Genomic characterization of mating type loci and mating type distribution in two apparently asexual plantation tree pathogens

J. Aylward[1,2*], M. Havenga[1,2], L.L. Dreyer[3], F. Roets[2], B. D. Wingfield[1], M. J. Wingfield[1]

[1] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20 Hatfield, 0028, South Africa.

[2] Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1 Matieland, 7602, South Africa

[3] Department of Botany and Zoology, Stellenbosch University, Private Bag X1 Matieland, 7602, South Africa.

*janneke.aylward@fabi.up.ac.za

Running head: Mating types of two plantation pathogens

Keywords: Capnodiales, heterothallic, mating type, stem canker, Teratosphaeria

Abstract

The *Eucalyptus* stem canker pathogens, *Teratosphaeria gauchensis* and *T. zuluensis* (Capnodiales, Teratosphaeriaceae) are found in many tropical regions of the world where their hosts are cultivated for plantation forestry. Population genetic analyses have suggested that some populations undergo recombination, even though their sexual states have never been observed. Against this background, the aim of this study was to characterise the mating type (*MAT*) locus of these species and thus to better understand the basis of their diversity. Known Mycosphaerellaceae *MAT* genes were used to identify and investigate the *MAT* locus in the *T. gauchensis* and *T. zuluensis* genomes. Both species were found to be heterothallic and primers were designed to amplify the opposite *MAT* idiomorphs as well as conserved regions within the *MAT1-1-1* and *MAT1-2-1* genes. Each *Teratosphaeria MAT* idiomorph was defined by either the *MAT1-1-1* or the *MAT1-2-1* gene, and an idiomorph-specific hypothetical protein (*MAT1-1-10* and *MAT1-2-12*). Populations of *T. zuluensis* from Asia and southern Africa were dominated by a single mating type, whereas the proportions of the different idiomorphs for *T. gauchensis* in South America and southern Africa were similar. There was no physical evidence of sexual reproduction for either species and we argue that although recombination may be possible, it is unlikely to form an important part of their lifecycles in diseased *Eucalyptus* plantations. Instead, continuous human-mediated multiple introductions of these species have likely resulted in the current genetic structure of their populations, which holds risk for future disease outbreaks and inter-specific hybridisation.

Introduction

Teratosphaeria gauchensis and *T. zuluensis* are Dothideomycete fungi that independently cause an important stem canker disease on *Eucalyptus* species cultivated in areas of the world with tropical and sub-tropical climates (Aylward *et al.*, 2019). The disease, referred to as Teratosphaeria stem canker, is expressed as small, necrotic stem lesions that coalesce over time, ultimately forming gum-filled cankers (Wingfield *et al.*, 1996). These cankers reduce stem growth, negatively affect the pulping process, and may girdle the stems, arresting apical tree growth in severe cases (Gezahgne *et al.*, 2003, Old *et al.*, 2003). *Teratosphaeria gauchensis* and *T. zuluensis* are the only known stem pathogens residing in a well-known genus of leaf-infecting pathogens and endophytes (Quaedvlieg *et al.*, 2014).

Teratosphaeria stem canker was first detected in a South African *Eucalyptus* plantation in the late 1980's (Wingfield *et al.*, 1996). The disease is now known from all major *Eucalyptus*-growing regions of the world, including Asia, Africa, South and Central America, Hawaii, and was most recently found in Europe (Aylward *et al.*, 2019). The two different species that cause the disease have relatively distinct geographic distributions (Cortinas *et al.*, 2006), with *T. gauchensis* occurring in South America, Hawaii, Europe and North Africa and *T. zuluensis* in Asia, Mexico and southern Africa. In Africa, however, the boundaries of distribution have become ambiguous, with both *T. gauchensis* and *T. zuluensis* detected in Uganda (Jimu *et al.*, 2014) and *T. gauchensis* in southern Africa (Jimu *et al.*, 2015, 2016a).

Current evidence suggests that *T. gauchensis* and *T. zuluensis* were initially introduced into plantations with the germplasm of their *Eucalyptus* hosts (Aylward *et al.*, 2019), a hypothesis supported by the presence of *T. zuluensis* in *E. grandis* seed (Jimu *et al.*, 2016c). In the case of *T. zuluensis*, population genetic differentiation is clear between different countries and at different geographic and temporal scales within a country, indicative of independent introductions (Cortinas *et al.*, 2010, Chen *et al.*, 2011). *Teratosphaeria gauchensis* follows the

same trend at a larger scale, with independent origins of populations in South America and Africa (Jimu *et al.*, 2016b).

The index of association suggests some level of recombination in both *T. gauchensis* and *T. zuluensis*. This index measures the degree to which loci are linked and thus estimates whether meiosis is occurring (Maynard Smith *et al.*, 1993). It detected marginally significant recombination in a *T. zuluensis* population from China (Cortinas *et al.*, 2010, Chen *et al.*, 2011) and South Africa (Jimu *et al.*, 2016a) and *T. gauchensis* from Argentina (Cortinas *et al.*, 2011, Jimu *et al.*, 2016b), suggesting that sexual reproduction may be taking place. Native populations of these stem canker pathogens have never been found and are, therefore, not available for comparisons.

In Ascomycete fungi, sexual reproduction is genetically determined by a pair of dissimilar alleles (idiomorphs), that typically occur at the mating type (*MAT*) locus (Kronstad & Staben, 1997). With rare exceptions (Wilson *et al.*, 2015), both the *MAT1-1* and *MAT1-2* idiomorphs are required to induce sexual reproduction. The two may occur within a single individual (homothallism) or different individuals (heterothallism) of a species (Turgeon & Yoder, 2000). Insight into the reproductive ability and strategy of any species is valuable, as the ability to generate genetic diversity often translates to greater environmental adaptability (McDonald & Linde, 2002). In the case of pathogens, this feature impacts the formulation of disease management strategies and informs predictions of future disease emergence and severity.

Only the asexual states of *T. gauchensis* and *T. zuluensis* are known, similar to many other species of *Teratosphaeria* (Wingfield *et al.*, 1996, Crous *et al.*, 2009). This does not preclude the existence of a cryptic sexual state in nature. Population genetic structure, considered in combination with the relative abundance of mating type idiomorphs, has provided evidence for cryptic sexual cycles in apparently asexual fungi (Goodwin *et al.*, 2003, Stergiopoulos *et al.*, 2007). In some cases, this information facilitated *in vitro* induction of sexual structures (*e.g.* Yilmaz *et al.*, 2016). The aim of this study was to investigate whether sexual recombination

provides a reasonable explanation for the moderate genetic diversity observed in some *T. gauchensis* and *T. zuluensis* populations. This was achieved by characterising their *MAT* loci and thus considering the potential of these species to undergo sexual recombination in culture.

Materials and Methods

Fungal isolates and DNA extraction

The ex-holotypes of *T. gauchensis* (CBS119465 and CBS119467), ex-epitype of *T. zuluensis* (CBS119470) and *T. zuluensis* isolate CBS117262 were used to characterize the *MAT* loci of the two species. Seventy-three additional isolates of *T. gauchensis* and 295 of *T. zuluensis* (Supplementary File 1), sourced from across their global distribution, were used to investigate the distribution of mating types in areas affected by the stem canker disease that they cause.

All isolates used in this study are maintained in the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI) at the University of Pretoria, South Africa. For DNA extraction, cultures were grown on MEA+Y medium (Malt Extract Agar supplemented with 3.0 g/L Yeast Extract Powder, both from Merck, South Africa). Cultures were incubated at 25°C in the dark for two to three weeks before harvesting the mycelial mats for lyophilisation.

