Diversity and spatial distribution of fungal endophytes in a *Eucalyptus grandis* tree

by

Kerry-Anne Pillay

Submitted in partial fulfillment of the requirements for the degree

Magister Scientiae

In the faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria

April 2012

Supervisor: Dr. Marieka Gryzenhout
Co-Supervisors: Prof Bernard Slippers
Prof Mike Wingfield

I dedicate this thesis to my father Morgan and my mother Perimala for all the hardship they had to endure to ensure the success of their children.

'A father's love is higher than a mountain and a mother's love is deeper than the sea'

(A Japanese Proverb)

Declaration

I, Kerry-Anne Pillay declare this thesis, 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.

Kerry-Anne Pillay

April 2012

TABLE OF CONTENTS

Acknowledgement

I accept full responsibility for this document, however with this I feel obliged to acknowledge certain people that made the completion of this thesis possible.

Firstly I would like to thank god for giving me the will power and strength especially during difficult times to complete this thesis. In addition I would like to thank the following people:

- My supervisor Dr. Marieka Gryzenhout and co-supervisors Prof Bernard Slippers and Prof Mike Wingfield for their guidance, advice and encouragement throughout my degree.

- My colleagues and lab mates that provided a good working environment.

- The sequencing facility (Gladys Shabangu and Renate Zipfel), the Bioinformatic department (Oliver Bezuid) and Dr Jeff Garnas at the University of Pretoria who provided endless help with the processing and analyses of my data.

- The University of Pretoria (UP), National Research Foundation (NRF), members of the Tree Protection Co-operative Programme (TPCP), Centre of Excellence for Tree Health Biotechnology (CTHB) for their funding support.

- My family and friends for their unconditional love and support during the course of this thesis.

Preface

DNA bar-coding is a technique where DNA sequence data of a standardised DNA sequencing protocol is used to identify eukaryotic organisms. DNA bar-coding has also been applied to environmental, metagenetic samples in order to identify micro-organisms present in mixed microbial populations. Recent advances in high throughput sequencing allow for the direct sequencing of the amplicons of a barcode locus from environmental samples, which significantly reduces labour and costs, and yields large volumes of sequence data. This is both because the need for cloning is eliminated and because next generation sequencing is much cheaper than Sanger sequencing. The next generation sequence data derived from the microbial community ñmini-barcodesò are then compared with quality controlled databases using bioinformatic tools to establish identities.

Fungal endophytes infect plant tissues without causing any apparent disease symptoms. Grass fungal endophytes are transmitted vertically from parent to offspring and have been relatively well studied. In contrast, the fungal endophytes present in woody trees such as *Eucalyptus* are transmitted horizontally from tree to tree, and are very poorly known. Endophyte studies on trees such as *Eucalyptus* have focussed on those occurring as latent pathogens, such as the Botryosphaeriaceae. It will only be through an understanding of the entire endophytic community within trees such as *Eucalyptus* that we will truly understand their role in the ecology of these trees. Much remains to be learned about the total diversity of these endophytes, cryptic endophytic latent phases of pathogens other than the Botryosphaeriaceae, changes in the communities over time and dominance of certain groups and more.

In this study we characterised the diversity of the fungal endophytes present in a *Eucalyptus grandis* tree using a metagenetic, 454-pyrosequencing approach of the total DNA barcode amplicons from the fungi colonizing the plant. An isolate-based DNA bar-coding of fungal isolates approach was also used. The aim was to identify the maximum number of the fungal endophytes, including unculturable and culturable species. These sequence data would then contribute to a local endophytic database to aid future endophytic identification. Furthermore, a 454-pyrosequencing bioinformatic identification workflow and isolate bar-coding identification workflow were established that can be used to assign identities to high throughput sequence data, and isolate barcodes. This is important because there is a limited baseline of knowledge regarding the taxonomy and diversity of tree endophytes, especially in South Africa.

In Chapter 1, the literature regarding DNA bar-coding was reviewed regarding bar-coding standards, the organisations involved in bar-coding and bar-coding projects that have thus far been completed. The review also considered high throughput sequencing and the potential to use these techniques to produce mini-barcodes. Furthermore, the applicability of combining both 454-pyrosequencing and traditional DNA bar-coding approaches for identification purposes was considered. Thus the current state of knowledge regarding fungal endophytes of woody trees and their ecology was explored. We also considered the potential to study this system using the complimentary identification ability of the isolate DNA barcodes and the metagenetic, 454-pyrosequencing data.

In Chapter 2 the fungal endophytes present in three *E. grandis* trees was identified using the two different identification approaches of metagenetic (environmental DNA bar-coding) using 454-pyrosequencing, and isolate DNA bar-coding. The purpose of the high throughput sequencing (454-pyrosequencing) technique was to characterize the entire endophytic community without the need for culturing, thus also identifying unculturable endophytes and slow growers. The culturing process used for the conventional DNA bar-coding technique provided a valuable comparison and voucher cultures that can be used to study these cultures further using multi-gene sequencing and other techniques. Most importantly, the study aimed to establish a reliable database and identification workflow for the future study of the community of *Eucalyptus* endophytes.

Chapter 3 of the thesis focussed on an important pathogen group of *Eucalyptus* that is known to be one of the dominant groups of endophytes, namely the Botryosphaeriaceae. Multi-gene sequencing was used to identify the Botryosphaeriaceae from the cultures obtained from the *Eucalyptus grandis* trees. Isolates from a native relative of *Eucalyptus*, *Syzygium cordatum*, from the same location in South Africa, were also available from a parallel study, and these isolates were included to identify overlapping occurrences of the Botryosphaeriaceae between these two trees.

.

List of tables

Chapter 1

Chapter 2

Chapter 3

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 1

Bar-coding and next generation sequencing as tools to characterize fungal endophyte communities in *Eucalyptus*

Introduction

DNA bar-coding and 454-pyrosequencing are modern techniques that can be used to identify pathogens that cause disease on *Eucalyptus. Eucalyptus* plantations are an important part of the South African forestry industry, in particular for pulp and paper production (Schonau 1991; Turnbull 1999; 2000).The *Eucalyptus* pathogens present in South African plantations have been the focus of intensive studies over the past two decades (Wingfield *et al.* 2008). These pathogens include bacteria such as *Pantoea* spp. (Coutinho *et al.* 2002; Coutinho and Venter 2009) and *Ralstonia solanacearum,* (Coutinho *et al.* 2000) fungal pathogens such as *Kirramyces zuluense* (Wingfield *et al.* 1996), Botryosphaeriaceae (Smith *et al.* 1994; Smith *et al.* 1996a; Smith *et al.* 2001; Slippers *et al.* 2004a; Slippers *et al.* 2004b; Slippers *et al.* 2004c), *Erythriceum salmonicolor* (Nicol *et al.* 1993; Roux *et al.* 2001), *Quambalaria eucalypti* (Wingfield *et al.* 1993; De Beer *et al.* 2006; Roux *et al.* 2006), *Cylindrocladium pauciramosum* (Crous *et al.* 1991; 1993), *Chrysoporthe austroafricana* (Wingfield *et al.* 1989), *Kirramyces epiccocoides* (Crous *et al.* 1988), *Mycosphaerella* spp. (Crous and Wingfield 1996; Crous 1998; Hunter *et al.* 2004a; Hunter *et al.* 2004b) and *Teratosphaeria* spp., (Hunter *et al.* 2009) and the oomycetes *Pythium* spp (Linde *et al.* 1994) and *Phytophthora* spp (Wingfield and Knox-Davies 1980; Linde *et al.* 1994; Maseko *et al.* 2007). These pathogens cause diseases that reduce wood quality and production, and in some cases tree death. Their impact has been so severe that they contributed in shaping the breeding programs of forestry companies in South Africa such as planting resistant or tolerant stock to reduce losses due to pathogens (Wingfield 2003a; Wingfield *et al.* 2008). Despite years of research, the pathogen threat to *Eucalyptus* continues due to the complexity of controlling them and new pathogens emerging via introductions or host jumps (Coutinho *et al.* 1998; Keane *et al.* 2000; Wingfield *et al.* 2001; Wingfield 2003b; Gryzenhout *et al.* 2004; Slippers *et al.* 2005a; Nakabonge *et al.* 2006; Pavlic *et al.* 2007).

Apart from the fungal pathogens, little is known about the other parts of fungal communities present in and on *Eucalyptus* trees. One component of these communities includes fungal endophytes that are able to inhabit plant tissue without causing any visible disease symptoms (Clay 1993; Wilson 1995; Saikkonen *et al.* 1998; Faeth and Fagan 2002; Saikkonen *et al.* 2004; Schulz and Boyle 2006; Arnold *et al.* 2007; Rodriguez *et al.* 2009). These fungi are thought to play an important role in tree health by either positive or negative interactions with the host (Blodgett *et al.* 2000; Swart *et al.* 2000; Cannon and Simmons

2002; Davis *et al.* 2003; Campanile *et al.* 2007; Cheplick 2007). Further, fungal endophytes contribute largely to fungal biodiversity associated with trees (Arnold *et al.* 2000; Froehlich *et al.* 2000; Arnold *et al.* 2007; Arnold and Lutzoni 2007).

A group of endophytic fungi that has been well studied from *Eucalyptus* is the Botryosphaeriaceae (Smith *et al.* 1994; Smith *et al.* 1996a; Smith *et al.* 1996b; Smith *et al.* 2001; Ahumada 2003; Gezahgne *et al.* 2004; Slippers *et al.* 2004b; Slippers *et al.* 2005b; Mohali *et al.* 2007; Pérez *et al.* 2008; Taylor *et al.* 2008; Rodas *et al.* 2009; Slippers *et al.* 2009; Chen *et al.* 2011). This fungal family includes a number of latent pathogens of great importance to eucalypts and many other tree species (Slippers and Wingfield 2007). These fungi have an endophytic phase in their life cycle and their pathogenic tendencies are triggered by stress factors, such as drought and hail damage (Schoeneweiss 1981; Swart *et al.* 2000; Saikkonen *et al.* 2004; Desprez-Loustau *et al.* 2006; Gonthier *et al.* 2006; Slippers and Wingfield 2007). In depth studies have for example considered the co-occurrence and possible host jumping of Botryosphaeriaceae between native *Syzygium cordatum* trees and exotic *Eucalyptus* occurring in the same area, because both tree species belong to the Myrtaceae (Pavlic *et al.* 2007; Pérez *et al.* 2008).

There has never been a systematic study of all the endophytic fungi in *Eucalyptus,* or any other tree species, in South Africa. Elsewhere in the world, only a limited number of studies have focused on endophytes present in *Eucalyptus* spp., and then mostly as a part of this community (Bettucci and Saravay 1993; Fisher *et al.* 1993; Smith *et al.* 1996b; Bettucci *et al.* 1999; Slippers *et al.* 2004c).This lack of baseline knowledge makes it very difficult to compare the diversity of endophytic communities between hosts, geographic regions and follow changes that might happen over time. Such broad studies of endophytic communities are, however, hampered by the large scale and cost of such studies. These accrue from the large amount of culturing required, and the known limits of culturing with regards to unculturable or difficult to culture species (Carroll 1995; Arnold *et al.* 2007). A technique that can be used to address this problem is the use of DNA bar-coding which is a culturable method using the conventional sequencing methods, especially when combined with high throughput tools such as 454-pyrosequencing (Tedersoo *et al.* 2010). This review considers these different approaches to fungal community identification and its applicability to studying endophytes of *Eucalyptus* in South Africa.

2. DNA Bar-coding

2.1 Bar-coding standards

DNA bar-coding is widely accepted as a system for universal identification of species (Ratnasingham and Hebert 2007; Frézal and Leblois 2008). This method uses standardized DNA sequence to identify all species of a particular eukaryotic kingdom (Hebert *et al.* 2003; DeSalle *et al.* 2005; Hebert and Gregory 2005a; Savolainen *et al.* 2005; Schindel and Miller 2005; Hajibabaei *et al.* 2007a; Hajibabaei *et al.* 2007b; Casiraghi *et al.* 2010). The two basic aims of the bar-coding system are to (i) identify and assign unknown specimens to those that have been previously described (ii) facilitate the discovery of unknown species and incorporate these into known taxonomies (Summerbell *et al.* 2005; Miller 2007).

The first level of a DNA bar-coding project intends to generate a library of DNA sequences of known species. This primarily involves Sanger chain termination sequencing from vouchered and stable specimens or microbial cultures (Hajibabaei *et al.* 2005; Karow 2008). A first phase bar-coding study aims to establish a preliminary database that contains sequence data that is linked to specimens. As more specimens are bar-coded, the database will be updated regularly to aid future identification for that specific group of organisms that will be available for the entire scientific community.

The Barcode of Life Data Systems (BOLD) (www.barcodinglife.org) was initiated in 2004 to collect validated DNA sequences of organisms generated by individuals or consortiums around the world (Ratnasingham and Hebert 2007; Frézal and Leblois 2008). Only sequences generated from five specimens of the same species can be deposited into BOLD. The BOLD system is linked to other DNA sequence deposition centres such as those of the National Centre of Bioinformatics Institute (NCBI), the European Bioinformatics Institute (EBI) or the DNA Data Bank of Japan (DDBJ). When depositing sequences, additional information regarding primer sets used, electrophenogram trace files, translations, taxonomic information linked to sequences and linkage to voucher specimen deposition must be provided. BOLD further requires information pertaining to the specimen identifier, the taxonomic information of the specimen, the specimen characteristics, as well as details regarding the specimen collection and the DNA barcode sequence. This additional information on ecological data, validity of the origin of the sequences and availability of the

original sequence data is what distinguishes bar-coding from normal sequence deposition (Ratnasingham and Hebert 2007).

For species identification using DNA barcodes, a DNA barcode gene needs to be identified for the unknown specimen. This gene region should be variable enough to distinguish closely related species, easily amplified and short enough to allow amplification from degraded environmental samples (Valentini *et al.* 2008). A typical bar-coding project would entail collecting species that has been identified by morphology. The barcode sequence of an unknown organism and the barcode sequences deposited into the bar-coding database are then compared to each other (Hajibabaei *et al.* 2007b). An organism will be identified if its barcode sequence matches another barcode within the bar-coding database. If no close matches are found, a new barcode may be given to that organism, representing a new haplotype, a new geographical variant or possibly a new species or cryptic species. If a new barcode sequence is identified, standard bar-coding submission requirements will have to be followed in order to deposit barcodes into BOLD.

The 600 bp fragment of the mitochondrial cytochrome c oxidase (COX1) gene has proven to be effective as the universal barcode for animal species and has been approved as the official bar-coding regions for animals (Hebert and Gregory 2005a). This region has been used widely in the identification in for example different bird species (Hebert *et al.* 2004b), fish (Ward *et al.* 2005; Zhang and Hanner 2011), odonta (Rach *et al.* 2008), fly species (Whitworth *et al.* 2007), arachnids (Barrett and Hebert 2005), butterflies (Hebert *et al.* 2004a; Hausmann *et al.* 2010) *Collembola* spp. (springtails) (Hogg and Hebert 2004), primates (Lorenz *et al.* 2005), dove lice (Whiteman *et al.* 2004), crustaceans (Havermans *et al.* 2010) and many more.

In plants, there is low mitochondrial DNA variation, and therefore the COX1 gene cannot be used as the primary bar-coding gene. Thus, supplementary genes, such as the chloroplast genes for example the coding regions *matK, rbcL, rpoB,* and *rpoC,* as well as the non-coding spacers *atpF-atpH, trnH-psbA* and *psbK-psbI* (Chase *et al.* 2005; Chase and Fay 2009; Hollingsworth *et al.* 2009). The ITS2 region can also be useful as a bar-coding gene for plant species (Luo *et al.* 2010). Recently it has been suggested that the *matK* and *rbcL* be used as core plant bar-coding genes due their high species resolution (Hollingsworth *et al.* 2011).

The COX1 gene fragment cannot be used as the universal barcode in fungi. One reason is that the COX1 gene region contains varying numbers of introns, as illustrated in a study on *Penicillium* (Seifert *et al.* 2007). Consequently, the amplification of the gene region using standard primers produces PCR products of varying lengths (Mouhamadou *et al.* 2008). In another study using COX1 on fungi, this region was used together with ITS as the bar-coding region for the hyphomycete *Leohumicola* and in an attempt to distinguish three *Leohumicola* species from United States and South Africa (Nguyen and Seifert 2008). In this study, both loci showed similar levels of sequence variation and species resolution, and suggested that both might be suitable as bar-coding genes for this group of fungi. However, universal primers to amplify the COX1 in all fungi are not available and this would make it an unsuitable general bar-coding gene.

For fungi the internal transcribed spacers (ITS) of the nuclear rRNA repeat units are the most widely accepted, and unofficially used, locus for bar-coding purposes (Summerbell *et al.* 2005; Karkouri *et al.* 2007). The ITS is most frequently sequenced and thus a large database on NCBI-GenBank has been generated for a variety of fungi, including pathogenic, endophytic and saprophytic fungi (Nilsson *et al.* 2006; Nilsson *et al.* 2008). Using the ITS region alone for fungal bar-coding is, however, also problematic because certain phyto-pathogenic groups, such as *Fusarium* spp., show poor resolution at species level based on ITS sequence data (O'Donnell *et al.* 1998). Species of Botryosphaeriaceae (Pavlic *et al.* 2009a), *Chrysoporthe* (Myburg *et al.* 2002; Gryzenhout *et al.* 2004), *Ceratocyctis* (Witthuhn *et al.* 1998) and *Ophiostoma* (Kim and Breuil 2001; Grobbelaar *et al.* 2009) represent further examples where the ITS fails to resolve closely related species. However, despite the potential shortcomings of the ITS region for bar-coding fungi, several fungal bar-coding projects have been initiated using this region. This is because the advantages of ITS, namely being able to amplify it easily from all fungi and the large existing database linked to it, outweighs its shortcomings. There is also no other locus currently available that can compete with the breadth and depth of information linked to the ITS. The ITS has been recommended as the fungal bar-coding gene, after comparing its performance as a fungal identified gene against four other genes including the large sub unit (LSU) the RNA ribosomal II subunit (RBP2) (Schoch *et al.* 2011) Examples of such bar-coding studies are those conducted on, *Aspergillus* spp. (Geiser *et al.* 2007), *Fusarium* spp. (Gilmore *et al.* 2009). *Melampsora* spp. (Feau *et al.* 2009), *Penicillium* spp. (Seifert 2009), aquatic

hyphomycetes (Letourneau *et al.* 2010) and arbuscular fungi (AMF, Glomeromycota) (Stockinger *et al.* 2010).

In cases where the ITS region cannot resolve species identities, other loci, such as housekeeping genes or mitochondrial genes, are used to add additional resolution. These loci are sometimes more informative and more variable at a species level, and are selected because they are non-orthologous, single copy genes that can be amplified with ease using universal primers. Examples are the translation elongation factor (TEF 1-alpha) and other protein-coding genes such as beta-tubulin, calmodulin and histone H3 (Geiser *et al.* 2004; Gazis *et al.* 2011). Such polygenic approaches to identify species are valuable, but it does add unwanted complexity, and is therefore not universally applicable.

2.2 Bar-coding Consortia

Bar-coding projects tend to be large and require collaboration between different research groups, and across borders. For this reason, large consortia have emerged that coordinate, and in some cases fund, bar-coding work around the globe. This approach has also promoted the establishment of standards for different groups of organisms. Here we consider structures of relevance to this thesis, namely the overall co-ordination, bar-coding in South Africa and bar-coding of fungi.

The core partners of DNA bar-coding initiatives include the International Barcode of Life project (iBOL) (www.barcodeoflife.org) and the Consortium for the Barcode of Life (CBOL) (www.cbol.org). iBOL is a non-profit co-operation that spans 25 nations and aims to develop a DNA sequence library to identify organisms. The research is driven by 20 working groups that can be divided into five working areas, namely DNA barcode library creation, the bar-coding method, informatics, technologies, and administration. The 25 nations are divided into groups, each responsible for different roles that guide the bar-coding process. There are four central nodes that include Canada, China, the European Union (France, Germany, Italy and Portugal) and the United States, whose main responsibility deals with funding and co-ordination. The 10 regional nodes include Argentina, Australia, Brazil, India, Mexico, New Zealand, Norway, Russia, Saudi Arabia and South Africa, whose primary responsibility are to manage the bar-coding efforts within a region. Eight national nodes have thus far been formalized, including Colombia, Costa Rica, Kenya, Korea Madagascar, Papua New Guinea,

Peru and Panama, which will survey biodiversity through barcodes and falls under a regional node. Researchers of these various nodes will assemble specimens and interpret results. CBOL was established in 2004 and intends to promote standardized identification of specimens based on barcodes through the working groups. There are currently four working groups fall within the CBOL initiative, including the Data Analysis Working Group (DAWG), Database Working Group (DBWG), Leading Labs Network (LLN) and Plant Working Group (PlantWG) (Fig1).

South Africa is home to one of 10 bar-coding regional nodes with a function to oversee bar-coding work in the region. The only two other African countries officially listed as nodes are Kenya and Madagascar, which are national nodes linked with International Barcode of Life project (iBOL*).* Africaôs primary bar-coding projects entail bar-coding tree and fish species. An African project that has received significant focus and has had much impact is the bar-coding research pertaining to the flora in the Kruger Park in South Africa (Hollingsworth ***et al.*** 2011). The *psbK-psbI*and the *atpF-tpH* chloroplast genes were tested as potential barcodes using the flora of the Kruger National Park as a model system, which is part of the Maputoland-Pondolandï Albany biodiversity hotspot (Lahaye ***et al.*** 2008). These studies evaluated the efficiency of combining the *matK* gene either with *trnH-psbA* and/or *atpF-atpH* and/or *psbK-psbI* genes for bar-coding to improve species differentiation. These studies found that combining *matK* to *trnH-psbA* and *psb-psbI* improved species identification, but also suggested that *matK* can stand as a single bar-coding gene for plant species (Lahaye ***et al.*** 2008). Fish-BOL Africa, part of the Fish Barcode of Life Initiative (FISH-BOL) aims to barcode all Africaôs marine, estuarine and freshwater fish to aid fish taxonomy and gain knowledge regarding lineage diversity among fish (www.SAIAB.ac.za).

There are several joint fungal bar-coding projects or consortia. These include IM-BOL (IndoorMycota Barcode of Life), that aims to create an ITS database of indoor moulds. Other emerging fungal campaigns include Cultureï BOL that endeavors to barcode established culture collections and fungal herbaria where cultures already exist in the Centraalbureau voor Schimmelcultures (CBS) and Canadian Collection of Fungal Cultures, Mushroom-BOL that aims to barcode all known species of mushrooms, ISHAMï BOL (Medical Mycology bar-coding working group that endeavors to barcode fungi of medical importance), Mycotoxï BOL that targets mycotoxin producing moulds and involves a collaboration with Italy and

Canada, and QBOL that attempts to barcode quarantine organisms in support of plant health that focused mostly on Europe.

3. Next generation sequencing

3.1 Next generation sequencing technologies

Next generation sequencing differs from Sanger chain-termination sequencing in its ability to process thousands of sequence reads in parallel (Mardis 2008; Reis-Filho 2009; Voelkerding *et al.* 2009). This sequencing revolution has primarily been driven by three platforms including Roche 454-pyrosequencing Genome Sequencer (GS), commonly known as pyrosequencing (Roche Applied Science), Genome Analyzer (Illumina Inc./Solexa) and Life Technologies SOLiD System (Applied Biosystems), but a number of other technologies are currently entering the market. Next generation sequencers have different sequencing chemistries and amplification approaches that directly influence the read lengths, time per run and mega bases of sequences generated per run that will influence its application (Table 1). Newer sequencing technologies include sequencing by synthesis HeliScope (Helicos) (http://www.heicosbio.com) (Braslavsky *et al.* 2003; Richardson 2010; Shumway *et al.* 2010) and PacBio (Pacific Biosciences) (http://www.pacificbiosciences.com) (Glenn 2011). The latter has only been introduced in late 2010 and, therefore, still gaining popularity. The newer technologies are less frequently used due to low sequence accuracy and the increased technical support required for the optimal functioning of the CCD camera (Korlach *et al.* 2008). But there are others that are gaining popularity such as the Ion Torrent (http://www.iontorrent.com) that decodes DNA using voltage detection (Glenn 2011).

The production of next generation sequence libraries in the various next generation sequence techniques is not complicated. DNA or RNA is fragmented into pieces. The libraries that are constructed from the fragmentation are then sequenced with high coverage and aligned to reference sequences using bioinformatic tools to make statistical inferences (de Magalhães *et al.* 2011; Horner *et al.* 2010; Mardis 2008; Pop and Salzberg 2008; Shendure and Ji 2008). Various bioinformatic pipelines can be developed for each next generation sequencing technology depending on the application of the technology.

There are several advantages of next generation sequencing when compared to traditional Sanger chain-termination sequencing. When it is needed to generate sequence data for particular DNA regions from environmental samples, cloning is usually necessary to isolate individual sequences in the Sanger chain-termination approach (Guazzaroni *et al.* 2009; Singh *et al.* 2009). This could introduce cloning bias where certain DNA sequences are more readily produced in *E. coli* cells, thus excluding certain sequences (Forns *et al.* 1997; Morgan *et al.* 2010). Circumvention of cloning steps makes sequencing less costly and labour intensive (Nowrousian 2010). Next generation sequencers employ a single PCR step whereby sequence fragments are generated and then sequenced simultaneously. Multiple samples can also be combined by using sequence tags for each sample (Parameswaran *et al.* 2007). Gel electrophoresis, size separation and labelled primers are also not required for next generation sequencing as opposed to Sanger chain-termination sequencing (Ronaghi and Elahi 2002). Sequencing reactions are performed in real time and the raw data are directly analysed through computational rapid DNA sequencing methods (Hill *et al.* 2002; Hudson 2008; Pop and Salzberg 2008). The generation of these large numbers of sequence fragments with next generation platforms are fairly accurate, with the Illumina having an accuracy of 98.5%, ABIï SOLiD an accuracy of 99.94% (Hert *et al.* 2008) and 454-pyrosequencing has an accuracy of 99.4% to 99.51% (Huse *et al.* 2007; Creer *et al.* 2010). All of the above thus reduces experimental time and costs.

There are also disadvantages to next generation sequencing that need to be addressed in bar-coding studies. The main restriction of all three platforms is the small individual fragment sequence read lengths of 35-400bp as compared to Sanger chain-termination sequencing that readily produces read lengths of 600bp or more. In addition, next generation systems are prone to sequencing errors. These errors are especially frequent in the 3ô region (Ansorge 2009). The 454-pyrosequencing platform has further disadvantages such as reduced accuracy in homopolymer regions, which are regions of a repeated nucleotides (Rothberg and Leamon 2008) and the sequencing of chimeras that are artificially generated during PCR (Huber *et al.* 2004). New chemistry and software tools, such as Chimera checker (Nilsson *et al.* 2010), are, however, continually being developed to address these concerns. The large amount of sequence data generated by next generation sequencing, in conjunction with the short reads, however, continues to create a challenge for bioinformatic software to interpret the sequence data into meaningful information (Pop and Salzberg 2008; Valdivia-Granda 2008; Petrosino *et al.* 2009; Rosen *et al.* 2009; Voelkerding *et al.* 2009).

3.2 Applications of 454-Pyrosequencing to environmental metagenetic studies

Metagenetic studies are a molecular locus based approach to identify specific microbial communities present in a common environment (Warnecke and Hess 2009; Creer *et al.* 2010). This is in contrast to opposed to metagenomics, which is a functional analyses of environmental DNA from a particular environment that contains unculturable micro-organisms (Handelsman 2004; Hugenholtz and Tyson 2008). Metagenetics involves a process of DNA extraction from the particular environmental sample and subsequent amplification of a species-specific identification gene and sequencing of the identification gene of the target species.

Next generation sequencing can be useful for metagenetic studies as it allows for simultaneous gene sequencing of different organisms co-inhabiting the same niche, but that can be differentiated by sequence data. As discussed above, the power of next generation sequencing is limited by the short read lengths. Loci used for species identification in most organisms tend to be longer than what is covered by individual next generation sequence read lengths, to provide enough nucleotide variation to differentiate species.

454-pyrosequencing is the next generation sequencing technique that allows for the recovery of the longest individual fragments (Droege and Hill 2008). Recent advances in this technique allow for even longer reads. For example, the GS FLX titanium platform promises sequence reads of average 500bp (Droege and Hill 2008; Zhou *et al.* 2010). These long reads are useful for community analysis in metagenetic studies because entire gene regions used for identification and phylogenetic studies, such as the sequence of the entire ITS region, can be covered by single reads. Pyrosequencing also allows for the simultaneous sequencing of DNA from different sources when different sequence tags are linked to specific groups (Parameswaran *et al.* 2007; Huse *et al.* 2008). This allows samples from different sources to be pooled in a single run, which significantly reduces sequencing costs.

An increasing number of environmental studies have employed pyrosequencing in metagenetic studies. Examples of such studies that focus specifically on the microbial component of environmental samples include 454 pyrosequencing of the meiofaunal

biosphere (Creer *et al.* 2010), the bacterial diversity in a glacier foreland of high Arctic and the bacterial community structure in the western Arctic ocean (Kirchman *et al.* 2010; Schuette *et al.* 2010), the bacterial communities in the human gut (Wu *et al.* 2010), bacterial biofilm communities in water meters of drinking water distribution systems (Hong *et al.* 2010), the bacterial community composition and structure of German grassland soils (Nacke *et al.* 2010; Will *et al.* 2010), sequencing the fungal communities present in *Quercus* spp. (Jumpponen and Jones 2009), sequencing of tropical mycorrhizal fungi (Tedersoo *et al.* 2010), sequencing of the indoor fungal composition in geographically patterned and temperate zones (Amend *et al.* 2010), characterisation of the oral fungal biome in healthy individuals (Ghannoum *et al.* 2010), and estimating the fungal diversity present in forest soils (Buee *et al.* 2009). All of these studies indicate a hyperdiversity of microbes that could not be detected before due to the absence of culturing.

Pyrosequencing has also been used to identify other microscopic organisms, other than bacteria and fungi Porazinska *et al.* (2009a), for example tested the suitability of 454 pyrosequencing for nematode identification in metagenomic studies using a set of controlled experiments containing a pool of 41 nematode taxa. The pyrosequencing run recovered more than 90% of the nematodes indicating the suitability of pyrosequencing to identify nematodes from environmental samples (Porazinska *et al.* 2009a; Porazinska *et al.* 2009b).

