome Project Information, US Department of Energy genome programs, Washington, DC, http://genomics.energy.gov). SNPs can be highly informative polymorphic markers since they are more abundant in the genome than other variable mutations such as microsatellites or indels (Fakhrai-Rad et al., 2002; Gupta et al., 2001). SNPs generally have low mutation rates, compared to microsatellites, and might be evolutionary more conserved, allowing identification of orthologous sites in further related species (Broders et al., 2011; Fakhrai-Rad et al., 2002). SNPs occur in coding as well as non-coding regions, including exons, introns and regulatory regions (Schork et al., 2000). The frequency of SNPs in a genome can vary significantly between different fungal species. For example, in *Ophiognomonia* species, the occurrence of SNPs is relatively low with 1 SNP found every 599 276 bp but in *Fusarium graminearum,* the SNP density is as high as 0–17.5 SNPs per kb (Broders et al., 2011).

The SNP differences between isolates can be used to identify SNPs that are conserved within a species but variable between species (Pérez, 2010). SNP identification can be performed through genome-wide sequence comparisons of numerous isolates. Sequencing costs can be lowered by sequencing reduced representations of a genome such as restriction enzyme digested libraries (AFLP and RAD) or transcriptome libraries. Each individual can also be uniquely labelled prior to sequencing which enables multiplexing of individuals (Helyar et al., 2012; Van Tassell et al., 2008). Sequencing technologies such as pyrosequencing and Illumina (RAD-sequencing) have been used effectively for identification of SNPs used in population and phylogeographic studies (Broders et al., 2011; Chouvarine et al., 2012; Decker et al., 2009; Fakhrai-Rad et al., 2002; Hohenlohe et al., 2011).

Interspecies comparisons of diagnostic SNP markers have confirmed the utility of SNP data to distinguish between plant as well as animal species (Chouvarine et al., 2012; Garvin et al., 2011; Jones et al., 2013; Wang et al., 2013). There are few reports describing the utility of SNP markers for fungal species delineation but a recent study on three fungal genera, *Neofusicoccum, Fusarium* and *Teratosphaeria*, suggested these markers are useful in fungi as well (Pérez, 2010; Pérez et al., 2012). In this study an average of 18 polymorphic markers were developed for each genus and could effectively distinguish between cryptic species that were not separated by single gene phylogenetic markers. The SNP diversity within and between species in the *C. fimbriata s.l.* complex have not been investigated and provides great potential to assist in the identification of cryptic species.

5. Conclusions

This literature review provides a summary of the known species in *Ceratocystis*, the species complexes into which they are classified and the methodologies that have been used for the identification and taxonomy of the species. The *C. fimbriata s.l.* species complex is currently the largest complex in the genus and most of the species in the complex are pathogenic to their hosts (Engelbrecht and Harrington, 2005; Van Wyk et al., 2011b; Wingfield et al., 2013). Some species cause death of forest plantation trees, ornamental and fruit trees and have had a significant economic impact on plantation and fruit orchard industries (Al Adawi et al., 2006; Engelbrecht and Harrington, 2005; Heath et al., 2009). Species in this complex have rapidly spread throughout the globe as a result of anthropogenic activity and now occur in countries on almost every continent. Some of the species have also displayed the ability to jump from one host, such as fruit trees, to native leguminous trees, increasing their potential economic threat (Al Adawi et al., 2013; Chen et al., 2013; Tarigan et al., 2010). In order to investigate the full host range and geographical distribution of such pathogens it is essential to have reliable tools for identification of species in the *C. fimbriata s.l.* complex.

A diverse collection of species identification methods have been used in *Ceratocystis* species. Morphological species recognition criteria represents the historically oldest method applied (Upadhyay, 1981) but this review illustrates that morphology provides limited or no variation between closely related species (Van Wyk et al., 2011a; Van Wyk et al., 2011b). Even though biological species recognition has been useful for the delineation of some species, this method also has technical constraints and might not be very reliable in certain cases (Taylor et al., 2000). DNA sequence data have had a profound impact on the rate and accuracy of species identification in *Ceratocystis* taxonomy. The combination of phylogenetic gene regions for GCPSR has enabled the identification of numerous cryptic species in the genus (Mbenoun et al., 2013; Van Wyk et al., 2007a).

This review has highlighted the fact that the gene regions currently used to distinguish between *Ceratocystis* species have limitations in their phylogenetic utility. Some of the cryptic species identified are primarily based on sequence differences in their ITS region. The recent discovery of multiple ITS sequences in single isolates (Naidoo et al., 2013) has caused uncertainty about the use of this marker for future species descriptions (Harrington et al., 2014). There is, therefore, a need to screen existing alternative markers or develop new molecular markers to support the ITS phylogeny in *Ceratocystis*.

The aims of this thesis will be to investigate the potential use of alternative markers for species delineation in the *C. fimbriata s.l.* complex. For this purpose, a number of gene regions identified in this literature review will be investigated. The regions include those used in AFTOL and FBoL studies and genes identified from the FUNYBASE database (Marthey et al., 2008). In addition, the potential of SNP markers to provide significant support for species discrimination in *Ceratocystis* will be investigated. This method has not been studied extensively in fungal species and could prove to be informative. The thesis will then focus more specifically on addressing the species concept problems associated with the economically important species *C. manginecans* and *C. acaciivora* and investigate potential new areas where the species occur. The use of microsatellite markers to detect variation within and between these two species will also be investigated. It is envisaged that the new data presented in this thesis will help resolve the dilemma which now faces *C. fimbriata s.l.* taxonomy by providing additional insights on the current classification of the species in this species complex.

6. References

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

Aguileta, G., Marthey, S., Chiapello, H., Lebrun, M.H., Rodolphe, F., Fournier, E., Gendrault-Jacquemard, A., Giraud, T., 2008. Assessing the Performance of Single-Copy Genes for Recovering Robust Phylogenies. Syst. Biol. 57, 613-627.

Al-Subhi, A.M., Al-Adawi, A.O., Van Wyk, M., Deadman, M.L., Wingfield, M.J., 2006. *Ceratocystis omanensis*, a new species from diseased mango trees in Oman. Mycol. Res. 110, 237-245.

Al Adawi, A.O., Barnes, I., Khan, I.A., Al Subhi, A.M., Al Jahwari, A.A., Deadman, M.L., Wingfield, B.D., Wingfield, M.J., 2013. *Ceratocystis manginecans* associated with a serious wilt disease of two native legume trees in Oman and Pakistan. Australas. Plant Pathol. 42, 179-193.

Al Adawi, A.O., Deadman, M.L., Al Rawahi, A.K., Al Maqbali, Y.M., Al Jahwari, A.A., Al Saadi, B.A., Al Amri, I.S., Wingfield, M.J., 2006. Aetiology and causal agents of mango sudden decline disease in the Sultanate of Oman. Eur. J. Plant Pathol. 116, 247-254.

Al Adawi, A.O., Deadman, M.L., Al Rawahi, A.K., Khan, A.J., Al Maqbali, Y.M., 2003. *Diplodia theobromae* associated with sudden decline of mango in the Sultanate of Oman. Plant Pathol. 52, 419-419.

Amalfi, M., Raymundo, T., Valenzuela, R., Decock, C., 2012. *Fomitiporia cupressicola* sp. nov., a parasite on *Cupressus arizonica*, and additional unnamed clades in the southern USA and northern Mexico, determined by multilocus phylogenetic analyses. Mycologia 104, 880-893.

Avise, J.C., Wollenberg, K., 1997. Phylogenetics and the origin of species. Proc. Natl. Acad. Sci. USA 94, 7748-7755.

Baccelli, I., Comparini, C., Bettini, P.P., Martellini, F., Ruocco, M., Pazzagli, L., Bernardi, R., Scala, A., 2012. The expression of the cerato-platanin gene is related to hyphal growth and chlamydospores formation in *Ceratocystis platani*. FEMS Microbiol. Lett. 327, 155-163.

Baker, C.J., Harrington, T.C., Krauss, U., Alfenas, A.C., 2003. Genetic Variability and Host Specialization in the Latin American Clade of *Ceratocystis fimbriata*. Phytopathology 93, 1274-1284.

Bakshi, B.K., 1950. Fungi associated with ambrosia beetles in Great Britain. Trans. Brit. Mycol. Soc. 33, 111-120.

Bakshi, B.K., 1951. Studies on four species of *Ceratocystis*, with a discussion on fungi causing sap-stain in Britain. Mycol. Pap. 35, 1-16.

Balajee, S.A., Borman, A.M., Brandt, M.E., Cano, J., Cuenca-Estrella, M., Dannaoui, E., Guarro, J., Haase, G., Kibbler, C.C., Meyer, W., O'Donnell, K., Petti, C.A., Rodriguez-Tudela, J.L., Sutton, D., Velegraki, A., Wickes, B.L., 2009. Sequence-Based Identification of *Aspergillus, Fusarium*, and *Mucorales* Species in the Clinical Mycology Laboratory: Where Are We and Where Should We Go from Here? J. Clin. Microbiol. 47, 877-884.

Barnes, I., Guar, A., Burgess, T., Roux, J., Wingfield, B.D., Wingfield, M.J., 2001. Microsatellite markers reflect intra-specific relationships between isolates of the vascular wilt pathogen *Ceratocystis fimbriata*. Mol. Plant Pathol. 2, 319-325.

Barnes, I., Roux, J., Wingfield, B.D., Dudzinski, M.J., Old, K.M., Wingfield, M.J., 2003. *Ceratocystis pirilliformis*, a new species from *Eucalyptus nitens* in Australia. Mycologia 95, 865-871.

Begerow, D., Nilsson, H., Unterseher, M., Maier, W., 2010. Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Appl. Microbiol. Biotechnol. 87, 99-108.

Benson, D., Karsch-Mizrachi, I., Clark, K., Lipman, D.J., Ostell, J., Sayers, E.W., 2012. GenBank. Nucleic Acids Res. 40(Database issue), D48-53.

Bessey, E.A., 1950. Morphology and taxonomy of fungi. The Blakiston Company, Philadelphia.

Bhargava, A., Fuentes, F.F., 2010. Mutational Dynamics of Microsatellites. Mol. Biotechnol. 44, 250-266.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22, 148-155.

Blackwell, M., 2010. Fungal evolution and taxonomy. BioControl 55, 7-16.

Blackwell, M., Hibbett, D.S., Taylor, J.W., Spatafora, J.W., 2006. Research Coordination Networks: a phylogeny for kingdom Fungi (Deep Hypha). Mycologia 98, 829-837.

Bohar, G., 1996. *Ceratocystis erinaceus*: A new endophyte in the heartwood of oak. Acta Phytopathologica et Entomologica Hungarica 31, 213-218.

Broders, K.D., Woeste, K.E., Sanmiguel, P.J., Westerman, R.P., Boland, G.J., 2011. Discovery of single-nucleotide polymorphisms (SNPs) in the uncharacterized genome of the ascomycete *Ophiognomonia clavigignenti-juglandacearum* from 454 sequence data. Mol. Ecol. Resour. 11, 693–702.

Burgess, T., Wingfield, M.J., Wingfield, B.W., 2001. Simple Sequence Repeat Markers Distinguish among Morphotypes of *Sphaeropsis sapinea*. Appl. Environ. Microbiol. 67, 354-362.

Cai, L., Giraud, T., Zhang, N., Begerow, D., Cai, G., Shivas, R., 2011. The evolution of species concepts and species recognition criteria in plant pathogenic fungi. Fungal Divers. 50, 121-133.

Chakrabarti, A., Rep, M., Wang, B., Ashton, A., Dodds, P., Ellis, J., 2011. Variation in potential effector genes distinguishing Australian and non-Australian isolates of the cotton wilt pathogen *Fusarium oxysporum* f.sp. *vasinfectum*. Plant Pathol. 60, 232-243.

Chapin, E., Arcangioli, D., 2007. Evolution et situation du Chancre Colore Dans Le Monde, en Europe et en France. AFPP- Chancre Colore du Platane, Toulouse.

Chase, M.W., Fay, M.F., 2009. Barcoding of Plants and Fungi. Science 325, 682-683.

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

Chouvarine, P., Cooksey, A.M., McCarthy, F.M., Ray, D.A., Baldwin, B.S., Burgess, S.C., Peterson, D.G., 2012. Transcriptome-Based Differentiation of Closely-Related *Miscanthus* Lines. PLoS ONE 7, e29850.

Coetzee, M.P.A., Wingfield, B.D., Coutinho, T.A., Wingfield, M.J., 2000. Identification of the Causal Agent of Armillaria Root Rot of *Pinus* Species in South Africa. Mycologia 92, 777-785.

Comparini, C., Carresi, L., Pagni, E., Sbrana, F., Sebastiani, F., Luchi, N., Santini, A., Capretti, P., Tiribilli, B., Pazzagli, L., Cappugi, G., Scala, A., 2009. New proteins orthologous to cerato-platanin in various *Ceratocystis* species and the purification and characterization of cerato-populin from *Ceratocystis populicola*. Appl. Microbiol. Biotechnol. 84, 309-322.

De Queiroz, K., 2007. Species Concepts and Species Delimitation. Syst. Biol. 56, 879-886.

De Wet, J., Burgess, T., Slippers, B., Preisig, O., Wingfield, B.D., Wingfield, M.J., 2003. Multiple gene genealogies and microsatellite markers reflect relationships between morphotypes of *Sphaeropsis sapinea* and distinguish a new species of *Diplodia*. Mycol. Res. 107, 557-566.

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., Johnson, G.S., Brenneman, R.A., Hanotte, O., Eggert, L.S., Wiener, P., Kim, J.-J., Kim, K.S., Sonstegard, T.S., Van Tassell, C.P., Neibergs, H.L., McEwan, J.C., Brauning, R., Coutinho, L.L., Babar, M.E., Wilson, G.A., McClure, M.C., Rolf, M.M., Kim, J., Schnabel, R.D., Taylor, J.F., 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. 106, 18644-18649.

Dettman, J.R., Taylor, J.W., 2004. Mutation and evolution of microsatellite loci in *Neurospora*. Genetics 168, 1231-1248.

Duong, T.A., De Beer, Z.W., Wingfield, B.D., Wingfield, M.J., 2012. Phylogeny and taxonomy of species in the *Grosmannia serpens* complex. Mycologia 104, 715-732.

Engelbrecht, C.J.B., Harrington, T.C., 2005. Intersterility, morphology and taxonomy of *Ceratocystis fimbriata* on sweet potato, cacao and sycamore. Mycologia 97, 57-69.

Fakhrai-Rad, H., Pourmand, N., Ronaghi, M., 2002. Pyrosequencing™: An accurate detection platform for single nucleotide polymorphisms. Human Mutation 19, 479-485.

FAOSTAT, 2011. FAOSTAT On-line. Rome. United Nations Food and Agriculture Organization. Available at: http://faostat3.fao.org/faostat-gateway/go/to/home/E (accessed 9 September 2013).

Ferreira, E.M., Harrington, T.C., Thorpe, D.J., Alfenas, A.C., 2010. Genetic diversity and interfertility among highly differentiated populations of *Ceratocystis fimbriata* in Brazil. Plant Pathol. 59, 721-735.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., Butler, G., 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol. Biol. 6, 99.

Garvin, M.R., Marcotte, R.W., Palof, K.J., Riley, R.J., Kamin, L.M., Gharrett, A.J., 2011. Diagnostic Single-Nucleotide Polymorphisms Identify Pacific Ocean Perch and Delineate Blackspotted and Rougheye Rockfish. Trans. Am. Fish. Soc. 140, 984-988.

Garvin, M.R., Saitoh, K., Gharrett, A.J., 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. Mol. Ecol. Resour. 10, 915-934.

Gilmore, S.R., GrÄfenhan, T.O.M., Louis-Seize, G., Seifert, K.A., 2009. Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fusarium*. Mol. Ecol. Resour. 9, 90-98.

Giraud, T., Refrégier, G., Le Gac, M., De Vienne, D.M., Hood, M.E., 2008. Speciation in fungi. Fungal Genet. Biol. 45, 791-802.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., Petit, R.J., 2011. Current trends in microsatellite genotyping. Mol. Ecol. Resour. 11, 591-611.

Gupta, P.K., Roy, J.K., Prasad, M., 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr. Sci. 80, 524-535.

Hadden, G., Hugh-Jones, R., 2011. The disease killing Europe's plane trees. PRI's The World, BBC News. BBC news, Boston.

Halsted, B.D., 1890. Some fungous disease of the sweet potato. Agricultural College Experiment Station Bulletin 76, 1-32.

Harrington, T.C., 2000. Host specialization and speciation in the American wilt pathogen *Ceratocystis fimbriata*. Fitopatologia Brasileira 25S, 262-263.

Harrington, T.C., 2008. The genus *Ceratocystis*: where does the oak wilt fungus fit? In: Billings, R.F., Appel, D.N. (Eds.), National Oak wilt symposium, Austin, Texas.

Harrington, T.C., Kazmi, M.R., Al-Sadi, A.M., Ismail, S.I., 2014. Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbirata sensu stricto*. Mycologia In press. doi: 10.3852/13-189.

Harrington, T.C., McNew, D.L., 1997. Self fertility and uni-directional mating-type switching in *Ceratocystis coerulescens,* a filamentous ascomycete. Curr. Genet. 32, 52-59.

Harrington, T.C., McNew, D.L., 1998. Partial Interfertility Among the *Ceratocystis* Species on Conifers. Fungal Genet. Biol. 25, 44-53.

Harrington, T.C., Rizzo, D.M., 1999. Defining species in the fungi. In: Worrall, J.J. (Ed.), Structure and dynamics of fungal populations. Kluwer academic publishers, Dordrecht, Netherlands, pp. 43-71.

Harrington, T.C., Steimel, J.P., Wingfield, M.J., Kile, G.A., 1996. Isozyme Variation and Species Delimitation in the *Ceratocystis coerulescens* Complex. Mycologia 88, 104-113.

Harrington, T.C., Wingfield, M.J., 1998. The *Ceratocystis* species on conifers. Can. J. Bot. 76, 1446-1457.

Hausner, G., Reid, J., Klassen, G.R., 1993. Grouping of isolates and species of *Ceratocystis sensu lato* on the basis of molecular and morphological characters. In: Wingfield, M.J., Seifert, K.A., Webber, J.F. (Eds.), *Ceratocystis* and *Ophiostoma* - Taxonomy, Ecology, and Pathogenicity. American Phytopathological Society Press, Minnesota, pp. 93-104.

Heath, R.N., Wingfield, M.J., Wingfield, B.D., Meke, G., Mbaga, A., Roux, J., 2009. *Ceratocystis* species on *Acacia mearnsii* and *Eucalyptus* spp. in eastern and southern Africa including six new species. Fungal Divers. 34, 41-67.

Helyar, S.J., Limborg, M.T., Bekkevold, D., Babbucci, M., Van Houdt, J., Maes, G.E., Bargelloni, L., Nielsen, R.O., Taylor, M.I., Ogden, R., Cariani, A., Carvalho, G.R., Consortium, F., Panitz, F., 2012. SNP Discovery Using Next Generation Transcriptomic Sequencing in Atlantic Herring (*Clupea harengus*). PLoS ONE 7, e42089.

Hey, J., 2006. On the failure of modern species concepts. Trends Ecol. Evol. 21, 447-450.

Hibbett, D.S., Taylor, J.W., 2013. Fungal systematics: is a new age of enlightenment at hand? Nat. Rev. Microbiol. 11, 129-133.

Hogenhout, S.A., Van der Hoorn, R.A.L., Terauchi, R., Kamoun, S., 2009. Emerging Concepts in Effector Biology of Plant-Associated Organisms. Mol. Plant Microbe Interact. 22, 115-122.

Hohenlohe, P.A., Amish, S.J., Catchen, J.M., Allendorf, F.W., Luikart, G., 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol. Ecol. Resour. 11, 117-122.

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4, 275-284.

Hunt, J., 1956. Taxonomy of the genus *Ceratocystis*. Lloydia - A Quarterly Journal of Biological Science 19, 1-58.

Johnson, J.A., Harrington, T.C., Engelbrecht, C.J.B., 2005. Phylogeny and taxonomy of the North American clade of the *Ceratocystis fimbriata* complex. Mycologia 97, 1067-1092.

Jones, J.C., Fan, S., Franchini, P., Schartl, M., Meyer, A., 2013. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Mol. Ecol. 22, 2986-3001.

Kajitani, Y., Masuya, H., 2011. *Ceratocystis ficicola* sp. nov., a causal fungus of fig canker in Japan. Mycoscience 52, 349-353.

Kamgan, N.G., Wingfield, M.J., Mohammed, C., Carnegie, A.J., Pegg, G.S., Roux, J., 2012. *Ceratocystis* species, including two new species associated with nitidulid beetles, on eucalypts in Australia. Anton. Leeuw. 101, 217-241.

Kile, G.A., 1993. Plant diseases caused by species of *Ceratocystis sensu stricto* and *Chalara*. In: Wingfield, M.J., Seifert, K.A., Webber, J.F. (Eds.), *Ceratocystis* and *Ophiostoma* - Taxonomy, Ecology, and Pathogenicity. American Phytopathological Society Press, Minnesota, pp. 173-183

Kile, G.A., Harrington, T.C., Yuan, Z.Q., Dudzinski, M.J., Old, K.M., 1996. *Ceratocystis eucalypti* sp. nov., a vascular stain fungus from eucalypts in Australia. Mycol. Res. 100, 571-579.

Kirk , P.M., Cannon, P.F., Minter, D.W., Stalpers, J.A., 2008. Dictionary of the Fungi, 10th ed. CABI, Wallingford, UK.

Kõljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Lücking, R., Martín, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Põldmaa, K., Saag, L., Saar, I., Schüßler, A., Scott, J.A., Senés, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T., Weiss, M., Larsson, K.-H., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22, 5271-5277.

Krüger, M., Krüger, C., Walker, C., Stockinger, H., Schüßler, A., 2012. Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. New Phytol. 193, 970-984.

Kuramae, E.E., Robert, V., Snel, B., Weiß, M., Boekhout, T., 2006. Phylogenomics reveal a robust fungal tree of life. FEMS Yeast Res. 6, 1213-1220.

Lindner, D.L., Carlsen, T., Nilsson, R.H., Davey, M., Schumacher, T., Kauserud, H., 2013. Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. Ecol. Evol. 3, 1751-1764.

Loppnau, P.A., Breuil, C., 2003. Species level identification of conifer associated *Ceratocystis* sapstain fungi by PCR-RFLP on a β-tubulin gene fragment. FEMS Microbiol. Lett. 222, 143-147.

Lutzoni, F., Kauff, F., Cox, C.J., McLaughlin, D., Celio, G., Dentinger, B., Padamsee, M., Hibbett, D., James, T.Y., Baloch, E., Grube, M., Reeb, V., Hofstetter, V., Schoch, C., Arnold, A.E., Miadlikowska, J., Spatafora, J., Johnson, D., Hambleton, S., Crockett, M., Shoemaker, R., Sung, G.-H., Lücking, R., Lumbsch, T., O'Donnell, K., Binder, M., Diederich, P., Ertz, D., Gueidan, C., Hansen, K., Harris, R.C., Hosaka, K., Lim, Y.-W., Matheny, B., Nishida, H., Pfister, D., Rogers, J., Rossman, A., Schmitt, I., Sipman, H., Stone, J., Sugiyama, J., Yahr, R., Vilgalys, R., 2004. Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. Am. J. Bot. 91, 1446-1480.

Malloch, D., Blackwell, M., 1993. Dispersal biology of the Ophiostomatoid fungi. In: Wingfield, M.J., Seifert, K.A., Webber, J.F. (Eds.), *Ceratocystis* and *Ophiostoma* - Taxonomy, Ecology, and Pathogenicity. APS Press, Minnesota, pp. 207-218.

Marin, M., Preisig, O., Wingfield, B.D., Kirisits, T., Yamaoka, Y., Wingfield, M.J., 2005. Phenotypic and DNA sequence data comparisons reveal three discrete species in the *Ceratocystis polonica* species complex. Mycol. Res. 109, 1137-1148.

Marthey, S., Aguileta, G., Rodolphe, F., Gendrault, A., Giraud, T., Fournier, E., Lopez-Villavicencio, M., Gautier, A., Lebrun, M.-H., Chiapello, H., 2008. FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics doi: 10.1186/1471-2105-9-456.

Mayden, R.L., 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Chapman and Hall, London.

Mbenoun, M., Wingfield, M.J., Begoude Boyogueno, A., Wingfield, B.D., Roux, J., 2013. Molecular phylogenetic analyses reveal three new *Ceratocystis* species and provide evidence for geographic differentiation of the genus in Africa. Mycol. Prog. doi: 10.1007/s11557-013-0907-5.

Naidoo, K., Steenkamp, E.T., Coetzee, M.P.A., Wingfield, M.J., Wingfield, B.D., 2013. Concerted Evolution in the Ribosomal RNA Cistron. PLoS ONE 8, e59355. doi: 59310.51371/journal.pone.0059355.

Nkuekam, G.K., Wingfield, M.J., Roux, J., 2013. *Ceratocystis* species, including two new taxa, from *Eucalyptus* trees in South Africa. Australas. Plant Pathol. 42, 283-311.

O'Donnell, K., Cigelnik, E., 1997. Two Divergent Intragenomic rDNA ITS2 Types within a Monophyletic Lineage of the Fungus *Fusarium* Are Nonorthologous. Mol. Phylogenet. Evol. 7, 103-116.

Ocasio-Morales, R.G., Tsopelas, P., Harrington, T.C., 2007. Origin of *Ceratocystis platani* on native *Platanus orientalis* in Greece and its impact on natural forests. Plant Disease 91, 901-904.

Pavlic, D., 2009. Taxonomy and population diversity of Botryosphaeriaceae associated with woody hosts in South Africa and Western Australia. Department of Microbiology and Plant Pathology. University of Pretoria, Pretoria.

Pérez, G., 2010. Chapter 6: Inter-Specific SNPs for cryptic species delimitation and phylogenetic reconstructions in closely related species. Global genetic diversity of the *Eucalyptus* leaf pathogen *Teratosphaeria nubilosa* species complex in native forests and sommercial plantations, Microbiology and Plant Pathology, University of Pretoria, Pretoria pp. 185 - 205.

Pérez, G., Slippers, B., Wingfield, M.J., Wingfield, B.D., Carnegie, A.J., Burgess, T.I., 2012. Cryptic species, native populations and biological invasions by a eucalypt forest pathogen. Mol. Ecol. 21, 4452-4471.

Raja, A.H., Schoch, C.L., Hustad, V.P., Shearer, C.A., Miller, A.N., 2011. Testing the phylogenetic utility of *Mcm7* in the Ascomycota. MycoKeys 1, 63-94.

Reid, J., Iranpour, M., Rudski, S.M., Loewen, P.C., Hausner, G., 2010. A new conifer-inhabiting species of *Ceratocystis* from Norway. Botany 88, 971-983.

Rintoul, T.L., Eggertson, Q.A., Lévesque, C.A., 2012. Multigene phylogenetic analyses to delimit new species in fungal plant pathogens. In: Bolton, M.D., Thomma, B.P.H.J. (Eds.), Plant Fungal Pathogens: Methods and Protocols (Methods in Molecular Biology) Humana Press, pp. 549-569.

Rizatto, S., De Araújo Batista, C., Bajay, M., Sigrist, M., Ito, M., Monteiro, M., Cavallari, M., Pinheiro, J., Zucchi, M., 2010. A new set of microsatellite markers for the genetic characterization of *Ceratocystis fimbriata*, an economically important plant pathogen. Conserv. Genet. Resour. 2, 55-58.

Roe, A.D., Rice, A.V., Bromilow, S.E., Cooke, J.E.K., Sperling, F.A.H., 2010. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. Mol. Ecol. Resour. 10, 946-959.

Roux, J., Wingfield, M.J., 2009. *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. South. For. 71, 115-120.

Ruse, M., 1969. Definitions of Species in Biology. The Brit. J. Philos. Sci. 20, 97-119.

Santamaria, M., Vicario, S., Pappadà, G., Scioscia, G., Scazzocchio, C., Saccone, C., 2009. Towards barcode markers in Fungi: An intron map of Ascomycota mitochondria. BMC Bioinformatics 10.

Schmitt, I., Crespo, A., Divakar, P.K., Fankhauser, J.D., Herman-Sackett, E., Kalb, K., Nelsen, M.P., Nelson, N.A., Rivas-Plata, E., Shimp, A.D., Widhelm, T., Lumbsch, H.T., 2009. New primers for promising single-copy genes in fungal phylogenetics and systematics. Persoonia 23, 35-40.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. 109, 6241-6246.

Schoch, C.L., Seifert, K.A., Spouge, J., Robert, V.A., Bolchacova, E., Crous, P.W., 2011. An official barcode for fungi. Fourth International Barcode of Life Conference, University of Adelaide, Plenary Talk, Closing Plenary, Elder Hall.

Schork, N.J., Fallin, D., Lanchbury, S., 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. genet. 58, 250–264.

Seifert, K.A., 2009. Progress towards DNA barcoding of fungi. Mol. Ecol. Resour. 9, 83-89.

Seifert, K.A., Samson, R.A., deWaard, J.R., Houbraken, J., Lévesque, C.A., Moncalvo, J.-M., Louis-Seize, G., Hebert, P.D.N., 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. Proc. Natl. Acad. Sci. 104, 3901-3906.

Shivas, R.G., Cai, L., 2012. Cryptic fungal species unmasked. Microbiol. Aust., 36-37.

Simpson, M.C., Wilken, P.M., Coetzee, M.P.A., Wingfield, M.J., Wingfield, B.D., 2013. Analysis of microsatellite markers in the genome of the plant pathogen *Ceratocystis fimbriata*. Fungal Biol. 117, 545–555.

Spatafora, J.W., Blackwell, M., 1994. The polyphyletic origins of ophiostomatoid fungi. Mycol. Res. 98, 1-9.

Sprecher, E., Hanssen, H.-P., 1983. Distribution and strain-dependent formation of volatile metabolites in the genus *Ceratocystis*. Anton. Leeuw. 49, 493-499.

Steimel, J., Engelbrecht, C.J.B., Harrington, T.C., 2004. Development and characterization of microsatellite markers for the fungus *Ceratocystis fimbriata*. Mol. Ecol. Notes 4, 215-218.

Stergiopoulos, I., De Wit, P.J.G.M., 2009. Fungal Effector Proteins. Annu. Rev. Phytopathol. 47, 233-263.

Stipes, R.J., 1970. Comparative Mycelial Protein and Enzyme Patterns in Four Species of *Ceratocystis*. Mycologia 62, 987-995.

Tarigan, M., Van Wyk, M., Roux, J., Tjahjono, B., Wingfield, M.J., 2010. Three new *Ceratocystis* spp. in the *Ceratocystis moniliformis* complex from wounds on *Acacia mangium* and *A. crassicarpa*. Mycoscience 51, 53-67.

Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S., Fisher, M.C., 2000. Phylogenetic Species Recognition and Species Concepts in Fungi. Fungal Genet. Biol. 31, 21-32.

Upadhyay, H.P., 1981. A monograph of *Ceratocystis* and *Ceratocystiopsis*. University of Georgia Press, Athens.

Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Methods 5, 247-252.

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

Van Wyk, M., Pegg, G., Lawson, S., Wingfield, M.J., 2007b. *Ceratocystis atrox sp. nov*. associated with *Phoracantha acanthocera* infestations on *Eucalyptus grandis* in Australia. Australas. Plant Pathol. 36, 407-414.

Van Wyk, M., Roux, J., Barnes, I., Wingfield, B.D., Chhetri, D.B., Kirisits, T., Wingfield, M.J., 2004. *Ceratocystis bhutanensis* sp. nov., associated with the bark beetle *Ips schmutzenhoferi* on *Picea spinulosa* in Bhutan. Stud. Mycol. 50, 365-379.

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

Van Wyk, M., Roux, J., Kamgan, N.G., Wingfield, B.D., Wingfield, M.J., 2012. *Ceratocystis eucalypticola sp. nov*. from *Eucalyptus* in South Africa and comparison to global isolates from this tree. IMA Fungus 3, 45-58.

Van Wyk, M., Wingfield, B.D., Al-Adawi, A.O., Rossetto, C.J., Ito, M.F., Wingfield, M.J., 2011a. Two new *Ceratocystis* species associated with mango disease in Brazil. Mycotaxon 117, 381–404.

Van Wyk, M., Wingfield, B.D., Clegg, P.A., Wingfield, M.J., 2009a. *Ceratocystis larium* sp. nov., a new species from Styrax benzoin wounds associated with incense harvesting in Indonesia. Persoonia. 22, 75–82.