Genomic DNA extraction was based on the protocol of Damm *et al.* (2008). Approximately 30 mg mycelium was crushed in an MM301 tissue lyzer (Retsch, Germany), using six glass beads and 600 µl extraction buffer (2% CTAB, 100 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA and 1.5 µl mercaptoethanol). After incubation at 65°C for 15 min, the cell lysate was extracted once by adding 400 µl chloroform:isoamyl alcohol (42:1), centrifuging for 15 min at maximum speed and retaining the supernatant. Polysaccharides were precipitated by adding 200 µl 5 M KAc, incubating at -20°C and repeating the centrifugation. DNA was precipitated from the

supernatant by adding 600 µl isopropanol and centrifuging immediately as before. The DNA pellet was washed twice with 70% ethanol, dried and resuspended in low-TE buffer (10 mM Tris-HCl, 0.1 mM EDTA; pH 8).

MAT locus identification and characterization

The *MAT* loci of *T. gauchensis* (CBS119465) and *T. zuluensis* (CBS119470) were identified using their recently sequenced genomes (Wingfield *et al.*, in press; GenBank® Accession numbers VCMR00000000 and VCMQ00000000). The *MAT1-1-1* and *MAT1-2-1* protein sequences of representative species from Capnodiales genera with characterised *MAT* loci (Table 1) were obtained from GenBank® and used as query sequences in local tBLASTn searches in CLC Genomics Workbench 11.0.2 (www.qiagenbioinformatics.com). A maximum e-value of 1.0^{-5} and minimum query coverage of 50% were applied as thresholds.

Open reading frames (ORFs) were predicted in the putative *MAT* regions, including approximately 7 kb of upstream and downstream flanking regions, using WebAUGUSTUS (Hoff & Stanke, 2013) and FGENESH (Solovyev *et al.*, 2006). Since no Dothideomycete model species are available on WebAUGUSTUS, three predictions were run using Eurotiomycete (*Aspergillus fumigatus*), Leotiomycete (*Botrytis cinerea*) and Sordariomycete (*Fusarium graminearum*) species as models. In FGENESH, generic *Mycosphaerella* gene-finding parameters as well as parameters for the Dothideomycete order Pleosporales are available and were used for prediction. The predicted ORFs were annotated by comparing their predicted proteins against the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov) database with BLASTp and identifying conserved domains with Reverse-Position-Specific (RPS)-BLAST and the Protein Family (PF) database (Bateman *et al.*, 2004).

Table 1 GenBank® accession numbers of the mating type idiomorphs of Capnodiales species used in this study.

Species	GenBank Accession	
	MAT1-1	**MAT1-2**
Cercospora beticola	KC960688	KC960689
Dothistroma pini	DQ915449	DQ915452
Passalora fulva	DQ659350	DQ659351
Pseudocercospora eumusae	GU046393	GU046394
Pseudocercospora fijiensis	DQ787015	DQ787016
Zymoseptoria passerinii	AF483193	AF483194
Zymoseptoria tritici	AF440399	AF440398

Since a single *MAT* idiomorph was identified in the sequences of each of the *T. gauchensis* and *T. zuluensis* genomes (Results section), the opposite idiomorph for each species was identified using long-range PCR. The *MAT* loci from the *T. gauchensis* and *T. zuluensis* genomes were aligned in CLC Genomics Workbench and primers were designed in the conserved flanking regions using Primer3Plus 2.4.2 (Untergasser *et al.*, 2012). Primers Ts_APN-F (5'- CGTAGCTTTTGGATGTGTGC -3') and Ts_g4-R (5'- GTCAAGATCAGCGGTGACG -3') were subsequently used to amplify approximately 4 kb of the complementary *MAT* idiomorph in *T. gauchensis* (CBS119467) and *T. zuluensis* (CBS117262). The long-range PCR was performed with the KAPA LongRange HotStart PCR kit (KAPA Biosystems, Inc., Massachusetts) following the protocol described by Aylward *et al.* (2016), using an annealing temperature of 58°C. The 4 kb PCR products were cleaned and sequenced at Inqaba Biotec (Pretoria, South Africa), using a combination of Sanger sequencing and primer-walking. ORFs were predicted and annotated in these sequences as described above.

MAT distribution in *Eucalyptus* plantations

Two primer sets that target conserved regions in either the *MAT1-1-1* or the *MAT1-2-1* genes, and that amplify products of distinct size for each idiomorph, were designed. T_Ma1-F (5'- GGTCCAGAGCAGTTTGAAGR -3') and T_Ma1-R (5'- AGCCCATCATCTCCTGGTACT -3') target the *MAT1-1-1* alpha-box domain and yield a product of ~450 bp, whereas T_Ma1-1 (5'- TCGCTCTCAGCTCTCCACTT -3') and T_Ma2-2 (5'- GGTCACTCTGATGCCACTTG -3') target the *MAT1-2-1* HMG-box, yielding a ~250 bp product. These primer sets were applied, in a multiplex reaction, to determine the mating type idiomorph present in 47 *T. gauchensis* and 295 *T. zuluensis* isolates, sampled from diseased *Eucalyptus* plantations in 15 countries (Table S1). The multiplex PCR reactions were performed with the Ampliqon *Taq* DNA Polymerase Master Mix RED (Ampliqon, Denmark) and comprised 10 µl Ampliqon

master mix, a final concentration of 2 mM $MgCl_2$, 0.5 µM of each primer (T_Ma1F, T_Ma1-R, T_Ma2-1 and T_Ma2-2), approximately 100 ng of template DNA and water to a final volume of 20 µl. Reaction conditions were 94°C for 3 min, 40 cycles of 94°C for 30 s, 54.5°C for 30 s and 72°C for 45 s and a final extension of 72°C for 10 min.

During development, the *MAT* amplicons from several isolates were sequenced at the Central Analytical Facilities (CAF), Stellenbosch University, South Africa, to confirm amplification of the correct region. In countries where both mating types were present, the hypothesis of a 1:1 ratio was tested with a Pearson's Chi-Square test in R version 3.5.1 (R Core Team, 2018), applying the Benjamini & Hochberg (1995) method for adjusting p-values to account for multiple testing. The identity of "outlier" individuals, *e.g.* a single *MAT1-1* individual identified in a population of otherwise *MAT1-2* individuals, was confirmed by sequencing the ribosomal RNA Internal Transcribed Spacer (ITS) region following the methods proposed by White *et al.* (1990).

Attempts to induce sexual structures

An attempt was made to induce the sexual states of *T. gauchensis* and *T. zuluensis* in culture, even though they have never been observed. For each species, two isolates representing each mating type were paired under laboratory conditions in all six possible combinations. This resulted in four *MAT1-1* x *MAT1-2* crosses and two controls of *MAT1-1* x *MAT1-1* and *MAT1-2* x *MAT1-2*. For each pair, two crossings were performed in separate 65 mm Petri dishes by (i) placing five mm mycelial plugs adjacent to one another and (ii) vigorously mixing mycelium of the two isolates in 250 µl sterile water and spreading the mixture onto the surface of a Petri dish. This experiment was completed on MEA+Y medium, as well as MEA and MEA+Y supplemented with 10% (w/v) dried and finely ground *Eucalyptus grandis* leaves. Plates were incubated for eight weeks at 25°C in the dark.