4. Combining bar-coding and pyrosequencing to generate barcodes

Sanger sequencing remains the most practical method to sequence fragments greater than 650 bp. However, bar-coding studies that use next generation sequencers, including 454-pyrosequencing, to produce barcodes produce shorter sequence reads that contravene bar-coding standards (Karow 2008). While the rapid improvement in pyrosequencing technology, is expected to soon overcome this shortcoming, the individual sequences from community samples (where pyrosequencing can be applied, still not be linked to vouchered specimens (Karow 2008; Hibbett *et al.* 2011). There is, however, potential strength in combining strengths of Sanger and pyrosequencing. Targeted bar-coding of microscopic organisms based on a Sanger sequences to create a voucher-linked barcode database could support pyrosequencing data that will enable large scale characterization and comparison of the diversity in specific niches or ecosystems. Thus pyrosequence-based sampling could also be applied in more targeted surveys of areas or niches with barcode data as a basis.

Likewise a culture-based barcode approach could follow where pyrosequencing has indicated that a specific sample, area, host or niche contains numerous unknown organisms (Fig. 2).

5. Fungal Endophytes

Fungal endophytes constitute a large species group with the ability to inhabit plant tissue without causing disease (Clay 1993; Wilson 1993; Saikkonen *et al.* 1998; Faeth and Fagan 2002; Saikkonen *et al.* 2004; Rodriguez *et al.* 2009). Endophytic fungi are found in all plant species investigated to date such as ferns, conifers and angiosperms (Arnold and Lutzoni 2007; Saikkonen 2007; Hoffman and Arnold 2008), including also algae and, mosses, ferns, conifers and angiosperms, and exhibit a high diversity within these hosts (Arnold *et al.* 2000).

Fungal endophytes display a range of ecological interactions with their hosts. These interactions include mutualism, antagonism and commensalism (Sieber 2007). For example, in grasses endophytes belonging to the Ascomycete family Clavicipitaceae grow actively within the above ground tissue of grasses (Cannon and Simmons 2002). They display a mutualistic interaction with their hosts as they can restrain insect and fungal pathogen invasions in the plant, increase heavy metal tolerance, and increase drought resistance (Schoeneweiss 1981; Saikkonen *et al.* 1998; Blodgett *et al.* 2000; Clay and Schardl 2002; Arnold *et al.* 2003). The antagonistic ability of fungal endophytes against pathogenic fungi has also been tested in trees, for instance to control *Diplodia corticola,* the casual agent of cankers and vascular necrosis in oak tree species (Campanile *et al.* 2007). Other endophytic species influence their hosts negatively. For example, studies conducted on a globally important forage and turfgrass, *Loliumperrenne* (perennial ryegass), revealed a significantly reduced root:shoot ratio and a reduction of photosynthetic shoots after the grass was infected with endophytic *Neotyphodiumlolii* (Cheplick 2007). Endophytes may also increase plant survival by producing mycotoxins that can repel grazers and insects as in the case of several grass species (Rowan 1993; Cheeke 1995; Azevedo *et al.* 2000). Many other endophytes are also latent pathogens that result in disease when the plant becomes stressed due to adverse environmental conditions (Wilson 1995). Numerous endophytic fungi are, however, thought to be commensalistic, because they appear to merely inhabit plant tissue without affecting the host plant in any known way (Davis *et al.* 2003). The latter

view could simply be influenced by the lack of knowledge regarding the roles that they play in nature.

Endophytes are transmitted to trees from neighbouring trees via airborne spores that germinate and penetrate into the plant material and grow between the plant cells (Lodge *et al.* 1996; Arnold *et al.* 2003; Herre *et al.* 2007). In this way, various endophytes infect the host plant tissue with increasing abundance and diversity as the plant matures (Stone 1986). This non-systematic, horizontal mode of transmission is thought to occur in most plants, including woody gymnosperms and angiosperms (Herre *et al.* 2007). Endophytes of grasses also follow a vertical mode of transmission (Clay and Schardl 2002) where they are transmitted directly from parents to offspring in a systemic manner (Carroll 1988; Cannon and Simmons 2002; Saikkonen *et al.* 2004), thus resulting in an analogous endophyte assemblage for parent and offspring.

Species diversity of endophytic mycobiota is generally high in trees (Arnold *et al.* 2001; Arnold 2007; Arnold *et al.* 2007; Arnold and Lutzoni 2007). The number of fungal endophytes present within host tissues is largely determined by biotic and abiotic factors such as host species, plant organ and climatic conditions (Sieber 2007). The species composition of the endophytic species within plants appear to normally consists of a few dominant species together with various other less dominant species (Arnold *et al.* 2000; Arnold *et al.* 2003). The dominant species may be more abundant in spore loads, better adapted to survive in the plant tissue or able to colonize various tissue within the tree (Kumar and Hyde 2004).

Host specificity of endophytes remains unclear as a general concept. It has been suggested that dominant fungal species are specific to host species or related host species (Gennaro *et al.* 2003). The latter authors concluded that the endophytic communities of related host species are similar, in contrast to distantly related host species that harbour different types of endophytic communities (Gennaro *et al.* 2003). Studying host preference, Lodge (1997) showed that many endophytic fungi have the ability to infect a wide range of diverse hosts in the same environment. A recent study, however, identified distinct host related fungal communities in tropical leaves (Arnold *et al.* 2000).

The geographic origin and locality of the plant species often determines the composition of the endophyte assemblage (Fisher *et al.* 1993; Collado and Platas 1999; 2001; Hoffman and Arnold 2008). It has been noted that endophytic assemblages of the same species of tree growing in diverse geographical regions are different and thus more a reflection of the fungi present in the particular environment. For example, studies conducted by Fisher *et al*. (1994) on *Quercus ilex* revealed that endophytic assemblages outside the natural environment of the tree in England, Majorca and Switzerland consisted of different endophytic species compared to those in its native Mediterranean region. This suggests that trees growing outside their natural region can become colonized with indigenous endophytic fungi of those regions (Fisher *et al.* 1994). Other studies by Collado and Platas. (1999) confirmed the geographical influence of the endophyte assemblage by sampling *Quercus ilex* from four different sites in Switzerland and found the frequencies of dominant and less dominant species were dependent on the sampling sites (Collado and Platas 1999). Fungal endophytes present in *Trachycarpus fortune* (palm) within and beyond its natural geographical range were investigated by Taylor *et al*. (1999), who showed similar endophytic assemblages in palm trees that were continuously distributed in China were the palm is native, but palms growing in Australia and Switzerland showed different endophytic assemblages (Taylor *et al.* 1999).

Endophytic fungi have been studied intensively over the last three decades (Arnold 2007). Numerous studies have focused on listing endophytes in tree hosts, plant tissues and countries (Rodrigues and Samuels 1990; Arnold *et al.* 2000; Fruhlich *et al.* 2000; Gamboa and Bayman 2001; Photita *et al.* 2001; Guo *et al.* 2003; Arnold 2007; Arnold *et al.* 2007; Finlay and Clay 2007; Hoffman and Arnold 2008). In addition to compiling inventories, these studies also aim to understand the factors influencing the assemblages of endophytes within plant tissues, endophytic colonization patterns, the interaction with their hosts, characterization of metabolites produced by endophytes, the production of mycotoxins and their prospective roles in biological control with regards to invasive plant species and pest insects and pathogens (Rowan 1993; Wilson 1993; Cheeke 1995; Azevedo *et al.* 2000; Wang *et al.* 2007). Endophytes can also be exploited for human benefit, because some endophytes have been noted to exhibit anti-microbial, anti-cancer and anti-malarial properties (Strobel 2003; Wiyakrutta *et al.* 2004). Some endophytes are also pathogens during a part of the life-cycle that can cause disease in economically important tree species

and have thus been targeted for study for this reason (Carroll 1988; Romero *et al.* 2001; Photita *et al.* 2004; Slippers and Wingfield 2007).

Clearly, insight into the context of endophytic interactions will have important consequences for ecology and for potential applications such as bio-control and understanding host-pathogen interactions. By investigating and identifying endophytes in different hosts and tissues at different localities, a vast array of endophytic diversity is expected to be revealed. It is, therefore, imperative to use an accurate tool to identify endophytes, including rare species.

6. Identification of fungal endophytes

6.1 Conventional culture-based identification

Fungal endophytes are usually isolated from surface-sterilized plant material and thereafter individual isolates are purified by sub-culturing (Schulz *et al.* 1993). Identification of these fungi is then based on morphological studies based on culture morphology and morphological characteristics of fruiting structures. Often, specialized techniques are required for specific fungal groups, e.g. identification of *Fusarium* species requires specific media for sporulation (Toussoun and Nelson 1976).

Identification of diverse endophytes based solely on morphological characters can be problematic due to a lack of mycological expertise for certain fungal groups, inadequate or scattered taxonomic literature, and a lack of resolution due to morphologically indistinctive characteristics (Harrington and Rizzo 1999; Arnold *et al.* 2000; Arnold 2007). Morphological identification of a diverse collection of fungi is a difficult and time-consuming skill to acquire. Consequently, very few people are taxonomically adept and capable of designating a specimen or culture identity to species level (Tautz *et al.* 2003).

There are a number of limitations to the identification of endophytic fungi by using only morphological characters. For example, it has been estimated that only 1% of microbes are readily culturable (Pace 1997; Warnecke and Hess 2009). Numerous endophytes are also

sterile in culture, which is the part of the life cycle most commonly used in morphological identifications (Arnold 2007). Slow growing endophytes that are often out competed by the dominant endophytic groups, are especially difficult to characterize as their presence is usually concealed in the culturing process. It is also possible that the competitive interactions of the dominant endophytes with less dominant endophytes may result in different phenotypes of the less dominant endophytes depending on the competitive interaction exhibited by the dominant endophyte (Cannon and Simmons 2002). Such limitations require a combination of morphological and molecular techniques to more efficiently identify species.

6.2 Identification of fungi using DNA sequence data and DNA databases

Contemporary species identification in fungi relies strongly on DNA sequences, because the morphological species concept (species differentiated according to morphological characteristics) appear to fail commonly for these organisms (Nilsson *et al.* 2006). Even a few hundred base pairs can have high levels of polymorphisms that corresponds to the identities of various taxonomical units and can, therefore, be powerful identification tools (Mallet and Willmott 2003). It is consequently not surprising to find that virtually all species descriptions today include sequence data. In many cases where morphological data cannot distinguish species, they can be clearly separated based on sequence data (Harrington and Rizzo 1999; Bickford *et al.* 2007; Shenoy *et al.* 2007; Pavlic *et al.* 2009b).

Sequence data obtained from gene regions are compared to sequences available on public databases to establish species identity. Online databases such as GenBank, European Molecular Biology Laboratory (EMBL) and DNA Data Bank of Japan (DDBJ), are used extensively in systematic studies of fungi in general. Due to the growing databases of DNA sequences used in phylogenetic and taxonomic studies, it is thus increasingly possible to use DNA sequences routinely for identification purposes. For some groups of fungi, such as pathogenic *Candida* spp. systematic studies have been done so extensively that virtually all species have been typed based on DNA sequences, facilitating identifications based on DNA sequences (Odds and Jacobsen 2008).

The DNA databases maintained at the National Centre of Bioinformatics Institute (NCBI) or the European Bioinformatics Institute (EBI) are not for taxonomic purposes, as there is no taxonomic standard under which submissions must be made (Tautz *et al.* 2003). These databases also do not stipulate morphological, biogeographical and ecological information to be linked with the submissions. There is also a vast number of environmental fungal sequences available on NCBI-GenBank that lack any taxonomic information (Arnold 2008). Records such as sequencing methods, collection information and linked publications of these sequences is usually incomplete and often these sequences cannot be linked to any other related organism, which makes this information of little use to taxonomists.

Due to the great number of fungal sequences deposited in GenBank and the lack of imperative information linked to sequence submissions, there is a need to create high quality databases dedicated to specific fungal groups. In this regard there are already several databases dedicated to specific groups of fungi based on vouchered sequences. For example, UNITE is an rDNA sequence database devoted to ectomycorrhizal asco- and basidiomycetes (Koljalg *et al.* 2005). Several tools are linked to the database to aid identification of unknown species, including simple BLAST (Basic Alignment tool) methods and Galaxies that allows for web based phylogenetic analyses (Nilsson *et al.* 2004). Furthermore, the Fungal Tree of Life (AFTOL) project aims to generate large amount of sequence data from multiple loci for vouchered specimens in order to test the hypotheses of fungal phylogenies (Schoch *et al.* 2006; Hibbett *et al.* 2007). Sequences linked to the AFTOL publications are available for BLAST comparisons, given broad coverage of the majority of classes, orders and most families of fungi. Other databases focus on specific genera. For example, TrichOkey version2 is an online database for *Trichoderma* spp. based on vouchered *Trichoderma* ITS sequence (Druzhinina *et al.* 2006). FUSARIUM-ID is a database for the identification of *Fusarium* spp. based on partial Translation Elongation Factor 1-alpha. DNA sequences also only contains vouchered sequences attached to publicly available cultures and validly described species (Geiser *et al.* 2004). Furthermore, the CBS- KNAW (www.cbs.knaw.nl) fungal biodiversity centre has databases dedicated to filamentous fungi and yeasts as well as other databases such as the Actinomycetes strain database, Aphyllophorales database, *Fusarium* database, Anamorph-Teleomorph database, *Penicillium* database, the Medical fungi database, the *Phaeoacremonium* database, the *Russula* database and the *Mycosphaerella* database.

Databases dedicated to certain ecological niches such as UNITE for mycorrhizal fungi (Koljalg *et al.* 2005), are extremely useful for ecological studies focusing on fungal biodiversity and to elucidate the interactions in these communities in natural ecosystems. Recent bar-coding consortia focused on certain niches such as INDOOR MYCOTA, with general fungal barcodes deposited in BOLD. No database, however, exists for other ecological groups such as endophytes. Such a database would greatly improve the taxonomic status of taxa in the group as many fungal endophytes are not assigned to meaningful taxonomical groups. Such a database would also represent a tool to study of the hyper-diversity of endophytes, and would enable cross references between studies.

There are large numbers of incorrect sequence data deposited in public databases. A common sequencing error is that arising from of chimeras. A chimera is a sequence that is formed from more than one DNA template (Shuldiner *et al.* 1989; Mylvaganam and Dennis 1992). These occur during the PCR reaction where synthesis begins in one template, halted and then begins again from another template that shares significant homology from the original template. Therefore, chimera consists of two phylogenetically distinct sequences. The formation of chimeras is higher for mixed DNA templates in a PCR reaction as in the case for DNA extracted from environmental samples. One of the simplest ways to deal with the presence of chimeras is to sequence bi-directionally (forward and reverse) and build a consensus sequence for traditional sequencing. For the detection of chimeras generated from next generation sequencers, programs such as Chimera checker are able to identify chimeras by BLASTING the 5ôand 3ôend of the sequence separately to determine if they originate from two different templates. If Chimeraôs are not detected and these erroneous sequences are deposited in a database, a lowered quality database will arise (Bridge *et al.* 2003) and this can severely hamper correct species identification based on BLAST.

7. Conclusions

Studying endophytes on woody plants and trees presents a number of challenges. Firstly endophytes associated with trees have been poorly studied (Arnold *et al.* 2007) and many tree species have been considered only peripherally. Endophyte surveys on these plants are complicated because of culturing bias and often it is complex to delimit taxonomic groups for certain endophytes. This is because identification of endophytes is complicated as they

usually represent ill-studied taxonomic groups, often does not sporulate in culture and frequently fall in the fungal groups that are referred to as ñunculturedò, ñunknownò, ñsp.ò or that have emerged from environmental surveys that did not characterize the organisms found.

Many endophytes are notorious pathogens causing devastation to agricultural crops and plantations. Such endophytes include the Botryosphaeriaceae that cause severe disease symptoms to valuable eucalypt plantations during adverse environmental conditions (Slippers *et al.* 2004b; Slippers and Wingfield 2007). Furthermore, it is possible that many pathogens may have an unknown endophytic stage or may exist as endophytes in a different host that has yet to be described. This would not be detected in the absence of endophyte surveys.

Because of the difficulties of identifying endophytes, modern technologies can enhance studies to characterize endophyte communities. This would largely involve identifying endophytes based on sequence data. Combining DNA bar-coding to identify endophytes based on vouchered cultures; with a metagenetic approach using 454-pyrosequencing will enable studies of the entire endophytic community. In this way, a comprehensive database containing both the cultured and uncultured endophytes can be built. Sequence data can also be supplemented with more targeted studies employing fungus-specific probes or multi-gene phylogenies. These databases will thus form the foundation of future surveys, whereby pathogens occurring as endophytes that are potentially important in forests and to forestry, can be better studied.

References

Ahumada R (2003) *Pathogens in commercial Eucalyptus plantations in Chile, with special reference to Mycosphaerella and Botryosphaeria species* Master's Dissertation, University of Pretoria.

Amend AS, Seifert KA, Samson R and Bruns TD (2010) Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proceedings of the National Academy of Sciences* **107**, 13748-13753.

Ansorge WJ (2009) Next-generation DNA sequencing techniques. *New Biotechnology* **25**, 195-203.

Arnold AE (2007) Understanding the diversity of foliar endophytic fungi: Progress, Challenges, and Frontiers. *Fungal Biology Reviews* **21**, 51-66.

Arnold AE (2008) *FESIN.* August 3, *Metadata for environmental sequences*. University of Arizona, USA.

Arnold AE, Henk DA, Eells RL, Lutzoni F and Vilgalys R (2007) Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia* **99**, 185-206.

Arnold AE and Lutzoni F (2007) Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* **88**, 541-549.

Arnold AE, Maynard Z and Gilbert GS (2001) Fungal endophytes in dicotyledonous neotropical trees: patterns of abundance and diversity. *Mycological Research* **105**, 1502-1507.

Arnold AE, Maynard Z, Gilbert GS, Coley DC and Kursar TA (2000) Are tropical fungal endophytes hyperdiverse? *Tropical Endophyte Diversity* **3**, 267-274.

Arnold AE, Mejía LC, Kyllo D, Rojas EI, Maynard Z, Robbins N and Herre EA (2003) Fungal endophytes limit pathogen damage in a tropical tree. *Proceedings of the National Academy of Sciences* **100**, 15649-15654.

Azevedo JL, Maccheroni Jr W, Pereira JO and De Araújo WL (2000) Endophytic microorganisms: a review on insect control and recent advances on tropical plants. *Electronic Journal of Biotechnology* **3**, 15-16.

Barrett RDH and Hebert PDN (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology* **83**, 481-491.

Bettucci L, Alonso R and Tiscornia S (1999) Endophytic mycobiota of healthy twigs and the assemblage of species associated with twig lesions of *Eucalyptus globulus* and *E. grandis* in Uruguay. *Mycological Research.* **103**, 468-472.

Bettucci L and Saravay M (1993) Endophytic fungi of *Eucalyptus globulus*: a preliminary study. *Mycological Research* **97**, 679-682.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK and Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* **22**, 148-155.

Blodgett JT, Swart WJ, Louw SM and Weeks WJ (2000) Species composition of endophytic fungi in *Amaranthus hybridus* leaves, petioles, stems, and roots. *Mycologia* **92**, 853-859.

Braslavsky I, Hebert B, Kartalov E and Quake SR (2003) Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences* **100**, 3960-3964.

Bridge PD, Roberts PJ, Spooner BM and Panchal G (2003) On the unreliability of published DNA sequences. *New Phytologist* **160**, 43-48.

Buee M, Reich M, Murat C, Morin E, Nilsson RH, Uroz S and Martin F (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist* **184**, 449-456.

Campanile G, Ruscelli A and Luisi N (2007) Antagonistic activity of endophytic fungi towards *Diplodia corticola* assessed by in vitro and in planta tests. *European Journal of Plant Pathology* **117**, 237-246.

Cannon PF and Simmons CM (2002) Diversity and host preference of leaf endophytic fungi in the Iwokrama Forest Reserve, Guyana. *Mycol Soc America* **94**, 210-220.

Carroll G (1988) Fungal endophytes in stems and leaves: from latent pathogen to mutualistic symbiont. *Ecology* **69**, 2-9.

Carroll G (1995) Forest endophytes: pattern and process. *Canadian Journal of Botany* **73**, 1316-1324.

Casiraghi M, Labra M, Ferri E, Galimberti A and De Mattia F (2010) DNA barcoding: a six-question tour to improve users' awareness about the method. *Briefings in bioinformatics* **11**, 440-453.

Chase MW and Fay MF (2009) Barcoding of Plants and Fungi. *Science* **325**, 682-683.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N and Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals.

Philosophical Transactions of the Royal Society B: Biological Sciences **360**, 1889-1895.

Cheeke PR (1995) Endogenous toxins and mycotoxins in forage grasses and their effects on livestock. *Journal of Animal Science* **73**, 909-918.

Chen SF, Pavlic D, Roux J, Slippers B, Xie YJ, Wingfield MJ and Zhou XD (2011) Characterization of Botryosphaeriaceae from plantation grown *Eucalyptus* species in South China. *Plant Pathology* **60**, 739-751.

Cheplick GP (2007) Costs of fungal endophyte infection in *Lolium perenne* genotypes from Eurasia and North Africa under extreme resource limitation. *Environmental and Experimental Botany* **60**, 202-210.

Clay K (1993) Fungal endophytes of plants: biological and chemical diversity. *Natural toxins* **1**, 147-149.

Clay K and Schardl C (2002) Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist,* **160**, 99-127.

Collado J and Platas G (1999) Geographical and seasonal influences on the distribution of fungal endophytes in *Quercus ilex*. *New Phytologist* **144**, 525-532.

Collado J and Platas G (2001) Identification of an endophytic *Nodulisporium* sp. from *Quercus ile*x in central Spain as the anamorph of *Biscogniauxia mediterranea* by rDNA sequence analysis and effect of different ecological factors on distribution of the fungus. *Mycologia* **93**, 875-886.

Coutinho TA, Preisig O, Mergaert J, Cnockaert MC, Riedel KH, Swings J and Wingfield MJ (2002) Bacterial blight and dieback of *Eucalyptus* species, hybrids, and clones in South Africa. *Plant disease* **86**, 20-25.

Coutinho TA, Roux J, Riedel K, Terblanche J and Wingfield MJ (2000) First report of bacterial wilt caused by *Ralstonia solanacearum* on eucalypts in South Africa. *Forest Pathology* **30**, 205-210.

Coutinho TA and Venter SN (2009) *Pantoea ananatis*: an unconventional plant pathogen. *Molecular Plant Pathology* **10**, 325-335.

Coutinho TA, Wingfield MJ, Alfenas AC and Crous PW (1998) *Eucalyptus* rust: a disease with the potential for serious international implications. *Plant Disease* **82**, 819-825.

Creer S, Fonseca VG, Porazinska DL, Giblin Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML and Lambshead PJD (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology* **19**, 4-20.

Crous PW (1998) *Mycosphaerella* spp. and their anamorphs associated with leaf spot diseases of *Eucalyptus*. *Mycologia Memoir* **21**, 1-21.

Crous PW, Knox-Davies PS and Wingfield MJ (1988) *Phaeoseptoria eucalypti* and *Coniothyrium ovatum* on *Eucalyptus*. *Phytophylactica* **20**, 337-340.

Crous PW, Phillips AJL and Wingfield MJ (1991) The genera *Cylindrocladium* and *Cylindrocladiella* in South Africa, with special reference to forest nurseries. *South African Forestry Journal* **157**, 69-85.

Crous PW, Phillips AJL and Wingfield MJ (1993) New records of *Cylindrocladium* and *Cylindrocladiella* spp. in South Africa. *Plant Pathology* **42**, 302-305.

Crous PW and Wingfield MJ (1996) Species of *Mycosphaerella* and their anamorphs associated with leaf blotch disease of *Eucalyptus* in South Africa. *Mycologia*, 441-458.

Davis EC, Franklin JB, Shaw AJ and Vilgalys R (2003) Endophytic *Xylaria* (Xylariaceae) among liverworts and angiosperms: phylogenetics, distribution, and symbiosis 1. *American Journal of Botany* **90**, 1661-1667.

De Beer ZW, Begerow D, Bauer R, Pegg GS, Crous PW and Wingfield MJ (2006) Phylogeny of the Quambalariaceae fam. nov., including important *Eucalyptus* pathogens in South Africa and Australia. *Studies in Mycology* **55**, 289-298.

de Magalhães JP, Finch CE and Janssens G (2011) Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews* **9**, 315-323.

DeSalle R, Egan MG and Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B:* **360**, 1905-1916.

Desprez-Loustau ML, Marçais B, Nageleisen LM, Piou D and Vannini A (2006) Interactive effects of drought and pathogens in forest trees. *Annals of Forest Science* **63**, 597-612.

Droege M and Hill B (2008) The Genome Sequencer FLXÊ Systemð Longer reads, more applications, straight forward bioinformatics and more complete data sets. *Journal of Biotechnology* **136**, 3-10.

Druzhinina IS, Kopchinskiy AG and Kubicek CP (2006) The first 100 *Trichoderma* species characterized by molecular data. *Mycoscience* **47**, 55-64.

Faeth SH and Fagan WF (2002) Fungal endophytes: common host plant symbionts but uncommon mutualists. *Integrative and Comparative Biology* **42**, 360-368.

Feau N, Vialle A, Allaire M, Tanguay P, Joly DL, Frey P, Callan BE and Hamelin RC (2009) Fungal pathogen (mis-) identifications: A case study with DNA barcodes on *Melampsora* rusts of aspen and white poplar. *Mycological Research* **113**, 713-724.

Finlay RD and Clay K (2007) Fungal endophytes in forests, woody plants and grassland ecosystems: diversity, functional ecology and evolution. *Fungal Biology Reviews* **21**, 49-50.

Fisher PJ, Petrini O, Petrini LE and Sutton BC (1994) Fungal endophytes from the leaves and twigs of *Quercus ilex* from England, Majorca and Switzerland. *New Phytologist* **127**, 133-137.

Fisher PJ, Petrini O and Sutton BC (1993) A comparative study of fungal endophytes in leaves, xylem and bark of *Eucalyptus nitens* in Australia and England. *Sydowia* **45**, 1-14.

Forns X, Bukh J, Purcell RH and Emerson SU (1997) How *Escherichia coli* can bias the results of molecular cloning: preferential selection of defective genomes of *hepatitis C* virus during the cloning procedure. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13909-13914.

Frézal L and Leblois R (2008) Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution* **8**, 727-736.

Fruhlich J, Hyde KD and Petrini O (2000) Endophytic fungi associated with palms. *Mycological Research* **104**, 1202-1212.

Gamboa MA and Bayman P (2001) Communities of Endophytic Fungi in Leaves of a Tropical Timber Tree (*Guarea guidonia*: Meliaceae) 1. *Biotropica* **33**, 352-360.

Gazis R, Rehner S and Chaverri P (2011) Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology* **20**, 3001-3013.

Geiser DM, del Mar Jimenez-Gasco M, Kang S, Makalowska I, Veeraraghavan N, Ward TJ, Zhang N, Kuldau GA and O'Donnell K (2004) FUSARIUM-ID v. 1.0: A DNA sequence database for identifying *Fusarium*. *European Journal of Plant Pathology* **110**, 473-479.

Geiser DM, Klich MA, Frisvad JC, Peterson SW, Varga J and Samson RA (2007) The current status of species recognition and identification in *Aspergillus*. *Studies in Mycology* **59**, 1-10.

Gennaro M, Gonthier P and Nicolotti G (2003) Fungal endophytic communities in healthy and declining *Quercus robur L.* and *Q. cerris L.* trees in northern Italy. *Journal of Phytopathology* **151**, 529-534.

Gezahgne A, Roux J, Slippers B and Wingfield MJ (2004) Identification of the causal agent of *Botryosphaeria* stem canker in Ethiopian *Eucalyptus* plantations. *South African Journal of Botany* **70**, 241-248.

Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A and Gillevet PM (2010) Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathogens* **6**, e1000713

Gilmore SR, Grafenhan TOM, Louis-Seize G and Seifert KA (2009) Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fusarium*. *Molecular Ecology Resources* **9**, 90-98.

Glenn TC (2011) Field guide to next generation DNA sequencers. *Molecular Ecology Resources* **11**, 759-769.

Gonthier P, Gennaro M and Nicolotti G (2006) Effects of water stress on the endophytic mycota of *Quercus robur*. *Fungal Diversity* **21**, 69-80.

Grobbelaar JW, Aghayeva DN, de Beer ZW, Bloomer P, Wingfield MJ and Wingfield BD (2009) Delimitation of *Ophiostoma quercus* and its synonyms using multiple gene phylogenies. *Mycological Progress* **8**, 221-236.

Gryzenhout M, Myburg H, Van der Merwe NA, Wingfield BD and Wingfield MJ (2004) *Chrysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Studies in Mycology* **50**, 119-142.

Guazzaroni ME, Beloqui A, Golyshin PN and Ferrer M (2009) Metagenomics as a new technological tool to gain scientific knowledge. *World Journal of Microbiology and Biotechnology* **25**, 945-954.

Guo LD, Huang GR, Wang Y, He WH, Zheng WH and Hyde KD (2003) Molecular identification of white morphotype strains of endophytic fungi from *Pinus tabulaeformis*. *Mycological Research* **107**, 680-688.

Hajibabaei M, dewaard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM and Hebert PDN (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B:* **360**, 1959-1967.

Hajibabaei M, Singer GAC, Clare EL and Hebert PDN (2007a) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC biology* **5**, 24-31.

Hajibabaei M, Singer GAC, Hebert PDN and Hickey DA (2007b) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* **23**, 167-172.

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68**, 669-685.

Harrington TC and Rizzo DM (1999) *Defining species in the fungi.* In: Worrall J.J, (eds,) Structure and Dynamics of Fungal Populations. Kluwer Academic Press, Netherlands, 43-72.

Hausmann A, Haszprunar G and Hebert PDN (2010) DNA barcoding the Geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. *PloS one* **6**, e17134.

Havermans C, Nagy ZT, Sonet G, De Broyer C and Martin P (2010) DNA barcoding reveals new insights into the diversity of Antarctic species of *Orchomene sensu lato* (Crustacea: Amphipoda: Lysianassoidea). *Deep Sea Research Part II: Topical Studies in Oceanography* **58**, 230-241

Hebert PDN, Cywinska A, Ball SL and DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 313-321.

