Comparative genomics of *Knoxdaviesia* species in the Core Cape Subregion

by
Janneke Aylward

Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Science at Stellenbosch University

Supervisor: Prof. Léanne L. Dreyer
Co-supervisors:
Dr. Francois Roets
Prof. Emma T. Steenkamp
Prof. Brenda D. Wingfield
Prof. Michael J. Wingfield

March 2017

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Janneke Aylward Date: March 2017

ABSTRACT

Knoxdaviesia capensis and *K. proteae* are saprotrophic fungi that inhabit the seed cones (infructescences) of *Protea* plants in the Core Cape Subregion (CCR) of South Africa. Arthropods, implicated in the pollination of *Protea* species, disperse these native fungi from infructescences to young flower heads (inflorescences). *Knoxdaviesia proteae* is a specialist restricted to one *Protea* species, while the generalist *K. capensis* occupies a range of *Protea* species. Within young flower heads, *Knoxdaviesia* species grow vegetatively, but switch to sexual reproduction once flower heads mature into enclosed infructescences. Nectar becomes depleted and infructescences are colonised by numerous other organisms, including the arthropod vectors of the fungi.

The aim of this dissertation was to study the ecology of *K. capensis* and *K. proteae* by making use of their genome sequences. *Knoxdaviesia* belongs to the family Gondwanamycetaceae, for which no genomes were available at the start of this study. Using Illumina technology, we determined the genome sequences of both CCR *Knoxdaviesia* species and applied them to investigate reproduction, substrate use and tolerance to competition.

Population genetic studies on *K. capensis* and *K. proteae* have revealed massive diversity, suggesting an outcrossing reproductive strategy. To determine the genetic basis for this diversity, we used the genomes to identify and characterise their mating type (*MAT*) loci. Each species contained only a single *MAT* idiomorph per isolate, indicating that they require an individual of the opposite mating type for sexual reproduction. The *MAT* loci of the two *Knoxdaviesia* species were almost identical, reflecting their phylogenetic relatedness. The features of the *Knoxdaviesia MAT1-2-7* gene also suggested a historic recombination event between the *MAT1-1* and *MAT1-2* idiomorphs.

The carbon resources that *Knoxdaviesia* species utilise were investigated with phenome assays and compared to the carbon usage profile of a *Protea* pathogen, *Ceratocystis albifundus*. *Knoxdaviesia capensis*, the generalist, utilised the widest range of substrates, whereas the pathogen utilised the least. The *Knoxdaviesia* species were able to grow on all monosaccharides that occur in *Protea* nectar. The predicted proteins in the *Knoxdaviesia* and *C. albifundus* genomes suggested that cell wall degradation is important to the nutrition of *Knoxdaviesia* species in infructescences, whereas the pathogen prefers plant storage

polysaccharides. Overall, carbon metabolism in three ecologically different, but related fungi reflected their ecological adaptations.

Knoxdaviesia species appear to be effective competitors in infructescences. Few secondary metabolite biosynthesis clusters were, however, detected from the *K. capensis* and *K. proteae* genomes. This may suggest that it is the antimicrobial products of *Streptomyces* bacteria that rid infructescences of fungal competitors. The few secondary metabolite clusters of the *Knoxdaviesia* species likely produce compounds that enable them to tolerate these bacteria and arthropod and nematode predation. Proteins involved in cell defence were also detected among the predicted secreted proteins of *K. capensis* and *K. proteae*. *Knoxdaviesia proteae* appears to have some non-functional secondary metabolite clusters and secretes less cell defence proteins than *K. capensis*, suggesting that its specialisation on a single host has resulted in the loss of some functions.

OPSOMMING

Knoxdaviesia capensis en *K. proteae* is saprofitiese fungi wat die saadkeëls van *Protea* plante in die Kern Kaapse Substreek (KKS) van Suid-Afrika bewoon. Artropodes, geïmpliseer in die bestuiwing van *Protea* spesies, verprei hierdie inheemse fungi van saadkeëls na jong blomkoppe. *Knoxdaviesia proteae* is 'n spesialis wat beperk is tot een *Protea* spesie, terwyl die meer algemene *K. capensis* in 'n reeks *Protea* spesies voorkom. Binne jong blomkoppe groei *Knoxdaviesia* spesies vegetatief, maar skakel om na seksuele voortplanting sodra die blomkoppe tot geslote saadknoppe ontwikkel. Nektar raak uitgeput, en saadknoppe word gekoloniseer deur verskeie ander organismes, insluitend die artropode vektore van die fungi.

Die mikpunt van hierdie dissertasie was om die ekologie van *K. capensis* en *K. proteae* te bestudeer deur van genoom basisvolgordes gebruik te maak. *Knoxdaviesia* behoort aan die familie Gondwanamycetaceae, waarvoor geen genome beskikbaar was aan die begin van hierdie studie nie. Deur van Illumina tegnologie gebruik te maak het ons die genoom basisvolgordes van beide KKS *Knoxdaviesia* spesies bepaal, en hulle toegepas om die reproduksie, substraat-gebruik en toleransie teen kompetisie te ondersoek.

Populasie genetiese studies op *K. capensis* en *K. proteae* het massiewe diversiteit geopenbaar, wat 'n kruistelende reproduktiewe strategie voorstel. Om die genetiese basis vir hierdie diversiteit te bepaal het ons die genome gebruik om hul paringsmaat-tipe (*MAT*) lokusse te bepaal en te karakteriseer. Elke spesie het slegs een *MAT* idiomorf per isolaat bevat, wat aandui dat hulle 'n individu van die teenoorgestelde paringstipe benodig vir seksuele voortplanting. Die *MAT* lokusse van die twee *Knoxdaviesia* spesies was amper identies, wat hulle filogenetiese verwantskap reflekteer. Die kenmerke van die *Knoxdaviesia MAT1-2-7* geen suggereer ook 'n historiese rekombinasie gebeurtenis tussen die *MAT1-1* en *MAT1-2* idiomorfe.

Die koolstofbronne wat *Knoxdaviesia* spesies gebruik is ondersoek met fenoom toetse en vergelyk met die koolstofgebruik-profiel van 'n *Protea* patogeen, *Ceratocystis albifundus*. Die algemene *Knoxdaviesia capensis* het die wydste reeks substrate gebruik, terwyl die patogeen minste gebruik het. Die *Knoxdaviesia* spesies kon groei op al die monosakkariede wat in *Protea* nektar voorkom. Die voorspelde proteïene in die *Knoxdaviesia* en *C. albifundus* genome stel voor dat selwand degradasie belangrik is vir voeding van *Knoxdaviesia* spesies in saadkoppe, terwyl die patogeen plant storingspolisakkariede verkies. Oor die algemeen,

reflekteer die koolstofmetabolisme in drie ekologies verskillend, maar verwante fungi hulle ekologiese aanpassings.

Dit blyk of *Knoxdaviesia* effektiewe kompeteerders is in saadknoppe. Min sekondêre metaboliet biosintese groeperings is egter in die *K. capensis* en *K. proteae* genome opgespoor. Dit mag suggereer dat dit die antimikrobiese produkte van *Streptomyces* bakterieë is wat saadknoppe van fungus-kompeteerders ontneem. Die paar sekondêre metaboliet biosintese groeperings van die *Knoxdaviesia* spesies produseer waarskynlik verbindings wat hulle in staat stel om hierdie bakterieë sowel as artropode- en nematode-predasie te verdra. Proteïene betrokke by selbeskerming is ook waargeneem tussen die voorspelde gesekreteerde proteïene van *K. capensis* en *K. proteae*. *Knoxdaviesia proteae* blyk ook om nie-funksionele sekondêre metaboliet groeperings te hê, en sekreteer minder selbeskermingsproteïene as *K. capensis*, wat voorstel dat sy spesialisasie op ’n enkele gasheer die verlies van sekere funksies tot gevolg gehad het.

ACKNOWLEDGEMENTS

Like me, this dissertation is a product of grace. Were it not for those mentioned below, this document would not exist. I appreciate you all immensely.

Firstly, my sincere thanks to the National Research Foundation (NRF) and the Department of Science and Technology-NRF Centre of Excellence in Tree Health Biotechnology for providing research funding and personal financial support.

I am convinced that I have the best supervision team. Other students often offer their sympathies when they find out I have five supervisors, but Léanne, Francois, Emma, Brenda and Mike: thank you for showing me what professional relationships look like. Thank you for treating me like a colleague and thank you for investing time to develop me into a researcher.

When I started this project, I had little experience in genomics, but Lieschen Bahlman and Quentin Santana gave support and hours of patience! Lieschen, thank you for your guidance during the couple of days I sat in your office. Quentin, thank you for everything you taught me and the technical help you were always willing to provide.

On a more personal note, I have been blessed with friends and family who played a big part in my emotional stability, especially during the final few months. My husband Ernie, I love being on your team! Thank you for being a patient listener (as evidenced by your intimate knowledge of PCR) and for being steadfast. Our friends, some of who know the PhD journey first hand: thank you for good coffee, always being interested, always asking us to stay for dinner. Nandré and Wiehan, I could not ask for better siblings. Not long ago, Nandré said to me: “People don’t know what they’re getting into when they ask me about my sister.” Your underserved admiration is overwhelming. My parents, Wynand and Riétte, thank you for your prayers and support. Thank you for the peace and safety that is in your home; thank you for being excited about milestones with me. What a privilege to have parents like you!

Most importantly, thank You Lord Jesus, for providing all that was necessary for me to complete this PhD. You are the Giver of grace.

But You, O Lord, are a God merciful and gracious, slow to anger and abounding in steadfast love and faithfulness. – Psalms 86:15

TABLE OF CONTENTS

CHAPTER 1

Introduction

A host in a biodiversity hotspot

The southwestern region of South Africa encompasses the Greater Cape Floristic Region (GCFR), an area known for its winter rainfall and distinct plant composition (Born *et al.*, 2007). The GCFR is also subdivided into biomes. The largest subdivision, the Core Cape Subregion (CCR), previously referred to as the Cape Floristic Region (CFR), largely coincides with the border of the iconic Fynbos Biome (Bergh *et al.*, 2014). The CFR has also been described as the smallest of the world's six floral kingdoms (Cowling & Richardson, 1995; Manning & Goldblatt, 2012) and comprises the highest level of endemism and species richness within the GFCR (Born *et al.*, 2007). It is believed that the Mediterranean-type climate coupled with nutrient-poor soils and regular fires partly facilitates the high levels of biodiversity observed within this region (Cowling *et al.*, 1996).

The fynbos ("fine bush") vegetation in the Fynbos Biome is a predominantly shrubland vegetation type that is dominant within the CCR (Manning & Goldblatt, 2012). Of the three plant families that define fynbos (Ericaceae, Proteaceae and Restionaceae; Manning & Goldblatt, 2012), Proteaceae has its centre of diversity within the Cape (Linder, 2003; Valente *et al.*, 2010). Members of this family, specifically of the genus *Protea* L., may also occur beyond the borders of the CCR (Rebelo, 2001), including tropical Africa (Beard, 1962) and Australia (Cowling & Lamont, 1998). Several *Protea* species maintain their seeds in special storage structures on the plant until the water supply is interrupted (Rebelo, 2001), typically through fire or severe drought. These seed cones, known as infructescences, are formed from the involucral bracts that close around the inflorescences, forming a seed-storage organ. The enclosed structure is moist and protected from external factors such as wind and rain, proving an ideal niche for microorganisms, arthropods and nematodes (Coetzee & Giliomee, 1985; Human, 2013; Lee *et al.*, 2003; 2005; Roets *et al.*, 2005; 2006b; Theron *et al.*, 2012).

Ophiostomatoid fungi associated with *Protea*

Of the numerous fungi that have been found associated with Proteaceae species (Marincowitz *et al.*, 2008), the dominant colonizers of *Protea* infructescences are ophiostomatoid fungi (Lee *et al.*, 2005; Marais & Wingfield, 1994). Extensive surveys of infructescences of different *Protea* species have identified three *Knoxdaviesia* M.J. Wingf., P.S. van Wyk & Marasas and nine *Sporothrix* Hektoen & C.F. Perkins species as the dominant ophiostomatoid inhabitants, collectively occurring in at least 13 *Protea* hosts across southern Africa (De Beer *et al.*, 2016; Marais & Wingfield, 1994; 2001; 2006a; Roets *et al.*, 2005; 2011a; 2010; Wingfield & Van Wyk, 1993). *Protea*-associated ophiostomatoid fungi reside in two different orders: *Knoxdaviesia* species are in the Microascales, while *Sporothrix* is in the Ophiostomatales (De Beer *et al.*, 2013). Additionally, the *Sporothrix* symbionts of *Protea* are not monophyletic, suggesting that the association of ophiostomatoid fungi with serotinous *Protea* species has multiple origins (Roets *et al.*, 2006a; Wingfield *et al.*, 1999).

Ecological research

Since their discovery, several studies have investigated the ecology of *Protea*-associated ophiostomatoid fungi, elucidating their unique niche and dispersal mechanisms. Surveys of numerous *Protea* species suggest that ophiostomatoid fungi are confined to serotinous hosts (Roets *et al.*, 2005; 2009b). The enclosed infructescences provide a favourable micro-habitat (Zwölfer, 1979) in which the fungi flourish, especially during moist conditions (Roets *et al.*, 2005). In some *Protea* species with tightly closed infructescences, e.g. *P. repens* L., ophiostomatoid fungi may persist for several years (Roets *et al.*, 2005), likely because of effective moisture retention. Ascomata have not been observed on living tissue or seeds of *Protea* species. Coupled with the lack of disease symptoms, it appears that *Protea*-associated ophiostomatoid fungi have a saprotrophic rather than a pathogenic relationship with their *Protea* hosts (Marais & Wingfield, 1994; Roets *et al.*, 2005).

The infructescence niche has a limited life span, since older infructescences dry out due to mechanical damage by insects and animals and may disappear completely in the case of fire (Coetzee & Giliomee, 1987). Continual fungal dispersal to new flower heads is, therefore, necessary for survival (Aylward *et al.*, 2014; Roets *et al.*, 2009a). The polyphyletic assemblage of ophiostomatoid fungi have a convergent morphological adaptation to insect dispersal and form sexual structures that carry spore droplets on long ascomatal necks

(Spatafora & Blackwell, 1994). As such, it was expected that the ophiostomatoid associates of *Protea* would also be dispersed by arthropods. Although beetles are important secondary vectors (Roets *et al.*, 2009a), mites are the primary vectors of *Sporothrix* and *Knoxdaviesia* spores in *Protea* infructescences (Roets *et al.*, 2007; 2011b). Mites move between infructescences on the same *Protea* plant and are phoretic on beetles to accomplish secondary dispersal between plants (Roets *et al.*, 2009a). *Trichouropoda* Berlese and *Tarsonemus* Canestrini and Fonzago mite species have specialised spore-carrying structures and *Trichouropoda* mites can survive on a diet of their *Sporothrix* symbiont alone (Roets *et al.*, 2007; 2011b). Overall these observations suggest a mutualistic relationship between the fungi and their primary mite vectors and show an overlap in the dispersal mechanisms of the different ophiostomatoid genera in *Protea* infructescences.

The most recent ecological studies have considered the reason for the difference in *Protea* host specificity among the ophiostomatoid fungi. *Knoxdaviesia capensis* M.J. Wingf. & P.S. van Wyk is known from nine different *Protea* hosts (Aylward *et al.*, 2015b; Marais & Wingfield, 1994; Roets *et al.*, 2005; 2011a; Wingfield & Van Wyk, 1993) and *O. splendens* G.J. Marais & M.J. Wingf., *O. phasma* Roets, Z.W. de Beer & M.J. Wingf. and *O. africanum* G.J. Marais & M.J. Wingf. are each known from at least two hosts, but only *P. neriifolia* R. Br. is shared between the first two *Sporothrix* species (Marais & Wingfield, 1994; 2001; Roets *et al.*, 2006a). The remaining ophiostomatoid species have only been isolated from a single host despite extensive surveys (2006a; 2008; Roets *et al.*, 2005; 2010; Wingfield *et al.*, 1988). *In vitro* growth studies showed that *Protea*-associated ophiostomatoid fungi grow optimally on media supplemented with their natural hosts (Roets *et al.*, 2011a). Although factors such as temperature and relative humidity influence the occurrence of these fungi, the host exclusivity of each species is linked to the chemistry of its host rather than the associated abiotic factors (Roets *et al.*, 2011a).

Population research

Although the vectors of ophiostomatoid fungi had been identified, the frequency and distance at which these species disperse remained unknown, prompting investigation to continue at a population genetics level. Population genetic studies with microsatellite markers have been completed for two species in the *Knoxdaviesia* lineage of *Protea*-associated ophiostomatoid fungi. This genus was selected first, since it has proved cumbersome in ecological studies due to slow-growth and intolerance to cycloheximide (Wingfield & Van Wyk, 1993; Wingfield *et*

al., 1988). Populations of both *K. proteae* M.J. Wingf., P.S. van Wyk & Marasas and *K. capensis* show high genetic and almost maximal genotypic diversity (Aylward *et al.*, 2014; 2015a; in prep.). Since ascomata are the most prominent feature of ophiostomatoid fungi in infructescences (Lee *et al.*, 2005; Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988), this diversity suggests that the response of *Knoxdaviesia* species to the *Protea* infructescence environment is recombination, and consequently, increased genetic variation.

Protea-associated *Knoxdaviesia* species also disperse frequently and may cover long distances. Within a stand of its only host, *P. repens*, *K. proteae* maintains a panmictic population structure (Aylward *et al.*, 2014). Similarly, *K. capensis* populations in close proximity are not differentiated, even when these populations originate from different host species (Aylward *et al*., 2017). Dispersal is, therefore, limited by geographic distance, rather than by host identity. Once distances between *Knoxdaviesia* populations increase, genetic differentiation remains minimal and gene flow can still be detected (Aylward *et al.*, 2015a; 2017). The long-distance dispersal pattern shown by these fungi emphasises the role of mite phoresy in dispersing these fungi. Other than beetles, birds may be additional vehicles for dispersing ophiostomatoid-spore-carrying mites (Aylward *et al.*, 2015a; N. Theron, pers. comm.) and are the subject of current research. Since the beetle and the possible bird vectors also pollinate *Protea* species (Coetzee & Giliomee, 1985; Littlejohn, 2001), fungal dispersal may also reflect gene flow between *Protea* plants and the flower constancy of the pollinators.

Aims of this dissertation

The purpose of this dissertation is to use genomics to expand current knowledge on the *Knoxdaviesia* lineage of *Protea*-associated species. Population genetic studies have indicated that recombination is common in both *K. capensis* and *K. proteae*, but it remains unknown whether recombination is due to obligate or facultative outcrossing. We are further interested in the substances that *Protea*-associated ophiostomatoid species use as energy sources and the factors underlying the host range disparity between *K. capensis* and *K. proteae*.

I aim to address the following broad questions:

1. What is the genetic basis for the observed outcrossing sexual reproduction in *K. capensis* and *K. proteae*?
2. What carbon sources in *Protea* infructescences do *Knoxdaviesia* species exploit?
3. How are *Protea*-associated *Knoxdaviesia* species adapted to contend with competitors in *Protea* infructescences?
4. What factors may influence the difference in host range between *K. capensis* and *K. proteae*?

Outline of this dissertation

In **Chapter 2**, an overview of the current status of fungal genome sequencing is provided, with specific reference to plant pathogenic fungi with available genome sequences. This chapter will appear in the June 2017 volume of *IMA Fungus*. The sequencing, assembly and annotation of the *K. capensis* and *K. proteae* genomes are reported in **Chapter 3** as published in *Standards in Genomic Sciences* (Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D. & Wingfield, M. J. (2016). *SIGS* 11, 1-7). **Chapter 4** uses these genomes to address the first aim and was recently published in *Fungal Genetics and Biology* (Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, M. J. & Wingfield, B. D. (2016). *Fungal Genet. Biol.* 96, 47-57). The carbon sources used by the saprotrophic *Knoxdaviesia* species are explored in **Chapter 5** and compared to the carbon use of a pathogen that also inhabits *Protea* plants. Finally, the genome divergence between the *Knoxdaviesia* species and the potential competitive abilities of each are investigated in **Chapter 6**. Both chapters 5 and 6 consider the difference in host range between *K. capensis* and *K. proteae*.

REFERENCES

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2014). Panmixia defines the genetic diversity of a unique arthropod-dispersed fungus specific to *Protea* flowers. *Ecology and Evolution* **4**, 3444-3455.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015a). Long-distance dispersal and recolonization of a fire-destroyed niche by a mite-associated fungus. *Fungal Biology* **119**, 245-256.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015b). *Knoxdaviesia proteae* is not the only *Knoxdaviesia*-symbiont of *Protea repens*. *IMA Fungus* **6**, 471-476.

Aylward, J., Dreyer, L. L., Laas, T., Smit, L., & Roets, F. (2017). *Knoxdaviesia capensis*: dispersal ecology and population genetics of a flower-associated fungus. *Fungal Ecology* **26**, 28-36.

Beard, J. S. (1962). The genus *Protea* in tropical Africa. *Kirkia* **3**, 138-206.

Bergh, N. G., Verboom, G., Rouget, M. & Cowling, R. M. (2014). Vegetation types of the Greater Cape Floristic Region. In *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region*, pp. 26-46. Edited by N. Allsopp, J. F. Colville & G. A. Verboom. New York: Oxford University Press

Born, J., Linder, H. & Desmet, P. (2007). The Greater Cape Floristic Region. *Journal of Biogeography* **34**, 147-162.

Coetzee, J. & Giliomee, J. (1987). Seed predation and survival in the infructescences of *Protea repens* (Proteacae). *South African Journal of Botany* **53**, 61-64.

Coetzee, J. H. & Giliomee, J. H. (1985). Insects in association with the inflorescence of *Protea repens* (L.) (Proteaceae) and their role in pollination. *Journal of the Entomological Society of South Africa* **48**, 303-314.

Cowling, R. M. & Richardson, D. M. (1995). *Fynbos: South Africa's unique floral kingdom*. Vlaeberg: Fernwood Press.

Cowling, R. M., Rundel, P. W., Lamont, B. B., Arroyo, M. K. & Arianoutsou, M. (1996). Plant diversity in Mediterranean-climate regions. *Trends in Ecology & Evolution* **11**, 362-366.

Cowling, R. M. & Lamont, B. B. (1998). On the Nature of Gondwanan Species Flocks: Diversity of Proteaceae in Mediterranean South-western Australia and South Africa. *Australian Journal of Botany* **46**, 335-355.

De Beer, Z. W., Seifert, K. A. & Wingfield, M. J. (2013). The ophiostomatoid fungi: their dual position in the *Sordariomycetes*. In *Ophiostomatoid fungi: expanding frontiers*, pp. 1-20. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield: CBS Biodiversity Series.

De Beer, Z. W., Duong, T. A. & Wingfield, M. J. (2016). The divorce of *Sporothrix* and *Ophiostoma*: solution to a problematic relationship. *Studies in Mycology* **83**, 165-191.

Human, Z. R. (2013). The diversity and ecology of actinomycetes associated with environments dominated by ophiostomatoid fungi. MSc Thesis, University of Pretoria, Pretoria, South Africa.

Lee, S., Groenewald, J. Z., Taylor, J. E., Roets, F. & Crous, P. W. (2003). Rhynchostomatoid fungi occurring on Proteaceae. *Mycologia* **95**, 902-910.

Lee, S., Roets, F. & Crous, P. W. (2005). Biodiversity of saprobic microfungi associated with the infructescences of *Protea* species in South Africa. *Fungal Diversity* **19**, 69-78.

Linder, H. P. (2003). The radiation of the Cape flora, southern Africa. *Biological Reviews* **78**, 597-638.

Littlejohn, G. (2001). The challenges of breeding wild flower cultivars for use in commercial floriculture: African Proteaceae. In *XX International Eucarpia Symposium, Section Ornamentals, Strategies for New Ornamentals-Part I 552*, pp. 25-38.

Manning, J. & Goldblatt, P. (2012). Plants of the Greater Cape Floristic Region. 1: The Core Cape Flora. In *Strelitzia 29*. Pretoria: South African National Biodiversity Institute.

Marais, G. J. & Wingfield, M. J. (1994). Fungi associated with infructescences of *Protea* species in South Africa, including a new species of *Ophiostoma*. *Mycological Research* **98**, 369-374.

Marais, G. J. & Wingfield, M. J. (2001). *Ophiostoma africanum* sp. *nov*., and a key to ophiostomatoid species from *Protea* infructescences. *Mycological Research* **105**, 240-246.

Marincowitz, S., Crous, P. W., Groenewald, J. Z. & Wingfield, M. J. (2008). *Microfungi occurring on Proteaceae in the fynbos*. Utrecht: CBS Fungal Biodiversity Centre.

Rebelo, T. (2001). *Proteas: A field guide to the Proteas of Southern Africa*. Vlaeberg, South Africa: Fernwood Press.

Roets, F., Dreyer, L. L. & Crous, P. W. (2005). Seasonal trends in colonisation of *Protea* infructescences by *Gondwanamyces* and *Ophiostoma* spp. *South African Journal of Botany* **71**, 307-311.

Roets, F., de Beer, Z. W., Dreyer, L. L., Zipfel, R., Crous, P. W. & Wingfield, M. J. (2006a). Multi-gene phylogeny for *Ophiostoma* spp. reveals two new species from *Protea* infructescences. *Studies in Mycology* **55**, 199-212.

Roets, F., Dreyer, L. L., Geertsema, H. & Crous, P. W. (2006b). Arthropod communities in Proteaceae infructescences: seasonal variation and the influence of infructescence phenology. *African Entomology* **14**, 257-265.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2007). Discovery of Fungus-Mite Mutualism in a Unique Niche. *Environmental Entomology* **36**, 1226-1237.

Roets, F., de Beer, Z. W., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2008). *Ophiostoma gemellus* and *Sporothrix variecibatus* from Mites Infesting *Protea* Infructescences in South Africa. *Mycologia* **100**, 496-510.

Roets, F., Crous, P. W., Wingfield, M. J. & Dreyer, L. L. (2009a). Mite-mediated hyperphoretic dispersal of *Ophiostoma* spp. from the Infructescences of South African *Protea* spp. *Environmental Entomology* **28**, 143-152.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2009b). Fungal radiation in the Cape Floristic Region: An analysis based on *Gondwanamyces* and *Ophiostoma*. *Molecular Phylogenetics and Evolution* **51**, 111-119.

Roets, F., Wingfield, B. D., de Beer, Z. W., Wingfield, M. & Dreyer, L. L. (2010). Two new *Ophiostoma* species from *Protea caffra* in Zambia. *Persoonia* **24**, 18-28.

Roets, F., Theron, N., Wingfield, M. J. & Dreyer, L. L. (2011a). Biotic and abiotic constraints that facilitate host exclusivity of *Gondwanamyces* and *Ophiostoma* on *Protea*. *Fungal Biology* **116**, 49-61.

Roets, F., Wingfield, M. J., Wingfield, B. D. & Dreyer, L. L. (2011b). Mites are the most common vectors of the fungus *Gondwanamyces proteae* in *Protea* infructescences. *Fungal Biology* **115**, 343-350.

Spatafora, J. W. & Blackwell, M. (1994). The polyphyletic origins of ophiostomatoid fungi. *Mycological Research* **98**, 1-9.

Theron, N., Roets, F., Dreyer, L. L., Esler, K. J. & Ueckermann, E. A. (2012). A new genus and eight new species of Tydeoidea (Acari: Trombidiformes) from *Protea* species in South Africa. *International Journal of Acarology* **38**, 257-273.

Valente, L. M., Reeves, G., Schnitzler, J., Mason, I. P., Fay, M. F., Rebelo, T. G., Chase, M. W. & Barraclough, T. G. (2010). Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* **64**, 745-760.

Wingfield, B. D., Viljoen, C. D. & Wingfield, M. J. (1999). Phylogenetic relationships of ophiostomatoid fungi associated with *Protea* infructescences in South Africa. *Mycological Research* **103**, 1616-1620.

Wingfield, M. J., Wyk, P. S. V. & Marasas, W. F. O. (1988). *Ceratocystiopsis proteae* sp. *nov*., with a new anamorph genus. *Mycologia* **80**, 23-30.

Wingfield, M. J. & Van Wyk, P. S. (1993). A new species of *Ophiostoma* from *Protea* infructescences in South Africa. *Mycological Research* **97**, 709-716.

Zwölfer, H. (1979). Strategies and counter strategies in insect population systems competing for space and food in flower heads and plant galls. *Fortschritte der Zoologie* **25**, 331-353.

CHAPTER 2

A plant pathology perspective of fungal genome sequencing

ABSTRACT

The majority of plant pathogens are fungi and many of these adversely affect food security. The aim of this mini-review is to provide an analysis of the plant pathogenic fungi for which genome sequences are publically available, to assess their general genome characteristics and to consider how genomics has impacted plant pathology. A list of sequenced fungal species was assembled, the taxonomy of all species verified and the potential reason for sequencing each of the species considered. The genomes of 1090 fungal species are currently (October 2016) in the public domain and this number is rapidly rising. Pathogenic species comprised the largest category (35.5 %) and amongst these plant pathogens are predominant. Of the 191 plant pathogenic fungal species with available genomes, 61.3 % cause diseases on food crops, more than half of which are staple crops. The genomes of plant pathogens are slightly larger than those of other fungal species sequenced and they contain fewer coding sequences in relation to their genome size. Both of these factors can be attributed to the expansion of repeat elements. Sequenced genomes of plant pathogens provide blueprints from which potential virulence factors were identified and from which genes associated with different pathogenic strategies could be predicted. These have also made it possible to evaluate adaptability of pathogen genomes and genomic regions that experience selection pressures. Some genomic patterns, however, remain poorly understood and plant pathogen genomes alone are not sufficient to unravel complex pathogen-host interactions. Genomes, therefore, cannot replace experimental studies that can be complex and tedious. Ultimately, the most promising opportunity lies in using fungal plant pathogen genomics to inform disease management and risk assessment strategies. This will ultimately minimize the risks of future disease outbreaks and assist in preparation for emerging pathogen outbreaks.

INTRODUCTION

Sequencing of fungal genomes is being driven by various groups of scientists having different interests and needs from the resultant genomic data. Mycologists desire genome data to understand how fungi live and evolve, while industries require improved metabolic pathways or how to find new sources of natural products. The medical and plant pathology sectors need this information to understand diseases; to improve diagnoses, understand how they function and ultimately prevent or at least manage disease outbreaks (Kelman, 1985). By 2007, the genomes of 42 eukaryotes were available (Cornell *et al.*, 2007) and by 2008 the number of fungal genomes exceeded 90 (Park *et al.*, 2008). Today, more than 3 000 fungi are targeted in completed or ongoing genome projects and the genomes of more than 900 fungal species have been released. The substantial and growing investment in determining genome sequences clearly reflects the positive impact that this field is having on research. Our question here is what the impact has been for plant pathogenic fungi.

This mini-review aims to summarise the number of available fungal plant pathogen genomes, determine their general characteristics, and consider the impact that the availability of these genomes is having on the study of plant pathology. In order to determine which fungal plant pathogens have been sequenced, we surveyed fungal species (including Microsporidia, but excluding Oomycota) listed in 11 online genome repositories (Table 2.1), including MycoCosm (Grigoriev *et al.*, 2012, 2013), NCBI Genome (www.ncbi.nlm.nih.gov/genome), the Broad Institute (www.broadinstitute.org) and the universal cataloguing database, Genomes OnLine Database (GOLD; Reddy *et al.*, 2014). Fungal species that were found in more than one database were clustered and the current classification of all species was verified up to ordinal level using MycoBank (Robert *et al.*, 2013) and Index Fungorum (www.IndexFungorum.org). Most recent scientific literature was consulted where the two online reference data bases were not in agreement. Synonymous names associated with each species were noted by consulting MycoBank. We used this non-redundant list to accurately determine the number of fungal species with available genome sequences and, specifically, the extent to which fungal plant pathogens have been sequenced.

Table 2.1 Fungal genome resources used to populate the genome species list.

Name	Abbreviation	Number of fungal genome projects	URL
Aspergillus Genome Database	AspGD	4	www.aspgd.org
Candida Genome Database	CGD	4	www.candidagenome.org
EnsemblFungi		53	fungi.ensembl.org
Fungal Genome Resource		4	gene.genetics.uga.edu
Genomes OnLine Database	GOLD	3362	genomesonline.org
JGI Genome Portal: MycoCosm		>1200	genome.jgi.doe.gov/fungi
NCBI Genome		>1000	www.ncbi.nlm.nih.gov/genome
PomBase		1	www.pombase.org
Saccharomyces Genome Database	SGD	>50	www.yeastgenome.org
The Broad Institute		>100	www.broadinstitute.org/science/projects/fungal-genome-initiative
The Institute of Bioinformatics and Systems Biology	IBIS	20	www.helmholtz-muenchen.de/en/ibis/institute/groups/fungal-microbial-genomics/resources/index.html
University of Kentucky		29	www.endophyte.uky.edu/

BEYOND THE 1 000 MARK

The lower cost of genome sequencing, due to high-throughput technologies, has encouraged large scale genome initiatives. These include the 5,000 Insect Genome Project (i5k; Robinson *et al.*, 2011), Genome 10K (Genome 10K Community of Scientists, 2009) and 1000 Plants (www.onekp.com). These projects aim to sample species diversity by sequencing, respectively, whole genomes of insects and vertebrates and the transcriptomes of plant species. Similarly, fungal genome sequencing programmes such as the Fungal Genome Initiative (Fungal Research Community, 2002; The Fungal Genome Initiative Steering Commitee, 2003), the Fungal Genomics Program (Grigoriev *et al.*, 2011; Martin *et al.*, 2011) and its extension, the 1 000 Fungal Genomes (1KFG) Project (Spatafora, 2011), have contributed significantly to the number of fungal genomes currently available and continue to do so.

The prevalence of fungi as study organisms is clear when considering the on-going and completed genome sequencing projects. A catalogue of genome projects, GOLD (Reddy *et al.*, 2014), began in 1997 with six genome entries (Bernal *et al.*, 2001) and currently (14 October 2016) includes 7 422 eukaryal whole genome sequencing projects, of which 3 515 (47.4 %) are fungal. The genome databases (Table 2.1) list numerous completed and on-going fungal projects, although many entries do not represent different species. Of the 1 459 completed fungal genome projects in GOLD, slightly more than half (*ca.* 775) are unique species, whilst the remainder comprise additional strains of already-sequenced species. To illustrate the extent and prevalence of sequenced genomes in the fungal kingdom, we mapped species with publically available genomes onto ordinal consensus trees (Fig. 2.1A-C).

The most recent of the fungal genome sequencing initiatives, the five-year international collaborative 1KFG Project, aims to sequence and annotate two species from each of the more than 500 known fungal families (Spatafora, 2011). In three years, there has been a shift from obtaining representative genomes for all the fungal phyla (Buckley, 2008) to targeting genome sequencing at the family level. By October 2016, the genomes of 1 090 different fungal species were publically available (Supplementary file 2.1). Of this number, the 1KFG Project has released approximately 60 % of the total fungal species genomes available.

Although the target of 1 000 sequenced fungal species has been reached, the goal of two genomes from each family is a bigger task than the number 1 000 (Fig. 2.1). The goal of having the genome sequences for two representatives have only been achieved in 85 families

in the Ascomycota, 66 in the Basidiomycota and 11 in the remainder of the fungi. Not surprisingly, some economically and medically important families (e.g. Aspergillaceae, Clavicipitaceae, Mucoraceae, Mycosphaerellaceae, Saccharomycetaceae, Tremellaceae, Ustilaginaceae) have many more than two representatives. Additionally, taxonomic revision and species descriptions continue to generate new fungal families and orders. In the almost ten years since the publication of the Hibbett *et al.* (2007) consensus tree, more than 50 fungal orders have been described, somewhat increasing the workload of the 1KFG Project. Additionally, less than 10 % of the conservative estimate of 1.5 million total fungal species are known (Hawksworth, 2012) and new species descriptions continuously emerge. Therefore, the combined goals of sampling fungal biodiversity and sequencing the genomes of representative species are a continuous process.