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

Van Wyk, M., Wingfield, B.D., Mohali, S., Wingfield, M.J., 2009b. *Ceratocystis fimbriatomima*, a new species in the *C. fimbriata sensu lato* complex isolated from *Eucalyptus* trees in Venezuela. Fungal Divers. 34, 175-185.

Van Wyk, M., Wingfield, B.D., Wingfield, M.J., 2011b. Four new *Ceratocystis* spp. associated with wounds on *Eucalyptus, Schizolobium* and *Terminalia* trees in Ecuador. Fungal Divers. 46, 111-131.

Voigt, K., Kirk, P., 2011. Recent developments in the taxonomic affiliation and phylogenetic positioning of fungi: impact in applied microbiology and environmental biotechnology. Appl. Microbiol. Biotechnol. 90, 41-57.

Walker, D.M., Castlebury, L., Rossman, A., Mejía, L., White, J., 2012a. Phylogeny and taxonomy of *Ophiognomonia* (*Gnomoniaceae*; Diaporthales), including twenty-five new species in this highly diverse genus. Fungal Divers. 57, 85-147.

Walker, D.M., Castlebury, L.A., Rossman, A.Y., White, J.F., 2012b. New molecular markers for fungal phylogenetics: Two genes for species-level systematics in the Sordariomycetes (Ascomycota). Mol. Phylogenet. Evol. 64, 500-512.

Wang, X.Q., Zhao, L., Eaton, D.A.R., Li, D.Z., Guo, Z.H., 2013. Identification of SNP markers for inferring phylogeny in temperate bamboos (*Poaceae: Bambusoideae*) using RAD sequencing. Mol. Ecol. Resour. 13, 938-945.

Wingfield, B.D., Grant, W.S., Wolfaardt, J.F., Wingfield, M.J., 1994. Ribosomal RNA sequence phylogeny is not congruent with ascospore morphology among species in *Ceratocystis sensu stricto*. Mol. Biol. Evol. 11, 376-383.

Wingfield, B.D., Van Wyk, M., Roos, H., Wingfield, M.J., 2013. *Ceratocystis*: Emerging evidence for discrete generic boundries. In: Seifert, K.A., De Beer, W., Wingfield, M.J. (Eds.), Ophiostomatoid Fungi: Expanding Frontiers. CBS-KNAW Fungal Biodiversity Centre, AD Utrecht, The Netherlands, pp. 57-64.

Wingfield, M.J., De Beer, C., Visser, C., Wingfield, B.D., 1996. A new *Ceratocystis* species defined using morphological and ribosomal DNA sequence comparisons. Syst. Appl. Microbiol. 19, 191-202.

Wingfield, M.J., Harrington, T.C., Solheim, H., 1997. Two species in the *Ceratocystis coerulescens* complex from conifers in western North America. Can. J. Bot. 75, 827-834.

Witthuhn, R.C., Harrington, T.C., Steimel, J.P., Wingfield, B.D., Wingfield, M.J., 2000. Comparison of isozymes, rDNA spacer regions and MAT-2 DNA sequences as phylogenetic characters in the analysis of the *Ceratocystis coerulescens* complex. Mycologia 92, 447-452.

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

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

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. 13, 303-314.

Yuan, Z.-Q., Mohammed, C., 2002. *Ceratocystis moniliformopsis* sp. nov., an early coloniser of *Eucalyptus obliqua* logs in Tasmania, Australia. Aust. Syst. Bot. 15, 125-133.

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

Chapter 2

Molecular markers delimit cryptic species in the *Ceratocystis fimbriata sensu lato* complex

ABSTRACT

Ceratocystis fimbriata sensu lato represents a complex of closely related fungi that are mostly plant pathogens. Morphological variation between the species in this complex is limited and species delineation is strongly reliant on phylogenetic inference. Primary support for many of the species is based on the ITS region which, on its own, has been used to delineate all species described in the complex. However, the recent discovery of two ITS types in a single isolate of *C. fimbriata s.l.* questions the use of this marker in taxonomic studies. The aim of this study was to consider the potential use of alternative gene regions to support the species boundaries in this complex. The phylogenetic value of the β-tubulin 1 (βT 1) and Translation Elongation Factor 1-α (EF 1-α) gene regions, generally used in combination with ITS, were re-evaluated and compared to five single copy protein coding genes (Calmodulin, RPBII, MS204, FG1093 and Mcm7). As an alternative approach, genome-wide single nucleotide polymorphisms (SNPs) were identified and evaluated as diagnostic markers to distinguish between the species. Fifteen species in the *C. fimbriata s.l.* complex were used in this study. None of the protein coding genes could be used to distinguish all species but a combination of the βT 1, MS204 and RPBII gene regions resolved 11 of the 15 described species. Unique SNP markers were identified for 13 of the species and these provided significant additional support for most of the established taxon boundaries. Other than ITS, none of the markers from this study could distinguish between *C. acaciivora* and *C. manginecans* and these species could be reduced to synonymy. Results of this study also revealed the likely existence of additional species in the *C. fimbriata s.l.* complex.

1. Introduction

A species is defined as a group of individuals that have a unique evolutionary history and that have evolved independently from other lineages (Giraud et al., 2008). In the fungal Kingdom, support for distinct species is based primarily on morphological, biological and phylogenetic characters (Cai et al., 2011; Taylor et al., 2000). However, the classical morphological species recognition approach lacks the capacity to distinguish between closely related taxa. Furthermore, the biological species concept is not possible for all fungi and it can be unreliable and misleading in laboratory environments (Harrington, 2000; Taylor et al., 2000). During the course of the past 25 years, phylogenetic analysis and genealogical concordance phylogenetic species recognition (GCPSR) has gained acceptance for the recognition of many new taxa that would not have been possible based on other criteria (Bridge et al., 2005; Rintoul et al., 2012; Taylor et al., 2000).

Ceratocystis, a genus of insect-associated fungi, provides a useful example of how the morphological, biological and phylogenetic species recognition criteria have been applied in taxonomic studies. The type species, *Ceratocystis fimbriata sensu stricto (s.s.)* was first described more than 120 years ago as the causal agent of root rot on sweet potato (Halsted, 1890). Since then, 61 distinct species have been described in *Ceratocystis* based on a combination of various recognition criteria. The majority of these species reside in three well-supported species complexes defined as *C. coerulescens, C. moniliformis* and *C. fimbriata sensu lato* (Mbenoun et al., 2013; Wingfield et al., 2013).

Up until the early 2000's, species of the *C. fimbriata s.l.* complex were primarily defined based on unique morphological characteristics (Barnes et al., 2003; Kile et al., 1996; Upadhyay, 1981; Wingfield et al., 1996). In some cases identification has been supplemented using mating studies (Engelbrecht and Harrington, 2005; Ferreira et al., 2010) which, for example, could be used to confirm the species boundaries of *C. fimbriata s.s., C. platani* and *C. cacaofunesta* (Engelbrecht and Harrington, 2005). However, the recognition of the majority of cryptic species in *Ceratocystis* was based on DNA sequence data and phylogenetic inference (Van Wyk et al., 2011a; Wingfield et al., 1996).

Since the incorporation of DNA sequence data for species delineation in 1996 (Wingfield et al., 1996), 77% of the species currently known in the Genus *Ceratocystis* have been described. Initial phylogenetic species delineation relied heavily on gene genealogies of the 5.8S and surrounding internal transcribed spacer regions 1 and 2 (ITS1 and ITS2) (Barnes et al., 2003; Roux et al., 2004). In an attempt to increase the resolution of species boundaries, sequences of the β-tubulin 1 (βT 1) and Translation Elongation Factor 1-α (EF 1-α) gene regions have generally been used in combination with the ITS region for species delineation (Van Wyk et al., 2012; Van Wyk et al., 2011b).

The Fungal Barcode of Life (FBoL) consortium has defined the ITS region as the barcode for fungal species identification (Schoch et al., 2012), although it is recognised that the region is not reliable for delineation of all species. Intragenomic ITS variation has been reported in single isolates of some fungi such as species of *Fusarium* (O'Donnell and Cigelnik, 1997), *Laetiporus* (Lindner and Banik, 2011), and more recently in a species of *Ceratocystis* (Al Adawi et al., 2013; Naidoo et al., 2013). In the case of *Ceratocystis*, the one ITS sequence obtained from an isolate was identical to that of *C. manginecans,* while the second ITS type differed by seven nucleotides and had the same sequence as *C. acaciivora.* Whether similar problems exist for all species in the *C. fimbriata s.l.* complex has not been confirmed, although it has recently been suggested in species closely related to *C. acaciivora* and *C. manginecans* (Harrington et al., 2014).

Most of the cryptic species in the *C. fimbriata s.l.* species complex have been separated on the basis of genealogical concordance but a few of these receive the majority of phylogenetic signal from the ITS region (Harrington et al., 2014; Tarigan et al., 2011; Van Wyk et al., 2011a; Van Wyk et al., 2011b). The two additional gene regions used in GCPSR, βT 1 and EF 1-α, contain limited variation and provide very low or no significant bootstrap support for these species. This situation occurs in species such as *C. curvata, C. ecuadoriana, C. mangicola, C. mangivora, C. acaciivora, C. manginecans* and *C. eucalypticola* (Tarigan et al., 2011; Van Wyk et al., 2012; Van Wyk et al., 2011a; Van Wyk et al., 2011b).

Alternative gene regions that can be used for phylogenetic analysis of *C. fimbriata s.l.* species include regions used in fungal phylogenetic studies such as the Assembling the Fungal Tree of Life (AFTOL) and FBoL projects (Seifert, 2009; Spatafora, 2005). These include the ribosomal small subunit (SSU), 1st and 2nd largest subunits of RNA polymerase II (RPBI and II), ATP synthase 6 (ATP6), β-tubulin 2 (βT 2) and Calmodulin gene regions (O'Donnell et al., 2000; Spatafora, 2005). In addition, various single copy protein coding genes, conserved in most fungi, have been identified from genome comparisons (Aguileta et al., 2008; Marthey et al., 2008). Some of these genes that might be useful as phylogenetic markers include the pre-rRNA processing protein (Tsr1), mini-chromosome maintenance protein (Mcm7), 60S ribosomal protein L37 (FG1093) and guanine nucleotide-binding protein subunit beta-like protein (MS204) and the latter two showed good potential to delineate closely related species (Raja et al., 2011; Schmitt et al., 2009; Tretter et al., 2013; Walker et al., 2012b).

Single nucleotide polymorphisms (SNPs) that differ between species also provide opportunities as taxonomic markers. SNPs can occur at high frequencies in a genome and are more conserved than other polymorphisms since they do not change the reading frame in coding regions (Fakhrai-Rad et al., 2002). Diagnostic SNP markers have consequently

been applied in a wide range of studies on, for example, plants (Chouvarine et al., 2012; Wang et al., 2013), animals (Garvin et al., 2011; Jones et al., 2013) and fungi (Pavlic, 2009; Pérez, 2010; Pérez et al., 2012) to distinguish between species.

There is a significant need for additional molecular markers to clearly and accurately differentiate between closely related species in the *C. fimbriata s.l.* complex, especially due to their importance as plant pathogens. Many species in this complex are the cause of important diseases including root rot, wilt and cankers, especially on trees but also on various root crops (Al Adawi et al., 2013; Engelbrecht and Harrington, 2005; Halsted and Fairchild, 1891; Kamgan et al., 2012; Tarigan et al., 2011; Van Wyk et al., 2010). The aim of this study was, therefore, to reassess the phylogenetic value of the βT 1 and EF 1-α gene regions for taxonomic purposes and, in addition, to consider the potential of additional genetic markers to distinguish between the species in this complex.

2. Materials and methods

2.1. Fungal isolates and DNA extraction

Fungal isolates of 15 well-defined species in the Latin-American clade of the *C. fimbriata s. l.* species complex (Mbenoun et al., 2013), were obtained from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa. Species included *C. fimbriata s.s., C. cacaofunesta, C. manginecans, C. platani, C. acaciivora, C. colombiana, C. curvata, C. diversiconidia, C. ecuadoriana, C. eucalypticola, C. fimbriatomima, C. mangicola, C. mangivora, C. neglecta* and *C. papillata* (Table 1). Two *C. pirilliformis* isolates were included as outgroups for the phylogenetic studies. Depending on availability, between two and five isolates from different geographic locations and/or hosts were chosen per species, including the type strain of each species. All isolates were grown in culture on 2% malt extract agar (MEA) supplemented with 50 mg/l streptomycin (Sigma-Aldrich, Germany) and 100 µg/l thymine (Sigma-Aldrich, Germany).

For DNA extraction, mycelium was scraped from the surface of MEA plates and freeze dried. Samples were crushed to a powder with sterile metal beads in a Mixer Mill type MM 301, Retsch® tissue grinder (Retsch, Germany). DNA was extracted using a phenol/chloroform method (Goodwin et al., 1992). Extracted DNA was quantified using a Nanodrop ND_1000 instrument (Nanodrop, Wilmington, Delaware) and the quality assessed by gel electrophoresis on a 1% agarose gel (AGE). For AGE, 5 µl DNA was combined with 2 µl GelRed™ (Biotium, California) and the DNA visualized under UV illumination. DNA concentrations were standardized to a working dilution of 30 ng/µl for subsequent reactions.

2.2. Single copy phylogenetic gene regions

2.2.1. Primer design, PCR amplification and sequencing

Regardless of whether or not ITS sequence data were available on GenBank, amplification and sequencing of all isolates were repeated in this study to avoid discrepancies. The ITS1 and ITS4 primers (White et al., 1990) were used for amplification. Due to a long poly-A repeat in the sequence, primers ITS2 and ITS3 were used for additional sequencing of the internal regions in some isolates (White et al., 1990).

Seven additional gene regions were tested for amplification and their potential to be used as phylogenetic markers in the *C. fimbriata s.l.* complex evaluated. These were Mcm7, Tsr1, Calmodulin (CAL), RPBII, β-tubulin 2 (βT 2), FG1093 and MS204. Information for all primers used in PCR and sequence reactions in this study are summarised in Table 2. The Mcm7 region was amplified with primers *Mcm7*-709 and *Mcm7*-1348 (Schmitt et al., 2009) and CAL with CAL2F and CAL2R2 primers (Duong et al., 2012). Some primers were modified at a few nucleotide sites to be more specific for species in the *C. fimbriata s.l.* complex by aligning them to the *C. fimbriata s.s.* genome (GenBank Accession number APWK01000000)(Wilken et al., 2014). For amplification of the FG1093 region, the FG1093F.cerato and FG1093R.cerato primers were modified from the original FG1093 E1F1 and FG1093 E3R1 primers (Walker et al., 2012b). For the MS204 region the MS204F.cerato and MS204R.cerato primers were modified from the MS204 E1F1 and MS204 E5R1 primers. For species where amplification proved problematic, a smaller region of MS204 was amplified with the primers MS204F.ceratoB and MS204R.ceratoB. Various primer combinations were tested for the Tsr1 region. This included testing the original primers (Schmitt et al., 2009), primers modified from the original (Tsr1.cerato) as well as a completely new forward primer designed in this region. Primers RPB2-5Fb and RPB2-7Rb, used to amplify the RPBII region, were modified from the original RPB2-5F and RPBI2-7R primers and T1d (Duong et al., in preparation), Bt1d, and Bt2d were used for amplification of the βT 2 region (Duong et al., in preparation). For isolates where there were no reliable data available for the βT 1 and EF 1-α regions on GenBank, these were amplified using primers βt1a and βt1b (Glass and Donaldson, 1995) for βT 1 and EF1-728F and EF1-986R (Jacobs et al., 2004) for EF 1-α.

PCR reaction mixtures with a total volume of 25 µl, consisted of 30 ng DNA, 1.5 units FastStart *Taq* DNA polymerase with 2.5 µl 10x FastStart *Taq* DNA polymerase PCR buffer (with 1.5 mM $MgCl_2$) (Roche Applied Science, South Africa), 250 µM dNTP and 0.2 µM forward and reverse primer or 0.8 µM in the case of the degenerate primers. $MgCl_2$ concentration differed for different primer sets (Table 2). An Expand-PCR program was used for amplification of the following primer sets: ITS, Mcm7, βt 1, EF1-α, MS204 (MS204.cerato)

and FG1093. The program was as follows: 96 °C 10 min, (94 °C 30 s, 55 °C 45 s, 72 °C 1 min) x 10 cycles, (94 °C 30 s, 55 °C 45 s, 72 °C 1 min + 5 s/cycle increase) x 30 cycles, 72 °C 10 min. Annealing temperatures differed for each primer set (Table 2). The PCR program for CAL, RPBII, βT 2 and MS204.ceratoB was 95 °C 5 min, (95 °C 30 s, 64 °C 30 s, 72 °C 1 min) x 35 cycles, 72 °C 8 min.

To confirm PCR amplification success of all primer sets, AGE (Agarose Gel Electrophoresis) was performed by combining 2 µl PCR product and 2 µl GelRed™ (Biotium, California) and separating DNA fragments on a 2% agarose gel. DNA bands were visualized under UV illumination. PCR and sequence products were purified with 6% Sephadex G-50 columns using the manufacturer's protocols (Sigma-Aldrich, Germany). Amplification reactions for sequencing were performed using the ABI PRISM™ Big DYE Terminator Cycle Sequencing Ready Reaction Kit (Applied BioSystems, Foster City, California), following the manufacturer's protocols. Sequences were generated on an ABI PRISM™ 3500xl Autosequencer (Applied BioSystems, Foster City, California, USA). All sequence data generated in this study were submitted to GenBank (see Table 1). For the FG1093 region, some isolates required cloning to obtain optimal sequence results. Cloning was performed with a pGEM®-T Easy Vector System, following the manufacturer's instructions (Promega, Madison, USA).

2.2.2. Sequence alignment and phylogenetic analysis

The quality of the raw sequence reads were evaluated and assembled in CLC Bio Main workbench v.6 (CLC Bio, www.clcbio.com). Consensus sequences for each gene region of all isolates were aligned in MAFFT v.6 with the alignment strategy set to E-INS-i for the ITS data set and L-INS-i for all other data sets (Katoh et al., 2005). Alignments of the data sets were also manually inspected and edited in MEGA 5 (Tamura et al., 2011). Two of the gene regions contained a long poly-A repeat region and this was excluded from the analyses. Maximum parsimony analysis (MP), Maximum likelihood (ML) and Bayesian inference (BI) were applied to each data set individually for tree construction. Parsimony analysis was performed in PAUP v. 4.0 (Swofford, 2002) and trees were obtained using the heuristic search option with 1000 replicates with random addition of sequences and a tree bisection reconnection (TBR) branch swapping strategy. Both introns and exons were considered for each gene region. Indels were treated as a 5th character. For application in ML and BI analyses, the best model of evolution for each gene region was identified using jModelTest 0.1 and applying the Akaike Information Criterion (Posada, 2008). ML tree construction was performed in PhyML 3.0 (Guindon and Gascuel, 2003) with the following criteria: proportion of invariable sites was 0, gamma shape was estimated by the program

and the number of substitution sites were set to 6 (except for βT 1 where nst=2). The starting tree was obtained using a BioNJ approach and the branch swopping strategy was set to select the best of either NNI or SPR algorithms. Statistical support for the branches of both MP an ML trees was obtained by 1000 replicates of non-parametric bootstrap analysis of the sequence data. *Ceratocystis pirilliformis* forms part of the *C. fimbriata s.l.* complex but is distantly related to the species in the Latin-American clade and was thus selected as the outgroup to root the trees.

Additional branch support was obtained through Bayesian analysis applying a Markov Chain Monte Carlo (MCMC) algorithm in the program MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003). Tree searching was performed using four independent chains and were run for 6 000 000 generations, sampling every 100th tree. Analyses were run twice and concordance between the two sets was investigated by comparing the log likelihoods in Tracer v.1.5 (Rambaut and Drummond, 2009). The burn-in for each dataset was performed in MrBayes and set to 10 000 generations. The posterior probabilities for the tree topology were obtained by constructing a consensus tree from the data using MrBayes and viewing it in TreeView X (Page, 1996).

2.2.3. Phylogenetic value of single genes and gene combinations

A combination of criteria was used to select the gene regions with the most potential to be used as phylogenetic markers. A three and four gene region combination was evaluated. This was based on i) the number of species that could be distinguished with significant bootstrap and Bayesian support values (>70 BS and >95 BI) by the gene region, ii) the number of species shown to be monophyletic based on the genealogical sorting index (gsi) value (Cummings et al., 2008) and iii) the congruence in tree topology as compared to a combined reference tree (Nye topological score) (Nye et al., 2006).

The gsi_T is a statistical support value, additional to BS and BI values, that indicates the exclusive ancestry of a group of organisms in a genealogy and has proven informative in recent fungal phylogenetic studies (Sakalidis et al., 2011; Taole et al., 2012; Walker et al., 2012a). The analysis produces a value on a scale from 0 to 1 for each identified group, with 0 indicating no exclusive ancestry from other groups in the genealogy and 1 representing monophyly. For each gene region the gsi value was calculated for 100 ML bootstrap trees randomly selected and 10 000 permutation tests were performed for statistical support of each gsi value. From these values, the gsi_T was calculated as a weighted average of all 100 gsi values for each gene region and was considered statistically significant if it had a P-value<0.05 (Sakalidis et al., 2011). All calculations were performed online at http://www.genealogicalsorting.org/index.php.

To determine how accurately each individual gene region represents the relationships among all taxa, a Bayesian consensus tree for each gene region was compared to a combined reference tree of all the gene regions (Aguileta et al., 2008). The combined reference tree was constructed from all eight gene regions using a Bayesian approach as described in Aguileta et al. (2008), incorporating the corresponding nucleotide substitution model for each gene region. The topological difference in tree topologies were compared in the online program 'compare2trees' (http://www.mas.ncl.ac.uk/~ntmwn/compare2trees/index.html), based on an algorithm that compares the branches and partition of nodes between two trees and gives an overall topological congruence score (Nye et al., 2006). An overall score of 85% was selected as the cut-off point for a marker to be compatible with the other regions considered.

The five most informative gene regions were selected on the basis of the three criteria as stated before. A partition homogeneity test (PHT), with 1000 repeats, was performed on different arrangements of three and four gene region combinations in PAUP 4.0 to determine whether the sequences could be combined (Swofford, 2002). The combined tree, based on three or four gene regions, was constructed using MP, ML and BI analyses. Identical conditions, as those applied to the individual gene regions, were used in the different tree construction methods.

2.3. Development of SNP markers

2.3.1. Sequence data generation for SNP calling

SNP markers were developed from 454 sequence data for six species in the *C. fimbriata s.l.* species complex. These markers could subsequently be used to consider variation in the rest of the species in this species complex. This method was shown to be effective during a previous study on fungal species complexes (Pérez, 2010; Pérez et al., 2012). To generate sequence data, reduced representations of genome sequences were generated with a protocol similar to the initial steps of an AFLP protocol up to the pre-amplification step (Myburg and Remington, 2000) and sequenced with 454 pyrosequencing. Isolates included for sequencing were *C. fimbriata s.s*. (CMW 1547, 15049, 14799) (Table 1), *C. cacaofunesta* (CMW 14809, 15051, 14798), *C. platani* (CMW 14802, 23918, 23450) and a combined group consisting of *C. manginecans, C. acaciivora, C. mangicola* and *C. mangivora* (CMW 13851, 13852, 21123, 22563, 17568, 17570, 23623, 14797, 15052), which are referred to as the *C. manginecans* group in this study.

DNA digestion and adapter ligation: All genomic DNA was digested with a frequent and a rare cutting restriction enzyme, followed by ligation of restriction enzyme specific adapters to the DNA fragments. A master mix digestion reaction consisting of 1x R/L

buffer, 2 units *Eco*RI, 2 units *Mse*I and ddH_2O to a final volume of 10 µl per sample was used to digest 150 ng genomic DNA. The DNA solution was made up to a 20 µl reaction volume with ddH_2O and this was mixed with 10 µl of master mix. The reaction was incubated for 3 hrs at 37 °C, followed by 15 min at 65 °C. The ligation reaction was performed directly afterwards by combining 30 µl digested DNA with 10 µl ligation master mix (1x R/L buffer, 1mM ATP pH7, 1 pmol *Eco*RI- and 10 pmol *Mse*I adaptors, 1 unit T4 DNA ligase up to 40 µl final volume with ddH_2O) and was incubated for 3 hrs at 22 °C.

PCR preamplification: Preamplification reactions were performed with 5 µl of the ligated products. Total reaction volumes of 30 µl consisted of 1x PCR buffer (+1.5 mM Mg), 0.2 mM dNTP, 0.3 µM *Eco*RI+A primer, 0.3 µM *Mse*I+C primer, 0.6 units *Taq* polymerase (Expand *Taq*). The PCR program was as follow: 94 °C for 4 min, (94 °C for 30 s; 56 °C for 30 s; 72 °C for 1 min + 1 s/cycle extra) x 25 cycles, 72 °C for 2 min. Amplification smears were analysed by AGE on a 2% gel.

Pooling of isolates into species groups: Amplicons of all the isolates of the same species were pooled in four sample sets representing *C. fimbriata s.s*., *C. cacaofunesta*, *C. platani* and *C. manginecans*. Amplicons were pooled by combining 25 µl PCR product of each isolate and then precipitated with 0.1 vol of 10 M NaOAc and 2.5 vol absolute ethanol and incubated on ice for 10 min. Samples were then centrifuged and washed with 70% ethanol and the dried product resuspended in 30 µl H_2O.

Size separation: Sample sets were size-separated on a 1.2% agarose gel at 60 V for 1 hr. Bands in the size range of 150-450 bp were excised and purified using NucleoSpin® ExtractII kit (Macherey-Nagel, Germany).

DNA tagging and 454 sequencing: 454-adapters with identity tags for each of the four species were added to the DNA fragments by means of a PCR reaction. The primer sequences for the forward reaction consisted of the 454 adaptor A sequence plus the species specific sequence tag plus the *EcoRI* adaptor specific sequence (Sequence: 5'GCCTCCCTCGCGCCATCAG-NNNN-GACTGCGTACCAATTC3'). The reverse primer sequence consisted of the 454 adaptor B sequence, the species specific sequence tag, and the *MseI* adaptor specific sequence (Sequence: 5'GCCTTGCCAGCCCGCTCAG–NNNN-GATGAGTCCTGAGTAA3') (Pérez, 2010). The species specific identification tag for the *C. cacaofunesta* sample set was ATCG, CTAG for the *C.fimbriata s.s*. sample set, AGCT for the *C. platani* sample set and CAGT for the *C. manginecans* samples. The program for the PCR amplification was 94° 2 min, (94 °C 30 s, 60 °C 30 s, 72 °C 60 s) x 25 cycles, 72 °C 2 min.

PCR amplicons from all four sample sets were precipitated using 10 µl NaOAc (10 M) and 200 µl absolute ethanol (100%). DNA concentrations of all amplicons were adjusted to 30 ng/µl and 20 µl of each of the four samples were pooled to a final volume of 80 µl. The

product was sequenced using a Genome Sequencer 454 FLX (Roche, Inqaba Biotec, Pretoria, South Africa).

2.3.2. SNP identification, marker development and application to C. fimbriata s.l. species

Raw reads from the 454 sequencing were assembled using CLC Genomics workbench 5.0 (CLC Bio) and contigs were generated for orthologous regions, containing sequences from all four species. Contigs were constructed with the following parameters: similarity = 0.8, length fraction = 0.5, insertion and deletion costs = 3, mismatch cost = 2 and minimum contig length = 200. Each contig was investigated individually to determine the presence of SNPs that were conserved within a species but able to differentiate between species. Informative contigs were identified based on the number of SNPs present in a contig (minimum of 4 SNPs) and the number of species between which the SNPs could differentiate. For the purpose of this study, a nucleotide difference was considered as a species specific SNP only where it occurred in the majority of reads of at least one of the species. Regions that were present in the *C. fimbriata s.s.* genome more than once, based on BLAST results, and regions too variable for primer design were excluded.

For the selected contigs, primers were designed in the non-variable regions flanking the informative SNP region, using CLC Main workbench 6.0 (CLC Bio). Where SNPs were located inordinately close to the 3' or 5' end of the contig the *Ceratocystis fimbriata s.s.* genome (GenBank Accession number APWK01000000) was used to design primers located upstream or downstream from the SNP regions. Parameters were set for a maximum primer length of 22 bp, minimum primer length of 18 bp, maximum G/C content of 0.6, minimum G/C content of 0.4, maximum melting temperature of 58 °C and a minimum melting temperature of 48 °C. All primers pairs were designed with Tm temperatures as close to one another as possible in order to simplify multiplex PCR reactions. The designed primers (Table 3) were synthesized by Inqaba Biotech (Pretoria, South Africa).

Each of the designed primer sets were first tested on the four species used for 454 sequencing. This was to ensure PCR success and to confirm the presence of the SNPs as predicted using the 454 data. The regions that were most informative and amplified well in the majority of species were selected and amplified in three to five additional isolates of all other species included in this study (Table 1). PCR conditions were identical to those used for the single gene region amplification. The PCR program was as follow: 95° 5 min, (94 °C 30 s, 55 °C 30 s, 72 °C 90 s) x 38 cycles and a final extension of 72 °C for 10 min.

PCR products were amplified in 96-well PCR plates and purified by means of an ExoSAP method (Glenn and Schable, 2005). Purified PCR products were used for amplification of sequencing products performed in 96-well MicroAmp® reaction plates.

Sequence products were purified using ethanol precipitation (Glenn and Schable, 2005). The dried product was sequenced on an ABI PRISM™ 3500xl Autosequencer (Applied BioSystems, Foster City, California, USA).

2.3.3. Evaluation of SNP markers for species delimitation

Sequences from amplified SNP regions were assembled, analysed and edited in CLC Main workbench 6.0 (CLC Bio). Two different approaches were utilised to investigate the SNP variation between the isolates. First, the entire sequenced region for each SNP primer set was considered and a combined dataset of all SNP regions was generated. A cladogram was constructed from the dataset based on MP and BI analysis using settings similar to those used to analyse the single gene regions. Due to the presence of large indels, gaps were coded as a 5th character using FastGap v.1.0.7 (Borchsenius, 2007). The best model of evolution for each SNP region was determined using jModelTest 0.1 (Posada, 2008), and implemented in the BI analysis.

In the second approach, only the SNP sites (SNPs and indels) from each region were considered to construct a haplotype network. All of the SNPs, from the selected SNP regions, were combined into a single concatenated SNP haplotype for each of the isolates. This was constructed by aligning the sequences of all isolates for each SNP region separately in MEGA 5 using MUSCLE alignment (Edgar, 2004) and removing the constant sites. The variable sites from all SNP regions were then concatenated. Haplotypes were determined from the aligned SNP data in DnaSP v. 5 (Librado and Rozas, 2009). Gaps were included for haplotype construction. The identified haplotypes were used as input data to construct a haplotype network, based on a median-joining algorithm, on NETWORK v. 4.6.1.1 (Bandelt et al., 1999).

3. Results

3.1. Fungal isolates and DNA extraction

DNA extraction was successful for all of the 64 isolates considered and the concentrations ranged from 100 ng/µl to 1 500 ng/µl. Analysis of the DNA quality, using AGE, indicated that the DNA was of acceptable quality with low levels of RNA contamination and the DNA for all the samples was diluted to a concentration of 30 ng/µl for PCR amplification reactions.

3.2. Single copy phylogenetic gene regions

3.2.1. Primer design, PCR amplification and sequencing

PCR amplifications were successful for the majority of isolates for all gene regions, other than Tsr1 and βT 2. Tsr1 had a very low PCR success level and, even though the βT 2

region amplified successfully, sequencing of this region was problematic due to a 12-16 nucleotide poly-A repeat in the intron region between exon 1 and 2. These gene regions were thus excluded from further analyses. Expected amplicon sizes were obtained from PCR reactions of the eight other gene regions resulting in fragments of ± 550 bp for ITS, 600 bp for βT 1, 800 bp for EF 1-α, 600 bp for CAL, 600 bp for FG1093, 1200 bp for MS204 with MS204.cerato primer set and 800 bp with MS204.ceratoB primer set, 700 bp for Mcm7 and 1100 bp for RPBII.