Hebert PDN and Gregory TR (2005a) The promise of DNA barcoding for taxonomy. *Systematic Biology* **54**, 852-859.

Hebert PDN, Penton EH, Burns JM, Janzen DH and Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator. Proceedings of the National Academy of Sciences of the United States of America* **101**, 14812-14817.

Hebert PDN, Stoeckle MY, Zemlak TS and Francis CM (2004b) Identification of birds through DNA barcodes. *PLoS Biology* **2**, 1657-1663.

Herre EA, Mejia LC, Kyllo DA, Rojas E, Maynard Z, Butler A and Van Bael SA (2007) Ecological implications of anti-pathogen effects of tropical fungal endophytes and mycorrhizae. *Ecology* **88**, 550-558.

Hert DG, Fredlake CP and Barron AE (2008) Advantages and limitations of next-generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* **29**, 4618-4626.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM and Lucking R (2007) A higher-level phylogenetic classification of the fungi. *Mycological Research* **111**, 509-547.

Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P and Nilsson RH (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. *Fungal Biology Reviews* **25**, 38-47

Hill JE, Seipp RP, Betts M, Hawkins L, Van Kessel AG, Crosby WL and Hemmingsen SM (2002) Extensive profiling of a complex microbial community by high-throughput sequencing. *Applied and Environmental Microbiology* **68**, 3055-3066.

Hoffman MT and Arnold AE (2008) Geographic locality and host identity shape fungal endophyte communities in cupressaceous trees. *Mycological Research* **112**, 331-344.

Hogg ID and Hebert PDN (2004) Biological identification of springtails (*Hexapoda: Collembola*) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology* **82**, 749-754.

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, Van Der Bank M, Chase MW, Cowan RS, Erickson DL and Fazekas AJ (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* **106**, 12794-11297.

Hollingsworth PM, Graham SW and Little DP (2011) Choosing and Using a Plant DNA Barcode. *PloS one* **6**, e19254.

Hong PY, Hwang C, Ling F, Andersen GL, LeChevallier MW and Liu WT (2010) Analysis of bacterial biofilm communities in water meters of a drinking water distribution system via pyrosequencing. *Applied and Environmental Microbiology* **76**, 5631-5635.

Horner DS, Pavesi G, Castrignan ÃT, De Meo PDO, Liuni S, Sammeth M, Picardi E and Pesole G (2010) Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in bioinformatics* **11**, 181-197.

Huber T, Faulkner G and Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**, 2317-2319.

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**, 3-17.

Hugenholtz P and Tyson GW (2008) Microbiology: metagenomics. *Nature* **455**, 481-483.

Hunter GC, Crous PW, Carnegie AJ and Wingfield MJ (2009) *Teratosphaeria nubilosa*, a serious leaf disease pathogen of *Eucalyptus* spp. in native and introduced areas. *Molecular plant pathology* **10**, 1-14.

Hunter GC, Crous PW, Roux J, Wingfield BD and Wingfield MJ (2004a) Identification of *Mycosphaerella* species associated with *Eucalyptus nitens* leaf defoliation in South Africa. *Australasian Plant Pathology* **33**, 349-355.

Hunter GC, Roux J, Wingfield BD, Crous PW and Wingfield MJ (2004b) *Mycosphaerella* species causing leaf disease in South African *Eucalyptu*s plantations. *Mycological Research* **108**, 672-681.

Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA and Sogin ML (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genetics* **4**, e1000255.

Huse SM, Huber JA, Morrison HG, Sogin ML and Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome biology* **8**, R143.

Jumpponen A and Jones KL (2009) Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist* **184**, 438-448.

Karkouri KE, Murat C, Zampieri E and Bonfante P (2007) Identification of internal transcribed spacer sequence motifs in truffels: a first step toward their bar coding. *Applied and Environmental Microbiology* **73**, 5320-5330.

Karow J (2008) *International Barcode of Life Project Sees Room for Sanger, New Sequencing Techs*. [online] Available: www.genomeweb.com/sequencing/international-barcode-life-project-sees-room-sanger-new-sequencing-techs-0 [Acessed 3 march 2009].

Keane PJ, Kile GA and Podger FD (2000) *Diseases and pathogens of eucalypts* Csiro.

Kim SH and Breuil C (2001) Common nuclear ribosomal internal transcribed spacer sequences occur in the sibling species *Ophiostoma piceae* and *O. quercus*. *Mycological Research* **105**, 331-337.

Kirchman DL, Cottrell MT and Lovejoy C (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environmental Microbiology* **12**, 1132-1143.

Koljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R and Larsson E (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* **166**, 1063-1068.

Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M and Turner SW (2008) Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences* **105**, 1176-1181.

Kumar DSS and Hyde KD (2004) Biodiversity and tissue-recurrence of endophytic fungi in *Tripterygium wilfordii*. *Fungal Diversity* **17**, 69-90.

Lahaye R, Savolainen V, Duthoit S, Maurin O and van der Bank M (2008) A test of psbK-psbI and atpF-atpH as potential plant DNA barcodes using the flora of the Kruger National Park as a model system (South Africa). *Nature Precedings (http://0-hdl.handle.net.innopac.up.ac.za/10101/npre.2008.1896.1).*

Letourneau A, Seena S, Marvanová L and Bärlocher F (2010) Potential use of barcoding to identify aquatic hyphomycetes. *Fungal Diversity* **40**, 51-64.

Linde C, Kemp GHJ and Wingfield MJ (1994) Diseases of pines and eucalypts in South Africa associated with *Pythium* and *Phytophthora* species. *South African Forestry Journal* **169**, 25-32.

Lodge DJ, Fisher PJ and Sutton BC (1996) Endophytic fungi of *Manilkara bidentata* leaves in Puerto Rico. *Mycologia* **88**, 733-738.

Lorenz JG, Jackson WE, Beck JC and Hanner R (2005) The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1869-1877.

Luo K, Chen SL, Chen KL, Song JY, Yao H, Ma X, Zhu YJ, Pang XH, Yu H and Li XW (2010) Assessment of candidate plant DNA barcodes using the Rutaceae family. *Science China Life Sciences* **53**, 701-708.

Mallet J and Willmott K (2003) Taxonomy: renaissance or Tower of Babel? *Trends in Ecology & Evolution* **18**, 57-59.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-141.

Maseko B, Burgess TI, Coutinho TA and Wingfield MJ (2007) Two new *Phytophthora* species from South African *eucalyptus* plantations. *Mycological Research* **111**, 1321-1338.

Miller SE (2007) DNA barcoding and the renaissance of taxonomy. *Proceedings of the National Academy of Sciences* **104**, 4775-4776.

Mohali S, Slippers B and Wingfield MJ (2007) Identification of Botryosphaeriaceae from *Eucalyptus*, *Acacia* and *Pinus* in Venezuela. *Fungal Diversity* **25**, 103-125.

Morgan JL, Darling AE and Eisen JA (2010) Metagenomic sequencing of an in vitro-simulated microbial community. *PloS one* 5:e10209.

Mouhamadou B, Carriconde F, Gryta H, Jargeat P, Manzi S and Gardes M (2008) Molecular evolution of mitochondrial ribosomal DNA in the fungal genus *Tricholoma*: Barcoding implications. *Fungal Genetics and Biology* **45**, 1219-1226.

Myburg H, Gryzenhout M, Wingfield BD and Wingfield MJ (2002) -Tubulin and Histone H3 gene sequences distinguish *Cryphonectria cubensis* from South Africa, Asia, and South America. *Botany* **80**, 590-596.

Mylvaganam S and Dennis PP (1992) Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui*. *Genetics* **130**, 399-410.

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Daniel R and Gilbert J (2010) Pyrosequencing-Based Assessment of Bacterial Community

Structure Along Different Management Types in German Forest and Grassland Soils. *PLoS ONE* **6**, 1387-1390.

Nakabonge G, Roux J, Gryzenhout M and Wingfield MJ (2006) Distribution of Chrysoporthe canker pathogens on *Eucalyptus* and *Syzygium* spp. in eastern and southern Africa. *Plant disease* **90**, 734-740.

Nguyen HDT and Seifert KA (2008) Description and DNA barcoding of three new species of *Leohumicola* from South Africa and the United States. *Persoonia: Molecular Phylogeny and Evolution of Fungi* **21**, 57-69.

Nicol N, Kemp GHJ and Wingfield MJ (1993) *Corticium salmonicolor* associated with a serious canker disease of *Eucalyptus* in South Africa. *Phytophylactica* **25**, 198.

Nilsson RH, Abarenkov K, Veldre V, Nylinder S and De WIT (2010) An open source chimera checker for the fungal ITS region. *Molecular Ecology Resources* **10**, 1076-1081

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N and Larsson KH (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics Online* **4**, 193-201.

Nilsson RH, Larsson KH and Ursing BM (2004) Galaxieð CGI scripts for sequence identification through automated phylogenetic analysis. *Bioinformatics* **20**, 1447-1452.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH and Kõljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS One* **1** e59. doi:10.1371.

Nowrousian M (2010) Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryotic cell* **9**, 1300-1310.

O'Donnell K, Cigelnik E and Nirenberg HI (1998) Molecular systematics and phylogeography of the *Gibberella fujikuroi* species complex. *Mycologia* **90**, 465-493.

Odds FC and Jacobsen MD (2008) Multilocus sequence typing of pathogenic *Candida* species. *Eukaryotic Cell* **7**, 1075-1084.

Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**, 734-740.

Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M and Fire AZ (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research* **35**, e130.

Pavlic D, Slippers B, Coutinho T and Wingfield M (2009) Molecular and phenotypic characterisation of three phylogenetic species discovered within the *Neofusicoccum parvum/N. ribis* complex. *Mycologia* **101**, 636-647.

Pavlic D, Slippers B, Coutinho TA and Wingfield MJ (2007) Botryosphaeriaceae occurring on native *Syzygium cordatum* in South Africa and their potential threat to *Eucalyptus. Plant Pathology* **56**, 624-636.

Pavlic D, Slippers B, Coutinho TA and Wingfield MJ (2009b) Multiple gene genealogies and phenotypic data reveal cryptic species of the Botryosphaeriaceae: A case study on the *Neofusicoccum parvum/N. ribis* complex. *Molecular Phylogenetics and Evolution* **51**, 259-268.

Pérez CA, Altier N, Simeto S, Wingfield MJ, Slippers B and Blanchette RA (2008) Botryosphaeriaceae from *Eucalyptus* and native Myrtaceae in Uruguay. *Agrociencia* **12**, 19-30.

Petrosino JF, Highlander S, Luna RA, Gibbs RA and Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. *Clinical chemistry* **55**, 856-866.

Photita W, Lumyong S, Lumyong P and Hyde KD (2001) Endophytic fungi of wild banana (*Musa acuminata*) at Doi Suthep Pui National Park, Thailand. *Mycological Research* **105**, 1508-1513.

Photita W, Lumyong S, Lumyong P, McKenzie EHC and Hyde KD (2004) Are some endophytes of *Musa acuminata* latent pathogens. *Fungal Diversity* **16**, 131-140.

Pop M and Salzberg SL (2008) Bioinformatics challenges of new sequencing technology. *Trends in Genetics* **24**, 142-149.

Porazinska DL, Giblin-Davis RM, Faller L, Farmerie W, Kanzaki N, Morris K, Powers TO, Tucker AE, Sung WAY and Thomas WK (2009a) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources* **9**, 1439-1450.

Porazinska DL, Sung WAY, Giblin Davis RM and Thomas WK (2009b) Reproducibility of read numbers in high throughput sequencing analysis of nematode community composition and structure. *Molecular Ecology Resources* **10**, 666-676.

Rach J, DeSalle R, Sarkar IN, Schierwater B and Hadrys H (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in *Odonata*. *Proceedings of the Royal Society B: Biological Sciences* **275**, 237-247.

Ratnasingham S and Hebert PDN (2007) bold: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Notes* **7**, 355-364.

Reis-Filho J (2009) Next-generation sequencing. *Breast Cancer Research* **11**, S12-19.

Richardson P (2010) Special issue: next generation DNA sequencing. *Genes* **1**, 385-387.

Rodas CA, Slippers B, Gryzenhout M and Wingfield MJ (2009) Botryosphaeriaceae associated with *Eucalyptus* canker diseases in Colombia. *Forest Pathology* **39**, 110-123.

Rodrigues KF and Samuels GJ (1990) Preliminary study of endophytic fungi in a tropical palm. *Mycological Research* **94**, 827-830.

Rodriguez RJ, White Jr JF, Arnold AE and Redman RS (2009) Fungal endophytes: diversity and functional roles. *New Phytologist* **182**, 314-330.

Romero A, Carrion G and Rico-Gray V (2001) Fungal latent pathogens and endophytes from leaves of *Parthenium hysterophorus* (Asteraceae). *Fungal Diversity* **7**, 81-87.

Ronaghi M and Elahi E (2002) Pyrosequencing for microbial typing. *Journal of Chromatography* **782**, 67-72.

Rosen GL, Sokhansanj BA, Polikar R, Bruns MA, Russell J, Garbarine E, Essinger S and Yok N (2009) Signal processing for metagenomics: extracting information from the soup. *Current genomics* **10**, 493-510.

Rothberg JM and Leamon JH (2008) The development and impact of 454 sequencing. *Nature Biotechnology* **26**, 1117-1124.

Roux J, Heath RN, Van der Hoef A and Wingfield MJ (2001) First report of pink disease on *Eucalyptus* and *Podocarpus* in South Africa. *Phytopathology* **78**, 427-433.

Roux J, Mthalane ZL, De Beer ZW, Eisenberg B and Wingfield MJ (2006) *Quambalaria* leaf and shoot blight on *Eucalyptus nitens* in South Africa. *Australasian Plant Pathology* **35**, 427-433.

Rowan DD (1993) Lolitrems, peramine and paxilline: mycotoxins of the ryegrass/endophyte interaction. *Agriculture, Ecosystems & Environment* **44**, 103-122.

Saikkonen K (2007) Forest structure and fungal endophytes. *Fungal Biology Reviews* **21**, 67-74.

Saikkonen K, Faeth SH, Helander M and Sullivan TJ (1998) Fungal endophytes: a continuum of interactions with host plants. *Annual Review of Ecology and Systematics* **29**, 319-343.

Saikkonen K, Wäli P, Helander M and Faeth SH (2004) Evolution of endophyteï plant symbioses. *Trends in Plant Science* **9**, 275-280.

Santana QC, Coetzee MPA, Steenkamp ET, Mlonyeni OX, Hammond GNA, Wingfield MJ and Wingfield BD (2009) Microsatellite discovery by deep sequencing of enriched genomic libraries. *BioTechniques* **46**, 217-223.

Savolainen V, Cowan RS, Vogler AP, Roderick GK and Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B:* **360**, 1805-1811.

Schindel DE and Miller SE (2005) DNA barcoding a useful tool for taxonomists. *Science* **435**, 17-22.

Schoch C, Seifert KA and Crous PW (2011) Progress on DNA Barcoding of Fungi *IMA fungus* **2**, 83-89.

Schoch CL, Kohlmeyer J, Volkmann-Kohlmeyer B, Tsui CKM and Spatafora JW (2006) The halotolerant fungus *Glomerobolus gelineus* is a member of the Ostropales. *Mycological research* **110**, 257-263.

Schoeneweiss DF (1981) The Role of Environmental Stress in Diseases of Woody Plants. *Plant Disease* **65**, 308-314.

Schonau APG (1991) Role of eucalypt plantations in timber supply and forest conservation in sub-Saharan Africa. *South African Forestry Journal* **156**, 56-60.

Schuette UME, Abdo Z, Foster J, Ravel J, Bunge J, Solheim B and Forney LJ (2010) Bacterial diversity in a glacier foreland of the high Arctic. *Molecular Ecology* **19**, 54-66.

Schulz B and Boyle C (2006) What are Endophytes? *Microbial Root Endophytes* **9**, 1-13.

Schulz B, Wanke U, Draeger S and Aust HJ (1993) Endophytes from herbaceous plants and shrubs: effectiveness of surface sterilization methods. *Mycological Research* **97**, 1447-1450.

Seifert KA (2009) Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* **9**, 83-89.

Seifert KA, Samson RA, deWaard JR, Houbraken J, Lévesque CA, Moncalvo JM, Louis-Seize G and Hebert PDN (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences* **104**, 3901-3906.

Shendure J and Ji H (2008) Next-generation DNA sequencing. *Nature biotechnology* **26**, 1135-1145.

Shenoy BD, Jeewon R and Hyde KD (2007) Impact of DNA sequence-data on the taxonomy of anamorphic fungi. *Fungal Diversity* **26**, 1-54.

Shuldiner AR, Nirula A and Roth J (1989) Hybrid DNA artifact from PCR of closely related target sequences. *Nucleic acids research* **17**, 4409.

Shumway M, Cochrane G and Sugawara H (2010) Archiving next generation sequencing data. *Nucleic acids research* **38**, D870-D871.

Sieber TN (2007) Endophytic fungi in forest trees: are they mutualists? *Fungal Biology Reviews* **21**, 75-89.

Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, Bali V and Batra N (2009) Metagenomics: Concept, methodology, ecological inference and recent advances. *Biotechnology Journal* **4**, 480-494.

Slippers B, Burgess T, Pavlic D, Ahumada R, Maleme H, Mohali S, Rodas C and Wingfield MJ (2009) A diverse assemblage of Botryosphaeriaceae infect *Eucalyptus* in native and non-native environments. *Southern Forests: a Journal of Forest Science* **71**, 101-110.

Slippers B, Crous PW, Denman S, Coutinho TA, Wingfield BD and Wingfield MJ (2004a) Combined multiple gene genealogies and phenotypic characters differentiate several species previously identified as *Botryosphaeria dothidea*. *Mycologia* **96**, 83-101.

Slippers B, Fourie G, Crous PW, Coutinho TA, Wingfield BD, Carnegie AJ and Wingfield MJ (2004b) Speciation and distribution of *Botryosphaeria* spp. on native and introduced *Eucalyptus* trees in Australia and South Africa. *Studies in Mycology* **50**, 343-358.

Slippers B, Fourie G, Crous PW, Coutinho TA, Wingfield BD and Wingfield MJ (2004c) Multiple gene sequences delimit *Botryosphaeria australis* sp. nov. from *B. lutea*. *Mycologia* **96**, 1030-1041.

Slippers B, Stenlid J and Wingfield MJ (2005a) Emerging pathogens: fungal host jumps following anthropogenic introduction. *Trends in ecology & evolution (Personal edition)* **20**, 420-428.

Slippers B, Summerel BA, Crous PW, Coutinho TA, Wingfield BD and Wingfield MJ (2005b) Preliminary studies on *Botryosphaeria* species from Southern Hemisphere conifers in Australasia and South Africa. *Australasian Plant Pathology* **34**, 213-220.

Slippers B and Wingfield MJ (2007) Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biology Reviews* **21**, 90-106.

Smith H, Crous PW, Wingfield MJ, Coutinho TA and Wingfield BD (2001) *Botryosphaeria eucalyptorum* sp. nov., a new species in the *B. dothidea*-complex on *Eucalyptus* in South Africa. *Mycologia* **93**, 277-285.

Smith H, Kemp GHJ and Wingfield MJ (1994) Canker and die back of *Eucalyptus* in South Africa caused by *Botryosphaeria dothidea*. *Plant pathology* **43**, 1031-1034.

Smith H, Wingfield MJ, Crous PW and Coutinho TA (1996a) *Sphaeropsis sapinea* and *Botryosphaeria dothidea* endophytic in *Pinus* spp. and *Eucalyptus* spp. in South Africa. *South African Journal of Botany,* **62**, 86-88.

Smith H, Wingfield MJ and Petrini O (1996b) *Botryosphaeria dothidea* endophytic in *Eucalyptus grandis* and *Eucalyptus nitens* in South Africa. *Forest Ecology and Management* **89**, 189-195.

Stockinger H, Krüger M and Schübler A (2010) DNA barcoding of arbuscular mycorrhizal fungi. *New Phytologist* **187**, 461-474.

Stone JK (1986) Foliar endophytes of *Pseudotsuga menziesii (Mirb.) Franco*. Cytology and physiology of the host-endophyte relationship. Ph.D. Thesis University of Oregon, Eugene, U.S.A.

Strobel GA (2003) Endophytes as sources of bioactive products. *Microbes and Infection* **5**, 535-544.

Summerbell RC, Levesque CA, Seifert KA, Bovers M, Fell JW, Diaz MR, Boekhout T, De Hoog GS, Stalpers J and Crous PW (2005) Microcoding: the second step in DNA barcoding. *Philosophical Transactions of the Royal Society B:* **360**, 1897-1903.

Swart L, Crous PW, Petrini O and Taylor JE (2000) Fungal endophytes of proteaceae, with particular emphasis on *Botryosphaeria proteae*. *Mycoscience* **41**, 123-127.

Tautz D, Arctander P, Minelli A, Thomas RH and Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology & Evolution* **18**, 70-74.

Taylor JE, Hyde KD and Jones EBG (1999) Endophytic fungi associated with the temperate palm, *Trachycarpus fortunei*, within and outside its natural geographic range. *New Phytologist* **142**, 335-346.

Taylor K, Barber PA, St J. Hardy GE and Burgess TI (2008) Botryosphaeriaceae from tuart (*Eucalyptus gomphocephala*) woodland, including descriptions of four new species. *Mycological Research*, **113**, 337-353.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G and Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* **188**, 291-301.

Toussoun TA and Nelson PE (1976) A pictorial guide to the identification of Fusarium species according to the taxonomic system of Snyder and Hansen. Pennsylvania State University Press, University Park, 1-42.

Turnbull JW (1999) Eucalypt plantations. *New Forests* **17**, 37-52.

Turnbull JW (2000) Diseases and Pathogens of Eucalyptus. In: Keane PJ, Podger FD, Brown BN (eds.) *Economic and social importance of eucalypts.* CSIRO, Australia, pp 1-7.

Valdivia-Granda W (2008) The next meta-challenge for Bioinformatics. *Bioinformation* **2**, 358-362.

Valentini A, Pompanon F and Taberlet P (2008) DNA barcoding for ecologists. *Trends in Ecology & Evolution* 24, 110-117.

Voelkerding KV, Dames SA and Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry* **55**, 641-658.

Wang B, Priest MJ, Davidson A, Brubaker CL, Woods MJ and Burdon JJ (2007) Fungal endophytes of native *Gossypium* species in Australia. *Mycological Research* **111**, 347-354.

Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* **360**, 1847-1857.

Warnecke F and Hess M (2009) A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *Journal of Biotechnology* **142**, 91-95.

Whiteman NK, Santiago-Alarcon D, Johnson KP and Parker PG (2004) Differences in straggling rates between two genera of dove lice (Insecta: Phthiraptera) reinforce population genetic and cophylogenetic patterns. *International Journal for Parasitology* **34**, 1113-1119.

Whitworth TL, Dawson RD, Magalon H and Baudry E (2007) DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B: Biological Sciences* **274**, 1731-1739.

Will C, Thurmer A, Wollherr A, Nacke H, Herold N, Schrumpf M, Gutknecht J, Wubet T, Buscot F and Daniel R (2010) Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes. *Applied and Environmental Microbiology* **76**, 6751-6759.

Wilson D (1993) Fungal endophytes: out of sight but should not be out of mind. *Oikos* **68**, 379-384.

Wilson D (1995) Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos* **73**, 274-276.

Wingfield MJ (2003) Increasing threat of diseases to exotic plantation forests in the Southern Hemisphere: lessons from *Cryphonectria canker*. *Australasian Plant Pathology* **32**, 133-139.

Wingfield MJ, Crous PW and Coutinho TA (1996) A serious canker disease of *Eucalyptus* in South Africa caused by a new species of *Coniothyrium*. *Mycopathologia* **136**, 139-145.

Wingfield MJ, Crous PW and Swart WJ (1993) *Sporothrix eucalypti* (sp. nov.), a shoot and leaf pathogen of *Eucalyptus* in South Africa. *Mycopathologia* **123**, 159-164.

Wingfield MJ and Knox-Davies PS (1980) Observations on diseases in pine and Eucalyptus plantations in South Africa. *Phytophylactica* **12**, 57-63.

Wingfield MJ, Slippers B, Hurley BP, Coutinho TA, Wingfield BD and Roux J (2008) Eucalypt pests and diseases: growing threats to plantation productivity. *Southern Forests: a Journal of Forest Science* **70**, 139-144.

Wingfield MJ, Slippers B, Roux J and Wingfield BD (2001) Worldwide movement of exotic forest fungi, especially in the tropics and the Southern Hemisphere. *BioScience* **51**, 134-140.

Wingfield MJ, Swart WJ and Abear BJ (1989) First record of Cryphonectria canker of *Eucalyptus* in South Africa. *Phytophylactica* **21**, 311-313.

Witthuhn RC, Wingfield BD, Wingfield MJ, Wolfaardt M and Harrington TC (1998) Monophyly of the conifer species in the *Ceratocystis coerulescens* complex based on DNA sequence data. *Mycologia* **1**, 96-101.

Wiyakrutta S, Sriubolmas N, Panphut W, Thongon N, Danwisetkanjana K, Ruangrungsi N and Meevootisom V (2004) Endophytic fungi with anti-microbial, anti-cancer and anti-malarial activities isolated from Thai medicinal plants. *World Journal of Microbiology and Biotechnology* **20**, 265-272.

Wu GD, Lewis JD, Hoffmann C, Chen YY, Knight R, Bittinger K, Hwang J, Chen J, Berkowsky R and Nessel L (2010) Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16 S sequence tags. *BMC microbiology* **10**, 206-220.

Zhang JB and Hanner R (2011) DNA barcoding is a useful tool for the identification of marine fishes from Japan. *Biochemical Systematics and Ecology* **39**, 31-42.

Zhou X, Ren L, Meng Q, Li Y, Yu Y and Yu J (2010) The next-generation sequencing technology and application. *Protein and Cell* **1**, 520-536.

Table 1. A comparison of the three dominant next generation sequencing platforms available

	454 Roche	**Illumina**	**SOLiD**
Sequencing chemistry	Pyrosequencing	Polymeraseï based sequencing by synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Bridge PCR	Emulsion PCR
Mb/run	400-600 Mb	>25 000 Mb	>20 000 Mb
Time/run	7 hrs	4 days	5 days
Read length	400- 500 bp	35 ï 75bp	50bp
Applications	- Genome sequencing - Metagenetics - Mini satellite discovery (Santana *et al.* 2009) - Genotyping, single nucleotide polymorphism	- Genome sequencing - Studying protein interactions - Small RNAôs - Genotyping, Single nucleotide polymorphisms - Gene expression by RNA-sew - His tone modifications - Transcript profiling	Same as Illumina

Table adapted from (Mardis 2008; Zhou *et al.* 2010; de Magalhães *et al.* 2011)

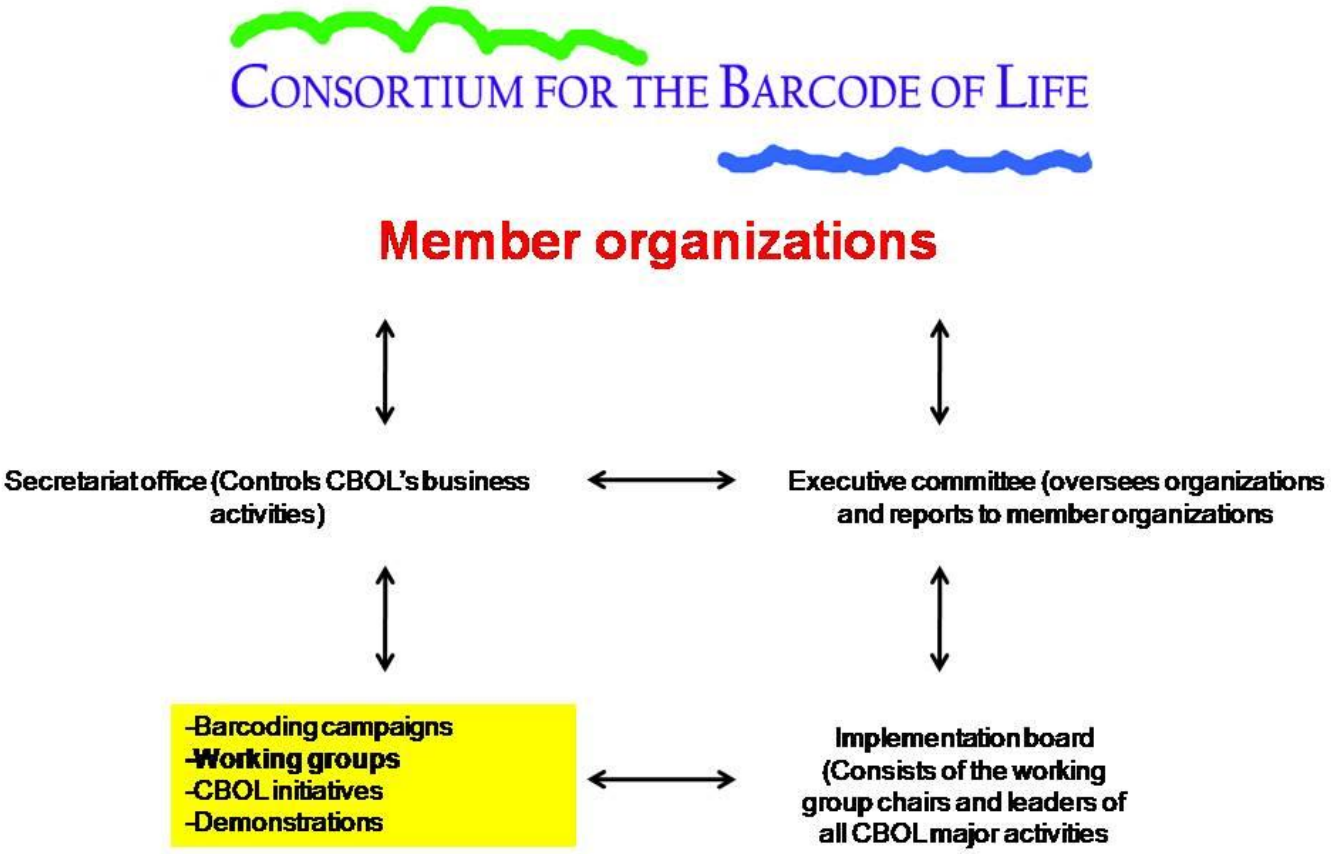


Fig. 1 Structure and management of Consortium for the Barcode of life

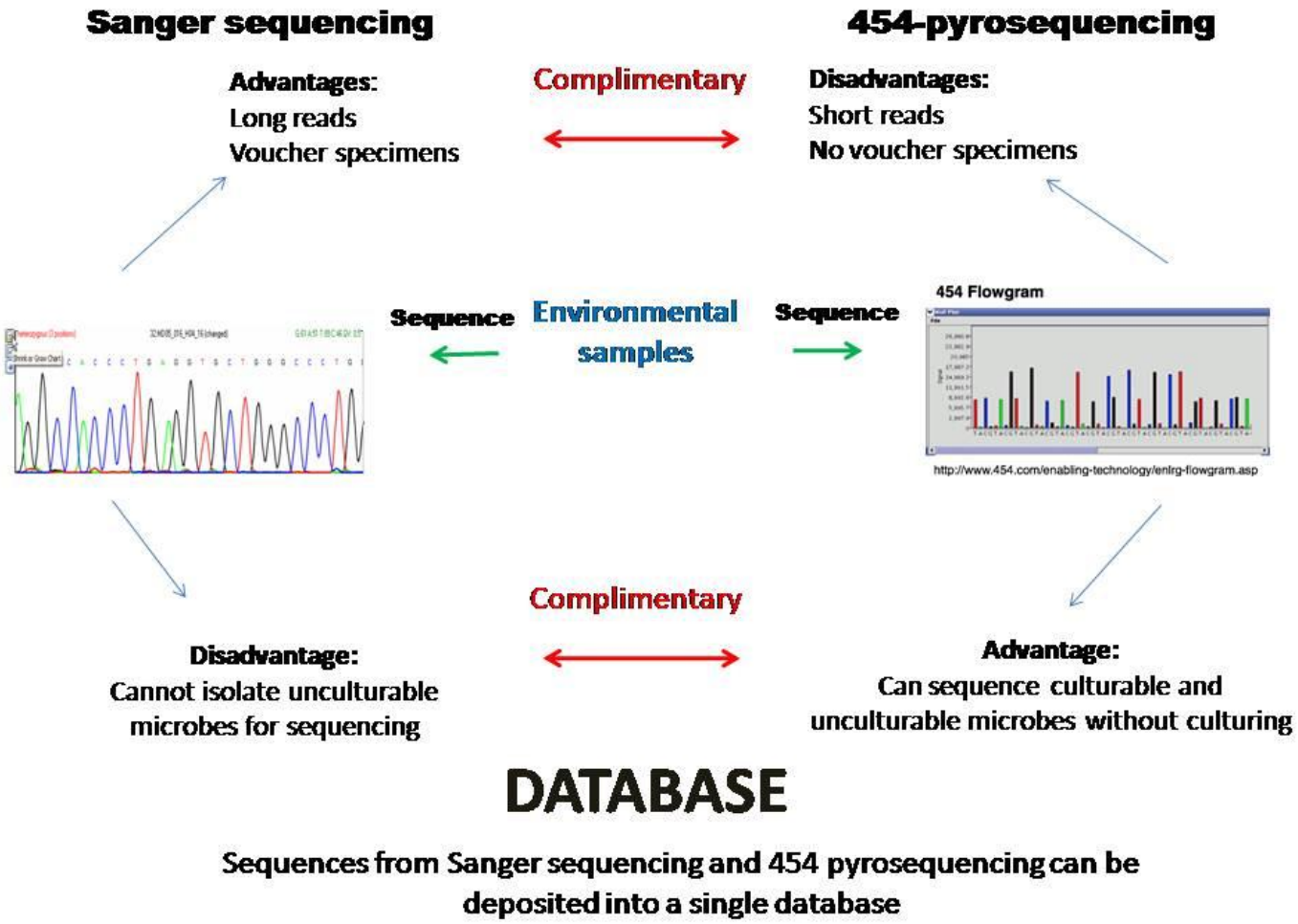


Fig. 2 a comparison of Sanger chain termination sequencing with 454 pyrosequencing for bar-coding environmental samples.

