# CHARACTERIZATION OF THE *TERATOSPHAERIA DESTRUCTANS WINGFIELD & CROUS (1996)* GENOME

C. T. TATHAM

# CHARACTERIZATION OF THE *TERATOSPHAERIA DESTRUCTANS WINGFIELD AND CROUS (1996)* GENOME

By

**Catherine T. Tatham**

Submitted in partial fulfilment of the requirement of the degree

**Magister Scientiae**

In the Faculty of Natural and Agricultural Sciences
Department of Biochemistry, Genetics and Microbiology
University of Pretoria
Pretoria

May 2019

**Primary Supervisor**
Prof. B. D. Wingfield

**Co-supervisors**
Prof. M. J. Wingfield
Dr. M. A. van der Nest
Dr. P. M. Wilken

# DECLARATION

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

______________________ ______________

**Catherine T. Tatham** **Date**

*Anything’s possible if you’ve got enough nerve.*

*J. K. Rowling*

*To Mum and Dad.*

# ACKNOWLEDGEMENTS

To Brenda, thank you for an incredible opportunity to learn. Thank you for the endless opportunities you've provided me to travel, to meet new people, and to be challenged.

To Mike, thank you for your invaluable knowledge and insight, not only during this project, but throughout my whole journey with FABI.

To Magriet, thank you for your insight and guidance throughout this project.

To Markus, thank you for you endless mentoring and friendship over the last seven years. Our daily coffee catch ups will forever be treasured.

To Tuan and Aquillah, thank you for your willingness to help me, no matter how small or big the task. Your knowledge and patience has been invaluable.

To the members of the 6th Floor, thank you for your support, and for sharing my frustrations and excitement over the years.

To the NRF, CTHB and FABI, thank you for the funding that made this research possible.

To Andi and Frances, thank you for all of your help behind the scenes.

To Bianca, Cait, Henry and Oreo, thank you for being there through it all, and for keeping a smile on my face.

To Mum, Dad, and Fats, thank you for your constant support, and for encouraging me to keep going. Without you, none of this would have been possible.

# PREFACE

*Teratosphaeria destructans* is a devastating *Eucalyptus* leaf and stem pathogen. Initially described from dying *Eucalyptus* plantations in North Sumatra two decades ago, this pathogen has since been identified in other eucalypt growing regions of the world, including South Africa. Despite the severity of this pathogen and the increase in its geographic range, nothing is known about the biology of *T. destructans*. This dissertation aims to address this.

Chapter 1 is presented as a literature study on the role that genomics plays in understanding the biology of emerging Ascomycete tree pathogens. This chapter briefly summarises the history of genome sequencing, with a focus on how the technology has developed. This is followed by a discussion on how the increased availability and accessibility of genome data has impacted our understanding of mating and population genetics within Ascomycete tree pathogens.

Sexual reproduction plays a vital role in generating genetic variation within populations, and thus in assisting a pathogen in adapting to changing environments. The second chapter of this dissertation aims to determine the sexual strategy of *T. destructans* by studying the mating type locus of this pathogen. To do this, the genomes of two *T. destructans* isolates from different geographic locations were sequenced and assembled, and the mating type loci were identified and described.

The final chapter of this dissertation focuses on developing microsatellite markers for *T. destructans*. These markers will serve as tools for discerning genetic variation both within populations and between populations. To develop the microsatellite markers, three *T. destructans* genomes were mined for microsatellites. The identified microsatellites were compared, and polymorphic loci were identified.

*Teratosphaeria destructans* is of great significance to *Eucalyptus* producing countries around the world. Understanding how this pathogen reproduces, as well as the distribution of genetic variation across populations will aid in the development of effective management and prevention strategies. Each of the chapters in this dissertation have been compiled as a separate entity, and thus content may overlap.

# CONTENTS

**ACKNOWLEDGEMENTS** i
**PREFACE** ii

**CHAPTER 1**
**Genomics as a tool for understanding Sexual Reproduction and Genetic Variation in Emerging Ascomycete Tree Pathogens** 1
1. INTRODUCTION 2
2. GENOME SEQUENCING 3
3. MATING STRATEGIES IN FUNGI AND GENOME SEQUENCING 5
   3.1 Mating in Ascomycetes 5
   3.2 Identification of Mating Types 6
4. POPULATION GENETICS OF FUNGI AND GENOME SEQUENCING 7
   4.1 Microsatellite Markers 8
5. DOTHIDEOMYCETES AND AVAILABLE GENOME SEQUENCES 9
6. CONCLUSIONS 10
7. REFERENCES 12
8. FIGURES 19

**CHAPTER 2**
**The *MAT1* locus structure of *Teratosphaeria destructans* indicates a Heterothallic Mating Strategy** 22
1. INTRODUCTION 24
2. MATERIALS AND METHODS 26
   2.1 Genome sequencing and annotation 26
   2.2 Reconstruction of the *MAT1* locus 27
   2.3 Geographic distribution of mating types 27
3. RESULTS 28
   3.1 Genome sequencing and annotation 28
   3.2 Reconstruction of the *MAT1* locus 29
   3.3 Geographic distribution of mating types 30
4. DISCUSSION 31
5. REFERENCES 34
6. TABLES AND FIGURES 38

**CHAPTER 3**

**Characterisation of Microsatellite Markers in *Teratosphaeria destructans*** 50

1. INTRODUCTION 52
2. MATERIALS AND METHODS 54
   - 2.1 A comparison of microsatellite markers in three *Teratosphaeria destructans* genomes 54
   - 2.2 Identification and analysis of previously published *Teratosphaeria destructans* microsatellites 54
   - 2.3 *In silico* evaluation of novel microsatellite markers 54
   - 2.4 Identification and analysis of previously published microsatellites for *Teratosphaeria* species other than *Teratosphaeria destructans* 55
3. RESULTS 55
   - 3.1 A comparison of microsatellite markers in three *Teratosphaeria destructans* genomes 55
   - 3.2 Identification and analysis of previously published *Teratosphaeria destructans* microsatellites 55
   - 3.3 *In silico* evaluation of novel microsatellite markers 56
   - 3.4 Identification and analysis of previously published microsatellites for *Teratosphaeria* species other than *Teratosphaeria destructans* 56
4. DISCUSSION 57
5. REFERENCES 59
6. TABLES AND FIGURES 62

**SUMMARY** 72

# Chapter 1

## Genomics as a tool for understanding Sexual Reproduction and Genetic Variation in Emerging Ascomycete Tree Pathogens

# Genomics as a tool to study Sexual Reproduction and Genetic Variation in Emerging Ascomycete Tree Pathogens

## 1. INTRODUCTION

Fungi and plants have evolved in close proximity over millions of years leading to the development of many relationships between them. In the case of Ascomycetes many of these relationships relate to plant diseases (Knief 2014; Saunders 2015). Historically, Ascomycete plant pathogens have caused substantial damage to plants (Berbee 2001; Wingfield *et al.* 2017). Some of the most notable examples are the causal agents of Dutch Elm Disease *Ophiostoma ulmi* and *O. novo-ulmi* (Lodge 1991), and *Cryphonectria parasitica* which is responsible for the devastating disease known as Chestnut Blight (Anagnostakis 2007). In more recent times, there has been a steady increase in the emergence diseases caused by Ascomycete plant pathogens, highlighting a need to understand the biology of these novel and emerging diseases (Cairns *et al.* 2016).

The concept of an "emerging plant disease" has been adapted from the definition of emerging infectious diseases used in the medical and veterinary fields (Anderson *et al.* 2004). Here, an emerging infectious disease is defined as a disease where the incidence has increased in the last 20 years or where it has the potential to increase in the future (Anderson *et al.* 2004). In many instances, emerging diseases are caused by pathogens that have one or more of four characteristics. These include (1) an increase in incidence, geographic or host range; (2) diseases that show a change in pathogenesis levels; (3) a newly evolved pathogen; or (4) pathogens that have been recently discovered or that are newly recognized (Vurro *et al.* 2010; Picco *et al.* 2011). Such characteristics have been observed in Ascomycete plant pathogens (Desprez-Loustau *et al.* 2009; Picco *et al.* 2011; Lione *et al.* 2018).

When studying emerging pathogens, researchers commonly target specific aspects of their biology, the most important of which is genetic diversity. This diversity is largely responsible for determining the adaptability of a pathogen to its environment (Dobzhansky 1955), with a greater population genetic diversity translating to better adaptability to the surrounding environment (Dobzhansky 1955; Gladieux *et al.* 2015). Therefore, it is imperative to determine how genetic diversity arises within a population, and to understand the distribution of genetic diversity across populations.

Sexual reproduction plays an important role in both the generation and maintenance of population-level genetic variation. In its simplest form, sexual reproduction in fungi occurs when the genetic material of two compatible mating partners undergoes fusion, followed by recombination, and meiosis and mitosis to form progeny (Lee *et al.* 2010). In most diploid fungi, this involves the formation of offspring through the fusion of two haploid gametes that were produced during meiotic process (Heitman *et al.* 2013). This is different to the sexual process found in most Ascomycete fungi that have predominately haploid lifecycles. In these species, sexual reproduction is facilitated by a diploid cell known as a heterokaryon which arises from the fusion of two haploid parental nuclei. Fusion is subsequently followed by meiotic division which not only restores the haploid nature of the cells to the progeny, but also generates genetic variation in the progeny (Billiard *et al.* 2012).

An ability to quantify genetic variation at the population level is able to provide valuable insight into the dynamics of pathogen populations (Dobzhansky 1955). Such genetic variation can be assessed at a genic or genotypic level within populations by analysing the distribution and polymorphism of genetic markers within populations (Sunnucks 2000). The level of polymorphism within a population can be indicative of a number of situations. For example, a low level of genetic variation could signify a recently introduced population currently experiencing a genetic bottle neck (Sakai *et al.* 2001; Gladieux *et al.* 2015; Wingfield *et al.* 2017). In contrast, a high level of variation could indicate that a population is well-established and potentially able to reproduce sexually (Gladieux *et al.* 2015).

A number of laboratory techniques exist to assess the mating strategy utilized by an Ascomycete plant pathogens, as well as the genetic variation in populations (Zane *et al.* 2002). However, these are often time consuming and costly. The popularity and increasing availability of genome sequencing has made determining these aspects significantly easier (Aylward *et al.* 2017; Wilken *et al.* 2017). This review highlights the role that genomics has played in studying emerging Ascomycete plant pathogens. Most notably, the focus is on the use of full genome sequences to rapidly identify and characterize a pathogen's mode of reproduction. In addition, the review will also consider how genomics has simplified the development of tools to study genetic variation within fungal populations.

## 2. GENOME SEQUENCING

Full genome sequencing has a surprisingly long history that is intimately tied to advances in sequencing technologies. The first technology used for full genome sequencing was

available as early as 1979 (Staden 1979). This made use of Shotgun Sequencing technology where fragmented genomic DNA was sequenced using the chain termination method (Staden 1979). However, this technology was useful only for genomes smaller than 7000 base pairs in size (Staden 1979). This remained the status quo for more than a decade, before technological advances in the 1990s made it possible to sequence larger genomes (Edwards *et al.* 1991; Besser *et al.* 2018). The largest contributing factor to this advancement was the recognition that DNA fragments could be sequenced in both directions, increasing the number of reads, and thus increasing the accuracy of the genome assembly (Edwards *et al.* 1991). Since then, there have been many advances in genome sequencing technology, with an associated shift towards high-throughput sequencing (Knief 2014; Besser *et al.* 2018).

All high-throughput sequencing technologies are characterized by the production of massive amounts of data in parallel (Church 2006; Schuster 2008; Knief 2014; Besser *et al.* 2018). This can be achieved through either short read methods, or long read methods, which have been defined as second and third generation technologies, respectively (Knief 2014). Second generation technologies rely on a PCR based method to sequence each genome multiple times in small random fragments. Although there are a number of technologies available, they all share a similar workflow whereby DNA is first extracted, followed by library preparation where the DNA is sheared into fragments and adaptors or barcodes are added and fragments are amplified. This is then followed by template preparation by bridge amplification or emulsion PCR after which fragments are finally subjected to automated sequencing (Vincent *et al.* 2017; Besser *et al.* 2018). Second generation technologies currently available differ greatly in terms of accuracy, engineering, output and cost (Buermans and den Dunnen 2014; Besser *et al.* 2018).

Third generation technologies target a single molecule of DNA without making use of PCR, allowing for the production of longer reads (Besser *et al.* 2018). This is particularly advantageous when sequencing complex genomic regions, such as regions with high levels of repeats, that would be difficult to assemble using short reads generated by second generation technologies (Schadt *et al.* 2010; Besser *et al.* 2018). There are currently two third generation sequencing instruments that are widely used. The first, PacBio RSII system from Pacific Biosciences makes use of single molecule, real-time sequencing technology (Rhoads and Au 2015). While this system does allow for extremely long fragments of 20kb to be sequenced in a matter of hours, run throughput is still low and costly (Besser *et al.*

2018). To alleviate this problem, PacBio have since released the Sequel, a cost effective version of the RSII with a higher throughput (Besser *et al.* 2018). The second platform, known as the MinION, was developed by Oxford Nanopore Technologies and is the first commercially available platform making use of Nanopore technology (Stoddart *et al.* 2009). This technology allows for nucleotide bases to be identified as they base through a biological or solid state pore a few nanometres in diameter (Stoddart *et al.* 2009).

## 3. MATING STRATEGIES IN FUNGI AND GENOME SEQUENCING

In Ascomycetes, sexual reproduction can take place in a number of ways (Billiard *et al.* 2011; Ni *et al.* 2011). Most basically, species either require a partner in order to reproduce or they are able to reproduce on their own in the absence of a mating partner. Understanding the mechanism by which sexual reproduction is achieved in Ascomycete fungi is largely achieved by determining the distribution of genes responsible for sexual reproduction at the mating type locus (Blakeslee 1904; Coppin *et al.* 1997; Ni *et al.* 2011).

### 3.1 Mating in Ascomycetes

Sexual reproduction in filamentous ascomycetes is mediated by the mating-type or *MAT1* locus (Billiard *et al.* 2011; Ni *et al.* 2011). This locus is comprised of several open reading frames and is flanked on either side by conserved regions (Glass *et al.* 1990; Turgeon and Yoder 2000; Wilken *et al.* 2017). The combination of genes within the locus itself determine the structure of the mating-type locus. Two allelic variants termed idiomorphs have been identified in Ascomycetes, and these are known as the *MAT1-1* and *MAT1-2* idiomorphs (Metzenberg and Glass 1990; Ni *et al.* 2011; Wilken *et al.* 2017). Genic variation within the *MAT1* locus has resulted in the description of two distinct sexual reproduction strategies, namely heterothallism and homothallism (Cisar and TeBeest 1999; Alby *et al.* 2009; Billiard *et al.* 2011).

Heterothallism and homothallism are two contrasting sexual reproductive strategies. Heterothallic fungi by definition possess only one mating-type idiomorph in their genome (Yun *et al.* 2000; Billiard *et al.* 2011). Genes in the idiomorph confer mating specificity, i.e. the *MAT1-1-1* gene in the *MAT1-1* idiomorph confers a MAT1 mating specificity while the *MAT1-2-1* gene in the *MAT1-2* idiomorph confers a MAT2 mating specificity. In order for sexual reproduction to occur, heterothallic fungi require a mating partner with the opposite but compatible mating type (Ni *et al.* 2011). Homothallic fungi, in contrast, possess all the genes required for sexual reproduction to take place in a single haploid genome and are

thus able to reproduce sexually in the absence of a partner (Cisar and TeBeest 1999; Ni *et al.* 2011).

Many variations of homothallism have been observed in fungi, and these are distinguished based on whether the genes of the mating idiomorphs are present within a single cell of multiple nuclei (Lin and Heitman 2007; Wilson *et al.* 2015b). Reproduction that occurs as a result of both idiomorphs being present in a single genome is known as primary homothallism (Lin and Heitman 2007; Ni *et al.* 2011). In this instance, genes may be linked and appear at a single locus, or unlinked and occur at separate regions within the genome (Yun *et al.* 2000). The second form of homothallism, secondary homothallism, functionally resembles heterothallism as two compatible mating partners are required for sexual reproduction to take place. However, these compatible mating types are contributed by a single individual. The production of these mating partners occurs either through pseudo-homothallism (the presence of multiple, distinct nuclei in a single cell) (Nelson 1996) or mating-type switching (the interconversion of mating types) (Perkins 1987; Coppin *et al.* 1997; Ni *et al.* 2011; Wilken *et al.* 2014).