Amplification of the MS204 region with the MS204.cerato primer set (Table 2) resulted in poor sequence results for 13 of the isolates but amplification with the redesigned primer set (MS204 ceratoB) produced high quality sequences. The FG1093 region contained a 12bp long poly-A repeat in some isolates, making downstream sequencing reactions challenging. Additionally, in isolates of *C. colombiana* and *C. ecuadoriana,* conflicting base calling was found at ten nucleotide sites in the sequence chromatogram. For this reason, PCR products were cloned and both FG1093 haplotypes obtained for each species were included in the phylogenetic analyses. Sequence data for the ITS region of the majority of isolates was of good quality and did not contain any ambiguous sites. However, for 15 of the isolates the data was unusable due to a long poly-A repeat in the region and GenBank data were used for these isolates in downstream analyses (Table 1). Reliable sequence data for the βT 1 and EF 1-α regions could also be obtained from GenBank for the majority of isolates (Table 1).

3.2.2. Sequence alignment and phylogenetic analysis

Sequence alignment of the gene regions showed that the FG1093, MS204 and RPBII gene regions contained the greatest number of variable sites, apart from ITS, while CAL and Mcm7 contained the least (Table 4). From the aligned sequence data of each gene region the total number of characters that could be used for phylogenetic analysis and the number of parsimony informative characters were determined (Table 5). Even though RPBII had the greatest number of characters (1129 bp), ITS and EF 1-α had the most parsimony informative characters (197 and 42 bp).

The number of parsimonious trees produced from MP analysis as well as the consistency index (CI) and retention index (RI) values obtained for each gene region are summarised in Table 5. jModelTest suggested different evolutionary models for each gene region (Table 5) which were incorporated accordingly in ML and BI analyses. The log likelihood value for the most likely phylogenetic tree produced from ML analysis is summarised in Table 5. The tree topology for each region was congruent in all three methods of tree construction (MP, ML and BI) and one of the most parsimonious trees was

selected for representation for each gene region (Fig. 1). Maximum parsimony and Maximum likelihood BS values and Bayesian PP values are indicated on the branches.

The ITS region was the only region that provided statistically significant support for all fifteen *C. fimbriata s.l.* species investigated (Fig. 1a). This was supported by either >70% bootstrap (BS) values from MP or ML analysis, or >95% posterior probabilities (PP) produced by Bayesian analysis or both. Phylogenetic trees produced from the regions Mcm7 (Fig. 1b) and EF 1-α (Fig. 1c) resulted in the lowest resolution, distinguishing only one (*C. diversiconidia*) and three (*C. colombiana, C. cacaofunesta* and *C. diversiconidia*) species respectively. The CAL gene region (Fig. 1d) provided support for four distinct species. FG1093, RPBII, βT 1 and MS204 phylogenetic trees could distinguish the greatest number of species. The FG1093 region enabled the recognition of five distinct species (Fig. 1e) and the RPBII region distinguished between four to seven species (Fig. 1f). High Bayesian PP values were obtained for seven monophyletic groups in RPBII but some BS values were only between 60% and 70%. The βT 1 region (Fig. 1g) provided support for between five to six distinct species and MS204 delineated eight species (Fig. 1h).

The species that were only resolved by the ITS gene region included *C. acaciivora, C. manginecans*, *C. mangicola, C. mangivora, C. curvata* and *C. eucalypticola* (Fig. 1a). Even though *C. mangicola* and *C. mangivora* could not be distinguished using the other gene regions, the two species were supported as a single clade, separate from all other *C. fimbriata s.l.* species, by βT 1 (87/63% BS and 95% PP) and RPBII (70% MP BS) (Fig. 1f and 1g). *Ceratocystis curvata* and *C. eucalypticola*, also not distinguished using other gene regions, had support values above 60% in the EF 1-α, FG1093 and MS204 gene genealogies (Fig. 1c, 1e and 1h).

The *C. cacaofunesta* isolates included in this study, grouped in two distinct clades in the phylogenetic trees of six of the gene regions. The βT 1, CAL, MS204, EF 1-α, RPBII and ITS gene genealogies consistently grouped the isolates CMW 14803, 14809 and 15051 in one clade (Clade A) and CMW 14798 and 26375 in another clade (Clade B). The βT 1, CAL, MS204 and EF 1-α regions only provided significant branch support for Clade A to be distinct, RPBII only supported Clade B and ITS significantly supported both clades. These isolates appear to represent distinct taxa.

3.2.3. Phylogenetic value of single genes and gene combinations

The gsi_T values generally showed monophyly for the same species that had high phylogenetic branch support values but could additionally suggest the degree of monophyly for the species that had low branch support values. The gsi_T values for all species and all eight gene regions are summarised in Table 6. The gsi analysis provided significantly

stronger support than traditional phylogenetic analysis for at least two species in every gene region (highlighted in Table 6).

The species *C. acaciivora* had low, but statistically significant gsi_T values from the βT 1 (gsi_T=0.27) and ITS (gsi_T=0.47) gene regions. This indicated incomplete lineage sorting. *Ceratocystis curvata* and *C. ecuadoriana* had low branch support values in traditional phylogenetic analyses. The gsi_T values were also low for *C. curvata* in the FG1093 and MS204 gene regions with values of 0.37 and 0.4 but higher in EF 1-α with gsi_T=0.7 (Table 6). For *C. eucalypticola* the gsi_T value in EF 1-α was not statistically significant but FG1093 and MS204 indicated significantly higher values of 0.71 and 0.87, respectively. The separation of the *C. cacaofunesta* isolates into two clades was supported by five gene regions (ITS, CAL, βT 1, MS204 and Mcm7) for Clade A with gsi_T values ranging from 0.753 to 0.976, and RPBII supported Clade B with a value of 0.923.

Comparison of tree topologies, based on Nye topological score, showed a minor level of conflict between each region as compared to the eight gene region combined tree (Fig. 2). Conflict was considered significant only where the specific branch was supported by >70% MP or ML bootstrap support or >95% Bayesian PP. The EF 1-α region showed the most conflict (80.4%), while CAL had the highest level of congruence (96.8%), followed by RPBII (94.4%) and βT 1 (91%) (Fig.1).

The most significant conflict in the EF 1-α genealogy was the grouping of *C. ecuadoriana* and *C. cacaofunesta* as closest relatives (Fig. 1c). FG1093 showed conflict in the tree topology for two specific clades indicating *C. ecuadoriana* and *C. colombiana* (rather than *C. ecuadoriana* and *C. neglecta*) to be closely related and also grouped *C. curvata* in this clade (Fig. 1e). The positions of the second haplotypes of *C. ecuadoriana* and *C. colombiana,* obtained from the cloned sequences, were located more accurately when compared to the tree topologies of other gene regions. The second haplotype in *C. ecuadoriana* grouped with *C. neglecta.* The second *C. colombiana* haplotype was also supported as a distinct species. For this reason, the cloned sequences were used for the combined gene region analyses. Mcm7 analyses showed that both *C. cacaofunesta* clades were most closely related to *C. colombiana* and *C. platani* (Fig. 1b). MS204 grouped *C. platani* together in a clade with *C. ecuadoriana* and *C. neglecta* (Fig. 1h). RPBII grouped Clade B of *C. cacaofunesta* and *C. colombiana* together but the internal node was not significantly supported (Fig. 1f).

The five gene regions that had the greatest potential to be applied as additional phylogenetic markers for delimitation of *C. fimbriata s.l.* species were identified based on the three criteria mentioned in 2.2.3 (Fig. 2). The RPBII, βT 1 and MS204 gene regions were identified as the three most informative gene regions to be used in combination, with CAL and FG1093 as additional gene regions. Various combinations of the five gene regions were

considered in order to identify the least number of genes required to distinguish all described species in the *C. fimbriata s.l.* complex. The partition homogeneity test (PHT) performed on the gene combinations had values as follow: 0.063 for the RPBII, βT 1 and MS204 gene combination, a value of 0.099 for the four gene combination where CAL was included and 0.01 for the four gene combination where FG1093 was included. Even though the latter value is low, it is still acceptable and the gene combination would not decrease the phylogenetic accuracy (Cunningham, 1997).

The three gene combination including RPBII, βT 1 and MS204 provided effective delineation between most species in *C. fimbriata s.l..* However, they could not be used to distinguish between *C. eucalypticola*, *C. curvata, C. acaciivora, C. manginecans, C. mangicola* and *C. mangivora* from each other. The four gene combination of RPBII, βT 1, MS204, including CAL did not add any significance to the phylogenetic tree and CAL was excluded from further analysis. Including the FG1093 region in the four gene combination increased the support values for *C. eucalypticola* and *C. curvata* as distinct species. In order to illustrate the resolution obtained from the four most informative gene regions this combination was selected to construct the tree for presentation (Fig. 3; Table 5).

3.3. Development of SNP markers

3.3.1. Sequence data generation for SNP calling

The 454 pyrosequencing produced 33 418 reads, with and average read length of 200 bases. The number of reads per species group ranged from 3000 (*C. manginecans*) to 13 000 (*C. platani*) and could be assembled into 867 contigs. About 65 contigs were discarded either because they consisted of a single read or because the contigs showed no variation between species. The contigs that were retained for further use, had a coverage ranging from 3 to 270 reads per contig.

3.3.2. SNP identification, marker development and application to C. fimbriata s.l. species

A total of 29 primer pairs were designed (Table 3), with an annealing temperature of 55 °C. The majority of the primer sets resulted in successful amplification in all 62 isolates, producing a single band with expected fragment sizes for the different primer sets (Table 3). SNP region 8 could not be amplified in *C. diversiconidia* and SNP region 12 could not be amplified in *C. mangivora,* despite several attempts at optimisation. These two regions were represented by N's for subsequent analyses. Sequencing was successful for the majority of samples after the first attempt but some required optimisation by increasing the annealing temperature to 57-58 °C in the amplification reaction for sequencing. Amplification and

sequencing for SNP regions 1, 2, 14 and 29 could not be optimized and were excluded from subsequent analyses.

3.3.4. Evaluation of SNP markers for species delimitation

The nine most informative SNP regions identified were SNP 8,10,12,15,16,18,24,26 and 32 (Table 7). The full length sequence data of all nine combined SNP regions resulted in 2518 characters after alignment, which were used for cladogram construction. A total of 23 indels were present in the alignment and these were coded and included in analyses, resulting in a total of 354 parsimony informative characters. The SNPs thus provided a greater number of informative characters than the 305 sites obtained from the seven single copy genes (Table 4 and 7). Since the nine most informative SNP regions were amplified and sequenced in all the *C. fimbriata s.l*. species investigated, additional SNPs not initially detected in the 454 sequence data, were identified.

Maximum parsimony analysis showed a CI of 0.770 and a RI of 0.960 and four most parsimonious trees were produced from the analysis. jModelTest indicated the HKY evolutionary model for SNP8 and SNP18, TIM1+G for SNP10, TrN+G for SNP12, TPM2uf+I for SNP15, TIM1+I for SNP16, TIM1+I+G for SNP24, TPM3uf+G for SNP32 and HKY+I for SNP26. The MP and BI approaches resulted in the same tree topology and both had a high level of branch support for each clade. One of the cladograms, based on MP analysis, was selected for presentation (Fig. 4A) and support values are shown above the branches. Strongly supported, distinct groupings could be seen in the cladogram for *C. colombiana*, *C. curvata, C. diversiconidia, C. ecuadoriana, C. eucalypticola*, *C. fimbriata*, *C. fimbriatomina*, *C. mangicola*, *C. mangivora*, *C. manginecans, C. neglecta, C. papillata* and *C. platani.* As was true for the protein coding gene regions, *C. cacaofunesta* isolates resided in two different clades. Most *C. acaciivora* isolates could not be distinguished from *C. manginecans* but isolates CMW 22563 and 22562 were located further away from the others, since they had nucleotide differences in two SNP regions at three SNP sites (Table 7).

Analyses based only on the SNP sites resulted in a total of 360 characters, consisting of 181 SNPs and 23 indels (Table 7). The SNP data provided fixed SNP differences between all of the species except for some of the *C. acaciivora* isolates that were identical to *C. manginecans*. A total of 29 haplotypes were revealed in the SNP data. The network analysis combined some of the haplotypes at a single node, thus resulting in 25 external nodes (Fig. 4B). Most isolates of a single species showed minor levels of variation. However, a higher level of variation was clear within the *C. platani* isolates. The isolates from USA represented two separate haplotypes and the isolates from Europe (Greece and Switzerland) represented a third haplotype. Despite the slight variation observed within a species, 13 of

the species defined using ITS sequence data, were clearly separated in the network and the relatedness between species correlated with what is seen in the cladogram.

4. Discussion

The delimitation of many species in the *C. fimbriata s.l*. complex has been strongly reliant on sequence data from the ITS region. In this study seven gene regions and nine SNP markers were used to reconsider the boundaries of 15 of these taxa described in the Latin American clade of the species complex. The resolution of these markers was compared with that provided by the ITS, βT 1 and EF 1-α gene regions (Mbenoun et al., 2013; Tarigan et al., 2011; Van Wyk et al., 2011b). While none of these alternative regions were able to support all the species defined in ITS phylogenies, a combination of four gene regions (βT 1, RPBII, MS204 and FG1093) provided significant support for the delineation of most species. The SNP markers developed were more informative than the gene regions, making it possible to clearly distinguish between all but two species and also revealed the presence of at least one cryptic species.

Of the seven gene regions evaluated in this study, RPBII and MS204 distinguished the greatest number of species in the *C. fimbriata s.l.* complex. These two gene regions were as informative as the βT 1 region and significantly more useful than EF 1-α (Fig. 2). A combination of RPBII, MS204 and βT 1 provided the best resolution and supported the delineation of 11 of the 15 species in the *C. fimbriata s.l*. complex. Even though the FG1093 region was not as informative as the aforementioned regions, it provided support for differentiation of one additional species, *C. curvata*, not distinguished by the combined three-gene phylogeny (Fig 3). The apparent existence of two copies of the FG1093 region in some species requires careful consideration of the sequence quality obtained from amplicons before it is used in analyses.

Four of the gene regions, CAL, βT 2, Mcm7 and Tsr1, considered in this study were not useful for species delineation. Even though the tree topology for the CAL gene region was highly congruent with that of other gene regions (Fig. 2), it did not increase the phylogenetic resolution in the combined species tree. The Tsr1 gene region had a low PCR success rate and the Mcm7 region provided negligible phylogenetic resolution. Such technical difficulties for the use of these gene regions have been reported previously (Raja et al., 2011; Schmitt et al., 2009; Tretter et al., 2013). The βT 2 region should also be avoided in phylogenetic studies in the *C. fimbriata s.l.* complex as it is difficult to obtain sequences for this gene region.

The SNP markers developed in this study provide a powerful diagnostic tool to distinguish between species in the *C. fimbriata s.l.* complex. The markers provided support for 13 distinct species and are thus more informative than the protein coding gene regions

investigated. The SNPs provided support for *C. mangicola* and *C. mangivora* as distinct species and provided a higher level of support for the monophyly of *C. curvata* and *C. eucalypticola* than the protein coding gene regions. The fact that all isolates for a specific species contained fixed SNPs and consistently grouped together in the network and cladogram constructed improves confidence in the SNP regions.

The collective data derived from the evaluated gene regions and SNP markers, provided new insights to the boundaries of three species in the *C. fimbriata s.l.* complex. The *C. cacaofunesta* isolates were differentiated into two separate groups, with significant branch support values (Fig. 1) obtained from six of the gene regions (Table 4), as well as the combined SNP markers (Fig. 4). ITS sequence data and mating studies, initially used to identify this species (Engelbrecht and Harrington, 2005; Ferreira et al., 2010), suggested that there are two distinct lineages (Ecuador and Brazil/Costa Rica/Colombia) for South American isolates of *C. fimbriata s.l.* from cacao. Results from this study and all other currently available data strongly suggest that the isolates CMW 14798 and 26375 of *C. cacaofunesta* represent an undescribed species in the *C. fimbriata s.l.* complex. *Ceratocystis acaciivora* and *C. manginecans* could not be distinguished using any of the seven markers considered in this study and they most likely represent the same species. These species can only be distinguished using the ITS region, which has clearly emerged as unusable since two ITS types can occur in a single species such as *C. manginecans* (Al Adawi et al., 2013; Harrington et al., 2014; Naidoo et al., 2013).

Results of this study have shown that most species described in the *C. fimbriata s.l.* complex can be distinguished based on the combined sequence data of the βT 1, RPBII, MS204 and FG1093 gene regions. This is supported by the suite of SNP markers, which provided data congruent with the gene region phylogenies. As taxonomists are not in agreement on the best criterion for species delineation, alternative approaches to the GCPSR could propose less species in the complex, as recently suggested by Harrington et al. (2014). Species delimitation was based on a single gene region, the MAT gene, and intersterility tests between individual isolates in the *C. fimbriata s.l.* complex. However, phylogenetic analyses from the current study were based on a greater number of molecular markers which suggests the MAT gene is less informative for recently emerging taxa. The interfertility observed between some of the species should also be interpreted with caution as phylogenetically distinct species have been able to mate in laboratory conditions in various studies (Harrington and McNew, 1998; Hibbett and Donoghue, 1996; O'Donnell et al., 2004). There is a strong case to defend the species boundaries presented in this study and accepting these will provide opportunities to better understand the biology and ecology of an important group of fungi.

5. References

Aguileta, G., Marthey, S., Chiapello, H., Lebrun, M.H., Rodolphe, F., Fournier, E., Gendrault-Jacquemard, A., Giraud, T., 2008. Assessing the Performance of Single-Copy Genes for Recovering Robust Phylogenies. Syst. Biol. 57, 613-627.

Al Adawi, A.O., Barnes, I., Khan, I.A., Al Subhi, A.M., Al Jahwari, A.A., Deadman, M.L., Wingfield, B.D., Wingfield, M.J., 2013. *Ceratocystis manginecans* associated with a serious wilt disease of two native legume trees in Oman and Pakistan. Australas. Plant Pathol. 42, 179-193.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37-48.

Barnes, I., Roux, J., Wingfield, B.D., Dudzinski, M.J., Old, K.M., Wingfield, M.J., 2003. *Ceratocystis pirilliformis*, a new species from *Eucalyptus nitens* in Australia. Mycologia 95, 865-871.

Borchsenius, F., 2007. FastGap 1.0.7. Software distributed by the author (http://192.38.46.42/aubot/fb/FastGap_home.htm).

Bridge, P.D., Spooner, B.M., Roberts, P.J., 2005. The Impact of Molecular Data in Fungal Systematics. In: Callow, J.A. (Ed.), Advances in Botanical Research. Academic Press, pp. 33-67.

Cai, L., Giraud, T., Zhang, N., Begerow, D., Cai, G., Shivas, R., 2011. The evolution of species concepts and species recognition criteria in plant pathogenic fungi. Fungal Divers. 50, 121-133.

Chouvarine, P., Cooksey, A.M., McCarthy, F.M., Ray, D.A., Baldwin, B.S., Burgess, S.C., Peterson, D.G., 2012. Transcriptome-Based Differentiation of Closely-Related *Miscanthus* Lines. PLoS ONE 7, e29850.

Cummings, M.P., Neel, M.C., Shaw, K.L., 2008. A genealogical approach to quantifying lineage divergence. Evolution 62, 2411-2422.

Cunningham, C.W., 1997. Can three incongruence tests predict when data should be combined? Molecular Biology and Evolution 14, 733-740.

Duong, T.A., De Beer, Z.W., Wingfield, B.D., Wingfield, M.J., 2012. Phylogeny and taxonomy of species in the *Grosmannia serpens* complex. Mycologia 104, 715-732.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797.

Engelbrecht, C.J.B., Harrington, T.C., 2005. Intersterility, morphology and taxonomy of *Ceratocystis fimbriata* on sweet potato, cacao and sycamore. Mycologia 97, 57-69.

Fakhrai-Rad, H., Pourmand, N., Ronaghi, M., 2002. Pyrosequencing™: An accurate detection platform for single nucleotide polymorphisms. Hum. Mutat. 19, 479-485.

Ferreira, E.M., Harrington, T.C., Thorpe, D.J., Alfenas, A.C., 2010. Genetic diversity and interfertility among highly differentiated populations of *Ceratocystis fimbriata* in Brazil. Plant Pathol. 59, 721-735.

Garvin, M.R., Marcotte, R.W., Palof, K.J., Riley, R.J., Kamin, L.M., Gharrett, A.J., 2011. Diagnostic Single-Nucleotide Polymorphisms Identify Pacific Ocean Perch and Delineate Blackspotted and Rougheye Rockfish. Trans. Am. Fish. Soc. 140, 984-988.

Giraud, T., Refrégier, G., Le Gac, M., De Vienne, D.M., Hood, M.E., 2008. Speciation in fungi. Fungal Genet. Biol. 45, 791-802.

Glass, N.L., Donaldson, G.C., 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. Appl. Environ. Microbiol. 61, 1323-1330.

Glenn, T.C., Schable, N.A., 2005. Isolating Microsatellite DNA Loci. In: Zimmer, E.A., Roalson, E.H. (Eds.), Methods in Enzymology. Academic Press, San Diego, CA, pp. 202-222.

Goodwin, S.B., Drenth, A., Fry, W.E., 1992. Cloning and genetic analysis of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. Curr. Genet. 22, 107-115.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696-704.

Halsted, B.D., 1890. Some fungous disease of the sweet potato. Agricultural College Experiment Station Bulletin 76, 1-32.

Halsted, B.D., Fairchild, D.G., 1891. Sweet-Potato black rot. J. Mycol. 7, 1-11.

Harrington, T.C., 2000. Host specialization and speciation in the American wilt pathogen *Ceratocystis fimbriata*. Fitopatologia Brasileira 25S, 262-263.

Harrington, T.C., Kazmi, M.R., Al-Sadi, A.M., Ismail, S.I., 2014. Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbirata sensu stricto*. Mycologia In press. doi: 10.3852/13-189.

Harrington, T.C., McNew, D.L., 1998. Partial Interfertility Among the *Ceratocystis* Species on Conifers. Fungal Genet. Biol. 25, 44-53.

Hibbett, D.S., Donoghue, M.J., 1996. Implications of Phylogenetic Studies for Conservation of Genetic Diversity in Shiitake Mushrooms. Conserv. Biol. 10, 1321-1327.

Jacobs, K., Bergdahl, D.R., Wingfield, M.J., Halik, S., Seifert, K.A., Bright, D.E., Wingfield, B.D., 2004. *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. Mycol. Res. 108, 411-418.

Jones, J.C., Fan, S., Franchini, P., Schartl, M., Meyer, A., 2013. The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. Mol. Ecol. 22, 2986-3001.

Kamgan, N.G., Wingfield, M.J., Mohammed, C., Carnegie, A.J., Pegg, G.S., Roux, J., 2012. *Ceratocystis* species, including two new species associated with nitidulid beetles, on eucalypts in Australia. Anton. Leeuw. 101, 217-241.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511-518.

Kile, G.A., Harrington, T.C., Yuan, Z.Q., Dudzinski, M.J., Old, K.M., 1996. *Ceratocystis eucalypti* sp. nov., a vascular stain fungus from eucalypts in Australia. Mycol. Res. 100, 571-579.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452.

Lindner, D.L., Banik, M.T., 2011. Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. Mycologia 103, 731-740.

Marthey, S., Aguileta, G., Rodolphe, F., Gendrault, A., Giraud, T., Fournier, E., Lopez-Villavicencio, M., Gautier, A., Lebrun, M.-H., Chiapello, H., 2008. FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics doi: 10.1186/1471-2105-9-456.

Mbenoun, M., Wingfield, M.J., Begoude Boyogueno, A., Wingfield, B.D., Roux, J., 2013. Molecular phylogenetic analyses reveal three new *Ceratocystis* species and provide evidence for geographic differentiation of the genus in Africa. Mycol. Prog. doi: 10.1007/s11557-013-0907-5.

Myburg, A.A., Remington, D.L., 2000. Protocol for high-throughput AFLP analysis using LI-COR IR2 automated sequencer. North Carolina State Forest Biotechnology AFLP protocol. N C State Univ For Dep, Raleigh N.C.

Naidoo, K., Steenkamp, E.T., Coetzee, M.P.A., Wingfield, M.J., Wingfield, B.D., 2013. Concerted Evolution in the Ribosomal RNA Cistron. PLoS ONE 8, e59355. doi: 59310.51371/journal.pone.0059355.

Nye, T.M.W., Liò, P., Gilks, W.R., 2006. A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. Bioinformatics 22, 117-119.

O'Donnell, K., Cigelnik, E., 1997. Two Divergent Intragenomic rDNA ITS2 Types within a Monophyletic Lineage of the Fungus *Fusarium* Are Nonorthologous. Mol. Phylogenet. Evol. 7, 103-116.

O'Donnell, K., Nirenberg, H., Aoki, T., Cigelnik, E., 2000. A multigene phylogeny of the *Gibberella fujikuroi* species complex: Detection of additional phylogenetically distinct species. Mycoscience 41, 61-78.

O'Donnell, K., Ward, T.J., Geiser, D.M., Corby Kistler, H., Aoki, T., 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. Fungal Genet. Biol. 41, 600-623.

Page, R.D.M., 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357-358.

Pavlic, D., 2009. Taxonomy and population diversity of Botryosphaeriaceae associated with woody hosts in South Africa and Western Australia. Department of Microbiology and Plant Pathology. University of Pretoria, Pretoria.

Pérez, G., 2010. Chapter 6: Inter-Specific SNPs for cryptic species delimitation and phylogenetic reconstructions in closely related species. Global genetic diversity of the *Eucalyptus* leaf pathogen *Teratosphaeria nubilosa* species complex in native forests and sommercial plantations, Microbiology and Plant Pathology, University of Pretoria, Pretoria pp. 185 - 205.

Pérez, G., Slippers, B., Wingfield, M.J., Wingfield, B.D., Carnegie, A.J., Burgess, T.I., 2012. Cryptic species, native populations and biological invasions by a eucalypt forest pathogen. Mol. Ecol. 21, 4452-4471.

Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253-1256.

Raja, A.H., Schoch, C.L., Hustad, V.P., Shearer, C.A., Miller, A.N., 2011. Testing the phylogenetic utility of *Mcm7* in the Ascomycota. MycoKeys 1, 63-94.

Rambaut, A., Drummond, A.J., 2009. Tracer v1.5. Institute of Evolutionary Biology.

Rintoul, T.L., Eggertson, Q.A., Lévesque, C.A., 2012. Multigene phylogenetic analyses to delimit new species in fungal plant pathogens. In: Bolton, M.D., Thomma, B.P.H.J. (Eds.), Plant Fungal Pathogens: Methods and Protocols. Humana Press, pp. 549-569.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572-1574.

Roux, J., Van Wyk, M., Hatting, H., Wingfield, M.J., 2004. *Ceratocystis* species infecting stem wounds on *Eucalyptus grandis* in South Africa. Plant Pathol. 53, 414-421.

Sakalidis, M.L., Hardy, G.E.S.J., Burgess, T.I., 2011. Use of the Genealogical Sorting Index (GSI) to delineate species boundaries in the *Neofusicoccum parvum–Neofusicoccum ribis* species complex. Mol. Phylogenet. Evol. 60, 333-344.

Schmitt, I., Crespo, A., Divakar, P.K., Fankhauser, J.D., Herman-Sackett, E., Kalb, K., Nelsen, M.P., Nelson, N.A., Rivas-Plata, E., Shimp, A.D., Widhelm, T., Lumbsch, H.T., 2009. New primers for promising single-copy genes in fungal phylogenetics and systematics. Persoonia 23, 35-40.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. 109, 6241-6246.

Seifert, K.A., 2009. Progress towards DNA barcoding of fungi. Mol. Ecol. Resour. 9, 83-89.

Spatafora, J.W., 2005. Assembling the fungal tree of life (AFTOL). Mycol. Res. 109, 755-756.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer Associates, Sunderland, MA.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28, 2731-2739.

Taole, M., Burgess, T., Gryzenhout, M., Wingfield, B.D., Wingfield, M.J., 2012. DNA sequence incongruence and inconsistent morphology obscure species boundaries in the *Teratosphaeria suttonii* species complex. Mycoscience 53, 270-283.

Tarigan, M., Roux, J., Van Wyk, M., Tjahjono, B., Wingfield, M.J., 2011. A new wilt and die-back disease of *Acacia mangium* associated with *Ceratocystis manginecans* and *C. acaciivora sp*. nov. in Indonesia. SA J. Bot. 77, 292-304.

Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S., Fisher, M.C., 2000. Phylogenetic Species Recognition and Species Concepts in Fungi. Fungal Genet. Biol. 31, 21-32.

Tretter, E.D., Johnson, E.M., Wang, Y., Kandel, P., White, M.M., 2013. Examining new phylogenetic markers to uncover the evolutionary history of early-diverging fungi: comparing MCM7, TSR1 and rRNA genes for single- and multi-gene analyses of the *Kickxellomycotina*. Persoonia 30, 106–125.

Upadhyay, H.P., 1981. A monograph of *Ceratocystis* and *Ceratocystiopsis*. University of Georgia Press, Athens.

Van Wyk, M., Roux, J., Kamgan, N.G., Wingfield, B.D., Wingfield, M.J., 2012. *Ceratocystis eucalypticola sp. nov*. from *Eucalyptus* in South Africa and comparison to global isolates from this tree. IMA Fungus 3, 45-58.

Van Wyk, M., Wingfield, B.D., Al-Adawi, A.O., Rossetto, C.J., Ito, M.F., Wingfield, M.J., 2011a. Two new *Ceratocystis* species associated with mango disease in Brazil. Mycotaxon 117, 381–404.

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

Van Wyk, M., Wingfield, B.D., Wingfield, M.J., 2011b. Four new *Ceratocystis* spp. associated with wounds on *Eucalyptus, Schizolobium* and *Terminalia* trees in Ecuador. Fungal Divers. 46, 111-131.

Walker, D.M., Castlebury, L., Rossman, A., Mejía, L., White, J., 2012a. Phylogeny and taxonomy of *Ophiognomonia* (*Gnomoniaceae*; Diaporthales), including twenty-five new species in this highly diverse genus. Fungal Divers. 57, 85-147.

Walker, D.M., Castlebury, L.A., Rossman, A.Y., White, J.F., 2012b. New molecular markers for fungal phylogenetics: Two genes for species-level systematics in the Sordariomycetes (Ascomycota). Mol. Phylogenet. Evol. 64, 500-512.

Wang, X.Q., Zhao, L., Eaton, D.A.R., Li, D.Z., Guo, Z.H., 2013. Identification of SNP markers for inferring phylogeny in temperate bamboos (*Poaceae: Bambusoideae*) using RAD sequencing. Mol. Ecol. Resour. 13, 938-945.

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

Wilken, P.M., Steenkamp, E.T., Wingfield, M.J., De Beer, Z.W., Wingfield, B.D., 2014. Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. IMA Fungus 4, 357–358.

Wingfield, B.D., Van Wyk, M., Roos, H., Wingfield, M.J., 2013. *Ceratocystis*: Emerging evidence for discrete generic boundries. In: Seifert, K.A., De Beer, W., Wingfield, M.J. (Eds.), Ophiostomatoid Fungi: Expanding Frontiers. CBS-KNAW Fungal Biodiversity Centre, AD Utrecht, The Netherlands, pp. 57-64.

Wingfield, M.J., De Beer, C., Visser, C., Wingfield, B.D., 1996. A New *Ceratocystis* Species Defined Using Morphological and Ribosomal DNA Sequence Comparisons. Syst. Appl. Microbiol. 19, 191-202.

6. Tables and figures

Table 1

Details of the *Ceratocystis fimbriata s.l.* isolates used in this study, including GenBank accession numbers of the sequences generated in this study.