Chapter 2

Metagenetics and isolate bar-coding reveals high fungal endophyte diversity in *Eucalyptus grandis*

Abstract

Very little is known about fungal endophyte diversity in trees in South Africa. Furthermore, the prevalence of latent pathogens occurring as endophytes is poorly understood. Identifying fungal endophytes is difficult as traditional identification involves expertise, time and capacity consuming culturing techniques and some endophytes that are unculturable and slow growing are omitted. In this study we characterise the fungal endophytic community and latent pathogens in *Eucalyptus grandis* trees from a single location using a combination of metagenetics involving 454-pyrosequencing and complimentary isolate-based DNA bar-coding. Three *E. grandis* trees were chosen for the 454-pyrosequencing metagenetics approach. DNA was extracted from endophyte containing plant material, from which the ribosomal internal transcribed spacer (ITS) locus was amplified as a mini-barcode using fungal specific primers. Endophyte isolations were made from one of the *E. grandis* trees using a high density sampling approach. A detailed bar-coding workflow was developed whereby the most accurate identities could be assigned for the barcodes of these cultures, and for the mini-barcodes. A total of 27 878 ITS pyrosequencing reads were generated, but only 9 890 reads were used for analysis after quality control, representing 1 280 Molecular Operational Taxonomic Units (MOTUôs). Four hundred and sixty ITS sequences were generated from isolates, which represented 85 species. The 454-pyrosequencing thus revealed at least 15 times more endophyte taxa than the isolate based method. Accumulation curves showed that this large number of species did not represent the full diversity, and additional species would have been found with deeper sampling or sequencing. Results from our study showed that the endophytic community was dominated by fungal families that are known to be pathogens, including Mycosphaerellaceae, Teratosphaeriaceae and Botryosphaeriaceae.

Introduction

Fungal endophytes are a diverse group of microscopic organisms that inhabit various plant tissues without producing disease symptoms (Clay 1993; Wilson 1995; Saikkonen *et al.* 1998; Saikkonen *et al.* 2004; Schulz and Boyle 2006; Rodriguez *et al.* 2009). These endophytes have been found abundantly in all plants investigated (Stone *et al.* 2000; Arnold *et al.* 2007; Saikkonen 2007; Hoffman and Arnold 2008) and represent a significant contribution to global fungal diversity. Understanding this diversity has been hampered by the lack of methods to comprehensibly characterise entire endophytic communities. Recent advances in sequencing technologies and DNA bar-coding are likely to assist in exposing this diversity and promote an understanding of the ecological roles of endophytic fungi.

Traditional endophyte characterization is hindered by numerous constraints. For example, this approach requires the isolation and purification of cultures that can be a lengthy and time-intensive process. Furthermore, most identifications of endophytes in the past were based on characteristics of the cultures and fruiting structures formed on artificial media (Tautz *et al.* 2003), which has been limited by the fact that many endophytes are unculturable or sterile in culture and cannot be identified based on morphology (Arnold *et al.* 2007). In addition, many endophytes share morphological characters making identification in culture difficult (Harrington and Rizzo 1999). Other endophytes are slow growing and are often out-competed on medium by fast growing fungi and consequently overlooked. It is, therefore, likely that endophytic diversity studies based on culturing methods has underestimated true diversity (Arnold 2007). Species identification based on DNA sequence data is available to overcome some of these constraints.

DNA bar-coding aims to use a short standardized DNA sequence motif to identify species based on genetic variation on a sequence level (Hebert *et al.* 2003; Hebert and Gregory 2005; Savolainen *et al.* 2005). This approach for identification has been shown to be rapid and accurate, for example to identify bird species (Hebert *et al.* 2004b), fish species (Ward *et al.* 2005; Zhang and Hanner 2011), several insects such as butterflies (Hebert *et al.* 2004a; Hausmann *et al.* 2010) and many more. For animals the mitochondrial Cytochrome c oxidase (COX1) has been set as the standard bar-coding gene (Hebert and Gregory 2005). In plants the chloroplast genes such as *matK* and *rbcL* are used as bar-coding genes (Chase *et al.* 2007; Chase and Fay 2009; Hollingsworth *et al.* 2011).

The COX1 locus is unsuitable for fungi as it contains introns that result in variable sizes of amplified products (Seifert *et al.* 2007; Begerow *et al.* 2010; Stockinger *et al.* 2010). The internal transcribed spacer (ITS) region has now been recommended as the fungal barcode gene as it has been shown to be the best available locus to resolve up to 72% of fungal species (Schoch *et al.* 2011, 2012). This locus has also been commonly used in the past as the fungal species identification gene (Nilsson *et al.* 2009) and in fungal community studies (Dahlberg 2001; Buchan *et al.* 2002; Kennedy and Clipson 2003; Anderson and Cairney 2004; O'Brien *et al.* 2005; Peay *et al.* 2008).

Next generation sequencing techniques allow for thousands to millions of reads to be processed in parallel (Mardis 2008; Reis-Filho 2009; Voelkerding *et al.* 2009). This tool can be applied to environmental studies where individual species in mixed microbial populations can be differentiated, identified or compared in parallel based on sequence data. This approach significantly reduces the time and cost compared to conventional methods (Jarvie 2005; Hert *et al.* 2008; Mardis 2008; Ansorge 2009; Harismendy *et al.* 2009; Creer *et al.* 2010). Next generation sequencing tools also overcome the problem of identifying non-culturable microbes (Mardis 2008; Petrosino *et al.* 2009). Due to the high number of microbes in environmental samples, these studies require large amounts of sequence data to characterize all the diversity. It is costly and time consuming to produce such large data sets using traditional Sanger sequencing methods (Chan 2005; Metzker 2005; Delseny *et al.* 2010). Therefore next generation sequencing technologies have become a valuable tool to study microbial community diversity and structure.

Eucalyptus plantations are extremely valuable to the forestry industry of South Africa as a source of timber and pulp (Schonau 1991; Turnbull 1999; 2000). The majority of the mycological studies on *Eucalyptus* thus far have been focused on pathogenic fungi such as species of *Mycosphaerella* (Crous 1998; Crous *et al.* 2001; Crous *et al.* 2004a; Crous *et al.* 2006c; Hunter *et al.* 2006), *Cylindrocladium* (Crous *et al.* 2004c; Crous *et al.* 2006a), *Chrysoporthe* (Gryzenhout *et al.* 2004), Botryosphaeriaceae (Slippers *et al.* 2004a; Slippers *et al.* 2004b; Slippers *et al.* 2004c), and to a lesser degree species of *Cytospora* (Adams *et al.* 2004), *Coniella* (Van Niekerk and Groenewald 2004), *Quambalaria* (De Beer *et al.* 2006) and *Harknessia* (Lee *et al.* 2004). Various fungi other than pathogens were identified by Sankaran *et al*. (1995) amounting to 350 species reported on *Eucalyptus* between 1995-2002 (Sankaran *et al.* 1995; Hyde *et al.* 2007), while a number of recent studies have added more species (Summerbell *et al.* 2006; Crous *et al.* 2007; Cheewangkoon *et al.* 2009),

including 80 new species of *Mycosphaerella* (Crous 1998) and 20 new species of *Calonectria* specific to *Eucalyptus* (Hyde *et al.* 2007). Overall endophytic communities occurring on *Eucalyptus* (Myrtaceae) have, however, been poorly studied (Bettucci and Saravay 1993; Fisher *et al.* 1993; Bettucci and Alonso 1997; Bettucci *et al.* 1999) and no such studies have been undertaken in South Africa. Additional endophyte studies include those only focusing on latent pathogens with an endophytic stage in their life cycle, such as Botryosphaeriaceae (Slippers and Wingfield, 2007).

In this study, the fungal endophytic communities present in three *Eucalyptus grandis* trees were characterised using a metagenetic approach, combined with isolate based bar-coding. The primary aims were to (i) compare the efficiency of the two techniques to identify endophytes, (ii) characterise a proportion of the endophytic diversity and its spatial structure in the tree, and (iii) establish a database that can aid future endophytic studies on this and other hosts. Through this exhaustive approach a comprehensive overview of the fungi occurring as endophytes in the *Eucalyptus* trees sampled could be obtained and the potential pathogens with an endophytic stage could be identified.

2. Materials and Methods

2.1 Site and Sampling

Three *Eucalyptus grandis* trees, approximately 50 m apart, were sampled in April 2009 during autumn in Mtubatuba (KwaZulu-Natal province), South Africa (E 32ô 05ô 54,0 S28ô 29ô 53.0, 33m above sea level). The *Eucalyptus* plantations are common on the east coast of South Africa due to its sub-tropical climate and reasonably high rainfall. A total of 52 leaf samples (20 leaves from tree 1, 17 leaves from tree 2, and 15 leaves from tree 3), 42 petiole samples (12 petioles from tree 1, 16 petioles from tree 2 and 14 petioles from tree 3) 39 twig samples (9 twigs from tree 1, 10 twigs from tree 2 and 10 twigs from tree 3) and 14 trunk increment core samples (6 increments core from tree 1, 4 increment cores from tree 2 and 6 increment cores from tree3) were collected from the three trees for the 454-pyrosequencing study. From one of the *Eucalyptus grandis* trees (tree 1); the tissues sampled were collected and processed doubly so that half of the pieces were used to isolate endophytes. The other half of the pieces from tree 1, together with those with equal number of pieces of the other two trees, was used for DNA extraction and direct amplification of fungal DNA. Samples were placed in brown paper bags and transported to the laboratory for processing within two days of collection.

To remove fungal propagules and epiphytic fungi on the plant tissue surfaces (leaves, twigs, petioles and trunk increments) all the substrate tissue samples used in this study were surface-disinfested using 10 % hydrogen peroxide for 3 minutes after which samples were washed twice with sterile water for 1 minute. The samples used for the 454-pyrosequencing were dissected into 5 mm by 5 mm subsections, placed in 2 ml Eppendorf tubes and stored at -40 ºC until DNA extraction. The samples used for the bar-coding isolations were similarly dissected, where ten discs (5mm in diameter) from each leaf and five pieces (3 mm) were taken from each petiole, twig and bark increments. Tissue samples where then plated on malt extract agar (MEA) for fungal isolations.

2.2 Metagenetic 454-pyrosequencing

2.2.1 DNA extraction and amplification

Plant material was freeze dried and ground to a powder using sterile metal beads on a Mixer Mill type MM 301 Retsch^{R}tissue lyser (Retsch, Germany) for 10 min at a frequency of 30 cycles. Genomic DNA was extracted from leaf, twig, petiole and trunk increment samples separately using the Zymo plant/seed extraction Kit^{TM} (Zymo Research, USA) following the manufacturerôs instructions. The DNA concentrations for the tissue samples were not standardized because the DNA concentrations reflected those of the plant and not the fungal DNA.

The ITS rDNA regions of the endophytes were amplified directly from the mixture with plant DNA using the fungal specific forward primer ITS1F (Carbone and Kohn 1999) and reverse primer ITS4 (White *et al.* 1990). The PCR conditions were as follows: 94 ºC for 4 min, followed by 35 cycles of 94 ºC for 30 sec (denaturation), 55 ºC for 45 sec (annealing), 72 ºC for 1 min (elongation) and 72 ºC for 4 min (final elongation). The PCR products were then visualized in 2 % agarose gels using Gel Red (Biotium, Haward, California, USA) in 1x TAE buffer (Tris base 0.4 M, acetic acid 1 %, EDTA 0.5 M, pH 8.0).

The Zymoclean^{TM} Gel DNA Recovery Kit (Zymo Research) was used to purify 20 µl of the PCR products from the gel following manufacturerôs instructions. DNA concentrations for these subsamples were determined and standardized to 5 µg/µl using a NDï 1000 spectrophotometer V3.7.1 thermo (SCIENTIFIC, USA). A nested PCR was executed on the subsamples to add the required pyrosequencing adapted primers (fusion primers) to the amplified products (Table 1). The different tissue samples from all three trees were tagged

individually (Table 1), using 12 different tags (Inqaba Biotech, Pretoria, Gauteng, South Africa). The ITS1F primer contained the pyrosequencing A-adapter and the multiplex tag, while the ITS4 primer contained only the reverse B adapter (Table 1).

The nested PCR reactions were carried out in a final volume of 50 µl using 2 µl of gel purified template DNA (5 ng/µl), 2 µl of each primer (10 mM), 5 µl (10 mM) dNTPôs, 5 µl of 10x PCR buffer with $MgCl_2$, 1 unit *taq* polymerase (Roche Molecular Biochemicals, Almeda, California) and 32 µl sterile distilled water (SABAX water; Adcock Ingrams, Bryanston), with the PCR conditions the same as that of the initial PCR. After gel electrophoresis the PCR products were gel purified and pooled according to tissue type and tree number. This resulted in 11 amplicon libraries after the trunk increments from tree 2 failed to amplify. The final concentrations were determined using a Micro-Volume Full-Spectrum Fluorospectrometer thermo SCIENTIFIC at Inqaba Biotech (Pretoria, Gauteng, South Africa). The samples were then pooled at equimolar concentrations and sequenced on the Genome Sequencer FLX 454 Titanium (454 Life Sciences/ Roche Applied Biosystems, Bradford, USA) at Inqaba Biotech.

2.2.2 Pyrosequencing data analyses

2.2.2a. Pre-analyses filtering

A significant portion of pre-analyses were done on the data by the sequencing service provider (InqabaBiotechï www.inqababiotec.co.za). Poor quality sequences were discarded using stringent filtering algorithms in the 454-sequencing system software developed by Roche Applied Biosystems as prescribed in the Genome Sequencer FLX System Software Manual version 2.5p1 (Anonymous 2010). Read quality filters (signal quality) and read trimmings were also applied to the sequences. These filters included a key pass filter, a dot filter, a mixed filter, a signal intensity filter, a primer filter and a trim back filter. The resultant data were compiled in a composite FASTA FNA file where the individual files represented each tag to identify the origin of the sequence.

2.2.2b Sequence data filtering and sequence data analyses

A workflow was developed to analyse the 454-pyrosequence data obtained from Inqaba (Fig. 1). Sequence analyses were performed separately for the data sets of each tag and thereafter an analysis was done for the complete dataset. Data were analysed separately in

order to make tissue comparisons from the different trees. Only high quality sequences determined by the standard flow gram format (SFF) files were retained and sequence error reads were minimised by removing the multiplex tags and all sequences that contained mismatches within the priming site. Thereafter, primer sequences were removed using Python regular expression scripts (http://www.python.org/). Python scripts were also used to remove sequences shorter than 100 bp because they were considered too short for correct BLAST identifications (Wommack *et al.* 2008). The remaining sequences were passed through a chimera checker (Nilsson *et al.* 2010) to filter out chimeric sequences that potentially occurred as a pyrosequencing artefact.

Molecular Operational Taxonomic Units (MOTUôS) were assigned to the pyrosequences in step 3 of the pyrosequencing workflow (Fig. 1). In this step, all sequences that passed the filtering were clustered into MOTUôS using CD-HIT-EST (Li and Godzik 2006; Huang *et al.* 2010). MOTU's were clustered at 98% sequence similarity for that covered 98% of the sequence (c=-0.98). This threshold was also used for analysing the microbiota in the human intestine (Claesson *et al.* 2009).

Tentative identities were assigned to the individual MOTU's assigned at 98% similarity. To assign MOTUôs to order, family or genus, sequences belonging to each cluster were manually subjected to BLASTN against the non-redundant NCBI GenBank database (www.ncbi.nlm.nih.gov) after exclusion of environmental samples. The sequences within the cluster were then compared using phylogenetic applications, as performed in the bar-coding study (Fig. 2; described below) that included the closest reference sequences from GenBank in the analysis, to get a possible identification. Programmes such as MEGAN for the analysis of metagenomic data (Huson *et al.* 2007) that assign identities to the sequences directly from BLASTN of GenBank, were not used because this system is entirely dependent on the results determined from GenBank, which have numerous problems in assigning identities (Bridge *et al.* 2003; Vilgalys 2003; Nilsson *et al.* 2006). A quality controlled database for endophytic sequences does not yet exist.

Comparisons of the metagenetic 454-pyrosequencing MOTU sequences with the isolate barcodes (see below) were undertaken to verify taxonomic identities of the MOTU's obtained with pyrosequencing. Each cluster representing an MOTU from the 454-pyrosequencing sequence reads were compared against the local endophytic database containing ITS sequences from the isolated endophytes using the formatdb algorithm (Korf *et al.* 2003).

Only MOTU sequences with BIT scores of more than 250 when compared to the local endophytic database, were further analyzed because pyrosequencing reads ranged from 100 bp to 350 bp, while the ITS barcode sequences from isolates were always greater than 350 bp. The combined datasets of the isolate ITS barcode sequences, pyrosequencing MOTUôs and the closest reference sequences from GenBank were then subjected to further phylogenetic analysis similar to those done in the bar-coding workflow (see below). MOTU's have not been submitted as mini-barcodes (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008) to BOLD, the database for barcodes (www.barcodinglife.org) due to the processes that are still being finalised for fungi.

2.2.2c Statistical analyses

The data obtained for the endophyte community with the pyrosequencing analyses were analysed to test the significance of endophyte community structure. A nonparametric MANOVA was used to test the similarity of the MOTUôs recovered in each of the plant tissue types and results were visualized using non-metric multidimensional scaling (NMDS) (Prentice 1977). The NMDS analyses were performed on R2.12 using the *vegan* package (Oksanen *et al.* 2010). MOTU-accumulation curves and 95% confidence intervals were created via re-sampling (n=1000 iterations). Given the small number of independent samples (n=3 trees) the degree of nestedness in the data was ignored, which would tend to underestimate total MOTU diversity.

2.3 Isolate bar-coding

2.3.1 Fungal isolation

The surface sterilized tissue sections from tree 1 were placed on malt extract (MEA) agar (20 g malt extract, 20 g agar; Biolab, Midrand, SA) with four to six subsections from the same sample placed approximately 4 cm apart. The plates were incubated at 25 ºC for approximately 10 days. Growth of endophytic fungi from the plant tissue was checked daily to isolate slow growing fungi before they were overgrown by other fungi. Isolates were sub-cultured to obtain pure cultures by transferring single hyphae onto new MEA agar plates using a sterile needle. Plant tissue that did not show any initial fungal growth was monitored for a month. Purified cultures were incubated for two weeks under near-UV light and grouped in morphotypes according to colony shape, colour, texture, mycelium type, medium discolouration and colony density (Arnold et al., 2007).

2.3.2. Generation of DNA barcodes

Fungal mycelium was scraped from newly grown fungal cultures and transferred to 2 ml Eppendorf tubes, freeze dried and thereafter ground to a powder using sterile 2 mm metal beads on a Mixer Mill type MM 301 Retsch^{R} tissue lyser (Retsch, Germany) for 3 min at a frequency of 30 cycles. DNA was extracted from the endophytic cultures using a DNA extraction protocol described by Moller *et al.* (2002). DNA pellets were re-suspended in 50 µl SABAX water. The DNA concentrations were determined using the NDï 1000 spectrophotometer V3.7.1 thermo Fischer SCIENTIFIC, USA and DNA was diluted to 50 ng/µl concentrations for subsequent amplifications.

The full length ITS region, including parts of the small ribosomal subunit (SSU) and large ribosomal subunit (LSU), were PCR amplified using the forward V9G primer (Hoog and Ende 1998) and reverse LR5 (Vilgalys and Hester 1990) primer. The PCR reaction consisted of a 25 µl final volume and included 0.5 µl DNA template (50 ng/µl), 1 µl of each primer (10 mM), 2.5 µl (10 mM) dNTPôs, 2.5 µl of 10x PCR buffer with $MgCl_2$, 1 unit *taq* polymerase (Roche Molecular Biochemicals, Almeda, California) and 17 µl sterile distilled water (SABAX water; Adcock Ingrams, Bryanston). The PCR conditions and gel electrophoreses were identical to those described above.

PCR products were purified using Sephadex^{R}G-50 columns (Sigma-Aldrich, Germany). The purified PCR products were sequenced in both directions with the Big Dye terminator cycle sequencing kit (PE applied Biosystems - *Perkin Elmer, Foster City, California*) on the ABI 3130x1 genetic analyzer using the pop7 polymer (Applied PRISM, Foster City, California). The full length ITS regions using the forward V9G primer (Hoog and Ende 1998) and reverse ITS4 (White *et al.* 1990) internal primer was sequenced. Each sequencing PCR reaction contained 2.5 µl purified DNA, 2.1 µl reaction buffer, 0.5 µl ready reaction buffer (BigDye), 1.5 µl primer (10mM) and 5.4 µl sterile distilled water (SABAX water; Adcock Ingrams, Bryanston). The sequencing PCR conditions were as follows: 96 ºC for 10 sec followed by 25 cycles of 53 ºC for 5 sec, and 60 ºC for 4 min.

Nucleotide consensus sequences from the forward and reverse sequences were built using CLC Bio Workbench version 5 (CLC bio, Aarhus, Denmark). Contigs were checked manually for inconsistencies. All sequences generated in this study were deposited into a searchable local sequence database created for endophytes that was established using the formatdb

algorithm of BLAST (Korf *et al.* 2003). The barcodes could not be uploaded into BOLD because the fungal barcode has just been published by the time of completion of this thesis and protocols for incorporating data into BOLD has not yet been finalised.

In order to arrive at a trusted identity for the barcodes, a taxonomic approach was followed (see the workflow in Fig. 2). In the first step consensus sequences were BLASTed against the non-redundant NCBI-GenBank database (www.ncbi.nlm.nih.gov) after the exclusion of environmental samples, to determine generic affiliations. For species affiliations, sequences were manually grouped according to their BLAST results into tentative genera or MOTU groups. The sequences within these individual groups were then subjected to a basic phylogenetic analysis, to determine species delimitations. These were subjected to new BLAST searches coupled with rudimentary phylogenetic comparisons to establish MOTU identities and relatedness to other fungal taxa. Closest reference sequences of related species obtained from GenBank were added in an alignment using Molecular Evolutionary Genetic Analysis MEGA version 5 (Tamura *et al.* 2011). The matching part of the consensus sequences was aligned using the MUSCLE alignment option and phylogenetic analyses were performed using maximum parsimony. Trees were constructed in MEGA 5 using the standard default settings for each group to establish their relation to each other and phylogenetic distinctiveness. Species identifications were used to confirm phylogenetic constructs.

It was not possible to decide whether closely related groups that differed from each other with only a small number of base pairs represented different species. A cut off of 98 % similarity was used in these cases to distinguish between different phylogenetic groups (Step 3, Fig. 2). The possibility of further cryptic species existing within groups was not further investigated in this study. Species of genera such as *Penicillium*, where ITS is insufficient to distinguish between species, could also not be differentiated and were then designated as "sp." only.

3. Results

3.1 Metagenetic 454-Pyrosequencing

Pyrosequence data analysis

A total of 27 878 reads (Table 2) were generated from the pyrosequencing run (4 020 reads from tree 1, 14 539 reads from tree 2 and 19 289 reads from tree 3). After the filtering process, 9 890 reads were retained for bar-coding analyses across the three trees. Tree 3 accounted for most of the reads (69 %) used for the downstream analysis, followed by tree 1 (19 %) and tree 2 (11 %).

The pyrosequencing reads represented two fungal phyla, including the Ascomycota (99 %), and Basidiomycota (1 %). Approximately 60 % of the total reads grouped into MOTU's that could not be identified to family, order or class level and were labelled as ôunknownô These groups represented 46 % of reads in tree 1, 59 % in tree 2 and 74 % in tree 3. The other MOTUôs were distributed across 28 families (Table 3) and 14 orders (Fig. 3) of the Ascomycota and Basidiomycota. These orders included the Capnodiales (28 %), Hypocreales (6 %), Pleosporales (3 %), Dothideales (1.5 %), Xylariales (1 %), Agaricales (0.8 %), Botryosphaeriales (0.62 %), Eurotiales (0.5 %), mitosporic ascomycota (0.4 %), Lecanorales (0.1 %), Diaporthales (0.07 %), Calosphaeriales (0.04 %), Phyllachorales (0.02 %), and Saccharomycetales (0.01 %)

Approximately 35 % of the pyrosequencing read identities were present in the local endophytic database developed in the isolate bar-coding study (Fig. 4). Of the remaining 65 % of the pyrosequencing reads, only 5 % could be identified based on public databases. Of the isolate barcodes, 95 % were present in the 454-pyrosequencing run.

Statistical analysis

The distribution of MOTU's amongst the different plant tissues in the three trees varied significantly. Tree 1 had the highest MOTU richness in the twig tissue, followed by those in the petioles, leaves and twigs. Tree 2 had the highest MOTU richness in leaves followed by those in petioles and twigs (there were no trunk increment samples for tree 2). Tree 3 had the highest MOTU richness in the petioles, followed by those in twigs, trunk increments and leaves (Fig. 5).

The MOTUôs within each tree were more similar to each other than those between the tissues of the three different trees (Fig. 6). This was evident from the similarity of the

MOTUôs between trees (F_{25} 1.38, 0.0009), which was more significant than the similarity between tissue types ($F_{(approx)\ 35}$ 0.98, 0.5). The MOTU community was thus predicted more accurately by tree ($F_{2,10} = 1.39$; $P < 0.001$; $R^2 = 0.26$) than by tissue type ($F_{3,10} = 0.98$; $P < 0.5$; $R^2 = 0.27$).

The species accumulation curve was approximately linear up to a predicted eight samples where the number of MOTU's was approximately 1 200 with no observable asymptote (Fig. 7). This suggests the existence of a highly diverse endophyte pool. However, because the curve bent only slightly at sample 8, it is expected that the curve is not close to saturation and that additional MOTUôs would be identified with increased sampling.

3.2 Isolate Bar-coding

A total of 630 endophytic cultures were isolated from the four different tissue types obtained from a single *E. grandis* tree (tree 1). Only 460 ITS sequences were obtained because some isolates failed to amplify or produced poor sequence reads after repeated purification. There was a vast difference in the number of endophyte isolates from the different tissues sampled. The majority of the isolates (78 %) were from leaf tissue, followed by those from twigs (13 %), wood increments (5 %) and petioles (3 %) (Fig. 8). The 460 ITS sequences represented 85 species, distributed across 40 families and 19 orders (Table 4). Of the 85 cultured species, 19 could not be grouped into families because no sequences with definite and consistent generic, family or order name existed in GenBank or any other database searched. The Ascomycota accounted for 95% of the species identified. The top six most frequently isolated groups represented 70 % of the total isolates and included pestalotioid fungi (*Pestalotiopsis* spp. and related genera such as *Truncatella*), *Alternaria* spp., Mycosphaerellaceae, Teratosphaeriaceae, *Sydowia* spp. and Botryosphaeriaceae. Basidiomycota accounted for the remaining 5 % of the species and included *Athelia* sp., *Resincium* spp., *Schizophyllum commune*, *Phlebiopsis* spp. and an unknown Basidiomycete species (Fig. 9).