### 3.2 Identification of Mating Types

Initially, mating types and reproductive strategies were assigned to fungi according to laboratory based mating studies (Blakeslee 1904; Lin *et al.* 2005). In these studies, fungal isolates of a single species were screened for the production of sexual structures. If these were produced by isolates derived from single spores, the fungus was determined to be homothallic (Ni *et al.* 2011). If no sexual structures were observed, isolates were co-cultured to determine whether sexual structures would form between two potentially compatible mating partners. If this resulted in successful production of sexual structures, the fungus was deemed to be heterothallic (Cisar and TeBeest 1999). In instances where no sexual structures were observed in either situation, the fungus was determined to be asexual (Taylor *et al.* 1999).

The mode of sexual reproduction determined in laboratory mating studies can be confirmed using molecular techniques. Typically, this includes identifying the mating-type genes present at the *MAT1* locus and determining their organization within the locus (Glass *et al.* 1988; Dyer *et al.* 2003; Wilken *et al.* 2017). This can be achieved by using a workflow that includes techniques such as PCR, sequencing and computational analysis (Glass *et al.* 1988; Pöggeler 2002; Dyer *et al.* 2003; Bihon *et al.* 2014; Wilken *et al.* 2014, 2017; Wilson

*et al.* 2015a). While these techniques are seemingly simple, working in the absence of genome data and reliable sequence data can be challenging.

Amplification of genes within the mating locus is possible only with adequate primer design. In the event that sequence data are not available, the mating-type loci of closely related species can be used to design degenerate primers for PCR amplification (Coppin *et al.* 1997). Additionally, these loci can provide an indication the reproductive strategy in a fungal species as closely related species are known to share similar reproductive strategies (Coppin *et al.* 1997). This can be helpful in devising a strategy to target the mating region, as significant differences exist in the nucleotide variation within the locus of heterothallic and homothallic species. In the case where a heterothallic species is suspected, the use of degenerate primers that can amplify across the complete locus is advisable. A strategy targeting the individual gene regions within the locus is advisable for homothallic species.

Once successful amplification has been achieved, the fragment can be sequenced to ensure the correct fragment has been amplified (Arie *et al.* 1997; Pöggeler 2002; Bihon *et al.* 2014; Wilken *et al.* 2014). Based on this sequence data, new primers can be designed, and the process can be repeated. Although such a degenerate primer-based PCR approach is certainly useful in instances where genome data are not available, this method can be time consuming. It also requires a large amount of optimization to correctly determine the presence of genes and their organization within the locus. With the advent of genome sequencing, identifying the mating type genes and loci has become routine and practical (Figure 1) (Galagan *et al.* 2005; Wilken *et al.* 2014, 2017).

## 4. POPULATION GENETICS OF FUNGI AND GENOME SEQUENCING

Population genetics is a subfield of genetics that deals with the genetic variation between populations (Dobzhansky 1955). Understanding this aspect of a population provides insights into the mechanism by which genetic variation is introduced into the population (Wright and Bentzen 1994). More genetically diverse populations are more difficult to manage and prevention strategies are more complicated (Ellegren 2004). Genetic markers such as microsatellite markers are commonly used to identify and quantify the genetic variation within and between populations (Ellegren 2004).

There are a number of important factors to consider when selecting a marker type to use for population genetic studies (Wright and Bentzen 1994). The sensitivity of the marker is

important to ensure that appropriate levels of information are generated (Sunnucks 2000). In many instances, markers can be identified in both nuclear and organellar DNA, which will be inherited in a biparental and a uniparental fashion, respectively (Sunnucks 2000). It is also important to consider whether multilocus DNA techniques (randomly amplified polymorphic DNA or amplified polymorphic length DNA) or single-locus DNA techniques (microsatellites or single copy nuclear DNA regions) will be best suited for the given study. Typically, single-locus techniques are favoured as they are more flexible, informative and connectable as they can be analysed using multiple techniques (Sunnucks 2000; Zane *et al.* 2002). Due to their sensitivity, presence in both nuclear and mitochondrial DNA, and single-locus analysis approach, microsatellites are frequently used in population genetics studies (Jarne and Lagoda 1996; Zane *et al.* 2002).

Microsatellites, also known as short tandem repeats or simple sequence repeats, are variable DNA fragments that consist of repeated nucleotide sequence found in both prokaryotic and eukaryotic genomes (Zane *et al.* 2002). These repeats consist of short motifs of one to six nucleotides that are tandemly repeated, and are flanked by unique sequences (Wright and Bentzen 1994; Field and Wills 1998; Richard and Paques 2000; Toth *et al.* 2000). Microsatellites can be classified into four classes depending on the nature of the repeat, namely perfect, imperfect, interrupted, or composite (Oliveira *et al.* 2006). In order for a microsatellite to be classified as a perfect microsatellite, the repeat sequence should not be interrupted by a nucleotide that does not form part of the repeat motif. In contrast, an imperfect microsatellite contains a pair of bases between the repeat motif that do not form part of the motif. An interrupted microsatellite contains a short sequence within the repeat that does not form part of the motif. Composite microsatellites are made up of two distinct adjacent repeats (Field and Wills 1998; Oliveira *et al.* 2006; Simpson *et al.* 2013).

Due to their high mutation rate as a result of slippage during replication (Ellegren 2004), and simple Mendelian mode of inheritance, microsatellite markers are often used to study of population structure, mating systems, and pedigrees. However, a major drawback of using microsatellites is that they often need to be developed *de novo* for most species (Zane *et al.* 2002). Classical methods used to develop microsatellites such as random amplification of microsatellite regions (RAMs) (Hantula *et al.* 1996) and fast amplification by AFLP of sequences containing repeats (FIASCO) (Zane *et al.* 2002), require a number of laboratory based steps. These steps include the generation of a genomic library, isolation and

sequencing of microsatellite-containing clones, optimization of PCRs for multiple primer pairs, and finally testing the primer pairs on a set of individuals.

The development of computer-based methods has dramatically reduced not only the time taken to develop microsatellite markers, but also the cost to do so. Such computer based techniques often rely on a genome sequence for the development of microsatellite markers (Zane *et al.* 2002; Simpson *et al.* 2013; Santana *et al.* 2016). There are a number of programs that can be used to do this, including MSatCommander and SSR Locator (Carlos *et al.* 2008; Faircloth 2008), and are readily available online. This software can be used to determine the location of microsatellites within the genome, and primers can then be designed for amplification. If only one genome is available, microsatellites can be tested for polymorphism in the population of individuals. This can be achieved using PCR amplification and Sanger Sequencing. This can be difficult as many of the microsatellites might not be polymorphic in the population, but this is only evident after screening in several isolates. If more than one genome of the species is available, variation within the microsatellites can be assessed *in silico* by comparing the markers across the genomes. This ultimately significantly reduces the time taken to develop microsatellite markers for population analysis (Barnes *et al.* 2001; Abdelkrim *et al.* 2009; Simpson *et al.* 2013).

## 5. DOTHIDEOMYCETES AND AVAILABLE GENOME SEQUENCES

The Dothideomycetes is the largest and most ecologically diverse group of the Ascomycetes, with over 19000 species accommodated in 1300 genera and 12 orders (Kirk *et al.* 2008; Schoch *et al.* 2009). The diversity of species in this Class is highlighted by the fact that a wide range of food and biofuel crops are infected by at least one species of the Dothideomycetes (Schoch *et al.* 2009; Ohm *et al.* 2012). The great negative economic impact and biological diversity of this group, coupled with the increased availability and affordability of whole genome sequencing has made it possible to study many plant pathogens in this group. Although all 12 Orders of the fungi in Dothideomycetes include plant pathogens, the Pleosporales and Capnodiales have a high number of destructive pathogens and have thus been the focus of numerous studies (Hunter *et al.* 2009; Schoch *et al.* 2009; Ohm *et al.* 2012; Yun *et al.* 2013) (Figure 2).

The genus *Teratosphaeria* includes species that are particularly destructive members of the Capnodiales Order (Hunter *et al.* 2011). In South Africa, three species of *Teratosphaeria* are known, all of which pose potential threats to the South African forestry industry. These

species include the stem canker pathogen, *T. zuluensis* (Jimu *et al.* 2016*;* Cortinas *et al.* 2006, 2008*;* Aylward *et al.* 2019), and the leaf and shoot pathogens *T. epicoccoides* and *T. destructans* (Taole *et al.* 2012; Greyling *et al.* 2016). Although several studies have focused on the stem canker pathogens in this group, relatively little is known about the leaf and shoot pathogens. Of particular interest is *T. destructans* as this pathogen was recently identified in South Africa (Greyling *et al.* 2016), and has previously had devastating impacts on plantations in North and South Sumatra, and China (Wingfield *et al.* 1996; Old *et al.* 2003; Burgess *et al.* 2007; Greyling *et al.* 2016).

*Teratosphaeria destructans*, formerly known as *Kirramyces destructans*, is a leaf and shoot pathogen of *Eucalyptus* trees (Wingfield *et al.* 1996). This pathogen was first isolated from *Eucalyptus* trees in North Sumatra (Wingfield *et al.* 1996), however the origin of the fungus is unknown. One possibility is that the fungus originated from Australia and was transported with the trees from their native environment to North Sumatra (Burgess *et al.* 2007). However, this theory has been re-evaluated (Andjic *et al.* 2016). Alternatively, it is possible that the pathogen originated from a Myrtaceae native to Sumatra (Wingfield *et al.* 1996; Andjic *et al.* 2011, 2016). Until recently, *T. destructans* outbreaks had not been reported outside of Sumatra, but the fungus was isolated from *Eucalyptus* trees in China, Thailand, Vietnam and Laos, as well as in Zululand, South Africa (Old *et al.* 2003; Greyling *et al.* 2016). Given the devastating nature of this pathogen, the rapid spread of *T. destructans* is concerning for eucalypt populations around the world.

Although the morphology, symptoms and hosts of *T. destructans* are known, a great deal is still not known about this pathogen when compared to what is known about other emerging pathogens. For example, the mating strategy utilized by *T. destructans* has yet to be determined, and effective population markers have not yet been developed. Both of these aspects, as well as several others, can be determined by making use of the *T. destructans* genome sequence.

## 6. CONCLUSIONS

The increased availability and accessibility of genome sequence data has undoubtedly made studying emerging plant pathogens easier, particularly in the case of devastating pathogens such as *T. destructans*. In particular, determining important biological aspects of a pathogen, such as mating strategy, has become an easy and routine procedure.

Determining these crucial aspects as early as possible could play a detrimental role in the development of prevention and management strategies of emerging pathogens.

## 7. REFERENCES


Abdelkrim J., Robertson B. C., Stanton J. A. L., Gemmell N. J., 2009 Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. Biotechniques 46: 185–192.

Alby K., Schaefer D., Bennett R. J., 2009 Homothallic and heterothallic mating in the opportunistic pathogen *Candida albicans*. Nature 460: 890–894.

Anagnostakis S. L., 2007 Chestnut Blight: The Classical Problem of an Introduced Pathogen. Mycologia 79: 23.

Anderson P. K., Cunningham A. A., Patel N. G., Morales F. J., Epstein P. R., *et al.*, 2004 Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 19: 535–544.

Andjic V., Glen M., Dell B., Wingfield M. J., Hardy G. E. S., *et al.*, 2011 *Teratosphaeria destructans* in Australia; Biosecurity threat or elusive native pathogen?

Andjic V., Maxwell A., Hardy G. E. S., Burgess T. I., 2016 New cryptic species of *Teratosphaeria* on *Eucalyptus* in Australia. IMA Fungus 7: 253–263.

Arie T., Christiansen S. K., Yoder O. C., Turgeon B. G., 1997 Efficient cloning of ascomycete mating type genes by PCR amplification of the conserved MAT HMG box. Fungal Genet. Biol. 21: 118–130.

Aylward J., Steenkamp E. T., Dreyer L. L., Roets F., Wingfield B. D., *et al.*, 2017 A plant pathology perspective of fungal genome sequencing. IMA Fungus 8: 1–15.

Aylward J., Roets F., Dreyer L. L., Wingfield M. J., 2019 Teratosphaeria stem canker of *Eucalyptus*: Two pathogens, one devastating disease. Mol. Plant Pathol. 20: 8–19.

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

Berbee M. L., 2001 The phylogeny of plant and animal pathogens in the Ascomycota. Physiol. Mol. Plant Pathol. 59: 165–187.

Besser J., Carleton H. A., Gerner-Smidt P., Lindsey R. L., Trees E., 2018 Next-generation sequencing technologies and their application to the study and control of bacterial infections. Clin. Microbiol. Infect. 24: 335–341.

Bihon W., Wingfield M. J., Slippers B., Duong T. A., Wingfield B. D., 2014 *MAT* gene idiomorphs suggest a heterothallic sexual cycle in a predominantly asexual and important pine pathogen. Fungal Genet. Biol. 62: 55–61.

Billiard S., López-Villavicencio M., Devier B., Hood M. E., Fairhead C., *et al.*, 2011 Having

sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. 86: 421–442.

Billiard S., López-Villavicencio M., Hood M. E., Giraud T., 2012 Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. J. Evol. Biol. 25: 1020–1038.

Blakeslee A. F., 1904 Sexual Reproduction in the Mucorineae. Proc. Natl. Acad. Arts Sci. 40: 205–319.

Buermans H. P. J., Dunnen J. T. den, 2014 Next generation sequencing technology: Advances and applications. Biochim. Biophys. Acta - Mol. Basis Dis. 1842: 1932–1941.

Burgess T. I., Andjic V., Wingfield M. J., Hardy G. E. S. J., 2007 The eucalypt leaf blight pathogen *Kirramyces destructans* discovered in Australia. Australas. Plant Dis. Notes 2: 141–144.

Cairns T. C., Studholme D. J., Talbot N. J., Haynes K., 2016 New and Improved Techniques for the Study of Pathogenic Fungi. Trends Microbiol. 24: 35–50.

Carlos L., Palmieri D. A., Souza V. Q. De, Kopp M. M., Carvalho D., *et al.*, 2008 SSR Locator: Tool for Simple Sequence Repeat Discovery Integrated with Primer Design and PCR Simulation. Int. J. Plant Genomics: 1–9.

Church G. M., 2006 Genomes for All. Sci. Am. 294: 46–54.

Cisar C. R., TeBeest D. O., 1999 Mating system of the filamentous ascomycete, *Glomerella cingulata*. Curr. Genet. 35: 127–133.

Coppin E., Debuchy R., Arnaise S., Picard M., 1997 Mating types and sexual development in filamentous ascomycetes. Microbiol. Mol. Biol. Rev. 61: 411–28.

Cortinas M., Crous P. W., Wingfield B. D., Wingfield M. J., 2006 Multi-gene phylogenies and phenotypic characters distinguish two species within the *Colletogloeopsis zuluensis* complex associated with *Eucalyptus* stem cankers. Stud. Mycol. 55: 133–146.

Cortinas M. N., Wingfield B. D., Wingfield M. J., 2008 Microsatellite markers for the *Eucalyptus* stem canker fungal pathogen *Kirramyces gauchensis*. Mol. Ecol. Resour. 8: 590–592.

Desprez-Loustau M. L., Courtecuisse R., Robin C., Husson C., Moreau P. A., *et al.*, 2009 Species diversity and drivers of spread of alien fungi (sensu lato) in Europe with a particular focus on France. Biol. Invasions 12: 157–172.

Dobzhansky T., 1955 A review of some fundamental concepts and problems of population genetics. Cold Spring Harb. Symp. Quant. Biol. 20: 1–15.

Dyer P. S., Paoletti M., Archer D. B., 2003 Genomics reveals sexual secrets of *Aspergillus*. Microbiology 149: 2301–2303.

Edwards A., Caskey C., 1991 Closure strategies for random DNA sequencing. Methods 3:

41–47.

Ellegren H., 2004 Microsatellites: Simple sequences with complex evolution. Nat. Rev. Genet. 5: 435–445.

Faircloth B. C., 2008 MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol. Ecol. Resour. 8: 92–94.

Field D., Wills C., 1998 Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad. Sci. United.

Galagan J. E., Hynes M., Pain A., Machida M., Purcell S., *et al.*, 2005 Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature 438: 1105–1115.

Gladieux P., Feurtey A., Hood M., Snirc A., Clavel J., *et al.*, 2015 The population biology of fungal invasions. Mol Ecol 24: 1969–1986.

Glass N. L., Vollmer S. J., Staben C., Grotelueschen J., Metzenberg R. L., *et al.*, 1988 DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar. Science (80- . ). 241: 570–573.