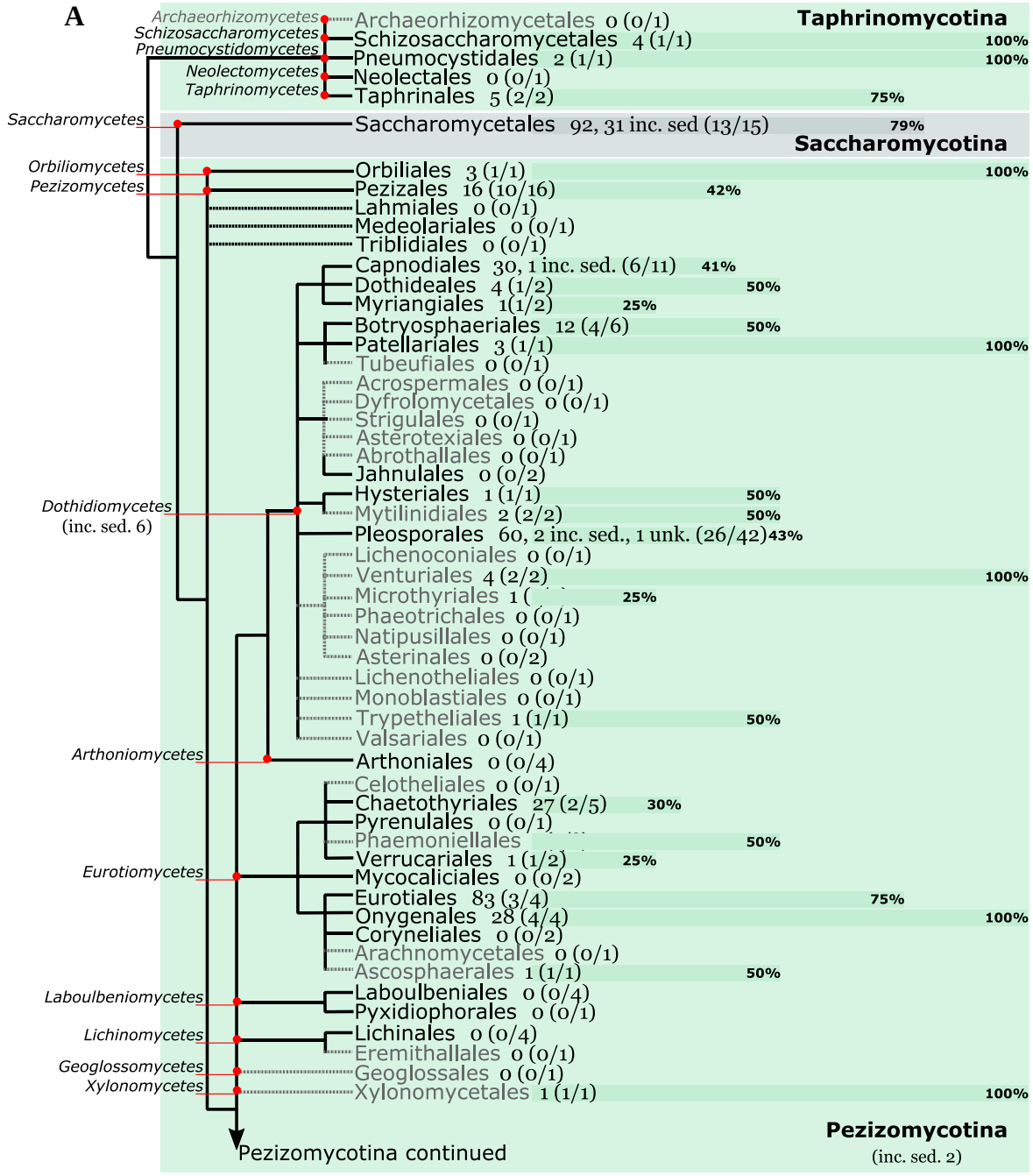

Figure 2.1A (legend on page 17)

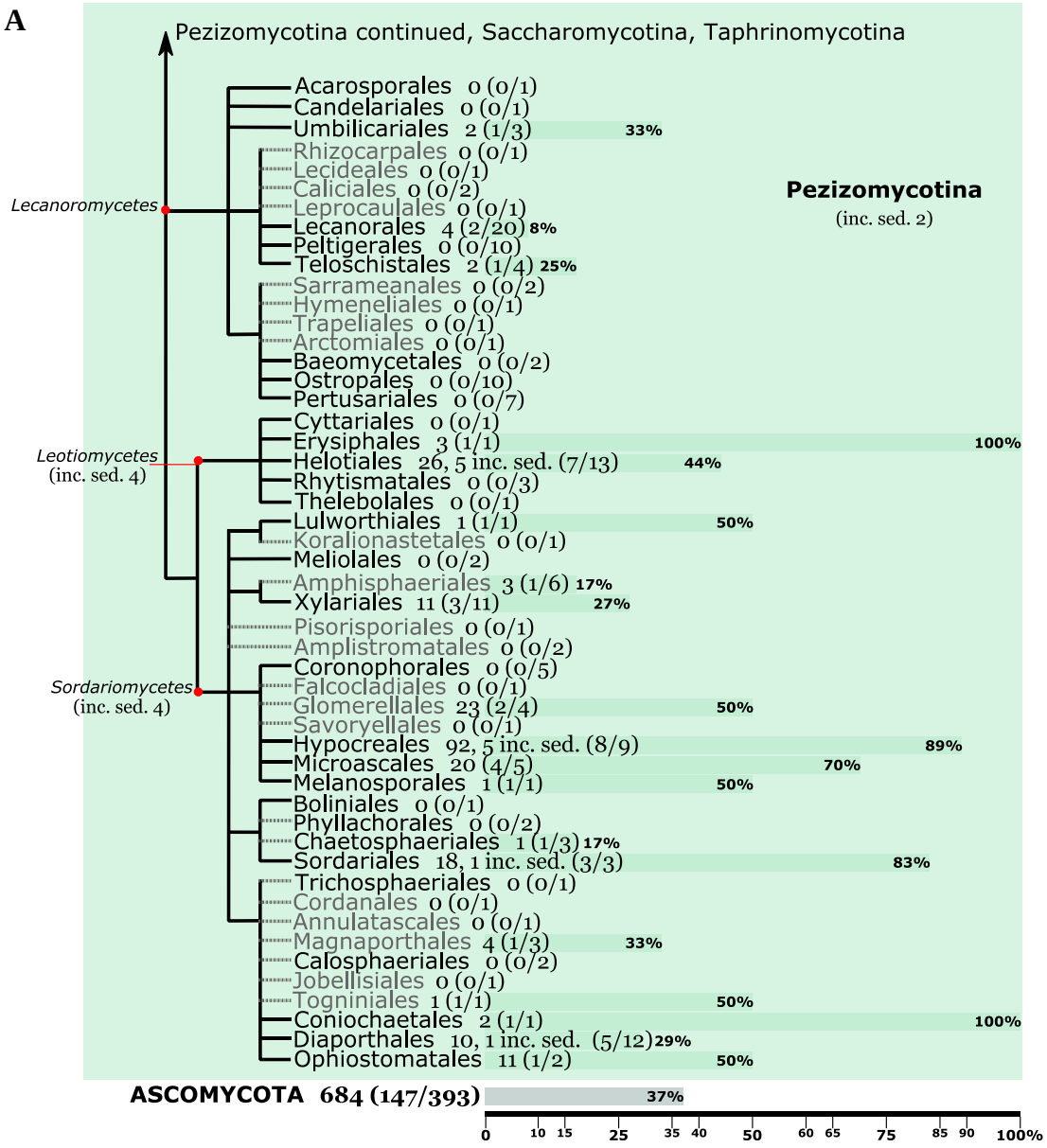

Figure 2.1A (continued; legend on page 17)

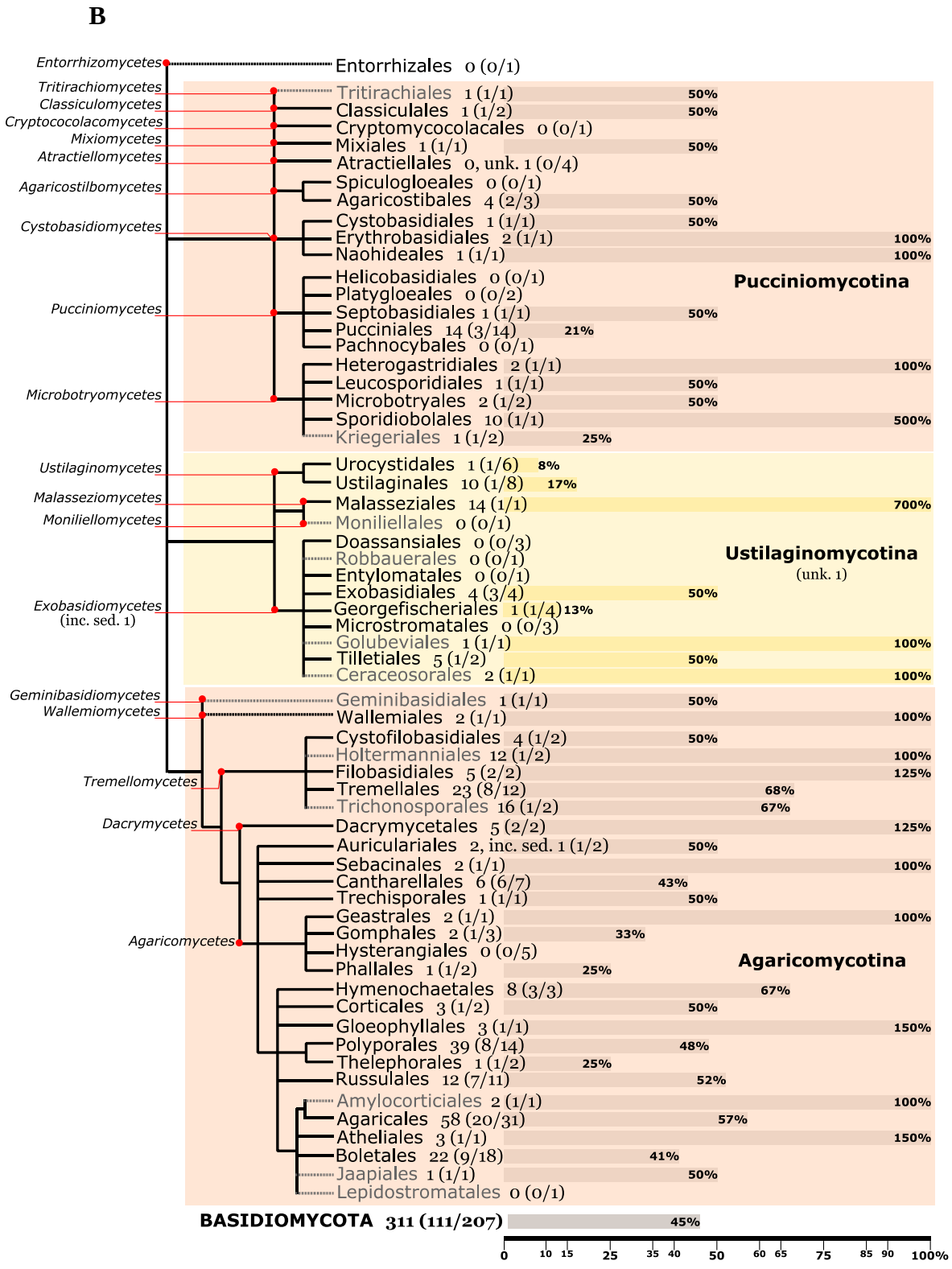

Figure 2.1B (legend on page 17)

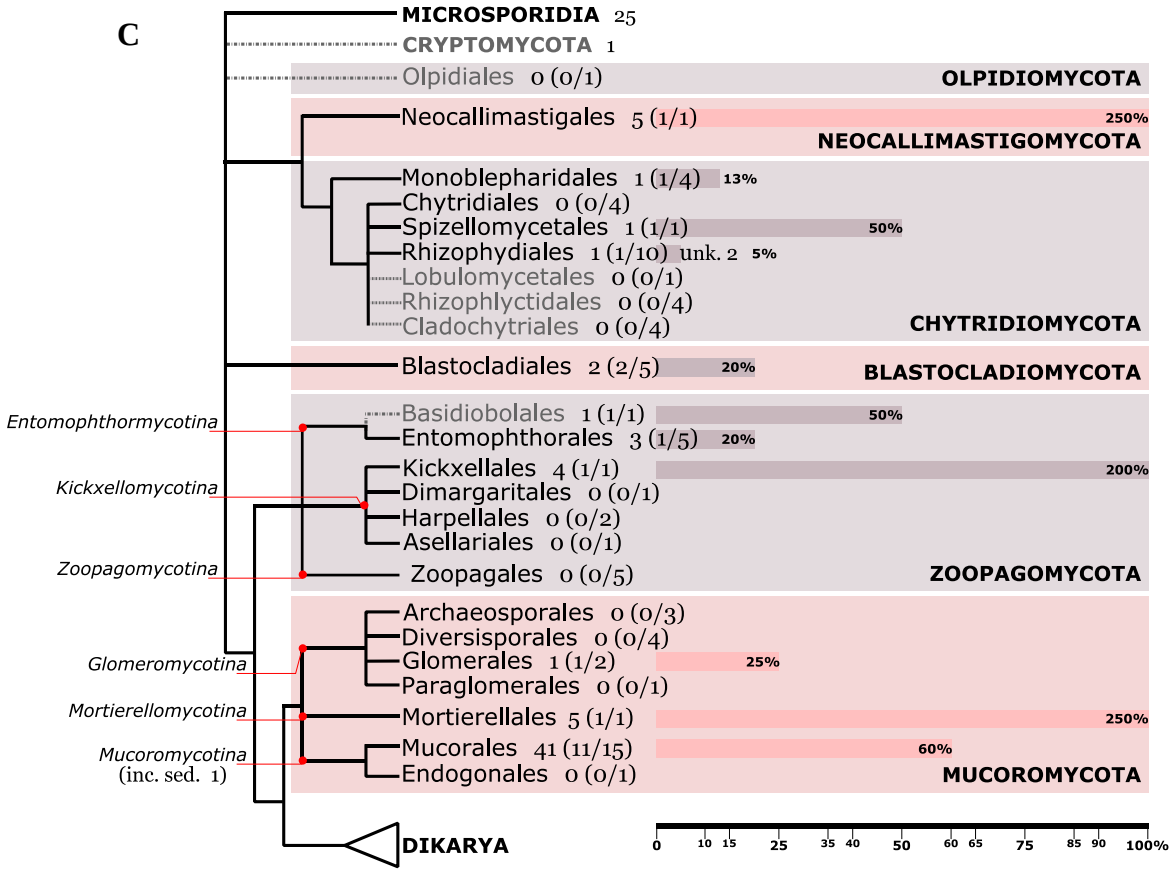

Figure 2.1 Ordinal consensus trees depicting the taxonomic (subphylum, class and order) distribution of publically available genomes for the Ascomycota (A), Basidiomycota (B) and early-diverging fungi (C). The number of sequenced genomes from each order is indicated after the order name. Where sequenced species are not classified into a family or have not been described, these are indicated as *incertae sedis* (inc. sed.) or unknown (unk.), respectively. The number of families with sequenced representatives out of the total number of described families is indicated in brackets. For each order, horizontal bars show the current progress of sequencing two genomes per family, indicated according to the scale bar below the figure. The Dikarya consensus trees are according to Hibbett *et al.* (2007), while the classification of Spatafora *et al.* (2016) was included in the tree of early-diverging fungi. Orders described after Hibbett *et al.* (2007) have been added in grey (see supplementary file 3 for references). The figures do not include unclassified fungi that have not been sequenced.

ARE PATHOGENS PREFERENTIALLY SEQUENCED?

More than 90 % of known fungal species reside in the subkingdom Dikarya (Kirk *et al.*, 2008) comprised of the two largest phyla, Ascomycota and Basidiomycota. The large number of ascomycete and basidiomycete species for which genome sequences have been determined (Table 2.2) is, therefore, not an over-emphasis of these common phyla, but rather reflects the size and diversity of the Dikarya (Fig. 2.2). In the majority of cases, the proportion of sequenced species in the phyla of early-diverging fungi is congruent with the known species, suggesting that genome projects have not neglected these phyla (Fig. 2.2). The Mucoromycota has a larger proportion of sequenced species than known species due to the sequencing of several Mucoraceae species that cause human mucoromycosis. One phylum (Olpidiomycota) and one subphylum (Zoopagomycotina) of early-diverging fungi, however, do not have any sequenced representatives and do not have "targeted" or "in progress" projects listed on GOLD. Olpidiomycota was only described recently (Doweld, 2013) and its members appear to be poorly known. The lack of Zoopagomycotina sequences can likely be ascribed to very few available pure cultures of this predominantly parasitic group of fungi (Spatafora *et al.*, 2016).

Within the subphyla of the Ascomycota and Basidiomycota, the proportion of sequenced species also largely corresponds to the number of known species (Fig. 2.3), with the major exception being the Saccharomycotina (budding yeasts; Fig. 2.3A). The emphasis placed on this subphylum is even more pronounced when considering not only the number of species, but also the number of strains that have been sequenced. Although the Pezizomycotina (filamentous fungi) is by far the most species-rich subphylum on the genome list (547 spp.), most sequenced strains are in the Saccharomycotina (416), maintaining its previous status as the most sequenced subphylum in the kingdom (Cuomo & Birren, 2010). Other than members of the Saccharomycotina, seven highly sequenced (≥ 10 strains) species are listed in GOLD (Table 2.3). One is of industrial importance (*Rhodotorula toruloides* (Banno) Q.M. Wang, F.Y. Bai, M. Groenewald & T. Boekhout), whereas the remainder influence food security (*Aspergillus flavus* Link, *Fusarium oxysporum* Schltdl. and *Magnaporthe oryzae* B.C. Couch) or human health (*Cryptococcus gattii* (Vanbreus. & Takashio) Kwon-Chung & Boekhout, *Coccidioides posadasii* M.C. Fisher and *Trichophyton rubrum* (Castell.) Sabour.).

Table 2.2 Number of fungal species from each phylum and subphylum with at least one available genome.

Phylum and subphylum	Sequenced fungal species	Known fungal species[1]
ASCOMYCOTA	**684**	**> 64 000**
Pezizomycotina	*547*	
Saccharomycotina	*123*	
Taphrinomycotina	*11*	
Incertae sedis	*3*	
BASIDIOMYCOTA	**311**	**> 31 000**
Agaricomycotina	*227*	
Pucciniomycotina	*43*	
Ustilaginomycotina	*41*	
BLASTOCLADIOMYCOTA	**2**	**> 175**
CHYTRIDIOMYCOTA	**5**	**> 700**
CRYPTOMYCOTA	**1**	**?**
MICROSPORIDIA	**25**	**> 1 300**
MUCOROMYCOTA	**47**	
Glomeromycotina	*1*	> 165
Mortierellomycotina	*5*	
Mucoromycotina	*41*	> 325
NEOCALLIMASTIGOMYCOTA	**5**	**> 20**
ZOOPAGOMYCOTA	**8**	
Entomophthoromycotina	*4*	> 275
Kickxellomycotina	*4*	> 260
UNKNOWN	**2**	
Total sequenced	***1090***	

[1] According to Kirk *et al.* (2008)

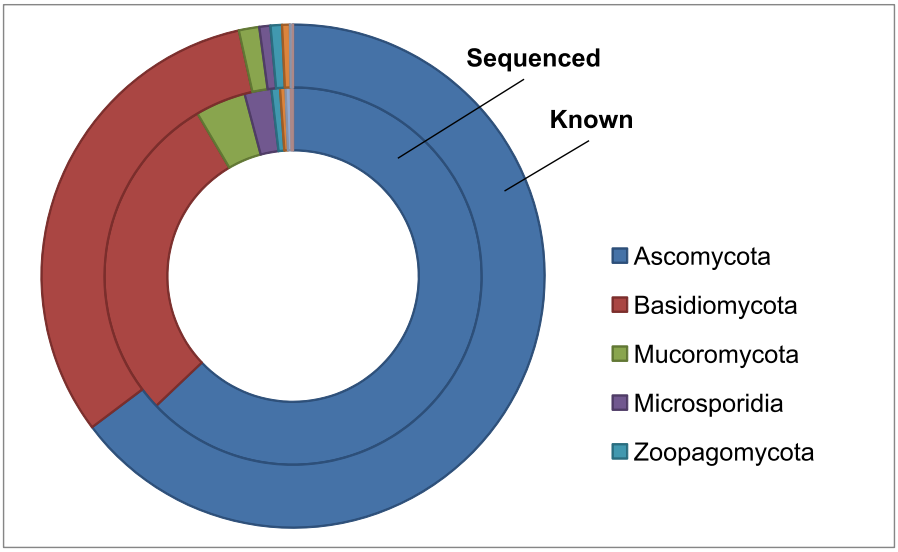

Figure 2.2 Comparison between the proportion of known and sequenced fungal species in the major fungal taxonomic divisions. The number of known species were obtained from Kirk *et al.* (2008).

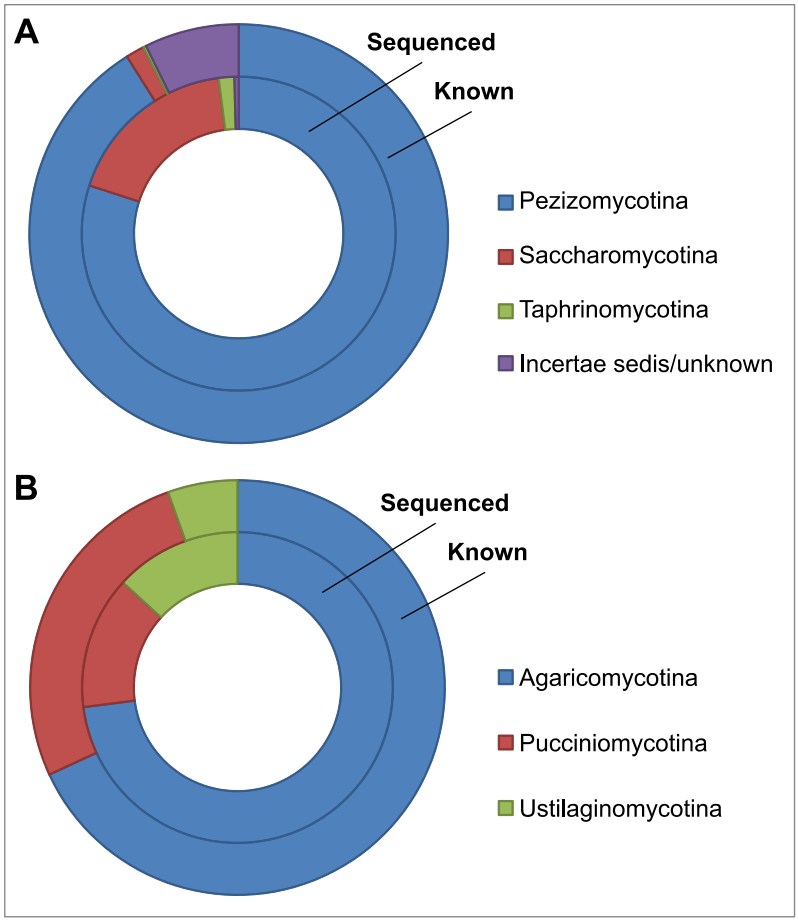

Figure 2.3 Comparison between the proportion of known and sequenced fungal species in the subphyla of the Ascomycota (A) and Basidiomycota (B). The number of known species were obtained from Kirk *et al.* (2008).

Table 2.3 Fungal species on the Genomes OnLine Database (Bernal *et al.*, 2001) with 10 or more completed whole-genome sequencing projects.

Species	Strains	Phylum	Subphylum
Saccharomyces cerevisiae	166	Ascomycota	Saccharomycotina
Magnaporthe oryzae	48	Ascomycota	Pezizomycotina
Candida albicans	35	Ascomycota	Saccharomycotina
Komagataella pastoris	32	Ascomycota	Saccharomycotina
Saccharomyces kudriavzevii	20	Ascomycota	Saccharomycotina
Cryptococcus gattii	18	Basidiomycota	Agaricomycotina
Fusarium oxysporum	17	Ascomycota	Pezizomycotina
Trichophyton rubrum	12	Ascomycota	Pezizomycotina
Aspergillus flavus	10	Ascomycota	Pezizomycotina
Coccidioides posadasii	10	Ascomycota	Pezizomycotina
Rhodotorula toruloides	10	Basidiomycota	Pucciniomycotina
Saccharomyces pastorianus	10	Ascomycota	Saccharomycotina

A 2008 report by the American Academy of Microbiology (Buckley, 2008) stated that fungal genome sequencing is “heavily” skewed in the favour of human pathogens. At that time whole-genome sequencing of eukaryotes, especially fungi, was in its infancy and the statement was based on “100-150 fungal representatives”. The initial high cost of genome sequencing would have favoured fungi of medical importance, but as the number of sequenced fungi grew and the cost decreased, this pattern was bound to change. We assessed whether pathogens are highly sequenced by consulting recent scientific literature (where available) on each fungal species on the genome list and categorising them according to their significance and reason for being sequenced. The largest (41.4 %) category consisted of pathogenic fungi and fungi of medical importance (Fig. 2.4), of which plant pathogens were the most prevalent group (49.4 %). Currently, 191 plant pathogenic species have publically available genomes (Supplementary file 2.2) and all belong to the Dikarya. Of these, 117 are pathogens of at least one food crop and 43 affect gymnosperms, the majority of which are commercially important (Table 2.4). The 117 food crop species include pathogens of cereals, fruit, vegetables and legumes. At least 60 of these species are responsible for diseases on 10 of the 15 global staple food crops (FAO, 1995).

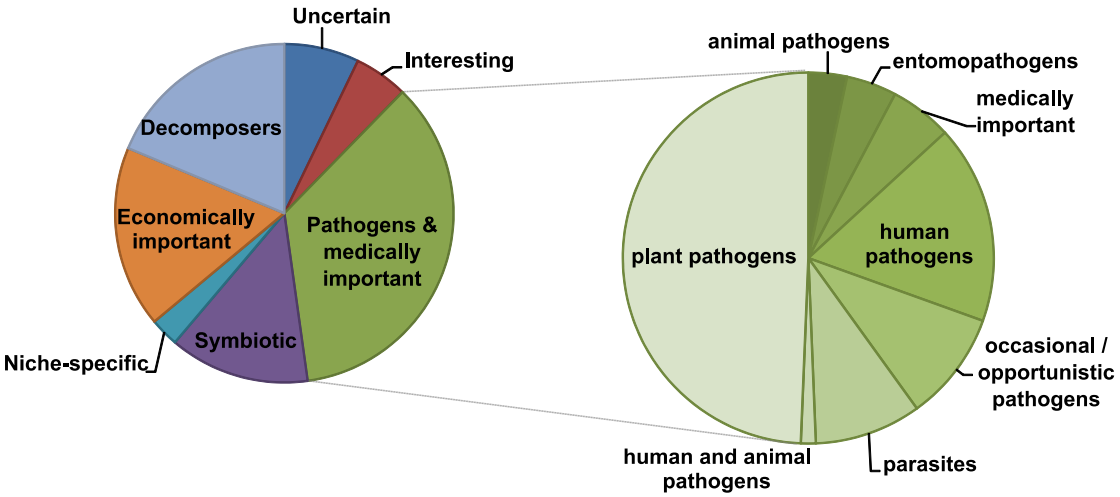

Figure 2.4 Categories of significance identified in the 1090 sequenced fungal species. Pathogens comprise the largest category within which plant pathogens are predominant. "Medically important" species represent fungi that are not directly pathogenic, but cause food or environmental contamination. "Interesting" species are studied for their development or metabolism. "Niche-specific" refers to species occupying abiotic niches, whereas "symbiotic" species are associated with other organisms. "Economically important" species have a use in the economy, *e.g.* culinary, biocontrol or pharmaceutical industries. Most of the parasites belong to the Microsporidia.

Table 2.4 Categories of plants affected by the 191 sequenced fungal plant pathogens.

Plant pathogen categories	Genomes available	%
Cash Crop Pathogens	8	4.2
Food Crop Pathogens	117	61.3
Grains	*48*	*25.2*
Fruit	*37*	*19.4*
Vegetables	*10*	*5.2*
Legumes	*11*	*5.8*
Multiple crop types	*11*	*5.8*
Gymnosperm Pathogens	43	22.5
Other[1]	23	12.0
TOTAL	*171*	*100*

[1] Non-gymnosperms not cultivated for food.

Clearly, genome sequencing projects have placed an emphasis on plant pathogenic species, specifically those affecting food security or commercial forestry. In general, fungal pathogens are highly represented in the genome list. Furthermore, as the number of available genomes has increased, plant pathogens have replaced human pathogens as the predominantly sequenced category of species. Since most plant pathogens are fungal (Carris *et al.*, 2012), the emphasis placed on fungal genome sequencing may (at least partly) be attributed to food security. For example, *M. oryzae*, a plant pathogen for which numerous strains have been sequenced (Table 2.3), is predicted to increase its distribution range and impact due to increased temperature and carbon dioxide levels (Gautam *et al.*, 2013). The sheer number of plant species and their associated disease-causing fungi makes this change in the focus of genome sequencing understandable. Sequencing a large number of plant pathogens that affect a range of plant species is, after all, less of a bias than sequencing many pathogenic species associated with a single species (humans).

GENOME SIZE AND GENE NUMBERS IN PLANT PATHOGENIC FUNGAL GENOMES

As far as we are aware, this review includes the first comprehensive list of plant pathogenic fungal genomes that have been sequenced to date. We, therefore, briefly present an overview of the genome characteristics of these species in comparison to other sequenced fungal species. We specifically looked at genome size and the numbers of genes encoded, because previous studies have revealed a link between plant pathogenicity and genome size and gene content (Duplessis *et al.*, 2011; Ohm *et al.*, 2012; Spanu *et al.*, 2010).

The 1 090 fungal species with publically available genome sequences have genome sizes ranging between two and 336 million base pairs (Mbp; Fig. 2.5A). The majority of these genomes fall within the 30-40 Mbp range (average = 37.2, median = 33.6), consistent with the size distribution of the 1 940 entries in the Fungal Genome Size Database (Kullman *et al.*, 2005). The genome sizes of sequenced plant pathogens are only slightly, but significantly, larger compared to this “norm”. This difference was most apparent in the pathogenic Ascomycetes for which Mann-Whitney U tests indicated the highest level of significance (Median = 38.0; U = 48131, $P < 0.01$). The average genome size of plant pathogenic Basidiomycetes (57.3 Mb) was much larger than that of the plant pathogenic ascomycetes

(39.4 Mb) and the remainder of the fungal genomes (34.8 Mb), owing to several pathogenic pucciniomycete (rust) species with genomes larger than 100 Mb.

Somewhat larger genome sizes in plant pathogens are congruent with the hypothesis that these often contain more repeated elements than other species (discussed below) (Castanera *et al.*, 2016; Ma *et al.*, 2010). Sequenced plant pathogens also have larger genomes than human, animal and opportunistic fungal pathogens (Fig. 2.5B). Although sequencing has thus far sampled the genome size distribution of the majority of the fungal kingdom, species with excessively large genome sizes have been omitted. This is not necessarily only due to the higher cost of sequencing large genomes, but likely also the complexity of obtaining sufficient biomaterial from a single obligately parasitic individual cultured on a live host (Barnes & Szabo, 2008). Since the majority of species in the Pucciniomycotina reside in the Pucciniales order of obligate plant pathogens (Kirk *et al.*, 2008), the latter may also explain why the proportion of sequenced species in this group is slightly less than the known species (Fig. 3B). Therefore, genome-sequencing efforts so far, most likely underestimate the maximum size of plant pathogen genomes.

Considering the number of predicted open reading frames (ORFs), 714 of the available genomes have publically accessible gene annotations. The sequenced fungi have, on average, 11 256 ± 3 873 total predicted ORFs at a density of 351.8 ± 104.0 ORFs per Mb (Fig. 2.5C). In comparison to the other genomes, plant pathogenic fungi do not differ in the number of predicted ORFs, but they do have significantly fewer ORFs when accounting for genome size ($U = 35647$; $P < 0.01$). This trend was also observed in the animal pathogenic fungi (including entomopathogens; $U = 8847$; $P < 0.01$). Previous whole-genome studies suggest that the number of coding genes does not necessarily increase with genome size, since transposable elements and repetitive sequences proliferate in large genomes (Kidwell, 2002). The lower number of ORFs/Mb in the genome of plant pathogenic ascomycetes is, therefore, consistent with their larger genome size being due to repetitive elements. Additionally, some pathogens have lost genes redundant in their lifecycles (Spanu *et al.*, 2010), which may also decrease their ORFs/Mb. This trend could, however, not be detected in the genomes of human and opportunistic pathogens (Fig. 2.5D).

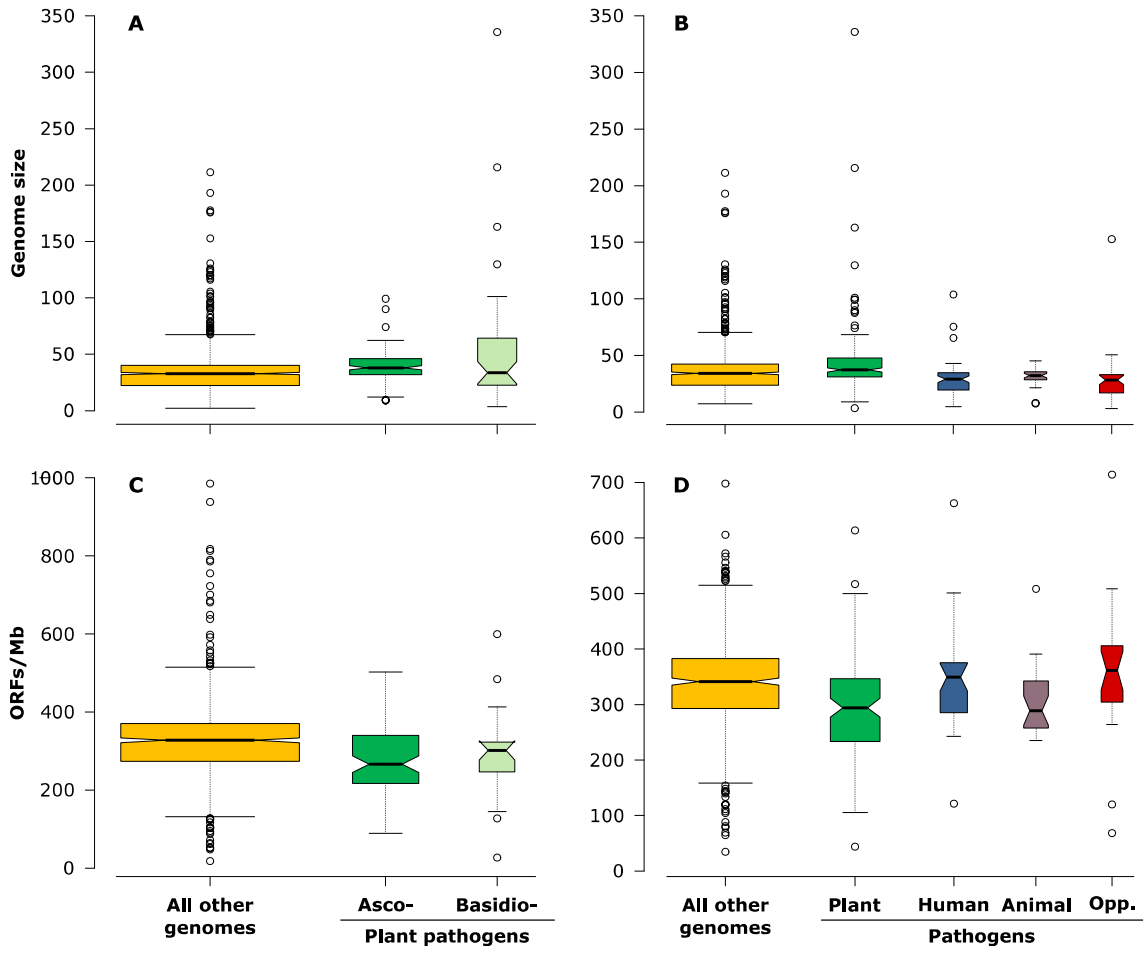

Figure 2.5 Genome size (A and B) and number of open reading frames (ORFs) per million base pairs (Mb; C and D) in the plant pathogen genomes compared to the remainder of the genome list and other pathogens. Boxplots were drawn in BoxplotR (Spitzer *et al.*, 2014) using the Tukey whisker extent. Width of the boxes is proportional to the square root of the sample size; notches show the 95 % confidence interval of the median. Opp. = opportunistic pathogens. In B and D, animal pathogens include entomopathogenic fungi and Microsporidia are excluded from "other" genomes, since they have small genomes with many ORFs/Mb.

IMPACT OF GENOMES ON PLANT PATHOLOGY

Ever since the advent of plant pathology, researchers have been interested in the biology of plant pathogens and how this can be translated into means for disease control. In this regard, genome data are not used in isolation, but provide context for observational and experimental data, thereby accelerating the pace of traditional research methods. For emerging pathogens, a disease may be known, but the mechanisms relating to infection biology and virulence not necessarily understood. In these cases, a genome can provide the first glimpse of the potential effectors and toxins that are present (*e.g*. Ellwood *et al.*, 2010). In some plant pathogens, genomes have resulted in a shift of conventional paradigms. Here a classic example is the discovery of entire horizontally transferrable chromosomes related to pathogenicity in *Fusarium oxysporum* f. sp. *lycopersici* (Sacc.) W.C. Snyder & H.N. Hansen (Ma *et al.*, 2010). The primary impacts of genome sequences on plant pathology have been a better understanding of the pathogenicity, lifestyle and genome evolution of pathogens. Furthermore, genomes are also resources from which genetic tools can be used to mine information.