4. Discussion

This study revealed a very high diversity of fungal endophytes from a limited number of trees. Eighty-five species were identified from the isolate bar-coding using full length ITS rDNA barcodes, representing 19 orders and 40 families. The metagenetic approach, using pyrosequencing and a portion of the ITS rDNA locus as a mini-barcode, identified 1 280 MOTUs. This represented 15 times more species than those found using the isolation bar-coding, but consisting of fewer orders and families, namely 14 orders and 28 families. This is possibly due to the large number of unknowns, representing 84 % of the diversity of the metagenetic pyrosequencing reads that could not be grouped into families or genera. Search results were also influenced by the shorter sequences (up to 350 bp) obtained during pyrosequencing, compared to those from the fungal isolation based identifications where the full ITS1 and ITS2 region (approximately 650bp) were used.

The dominant groups found in both the metagenetic and isolate bar-coding study were congruent. These included species of Mycosphaerellaceae, Teratosphaeriaceae and Botryosphaeriaceae, *Sydowia eucalypti,* and *Alternaria* spp. Of these the Mycosphaerellaceae, Teratosphaeriaceae, Botryosphaeriaceae *Sydowia eucalypti* and *Sydowia eucalypti* include known pathogens of *Eucalyptus* (Burgess and Wingfield, 2004; Crous *et al*., 1995). Although the Mycosphaerellaceae, Teratosphaeriaceae and Botryosphaeriaceae are known as pathogens and have an endophytic stage in their life cycle (Crous and Wingfield 1996; Crous *et al.* 1998; Ganley *et al.* 2004; Slippers and Wingfield 2007; Swart *et al.* 2000) the fact that they were also dominant as endophytes within these trees have not been shown before.

The Mycosphaerellaceae, Teratosphaeriaceae and Botryosphaeriaceae species residing within the dominant groups require additional scrutiny to accurately determine their species identities. Detailed phylogenetic comparisons have been published within the specific genera (e.g. Crous *et al.* 2004a for the *Mycosphaerellaceae* and *Teratosphaeriaceae,* Crous *et al.* 2006d for the Botryosphaeriaceae) and usually includes additional gene sequences. For example, multi-gene analyses of the Translation Elongation Factor (TEF-Ŭ) and the RNA polymerase II subunit (RPB2) were required by Pavlic *et al.* (2008) to distinguish between cryptic Botryosphaeriaceae species from *Syzygium cordatum* trees in South Africa, which are closely related to *Eucalyptus*. This was because the ITS region is unable to determine species identities for this group of fungi (see Chapter 3 of this thesis). This, however, falls outside the scope of this study.

There was one significant difference in the dominant groups identified by the isolation bar-coding versus those identified by the metagenetics approach. Pestalotioid fungi such as species of *Pestalotiopsis* were abundantly isolated, but they did not form one of the dominant groups using 454-pyrosequencing. Only five reads were identified as pestalotioid, and the family to which this group of fungi belong, the Amphisphaeriaceae, only made up 0.53% of the total number of reads (Table 3). The high isolation trend of pestalotioid species was also observed by Lodge *et al.* (1996) from *Manilkara bidentata*, possibly because these species grow rapidly and are likely to have a strong competitive ability in culture. In addition many *Pestalotiopsis* species produce antifungal compounds that may limit other fungal growth (Maharachchikumbura *et al*. 2011) Culturing is thus not necessarily reflective of dominance or frequency within the endophytic community (Hyde and Soytong, 2008). Conversely, bias possibly introduced by pyrosequencing could include that amplifications to prepare samples for pyrosequencing could be biased to certain groups of fungi, and more dominant groups may be preferentially amplified (Kanagawa 2003; Bellemain *et al.* 2010). It will, however, be difficult to determine how well the PCR amplifications have worked for all the groups because the amplicons of all the species are mixed in each reaction.

Human and plant pathogens were identified from the endophyte isolates from *Eucalyptus*. Besides the Mycosphaerellaceae, Teratosphaeriaceae, Botryosphaeriaceae and *Sydowia eucalypti* that are known as *Eucalyptus* pathogens, other plant pathogens from completely different types of plants were also found. These included *Stemphylium solani*, a pathogen of several crops such as cotton, garlic and peppers (Hwang 2004; Mehta and Brogin 2000; Mehta 1998; Zheng *et al.* 2008), *Leptosphaerulina charatarum* that is known as a pathogen of wheat (Roux 1986; Toth *et al.* 2007), and the cob rot pathogen of maize *Nigrospora oryzae* (Hudson 1963). *Schizophyllum commune* (Rihs *et al.* 1996; Buzina *et al.* 2001), and *Meyerozyma guilliermondii* (Girmenia *et al.* 2006; Pfaller *et al.* 2006; Yamamura *et al.* 2009) are both implicated in human diseases. These sequences were carefully identified as these species based on ITS sequences available in GenBank, but whether the fungi isolated from *Eucalyptus* indeed represent these pathogens most likely will only be established with phylogenetic comparisons to pathogen strains of these species. Other species that are not known to be pathogenic at this stage, but that reside in genera that include plant pathogens included a *Devriesia* sp., a *Colletotrichum* sp., a *Pilidiella* sp., a *Phaemoniella* sp., a *Phaeoacremonium* sp., a *Paraphaeosphaeria* sp., several fungi in the Nectriaceae and Bionectriaceae, a *Fusarium* sp. and a *Cladosporium* sp. The fact that fungi closely resembling plant pathogens of other plants occur cryptically in such an unrelated host as

Eucalyptus, may represent unknown caches of these pathogens, a point important for disease control.

The limitation of relying on existing public databases to identify 454-pyrosequences and even the full barcodes for these endophytes was evident in this study. This is especially so due to the few sequences of South African fungi publically available, especially those of endophytes. Furthermore, NCBI GenBank contains large numbers of unconfirmed and erroneous sequences (Bridge *et al.* 2003; Vilgalys 2003; Nilsson *et al.* 2006). Confirmed sequences in public databases also have a bias towards economically important fungi that can be studied in culture (Crous *et al.* 2006b). Ideally, specialized databases need to be created that contain only quality-controlled sequences and identifications. UNITE (URL http://unite.zbi.ee) is such a database dedicated to the identification of ectomycorrhizal fungi with mechanisms in place to discard weakly matching sequences that can distort BLAST hits (Koljalg *et al.* 2005). DNA bar-coding aims to create a similar database, where sequences need to adhere to certain quality criteria and can also be linked to vouchered specimens. In this study, we thus initiated a database based on the more tedious isolation approach to compliment the metagenetic 454-pyrosequencing approach. This complimentary approach acknowledges the advantages and disadvantages of these two approaches, and have been recognised before (Moritz and Cicero 2004; Dasmahapatra and Mallet 2006; Parameswaran *et al.* 2007; Frézal and Leblois 2008; Hert *et al.* 2008; Karow 2008; Ansorge 2009; Warnecke and Hess 2009; Tedersoo *et al.* 2010; Hajibabaei *et al.* 2005; Ratnasingham and Hebert 2007).

The large numbers of reads produced by 454-pyrosequencing requires automated clustering and identification that is based on predetermined similarity threshold values. In our study we have opted for a 98% cut off value as used by Claesson *et al.* 2009. However, a single cut-off threshold value is not sufficient to distinguish between interspecific and intraspecific variability for all fungi (Nilsson *et al.* 2008). A 3% threshold is usually used as the standard for SSU in bacteria (Rosselló-Mora and Amann 2001) and has also been adopted in fungi (Nilsson *et al.* 2008), but a 3% threshold is too high for certain species such as *Aspergillus* and *Penicillium* (Nilsson *et al.* 2008; Nilsson *et al.* 2009a; Seifert 2009). A 1% threshold is too low for certain species and may over estimate diversity, as in the case of *Xylaria hypoxylon* for which a 24% intraspecific variability has been reported in the ITS region (Nilsson *et al.* 2008). ITS variability introduced by multiple copies of the ITS can also result in single nucleotide polymorphisms (SNPs) within the ITS region, as in the case with *Fusarium*

spp. (O'Donnell 1992). The threshold used in metagenetic studies thus needs to be determined with caution as it might either over or underestimate true species diversity, depending on the specific evolutionary rate and diversity in a specific group (Nilsson *et al.* 2011). For metagenetic studies where a wide diversity of fungi is expected to be found in natural environments, this is problematic because no single cut-off has as yet been acknowledged and it will have to be accepted that results are to be interpreted related to the chosen cut-off value.

Despite the large number of 454-pyrosequencing reads compared to those sequences generated for the isolate bar-coding, a number of species found in the cultural approach (5 % of the total) were not identified by the pyrosequencing. This could be due to the poor resolution during BLAST searches obtained with the shorter ITS sequences generated by the 454-pyrosequencing (Hillis and Dixon 1991; Hershkovitz and Lewis 1996). Technical biases may have been introduced during the sample preservation, DNA isolation methods and the subsequent PCR reactions from different plant tissue samples (Tedersoo *et al.* 2010). Many of the shortened reads were also discarded to prevent possible misidentifications and these could possibly represent some of the isolates. Singleton sequence reads were also not used in metagenetic analyses because it was impossible to determine whether these singletons were pyrosequencing-derived artefacts or true MOTUs originating from single individuals of a species (Tedersoo *et al.* 2010). Singletons can thus populate sequence databases with non-biological data if they emanate from sequencing artefacts (Tedersoo *et al.* 2010; Behnke *et al.* 2011; Nilsson *et al.* 2011) or they overestimate species diversity (Tedersoo *et al.* 2010; Nilsson *et al.* 2011). Singletons are, however, not a problem when working with isolates, as sequences can be repeated and confirmed. However, the singletons represented 976 reads (9.9 % of the total) in our study and they had had no affiliations with any sequence available on NCBI GenBank or in the local database created. Singletons thus do not present an explanation for the missing bar-coded taxa in the metagenetic data.

The sampling strategy undertaken in this study was not intended to discover all endophytes linked to *Eucalyptus*, but rather to establish and compare the different identification techniques of DNA bar-coding and pyrosequencing. Through this approach we hoped to estimate the maximum diversity in a single tree without having to sample excessively and to initiate a methodology for future surveys. However, species accumulation curves suggest that the existing biodiversity on *Eucalyptus* in that area has not nearly been reached. Even

so, the 630 endophytic isolates obtained from the one *Eucalyptus* tree is typical of the large scale isolations usually necessary for endophyte community studies (Arnold *et al.* 2000; Fruhlich *et al.* 2000; Guo *et al.* 2000; Gamboa and Bayman 2001; Photita *et al.* 2001; Arnold 2007; Arnold *et al.* 2007; Arnold and Lutzoni 2007; Hoffman and Arnold 2008). Such large numbers make it difficult and impractical to isolate from more trees. Furthermore, the sequencing costs to identify these endophyte cultures with full barcodes would be significantly higher than that required to achieve the equivalent output obtained by 454-pyrosequencing. Our approach of building a growing database based on isolate bar-coding, together with the power of 454 pyrosequencing, will likely make it possible to sample many more trees in future to fully capture and identify the diversity of the endophytes. However, it is recommended that samples used for 454-pyrosequencing and isolation should be processed and analysed in a standardised manner in order to compare results from different studies more accurately.

Acknowledgements

We thank the Thuthuka Programme of the National Research Foundation (NRF), members of the Tree Protection Co-operative Programme (TPCP) and the Department of Science and Technology (DST)/NRF Centre of Excellence in Tree Health Biotechnology (CTHB), South Africa, for financial support. Additionally, we thank Dr Jeff Garnas (FABI, University of Pretoria) for the ecological statistical analyses and Mr. Oliver Bezuid (University of Pretoria) worked closely with us in analysing the pyrosequencing data and developing the bioinformatics pipeline.

References

Adams GC, Wingfield MJ, Common R and Roux J (2004) Phylogenetic relationships and morphology of *Cytospora* species and related teleomorphs (Ascomycota, Diaporthales, Valsaceae) from *Eucalyptus*. *Studies in Mycology* **52**, 1-144.

Anderson IC and Cairney JWG (2004) Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. *Environmental Microbiology* **6**, 769-779.

Anonymous (2010) 454 Sequencing System Software Manual, v 2.5p1 Part B ï GS Run Processor, GS Reporter, GS Run Browser, GS Support Tool. 454 Life Sciences Corp., Branford, United Kingdom, pp. 1-96.

Ansorge WJ (2009) Next-generation DNA sequencing techniques. *New Biotechnology* **25**, 195-203.

Arnold AE (2007) Understanding the diversity of foliar endophytic fungi: progress, challenges, and frontiers. *Fungal Biology Reviews* **21**, 51-66.

Arnold AE, Henk DA, Eells RL, Lutzoni F and Vilgalys R (2007) Diversity and phylogenetic affinities of foliar fungal endophytes in loblolly pine inferred by culturing and environmental PCR. *Mycologia* **99**, 185-206.

Arnold AE and Lutzoni F (2007) Diversity and host range of foliar fungal endophytes: are tropical leaves biodiversity hotspots? *Ecology* **88**, 541-549.

Arnold AE, Maynard Z, Gilbert GS, Coley DC and Kursar TA (2000) Are tropical fungal endophytes hyperdiverse? *Tropical Endophyte Diversity* **3**, 267-274.

Begerow D, Nilsson H, Unterseher M and Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Microbiology and Biotechnology* **87**, 99-108.

Behnke A, Engel M, Christen R, Nebel M, Klein RR and Stoeck T (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology* **13**, 340-349.

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P and Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology* **10**, 189-120.

Bettucci L and Alonso R (1997) A comparative study of fungal populations in healthy and symptomatic twigs of *Eucalyptus grandis* in Uruguay. *Mycological Research* **1 01**, 1060-1064.

Bettucci L, Alonso R and Tiscornia S (1999) Endophytic mycobiota of healthy twigs and the assemblage of species associated with twig lesions of *Eucalyptus globulus* and *E. grandis* in Uruguay. *Mycological Research* **103**, 468-472.

Bettucci L and Saravay M (1993) Endophytic fungi of *Eucalyptus globulus*: a preliminary study. *Mycological Research* **97**, 679-682.

Bridge PD, Roberts PJ, Spooner BM and Panchal G (2003) On the unreliability of published DNA sequences. *New Phytologist* **160**, 43-48.

Buchan A, Newell SY, Moreta JIL and Moran MA (2002) Analysis of internal transcribed spacer (ITS) regions of rRNA genes in fungal communities in a southeastern US salt marsh. *Microbial Ecology* **43**, 329-340.

Burgess, T. and Wingfield, M.J. (2004) *Impact of fungi in natural forest ecosystems: A focus on Eucalyptus.* In: Sivasithamparam, K.,Dixon, K.W. and Barrett, R.L., (eds.) Microorganisms in plant conservation and biodiversity. Kluwer Academic Publishers, Dordrecht, 285-306.

Buzina W, Lang-Loidolt D, Braun H, Freudenschuss K and Stammberger H (2001) Development of molecular methods for identification of *Schizophyllum commune* from clinical samples. *Journal of Clinical Microbiology* **39**, 2391-2396.

Carbone I and Kohn LM (1999) A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia* **91**, 553-556.

Chan EY (2005) Advances in sequencing technology. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 13-40.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinan S, Petersen G, Seberg O, Jorgsensen T, Cameron KM and Carine M (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon* **56**, 295-299.

Chase MW and Fay MF (2009) Barcoding of Plants and Fungi. *Science* **325**, 682-683.

Cheewangkoon R, Groenewald JZ, Summerell BA, Hyde KD, To-Anun C and Crous PW (2009) Myrtaceae, a cache of fungal biodiversity. *Persoonia* **23**, 55-85.

Claesson MJ, O'Sullivan O, Wang Q, Nikkila J, Marchesi JR, Smidt H, De Vos WM, Ross RP and O'Toole PW (2009) Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS One* **4**, e6669.

Clay K (1993) Fungal endophytes of plants: biological and chemical diversity. *Natural Toxins* **1**, 147-149.

Creer S, Fonseca VG, Porazinska DL, Giblin Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML and Lambshead PJD (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology* **19**, 4-20.

Crous PW (1998) *Mycosphaerella spp. and their anamorphs associated with leaf spot diseases of Eucalyptus. Mycologia Memoir* **21**, 1-170.

Crous PW and Wingfield MJ (1996) Species of *Mycosphaerella* and their anamorphs associated with leaf blotch disease of *Eucalyptus* in South Africa. *Mycologia*, 441-458.

Crous PW, Groenewald JZ, Mansilla JP, Hunter GC and Wingfield MJ (2004a) Phylogenetic reassessment of *Mycosphaerella* spp. and their anamorphs occurring on *Eucalyptus*. *Studies in Mycology* **50**, 195-214.

Crous PW, Groenewald JZ, Pongpanich K, Himaman W, Arzanlou M and Wingfield MJ (2004b) Cryptic speciation and host specificity among *Mycosphaerella* spp. occurring on Australian *Acacia* species grown as exotics in the tropics. *Studies in Mycology* **50**, 457-469.

Crous PW, Groenewald JZ, Risède JM, Simoneau P and Hyde KD (2006a) *Calonectria* species and their *Cylindrocladium* anamorphs: species with clavate vesicles. *Studies in Mycology* **55**, 213-226.

Crous PW, Groenewald JZ, Risède JM, Simoneau P and Hywel-Jones NL (2004c) *Calonectria* species and their *Cylindrocladium* anamorphs: species with sphaeropedunculate vesicles. *Studies in Mycology* **50**, 415-430.

Crous PW, Kang JC and Braun U (2001) A phylogenetic redefinition of anamorph genera in *Mycosphaerella* based on ITS rDNA sequence and morphology. *Mycologia* **93**, 1081-1101.

Crous PW, Mohammed C, Glen M, Verkley GJM and Groenewald JZ (2007) *Eucalyptus* microfungi known from culture. 3. *Eucasphaeria* and *Sympoventuria* genera nova, and new species of *Furcaspora, Harknessia,Heteroconium* and *Phacidiella*. *Fungal Diversity* **25**, 19-36.

Crous PW, Rong IH, Wood A, Lee S, Glen H, Botha W, Slippers B, Beer WZ, Wingfield MJ and Hawksworth DL (2006b) How many species of fungi are there at the tip of Africa? *Studies in Mycology* **55**, 13-33.

Crous PW, Wingfield MJ, Mansilla JP, Alfenas AC and Groenewald JZ (2006c) Phylogenetic reassessment of *Mycosphaerella* spp. and their anamorphs occurring on *Eucalyptus*. II. *Studies in Mycology* **55**, 99-131.

Crous, P.W., Slippers, B., Wingfield, M.J., Rheeder, J., Marasas, W.F.O., Philips, A.J.L., Alves, A., Burgess, T., Barber, P., Groenewald, J.Z., (2006d) Phylogenetic lineages in the Botryosphaeriaceae. *Studies in Mycology* **55**, 235-253.

Crous, P.W., Lennox, C.L., Sutton, B.C., (1995) *Selenophoma eucalypti* and *Stigmina robbenensis* spp. nov. from *Eucalyptus* leaves on Robben Island. *Mycological Research* **99**, 648-652.

Dahlberg A (2001) Community ecology of ectomycorrhizal fungi: an advancing interdisciplinary field. *New Phytologist* **150**, 555-562.

Dasmahapatra KK and Mallet J (2006) Taxonomy: DNA barcodes: recent successes and future prospects. *Heredity* **97**, 254-255.

De Beer ZW, Begerow D, Bauer R, Pegg GS, Crous PW and Wingfield MJ (2006) Phylogeny of the Quambalariaceae fam. nov., including important *Eucalyptus* pathogens in South Africa and Australia. *Studies in Mycology* **55**, 289-298.

Delseny M, Han B and Hsing YI (2010) High throughput DNA sequencing: The new sequencing revolution. *Plant Science* **179**, 407-422.

Fisher PJ, Petrini O and Sutton BC (1993) A comparative study of fungal endophytes in leaves, xylem and bark of *Eucalyptus nitens* in Australia and England. *Sydowia* **45**, 1-14.

Frézal L and Leblois R (2008) Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution* **8**, 727-736.

Fruhlich J, Hyde KD and Petrini O (2000) Endophytic fungi associated with palms. *Mycological Research* **104**, 1202-1212.

Gamboa MA and Bayman P (2001) Communities of Endophytic Fungi in Leaves of a Tropical Timber Tree (*Guarea guidonia*: Meliaceae) 1. *Biotropica* **33**, 352-360.

Ganley RJ, Brunsfeld SJ and Newcombe G (2004) A community of unknown, endophytic fungi in western white pine. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10107-10112.

Girmenia C, Pizzarelli G, Cristini F, Barchiesi F, Spreghini E, Scalise G and Martino P (2006) *Candida guilliermondii fungemia* in patients with hematologic malignancies. *Journal of Clinical Microbiology* **44**, 2458-2464.

Gryzenhout M, Myburg H, Van der Merwe NA, Wingfield BD and Wingfield MJ (2004) *Chrysoporthe*, a new genus to accommodate *Cryphonectria cubensis*. *Studies in Mycology* **50**, 119-142.

Guo LD, Hyde KD and Liew ECY (2000) Identification of endophytic fungi from *Livistona chinensis* based on morphology and rDNA sequences. *New Phytologist* **147**, 617-630.

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB and Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* **6**, 959-964.

Hajibabaei M, dewaard JR, Ivanova NV, Ratnasingham S, Dooh RT, Kirk SL, Mackie PM and Hebert PDN (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B:* **360**, 1959-1967

Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ and Levy S (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology* **10**, 32-45.

Harrington TC and Rizzo DM (1999) *Defining species in the fungi.* In: Worrall J.J, (eds,) Structure and Dynamics of Fungal Populations. Kluwer Academic Press, Netherlands, 43-72.

Hausmann A, Haszprunar G and Hebert PDN (2010) DNA barcoding the Geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. *PloS One* **6**, e17134.

Hebert PDN and Gregory TR (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology* **54**, 852-859.

Hebert PDN, Cywinska A, Ball SL and DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, 313-321.

Hebert PDN, Penton EH, Burns JM, Janzen DH and Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences* **101**, 14812-14817.

Hebert PDN, Stoeckle MY, Zemlak TS and Francis CM (2004b) Identification of birds through DNA barcodes. *PLoS Biology* **2**, 1657-1663.

Hershkovitz MA and Lewis LA (1996) Deep-level diagnostic value of the rDNA-ITS region. *Molecular Biology and Evolution* **13**, 1276-1295.

Hert DG, Fredlake CP and Barron AE (2008) Advantages and limitations of next-generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* **29**, 4618-4626.

Hillis DM and Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology* **66**, 411-453.

Hoffman MT and Arnold AE (2008) Geographic locality and host identity shape fungal endophyte communities in cupressaceous trees. *Mycological Research* **112**, 331-344.

Hollingsworth PM, Graham SW and Little DP (2011) Choosing and Using a Plant DNA Barcode. *PloS One* **6**, e19254.

Hoog GS and Ende A (1998) Molecular diagnostics of clinical strains of filamentous Basidiomycetes. *Mycoses* **41**, 183-189.

Huang Y, Niu B, Gao Y, Fu L and Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680-682.

Hudson HJ (1963) The perfect state of *Nigrospora oryzae. Transactions of the British Mycological Society* **46**, 355-360.

Hunter GC, Wingfield BD, Crous PW and Wingfield MJ (2006) A multi-gene phylogeny for species of *Mycosphaerella* occurring on *Eucalyptus* leaves. *Studies in Mycology* **55**, 147-161.

Huson DH, Auch AF, Qi J and Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Research* **17**, 377-386.

Hwang HS (2004) Gray leaf spot in peppers caused by *Stemphylium solani* and *S. lycopersici*. *The Plant Pathology Journal* **20**, 85-91.

Hyde KD, Bussaban B, Paulus B, Crous PW, Lee S, McKenzie EHC, Photita W and Lumyong S (2007) Diversity of saprobic microfungi. *Biodiversity and Conservation* **16**, 7-35.

Hyde KD, Soytong K (2008) The fungal endophyte dilemma. *Fungal Diversity* **33**, 163-173.

Jarvie T (2005) Next generation sequencing technologies. *Drug Discovery Today: Technologies* **2**, 255-260.

Kanagawa T (2003) Bias and artefacts in multi template polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering* **96**, 317-323.

Karow J (2008) International Barcode of Life Project Sees Room for Sanger, *New Sequencing Techs*. [online] Available: www.genomeweb.com/sequencing/international-barcode-life-project-sees-room-sanger-new-sequencing-techs-0 [Acessed 3 march 2009].

Kennedy N and Clipson N (2003) Fingerprinting the fungal community. *Mycologist* **1 7**, 158-164.

Koljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R and Larsson E (2005) UNITE: a database providing web-based

methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist* **166**, 1063-1068.

Korf I, Yandell M and Bedell J (2003) BLAST: An Essential Guide to the Basic Local Aignment Search Tool. In: LeJune L, (eds.) *Blast.* O'Reilly & Associates, Inc. Sebastopol, California, USA.

Lee S, Groenewald JZ and Crous PW (2004) Phylogenetic reassessment of the coelomycete genus *Harknessia* and its teleomorph Wuestneia (Diaporthales), and the introduction of *Apoharknessia* gen. nov. *Studies in Mycology* **50**, 235-252.

Li W and Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659.

Lodge DJ, Fisher PJ and Sutton BC (1996) Endophytic fungi of *Manilkara bidentata* leaves in Puerto Rico. *Mycologia* **88**, 733-738.

Maharachchikumbura SSN, Guo L, Chukeatirote E, Bahkali A, Hyde K (2011) Pestalotiopsis-morphology, phylogeny, biochemistry and divesity. *Fungal Diversity* **50**, 167-187

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-141.

Mehta YR (1998) Severe outbreak of *Stemphylium* leaf blight, a new disease of cotton in Brazil. *Plant Disease* **82**, 333-336.

Mehta YR and Brogin RL (2000) Phytotoxicity of a culture filtrate produced by *Stemphylium solani* of cotton. *Plant Disease* **84**, 838-842.

Metzker ML (2005) Emerging technologies in DNA sequencing. *Genome Research* **15**, 1767-1776.

Meusnier I, Singer G, Landry JF, Hickey D, Hebert P and Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**, 214-217.

Moller EM, Bahnweg G, Sanderman H, Geiger HH (1992) A simple and efficient protocol for isolation of high molecular weight filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic acids research* **20**, 6115-6116.

Moritz C and Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biology* **2**, e354.

Nilsson R, Tedersoo L, Lindahl BD, Kjoller R, Carlsen T, Quince C, Abarenkov K, Pennanen T, Stenlid J and Bruns T (2011) Towards standardization of the description and publication of next generation sequencing datasets of fungal communities. *New Phytologist* **191**, 314-318.

Nilsson RH, Abarenkov K, Veldre V, Nylinder S, De Wit P, Brosche S, Alfredsson JF, Ryberg M and Kristiansson E (2010) An open source chimera checker for the fungal ITS region. *Molecular Ecology Resources* **10**, 1076-1081.

Nilsson RH, Bok G, Ryberg M, Kristiansson E and Hallenberg N (2009a) A software pipeline for processing and identification of fungal ITS sequences. *Source Code for Biology and Medicine* **4**, 1-6.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N and Larsson KH (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics Online* **4**, 193-201.

Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E and Kristiansson E (2009b) The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiology Letters* **296**, 97-101.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH and Kõljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS One* **1,** e59.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM and Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology* **71**, 5544-5550.

O'Donnell K (1992) Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*). *Current Genetics* **22**, 213-220.

Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara RG, Simpson GL, Solymos P, Stevens MHH and Wagner H (2010) *vegan: Community Ecology Package. R package*. [online] Available at: http://CRAN.R-project.org/package=vegan

Parameswaran P, Jalili R, Tao L, Shokralla S, Gharizadeh B, Ronaghi M and Fire AZ (2007) A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. *Nucleic Acids Research* **35**, e130.

Pavlic D, Slippers B, Coutinho TA and Wingfield MJ (2008) Multiple gene genealogies and phenotypic data reveal cryptic species of the Botryosphaeriaceae: A case study on the *Neofusicoccum parvum/N. ribis* complex. *Molecular Phylogenetics and Evolution* **51**, 259-268.

Peay KG, Kennedy PG and Bruns TD (2008) Fungal community ecology: a hybrid beast with a molecular master. *BioScience* **58**, 799-810.

Petrosino JF, Highlander S, Luna RA, Gibbs RA and Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry* **55**, 856-866.

Pfaller MA, Diekema DJ, Mendez M, Kibbler C, Erzsebet P, Chang SC, Gibbs DL and Newell VA (2006) *Candida guilliermondii,* an opportunistic fungal pathogen with

decreased susceptibility to fluconazole: geographic and temporal trends from the ARTEMIS DISK antifungal surveillance program. *Journal of Clinical Microbiology* **44**, 3551-3556.

Photita W, Lumyong S, Lumyong P and Hyde KD (2001) Endophytic fungi of wild banana (*Musa acuminata*) at Doi Suthep Pui National Park, Thailand. *Mycological Research* **105**, 1508-1513.

Prentice IC (1977) Non-metric ordination methods in ecology. *The Journal of Ecology* **65**, 85-94.

Ratnasingham S and Hebert PDN (2007) bold: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Notes* **7**, 355-364.

Reis-Filho J (2009) Next-generation sequencing. *Breast Cancer Research* **11**, S12-19.

Rihs JD, Padhye AA and Good CB (1996) Brain abscess caused by *Schizophyllum commune*: an emerging basidiomycete pathogen. *Journal of Clinical Microbiology* **34**, 1628-1632.

Rodriguez RJ, White Jr JF, Arnold AE and Redman RS (2009) Fungal endophytes: diversity and functional roles. *New Phytologist* **182**, 314-330.

Rosselló-Mora R and Amann R (2001) The species concept for prokaryotes. *FEMS Microbiology Reviews* **25**, 39-67.

Roux C (1986) *Leptosphaerulina chartarum* sp. nov., the teleomorph of *Pithomyces chartarum*. *Transactions of the British Mycological Society* **86**, 319-323.