Glass N. L., Grotelueschen J., Metzenberg R. L., 1990 *Neurospora crassa*: A mating-type region. Proc. Natl. Acad. Sci. U. S. A. 87: 4912–4916.

Greyling I., Wingfield M. J., Coetzee M. P., Marincowitz S., Roux J., 2016 The *Eucalyptus* shoot and leaf pathogen *Teratosphaeria destructans* recorded in South Africa. South. For. a J. For. Sci. 78: 123–129.

Hantula J., Dusabenyagasani M., Hamelin R. C., 1996 Random amplified microsatellites (RAMS) — a novel method for characterizing genetic variation within fungi. For. Pathol. 26: 159–166.

Heitman J., Sun S., James T. Y., 2013 Evolution of fungal sexual reproduction. Mycologia 105: 1–27.

Hunter G. C., Crous P. W., Carnegie A. J., Wingfield M. J., 2009 *Teratosphaeria nubilosa*, a serious leaf disease pathogen of *Eucalyptus* spp. in native and introduced areas. Mol. Plant Pathol. 10: 1–14.

Hunter G. C., Crous P. W., Carnegie A. J., Burgess T. I., Wingfield M. J., 2011 Mycosphaerella and Teratosphaeria diseases of *Eucalyptus*; easily confused and with serious consequences. Fungal Divers. 50: 145–166.

Jarne P., Lagoda P. J. L., 1996 Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11: 424–429.

Jimu L., Chen S., Wingfield M. J., Mwenje E., Roux J., 2016 Three genetic groups of the *Eucalyptus* stem canker pathogen *Teratosphaeria zuluensis* introduced into Africa from an unknown source. Antonie Van Leeuwenhoek 109: 21–33.

Kirk P., Cannon P., Minter D., Stalpers L., 2008 *Ainsworth and Bisby's dictionary of the Fungi*. CAB International, Wallingford, UK.

Knief C., 2014 Analysis of plant microbe interactions in the era of next generation sequencing technologies. Front. Plant Sci. 5: 1–24.

Lee S. C., Ni M., Li W., Shertz C., Heitman J., 2010 The Evolution of Sex: a Perspective from the Fungal Kingdom. Microbiol. Mol. Biol. Rev. 74: 298–340.

Lin X., Hull C. M., Heitman J., 2005 Sexual Reproduction between partners of the same mating type in *Cryptococcus neoformans*. Nature 434: 1017–1021.

Lin X., Heitman J., 2007 Mechanisms of homothallism in fungi and transitions between heterothallism and homothallism. In: Heitman J, Kronstad J, Taylor J, Casselton L (Eds.), *Sex in Fungi: Molecular determination and evolutionary implication*, American Society for Microbiology Press, Washington DC, pp. 35–57.

Lione G., Danti R., Fernandez-Conradi P., Ferreira-Cardoso J. V., Lefort F., *et al.*, 2018 The emerging pathogen of chestnut *Gnomoniopsis castaneae*: The challenge posed by a versatile fungus. Eur. J. Plant Pathol.: 1–15.

Lodge A. H., 1991 *Ophiostoma novo-ulmi sp. nov.*, causative agent of current Dutch elm disease pandemics. Mycopathologia 115: 151–161.

Metzenberg R. L., Glass N. L., 1990 Mating type and mating strategies in *Neurospora*. Bioassays 12: 53–59.

Nelson M. A., 1996 Mating systems in ascomycetes: A romp in the sac. Trends Genet. 12: 69–74.

Ni M., Feretzaki S. S., Wang X., Heitman J., 2011 Sex in Fungi. Annu. Rev. Genet. 45: 405–430.

Ohm R. A., Horwitz B. A., Zhong S., Lucas S., Barry K. W., *et al.*, 2012 Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. PLoS Pathog. 8: e1003037.

Old K. M., Pongpanich K., Thru P. Q., Wingfield M. J., Yuan Z. Q., 2003 *Phaeophleospora destructans* causing leaf blight epidemics in South East Asia. In: *8th International Congress of Plant Pathology*, Christchurch, New Zealand, p. 165.

Oliveira E. J., Padua J. G., Zucchi M. I., Vencovsky R., Vieira M. L. C., 2006 Origin, evolution and genome distribution of microsatellites. Genet. Mol. Biol. 29: 294–307.

Perkins D. D., 1987 Mating-type switching in filamentous ascomycetes. Genetics 115: 215–

216.

Picco A. M., Angelini P., Ciccarone C., Franceschini A., Ragazzi A., *et al.*, 2011 Biodiversity of emerging pathogenic and invasive fungi in plants, animals and humans in Italy. Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol. 145: 988–996.

Pöggeler S., 2002 Genomic evidence for mating abilities in the asexual pathogen *Aspergillus fumigatus*. Curr. Genet. 42: 153–160.

Rhoads A., Au K. F., 2015 PacBio Sequencing and Its Applications. Genomics, Proteomics Bioinforma. 13: 278–289.

Richard G. F., Paques F., 2000 Mini- and microsatellite expansions: The recombination connection. EMBO Rep. 11: 122–126.

Sakai A. K., Allendorf F. W., Holt J. S., Lodge D. M., Molofsky J., *et al.*, 2001 The Population Biology of Invasive Species. Annu. Rev. Ecol. Syst. 32: 305–332.

Santana Q. C., Coetzee M. P. A., Wingfield B. D., Wingfield M. J., Steenkamp E. T., 2016 Nursery-linked plantation outbreaks and evidence for multiple introductions of the pitch canker pathogen *Fusarium circinatum* into South Africa. Plant Pathol. 65: 357–368.

Saunders D. G. O., 2015 Hitchhiker's guide to multi-dimensional plant pathology. New Phytol. 205: 1028–1033.

Schadt E. E., Turner S., Kasarskis A., 2010 A window into third-generation sequencing. Hum. Mol. Genet. 19: R227–R240.

Schoch C. L., Crous P. W., Groenewald J. Z., Boehm E. W. A., Burgess T. I., *et al.*, 2009 A class-wide phylogenetic assessment of Dothideomycetes. Stud. Mycol. 64: 1–15.

Schuster S. C., 2008 Next-generation sequencing transforms today's biology. Nat. Methods 5: 16–18.

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

Staden R., 1979 A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 6: 2601–2610.

Stoddart D., Heron A. J., Mikhailova E., Maglia G., Bayley H., 2009 Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. Proc. Natl. Acad. Sci. 106: 7702 LP-7707.

Sunnucks P., 2000 Efficient genetic markers for population biology. Tree 15: 199–203.

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

Taylor J. W., Jacobson D. J., Fisher M. C., 1999 The evolution of asexual fungi: Reproduction, speciation and classification. Annu. Rev. Phytopathol. 37: 197–246.

Toth G., Gaspari Z., Jurka J., 2000 Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res. 10: 967–981.

Turgeon B. G., Yoder O. C., 2000 Proposed Nomenclature for Mating Type Genes of Filamentous Ascomycetes. Fungal Genet. Biol. 31: 1–5.

Vincent A. T., Derome N., Boyle B., Culley A. I., Charette S. J., 2017 Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. J. Microbiol. Methods 138: 60–71.

Vurro M., Vannacci G., Pisa U., 2010 Emerging infectious diseases of crop plants in developing countries: Impact on agriculture and socio-economic consequences. Food Secur. 2: 113–132.

Wilken P. M., Steenkamp E. T., Wingfield M. J., Beer Z. W. De, Wingfield B. D., 2014 DNA loss at the *Ceratocystis fimbriata* mating locus results in self-sterility. PLoS One 9: e92180.

Wilken P. M., Steenkamp E. T., Wingfield M. J., Beer Z. W. de, Wingfield B. D., 2017 Which MAT gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. Fungal Biol. Rev. 31: 199–211.

Wilson A. M., Godlonton T., Nest M. A. Van Der, Wilken P. M., Wingfield M. J., *et al.*, 2015a Unisexual Reproduction in *Huntiella moniliformis*. Fungal Genet. Biol. 80: 1–9.

Wilson A. M., Wilken P. M., Nest M. A. van der, Steenkamp E. T., Wingfield M. J., *et al.*, 2015b Homothallism: an umbrella term for describing diverse sexual behaviours. IMA Fungus 6: 207–214.

Wingfield M. J., Crous P. W., Boden D., 1996 *Kirramyces destructans sp. nov.*, a serious leaf pathogen of *Eucalyptus* in Indonesia. South African J. Bot. 62: 325–327.

Wingfield M. J., Slippers B., Wingfield B. D., 2017 The unified framework for biological invasions: A forest fungal pathogen perspective. Biol. Invasions 19: 3201–3214.

Wright J. M., Bentzen P., 1994 Microsatellites: Genetic markers for the future. Rev. Fish Biol. Fish. 4: 384–388.

Yun S.-H., Arie T., Kaneko I., Yoder O., Turgeon B. G., 2000 Molecular organization of mating type loci in heterothallic, homothallic, and asexual *Gibberella/Fusarium* species. Fungal Genet. Biol. 31: 7–20.

Yun S. H., Yoder O. C., Gillian Turgeon B., 2013 Structure and function of the mating-type locus in the homothallic ascomycete, *Didymella zeae-maydis*. J. Microbiol. 51: 814–820.

Zane L., Bargelloni L., Patarnello T., 2002 Strategies for microsatellite isolation: A review. Mol. Ecol. 11: 1–16.

# Chapter 1

## Figures

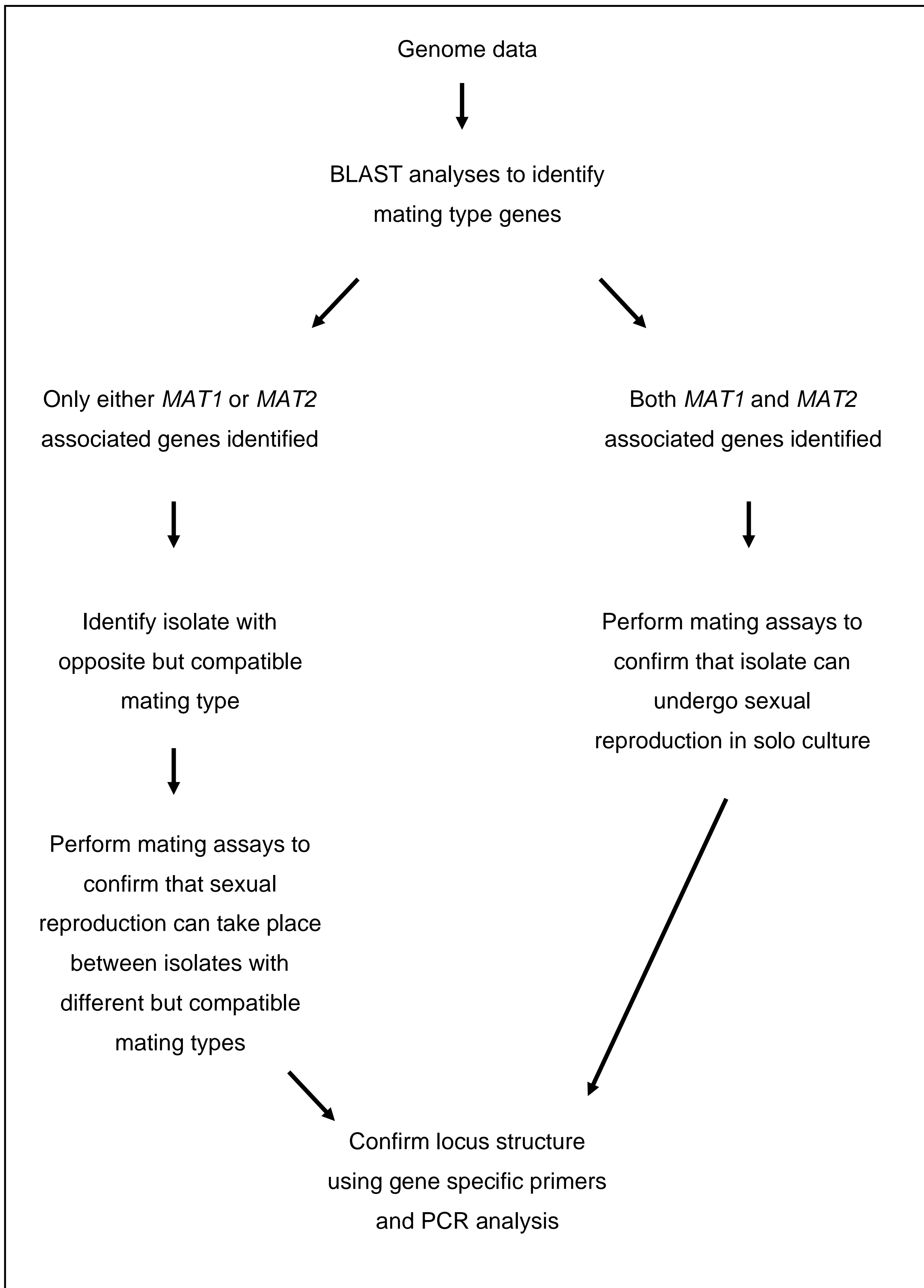


**Figure 1** The process undertaken to determine the mating strategy utilized by an Ascomycete plant pathogen.

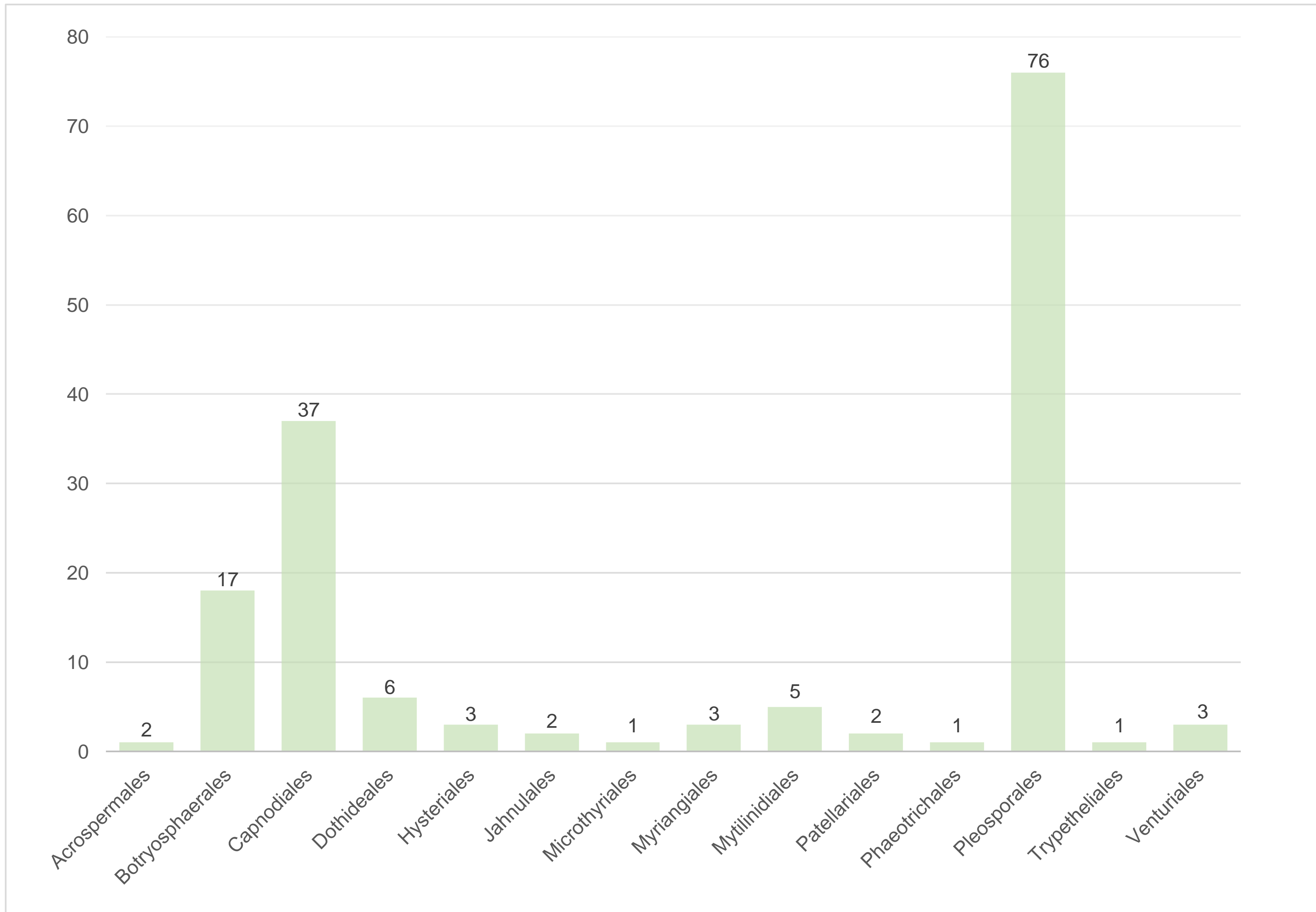


**Figure 2** The distribution of complete and permanent draft genome sequences of unique species available for the Dothideomycetes based on the Genomes Online Database as of the 25 May 2019.