Pathogenicity and lifestyle

Secreted and cell surface proteins mediate the interaction between pathogen and host and are often first to be characterised from plant pathogen genomes. Genome sequences enable *in silico* predictions of secreted virulence proteins (effectors), bypassing traditional enzyme assays or chromatography/spectrometry techniques that were ineffective at detecting less abundant effectors. For the corn smut fungus, *Ustilago maydis* (DC.) Corda, previous experimental studies were not able to identify the virulence factors eventually highlighted by interrogating the genome sequence (Kamper *et al.*, 2006). The proteins identified *in silico* could then be used to experimentally determine the function of specific effectors in the infection process (Liu *et al.*, 2015). This genome-based effector identification and subsequent phenotype determination has contributed significantly to the online Pathogen Host Interactions Database (Urban *et al.*, 2015). Similarly, genomic prediction of secondary metabolite biosynthesis genes have been well established (Keller *et al.*, 2005) and gene deletion systems could subsequently be used to determine the phenotypes that they confer (Lee *et al.*, 2005). Gene inventories of plant pathogens have also revealed proteins not previously known to be involved in pathogenicity: for example the high diversity of membrane transporters in the *F. oxysporum* and *Pyrenochaeta lycopersici* R.W. Schneid. &

Gerlach genomes strongly implicates them in the pathogenicity of these fungi (Aragona *et al.*, 2014).

Intuitively, cell wall degrading enzymes can be expected to be important in plant pathogenesis and the presence of these enzymes in plant pathogens was well established before the genomic era (Jones *et al.*, 1972; Schneider & Collmer, 2010). Genome sequences have confirmed that most plant pathogens encode an array of cell wall degrading enzymes, specifically pectinolytic enzymes in dicot pathogens (Klosterman *et al.*, 2011; Olson *et al.*, 2012). The diversity of cell wall degrading enzymes in a genome appears to increase with host range, as exemplified by the massive number of carbohydrate degrading enzymes in *Macrophomina phaseolina* (Tassi) Goid. that infects over 500 plant species (Islam *et al.*, 2012). Exceptions to this perceived norm typically occur in specialised modes of pathogenesis. For example, cell wall degrading enzymes are absent from the genome of the anther smut fungus *Microbotryum lychnidis-dioicae* (DC.) G. Deml & Oberw. (Perlin *et al.*, 2015). Rather than attacking plant cells, this pathogen has an array of enzymes to influence host development, enabling fungal spores to be substituted for pollen. In contrast, gene inventories suggest that necrotrophic pathogens induce apoptosis in host cells rather than breaking down their cell walls (McDonald *et al.*, 2015). Metabolism-related enzymes in fungal genomes, therefore, have great potential to predict infection strategies and lifestyle.

Beyond the analysis of single genome sequences, comparing the genomes of ecologically different strains and species has substantial value. For example, analysis of the *Rhizoctonia solani* J.G. Kühn AG2-2IIIB genome would have revealed only an abundance of cell wall degrading enzymes. However, comparisons with less aggressive *R. solani* strains revealed that its virulence can be linked to a significant expansion of polysaccharide lyase enzymes (Wibberg *et al.*, 2016). Similarly, comparisons of resistant and non-resistant *Penicillium digitatum* strains has enabled identification of mutations conferring tolerance to antifungal compounds (Marcet-Houben *et al.*, 2012).

Comparisons between the genomes of 18 dothideomycete species has suggested that the number of effectors encoded by these fungi are linked to their pathogenic lifestyle (Ohm *et al.*, 2012). The greatest number of effectors was identified from necrotrophic pathogens, whereas hemibiotrophs have apparently reduced their effector arsenal to evade plant defences before they switch to necrotrophy (Ohm *et al.*, 2012). Furthermore, multiple genome comparisons have been used to highlight specific genes under diversifying selection, revealing that evolutionary pressure on plant pathogen effector proteins drive their adaptation

(Schirawski *et al.*, 2010; Stukenbrock *et al.*, 2011). Comparison between plant and fungal genomes have also become essential tools to tease apart pathogen and plant RNA sequences when analysing *in planta* transcript data (McDonald *et al.*, 2015). The value of multiple genome comparisons has prompted projects such as the Fungal Genome Initiative and the 1KFP to focus not on sequencing single species, but groups of species useful in a comparative context (Grigoriev *et al.*, 2013; The Fungal Genome Initiative Steering Committee, 2004).

The evolution of different fungal lifestyles is a fascinating topic considered by many comparative genomics studies. Some plant pathogenic fungi with different lifestyles have surprisingly similar gene contents (De Wit *et al.* 2012), yet unique genes mediate their host interactions. The large proportion of unique secreted effector proteins and host-specific hydrolytic enzymes in plant pathogenic fungi implies that host association drives their adaptation and, therefore, evolution (De Wit *et al.* 2012, Duplessis *et al.* 2011, O'Connell *et al.* 2012, Spanu *et al.* 2010). The effect of host association is further emphasised by the diversification of both effectors and hydrolytic enzymes in broad host range pathogens such as *Colletotrichum higginsianum* Sacc. (O'Connell *et al.* 2012). In contrast, host association cannot explain the loss of primary metabolism genes that led to obligate biotrophy in powdery mildew fungi (Spanu *et al.* 2010). Similarly, selective pressures that mediate the evolution of a hemibiotrophic strategy, where the pathogen transitions between a biotrophic and necrotrophic lifestyle, are poorly understood.

Genome evolution of plant pathogens

The arms race between pathogen and host (Stahl & Bishop, 2000) makes pathogen adaptability, or evolutionary potential, particularly interesting (McDonald & Linde, 2002). Reproduction and gene diversity are two of the factors that influence evolutionary potential (McDonald & Linde, 2002) and these can be estimated from genome sequences. For example, an analysis of the mating type genes that govern sexual reproduction can provide insights into the mating strategy of a fungus. Heterothallic ascomycete fungi are identified by the occurrence of a single mating type in a genome (Kronstad & Staben, 1997), whereas homothallic fungi contain both mating types, either in the same genome or in a dikaryotic cellular state (Wilson *et al.*, 2015). Many fungi propagate only vegetatively or sexual reproduction is difficult to observe. In such cases, genomic analyses have been able to reveal the presence of mating type genes (*e.g.* Bihon *et al.*, 2014; Marcet-Houben *et al.*, 2012),

suggesting that these species could have a cryptic sexual cycle (Bihon *et al.*, 2014). The mating type sequence information can subsequently be used to determine the distribution of different mating types in a population (Aylward *et al.*, 2016, Chapter 4; Haasbroek *et al.*, 2014). In contrast, genomes can also reveal the importance of mitotic recombination for generating new allelic combinations. For example, a whole genome survey concluded that mating type genes are completely absent from the tomato pathogen *P. lycopersici* and that it has an expansion of gene modules associated with heterokaryon incompatibility (Aragona *et al.*, 2014).

Adaptability to a changing environment or a resistant host can also be mediated by genome plasticity. Transposons are repetitive elements in DNA known to contribute to genome plasticity and evolution (Wöstemeyer & Kreibich, 2002). The expansion of repeats in many plant pathogen genomes points to their role in diversification and adaptation (Raffaele & Kamoun, 2012; Spanu *et al.*, 2010; Thon *et al.*, 2006) and has been directly implicated in the pathogenicity of the wheat necrotroph *Pyrenophora tritici-repentis* (Died.) Drechsler (Manning *et al.*, 2013). Surveys of transposons across plant pathogen genomes have revealed differences in their number and activity between the essential core and dispensable supernumerary chromosomes (Ohm *et al.*, 2012; Vanheule *et al.*, 2016). In *F. poae* (Peck) Wollenw., repeat expansion in the core chromosomes is contained, while the non-essential supernumerary chromosomes have many active transposons that invade the core chromosomes (Vanheule *et al.*, 2016). The supernumerary chromosomes also provide opportunity for duplication and diversification of core genes, thereby facilitating adaptation. An example of such diversification and adaptation in the postharvest pathogen *Penicillium digitatum* (Pers.) Sacc. is the association of DNA transposons and ABC transporters in drug resistant strains (Sun et al., 2013).

Other than facilitating whole-genome related studies, genome sequences have become ideal resources from which to mine genetic tools. Previously, species-specific population genetic tools like microsatellites had to be developed painstakingly by cloning and genome walking (Barnes *et al.*, 2001; Burgess *et al.*, 2001). Now, any genome sequence enables rapid identification of such genetic markers (*e.g.* Haasbroek *et al.*, 2014). This holds true for diagnostic markers: genome regions that unambiguously and rapidly identify a pathogen and / or differentiate between pathogens can be designed by inspecting whole genome sequences. Although development of such markers in fungi is lagging behind viral and bacterial pathogens, some examples have recently become available. A pathotype specific marker has

been developed from the genome of *M. oryzae triticum* (Pieck *et al.*, 2016) and comparative genomics has detected diagnostic regions in two *Calonectria* De Not. species (Malapi-Wight *et al.*, 2016) and in *Pseudoperonospora cubensis* (Berk. & M.A. Curtis) Rostovzev (Withers *et al.*, 2016). Continued application of fungal genomes to generate identification tools is bound to increase the efficiency of quarantine procedures (McTaggart *et al.*, 2016).

CHALLENGES

Although the availability of fungal genomes has facilitated great strides in our knowledge and understanding of infection processes and genome evolution, there remains much to learn. For example, the regulatory elements in most genomes remain poorly annotated and require complex experimental methodologies for accurate identification (*e.g.* Shen *et al.*, 2012). In a recent review, Schatz (2015) commented that sequencing human genomes has been one of the greatest accomplishments of the past two decades but "one of the greatest pursuits for the next twenty years will be trying to understand what it all means". The same can be said for fungal genomes. The information that can be gleaned from a genome sequence is bound to increase as our understanding of these sequences grows.

Genome sequences should not be seen to provide "silver bullets", although they are often sold this way. They provide the blueprint of potential cellular activities, but are not sufficient to unravel the complexity of pathogen-host interactions. For example, in *Fusarium oxysporum*, the cell wall degrading enzymes secreted during infection of tomato displayed a clear succession (Jones *et al.*, 1972), an ecologically relevant process that could not be deduced from a gene inventory. In combination with transcriptome data, however, genomic data has revealed how pathogens tolerate host defences (DiGuistini *et al.*, 2011) and how hosts can resist pathogen infection (Zhu *et al.*, 2012). Experimental work, both *in vitro* and *in plantae*, will remain essential components in studying fungal plant pathogens.

The end goal of studying any host-pathogen relationship is clearly to inform disease management and control. Thus far, identifying specific molecular targets has had little impact on developing new antifungal inhibitors (Odds, 2005) and integrative management strategies must, therefore, be a priority. Ultimately, the elucidated effector proteins, host targets, and the overall insights gained into the biology of pathogens must inform disease management strategies (Maloy, 2005). It is also crucial that they inform risk assessment protocols governing biosecurity (McTaggart *et al.*, 2016). It is, therefore, essential that the ecological

significance of genome patterns is studied to ensure that this knowledge can be extrapolated to emerging pathogen threats.

As revealed by comparative genomics, deciphering plant pathogen evolution is in many cases dependent on comparisons with species having other lifestyles. A large scale example of this is the revised classification of species previously known as Zygomycetes (Spatafora *et al.*, 2016); an endeavour possible because of the availability of multiple genome sequences for this group. In this regard, filling in the gaps in the list of sequenced species is crucial to our understanding of relationships and pathogenesis. The challenge is, therefore, to continue sequencing apparently uninteresting or unimportant taxonomic groups along with the economically important in order to ultimately gain a holistic view.

CONCLUSIONS

The activities of independent research groups and several fungal sequencing initiatives (Fungal Research Community, 2002; Grigoriev *et al.*, 2011; Martin *et al.*, 2011; Spatafora, 2011; The Fungal Genome Initiative Steering Commitee, 2003), have resulted in the number of publically available fungal genomes growing exponentially since the first genome was sequenced in 1996 (Goffeau *et al.*). The taxonomic distribution of sequenced fungal genomes is currently roughly congruent with the number of species known from each phylum and subphylum. This is an important and impressive achievement in the goal of sampling biodiversity and representing the phylogenetic groups of the fungal kingdom (Fungal Research Community, 2002; The Fungal Genome Initiative Steering Commitee, 2003). Many of the genomes have been sequenced to sample environmental and ecological diversity. However, investment continues to be primarily focussed on projects that have direct anthropogenic importance. The emphasis on genomes of plant pathogenic fungi has specifically increased subsequent to the Buckley (2008) overview of sequenced fungal species.

The genomes of more than 1 000 fungal species are already publically available and this number is growing steadily. Fungal genomics has enabled rapid characterisation of plant pathogen genomes and revealed features that allow better understanding of the biology of these species. It has also made it possible to rapidly develop tools to study pathogen biology and genetics. In a field where delayed action has profound consequences for livelihoods and food security, genome sequences provide us with essential tools to prepare for the emergence

of new plant pathogens and future disease outbreaks. In this regard, the medical example provided by Bill Gates (2015) that the application of available technologies could significantly have reduced the impact of the recent Ebola epidemic also holds for plant pathology. Particularly in the era of genomics, we have significant tools to deal with the plant disease arms race and we must apply them more actively and aggressively.

ACKNOWLEDGEMENTS

We thank the National Research Foundation (NRF) and the NRF-Department of Science and Technology (DST) Centre of Excellence in Tree Health Biotechnology (CTHB) for financial support.

Supplementary files 2.1 - 2.3 are provided electronically.

REFERENCES

Aragona, M., Minio, A., Ferrarini, A., Valente, M. T., Bagnaresi, P., Orrù, L., Tononi, P., Zamperin, G., Infantino, A., Valè, G., *et al.* (2014). De novo genome assembly of the soil-borne fungus and tomato pathogen *Pyrenochaeta lycopersici*. *BMC Genomics* **15**, 1-12.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, M. J. & Wingfield, B. D. (2016). Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*. *Fungal Genetics and Biology* **96**, 47-57.

Barnes, C. & Szabo, L. (2008). A rapid method for detecting and quantifying bacterial DNA in rust fungal DNA samples. *Phytopathology* **98**, 115-119.

Barnes, I., Gaur, A., Burgess, T., Roux, J., Wingfield, B. D. & Wingfield, M. J. (2001). Microsatellite markers reflect intra-specific relationships between isolates of the vascular wilt pathogen *Ceratocystis fimbriata*. *Molecular Plant Pathology* **2**, 319-325.

Bernal, A., Ear, U. & Kyrpides, N. (2001). Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Research* **29**, 126-127.

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 Genetics and Biology* **62**, 55-61.

Buckley, M. (2008). *The Fungal Kingdom: diverse and essential roles in earth's ecosystem.* A report from the American Academy of Microbiology, Washington, DC. www.asm.org.

Burgess, T., Wingfield, M. J. & Wingfield, B. D. (2001). Simple Sequence Repeat Markers Distinguish among Morphotypes of *Sphaeropsis sapinea*. *Applied and Environmental Microbiology* **67**, 354-362.

Carris, L., Little, C. & Stiles, C. (2012). Introduction to Fungi. The Plant Health Instructor. doi: 10.1094/PHI-I-2012-0426-01.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., *et al.* (2016). Transposable Elements *versus* the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLOS Genetics* **12**, e1006108.

Cornell, M. J., Alam, I., Soanes, D. M., Wong, H. M., Hedeler, C., Paton, N. W., Rattray, M., Hubbard, S. J., Talbot, N. J. & Oliver, S. G. (2007). Comparative genome analysis across a kingdom of eukaryotic organisms: Specialization and diversification in the Fungi. *Genome Research* **17**, 1809-1822.

Cuomo, C. A. & Birren, B. W. (2010). Chapter 34 - The Fungal Genome Initiative and Lessons Learned from Genome Sequencing. In *Methods in Enzymology*, pp. 833-855. Edited by W. Jonathan, G. Christine & R. F. Gerald: Academic Press.

De Wit, P. J., Van Der Burgt, A, Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P. & Cox, M. P. (2012). The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. *PLOS Genetics* **8**, e1003088.

DiGuistini, S., Wang, Y., Liao, N. Y., Taylor, G., Tanguay, P., Feau, N., Henrissat, B., Chan, S. K., Hesse-Orce, U., Alamouti, S. M., *et al.* (2011). Genome and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*, a lodgepole pine pathogen. *Proceedings of the National Academy of Sciences* **108**, 2504-2509.

Doweld, A. B. (2013). Index Fungorum no. 42. *Index Fungorum* http://www.indexfungorum.org/Publications/Index%20Fungorum%20no.42.pdf.

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., *et al.* (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences* **108**, 9166-9171.

Ellwood, S. R., Liu, Z., Syme, R. A., Lai, Z., Hane, J. K., Keiper, F., Moffat, C. S., Oliver, R. P. & Friesen, T. L. (2010). A first genome assembly of the barley fungal pathogen *Pyrenophora teres* f. *teres*. *Genome Biology* **11**, 1-14.

FAO (1995). *Dimensions of need: an atlas of food and agriculture*. Rome, Italy: Food & Agriculture Organization of the United Nations. Available from www.fao.org.

Friesen, T. L., Stukenbrock, E. H., Liu, Z., Meinhardt, S., Ling, H., Faris, J. D., Rasmussen, J. B., Solomon, P. S., McDonald, B. A. & Oliver, R. P. (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics* **38**, 953-956.

Fungal Research Community (2002). Fungal Genome Initiative (white paper). *Whitehead Institute*, viewed 24 February 2014, http://www.broadinstitute.org/annotation/fungi/fgi/FGI_01_whitepaper_2002.pdf.

Gates, B. (2015). The Next Epidemic — Lessons from Ebola. *New England Journal of Medicine* **372**, 1381-1384.

Gautam, H., Bhardwaj, M. & Kumar, R. (2013). Climate change and its impact on plant diseases. *Current Science* **105**, 1685-1691.

Genome 10K Community of Scientists (2009). Genome 10K: A Proposal to Obtain Whole-Genome Sequence for 10 000 Vertebrate Species. *Journal of Heredity* **100**, 659-674.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 Genes. *Science* **274**, 546-567.

Grigoriev, I. V., Cullen, D., Goodwin, S. B., Hibbett, D., Jeffries, T. W., Kubicek, C. P., Kuske, C., Magnuson, J. K., Martin, F. & Spatafora, J. W. (2011). Fueling the future with fungal genomics. *Mycology* **2**, 192-209.

Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R. A., *et al.* (2012). The Genome Portal of the Department of Energy Joint Genome Institute. *Nucleic Acids Research* **40**, D26-D32.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., *et al.* (2013). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* 10.1093/nar/gkt1183.

Haasbroek, M. P., Craven, M., Barnes, I. & Crampton, B. G. (2014). Microsatellite and mating type primers for the maize and sorghum pathogen, *Exserohilum turcicum*. *Australasian Plant Pathology* **43**, 577-581.

Hawksworth, D. (2012). Global species numbers of fungi: are tropical studies and molecular approaches contributing to a more robust estimate? *Biodiversity and Conservation* **21**, 2425-2433.

Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., Huhndorf, S., James, T., Kirk, P. M., Lücking, R., *et al.* (2007). A higher-level phylogenetic classification of the Fungi. *Mycological Research* **111**, 509-547.

Islam, M. S., Haque, M. S., Islam, M. M., Emdad, E. M., Halim, A., Hossen, Q. M. M., Hossain, M. Z., Ahmed, B., Rahim, S., Rahman, M. S., *et al.* (2012). Tools to kill: Genome of one of the most destructive plant pathogenic fungi *Macrophomina phaseolina*. *BMC Genomics* **13**, 1-16.

Jones, T. M., Anderson, A. J. & Albersheim, P. (1972). Host-pathogen interactions IV. Studies on the polysaccharide-degrading enzymes secreted by *Fusarium oxysporum* f. sp. *lycopersici*. *Physiological Plant Pathology* **2**, 153-166.

Kamper, J., Kahmann, R., Bolker, M., Ma, L.-J., Brefort, T., Saville, B. J., Banuett, F., Kronstad, J. W., Gold, S. E., Muller, O., *et al.* (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97-101.

Keller, N. P., Turner, G. & Bennett, J. W. (2005). Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology* **3**, 937-947.

Kelman, A. (1985). Plant pathology at the crossroads. *Annual Review of Phytopathology* **23**, 1-12.

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49-63.

Kirk, P. M., Cannon, P. F., Minter, D. W. & Stalpers, J. A. (2008). Dictionary of the Fungi (10th edition). UK: CAB International.

Klosterman, S. J., Subbarao, K. V., Kang, S., Veronese, P., Gold, S. E., Thomma, B., Chen, Z., Henrissat, B., Lee, Y.-H. & Park, J. (2011). Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLOS Pathogens* **7**, e1002137.

Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. & Turgeon, B. G. (2003). Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proceedings of the National Academy of Sciences* **100**, 15670-15675.

Kronstad, J. W. & Staben, C. (1997). Mating type in filamentous fungi. *Annual Review of Genetics* **31**, 245-276.

Kullman, B., Tamm, H. & Kullman, K. (2005). Fungal Genome Size Database. http://www.zbi.ee/fungal-genomesize.

Lee, B.-N., Kroken, S., Chou, D. Y. T., Robbertse, B., Yoder, O. C. & Turgeon, B. G. (2005). Functional Analysis of All Nonribosomal Peptide Synthetases in *Cochliobolus heterostrophus* Reveals a Factor, NPS6, Involved in Virulence and Resistance to Oxidative Stress. *Eukaryotic Cell* **4**, 545-555.

Liu, J., Yuan, Y., Wu, Z., Li, N., Chen, Y., Qin, T., Geng, H., Xiong, L. & Liu, D. (2015). A Novel Sterol Regulatory Element-Binding Protein Gene *sreA* Identified in *Penicillium digitatum* Is Required for Prochloraz Resistance, Full Virulence and *erg11* (*cyp51*) Regulation *PLOS ONE* **10**, e0117115.

Ma, L.-J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M. & Henrissat, B. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367-373.

Malapi-Wight, M., Demers, J. E., Veltri, D., Marra, R. E. & Crouch, J. A. (2016). LAMP Detection Assays for Boxwood Blight Pathogens: A Comparative Genomics Approach. *Scientific Reports* **6**, 2640.

Maloy, O. C. (2005). Plant Disease Management. *The plant health instructor* **10**, DOI: 10.1094/PHI-I-2005-0202-1001

Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., Figueroa, M., Freitag, M., Hane, J. K., Henrissat, B., *et al.* (2013). Comparative Genomics of a Plant-Pathogenic Fungus, *Pyrenophora tritici-repentis*, Reveals Transduplication and the Impact of Repeat Elements on Pathogenicity and Population Divergence. *Genes|Genomes|Genetics* **3**, 41-63.

Marcet-Houben, M. & Gabaldón, T. (2010). Acquisition of prokaryotic genes by fungal genomes. *Trends in Genetics* **26**, 5-8.

Marcet-Houben, M., Ballester, A.-R., de la Fuente, B., Harries, E., Marcos, J. F., González-Candelas, L. & Gabaldón, T. (2012). Genome sequence of the necrotrophic fungus *Penicillium digitatum*, the main postharvest pathogen of citrus. *BMC Genomics* **13**, 1-18.

Martin, F., Cullen, D., Hibbett, D., Pisabarro, A., Spatafora, J. W., Baker, S. E. & Grigoriev, I. V. (2011). Sequencing the fungal tree of life. *New Phytologist* **190**, 818-821.

McDonald, B. A. & Linde, C. (2002). Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annual Review of Phytopathology* **40**, 349-379.

McDonald, M. C., McDonald, B. A. & Solomon, P. S. (2015). Recent advances in the *Zymoseptoria tritici*–wheat interaction: insights from pathogenomics. *Frontiers in Plant Science* **6**.

McTaggart, A. R., van der Nest, M. A., Steenkamp, E. T., Roux, J., Slippers, B., Shuey, L. S., Wingfield, M. J. & Drenth, A. (2016). Fungal Genomics Challenges the Dogma of Name-Based Biosecurity. *PLOS Pathogens* **12**, e1005475.

O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., Damm, U., Buiate, E.A., Epstein, L. & Alkan, N. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nature Genetics* **44**, 1060-1065.

Odds, F. C. (2005). Genomics, molecular targets and the discovery of antifungal drugs. *Revista Iberoamericana de Micología* **22**, 229-237.

Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., Condon, B. J., Copeland, A. C., Dhillon, B. & Glaser, F. (2012). Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. *PLOS Pathogens* **8**, e1003037.

Olson, Å., Aerts, A., Asiegbu, F., Belbahri, L., Bouzid, O., Broberg, A., Canbäck, B., Coutinho, P. M., Cullen, D., Dalman, K., *et al.* (2012). Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytologist* **194**, 1001-1013.

Park, J., Park, B., Jung, K., Jang, S., Yu, K., Choi, J., Kong, S., Park, J., Kim, S. & Kim, H. (2008). CFGP: a web-based, comparative fungal genomics platform. *Nucleic Acids Research* **36**, D562-D571.

Perlin, M. H., Amselem, J., Fontanillas, E., Toh, S. S., Chen, Z., Goldberg, J., Duplessis, S., Henrissat, B., Young, S., Zeng, Q., *et al.* (2015). Sex and parasites: genomic and transcriptomic analysis of *Microbotryum lychnidis-dioicae*, the biotrophic and plant-castrating anther smut fungus. *BMC Genomics* **16**, 1-24.

Pieck, M. L., Ruck, A., Farman, M., Peterson, G. L., Stack, J. P., Valent, B. & Pedley, K. F. (2016). Genomics-Based Marker Discovery and Diagnostic Assay Development for Wheat Blast. *Plant Disease* http://dx.doi.org/10.1094/PDIS-04-16-0500-RE.

Raffaele, S. & Kamoun, S. (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* **10**, 417-430.

Reddy, T., Thomas, A. D., Stamatis, D., Bertsch, J., Isbandi, M., Jansson, J., Mallajosyula, J., Pagani, I., Lobos, E. A. & Kyrpides, N. C. (2014). The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. *Nucleic Acids Research*, gku950.

Robert, V., Vu, D., Amor, A. B. H., van de Wiele, N., Brouwer, C., Jabas, B., Szoke, S., Dridi, A., Triki, M. & ben Daoud, S. (2013). MycoBank gearing up for new horizons. *IMA Fungus* **4**, 371.

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., Lawson, D., Okamuro, J., Robertson, H. M. & Schneider, D. J. (2011). Creating a Buzz About Insect Genomes. *Science* **331**, 1386.

Schatz, M. C. (2015). *The next 20 years of genome research. doi: 10.1101/020289.*

Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., Di Stasio, M., Rössel, N., Mendoza-Mendoza, A., Pester, D., *et al.* (2010). Pathogenicity Determinants in Smut Fungi Revealed by Genome Comparison. *Science* **330**, 1546-1548.

Schneider, D. J. & Collmer, A. (2010). Studying plant-pathogen interactions in the genomics era: beyond molecular Koch's postulates to systems biology. *Annual Review of Phytopathology* **48**, 457-479.

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., *et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120.

Slot, J. C. & Rokas, A. (2011). Horizontal transfer of a large and highly toxic secondary metabolic gene cluster between fungi. *Current Biology* **21**, 134-139.

Soanes D. & Richards, T. A. (2014). Horizontal gene transfer in eukaryotic plant pathogens. *Annual Review of Phytopathology* **52**, 583-614.

Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., Loren van Themaat, E. V., Brown, J. K. M., Butcher, S. A., Gurr, S. J., *et al.* (2010). Genome Expansion and Gene Loss in Powdery Mildew Fungi Reveal Tradeoffs in Extreme Parasitism. *Science* **330**, 1543-1546.

Spatafora, J. (2011). 1000 fungal genomes to be sequenced. *IMA Fungus* **2**, 41.

Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., *et al.* (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* **108**, 1028-1046.

Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. (2014). BoxPlotR: a web tool for generation of box plots. *Nature Methods* **11**, 121-122.

Stahl, E. A. & Bishop, J. G. (2000). Plant–pathogen arms races at the molecular level. *Current Opinion in Plant Biology* **3**, 299-304.

Stukenbrock, E. H., Bataillon, T., Dutheil, J. Y., Hansen, T. T., Li, R., Zala, M., McDonald, B. A., Wang, J. & Schierup, M. H. (2011). The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research* **21**, 2157-2166.

Sun, B-F., Xiao, J-H., He, S., Liu, L., Murphy, R. W. & Huang, D-W. (2013a). Multiple interkingdom horizontal gene transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. *PLOS ONE* **8**, e60029.

Sun, X., Ruan, R., Lin, L., Zhu, C., Zhang, T., Wang, M., Li, H. & Yu, D. (2013b). Genomewide investigation into DNA elements and ABC transporters involved in imazalil resistance in *Penicillium digitatum*. *Fems Microbiology Letters* **348**, 11-18.

The Fungal Genome Initiative Steering Commitee (2003). A White Paper for Fungal Comparative Genomics. *Whitehead Institute*, viewed 24 February 2014, http://www.broadinstitute.org/annotation/fungi/fgi/FGI_02_whitepaper_2003.pdf.

The Fungal Genome Initiative Steering Committee (2004). Fungal Genome Initiative: A White Paper for Fungal Genomics. *The Broad Institute of Harvard and MIT*. Cambridge, Massachusetts.

Thon, M. R., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T. K. & Dean, R. A. (2006). The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biology* **7**, 1-9.

Thynne, E., McDonald, M. C. & Solomon, P. S. (2015). Phytopathogen emergence in the genomics era. *Trends in Plant Science* **20**, 246-255.

Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H. & Hammond-Kosack, K. E. (2015). The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Research* **43**, D645-D655.

Vanheule, A., Audenaert, K., Warris, S., van de Geest, H., Schijlen, E., Höfte, M., De Saeger, S., Haesaert, G., Waalwijk, C. & van der Lee, T. (2016). Living apart together: crosstalk between the core and supernumerary genomes in a fungal plant pathogen. *BMC Genomics* **17**, 1-18.

Wibberg, D., Andersson, L., Tzelepis, G., Rupp, O., Blom, J., Jelonek, L., Pühler, A., Fogelqvist, J., Varrelmann, M., Schlüter, A., *et al.* (2016). Genome analysis of the sugar beet pathogen *Rhizoctonia solani* AG2-2IIIB revealed high numbers in secreted proteins and cell wall degrading enzymes. *BMC Genomics* **17**, 1-12.

Wilson, A. M., Wilken, P. M., van der Nest, M. A., Steenkamp, E. T., Wingfield, M. J. & Wingfield, B. D. (2015). Homothallism: an umbrella term for describing diverse sexual behaviours. *IMA Fungus* **6**, 207.

Withers, S., Gongora-Castillo, E., Gent, D., Thomas, A., Ojiambo, P. S. & Quesada-Ocampo, L. M. (2016). Using Next-Generation Sequencing to Develop Molecular Diagnostics for *Pseudoperonospora cubensis*, the Cucurbit Downy Mildew Pathogen. *Phytopathology* **106**, 1105-1116.

Wöstemeyer, J. & Kreibich, A. (2002). Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. *Current Genetics* **41**, 189-198.

Zhu, S., Cao, Y.-Z., Jiang, C., Tan, B.-Y., Wang, Z., Feng, S., Zhang, L., Su, X.-H., Brejova, B. & Vinar, T. (2012). Sequencing the genome of *Marssonina brunnea* reveals fungus-poplar co-evolution. *BMC Genomics* **13**, 382.

Declaration by the candidate (for published chapter)

With regards to Chapter 3 (Complete genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa), the nature and scope of my contribution were as follows:

Nature of contribution	**Extent of contribution (%)**
Isolation of strains; DNA and RNA extraction	100
Assembly and annotation of genomes	100
Manuscript preparation	80

The following co-authors have contributed to Chapter 3:

Name	**e-mail address**	**Nature of contribution**	**Extent of contribution (%)**
Emma T. Steenkamp	emma.steenkamp@up.ac.za	Provided guidance and edited the manuscript	2.5
Leanne L. Dreyer	ld@sun.ac.za	"	2.5
Francois Roets	fr@sun.ac.za	"	2.5
Brenda D. Wingfield	brenda.wingfield@up.ac.za	"	2.5
Michael J. Wingfield	michael.wingfield@up.ac.za	Provided guidance and funded the genomes	10

Signature of candidate:  ……

Declaration by co-authors:

The undersigned hereby confirm that

1. The declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 3.
2. No other authors contributed to Chapter 3 than those specified above, and
3. Potential conflict of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 3 of this dissertation

Signature	**Institutional affiliation**	**Date**
Leanne L. Dreyer	Stellenbosch University	6 Oct. 2016
Francois Roets	Stellenbosch University	6 Oct. 2016
Emma T. Steenkamp	University of Pretoria	12/10/2016
Michael J. Wingfield	University of Pretoria	6 Oct 2016
Brenda D. Wingfield	University of Pretoria	6 Oct 2016

CHAPTER 3

Genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa

ABSTRACT

Two closely related ophiostomatoid fungi, *Knoxdaviesia capensis* and *K. proteae*, inhabit the fruiting structures of certain *Protea* species indigenous to southern Africa. Although *K. capensis* occurs in several *Protea* hosts, *K. proteae* is confined to *P. repens*. In this study, the genomes of *K. capensis* CBS139037 and *K. proteae* CBS140089 are determined. The genome of *K. capensis* consists of 35,537,816 bp assembled into 29 scaffolds and 7,940 predicted protein-coding genes of which 6,192 (77.98%) could be functionally classified. *Knoxdaviesia proteae* has a similar genome size of 35,489,142 bp that is comprised of 133 scaffolds. A total of 8,173 protein-coding genes were predicted for *K. proteae* and 6,093 (74.55%) of these have functional annotations. The GC-content of both genomes is 52.8%.

INTRODUCTION

Two lineages of the polyphyletic assemblage known as ophiostomatoid fungi (Spatafora & Blackwell, 1994) are associated with the fruiting structures (infructescences) of serotinous *Protea* L. plants (Wingfield *et al.*, 1999). *Protea* species are a key component of the fynbos vegetation in the Core Cape Subregion (CCR) of South Africa (Manning & Goldblatt, 2012) and the genus is predominantly encountered in South Africa (Gibbs Russell, 1984; Rebelo, 2001). The *Protea*-associated ophiostomatoid fungi are, therefore, believed to be endemic to this region, similar to their hosts. This association of ophiostomatoid fungi with a keystone plant genus in a biodiversity hotspot is intriguing (Mittermeier *et al.*, 1998), as many ophiostomatoid fungi are notorious pathogens of trees (Brasier, 1991; Harrington & Wingfield, 1998; Paine *et al.*, 1997; Roux & Wingfield, 2009), yet the *Protea* ophiostomatoid species are not associated with disease symptoms (Roets *et al.*, 2013).

Ophiostomatoid fungi are characterized by the flask-shaped morphology of their sexual fruiting structures and their association with arthropods (Malloch & Blackwell, 1993; Spatafora & Blackwell, 1994). The *Protea*-associated members of this assemblage are primarily dispersed by mites that come into contact with fungal spores in the *Protea* infructescences (Roets *et al.*, 2007; 2011b). These mites have limited dispersal ability, but use beetles and possibly larger vertebrates (such as birds) as vehicles for long-distance dispersal (Aylward *et al.*, 2015; Roets *et al.*, 2009).

The three *Knoxdaviesia* M.J. Wingf., P.S. van Wyk & Marasas species associated with *Protea* have intriguing host ranges. *Knoxdaviesia capensis* M.J. Wingf. & P.S. van Wyk occurs in at least eight different *Protea* hosts, whereas *K. proteae* M.J. Wingf., P.S. van Wyk & Marasas and *K. wingfieldii* (Roets & Dreyer) Z.W. de Beer & M.J. Wingf. are confined to single host species, respectively *P. repens* L. and *P. caffra* Meisn. (Crous *et al.*, 2012; Roets *et al.*, 2005; 2011a; Wingfield & Van Wyk, 1993). An investigation of the population biology of *K. proteae*, revealed that this fungus has a high level of intra-specific genetic diversity and that it is extensively dispersed within the CCR of South Africa (Aylward *et al.*, 2014b; 2015). However, other than host range and dispersal mechanisms, little is known about the biology and ecology of *Knoxdaviesia* in general (Roets *et al.*, 2013). Here we present the description of the first drafts of the genome sequences of the two CCR species, *K. capensis* and *K. proteae*, as well as their respective annotations.