Species[a]	Host	Country	CMW no.[c]	GenBank accession no.[b]							
				Cal	FG1093	Mcm7	MS204	RPBII	ITS	βT 1	EF 1-α
C. acaciivora	*Acaciivora mangium*	Indonesia	22564	KJ601633	KJ601669	KJ601525	KJ601561	KJ601597	*EU588657.1*	*EU588637.1*	EU588647
C. acaciivora[$]	*A. mangium*	Indonesia	22563	KJ601632	KJ601668	KJ601524	KJ601560	KJ601596	*EU588656.1*	*EU588636.1*	EU588646
*C. acaciivora**	*A. mangium*	Indonesia	22562	-	-	-	-	-	-	-	-
*C. acaciivora**	*A. mangium*	Indonesia	22595	-	-	-	-	-	-	-	-
*C. acaciivora**	*A. mangium*	Indonesia	22621	-	-	-	-	-	-	-	-
C. cacaofunesta	*Theobroma cacao*	Ecuador	14809	KJ601635	KJ601671	KJ601527	KJ601563	KJ601599	*DQ520637.1*	KJ601509	KJ601516
C. cacaofunesta	*T. cacao*	Costa Rica	14798	KJ601637	KJ601673	KJ601529	KJ601565	KJ601601	*AY157952.1*	KJ601511	KJ601518
C. cacaofunesta	*T. cacao*	Ecuador	14803	KJ601634	KJ601670	KJ601526	KJ601562	KJ601598	*AY157950.1*	KJ631108	KJ601515
C. cacaofunesta	*T. cacao*	Costa Rica	15051	KJ601636	KJ601672	KJ601528	KJ601564	KJ601600	*AY157951.1*	KJ601510	KJ601517
C. cacaofunesta[$]	*T. cacao*	Brazil	26375	KJ601638	KJ601674	KJ601530	KJ601566	KJ601602	*AY157953.1*	KJ601512	KJ601519
C. colombiana[$]	*Coffee arabica*	Colombia	5751	KJ601639	KJ601675/ KJ601676	KJ601531	KJ601567	KJ601603	*AY177233.1*	*AY177225.1*	*EU241493.1*
C. colombiana	*C. arabica*	Colombia	5761	KJ601640	KJ601677/ KJ601678	KJ601532	KJ601568	KJ601604	*AY177234.1*	*AY177224.1*	*EU241492*
*C. colombiana**	Soil in coffee plantation	Colombia	9565	-	-	-	-	-	-	-	-
*C. colombiana**	*Schizolobium parahybum*	Colombia	11280	-	-	-	-	-	-	-	-
C. curvata	*Eucalyptus deglupta*	Colombia	22435	KJ601641	KJ601679	KJ601533	KJ601569	KJ601605	*FJ151437.1*	*FJ151449.1*	*FJ151471.1*
C. curvata[$]	*E. deglupta*	Colombia	22442	KJ601642	KJ601680	KJ601534	KJ601570	KJ601606	*FJ151436.1*	*FJ151448.1*	*FJ151470.1*
*C. curvata**	*E. deglupta*	Colombia	22433	-	-	-	-	-	-	-	-
*C. curvata**	*E. deglupta*	Colombia	22432	-	-	-	-	-	-	-	-
C. diversiconidia[$]	*Terminalia ivorensis*	Colombia	22445	KJ601643	KJ601683	KJ601535	KJ601571	KJ601607	*FJ151440.1*	*FJ151452.1*	*FJ151474.1*
C. diversiconidia	*T.ivorensis*	Colombia	22448	KJ601644	KJ601684	KJ601536	KJ601572	KJ601608	*FJ151441*.1	FJ151453	*FJ151475*
*C. diversiconidia**	*T. ivorensis*	Colombia	22446	-	-	-	-	-	-	-	-

C. ecuadoriana[$]	*E. deglupta*	Colombia	22092	KJ601645	KJ601685/ KJ601686	KJ601537	KJ601573	KJ601609	*FJ151432.1*	*FJ151444.1*	FJ151466
C. ecuadoriana	*E. deglupta*	Colombia	22097	KJ601646	KJ601687/ KJ601688	KJ601538	KJ601574	KJ601610	FJ151434	*FJ151446.1*	FJ151468
*C. ecuadoriana**	*E. deglupta*	Colombia	22093	-	-	-	-	-	-	-	-
*C. ecuadoriana**	*E. deglupta*	Colombia	22405	-	-	-	-	-	-	-	-
C. eucalypticola	*Eucalyptus grandis*	South Africa	10000	KJ601647	KJ601689	KJ601539	KJ601575	KJ601611	*FJ236722.1*	*FJ236782.1*	*FJ236752*
C. eucalypticola [$]	*E. grandis*	South Africa	11536	KJ601648	KJ601690	KJ601540	KJ601576	KJ601612	*FJ236723.1*	*FJ236783.1*	*FJ236753.1*
*C. eucalypticola**	*E. grandis*	South Africa	9998	-	-	-	-	-	-	-	-
*C. eucalypticola**	*E. grandis*	South Africa	12663	-	-	-	-	-	-	-	-
C. fimbriata	*Imopoea batatas*	Papua New Guinea	1547	KJ601649	KJ601691	KJ601541	KJ601577	KJ601613	AF264904	*EF070443.1*	*EF070395.1*
C. fimbriata[$]	*I. batatas*	USA, California	14799	KJ601650	KJ601692	KJ601542	KJ601578	KJ601614	KC493160	KF302689	KJ631109
*C. fimbriata**	*I. batatas*	USA	15049	-	-	-	-	-	-	-	-
C. fimbriatomima[$]	*Eucalyptus sp.*	Venezuela	24174	KJ601651	KJ601693	KJ601543	KJ601579	KJ601615	*EF190963.1*	*EF190951.1*	*EF190957*
C. fimbriatomima	*Eucalyptus sp.*	Venezuela	24176	KJ601652	KJ601694	KJ601544	KJ601580	KJ601616	*EF190964.1*	*EF190952.1*	*EF190958*
C. fimbriatomima	*Eucalyptus sp.*	Venezuela	24377	KJ601653	KJ601695	KJ601545	KJ601581	KJ601617	*EF190966.1*	*EF190954.1*	KJ601520
*C. fimbriatomima**	*Eucalyptus sp.*	Venezuela	24378	-	-	-	-	-	-	-	-
C. mangicola	*Mangifera indica*	Brazil	28907	KJ601655	KJ601697	KJ601547	KJ601583	KJ601619	*FJ200257.1*	FJ200270	*FJ200283*
C. mangicola[$]	*M. indica*	Brazil	14797	KJ601654	KJ601696	KJ601546	KJ601582	KJ601618	*AY953382.1*	*EF433307.1*	*EF433316*
*C. mangicola**	*M. indica*	Brazil	28908	-	-	-	-	-	-	-	-
*C. mangicola**	*M. indica*	Brazil	28914	-	-	-	-	-	-	-	-
C. manginecans[$]	*M. indica*	Oman	13851	KJ601656	KJ601698	KJ601548	KJ601584	KJ601620	AY953383	*EF433308.1*	*EF433317*
C. manginecans	*Hypocryphalus mangiferae*	Oman	13852	KJ601657	KJ601699	KJ601549	KJ601585	KJ601621	AY953384	*EF433309.1*	*EF433318*
*C. manginecans**	*M. indica*	Oman	15314	-	-	-	-	-	-	-	-
*C. manginecans**	*M. indica*	Pakistan	23634	-	-	-	-	-	-	-	-
*C. manginecans**	*A. crassicarpa*	Indonesia	21123	-	-	-	-	-	-	-	-
*C. manginecans**	*M. indica*	Oman	13854	-	-	-	-	-	-	-	-
C. mangivora[$]	*M. indica*	Brazil	15052	KJ601658	KJ601700	KJ601550	KJ601586	KJ601622	*EF433298*.1	EF433306	EF433315
C. mangivora	*M. indica*	Brazil	27305	KJ601659	KJ601701	KJ601551	KJ601587	KJ601623	FJ200262	FJ200275	FJ200288

*C. mangivora**	*M. indica*	Brazil	27304	-	-	-	-	-	-	-	-
*C. mangivora**	*M. indica*	Brazil	27307	-	-	-	-	-	-	-	-
*C. mangivora**	*M. indica*	Brazil	28909	-	-	-	-	-	-	-	-
C. neglecta[$]	*E. grandis*	Colombia	17808	KJ601660	KJ601681	KJ601552	KJ601588	KJ601624	*EF127990.1*	*EU881898.1*	*EU881904*
C. neglecta	*E. grandis*	Colombia	18194	KJ601661	KJ601682	KJ601553	KJ601589	KJ601625	*EF127991.1*	*EU881899.1*	*EU881905*
C. papillata[$]	*Citrus limon*	Colombia	8856	KJ601662	KJ601702	KJ601554	KJ601590	KJ601626	*AY233867.1*	AY233874	*EU241484*
C. papillata	*C. arabica*	Colombia	10844	KJ601663	KJ601703	KJ601555	KJ601591	KJ601627	*AY177238.1*	*AY177229.1*	*EU241481*
*C. papillata**	*S. parahybum*	Colombia	28662	-	-	-	-	-	-	-	-
C. pirilliformis[$]	*Eucalyptus nitens*	Australia	6579	KJ601666	KJ601706	KJ601558	KJ601594	KJ601630	*AF427105.1*	*DQ371653.1*	*AY528983*
C. pirilliformis	*E.nitens*	Australia	6583	KJ601667	KJ601707	KJ601559	KJ601595	KJ601631	KJ601523	KJ601514	KJ601522
C. platani[$]	*Platanus occidentalis*	USA	14802	KJ601664	KJ601704	KJ601556	KJ601592	KJ601628	*DQ520630.1*	*EF070425.1*	*EF070396*
C. platani	*Platanus orientalis*	Greece	23450	KJ601665	KJ601705	KJ601557	KJ601593	KJ601629	KJ631107	KJ601513	KJ601521
*C. platani**	*Platanus* sp.	Switzerland	1896	-	-	-	-	-	-	-	-
*C. platani**	*P. occidentalis*	USA	26380	-	-	-	-	-	-	-	-
*C. platani**	*P. occidentalis*	Greece	23918	-	-	-	-	-	-	-	-
*C. platani**	*Platanus* sp.	Greece	23451	-	-	-	-	-	-	-	-

a All isolates listed, except *C. pirilliformis*, were used in the SNP study but those labelled with * were not considered in the single copy protein coding gene analysis.
b GenBank data generated during previous studies are in italic.
c CMW culture collection numbers refer to cultures obtained from the Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa. The type strains are indicated by $.

Table 2

Primers used for amplification of single copy gene regions investigated in this study.

Gene region	Primer name[a]	Direction	Sequence (5'-3')[b]	$MgCl_2$ (mM)	Tm (°C)	Successful amplification	Reference
βt 2	T1d	Forward	A**C**C ATG CG**Y** GA**R** AT**Y** GTA AGT	1	59	Yes	Duong et al., in prep.
	Bt1d	Internal Forward	AGG GTA ACC ARA TYG GTG C				
	Bt2d	Reverse	GTA YTG YCC CTT GGC CCA GTT G				
CAL	CAL2F	Forward	GAC AAG GAY GGY GAT GGT	0.5	64	Yes	Duong et al., 2012
	CAL2R2	Reverse	CTT CTC GCC RAT SGA SGT CAT				
FG1093	**FG1093F.cerato**	Forward	GCG CCA CA**A** CAA G**T**C **G**CA C**GT**	1	70	Yes	
	FG1093R.cerato	Reverse	TTC T**C**C GCT TG**C** CCT T**G**T CRS				
Mcm7	*Mcm7* -709	Forward	ACI MGI GTI TCV GAY GTH AAR CC	2.5	55	Yes	Schmitt et al., 2009
	Mcm7 -1348	Reverse	GAY TTD GCI ACI CCI GGR TCW CCC AT				
MS204	**MS204F.cerato**	Forward	AAG GGC ACC CT**C** GAG GGC CAC	1	56	Half of isolates	
	MS204R.cerato	Reverse	GAT GGT **R**AC GG**T** GTT GAT GTA				
	MS204F.ceratoB	Forward	GGC TGA GCA GCT GAT CCT T	1	56	Yes	
	MS204R.ceratoB	Reverse	ATG TCC GGG TAG TGT TAC CG				
RPBII	**RPB2-5Fb**	Forward	GAY GAY **C**G**T** GAT CA**C** TTY GG	0.5	61	Yes	Duong et al., in prep.
	RPB2-7Rb	Reverse	CCC AT**R** GCY TG**Y** TT**R** CCC AT				
Tsr1	*Tsr1* -1453for	Forward	GAR TTC CCI GAY GAR ATY GAR CT	1-2.5	54	Very few isolates	Schmitt et al., 2009
	Tsr1 -2308rev	Reverse	CTT RAA RTA ICC RTG IGT ICC				
	Tsr1F.cerato	Forward	GAR TT**Y** CC**N** GAY GAR ATY GAR CT	1-2.5	54	No	
	Tsr1R.cerato	Reverse	**Y**TT RAA RTA **N**CC RTG **N**GT **N**CC				
	Tsr1F.new	Forward	GAY GAY CAY CAY TAC TTC TC				

a Primers in bold were designed in this study.
b Bold, underlined nucleotides in the primer sequences indicates where primers were modified from the original primers.

Table 3

The twenty five primers designed to amplify SNP regions in the *Ceratocystis fimbriata sensu lato* isolates.

Primer number*	Forward sequence (5'-3')	Reverse sequence (5'-3')	Expected size (bp)	Tm (°C)	GenBank acc. no.
5	CATCCGCACAAAGTTTAACA	TACGGCGCTTCTTTTTCA	291	56	n/a
7	ATAGATATGCTCCACAAGAC	CAACACATAGCAACCAACAA	293	56	n/a
8	CATCACGGGAGAATCAAA	TATAGCACCACACAGCCA	309	55	KJ601500
9	GTGTTGCCGAAAAAATGTCT	GCGATGGTGAGTGAGATT	484	56	n/a
10	CAATGTTGCTCATCCAGT	CCGCCACACAAGATTTCA	266	55	KJ601501
11	GAGATGGGCTGTTGCTGA	AAGGACAAAACGGGACGA	289	58	n/a
12	GTGAAGTAATGCGAAGGT	GAATTCATGGGTATCAGGG	331	56	KJ601502
13	CCACTGTATGCTACTCTATT	TGATACAGCGTCGAGAAA	228	56	n/a
15	CTCCCTCGCTCTAGAATAA	GGTGGTAGGCTCAAGTGT	270	58	KJ601503
16	TTGCACAGGCTTAAGTGG	AGAGTTTTCGGTTGGTGG	283	57	KJ601504
17	TGTTCTACCGTTTTCCCT	GGCTTGTGCTTGTTATCT	188	55	n/a
18	GATCCCATACCAGTACTT	TTTTTCGCTCTATCTGCC	344	55	KJ601505
19	GGTTGCCATTATGTTAGTG	GATGGAAATGGTGTAGTG	287	55	n/a
20	ACAATGGCGTTACGGATA	CACTTTCACCATGTCTTCT	431	55	n/a
21	TTCCCTGTGTTTTTTGCC	GCTTTATTAGCGATTTCGTG	311	56	n/a
22	GCTTCTTCTTGGTTTCCT	GCCTATGCCATTCTATCT	367	55	n/a
23	GGAGCAGGGCAATAACAA	CCTCCCCTTTTTCATCAC	279	57	n/a
24	CAGCAGCGAGTTCAGACA	CAGGGAGTTTGGGCTAGT	351	59	KJ601506
25	AGATGAGATGAATAAGGTGG	GGAATTGTTGTCGTAGGC	398	56	n/a
26	GCCAGCTCTGCAAAATCAA	CACACCCAGTAGACCCTT	334	58	KJ601507
27	CGGCTATTCACGACTAATTT	CCTTCCCTTTTCTCTTCTTT	262	56	n/a
28	CCCACCAAAACTTTACCC	AAAAGAATTCTCAAAGCCGC	299	57	n/a
30	CAATGCACCGAATTCTCC	GCACCATGCTAATACGCT	244	57	n/a
31	TAGCAGCAGACATGGAAG	ACAGAAGGAAGAGTAGCA	316	57	n/a
32	GTTTCTTGAATTCCCTCAGT	GCTTGTGGACAGTGTATT	276	56	KJ601508

*Primer sets that did not amplify successfully were excluded from the table

Table 4

Representation of the variable sites between 16 *Ceratocystis fimbriata s.l.* species, in the Latin American clade, in seven gene regions investigated in this study. Gene regions include A) Mcm7 and EF 1-α, B) CAL and FG1093, C) βT 1 and RPBII, D) MS204 and E) ITS.

A

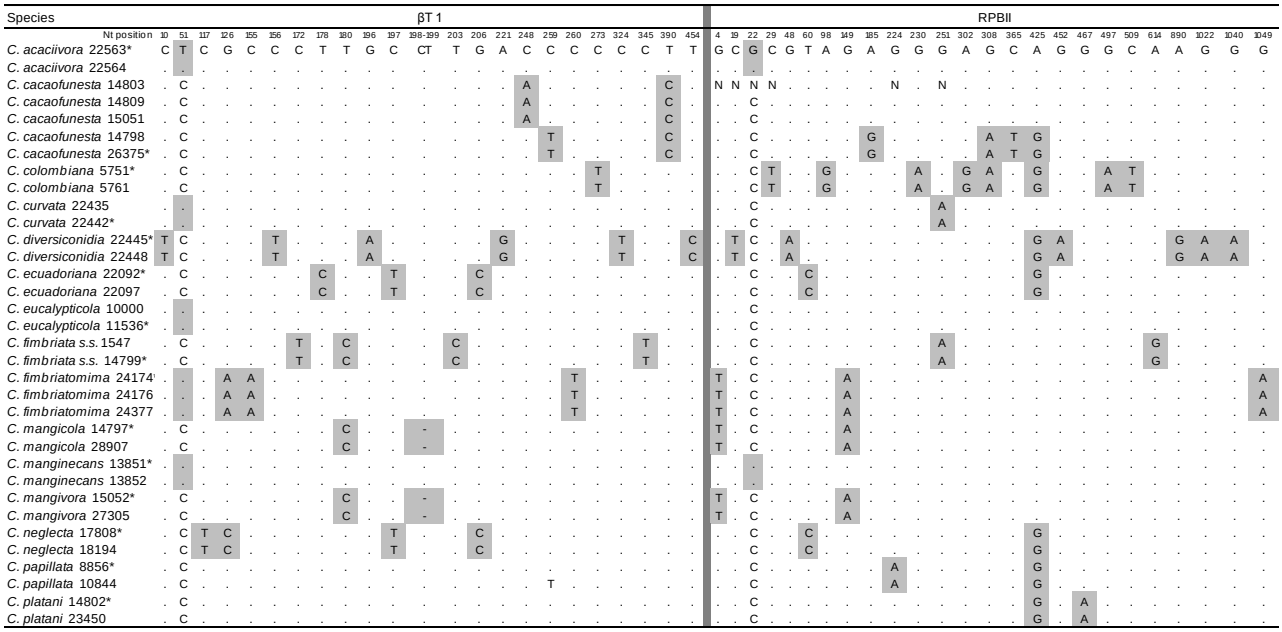
Species	Mcm7														EF 1-α																	
Nt position	13	33	40	55	91	109	139	184	277	319	382	451	469	489	67	154	155	194	197	206	329	382	388	390	391-393	394	395	597	656	666	690	716
C. acaciivora 22563*	T	G	T	C	C	G	A	C	A	A	T	G	C	C	.	-	T	.	T	A	C	C	T	T	-	T	A	C	A	T	C	C
C. acaciivora 22564	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. cacaofunesta 14803	C	A	.	.	T	T	.	.	.	G	.	.	T	.	G	-	.	.	C	.	.	.	C	.	TCA	.	G	.	T	C	T	.
C. cacaofunesta 14809	C	A	.	.	T	T	.	.	.	G	.	.	T	.	G	-	.	.	C	.	.	.	C	.	TCA	.	G	.	T	C	T	.
C. cacaofunesta 15051	C	A	.	.	T	T	.	.	.	G	.	.	T	.	G	-	.	.	C	.	.	.	C	.	TCA	.	G	.	T	C	T	.
C. cacaofunesta 14798	C	A	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. cacaofunesta 26375*	C	A	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. colombiana 5751*	C	A	.	.	T	T	.	T	.	G	.	.	.	.	.	-	.	.	.	C	.	.	.	-	-	-	-	.	.	.	.	.
C. colombiana 5761	C	A	.	.	T	T	.	T	.	G	.	.	.	.	.	-	.	.	.	C	.	.	.	-	-	-	-	.	.	.	.	.
C. curvata 22435	.	.	.	.	T	.	.	.	.	G	.	A	.	.	.	-	.	.	.	.	.	.	.	.	-	.	G	.	.	.	.	.
C. curvata 22442*	.	.	.	.	T	.	.	.	.	G	.	A	.	.	.	-	.	.	.	.	.	.	.	.	-	.	G	.	.	.	.	.
C. diversiconidia 22445*	C	.	C	G	T	T	G	.	G	G	C	.	.	.	G	-	.	G	C	.	T	A	.	.	-	.	.	.	.	.	.	.
C. diversiconidia 22448	C	.	C	G	T	T	G	.	G	G	C	.	.	.	G	-	.	G	C	.	T	A	.	.	-	.	.	.	.	.	.	.
C. ecuadoriana 22092*	.	.	.	.	T	T	.	.	.	G	.	.	.	.	G	-	.	.	C	.	.	.	C	.	TCA	.	.	T	T	C	T	.
C. ecuadoriana 22097	.	.	.	.	T	T	.	.	.	G	.	.	.	.	G	-	.	.	C	.	.	.	C	.	TCA	.	.	T	T	C	T	.
C. eucalypticola 10000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. eucalypticola 11536*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriata s.s. 1547	.	.	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriata s.s. 14799*	.	.	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriatomima 24174*	.	.	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriatomima 24176	.	.	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriatomima 24377	.	.	.	.	T	T	.	.	.	G	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. mangicola 14797*	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. mangicola 28907	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. manginecans 13851*	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A
C. manginecans 13852	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A
C. mangivora 15052*	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. mangivora 27305	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. neglecta 17808*	.	.	.	.	T	T	.	.	.	G	.	.	.	.	G	-	.	.	C	.	.	.	C	.	TCA	.	.	T	T	.	.	.
C. neglecta 18194	.	.	.	.	T	T	.	.	.	G	.	.	.	.	G	-	.	.	C	.	.	.	C	.	TCA	.	.	T	T	.	.	.
C. papillata 8856*	.	.	.	.	T	T	.	.	.	G	.	.	.	.	G	-	.	.	C	.	.	.	C	.	TCA	.	.	.	T	C	.	.
C. papillata 10844	.	.	.	.	T	T	.	.	.	G	.	.	.	.	G	-	.	.	C	.	.	.	C	.	TCA	.	.	.	T	C	.	.
C. platani 14802*	C	A	.	.	T	T	.	.	.	G	.	.	.	T	.	T	.	.	.	.	.	.	.	.	-	.	.	.	T	.	.	.
C. platani 23450	C	A	.	.	T	T	.	.	.	G	.	.	.	T	.	T	.	.	.	.	.	.	.	.	-	.	.	.	T	.	.	.

B

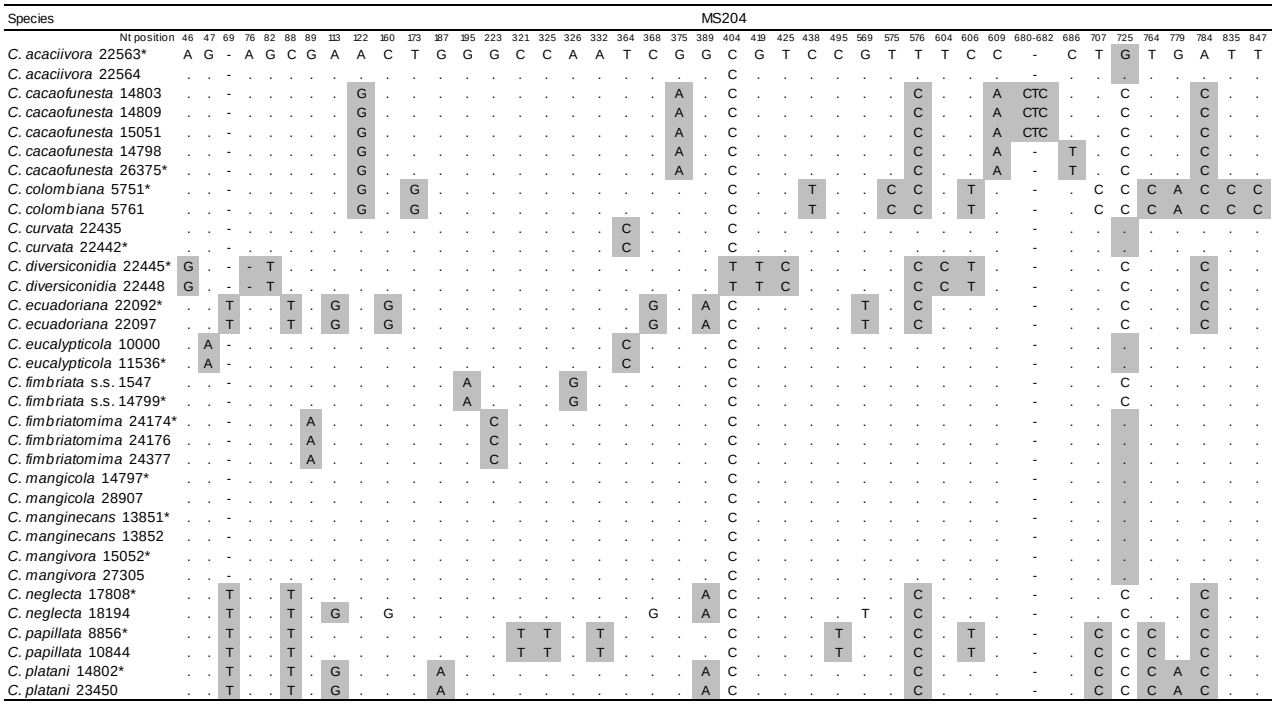
Species	CAL															FG1093																											
Nt position	41	65	89	163	195	266	283	302	397	420	455	474	554	558	564	14	15	23	27	84	86	97	98	108	116	132	168	174	177	178	196	202	216	217	218	219	221	253	409	442	450	491	528
C. acaciivora 22563*	G	G	A	T	G	T	G	G	G	.	C	C	C	G	G	A	T	A	G	T	C	T	T	A	G	A	G	G	A	T	A	G	G	G	T	G	C	A	C	A	C	T	G
C. acaciivora 22564	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. cacaofunesta 14803	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	.	.
C. cacaofunesta 14809	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	.	.
C. cacaofunesta 15051	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	.	.
C. cacaofunesta 14798	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	.	.
C. cacaofunesta 26375*	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	.	.
C. colombiana 5751*	.	.	.	.	.	.	.	.	.	C	A	T	G	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A
C. colombiana 5761	.	.	.	.	.	.	.	.	.	C	A	T	G	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A
C. colombiana 5751 clone																G	C	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.	.	.
C. colombiana 5761 clone																G	C	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	G	.	.	.
C. curvata 22435	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A
C. curvata 22442*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A
C. diversiconidia 22445*	A	A	G	C	A	C	.	.	.	.	.	.	G	.	.	G	.	.	.	.	T	.	C	G	.	T	A	A	G	C	.	.	.	.	C	A	T	.	.	G	.	.	.
C. diversiconidia 22448	A	A	G	C	A	C	.	.	.	.	.	.	G	.	.	G	.	.	.	.	T	.	C	G	.	T	A	A	G	C	.	.	.	.	C	A	T	.	.	G	.	.	.
C. ecuadoriana 22092*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	N	N	T	.	.	G	.	.	.	.	.	.	.	.	N	N	.	.	N	.	.	N	.	.	N
C. ecuadoriana 22097	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	.	.	A
C. ecuadoriana 22092 clone																G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
C. ecuadoriana 22097 clone																G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
C. eucalypticola 10000	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
C. eucalypticola 11536*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
C. fimbriata s.s. 1547	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	T	.	.
C. fimbriata s.s. 14799*	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	C	T	.	.	G	.	.	.	.	.	.	.	.	T	A	.	.	T	.	.	.	T	.	.
C. fimbriatomima 24174*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. fimbriatomima 24176	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. fimbriatomima 24377	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. mangicola 14797*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. mangicola 28907	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. manginecans 13851*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. manginecans 13852	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. mangivora 15052*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. mangivora 27305	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C. neglecta 17808*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
C. neglecta 18194	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
C. papillata 8856*	.	.	.	.	.	.	.	A	A	.	.	.	G	N	N	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	G	.	G	.	.	.
C. papillata 10844	.	.	.	.	.	.	.	A	A	.	.	.	G	.	.	G	.	.	.	.	T	.	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	G	.	G	.	.	.
C. platani 14802*	.	.	.	.	.	.	.	.	.	.	.	.	G	T	.	G	.	G	.	.	T	C	.	G	T	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	C	.
C. platani 23450	.	.	.	.	.	.	.	.	.	.	.	.	G	T	.	G	.	G	.	.	T	C	.	G	T	.	.	.	.	.	.	.	T	.	.	.	T	.	.	G	.	C	.

Table 4 (continued)

C



D



Identical nucleotides are indicated as dots, missing data as 'N' and indels as a dash. If a nucleotide difference was only present in one of the 34 isolates it was excluded from the table. The type strain of each species is indicated by *. Variable sites that distinguish between species are highlighted in grey.

Table 4 (continued)

E

Species	ITS																																									
Nt position	6	10	13	16	19-22	24	26	27	30	33	36	58	67	107	109	110	111-116	117	118	12-122	123	129	132	133	135	139	140	141	154	156	157	158	161	163	165-167	173	177	178	179	180	182
C. acaciivora 22563*	T	A	G	C	-	A	C	-	G	G	-	T	-	C	G	A	AGAG	A	G	C-	A	G	A	G	A	T	T	A	C	T	C	G	A	T	GT-	C	T	T	A	T	A
C. acaciivora 22564	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	AGAG	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	GT-	.	.	.	.	.	.
C. cacaofunesta 14803	.	.	.	-	TATA	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	C	T	.	.	A	.	.	AT-	.	.	.	.	.	.
C. cacaofunesta 14809	.	.	.	-	TATA	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	C	T	.	.	A	.	.	AT-	.	.	.	.	.	.
C. cacaofunesta 15051	.	.	.	-	TATA	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	C	T	.	.	A	.	.	AT-	.	.	.	.	.	.
C. cacaofunesta 14798	.	.	.	C	C-TA	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GTA	.	.	.	.	.	.
C. cacaofunesta 26375*	.	.	.	C	C-TA	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GTA	.	.	.	.	.	.
C. colombiana 5751*	.	.	A	C	-	.	.	-	.	.	A	.	T	.	.	.	AAGGGG	G	.	.	.	.	.	.	.	T	A	.	.	.	.	A	.	.	GTA	.	.	.	T	A	.
C. colombiana 5761	.	.	A	C	-	.	.	-	.	.	A	.	T	.	.	.	AAGGGG	G	.	.	.	.	.	.	.	T	A	.	.	.	.	A	.	.	GTA	.	.	.	T	A	.
C. curvata 22435	G	.	.	-	-	.	T	T	.	.	-	.	-	T	.	.	AGGG	.	.	.	C	.	.	A	.	T	.	.	T	.	T	A	C	C	GT-	T	.	A	T	A	T
C. curvata 22442*	G	.	.	-	-	.	T	T	.	.	-	.	-	T	.	.	AGGG	.	.	.	C	.	.	A	.	T	.	.	T	.	T	A	C	C	GT-	T	.	A	T	A	T
C. diversiconidia 22445*	.	-	.	-	-	.	.	-	.	A	-	C	T	.	A	T	-	.	A	.	.	A	.	.	.	-	.	.	.	G	.	A	.	.	---	.	A	A	T	.	T
C. diversiconidia 22448	.	-	.	-	-	.	.	-	.	A	-	C	T	.	A	T	-	.	A	.	.	A	.	.	.	-	.	.	.	G	.	A	.	.	---	.	A	A	T	.	T
C. ecuadoriana 22092*	.	.	.	-	-	G	.	-	A	.	-	.	-	.	.	.	-	.	.	.	.	.	G	.	.	T	.	.	.	C	.	A	.	.	---	.	.	.	T	C	.
C. ecuadoriana 22097	.	.	.	-	-	G	.	-	A	.	-	.	-	.	.	.	-	.	.	.	.	.	G	.	.	-	.	.	.	C	.	A	.	.	---	.	.	.	T	C	.
C. eucalypticola 10000	.	.	.	-	-	.	.	T	.	.	-	.	-	.	.	.	AGAG	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	GT-	.	.	.	.	.	.
C. eucalypticola 11536*	.	.	A	-	-	.	.	T	.	.	-	.	-	.	.	.	AGAG	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	AT-	.	.	.	.	.	.
C. fimbriata s.s. 1547	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	AGGG	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	AT-	.	.	.	.	.	.
C. fimbriata s.s. 14799*	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	AGGG	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	AT-	.	.	.	.	.	.
C. fimbriatomima 24174*	.	.	.	-	-	.	.	T	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	.	.
C. fimbriatomima 24176	.	.	.	-	-	.	.	T	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	.	.
C. fimbriatomima 24377	.	.	.	-	-	.	.	T	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	.	.
C. mangicola 14797*	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	G	-	.	.	.	.	.	A	.	.	ATA	.	.	.	.	.	.
C. mangicola 28907	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	G	-	.	.	.	.	.	A	.	.	ATA	.	.	.	.	.	.
C. manginecans 13851*	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	AGAG	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	ATA	.	.	.	.	.	.
C. manginecans 13852	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	AGAG	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	ATA	.	.	.	.	.	.
C. mangivora 15052*	.	.	A	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	C	.
C. mangivora 27305	.	.	A	-	-	.	.	T	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	C	.
C. neglecta 17808*	.	.	.	-	-	G	.	-	.	.	-	.	-	.	.	.	-	.	.	-	.	.	G	.	.	-	.	.	.	C	.	A	.	.	---	.	.	.	T	C	.
C. neglecta 18194	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	-	.	.	G	.	.	-	.	.	.	C	.	A	.	.	---	.	.	.	T	C	.
C. papillata 8856*	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	TA	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	C	.
C. papillata 10844	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	TA	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	.	.	T	C	.
C. platani 14802*	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	A	.	.	.	.
C. platani 23450	.	.	.	-	-	.	.	-	.	.	-	.	-	.	.	.	-	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	GT-	.	A	.	.	.	.