# Chapter 2

## The *MAT1* locus structure of *Teratosphaeria destructans* indicates a Heterothallic Mating Strategy

## The *MAT1* locus structure of *Teratosphaeria destructans* indicates a Heterothallic Mating Strategy


### ABSTRACT

Mating strategies in fungi are governed by genes located within the mating-type (*MAT1*) locus. These strategies can be determined by performing mating assays in a laboratory environment, or by using molecular methods to identify which genes are present at the *MAT1* locus. The decrease in the cost of whole genome sequencing has resulted in many more fungal genomes becoming available. This in turn has meant that characterising the *MAT1* locus is a reasonable strategy to assist in determining the mating strategy of a fungus. The aim of this study was to elucidate the mating strategy of the damaging *Eucalyptus* shoot and leaf pathogen *Teratosphaeria destructans* by utilizing the genomes of two *T. destructans* isolates. In addition, the distribution of mating-type genes in isolates of the pathogen collected in North Sumatra, South Sumatra and South Africa was determined. To achieve this goal, the two genomes were sequenced and assembled and mating-type genes within each isolate were identified through BLAST analyses. Primers to amplify mating type genes were designed and used to screen isolates from the above mentioned locations. The results revealed two mating idiomorphs (the *MAT1-1* idiomorph harbouring the *MAT1-1-1* gene, and the *MAT1-2* idiomorph harbouring the *MAT1-2-1* and an additional mating-type gene, named *MAT1-2-12*), suggesting that *T. destructans* is a heterothallic fungus. Furthermore, both mating types were present in isolates collected in North Sumatra, the area where the pathogen was first collected and described as well as in South Sumatra. This suggests that the pathogen was introduced into its invasive range with sufficient diversity to include two mating types. In contrast, only one mating type was identified within the South African isolates tested, suggesting a single recent introduction.

## 1. INTRODUCTION

Ascomycete fungi can be classified as utilizing one of two reproductive strategies, either heterothallism or homothallism, based on their mating-type (*MAT1*) locus (Coppin *et al.* 1997; Billiard *et al.* 2011; Wilken *et al.* 2017). This locus is comprised of a number of open reading frames (ORFs), some of which encode for transcription factors that regulate the expression of downstream sex-related genes (Kronstad and Staben 1997; Debuchy *et al.* 2010; Wilson *et al.* 2015b; Wilken *et al.* 2018). Expression of all mating genes is required for sexual reproduction to occur (Coppin *et al.* 1997). This has led to Ascomycete mating systems being classified based on the gene content and structure of the *MAT1* locus (Coppin *et al.* 1997; Wilken *et al.* 2017).

Heterothallism (or self-sterility) refers to a system where two compatible mating partners are required for sexual reproduction to proceed as each mating parent possess only a single *MAT1* idiomorph (Kronstad and Staben 1997). In this instance, there are two non-allelic gene combinations known for *MAT1* locus, and these are referred to as the *MAT1-1* and *MAT1-2* idiomorphs (Metzenberg and Glass 1990). In contrast, self-fertility, also known as homothallism is defined as the ability of a single fungal isolate to reproduce sexually in solo culture because all genes required for reproduction are present in a single haploid genome (Kronstad and Staben 1997). Homothallism has been characterized into a several sub-classes that have recently been reviewed (Wilson *et al.* 2015b). While the genes located within the *MAT1* locus vary, the genes flanking this region are known to remain consistent (Wilken *et al.* 2017).

Originally the mating strategy of a fungus was determined by observing sexual reproduction under artificial conditions (Cisar and TeBeest 1999; Ni *et al.* 2011). This is most often accomplished in a laboratory environment using mating assays and monitoring for the production of sexual structures known as ascomata (Cisar and TeBeest 1999). Fungi that are able to produce ascomata in cultures derived from single spores are hypothesized to be homothallic, while those that require the co-incubation of isolates of opposite mating type for ascomata production are considered to be heterothallic (Cisar and TeBeest 1999; Ni *et al.* 2011). Although these artificial mating assays remain widely used, they do not work with many fungi and molecular methods that are less time consuming and labour intensive are becoming more prominent.

Molecular methods are rapidly emerging as the technique of choice for inferring mating strategies in fungi (Bihon *et al.* 2014; Wilken *et al.* 2014; Wilson *et al.* 2015a; Yin *et al.* 2017; Kanzi *et al.* 2019) . In some cases, this can be done by using gene specific primers from closely related species, combined with a "primer walking" approach (Wilken *et al.* 2012; De Miccolis Angelini *et al.* 2016). In this approach, primers matching a known sequence flanking the unknown sequence are used to amplify a small region of unknown sequence. The resulting PCR product can be sequenced and used to design primers to amplify the next portion of unknown sequence. This process is repeated until the full fragment has been sequenced. This method relies on prior knowledge of the mating-type locus structure of a closely related species, the assumption of high sequence homology between the species as well as extensive optimization during the "primer walking" process. The process has become considerably simpler where full genome sequences are available. Local BLAST searches using known mating-type genes from closely, or even distantly related species can be used to evaluate genes present at the *MAT1* locus. The results can then be confirmed using conventional PCR and sequencing (Bihon *et al.* 2014; Wilson *et al.* 2015a; Wilken *et al.* 2017).

The aim of this study was to infer the mating strategy of the *Eucalyptus* pathogen *Teratosphaeria destructans* using whole genome sequencing. This Dothideomycete fungus is a devastating leaf and shoot pathogen of commercially planted, non-native *Eucalyptus* (Old *et al.* 1996; Wingfield *et al.* 1996; Burgess *et al.* 2007; Andjic *et al.* 2011b; Greyling *et al.* 2016). Since its discovery in North Sumatra and subsequent description two decades ago (Wingfield *et al.* 1996), *T. destructans* has expanded its geographic range throughout South East Asia into South Sumatra, China and Vietnam (Old *et al.* 1996; Wingfield *et al.* 1996; Burgess *et al.* 2007; Greyling *et al.* 2016). Most recently, this pathogen was reported in the sub-tropical regions of South Africa (Greyling *et al.* 2016). Very little is known regarding the biology of *T.* destructans and the study included three core objectives. These were 1) to sequence and assemble two genomes of *T. destructans*; 2) to identify the mating strategy utilized by *T. destructans*; and finally, 3) to determine the geographic distribution of mating types among isolates of this pathogen for isolates collected from three geographic areas.

# 2. MATERIALS AND METHODS

## 2.1 Genome sequencing and annotation

Two isolates of *T. destructans* (CMW 45982 and CMW 45661) were sourced from the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria. These isolates had been collected from North and South Sumatra, respectively (Table 1). Cultures were grown on 2% malt extract agar supplemented with streptomycin (MEA-S; 20 g/L malt extract (Biolab, Merck), 20 g/L agar (Biolab, Merck), 150 mg/L streptomycin (Biolab, Merck)) for 4 weeks. Mycelia were crushed in 1 mL of 2% malt extract liquid media (20 g/L malt extract (Biolab, Merck)). The resulting suspension was transferred to 250mL flasks containing 50 mL of 2% malt extract liquid media. Liquid cultures were incubated at 25˚C for 2 weeks, after which the resulting mycelial growth was plated on 2% MEA-S. The cultures were incubated at 25˚C for 8 weeks before mycelia were harvested, freeze dried and used for DNA extractions.

DNA was extracted from lyophilized fungal mycelia using a protocol described by Brunner *et al.* (2001). Following extraction, the resulting DNA pellet was resuspended in 100 µL TE (10 mM tris(hydroxymethyl)aminomethane-HCl pH 8.0 (Sigma-Aldrich, Missouri, USA); 1 mM ethylenediaminetetraacetic acid pH 8.0 (Sigma-Aldrich, Missouri, USA)) buffer. The DNA samples were stained using Biotium GelRed Nucleic Acid Stain (Anatech, Johannesburg, South Africa) before being separated and analysed using 2% (w/v) agarose gel electrophoresis, and visualized using the Bio-Rad Molecular Imager Gel Doc XR+ (Bio-Rad, California, USA) to assess their quality. The DNA quantity was measured fluorometrically using the Qubit instrument and dsDNA High Sensitivity (HS) Assay Kit (ThermoFisher Scientific, Massachusetts, USA).

These samples were submitted to Macrogen Korea for whole genome sequencing. A pair-end library with an insert size of 350 bp was generated using the TruSeq DNA PCR-free library preparation kit (Illumina, California, USA) for isolate CMW 45982. Additionally, a mate-pair library with an insert size of 5 kb was prepared for this isolate using the Nextera mate-pair library preparation kit (Illumina, California, USA). For isolate CMW 45661 a mate-pair library with an insert size of 5 kb was generated also using the Nextera mate-pair library preparation kit. Sequencing was conducted using the Illumina HiSeq 2500 Platform (California, USA). The resulting reads were subjected to quality filtering and trimming using

default parameters before assemblies were generated using CLC Genomics (version 10.0.1; CLC Bio, Denmark). For isolate CMW 45982 a *de novo* assembly was conducted using the pair-end reads as the input, and the mate-pair reads as the guidance reads. For isolate CMW 45661, only a subset of the 20 million longest reads was used for the *de novo* assembly. For both assemblies, automatic bubble and word size were used, with a mismatch cost of 2, an insertion cost of 3 and a deletion cost of 3. Contigs were joined using the Join Contigs function in the Genome Finishing Module available for CLC Genomics (Supplementary Tables 1 and 2).

### 2.2 Reconstruction of the *MAT1* locus

Local BLAST searches were used to identify the contigs on which the mating-type genes were located. Protein sequences for the *MAT1-1-1*, and *MAT1-2-1* genes of *Zymoseptoria tritici* (Table 2) were used as query sequences in a tBLASTn search against all the contigs of the two *T. destructans* genomes. In addition protein sequences for the DNA Lyase and Cytochrome C Oxidase genes from two *Zymoseptoria* species were used as query sequences as these genes have been shown to be closely associated with the mating-type locus (Wilken *et al.* 2017). All contigs identified through these searches were annotated using the online platform of Augustus (augustus.gobics.de) (Stanke and Morganstern 2005) as well as FgenesH+ (softberry.com/berry.phtml) (Solovyev *et al.* 2006). Predicted open reading frames were assigned putative identities by comparison to known protein sequences in the NCBI Fungal Database using BLASTn. Additionally, the open reading frames were converted to protein sequences using CLC Genomics and were analysed for the presence of conserved domains using the online interface of InterProScan version 72.0 (ebi.ac.uk/interpro) (Jones *et al.* 2014).

### 2.3 Geographic distribution of mating types

To determine the distribution of the mating types across three sub-populations of *T. destructans*, 16 isolates from North Sumatra (5), South Sumatra (9), and South Africa (2) were obtained from the CMW Culture Collection (Table 1). These isolates were grown on 2% MEA at 25˚C, before being subjected to DNA extractions used above. The quality of the extracted nucleic acid was assessed using 2% (w/v) agarose gel electrophoresis, before being used in PCR reactions targeting the various mating-type genes.

Primer pairs (Table 3) targeting the putative *MAT1* genes as well as the genes flanking the locus were designed using the online primer design tool Primer3web version 4.1.0 (http://primer3.ut.ee/) (Koressaar and Remm 2007; Untergasser *et al.* 2012). Each primer set was used in a 25 μL PCR reaction mixture that contained 1 U SuperTherm Polymerase (Hoffman-La Roche, Switzerland), 1 × SuperTherm reaction buffer, 2.5 mM $MgCl_2$, 0.25 mM of each dNTP, 0.2 mM of each primer and 20 – 50 ng of template DNA. Amplification was performed in the Eppendorf Thermal Cycler (Eppendorf, Hamburg, Germany) using a 5 minute initial denaturation step at 94˚C, followed by 35 cycles of denaturation for 30 seconds at 94˚C, annealing for 30 seconds at the annealing temperature corresponding to the primer pair, and extension for 1 minute at 68˚C, followed by a single extension step of 7 minutes at 68˚C. PCR products were stained using Biotium GelRed Nucleic Acid Stain (Anatech, Johannesburg, South Africa) before being separated using 2% (w/v) agarose gel electrophoresis, and visualized using the Bio-Rad Molecular Imager Gel Doc XR+. Selected amplicons were purified using Sephadex (Sigma-Aldrich, Missouri, USA), and sequenced using the original PCR primers, the Big Dye cycle sequencing kit with Amplitaq DNA polymerase (Perkin-Elmer, Warrington, USA) and an ABI PRISM 3300 Genetic Analyser (Applied Biosystems, Foster City, USA). All chromatograms were visualized and analysed using CLC Genomics Workbench.

## 3. RESULTS

### 3.1 Genome sequencing and annotation

Sequencing of the genomes of two *T. destructans* isolates (CMW 45982 and CMW 45661) produced reads with an average length of 113 bp and 150 bp, respectively. For isolate CMW 45982, the 350 bp pair end library contained 123 311 016 reads and the 5 kb mate-pair library contained 43 786 484 reads. After trimming, 122 703 767 and 21 891 505 reads remained in each library, respectively. The 5 kb mate-pair library for isolate CMW 45661 contained 391 352 348 reads, of which 391 352 327 remained after trimming. Twenty million of the longest trimmed reads were selected for use in the genome assembly. Initial assemblies yielded 867 contigs for isolate CMW 45982 and 2 689 contigs for isolate CMW 45661. These contigs were ordered into 799 and 2 356 scaffolds, respectively, using multiple rounds of scaffolding through the Genome Finishing Module in CLC Genomics Workbench (Supplementary Tables 1 and 2). The genome for CMW 45982 was calculated to be 27.6 Mb in size, while the genome for CMW 45661 was calculated to be 28.8 Mb.

### 3.2 Reconstruction of the *MAT1* locus

Local tBLASTn searches using mating-type genes and genes flanking the *MAT1* locus revealed a single contig that putatively contained mating-type genes from each of the two *T. destructans* genomes. Gene annotations of these contigs predicted four and five open reading frames for isolates CMW 45661 and CMW 45982, respectively. When compared to the NCBI database using BLASTn searches, it was found that many of the predicted ORFs were similar to known mating type genes previously described in the Ascomycetes (Table 4). Both isolates were found to possess open reading frames similar to DNA Lyase from *Hortaea werneckii*, a hypothetical protein from *Baudoinia panamericana* and Cytochrome C Oxidase from *Exophiala dermatitidis.* A *MAT1-1-1* gene similar to that of *Hortaea werneckii* was identified in isolate CMW 45661, while isolate CMW 45982 was found to possess a *MAT1-2-1* gene similar to that of *Cercospora zinea* and an additional hypothetical protein similar to one found in *Rhodotorula toruloides* (Table 4).

The InterProScan analysis identified conserved domains among several of the predicted protein coding sequences (Table 5). The domains present in DNA Lyase, Cytochrome C Oxidase, *MAT1-1-1* and *MAT1-2-1* were characteristic of these proteins (Wilken *et al.* 2017). For the hypothetical protein flanking the *MAT* idiomorphs of both isolates, a Synaptotagmin-Like Mitochondrial-Lipid-Binding domain was predicted. Although characterized proteins containing this domain have not been associated with the mating-type locus before, this domain has been implicated in controlling the transfer of lipids at membrane contact sites (Schauder *et al.* 2014). In contrast, no domain was predicted for the additional hypothetical protein identified in CMW 45982.

The length of the Augustus predicted *MAT1-1-1* gene of *T. destructans* predicted by Augustus was 2643 bp, and that of the *MAT1-2-1* gene was 1620 bp. The predicted *MAT1-1-1* ORF had seven introns and encoded for a protein of 695 amino acids in size (Figure 1). In contrast, the *MAT1-2-1* ORF had three introns, encoding for a 496 amino acid protein (Figure 1). An additional ORF encoding a hypothetical protein was identified alongside the *MAT1-2-1* gene but not the *MAT1-1-1* gene. This ORF was 909 bp in length and contained no introns (Figure 1). The presence and location of this hypothetical protein suggests that this ORF encodes a mating-type gene. The apparent lack of similarity between this ORF and that of previously described *MAT1-2* genes available on the NCBI Fungal Database

implies that this is a novel *MAT1-2* gene, and is the twelfth *MAT1-2* gene to be identified (Wilken *et al.* 2017). Thus, this gene should be referred to as *MAT1-2-12*. Downstream of the mating-type genes in both isolates a DNA Lyase gene of 1900 bp in size was identified. This gene contained one intron 58 bp in size. An additional two ORFs, one encoding a hypothetical protein, and one encoding Cytochrome C Oxidase, were identified upstream from the mating type genes in both isolates (Figure 1).