ORGANISM INFORMATION

Classification and features

The one lineage of *Protea*-associated ophiostomatoid fungi resides in the Ophiostomataceae (Ophiostomatales, Ascomycota), while the second resides in the Gondwanamycetaceae (Microascales, Ascomycota) (Réblová *et al.*, 2011; Roets *et al.*, 2013). The latter group includes three closely related *Protea*-associated species in the genus *Knoxdaviesia* (Fig. 3.1). This genus was initially described to accommodate the asexual state of the first species in the genus, *K. proteae* (Wingfield *et al.*, 1988). Under the dual nomenclature system of fungi, the sexual state of this fungus was described in the same paper as *Ceratocystiopsis proteae* M.J. Wingf., P.S. van Wyk & Marasas (Wingfield *et al.*, 1988). A new genus, *Gondwanamyces* G.J. Marais & M.J. Wingf., was later described to accommodate the sexual state of this species and that of another species, *Ophiostoma capense* M.J. Wingf. & P.S. van Wyk (Marais *et al.*, 1998). The asexual states of both remained to be treated as species of *Knoxdaviesia*. Since the abolishment of the dual nomenclature system of fungi, the oldest genus name takes preference, irrespective of morph (Hawksworth, 2011; McNeill *et al.*, 2012). The name *Knoxdaviesia*, therefore, has priority and all species previously treated in *Gondwanamyces* were transferred to *Knoxdaviesia* (De Beer *et al.*, 2013).

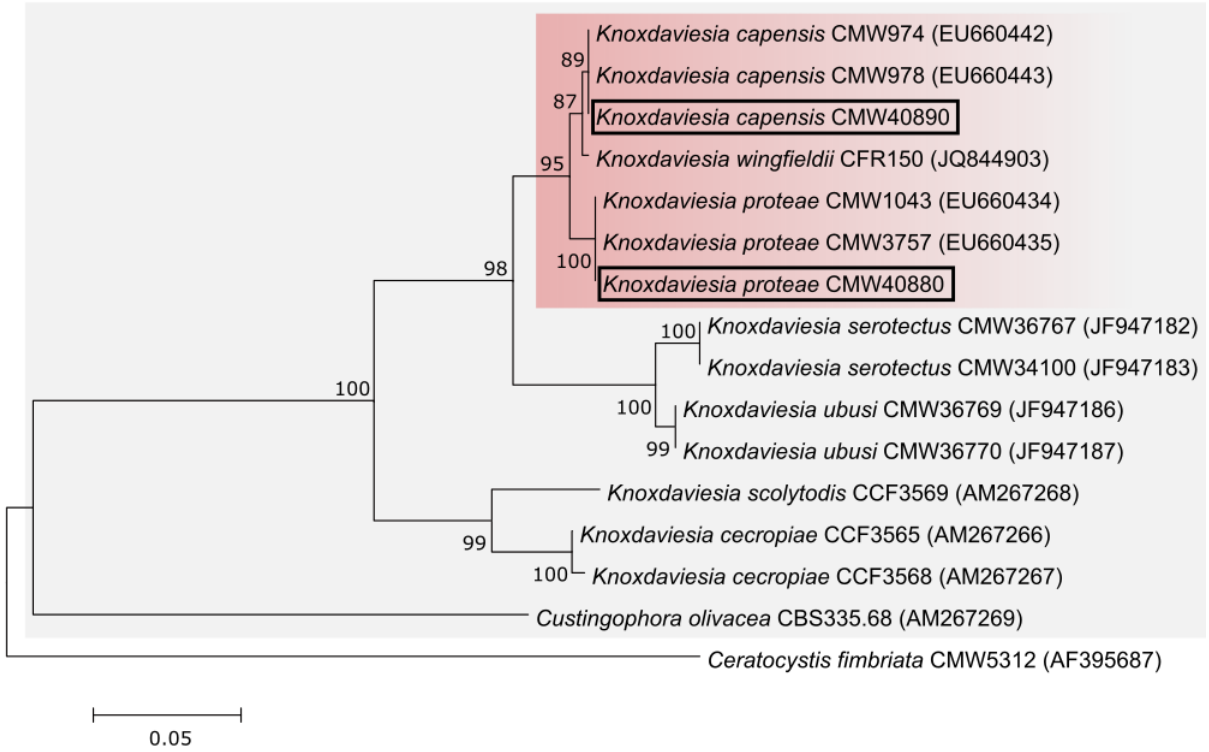

Figure 3.1 Maximum Likelihood tree illustrating the phylogenetic position of *K. capensis* and *K. proteae* in the Gondwanamycetaceae (grey block). The *Protea*-associated species are shaded red and the two isolates for which genome sequences were determined are indicated with a box. The sequences of the Internal Transcribed Spacer (ITS) region (available from GenBank®, accession numbers in brackets following isolate numbers) were aligned in MAFFT 7 (Katoh & Standley, 2013). The phylogeny was calculated in MEGA6 (Tamura *et al.*, 2013) using the Tamura-Nei substitution model (Tamura & Nei, 1993), 1,000 bootstrap replicates and *Ceratocystis fimbriata* (Ceratocystidaceae) as an outgroup.

In a study determining the genome sequence of any fungus, it is advisable to use a living isolate connected to the type specimen. However, the ex-type isolate of *K. proteae* (CMW738 = CBS486.88) is more than 20 years old and does not display the characteristic morphological features of the fungus in culture anymore. No living ex-type isolate exists for *K. capensis*. We thus collected fresh isolates of both species for this study in order to eliminate possible mutations or degradation that may have occurred though continual artificial propagation in culture media. The new isolates (Fig. 3.1 and 3.2) were collected from the same localities and hosts as the holotype specimens: *K. capensis* (CMW40890 =

CBS139037) from the infructescences of *P. longifolia* Andrews in Hermanus, and *K. proteae* (CMW40880 = CBS140089) from *P. repens* infructescences in Stellenbosch, both locations in the Western Cape Province of South Africa. General features of these isolates are outlined in Table 1.

Figure 3.2 Sexual sporing structures of the two *Knoxdaviesia* species sequenced in this study. *Knoxdaviesia capensis* (A) and *K. proteae* (B) were sampled from *Protea longifolia* and *P. repens* flowers, respectively. Scale bars = 1 mm

Table 3.1 Classification and general features of *Knoxdaviesia capensis* and *K. proteae* (Field *et al.*, 2008).

MIGS ID	Property	*K. capensis* Term	*K. proteae* Term	Evidence code[1]
	Classification	Domain Fungi	Domain Fungi	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
		Phylum Ascomycota	Phylum Ascomycota	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
		Class Sordariomycetes	Class Sordariomycetes	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
		Order Microascales	Order Microascales	TAS (Wingfield *et al.*, 1999)
		Family Gondwanamycetaceae	Family Gondwanamycetaceae	TAS (Réblová *et al.*, 2011)
		Genus *Knoxdaviesia*	Genus *Knoxdaviesia*	TAS (De Beer *et al.*, 2013)
		Species *capensis* Strain: CMW40890 = CBS139037	Species *proteae* Strain: CMW40880 = CBS140089	TAS (De Beer *et al.*, 2013)
	Cell shape	septate, smooth-walled hyphae	septate, smooth-walled hyphae	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
	Motility	Non-motile	Non-motile	NAS
	Sporulation	Unsheathed allantoid ascospores	Falcate ascospores	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
	Temperature range	15-30°C	15-30°C	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
	Optimum temperature	25°C	25°C	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)

	pH range; Optimum	Unknown	Unknown	
	Carbon source	Unknown	Unknown	
MIGS-6	Habitat	Seed cones (infructescences) of *Protea* spp.	Seed cones (infructescences) of *Protea repens* L.	TAS (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988)
MIGS-6.3	Salinity	Unknown	Unknown	
MIGS-22	Oxygen requirement	Aerobic; requirement/tolerance unknown	Aerobic; requirement/tolerance unknown	IDA
MIGS-15	Biotic relationship	Plant-associated	Plant-associated	TAS (Marais *et al.*, 1998)
MIGS-14	Pathogenicity	None known	None known	IDA
MIGS-4	Geographic location	Hermanus, South Africa	Stellenbosch, South Africa	IDA
MIGS-5	Sample collection	February 2014	January 2014	IDA
MIGS-4.1	Latitude	-34.4093	-33.9430	IDA
MIGS-4.2	Longitude	19.2150	18.8802	IDA
MIGS-4.4	Altitude	20 m	140 m	IDA

[1] Evidence codes - *IDA*: Inferred from Direct Assay; *TAS*: Traceable Author Statement (i.e., a direct report exists in the literature); *NAS*: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from http://www.geneontology.org/GO.evidence.shtml of the Gene Ontology project (Ashburner *et al.*, 2000)

GENOME SEQUENCING INFORMATION

Genome project history

Considering the lack of ecological information on the genus *Knoxdaviesia* and the close relationship these Microascalean fungi have to important plant pathogens, two *Protea*-associated *Knoxdaviesia* species, believed to be native to the CCR in South Africa, were selected for genome sequencing. Both species were sequenced at Fasteris in Switzerland. The genome projects are listed in the Genomes OnLine Database (GOLD) (Pagani *et al.*, 2012) and the whole genome shotgun (WGS) project has been deposited at DDBJ/EMBL/GenBank (Table 3.2). Table 3.2 presents the project information and its association with the minimum information about a genome sequence (MIGS) version 2.0 compliance (Field *et al.*, 2008). The full MIGS records for *K. capensis* and *K. proteae* are available in Tables S3.1 and S3.2, respectively.

Table 3.2 Project information.

MIGS ID	Property	*K. capensis* Term	*K. proteae* Term
MIGS 31	Finishing quality	High quality draft	High quality draft
MIGS-28	Libraries used	2x paired-end (PE) (350 and 550 bp) and 1x mate-pair (MP) (3 kbp)	2x paired-end (PE) (350 and 550 bp) and 1x mate-pair (MP) (3 kbp)
MIGS 29	Sequencing platforms	Illumina Hiseq 2500	Illumina Hiseq 2500
MIGS 31.2	Fold coverage	PE library 1: 91.6 x PE library 2: 80 x MP library: 17 x	PE library 1: 142 x PE library 2: 79.3 x MP library: 50.2 x
MIGS 30	Assemblers	ABySS 1.5.2; SSPACE 3.0	ABySS 1.5.2; SSPACE 3.0
MIGS 32	Gene calling method	MAKER 2.31.8	MAKER 2.31.8
	Genbank ID	LNGK00000000	LNGL00000000
	GenBank Date of Release	11 January 2016	11 January 2016
	GOLD ID	Gp0093999	Gp0110284
	BIOPROJECT	PRJNA246171	PRJNA275563
MIGS 13	Source Material Identifier	CMW40890/CBS139037	CMW40880/CBS 140089
	Project relevance	Biodiversity, evolution	Biodiversity, evolution

Growth conditions and nucleic acid isolation

Both *K. capensis* and *K. proteae* were cultured on Malt Extract Agar (MEA; Merck, Wadeville, South Africa) overlaid with sterile cellophane sheets (Product no. Z377597, Sigma-Aldrich, Steinham, Germany). After 10 days of growth at 25°C, mycelia was scraped from the cellophane and DNA was extracted according to Aylward *et al.* (2014a). Approximately 5 µg DNA from each species was used to prepare the three Illumina libraries (Table 3.2).

RNA was extracted from the *K. proteae* genome isolate to use as evidence for gene prediction. After growth on MEA at 25°C for approximately 10 days, total RNA was isolated from the mycelia with the PureLink™ RNA Mini Kit (Ambion, Austin, TX, USA). Quality control was performed on the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) using the RNA 6000 Nano Assay kit (Agilent Technologies, USA). The mRNA component of the total RNA was subsequently extracted with the Dynabeads® mRNA purification kit (Ambion, Austin, TX, USA).

Genome sequencing and assembly

The genomes of *K. capensis* and *K. proteae* were sequenced with the Illumina HiSeq 2500 platform at Fasteris, Switzerland, using two paired-end and one Nextera mate-pair library (Table 3.2). More than 60 million paired-end and 8 million mate-pair reads were obtained for each species. These reads were trimmed in CLC Genomics Workbench 6.5 (CLC bio, Aarhus, Denmark) so that the Phred *Q* (quality) score of each base was at least Q20. VelvetOptimiser (Gladman & Seeman, unpublished), a Perl script used as part of the Velvet assembler (Zerbino, 2010; Zerbino & Birney, 2008), was initially used to optimize the assembly parameters. Assembly of contigs was performed in ABySS 1.5.2 (Simpson *et al.*, 2009) using the optimal parameters suggested by VelvetOptimiser as a starting point. Several assemblies were computed using kmer-values slightly higher and lower than the kmer-value suggested by VelvetOptimiser. The assembly with the lowest number of contigs was used to build scaffolds in SSPACE 3.0 (Boetzer *et al.*, 2011), discarding scaffolds smaller than 1,000 bp. Automatic gap closure was performed in GapFiller 1.10 (Boetzer & Pirovano, 2012). The average genome coverage of each library was estimated using the Lander-Waterman equation

(total sequenced nucleotides/genome size) (Table 3.2), which yielded a combined average coverage for the three libraries of 188.5x (*K. capensis*) and 271.5x (*K. proteae*).

The *K. capensis* genome consists of 29 scaffolds ranging between 1,226 and 5,637,848 bp, whereas the 133 scaffolds of *K. proteae* are sized between 1,022 and 2,610,973 bp. A search for the 1,438 fungal universal single-copy ortholog genes with BUSCO 1.1b1 (Simão *et al.*, 2015) identified 1,355 complete and 67 partial genes in *K. capensis* and 1,366 complete and 57 partial genes in *K. proteae*. The two genomes are therefore estimated to be > 98 % complete.

The extracted mRNA of *K. proteae* was sequenced using an Ion PI™ Chip on the Ion Proton™ System (Life Technologies, Carlsbad, CA) at the Central Analytical Facility (CAF), Stellenbosch University, South Africa. The >49 million raw RNA-Seq reads were mapped to the *K. capensis* genome in CLC Genomics Workbench and assembled with Trinity 2.0.6 (Haas *et al.*, 2013) using the genome-guided option

Genome annotation

Genome annotation was performed with the MAKER 2.31.8 pipeline (Cantarel *et al.*, 2008; Holt & Yandell, 2011), using custom repeat libraries for each species constructed with RepeatScout 1.0.5 (Price *et al.*, 2005) and two *de novo* gene predictors, SNAP 2006-07-28 (Korf, 2004) and AUGUSTUS 3.0.3 (Stanke *et al.*, 2004). The assembled *K. proteae* RNA-Seq and predicted protein and/or transcript sequences from 22 sequenced Sordariomycete species (Table S3.3), including two Microascalean fungi, were provided as additional evidence. AUGUSTUS was trained with the assembled *K. proteae* RNA-Seq data and subsequently MAKER was used to annotate the largest scaffold of the *K. capensis* and the largest scaffold of the *K. proteae* assembly, independently. After manually curating all the gene predictions on these scaffolds with Apollo 1.11.8 (Lewis *et al.*, 2002), SNAP was trained with the curated gene predictions of each scaffold and the scaffolds were re-annotated. SNAP was re-trained for each species individually and subsequently both genomes were annotated. EuKaryotic Orthologous Group (KOG) classifications were assigned to the predicted proteins through the WebMGA (Wu *et al.*, 2011) portal that performs reverse-position-specific BLAST (Altschul *et al.*, 1990) searches on the KOG database (Tatusov *et al.*, 2003). Additional functional annotations were predicted with InterProScan 5.13-52.0

(Goujon *et al.*, 2010; Zdobnov & Apweiler, 2001), SignalP 4.1 (Petersen *et al.*, 2011) and TMHMM 2.0 (Krogh *et al.*, 2001).

GENOME PROPERTIES

Knoxdaviesia capensis and *K. proteae* have similar genome sizes at 35.54 and 35.49 Mbp, respectively. It was possible to assemble the *K. capensis* genome into 29 scaffolds larger than 1,000 bp, whereas the number of scaffolds above this threshold achieved for *K. proteae* was 133. Both genomes had a GC content of 52.8%.

A total of 7,940 protein-coding genes were predicted for *K. capensis* and 8,174 for *K. proteae*. Additionally 137 and 116 tRNA and 30 and 27 rRNA genes were predicted for each species, respectively. More than 74% of the protein-coding genes of each species could be assigned to a putative function via the KOG and Pfam databases. The content of the two genomes are summarized in Tables 3.3-3.4.

CONCLUSIONS

At least six Microascalean fungi currently have publically accessible genomes (Van der Nest *et al.*, 2014a; 2014b; Vandeputte *et al.*, 2014; Wilken *et al.*, 2013). *Knoxdaviesia capensis* and *K. proteae*, however, represent the first sequenced genomes from the Microascalean family *Gondwanamycetaceae*. The genomes of these two species will not only enable further understanding of the unique ecology of *Protea*-inhabiting fungi, but will also be valuable in taxonomic and evolutionary studies.

Table 3.3 Genome statistics.

Species	***K. capensis***		***K. proteae***	
Attribute	**Value**	**% of Total[1]**	**Value**	**% of Total[1]**
Genome size (bp)	35,537,816	100.00	35,489,142	100.00
DNA coding (bp)	12,640,368	35.57	12,542,580	35.34
DNA G+C (bp)	18,774,628	52.83	18,745,365	52.82
DNA scaffolds	29		133	
Total genes	8,107	100.00	8,316	100.00
Protein coding genes	7,940	97.94	8,173	98.28
RNA genes[2]	167	2.06	143	1.72
Pseudo genes	unknown		unknown	
Genes in internal clusters	unknown		unknown	
Genes with function prediction	6,192	77.98	6,093	74.55
Genes assigned to KOGs	6,059	76.31	6,015	73.60
Genes with Pfam domains	5,455	68.70	5,335	65.28
Genes with signal peptides	354	4.46	335	4.10
Genes with transmembrane helices	1,510	19.02	1,527	18.68

[1] The total is based on either the size of the genome in base pairs or the total number of protein-coding genes in the annotated genome.

[2] Based on tRNA and rRNA genes only

Table 3.4 Number of genes associated with the 25 general KOG functional categories.

Species	***K. capensis***		***K. proteae***		
Code	**Value**	**% of total[1]**	**Value**	**% of total[1]**	**Description**
J	359	4.52	371	4.54	Translation, ribosomal structure and biogenesis
A	280	3.53	273	3.34	RNA processing and modification
K	475	5.98	484	5.92	Transcription
L	196	2.47	198	2.42	Replication, recombination and repair
B	109	1.37	99	1.21	Chromatin structure and dynamics
D	209	2.63	227	2.78	Cell cycle control, cell division, chromosome partitioning
Y	34	0.43	32	0.39	Nuclear structure
V	32	0.40	32	0.39	Defence mechanisms
T	505	6.36	586	5.95	Signal transduction mechanisms
M	69	0.87	76	0.93	Cell wall/membrane/envelope biogenesis
N	6	0.08	6	0.07	Cell motility
Z	279	3.51	289	3.54	Cytoskeleton
W	10	0.13	12	0.15	Extracellular structures
U	539	6.79	543	6.64	Intracellular trafficking, secretion, and vesicular transport
O	502	6.32	495	6.06	Post-translational modification, protein turnover, chaperones
C	265	3.34	256	3.13	Energy production and conversion
G	202	2.54	202	2.47	Carbohydrate transport and metabolism
E	227	2.86	228	2.79	Amino acid transport and metabolism
F	76	0.96	74	0.91	Nucleotide transport and metabolism
H	87	1.10	85	1.04	Coenzyme transport and metabolism
I	234	2.95	234	2.86	Lipid transport and metabolism
P	144	1.81	151	1.85	Inorganic ion transport and metabolism
Q	139	1.75	137	1.68	Secondary metabolites biosynthesis, transport and catabolism
R	735	9.26	694	8.49	General function prediction only
S	344	4.33	330	4.04	Function unknown
X	2	0.03	1	0.01	Multiple functions
-	1881	23.69	2159	26.41	Not in KOGs

[1] The total is based on the total number of protein coding genes in the genome.

ACKNOWLEDGEMENTS

We are grateful to Quentin Santana and Dr. Lieschen Bahlmann for their guidance in the genome assembly and annotation procedures and to Dr. Wilhelm de Beer for the taxonomic information he contributed to this manuscript. This research was funded by the National Research Foundation (NRF) and the Department of Science and Technology-NRF Centre of Excellence in Tree Health Biotechnology. We also thank the Cape Nature Conservation Board for supplying the necessary collection permits.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. & Eppig, J. T. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2014a). Development of polymorphic microsatellite markers for the genetic characterisation of *Knoxdaviesia proteae* (Ascomycota: Microascales) using ISSR-PCR and pyrosequencing. *Mycological Progress* **13**, 439-444.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2014b). Panmixia defines the genetic diversity of a unique arthropod-dispersed fungus specific to *Protea* flowers. *Ecology and Evolution* **4**, 3444-3455.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015). Long-distance dispersal and recolonization of a fire-destroyed niche by a mite-associated fungus. *Fungal Biology* **119**, 245-256.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579.

Boetzer, M. & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology* **13**, R56.

Brasier, C. M. (1991). *Ophiostoma novo-ulm*i sp. nov., causative agent of current Dutch elm disease pandemics. *Mycopathologia* **115**, 151-161.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S. & Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**, 188-196.

Crous, P. W., Summerell, B. A., Shivas, R. G., Burgess, T. I., Decock, C. A., Dreyer, L. L., Granke, L. L., Guest, D. I., Hardy, G. E. S. T. J., Hausbeck, M. K., *et al.* (2012). Fungal Planet description sheets: 107-127. *Persoonia* **28**, 138-182.

De Beer, Z. W., Seifert, K. A. & Wingfield, M. J. (2013). A nomenclature for ophiostomatoid genera and species in the *Ophiostomatales* and *Microascales*. In *Ophiostomatoid fungi: expanding frontiers*, pp. 245-322. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield: CBS Biodiversity Series.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., Allen, M. J. & Angiuoli, S. V. (2008). The minimum information about a genome sequence (MIGS) specification. *Nature Biotechnology* **26**, 541-547.

Gibbs Russell, G. (1984). Analysis of the size and composition of the southern African flora. *Bothalia* **15**, 613-629.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J. & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research* **38**, W695-W699.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B. & Lieber, M. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494-1512.

Harrington, T. C. & Wingfield, M. J. (1998). The *Ceratocystis* species on conifers. *Canadian Journal of Botany* **76**, 1446-1457.

Hawksworth, D. L. (2011). A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. *IMA Fungus* **2**, 155.

Holt, C. & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491.

Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59.

Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

Lewis, S. E., Searle, S., Harris, N., Gibson, M., Lyer, V., Richter, J., Wiel, C., Bayraktaroglir, L., Birney, E. & Crosby, M. (2002). Apollo: a sequence annotation editor. *Genome Biology* **3**, 1-14.

Malloch, D. & Blackwell, M. (1993). Dispersal biology of the ophiostomatoid fungi. In *Ceratocystis and Ophiostoma: Taxonomy, Ecology and Pathology*, pp. 195-206. Edited by M. J. Wingfield, K. A. Seifert & J. F. Webber. St. Paul: APS Press.

Manning, J. & Goldblatt, P. (2012). Plants of the Greater Cape Floristic Region. 1: The Core Cape Flora. In *Strelitzia 29*. Pretoria: South African National Biodiversity Institute.

Marais, G. J., Wingfield, M. J., Viljoen, C. D. & Wingfield, B. D. (1998). A New Ophiostomatoid Genus from *Protea* Infructescences. *Mycologia* **90**, 136-141.

McNeill, J., Barrie, F., Buck, W., Demoulin, V., Greuter, W., Hawksworth, D., Herendeen, P., Knapp, S., Marhold, K. P., Prado, J., *et al.* (2012). *International Code of Nomenclature for algae, fungi, and plants (Melbourne Code) adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011*: Regnum Vegetabile 154, Koeltz Scientific Books.

Mittermeier, R. A., Myers, N., Thomsen, J. B., Da Fonseca, G. A. B. & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology* **12**, 516-520.

Pagani, I., Liolios, K., Jansson, J., Chen, I.-M. A., Smirnova, T., Nosrat, B., Markowitz, V. M. & Kyrpides, N. C. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research* **40**, D571-D579.

Paine, T. D., Raffa, K. F. & Harrington, T. C. (1997). Interactions among scolytid bark beetles, their associated fungi, and live host conifers. *Annual Review of Entomology* **42**, 179-206.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.

Price, A. L., Jones, N. C. & Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, i351-i358.

Rebelo, T. (2001). *Proteas: A field guide to the Proteas of Southern Africa*. Vlaeberg, South Africa: Fernwood Press.

Réblová, M., Gams, W. & Seifert, K. A. (2011). *Monilochaetes* and allied genera of the *Glomerellales*, and a reconsideration of families in the *Microascales*. *Studies in Mycology* **68**, 163-191.

Roets, F., Dreyer, L. L. & Crous, P. W. (2005). Seasonal trends in colonisation of *Protea* infructescences by *Gondwanamyces* and *Ophiostoma* spp. *South African Journal of Botany* **71**, 307-311.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2007). Discovery of Fungus-Mite Mutualism in a Unique Niche. *Environmental Entomology* **36**, 1226-1237.

Roets, F., Crous, P. W., Wingfield, M. J. & Dreyer, L. L. (2009). Mite-mediated hyperphoretic dispersal of *Ophiostoma* spp. from the Infructescences of South African *Protea* spp. *Environmental Entomology* **28**, 143-152.

Roets, F., Theron, N., Wingfield, M. J. & Dreyer, L. L. (2011a). Biotic and abiotic constraints that facilitate host exclusivity of *Gondwanamyces* and *Ophiostoma* on *Protea*. *Fungal Biology* **116**, 49-61.

Roets, F., Wingfield, M. J., Wingfield, B. D. & Dreyer, L. L. (2011b). Mites are the most common vectors of the fungus *Gondwanamyces proteae* in *Protea* infructescences. *Fungal Biology* **115**, 343-350.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2013). Taxonomy and ecology of ophiostomatoid fungi associated with *Protea* infructescences. In *Ophiostomatoid fungi: expanding frontiers*, pp. 177-187. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield. Utrecht, The Netherlands: CBS Biodiversity Series.

Roux, J. & Wingfield, M. J. (2009). *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. *Southern Forests* **71**, 115-120.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Supplementary Online Materials: http://busco.ezlab.org/files/BUSCO-SOM.pdf.

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. & Birol, İ. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research* **19**, 1117-1123.

Spatafora, J. W. & Blackwell, M. (1994). The polyphyletic origins of ophiostomatoid fungi. *Mycological Research* **98**, 1-9.

Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Research* **32**, W309-W312.

Tamura, K. & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**, 512-526.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L. & Nikolskaya, A. N. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**, 41.

Van der Nest, M. A., Beirn, L. A., Crouch, J. A., Demers, J. E., de Beer, Z. W., De Vos, L., Gordon, T. R., Moncalvo, J. M., Naidoo, K., Sanchez-Ramirez, S., *et al.* (2014a). IMA Genome-F 3: Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntiella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana*, and *Sclerotinia echinophila*. *IMA Fungus* **5**, 473-486.

Van der Nest, M. A., Bihon, W., De Vos, L., Naidoo, K., Roodt, D., Rubagotti, E., Slippers, B., Steenkamp, E. T., Wilken, P. M. & Wilson, A. (2014b). Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. *IMA Fungus* **5**, 135-140.

Vandeputte, P., Ghamrawi, S., Rechenmann, M., Iltis, A., Giraud, S., Fleury, M., Thornton, C., Delhaès, L., Meyer, W., Papon, N., *et al.* (2014). Draft Genome Sequence of the Pathogenic Fungus *Scedosporium apiospermum*. *Genome Announcements* **2**, e00988-00914.

Wilken, P. M., Steenkamp, E. T., Wingfield, M. J., De Beer, Z. W. & Wingfield, B. D. (2013). IMA Genome-F 1: *Ceratocystis fimbriata*: Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA Fungus* **4**, 357.

Wingfield, B. D., Viljoen, C. D. & Wingfield, M. J. (1999). Phylogenetic relationships of ophiostomatoid fungi associated with *Protea* infructescences in South Africa. *Mycological Research* **103**, 1616-1620.

Wingfield, M. J., Wyk, P. S. V. & Marasas, W. F. O. (1988). *Ceratocystiopsis proteae* sp. *nov*., with a new anamorph genus. *Mycologia* **80**, 23-30.

Wingfield, M. J. & Van Wyk, P. S. (1993). A new species of *Ophiostoma* from *Protea* infructescences in South Africa. *Mycological Research* **97**, 709-716.

Wu, S., Zhu, Z., Fu, L., Niu, B. & Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* **12**, 444.

Zdobnov, E. M. & Apweiler, R. (2001). InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848.

Zerbino, D. R. & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829.

Zerbino, D. R. (2010). Using the Velvet de novo Assembler for Short-Read Sequencing Technologies. *Current Protocols in Bioinformatics* **Unit 11.5**, 1-13.

SUPPLEMENTARY INFORMATION

Table S3.1 Associated MIGS record for *Knoxdaviesia capensis*.

MIGS-ID	Field Name	Description
MIGS-2	MIGS CHECK LIST TYPE	Eukaryote
MIGS-3	Project Name	*Knoxdaviesia capensis* H9 genome sequencing
MIGS-4	Geographic Location	South Africa: Western Cape Province: Hermanus
4.1	Latitude	-34.4093
4.2	Longitude	19.2150
MIGS-5	Time of Sample collection	2014-02-14
MIGS-6	Habitat (EnvO)	Plant-associated
MIGS-7	Subspecific genetic lineage	
MIGS-8	Ploidy	Haploid
MIGS-9	Number of replicons	Unknown
MIGS-10	Extrachromosomal elements	
MIGS-11	Estimated Size	35,000,000
MIGS-12	Reference for biomaterial or Genome report	(this manuscript)
MIGS-13	Source material identifiers	CBS 139037; CMW 40890
MIGS-14	Known Pathogenicity	Non-pathogen
MIGS-15	Biotic Relationship	Symbiont
MIGS-16	Specific Host	*Protea longifolia* (TaxID 73576)
MIGS-17	Host specificity or range (taxid)	*Protea* spp. (TaxID 4331)
MIGS-18	Health status of Host	Asymptomatic
MIGS-19	Trophic Level	Heterotroph
MIGS-20	Propagation	Sexual and asexual
MIGS-22	Relationship to Oxygen	Aerobic
MIGS-23	Isolation and Growth conditions	DOI: 10.1007/s11557-013-0951-1
MIGS-27	Nucleic acid preparation	PMCID: PMC334490
MIGS-28	Library construction	Paired-end
28.1	Library size	350, 550, 3 000
MIGS-29	Sequencing method	Illumina HiSeq 2500
MIGS-30	Assembly	
30.1	Assembly method	ABYSS 1.5.2, SSPACE 3.0
30.2	Estimated error rate	1 in 100 bp
30.3	Method of calculation	PHRED
MIGS-31	Finishing strategy	
31.1	Status	High-quality draft
31.3	Contigs	29

Table S3.2 Associated MIGS record for *Knoxdaviesia proteae*.

MIGS-ID	Field Name	Description
MIGS-2	MIGS CHECK LIST TYPE	Eukaryote
MIGS-3	Project Name	*Knoxdaviesia proteae* strain SB2.3 genome sequencing
MIGS-4	Geographic Location	South Africa: Western Cape Province: Stellenbosch
4.1	Latitude	-33.9430
4.2	Longitude	18.8802
MIGS-5	Time of Sample collection	2014-01-26
MIGS-6	Habitat (EnvO)	Plant-associated
MIGS-7	Subspecific genetic lineage	
MIGS-8	Ploidy	Haploid
MIGS-9	Number of replicons	Unknown
MIGS-10	Extrachromosomal elements	
MIGS-11	Estimated Size	35,000,000
MIGS-12	Reference for biomaterial or Genome report	(this manuscript)
MIGS-13	Source material identifiers	CBS 140089; CMW 40880
MIGS-14	Known Pathogenicity	Non-pathogen
MIGS-15	Biotic Relationship	Symbiont
MIGS-16	Specific Host	*Protea repens* (TaxID 4332)
MIGS-17	Host specificity or range (taxid)	*Protea repens* (TaxID 4332)
MIGS-18	Health status of Host	Asymptomatic
MIGS-19	Trophic Level	Heterotroph
MIGS-20	Propagation	Sexual and asexual
MIGS-22	Relationship to Oxygen	Aerobic
MIGS-23	Isolation and Growth conditions	DOI: 10.1007/s11557-013-0951-1
MIGS-27	Nucleic acid preparation	PMCID: PMC334490
MIGS-28	Library construction	Paired-end
28.1	Library size	350, 550, 3 000
MIGS-29	Sequencing method	llumina HiSeq 2500
MIGS-30	Assembly	
30.1	Assembly method	ABYSS 1.5.2, SSPACE 3.0
30.2	Estimated error rate	1 in 100 bp
30.3	Method of calculation	PHRED
MIGS-31	Finishing strategy	
31.1	Status	High-quality draft
31.3	Contigs	29

Table S3.3 is provided in the electronic supplementary materials.

Declaration by the candidate (for published chapter)

With regards to Chapter 4 (Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*), the nature and scope of my contribution were as follows:

Nature of contribution	**Extent of contribution (%)**
Identification and characterization of *MAT* loci	100
PCR and sequencing	100
Mating type crosses	100
Manuscript preparation	75

The following co-authors have contributed to Chapter 4:

Name	**e-mail address**	**Nature of contribution**	**Extent of contribution (%)**
Emma T. Steenkamp	emma.steenkamp@up.ac.za	Provided guidance and edited the manuscript	5
Leanne L. Dreyer	ld@sun.ac.za	"	5
Francois Roets	fr@sun.ac.za	"	5
Michael J. Wingfield	michael.wingfield@up.ac.za	"	5
Brenda D. Wingfield	brenda.wingfield@up.ac.za	"	5

Signature of candidate:  …….

Declaration by co-authors:

The undersigned hereby confirm that

1. The declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 4.
2. No other authors contributed to Chapter 4 than those specified above, and
3. Potential conflict of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 4 of this dissertation

Signature	**Institutional affiliation**	**Date**
Leanne L. Dreyer	Stellenbosch University	6 Oct. 2016
Francois Roets	Stellenbosch University	6 Oct. 2016
Emma T. Steenkamp	University of Pretoria	12/10/2016
Michael J. Wingfield	University of Pretoria	6 Oct 2016
Brenda D. Wingfield	University of Pretoria	6 Oct 2016

CHAPTER 4

Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*

ABSTRACT

Sexual reproduction is necessary to generate genetic diversity and, in ascomycete fungi, this process is controlled by a mating type (*MAT*) locus with two complementary idiomorphs. *Knoxdaviesia capensis* and *K. proteae* (Sordariomycetes; Microascales; Gondwanamycetaceae) are host-specific saprotrophic fungi that show high population diversity within their *Protea* plant hosts in the Core Cape Subregion of South Africa. We hypothesise that this diversity is the result of outcrossing driven by a heterothallic mating system and sought to describe the *MAT1* loci of both species. The available genome assembly of each isolate contained only one of the *MAT1* idiomorphs necessary for sexual reproduction, implying that both species are heterothallic. Idiomorph segregation during meiosis, a 1:1 ratio of idiomorphs in natural populations and mating experiments also supported heterothallism as a sexual strategy. Long-range PCR and shot-gun sequencing to identify the opposite idiomorph in each species revealed no sequence similarity between *MAT1-1* and *MAT1-2* idiomorphs, but the homologous idiomorphs between the species were almost identical. The *MAT1-1* idiomorph contained the characteristic *MAT1-1-1* and *MAT1-1-2* genes, whereas the *MAT1-2* idiomorph consisted of the genes *MAT1-2-7* and *MAT1-2-1*. This gene content was similar to that of the three species in the Ceratocystidaceae (Microascales) with characterised *MAT* loci. The *Knoxdaviesia MAT1-2-7* protein contained and alpha domain and predicted intron, which suggests that this gene arose from *MAT1-1-1* during a recombination event. In contrast to the Ceratocystidaceae species, *Knoxdaviesia* conformed to the ancestral Sordariomycete arrangement of flanking genes and is, therefore, a closer reflection of the structure of this locus in the Microascalean ancestor.

INTRODUCTION

Sexual reproduction is universal across eukaryotic life, despite being more biologically costly than asexual propagation (Lehtonen *et al.*, 2012; Ni *et al.*, 2011; Otto, 2009; Stearns, 1987). Most fungi maintain both of these reproductive strategies that are controlled by diverse genetic mechanisms (Billiard *et al.*, 2012; Ni *et al.*, 2011). Evidence suggests that many fungi exploit the low cost of clonal reproduction during favourable environmental conditions, but switch to sexual reproduction under stress when adaptation becomes necessary (Ni *et al.*, 2010; Nielsen & Heitman, 2007; Seymour *et al.*, 2005). Novel allele combinations are essential for adaptation and the re-shuffling of genetic material enables selection against harmful or unfavourable genotypes that may be propagated through clonal reproduction (Lynch *et al.*, 1993).