E

Species	ITS																																							
Nt position	184	187	189	190	191	192-200	202	203	204	206	208	213	228	276	280	286	292	314	328	336	342	361	365	376	384	395	398	403	410	419	420	421	422	423	424	425	427	437	473	479	486
C. acaciivora 22563*	T	C	A	G	A	-	T	T	T	T	A	C	C	C	A	G	G	T	G	T	A	A	C	A	A	T	G	T	A	G	A	C	T	-	C	T	T	C	G	T	C
C. acaciivora 22564	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. cacaofunesta 14803	A	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. cacaofunesta 14809	A	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. cacaofunesta 15051	A	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. cacaofunesta 14798	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	C	.
C. cacaofunesta 26375*	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	C	.
C. colombiana 5751*	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.
C. colombiana 5761	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.
C. curvata 22435	.	T	C	A	G	-	A	.	A	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. curvata 22442*	.	T	C	A	G	-	A	.	A	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. diversiconidia 22445*	.	T	T	T	G	TTAGAAATT	G	A	.	.	-	.	.	T	G	A	A	A	A	G	T	T	G	T	G	C	A	C	C	A	.	.	.	-	.	C	.	G	A	C	T
C. diversiconidia 22448	.	T	T	T	G	TTAAAAATT	G	A	.	.	-	.	.	T	G	A	A	A	A	G	T	T	G	T	G	C	A	C	C	A	.	.	.	-	.	C	.	G	A	C	T
C. ecuadoriana 22092*	.	.	T	A	.	-	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	.
C. ecuadoriana 22097	.	.	T	A	.	-	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	.
C. eucalypticola 10000	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. eucalypticola 11536*	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	.	.
C. fimbriata s.s. 1547	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	C	.
C. fimbriata s.s. 14799*	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	C	.
C. fimbriatomima 24174*	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriatomima 24176	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. fimbriatomima 24377	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. mangicola 14797*	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. mangicola 28907	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. manginecans 13851*	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. manginecans 13852	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
C. mangivora 15052*	.	.	T	A	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	C	.
C. mangivora 27305	.	.	T	A	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	-	A	.	C	.	.	C	.
C. neglecta 17808*	.	.	T	A	.	-	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	.
C. neglecta 18194	.	.	T	A	.	-	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	C	.	.	.	.	.
C. papillata 8856*	.	.	T	A	.	-	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.
C. papillata 10844	.	.	T	A	.	-	-	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.
C. platani 14802*	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	C	C	.	C	.	.	.	.	.
C. platani 23450	.	.	.	G	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	C	T	.	C	.	.	.	.	.

Identical nucleotides are indicated as dots, missing data as ‘N’ and indels as a dash. If a nucleotide difference was only present in one of the 34 isolates it was excluded from the table. The type strain of each species is indicated by *. Variable sites that can, in combinaiton, distinguish between species are highlighted in grey.

Table 5

Information on the aligned sequence data of each gene region investigated, including data from Maximum parsimony and Maximum likelihood phylogenetic analysis.

	ITS	Mcm7	EF	CAL	FG1093	βT 1	RPBII	MS204	Four gene combination
Total no. of characters (bp)	560	549	739	570	534	550	1129	880	3093
Parsimony informative characters	197	27	42	37	41	38	38	78	195
No. of constant characters	334	522	693	532	481	503	1089	783	2855
Characters excluded from analysis	14	-	-	-	12	-	-	14	26
Variable characters, uninformative	15	-	4	1	-	9	3	5	17
Number of MP trees obtained	12	2	9	6	67	16	6	3	16
Consistency index (CI)	0.766	0.9	0.904	1	0.957	0.961	0.976	0.892	0.831
Retention index (RI)	0.86	0.959	0.96	1	0.987	0.974	0.987	0.947	0.907
Nt substitution model (jModeltest)	TVM+G	TPM2+G	TPM2uf+G	TVMef+G	TPM3uf+I+G	K80	TIM2ef	TPM2uf+G	TPM2uf+G
-log likelihood for ML tree	1611.561	938.738	1256.811	997.603	1009.848	1083.923	1858.697	1736.852	5954.243

Table 6

Values obtained from the genealogical sorting index, providing alternative analysis and support values for the monophyly of *C. fimbriata s.l.* species.

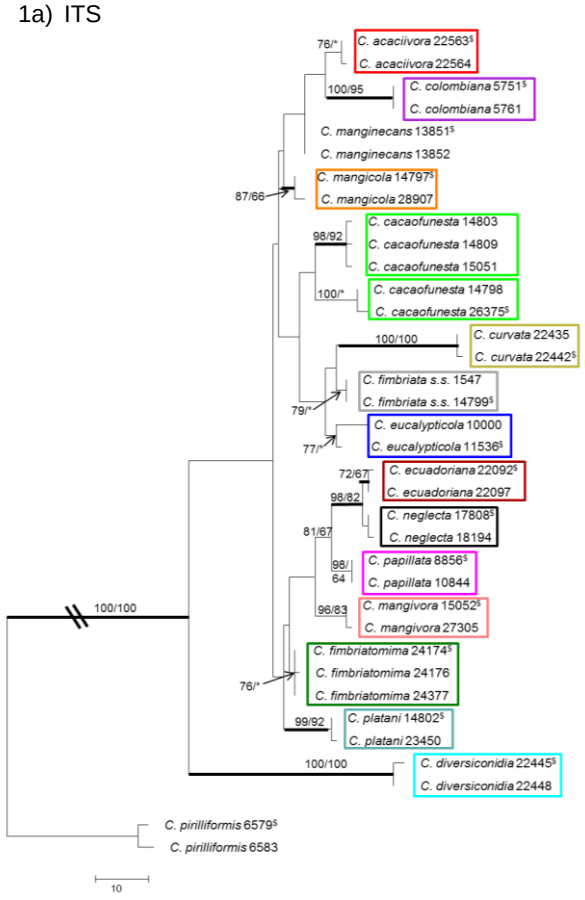
	ITS	Mcm7	EF	Cal	FG1093	βt 1	RPBII	MS204	Support for
Species	gsi_T [c]	gsi_T [c]	gsi_T [c]	gsi_T [c]	gsi_T [c]	gsi_T [c]	gsi_T [c]	gsi_T [c]	monophyly[e]
C. acaciivora	**0.473**	0.18	0.202	0.172	**0.285**	**0.268**	**0.431**	**0.241**	none
C. cacaofunesta [a]	**0.976**	**0.793**	**0.583**	**0.927**	**0.495**	**0.835**	**0.25**	**0.753**	5 genes
C. cacaofunesta [b]	**0.567**	**0.362**	0.191	0.11	**0.342**	**0.638**	**0.923**	**0.579**	1 gene
C. colombiana	**0.964**	**0.75**	**0.846**	**0.972**	**0.907**[d]	**0.687**	**0.985**	**1**	7 genes
C. curvata	**1**	**0.722**	**0.699**	0.109	**0.371**	0.243	**0.436**	**0.413**	2 genes
C. diversiconidia	**0.835**	**0.974**	**0.955**	**0.943**	**0.964**	**0.959**	**0.992**	**0.99**	8 genes
C. ecuadoriana	**0.81**	0.23	**0.71**	0.178	0.172	**0.851**	**0.39**	**0.614**	2 genes
C. eucalypticola	**0.334**	**0.499**	0.15	0.173	**0.709**	0.25	**0.266**	**0.869**	1 gene
C. fimbriata s.l.	**0.656**	0.215	**0.203**	**0.672**	**0.865**	**0.978**	**0.781**	**0.866**	4 genes
C. fimbriatomima	**0.53**	**0.23**	**0.192**	0.14	**0.331**	**0.95**	**0.762**	**0.934**	3 genes
C. mangicola	**0.774**	**0.28**	**0.642**	0.137	**0.375**	**0.456**	**0.411**	**0.232**	1 gene
C. manginecans	**0.431**	**0.242**	**0.711**	0.144	**0.272**	**0.249**	**0.413**	0.198	none
C. mangivora	**0.931**	**0.262**	0.182	0.171	**0.375**	**0.401**	**0.482**	**0.292**	1 gene
C. neglecta	**0.491**	0.211	**0.711**	0.142	**0.306**	**0.929**	**0.382**	**0.249**	1 gene
C. papillata	**0.805**	0.201	0	**0.868**	**0.905**	0.186	**0.684**	**1**	4 genes
C. pirilliformis	**0.65**	**0.485**	**0.49**	**0.485**	**0.523**	**0.485**	**0.501**	**0.485**	outgroup
C. platani	**0.947**	**0.807**	**0.382**	**0.729**	**0.975**	0.206	**0.744**	**0.995**	5 genes

a Isolates representing the Ecuadorian lineage of *C. cacaofunesta* isolates (CMW 14803, 14809 and 15051).
b Isolates representing the Brazil/Costa Rica lineage of *C. cacaofunesta* isolates (CMW 14798 and 26375).
c Gsi values that are statistically significant (P<0.05) are indicated in bold and values that provide higher support for a species than obtained from the phylogenetic analyses are highlighted in grey. Gsi values above 0.75, with a P< 0.05 were considered as significant support for the monophyly of a species.
d Value is for the second haplotype of the *C. colombiana* isolates obtained from the cloned sequences.
e The numbers indicate the amount of gene regions that support a species to be monophyletic based on gsi values.

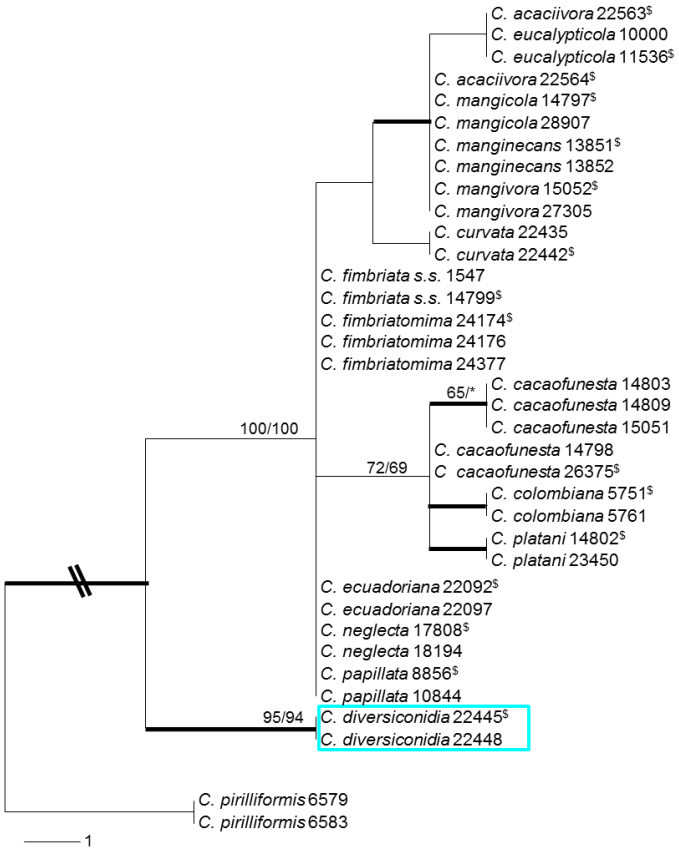
Table 7

A representation of all the variable sites identified from the nine SNP regions investigated in 16 *C. fimbriata s.l.* species - Available as an additional file on CD

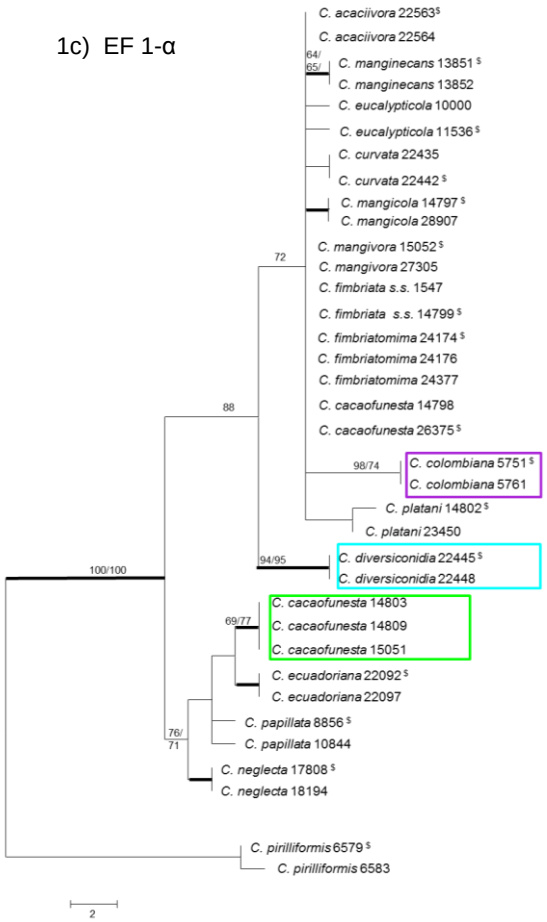
1a) ITS



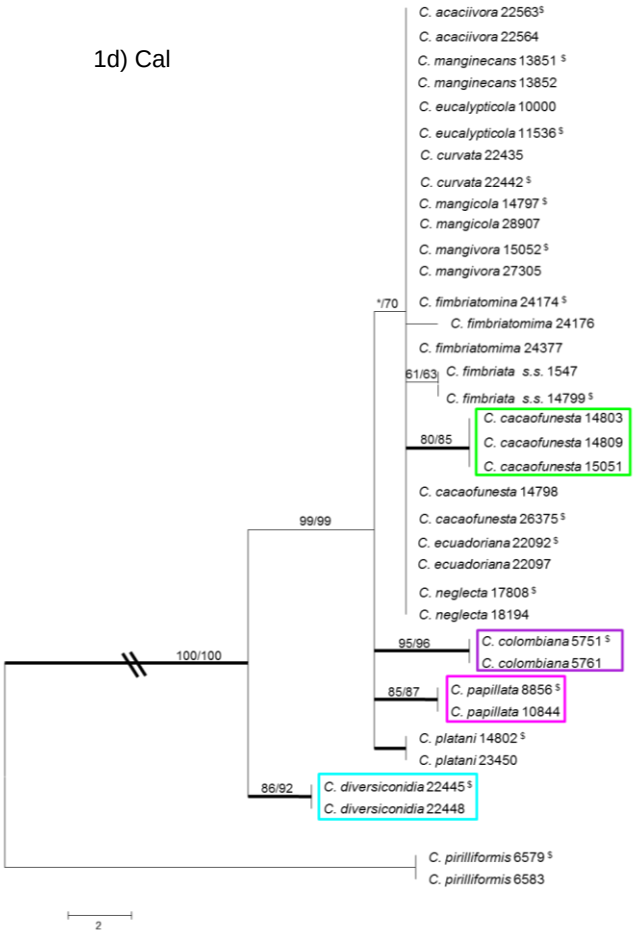
1b) Mcm7



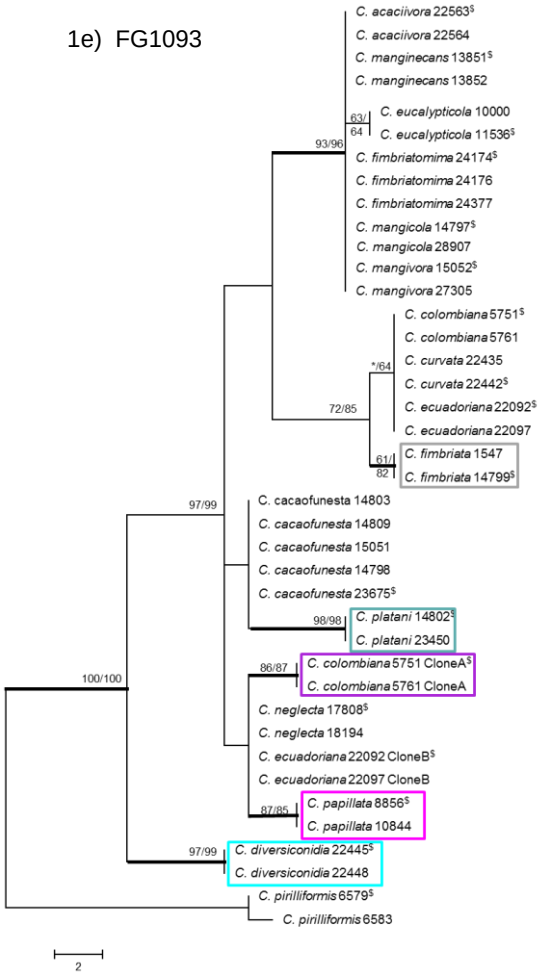
1c) EF 1-α



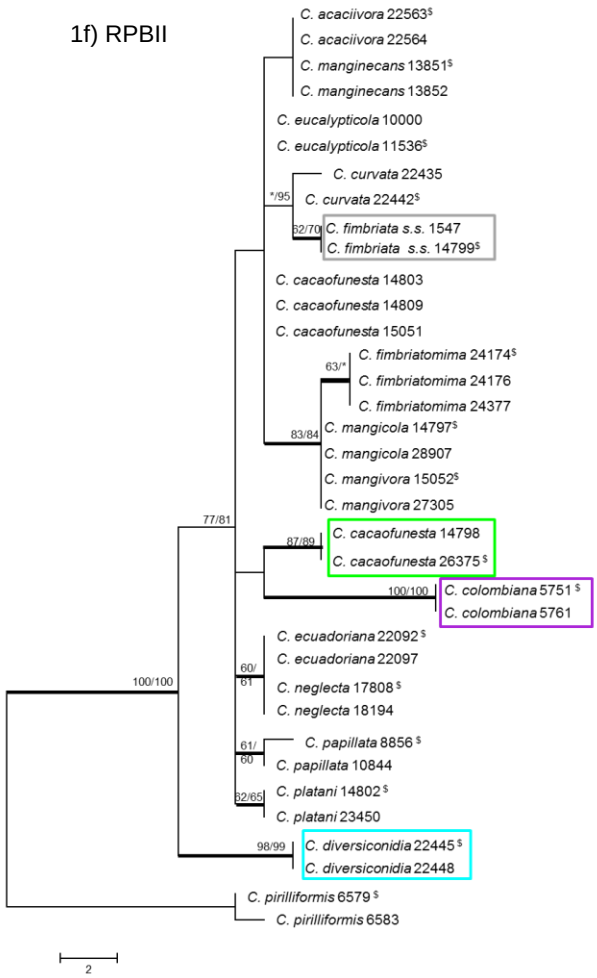
1d) Cal



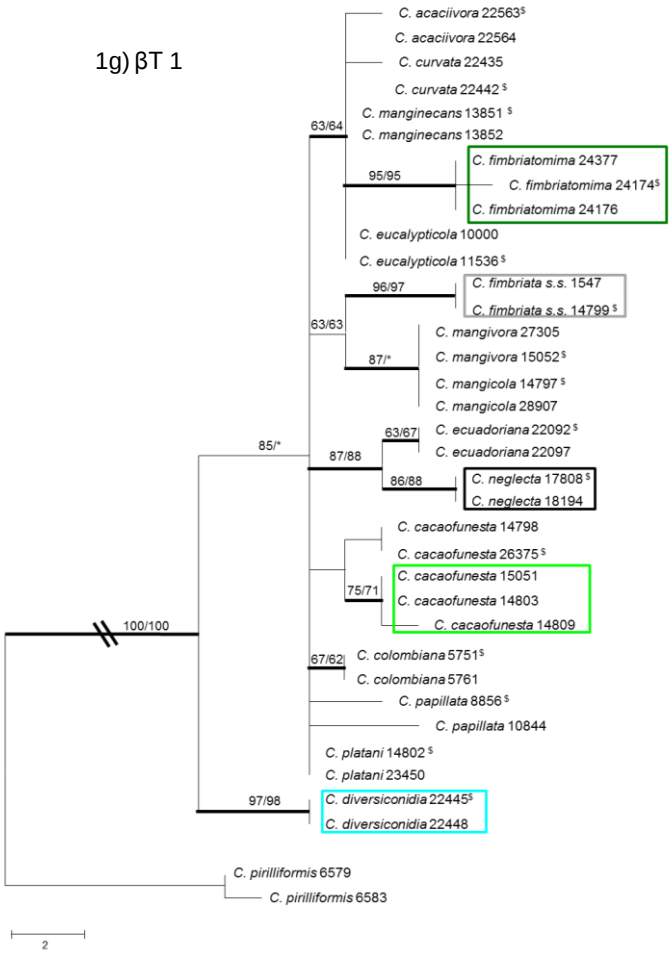
1e) FG1093



1f) RPBII



1g) βT 1



1h) MS204

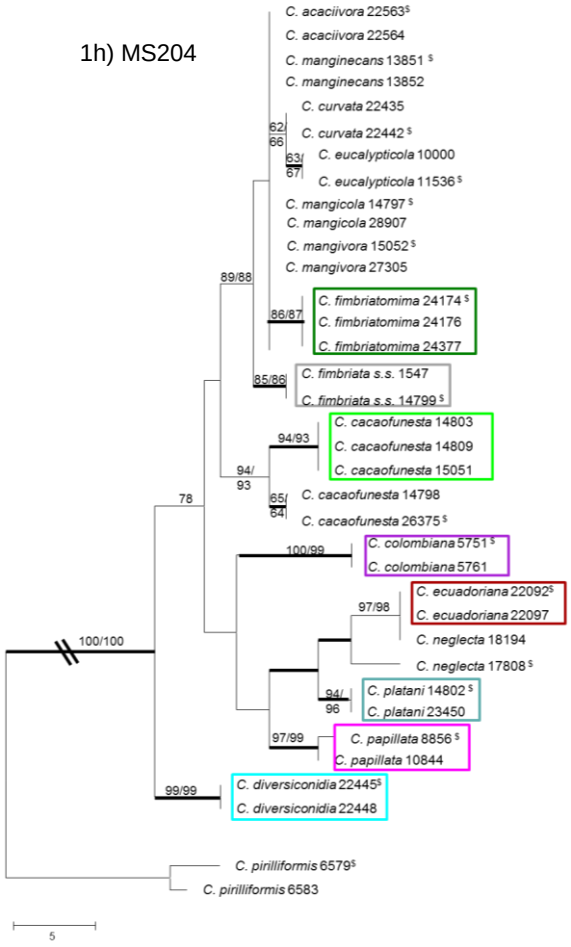


Fig. 1. One of the most parsimonious trees from Maximum parsimony analysis representing the *C. fimbriata s.l*. species phylogeny for each of the gene regions a) ITS, b) Mcm7, c) EF 1-α, d) Calmodulin, e) FG1093, f) RPBII, g) βT 1 and h) MS204 respectively. The type strain of each species is indicated by $. Branch support values are indicated above the branches with bootstrap values > 60% from MP followed by ML analysis. Bayesian posterior probabilities above 95% are indicated by thick branches. An * indicates support values below the threshold. Branch length indicates number of nucleotide changes.

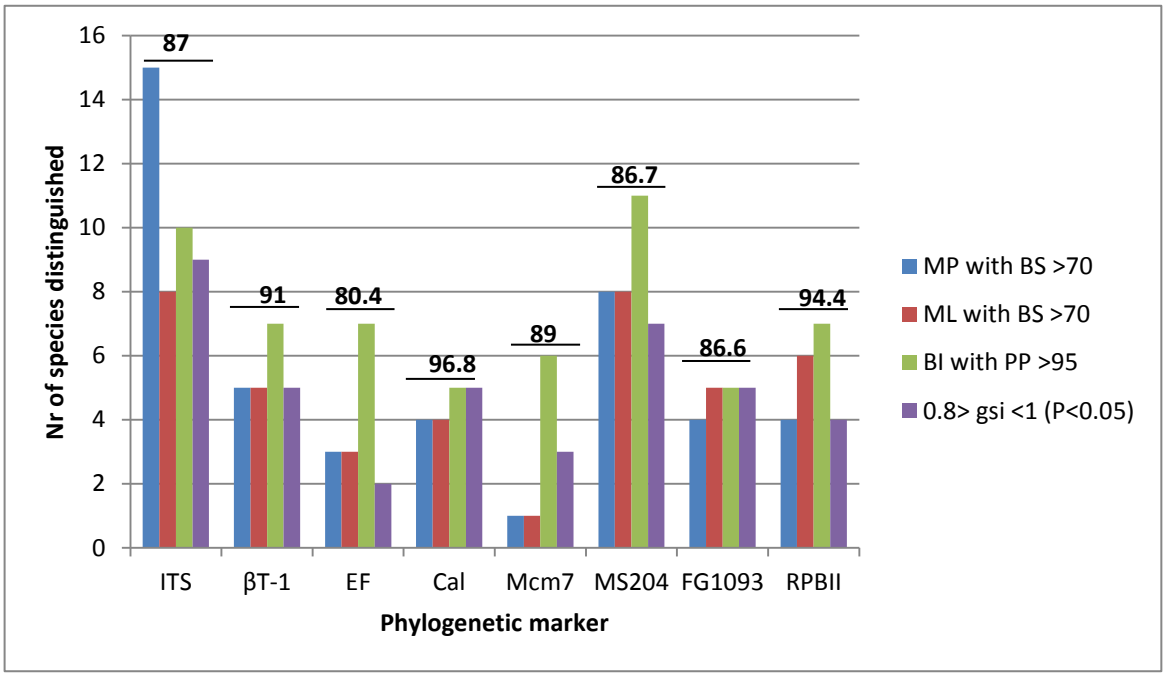


Fig. 2. Summary of the phylogenetic utility of all phylogenetic gene regions investigated in this study. The x-axis represents the different gene regions and the y-axis the number of species distinguished from a total of 15 distinct species defined in literature. The four columns for every gene region represent four alternative methods applied to the sequence data generated for species delineation. MP refers to Maximum Parsimony analysis with Bootstrap support >70%, ML to Maximum likelihood, BI to Bayesian inference with posterior probabilities > 95 and gsi refers to the genealogical sorting index with high values supporting the monophyly of a species. Underlined values above the bars of each gene region indicate how well the tree topology correlates with that of the combined tree constructed from all eight gene regions (Nei topological score).

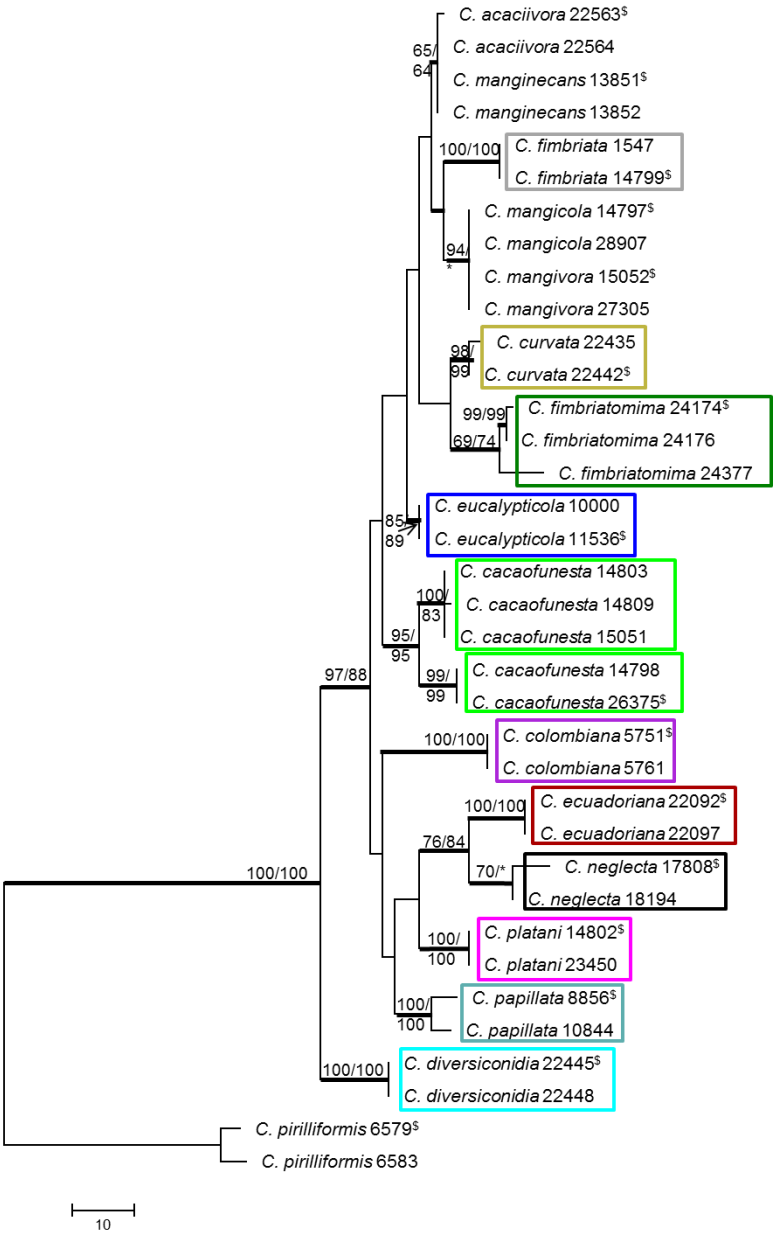


Fig. 3. One of the most parsimonious trees selected from Maximum parsimony analysis for representation of the *C. fimbriata s.l*. species phylogeny. This was constructed from the combined data from the gene regions βT 1, FG1093, MS204 and RPBII. The type strain of each species is indicated by $. Branch support values are indicated above the branches and include bootstrap values above 60 from MP/ ML analysis. Bayesian posterior probability above 95% is indicated by thick branches. An * indicates support values from any of the analyses below the threshold.

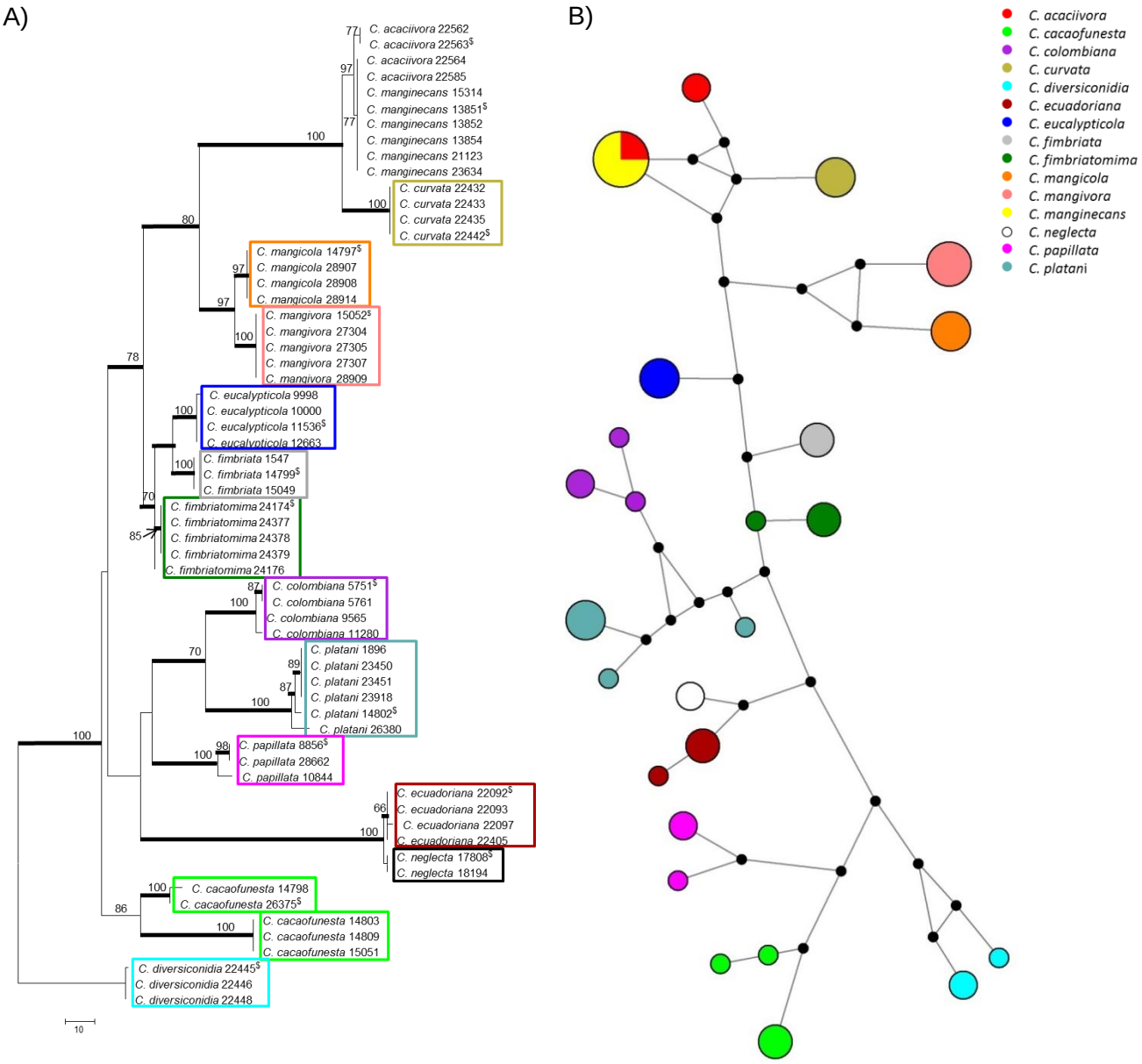


Fig. 4. Species delineation of *C. fimbriata s.l.* species based on SNP variation from the nine amplified SNP regions. A) Cladogram of *C. fimbriata s.l.* isolates based on all amplified SNP regions considering the sequence of the entire SNP region. The type strain of each species is indicated by $. Maximum parsimony analysis and Bayesian inference was performed on all represented isolates. Bootstrap support values are indicated at each branch and Bayesian posterior probability values >95 are indicated by thick branches. B) Haplotype network constructed based on a median-joining algorithm. Only the SNP sites, obtained from nine SNP regions, were considered. Nodes (circles) represent haplotypes and the size of the node corresponds to the number of isolates included in the haplotype. Coloured nodes represent existing individuals, coloured according to their species status as suggested by ITS sequence data. Black internal nodes represent missing haplotypes.