The presence of two mating-type idiomorphs, as well as corresponding flanking regions for *T. destructans* made it possible to define the boundaries of the *MAT1* locus (Figure 2). The *MAT1-1* idiomorph present in isolate CMW 45661 was 4528 bp in size and included only the *MAT1-1-1* gene. In contrast, the *MAT1-2* idiomorph identified in CMW 45892 was 4056 bp in size and contained the *MAT1-2-1* gene as well as the novel *MAT1-2* gene. Sequence similarity between the two isolates was restricted to the flanking regions of the locus, and no significant sequence similarity was observed between the mating type idiomorphs (Figure 2).

### 3.3 Geographic distribution of mating types

Using gene-specific primers, fragments of the DNA Lyase, Cytochrome C Oxidase, and the ORF encoding the hypothetical protein flanking the *MAT1* locus were amplified in all isolates used in this study. The primers designed to amplify regions of the *MAT1* locus produced fragments of the expected size in many of the *T. destructans* isolates tested. Isolates were found to possess either the *MAT1-1-1* gene, or both the *MAT1-2-1* gene and the novel *MAT1-2-12* gene (Figure 3). Both mating types were found in North Sumatra and South Sumatra, while only the MAT1-2 mating type was identified in South Africa (Table 6).

## 4. DISCUSSION

This study has provided insight into the reproductive strategy utilised by the destructive eucalypt pathogen *T. destructans*. The structure and gene content of the mating-type locus revealed the presence of two idiomorphs with variable gene content, supporting heterothallism as its mating strategy (Metzenberg and Glass 1990; Billiard *et al.* 2012; Wilken *et al.* 2017). Using PCR, these two idiomorphs representing the mating types of a heterothallic species were identified in isolates of *T. destructans* from North and South Sumatra indicating that sexual reproduction could be possible in these regions of Sumatra. In contrast only a single mating type was identified in South Africa inferring that sexual reproduction would not be possible in this region.

The *MAT1* locus of *T. destructans* had a similar structure to that of other heterothallic Ascomycetes (Bihon *et al.* 2014; De Miccolis Angelini *et al.* 2016), although there were noticeable differences. The *MAT1-1* idiomorph was found to possess a single matingtype gene, the *MAT1-1-1* gene. This gene defines the *MAT1-1* idiomorph (Turgeon and Yoder 2000; Wilken *et al.* 2017), and has been identified in other heterothallic Dothideomycetes (Bihon *et al.* 2014). The *MAT1-2* idiomorph in *T. destructans* was found to possess two genes. These were the *MAT1-2-1* gene as well as an additional gene encoding a hypothetical protein. As with the *MAT1-1-1* gene, the *MAT1-2-1* gene defines the *MAT1-2* idiomorph (Wilken *et al.* 2017). This additional gene located in the mating-type locus encoding for a hypothetical protein is classified as a mating-type gene due to its presence in only one of the idiomorphs (Wilken *et al.* 2017). This is a novel mating-type gene and we propose that it is named *MAT1-2-12*; thus the twelfth mating type gene described within the *MAT1-2* idiomorph in line with the mating system proposed by (Turgeon and Yoder 2000) and reviewed recently by (Wilken *et al.* 2017). As no domain was identified when analysed with InterProScan (Table 5), further analysis would need to be conducted to determine the function of this new mating-type gene.

The native range of *T. destructans* is currently unknown (Wingfield *et al.* 1996; Greyling *et al.* 2016), although the host specificity to *Eucalyptus* fuels speculation that the pathogen and host trees share the native range of Australia (Andjic *et al.* 2016). Consequently, as the isolates considered in this study were from areas outside of Australia, it is likely that the pathogen has been accidently introduced into these regions (Wingfield *et al.* 1996). Introductions of this pathogen have been attributed to the movement of *Eucalyptus*

germplasm used in the establishment of eucalypt nurseries and plantations (Andjic *et al.* 2011a). The results of this study showed that isolates with differing mating types are present in the same geographic region. This could have resulted from multiple introductions of isolates over time, as observed with other *Teratosphaeria* species (Taole *et al.* 2015), or a single introduction of multiple isolates. Further analysis of the genetic diversity within individual populations would be required to asses this further.

The presence of both mating idiomorphs in populations from North and South Sumatra suggest that *T. destructans* should be capable of sexual reproduction in these regions. This would have significant implications for the generation of genetic variation in these introduced populations, with more variation increasing the adaptability of the species to unfavourable environments, complicating efforts to manage the disease that it causes (Lee *et al.* 2010; Billiard *et al.* 2011). Although *T. destructans* might be capable of sexual reproduction in these regions based on both mating types are present in these populations, sexual structures for this fungus have not yet been identified. Cryptic sexual reproduction has been hypothesized to occur in other heterothallic Dothideomycetes (Groenewald *et al.* 2006; Bihon *et al.* 2014), and it is possible that this could also be true for *T. destructans*. For the time being, it is known only in its asexual state in which the mitosporic conidia are produced prolifically (Wingfield *et al.* 1996).

Although only a small sample size was tested in this study, the results suggest that only a single mating type (MAT1-2) is present in South Africa. This finding suggests that *T. destructans* could have been introduced as a single isolate in a single introduction event. Analysis of genetic variation within the South African population would provide insight into how long ago this introduction took place (Taole *et al.* 2015; Wingfield *et al.* 2017), and if it was in fact a single introduction of a single isolate. Furthermore, increasing the sample size and evaluating the mating type of more isolates in South Africa would provide clarification on how many mating types are present in this region.

Molecular genetics tools, particularly whole genome sequencing, are increasingly being used to determine mating strategies within Ascomycetes (Bihon *et al.* 2014; Wilson *et al.* 2015a; Aylward *et al.* 2017; Wilken *et al.* 2018). This study is the first to focus on mating strategies in *Teratosphaeria*, a group of diverse and pathogenic fungi. The results yielded the structure of the mating-type locus of heterothallic *T. destructans*, as well as identified a novel gene within the *MAT1-2* idiomorph. Not only do these findings provide insights into

the reproductive strategies utilized by the Teratosphaeriaceae, but also provides further understanding regarding mating in the Dothideomycetes.

## 5. REFERENCES


Andjic V., Dell B., Barber P., Hardy G., Wingfield M. J., *et al.*, 2011a Plants for planting; indirect evidence for the movement of a serious forest pathogen, *Teratosphaeria destructans*, in Asia. Eur. J. Plant Pathol. 131: 49–58.

Andjic V., Glen M., Dell B., Wingfield M. J., Hardy G. E. S., *et al.*, 2011b *Teratosphaeria destructans* in Australia*; Biosecurity threat or elusive native pathogen?*

Andjic V., Maxwell A., Hardy G. E. S., Burgess T. I., 2016 New cryptic species of *Teratosphaeria* on *Eucalyptus* in Australia. IMA Fungus 7: 253–263.

Aylward J., Steenkamp E. T., Dreyer L. L., Roets F., Wingfield B. D., *et al.*, 2017 A plant pathology perspective of fungal genome sequencing. IMA Fungus 8: 1–15.

Bihon W., Wingfield M. J., Slippers B., Duong T. A., Wingfield B. D., 2014 MAT gene idiomorphs suggest a heterothallic sexual cycle in a predominantly asexual and important pine pathogen. Fungal Genet. Biol. 62: 55–61.

Billiard S., López-Villavicencio M., Devier B., Hood M. E., Fairhead C., *et al.*, 2011 Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biol. Rev. 86: 421–442.

Billiard S., López-Villavicencio M., Hood M. E., Giraud T., 2012 Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. J. Evol. Biol. 25: 1020–1038.

Brunner I., Brodbeck S., Buchler U., Sperisen C., 2001 Molecular identification of fine roots of trees from the Alps: reliable and fast DNA extraction and PCR – RFLP analyses of plastid DNA. Mol. Ecol. 10: 2079–2087.

Burgess T. I., Andjic V., Wingfield M. J., Hardy G. E. S. J., 2007 The eucalypt leaf blight pathogen *Kirramyces destructans* discovered in Australia. Australas. Plant Dis. Notes 2: 141–144.

Cisar C. R., TeBeest D. O., 1999 Mating system of the filamentous ascomycete, *Glomerella cingulata*. Curr. Genet. 35: 127–133.

Coppin E., Debuchy R., Arnaise S., Picard M., 1997 Mating types and sexual development in filamentous ascomycetes. Microbiol. Mol. Biol. Rev. 61: 411–28.

Debuchy R., Berteaux-lecellier V., Silar P., 2010 Mating Systems and Sexual Morphogenesis in Ascomycetes. In: *Sexual Development*,, pp. 501–511.

De Miccolis Angelini R. M., Rotolo C., Pollastro S., Faretra F., 2016 Molecular analysis of the mating-type (*MAT1*) locus in strains of the heterothallic ascomycete *Botrytis cinerea*. Plant Pathol. 65: 1321–1332.

Grandaubert J., Bhattacharyya A., Stukenbrock E. H., 2015 RNA-seq based gene annotation and comparative genomics of four fungal grass pathogens in the genus *Zymoseptoria* identify novel orphan genes and species-specific invasions of transposable elements. G3 5: 1323–1333.

Greyling I., Wingfield M. J., Coetzee M. P., Marincowitz S., Roux J., 2016 The *Eucalyptus* shoot and leaf pathogen *Teratosphaeria destructans* recorded in South Africa. South. For. a J. For. Sci. 78: 123–129.

Groenewald M., Groenewald J. Z., Harrington T. C., Abeln E. C. A., Crous P. W., 2006 Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of cryptic sex. Fungal Genet. Biol. 43: 813–825.

Jones P., Binns D., Chang H.-Y., Fraser M., Li W., *et al.*, 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240.

Kanzi A. M., Steenkamp E. T., Merwe N. A. van der, Wingfield B. D., 2019 The mating system of the *Eucalyptus* canker pathogen *Chrysoporthe austroafricana* and closely related species. Fungal Genet. Biol. 123: 41–52.

Koressaar T., Remm M., 2007 Enhancements and modifications of primer design program Primer3. Bioinformatics 23: 1289–1291.

Kronstad J. W., Staben C., 1997 Mating type in filamentous fungi. Annu. Rev. Genet. 31: 245–276.

Lee S. C., Ni M., Li W., Shertz C., Heitman J., 2010 The Evolution of Sex: A Perspective from the Fungal Kingdom. Microbiol. Mol. Biol. Rev. 74: 298–340.

Metzenberg R. L., Glass N. L., 1990 Mating type and mating strategies in *Neurospora*. Bioassays 12: 53–59.

Ni M., Feretzaki S. S., Wang X., Heitman J., 2011 Sex in Fungi. Annu. Rev. Genet. 45: 405–430.

Old K. M., Pongpanich K., Thru P. Q., Wingfield M. J., Yuan Z. Q., 1996 *Phaeophleospora destructans* causing leafblight epidemics in South East Asia. 8th International Congress of Plant Pathology 27:165.

Schauder C. M., Wu X., Saheki Y., Narayanaswamy P., Torta F., *et al.*, 2014 Structure of a lipid-bound extended synaptotagmin indicates a role in lipid transfer. Nature 510: 552.

Solovyev V., Kosarev P., Seledsov I., Vorobyev D., 2006 Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7: 10.1-10.12.

Stanke M., Morganstern B., 2005 AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33: W465–W467.

Stukenbrock E. H., Jorgensen F. G., Zala M., Hansen T. T., McDonald B. A., *et al.*, 2010

Whole-genome and chromosome evolution associated with host adaption and speciation of the wheat pathogen *Mycosphaerella graminicola*. PLoS Genet. 6: E1001189.

Taole M., Bihon W., Wingfield B. D., Wingfield M. J., Burgess T. I., 2015 Multiple introductions from multiple sources: Invasion patterns for an important *Eucalyptus* leaf pathogen. Ecol. Evol. 5: 4210–4220.

Turgeon B. G., Yoder O. C., 2000 Proposed nomenclature for mating-type genes of filamentous Ascomycetes. Fungal Genet. Biol. 31: 1–5.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., *et al.*, 2012 Primer3 - new capabilities and interfaces. Nucleic Acids Res. 40: e155.

Waalwijk C., Mendes O., Verstappen E. C. P., Waard M. A. de, Kema G. H., 2002 Isolation and characterization of the mating-type idiomorphs from the wheat septoria leaf blotch fungus *Mycosphaerella graminicola*. Fungal Genet. Biol. 35: 277–286.

Wilken P. M., Steenkamp E. T., Hall T. A., Beer Z. W. De, Wingfield M. J., *et al.*, 2012 Both mating types in the heterothallic fungus *Ophiostoma quercus* contain *MAT1-1* and *MAT1-2* genes. Fungal Biol. 116: 427–437.

Wilken P. M., Steenkamp E. T., Wingfield M. J., Beer Z. W. De, Wingfield B. D., 2014 DNA loss at the *Ceratocystis fimbriata* mating locus results in self-sterility. PLoS One 9: e92180.

Wilken P. M., Steenkamp E. T., Wingfield M. J., Beer Z. W. de, Wingfield B. D., 2017 Which *MAT* gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. Fungal Biol. Rev. 31: 199–211.

Wilken P. M., Steenkamp E. T., Nest M. A. van der, Wingfield M. J., Beer Z. W. De, *et al.*, 2018 Unexpected placement of the *MAT1-1-2* gene in the MAT1-2 idiomorph of *Thielaviopsis*. Fungal Genet. Biol. 113: 32–41.

Wilson A. M., Godlonton T., Nest M. A. Van Der, Wilken P. M., Wingfield M. J., *et al.*, 2015a Unisexual Reproduction in *Huntiella moniliformis*. Fungal Genet. Biol. 80: 1–9.

Wilson A. M., Wilken P. M., Nest M. A. van der, Steenkamp E. T., Wingfield M. J., *et al.*, 2015b Homothallism: An umbrella term for describing diverse sexual behaviours. IMA Fungus 6: 207–214.

Wingfield M. J., Crous P. W., Boden D., 1996 *Kirramyces destructans sp. nov.*, a serious leaf pathogen of *Eucalyptus* in Indonesia. South African J. Bot. 62: 325–327.

Wingfield M. J., Slippers B., Wingfield B. D., 2017 The unified framework for biological invasions: A forest fungal pathogen perspective. Biol. Invasions 19: 3201–3214.

Yin Z., Ke X., Li Z., Chen J., Gao X., *et al.*, 2017 Unconventional recombination in the mating

type locus of heterothallic apple canker pathogen *Valsa mali*. G3 7: 1259–1265.

# Chapter 2

## Tables and Figures

**Table 1** *Teratosphaeria destructans* isolates used in this study.

| Culture Collection Number | Origin of the Isolate | Mating-type |
|---|---|---|
| CMW 45661[a] | Baserah, South Sumatra | MAT1-1 |
| CMW 45664 | Baserah, South Sumatra | MAT1-2 |
| CMW 45674 | Baserah, South Sumatra | MAT1-2 |
| CMW 45649 | Baserah, South Sumatra | MAT1-2 |
| CMW 45653 | Baserah, South Sumatra | MAT1-1 |
| CMW 45659 | Baserah, South Sumatra | MAT1-2 |
| CMW 45681 | Porsea Nursery, South Sumatra | MAT1-1 |
| CMW 45684 | Porsea Nursery, South Sumatra | MAT1-2 |
| CMW 45977 | Porsea Nursery, South Sumatra | MAT1-2 |
| CMW 45688 | Aek Raja, North Sumatra | MAT1-1 |
| CMW 45978 | Aek Raja, North Sumatra | MAT1-2 |
| CMW 45693 | Aek Raja, North Sumatra | MAT1-1 |
| CMW 45698 | Aek Raja, North Sumatra | MAT1-1 |
| CMW 45982[a] | Aek Raja, North Sumatra | MAT1-2 |
| CMW 44957 | Zululand, South Africa | MAT1-2 |
| CMW 44962 | Zululand, South Africa | MAT1-2 |

[a] denotes the isolates selected for genome sequencing.