The mating type (*MAT*) genes of fungi control the recognition between sexual partners and the subsequent development of sexual progeny (Coppin *et al.*, 1997; Perkins, 1987). In ascomycetes, mating type is determined by a single locus, *MAT1*, and two mating type idiomorphs (dissimilar alleles), *MAT1-1* and *MAT1-2* (Kronstad & Staben, 1997; Nelson, 1996; Turgeon & Yoder, 2000). Homothallic fungi have a *MAT1-1/2* genotype (Turgeon & Yoder, 2000); both *MAT1* idiomorphs occur in one genome, making them self-fertile. In heterothallic species, the absence of either idiomorph results in self-sterility that necessitates outcrossing between two individuals of opposite mating type for sexual reproduction (Kronstad & Staben, 1997; Nelson, 1996). Although it is widely accepted that the presence of both idiomorphs is necessary for sexual reproduction, fungal mating strategies are diverse and many exceptions to this rule have been observed (Heitman, 2015). For example, some species are self-fertile despite only possessing one *MAT1* idiomorph (unisexual reproduction; Alby & Bennett, 2011; Glass & Smith, 1994; Lin *et al.*, 2005; Wilson *et al.*, 2015).

The two idiomorphs of the ascomycete *MAT1* locus each contain at least one open reading frame (ORF) with a characteristic motif (Turgeon & Yoder, 2000). The *MAT1-1* idiomorph is defined by an ORF with an alpha domain (*MAT1-1-1*), although up to two additional "accessory" ORFs can occur in this idiomorph. The *MAT1-2* idiomorph generally has a single ORF (*MAT1-2-1*) with an HMG-box motif. The functions of each of these *MAT1* genes is not fully understood, but it is believed that each idiomorph encodes transcription factors (Herskowitz, 1989) that ultimately perform a dual function. Firstly, the transcription factors mediate a hormonal recognition mechanism between individuals by producing a pheromone

and receptors for the pheromone of the opposite mating type (Coppin *et al.*, 1997; Glass *et al.*, 1990; Ni *et al.*, 2011). Secondly, these genes are involved in the formation of sexual structures (Coppin *et al.*, 1997). Molecular studies have shown that the *MAT1-1-1* gene alone is able to induce fertilization, but in *Podospora anserina* (Rabenh.) Niessl, the accessory genes analogous to *MAT1-1-2* and *MAT1-1-3* are needed for the sexual structures to develop fully (Debuchy *et al.*, 1993), supporting the dual function. As the only consistently occurring ORF on the *MAT1-2* idiomorph, *MAT1-2-1* appears to be involved in ascomatal development and is the sole determinant of the necessary functions in this mating type (Coppin *et al.*, 1997; Staben & Yanofsky, 1990).

Of the five Microascales (Sordariomycetes) families (Maharachchikumbura *et al.*, 2015; Réblová *et al.*, 2011) only three species in the predominantly plant-associated, agriculturally important Ceratocystidaceae (De Beer *et al.*, 2014) have been studied extensively in terms of mating type genetics. The sweet potato pathogen, *Ceratocystis fimbriata s.s.* Ellis & Halst., is homothallic and undergoes unidirectional mating type switching whereby it loses its *MAT1-2-1* gene and becomes self-sterile (Harrington & McNew, 1997; Wilken *et al.*, 2014). The other two Ceratocystidaceae species studied are members of the genus *Huntiella* that typically show a saprotrophic association with tree wounds (Van Wyk *et al.*, 2006). Both of the studied species in this genus are heterothallic (Wilson *et al.*, 2015), although *H. moniliformis* (Hedgc.) Z.W. de Beer, T.A. Duong & M.J. Wingf. is also capable of unisexual reproduction; since it contains a single *MAT1* idiomorph, yet produces ascomata (Wilson *et al.*, 2015). One trait that unites the diverse mating strategies in Ceratocystidaceae is their deviation from the consensus gene order of the Sordariomycetes. The cytoskeleton assembly control (*SLA2*) and DNA lyase (*APN2*) genes that flank the *MAT* locus in almost all Sordariomycetes (Debuchy & Turgeon, 2006) have an altered order and orientation in the Ceratocystidaceae. The genes that typically flank the downstream region of *MAT1* loci have shifted to an upstream position in *C. fimbriata*. A similar shift is evident in the two studied *Huntiella* species, although the *APN2* gene has shifted to a genomic position far from the *MAT1* locus (Wilson *et al.*, 2015).

The aim of this study was to describe the *MAT1* locus of two saprotrophic, but host-specific species in the Gondwanamycetaceae, which is also a member of the Microascales (Réblová *et al.*, 2011). These fungi (*Knoxdaviesia capensis* M.J. Wingf. & P.S. van Wyk and *K. proteae* M.J. Wingf., P.S. van Wyk & Marasas) occur in the seed cones of *Protea*, a keystone plant genus in the Core Cape Subregion of South Africa (Bergh *et al.*, 2014; Cowling, 1992;

Manning & Goldblatt, 2012). The arthropod, and possibly bird, vectors of these fungi disperse ascospores between *Protea* flower heads (Roets *et al.*, 2011b). After flowering, the *Protea* inflorescence matures into an enclosed seed cone in which *K. capensis* and *K. proteae* are visible on decaying floral structures as ascomata that present spore droplets on long ostiolar necks (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988). Although conidiophores may also be present, *K. capensis* and *K. proteae* sexual structures are abundant within infructescences (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988), indicating that sexual reproduction is prevalent and likely the dominant mode of reproduction at this stage of their life-cycle. As ascomycete fungi, *K. capensis* and *K. proteae* are haploid during the vegetative state and sexual reproduction would thus only add genetic diversity if it is not a result of self-fertilization (Fincham & Day, 1963; Milgroom, 1996; Moore & Novak Frazer, 2002). High gene and genotypic diversity and random allele association within two populations of *K. proteae* (Aylward *et al.*, 2014; 2015a) and nine populations of *K. capensis* (Aylward *et al.*, 2017) strongly suggest that sexual reproduction in these species is non-selfing. *Protea*-associated *Knoxdaviesia* individuals, therefore, regularly recombine to produce genetically novel offspring, but whether outcrossing is a prerequisite for sexual reproduction in these species (*i.e.* heterothallism) or whether it is optional (*i.e.* homothallism) remains unknown.

The genomes of *K. capensis* and *K. proteae* have recently been sequenced (Aylward *et al.*, 2016, Chapter 3). In this study, we used these genomes to investigate the genetic basis of mating in *Knoxdaviesia*. In so doing, we tested the hypothesis that the genetic diversity observed in natural populations of these species is due to outcrossing driven by a heterothallic mating system. As a secondary aim, we compared the identified *Knoxdaviesia MAT1* loci to the three species in the Ceratocystidaceae with characterised *MAT1* loci.

MATERIALS AND METHODS

Fungal isolates and genome sequences used

The genomes of *K. capensis* CBS139037 (LNGK00000000.1) and *K. proteae* CBS140089 (LNGL00000000.1) were sequenced in a previous study (Aylward *et al.*, 2016, Chapter 3) and are available in GenBank® (Benson *et al.*, 2013). The *MAT1* locus of *C. fimbriata* CMW14799, previously characterised from its sequenced genome (Wilken *et al.*, 2014), was

also obtained from GenBank (KF033902.1; KF033903.1). Other than the *Knoxdaviesia* genome isolates, three additional strains of *K. capensis* (CMW40886, CMW40889, CMW40892) and *K. proteae* (CMW40879, CMW40882, CMW40883) were used in this study to perform crossing experiments. All isolates were routinely cultured on Potato Dextrose Agar (PDA; Merck, Wadeville, South Africa) for approximately seven days at 25°C and maintained at 4°C.

Identification of *MAT* loci from genome sequences

Ceratocystis fimbriata s.s. is currently the species most closely related to *Knoxdaviesia* that has a characterised *MAT1* locus with available gene models. The predicted proteins of the *MAT1* locus of *C. fimbriata* CMW14799 (AHV84683-84701) were used to search for the *MAT1* locus in the genomes of *K. capensis* and *K. proteae* by performing local BLASTx searches in CLC Genomics workbench 6 (CLC Bio, Denmark). Preliminary analyses identified a single *MAT1* idiomorph from each *Knoxdaviesia* genome and revealed that the two genomes contained opposite *MAT1* idiomorphs. Subsequently, we mapped the raw sequence reads from the *K. capensis* genome (GenBank Accession: SRX1453186, SRX1453795 and SRX1453796) to the *K. proteae MAT1* region in CLC Genomics workbench to identify the terminal ends of the *MAT1* locus in *K. proteae*. The procedure was repeated using the *K. proteae* raw sequence reads (GenBank Accession: SRX1453891, SRX1453905 and SRX1453906) and the *K. capensis MAT1* region.

Identification of the opposite *MAT* idiomorph in each species

Since each *Knoxdaviesia* genome contained a single *MAT1* idiomorph, the opposite idiomorph of each species had to be determined from a strain with the opposite mating type. Primers were designed in the conserved flanking regions of the *MAT1* locus in both *Knoxdaviesia* genomes (Table 4.1; Fig. 4.1) and used, in conjunction with internal primers (Table 4.1; Fig. 4.1), to amplify the opposite idiomorph in *K. capensis* CMW40886 and *K. proteae* CMW40882. The 50 µl reactions contained 1.75 mM $MgCl_2$, 0.3 mM of each dNTP, 0.5 µM of each primer, approximately 250 ng template DNA, 2 µl DMSO, 10 µl 5x KAPA LongRange Buffer and 1.25 units of KAPA LongRange HotStart DNA Polymerase (KAPA Biosystems, Inc., Wilmington, MA). Initial denaturation was performed at 94°C for 4

minutes, followed by 35 cycles of 94°C for 30 seconds, 55.5°C for 30 seconds and 68°C for one minute per kb to be amplified (Table 4.1). The final extension step was at 72°C for 10 minutes. The products were submitted for paired-end shotgun sequencing on the Illumina MiSeq platform at Inqaba Biotechnical Industries (Pty) Ltd., Pretoria, South Africa.

The next generation reads were trimmed with Trimmomatic (Bolger *et al.*, 2014), discarding the first 10 bases, those with an average phred score below 25 (using a sliding window of four base pairs), and reads < 30 bp in length. The quality of the sequences was confirmed with FastQC 0.11.4 (Babraham Bioinformatics, Babraham Institute, Cambridge). A *de novo* assembly of the trimmed sequences of each idiomorph was performed in CLC Genomics workbench. Subsequently, the sequence reads were mapped back onto the assembly of each idiomorph to detect potential errors. In order to compare the two idiomorphs in each species and the two pairs of homologous idiomorphs, dotplot comparisons of the nucleic acid sequences were computed with YASS (Noé & Kucherov, 2005) using the default parameters. Since repeat structures have been found in the *MAT1* locus of *C. fimbriata* (Wilken *et al.*, 2014), the *Knoxdaviesia MAT1* loci were interrogated for repeat clusters with RepFind (Betley *et al.*, 2002).

Table 4.1 Summary of the *Knoxdaviesia MAT1* primers used in this study.

Primers			**Products**	
Number	**Name**	**5'-3'**	**Alpha domain**	**Product size (bp)**
1a	KcM1-F	CCG CAC TGT ACA TCA CAA CA	1a↔2	835
1b	KpM1-F	CCT CGT CTC GAA TGA AGG AG	1b↔2	357
2	KcM1-R	GGT CAC CGA AAA GAA GAC CA	**HMG-box domain**	
3	KpHMG-F	ATC CTC ATG CCA CAA TAC CC	3↔4	516
4	KpHMG-R	GAA GTT GAA GTC CGC TTT GC	**Long range *MAT1-1***	
5	KxMAT-F	TGG TTG AAA GGG AAA TGA GG	5↔2	3914
6a	KxMAT-R	AAG ACA AAG GAC GGC CTA GC	1c↔6a	6268
1c	LR-KpM1F	GAG TGG CCT TGT CTT GAC CT	**Long range *MAT1-2***	
6b	Apn-R	ATC TGT GCC GGT ACT TCA ACC	5↔4	5751
			3↔6b	2825

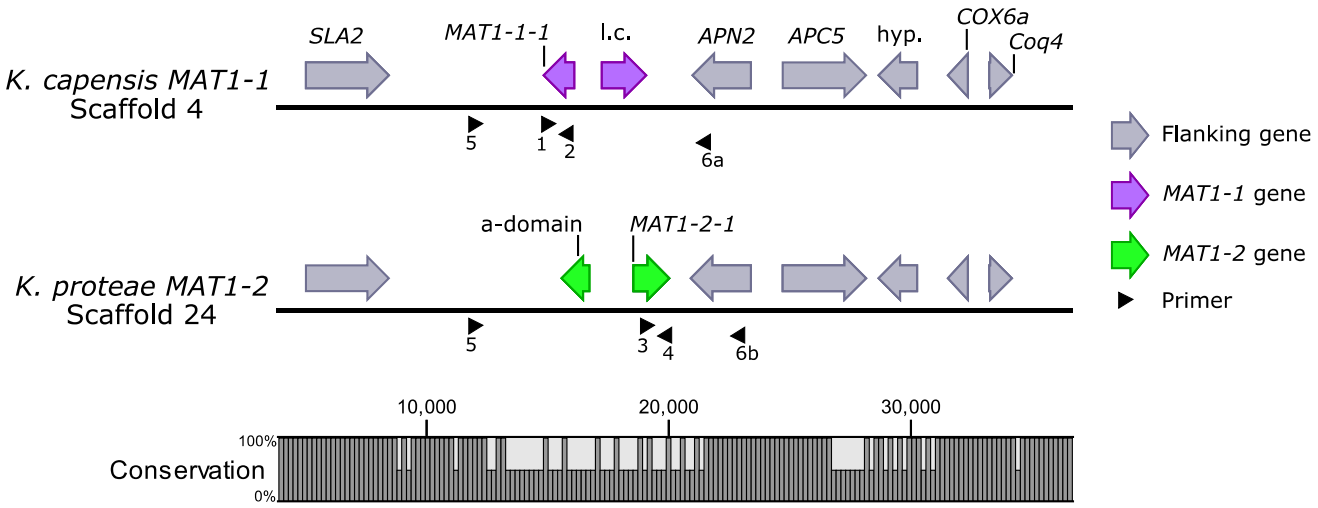

Figure 4.1 The *MAT1* mating type locus identified from the genomes of *Knoxdaviesia capensis* and *K. proteae*. The graph indicates nucleotide sequence conservation between the loci of the two fungi. Open reading frames are indicated by blocked arrows. The lack of conservation in the *APC5* gene is due to missing sequence data in the *K. proteae* genome sequence. l.c. = low complexity region protein, hyp. = hypothetical protein, a-domain = alpha domain-containing protein. See Table 4.1 for primer names and sequences. Primers 1a, 1b and 1c lie in the same approximate location.

Knoxdaviesia MAT genes

WEBAUGUSTUS (Hoff & Stanke, 2013) was used to predict open reading frames (ORFs) in the *MAT1* loci obtained from the *Knoxdaviesia* genomes and the amplified idiomorphs, using gene models from the closest available relative, *Fusarium graminearum* Schwabe. The putative identities of the predicted ORFs were determined by DELTA-BLAST protein searches (Boratyn *et al.*, 2012) at the National Centre for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Conserved domains in putative *MAT1* proteins were identified with searches at Interpro (Mitchell *et al.*, 2015) and NCBI's Conserved Domains Database (CDD; Marchler-Bauer *et al.*, 2015).

The *Knoxdaviesia MAT1* proteins were compared with those of *C. fimbriata* (KF033902), *H. moniliformis* and *H. omanensis* (Al-Subhi, M.J. Wingfield, M. van Wyk & Deadman) Z.W. De Beer, T.A. Duong & M.J. Wingf. (obtained from A. Wilson). In addition, we used the *MAT1-1-1*, *MAT1-1-2* and *MAT1-2-1* proteins of six other Sordariomycetes also used by Wilken *et al.* (2014): *Cryphonectria parasitica* (Murrill) M.E. Barr (AF380365/AF380364),

Cordyceps takaomontana Yakush. & Kumaz. (AB096216/AB084921), *Fusarium fujikuroi* Nirenberg (AF100925/AF100926), *Magnaporthe grisea* (T.T. Hebert) M.E. Barr (AB080670/AB080671), *Neurospora crassa* Shear & B.O. Dodge (M33876/ M54787) and *Podospora anserina* (X73830/X64194/X64195). These were aligned to the *Knoxdaviesia* proteins and to the consensus sequences of the alpha (PFAM04769) and HMG-box (PFAM00505) domains, using the accurate alignment algorithm in CLC Genomics workbench. The consensus sequences were obtained from the protein families (Pfam) database (Finn *et al.*, 2014).

Distribution of *MAT* idiomorphs in natural populations and ascomata

To determine the abundance of each *MAT1* idiomorph in natural populations, DNA previously extracted from 94 strains of *K. capensis* isolated from six different hosts (Aylward *et al.*, 2015b; 2017) and 184 strains of *K. proteae* from a single host (Aylward *et al.*, 2014; 2015a) in the Core Cape Subregion, were used (Table 4.2). Primers KcM1-F (1a in Fig. 4.1) and KcM1-R (2 in Fig. 4.1), that flank the conserved alpha domain of *MAT1-1-1*, and KpHMG-F (3 in Fig. 4.1) and KpHMG-R (4 in Fig. 4.1), that flank the HMG-box of *MAT1-2-1* in the two *Knoxdaviesia* species were designed with Primer3 (Table 4.1; Untergasser *et al.*, 2007). A new forward primer, Kp-M1-F (1b in Fig. 4.1), was later designed to bind to the *K. proteae* alpha domain region with greater specificity (Table 4.1). The sizes of the amplicons for these two conserved regions were calculated to differ by at least 150 bp, in order to be distinguished on an agarose gel and, therefore, to identify the *MAT1* idiomorph using a single multiplex PCR. The 20 µl PCRs contained 10 µl KAPA Taq ReadyMix (Kapa Biosystems, Inc., Boston, USA), $MgCl_2$ at a final concentration of 2.75 mM, 0.4 µM each of the alpha domain and HMG-box primer pairs and approximately 150 ng template DNA. PCR cycling conditions were 95°C for 3 minutes, followed by 35 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 1.15 minutes, and a final extension step of 72°C for 10 minutes. The hypothesis of a 1:1 distribution of *MAT1-1* and *MAT1-2* idiomorphs within different sampling localities and from different host species was tested with Pearson's Chi-Square and two-sided binomial tests in R version 3.2.3 (R Core Team, 2014).

Table 4.2 *Knoxdaviesia* strains and populations used in this study.

Species	Host	Location[1]	Number of strains	Reference
Knoxdaviesia capensis CMW40890	*Protea longifolia* Andrews	Hermanus	1	Aylward *et al*. (2016), Chapter 3
"	*P. coronata* Lam.	Ataraxia	15	Aylward *et al*. (2017)
"	"	Du Toits Kloof	11	"
"	"	Greyton	9	"
"	"	Kleinmond	5	"
"	"	Helderberg	9	"
"	*P. neriifolia* R. Br.	Betty's Bay	6	"
"	"	Kogelberg	8	"
"	*P. lepidocarpodendron* L.	Betty's Bay	7	"
"	*P. longifolia*	Kogelberg	9	"
	P. repens L.	Gouritz	7	Aylward *et al*. (2015b)
	"	Franschoek	10	"
Knoxdaviesia proteae CMW40880	*P. repens*	Stellenbosch	1	Aylward *et al*. (2016), Chapter 3
	"	Gouritz	83	Aylward *et al*. (2014)
	"	Franschoek	101	Aylward *et al*. (2015a)

[1] All localities are in the Western Cape Province, South Africa

The distribution of *MAT1* idiomorphs in ascospores derived from a single parent was also investigated. For both *K. capensis* and *K. proteae*, single ascospore progeny were cultured from a spore droplet taken from a single ascoma following the protocol of Wilson *et al.* (2015). Since *Knoxdaviesia* species do not readily form sexual structures in culture, a spore droplet from an ascoma on *P. coronata* (for *K. capensis*) and *P. repens* (for *K. proteae*) flowers were transferred to 20 µl Soltrol 130 oil (Chemfit, Gauteng, South Africa), vortexed and streaked out onto half-strength PDA (19.5g PDA/L; Merck, Wadeville, South Africa). For *K. proteae*, ascospores germinated only if an equal volume (20 µl) of sterile water was mixed with the Soltrol oil prior to streaking. Plates were incubated at 25°C and 30 germinating ascospores from each species were transferred onto fresh half-strength PDA after approximately 48 hours. After approximately 10 days at 25°C, DNA from the single ascospore progeny was extracted and amplified with the Extract-N-Amp™ Plant PCR Kit (Sigma-Aldrich, Steinham, Germany) according to the manufacturer's instructions. PCR reaction conditions and protocols followed those described above. The hypothesis of a 1:1 distribution of *MAT1-1* and *MAT1-2* idiomorphs within the spore drops of each species was again tested with Pearson's Chi-Square and two-sided binomial tests in R.

Knoxdaviesia mating type experiments

Knoxdaviesia species rarely produce sexual reproductive structures in culture and then only when an ascospore mass is cultured directly from the host (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988). Subsequent propagation in culture is exclusively asexual. We attempted to induce the formation of ascomata in culture by reciprocally pairing two *MAT1-1* and two *MAT1-2* strains from each species in all six possible combinations on half-strength PDA. The pairings, therefore, included four *MAT1-1* vs. *MAT1-2* crossings and two negative controls (*MAT1-1* vs. *MAT1-1* and *MAT1-2* vs. *MAT1-2*). All six pairings were made in triplicate by (1) sub-culturing two strains approximately two cm apart (Wilson *et al.*, 2015), (2) streaking two conidial suspensions on opposite sides of plates (Gilgado *et al.*, 2010) and (3) combining equal volumes of two conidial suspensions and streaking the mixture. We also paired *K. capensis* and *K. proteae* to observe whether recognition and subsequent ascomatal development would take place. Plates were incubated at room temperature and in the dark to simulate the infructescence environment. They were monitored for up to four months for the presence of ascomata.

RESULTS

Identification of the *MAT1* locus

The BLASTx searches conducted on the *K. capensis* and *K. proteae* genomes highlighted regions of *K. capensis* scaffold 4 as comprising sequences with homology to *MAT1-1-1* and common *MAT1* flanking regions. BLAST searches could not detect the commonly occurring *MAT1-1-2* gene at this locus or in the rest of the genome. Scaffold 24, 30 and 37 of *K. proteae* contained ORFs with an HMG-box domain, but only the predicted protein in scaffold 24 had homology to fungal *MAT1-2-1* proteins (Table S4.1): *Ustilaginoidea virens* (KDB17701), *Metarhizium brunneum* (KID72395) and *Metarhizium majus* (KID83388). Mapping the raw sequence reads of the *K. capensis MAT1-1* genome against the *K. proteae MAT1-2* genome, and *vice versa* (Fig. S4.1) illustrated that the *MAT1-1* idiomorph (7744 bp) is slightly larger than the *MAT1-2* idiomorph (4968 bp).

The opposite *MAT* idiomorph for each *Knoxdaviesia* species was amplified in two parts using the primers designed from the conserved *MAT1* flanking regions in the *Knoxdaviesia* genomes and internal primers designed from the *MAT1* loci. These overlapping sequences

were pooled before constructing Illumina paired-end libraries for each idiomorph. The *MAT1-1* locus of *K. proteae* was amplified with KxMAT-F (5) and KcM1-R (2), and LR-KpM1F (1c) and KxMAT-R (6a), respectively yielding amplicons of approximately 3.9 kb and 6.3 kb. The *K. capensis MAT1-2* locus was amplified with KxMAT-F (5) and KpHMG-R (4), and KpHMG-F (3) and Apn-R (6b), yielding 5.8 and 2.8 kb amplicons, respectively.

The *MAT1-1* idiomorph of *K. proteae* could be assembled into a 9374 bp contig and the *MAT1-2* idiomorph of *K. capensis* into a 6908 bp contig. For *K. capensis* and *K. proteae*, respectively, 93.8 % and 95.1 % of the trimmed sequence reads mapped back to the assembly. Both of these assemblies were larger than the estimated size of each idiomorph, reflecting the upstream and downstream flanking regions that were included in the long-range PCR. The idiomorph sizes were again estimated by mapping the genome reads of the opposite mating type to each idiomorph. The *K. proteae MAT1-1* idiomorph was estimated at 7795 bp and the *K. capensis MAT1-2* idiomorph at 4988 bp.

Knoxdaviesia MAT1 genes

MAT1-1 idiomorph

WEBAUGUSTUS predicted two ORFs in each *MAT1* idiomorph. In the *K. capensis* genome, a conserved motif search identified an alpha-box (PS51325)/*MAT* alpha 1 domain (IPR006856; PFAM04769) at amino acid residues 57-196 in the first *MAT* ORF (Kc1_g2). The second *MAT* ORF (Kc1_g3) contains a low complexity region and could not be matched to any protein on the NCBI protein database. After masking the low complexity region, a BLASTp search revealed poor similarity to the *MAT1-1-2* gene of *Cordyceps militaris* L. (Link). Similar results were obtained for the two ORFs predicted from the *K. proteae MAT1-1* locus. Detailed BLAST results are given in supplementary Table S4.1.

The predicted *MAT1-1-1* genes of both *Knoxdaviesia* species contained a single intron occurring within the conserved alpha domain (Fig. 4.2A). The position of this intron in the conserved protein domain was identical to the position of the intron in the *MAT1-1-1* proteins of other Sordariomycetes (Debuchy & Turgeon, 2006), including *C. fimbriata* and *H. omanensis* (Wilken *et al.*, 2014; Wilson *et al.*, 2015). This domain was well conserved between *Knoxdaviesia* and the Ceratocystidaceae species, although *C. fimbriata* displays several deviations from amino acids that remain conserved in *Knoxdaviesia* and *Huntiella* (Fig. 4.2A).

Alignment of the *Knoxdaviesia* low complexity-region proteins to the Ceratocystidaceae (Fig. 4.2B) and other Sordariomycete *MAT1-1-2* proteins (Fig. S4.2) showed that they contain the conserved HPG (Histidine-Proline-Glycine) domain (PFAM17043) that has been proposed for this protein by Debuchy and Turgeon (2006). In the Ceratocystidaceae this motif is, however, HYG (Histidine-Tyrosine-Glycine; Wilken *et al.*, 2014). Kanematsu *et al.* (2007) proposed a conserved PPF (Proline-Proline-Phenylalanine) motif, but in *Knoxdaviesia* only the two Prolines were present, whereas only the first Proline is conserved in the Ceratocystidaceae species. The gene models predicted for these low complexity-region proteins also agree with the *MAT1-1-2* gene model of *C. fimbriata* (Fig. 4.3), providing additional evidence that they represent the *MAT1-1-2* gene in *Knoxdaviesia*. These proteins were longer (548 in *K. capensis* and 547 in *K. proteae*) than any of the other *MAT1-1-2* proteins considered here, partly owing to the low complexity region of approximately 24 amino acids at the C-terminal ends.

MAT1-2 idiomorph

The first *MAT* ORF (Kp2_g2) in the genome of *K. proteae* is an ORF smaller than the putative *Knoxdaviesia MAT1-1-1* genes, but with homology to the *MAT* alpha 1 domain. The second *K. proteae* ORF (Kp2_g3) contained a HMG-box domain (IPR009071; PFAM00505) at residues 156-230, analogous to *MAT1-2-1* proteins. As with the *K. proteae MAT1-1* idiomorph, the ORFs predicted from the assembled *K. capensis MAT1-2* idiomorph were also identified as an alpha domain-containing protein and the *MAT1-2-1* gene (Table S4.1).

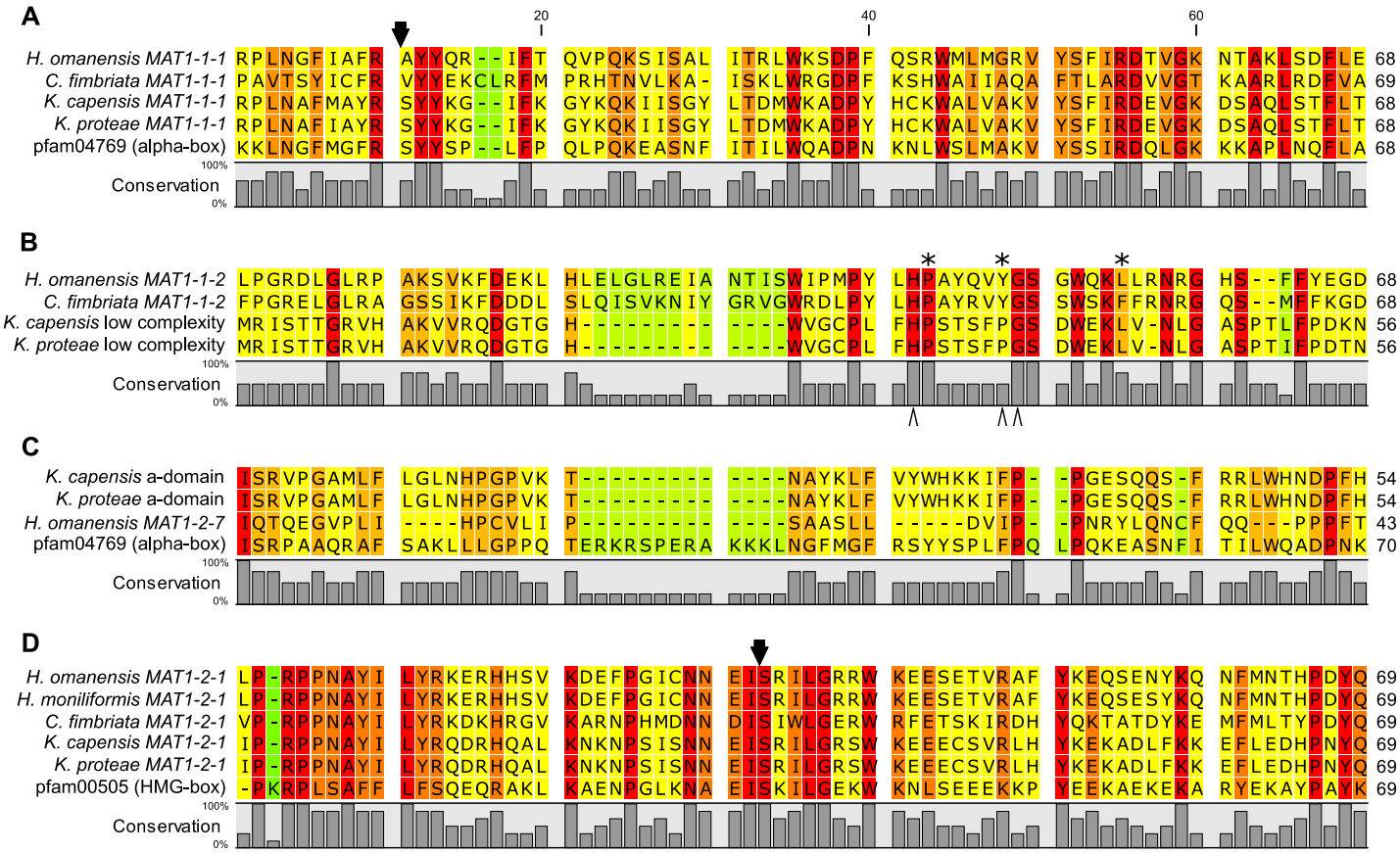

Figure 4.2 Alignment of conserved regions of the *MAT1* locus of *Knoxdaviesia*, *Huntiella* and *Ceratocystis fimbriata*. A) Alpha domain of *MAT1-1-1*, B) proposed PPF (*) and HPG (^) domains of *MAT1-1-2*, C) similarity between *Knoxdaviesia* alpha domain-containing proteins and *Huntiella MAT1-2-7* and D) HMG-box domain of *MAT1-2-1*. The conservation of residues is illustrated with a spectrum from green to red, with green representing < 30%, yellow < 65%, orange < 100% and red 100% amino acid identity. Conserved intron positions are indicated with a black arrow.

Whereas the *Knoxdaviesia MAT1-2-1* proteins showed a high level of conservation with known *MAT1-2-1* proteins (Figs. 4.2D, S4.2), including a conserved intron position (Debuchy & Turgeon, 2006), the first ORF in the *Knoxdaviesia MAT1-2* idiomorphs was unusual. These alpha domain-containing proteins were almost half the size of the *Knoxdaviesia MAT1-1-1* proteins (208 vs. 401 amino acids) and alignment indicated that the conserved domain is the only common region (Fig. S4.3). The conserved domain in these proteins also showed more deviation from the consensus *MAT* alpha 1 domain than *MAT1-1-1* proteins. Surprisingly, however, the alpha domain proteins had a predicted intron at the same position as the *MAT1-1-1* proteins (Fig. S4.3), suggesting that the two have a common origin.

Wilson *et al.* (2015) recently described a new *MAT1* gene from the *MAT1-2* idiomorph of *H. omanensis*. This gene, named *MAT1-2-7*, lies upstream of *MAT1-2-1* and does not have a known conserved domain, neither does it contain an intron. Alignment between the *Knoxdaviesia* alpha domain-containing proteins, the *H. omanensis MAT1-2-7* protein and the consensus *MAT* alpha 1 domain (Fig. 4.2C) showed that the *Knoxdaviesia* and *Huntiella* proteins are similar. This suggests that the *Huntiella MAT1-2-7* protein may also have arisen from an alpha domain-containing protein. We thus propose that these genes are homologous and therefore consider the *Knoxdaviesia MAT1-2* alpha domain-containing genes as *MAT1-2-7*.

Genes flanking the *MAT* locus

BLAST searches identified the ORFs flanking the *MAT1* locus in both *Knoxdaviesia* species as the *SLA2* and *APN2* genes (Fig 4.1; Table S4.1). Other genes are commonly associated with the *MAT* locus (*e.g.* the cytochrome C oxidase VIa subunit (*COX6a*), anaphase promoting complex subunit 5 (*APC5*) and coenzymeQ biosynthesis (*Coq4*) genes) and were also detected downstream of *APN2*. Unlike the Ceratocystidaceae, the *Knoxdaviesia MAT1* locus conformed to the ancestral Sordariomycete arrangement of flanking genes (Fig 4.3; Debuchy & Turgeon, 2006).

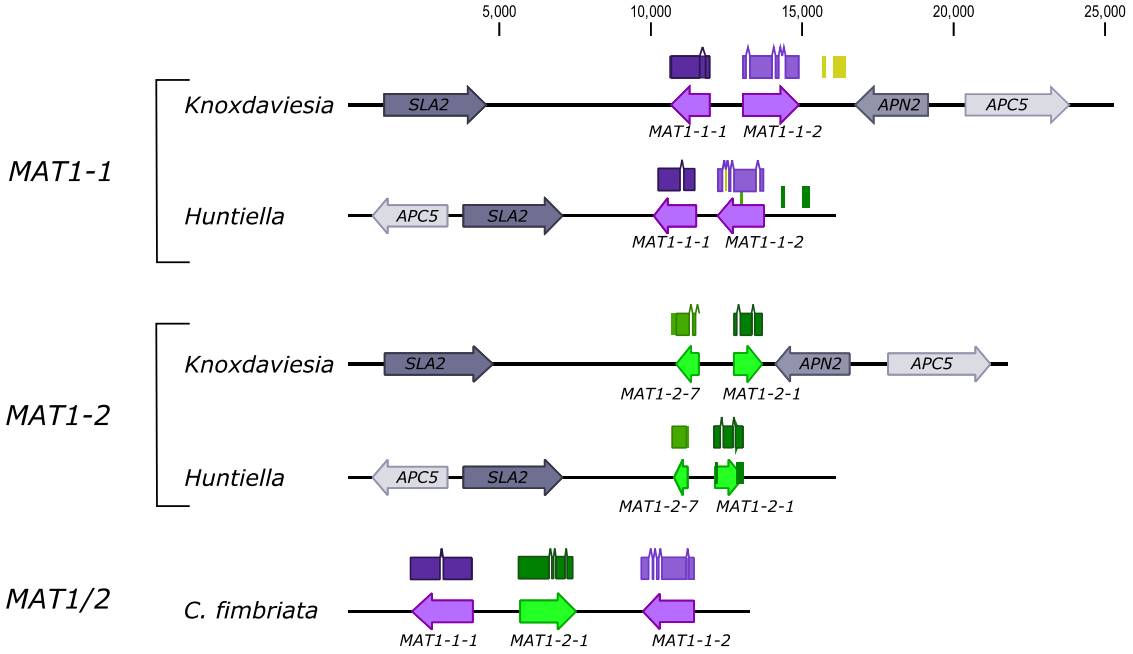

Figure 4.3 Comparison between the *MAT1* locus organisation of *Knoxdaviesia*, *Huntiella* and *Ceratocystis fimbriata*. Gene models are indicated above the *MAT1* genes as boxes (exons) and gaps (introns). *APN2* is not present in the regions flanking *MAT1* of *Huntiella*.