Results

Teratosphaeria MAT1 locus

The *MAT1-1* proteins of all seven Capnodiales species had significant (e-values < 1.0^{-50}) BLAST hits to a 1000 bp region in the *T. zuluensis* (CBS119470) genome, whereas the *MAT1-2* proteins had significant (e-values < 1.0^{-20}) hits to a region of similar size in the *T. gauchensis* (CBS119467) genome. From the *ca.* 16 kb portion extracted from each candidate scaffold, WebAUGUSTUS consistently predicted six ORFs in *T. zuluensis* and seven in *T. gauchensis*. With the exception of the final ORF that was split in two by the *Botrytis cinerea* gene model, the same ORFs were predicted with all three gene models. The FGENESH predictions, although based on the models of more closely related species, merged two of the *T. zuluensis* flanking ORFs, resulting in an uncharacteristically large gene of > 4 kb. WebAUGUSTUS and *F. graminearum* gene models were applied in subsequent analyses.

Long-range PCR of the putative *MAT* loci of both *T. gauchensis* and *T. zuluensis* yielded amplicons with the expected size of approximately four kb. A MATalpha_HMGbox (PF04769; IPR006856) ("alpha-box") domain was identified in the first predicted ORF of the putative *MAT1-1* idiomorphs. The predicted ORF, however, had a CDS (coding sequence) of 2160 bp, whereas most (~97%) of Ascomycete *MAT1-1-1* coding sequences in GenBank are < 1700 bp. The conserved alpha-box domain lay within the 3' region of the gene and only this C-terminal half of the long *MAT1-1-1* protein was homologous to the *MAT1-1-1* proteins of other Capnodiales species (Supplementary File 2, Fig. S1a). The intron linking the two halves of the gene was subsequently investigated and was unusually long (123 bp in *T. gauchensis* and 122 bp in *T. zuluensis*) compared to the six other predicted introns (47-58 bp). Unlike the other introns, it did not contain an intron branch site (lariat) sequence matching the fungal consensus (Kupfer *et al.*, 2004) (Supplementary File 2, Fig. S2). We, therefore, concluded that this is not a true intron and the large *MAT1-1-1* gene is not a true prediction. Rather, the *Teratosphaeria*

MAT1-1 locus contains two separate ORFs. This hypothesis is a better fit to known *MAT1-1-1* gene models, but nevertheless requires expression data for confirmation.

The *Teratosphaeria MAT1-1* locus (Fig. 1) was comprised of two genes with four exons each. These include a hypothetical *MAT* protein without a conserved domain or known homolog and a putative *MAT1-1-1* gene, similar to those of other Capnodiales species. The 1080 bp *MAT1-1-1* CDS encodes a protein of 360 amino acids. The only *MAT1-1-1* intron conserved in *Teratosphaeria* and other Capnodiales species is the third intron, located within the alpha-box domain (Supplementary File 2, Fig. S1a). The first intron was located upstream of the alpha-box and was absent in the other Capnodiales species, whereas the second intron was within the alpha-box, but six amino acids upstream of the typical Capnodiales intron splice site.

Two ORFs were predicted in the *MAT1-2* locus (Fig. 1). The first was a putative *MAT1-2-1* gene of 1323 bp with two introns and encoding a 441 amino acid protein with an HMG-box domain (PF00505). The second *MAT1-2* intron was within the HMG-box domain and is conserved in all seven Capnodiales species, whereas the first intron was present only in *Teratosphaeria*, *Passalora fulva*, *Pseudocercospora eumusae* and *Ps. fijiensis* (Supplementary File 2, Fig. S1b). The proteins of the remaining species did not align in this region, explaining the lack of an intron. Similar to the *MAT1-1* idiomorph, the *Teratosphaeria MAT1-2* idiomorph also contained a hypothetical *MAT* protein lacking a conserved domain (Fig. 1). The *MAT* loci of both *Teratosphaeria* species have been deposited in GenBank® linked to the following accession numbers: MN119556-MN119559.

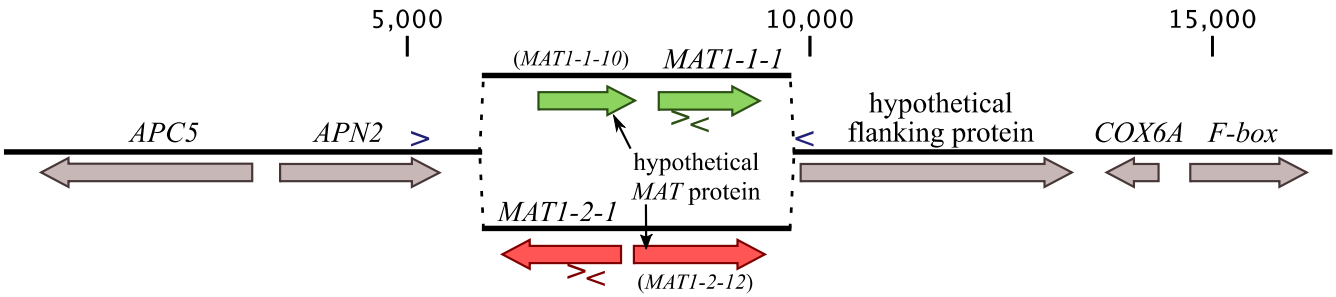


Figure 1 *MAT1* locus of *Teratosphaeria gauchensis* and *T. zuluensis*. The two single black lines represent conserved flanking regions. Each idiomorph comprised a known *MAT* gene (*MAT1-1-1* or *MAT1-2-1*) and an idiomorph-specific hypothetical *MAT* protein (*MAT1-1-10* or *MAT1-2-12*). Arrowheads (< and >) indicate the position of the primer pairs Ts_APN-F and Ts_g4-R in the conserved flanks of the *MAT1* locus, T_Ma1-F and T_Ma1-R in the *MAT1-1-1* gene and T_Ma2-1 and T_Ma2-2 in the *MAT1-2-1* gene.

Interspecific homology between Capnodiales *MAT* genes

The corresponding *MAT* idiomorphs of *T. gauchensis* and *T. zuluensis* were very similar to each other, with a nucleotide identity of > 93% between the respective *MAT1-1* and *MAT1-2* idiomorph alignments. This was higher than the 91.4% and 89.1% nucleotide identity in the *ca.* 5 kb upstream and downstream flanking regions, respectively. The *MAT* idiomorph identity increased to > 97% when only considering the *MAT1-1-1* and *MAT1-2-1* genes and to > 95% when considering the *MAT1-1* hypothetical ORF, whereas identities in the flanking genes remained below 95%. The *MAT1-2-1* hypothetical ORF alignment had a lower identity (92.4%) compared to the identity across the entire idiomorph. The respective *MAT1-1-1* and *MAT1-2-1* proteins were nearly identical with their amino acid identity exceeding 98%. Lower conservation was apparent in the two hypothetical *MAT* proteins, for which the respective amino acid identities were lower than their nucleotide identities. No inter-specific or intra-specific homology was detected between opposite *Teratosphaeria* idiomorphs. The *Teratosphaeria MAT1-1-1* and *MAT1-2-1* proteins had low similarity to the *MAT* proteins of the seven other members of Capnodiales included in this study for comparative purposes. Amino acid identity between *Teratosphaeria* and other Capnodiales species ranged from 20.2 – 47.6% for *MAT1-1-1* and 14.0 – 26.8% for *MAT1-2-1* and protein alignments indicated that sequence conservation between species is largely limited to the conserved alpha- and HMG-box domains (Supplementary File 2; Fig. S1).