**Table 2** Known protein sequences used to identify contigs of *T. destructans* containing mating type genes.

| Protein | Accession Number | Organism | Reference |
|---|---|---|---|
| *MAT1-1-1* | AAL30838.1 | *Zymoseptoria tritici* | (Waalwijk *et al.* 2002) |
| *MAT1-2-1* | AAL30836.1 | *Zymoseptoria tritici* | (Waalwijk *et al.* 2002) |
| DNA Lyase | ADU79051.1 | *Zymoseptoria tritici* | (Stukenbrock *et al.* 2010) |
| Cytochrome C Oxidase | KJY01430.1 | *Zymoseptoria brevis* | (Grandaubert *et al.* 2015) |

**Table 3** Primer pairs designed to amplify intragenic regions of the genes within and flanking in *MAT1* locus.

| Primer Name | Primer Sequence (5' – 3') | Tm (˚C) | Target Gene |
|---|---|---|---|
| MAT111F | ACGCAGAAGACGATCTCCAA | 60.4 | *MAT1-1-1* |
| MAT111R | CTCGCTAGTTGGGTCGAAGA | 62.45 | |
| 121F | TGGATGATGCATTGCCAGTG | 60.4 | *MAT1-2-1* |
| 121R | GAGAACCTCAGGCAGTGTCT | 62.45 | |
| HYP_F | TAGGGTCTTCGGCATCATCC | 62.45 | Hypothetical 1 |
| HYP_R | TTTCTGTATCCCTTTGCGCC | 60.4 | |
| HYP_F2 | CAGACTCCTCGCAAAGAACG | 62.45 | Hypothetical 2 |
| HYP_R2 | TGCCAAAAGTCACCTTGTCG | 60.4 | |

**Table 4** Identities assigned to predicted open reading frames using BLAST analyses and the NCBI Fungal Database.

| Isolate | Gene | Top Hit | E Value | % Identity |
|---|---|---|---|---|
| 45982 | DNA Lyase | *Hortaea werneckii* | 5e-104 | 75 |
| | *MAT1-2-1* | *Cercospora zinea* | 3e-35 | 80 |
| | Hypothetical Protein 1 [a] | *Rhodotorula toruloides* | 1.0 | 93 |
| | Hypothetical Protein 2 [a] | *Baudoinia panamericana* | 5e-62 | 66 |
| | Cytochrome C Oxidase | *Exophiala dermatitidis* | 5e-104 | 75 |
| 45661 | DNA Lyase | *Hortaea werneckii* | 3e-176 | 73 |
| | *MAT1-1-1* | *Hortaea werneckii* | 3e-57 | 72 |
| | Hypothetical Protein 2 [a] | *Baudoinia panamericana* | 5e-62 | 66 |
| | Cytochrome C Oxidase | *Exophiala dermatitidis* | 9e-31 | 77 |

[a] Hypothetical protein refers to a predict ORF to which no function can be readily assigned.

**Table 5** InterProScan analysis of the genes present in the mating-type locus and flanking regions.

| Isolate | Gene | Domain | Domain Number |
|---|---|---|---|
| 45982 | DNA Lyase | Endonuclease/ Exonuclease/ Phosphatase | PF03372 |
| | *MAT1-2-1* | High Mobility Group Box | PF0050 |
| | Hypothetical Protein 1 | No predicted domain | - |
| | Hypothetical Protein 2 | Synaptotagmin-Like Mitochondrial-Lipid-Binding | PS51847 |
| | Cytochrome C Oxidase | Cytochrome C Oxidase Subunit VIa | PF02046 |
| 45661 | DNA Lyase | Endonuclease/ Exonuclease/ Phosphatase | PF03372 |
| | *MAT1-1-1* | Alpha High Mobility Group Box | PF04769 |
| | Hypothetical Protein 2 | Synaptotagmin-Like Mitochondrial-Lipid-Binding | PS51847 |
| | Cytochrome C Oxidase | Cytochrome C Oxidase Subunit VIa | PF02046 |

**Table 6** Distribution of mating-type genes across South and North Sumatra, South Africa.

| **Region** | ***MAT1-1-1* Gene Present** | ***MAT1-2-1* Gene Present** | **Total** |
|---|---|---|---|
| South Sumatra | 3 | 6 | 9 |
| North Sumatra | 3 | 2 | 5 |
| South Africa | 0 | 2 | 2 |
| Total | 6 | 10 | 16 |

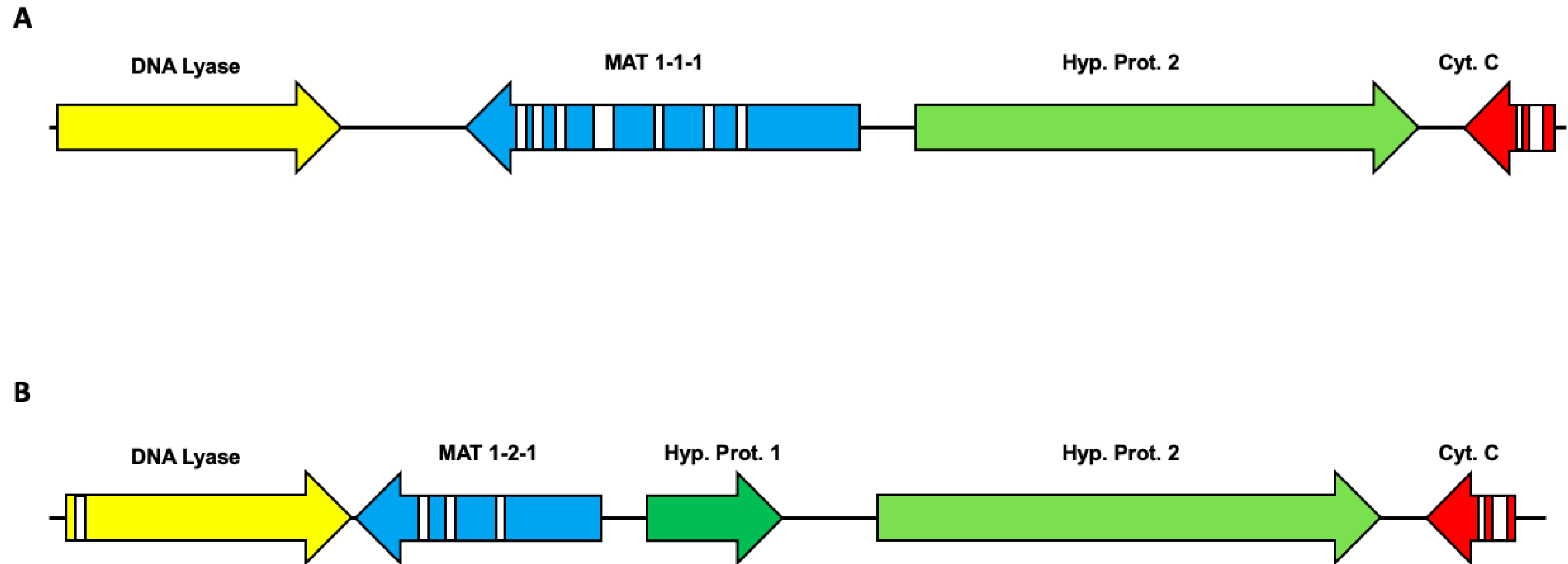


**Figure 1** The structure of the *MAT1-1* (A) and *MAT1-2* (B) mating-type idiomorphs and flanking regions for *T. destructans*. The *MAT1-1* idiomorph contains the *MAT1-1-1* gene, while the *MAT1-2* idiomorph contains the *MAT1-2-1* gene and a gene encoding a hypothetical protein (Hyp. Prot.). Genes are indicated by coloured regions, while introns are indicated by white regions within the genes.

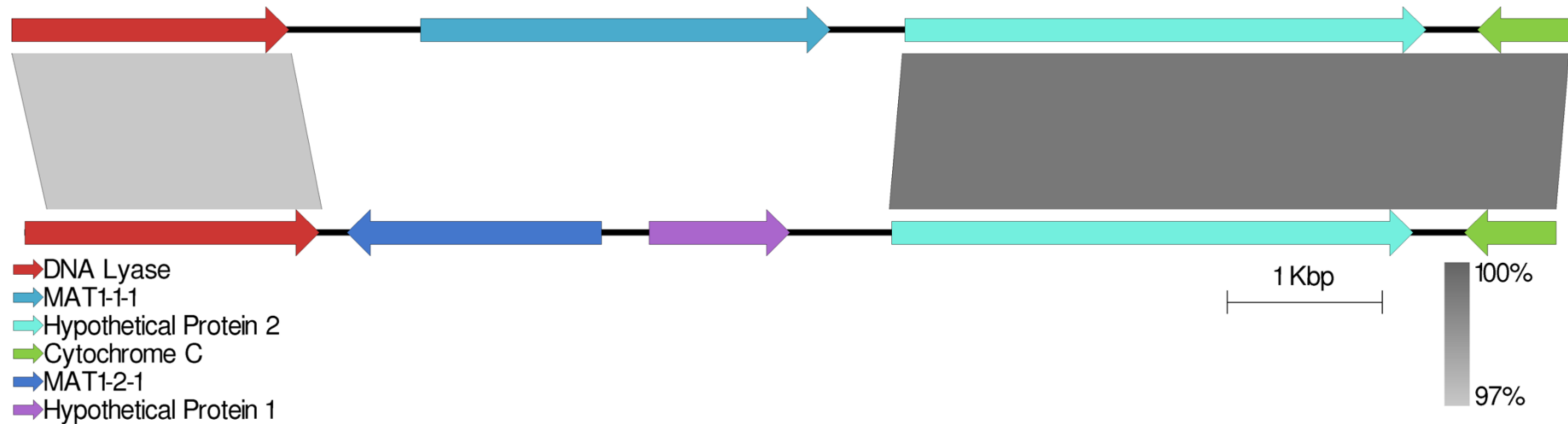


**Figure 2** Sequence similarity was observed between the flanking regions of the *MAT1* locus in both idiomorphs (indicated in grey). No similarity was observed between the *MAT1-1-1* gene in the *MAT1-1* idiomorph and the *MAT1-2-1* and *MAT1-2-12* genes in the *MAT1-2* idiomorph, allowing for boundaries of the idiomorph to be defined.

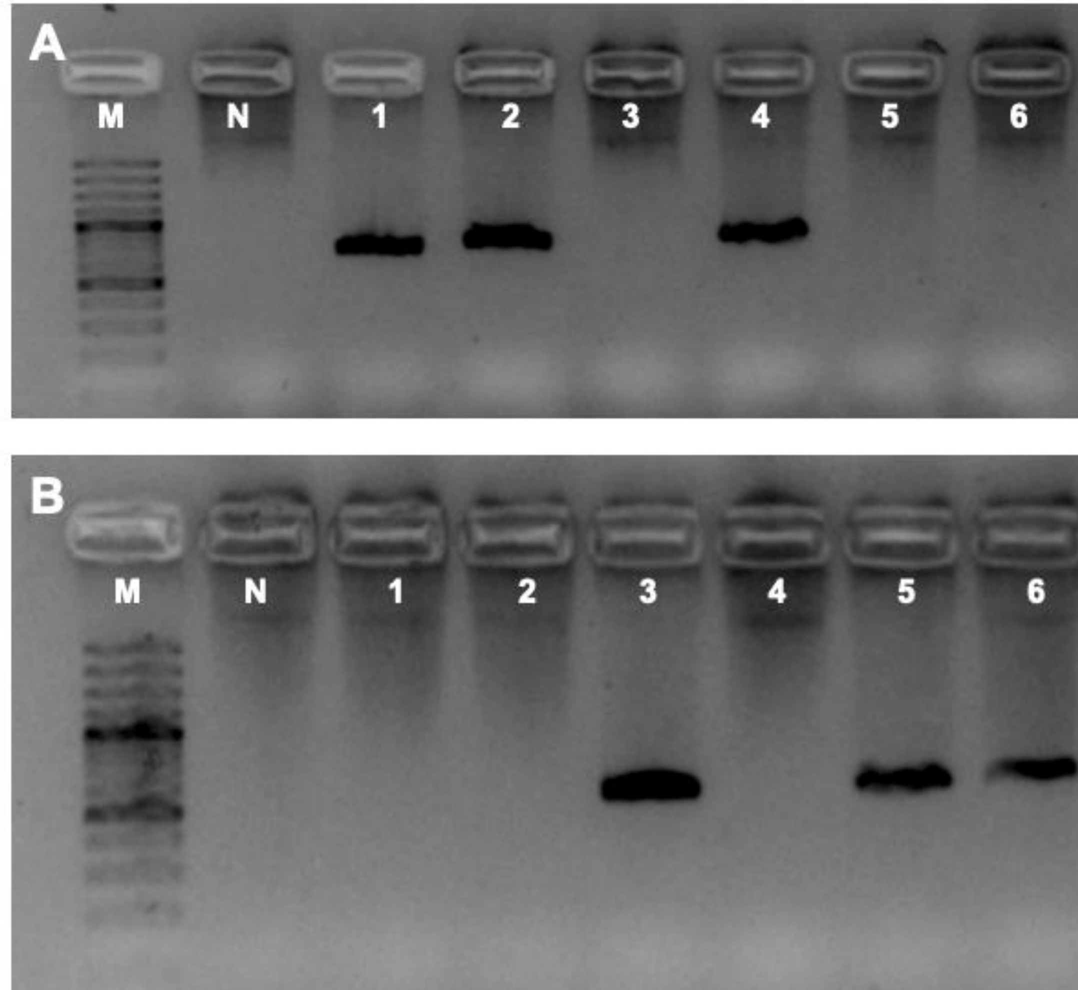


**Figure 3** The presence of the *MAT1-1-1* gene (Figure A) and *MAT1-2-1* gene (Figure B) in various isolates of *T. destructans*. 1: CMW 45661, 2: CMW 45653, 3: CMW 45681, 4: CMW 45688, 5: CMW 45982, 6: CMW 45659, M: a 100bp marker, N: Negative control.

**Supplementary Table 1** Assembly and scaffolding steps for the *T. destructans* CMW 45661 genome.

| Assembly/ Scaffold | Program | Input Data | Parameters | Resulting Number of Contigs |
|---|---|---|---|---|
| 1 | *De novo* assembly on CLC Genomics | Trimmed 20 million mate-pair reads | Default | 2689 |
| 2 | Join contigs on Genome Finishing Module | Assembly 1; Trimmed mate-pair reads | Use long reads | 2 466 |
| 3 | Join contigs on Genome Finishing Module | Assembly 2; Trimmed 20 million mate-pair reads | Use long reads | 2 400 |
| 4 | Join contigs on Genome Finishing Module | Assembly 3; Trimmed mate-pair reads | Use long reads | 2 356 |

**Supplementary Table 2** Assembly and scaffolding steps for the *T. destructans* CMW 45892 genome.

| **Assembly/ Scaffold** | **Program** | **Input Data** | **Parameters** | **Resulting Number of Contigs** |
|---|---|---|---|---|
| 1 | *De novo* assembly on CLC Genomics | | | 867 |
| 2 | Join contigs on Genome Finishing Module | Assembly 1; Trimmed pair end library | Use long reads | 836 |
| 3 | Join contigs on Genome Finishing Module | Assembly 2; Trimmed pair end paired reads | Use long reads | 823 |
| 4 | Join contigs on Genome Finishing Module | Assembly 3; Trimmed pair end library | Use long reads | 823 |
| 5 | Join contigs on Genome Finishing Module | Assembly 4; Trimmed pair end orphan reads | Use long reads | 799 |

# Chapter 3

## Characterisation of Microsatellite Markers in *Teratosphaeria destructans*

# CHARACTERISATION OF MICROSATELLITE MARKERS IN *TERATOSPHAERIA DESTRUCTANS*


## ABSTRACT

Microsatellite markers are commonly used to evaluate genetic variation at a population level. This study aimed to identify and evaluate the distribution of microsatellite markers across three genomes of *Teratosphaeria destructans*, which is an important *Eucalyptus* leaf pathogen. The resulting microsatellites were compared to published microsatellites for *T. destructans, T. gauchensis* and *T. zuluensis.* The data were also used to develop novel microsatellites to further distinguish between *T. destructans* isolates. The results yielded a total of 11 microsatellite markers that could be used to differentiate between *T. destructans* isolates from around the world, 5 previously published and 6 from this study. Microsatellites published for *T. gauchensis* and *T. zuluensis* were found to be useful only in discerning between the three species. This study is the first of its kind to analyse microsatellites within the *Teratosphaeria* genus at the genome level. The results yielded will assist in further understanding the structure of *T. destructans* populations.