Chapter 3

Genetic diversity of *Ceratocystis manginecans* populations on mango and legume trees

ABSTRACT

The fungal pathogen, *Ceratocystis manginecans,* causes a serious canker and wilt disease of trees. Hosts include mango and various leguminous trees in Oman and Pakistan in the Middle East and *Acacia* spp. in Indonesia. A *Ceratocystis* species, with similar morphology to *C. fimbriata sensu lato* species, has recently been reported in Vietnam, causing severe disease of *Acacia* trees. Preliminary population studies on isolates from mango in Oman and Pakistan have shown that the pathogen represents a single clonal haplotype, indicative of an introduced pathogen. The aim of this study was to definitively identify the *Ceratocystis* species, present in Vietnam, based on DNA sequence data. In addition, the genetic diversity and population structure of 161 isolates, from four host-associated populations of *C. manginecans* from Oman, Pakistan, Indonesia and Vietnam were analysed. This was done by applying a combination of 14 previously developed and a new set of 10 microsatellite markers, designed from the *C. manginecans* genome. The sequence data confirmed that the isolates in Vietnam are the same as those in Indonesia and were thus identified as *C. manginecans*. Unlike the populations in Oman and Pakistan, relatively high levels of genetic variation were found among the isolates from Indonesia and Vietnam. The Vietnam population was significantly differentiated from the other populations and isolates from this area had the highest level of genetic diversity thus far encountered for the pathogen.

1. Introduction

Ceratocystis manginecans is a fungal pathogen, first described from Oman and Pakistan, as the causal agent of a severe mango wilt disease (Al Adawi et al., 2013b; Al Adawi et al., 2006; Van Wyk et al., 2007). The species resides in the *C. fimbriata sensu lato* complex that includes numerous cryptic species, the majority of which are aggressive tree pathogens (Al Adawi et al., 2013b; Engelbrecht and Harrington, 2005; Ocasio-Morales et al., 2007; Roux and Wingfield, 2013; Wingfield et al., 2013). *Ceratocystis acaciivora* is morphologically identical to *C. manginecans* and both have been reported on *Acacia mangium* trees in Indonesia (Tarigan et al., 2011). An in-depth investigation of the taxonomic placement of species in the *C. fimbriata s.l.* complex, using eight phylogenetic markers and nine SNP markers, has shown that the two species are identical, except in the rRNA internal transcribed spacer regions 1 and 2 (ITS) (Chapter 2). However, two ITS haplotypes have been identified in a single isolate (Al Adawi et al., 2013b; Naidoo et al., 2013), and the two species must be treated as synonyms.

Since the first report of mango wilt in Oman in 1999 (Al Adawi et al., 2003), *C. manginecans* infection has expanded to 12 districts of Oman and to three in Pakistan (Al Adawi et al., 2013a; Al Adawi et al., 2006; Van Wyk et al., 2007). Concern regarding the pathogen has increased in recent years, as it has also been found on the native legume trees *Prosopis cineraria* (Ghaf) in Oman and *Dalbergia sissoo* (Shisham) in Pakistan (Al Adawi et al., 2013b). This species has also been reported on the widely planted *Acacia mangium* in South East Asian countries, including Indonesia (Tarigan et al., 2011) and Malaysia (Wingfield, unpublished), and on *Eucalyptus* species in South China (Chen et al., 2013). Isolates of *C. fimbriata s.l.,* which could represent the same fungus, have also been reported from *Eucalyptus* trees in Brazil and pomegranate in India (Harrington et al., 2014).

Disease symptoms associated with *C. manginecans* infection on mango and legume trees include staining of vascular tissue, stem cankers and wilt (Al Adawi et al., 2013b; Tarigan et al., 2011; Van Wyk et al., 2007). These symptoms can result in tree death within six months of infection (Al Adawi et al., 2006; Tarigan et al., 2011). *Acacia mangium* trees form an essential part of the paper and pulp industry in Indonesia and surrounding areas (FSIV, 2009; Maslin, 2013). Oman and Pakistan are also highly reliant on mango production for fruit consumption and exports (Al Adawi et al., 2006; FAOSTAT, 2011; Kumar et al., 2011). The disease is thus a serious threat to fruit crop and plantation tree industries in the Middle East and South East Asia.

In a recent survey of *Acacia* plantations in Vietnam, a species of *Ceratocystis* was identified causing stem cankers and crown wilt on *Acacia auriculiformis, A. mangium* and a hybrid of the two species in eight provinces in the country (Thu et al., 2012). This was the first report of a *Ceratocystis* species causing disease on *Acacia* in Vietnam. The geographic

proximity of Vietnam to Indonesia and the similarity of the disease on the same host plant (Tarigan et al., 2011) suggest that the pathogen could be *C. manginecans.*

A population genetic study was recently performed on *C. manginecans* isolates collected from mango trees from Oman and Pakistan (Al Adawi et al., 2014), using microsatellite markers (Barnes et al., 2001; Steimel et al., 2004). Isolates from the wood-boring beetle *Hypocryphalus mangiferae*, the major vector of *C. manginecans* on mango in Oman and Pakistan (Al Adawi et al., 2013a), were included. The data showed that the isolates in both countries, from the mango trees and bark beetles, represent a single clonal population, suggesting that the fungus is an introduced pathogen. This could have been introduced from South East Asia, where the fungus occurs on *A. mangium* (Tarigan et al., 2011), or from South-America, as suggested by Harrington et al. (2014). The rapid distribution of the pathogen in Oman and Pakistan was most likely facilitated by *Hypocryphalus mangiferae* (Al Adawi et al., 2013a; Al Adawi et al., 2013b; Masood et al., 2010). This insect is thought to be native to southern Asia but has not yet been connected to outbreaks of *C. manginecans* in this region (Atkinson and Peck, 1994; Masood et al., 2010; Masood et al., 2008).

The aim of this study was to consider the population genetic structure of *C. manginecans* isolates from the Middle East (Oman and Pakistan) and South East Asia, thus beyond those areas previously considered. In doing so, it was necessary to determine whether isolates from Vietnam represented the same species as those causing disease of *A. mangium* in Indonesia. The species identity was determined in the current study, using a sequence based approach. The study also included isolates from the legume trees *P. cineraria* and *D. sissoo* from Oman and Pakistan, respectively, as these were not considered in terms of their genetic structure in the study by Al Adawi et al. (2014). The genetic diversity of the populations was investigated using a combination of previously developed microsatellite markers (Barnes et al., 2001; Steimel et al., 2004), as well as a new set of markers designed in this study.

2. Materials and methods

2.1. Fungal isolates and DNA extraction

Isolates from Oman and Pakistan, used in this study, included those from mango and bark beetles infesting these trees, previously considered by Al Adawi et al. (2014). These isolates were included in this study to evaluate the newly developed microsatellite markers in this population and compare the results to the other populations included in this study. In addition, collections from the two leguminous tree species (Table 1) were also included. The isolates from *P. cineraria* trees were collected in Sohar in Oman and those from *D. sissoo* trees in Faisalabad and Shorkot in Pakistan (Al Adawi et al., 2013b). In Indonesia, isolates

were collected from wilting *A. mangium* trees in three plantation areas of the Riau province of Sumatra (Fig. 1a). Isolates were collected from diseased *A. mangium, A. auriculiformis* or hybrids of these two tree species in Vietnam during a previous disease survey in seven provinces (Thu et al., 2012; Fig. 1b). All isolates are maintained in the culture collection facility (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), Pretoria, South Africa.

All cultures were grown at room temperature (23-25°C) on 2% malt extract agar (MEA) (Biolab, Midrand, South Africa) supplemented with 50 mg/l streptomycin (Sigma-Aldrich, Germany) and 100 µg/l thymine (Sigma-Aldrich, Germany). Mycelium was collected from the surface of the MEA plates and freeze dried. Dried mycelium was crushed in a Mixer Mill type MM 301 Retsch® tissue lyser (Retsch, Germany), with metal beads. DNA was extracted using a phenol/chloroform method (Goodwin et al., 1992) and quantified using a Nanodrop ND_1000 instrument (Nanodrop, Wilmington, Delaware). The quality of the DNA was then analysed by means of gel electrophoresis (AGE) on a 1% agarose gel by combining a volume of 5 µl DNA with 2 µl GelRed™ (Biotium, California). The DNA was visualised with UV light.

2.2. Confirmation of isolate identity

2.2.1. PCR amplification, sequencing and cloning

Isolates from mango and legume trees in Oman and Pakistan, used in this study, had been identified as *C. manginecans* in previous studies (Al Adawi et al., 2013b; Al Adawi et al., 2014). Those from Indonesia (110 isolates) and Vietnam (24 isolates) were identified as belonging to the *C. fimbriata s.l.* complex, based only on morphology. These isolates were subjected to DNA sequence analyses by amplifying the ITS gene region in all the isolates, using primers ITS1 and ITS4 (White et al., 1990). This region was selected for identification of the isolates, as it remains the most informative region to date. However, to obtain a reliable identification of the isolates, the two additional regions β-tubulin 1 (βT 1) and Translation Elongation Factor 1-α (EF 1-α) were amplified in a subset of 24 isolates to confirm the ITS data. The primers βT1a and βT1b (Glass and Donaldson, 1995) and EF1F and EF1R (Jacobs et al., 2004) were used, respectively, for amplification and sequencing.

The PCR reaction mixture, used for the amplification of all gene regions, consisted of 30 ng DNA, 2.5 mM $MgCl_2$, 250 µM dNTP, 1 unit FastStart *Taq* DNA polymerase (Roche Applied Science, South Africa), 2.5 µl 10x FastStart *Taq* DNA polymerase PCR buffer (including 1.5 mM $MgCl_2$), between 0.2 – 0.4 µM of each forward and reverse primer and adjusted to a total reaction volume of 25 µl with autoclaved distilled H_2O. The PCR cycler program consisted of: 95 °C for 5 min, 10 cycles of 95 °C for 30 s, 55 °C for 45 s, 72 °C for 90 s, another 30 cycles of 95 °C for 30 s, 58 °C for 45 s and 72 °C for 90 s (5 s increase per

cycle at 72 °C) with a final step at 72 °C for 10 min. The primer annealing temperature for the ITS primers was 60 °C.

PCR amplification success was determined by AGE on a 2% agarose gel and DNA was visualised under UV illumination. PCR and sequence amplicons were purified with 6% Sephadex G-50 in Centricep filter columns (Sigma-Aldrich, Germany). The ABI PRISM™ Big DYE Terminator Cycle Sequencing Ready Reaction Kit (Applied BioSystems, Foster City, California) was used to perform the amplification reactions, according to manufacturer's protocols, and the sequence of the amplicons was determined on an ABI PRISM™ 3500xl Autosequencer (Applied BioSystems, Foster City, California). Where reads were ambiguous in the ITS region, due to multiple ITS copies (Al Adawi et al., 2013b; Naidoo et al., 2013), the ITS PCR amplicons were cloned using the pGEM®-T and pGEM®-T Easy Vector System (Promega, Madison, USA). The primers T7 and SP6 were used for amplification and sequencing (Invitrogen, Life technologies, Johannesburg, SA). For each ambiguous PCR amplicon, up to five clones were sequenced. Since the number of possible ITS haplotypes were not known, this number was arbitrarily selected to increase the possibility of observing more than one ITS haplotype. The amplification reaction mixture had a total volume of 25 µl, consisting of 30 ng DNA, 1.5 units MyTaq™ DNA Polymerase (Bioline Ltd. UK), 5 µl 5x MyTaq™ Buffer, 0.2 µM of each primer and distilled H_2O. The PCR cycler program consisted of 95 °C for 5 min followed by 35 cycles of 95 °C for 30 s, 56 °C for 30 s, 72 °C for 1 min and a final extension of 72 °C for 10 min.

2.2.2. Sequence alignment and phylogenetic analysis

The raw sequence reads for each of the isolates and all three gene regions were assembled in CLC Main Workbench V. 6.0 (CLC Bio, www.clcbio.com) to obtain consensus sequences for every isolate. The sequences obtained for each gene region were aligned to a previously generated data set of the ITS, βT 1 and EF 1-α gene regions for species in the *C. fimbriata s.l.* complex (Chapter 2, this dissertation). Alignments were performed using MUSCLE in MEGA v. 5 (Tamura et al., 2011) and used to construct a phylogenetic tree based on Maximum parsimony (MP). Due to the great number of isolates used in this study, all isolates with identical ITS sequences were represented by a single haplotype to reduce the number of taxa in the phylogenetic tree. Sequences with more than 1 bp difference from the consensus were considered as a different haplotype. The MP tree was constructed in PAUP 4.0, based on a heuristic search option, with random addition of sequences and Tree Bisection Reconstruction (TBR) branch swopping method (Swofford, 2002). Statistical support for the tree branches was confirmed with a 1000 bootstrap replicates.

2.3. Development of microsatellite markers and GeneScan analysis

Microsatellite markers, available for species in the *C. fimbriata* complex (Barnes et al., 2001; Steimel et al., 2004), were used in this study. However, in order to increase the rigor of population genetic analyses in this study, additional microsatellite or simple sequence repeat (SSR) markers were developed specifically for *C. manginecans*. This was achieved using the fully sequenced genomes of two isolates previously identified as *C. manginecans* (CMW 17570) and *C. acaciivora* (CMW22621). These genomes are available on request from the authors and readers are also referred to the genome of the closely related species *C. fimbriata s.s.* (Wilken et al., 2014). Initial microsatellite regions were identified from the CMW22621 genome by analysing the assembled Illumina sequence contigs in Msatfinder v. 2.0 on-line (Thurston and Field, 2005). Parameters were set to detect tri, tetra, penta or hexa nucleotide motifs with a minimum of 9 repeat units. Contigs containing microsatellite regions were imported into CLC Genomics workbench v. 6.0 (CLC Bio, www.clcbio.com). To identify polymorphic SSR regions, the contigs that contained perfect microsatellites, optimal for primer design, were selected and compared with the orthologous region in CMW 17570 by means of BLASTn analysis. For the SSR regions that were polymorphic between isolates, primers were designed to flank these regions. The parameters for primer design included a primer length between 18 bp and 22 bp, a G/C content between 0.4 and 0.6 and an annealing temperature between 50°C and 62°C. Primers were also designed to produce fragment lengths that ranged between 180 and 500 bp to allow for multiplexing during GeneScan analyses.

The amplification success of the 27 available *C. fimbriata s.l.* SSR primer sets (Barnes et al., 2001; Steimel et al., 2004), as well as the newly designed primers, was evaluated in three *C. manginecans* isolates (CMW 22563, 38022, 38002). PCR reaction mixtures for the previously developed primers were identical to those for the ITS described above with the exception of the addition of 2 mM $MgCl_2$ per reaction. A multiplex PCR procedure was developed for the new SSR markers by combining a maximum of three primer sets in a single PCR reaction. The PCR reaction mixture consisted of 30 ng DNA, 1.5 units MyTaq™ DNA Polymerase (Bioline Ltd., UK), 5 µl 5x MyTaq™ reaction buffer and a total primer concentration of 0.32 µM for both the forward and reverse primers. Reaction volume was a total of 25 µl. The PCR cycler was programmed as follows: 95 °C for 5 min, 10 cycles of 95 °C for 30 s, 55 °C for 45 s, 72 °C for 1:30 min, followed by 30 cycles of 95 °C for 30 s, 55 °C for 45 s (with 5 s/cycle time increase), 72 °C for 1:30 min and a final step at 60 °C for 35 min. PCR amplification success was evaluated using AGE. Individual PCR amplicons, obtained from the new set of SSR primers, were sequenced for all three isolates to confirm that the allelic variation observed is due to mutations within the microsatellite repeat region.

The primer sets that amplified successfully in all three isolates where resynthesized and the forward primer labelled with either PET, VIC, NED or FAM fluorescent dyes from the G5 labelling kit (Applied Biosystems, Warrington, UK). These primers where screened on all isolates used in the study, including those from Oman, Pakistan, Indonesia and Vietnam. Three panels were designed for fragment analyses using GeneScan (Applied Biosystems, Foster City, California), while taking amplicon size and fluorescent dye colour into consideration, such that the amplicons of different primer sets could be pooled and run together in a single lane. A 1/200 dilution of each microsatellite product in a particular panel was pooled together. Of the pooled reaction mix, 1 µl was then combined with 0.4 µl GeneScan-500 Liz size standard (Applied Biosystems, Foster City, California) and 10 µl formamide and run on an ABI PRISM$_{TM}$ 3500xl Autosequencer (Applied BioSystems, Foster City, California, USA). Genescan data were analysed with GeneMapper® v.4.1 software (Life Technologies, Foster City, CA) to score allele fragment sizes. The 27 isolates from mango and bark beetles, considered by Al Adawi et al. (2014) and also used in the present study, were re-run on the ABI PRISM$_{TM}$ 3500xl Autosequencer and re-scored to avoid shifts in allele sizes experienced when using different apparatus.

All unique alleles (an allele present only in a single isolate) were sequenced to confirm the existence of the allele. In addition, various different alleles per locus were sequenced to confirm that the variation observed was due to the expansion or contraction in the microsatellite region and not in the flanking regions. PCR and sequencing reactions were carried out with identical conditions as described above, except that non-fluorescent primers were used.

2.4. Statistical analyses of SSR data

All the isolates analysed in the study were considered as four groups, representing host and/or geographic origin. The first group included isolates from mango in Oman and Pakistan, the second from *D. sissoo* and *P. cineraria* in the same countries, the third from *Acacia* spp. in Indonesia and the fourth from *Acacia* spp. in Vietnam. The number of alleles, allele frequencies and the gene diversity (H) (Nei, 1973) of each microsatellite locus in all four populations were calculated using POPGENE v. 1.32 (Yeh et al., 1999). The existence of private alleles and the different haplotypes present in each population were determined using GenAlEx v.6.2 (Peakall and Smouse, 2006). The genotypic diversity (Ĝ) of each population was calculated according to the method of Stoddard and Taylor (1988). The maximum genotypic diversity that can be obtained in a population, when every isolate has a distinct genotype, will be equal to the population size. To consider the differences in sample sizes when calculating genotypic diversity, Ĝ was divided by the population size to calculate

the percentage of the maximum genotypic diversity (Ĝ%) that was obtained in the population (Boeger et al., 1993). A t-test was performed to determine whether the differences in diversity between populations is significant at a confidence level of 95% (Chen et al., 1994). Due to the significant difference in population sample sizes, an alternative analysis was performed to confirm if the genotypic differences are significant. The genotypic diversities of each population were compared, based on Nei's (1987) genetic diversity index, and Bootstrap analysis was performed for these comparisons, using 1000 permutations, in the program GenoDive (Meirmans and Van Tienderen, 2004).

For analysis of population genetic differentiation, the analogue for Wright's F_{ST} value in haploid individuals, ϕ_{PT} (Excoffier et al., 1992), was calculated in GenAlEx v.6.2 (Peakall and Smouse, 2006) using an analysis of molecular variance (AMOVA). The ϕ_{PT} values were compared in a pair-wise manner between each of the populations to determine if the differentiation between any of the populations is significant. The AMOVA analysis was used to compare the variation within and between populations. Statistical significance of the differentiation in both analyses were assessed based on 1 000 permutations. A null hypothesis of no genetic variation among the populations was assumed, at a P value of 0.001.

To determine whether population structure was present among the investigated isolates, the program STRUCTURE v 2.2 (Pritchard et al., 2000) was used. Analysis in this program uses a Bayesian clustering method to assign each individual to a potential population without prior information of its origin. To determine the most likely number of clusters or populations (K), preliminary runs were performed for K-values ranging from one to six, using 20 iterations. The best parameters for analysis of the dataset were determined by evaluating the effect on the K-value when considering both models of admixture and no admixture as well as correlated and non-correlated allele frequencies among populations. Two approaches were used to determine the most likely K-value from the data obtained. Firstly, the K-value with the highest posterior probability for the data (Ln P(D)) (Pritchard et al., 2000), was determined by plotting the mean log likelihood (from all 20 runs) for each K. Secondly, K was identified based on the rate of variation between the LnP(D) of successive K-values by calculating the statistic ΔK (Evanno et al., 2005). A final Bayesian analysis was performed with the most likely K-value, applying a burn-in value of 100 000 generations with 2 million Markov Chain Monte Carlo (MCMC) generations and repeated three times. A model of admixture and allelic correlation was assumed. The final results were processed for visual representation using the program Distruct 1.1 (Rosenberg, 2004).

In order to calculate the genetic distance between the isolates, the proportion of shared alleles (Chakraborty and Jin, 1993) was used as a distance measure and a distance matrix

was constructed using the software Populations v. 1.2.32 (Langella, 1999). Subsequently, a neighbor joining (NJ) tree was constructed in the same program to visualise the calculated distances. One thousand bootstrap replicates were performed to ensure the results are statistically significant. The NJ tree was viewed and edited in MEGA v. 5 (Tamura et al., 2011).

3. Results

3.1. Fungal isolates and DNA extraction

A total of 161 isolates were used in this study (Table 1). These included 110 isolates collected from three *Acacia* plantation areas of Indonesia and 24 isolates from seven regions in Vietnam. For the Oman collection, 10 isolates were from mango, four from *H. mangiferae* infesting these trees and three were from *P. cineraria.* Isolates from Pakistan included four from mango from two locations, two from *H. mangiferae* and three from *D. sissoo,* collected from two different locations in Pakistan.

3.2. Confirmation of isolate identity

3.2.1. PCR amplification, sequencing and cloning

PCR amplification of the ITS region in all isolates resulted in the expected 550 bp fragment size and good quality sequences. However, the sequence chromatograms of the ITS region for 35 isolates from Indonesia and 17 from Vietnam had clear peaks up to ~120 bp after which the base calling showed conflicting sequence data (Fig. 2). PCR amplification of the cloned ITS sequences produced fragment sizes of about 1000bp and the sequence reactions for these amplicons resulted in clear, readable chromatograms. PCR amplification for the subset of 24 isolates gave amplicon sizes of 600 bp for the βT 1 and 800 bp for the EF 1-α gene regions.

3.2.2. Sequence alignment and phylogenetic analysis

Alignment of all the ITS sequences, including the cloned sequences, showed that the isolates had several different ITS haplotypes. The Indonesian isolates contained either the *C. acaciivora* or the *C. manginecans* haplotype, or minor variants of it. Where more than one clone was sequenced for an isolate, both haplotypes could be found in a single individual (Table 2; Table 3). In addition to the ITS sequence for *C. acaciivora* and *C. manginecans*, the sequences of the isolates from Vietnam also corresponded to *C. eucalypticola*, and two of the three haplotypes were consistently found in the cloned sequences of individual isolates (Table 2). The species identity of the isolates was confirmed in the MP phylogenetic tree, with all the isolates clustering with one or more of these three defined species (Fig. 3).

Sequence data for the βT 1 and EF 1-α gene regions could not distinguish between the subset of 24 isolates considered, regardless of the quality of their ITS sequences (MP tree not shown). The βT 1 region was identical in *C. acaciivora, C. manginecans, C. eucalypticola* and the isolates used in this study. The EF 1-α region contained two variable sites between *C. acaciivora, C. manginecans* and *C. eucalypticola* but all the isolates investigated had identical sequences to the *C. acaciivora* sequence (Table 3). For the three gene regions considered, ITS was the only phylogenetic marker that showed significant sequence differences (17 bp) between *C. acaciivora, C. manginecans* and *C. eucalypticola* (Table 3). Since the sequences of the other gene regions could not distinguish between the isolates and the ITS region displayed intragenomic variation, all isolates in this study were considered to represent the single species *C. manginecans*.

3.3. Development of microsatellite markers and GeneScan analysis

Twelve polymorphic SSR regions were identified from the genome sequences of two isolates, previously identified as *C. manginecans* and *C. acaciivora*. The sequences generated for each SSR region confirmed the presence of the microsatellite polymorphisms for 10 of the 12 markers. Nine of the markers were tri-nucleotide repeats and one was a tetra-nucleotide repeat (marker AF5). These 10 markers were used for further population analyses (Table 4).

Of the 27 available SSR primers (Barnes et al., 2001; Steimel et al., 2004), 19 amplified successfully in all 161 isolates (Table 5, file available on CD). The markers GACA650, CF13/14 and AG15/16 were monomorphic in all isolates investigated. CAT1 resulted in a fingerprint profile for some isolates for which the allele size could not be determined and CF11/12 produced allele sizes differing by a single nucleotide, which were not considered informative and they were thus excluded from further analyses. The remaining 14 SSR markers and the 10 newly designed markers, together, resulted in 70 alleles ranging from two to eight (CAA38) alleles per locus with an average of three per locus (Table 6). In three of the markers, CAA10, CAA15 and CAT3K, the most dominant allele was present in more than 90% of the isolates, indicating the low level of variation displayed by these markers (Table 6).

A total of 27 of the 70 alleles were private. They were either observed at high frequency but were unique to a single population or they occurred only once in that specific population (Table 6). The population from legume hosts in Oman and Pakistan included one private allele (166) in marker T3, which occurred in a single isolate of the population. The mango isolates also included one allele (306) unique to the population, in marker T6, which occurred at a frequency of 0.095. The population from Indonesia included seven private alleles with two alleles occurring in single isolates (Table 6). Isolates from Vietnam included

a total of 18 private alleles with four alleles occurring only once in the population. The DNA sequences of the private alleles confirmed that the variation observed between alleles was within the microsatellite repeat region, following a stepwise mutation pattern (Fig. 4).

3.4. Statistical analyses of SSR data

Analysis of the data obtained from the microsatellite markers, in POPGENE v. 1.32 (Yeh et al., 1999), showed a gene diversity ranging from 0.127 to 0.624 per locus with an average gene diversity of 0.437 (Table 6). The genetic variation in the different populations, including the number of alleles present, the number of haplotypes and the gene diversity in each population, is presented in Table 7. The mango population had the lowest average gene diversity (H=0.007) and the *Acacia* population from Vietnam had the highest (H=0.44). A total of 41 haplotypes were identified in all the *C. manginecans* isolates screened. The greatest number of haplotypes (23) and unique haplotypes (12) were found in the Indonesian population (Table 7), but this was also the population with the largest sample size.

The isolates from mango had the lowest (5.8%) and the isolates from *Acacia* from Vietnam had the highest (32.5%) percentage of maximum genotypic diversity (Table 7) but this was not statistically different between the populations. There were inordinately few isolates from *P. cineraria* and *D. sissoo* (the legume trees from Oman and Pakistan respectively) to reliably calculate genotypic diversity for them and they were not included in the comparisons. In contrast, bootstrap analyses showed significant differences ($P<0.025$) in genotypic diversity between some of the populations with posterior probability values of P(mango ≤ legume trees) = 0.001, P(*Acacia* Indonesia ≥ mango) = 0.001, P(*Acacia* Vietnam ≥ mango) = 0.001 and P (*Acacia* Vietnam ≥ *Acacia* Indonesia) = 0.011.

Analysis of the population differentiation, based on AMOVA, supported the existence of genetically distinct populations with an overall ϕ_{PT}=0.905 (P=0.001). Of the total variation observed for all the isolates, 91% was due to genetic differences among populations and only 9% due to variation within the population (Table 8). Pairwise comparisons showed that most of this variation can be ascribed to the *Acacia* populations from Vietnam and Indonesia and these were also genetically differentiated from the other populations. The Vietnam population had the highest genetic distinction with ϕ_{PT} values ranging from 0.846 to 0.942 (P=0.001) (Table 9). The Indonesia population was less differentiated from the mango and legume populations and no differentiation was observed between the isolates from mango and legume trees from Oman and Pakistan.

Analysis of population structure, based on a Bayesian clustering method, showed the most likely number of genetic clusters (K) to be two, followed by K=3 and K=4 (Fig. 5).

Preliminary runs indicated no significant difference in the calculation of K whether a model assuming admixture was considered or not or whether correlation between alleles was considered. When assuming the existence of only two clusters, 99% of the isolates from mango and legume trees in Oman and Pakistan, were assigned to the same cluster (Fig. 6A). Almost 43% of the isolates from Indonesia (*Acacia*) belonged to this group but only 14% of the isolates from Vietnam (*Acacia*) could be considered from the same cluster. The majority (86%) of isolates from Vietnam were assigned to a single cluster but this cluster showed signs of admixture with the Indonesia cluster.

Although STRUCTURE indicated K=2 as the most likely number of clusters, K=3 and K=4 was also presented (Fig. 6B and 6C). The latter values might be more representative of the expected number of clusters when considering the gene diversity between the populations together with the results from the AMOVA analyses. Where the isolates were considered to consist of three genetic clusters (Fig. 6B), those from Oman and Pakistan still formed a single, distinct cluster but 40% of the isolates from *Acacia* from Indonesia and 15% from Vietnam grouped with this cluster. The majority of isolates from Indonesia (58%) formed a distinct cluster and 85% of the Vietnam isolates were assigned to another cluster with very little admixture between the two. When four clusters were considered (Fig. 6C), the isolates from Vietnam still grouped in the same cluster as before but the Indonesia isolates could be ascribed to three potential clusters.

The genetic distance calculated between all the isolates, displayed in the NJ tree (Fig. 7), showed distinct groupings, clustering the isolates from mango and legume trees in Oman and Pakistan in a single clade with little variation. The isolates from Indonesia grouped together but this group was also sub-divided into smaller clusters, indicating a higher level of variation in the population. Some of the clusters corresponded with collection sites; for example three clusters included only isolates from one plantation area (Teso) in Indonesia and another cluster, only isolates from the Pelalawan plantation area. One of the clusters consisting of isolates from Teso, had a high genetic relatedness to the isolates from Oman and Pakistan. Comparison of the original microsatellite data (Table 5, file available on CD) showed that these isolates differed at only one or two alleles from the Oman/Pakistan mango population. The Vietnam isolates grouped in a cluster significantly distinct from all other populations. The haplotypes of the Vietnam isolates showed no distinct clustering correlated to the collection sites with a single haplotype identified in three regions which are 800 km and 1700 km apart (Fig. 1B).

4. Discussion

Based on DNA sequence comparisons, this study confirmed that the isolates associated with severe wilt of *Acacia* spp. in Vietnam represent the same species as that

killing *A. mangium* in Indonesia and mango in Oman and Pakistan. Thus, all of the isolates investigated were confirmed to represent the single species *C. manginecans*. Furthermore, the microsatellite markers showed significantly higher genetic diversity among the isolates from South East Asia, compared to those from the Middle East. The highest genetic diversity was found in the isolates from Vietnam, suggesting the fungal pathogen has existed in that area for a long period of time and it might be considered native there.