Comparison between the hypothetical *MAT* proteins of *Teratosphaeria* and hypothetical proteins (referred to as "MATORFs") found in or associated with the *MAT* loci of *C. beticola*, *Pa. fulva* and *Ps. eumusae*, indicated significant similarity between the *Teratosphaeria MAT1-1* hypothetical ORFs and *MATORF2* genes. Similarly, the *Teratosphaeria MAT1-2* hypothetical ORFs were homologous to the *MATORF1* genes in these species. Pairwise nucleotide identity across the coding sequences of the *Teratosphaeria* hypothetical *MAT* proteins and the *MATORF* genes ranged between 30 and 40%, but decreased to 10 - 17% amino acid identity and 24 - 43%

amino acid similarity (Supplementary File 3, Table S1). Although overall protein similarity between each *Teratosphaeria* hypothetical *MAT* protein and the corresponding Capnodiales MATORF was low, numerous conserved positions in each alignment supported the notion that these proteins are homologous (Supplementary File 3). Additionally, whether the hypothetical *MAT* protein of *Teratosphaeria* lay upstream (in *MAT1-1*) or downstream (in *MAT1-2*) of the primary *MAT* gene corresponded to the position of the putative MATORF homolog in *Cercospora* and in the *MAT1-1* idiomorph of *Pseudocercospora* (Arzanlou *et al.*, 2010, Bolton *et al.*, 2014). We suggest that the *Teratosphaeria MAT1-1* hypothetical ORF (homologous to *MATORF2*) be named *MAT1-1-10* and the *MAT1-2* hypothetical ORF (homologous to *MATORF1*) named *MAT1-2-12*, following the nomenclature for *MAT* genes proposed by Wilken *et al.* (2017).

MAT flanking genes

Three of the ORFs flanking the *Teratosphaeria MAT* locus are commonly associated with the *MAT* locus of Ascomycete fungi. These were the two ORFs upstream of the *MAT* locus, namely *APC5* (anaphase-promoting complex subunit 5) and *APN2* (DNA purinic/apyrimidinic lyase 2) (Conde-Ferráez *et al.*, 2007) and the downstream *COX6A* (Cytochrome c oxidase subunit VIa) (Debuchy & Turgeon, 2006). A hypothetical protein without a detected conserved domain, but with homology to PH-domain (Pleckstrin homology, IPR001849) containing proteins, was found directly downstream of the *Teratosphaeria MAT* locus, and further down a protein with an F-box domain (PF12937). Since *SLA2* (cytoskeleton assembly control) was absent, we conducted tBLASTn searches with the SLA2 protein of *Zymoseptoria tritici* (XP_003847543.1). In *T. zuluensis*, *SLA2* was found on a different contig than the *MAT* locus, but in the contiguous *T. gauchensis* genome assembly, *SLA2* occurred on the same contig, 1.2 Mb from the *MAT* locus.