Saikkonen K (2007) Forest structure and fungal endophytes. *Fungal Biology Review s***21**, 67-74.

Saikkonen K, Faeth SH, Helander M and Sullivan TJ (1998) Fungal endophytes: a continuum of interactions with host plants. *Annual Review of Ecology and Systematics* **29**, 319-343.

Saikkonen K, Wäli P, Helander M and Faeth SH (2004) Evolution of endophyteï plant symbioses. *Trends in Plant Science* **9**, 275-280.

Sakalidis M (2004) *Resolving the Botryosphaeria ribis-B. parva species complex; a molecular and phenotypic investigation*. Honours thesis, Murdoch University, Australia.

Sankaran K, Sutton B and Minter D (1995) A checklist of fungi recorded on *Eucalyptus*. *Mycological papers* **170**, 1-376.

Savolainen V, Cowan RS, Vogler AP, Roderick GK and Lane R (2005) Towards writing the Encyclopaedia of Life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B:***360**, 1805-1811.

Schoch C, Seifert KA and Crous PW (2011) Progress on DNA Barcoding of Fungi *IMA fungus* **2**, 83-89.

Schoch C, Seifert KA, Huhndorf S, Robert V, Spouge J, Levesque A, Chen W Fungal working group (2012) Barcoding Consortium Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* **10**, 1073 -1079

Schonau APG (1991) Role of eucalypt plantations in timber supply and forest conservation in sub-Saharan Africa. *South African Forestry Journal* **156**, 56-60.

Schulz B and Boyle C (2006) What are Endophytes? *Microbial Root Endophytes* **9**, 1-13.

Seifert KA (2009) Progress towards DNA barcoding of fungi. *Molecular Ecology Resources* **9**, 83-89.

Seifert KA, Samson RA, deWaard JR, Houbraken J, Lévesque CA, Moncalvo JM, Louis-Seize G and Hebert PDN (2007) Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Sciences* **104**, 3901-3906.

Slippers B, Crous PW, Denman S, Coutinho TA, Wingfield BD and Wingfield MJ (2004a) Combined multiple gene genealogies and phenotypic characters differentiate several species previously identified as *Botryosphaeria dothidea. Mycologia* **96**, 83-101.

Slippers B, Fourie G, Crous PW, Coutinho TA, Wingfield BD, Carnegie AJ and Wingfield MJ (2004b) Speciation and distribution of *Botryosphaeria* spp. on native and introduced *Eucalyptus* trees in Australia and South Africa. *Studies in Mycology* **50**, 343-358.

Slippers B, Fourie G, Crous PW, Coutinho TA, Wingfield BD and Wingfield MJ (2004c) Multiple gene sequences delimit *Botryosphaeria australis* sp. nov. from *B. lutea*. *Mycologia* **96**, 1030-1041.

Slippers B and Wingfield MJ (2007) Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biology Reviews* **21**, 90-106.

Stockinger H, Krüger M and Schübler A (2010) DNA barcoding of arbuscular mycorrhizal fungi. *New Phytologist* **187**, 461-474.

Stone JK, Bacon CW and White JF (2000) An overview of endophytic microbes: endophytism defined. *Microbial endophytes* **3**, 29-33.

Summerbell BA, Groenewald JZ, Carnegie AJ, Summerbell RC and Crous PW (2006) *Eucalyptus* microfungi known from culture. 2. *Alysidiella*, *Fusculina* and *Phlogicylindrium* genera nova, with notes on some other poorly known taxa. *Fungal Diversity* **23**, 323-350.

Swart L, Crous PW, Petrini O and Taylor JE (2000) Fungal endophytes of proteaceae, with particular emphasis on *Botryosphaeria proteae*. *Mycoscience* **41**, 123-127.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28**, 2731-2739.

Tautz D, Arctander P, Minelli A, Thomas RH and Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution* **18**, 70-74.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G and Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* **188**, 291-301.

Turnbull JW (1999) Eucalypt plantations. *New Forests* **17**, 37-52.

Turnbull JW (2000) Diseases and Pathogens of *Eucalyptus*. In: Keane PJ, Podger FD, Brown BN (eds.) *Economic and social importance of eucalypts.* CSIRO, Australia, pp 1-7.

Toth B, Csosz M, Dijksterhuis J, Frisvad JC and Varga J (2007) Short communication *Pithomyces Chartarum* as a pathogen of wheat. *Journal of Plant Pathology* **89**, 405-408.

Van Niekerk JM and Groenewald JZ (2004) Systematic reappraisal of *Coniella* and *Pilidiella,* with specific reference to species occurring on *Eucalyptus* and *Vitis* in South Africa. *Mycological Research* **108**, 283-303.

Vilgalys R (2003) Taxonomic misidentification in public DNA databases. *New Phytologist* **160**, 4-5.

Vilgalys R and Hester M (1990) Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology* **172**, 4238-4246.

Voelkerding KV, Dames SA and Durtschi JD (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry* **55**, 641-658.

Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* **360**, 1847-1857.

Warnecke F and Hess M (2009) A perspective: metatranscriptomics as a tool for the discovery of novel biocatalysts. *Journal of Biotechnology* **142**, 91-95.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White

TJ, editors. (eds). *PCR Protocols: a guide to methods and applications*. Academic Press, USA *315–322.*

Wilson D (1995) Endophyte: the evolution of a term, and clarification of its use and definition. *Oikos* **73**, 274-276.

Wommack KE, Bhavsar J and Ravel J (2008) Metagenomics: read length matters. *Applied and Environmental Microbiology* **74**, 1453-1463.

Yamamura M, Makimura K, Fujisaki R, Satoh K, Kawakami S, Nishiya H and Ota Y (2009) Polymerase chain reaction assay for specific identification of *Candida guilliermondi*i (*Pichia guilliermondii*). *Journal of Infection and Chemotherapy* **15**, 214-218.

Zhang JB and Hanner R (2011) DNA barcoding is a useful tool for the identification of marine fishes from Japan. *Biochemical Systematics and Ecology* **39**, 31-42.

Zheng L, Huang J and Hsiang T (2008) First report of leaf blight of garlic (*Allium sativum*) caused by *Stemphylium solani* in China. *Plant Pathology* **57**, 380-380.

Table 1. List of primers for nested PCR with added 454-pyrosequencing primers and multiplex tags

Tree no-tissue	**Pyrosequencing A adapter**	**multiplex tag**	**ITS1F/ITS5**
Tree1- leaves	5' - CGTATCGCCTCCCTCGCGCCATCAGACGAGTGCGT*ggaagtaaaagtcgtaacaagg* - 3'		
Tree-1 petioles	5ô CGTATCGCCTCCCTCGCGCCATCAGACGCTCGACA*ggaagtaaaagtcgtaacaagg* - 3'		
Tree1- twigs	5ô CGTATCGCCTCCCTCGCGCCATCAGAGACGCACTC*ggaagtaaaagtcgtaacaagg* - 3'		
Tree1- increments	5ć- CGTATCGCCTCCCTCGCGCCATCAGCTCGCGTGTC*ggaagtaaaagtcgtaacaagg* - 3'		
Tree2- leaves	5ô CGTATCGCCTCCCTCGCGCCATCAGATCAGACACG*cttggtcatttagaggaagtaa*- 3'		
Tree2- petioles	5ô CGTATCGCCTCCCTCGCGCCATCAGATATCGCGAG*cttggtcatttagaggaagtaa* - 3'		
Tree2- twigs	5ô CGTATCGCCTCCCTCGCGCCATCAGCGTGTCTCTA*cttggtcatttagaggaagtaa*- 3'		
Tree2- increments	**N/A**		
Tree3- leaves	5ô CGTATCGCCTCCCTCGCGCCATCAGTAGTATCAGC*cttggtcatttagaggaagtaa* - 3'		
Tree3- petioles	5ô CGTATCGCCTCCCTCGCGCCATCAGTCTCTATGCGcttggtcatttagaggaagtaa- 3'		
Tree3- twigs	5ć- CGTATCGCCTCCCTCGCGCCATCAGTGATACGTCT*cttggtcatttagaggaagtaa* -3'		
Tree3- increments	5ô CGTATCGCCTCCCTCGCGCCATCAGTCTCTATGCG ggaagtaaaagtcgtaacaagg - 3'		
Reverse : adaptor B + reverse primer (ITS4)	5ôCTATGCGCCTTGCCAGCCCGCTCAGtcctccgcttattgatatgc 3ô		

Table 2. Results of 454-pyrosequencing data obtained from three *Eucalyptus grandis* trees growing in the same geographical location used in this study

	Tree1 leaves	Tree1 petioles	Tree1 twigs	Tree1 increments	Tree2 leaves	Tree2 petioles	Tree2 twigs	Tree3 leaves	Tree3 petioles	Tree3 twigs	Tree3 increments
Total reads	602	921	2324	173	1657	2138	774	8396	6863	2779	1251
Potential chimera	3	17	17	0	3	11	1	52	41	9	30
> 100bp	396	464	1045	129	1324	1465	680	5903	4124	2492	508
Reads for analysis	203	440	1262	44	330	662	93	2441	2698	1004	713
98% clustering	112	208	503	32	96	144	54	434	689	278	263

Summary	Total
Total reads	27 878
Potential chimera	184
> 100bp	18530
Reads for analysis	9890
98% clustering	2813

Table 3. Fungal families identified from the 454-pyrosequencing data of the three *Eucalyptus grandis* trees

Family	% reads
Bionectriaceae	0.2
Botryosphaeriaceae	0.63
Calosphaeriaceae	0.04
Cururbitariaceae	0.03
Debarymyataceae	0.34
Diaporthaceae	0.07
Dothideaceae	0.04
Dothioroaceae	0.83
Glommeraceae	0.02
Hypocreaceae	5
Lecanorales	0.1
Lophiostomataceae	0.02
Mitosporic ascomycota	0.35
Mitosporic Amphisphaeceae	0.53
Mitosporic Davidiellaceae	0.66
Mitosporic Pleosporaceae	3.03
Mitosporic Trichocomaceae	0.1
Mycosphaerellaceae	12.3
Nectriaceae	0.01
Phaeosphaeriaceae	0.03
Phanerochaetaceae	0.07
Pleosporaceae	0.08
Saccharomycetales	0.01
Schizophyllaceae	0.81
Sporomiaceae	0.1
Teratosphaeriaceae	15
Trichoriaceae	0.4
Xylariaceae	0.51

Table 4. Identities endophyte isolates obtained from a single *Eucalyptus grandis* tree where phylogroups were determined by sequence similarity based on Blast and phylogeny.

BLAST group	No of isolates	No of phylo groups
Annulohypoxylon sp.	10	3
Alternaria sp.	52	4
Bionectria sp.	13	1
Botryosphaeriaceae	58	5
Cladosporium spp.	14	4
Cochliobolus sp.	1	1
Colletotrichum sp.	3	3
Daldinia sp.	5	1
Didymellaceae	9	1
Guignardia sp.	6	1
Hypocrea sp.	13	5
Lecytophora spp.	1	1
Montagnulaceae sp.	3	2
Mycosphaerellaceae	37	3
Nectria sp.	1	1
Penicillium spp.	5	3
Pestalotiopsis/Bagadiella/Truncella spp.	110	6
Phaeoacremonium spp.	2	1
Phomopsis sp.	1	1
Sarcosomataceouseae	10	1
Sydowia eucalypti	42	2
Stemphylium solani	1	1
Teratosphaeriaceae	24	2
Tricholomataceae	1	1
Unknown1	2	1
Unknown2	1	1
Unknown3	2	1

Table 4: continued

Unknown4	6	1
Unknown5	4	1
Unknown6	2	1
Unknown7	1	1
Unknown8	1	1
Unknown9	1	1
Unknown10	1	1
Unknown11	5	1
Unknown12	1	1
Unknown13	1	1
Unknown14	1	2
Unknown15	3	1
Unknown16	3	1
Unknown17	2	1
Unknown18	2	1
Unknown19	2	1
Xylariaceae	5	4
Total	**476**	**85**

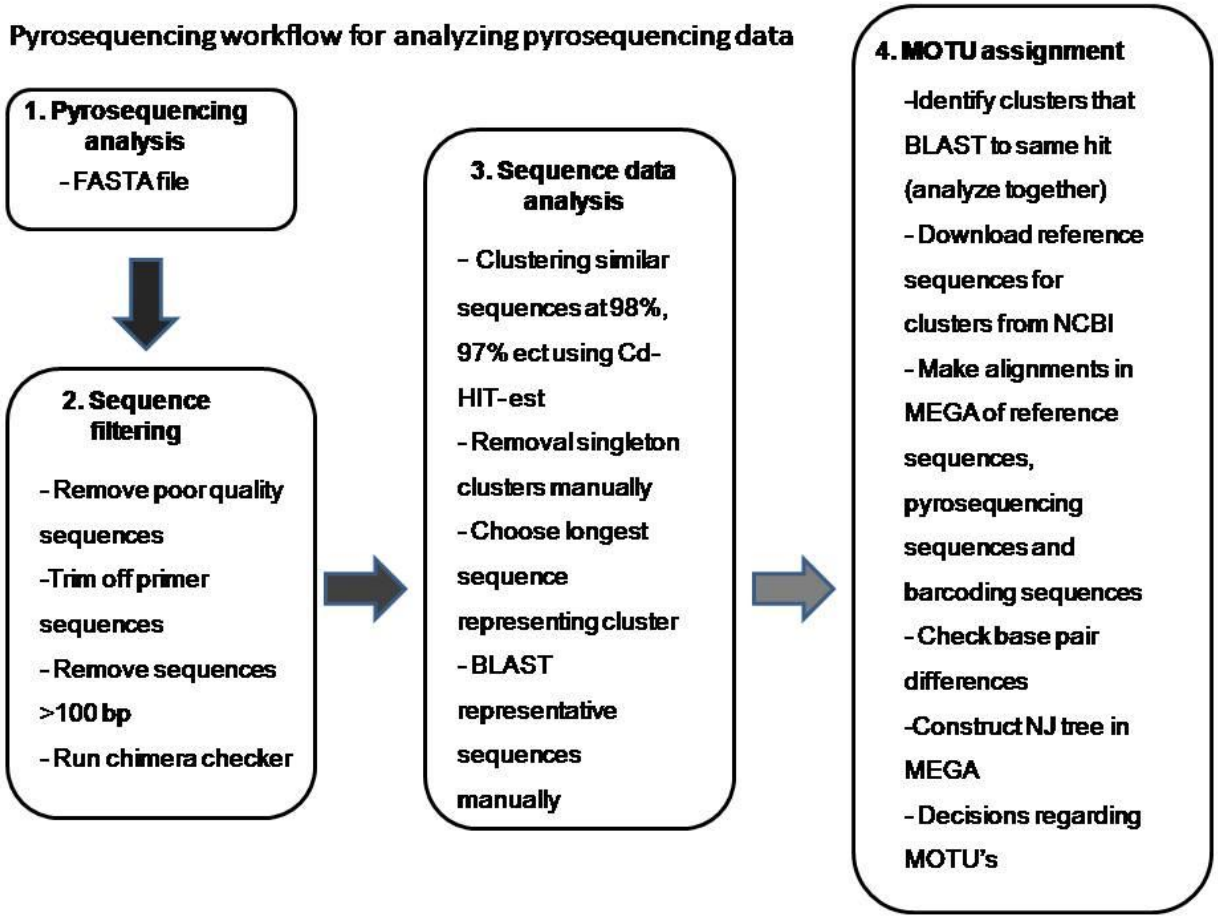


Fig. 1 Pyrosequencing workflow used to determine Molecular Operational Taxonomic Units (MOTU) identities.

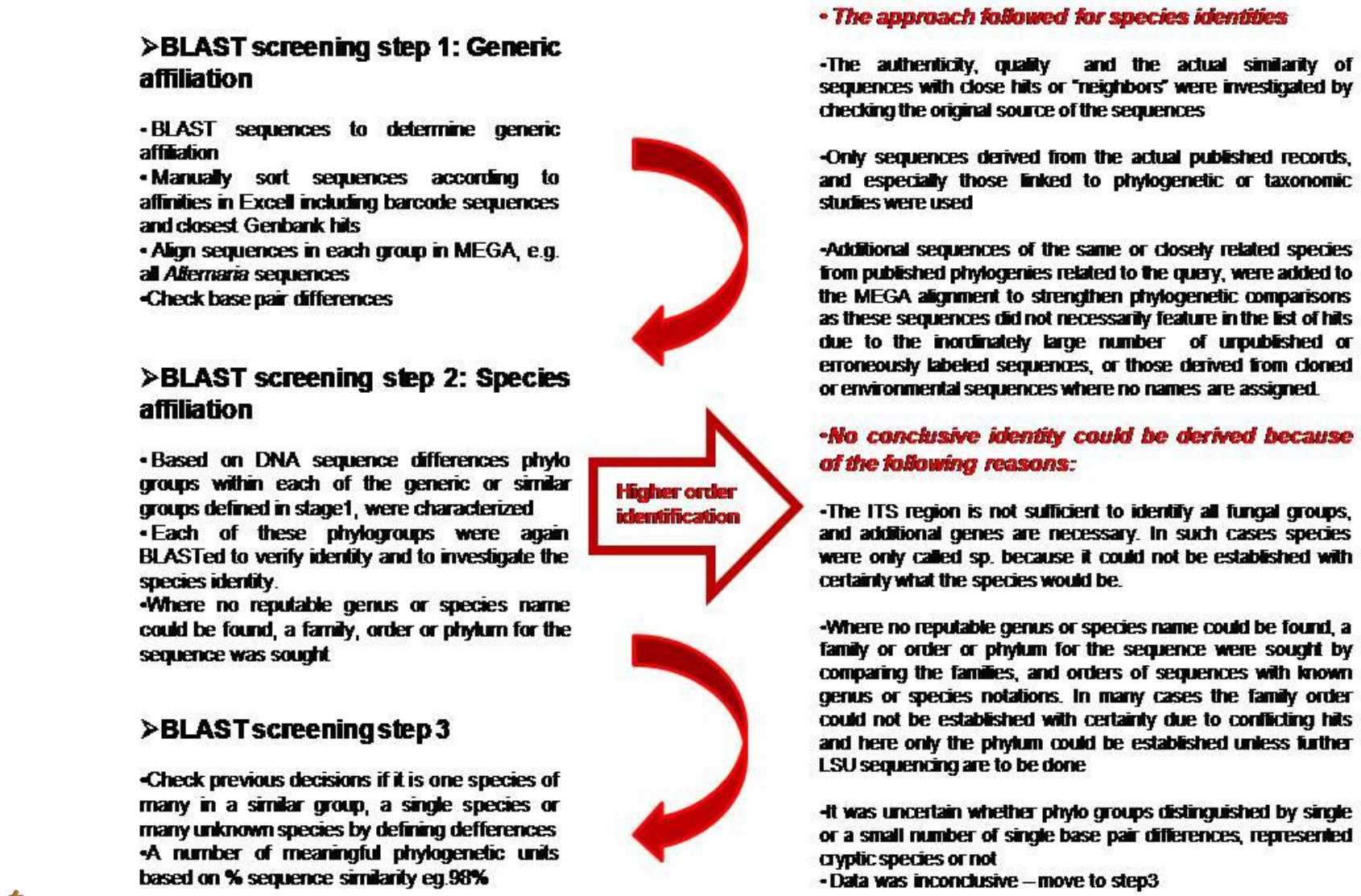


Fig. 2 Isolate bar-coding workflow used for identification based on full length ITS region.

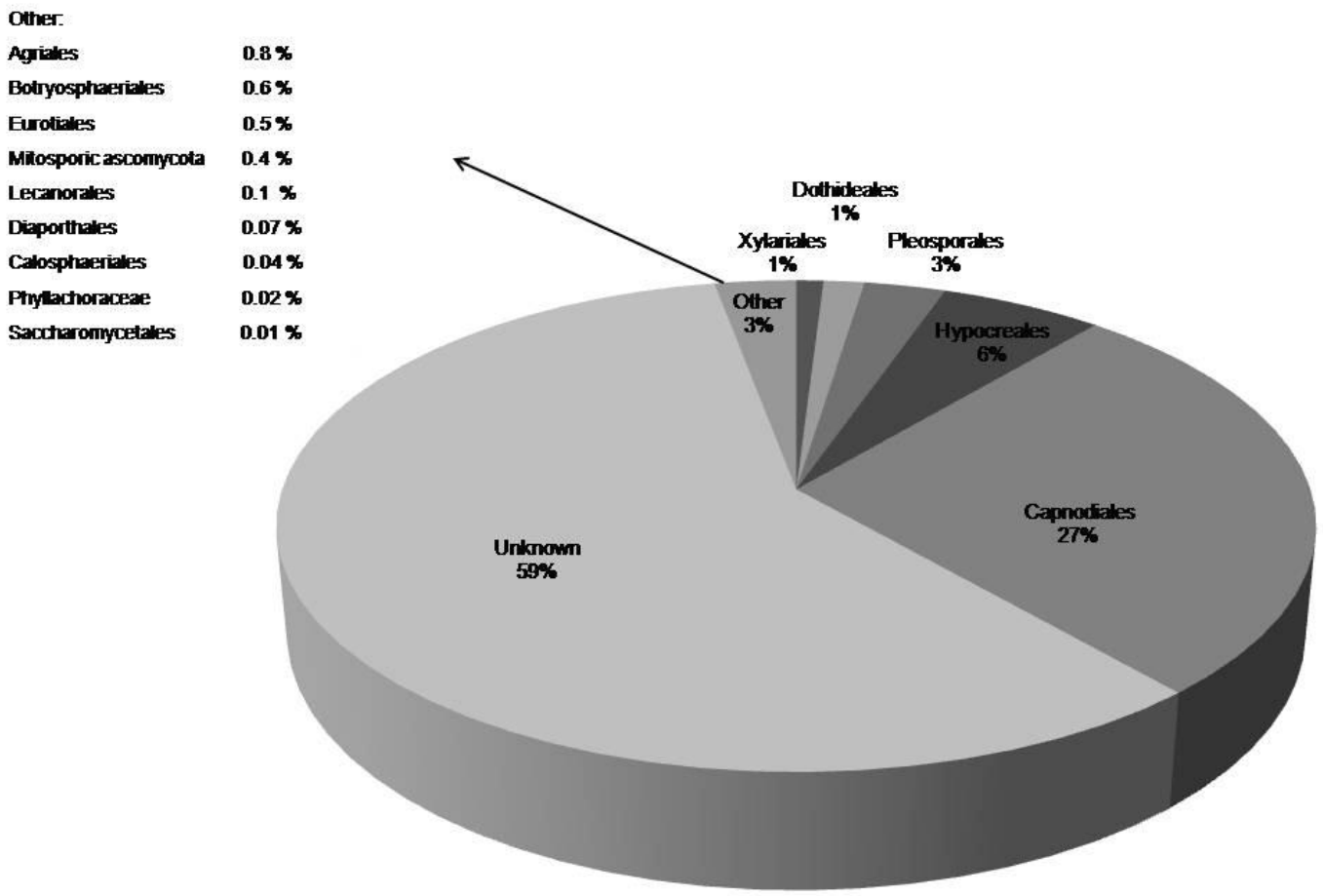


Fig. 3 Proportional distribution of the different fungal orders detected from the various tissues of three *Eucalyptus grandis* trees with 454-pyrosequencing of the ITS region.

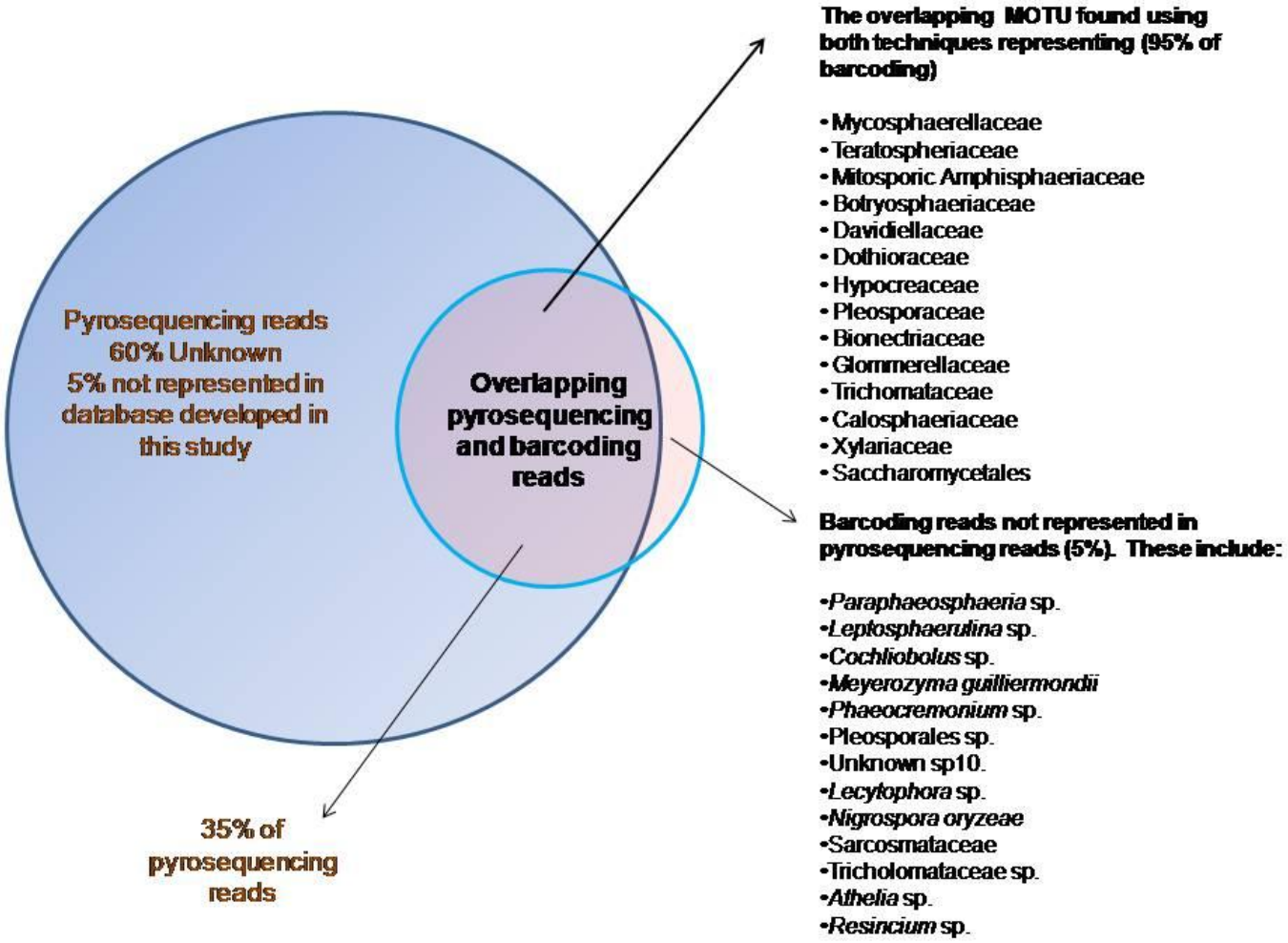


Fig. 4 Proportional distribution of the overlapping Molecular Operational Taxonomic Units (MOTUs) found using metagenetic 454-pyrosequening and isolate bar-coding. A total of 95 % of bar-coding reads were recovered in the 454-pyrosequencing reads, the other 5 % was only identified in the bar-coding study. The overlap accounted for 35 % of the 454-pyrosequencing reads. The remaining 65 % of the pyrosequencing was not represented in the bar-coding database developed in this study.

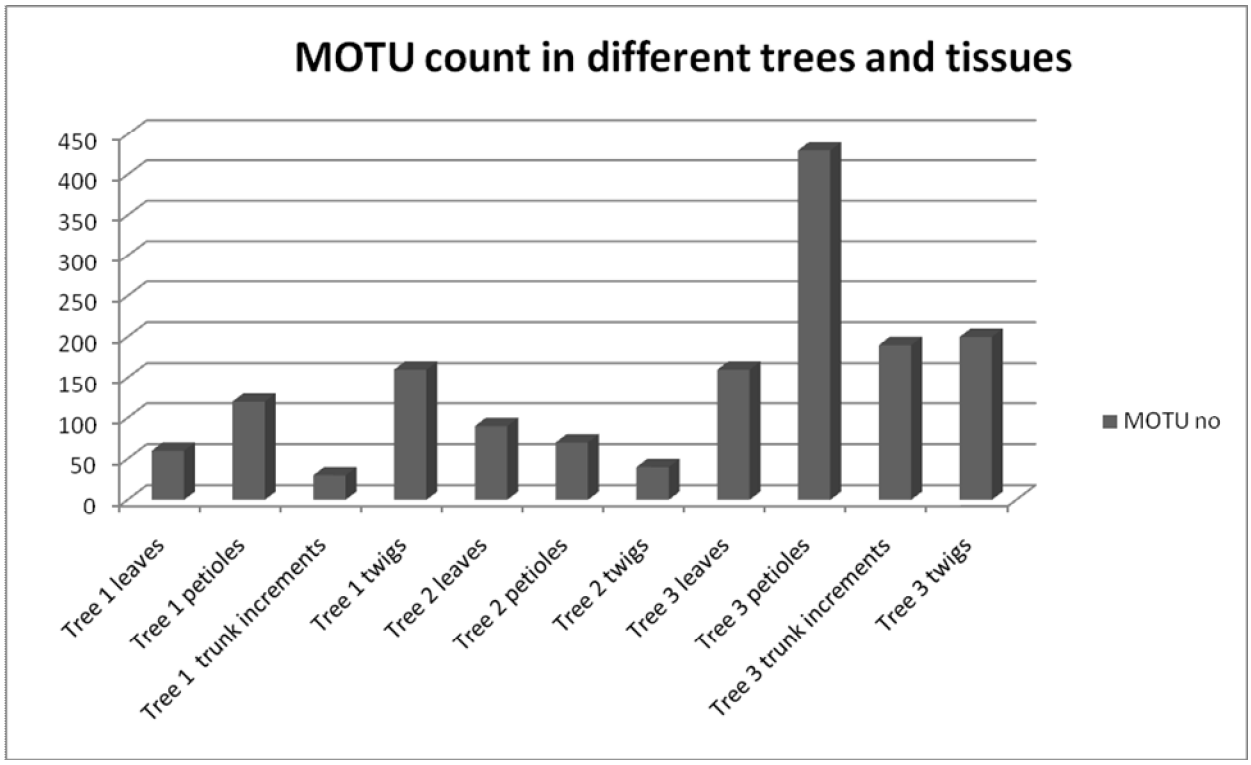


Fig. 5 Number of the MOTUôs (Molecular Operational Taxonomic Units) among the different tree tissues obtained from the 454-pyrosequencing run.