Knoxdaviesia capensis and *K. proteae MAT1* comparison

Dotplot comparisons between the opposite *MAT* idiomorphs in each of the species indicated that the two idiomorphs share little sequence similarity, with the only exception being two similar 160 bp fragments in the *K. proteae* idiomorphs (Fig. 4.4A, B). The size difference between the *MAT1-1* and *MAT1-2* idiomorphs can be ascribed to smaller ORFs in the *MAT1-2* idiomorph and a smaller intergenic region (386 bp in *MAT1-1* vs. >1.8 kb in *MAT1-2*) separating the *MAT1-2-1* and *APN2* genes. In contrast, the homologous *MAT1-1* and *MAT1-2* idiomorphs shared 84.3 % and 86.4 % sequence identity at the nucleotide level. The dotplot indicated two small inversions between each of the idiomorphs in *K. capensis* and *K. proteae* and illustrated the low complexity region of the *MAT1-2-7* gene (Fig. 4.4C, D). Significant repeats, as are present in the unswitched *C. fimbriata MAT1* locus (Wilken *et al.*, 2014), could not be identified in the *Knoxdaviesia* species. The *MAT1* idiomorphs of both *Knoxdaviesia* species have been deposited in GenBank: *K. capensis MAT1-1* (KX832965) and *MAT1-2* (KX832968); *K. proteae MAT1-1* (KX832967) and *MAT1-2* (KX832966).

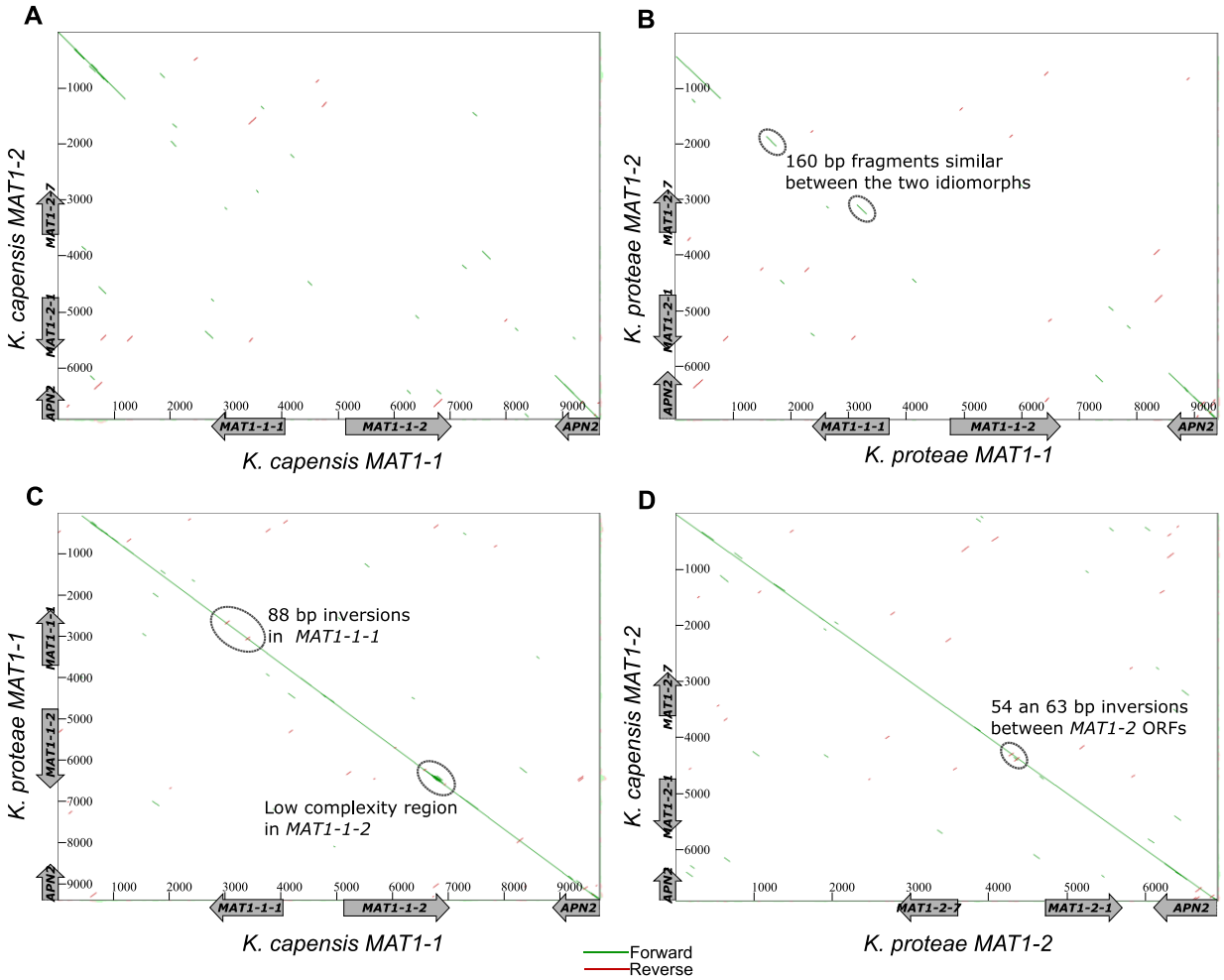

Figure 4.4 Pairwise dotplot comparisons between the *MAT1* idiomorphs of *Knoxdaviesia capensis* and *K. proteae*. Intra-specific comparisons between the A) *K. capensis* and B) *K. proteae* idiomorphs, as well as inter-specific comparisons between the homologous C) *MAT1-1* and D) *MAT1-2* idiomorphs are shown. The positions of the *MAT1-1*, *MAT1-2* and *APN2* genes are indicated on each plot.

Distribution of *MAT* idiomorphs in natural populations and ascomata

The Chi-Square and two-tailed binomial tests were conducted with the null hypothesis of a 1:1 ratio of mating types. The probability of obtaining either one of the idiomorphs was therefore set as 0.5. The P-values and 95% confidence intervals of the tests (Table S4.2) indicated that almost all populations had a ratio not significantly different from 1:1 (Fig. 4.5) Confidence intervals were smaller in populations with large (≥ 30) sample sizes. The only population that deviated from the 1:1 ratio was a small population (Franshoek; n = 10) of *K. capensis* sampled from *P. repens*, which is a non-preferred host of this fungus (Roets *et al.*, 2011a). This deviation also affected the overall distribution of *K. capensis* on this host (Fig. 4.5). The *P. longifolia* population appears to be skewed towards *MAT1-2* individuals, although this was not statistically significant. If we consider testing a single hypothesis multiple times as grounds for applying a multiple testing correction procedure, the new critical value, as defined by Bonferroni (Noble, 2009), would be P = 0.003. According to the binomial test, the Franschoek *K. capensis* population would not be significantly different from zero at this critical value, although it remains significant based on the Chi-Square test.

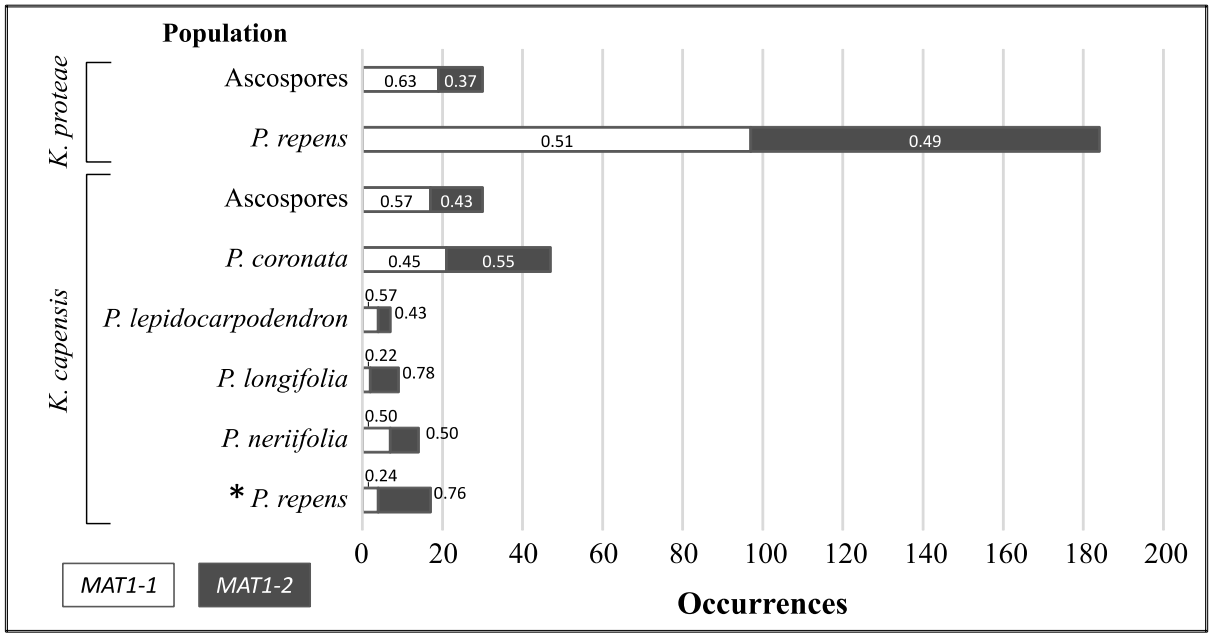

Figure 4.5 Distribution of the *MAT1* idiomorphs in ascospore droplets and *Protea* populations of *Knoxdaviesia capensis* and *K. proteae*. White bars represent the number of occurrences of *MAT1-1* and grey the occurrences of *MAT1-2*. The frequency of each idiomorph is indicated on the bars. The population that showed a significant deviation (P = < 0.003) from a 1:1 distribution is indicated with an asterisk.

Knoxdaviesia mating experiments

Only the pairing methods that involved conidial suspensions of opposite mating types showed any sign of a mating reaction. Strains sub-cultured close together did not form an interaction zone or give rise to ascomata. Ascomatal development was observed in spore mixture pairings of *K. proteae* (Fig. 4.6A), although ascomata were sparse and surrounded by prolific vegetative growth. Ascomata were observed in one culture of *K. capensis* grown from a droplet containing a mass of ascospores, but crosses between isolates representing opposite idiomorphs resulted only in large masses of what could have been an aggregation of developing ascomata (Fig. 4.6B, C). The *K. capensis* "aggregations" produced spore droplets containing conidia, whereas the *K. proteae* ascomata did not exude spore droplets.

Although attempts to hybridise *K. capensis* and *K. proteae* were unsuccessful, black aggregations, similar to those observed in the developing ascomata of *K. proteae*, were apparent in these interspecific crosses (Fig. 4.6D). It is, therefore, possible that some recognition occurs between the two species, although further ascomatal development is impeded.

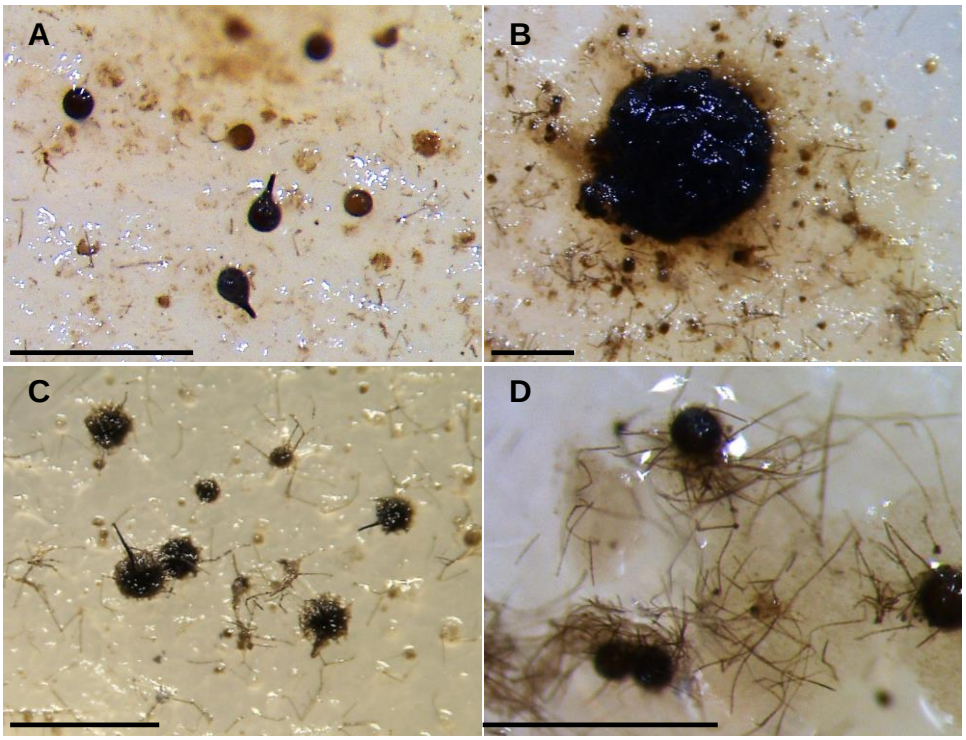

Figure 4.6 Mating reactions between opposite idiomorphs of *Knoxdaviesia capensis* and *K. proteae* in culture. A) Ascomata and developing ascomata of *K. proteae* (CMW408879 x CMW40883). B) Aggregations of *K. capensis* (CMW40886 x CBS139037) likely representing developing ascomata. C) Ascomata of *K. capensis* in a culture grown from an

ascospore mass. D) Aggregations formed during an inter-specific cross between *K. capensis* and *K. proteae*. Scale bars = 0.5 mm.

DISCUSSION

Knoxdaviesia: outcrossing via heterothallism

Only one *MAT1* idiomorph was identified in each of the two *Knoxdaviesia* genomes examined, but the opposite idiomorphs could be amplified in different strains of the species. *MAT1-1* and *MAT1-2* individuals occur at similar proportions in the *Knoxdaviesia* populations from which high genetic diversity was previously detected (Aylward *et al.*, 2014; 2015a; 2017), supporting the hypothesis that heterothallism underlies this diversity. The single *K. capensis* population that deviated from the 1:1 ratio was collected from *P. repens*, a non-preferred host (Roets *et al.*, 2011a) on which it occurs infrequently and likely only survives until *K. proteae* is introduced (Aylward *et al.*, 2015b). Furthermore, approximately half of the progeny of haploid crosses contained the *MAT1-1* and the other half the *MAT1-2* idiomorph.

Although the production of ascomata could be induced only poorly in some intra-specific crosses, single-spore isolates of *Knoxdaviesia* species do not form ascomata and a mating reaction could only be induced between strains containing opposite *MAT1* idiomorphs. The overall results of this study, therefore, show that both *K. capensis* and *K. proteae* have a heterothallic mating system, thereby requiring outcrossing between two individuals of opposite mating type for sexual reproduction.

Knoxdaviesia species are vectored by arthropods (Roets *et al.*, 2011b), and possibly passerine birds, and may spend a considerable amount of time on these animals before being deposited in a suitable habitat. Since most fungi are capable of switching between sexual and asexual reproduction in response to environmental conditions (Billiard *et al.*, 2012), the prevalence of sexual reproduction in *Protea*-associated *Knoxdaviesia* suggests that that this form of reproduction is beneficial in the *Protea* environment. Various trade-offs exist for sexual reproduction (Billiard *et al.*, 2012). Disadvantages include the energy cost of developing sexual structures, the risk of failing to find a suitable mate (in the case of non-selfing species) and the possibility that locally adapted allele combinations may be broken through recombination. In *Knoxdaviesia*, the cost of outcrossing is likely offset by, amongst others, preventing the accumulation of harmful alleles and enabling diversification through

recombination, a trait especially important for adaptation. Other than the creation of genetic diversity, production of dispersal propagules may be an essential factor underlying the prevalence of sexual reproduction in *Protea* infructescences. This is because ascospores are usually more resilient, and therefore suitable for dispersal, than their conidial counterparts (Aanen & Hoekstra, 2007).

Evolutionary conservation between *Knoxdaviesia MAT1* loci

The homologous *MAT1* idiomorphs of *K. capensis* and *K. proteae* were highly similar and their mating type proteins almost identical. Our interspecific mating type experiments also suggested that some determinants of recognition in strains of opposite mating type remain conserved between *K. capensis* and *K. proteae*. In contrast, the Sordariomycete *MAT1-1-1* and *MAT1-2-1* proteins used as a basis for comparison had low amino acid identities to each other and as well as to their *Knoxdaviesia* homologs (between 7.9 % and 38.5 %). Studies on the mating type loci of fungi have revealed great diversity in size and gene content. Despite their conserved domains, homologous *MAT1* genes in different species vary greatly and are believed to evolve more rapidly than other genes in the genome (Turgeon, 1998; Wik *et al.*, 2008). The high level of similarity between the *Protea*-associated *Knoxdaviesia MAT1* loci is, therefore, likely a reflection of the close phylogenetic relationship between these species (Roets *et al.*, 2009).

Despite the recognition that may occur between *K. capensis* and *K. proteae* and the sequence similarity in their *MAT1* regions, these species appear to be reproductively isolated. The initial prevention of gene flow in the common ancestor was likely facilitated by host association. *Knoxdaviesia proteae* occurs exclusively in *P. repens* (Wingfield *et al.*, 1988), whereas *K. capensis* is found in the infructescences of nine different *Protea* host species (Aylward *et al.*, 2015b; Roets *et al.*, 2009; Wingfield & Van Wyk, 1993). The occurrence of *K. capensis* in *P. repens* infructescences was discovered only recently, this being due to the low level at which it occurs on this host (Aylward *et al.*, 2015b). Therefore, despite the clear disparity in host range between *K. capensis* and *K. proteae*, host association does not completely isolate the two species, although the historical distribution of hosts may have facilitated isolation. We speculate that the *Protea*-associated *Knoxdaviesia* ancestor occupied a range of host species, but became isolated on *P. repens*. *Protea repens*, or at least one population of this host, may have been geographically separated from the other *Protea* hosts

historically. Fire, a common occurrence in Fynbos vegetation in which *Protea* species occur (Cowling, 1992), may clear vast areas, possibly providing a mechanism for isolating host populations. Should the *Knoxdaviesia* ancestor have been isolated in *P. repens* by such means, sufficient time without gene flow would have resulted in genetic drift between the *Knoxdaviesia* populations on *P. repens* and the *Knoxdaviesia* populations on other hosts such that speciation has occurred (Slatkin, 1985; 1987).

Knoxdaviesia MAT1 characteristics

The most obvious difference between the *MAT1* loci of the Gondwanamycetaceae and Ceratocystidaceae species is the order of genes flanking the *MAT1* locus. Although we expected that the rearranged order in the Ceratocystidaceae might be a phenomenon common to the Microascales, the *Knoxdaviesia* flanking genes follow the ancestral Sordariomycete arrangement where *SLA2* and *APN2* flank the *MAT1* locus (Debuchy & Turgeon, 2006). Since *Knoxdaviesia* represents the sister family of *Ceratocystis* and *Huntiella* in the Microascales (Réblová *et al.*, 2011), we suspect that this specific rearrangement is limited to the Ceratocystidaceae and probably occurred in an early ancestor of this family.

Although the *MAT1* genes in the two *Knoxdaviesia* species are virtually identical, the diversity of *MAT1* genes was exemplified in the comparison with the three Ceratocystidaceae species. Differences in protein length and amino acid sequence were apparent, yet regions associated with protein function, such as the conserved domains and intron positions, remain uniform. The accessory genes (*MAT1-1-2* and *MAT1-2-7*) in the *Knoxdaviesia* idiomorphs appeared particularly divergent in comparison to known *MAT1* genes, necessitating future expression analysis to determine whether they are transcribed. These could not be identified with BLAST searches, although alignments revealed similarity to Ceratocystidaceae accessory *MAT1* proteins. Until recently, *MAT1-1-2* genes lacked a universally conserved PFAM domain and the low complexity region in the *Knoxdaviesia MAT1-1-2* gene also hampered identification. An HPYG domain is conserved in the studied Ceratocystidaceae *MAT1-1-2* proteins (Wilken *et al.*, 2014), but in *Knoxdaviesia* the Tyrosine (Y) residue is replaced by a Proline. As with the flanking gene arrangement, the *MAT1* genes of *Knoxdaviesia* thus conform to the general pattern of the Sordariomycetes rather than to the derived state of the Ceratocystidaceae.

MAT1-2-7 gene descended from *MAT1-1-1*

To the best of our knowledge, the putative *Knoxdaviesia MAT1-2-7* genes represent the first report of Ascomycete *MAT1-2* idiomorph proteins containing an alpha domain. The presence of a truncated *MAT1-1-1* gene at the same position on the *MAT1-2* idiomorph is a recognised pattern in the Ophiostomatales (Comeau *et al.*, 2015; Duong *et al.*, 2013; Tsui *et al.*, 2013), but this truncated *MAT1-1-1* gene lacks the alpha domain. Other than the *MAT1-1-1* gene, *Ophiostoma quercus* (Georgévitch) Nannf. is also known to harbour fragments of the *MAT1-1-3* gene in its *MAT1-2* idiomorph (Wilken *et al.*, 2012) and, conversely, *Aspergillus fumigatus* Fresenius has part of the *MAT1-2-1* gene in its *MAT1-1* idiomorph (Paoletti *et al.*, 2005). In all of these cases, the various fragments are thought to have originated from one or more unequal recombination/crossover events between the two *MAT1* idiomorphs in a common heterothallic ancestor (Gioti *et al.*, 2012). However the truncated *MAT1-1-1* gene of the Ophiostomatales may be functional rather than a mere remnant of ancestral recombination. Tsui *et al.* (2013) showed that this gene is expressed in *Grosmannia clavigera* (Rob.-Jeffr. & R.W. Davidson) Zipfel, Z.W. de Beer & M.J. Wingf. and suggested that the alpha domain had been lost after the crossover event.

Although the alpha domain is not readily detectable in the *MAT1-2-7* gene of *H. omanensis*, it is nevertheless similar to its *Knoxdaviesia* counterpart. The intron in *Knoxdaviesia MAT1-2-7*, which lies at the identical position as the *MAT1-1-1* intron, provides compelling evidence for an evolutionary link between the two genes. *MAT* alpha 1 domains have a DNA-binding/bending function (Gene Ontology: 0008301) and act as transcriptional activators for pheromone and receptor genes involved in the mating process (Bender & Sprague, 1987; Casselton, 2002). If the *MAT1-2-7* genes of *Knoxdaviesia* and *Huntiella* had arisen from the *MAT1-1-1* gene via recombination and, therefore, contained a MAT alpha 1 domain, the constraint to conserve this domain would be negated by the presence of a functional *MAT1-1-1* protein during the mating reaction. Conversely, this domain may have mutated to prevent self-fertility, reflecting the essential benefits that outcrossing confers in these species. Expression analysis is, however, needed to verify whether the *Huntiella* and *Knoxdaviesia MAT1-2-7* genes are functional.

CONCLUSIONS

This study is the first to investigate fungal reproduction in the Core Cape Subregion and it revealed how the *Protea*-associated *Knoxdaviesia* species maintain high levels of genetic diversity on their native hosts. Additionally, it provides information on the extent to which the *MAT1* locus has diversified in the Ceratocystidaceae, a family that gave rise to many economically important pathogens. The *Knoxdaviesia* species considered here were shown to have a classic heterothallic *MAT1* locus with strong resemblance to the ancestral gene organisation and conserved domains of Sordariomycete fungi. As such, the *Knoxdaviesia MAT1* locus is probably a more accurate representation of the ancestral *MAT1* locus of Microascalean fungi than the *MAT1* loci of Ceratocystidaceae species.

ACKNOWLEDGEMENTS

We thank the National Research Foundation (NRF) and the Department of Science and Technology (DST)-NRF Centre of Excellence in Tree Health Biotechnology (CTHB) for financial assistance. We are also grateful to two anonymous reviewers for their comments that improved the clarity of this manuscript.

REFERENCES

Aanen, D. & Hoekstra, R. (2007). Why sex is good: on fungi and beyond. In *Sex in fungi: Molecular determination and evolutionary implications*, pp. 527-534. Edited by J. Heitman, J. W. Kronstad, J. W. Taylor & L. A. Casselton: ASM Press.

Alby, K. & Bennett, R. J. (2011). Interspecies pheromone signaling promotes biofilm formation and same-sex mating in *Candida albicans*. *Proceedings of the National Academy of Sciences* **108**, 2510-2515.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2014). Panmixia defines the genetic diversity of a unique arthropod-dispersed fungus specific to *Protea* flowers. *Ecology and Evolution* **4**, 3444-3455.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015a). Long-distance dispersal and recolonization of a fire-destroyed niche by a mite-associated fungus. *Fungal Biology* **119**, 245-256.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015b). *Knoxdaviesia proteae* is not the only *Knoxdaviesia*-symbiont of *Protea repens*. *IMA Fungus* **6**, 471-476.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D. & Wingfield, M. J. (2016). Genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa. *Standards in Genomic Sciences* **11**, 1-7.

Aylward, J., Dreyer, L. L., Laas, T., Smit, L., & Roets, F. (2017). *Knoxdaviesia capensis*: dispersal ecology and population genetics of a flower-associated fungus. *Fungal Ecology* **26**, 28-36.

Bender, A. & Sprague, G. F. (1987). MATα1 protein, a yeast transcription activator, binds synergistically with a second protein to a set of cell-type-specific genes. *Cell* **50**, 681-691.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research* **41**, D36-D42.

Bergh, N. G., Verboom, G., Rouget, M. & Cowling, R. M. (2014). Vegetation types of the Greater Cape Floristic Region. In *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region*, pp. 26-46. Edited by N. Allsopp, J. F. Colville & G. A. Verboom. New York: Oxford University Press

Betley, J. N., Frith, M. C., Graber, J. H., Choo, S. & Deshler, J. O. (2002). A ubiquitous and conserved signal for RNA localization in chordates. *Current Biology* **12**, 1756-1761.

Billiard, S., López-Villavicencio, M., Hood, M. & Giraud, T. (2012). Sex, outcrossing and mating types: unsolved questions in fungi and beyond. *Journal of Evolutionary Biology* **25**, 1020-1038.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

Boratyn, G. M., Schaffer, A., Agarwala, R., Altschul, S. F., Lipman, D. J. & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct* **7**, 12.

Casselton, L. (2002). Mate recognition in fungi. *Heredity* **88**, 142-147.

Comeau, A. M., Dufour, J., Bouvet, G. F., Jacobi, V., Nigg, M., Henrissat, B., Laroche, J., Levesque, R. C. & Bernier, L. (2015). Functional annotation of the *Ophiostoma novo-ulmi* genome: insights into the phytopathogenicity of the fungal agent of Dutch elm disease. *Genome Biology and Evolution* **7**, 410-430.

Coppin, E., Debuchy, R., Arnaise, S. & Picard, M. (1997). Mating types and sexual development in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews* **61**, 411-428.

Cowling, R. M. (1992). The ecology of fynbos: Nutrients, fire and diversity. Cape Town: Oxford University Press.

De Beer, Z., Duong, T., Barnes, I., Wingfield, B. & Wingfield, M. (2014). Redefining *Ceratocystis* and allied genera. *Studies in Mycology* **79**, 187-219.

Debuchy, R., Arnaise, S. & Lecellier, G. (1993). The *mat*- allele of *Podospora anserina* contains three regulatory genes required for the development of fertilized female organs. *Molecular and General Genetics* **241**, 667-673.

Debuchy, R. & Turgeon, B. (2006). Mating-type structure, evolution, and function in Euascomycetes. In *The Mycota I: Growth, Differentiation and Sexuality*, pp. 293-323. Edited by U. Kües & R. Fischer: Springer.

Duong, T. A., De Beer, Z. W., Wingfield, B. D. & Wingfield, M. J. (2013). Characterization of the mating-type genes in *Leptographium procerum* and *Leptographium profanum*. *Fungal Biology* **117**, 411-421.

Fincham, J. R. S. & Day, P. R. (1963). *Fungal Genetics*. Oxford: Blackwell Scientific Publications.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., *et al.* (2014). Pfam: the protein families database. *Nucleic Acids Research* **42**, D222-D230.

Gilgado, F., Gené, J., Cano, J. & Guarro, J. (2010). Heterothallism in *Scedosporium apiospermum* and description of its teleomorph *Pseudallescheria apiosperma* sp. nov. *Medical Mycology* **48**, 122-128.

Gioti, A., Mushegian, A. A., Strandberg, R., Stajich, J. E. & Johannesson, H. (2012). Unidirectional evolutionary transitions in fungal mating systems and the role of transposable elements. *Molecular Biology and Evolution* **29**, 3215-3226.

Glass, N. L., Grotelueschen, J. & Metzenberg, R. L. (1990). *Neurospora crassa* A mating-type region. *Proceedings of the National Academy of Sciences* **87**, 4912-4916.

Glass, N. L. & Smith, M. L. (1994). Structure and function of a mating-type gene from the homothallic species *Neurospora africana*. *Molecular and General Genetics* **244**, 401-409.

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

Heitman, J. (2015). Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biology Reviews* **29**, 108-117.

Herskowitz, I. (1989). A regulatory hierarchy for cell specialization in yeast. *Nature* **342**, 749-757.

Hoff, K. J. & Stanke, M. (2013). WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**, W123-W128.

Kanematsu, S., Adachi, Y. & Ito, T. (2007). Mating-type loci of heterothallic *Diaporthe* spp.: homologous genes are present in opposite mating-types. *Current Genetics* **52**, 11-22.

Kronstad, J. W. & Staben, C. (1997). Mating type in filamentous fungi. *Annual Review of Genetics* **31**, 245-276.

Lehtonen, J., Jennions, M. D. & Kokko, H. (2012). The many costs of sex. *Trends in Ecology & Evolution* **27**, 172-178.

Lin, X., Hull, C. M. & Heitman, J. (2005). Sexual reproduction between partners of the same mating type in *Cryptococcus neoformans*. *Nature* **434**, 1017-1021.

Lynch, M., Bürger, R., Butcher, D. & Gabriel, W. (1993). The mutational meltdown in asexual populations. *Journal of Heredity* **84**, 339-344.

Maharachchikumbura, S. S. N., Hyde, K. D., Jones, E. B. G., McKenzie, E. H. C., Huang, S.-K., Abdel-Wahab, M. A., Daranagama, D. A., Dayarathne, M., D'souza, M. J., Goonasekara, I. D., *et al.* (2015). Towards a natural classification and backbone tree for *Sordariomycetes*. *Fungal Diversity* **72**, 199-301.

Manning, J. & Goldblatt, P. (2012). Plants of the Greater Cape Floristic Region. 1: The Core Cape Flora. In *Strelitzia 29*. Pretoria: South African National Biodiversity Institute.

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., *et al.* (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research* **43**, D222-D226.

Milgroom, M. G. (1996). Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* **34**, 457-477.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., *et al.* (2015). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research* **43**, D213-D221.

Moore, D. & Novak Frazer, L. (2002). *Essential Fungal Genetics*. New York: Springer-Verlag.

Nelson, M. A. (1996). Mating systems in ascomycetes: a romp in the sac. *Trends in Genetics* **12**, 69-74.

Ni, M., Gao, N., Kwon, N.-J., Shin, K.-S. & Yu, J.-H. (2010). Regulation of *Aspergillus* conidiation. In *Cellular and Molecular Biology of Filamentous Fungi*, pp. 559-576. Edited by K. Borkovich & D. Ebbole. Washington DC: ASM Press.

Ni, M., Feretzaki, M., Sun, S., Wang, X. & Heitman, J. (2011). Sex in fungi. *Annual Review of Genetics* **45**, 405.

Nielsen, K. & Heitman, J. (2007). Sex and Virulence of Human Pathogenic Fungi. In *Fungal Genomics*, pp. 143-173. Edited by Dunlap, J. C. Amsterdam: Academic Press.

Noble, W. S. (2009). How does multiple testing correction work? *Nature Biotechnology* **27**, 1135-1137.

Noé, L. & Kucherov, G. (2005). YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Research* **33**, W540-W543.

Otto, S. P. (2009). The evolutionary enigma of sex. *The American Naturalist* **174**, S1-S14.

Paoletti, M., Rydholm, C., Schwier, E. U., Anderson, M. J., Szakacs, G., Lutzoni, F., Debeaupuis, J.-P., Latgé, J.-P., Denning, D. W. & Dyer, P. S. (2005). Evidence for Sexuality in the Opportunistic Fungal Pathogen *Aspergillus fumigatus*. *Current Biology* **15**, 1242-1248.

Perkins, D. D. (1987). Mating-type switching in filamentous ascomycetes. *Genetics* **115**, 215.

R Core Team (2014).R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Réblová, M., Gams, W. & Seifert, K. A. (2011). *Monilochaetes* and allied genera of the *Glomerellales*, and a reconsideration of families in the *Microascales*. *Studies in Mycology* **68**, 163-191.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2009). Fungal radiation in the Cape Floristic Region: An analysis based on *Gondwanamyces* and *Ophiostoma*. *Molecular Phylogenetics and Evolution* **51**, 111-119.

Roets, F., Theron, N., Wingfield, M. J. & Dreyer, L. L. (2011a). Biotic and abiotic constraints that facilitate host exclusivity of *Gondwanamyces* and *Ophiostoma* on *Protea*. *Fungal Biology* **116**, 49-61.

Roets, F., Wingfield, M. J., Wingfield, B. D. & Dreyer, L. L. (2011b). Mites are the most common vectors of the fungus *Gondwanamyces proteae* in *Protea* infructescences. *Fungal Biology* **115**, 343-350.

Seymour, F. A., Crittenden, P. D. & Dyer, P. S. (2005). Sex in the extremes: lichen-forming fungi. *Mycologist* **19**, 51-58.

Slatkin, M. (1985). Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics* **16**, 393-430.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science* **236**, 787-792.

Staben, C. & Yanofsky, C. (1990). *Neurospora crassa* a mating-type region. *Proceedings of the National Academy of Sciences* **87**, 4917-4921.

Stearns, S. C. (1987). The evolution of sex and its consequences. Basel: Birkhäuser.

Tsui, C. K.-M., DiGuistini, S., Wang, Y., Feau, N., Dhillon, B., Bohlmann, J. & Hamelin, R. C. (2013). Unequal recombination and evolution of the Mating-Type (*MAT*) loci in the pathogenic fungus *Grosmannia clavigera* and relatives. *G3: Genes|Genomes|Genetics* **3**, 465-480.

Turgeon, B. G. (1998). Application of mating type gene technology to problems in fungal biology. *Annual Review of Phytopathology* **36**, 115-137.

Turgeon, B. G. & Yoder, O. C. (2000). Proposed Nomenclature for Mating Type Genes of Filamentous Ascomycetes. *Fungal Genetics and Biology* **31**, 1-5.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35**, W71-W74.

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

Wik, L., Karlsson, M. & Johannesson, H. (2008). The evolutionary trajectory of the mating-type (mat) genes in *Neurospora* relates to reproductive behavior of taxa. *BMC Evolutionary Biology* **8**, 1.

Wilken, P. M., Steenkamp, E. T., Hall, T. A., De Beer, Z. W., Wingfield, M. J. & Wingfield, B. D. (2012). Both mating types in the heterothallic fungus *Ophiostoma quercus* contain *MAT1-1* and *MAT1-2* genes. *Fungal Biology* **116**, 427-437.

Wilken, P. M., Steenkamp, E. T., Wingfield, M. J., de Beer, Z. W. & Wingfield, B. D. (2014). DNA Loss at the *Ceratocystis fimbriata* Mating Locus Results in Self-Sterility. *PLOS ONE* **9**, e92180.

Wilson, A., Godlonton, T., Van der Nest, M. A., Wilken, P. M., Wingfield, M. J. & Wingfield, B. D. (2015). Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genetics and Biology* **80**, 1-9.

Wingfield, M. J., Wyk, P. S. V. & Marasas, W. F. O. (1988). *Ceratocystiopsis proteae* sp. *nov*., with a new anamorph genus. *Mycologia* **80**, 23-30.

Wingfield, M. J. & Van Wyk, P. S. (1993). A new species of *Ophiostoma* from *Protea* infructescences in South Africa. *Mycological Research* **97**, 709-716.