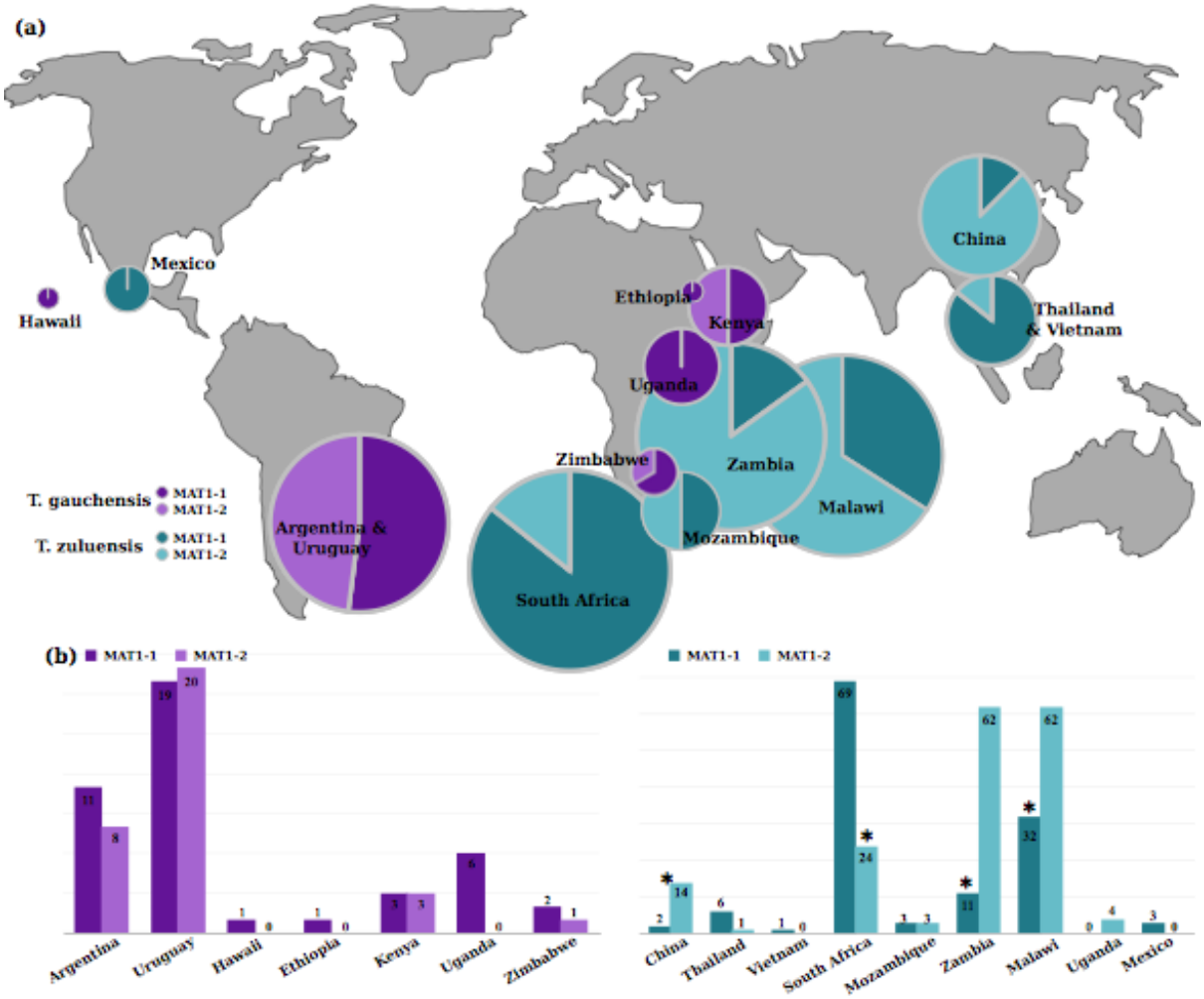


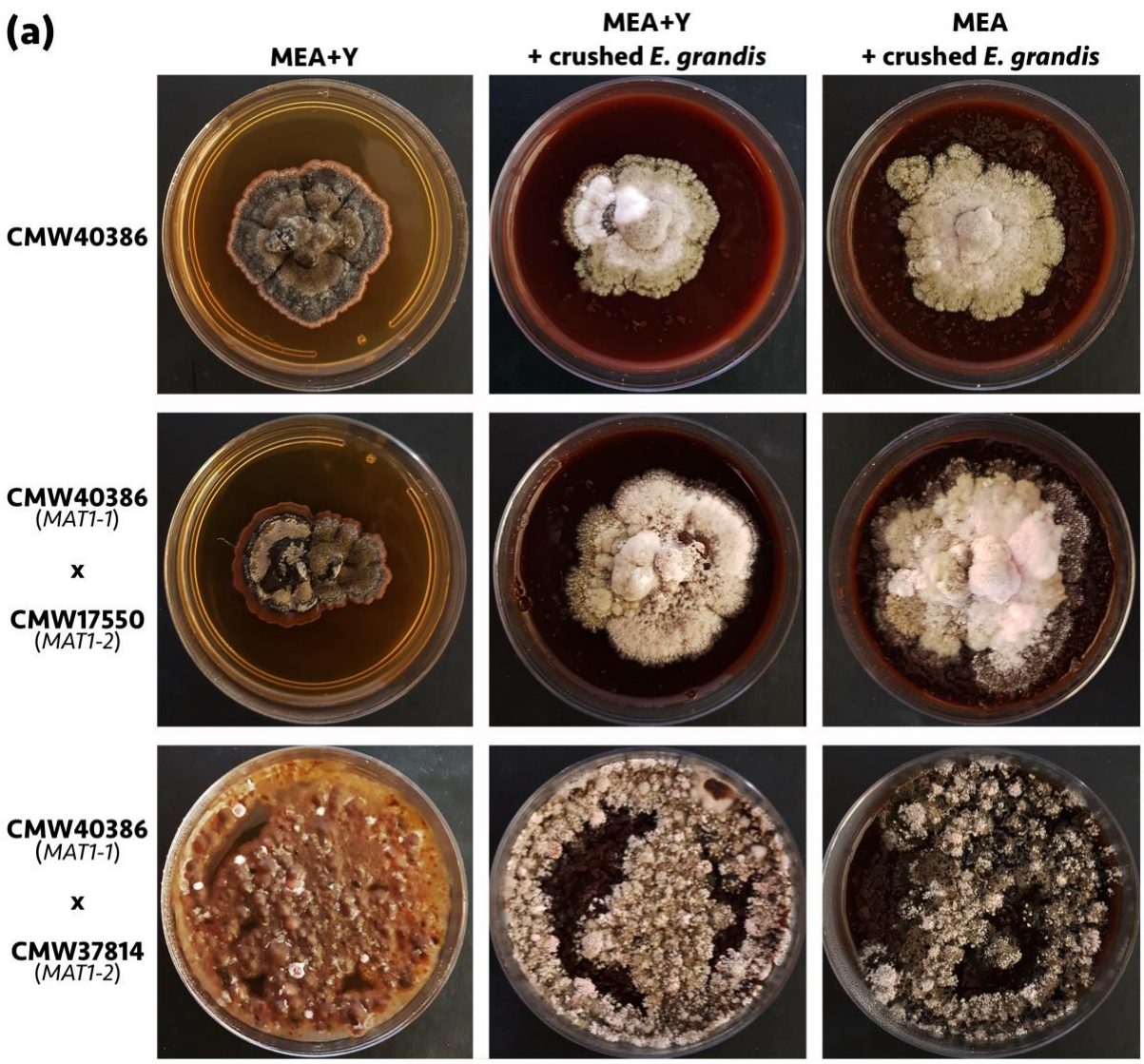
Figure 2 Distribution of *Teratosphaeria gauchensis* and *T. zuluensis* mating types. The size of each pie chart (a) corresponds to the number of tested isolates. Bar plots for *T. gauchensis* (b) and *T. zuluensis* (c) indicate the number of isolates of each mating type per country. Asterisks denote a statistically significant difference in the mating type ratio ($p < 0.05$).

MAT1-1:*MAT1-2* ratio of isolates from plantations

MAT1-1 and *MAT1-2* isolates were found to co-occur in the majority of countries for which samples of *T. gauchensis* and *T. zuluensis* were available (Fig. 2). Approximately equal frequencies of the two *T. gauchensis* mating types were found in Argentina, Kenya, and Uruguay (Chi-Square p-value > 0.05). Most Argentinian and Kenyan isolates were sampled from the same year and locality, but Uruguayan samples represented six areas and four different sampling years (Supplementary File 1). Some of these areas (e.g. Cofusa) appeared to have a dominant mating type, but too few isolates were available to form a trend. Low numbers of isolates were available from Africa and represented predominantly the *MAT1-1* idiomorph. Three *MAT1-2* individuals were identified from Kenya, one from Zimbabwe and none from Uganda.

Teratosphaeria zuluensis isolates from China, Malawi, Mozambique, South Africa and Zambia also included both mating types. In contrast to *T. gauchensis*, however, all countries for which > 10 *T. zuluensis* isolates were available had ratios deviating significantly from 1:1 (Chi-Square p-value < 0.05). The skewed ratio was especially apparent in populations from southern Africa, with *MAT1-1* dominating in South Africa and *MAT1-2* most common in Malawi and Zambia. Whereas most of the Malawian and Zambian isolates represented samples taken from the same site in one year, the South African samples were taken from 12 different sites between 1997 and 2012 (Supplementary File 1). With the exception of three 1997 sampling sites with less than six isolates each, the *MAT1-1* idiomorph was also most numerous within sites.

For both *T. gauchensis* and *T. zuluensis* the availability of isolates from some locations was limited, but those available were included for comparative purposes. The *T. gauchensis* isolate from Ethiopia and Hawaii were both of the *MAT1-1* type. The single *T. zuluensis* isolate from Vietnam and three Mexican isolates were *MAT1-1*, with four *MAT1-2 T. zuluensis* isolates from Uganda. To the best of our knowledge, Uganda is the only country where *T. gauchensis* and *T.*


(a)
MEA+Y
MEA+Y
+ crushed E. grandis
MEA
+ crushed E. grandis
CMW40386
CMW40386
(MAT1-1)
x
CMW17550
(MAT1-2)
CMW40386
(MAT1-1)
x
CMW37814
(MAT1-2)