## 1. INTRODUCTION

Microsatellite markers, also known as short tandem repeats or simple sequence repeats, are variable DNA loci made up of tandemly repeated nucleotide sequences (Field and Wills 1998; Richard and Paques 2000; Toth *et al.* 2000). They are present in both prokaryote and eukaryote genomes, and are found in coding and noncoding regions of the genome (Tautz and Renz 1984; Field and Wills 1996; Zane *et al.* 2002). Microsatellites are characterized by a high level of polymorphism which is thought to be the result of polymerase slippage during DNA replication and well as exchange of genetic material across chromosomes (Schlötterer and Tautz 1992; Buschiazzo and Gemmell 2006). This, coupled with their simple Mendelian mode of inheritance have made microsatellites the genetic marker of choice for strain typing, genetic mapping and studying population structures (Zane *et al.* 2002; Ellegren 2004). Despite the overwhelming popularity of microsatellites for population genetic studies, the major drawback to their use as genetic markers is the need for *de novo* isolation of microsatellites when a species is studied for the first time (Tautz and Renz 1984; Zane *et al.* 2002; Rico *et al.* 2013; Simpson *et al.* 2013).

Traditional methods to isolate microsatellite markers from genomes rely on cloning, as well as screening of massive partial genomic libraries with appropriate repeat containing probes (Rassmann *et al.* 1991; Zane *et al.* 2002). This is a time consuming process and has proven to be extremely tedious for genomes with a low frequency of microsatellites (Zane *et al.* 2002). In an attempt to devise a simpler, less time consuming and tedious method to isolate microsatellites, methods making use of enrichment protocols (such as FIASCO) were developed (Zane *et al.* 2002). While these methods provided an improvement, a number of drawbacks were still apparent (Zane *et al.* 2002; Santana *et al.* 2016). However, the recent surge in whole genome data has made the identification of microsatellite markers significantly simpler.

Whole genome sequencing has become an increasingly popular tool in biological studies (Simpson *et al.* 2013; Aylward *et al.* 2017). This is largely due to an increase in its affordability making it widely accessible (Church 2006). Coupling genome information with the development of reliable online platforms designed to predict microsatellites (Benson 1999; Dieringer and Schlötterer 2003; Faircloth 2008) makes for an ideal system to rapidly and accurately predict these markers across genomes (Richard and Dujon 1996; Toth *et al.* 2000; Rico *et al.* 2013). In many instances, regions flanking microsatellites can be extracted,

allowing for accurate primer design to amplify the marker. In the case where a single genome for a species is available, primers can then be tested on a number of isolates to determine variability of the markers (Simpson *et al.* 2013). This process is simplified further in instances where two or more genomes for a species are available, as variability can be determined *in silico* (Manawasinghe *et al.* 2018).

Reliable genetic markers have proven invaluable when studying Ascomycete pathogens that infect *Eucalyptus* (Harrington *et al.* 1998; Cortinas *et al.* 2006; Nakabonge *et al.* 2007; Andjic *et al.* 2011; van der Merwe *et al.* 2013; Aylward *et al.* 2019). Eucalypts are propagated globally, often outside of their native rage of Australia (Turnbull 1999; Wingfield *et al.* 2008; Aylward *et al.* 2019). Because of this, these trees are often susceptible to infection from novel pathogens (Wingfield *et al.* 2008). These plants are also commonly transported around the world, unknowingly spreading the pathogens to new environments (Wingfield *et al.* 2017). In these instances, being able to rapidly develop reliable microsatellite markers can be valuable.

The availability of microsatellite markers has allowed for the understanding of the global distribution of several *Teratosphaeria* species (Hunter *et al.* 2009; Andjic *et al.* 2011; Taole *et al.* 2015; Jimu *et al.* 2016a; b; Aylward *et al.* 2019). Species of this genus have been identified as the causal agents for devastating diseases including Teratosphaeria leaf blight (Wingfield *et al.* 1996; Andjic *et al.* 2011; Taole *et al.* 2015; Greyling *et al.* 2016) and Teratosphaeria stem canker (Jimu *et al.* 2016a; b; Aylward *et al.* 2019). This study aimed to develop microsatellite markers for the *Eucalyptus* leaf pathogen *Teratosphaeria destructans* (Wingfield *et al.* 1996). To do this, the genome data of three isolates differing in geographic location were mined for microsatellites. The resulting microsatellites were compared between the three isolates, and variable microsatellites identified. The *T. destructans* genomes were also used to characterise previously described microsatellites from *T. destructans* (Andjic *et al.* 2011), *T. gauchensis* (Cortinas *et al.* 2008) and *T. zuluensis* (Cortinas *et al.* 2006). The results of this study will serve as a starting point for generating microsatellite markers to be used to differentiate between global populations of *T. destructans*.

## 2. METHODS AND MATERIALS

### 2.1 A comparison of microsatellite markers in three *Teratosphaeria destructans* genomes

Genome data for three publicly available *T. destructans* isolates were used in this study (Wingfield *et al.* 2018; Chapter 1). The isolates originated from North Sumatra (CMW 45982; Chapter 1), South Sumatra (CMW 45661; Chapter 1) and South Africa (CMW 44962; Wingfield *et al.* 2018), respectively (Table 1). The genome data were imported into CLC Genomics (version 10.0.1; CLC Bio, Denmark) for further analysis.

Perfect microsatellites (Chambers *et al.* 2000) were mined from fasta sequence files of the three assembled genomes using MSATCommander with default search parameters (Faircloth 2008). The microsatellites identified in each genome were sorted using Microsoft Office Excel 2018 according to motif type. The number of each motif type, the total number of repeats per megabase (Mb), and the percentage of each microsatellite class within each genome was calculated and compared across the three genomes.

### 2.2 Identification and analysis of previously published *Teratosphaeria destructans* microsatellites

Fasta sequence files of previously described microsatellites for *T. destructans* (Andjic *et al.* 2011) were obtained from GenBank. Four of these microsatellites were known to be monomorphic and two were polymorphic amongst the isolates used in the study of Andjic *et al.* (2011; Table 2). Using local BLAST searches on CLC Genomics, the microsatellite content of the three *T. destructans* genomes was screened for the presence of these microsatellites. Primer sequences for the corresponding microsatellites were obtained from the study of Andjic *et al.* (2011), and the primer binding sites were identified in the three *T. destructans* isolates used in this study. To do this, the find primer binding sites option in CLC Genomics was used.

### 2.3 *In silico* evaluation of novel microsatellite markers

A subset of the microsatellites identified during the original genome wide searches for isolates CMW 44962 and CMW 45982 were selected. This set consisted of eight

microsatellites (four tri- and four tetranucleotide microsatellites) that were randomly selected from isolate CMW 44962 (Table 3). Using the local BLAST function in CLC Genomics, isolate CMW 45982 was screened for the presence of these eight microsatellites. Primers were designed that would putatively amplify the microsatellites from both isolates using the online primer design tool Primer3web version 4.1.0 (Untergasser *et al.* 2012) (http://primer3.ut.ee/) (Table 4).

### 2.4 Identification and analysis of previously published microsatellites for *Teratosphaeria* species other than *Teratosphaeria destructans*

Similarly to the *T. destructans* microsatellites, previously described microsatellites for *T. gauchensis* and *T. zuluensis* were obtained from GenBank (Table 5) (Cortinas *et al.* 2006, 2008). Using the method described above, the three *T. destructans* genomes were screened for the presence of microsatellites from *T. gauchensis* and *T. zuluensis*. Primer binding sites were also compared to determine the usefulness of these genetic markers to assess genetic diversity across species.

## 3. RESULTS

### 3.1 A comparison of microsatellite markers in three *Teratosphaeria destructans* genomes

The number of perfect microsatellites identified by MSatCommander varied across the three *T. destructans* isolates. Isolate CMW 45982 was found to possess 1882 microsatellites (68.2 microsatellites/Mb), while 1876 microsatellites (65.2 microsatellites/Mb) were identified in isolate CMW 45661 and 1830 microsatellites (56.7 microsatellites/Mb) in isolate CMW 44962 (Table 6). These microsatellites were spread across six classes of microsatellites, with mononucleotide repeats most abundant across all three of the genomes (Table 6).

### 3.2 Identification and analysis of previously published *Teratosphaeria destructans* microsatellites

BLAST analyses revealed that only three of the four monomorphic microsatellites (VA-6, VA-13 and VA-15) described by (Andjic *et al.* 2011) were present in the three *T. destructans*

isolates used in this study (Table 7). The primer sets for each microsatellite were found to bind as expected. While the locus VA-6 was identical between the *T. destructans* isolates used previously (Andjic *et al.* 2011) and the three isolates used in this study, variations were observed in VA-13 and VA-15 (Figure 1). Isolates CMW 45982 and CMW 44962 showed no variation at VA-13 but shared a common novel variant for VA-15 (Figure 1C). In contrast, isolate CMW 45661 showed variation from previously described VA-13 and VA-15 loci (Figures 1B and 1D), as well as the loci observed in the other two isolates used in this study.

Analyses of the two polymorphic loci revealed the three isolates used in this study grouping in a similar pattern as observed for the monomorphic loci. Isolates CMW 45982 and CMW 44962 shared a common VA-2 variant, VA-2(A) described by (Andjic *et al.* 2011) while isolate CMW 45661 was found to possess a novel variant (Figure 1A). For the VA-18 locus, all three isolates were found to possess variants that were identified previously, with isolates CMW 45982 and CMW 44962 both containing VA-18 (A), and isolate CMW 45661 containing VA-18 (B) (Andjic *et al.* 2011).

### 3.3 *In silico* evaluation of novel microsatellite markers for potential use in *Teratosphaeria destructans*

Eight microsatellites were compared between isolates CMW 45982 and CMW 44962 in an attempt to identify microsatellites that varied between the two isolates (Table 3). Two of the regions (CT-4, and CT-5) were found to be monomorphic between the isolates. Four of the microsatellites (CT-2, CT-6 CT-7 and CT-8) were found to be polymorphic. For regions CT-2, CT-6 and CT-8 variation was observed in the number of motif repeats present (Table 3). Variation of a different type was observed at region CT-7. While isolate CMW 44962 was found to possess twenty consecutive repeats of an AATC motif, isolate CMW 45982 contained two sets of five repeats of the motif separated by a three nucleotide long sequence (Figure 2). Regions CT-1 and CT-3 were not present in isolate CMW 45982.

### 3.4 Identification and analysis of previously published microsatellites for *Teratosphaeria* species other than *Teratosphaeria destructans*

While a number of microsatellites from *T. gauchensis* and *T. zuluensis* were identified in *T. destructans*, the resulting fragments would yield PCR products less than half of those for *T. gauchensis* and *T. zuluensis*. This is smaller than the size commonly considered detectable

for microsatellite analysis (Ellegren 2004). The remaining primer pairs originally designed to amplify these microsatellites were also found not to have binding sites within the regions flanking the microsatellites in any of the three *T. destructans* isolates to allow for amplification.

## 4. DISCUSSION

This is the first study to provide insight into the genetic variation observed in three *T. destructans* isolates using genome analysis. *In silico* analysis showed that previously described microsatellites (Andjic *et al.* 2011) would be useful to discern between the isolate from South Sumatra, and the isolates from North Sumatra and South Africa. The present study has added to the available microsatellite markers for the pathogen by identifying putative polymorphic microsatellites to differentiate between the North Sumatra and South African isolates.

Genome wide microsatellite comparison between the three *T. destructans* isolates used in this study revealed interesting variation in the abundance of microsatellite types. Variation in the number of microsatellites present across the three genomes suggests potential to differentiate between isolates based on the microsatellites present within a genome. The results were also consistent with the findings of other studies focusing on genome wide microsatellite analysis of Ascomycetes; more specifically, the microsatellite abundance was lower in the three *T. destructans* isolates when compared with other non-fungal Eukaryotes (Toth *et al.* 2000; Lim *et al.* 2004; Simpson *et al.* 2013). In contrast to other fungal species, tri- and hexanucleotides were not the most abundant motifs in the *T. destructans* genomes (Metzgar *et al.* 2000; Simpson *et al.* 2013). Rather, mononucleotides were the most abundant across the three *T. destructans* genomes, with trinucleotides being the second most abundant. This deviation from what has been found in other studies (Lim *et al.* 2004; Simpson *et al.* 2013) is possibly a result of this study relying on *in silico* analysis rather than an experimental approach to identify microsatellites.

BLAST analyses showed that the *T. destructans* microsatellites described and developed by (Andjic *et al.* 2011) are present in the three isolates considered in this study. Furthermore, the primer pairs were found to successfully bind *in silico*, implicating that these regions can be amplified *in vitro*. These results showed that the microsatellites would be useful for isolates originating outside of the geographic regions (China, Vietnam, Thailand and

Indonesia) tested by (Andjic *et al.* 2011). Additionally, the novel variations described in this study suggest that these microsatellites can be used to uniquely distinguish isolates from South Sumatra and South Africa. The novel variation identified in the isolate from North Sumatra (CMW 45982) in the current study when compared to the North Sumatran isolates used in the study by (Andjic *et al.* 2011) could be as a result of the time passed between sampling. Isolate CMW 45982 was collected in 2016, more than 12 years later than those collected for the study conducted by Andjic *et al.* (2011). This could be a result of rapid rate of evolution observed in microsatellite markers (Schlötterer and Tautz 1992; Jarne and Lagoda 1996; Ellegren 2004), or possible accidental introduction of new genotypes and thus different microsatellites to the region.

Isolates from North Sumatra and South Africa shared all published (Andjic *et al.* 2011) microsatellites. Although this was a small sample size, it is interesting to speculate that the isolates in South Africa could have originated from North Sumatra. However the variation observed in the microsatellites proposed in this study suggest that there is sufficient variation between the populations to differentiate between them.

The results of this study showed that microsatellites developed for the closely related species *T. gauchensis* (Cortinas *et al.* 2008) and *T. zuluensis* (Cortinas *et al.* 2006) were not transferable to *T. destructans*. Because closely related species have been shown to previously share microsatellites, this was unexpected (Simpson *et al.* 2013). However, while these microsatellites will not be useful to differentiate between *T. destructans* isolates, they do provide a tool to differentiate between *T. destructans* and isolates of *T. gauchensis* and *T. zuluensis*.

This study is the first of its kind to analyse microsatellites at a genomic level within *T. destructans*. Using this strategy, microsatellites were identified across three isolates allowing for putatively polymorphic markers to be rapidly identified. Additionally, these genomes were used to design novel microsatellites and accompanying primer pairs. While the results of this study are informative, further analysis is required to confirm the results *in vitro*.

## 5. REFERENCES

Andjic V., Dell B., Barber P., Hardy G., Wingfield M. J., *et al.*, 2011 Plants for planting; indirect evidence for the movement of a serious forest pathogen, *Teratosphaeria destructans*, in Asia. Eur. J. Plant Pathol. 131: 49–58.

Aylward J., Steenkamp E. T., Dreyer L. L., Roets F., Wingfield B. D., *et al.*, 2017 A plant pathology perspective of fungal genome sequencing. IMA Fungus 8: 1–15.

Aylward J., Roets F., Dreyer L. L., Wingfield M. J., 2019 Teratosphaeria stem canker of *Eucalyptus*: Two pathogens, one devastating disease. Mol. Plant Pathol. 20: 8–19.

Benson G., 1999 Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 27: 573–580.

Buschiazzo E., Gemmell N. J., 2006 The rise, fall and renaissance of microsatellites in Eukaryotic genomes. BioEssays 28: 1040–1050.

Chambers G., Chambers G. K., Macavoy E. S., 2000 Microsatellites: Consensus and controversy. Comp. Biochem. Physiol. Part B 126: 455–476.

Church G. M., 2006 Genomes for All. Sci. Am. 294: 46–54.

Cortinas M. N., Barnes I., Wingfield B. D., Wingfield M. J., 2006 Polymorphic microsatellite markers for the *Eucalyptus* fungal pathogen *Colletogloeopsis zuluensis*. Mol. Ecol. Notes 6: 780–783.

Cortinas M. N., Wingfield B. D., Wingfield M. J., 2008 Microsatellite markers for the *Eucalyptus* stem canker fungal pathogen *Kirramyces gauchensis*. Mol. Ecol. Resour. 8: 590–592.

Dieringer D., Schlötterer C., 2003 Microsatellite Analyser (MSA): A platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Resour. 3: 167–169.

Ellegren H., 2004 Microsatellites: Simple sequences with complex evolution. Nat. Rev. Genet. 5: 435–445.

Faircloth B. C., 2008 MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol. Ecol. Resour. 8: 92–94.