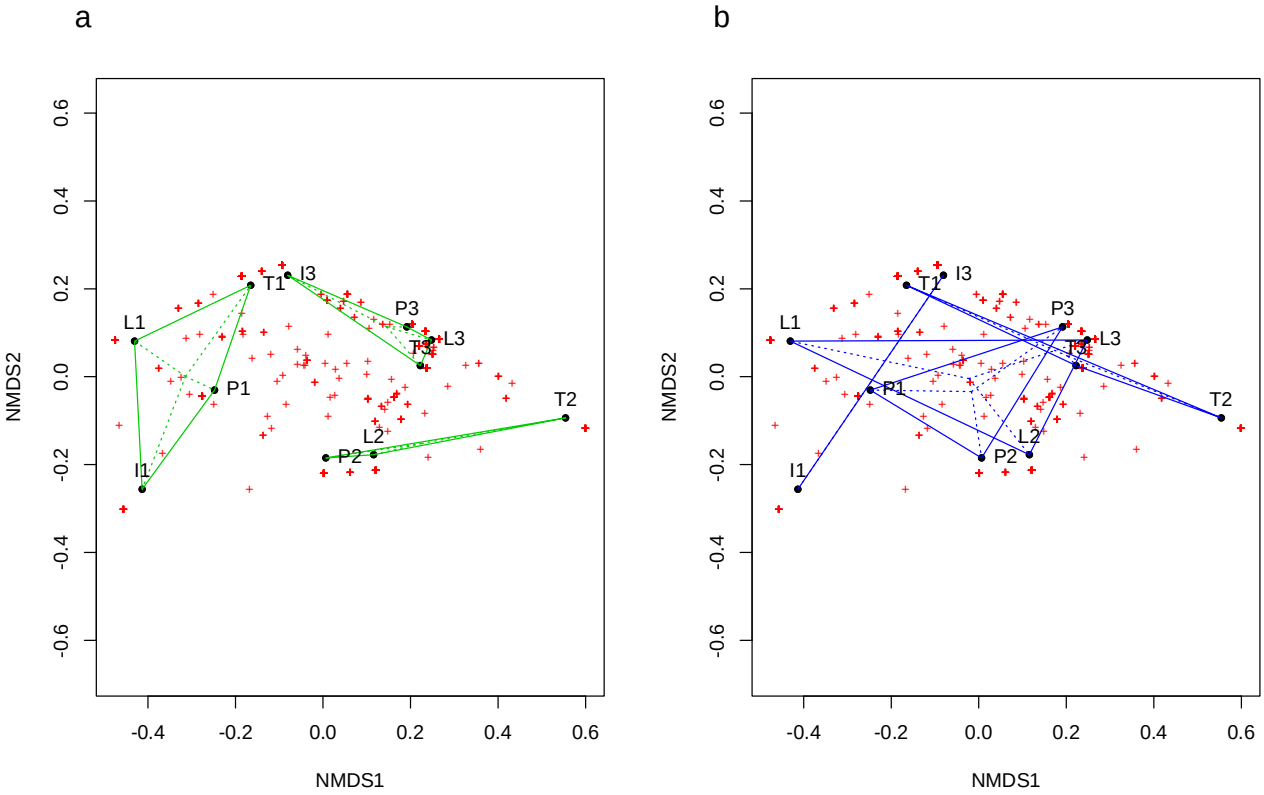


Fig. 6a Non-metric multi-dimensional scaling of the similarity between trees and tissue types. The red crosses indicate the MOTUôs in the study and the points represent the trees **(6a)** and tissue types **(6b)** (L1. L2. L3; leaves: P1, P2, P3; petioles: T1, T2, T3; twigs: I1 and I3; trunk

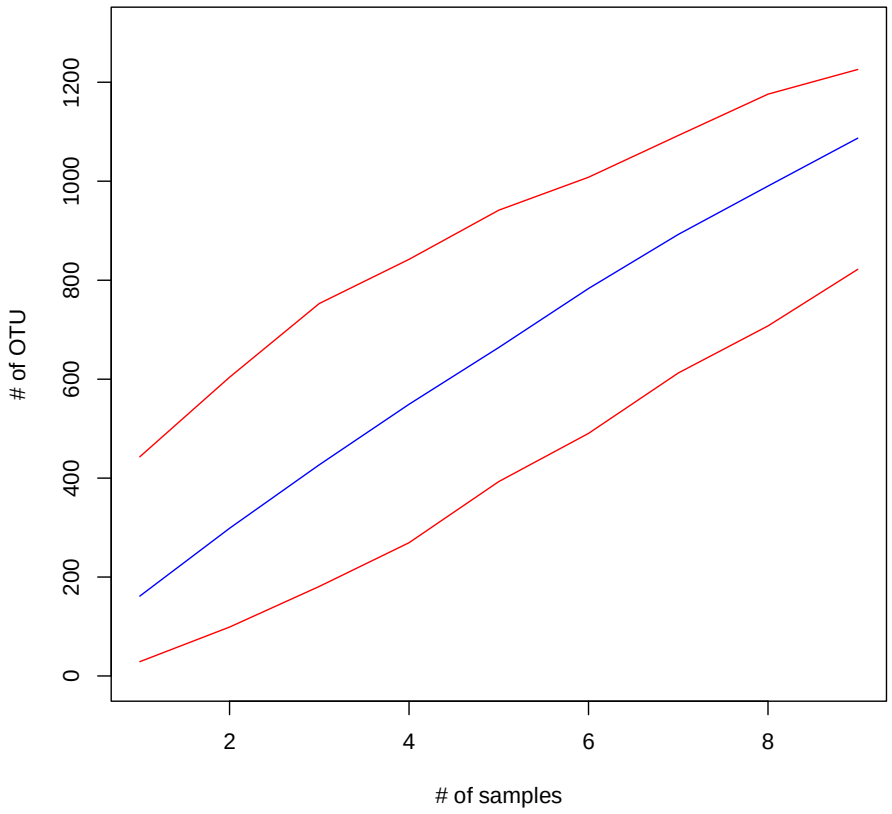


Fig. 7 Non-asymptotic species accumulation curve for fungal endophytes isolated from the above ground tissues of three *Eucalyptus grandis* trees growing in the same geographic locality depicting the number of MOTUs predicted from an increasing number of samples. The upper and lower curves indicate 95 % confidence intervals of the curve, while the blue curve shows the average MOTU at a given number of samples.

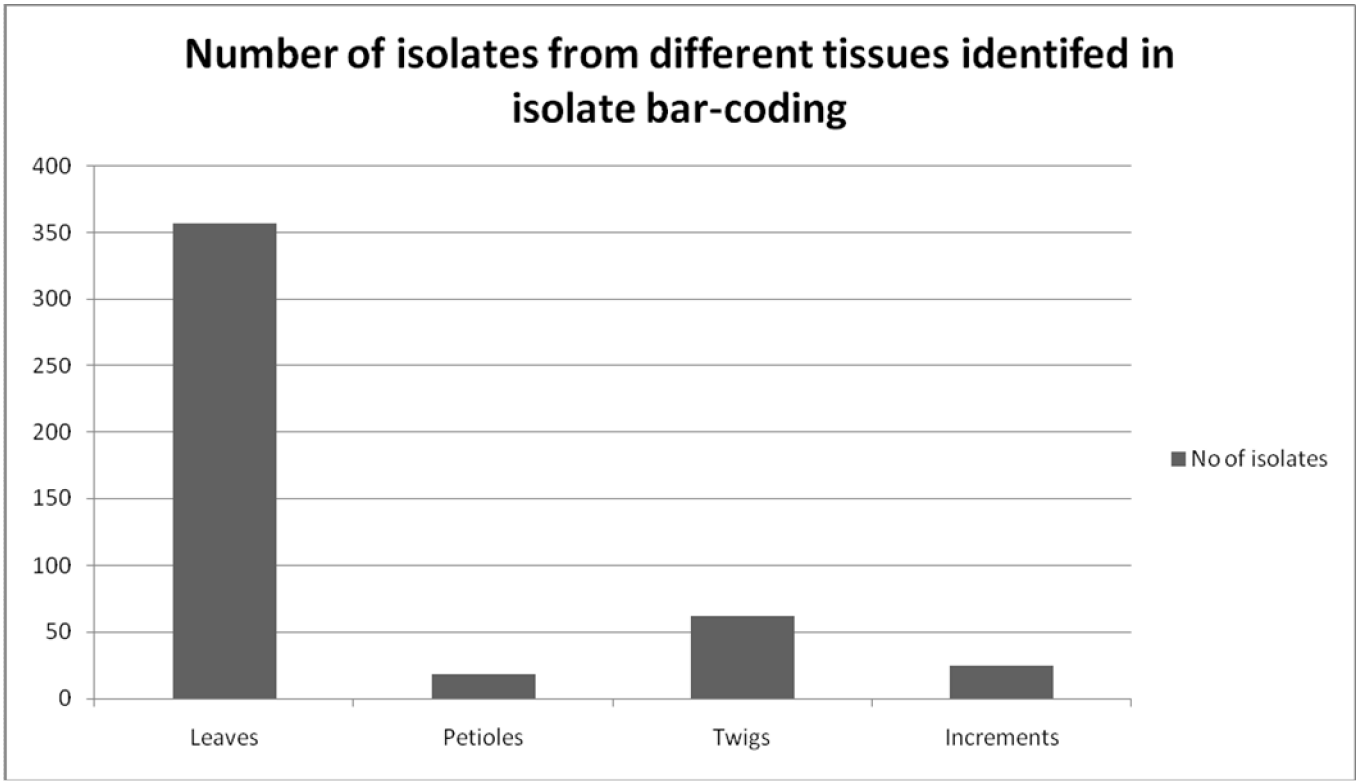


Fig. 8 Distribution of the number of endophytic isolates from different tissues obtained from the isolate bar-coding approach.

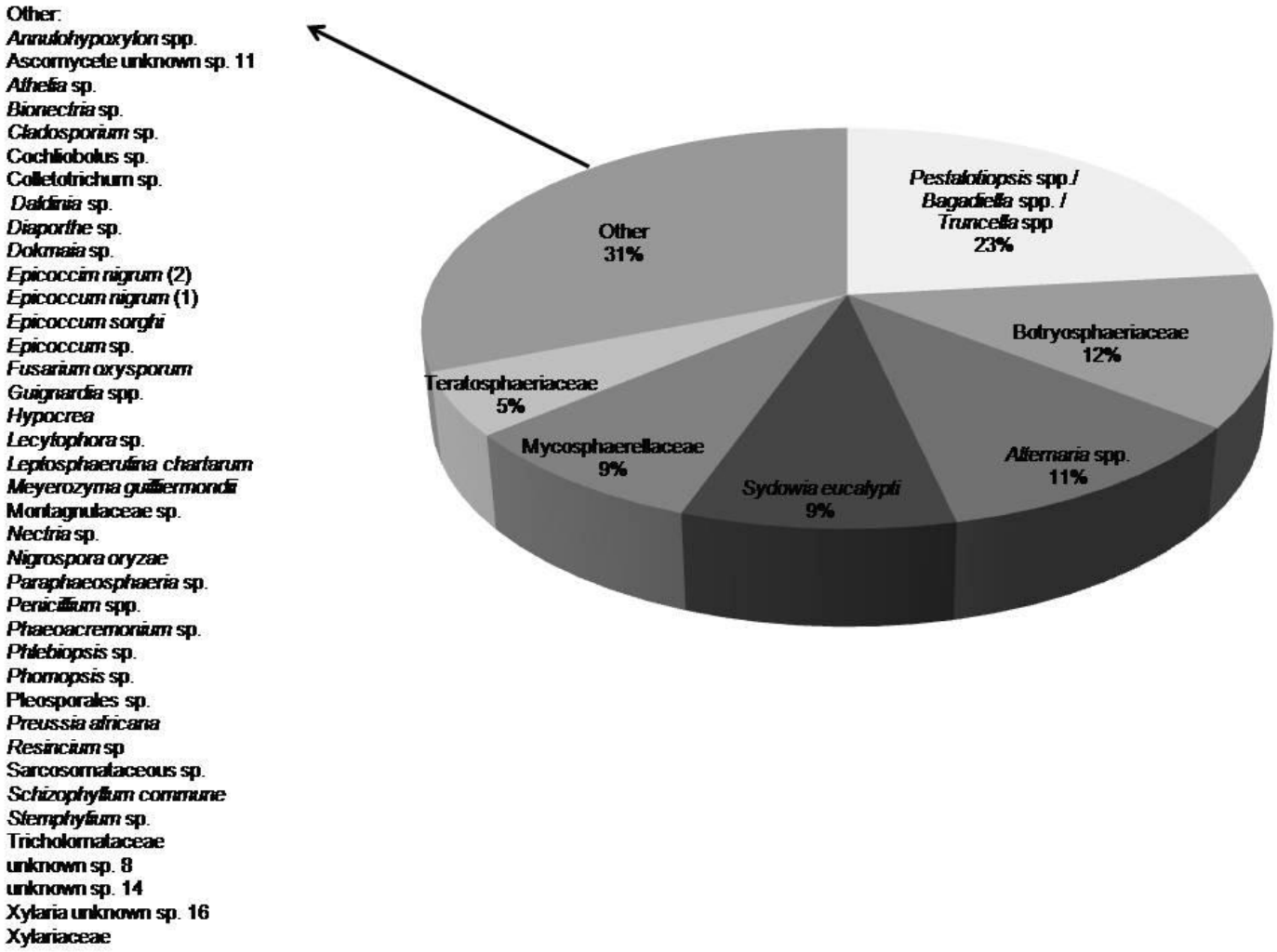


Fig. 9 Proportional distribution of the different fungal taxa based on isolate bar-coding of the ITS regions of a single *Eucalyptus grandis* tree.

Chapter 3:

Exploring the diversity and overlap of Botryosphaeriaceae from *Eucalyptus grandis* and *Syzygium cordatum* in South Africa

Chapter3:

Exploring the diversity and overlap of Botryosphaeriaceae from *Eucalyptus grandis* and *Syzygium cordatum* in South Africa

Abstract

Species in the fungal family Botryosphaeriaceae are latent pathogens on woody trees and have a wide host range, including native and introduced hosts. Multi-locus DNA sequence identification on a recent collection of Botryosphaeriaceae from *Eucalyptus grandis* and *Syzygium cordatum* trees in South Africa revealed cross-infectivity of several species, novel host associations and new country reports. *Neofusicoccum eucalyptorum*, *N. kwambonambiense, N. parvum, N. australe* and *Lasiodiplodia pseudotheobromae* were identified from both tree species, with L. *pseudotheobromae* and *N. eucalyptorum* isolated for the first time from *S. cordatum*, similar to *N. kwambonambiense* from *Eucalyptus*. This also represents the first report of *L. pseudotheobromae* from South Africa. Botryosphaeriaceae species on *Eucalyptus* species and *Syzygium cordatum* are fairly well known from South Africa. However, our study revealed new associations, indicating that intensive and continuous sampling is needed to fully comprehend complete host and country associations of these fungi. This will ensure the report of t new or segregated species that may have been detected before but under other species names.

Introduction

Botryosphaeriaceae species associated with *Eucalyptus* (Myrtales, Myrtaceae) trees in plantations have been well studied worldwide where they occur as endophytes and, in some cases, opportunistic latent pathogens (Burgess *et al*., 2005; Mohali *et al*., 2007; Slippers *et al*., 2009; Slippers *et al*., 2004b; Smith *et al*., 1994). Countries where Botryosphaeriaceae have been characterised on *Eucalyptus* include Congo (Roux *et al*., 2000), Uganda (Nakabonge. 2002), Chile (Ahumada, 2003), Australia (Slippers et al., 2004b), South Africa (Slippers *et a*l., 2009; Slippers *et al*., 2004a; Burgess *et al*., 2005; Burgess *et al*., 2006a;), Ethiopia (Gezahgne *et al*., 2004), Venezuela (Mohali *et al*., 2006), Colombia (Rodas *et al*., 2009), Uruguay (Pérez *et al*., 2008; Pérez *et al*., 2009) and China (Chen *et al*., 2011). The Botryosphaeriaceae that occur on these *Eucalyptus* trees in different parts of the world vary considerably. For example, in Venezuela the dominant Botryosphaeriaceae include *Botryosphaeria mamane, Neofusicoccum andium*, *N. parvum, N. Pseudofusicoccum N. stromaticum, Lasiodiplodia theobromae*, *L. crassispora* and *L. pseudotheobromae* (Mohali *et al*., 2006, 2007). This is different from species combinations present in western Australia that include *Fusicoccum ramsorum, N. parvum, N. australe, N. macroclavatum, P. adansoniae, P. ardesiarum, P. kimberleyense,* and *L. theobromae* (Burgess *et al*., 2005; Burgess *et al*., 2006a; Pavlic *et al*., 2008). In South Africa, *N. parvum*, *N. australe*, *N. eucalyptorum* and *N. eucalypticola* are dominant species (Slippers *et al*., 2009; Slippers *et al*., 2004b; Slippers *et al*., 2004c). These varying species compositions are indicative of a rich diversity of Botryosphaeriaceae on this host tree in different parts of the world.

Many Botryosphaeriaceae that occur on *Eucalyptus* also occur on other hosts. For example, *N. parvum* that is one of the dominant species on *Eucalyptus* in many parts of the world and also on *Populus nigra* (black poplar) in New Zealand (Slippers *et al*., 2004a), *Actinidia deliciosa* (kiwifruit) in New Zealand (Slippers *et al*., 2004a)*, Malus sylvestris* (wild apple) in New Zealand (Zhou and Stanosz, 2001)*, Ribes* sp. (currents) in Australia (Slippers *et al*., 2004a), *Tibouchina* sp. in Australia (Slippers *et al*., 2004b) *Heteropyxis natalensis* (lavender) (Smith *et al*., 2001), *Terminalia catappa* (Begoude *et al*., 2010) and *S. cordatum* (Pavlic *et al*., 2007) in South Africa. *N. australe* has been previously isolated from *Acacia* sp. in Australia (Slippers *et al*., 2004c), *Wollemia nobilis* in Australia (Slippers *et al*., 2005), and *Widdringtonia nodiflora* (mountain cypress) in South Africa (Slippers *et al*., 2005). *L. pseudotheobromae* was found on *Rosa* sp. (rose) in Netherlands (Alves *et al*., 2008), *Coffea* sp. (coffee) in Zaire (Alves *et al*., 2008), *Citrus aurantium* (sour orange) in Suriname (Alves

et al., 2008), *Gmelina arborea* (Beechwood) in Costa Rica (Alves *et al*., 2008), *Acacia mangium* (black wattle) in Costa Rica (Alves *et al*., 2008) and *Terminalia catappa* (Bengal almond) in Cameroon (Begoude *et al*., 2010). In Uruguay the Botryosphaeriaceae on a non-native *Eucalyptus* sp. and various species of native Myrtaceae (Pérez *et al*., 2008) included *N. parvum/ N. ribis* and *B. dothidea* on all the Myrtaceae, while *N. eucalyptorum* was found exclusively on *Eucalyptus*, however after further sampling *N. eucalyptorum* was also identified on other Myrtaceae such as *S. cordatum* (Pérez *et al*., 2009). These indicate that these fungi have the ability to infect a great diversity of hosts, and their occurrence between species is thus not surprising.

In South Africa Botryosphaeriaceae on native *S. cordatum* have been well characterised (Burgess and Wingfield, 2002; Pavlic *et al*., 2004; Pavlic *et al*., 2007, 2008, 2009; Smith *et al*., 1994). Species that have been described from this tree include *N. ribis, N. kwambonambiense, N. umdonicola, N. cordaticola, N. australe, N. mangiferae, N. parvum, N. luteum, B. dothidea, L. theobromae* and *L. gonubiense* (Pavlic *et al*., 2009; Pavlic *et al*., 2007a, 2008). These species have been collected from various locations in South Africa across the natural range of *S. cordatum* and were mostly isolated as endophytes or associated with disease symptoms such as die-back (Pavlic *et al*., 2004; Pavlic *et al*., 2007, 2009). Pathogenicity tests indicated that *N. ribis* and *L. theobromae* were the most pathogenic species on *S. cordatum* (Pavlic *et al*., 2007). Only *N. parvum* and *N. australe* are known to co-infect *Eucalyptus* spp. and *S. cordatum* in South Africa (Pavlic *et al*., 2007).

Previous studies identifying the Botryosphaeriaceae on *Eucalyptus* spp. and *S. cordatum* took a broad approach where a large number of trees were sampled over a broad geographic area. Furthermore, the samples were from both disease symptoms and healthy twigs and leaves. The current study followed a high density sampling approach in order to directly compare the Botryosphaeriaceae species assemblages from an *E. grandis* and a *S. cordatum* tree collected at the same time and from the same geographical location. A multi-gene sequencing approach was followed to identify the various isolates because the ITS region (Internal Transcribed Spacer) alone, which is commonly used for fungal species identification, is insufficient to resolve certain species complexes in the Botryosphaeriaceae. Examples include the *N. parvum/ N. ribis* complex and *N. luteum/ N. australe* complex (Pavlic *et al*., 2008; Slippers *et al*., 2004c). Additional DNA sequences of the elongation factor 1-alpha (EF-1Ŭ) and the RNA polymerase II subunit (RPB2) genes were thus also used.

2. Materials and Methods

2.1 Sampling site and fungal isolations

Sampling was performed on the eastern coast (Mtubatuba, KwaZulu Natal) of South Africa, where *Eucalyptus* are grown as non-natives in plantations surrounded by patches of natural vegetation that include *S. cordatum* trees. This particular site (E 32º54,ôS28º29ô53.0, 33m above sea level) was chosen because a comprehensive survey of the Botryosphaeriaceae on *S. cordatum* throughout South Africa (Pavlic *et al.* 2009) had previously shown that trees from this region had a high diversity of Botryosphaeriaceae species. An *E. grandis* and a *S. cordatum* tree were sampled in April 2009., following a high density sampling approach where asymptomatic plant tissues (leaves, increment cores of wood, twigs and petioles were taken from both trees, with the exception that for the *S. cordatum* tree, no petioles were sampled because the leaves are sessile. Four leaves per branch, four branches, one increment, and one tree per species were placed in paper bags and transferred to the laboratory.

To remove fungal propagules and epiphytic fungi on the plant tissue surfaces (leaves, twigs, petioles and trunk increments) all the substrate tissue samples used in this study were surface sterilized using 10 % hydrogen peroxide for 3 minutes after samples were washed twice with sterile water for 1 minute each. The surface sterilized tissue was dissected into smaller subsections. Leaf discs (5 mm diameter) and petioles, twigs and trunk increments (3 mm) from the trees were placed on full strength malt extract (MEA) agar (20 g malt extract, 20 g agar; Biolab, Midrand, SA) with four to six subsections from the same sample placed approximately 4 cm apart. The plates were incubated at 25 ºC for approximately 10 days. Growth of endophytic fungi from the plant tissue was checked daily to isolate slow growing fungi before they were overgrown by other fungi. Plant tissues that did not show any initial fungal growth were monitored for a month. Cultures morphologically resembling the Botryosphaeriaceae (grey to dark in colour with fluffy mycelium, or with black pigment visible from the reverse side of the Petri dish) were sub-cultured to obtain pure cultures by transferring single hyphae onto new MEA agar plates using a sterile needle. Purified cultures were incubated for two weeks under near-UV light and grouped in morphotypes according to colony shape, colour, texture, mycelium type, medium discolouration and colony density. ITS data were generated for these isolates and groupings based on the ITS sequence data were verified with culture morphology (texture, margin, colour, discolouration of the medium) to

ensure that the subsets of all groups observed with colony morphology and DNA sequences were included in additional multi-locus sequencing. All cultures are maintained in the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa.

2.2 DNA sequence analysis

Mycelium was scraped from the surface of cultures using a sterile scalpel, transferred to 2 ml Eppendorf tubes, freeze-dried and ground to a fine powder using sterile 2 mm metal beads on a Mixer Mill type MM 301 Retsch^{R} tissue lyser (Retsch, Germany) for 3 min at a frequency of 30 cycles. Total fungal genomic DNA was extracted following a method described by Moller *et al*. (1992). DNA pellets were re-suspended in 50 µl sterile SABAX water (Adcock Ingrams, Bryanston, South Africa). DNA concentrations were determined using the NDï 1000 spectrophotometer V3.7.1 (Thermo Fisher Scientific, USA). The DNA was diluted to 50 ng/µl for use in subsequent polymerase chain reactions (PCRõs).

The full length ITS region, that included parts of the small ribosomal subunit (SSU) and large ribosomal subunit (LSU), were amplified using the forward V9G primer (Hoog and Ende, 1998) and reverse LR5 (Vilgalys and Hester, 1990) primer. The PCR reaction consisted of a 25 µl final volume and included 0.5 µl DNA template (50 ng/µl), 1 µl of each primer (10 mM), 2.5 µl (10 mM) dNTPõs, 2.5 µl of 10x PCR buffer with $MgCl_2$, 1 unit *taq* polymerase (Roche Molecular Biochemicals, Almeda, California) and 17 µl sterile distilled water (SABAX water; Adcock Ingrams, Bryanston).The PCR conditions were as follows: 94 ºC for 4 min, followed by 35 cycles of 94 ºC for 30 sec (denaturation), 55 ºC for 45 sec (annealing), 72 ºC for 1 min (elongation) and 72 ºC for 4 min (final elongation). The PCR products were then visualized in 2 % agarose gels using Gel Red (Biotium, Haward, California, USA) in 1x TAE buffer (Tris base 0.4 M, acetic acid 1 %, EDTA 0.5 M, pH 8.0). A subset of isolates representing the various species identified based on ITS data (Table 1), were chosen for subsequent multi-gene sequencing. The EF-1Ŭ gene region was amplified using the primer pair EF1F and EF2R (Jacobs et al., 2004) and the RBP2 region using the primers RPB2Bot6F and RPB2Bot6R (Sakalidis, 2004). The EF-1Ŭ PCR protocol and program parameters were the same as that used for the ITS amplifications, while the protocol for amplifying RBP2 was that of (Pavlic et al., 2008). Amplified products were visualized on a 2 % agarose gel using Gel red (Biotium, Haward, California, USA) in 1x TAE buffer (Tris base 0.4M, acetic acid 1 %, EDTA 0.5M, pH 8.0).

Amplified DNA products from the three gene regions were purified using the Sephadex^{R} G-50 columns (Sigma-Aldrich). Sequencing was performed in both directions using the same forward and reverse primers used in the PCR reaction except for the ITS where the internal primer set V9G (Hoog and Ende, 1998) and ITS4 (White *et al*., 1990) was used. Each sequencing PCR reaction contained 2.5 µl purified DNA, 2.1 µl reaction buffer, 0.5 µl ready reaction buffer (BigDye), 1.5 µl primer (10mM) and 5.4 µl sterile distilled water (SABAX water; Adcock Ingrams, Bryanston). The sequencing PCR conditions were as follows: 96 ºC for 10 sec followed by 25 cycles of 53 ºC for 5 sec, and 60 ºC for 4 min.

Consensus sequences from the forward and reverse sequences were built using the CLC Bio Workbench version 5 (CLC bio, Aarhus, Denmark) and sequence inconsistencies were checked manually. ITS sequences generated and the additional genes sequenced were added to the datasets (Table 1) that included sequences of species of Botryosphaeriaceae found on *S. cordatum* and *E. grandis* in South Africa and other species known for these genera (Alves *et al*., 2008; Begoude *et al*., 2010; Burgess *et al*., 2006a; Damm *et al*., 2007; Pavlic *et al*., 2009a; Pavlic *et al*., 2004; Pavlic *et al*., 2007a, 2008; Slippers *et al*., 2004b; Slippers *et al*., 2004c). These sequences were aligned with the online programme MAFFT version 6 (Katoh *et al*., 2002) and alignments were verified manually.

A most parsimonious phylogenetic tree (MP) was inferred in PAUP (Phylogenetic Analysis Using Parsimony) version 4.0 (Swoffold 2000) for the three gene regions (ITS, EF-1Ŭ and RBP2) separately and for the combined ITS and EF-1Ŭ sequences. The RBP2 sequences could not be used in the combined dataset because they were only generated for species in the *Neofusicoccum* species complex. Heuristic searches were completed using random stepwise addition with 100 replicates, and the tree bisection and reconstruction (TBR) algorithm as branch swapping algorithm. Gaps were treated as 5th character and nucleotides were defined as unordered and unweighted. A 1000 replicate bootstrap analysis (Felsenstein 1985) was executed to assess the confidence levels of the branch nodes in the phylogenetic tree. A 1000 replicate partition homogeneity test was applied to the ITS and EF-1Ŭ sequence data sets to determine the congruency between the ITS and EF-1Ŭ sequence data after the exclusion of uninformative sites (Farris *et al*. 1995). Maximum likelihood (ML) phylogenetic analyses were performed on the DNA sequence data for the gene regions separately and combined to confirm the groupings obtained with the MP. The online program ATCG phyML 3.0 (http://atgc.lirmm.fr/phyml/) was used. Likelihood substitution models were determined by JModelTest: phylogenetic model averaging version

0.1.1 (2008) using the Akaike information criterion (AIC). The invariable sites were assumed to have a gamma distribution. Confidence levels of the branches were estimated using bootstrap analysis (1000 replicates).

3 Results

3.1 DNA sequence analysis

The ITS dataset consisted of 135 taxa, including 68 reference sequences and 67 generated sequences, thus representing all known genera in the Botryosphaeriaceae. The EF-1Ŭ dataset comprised of 103 taxa, including 36 reference sequences representing all genera of Botryosphaeriaceae and 67 sequences from this study. Sequences representing all known genera in the Botryosphaeriaceae were included in the dataset but not necessarily representing all species. Reference sequences of all species belonging to *Neofusicoccum* and *Lasiodiplodia* were selected for the combined datasets of the ITS and EF-1Ŭ because all the sequences generated in this study grouped into those clades. The combined ITS/EF-1Ŭ datasets consisted of 63 taxa, including 40 reference sequences and 23 sequences generated from this study. The RBP2 datasets contained only 47 taxa, consisting of 32 reference sequences and 15 sequences from this study, and only represented sequences of *Neofusicoccum* species, with the aim to distinguish isolates belonging to the *N. parvum/ N. ribis* species complex (Pavlic *et al*., 2008).