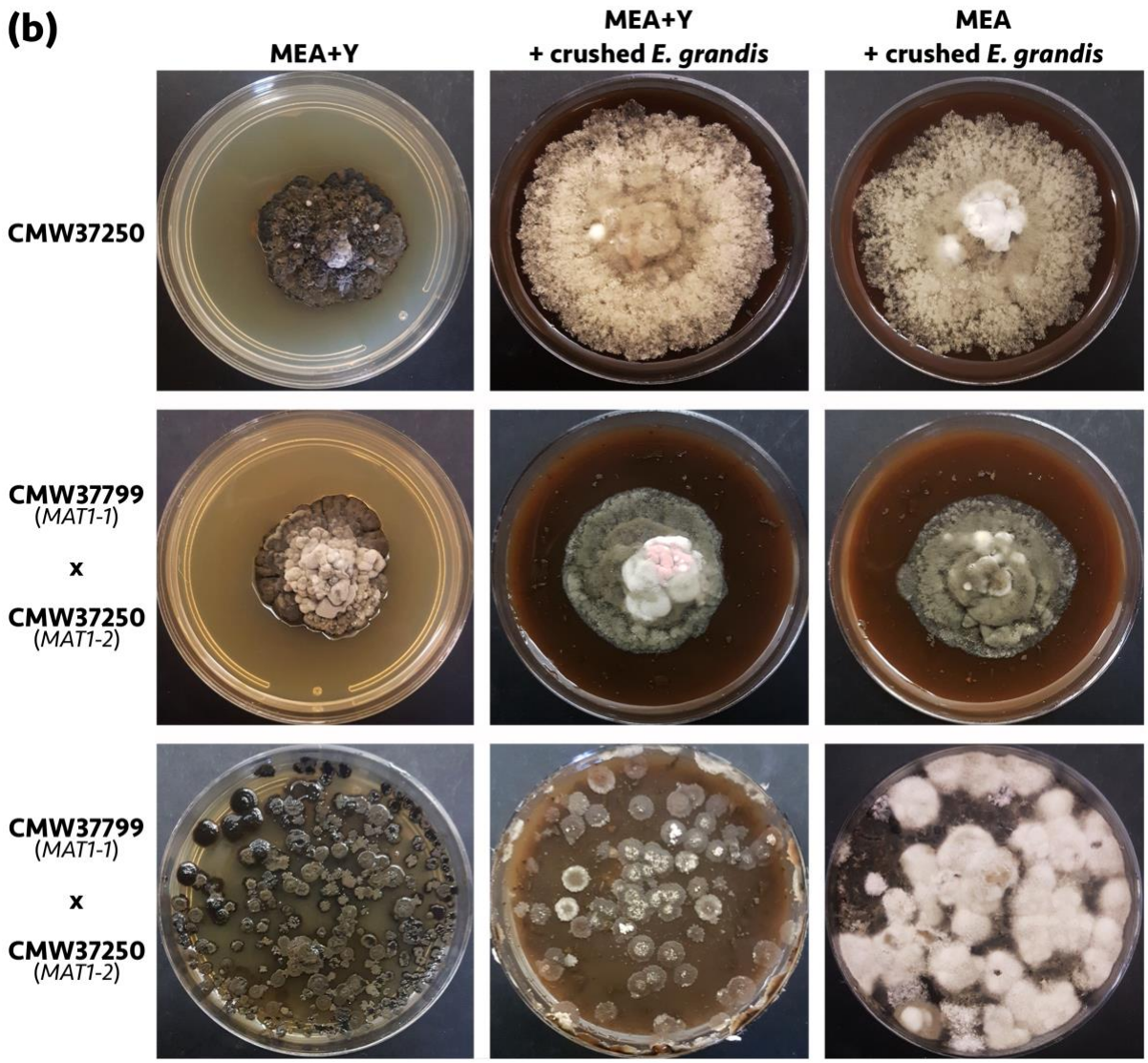


Figure 3 Attempted crossings of (a) *Teratosphaeria gauchensis* and (b) *T. zuluensis* on three different media types. The first row shows the growth of a single individual. The second and third rows, respectively, show crossing of opposite mating types by placing 5 mm agar plugs adjacent to each other and mixing mycelia. Interactions are similar to those of the *MAT1-1*x*MAT1-1* and *MAT1-2*x*MAT1*-2 controls (not shown). First column (MEA+Y) = Malt Extract Agar supplemented with yeast; second column = MEA+Y, supplemented with crushed *Eucalyptus grandis* leaves; third column = MEA + crushed *E. grandis* leaves.

zuluensis co-occur and, notably, opposite mating types of these species were identified in Ugandan plantations.

Attempts to induce sexual structures

None of the *in vitro* crosses between isolates of *T. gauchensis* or *T. zuluensis* showed signs of a mating interaction after eight weeks of incubation (Fig. 3). The *MAT1-1* x *MAT1-2* crosses of each species showed a similar vegetative growth pattern to the *MAT1-1* x *MAT1-1*- and *MAT1-2* x *MAT1-2* controls. Visualisation of overlapping growth under a light microscope did not reveal sexual or even asexual spores and confirmed the absence of a mating reaction. Additionally, vegetative incompatibility interactions were also absent and the cultures in all pairings grew freely towards each other and eventually merged.

Discussion

Teratosphaeria gauchensis and *T. zuluensis* are economically important tree pathogens lacking known sexual states. Population genetic studies using microsatellite markers have suggested that recombination could be taking place in both species at low levels (Cortinas *et al.*, 2010, 2011, Jimu *et al.*, 2016a), but their ability to reproduce sexually has not previously been explored. In this study, we confirmed that the two stem canker pathogens have a heterothallic mating system and that sexual recombination would be possible in areas where individuals of both mating type co-occur. There is, however, no physical evidence to suggest that recombination is taking place in diseased *Eucalyptus* plantations. Additionally, comparisons of the distribution of mating types and the genetic diversity in different countries (see below) showed that recombination is unlikely to occur. Nevertheless, the absence of sexual structures

in these species does not preclude their existence. It is feasible that these structures might yet be identified in nature or induced in the laboratory where conditions make this possible.

Characterisation of the *MAT* loci in *T. gauchensis* and *T. zuluensis* represents the first time that *MAT* idiomorphs have been characterised in any member of the Teratosphaeriaceae. In contrast, *MAT* loci have been characterised in species representing five genera of Dothideomycetes (Capnodiales), namely *Cercospora*, *Dothistroma*, *Passalora*, *Pseudocercospora* and *Zymoseptoria*, which all reside in the Mycosphaerellaceae (Quaedvlieg *et al.*, 2014). Like *T. gauchensis* and *T. zuluensis*, these species also have heterothallic mating systems (Waalwijk *et al.*, 2002, Groenewald *et al.*, 2006, Conde-Ferráez *et al.*, 2007, Groenewald *et al.*, 2007, Stergiopoulos *et al.*, 2007).

Species in the Mycosphaerellaceae can have two hypothetical MATORFs in an idiomorph and signatures of inverted homology between idiomorphs (Arzanlou *et al.*, 2010). In comparison, the structure of the *Teratosphaeria MAT* locus appears uncomplicated. Each of the *Teratosphaeria* idiomorphs comprised a single defining *MAT* gene (*MAT1-1-1* or *MAT1-2-1*) accompanied by one idiomorph-specific hypothetical *MAT* protein (*MAT1-1-10* or *MAT1-2-12*). The genomic position of the locus was conserved, being associated with the *APC5* and *APN2* genes upstream of the locus. This is similar to all known Mycosphaerellaceae examples (Stergiopoulos *et al.*, 2007, Arzanlou *et al.*, 2010, Bolton *et al.*, 2014). The downstream flanking genes of the Mycosphaerellaceae have not been fully described, but *Teratosphaeria* contains a *COX6a* gene in this region that is commonly associated with the *MAT* locus in other Ascomycetes (*e.g*. Simpson *et al.*, 2018). Although the *SLA2* gene is commonly found next to the *MAT* locus in Sordariomycete fungi (Debuchy & Turgeon, 2006), it was not found in the vicinity of the *Teratosphaeria MAT* locus. Other studies have also reported an absence of this gene around the *MAT* locus in species of Capnodiales (Conde-Ferráez *et al.*, 2007). Analysis of the *T. gauchensis* genome showed that the *SLA2* gene is > 1 Mb from the *MAT* locus.

There was a low level of similarity between the *Teratosphaeria* and Mycosphaerellaceae *MAT* genes, as described by nucleotide and protein identity. Nevertheless, the *MAT1-1-1* and *MAT1-2-1* genes of the studied species were very similar in the region of the conserved alpha- and HMG-box domains, respectively. One of the intron positions was also conserved across all species in both genes. In *MAT1-1-1*, the position of intron 1 differed between the two groups, whereas the third intron was missing in *Teratosphaeria*, something that is also true for *Zymoseptoria* (Groenewald *et al.*, 2006).

Both *Teratosphaeria MAT* idiomorphs contained hypothetical *MAT* proteins with similarity to the *MATORF1* and *MATORF2* genes previously predicted in some, although not all, of the studied Mycosphaerellaceae species. In the Mycosphaerellaceae, these MATORFs are unique in that they can occur in both the *MAT1-1* and *MAT1-2* idiomorphs, *e.g*. in *Pseudocercospora* (Arzanlou *et al.*, 2010) and *Pa. fulva* (Stergiopoulos *et al.*, 2007), or form part of the conserved flanking regions of the *MAT* locus, *e.g*. in *Cercospora* (Bolton *et al.*, 2014). The association of these genes with the *MAT* locus in the Mycosphaerellaceae requires further investigation, but their location within the *Teratosphaeria MAT* locus is clear. In this study we have provided them with the notations *MAT1-1-10* and *MAT1-2-12*.