SUPPLEMENTARY INFORMATION

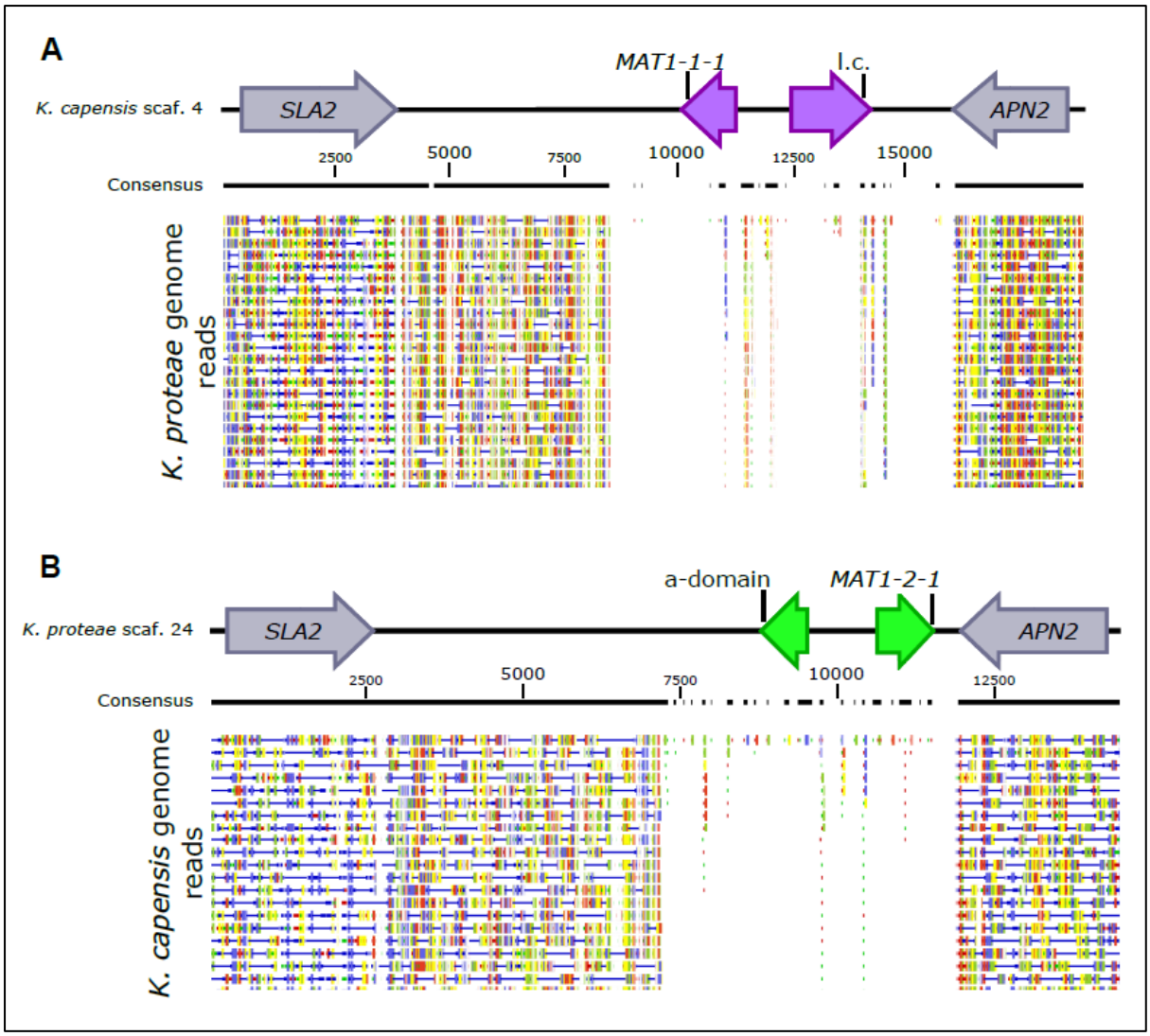

Figure S4.1. Estimation of *MAT1* idiomorph sizes by read mapping. A) The *MAT1-2* genome reads of *K. proteae* were mapped to the *MAT1-1* idiomorph of *K. capensis* and B) the *MAT1-1* genome reads of *K. capensis* were mapped to the *MAT1-2* genome of *K. proteae*. The consensus sequence represents the consensus according to the read mappings. l.c. = low complexity region protein, a-domain = alpha domain-containing protein.

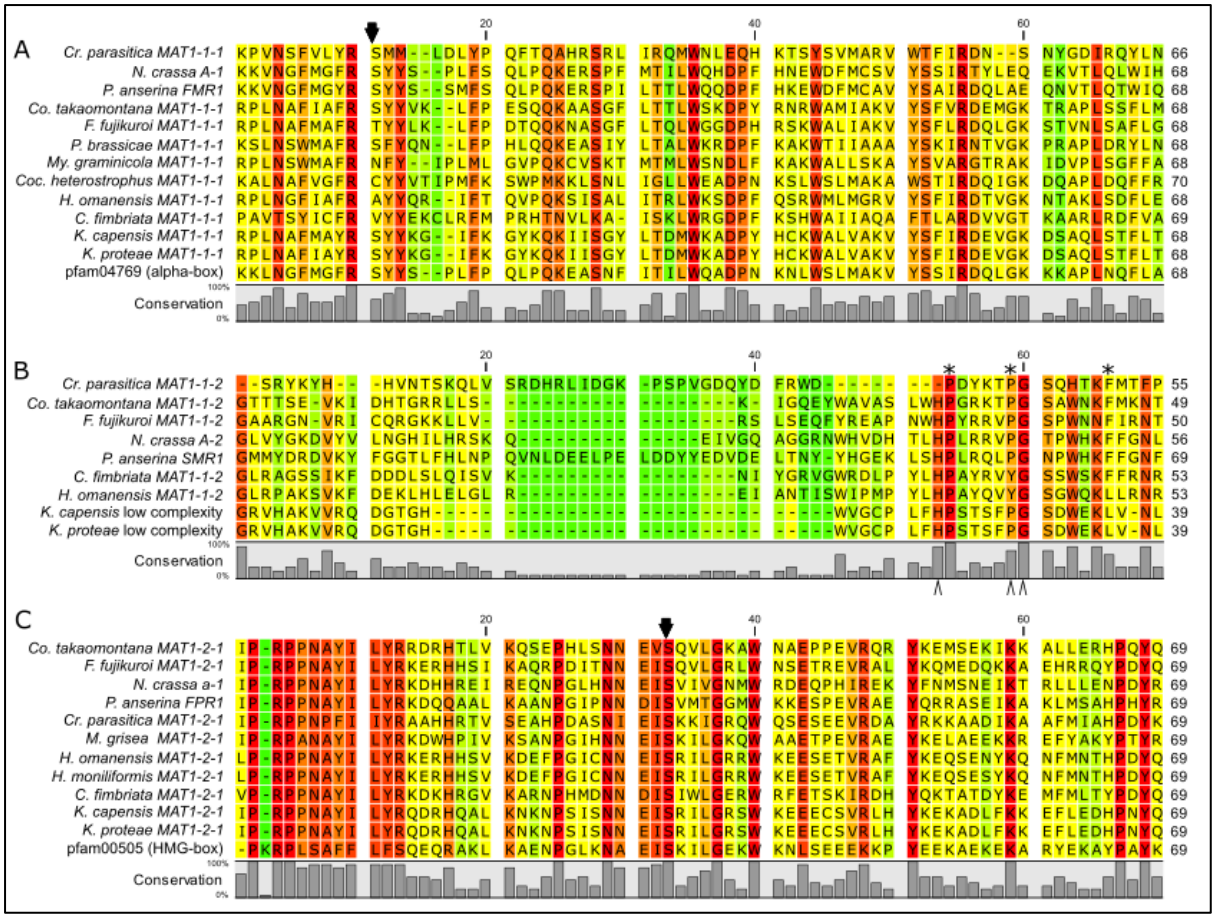

Figure S4.2. Alignment of *Knoxdaviesia MAT1* proteins to those of other Sordariomycetes. A) Alpha domain of *MAT1-1-1*, B) proposed PPF (*) and HPG (^) domains of *MAT1-1-2* and C) HMG-box domain of *MAT1-2-1*. As for Fig. 4.2, the conservation of residues is illustrated with a spectrum from green to red, with green representing < 30%, yellow < 65%, orange < 100% and red 100% amino acid identity. Conserved intron positions are indicated with black arrows. GenBank accession numbers are given in the text.

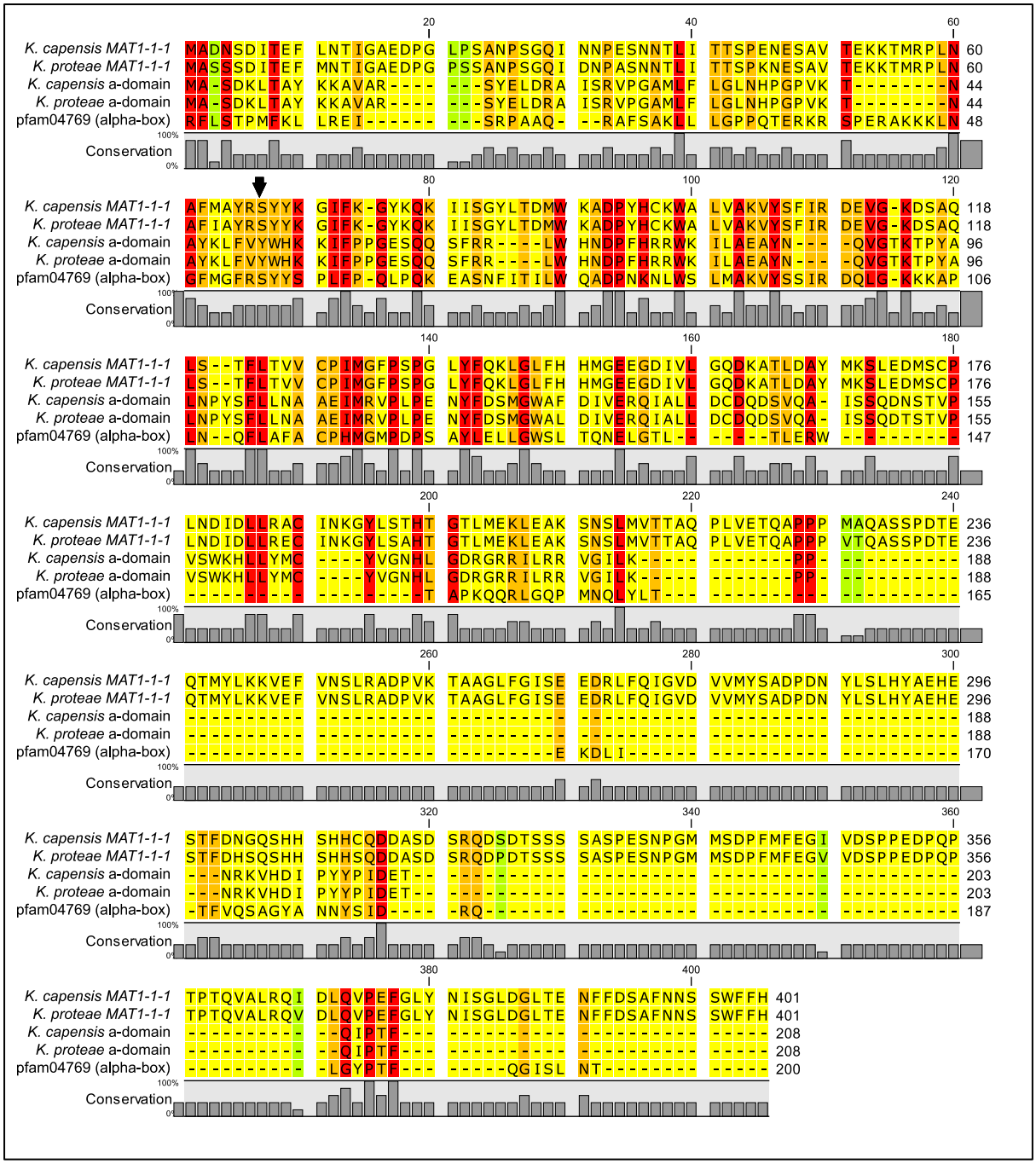

Figure S4.3. Alignment between *Knoxdaviesia MAT1-1-1* and *MAT1-2-7* (alpha domain-containing) proteins. The consensus MAT alpha 1 sequence (PFAM04769) is included as a reference. Green = 0%, yellow = 40%, orange = 60% and red ≥ 80% amino acid conservation. The conserved intron position is indicated with a black arrow.

Table S4.1 and **Table S4.2** are provided in the electronic supplementary materials.

Table S4.1 Identification of the predicted open reading frames (ORFs) in and around the *Knoxdaviesia capensis* and *K. proteae MAT1* loci.

Table S4.2 Distribution of mating types in *Knoxdaviesia* ascospore droplets and natural populations.

CHAPTER 5

Contrasting carbon metabolism in saprotrophic and pathogenic Microascalean fungi from *Protea* trees

ABSTRACT

Knoxdaviesia (Gondwanamycetaceae) species associated with *Protea* trees occupy a specialist niche, growing on decaying flowers in seed cones. This saprotrophic association is in sharp contrast to the pathogenicity often encountered in their sister family, the Ceratocystidaceae. The African indigenous *Ceratocystis albifundus*, for example, is the causal agent of wattle wilt, but has been isolated from diseased *Protea* species in South Africa. Other than being phylogenetically related, both *Knoxdaviesia* and *Ceratocystis* are vectored by arthropods, yet the two occupy distinctly different niches. The aim of this study was to compare substrate utilisation in two saprotrophic *Knoxdaviesia* species, one generalist and one specialist, and the pathogen *C. albifundus* by integrating phenome and whole-genome data. The generalist *K. capensis* grew on the widest range of carbon sources, alluding to how it has maintained its large *Protea* host range. *Knoxdaviesia proteae* used fewer substrates, suggesting that it lost some metabolic capacity due to its specialisation on a single host. Monosaccharides in *Protea* nectar appear to be important carbon sources for early colonising *Knoxdaviesia* species in flower heads. Once the flower head ages, these species would be able to switch to degrading the cellulose and hemicellulose components of cell walls. Contrary to expectations, the pathogen *C. albifundus* grew on few carbon substrates and had limited ability for cell wall degradation. This species did not utilise sucrose, but may prefer other phloem sugars such as stachyose. Rather than feeding on cell wall components, it likely uses its limited cell wall-degrading arsenal to gain entry and exploit resources within plant cells. Overall, carbon metabolism in three ecologically different, but related fungi reflected their ecological adaptations.

INTRODUCTION

Arthropod-vectored fungi in the Microascales have been the source of great economic losses world-wide. These include various pathogens of food crops (e.g., *Ceratocystis fimbriata s.s.* Ellis & Halst, *C. cacaofunesta* Engelbrecht & Harrington and *C. manginecans* M. van Wyk, Al Adawi & M.J. Wingf.) (Engelbrecht *et al.*, 2007; Van Wyk *et al.*, 2007) and of plantation trees (e.g. *C. albifundus* M.J. Wingf., De Beer & M.J. Morris) (Roux & Wingfield, 2013; Wingfield *et al.*, 1996), as well as saprotrophic or weakly pathogenic species that cause sap stain of plantation trees (Uzunovic *et al.*, 1999). Arthropod-vectored fungi in this order have also been reported to colonize the flower heads (infructescences) of South African *Protea* L. species without showing any obvious signs or symptoms of disease (Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988). The fungi from this unusual niche were identified as species of *Knoxdaviesia* M.J. Wingf., P.S. van Wyk & Marasas (De Beer *et al.*, 2013; Wingfield & Van Wyk, 1993; Wingfield *et al.*, 1988).

In addition to *Protea*, *Knoxdaviesia* species have been reported from a range of dead, decaying or diseased wood and plant tissues (De Beer *et al.*, 2013; Morgan-Jones & Sinclair, 1980; Pinnoi *et al.*, 2003; Van der Linde *et al.*, 2012), as well as weevil galleries (Kolařík & Hulcr, 2009). The genus includes nine species, of which only three are associated with *Protea*. One of these is *K. wingfieldii* Roets & Dreyer that was discovered on *P. caffra* Meisn. in KwaZulu-Natal (Crous *et al.*, 2012). The remaining two species are known only from the Core Cape Subregion (CCR) biodiversity hotspot in the Western Cape Province of South Africa (Bergh *et al.*, 2014; Mittermeier *et al.*, 1998). Here, *K. proteae* M.J. Wingf., P.S. van Wyk & Marasas occurs exclusively on *P. repens* L. (Wingfield *et al.*, 1988), while the closely related *K. capensis* M.J. Wingf. & P.S. van Wyk occurs on several *Protea* species, including *P. repens* (Aylward *et al.*, 2015; Roets *et al.*, 2009b; Wingfield & Van Wyk, 1993).

Protea-associated *Knoxdaviesia* species are primarily vectored by mites (Roets *et al.*, 2011b), and these mites, in turn, are phoretic on the beetle, and possibly bird, pollinators of *Protea* species (Roets *et al.*, 2009a; N. Theron pers. comm.). The fungi enter the *Protea* inflorescences early in the flowering stage, when flowers are pollen receptive (Roets *et al.*, 2009a). During this time, the inflorescences produce copious amounts of nectar (Cowling & Mitchell, 1981; Wiens *et al.*, 1983) that likely support the growth of the fungi. Once the inflorescences of serotinous *Protea* plants have completed flowering, they are enclosed by involucral bracts to form fire-safe, seed storage structures known as infructescences (Rebelo,

2001). Within these enclosed structures, *Knoxdaviesia* species flourish and can occur on the styles, pollen presenters and perianth of the dead flowers as well as on the inner surface of the involucral bracts (Lee *et al.*, 2005). Their sporulating ascomata are some of the most dominant fungal features within infructescences (Lee *et al.*, 2005; Marais & Wingfield, 1994) and can proliferate for more than a year after pollination (Roets *et al.*, 2005). At this stage, nectar sugars are likely depleted and it is expected that the *Protea*-associated fungi rely on the decaying floral parts as a food source.

The decaying infructescence habitat occupied by *Knoxdaviesia* species varies greatly from that of many other Microascalean fungi that typically invade living vascular plant tissue (Morris *et al.*, 1993; Roux & Wingfield, 2009; Van Wyk *et al.*, 2007). For example, the black wattle (*Acacia mearnsii* De Wild) pathogen, *Ceratocystis albifundus*, infects vascular tissue through wounds caused by other biotic or abiotic factors (Roux *et al.*, 2007). Easy re-isolation of the pathogen from discoloured sapwood suggests that it has direct contact with both xylem and phloem vessels, although the rapid upward infection (Roux *et al.*, 1999a) could implicate xylem as the primary mode of spread. Phloem sap is rich in carbohydrates, of which sucrose is frequently the dominant component (Pritchard, 2007). Xylem does not contain sugars, but transports mineral nutrients to which *C. albifundus* would have direct access (Lucas *et al.*, 2013). Although the eventual death of the plant results from wilting, the observed lesions and gummosis caused by this pathogen suggest that *C. albifundus* also attacks plant cells (Morris *et al.*, 1993; Roux *et al.*, 1999a).

In the Western Cape Province, tree hosts of the pathogenic Microascalean fungi are widespread across commercial plantations that often border areas of natural fynbos vegetation and several plantation species have invaded fynbos (Van Wilgen, 2009). However, the *Protea*-associated species have not been found on other trees. The converse is also true, where both apparently native and introduced Microascalean pathogens have been found on plantation and even indigenous forest trees (Kamgan *et al.*, 2008; Roux *et al.*, 2007; Roux & Wingfield, 2009), yet these pathogens have never been found in *Protea* infructescences. However, *C. albifundus* has been reported from stem cankers of several indigenous trees, including *Protea* species (Crous *et al.*, 2004; Lee *et al.*, 2016; Roux *et al.*, 2007), in South Africa where the fungus is believed to be native (Roux *et al.*, 2001; 2007). Even though the *Protea* host is shared in this case, a distinct difference in niche occupation is apparent, with the pathogenic species occupying the vascular tissue associated with wounds (Morris *et al.*, 1993; Roux *et al.*, 1999b) and the saprotrophic species decaying floral parts in seed cones.

In this study, we integrate phenome and whole-genome data to consider the substrate utilisation of the generalist *K. capensis*, the specialist *K. proteae* and the pathogen *C. albifundus*. Our first objective was to investigate the discrepancy in host range between the two *Knoxdaviesia* species. Roets *et al.* (2011a) found that although the chemistry of *Protea* is not the only factor influencing the occurrence of *Protea*-associated ophiostomatoid fungi, it is the dominant factor in *Protea* species that are known hosts of *Knoxdaviesia* species. These authors also observed that both *K. capensis* and *K. proteae* grow optimally in culture when flowers of their natural *Protea* host are added to the medium. To explain these observations, we hypothesise that *K. capensis* is capable of using a wider range of substrates than *K. proteae*, enabling it to associate with different *Protea* hosts with variation in host chemistries. Our second objective was to characterise metabolic pathways that facilitate life in *Protea* infructescences in contrast to a lifestyle involving pathogenic wound association. We hypothesise that the vascular tissue habitat of *C. albifundus* enables it to persistently exploit the availability of sugars, whereas *Knoxdaviesia* species may only be exposed to nectar sugars for a short period after which they must switch to breaking down complex polysaccharides from decaying plant cells.

MATERIALS AND METHODS

Fungal isolates

One isolate of each of *K. capensis* (CBS139037), *K. proteae* (CBS140089) and *C. albifundus* (CMW17620) were used. These isolates were specifically selected because of the availability of their genome sequences in the public domain (Aylward *et al.*, 2016; Van der Nest *et al.*, 2014). For the analysis of glucose, fructose and sucrose usage, we also included four additional isolates of *C. albifundus* (CMW4068, CMW13980, CMW17274, CMW24685) and all isolates have been conserved in the culture collection (CMW) of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa. All cultures were routinely grown on Malt Extract Agar (MEA; Merck, Wadeville, South Africa) at 25°C for approximately seven days and thereafter, maintained at 4°C.

Phenotype microarrays

Biolog FF MicroPlates PM1 and PM2 (Biolog Inc., Hayward, CA) were used to investigate carbon metabolism in *K. capensis*, *K. proteae* and isolate CMW17620 of *C. albifundus*. Each microplate contained 95 distinct carbon sources and one water control (Table S5.1). Fungal cultures were grown on MEA plates overlaid with sterile cellophane (Product no. Z377597, Sigma-Aldrich, Steinham, Germany) until the fungal growth covered two thirds of the cellophane. Fungal material was then scraped from the cellophane, placed in 1.5 ml sterile water and homogenised with three 3 mm glass beads in an MM301 TissueLyser (Retsch, Inc., Germany). The mycelial debris was pelleted by centrifugation at 1.1 rcf for 1 minute and the cell suspension transferred to a new Eppendorf tube. A dilution series of the cell suspension was prepared and the transmittance measured using a PowerWave^{TM} HT Scanning Microplate Spectrophotometer (BioTek Instruments, Winooski, VT). The dilutions were adjusted so that a cell suspension with a transmittance (T) of 62% (absorbance = 0.21) was obtained. The 62% T suspension was used to prepare the fungal inoculum as described in the Biolog protocol (PM Procedures for Filamentous Fungi, 25-Aug-07) and 100 µl of the inoculum was pipetted into each well of microplate PM1 and PM2. Three technical replicates were performed several weeks apart. A new inoculum was prepared for every replicate, whereas the same inoculum was used on both PM1 and PM2 for each individual replicate of a species. The plates were incubated at 25°C and the amount of mycelial biomass produced in each well was measured at regular intervals (*ca*. five times daily) for seven days at 750 nm (Atanasova & Druzhinina, 2010; Mchunu *et al.*, 2013), using the PowerWave HT Microplate Spectrophotometer.

Glucose, fructose and sucrose utilisation was subsequently verified using the two *Knoxdaviesia* isolates and all five of the *C. albifundus* isolates included in this study. For this purpose, 15 µl of a spore suspension (prepared according to the BioLog protocol) was inoculated onto agar (15g/L) medium supplemented with 20% (weight:volume) of the relevant sugar and 11 mg/L phenol red sodium salt (Sigma-Aldrich, Steinham, Germany). Potato Dextrose Agar (Merck, Wadeville, South Africa) medium with phenol red sodium salt was used as a positive control, and medium containing agar only (15g/L), also with phenol red sodium salt, was used as negative control. The experiment was performed in triplicate. Because phenol red is a pH indicator that is yellow at pH < 6.5 and red at pH > 8, in an initially basic medium, it acted as a visual indicator of the accumulation of acidic metabolic by-products. The initial pH of the culture media was increased by adding 1-2 µl of 5M NaOH

per millilitre agar until a red colour was achieved. Cultures were grown for two weeks at 25°C after which growth was investigated with the aid of a dissection microscope.

Phenome analysis

Growth curves from the absorbance values recorded from the inoculated Biolog FF MicroPlates PM1 and PM2 were drawn in R 3.2.3 (R Core Team, 2014) using the *opm* package (Vaas *et al.*, 2013). Individual growth curves were manually analysed as well as grouped into active and inactive categories with the Vehkala *et al.* (2015) R pipeline, while considering the negative control (well A01), but not subtracting its signal. Only curves that showed activity in at least two of the replicates were considered active. The area under the curve (AUC) and lag phase (λ) parameters were estimated from the growth curves using a spline-based fitting procedure with 1 000 bootstrap replicates (Vaas *et al.*, 2012).

Genome analysis

The genomes of *Knoxdaviesia capensis* (LNGK00000000.1), *K. proteae* (LNGL00000000.1) and *C. albifundus* (JSSU00000000.1) are available on GenBank® (Benson *et al.*, 2013). We annotated the *Knoxdaviesia* genomes with MAKER (Cantarel *et al.*, 2008; Holt & Yandell, 2011) in a previous study (Aylward *et al.*, 2016), whereas the MAKER annotations of the *C. albifundus* genome (Van der Nest *et al.*, 2014) were provided by D. Roodt. The proteomes of these species were then annotated with KEGG (Kyoto Encyclopedia of Genes and Genomes) Orthology (KO) identifiers using the online BlastKOALA version 2.0 server (Kanehisa *et al.*, 2016; www.kegg.jp/blastkoala/). Via KEGG Mapper analysis, each protein with a KO annotation was mapped to its KEGG pathway and BRITE functional hierarchy (Kanehisa *et al.*, 2011). KO numbers associated with carbon and carbohydrate metabolism were investigated to identify functions unique to each species. Local tBLASTn and BLASTp searches were conducted against the genomes and predicted proteomes, respectively, of the opposite species to ensure that each unique protein was truly absent in the other two species and not the result of incomplete or incorrect annotation. Based on sequence similarity only, protein homology can be inferred confidently above 40 % amino acid identity (Rost, 1999). In doing BLAST searches, however, we were interested not only in protein homology, but more so in functional conservation as described by the Enzyme Commission (EC) numbers

associated with KO annotations (Bairoch, 2000; Kanehisa *et al.*, 2004). As suggested previously (Addou *et al.*, 2009), proteins were considered EC functional homologs when they shared an amino acid sequence similarity of at least 60 % over ≥ 60 % of the query sequence length. The identified polysaccharide degrading enzymes were interrogated with SignalP 4.1 (Petersen *et al.*, 2011) to predict whether they are secreted or intracellularly localized.

RESULTS

Substrate utilisation of *Knoxdaviesia* and *C. albifundus*

Knoxdaviesia capensis had a positive reaction, albeit at varying levels, to 51 of the 190 substrates, while *K. proteae* only showed significant growth on 32 substrates of which only one (di-peptide Gly-Pro) was not shared with *K. capensis* (Table 5.1, Fig. 5.1). The unique substrate use of *K. capensis* is, therefore, 45.5 % and only 3.1 % for *K. proteae*, suggesting that *K. capensis* is able to utilise a wider range of substrates than *K. proteae*. In contrast to the two *Knoxdaviesia* species, we observed positive growth for *C. albifundus* on only 15 of the 190 substrates tested. Of these, the utilisation of only 6 substrates was shared between *C. albifundus* and the two *Knoxdaviesia* species; four with only *K. capensis* and two with both *K. capensis* and *K. proteae*. In all six cases of common substrate utilisation, *C. albifundus* showed the lowest level of growth and the longest lag phase (see supplementary materials Fig. S5.1 for growth curves).

Fungal growth was expressed and compared using the area under the curve (AUC) parameter estimated from the growth curves (Fig. 5.1 and supplementary materials Fig. S5.1). For *Knoxdaviesia*, substrates that resulted in the most growth were primarily simple carbon sources (i.e., monosaccharides and disaccharides). Overall, *K. capensis* showed greater biomass production on the same substrates than *K. proteae*, although *K. proteae* grew better on two monosaccharides (D-xylose and β-methyl-D-galactoside), a peptide (gelatin), a polysaccharide (laminarin), and a glycoside (arbutin).

Table 5.1 Results of the phenome analysis[1] in *Knoxdaviesia capensis* (*Kc*), *K. proteae* (*Kp*) and *Ceratocystis albifundus* (*Ca*).

Substrate	Compound category	Growth[2]		
		Ca	***Kc***	***Kp***
L-Histidine	amino acid	x		
L-Isoleucine	amino acid	x	x	
L-Leucine	amino acid	x	x	
L-Alaninamide	amino acid		x	
L-Aspartic Acid	amino acid		x	
L-Glutamic Acid	amino acid		x	x
L-Alanine	amino acid		x	x
L-Proline	amino acid		x	x
L-Asparagine	amino acid		x	
L-Glutamine	amino acid		x	x
L-Phenylalanine	amino acid		x	
L-Arginine	amino acid		x	
L-Ornithine	amino acid		x	x
γ-Amino-n-Butyric Acid	amino acid		x	
Ethanolamine	amino alcohol		x	
4-Hydroxy-Benzoic Acid	carboxylic acid		x	x
Butyric Acid	carboxylic acid		x	x
Propionic Acid	carboxylic acid		x	x
D-Fructose	monoshaccaride		x	x
D-Fucose	monosaccharide	x		
D-Galactose	monosaccharide	x	x	x
D-Glucose	monosaccharide		x	x
D-Mannose	monosaccharide		x	x
D-Ribose	monosaccharide	x		
D-Xylose	monosaccharide		x	x
L-Arabinose	monosaccharide		x	x
L-Lyxose	monosaccharide		x	x
N-Acetyl-D-Glucosamine	monosaccharide		x	x
β-Methyl-D-Galactoside	monosaccharide		x	x
β-Methyl-D-Glucoside	monosaccharide		x	x
D-Cellobiose	disaccharide		x	x
D-Trehalose	disaccharide		x	x
Lactulose	disaccharide		x	x
α-D-Lactose	disaccharide		x	x
β-Gentiobiose	disaccharide		x	x
Maltotriose	trisaccharide	x		
Stachyose	tetrasaccharide	x		
Dextrin	polysaccharide		x	
Glycogen	polysaccharide		x	
Inulin	polysaccharide	x		
Laminarin	polysaccharide		x	x

Pectin	polysaccharide		x	
Gly-Pro	dipeptide			x
Gelatin	peptide		x	x
Adonitol	sugar alcohol		x	x
D-Arabitol	sugar alcohol	x	x	
D-Mannitol	sugar alcohol		x	x
D-Sorbitol	sugar alcohol		x	
Dulcitol	sugar alcohol		x	
L-Arabitol	sugar alcohol	x		
Xylitol	sugar alcohol		x	
D-Salicin	glucoside, alcoholic		x	x
L-Galactonic Acid-γ-Lactone	ester	x		
Methyl Pyruvate	ester	x		
Arbutin	glycoside	x	x	x
α-Keto-Glutaric Acid	dicarboxylic acid		x	
α-Keto-Valeric Acid	keto acid	x	x	
3-Hydroxy-2-Butanone (Acetoin)	ketone		x	
Uridine	nucleoside		x	
Tween 40	polysorbate		x	x
Tween 80	polysorbate		x	x
Total	*61*	*15*	*51*	*32*

[1.] Phenome assays conducted with Biolog FF MicroPlates PM1 and PM2 (Biolog Inc., Hayward, CA)

[2] Growth is indicated with "x"; empty cells indicate no growth.

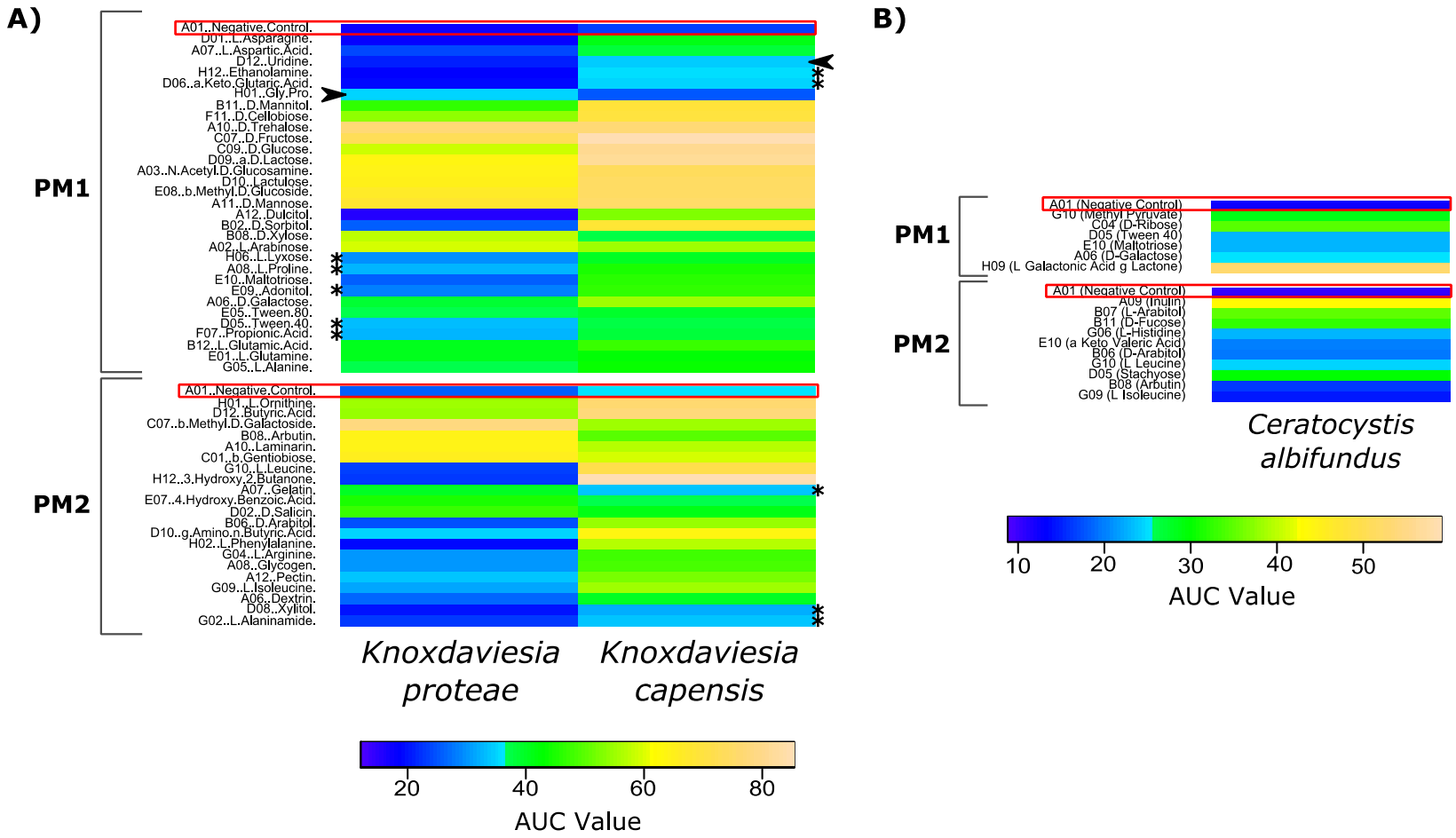

Figure 5.1 Heatmaps representing the average area under the curve (AUC). Relative to *Knoxdaviesia* (A), *Ceratocystis albifundus* (B) showed low growth on all substrates and is consequently depicted on its own. In the *Knoxdaviesia* heatmap, substrates in the blue spectrum are regarded as inactive, however, arrows and asterisks indicate potentially active substrates with a long lag phase and very low level of growth, respectively. In the *C. albifundus* heatmap, the blue spectrum represents a low level of growth. Negative controls are indicated with red boxes.

For carbohydrates, the two *Knoxdaviesia* species differed only in their utilisation of polysaccharides. Both grew on laminarin, a β-1,3-glucan, but *K. capensis* also grew on the α-1,4 and α-1,6 glucans, dextrin and glycogen and on the complex plant cell wall polysaccharide, pectin. Surprisingly, however, *K. capensis* did not grow on D-galacturonic acid, which forms part of the pectin backbone (i.e., pectin side chains can vary, while its backbone is composed of galacturonic acid homopolymers with α-1,4 linkages) (Mohnen, 2008). These differences in polysaccharide utilisation suggest that both *Knoxdaviesia* species have β-D-1,3-glucanase activity (Nelson *et al.*, 1963), but only *K. capensis* appears able to hydrolyse α-glycosidic bonds in polysaccharides.

Ceratocystis albifundus grew on very few carbohydrates. Surprisingly, no growth was observed on glucose, fructose and sucrose. It grew on three monosaccharides, D-fucose, D-galactose and D-ribose, as well as two oligosaccharides with α-1,4 and α-1,6, glycosidic bonds, respectively, maltotriose and stachyose. Although no fructose utilisation was apparent, *C. albifundus* also grew on a β-2,1 fructan, inulin, suggesting that this fungus is capable of breaking at least three types of glycosidic bonds.