Special effort was made in this study to confirm the identity of the isolates collected from *Acacia* spp. in South East Asia. Sequence data for the βT 1 and EF 1-α gene regions could not separate the isolates representing *C. acaciivora, C. eucalypticola* and *C. manginecans*. Furthermore, in a previous study, data from various other genetic markers have shown that *C. manginecans* and *C. acaciivora* represent the same species (Chapter 2). The three species could be separated based on the ITS region but we have shown in this study and consistent with the results of Naidoo et al (2013), that more than one of the three ITS haplotypes can occur in a single isolate. The emerging data strongly suggests that the isolates from *Acacia* spp. in Indonesia and Vietnam and those from mango and legume trees in Oman and Pakistan represent a single species. Following chronological precedence, *C. manginecans* is currently the most appropriate name for this fungus.

The new microsatellite markers designed in this study detected similar values of genetic variation for the *C. manginecans* isolates as the previously designed (Barnes et al., 2001; Steimel et al., 2004), broad range markers. It has been suggested that the cross-species transfer of microsatellite markers could show less polymorphism in a different species (Barbará et al., 2007). However, the gene diversity for the entire set of samples ranged from 0.254 to 0.624 for each of the new markers with an average gene diversity of 0.393, which is barely higher than for the markers developed by Barnes et al. (2001) and Stiemel et al. (2004). This provides support that either the new or the previously designed markers would give a fair representation of the actual diversity in the populations being studied.

The identical haplotype present in the isolates from mango was found on the native legume trees in Oman and Pakistan (Fig. 7). This raises the prospect that *C. manginecans*, clearly an invasive alien pathogen, has undergone a host jump from non-native and commercially propagated fruit trees to native trees. The low levels of genetic variation, previously observed in the isolates from mango and *H. mangiferae* in Oman and Pakistan (Al Adawi et al., 2014), were consistent in the new microsatellite markers utilised in the present study. The similarly low values identified in the isolates from legume trees supports the hypothesis that the species was introduced and that a single population has spread widely throughout the countries to different hosts (Al Adawi et al., 2014).

The *C. manginecans* population from *Acacia* spp. in Vietnam had high levels of genetic diversity. The values observed in these isolates (H=0.437 and %G=32.4) are comparable with those for other populations of *C. fimbriata s.l.* that are considered native to Brazil (H=0.382) and Africa (H=0.41 and %G=44), respectively (Barnes et al., 2005; Ferreira et al., 2010). This indicates that the pathogen has occurred in this area for a long period of time. The possible existence of an indigenous pathogen would explain the serious disease that it has begun to cause on a non-native *Acacia* spp., introduced from Australia (Roux and Wingfield, 2009; Sein and Mitlöhner, 2011).

The genetic diversity observed for *C. manginecans* in Vietnam was higher than that found for the same fungus in Indonesia. This is substantiated by the high proportion of unique haplotypes found in the isolates from Vietnam (29% of the isolates) compared to Indonesia (12%). A higher number of sites were investigated in Vietnam (eight) and these were widely distributed across the country, whereas the three regions in Indonesia were in closer proximity to one another. This could influence the level of diversity observed in the countries but will require additional sampling to confirm. Overall, the populations in South East Asia are genetically diverse and the results suggest that the fungus has been present in this area for a long period, thus accumulating variation through events such as mutation or recombination (McDonald, 1997). This would increase the possibility of an association emerging between the pathogen and its insect vector, *H. mangiferae*, as the beetle is native to South East Asia (Atkinson and Peck, 1994; Doran and Skelton, 1982; Moran et al., 1989; Mukherjee, 1972).

All the isolates used in this study, collected from the Middle East and South East Asia, could reliably be separated into three to four distinct clusters. The AMOVA analysis and ϕ_{PT} values provided support for the populations from Indonesia and Vietnam to be genetically distinct from one another and from the isolates collected from the Middle East. Nonetheless, the population from Indonesia appears less differentiated from the populations in Oman and Pakistan. The haplotype of one of the isolates found in Indonesia differed from the Oman and Pakistan isolates by a single allele. This difference might have been a recent mutation event in a single isolate and it is likely that these populations shared a common isolate in the past.

Results of this study have substantially expanded available knowledge regarding the geographical distribution of the important tree pathogen, *C. manginecans*. Previously reported on mango and legume trees in the Middle East as well as on non-native *Acacia* spp. in Indonesia (Al Adawi et al., 2013b; Tarigan et al., 2011), this study confirmed that the pathogen also occurs on *Acacia* spp. in Vietnam. The same fungus occurs on *Acacia* spp. in Malaysia (Wingfield, unpublished) and on *Eucalyptus* trees in South China (Chen et al.,

2013) and Brazil (Harrington et al., 2014). *Ceratocystis manginecans* has become a serious threat to fruit and forest plantation industries in the Middle East and South East Asian countries. The SSR markers used in this study should in future be screened on isolates from these countries to determine the genetic structure of additional populations. This information could assist in determining the distribution of the pathogen as well as in establishing effective quarantine and trade regulations between countries (Barney, 2005; Oberndorf, 2013).

5. References

Al Adawi, A.O., Al Jabri, R.M., Deadman, M.L., Barnes, I., Wingfield, B., Wingfield, M.J., 2013a. The mango sudden decline pathogen, *Ceratocystis manginecans*, is vectored by *Hypocryphalus mangiferae* (Coleoptera: Scolytinae) in Oman. Eur. J. Plant Pathol. 135, 243-251.

Al Adawi, A.O., Barnes, I., Khan, I.A., Al Subhi, A.M., Al Jahwari, A.A., Deadman, M.L., Wingfield, B.D., Wingfield, M.J., 2013b. *Ceratocystis manginecans* associated with a serious wilt disease of two native legume trees in Oman and Pakistan. Australas. Plant Pathol. 42, 179-193.

Al Adawi, A.O., Barnes, I., Khan, I.A., Deadman, M.L., Wingfield, B.D., Wingfield, M.J., 2014. Clonal structure of *Ceratocystis manginecans* populations from mango wilt disease in Oman and Pakistan. Australas. Plant Pathol. 10.1007/s13313-014-0280-0.

Al Adawi, A.O., Deadman, M.L., Al Rawahi, A.K., Al Maqbali, Y.M., Al Jahwari, A.A., Al Saadi, B.A., Al Amri, I.S., Wingfield, M.J., 2006. Aetiology and causal agents of mango sudden decline disease in the Sultanate of Oman. Eur. J. Plant Pathol. 116, 247-254.

Al Adawi, A.O., Deadman, M.L., Al Rawahi, A.K., Khan, A.J., Al Maqbali, Y.M., 2003. *Diplodia theobromae* associated with sudden decline of mango in the Sultanate of Oman. Plant Pathol. 52, 419-419.

Atkinson, T.H., Peck, S.B., 1994. Annotated Checklist of the Bark and Ambrosia Beetles (Coleoptera: Platypodidae and Scolytidae) of Tropical Southern Florida. Fla. Entomol. 77, 313-329.

Barbará, T., Palma-Silva, C., Paggi, G.M., Bered, F., Fay, M.F., Lexer, C., 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. Mol. Ecol. 16, 3759-3767.

Barnes, I., Guar, A., Burgess, T., Roux, J., Wingfield, B.D., Wingfield, M.J., 2001. Microsatellite markers reflect intra-specific relationships between isolates of the vascular wilt pathogen *Ceratocystis fimbriata*. Mol. Plant Pathol. 2, 319-325.

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

Barney, K., 2005. Central Plans and Global Exports: Tracking Vietnam's Forestry Commodity Chains and Export Links to China. Forest Trends, Washington.

Boeger, J.M., Chen, R.S., McDonald, B.A., 1993. Gene Flow Between Geographic Populations of *Mycosphaerella graminicola* (Anamorph *Septoria tritici*) Detected with Restriction Fragment Length Polymorphism Markers. Phytopathology 83, 1148-1154.

Chakraborty, R., Jin, L., 1993. A unified approach to study hypervariable polymorphisms: Statistical considerations of determining relatedness and population distances. In: Pena, S.D.J., Chakraborty, R., Epplen, J.T., Jeffreys, A. (Eds.), DNA Fingerprinting: State of the Science. Birkhäuser Basel, pp. 153-175.

Chen, R.S., Boeger, J.M., McDonald, B.A., 1994. Genetic stability in a population of a plant pathogenic fungus over time. Mol. Ecol. 3, 209-218.

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

Doran, J.C., Skelton, D.J., 1982. *Acacia mangium* seed collections for international provenance trials. Forest Genetics Resource Information, Food and Agriculture Organization, Rome, Italy, pp. 47-53.

Engelbrecht, C.J.B., Harrington, T.C., 2005. Intersterility, morphology and taxonomy of *Ceratocystis fimbriata* on sweet potato, cacao and sycamore. Mycologia 97, 57-69.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611-2620.

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479-491.

FAOSTAT, 2011. FAOSTAT On-line. Rome. United Nations Food and Agriculture Organization. Available at: http://faostat3.fao.org/faostat-gateway/go/to/home/E. Accessed: 9 September 2013.

Ferreira, E.M., Harrington, T.C., Thorpe, D.J., Alfenas, A.C., 2010. Genetic diversity and interfertility among highly differentiated populations of *Ceratocystis fimbriata* in Brazil. Plant Pathol. 59, 721-735.

FSIV, 2009. Asia-Pacific Forestry Sector Outlook Study II: Vietnam Forestry Outlook Study. Food and Agriculture Organization of the United Nations, Bangkok.

Glass, N.L., Donaldson, G.C., 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. Appl. Environ. Microbiol. 61, 1323-1330.

Goodwin, S.B., Drenth, A., Fry, W.E., 1992. Cloning and genetic analysis of two highly polymorphic, moderately repetitive nuclear DNAs from *Phytophthora infestans*. Curr. Genet. 22, 107-115.

Harrington, T.C., Kazmi, M.R., Al-Sadi, A.M., Ismail, S.I., 2014. Intraspecific and intragenomic variability of ITS rDNA sequences reveals taxonomic problems in *Ceratocystis fimbirata sensu stricto*. Mycologia In press. doi: 10.3852/13-189.

Jacobs, K., Bergdahl, D.R., Wingfield, M.J., Halik, S., Seifert, K.A., Bright, D.E., Wingfield, B.D., 2004. *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. Mycol. Res. 108, 411-418.

Kumar, B., Mistry, N.C., Singh, B., Gandhi, C.P., 2011. Indian Horticulture Database National Horticulture Board, pp. 1-296.

Langella, O., 1999. Populations 1.2.30. http://www.bioinformatics.org/tryphon/populations/.

Maslin, B., 2013. KEW Royal Botanical Gardens (*Acacia mangium*). Available at http://www.kew.org/science-conservation/plants-fungi/acacia-mangium-brown-salwood. Accessed: 9 July 2013.

Masood, A., Saeed, S., Erbilgin, N., Jung Kwon, Y., 2010. Role of stressed mango host conditions in attraction of and colonization by the mango bark beetle *Hypocryphalus mangiferae* Stebbing (Coleoptera: Curculionidae: Scolytinae) and in the symptom development of quick decline of mango trees in Pakistan. Entomol. Res. 40, 316-327.

Masood, A., Saeed, S., Sajjad, A., 2008. Characterization and damage patterns of different bark beetle species associated with mango sudden death syndrome in Punjab, Pakistan. Pakistan Entomol. 30, 163-168.

McDonald, B.A., 1997. The population genetics of fungi: Tools and techniques. Phytopathology 87, 448-453.

Meirmans, P.G., Van Tienderen, P.H., 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. 4, 792–794.

Moran, G.F., Muona, O., Bell, J.C., 1989. *Acacia mangium*: A tropical forest tree of the coastal lowlands with low genetic diversity. Evolution 43, 231-235.

Mukherjee, S.K., 1972. Origin of mango (*Mangifera indica*). Econ. Bot. 26, 260-264.

Naidoo, K., Steenkamp, E.T., Coetzee, M.P.A., Wingfield, M.J., Wingfield, B.D., 2013. Concerted Evolution in the Ribosomal RNA Cistron. PLoS ONE 8, e59355. doi: 59310.51371/journal.pone.0059355.

Nei, M., 1973. Analysis of Gene Diversity in Subdivided Populations. Proc. Natl. Acad. Sci. 70, 3321-3323.

Nei, M., 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Oberndorf, R.B., 2013. A Review of Progress and the Role of the Asia Forest Partnership 2002-2012. Combating Illegal Logging in Asia. Forest Trends.

Ocasio-Morales, R.G., Tsopelas, P., Harrington, T.C., 2007. Origin of *Ceratocystis platani* on native *Platanus orientalis* in Greece and its impact on natural forests. Plant Disease 91, 901-904.

Peakall, R.O.D., Smouse, P.E., 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288-295.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155, 945-959.

Rosenberg, N.A., 2004. Distruct: a program for the graphical display of population structure. Mol. Ecol. Notes 4, 137–138.

Roux, J., Wingfield, M.J., 2009. *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. South. For. 71, 115-120.

Roux, J., Wingfield, M.J., 2013. *Ceratocystis* species on the African continent, with particular reference to *C. albifundus*, an African species in the *C. fimbriata sensu lato* species complex. In: Seifert, K.A., De Beer, Z.W., Wingfield, M.J. (Eds.), Ophiostomatoid Fungi: Expanding Frontiers. CBS-KNAW Fungal Biodiversity Centre, AD Utrecht, The Netherlands, pp. 131-138.

Sein, C.C., Mitlöhner, R., 2011. *Acacia mangium* Willd. Ecology and silviculture in Vietnam. CIFOR, pp. 1-26.

Steimel, J., Engelbrecht, C.J.B., Harrington, T.C., 2004. Development and characterization of microsatellite markers for the fungus *Ceratocystis fimbriata*. Mol. Ecol. Notes 4, 215-218.

Stoddart, J.A., Taylor, J.F., 1988. Genotypic diversity: estimation and prediction in samples. Genetics 118, 705-711.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sinauer Associates, Sunderland, MA.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28, 2731-2739.

Tarigan, M., Roux, J., Van Wyk, M., Tjahjono, B., Wingfield, M.J., 2011. A new wilt and die-back disease of *Acacia mangium* associated with *Ceratocystis manginecans* and *C. acaciivora sp*. nov. in Indonesia. SA J. Bot. 77, 292-304.

Thu, P.Q., Qynh, D.N., Dell, B., 2012. *Ceratocystis* sp. causes crown wilt of *Acacia* spp. planted in some ecological zones of Vietnam. Unpublished.

Thurston, M.I., Field, D., 2005. Msatfinder on-line v. 2.0 (http://www.genomics.ceh.ac.uk/msatfinder/).

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

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

Wilken, P.M., Steenkamp, E.T., Wingfield, M.J., De Beer, Z.W., Wingfield, B.D., 2014. Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. IMA Fungus 4, 357–358.

Wingfield, M.J., Roux, J., Wingfield, B.D., Slippers, B., 2013. *Ceratocystis* and *Ophiostoma*: international spread, new associations and plant health. In: Keith A. Seifert, De Beer, W., Wingfield, M.J. (Eds.), Ophiostomatoid Fungi: Expanding Frontiers. CBS-KNAW Fungal Biodiversity Centre, Netherlands, pp. 191-200.

Yeh, F.C., Yang, R., Boyle, T., 1999. POPGENE Version 1.31: Microsoft window-based freeware for population genetic analysis. University of Alberta and Centre for International Forestry Research.

6. Tables and Figures

Table 1

Details on the *C. manginecans* isolates collected in Oman, Pakistan, Indonesia and Vietnam, used in this study, which includes the host, geographic location and sample sizes obtained from different regions.

Country	CMW number[a]	Host	Location	Collector and date collected	Sample size
Oman	15377, 15353	*Mangifera indica*	Shinas	A. O. Al Adawi (2003-2004) [b]	2
	15366, 15369, 15385, 15391	"	Liwa		4
	15313-15316	"	Sohar		4
	15371	"	Quariyat		1
	15317	*Hypocryphalus mangiferae*	Sohar		1
	15384	"	Liwa		1
	15381, 15382	"	Quariyat		2
	17225, 17568, 17570	*Prosopis cineraria*	Sohar	A. O. Al Adawi (2004) [c]	3
					Total 18
Pakistan	17567	*M. indica*	Faisalabad	A. O. Al Adawi (2003-2004) [b]	1
	23637, 23642, 23643	"	Multan		3
	23628, 23630	*H. mangiferae*	Faisalabad		2
	23623, 23624	*Dalbergia sissoo*	"	A. O. Al Adawi (2004) [c]	2
	23625	"	Shorkot		1
					Total 9
Indonesia	22560, 22561, 22563-22580	*Acacia mangium*	Pelalawan	M. Tarigan (2006)	20
	22581-22585	"	Logas		5
	22587-22591, 22593-22619, 22621-22623,22625,22626	"	Teso		37
	37825-37840, 37842-37871, 37925, 37926	"	Riau province		48
					Total 110
Vietnam	38000-38004	*A. auriculiformis*	Thua Thien Hue	P. Q. Thu (2008) [d]	5
	38005, 38012-38019	*A. mangium*	"		9
	38020	"	Quang Ninh		1

	38022, 38023	"	Tuyen Quang	2
	38006	*Acacia hybrid*	Binh Duong	1
	38007	"	Thua Thien Hue	1
	38008	"	Binh Phuoc	1
	38009-38011	"	Dong Nai	3
	38021	"	Lam Dong	1
				Total 24

a All the *C. manginecans* isolates are maintained in the culture collection (CMW) of the Forestry and Agriculture Biotechnology Institute (FABI), University of Pretoria, South Africa.
b Al Adawi et al., 2014
c Al Adawi et al., 2013b
d Thu et al., 2012

Table 2

Results for the sequences obtained from the cloned ITS amplicons in a subset of isolates from Indonesia and Vietnam.

CMW number[a]	Host	Location	No of Clones	Haplotypes for ITS[b]	*C. acaciivora* haplotype	*C. manginecans* haplotype	*C. eucalypticola* haplotype	Different haplotype[c]
Vietnam								
38001	*A. auriculiformis*	Thua Thien Hue	3	E			3	
38004	*A. auriculiformis*	Thua Thien Hue	4	E, M		1	3	
38007	*Acacia hybrid*	Thua Thien Hue	4	A, M	3	1		
38008	*Acacia hybrid*	Binh Phuoc	4	E, M		3	1	
38009	*Acacia hybrid*	Dong Nai	4	E			4	
38010	*Acacia hybrid*	Dong Nai	4	A, E	2		2	
38011	*Acacia hybrid*	Dong Nai	4	A, E	3		1	
38012	*Acacia mangium*	Thua Thien Hue	4	A, E	2		2	
38013	*Acacia mangium*	Thua Thien Hue	4	A, E	1		3	
38014	*Acacia mangium*	Thua Thien Hue	5	A, E	3		1	1
38015	*Acacia mangium*	Thua Thien Hue	5	A, E	3		2	
38017	*Acacia mangium*	Thua Thien Hue	5	A, E	1		3	1
38018	*Acacia mangium*	Thua Thien Hue	5	A, E	3		2	
38019	*Acacia mangium*	Thua Thien Hue	5	A, E	3		2	
Indonesia								
22565	*Acacia mangium*	Pelelawan	4	A, M	3	1		
22566	*Acacia mangium*	Pelelawan	4	A, M	1	3		
22567	*Acacia mangium*	Pelelawan	3	A, M	2	1		
22568	*Acacia mangium*	Pelelawan	5	A, M	4	1		
22569	*Acacia mangium*	Pelelawan	3	A, M	2	1		
22608	*Acacia mangium*	Teso	5	A	5			
22610	*Acacia mangium*	Teso	4	M		4		
22611	*Acacia mangium*	Teso	4	A	4			

a All the *C. manginecans* isolates are maintained in the culture collection (CMW) of the Forestry and Agriculture Biotechnology Institute (FABI), University of Pretoria, South Africa.
b The cloned ITS sequences of an individual isolate corresponded to one of three described species and these are indicated for every isolate by three different letters; *C. acaciivora* (A), *C. manginecans* (M) and *C. eucalypticola* (E).
c Some of the cloned sequences had more than 1% nucleotide differences from the described species and these were then considered as a different haplotype.

Table 3

Nucleotide differences between *C. acaciivora, C. manginecans, C. eucalypticola* and three isolates from Vietnam in the two gene regions ITS and EF 1-α, used for identification of *C. fimbriata s.l.* isolates.

CMW no.	Species[a]	Genbank acc. no.	ITS															EF 1-α	
			13	131	132	135	136	139	321	353	375	390	392	398	401	417	521	74	714
22563	*C. acaciivora*	EU588656.1 EU588646.1	-	C	A	A	G	G	T	A	A	-	C	C	A	A	-	C	C
22564	*C. acaciivora*	EU588657.1 EU588647.1	-	C	A	A	G	G	T	A	A	-	C	C	A	A	-	C	C
10000	*C. eucalypticola*	FJ236722.1 FJ236752	-	A	G	T	A	-	T	-	A	-	C	T	-	-	A	C	C
11536	*C. eucalypticola*	FJ236723.1 FJ236753.1	-	A	G	T	A	-	T	-	A	-	T	T	-	-	A	T	C
13851	*C. manginecans*	AY953383.1 EF433317	G	C	A	A	G	G	-	A	-	T	T	T	-	-	-	C	A
13852	*C. manginecans*	AY953384.1 EF433318	G	C	A	A	G	G	-	A	-	T	T	T	-	-	-	C	A
38007	Clone C of 38007		G	C	A	A	G	G	-	A	-	T	T	T	-	-	-	C	C
38007	Clone D of 38007		-	C	A	A	G	G	T	A	A	-	C	C	A	A	-		
38019	Clone A of 38019		-	A	G	T	A	-	T	-	A	-	T	T	-	-	A	C	C
38019	Clone B of 38019		-	C	A	A	G	G	T	A	A	-	C	C	A	A	-		
38004	Clone D of 38004		G	C	A	A	G	G	-	A	-	T	T	T	-	-	-	C	C
38004	Clone A of 38004		-	A	G	T	A	-	T	-	A	-	T	T	-	-	A		

a The last six sequences in the table represent two cloned sequences, each from three Vietnam isolates CMW38004, 38007 and 38019. Each of the two cloned sequences in a single isolate corresponded to that of two different species.

Table 4
Details of the new polymorphic microsatellite markers, developed from the genomes of *C. acaciivora* and *C. manginecans.*

SSR locus	SSR motif in *C. acaciivora* / *C. manginecans*	Forward primer sequence 5'-3' Reverse primer sequence 5'-3'	Fluores-cent label	Tm (°C)	Multiplex PCR no.[a]	GenBank acc. no.
AF2	(AGA)10/ (AGA)12	CATTCTCGAAACACTAGCG AGGAGAGAAAGGATGGTGG	VIC	57	1	KJ601490
AF3	(CTG)12/ (CTG)13	AAGAAGAAGGAAAGCATCCG ACATCAACATCGTTTCTAGCCA	NED	57	3	KJ601491
AF4	(ACA)10/ (ACA)8	CTGTTTGACGGCTTTGGA TGCTAATGGAGGTCGGTG	PET	57	1	KJ601492
AF5	(GTCA)9/ (GTCA)6	TGTTCTTCTGATTGTGCACT GAGGTTGGCGTTGGTTAG	FAM	57	5	KJ601493
AF6	(GAG)13/ (GAG)11	CTATTGCGAGTTCAAGGC ACCCCTCATGATTCACTTAC	VIC	55	2	KJ601494
AF7	(AGC)10/ (AGC)9	CCTACATCTTCTTTGAGCCCTT GTTGTGGCTGCTGGGTTT	NED	60	1	KJ601495
AF8	(GAG)12/ (GAG)11	CTATCTGTCCTTGCCCCT CGGGCCTTTCTTTTGTTTCTT	PET	59	5	KJ601496
AF9	(GCA)9/ (GCA)11	ACTCTACTACCCTCACAC GACTAGGCCTCCATTGAA	FAM	55	2	KJ601497
AF11	(ACA)10/ (ACA)7	TTGGACATACTTGGACGGG ATTTAGTGGGAATCTGCGG	VIC	57	3	KJ601498
AF12	(ACA)17/ (ACA)12	ACAGGTAAGAAGGGACAGAA GATAAGGAGAGTGGGAAAGG	PET	60	4	KJ601499

a PCR amplifications of the microsatellite regions were performed as multiplex reactions by pooling two to three primer sets in a single reaction. Primer sets with the same multiplex number were combined in a reaction.

Table 5 – Available as additional excel file on CD

GeneScan results obtained for all 161 *C. manginecans* isolates and all 24 microsatellite gene regions investigated in this study, indicating the allele sizes obtained for each isolate.

Table 6

Summary statistics, including the number of alleles, allele frequencies, allelic range and gene diversity for each microsatellite marker screened on 161 *C. manginecans* isolates.

		Private alleles [Frequency in population][b]						
SSR locus[a]	Number of alleles	Legumes (Oman and Pakistan)	Mango (Oman and Pakistan)	*Acacia* (Indonesia)	*Acacia* (Vietnam)	Allelic range	Highest allele frequency	Gene diversity[c]
AF2	3				**198** [0.042]	198-206	0.851	0.256
AF3	3				**233** [0.042]	218-233	0.503	0.506
AF4	3				**252** [0.292]	243-252	0.857	0.254
AF5	3				**270** [0.042]	250-270	0.721	0.406
AF6	2			**309** [0.009]		294-300	0.596	0.486
AF7	2					322-325	0.658	0.45
AF8	2					346-349	0.665	0.446
AF9	2					412-418	0.839	0.271
AF11	3				**450** [0.375]	447-455	0.87	0.236
AF12	6			**456** [0.303], **459** [0.028]	**444** [0.042], **447** [0.333]	438-459	0.55	0.624
AG7/AG8	3				**293** [0.083]	284-293	0.528	0.51
CF15/CF16	2					474-476	0.5432	0.498
CF23/CF24	2					157-160	0.5217	0.499
AAG8	2				**180** [0.875]	177-180	0.87	0.227
AAG9	3			**400** [0109]		397-403	0.503	0.563
CAA9	4	**166** [0.167]			**172** [0.667]	166-223	0.8571	0.254
CAA10	3				**135** [0.250], **138** [0.417]	126-138	0.901	0.184
CAA15	2				**320** [0.458]	320-323	0.926	0.138
CAA38	8		**306** [0.095]	**288** [0.009], **324** [0.136], **369** [0.027]	**207** [0.583], **234** [0.375]	207-369	0.603	0.602
CAA80	3				**316** [0.5]	304-316	0.696	0.458
CAT3K	2				**306** [0.458]	306-310	0.932	0.127
CAT9X	3				**279** [0.792]	279-286	0.851	0.261
CAT1200	2					376-379	0.857	0.245
CAG5	2					319-322	0.894	0.189

a Microsatellite markers indicated in bold were developed in this study.
b Private alleles are indicated in bold. These refer to alleles observed only once in a population and to alleles unique to a specific population, regardless of the frequency of the allele.
c Gene diversity was calculated for each microsatellite marker, based on Nei (1973) gene diversity calculations.

Table 7

Summary statistics of isolates collected from Oman, Pakistan, Indonesia and Vietnam indicating variation within and between populations.

Population	No. of isolates	No. of alleles	No. of alleles/ locus	No. of private alleles	No. of haplotypes (Private haplotypes[a])	Most frequent haplotype[b]	Gene diversity[c]	Genotypic diversity (Ĝ)[d]	Max % of Ĝ
Mango (Oman and Pakistan)	21	25	1.04	1	2 (0)	92%	0.007	1.208	5.753
Legumes (Oman and Pakistan)	6	27	1.13	1	4 (3)	50%	0.035	n/a	n/a
Acacia (Indonesia)	110	50	2.08	7	23 (12)	31%	0.233	6.402	5.82
Acacia (Vietnam)	24	60	2.5	18	12 (6)	21%	0.437	7.784	32.43

a Refers to haplotypes that are unique to the population and occurs only once in the population.
b Indicates the percentage of isolates in the population that contain the most frequent haplotype.
c The average gene diversity, obtained from all the microsatellite markers combined, was calculated in each population, based on Nei (1973) gene diversity calculations.
d Genotypic diversity was calculated using the method of Stoddard and Taylor (1988) and the %Ĝ was obtained by dividing the Ĝ value with the population size and multiplied by 100.

Table 8

Summary of the analysis of molecular variance (AMOVA) of the *C. manginecans* isolates collected from Oman, Pakistan, Indonesia and Vietnam to evaluate the genetic variation within and between populations.

Source	df[a]	SS[b]	Estimated variance	Proportion of variance components (%)
Among populations	3	284619.348	3574.825	91%
Within populations	157	58624.254	373.403	9%
Total	160	343243.602	3948.228	100%

a=degrees of freedom
b=sum of squares

Table 9

Pairwise comparison of population differentiation (ϕPT) among the four populations investigated in this study.

Sampling group	Oman and Pakistan Legumes	Oman and Pakistan Mango	Indonesia *Acacia*
Oman and Pakistan Legumes	-	-	-
Oman and Pakistan Mango	0.135*	-	-
Indonesia *Acacia*	0.498	0.523	-
Vietnam *Acacia*	0.846	0.905	0.942

*Value is not statistically significant with a P-value=0.001

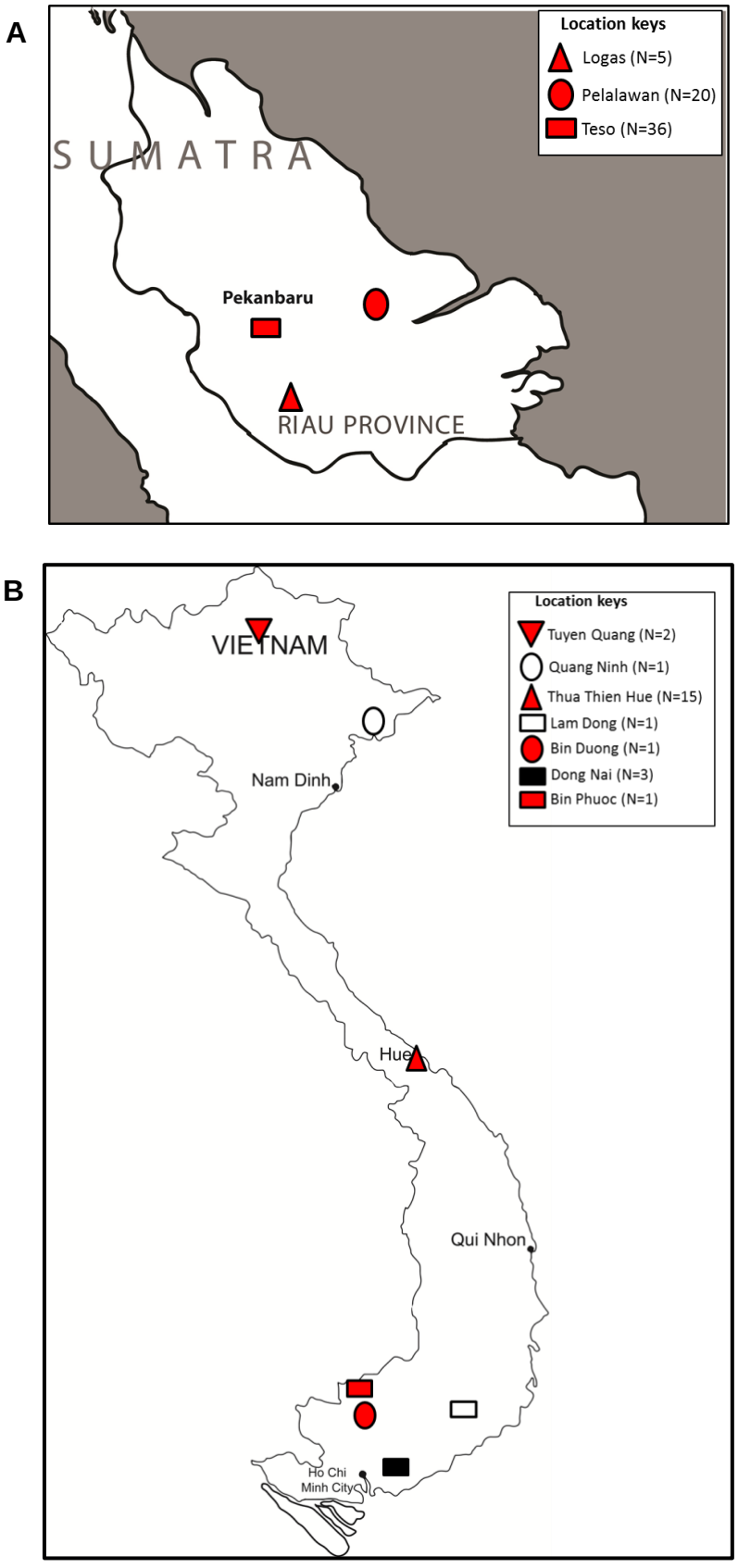


Fig. 1. Maps of the Riau province in the Sumatra island of Indonesia (A) and Vietnam (B) in South East Asia, indicating the collection sites and the number of *C. manginecans* isolates ("N") used in this study. The different collection sites are indicated by different keys on the maps.