The results of the PHT test (ITS/EF-1Ŭ) revealed that the datasets were incongruent (P-value = 0.001). This was because some isolates within the *N. parvum/N. ribis* complex grouped together in the EF-1Ŭ tree, but formed distinct groups in the ITS tree (Pavlic *et al*., 2008). The ITS and EF-1Ŭ datasets were thus combined, keeping the incongruence in the specific group in mind, in order to increase the number of informative sites. Some species could still not be distinguished with confidence in phylogenetic sub-clades based on the ITS and EF-1Ŭ data, but these isolates were distinguished based on the RBP2 genes (Pavlic *et al*., 2008). Results of these analyses are summarised in Table 2.

The 436 trees generated from the combined ITS and EF-1Ŭ datasets differed with respect to the grouping between clades, but were consistent with respect to isolates comprising terminal clades (data not shown). There was strong bootstrap support for clades representing known species in both the parsimony and likelihood analyses (data not shown). The ML tree was chosen for presentation (Figure 1a) and showed that isolates sequenced in

this study grouped with isolates of *N. parvum, N. eucalyptorum, N. australe, N. kwambonambiense* and *L. pseudotheobromae. Neofusicoccum parvum* and *N. kwambonambiense* could not be resolved based on the EF-1Ŭ data alone, but were separated in a combined dataset of EF-1Ŭ and ITS. The RBP2 sequences further confirmed the identity of these isolates as *N. parvum* and *N. kwambonambiense* (Figure 1b).

Most *Neofusicoccum* species were present *Eucalyptus grandis* and *Syzygium cordatum*. The exception was *N. australe* that was only isolated from *E. grandis* leaves and twigs in this study. The distribution of the Botryosphaeriaceae species within the trees was variable. *Neofusicoccum eucalyptorum* was isolated from *E. grandis* leaves, petioles and twigs as well as a single leaf of *S. cordatum. Neofusicoccum kwambonambiense* was isolated from the leaves and twigs of *E. grandis* and also from the leaves of *S. cordatum. Neofusicoccum parvum* was isolated from trunk increment cores of both tree species, and the petioles of the *E. grandis* tree. The *Lasiodiplodia* clade included only *L. pseudotheobromae* that was found in *S. cordatum* trunk increments.

4. Discussion

This study reports a number of species of the important tree pathogen family, the Botryosphaeriaceae, for the first time from South Africa and certain hosts. Five Botryosphaeriaceae species were identified from various plant tissues and occurring on a native *S. cordatum* and a non-native *E. grandis* tree at that particular moment in time. These include *N. eucalyptorum, N. kwambonambiense, N. australe, N. parvum,* and *L. pseudotheobromae*. Three of these species, including *N. eucalyptorum, N. kwambonambiense* and *N. parvum*, were isolated from both trees, while *N. australe* and *L. pseudotheobromae* were found only on *E. grandis* or *S. cordatum*, respectively. *Neofusicoccum eucalyptorum* and *L. pseudotheobromae* were found on *S. cordatum*, and *N. kwambonambiense* on *Eucalyptus*, for the first time in South Africa despite relatively wide and thorough surveys on these trees in the past.

The most abundant species of the Botryosphaeriaceae in the two trees sampled in this study was *N. eucalyptorum*, which represented 38 % of the total Botryosphaeriaceae isolates. This fungus was first described by Smith *et al*. (2001) from cankers on the main stems of *E. grandis* and *E. nitens* in South Africa. In our study it was also isolated from leaves, twigs and wood increments of *E. grandis*, and from a leaf of *S. cordatum*. This is the first time it is

reported from this host despite previous surveys in South Africa. This fungus is, however, known from other members of the Myrtaceae, such as *Blepharocalyx salicifolius, Myreceugenia glaucescens* and *Myrrhinium atropurpureum* var. *octandrum*in from Uruguay (Pérez *et al*., 2008*).* A previous study (Slippers *et al*., 2004b) suggested that *N. eucalyptorum* is native to *Eucalyptus* in Australia based on its abundance, distribution and its association with *Eucalyptus* spp. on that continent and possibly introduced to South Africa (Slippers *et al*., 2004c). This fungus clearly has the ability to infect at least various members in the Myrtaceae and could move between native and non-native hosts of these families in countries where it is introduced.

Neofusicoccum kwambonambiense was the second most abundant species, representing 33 % of the total Botryosphaeriaceae isolated in this study. This species was isolated from leaves of *S. cordatum* and the leaves and twigs of *E. grandis*. This fungus was previously reported from asymptomatic branches and leaves, dying branches and pulp of the ripe fruit of *S. cordatum* (Pavlic *et al*., 2009), and it only known from *S. cordatum* and close to the sampling location of this study in South Africa. Studies conducted by Sakalidis *et al*., 2011 also identified *N. kwambonambiense* on *E. dunni* in eastern Australia and *Corymbia torelliana* in northern Australia (Sakalidis *et al*., 2011). Therefore the origin and complete host and geographical range of *N. kwambonambiense* are widely unknown. Pathogenicity tests undertaken on both hosts by Pavlic *et al*. (2009) suggested that *N. kwambonambiense* is more pathogenic than isolates of *N. ribis* and *N. parvum,* and it was more aggressive on *Eucalyptus* spp. than native *S. cordatum.*

Neofusicoccum australe was isolated from the leaves and twigs of *Eucalyptus* and represented 7% of the Botryosphaeriaceae isolates. *Neofusicoccum australe* is a recently described species from diseased stems of native *Acacia* spp. in Australia (Slippers *et al*., 2004b; Slippers *et al*., 2004c) and has since been found on *Eucalyptus* spp. in western Australia and South Africa (Slippers *et al*., 2004b). Pavlic *et al*. (2007) found low levels of *N. australe* from *S. cordatum.* Those isolates were shown to be pathogenic in greenhouse trials on *S. cordatum* and an *E. grandis x camaldulensis* clone. Results of this study confirm that *N. australe* is yet another species that can naturally infect *Eucalyptus* and *S. cordatum* in South Africa.

Neofusicoccum parvum represented 12 % of the isolates and was predominantly isolated from the leaves, twigs and petioles of *Eucalyptus*. Only a single isolate came from a *S.*

cordatum leaf. Studies have shown that *N. parvum* is broadly distributed around the world with the potential to cause diseases of *Eucalyptus* (Ahumada, 2003; Crous *et al*., 1989; Slippers *et al*., 2004c; Slippers *et al*., 2004d). This fungus is also known from various hosts, in South Africa, including non-native *Tibouchina* sp. (Heath *et al*., 2011) and native *Heteropyxis natalenis* (Smith *et al*., 2001) and *S. cordatum* (Pavlic *et al*., 2007).

Lasiodiplodia pseudotheobromae represented 11 % of the Botryosphaeriaceae isolates and it was isolated from *S. cordatum* bark and trunk increments. This is the first report of *L. pseudotheobromae* on *S. cordatum*. *Lasiodiplodia pseudotheobromae* has been previously reported on *Eucalyptus* in Eastern Australia and Venezuela (Alves *et al*., 2008; Mohali *et al*., 2005), thus indicating that this species also has the ability to infect *Eucalyptus* spp. *Lasiodiplodia pseudotheobromae* has only recently been separated as a cryptic species from its sister species *L. theobromae* (Alves *et al*., 2008), a species known across the world from many. Reports of *L. theobromae* earlier than this can thus also represent *L. pseudotheobromae*. The extent that this species occur on *Eucalyptus* in South Africa and its ability to cause disease also needs to be considered.

The occurrence of the Botryosphaeriaceae species encountered in this study on both *Eucalyptus* and *S. cordatum* was not surprising. Previously *N. parvum*, *N. australe* and *L. theobromae* were known to co-infect both hosts (Pavlic *et al*., 2007), while this study also show that *N. eucalyptorum* and *N. kwambonambiense* can infect both trees. It thus appears to be a general characteristic of Botryosphaeriaceae in South Africa, and from early evidence also elsewhere, to be able to infect different hosts in the Myrtaceae, irrespective or their origin. There thus appears to be little limitation for invasive fungi to move from non-native to native hosts, and for endemic fungi in this group to infect non-native hosts. These findings re-affirm the previous concerns that such movement of Botryosphaeriaceae could be common and should be considered in quarantine and disease management programs (Burgess and Wingfield, 2002; Burgess *et al*., 2006b; Slippers and Wingfield, 2007; Wingfield *et al*., 2001, 2011).

The sampling strategy used in this study was limited as only a single tree of each species was sampled, albeit relatively intensively. Our results are thus not indicative of the full range of infection, geographical distribution, host range and tissue specificity of Botryosphaeriaceae on *E. grandis* and *S. cordatum* in South Africa, or even in the KwaMbonambi/ Matubatuba area. It rather represents a snapshot of these species in a

single tree at a particular time. Yet, despite this limited extent this study still yielded new reports from hosts well studied in the past. This indicates that continued monitoring, with the latest identification tools and taxonomic framework, is necessary to fully appreciate the full geographical and host ranges of these species and their potential to contribute to Botryosphaeria canker on these hosts.

Acknowledgements

We thank the Thuthuka Programme of the National Research Foundation (NRF), members of the Tree Protection Co-operative Programme (TPCP) and the Department of Science and Technology (DST)/NRF Centre of Excellence in Tree Health Biotechnology (CTHB), South Africa, for financial support. Additionally, we thank Jan Nagel, Fahimeh Jami and Tuan Duong for their assistance with the phylogenetic analysis.

Reference

Ahumada R (2003) *Pathogens in commercial Eucalyptus plantations in Chile, with special reference to Mycosphaerella and Botryosphaeria species* Master's Dissertation, University of Pretoria.

Alves A, Crous PW, Correia A and Phillips AJL (2008) Morphological and molecular data reveal cryptic speciation in *Lasiodiplodia theobromae*. *Fungal Diversity* **28**, 1-13.

Begoude BAD, Slippers B, Wingfield MJ and Roux J (2010) Botryosphaeriaceae associated with *Terminalia catappa* in Cameroon, South Africa and Madagascar. *Mycological Progress* **9**, 101-123.

Burgess T and Wingfield MJ (2002) Quarantine is important in restricting the spread of exotic seed-borne tree pathogens in the southern hemisphere. *International Forestry Review* **4**, 56-65.

Burgess, T. and Wingfield, M.J. (2002) *Impact of fungi in natural forest ecosystems: A focus on Eucalyptus.* In: Sivasithamparam, K.,Dixon, K.W. and Barrett, R.L., (eds.) Microorganisms in plant conservation and biodiversity. Kluwer Academic Publishers, Dordrecht, . 285-306..

Burgess TI, Barber PA and Hardy G (2005) *Botryosphaeria* spp. associated with eucalypts in Western Australia, including the description of *Fusicoccum macroclavatum* sp. nov. *Australasian Plant Pathology* **34**, 557-567.

Burgess TI, Barber PA, Mohali S, Pegg G, de Beer W and Wingfield MJ (2006a) Three new *Lasiodiplodia* spp. from the tropics, recognized based on DNA sequence comparisons and morphology. *Mycologia* **98**, 423-435.

Burgess TI, Sakalidis ML and Hardy GE (2006b) Gene flow of the canker pathogen *Botryosphaeria australis* between *Eucalyptus globulus* plantations and native eucalypt forests in Western Australia. *Austral Ecology* **31**, 559-566.

Chen SF, Pavlic D, Roux J, Slippers B, Xie YJ, Wingfield MJ and Zhou XD (2011) Characterization of Botryosphaeriaceae from plantation grown *Eucalyptus* species in South China. *Plant Pathology* **60**, 739-751.

Crous PW, Knox-Davies PS and Wingfield MJ (1989) Newly-recorded foliage fungi of *Eucalyptus* spp. in South Africa. *Phytophylactica* **21**, 85-88.

Damm U, Crous PW and Fourie PH (2007) Botryosphaeriaceae as potential pathogens of *Prunus* species in South Africa, with descriptions of *Diplodia africana* and *Lasiodiplodia plurivora* sp. nov. *Mycologia* **99**, 664-680.

Farris JS, Kallersjo M, Kluge AG and Bult C (1995) Constructing a significance test for incongruence. *Systematic Biology* **44**, 570-572.

Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist*, 1-15.

Gezahgne A, Roux J, Slippers B and Wingfield MJ (2004) Identification of the causal agent of *Botryosphaeria* stem canker in Ethiopian *Eucalyptus* plantations. *South African Journal of Botany* **70**, 241-248.

Heath RN, Roux J, Slippers B, Drenth A, Pennycook SR, Wingfield BD and Wingfield MJ (2011) Occurrence and pathogenicity of *Neofusicoccum parvum* and *N. mangiferae* on ornamental *Tibouchina* species. *Forest Pathology* **41**, 48-51.

Hoog GS and Ende A (1998) Molecular diagnostics of clinical strains of filamentous Basidiomycetes. *Mycoses* **41**, 183-189.

Jacobs K, Bergdahl DR, Wingfield MJ, Halik S, Seifert KA, Bright DE and Wingfield BD (2004) *Leptographium wingfieldii* introduced into North America and found associated with exotic *Tomicus piniperda* and native bark beetles. *Mycological Research* **108**, 411-418.

Katoh K, Misawa K, Kuma K and Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059-3066.

Moller EM, Bahnweg G, Sanderman H, Geiger HH (1992) A simple and efficient protocol for isolation of high molecular weight filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic acids research* **20**, 6115-6116.

Mohali S, Burgess TI and Wingfield MJ (2005) Diversity and host association of the tropical tree endophyte *Lasiodiplodia theobromae* revealed using simple sequence repeat markers. *Forest Pathology* **35**, 385-396.

Mohali S, Slippers B and Wingfield MJ (2006) Two new Fusicoccum species from *Acacia* and *Eucalyptus* in Venezuela, based on morphology and DNA sequence data. *Mycological Research* **110**, 405-413.

Mohali S, Slippers B and Wingfield MJ (2007) Identification of Botryosphaeriaceae from *Eucalyptus*, *Acacia* and *Pinus* in Venezuela. *Fungal Diversity* **25**, 103-125.

Nakabonge G. 2002. Diseases associated with plantation forestry in Uganda. MSc thesis, University of Pretoria

Pavlic D, Slippers B, Coutinho T and Wingfield M (2009) Molecular and phenotypic characterisation of three phylogenetic species discovered within the *Neofusicoccum parvum/N. ribis* complex. *Mycologia* **101**, 636-647.

Pavlic D, Slippers B, Coutinho TA, Gryzenhout M and Wingfield MJ (2004) *Lasiodiplodia gonubiensis* sp. nov., a new *Botryosphaeria* anamorph from native *Syzygium cordatum* in South Africa. *Studies in Mycology* **50**, 313-322.

Pavlic D, Slippers B, Coutinho TA and Wingfield MJ (2007) Botryosphaeriaceae occurring on native *Syzygium cordatum* in South Africa and their potential threat to *Eucalyptus*. *Plant Pathology* **56**, 624-636.

Pavlic D, Slippers B, Coutinho TA and Wingfield MJ (2008) Multiple gene genealogies and phenotypic data reveal cryptic species of the Botryosphaeriaceae: A case study on the *Neofusicoccum parvum/N. ribis* complex. *Molecular Phylogenetics and Evolution* **51**, 259-268.

Pérez CA, Altier N, Simeto S, Wingfield MJ, Slippers B and Blanchette RA (2008) Botryosphaeriaceae from *Eucalyptus* and native Myrtaceae in Uruguay. *Agrociencia* **12**, 19-30.

Pérez CA, Wingfield MJ, Slippers B, Altier NA and Blanchette RA (2009) *Neofusicoccum eucalyptorum*, a *Eucalyptus* pathogen, on native Myrtaceae in Uruguay. *Plant Pathology* **58**, 964-970.

Rodas CA, Slippers B, Gryzenhout M and Wingfield MJ (2009) Botryosphaeriaceae associated with *Eucalyptus* canker diseases in Colombia. *Forest Pathology* **39**, 110-123.

Roux J, Coutinho TA, Wingfield MJ and Bouillet JP (2000) Diseases of plantation *Eucalyptus* in the Republic of Congo. *South African Journal of Science* **96**, 454-456.

Sakalidis M (2004) *Resolving the Botryosphaeria ribis-B. parva species complex; a molecular and phenotypic investigation*. Honours thesis, Murdoch University, Australia.

Sakalidis ML, Hardy GESJ and Burgess TI (2011) Use of the Genealogical Sorting Index (GSI) to delineate species boundaries in the *Neofusicoccum parvum-Neofusicoccum ribis* species complex. *Molecular Phylogenetics and Evolution* **60**, 333-344.

Slippers B, Burgess T, Pavlic D, Ahumada R, Maleme H, Mohali S, Rodas C and Wingfield MJ (2009) A diverse assemblage of *Botryosphaeriaceae* infect *Eucalyptus* in native and non-native environments. *Southern Forests: a Journal of Forest Science* **71**, 101-110.

Slippers B, Crous PW, Denman S, Coutinho TA, Wingfield BD and Wingfield MJ (2004a) Combined multiple gene genealogies and phenotypic characters differentiate several species previously identified as *Botryosphaeria dothidea*. *Mycologia* **96**, 83-101.

Slippers B, Fourie G, Crous PW, Coutinho TA, Wingfield BD, Carnegie AJ and Wingfield MJ (2004b) Speciation and distribution of *Botryosphaeria spp*. on native and introduced *Eucalyptus* trees in Australia and South Africa. *Studies in Mycology* **50**, 343-358.

Slippers B, Fourie G, Crous PW, Coutinho TA, Wingfield BD and Wingfield MJ (2004c) Multiple gene sequences delimit *Botryosphaeria australis* sp. nov. from *B. lutea*. *Mycologia* **96**, 1030-1041.

Slippers B, Summerel BA, Crous PW, Coutinho TA, Wingfield BD and Wingfield MJ (2005) Preliminary studies on *Botryosphaeria* species from Southern Hemisphere conifers in Australasia and South Africa. *Australasian Plant Pathology* **34**, 213-220.

Slippers B and Wingfield MJ (2007) Botryosphaeriaceae as endophytes and latent pathogens of woody plants: diversity, ecology and impact. *Fungal Biology Reviews* **21**, 90-106.

Smith H, Crous PW, Wingfield MJ, Coutinho TA and Wingfield BD (2001) Botryosphaeria eucalyptorum sp. nov., a new species in the *B. dothidea*-complex on *Eucalyptus* in South Africa. *Mycologia* **93**, 277-285.

Smith H, Kemp GHJ and Wingfield MJ (1994) Canker and die-back of Eucalyptus in South Africa caused by Botryosphaeria dothidea. *Plant Pathology* **43**, 1031-1034.

Vilgalys R and Hester M (1990) Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology* **172**, 4238-4246.

White TJ, Bruns T, Lee S and Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols a guide to methods and applications*, 315-322.

Wingfield MJ, Slippers B, Roux J and Wingfield BD (2001) Worldwide movement of exotic forest fungi, especially in the tropics and the Southern Hemisphere. *BioScience* **51**, 134-140.

Wingfield MJ, Slippers B, Roux J and Wingfield BD (2011) Fifty years of tree pest and pathogen invasions, increasingly threatening world forests. In: *Fifty Years of Invasion Ecology*, pp. 89-99. Wiley Online Library.

Zhou S and Stanosz GR (2001) Relationships among *Botryosphaeria* species and associated anamorphic fungi inferred from the analyses of ITS and 5.8 S rDNA sequences. *Mycologia* **93**, 516-527.

Table 1. Isolates representing Botryosphaeriaceae used in the phylogenetic study

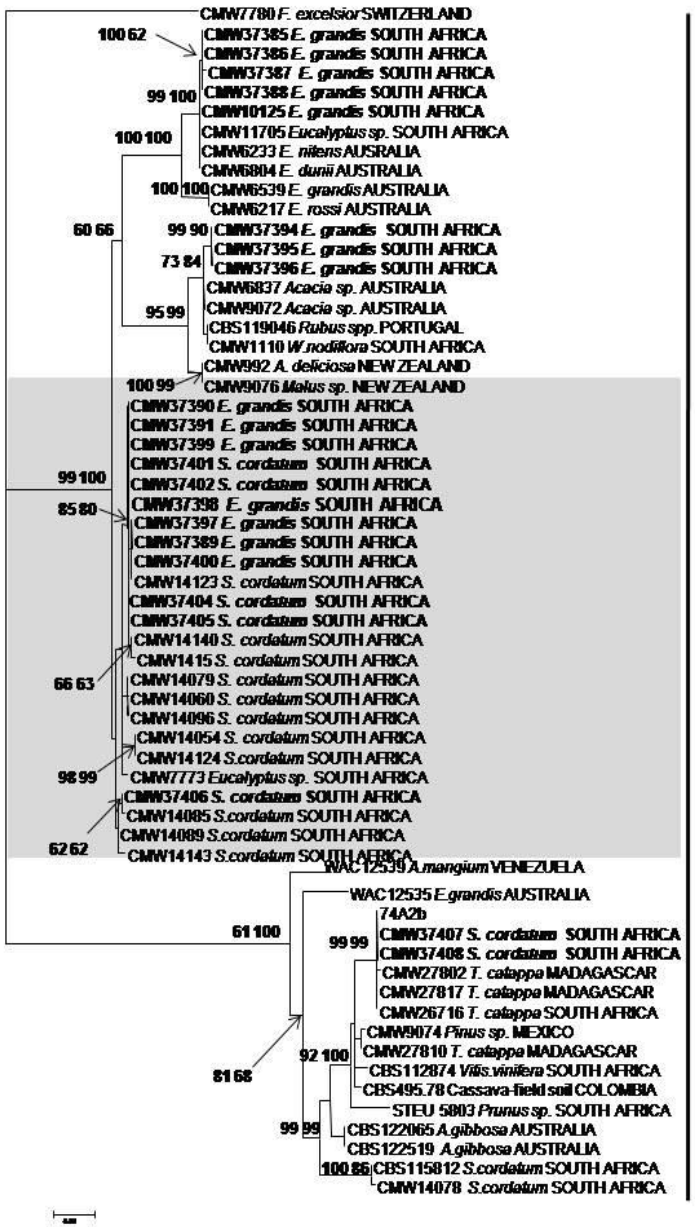
Isolate name	Identification	Host	Country	Reference	ITS	EF	RBP2
CMW37407	*Lasiodiplodia pseudotheobromae*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744583	JQ744604	
CMW37408	*Lasiodiplodia pseudotheobromae*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744584	JQ744605	
CMW37387	*Neofusicoccum eucalyptorum*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744579	JQ744600	
CMW37388	*Neofusicoccum eucalyptorum*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744580	JQ744601	
CMW37386	*Neofusicoccum eucalyptorum*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744581	JQ744602	
CMW 37385	*Neofusicoccum eucalyptorum*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744603	JQ744582	
CMW37396	*Neofusicoccum australe*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744576	JQ744597	
CMW37395	*Neofusicoccum australe*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744577	JQ744598	
CMW37394	*Neofusicoccum australe*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744578	JQ744599	
CMW37406	*Neofusicoccum parvum*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744564	JQ744585	JQ744609
CMW37400	*Neofusicoccum kwambonambiense*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744582	JQ744603	JQ744606
CMW37399	*Neofusicoccum kwambonambiense*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744566	JQ744587	JQ744614
CMW37401	*Neofusicoccum kwambonambiense*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744567	JQ744588	JQ744611
CMW37402	*Neofusicoccum kwambonambiense*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744568	JQ744589	JQ744612
CMW37389	*Neofusicoccum kwambonambiense*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744569	JQ744590	
CMW37398	*Neofusicoccum kwambonambiense*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744570	JQ744591	JQ744606
CMW37391	*Neofusicoccum kwambonambiense*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744571	JQ744592	JQ744615
CMW37405	*Neofusicoccum kwambonambiense*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744572	JQ744593	JQ744610
CMW37397	*Neofusicoccum kwambonambiense*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744573	JQ744594	JQ744607
CMW37395	*Neofusicoccum kwambonambiense*	*Eucalyptus grandis*	South Africa	Gryzenhout,M.	JQ744574	JQ744598	JQ744613
CMW37404	*Neofusicoccum kwambonambiense*	*Syzygium cordatum*	South Africa	Gryzenhout,M.	JQ744575	JQ744596	JQ744608

Abbreviations of isolates and culture collection: CBS, Centraalbureau voor Schimmelcultures Utrecht, Netherlands; CMW, Forestry and Agricultural Biotechnology Institution, University of Pretoria

Table 2. Statistics related to phylogenetic analyses

	ITS	EF-1Ů	RBP2	Combined (ITS and EF-1Ů)
Amplified region size (bp)	700	600	700	1 300
No. of Characters	500	244	613	744
Parsimony informative	100	132	182	230
Parsimony uninformative	376 constant, 24 variable	97 constant, 15 variable	421 constant, 10 variable	472 constant, 41 variable
No. of trees retained	6	99	2	436
g1	-0.65	-0.7	-0.7	-0.5
Consistency index (CI)	0.8	0.7	1	0.7
Retention index (RI)	0.9	0.9	1	0.9
Substitution model (AIC)	GTR+G	HKY+G	GTR	GTR+G
Gamma shape	0.2	0.4	0.6	0.2

a



b

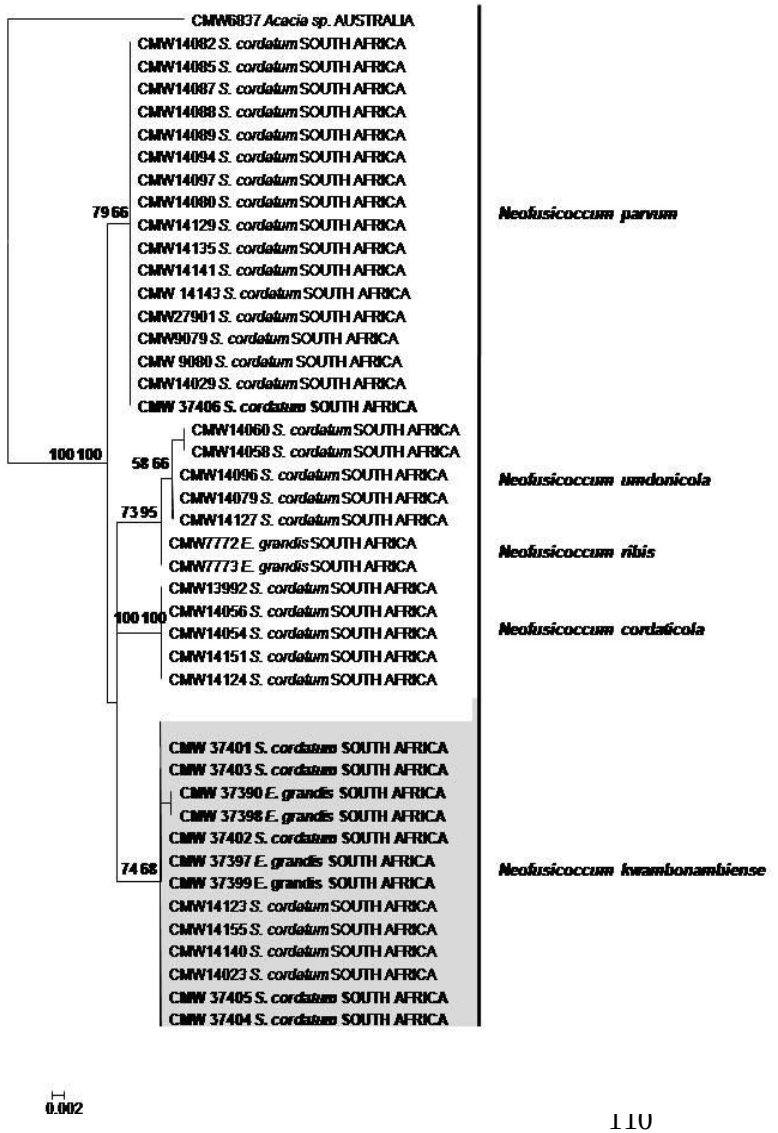


Fig1a. Phylogram produced with the TBR algorithm of a heuristic search on a combined dataset of the ITS and the EF1Ŭ. Group frequencies and bootstrap values are indicated (maximum likelihood bootstrap followed by maximum parsimony. Fig1b. Phylogram depicting the relationship amongst the *Neofusicoccum parvum*, *Neofusicoccum kwambonambiense, Neofusicoccum umdonicola* and *Neofusicoccum cordaticola* based on maximum likelihood and maximum parsimony analysis of the RBP2 gene.

Summary

Endophytic fungi of most woody trees are poorly studied due to their cryptic existence, the complexity of the community and the poor ability of traditional tools to characterize them. This study characterised the endophytic community present in commercially important *Eucalyptus grandis* trees in South Africa. Two different experimental approaches were used. Firstly environmental bar-coding, or metagenetics, was done using 454-pyrosequencing parallel sequencing of the barcode amplicons of all the fungal isolates present in the plants from the total DNA of the plant. Secondly, conventional DNA bar-coding of was done of fungal endophyte isolates. Isolates of the Botryosphaeriaceae family of latent endophytic pathogens were further characterised using a multi-gene phylogenetic approach from both *E. grandis* and related native *S. cordatum* that grew in close proximity.

The endophytes within the three *E. grandis* trees were hyper diverse. A total of 1 281 Molecular Operational Taxonomic Units (MOTU) was identified based on 454-pyrosequencing of the *E. grandis* fungal endophyte infections. Only 85 fungal endophytic species were identified amongst isolates from one of these trees, using the conventional DNA bar-coding approach. Fifteen times more species/MOTU was thus recovered using a metagenetics compared to an isolation approach. Despite this high diversity the species accumulation curves indicate that more endophytic diversity is to be discovered. The multi-gene analysis of Botryosphaeriaceae isolates obtained from the *E. grandis* and *S. cordatum* trees show that three species co-infect both these hosts. Two novel host associations are also reported. This approach of verifying identities of cryptic species with appropriate multi-gene analyses is most likely needed for other diverse species complexes associated with these trees.

A very thorough sampling strategy is required to adequately characterize the endophyte diversity in trees. The experimental approach, the 454-pyrosequening identification workflow and database described in this study will be useful to study these endophyte communities over time and space in future. Using these techniques and workflows described, questions related to host association, diversity and spatial distribution within hosts, and geographical delimitation of endophytes can be addressed.

Summary