Field D., Wills C., 1996 Long, polymorphic microsatellites in simple organisms. Proc. Biol. Sci. 263: 209–215.

Field D., Wills C., 1998 Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad. Sci. United.

Greyling I., Wingfield M. J., Coetzee M. P., Marincowitz S., Roux J., 2016 The *Eucalyptus*

shoot and leaf pathogen *Teratosphaeria destructans* recorded in South Africa. South. For. a J. For. Sci. 78: 123–129.

Harrington T. C., Steimel J. P., Kile G. A., 1998 Genetic variation in three *Ceratocystis* species with outcrossing, selfing and asexual reproductive strategies. Eur. J. For. Pathol. 28: 217–226.

Hunter G. C., Crous P. W., Carnegie A. J., Wingfield M. J., 2009 *Teratosphaeria nubilosa*, a serious leaf disease pathogen of *Eucalyptus* sp. in native and introduced areas. Mol. Plant Pathol. 10: 1–14.

Jarne P., Lagoda P. J. L., 1996 Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11: 424–429.

Jimu L., Chen S. F., Wingfield M. J., Mwenje E., Roux J., *et al.*, 2016a The Eucalyptus stem canker pathogen *Teratosphaeria gauchensis* represents distinct genetic groups in Africa and South America. For. Pathol. 46: 229–239.

Jimu L., Chen S., Wingfield M. J., Mwenje E., Roux J., 2016b Three genetic groups of the *Eucalyptus* stem canker pathogen *Teratosphaeria zuluensis* introduced into Africa from an unknown source. Antonie Van Leeuwenhoek 109: 21–33.

Lim S., Notley-McRobb L., Lim M., Carter D. A., 2004 A comparison of the nature and abundance of microsatellites in 14 fungal genomes. Fungal Genet. Biol. 41: 1025–1036.

Manawasinghe I. S., Zhang W., Li X., Zhao W., Chethana K. W. T., *et al.*, 2018 Novel microsatellite markers reveal multiple origins of *Botryosphaeria dothidea* causing the Chinese grapevine trunk disease. Fungal Ecol. 33: 134–142.

Merwe N. A. van der, Gryzenhout M., Steenkamp E. T., Wingfield B. D., Wingfield M. J., 2013 Multigene phylogenetic and population differentiation data confirm the existence of a cryptic species within *Chrysoporthe cubensis*. Plant Pathol. 62: 642–648.

Metzgar D., Bytof J., Wills C., 2000 Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res. 10: 72–80.

Nakabonge G., Gryzenhout M., Wingfield B. D., Wingfield M. J., Roux J., 2007 Genetic diversity of *Chrysoporthe cubensis* in eastern and southern Africa. S. Afr. J. Sci. 103: 261–264.

Rassmann K., Schlötterer C., Tautz D., 1991 Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. Electrophoresis 12: 113–115.

Richard G.-F., Dujon B., 1996 Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. Gene 174: 165–174.

Richard G. F., Paques F., 2000 Mini- and microsatellite expansions: The recombination connection. EMBO Rep. 11: 122–126.

Rico C., Normandeau E., Dion-Côté A. M., Rico M. I., Côté G., *et al.*, 2013 Combining next-generation sequencing and online databases for microsatellite development in non-model organisms. Sci. Rep. 3: 1–8.

Santana Q. C., Coetzee M. P. A., Wingfield B. D., Wingfield M. J., Steenkamp E. T., 2016 Nursery-linked plantation outbreaks and evidence for multiple introductions of the pitch canker pathogen *Fusarium circinatum* into South Africa. Plant Pathol. 65: 357–368.

Schlötterer C., Tautz D., 1992 Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20: 211–215.

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

Taole M., Bihon W., Wingfield B. D., Wingfield M. J., Burgess T. I., 2015 Multiple introductions from multiple sources: Invasion patterns for an important *Eucalyptus* leaf pathogen. Ecol. Evol. 5: 4210–4220.

Tautz D., Renz M., 1984 Simple sequences are ubiquitous repetitve components of eukaryotic genomes. Nucleic Acids Res. 12: 4127–4138.

Toth G., Gaspari Z., Jurka J., 2000 Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res. 10: 967–981.

Turnbull J. W., 1999 Eucalypt plantations. New For. 17: 37–52.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., *et al.*, 2012 Primer3 - new capabilities and interfaces. Nucleic Acids Res. 40: e155.

Wingfield M. J., Crous P. W., Boden D., 1996 *Kirramyces destructans sp. nov.*, a serious leaf pathogen of *Eucalyptus* in Indonesia. South African J. Bot. 62: 325–327.

Wingfield M. J., Slippers B., Hurley B. P., Coutinho T. A., Wingfield B. D., *et al.*, 2008 Eucalypt pests and diseases: Growing threats to plantation productivity. South. For. 70: 139–144.

Wingfield M. J., Slippers B., Wingfield B. D., 2017 The unified framework for biological invasions: A forest fungal pathogen perspective. Biol. Invasions 19: 3201–3214.

Wingfield B. D., Liu M., Nguyen H. D. T., Lane F. A., Morgan S. W., *et al.*, 2018 IMA Genome-F 10: Nine draft genome sequences of *Claviceps purpurea s.lat.*, including *C. arundinis, C. humidiphila,* and *C. cf. spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti, Grosmannia.* IMA Fungus 9: 401–418.

Zane L., Bargelloni L., Patarnello T., 2002 Strategies for microsatellite isolation: A review. Mol. Ecol. 11: 1–16.

# Chapter 3

## Tables and Figures

**Table 1** *Teratosphaeria destructans* isolates whose genomes were used in this study.

| Isolate | Geographic Location | Genome Size (Mb) | Reference |
|---|---|---|---|
| CMW 45982 | North Sumatra | 27.6 | Chapter 1 |
| CMW 45661 | South Sumatra | 28.8 | Chapter 1 |
| CMW 44962 | South Africa | 32.3 | (Wingfield *et al.* 2018) |

**Table 2** Published microsatellites for use in *T. destructans*.

| GenBank Accession Number | Region | Reference |
|---|---|---|
| EU620603.1 | VA-1 [a] | (Andjic *et al.* 2011) |
| EF686346.2 | VA-2 [b] | (Andjic *et al.* 2011) |
| EF686349.2 | VA-2 [b] | (Andjic *et al.* 2011) |
| EF686353.2 | VA-2 [b] | (Andjic *et al.* 2011) |
| EU620611.1 | VA-6 [a] | (Andjic *et al.* 2011) |
| EF686398.1 | VA-13 [a] | (Andjic *et al.* 2011) |
| EU620619.1 | VA-15 [a] | (Andjic *et al.* 2011) |
| EF686436.2 | VA-18 [b] | (Andjic *et al.* 2011) |
| EF686444.2 | VA-18 [b] | (Andjic *et al.* 2011) |

[a] denotes monomorphic loci

[b] denotes polymorphic loci

**Table 3** Microsatellites selected for comparison between *T. destructans* isolates CMW 44962 and CMW 45982.

| Nucleotide Class | Region | Motif | Number of Repeats in CMW 44962 | Number of Repeats in CMW 45982 |
|---|---|---|---|---|
| Trinucleotide | CT-1 | ACT | $(ACT)_{29}$ | Absent |
| Trinucleotide | CT-2 | AAG | $(AAG)_{14}$ | $(AAG)_{10}$ |
| Trinucleotide | CT-3 | CCG | $(CCG)_{8}$ | Absent |
| Trinucleotide | CT-4 | ACG | $(ACG)_{8}$ | $(ACG)_{8}$ |
| Tetranucleotide | CT-5 | ACCT | $(ACCT)_{10}$ | $(ACCT)_{10}$ |
| Tetranucleotide | CT-6 | AGGC | $(AGGC)_{13}$ | $(AGGC)_{14}$ |
| Tetranucleotide | CT-7 | AATC | $(AATC)_{20}$ | $(AATC)_{10}$ (interrupted) |
| Tetranucleotide | CT-8 | AAGC | $(AAGC)_{8}$ | $(AAGC)_{11}$ |

**Table 4** Primers designed to amplify microsatellite regions in *T. destructans.*

| Region | Primer Name | Primer Sequence (5' – 3') | Tm (˚C) | Amplicon Size (bp) |
|---|---|---|---|---|
| CT-1 | CT-1F | CGCCTCATCCTCATCCTCAT | 59.02 | 344 [b] |
| | CT-1R | CCGCAATCCATACATCGACC | 58.85 | |
| CT-2 | CT-2F | TCGGTGGTGGTGTATGTGAA | 58.95 | 333 [a] |
| | CT-2R | ACAACAGCAACAACGACCAG | 59.27 | 345 [b] |
| CT-3 | CT-3F | TGATCTGCTCGAGGACCTTC | 58.89 | 218 [b] |
| | CT-3R | CACATCCATCCAACTCCCCT | 59.08 | |
| CT-5 | CT-5F | CGCTTGAAGTCATTGTCGA | 57.85 | 348 [c] |
| | CT-5R | CAATGGCAGTGTCAGCGG | 59.13 | |
| CT-6 | CT-6F | CAACGAACCACTTCCCGAAC | 59.42 | 320 [a] |
| | CT-6R | CCCTGCTCTTTCTCGTTTCC | 58.56 | 313 [b] |
| CT-7 | CT-7F | TGGGGTTTTATTGGGATGCG | 58.52 | 212 [a] |
| | CT-7R | CACGCAATTGTCCCAGATCC | 59.27 | 353 [b] |
| CT-8 | CT-8F | TGGTCTGGCTTGGTCTTTCT | 58.86 | 210 [a] |
| | CT-8R | TCTTCGCAAAAGCGGTCATC | 59.20 | 198 [b] |

[a] denotes amplicon size for isolate CMW 45982

[b] denotes amplicon size for CMW 44962

[c] denotes amplicon size for both isolates.

**Table 5** Published microsatellites for *T. gauchensis and T. zuluensis* considered in this study.

| Species | GenBank Accession Number | Reference |
|---|---|---|
| *T. gauchensis* | DQ75190.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75191.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75192.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75193.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75194.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75195.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75196.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75197.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75198.1 | (Cortinas *et al.* 2008) |
| *T. gauchensis* | DQ75199.1 | (Cortinas *et al.* 2008) |
| *T. zuluensis* | DQ156110.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156111.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156112.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156113.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156114.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156115.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156116.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156117.1 | (Cortinas *et al.* 2006) |
| *T. zuluensis* | DQ156118.1 | (Cortinas *et al.* 2006) |

**Table 6** The number of nucleotide repeats identified in each *T. destructans* genome used in this study.

| Microsatellite class | Total Motifs (% of total) | Abundancy (microsatellite/mb) |
|---|---|---|
| | **CMW 45982** | |
| Mononucleotides | 1121 (59.6%) | 40.6 |
| Dinucleotides | 215 (11.4%) | 7.8 |
| Trinucleotides | 481 (25.6%) | 17.4 |
| Tetranucleotides | 36 (1.9%) | 1.3 |
| Pentanucleotides | 10 (0.5%) | 0.4 |
| Hexanucleotides | 19 (1.0%) | 0.7 |
| Total | 1882 (100%) | 68.2 |
| | **CMW 45661** | |
| Mononucleotides | 1082 (58.0%) | 37.6 |
| Dinucleotides | 231 (12.4%) | 8.0 |
| Trinucleotides | 487 (26.1%) | 16.9 |
| Tetranucleotides | 43 (2.3%) | 1.5 |
| Pentanucleotides | 8 (0.4%) | 0.3 |
| Hexanucleotides | 16 (0.9%) | 0.6 |
| Total | 1867 (100%) | 65.2 |
| | **CMW 44962** | |
| Mononucleotides | 1000 (54.6%) | 31.0 |
| Dinucleotides | 230 (12.6%) | 7.1 |
| Trinucleotides | 508 (27.8%) | 15.7 |
| Tetranucleotides | 43 (2.3%) | 1.3 |
| Pentanucleotides | 14 (0.8%) | 0.4 |
| Hexanucleotides | 35 (1.9%) | 1.1 |
| Total | 1830 (100%) | 56.7 |

**Table 7** Distribution of previously described *T. destructans* microsatellites across the three *T. destructans* genomes analysed here.

| GenBank Accession Number | Region | CMW 45982 | CMW 45661 | CMW 44982 |
|---|---|---|---|---|
| EU620603.1 | VA-1 | Absent | Absent | Absent |
| EF686346.2 | VA-2 | Present | Present [a] | Present |
| EF686349.2 | VA-2 | Absent | Absent | Absent |
| EF686353.2 | VA-2 | Absent | Absent | Absent |
| EU620611.1 | VA-6 | Present | Present | Present |
| EF686398.1 | VA-13 | Present | Present [a] | Present |
| EU620619.1 | VA-15 | Present [a] | Present [a] | Present [a] |
| EF686436.2 | VA-18 | Present | Absent | Present |
| EF686444.2 | VA-18 | Absent | Present | Absent |

[a] denotes a microsatellite variation not described by (Andjic *et al.* 2011).

**Figure 1** Novel variants of microsatellites described by (Andjic *et al.* 2011) observed in *T. destructans* isolates used in this study.

**A** Variant at region VA-2 in isolate CMW 45661.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VA-2 (A) | C | A | C | A | C | A | C | A | C | A | C | A | C | A | C | C | A | C | A |
| VA-2 (B) | C | A | C | A | C | A | C | A | C | A | C | A | C | - | - | C | A | C | A |
| VA-2 (C) | C | A | C | A | C | A | C | A | C | A | C | - | - | - | - | C | A | C | A |
| CMW 45661 | C | A | C | A | C | A | C | A | C | - | - | - | - | - | - | C | A | C | A |

**B** Variant at region VA-13 in isolate CMW 45661.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VA-13 | A | C | C | A | C | C | A | G | C | A | G | A | A | C | C | A | C | C | A | C | C | A | C | C | A | C | T | A | G | C | A | C | C | A | C |
| CMW 45661 | A | C | C | A | C | C | A | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | C | T | A | G | C | A | C | C | A | C |

**C** Variant at region VA-15 in isolates CMW 45982 and CMW 44962.

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VA-15 | C | C | A | C | A | C | A | C | A | C | A | C | A | C | A | A | C |
| CMW 45982 | C | C | A | C | A | C | A | C | A | C | A | C | - | C | A | A | C |
| CMW 44962 | C | C | A | C | A | C | A | C | A | C | A | C | - | C | A | A | C |

**D** Variant at region VA-15 in isolate CMW 45661.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VA-15 | C | C | A | G | G | C | A | A | A | G | C | C | A | G | C | A | A | T | T | C | T | G | C | C | T | A | G | A | C | T | C | G | A |
| CMW 45661 | C | C | A | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | C | G | A |

**Figure 2** Variation observed at region CT-7 between isolates CMW 44962 and CMW 45982.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CMW 45982 | C | A | A | T | C | A | A | T | G | A | T | C | A | A | T | C | A | A | T |
| CMW 44962 | C | A | A | T | C | A | A | T | - | - | - | C | A | A | T | C | A | A | T |

## SUMMARY

Whole genome sequencing is an increasingly favoured tool used to study plant pathogens. This has allowed for a dramatic decrease in the time and money required to identify and study various biological characteristics of pathogens, particularly mating strategy and population genetics. This study focused on a catastrophic eucalypt leaf and shoot pathogen, *Teratosphaeria destructans*, and aimed to elucidate the mating strategy of the pathogen and develop microsatellite markers for future population studies. The genomes of two *T. destructans* isolates from different geographic locations were sequenced and assembled. Local BLAST analyses revealed two distinct mating idiomorphs present at the mating type locus of *T. destructans*. The MAT1-1 idiomorph was found to possess a single mating type gene, the *MAT1-1-1* gene, while the MAT1-2 idiomorph was found to harbour the *MAT1-2-1* gene as well as an additional mating type gene which has putatively been named *MAT1-2-12*. PCR analysis of 16 isolates from North Sumatra, South Sumatra and South Africa using primers designed to amply fragments of these genes revealed that both mating types were found in both regions of Sumatra, while only a single mating type was identified in South Africa. These results show that *T. destructans* reproduces in a heterothallic manner, and is capable of doing so in North and South Sumatra, but not in South Africa. In order to develop microsatellite markers, both genomes and an additional genome from a South African isolate were mined for microsatellites. The resulting microsatellites were compared between the three isolates, and the polymorphic loci were extracted. Additionally, the presence of published microsatellites within the genomes was determined. Combined, these analyses yielded a total of 11 microsatellites that could be used to distinguish between the three isolates. Together, these results provide valuable insight into the biology of *T. destructans*, and will be of use when developing management and control strategies for this pathogen.