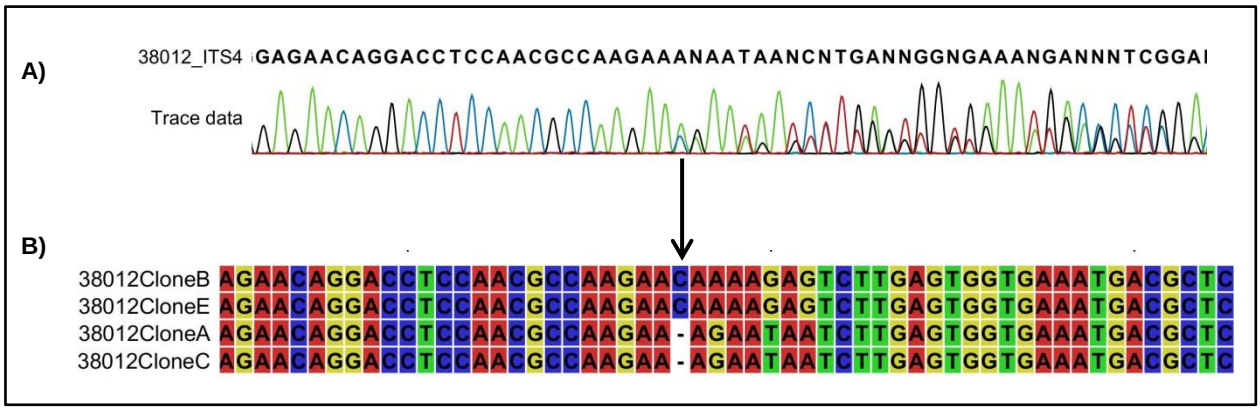


Fig. 2. A sequence chromatogram of the ITS region amplified in an isolate from Vietnam. A) A good quality sequence up to 120 bp after which there was conflict in the base calling. This is a clear illustration of the presence of more than one ITS type in a single isolate. B) Sequence data from the cloned sequences show that the presence of a single nucleotide indel (indicated by the arrow) resulted in a frame shift of the downstream sequence in the one ITS type.

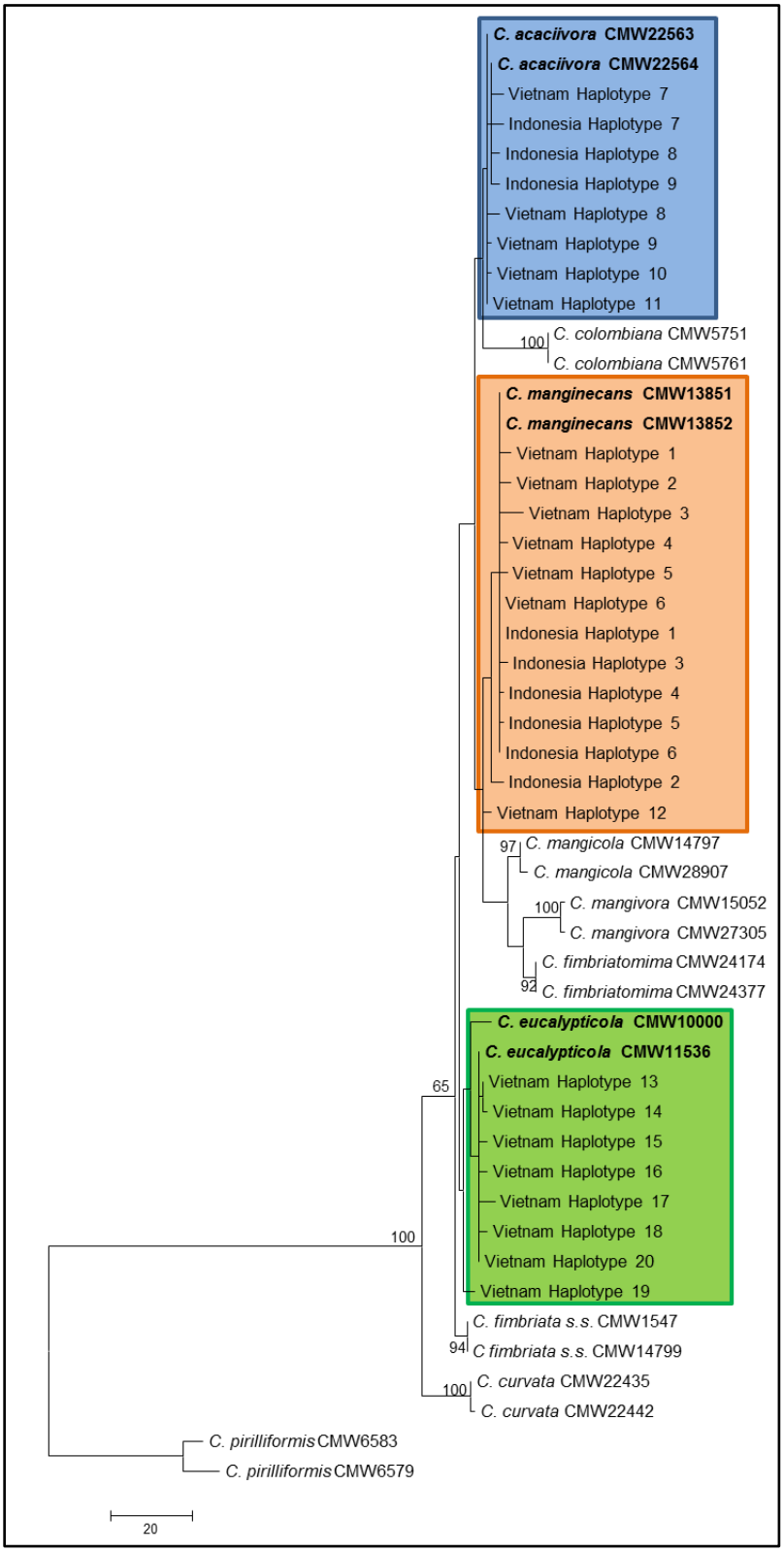


Fig. 3. One of the most parsimonious trees obtained from Maximum parsimony analysis of the ITS sequences of all isolates collected in Indonesia and Vietnam. To reduce the number of taxa in the tree, isolates with identical haplotypes were represented by a single sequence/haplotype. The sequences were compared to that of other, well-defined, species in the *C. fimbriata s.l.* species complex. All the Indonesia isolates clustered either with the *C. acaciivora* or *C. manginecans* species but some Vietnam isolates also clustered with *C. eucalypticola* species. Branch support values are indicated above the branches. The branch length represents the number of nucleotide differences.

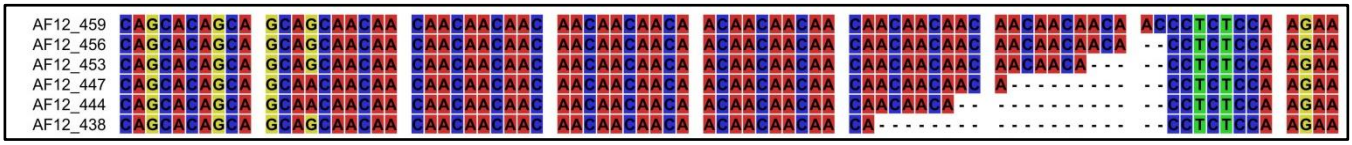


Fig. 4. Alignment of sequences representing all the alleles observed in the AF12 microsatellite marker. Some of these alleles were only observed in a single isolate and had to be sequenced to confirm that the length variation between the alleles are due to a step-wise mutation in the microsatellite region. The represented alignment confirms this. Alignment was performed in CLC Bio Main Workbench v. 6 (CLC Bio, www.clcbio.com).

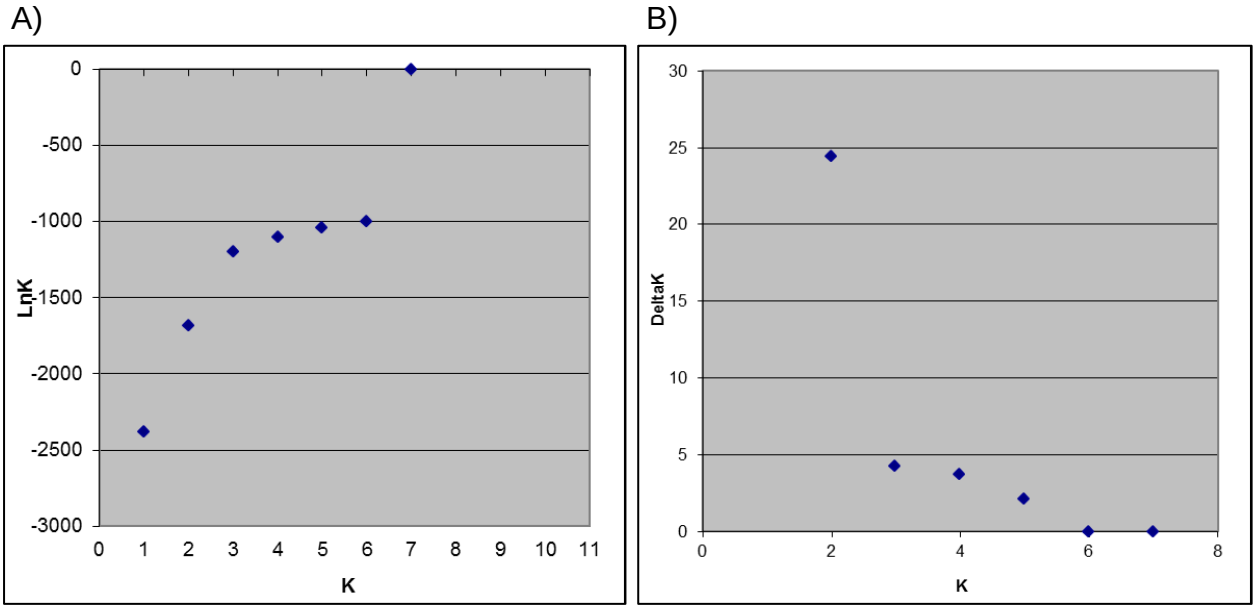


Fig. 5. Plot data of the A) K against LnK values for K=1-7 and the B) K against ΔK values to determine the most likely number of genetic clusters (K) all the isolates can be grouped in. The optimal K value is the value with the biggest rate of variation when comparing the LnP(D) to successive K-values. This result was used for a final population structure analysis in STRUCTURE v2.2.

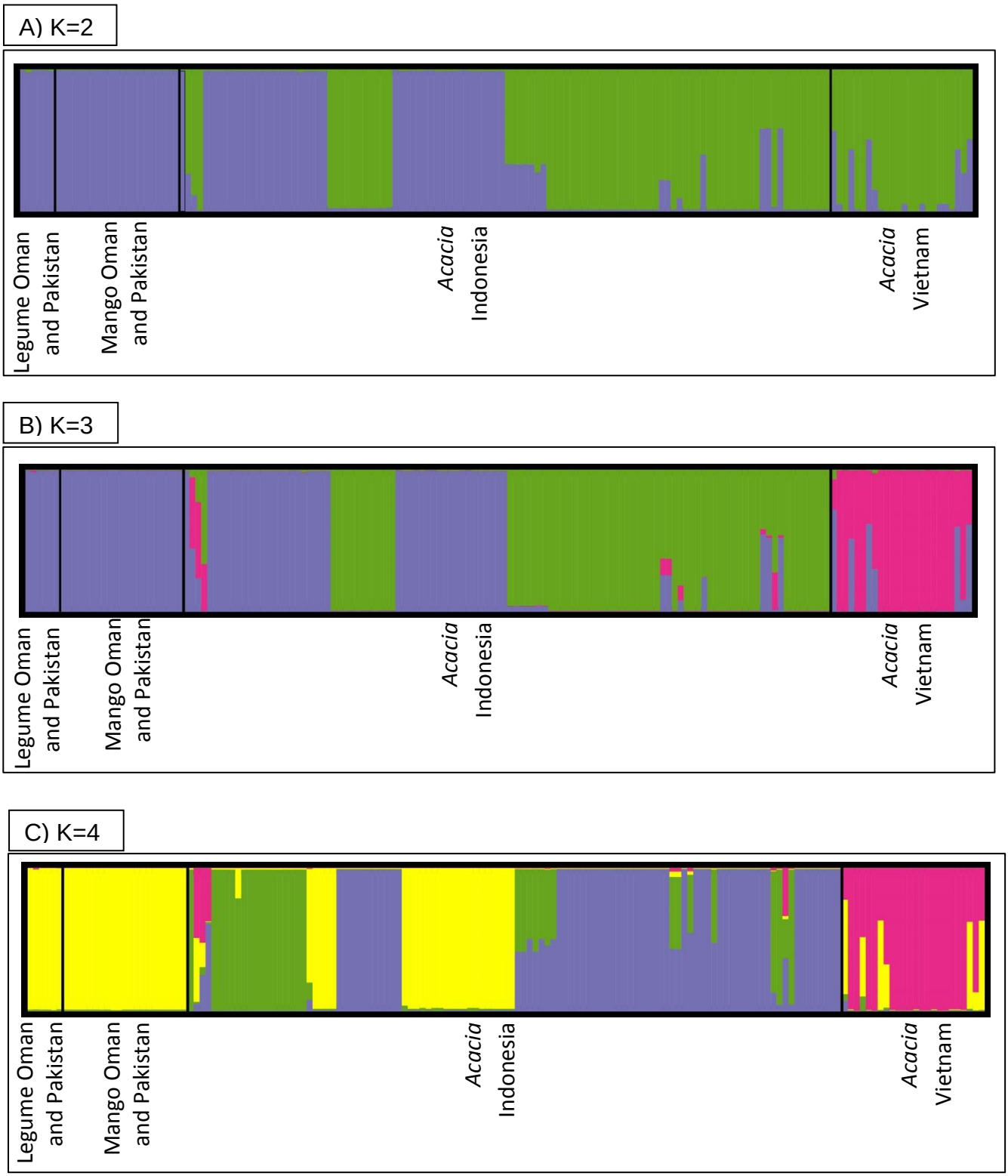
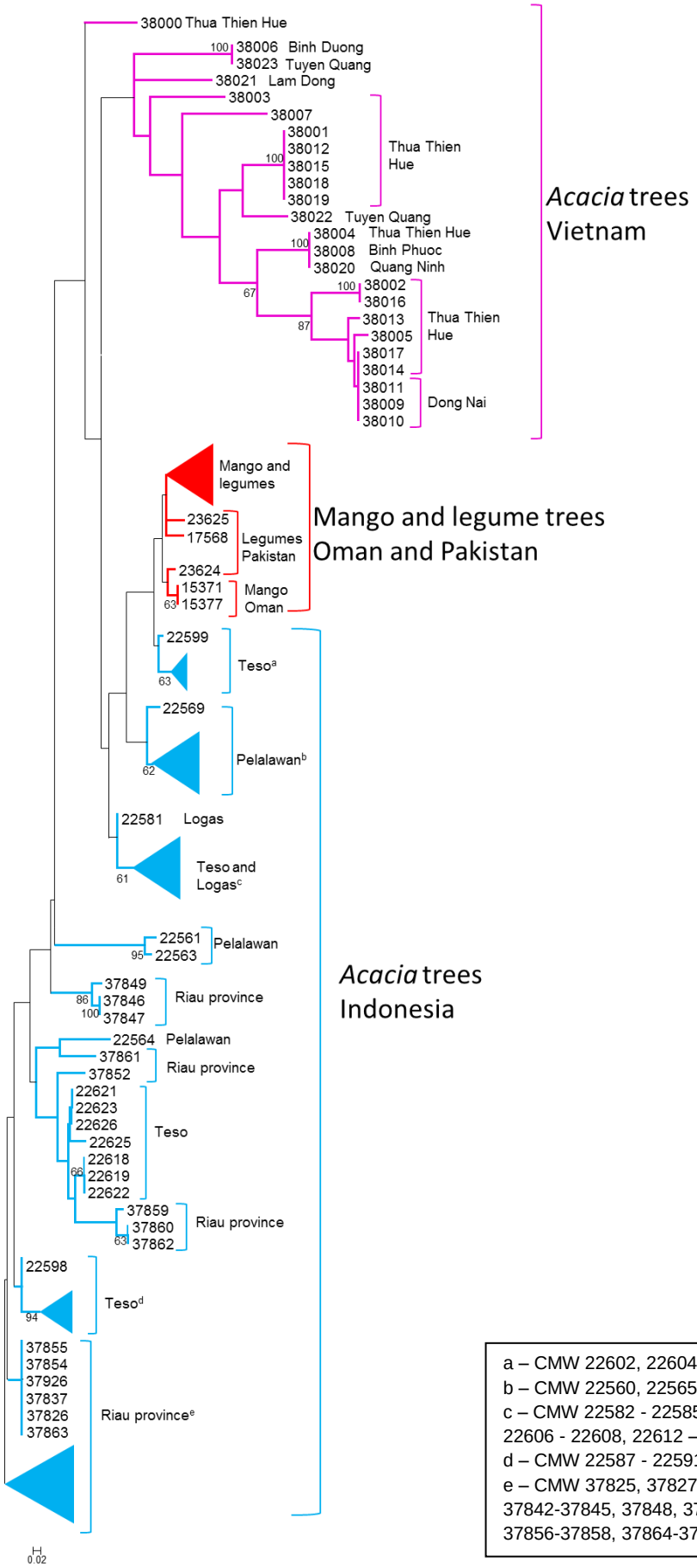


Fig. 6. Analysis of population structure (K) among the *C. manginecans* isolates collected from legume trees and mango in Oman and Pakistan and from *Acacia* trees in Indonesia and Vietnam. Each individual is represented by a single vertical line and the colours indicate the relatedness of an isolate to a specific cluster. A) STRUCTURE analysis suggested the isolates most likely consist of two clusters (K=2) (blue and green). B) An alternative number of clusters (K=3) were also considered, as these are more representative of the population structures obtained from other analyses.



a – CMW 22602, 22604, 22605, 22610 and 22611
b – CMW 22560, 22565 - 22568, 22570 – 22580
c – CMW 22582 - 22585, 22600, 22601, 22603, 22606 - 22608, 22612 – 22617
d – CMW 22587 - 22591, 22593 - 22597
e – CMW 37825, 37827-37836, 37838-37840, 37842-37845, 37848, 37850, 37851, 37853, 37856-37858, 37864-37871, 37925

Fig. 7. A neighbor joining tree that represents the genetic distance between the isolates investigated in this study. Distances were determined based on the proportion of shared

alleles (Chakraborty and Jin, 1993) and the statistical support for each branch was determined by 1000 bootstrap replicates. Bootstrap values higher than 60% are indicated in the figure. Each of the four pre-defined populations, based on host-association and country of origin, is indicated in a different colour. Distinct groupings could be observed among the isolates, with all the isolates from *Acacia* from Indonesia forming one cluster (blue), and those from Vietnam forming another cluster (pink). All the isolates from Oman and Pakistan were very closely related (red).

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Final Discussion and Future Prospects

The first chapter of this dissertation provided an overview of the taxonomy of the *Ceratocystis* genus. The methods currently applied for species identification were critically reviewed to determine their usefulness and limitations in *Ceratocystis*. This study emphasised the limited amount of information obtained from morphological and biological methods for identification of cryptic species. Analyses of the literature showed that more than 70% of the species described in *Ceratocystis* were significantly supported by DNA sequence data. This indicates the substantial impact that phylogenetics have had in cryptic species identification. The limitations of the gene regions, ITS, βT 1 and EF 1-α, currently used for the taxonomy of the *C. fimbriata s.l.* complex were also identified. This review highlighted the need for additional tools that can accurately classify species in this complex and identified a number of alternative approaches that could be investigated further.

The second chapter re-evaluated the efficacy of the phylogenetic markers currently applied for *C. fimbriata s.l.* classification. In addition, the potential of five additional gene regions to identify species in the complex were investigated. From the phylogenetic analyses of all eight gene regions, no single region could be identified as being equally informative as ITS. A combination of three regions, namely βT 1, RPBII and MS204, provided significant support for the majority of species and are proposed for future identification studies. An additional approach considered, which is relatively new to fungal species identification, was the development and use of species-specific SNP markers. The data obtained from these markers were congruent with the data obtained using the phylogenetic gene regions. Both sets of markers clearly distinguished all species in the *C. fimbriata s.l.* complex except *C. acaciivora* and *C. manginecans.* These two species were, therefore, proposed as synonyms of each other in this chapter. The markers proved useful and informative and a potentially new *C. fimbriata s.l.* species from cacao trees was recognised.

In chapter 3, the organism causing the disease on *Acacia* spp. in Vietnam was identified and confirmed to be *C. manginecans*. Cloned sequences of the ITS gene region from isolates obtained from Oman, Pakistan, Indonesia and Vietnam indicated the presence of multiple ITS haplotypes in numerous individuals. This corresponded to the *C. acaciivora* and *C. manginecans* ITS haplotypes. With the synonymy of the two species confirmed, *C. manginecans* has a broader geographic range than originally thought. The genetic diversity and population structure of isolates collected in Indonesia and Vietnam were compared to the clonal ones present in Oman and Pakistan. Ten additional microsatellite markers were developed for this purpose. The analyses indicated that the Vietnam population has the highest genetic diversity and the species have likely existed in this country for an extended period of time.

The studies from this dissertation have highlighted the importance of accurate species identification, specifically with regard to fungal pathogens. Reliable identification is

essential in order to investigate and prevent the spread of pathogens but also for scientific research considering their distribution and population diversity. This is dependent on the application of effective species identification methods, of which DNA sequence data was clearly shown in this study to be the most informative. These studies enabled the identification of a set of reliable molecular markers that can, in combination, be used for the identification of *C. fimbriata s.l.* species. Additionally, a new set of microsatellite markers were developed specifically for these pathogens. The markers developed in this study should prove useful also in taxonomic and population studies of other *Ceratocystis* species not yet investigated.

Future prospects

The ITS region still remains the only gene region that provides support for all the *C. fimbriata s.l.* species. Even after the analyses of five new gene regions, these will still need to be used in combination to give the same resolution as ITS. Although the SNP regions are effective, nine different regions have to be amplified per isolate to get the same data, making this a cumbersome task. The search for more informative phylogenetic markers to define cryptic species in the *C. fimbriata s.l.* complex is still needed. Identification of more gene regions can be initiated by screening the FUNYBASE database (http://genome.jouy.inra.fr/funybase), using the program POLYORPH (Feau et al., 2011; Marthey et al., 2008). The database contains hundreds of potential fungal phylogenetic markers and the program can identify genes with the potential to resolve lower level phylogenies. This potential can be confirmed by calculating the Townsend phylogenetic informativeness (PI) of the regions and plotting their profiles, against time, on a single graph (Townsend, 2007). The PI profile is an indication of the informativeness per site for each gene region to resolve tree topologies. These analyses can be performed *in silico* and thus identify genes with similar evolutionary rates to the ITS gene region, prior to any laboratory work.

One alternative diagnostic marker presented in the literature study, but not extensively investigated in this dissertation, is cerato-platanin (CP). This is a pathogenicity factor found in the mycelia of *Ceratocystis* species (Comparini et al., 2009). The amino acid sequences were significantly different between three distantly related *Ceratocystis* species investigated but were identical among closely related species such as *C. fimbriata s.s.* and *C. cacaofunesta*. During the course of this dissertation, primers were designed for this region and the nucleotide sequences were generated for six species. Multiple nucleotide differences were found between *C. acaciivora, C. manginecans, C. fimbriata s.s., C. platani, C. cacaofunesta* and *C. fimbriata s.l.* from coffee (Fig. 1). *Ceratocystis acaciivora* and *C.*

eucalypticola, however, were identical. The full potential of CP as a diagnostic marker can be investigated in future studies by obtaining the sequence of this gene in at least two isolates of all species of the *C. fimbriata s.l.* complex and evaluating the resolution it provides between cryptic species.

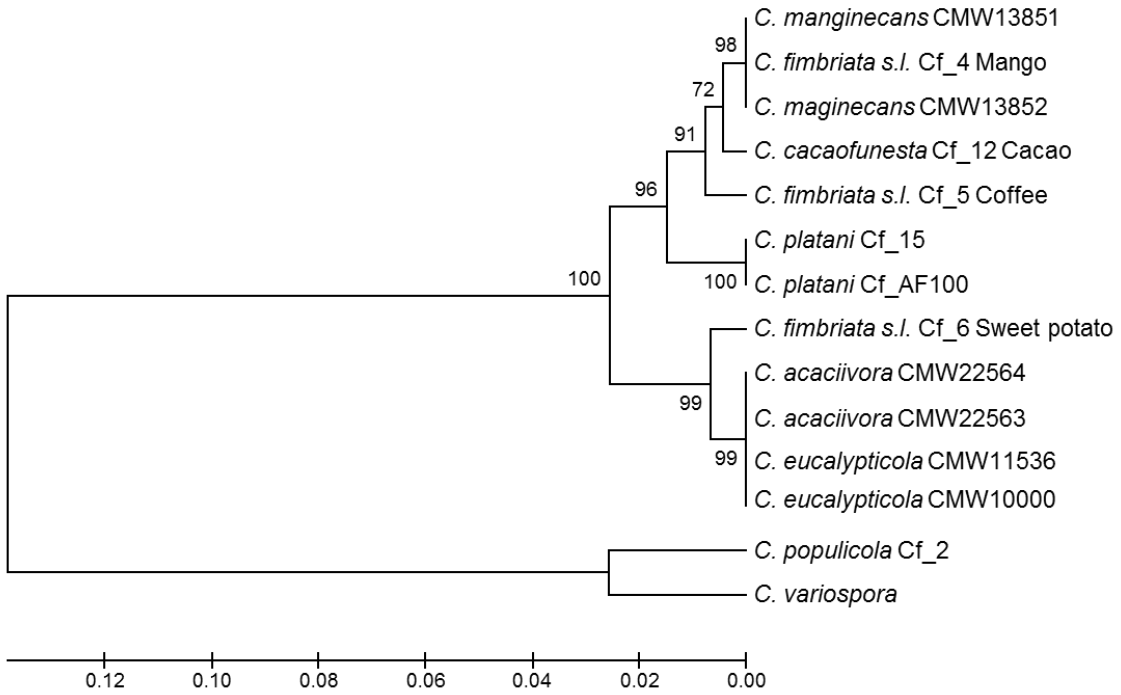


Fig. 1. Cladogram of representative species in the *C. fimbriata s.l.* species complex, based on sequence data of the ceratoplatanin gene. An Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was used for tree construction and branch support values were determined based on 1000 bootstrap replicates (indicated above the branches).

The identical sequence of the CP gene observed in *C. acaciivora* and *C. eucalypticola* but found to differ in *C. manginecans,* contrasts what was found in chapter 2 of this dissertation. Among the five gene regions considered in chapter 2, *C. manginecans* and *C. acaciivora* were identical but distinct from *C. eucalypticola*. This could suggest that a hybridization event occurred among the three species. The occurrence of the *C. eucalypticola* ITS haplotype along with the haplotype of either *C. manginecans* or *C. acaciivora* in a single isolate (chapter 3) also questions whether the ITS region was affected by hybridization. Even though it seems that *C. acaciivora* and *C. manginecans* are the same species, based on the data from the first two chapters, it was decided to not synonymise them officially due to the potential of hybrids. Further investigation is required before this can be ascertained.

With the increase in high throughput sequencing technologies and the decrease in cost for genome sequencing, fungal genome comparisons have become a greater possibility for species classification (Glenn, 2011). A phylogenomics approach is based on genome comparisons of a number of species and identification of the orthologous protein coding

genes they all have in common. The hundreds of genes identified in this manner can then be compared and analysed simultaneously between the species (Fitzpatrick et al., 2006; Rannala and Yang, 2008). This should become more feasible as additional projects, such as the 1000 fungal genome project, are completed as it will exponentially increase the availability of genomic data (http://1000.fungalgenomes.org). In recent studies at FABI, the genome sequences for five species in the *C. fimbriata s.l.* complex and 13 additional species in the *Ceratocystis* genus have been obtained (Unpublished). These genomes provide a basis for identification of numerous single copy protein coding genes that can be combined for a phylogenomic analysis of the genus.

References:

Comparini, C., Carresi, L., Pagni, E., Sbrana, F., Sebastiani, F., Luchi, N., Santini, A., Capretti, P., Tiribilli, B., Pazzagli, L., Cappugi, G., Scala, A., 2009. New proteins orthologous to cerato-platanin in various *Ceratocystis* species and the purification and characterization of cerato-populin from *Ceratocystis populicola*. Appl. Microbiol. Biotechnol. 84, 309-322.

Feau, N., Decourcelle, T., Husson, C., Desprez-Loustau, M.-L., Dutech, C., 2011. Finding Single Copy Genes Out of Sequenced Genomes for Multilocus Phylogenetics in Non-Model Fungi. PLoS ONE 6, e18803. doi: 10.1371/journal.pone.0018803.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., Butler, G., 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol. Biol. 6, 99.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 11, 759–769.

Marthey, S., Aguileta, G., Rodolphe, F., Gendrault, A., Giraud, T., Fournier, E., Lopez-Villavicencio, M., Gautier, A., Lebrun, M.-H., Chiapello, H., 2008. FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics 9:456. doi: 10.1186/1471-2105-9-456.

Rannala, B., Yang, Z., 2008. Phylogenetic Inference Using Whole Genomes. Annu. Rev. Genomics Hum. Genet. 9, 217-231.

Townsend, J.P., 2007. Profiling Phylogenetic Informativeness. Syst. Biol. 56, 222-231.

Summary

The fungal species *Ceratocystis acaciivora* and *C. manginecans* are economically important plant pathogens, reported from the Middle East and South East Asia. These fungi cause severe wilt of *Acacia* and mango trees, respectively. The species form part of the *C. fimbriata sensu lato* complex that consists of numerous closely related plant pathogens with a global distribution. *Ceratocystis acaciivora* and *C. manginecans* are morphologically identical and both have been reported on the same host, *A. mangium*, in Indonesia. Identification of these, and other species in *C. fimbriata s.l.* complex, is strongly reliant on phylogenetic analyses of three gene regions and the congruence of their phylogenies. However, the reliability of this approach has recently come into question. The ITS haplotypes of both *C. acaciivora* and *C. manginecans* have been identified in a single isolate. The two gene regions, β-tubulin 1 (βT 1) and Translation Elongation Factor 1-α (EF 1-α), are generally combined with ITS data but their phylogenies provide no resolution between these two species. The βT 1 and EF 1-α gene regions have also failed to resolve various other well-characterised and cryptic species in the *C. fimbriata s.l*. complex.

The aim of this dissertation was to develop and identify alternative genetic markers that can be utilized for species identification in the *C. fimbriata s.l.* species complex, particularly with regard to distinguishing between *C. acaciivora* and *C. manginecans*. An additional aim was to obtain a better understanding of the distribution and population genetic diversity of *C. manginecans* and *C. acaciivora* on various hosts in different geographical areas. Five phylogenetic gene regions, effectively used in the taxonomic studies of other ascomycetes, were considered for species identification in the *C. fimbriata s.l*. complex. Species-specific SNP markers were also developed and tested as diagnostic markers for species identification. The population studies were performed by screening a set of previously designed, as well as newly developed microsatellite markers in isolates obtained from four countries.

None of the five gene regions tested, allowed for better resolution of the species boundaries than that which can be achieved using the ITS region. In combination, three gene regions (βT 1, MS204 and RPBII) resolved the majority of species in the complex and should be useful for future taxonomic studies. The application of nine SNP markers proved to be highly informative and distinguished all species in the species complex. None of the genetic markers considered in this study allowed for the differentiation between *C. acaciivora* and *C. manginecans,* supporting the view that these are synonyms. The population study expanded the current knowledge of the distribution of *C. manginecans*. The microsatellite data revealed a high genetic diversity among the isolates in South East Asian countries and confirmed the presence of a genetically distinct population in Vietnam. *Ceratocystis*

manginecans has likely existed in this region for a long period of time. This study has provided the data for resolving taxonomic uncertainties of cryptic species in *Ceratocystis* and lays the foundation for future studies within this genus.