The four geographic areas for which a reasonable number *T. zuluensis* isolates were available for meaningful mating type comparisons displayed an obviously skewed ratio between *MAT1-1* and *MAT1-2* isolates. Isolates from South Africa, where this pathogen was first discovered (Wingfield *et al.*, 1996), were predominantly of the *MAT1-1* type. This is in contrast to those from China and east Africa that were predominated by *MAT1-2* individuals. The different dominant mating types in South Africa compared to other African countries was congruent with the population differentiation between these regions reported by Jimu *et al.* (2016a). A single mating type was recovered from each of the three remaining areas (Mexico, Uganda and Vietnam) considered, but this result was based on a limited number of isolates and could represent an anomaly. The single Vietnamese isolate was of the *MAT1-1* type, similar to isolates

from Thailand. This suggests a dominance of *MAT1-1 T. zuluensis* individuals on the Indochina Peninsula.

Areas that were represented by very low sample sizes of *T. gauchensis* (Ethiopia, Hawaii and Zimbabwe) had predominantly *MAT1-1* individuals, which was also the dominant mating type in north-east Africa. Although Kenyan isolates showed that the *MAT1-2* genotype is also present in this region, the dominance of one mating type in north-east Africa is congruent with the high levels (> 90%) of clonality in Ethiopia and Uganda, described by Jimu *et al.* (2016b). Contrary to our expectations and also in contrast to the case for *T. zuluensis*, the ratio of the two mating types was not skewed in the two larger South American *T. gauchensis* populations. These populations showed a lower incidence of clonality (<50%), as well as an even distribution of haplotypes (Cortinas *et al.*, 2011, Jimu *et al.*, 2016b). The maximum genotypic diversity (approximately 50%) was lower than for some *T. zuluensis* populations, but nevertheless supports a suggestion (Cortinas *et al.*, 2010, 2011) that sexual recombination could be taking place.

Randomly mating populations of heterothallic species are expected to comprise equal frequencies of the two mating types (Milgroom, 1996). The skewed distribution of *T. zuluensis* mating types, therefore, suggests a low likelihood of recombination in *Eucalyptus* plantations where this species causes disease. Some southern Chinese populations of *T. zuluensis* and isolates collected from Malawi, however, showed high levels of genotypic (84-100%) and gene (3–14 genes/locus) diversity (Cortinas *et al.*, 2010, Chen *et al.*, 2011), much higher than the ~50% genotypic diversity and 2–8 genes/locus reported for *T. gauchensis* (Cortinas *et al.*, 2011). Despite this diversity, the index of association indicated only weak and marginally significant recombination in one Chinese population (Cortinas *et al.*, 2010, Chen *et al.*, 2011). Combined with the clear deviation from a 1:1 mating ratio, this suggests that the contribution of recombination to the overall genetic structure of *T. zuluensis* populations in diseased *Eucalyptus* plantations could be minimal.

The results of this study suggest a difference in the frequency of recombination between *T. gauchensis* and *T. zuluensis*. These pathogens are closely related, represent the only known stem canker pathogens in their otherwise leaf-associated genus and produce indistinguishable disease symptoms (Aylward *et al.*, 2019). We thus also expected the pattern of their mating type distribution to be similar. The even distribution of mating types in *T. gauchensis* compared to the skewed distribution in *T. zuluensis* was thus surprising. However, this disparity is not necessarily due to biology. The genetics of both species indicated that recombination could be possible where these species are native, but in *Eucalyptus* plantations, the genetic diversity and mating type distribution may rather reflect different patterns of introduction and re-introduction across the globe.

Separate multiple introductions of each species could explain the difference between the mating type distribution of *T. gauchensis* and *T. zuluensis* observed in this study. For *T. zuluensis*, one mating type could have been introduced at a higher frequency than the other, perhaps sourced from other diseased populations. Following this hypothesis, introduced strains of *T. gauchensis* would have comprised similar proportions of the two mating types, reflecting the possibility of a randomly recombining source population (Milgroom, 1996). An alternative explanation for the *T. zuluensis* distribution is that one mating type was able to outcompete the other. Such competition between mating types due to different fitness levels is known in, for example, *Ceratocystis albifundus*, where *MAT1-2* individuals grow faster than *MAT1-1* individuals (Lee *et al.*, 2015). However, the similar proportions of *T. gauchensis* mating types and the different mating types of *T. zuluensis* that dominate in different areas, make it unlikely that a difference in fitness has played a role.

Multiple introductions could also account for the difference in gene and genotypic diversity of *T. gauchensis* and *T. zuluensis* populations. The mating type distribution that has emerged from this study suggests that recombination, and therefore genetic diversity, should be highest in *T. gauchensis*. In contrast, in *T. gauchensis* the genetic diversity is lower than in some *T. zuluensis*

populations (Cortinas *et al.*, 2010, Chen *et al.*, 2011, 2011). This suggests that, together with some possible level of recombination, multiple introductions of different genotypes have played an important role in defining the genetic diversity of *T. zuluensis* and *T. gauchensis* populations in *Eucalyptus* plantations. The fact that these pathogens can apparently be moved with seed (Jimu *et al.*, 2016c), which is a widely traded source of germplasm for forestry companies (Koskela *et al.*, 2014), would have facilitated this situation.

Identification of the area or origin of *T. gauchensis* and *T. zuluensis,* and thus collections of isolates from native populations, would allow for a much deeper understanding of their biology. Surveys have not yet detected these species in natural forests in either Australia or on native Myrtaceae elsewhere in the world (Pérez *et al.*, 2013, Burgess & Wingfield, 2017). Although we believe that recombination plays a small role where these species infect *Eucalyptus* plantations, the possibility of inter-specific recombination (hybridisation) between the two species also exists. Opposite mating types of *T. gauchensis* and *T. zuluensis* co-occur in Uganda and there is consequently a risk of hybridisation. This could lead to increased disease severity or the ability to withstand a wider range of environmental conditions (Aylward *et al.*, 2019). The potential of these species to recombine, at both an inter- and intra-specific level, should be further investigated.

Acknowledgements

We thank members of the Tree Protection Co-operative program (TPCP), the Department of Science and Technology (DST) - National Research Foundation (NRF) Centre of Excellence in Tree Health Biotechnology (CTHB) and the SARChI chair in Fungal Genomics for financial support.

References

Arzanlou M, Crous PW, Zwiers L-H, 2010. Evolutionary dynamics of mating-type loci of *Mycosphaerella* spp. occurring on banana. *Eukaryotic Cell* **9**, 164-72.

Aylward J, Roets F, Dreyer LL, Wingfield MJ, 2019. Teratosphaeria stem canker of *Eucalyptus*: two pathogens, one devastating disease. *Molecular Plant Pathology* **20**, 8-19.

Aylward J, Steenkamp ET, Dreyer LL, Roets F, Wingfield MJ, Wingfield BD, 2016. Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*. *Fungal Genetics and Biology* **96**, 47-57.

Bateman A, Coin L, Durbin R, *et al.*, 2004. The Pfam protein families database. *Nucleic Acids Research* **32**, D138-D41.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300.

Bolton MD, De Jonge R, Inderbitzin P, *et al.*, 2014. The heterothallic sugarbeet pathogen *Cercospora beticola* contains exon fragments of both *MAT* genes that are homogenized by concerted evolution. *Fungal Genetics and Biology* **62**, 43-54.

Burgess TI, Wingfield MJ, 2017. Pathogens on the Move: A 100-Year Global Experiment with Planted Eucalypts. *Bioscience* **67**, 14-25.

Chen SF, Barnes I, Chungu D, *et al.*, 2011. High population diversity and increasing importance of the *Eucalyptus* stem canker pathogen, *Teratosphaeria zuluensis*, in South China. *Australasian Plant Pathology* **40**, 407.

Conde-Ferráez L, Waalwijk C, Canto-Canché BB, *et al.*, 2007. Isolation and characterization of the mating type locus of *Mycosphaerella fijiensis*, the causal agent of black leaf streak disease of banana. *Molecular Plant Pathology* **8**, 111-20.

Cortinas MN, Barnes I, Wingfield BD, Wingfield MJ, 2011. Unexpected genetic diversity revealed in the *Eucalyptus* canker pathogen *Teratosphaeria gauchensis*. *Australasian Plant Pathology* **40**, 497-503.

Cortinas MN, Barnes I, Wingfield MJ, Wingfield BD, 2010. Genetic diversity in the *Eucalyptus* stem pathogen *Teratosphaeria zuluensis*. *Australasian Plant Pathology* **39**, 383-93.

Cortinas MN, Crous PW, Wingfield BD, Wingfield MJ, 2006. Multi-gene phylogenies and phenotypic characters distinguish two species within the *Colletogloeopsis zuluensis* complex associated with *Eucalyptus* stem cankers. *Studies in Mycology* **55**, 133-46.

Crous PW, Groenewald JZ, Summerell BA, Wingfield BD, Wingfield MJ, 2009. Co-occurring species of *Teratosphaeria* on Eucalyptus. *Persoonia* **22**, 38-48.

Damm U, Verkley GJM, Crous PW, Fourie PH, Haegi A, Riccioni L, 2008. Novel *Paraconiothyrium* species on stone fruit trees and other woody hosts. *Persoonia* **20**, 9-17.

Debuchy R, Turgeon BG, 2006. Mating-type structure, evolution, and function in Euascomycetes. In: Kües U, Fischer R, eds. *The Mycota I Growth, Differentiation and Sexuality*. Berlin: Springer, 293-323.

Gezahgne A, Roux J, Wingfield MJ, 2003. Diseases of exotic plantation *Eucalyptus* and *Pinus* species in Ethiopia. *South African Journal of Science* **99**, 29-33.

Goodwin SB, Waalwijk C, Kema GHJ, Cavaletto JR, Zhang G, 2003. Cloning and analysis of the mating-type idiomorphs from the barley pathogen *Septoria passerinii*. *Molecular Genetics and Genomics* **269**, 1-12.

Groenewald M, Barnes I, Bradshaw RE, *et al.*, 2007. Characterization and distribution of mating type genes in the *Dothistroma* needle blight pathogens. *Phytopathology* **97**, 825-34.

Groenewald M, Groenewald JZ, Harrington TC, Abeln ECA, Crous PW, 2006. Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of cryptic sex. *Fungal Genetics and Biology* **43**, 813-25.

Hoff KJ, Stanke M, 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**, W123-W8.

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

Jimu L, Chen SF, Wingfield MJ, Mwenje E, Roux J, 2016b. The *Eucalyptus* stem canker pathogen *Teratosphaeria gauchensis* represents distinct genetic groups in Africa and South America. *Forest Pathology* **46**, 229-39.

Jimu L, Kemler M, Wingfield MJ, Mwenje E, Roux J, 2016c. The *Eucalyptus* stem canker pathogen *Teratosphaeria zuluensis* detected in seed samples. *Forestry* **9**, 316-24.

Jimu L, Wingfield MJ, Mwenje E, Roux J, 2014. First report of *Teratosphaeria zuluensis* causing stem canker of *Eucalyptus grandis* in Uganda. *Forest Pathology* **44**, 242-5.

Jimu L, Wingfield MJ, Mwenje E, Roux J, 2015. Diseases on *Eucalyptus* species in Zimbabwean plantations and woodlots. *Southern Forests* **77**, 221-30.

Koskela J, Vinceti B, Dvorak W, *et al.*, 2014. Utilization and transfer of forest genetic resources: A global review. *Forest Ecology and Management* **333**, 22-34.

Kronstad JW, Staben C, 1997. Mating type in filamentous fungi. *Annual review of genetics* **31**, 245-76.

Kupfer DM, Drabenstot SD, Buchanan KL, *et al.*, 2004. Introns and splicing elements of five diverse fungi. *Eukaryotic Cell* **3**, 1088-100.

Lee DH, Roux J, Wingfield BD, Wingfield MJ, 2015. Variation in growth rates and aggressiveness of naturally occurring self-fertile and self-sterile isolates of the wilt pathogen *Ceratocystis albifundus*. *Plant Pathology* **64**, 1103-9.

Maynard Smith J, Smith NH, O'rourke M, Spratt BG, 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences* **90**, 4384-8.

Milgroom MG, 1996. Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* **34**, 457-77.

Old KM, Wingfield MJ, Yuan ZQ, 2003. *A manual of diseases of Eucalypts in South-East Asia*. Jakarta, Indonesia: CIFOR.

Pérez CA, Wingfield MJ, Altier N, Blanchette RA, 2013. Species of Mycosphaerellaceae and Teratosphaeriaceae on native Myrtaceae in Uruguay: evidence of fungal host jumps. *Fungal Biology* **117**, 94-102.

Quaedvlieg W, Binder M, Groenewald JZ, *et al.*, 2014. Introducing the Consolidated Species Concept to resolve species in the *Teratosphaeriaceae*. *Persoonia* **33**, 1-40.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Simpson MC, Coetzee MP, Van Der Nest MA, Wingfield MJ, Wingfield BD, 2018. *Ceratocystidaceae* exhibit high levels of recombination at the mating-type (*MAT*) locus. *Fungal Biology* **122**, 1184-91.

Solovyev V, Kosarev P, Seledsov I, Vorobyev D, 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biology* **7**, S10.

Stergiopoulos I, Groenewald M, Staats M, Lindhout P, Crous PW, De Wit PJGM, 2007. Mating-type genes and the genetic structure of a world-wide collection of the tomato pathogen *Cladosporium fulvum*. *Fungal Genetics and Biology* **44**, 415-29.

Turgeon BG, Yoder OC, 2000. Proposed Nomenclature for Mating Type Genes of Filamentous Ascomycetes. *Fungal Genetics and Biology* **31**, 1-5.

Untergasser A, Cutcutache I, Koressaar T, *et al.*, 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115-e.

Waalwijk C, Mendes O, Verstappen ECP, De Waard MA, Kema GHJ, 2002. Isolation and characterization of the mating-type idiomorphs from the wheat septoria leaf blotch fungus *Mycosphaerella graminicola*. *Fungal Genetics and Biology* **35**, 277-86.

White TJ, Bruns T, Lee S, Taylor JL, 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds. *PCR protocols: a guide to methods and applications.* San Diego, California: Academic Press, 315-22. (18.)

Wilken PM, Steenkamp ET, Wingfield MJ, De Beer ZW, Wingfield BD, 2017. Which *MAT* gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. *Fungal Biology Reviews* **31**, 199-211.

Wilson AM, Wilken PM, Van Der Nest MA, Steenkamp ET, Wingfield MJ, Wingfield BD, 2015. Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus* **6**, 207-14.

Wingfield BD, Fourie A, Simpson MC, *et al.*, in press. IMA Genome-F 11 draft genome sequences of *Fusarium xylarioides*, *Teratosphaeria gauchensis* and *T. zuluensis* and genome annotation for *Ceratocystis fimbriata*. *IMA Fungus* **10**.

Wingfield MJ, Crous PW, Coutinho TA, 1996. A serious canker disease of *Eucalyptus* in South Africa caused by a new species of *Coniothyrium*. *Mycopathologia* **136**, 139-45.

Yilmaz N, Hagen F, Meis JF, Houbraken J, Samson RA, 2016. Discovery of a sexual cycle in *Talaromyces amestolkiae*. *Mycologia* **108**, 70-9.

Data Availability

The data that support the findings of this study are openly available in GenBank® at https://www.ncbi.nlm.nih.gov/genbank, genome accession numbers VCMR00000000 and VCMQ00000000, *MAT* locus accession numbers MN119556-MN119559.