Ceratocystis albifundus generally produced less biomass than the two *Knoxdaviesia* species on the PM plates because of an apparently extended lag phase (Fig. 5.2). The lag phase (λ; the time between inoculation and growth onset), of all three species was evaluated for substrates showing positive growth. The modelled λ was often negative in one or more of the replicates, independent of the spline-fitting method used. A lag phase of zero is, however, the lowest possible value since initial optical density measurements were made shortly after inoculation. Negative λ values have also been observed in other studies that considered optical density values as growth indicators and appear to be the result of a poor fit to the data (Lindqvist, 2006; Standaert *et al.*, 2007). For the substrates showing positive growth, replicate two had predominantly positive λ values that corresponded with what was manually observed on the growth curves. Lag phase was, therefore, investigated only for this replicate. The plot of λ values and 95% confidence intervals (Fig. 5.2) showed that *C. albifundus* had the longest lag phase of the three species, lasting between 74 and 125 hours (mean = 101.50). Both *Knoxdaviesia* species had a lag phase that lasted less than 60 hours in the majority of cases. For the 31 substrates utilised by both *Knoxdaviesia* species, *K. capensis* took longer to begin its growth period in all but five cases. The average λ was 64.13 hours for *K. capensis* (range = 14.5-127) and 59.38 hours for *K. proteae* (range = 16-116). *Knoxdaviesia capensis* took longest to begin assimilating D-xylose, butyric acid and uridine (in decreasing order),

whereas *K. proteae* had the longest λ on the dipeptide Gly-Pro (its only unique substrate) and, similar to *K. capensis*, butyric acid.

Verification of the glucose, fructose and sucrose utilisation showed results congruent with those found using the BioLog plates. For *K. capensis* and *K. proteae*, the agar only and sucrose plates retained their pink colour. Some spore germination and sparse hyphal growth was apparent on the agar only control. This growth pattern is known in *Knoxdaviesia* species when they are cultivated on water agar and is exploited to store cultures for extended periods of time. *Knoxdaviesia* hyphal growth occurred on the glucose and fructose plates and sporulation was observed at the points of inoculation. All *C. albifundus* strains demonstrated the same phenotypes. The only *C. albifundus* treatment that retained its initial pink colour was the agar control, although the growth form on this control could not be distinguished from the glucose, sucrose and fructose treatments. In each of the three treatments, limited spore germination, similar to that of the negative control, was observed and growth did not seem to occur beyond initial germination. A small number of ascomata developed on the fructose, glucose and sucrose treatments in isolate CMW17274. The yellow colour development on the glucose, fructose and sucrose plates was indicative of metabolic activity, although the lack of growth suggested limited exploitation of these substrates after germination. Even if these sugars facilitate initial spore germination and metabolic activity, the lack of further growth suggests that *C. albifundus* could not utilise them as a sole carbon source.

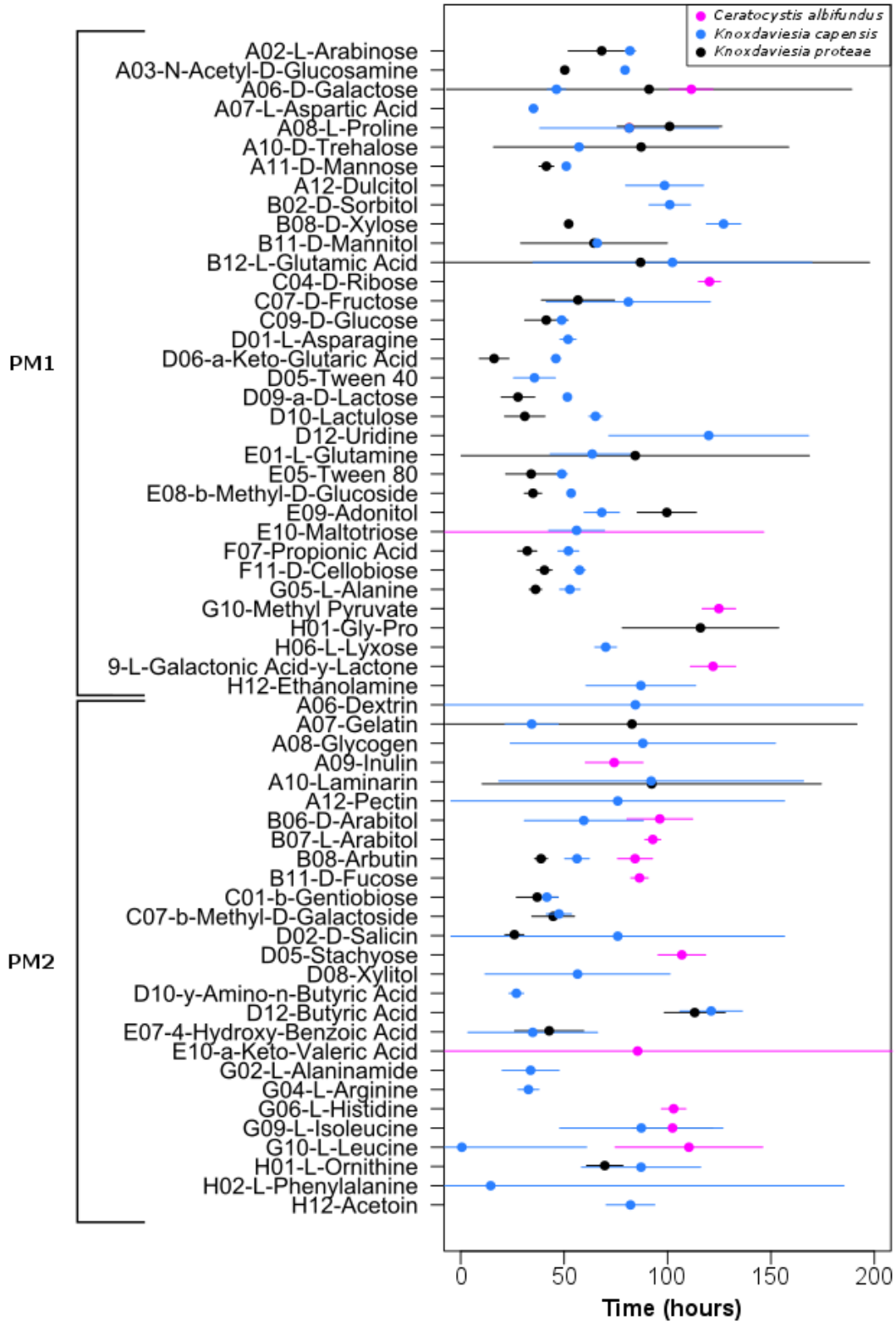

Figure 5.2 Length of the lag phase of *Knoxdaviesia capensis*, *K. proteae* and *Ceratocystis albifundus* on PM1 and PM2 substrates that induced growth. Horizontal lines represent the 95 % confidence interval.

Predicted proteome-based analysis of carbon and carbohydrate metabolism

Although having larger predicted proteomes than *C. albifundus*, < 37 % of the proteomes of each *Knoxdaviesia* species could be mapped to KEGG, in contrast to the 41.5 % of *C. albifundus* proteins (Table S5.2). Although a large portion of all three genomes, therefore, remains unclassified according to KEGG nomenclature, the KEGG database is sufficient to classify well-characterised proteins involved in metabolism (Kanehisa *et al.*, 2011; 2016). All three species had similar proportions of their annotated proteins devoted to carbohydrate metabolism (7.6 -7.9 %), although the actual number (216 proteins) is the lowest in *K. proteae*.

The primary differences in gene content related to carbon metabolism were observed between *Knoxdaviesia* and *C. albifundus*. The two *Knoxdaviesia* species, although displaying differences in gene copy number (see Table 5.2), had similar gene identities for carbon metabolism. The primary carbohydrate utilisation pathways predicted for these fungi are summarized in Fig. 5.3.

Structural plant polysaccharides as well as photosynthates are potential sources of carbon for plant-associated fungi (Caffall & Mohnen, 2009). We, therefore, first investigated the presence of glycosidases, polysaccharide lyases and carbohydrate esterases that degrade polysaccharides (Table 5.2). Glycosidases (EC 3.2.1-), enzymes that free monosaccharide and oligosaccharide sugars through hydrolysis of glycosidic bonds (Divakar, 2013; Kötzler *et al.*, 2014), represented the largest of these three groups in the studied species. Thirty-three different glycosidases were identified in the *Knoxdaviesia* and *C. albifundus* genomes (Table 5.2). Of these putative genes, five were unique to *Knoxdaviesia* and seven to *C. albifundus*. Functionally, *C. albifundus* had a slightly greater range of glycosidase enzymes. However, considering the copy number of these enzymes, *Knoxdaviesia* species ultimately had 60 identified glycosidases each, whereas *C. albifundus* had only 47.

A single category of carbohydrate esterases that degrade plant polysaccharides, as defined by Coutinho *et al.* (2009), was identified in the predicted proteome of all three species. Three secreted acetylxylan esterases were present in *Knoxdaviesia* and one in *C. albifundus* (Table 5.2). These enzymes support glycosidases by deacetylating the acetylxylan component of hemicellulose (Biely, 2012). Polysaccharide lyases could be identified only in *C. albifundus*, where only four enzymes in three classes were predicted. All three classes are associated with pectin degradation.

Table 5.2 Identified glycosidase, carbohydrate esterase and polysaccharide lyase enzymes associated with carbohydrate degradation.

EC number[1]	Enzyme name[2]	Substrate	Number of proteins (Proteins with secretion signal)[3]		
			Kc	***Kp***	***Ca***
Cellulose degradation					
3.2.1.4	cellulase	cellulose	5(3)	6(6)	2(2)
3.2.1.91	cellulose 1,4-β-cellobiosidase	cellulose	5(5)	4(4)	1(1)
3.2.1.21	β-glucosidase	glucans, e.g. cellobiose	7(2)	6(3)	5(0)
3.2.1.6	endo-1,3(4)-β-glucanase	glucans	1(0)	1(0)	1(0)
Hemicellulose degradation					
3.2.1.8	endo-1,4-β-xylanase	xylans	6(5)	6(4)	2(2)
3.2.1.37	β-xylosidase	xylans	1(1)	1(0)	0
3.2.1.177	α-xylosidase	xyloglucans	2(0)	2(0)	0
3.2.1.155	xyloglucan-specific exo-β-1,4-glucanase	xyloglucans	0	0	1(1)
3.1.1.72	acetylxylan esterase	acetylated xylan	4(3)	4(3)	1(1)
3.2.1.25	β-mannosidase	mannosyl-oligosaccharides	1(0)	1(0)	1(0)
3.2.1.78	β-mannanase	mannan	1(1)	1(1)	0
3.2.1.55	α-L-arabinofuranosidase	arabinan in xylan/pectin	1(0)	1(0)	2(1)
3.2.1.22	α-galactosidase (melibiase)	galactans	1(0)	1(0)	0
3.2.1.23	β-galactosidase	galactans	3(1)	3(1)	1(0)
Pectin degradation					
3.2.1.67	galacturan 1,4-α-galacturonidase	pectin backbone	0	0	1(1)
3.2.1.89	arabinogalactan endo-β-1,4-galactanase	arabinogalactans	0	0	1(1)
3.2.1.99	arabinan endo-1,5-α-L-arabinanase	arabinans	0	0	1(1)
4.2.2.2	pectate lyase	homogalacturonan	0	0	1(1)
4.2.2.10	pectin lyase	homogalacturonan	0	0	2(2)
4.2.2.23	rhamnogalacturonan endolyase	rhamnogalacturonan	0	0	1(1)

Storage polysaccharide degradation					
3.2.1.1	α-amylase	starch	0	0	1(1)
3.2.1.3	glucoamylase	glucans	1(0)	1(0)	1(1)
3.2.1.20	α-glucosidase	e.g. sucrose, starch	1(0)	1(0)	0
3.2.1.33	amylo-α-1,6-glucosidase	glycogen	1(0)	1(0)	1(0)
3.2.1.26	β-fructofuranosidase (invertase)	e.g. sucrose & maltose	0	0	2(1)
3.2.1.28	α,α-trehalase	trehalose	2(1)	2(0)	2(1)
3.2.1.39	glucan endo-1,3-β-D-glucosidase	glycans	1(1)	1(1)	1(0)
3.2.1.58	glucan 1,3-β-glucosidase	β-1,3-D-glucans	1(1)	1(1)	2(1)
Cell surface carbohydrates (glycoproteins / glycolipids)					
3.2.1.24	α-mannosidase	mannan	1(0)	1(0)	1(0)
3.2.1.52	β-N-acetylhexosaminidase	N-acetylglucosides / N-acetylgalactosides	2(1)	2(1)	2(1)
3.2.1.84	glucan 1,3-α-glucosidase	α-1,3-D-glucans	1(1)	1(0)	1(1)
3.2.1.106	mannosyl-oligosaccharide glucosidase	mannosyl-oligosaccharides	1(1)	1(1)	1(1)
3.2.1.113	mannosyl-oligosaccharide 1,2-α-mannosidase	mannosyl-oligosaccharides	5(2)	5(2)	4(1)
3.2.1.164	galactan endo-1,6-β-galactosidase	arabinogalactans	0	0	1(1)
Other					
3.2.1.14	chitinase	chitin	4(2)	5(0)	4(3)
3.2.1.-	SUN family β-glucosidase		1(1)	1(1)	1(1)
3.2.1.101	mannan endo-1,6-α-mannosidase	α-1,6-mannan	4(4)	4(4)	3(0)
	Total number of different polysaccharide-degrading enzymes				
	Glycosidases (EC 3.2.1.)*		*60(33)*	*60(30)*	*47(24)*
	Carbohydrate esterases (EC 3.1.1.)*		*4(3)*	*4(3)*	*1(1)*
	Polysaccharide lyases (EC 4.2.2.)*		*0*	*0*	*4(4)*
	Total number of enzymes that degrade specific plant structures				

Cell wall	*38(21)*	*37(22)*	*24(15)*
Storage polysaccharides	*7(3)*	*7(2)*	*10(5)*
Cell surface carbohydrates	*10(5)*	*10(4)*	*10(5)*

[1] EC = Enzyme Commission (Bairoch, 2000)

[2] Enzyme names according to the KEGG ENZYME database (Kanehisa *et al.*, 2004). The total number of glycosidases, carbohydrate esterases and polysaccharide lyases as well as the total number of enzymes associated with degrading specific plant structures are summarised below.

[3] Number of proteins with specific EC functionalities predicted with BlastKOALA (Kanehisa *et al.*, 2016). The number of these proteins that are putatively secreted, as determined by SignalP (Petersen *et al.*, 2011), is indicated in brackets. *Kc* = *Knoxdaviesia capensis*, *Kp* = *K. proteae*, *Ca* = *Ceratocystis albifundus*.

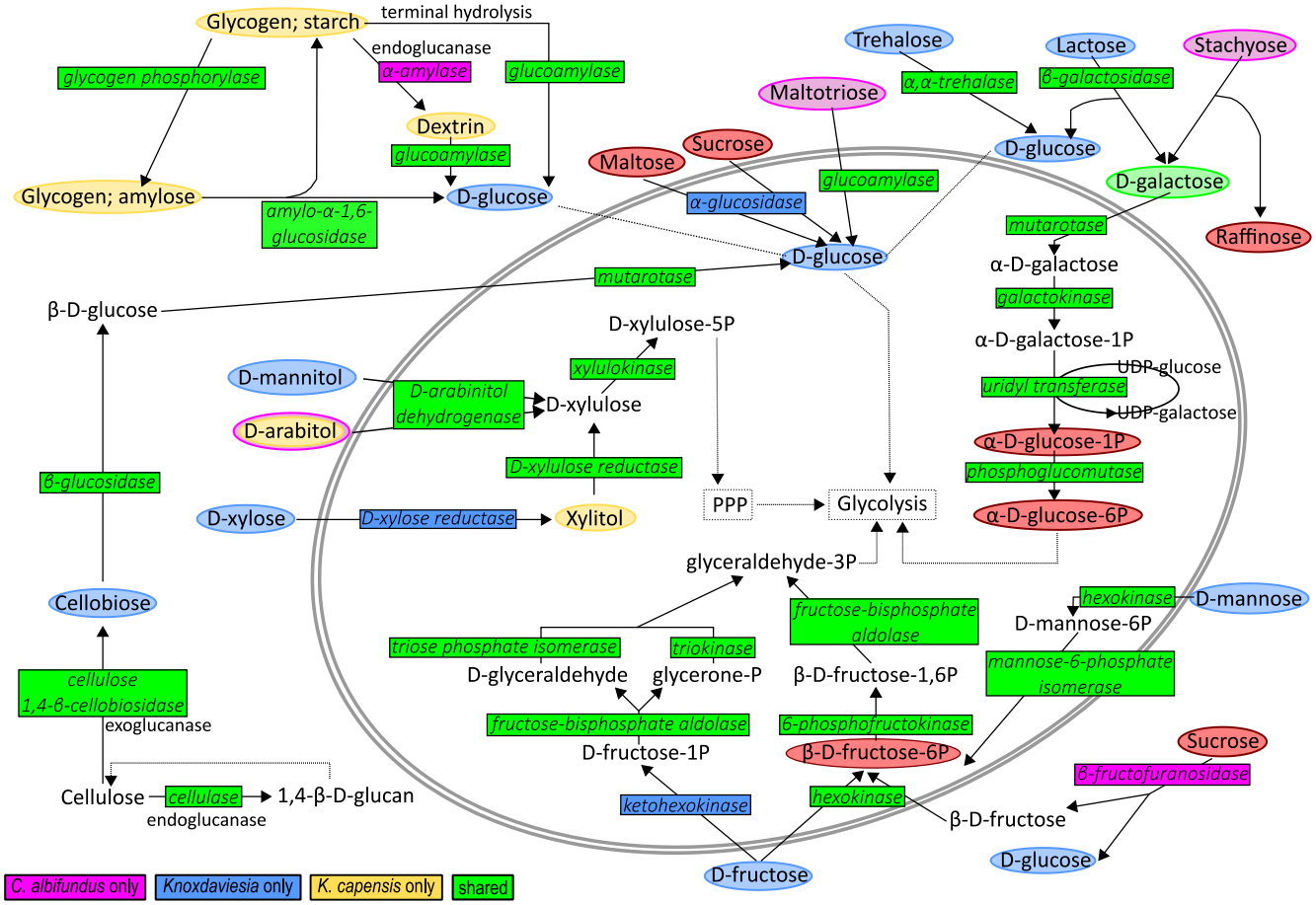

Figure 5.3 Carbohydrate utilisation pathways in *Knoxdaviesia capensis*, *K. proteae* and *C. albifundus*. Substrates tested in the PM assays are circled and coloured according to usage: used by all (green), used by none (red), used by both *Knoxdaviesia* species (blue), used by *K. capensis* only (yellow) and used by *C. albifundus* (purple). Both *K. capensis* and *C. albifundus* grew on D-arabitol. The same colouring scheme applies to the presence of enzymes. The grey double-lined oval represents the fungal cell.

A repertoire of enzymes for degrading plant cell wall components, storage polysaccharides and cell surface carbohydrates were predicted from both *Knoxdaviesia* and *C. albifundus* (Table 5.2). The discrepancy in the number of proteins between these species was primarily due to the number of different cell wall-degrading enzymes they encode. In terms of gene content, *K. capensis* and *K. proteae*, respectively, had 34 and 33 different proteins predicted to act on plant cell wall components, whereas *C. albifundus* had 23. Included among these 23 were three putatively secreted glycosidases, unique to *C. albifundus*, that degrade arabinans, arabinogalactans, and galacturans, which are all components of pectic polysaccharides (Caffall & Mohnen, 2009; Wong, 2008). *Knoxdaviesia* species are theoretically capable of degrading xylans, xyloglucans, galactans and mannans in hemicellulose, in addition to the cell wall-degrading enzymes shared with *C. albifundus*. The *Knoxdaviesia* species, however, had three fewer glycosidases associated with degrading plant storage polysaccharides than *C. albifundus*. Both *Knoxdaviesia* species and *C. albifundus* encoded ten enzymes that target cell surface polysaccharides. However, an additional enzyme class (EC 3.2.1.164) that degrades arabinogalactans on the cell surface was identified in *C. albifundus*.

Comparison of the phenome assays and predicted carbohydrate metabolism of *Knoxdaviesia* and *C. albifundus*

Galactose metabolism

Galactose was the only monosaccharide used by all three species studied (Table 5.1, Fig. 5.1). The proteome analysis indicated that this sugar enters glycolysis through the Leloir pathway of enzymes, also known in yeast as *GAL* genes (Sellick *et al.*, 2008). Initial KO annotations did not identify the second enzyme in this pathway, galactokinase (EC 2.7.1.6), in *K. proteae*. However, the BLASTp and tBLASTn searches with the *K. capensis* galactokinase protein had significant BLAST hits in the *K. proteae* proteome and genome, respectively. Since *K. proteae* grew on galactose in the phenome assays, the presence of this enzyme was expected. All species are also theoretically able to source galactose from galactan-containing oligo- and polysaccharides, such as lactose, via their β-galactosidase enzymes, although only the *Knoxdaviesia* species grew on lactose.

Xylan and D-xylose metabolism

Both *Knoxdaviesia* species were able to grow on D-xylose as the sole carbon source (Table 5.1, Fig. 5.1), a particularly intriguing ability since the nectar of some of the *Protea* hosts of

Knoxdaviesia species contain small amounts (0.1 – 5 %) of D-xylose (Nicolson & Van Wyk, 1998). The results of our proteome analysis (Table 5.2), indicated that this ability is likely linked to the D-xylose reductase enzyme (Verduyn *et al.*, 1985) encoded by *Knoxdaviesia*, but not identified in *C. albifundus*. The use of xylose as an energy source appears essential because the *Knoxdaviesia* species also encode α- and β-xylosidases that hydrolyse xyloglucans and xylans, respectively, in the plant cell wall, releasing D-xylose.

Fructose and mannose metabolism

A single difference between *Knoxdaviesia* and *C. albifundus* was noted in the fructose metabolism pathway – the absence of ketohexokinase (fructokinase) in *C. albifundus* (Table S5.3) and its inability to grow on D-fructose (Tables 5.1, Fig. 5.1). Although the enzymatic machinery for both of the known fructose degradation pathways is present in most organisms, the preferred pathway depends on the source of fructose, in microbes, or the organ, in animals (Kelker *et al.*, 1970; Tappy & Lê, 2010). The first pathway follows the well-known steps in glycolysis whereby fructose is sequentially phosphorylated at carbon 6 and 1 to form the glycolytic intermediate fructose-1,6-diP. The enzymes for this pathway were identified in all three species studied. The second pathway requires the ketohexokinase enzyme that is missing in *C. albifundus*, but needed to phosphorylate fructose to fructose-1P, which is subsequently broken down to glyceraldehyde-3P that enters glycolysis. *Ceratocystis albifundus* may, therefore, be capable of using the fructose-1,6-diP pathway to degrade fructose generated as a metabolic intermediate, but its inability to use exogenous fructose can be linked to the absence of ketohexokinase.

Utilisation of D-mannose as a carbon source presumably proceeds by phosphorylation and subsequent isomerization of mannose-6P to fructose-6P (Valentine & Bainbridge, 1978). Although all three species appeared to encode the necessary enzymatic machinery to accomplish these reactions, only the *Knoxdaviesia* species grew on D-mannose. This may be due to the fact that the *Knoxdaviesia* species encode a putatively secreted endoglucanase enzyme (not identified in *C. albifundus*) to degrade the β-1,4 linked polysaccharides of D-mannose that form the primary chain of this sugar in hemicellulose (Moreira & Filho, 2008).

Starch and cellulose metabolism

The exoglucanase enzymes, glucoamylase and amylo-α-1,6-glucosidase, that can cleave α-1,4 and α-1,6 linked glucose residues from the terminal ends of starch and glycogen molecules (Divakar, 2013), were identified in the proteomes of *Knoxdaviesia* and *C.*

albifundus (Table 5.2). Starch and glycogen are large storage polysaccharides of α-D-glucose that differ in length and the extent to which they are branched (Buléon *et al.*, 1998; Manners, 1991), and the various enzymes identified hydrolyse successive glucose residues from the terminal ends of these polymers. An additional endoglucanase, α-amylase, that cleaves internal α-1,4 linkages (Divakar, 2013) was predicted from *C. albifundus*, although only *K. capensis* grew on glycogen and dextrin, a hydrolysed form of starch (Table 5.1, Fig. 5.1).

Cellulose is another polymer of glucose, but is defined by β-1,4 linkages and functions as a structural rather than a storage polysaccharide (Brown, 2004). Secreted cellulolytic enzymes with endoglucanase and exoglucanase activity were predicted from all three proteomes (Table 5.2). These may act synergistically on the plant cellulose fibres because the endoglucanases expose substrates for the exoglucanases (Teeri, 1997). The exoglucanases cleave the disaccharide cellobiose from cellulose terminals and β-glucosidase completes the hydrolysis to yield β-D-glucose. Although all three species possessed several copies of the β-glucosidase enzyme, secretion signals were identified in only the *Knoxdaviesia* β-glucosidase enzymes. This could explain why *C. albifundus* did not grow on cellobiose during the phenome assays (Table 5.1 and 5.2). The *Knoxdaviesia* species encoded approximately double the number of cellulolytic and hemicellulolytic enzymes predicted from the *C. albifundus* proteome (Table 5.2). This disparity most likely reflects the relative importance of cellulose and hemicellulose as energy sources between the two groups.

Sucrose metabolism

Although they were not able to grow on sucrose, the *Knoxdaviesia* species grew on its glucose and fructose monomers (Table 5.1, Fig. 5.1). The α-glucosidase enzymes identified in the predicted proteomes of these species could hydrolyse sucrose, but did not contain secretion signals and have also previously been described as intracellularly localized enzymes (Batista *et al.*, 2005; Sutton & Lampen, 1962). The two *Knoxdaviesia* species would, therefore, not be able to degrade sucrose outside the cell, but would have to transport the molecule and cleave it intracellularly. The same sucrolytic α-glucosidase enzymes would also be responsible for cleaving the α-1,4 bond between the two glucose residues in maltose, explaining the lack of growth on this substrate.

The lack of *C. albifundus* growth on sucrose was surprising since two β-fructofuranosidases (commonly known as invertases) that hydrolyse the β-1,2 bond in sucrose were identified in

this species (Tables 5.1 and 5.2, Fig. 5.1). β-1,2 bonds are also found in other fructans, such as inulin, on which *C. albifundus* growth was observed.

Pectin degradation

Our proteome analyses (Table 5.2) showed that the enzymes involved in first half of the pectin degradation pathway (pectin to D-galacturonic acid) appears to be present only in *C. albifundus*, while those involved in the second half (D-galacturonic acid to L-glyceraldehyde) were present only in the *Knoxdaviesia* species (Fig. 5.4). All three of the species appeared to lack the final enzyme to convert L-glyceraldehyde to glycerol, although *Knoxdaviesia* should be capable of producing pyruvate by cleaving the preceding substrate, 2-keto-3-deoxy-L-galactonate (Martens-Uzunova & Schaap, 2008). None of the species were, however, capable of growing on D-galacturonic acid alone (Table 5.1, Fig. 5.1), implying that the downstream reactions of pectin degradation were non-functional.

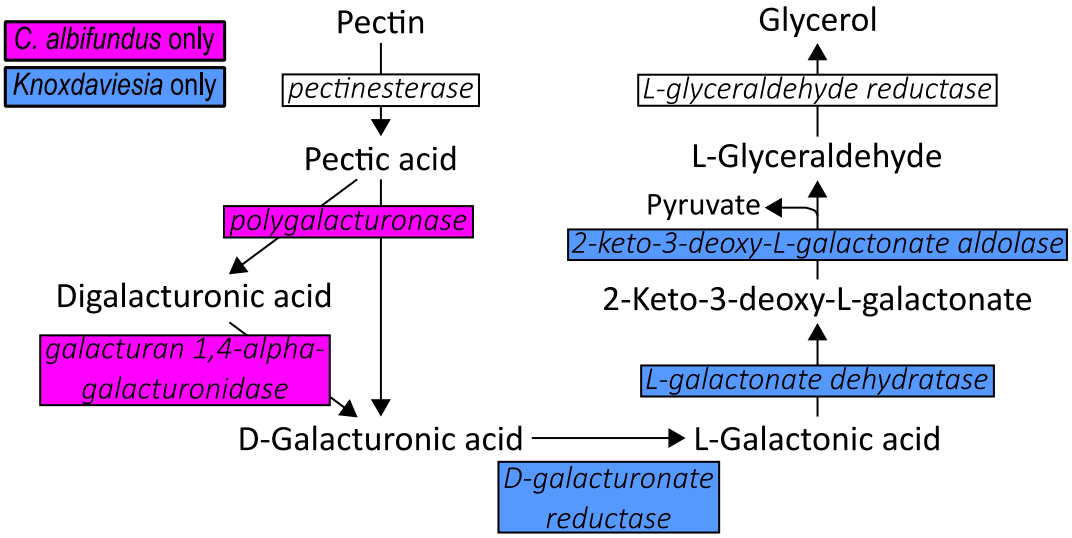

Figure 5.4 *Ceratocystis albifundus* and *Knoxdaviesia* enzymes associated with pectin backbone (homogalacturonan) degradation. None of the species are able to complete the pathway.

The only polysaccharide lyases (EC 4.2.2.-) that could be identified in the three proteomes examined belong to *C. albifundus* and are implicated in pectin degradation (Table 5.2). This species also encoded a collection of three different pectin-degrading glycosidases (Table 5.2). Although *C. albifundus* was apparently incapable of utilising pectin (Table 5.1, Fig. 5.1), the presence of these putatively secreted enzymes in its predicted proteome suggests that degradation of pectin may be an important process during *C. albifundus* infection.

Based on the proteome data, neither of the *Knoxdaviesia* species was expected to be capable of degrading pectin, yet *K. capensis* grew on pectin as a sole carbon source (Table 5.1, Fig. 5.1). This discrepancy may be due to the type of pectin used in the phenome assay. Pectin is a complex compound that can be constructed in several different ways: linear homogalacturonan polymers, xylogalacturonans branched by β-1,3-linked D-xylose and complex rhamnogalacturonans (Caffall & Mohnen, 2009; Wong, 2008). The pectin used in the phenome assay was purified from apple (Sigma Aldrich, Product P8471), a fruit that contains xylogalacturonans in its pectic fraction (Schols *et al.*, 1995). Although the major pectin-degrading enzymes were absent from *K. capensis*, its growth on this substrate could be explained by metabolism of the xylogalacturonan component, since the *Knoxdaviesia* species are potentially capable of releasing 1,4 linked D-xylose residues via β-xylosidase (Van den Brink & de Vries, 2011). *Knoxdaviesia capensis* is also the only species with a predicted secretion signal in its β-xylosidase enzyme (Table 5.2), explaining why *K. proteae* could not grow on pectin. Likewise, although *C. albifundus* appeared to be enzymatically incapable of hydrolysing the pectin backbone, its inability to utilise pectin sidechains may also be due to the specific type of the substance used in the phenome assay.

Sugar transporters

Although the glycolytic intermediates D-glucose-6P and fructose-6P are ubiquitous in living organisms, they did not stimulate growth in *Knoxdaviesia* or *C. albifundus*. Similarly, several other substrates are not used, despite the apparent presence of the necessary enzymes (e.g. sucrose in *C. albifundus*). However, the utilisation of a substrate depends on a range of factors: the physical presence of an enzyme to metabolise it, whether the enzyme is expressed and biologically active, and whether the necessary transport machinery is available to translocate the substrate into the cell where the enzyme can access it (Saier, 2000). We, therefore, investigated the proteins of each species that have KO identifiers associated with membrane transporters. Only five families associated with sugar transport were identified by KEGG (Table 5.3), although the single identified member of the GLUT family transports

myo-inositol rather than hexose sugars. The remaining four families are potentially able to transport glucose, galactose, trehalose, maltose and fucose, although, the transporter for the latter is predicted only in the *Knoxdaviesia* proteomes and not in *C. albifundus*. Additionally, each species was predicted to encode an unknown major facilitator transporter with unknown substrate affinity.

Table 5.3 Carbohydrate transport proteins identified in *Knoxdaviesia capensis* (*Kc*), *K. proteae* (*Kp*) and *Ceratocystis albifundus* (*Ca*).

				Number of proteins		
KO identifier[1]	**Transporter family**[2]	**Description**	**Potential substrate(s)**	***Kc***	***Kp***	***Ca***
K08150	SLC2: Facilitative GLUT transporter	myo-inositol member 13	myo-inositol	1	1	1
K15378	SLC45: Putative sugar transporter	member 1/2/4	glucose, galactose	2	2	2
K08141 (TC:2.A.1.1)	Sugar porter (SP) family	general alpha glucoside:H+ symporter	trehalose, maltose, turanose, isomaltose, α-methyl-glucoside, maltotriose, palatinose, melezitose	4	4	2
K02429 (TC:2.A.1.7)	Fucose:H+ symporter (FHS) family	L-fucose permease	fucose	1	1	0
K06902 (TC:2.A.1.24)	Unknown major facilitator-1	UMF1 family	uknown	1	1	1

[1] KO = KEGG Orthology identifiers predicted by BlastKOALA (Kanehisa *et al.*, 2016). TC numbers refer to the Transporter Classification Database (www.tcdb.org).

[2] Transporter family as described in the BRITE functional hierarchy (Kanehisa *et al.*, 2011). SLC = Solute Carrier Family; GLUT = Glucose Transporter

DISCUSSION

Cell-wall degrading enzymes: key to the saprotrophic lifestyle of *Knoxdaviesia*

The phenome assays have shown that *Protea*-associated *Knoxdaviesia* species are able to feed on exogenous glucose, fructose and xylose; three of the four sugars often present in *Protea* nectar (Nicolson & Van Wyk, 1998). As such, they would likely be the dominant source of energy during the early colonization of young *Protea* flowers (Roets *et al.*, 2009a). Based on our findings, we hypothesise that the reason *Knoxdaviesia* species do not metabolise sucrose is that they are unable to translocate it into the cytoplasm. Consequently, they cannot utilise it, since their only sucrolytic enzyme (α-glucosidase) is not secreted. Initial colonization of living *Protea* flower heads would be the only period during which sucrose is available. Plant invertases cleave sucrose into glucose and fructose, which causes sucrose levels to decrease with flower age in *Protea* nectar (Nicolson & Van Wyk, 1998). Therefore, due to their direct access to glucose and fructose in inflorescences and the presumed absence of sucrose in infructescences, sucrose transport and utilisation would be redundant for *Knoxdaviesia* species.

Results of the phenome study suggests that *Knoxdaviesia* species predominantly exploit arabinose, glucose, galactose, mannose and xylose, monomers that are abundant in cellulose and hemicellulose (Brown, 2004; Scheller & Ulvskov, 2010). Once the *Protea* infructescences have formed, nectar is depleted and the cell walls of flowers, the involucral bracts and any storage polysaccharides that may remain will likely be the only sources of carbon for these fungi. At this stage, the *Knoxdaviesia* species would need to switch from the easy uptake and metabolism of simple nectar sugars to the utilisation of complex carbohydrates. The carbohydrate active enzymes of few saprotrophic Sordariomycetes have been investigated, but available data suggests that the number of glycosidases and carbohydrate esterases identified in *Knoxdaviesia* is far below average (Martinez *et al.*, 2008; Van den Brink & de Vries, 2011; Zhao *et al.*, 2014). However, since *Protea* infructescences represent the only niche from which these fungi are known, their cell wall-degrading functionality has likely become focused on releasing cell wall monomers abundant in the *Protea* environment. Strong evidence for this is found in the results of the proteome analysis that displayed many, although not a great diversity, of cell wall degrading enzymes.

A comparison with *C. albifundus*, which uses stem wounds to establish on its hosts (Roux *et al.*, 2007; Roux & Wingfield, 2009), illustrates the importance of cell wall degrading

enzymes in the saprotrophic *Knoxdaviesia* species. *Ceratocystis albifundus* uses few monosaccharides and does not grow on any plant cell wall components. The cankers and lesions that develop as the *C. albifundus* infection spreads through the vascular tissue (Kubicek *et al.*, 2014; Morris *et al.*, 1993; Roux *et al.*, 1999a), may be partially explained by limited arsenal of cell wall degrading enzymes in this pathogen. As an example, pectin comprises approximately one third of dicot cell walls (Wong, 2008) and *C. albifundus* may be capable of weakening the cell wall structure if its pectinolytic enzymes are functional. Pathogen polygalacturonase enzymes (PGs) that attack the pectin backbone structure are well-documented as virulence proteins (Huang & Allen, 2000; Roper *et al.*, 2007; Rowe & Kliebenstein, 2007). In turn, plants often produce corresponding polygalacturonase inhibitor proteins (PGIPs) that elicit a defence response (Di *et al.*, 2006; Howell & Davis, 2005). It is, therefore, possible that *C. albifundus* evades plant defences by not producing PGs, but rather attacking the sidechains of pectin to weaken the structure. In contrast, *Knoxdaviesia* species apparently use their range of cell wall degrading enzymes to release monosaccharides from cellulose and hemicellulose. Gaining access to the cell is unnecessary for *Knoxdaviesia*, negating the need for this species to degrade pectin.

Evolution of specialist and generalist *Protea*-associated *Knoxdaviesia*

The results of our phenome analyses support the hypothesis that the generalist, *K. capensis*, has a greater capacity to utilise diverse substrates than its specialist counterpart. However, when considering substrates shared between the two species, *K. proteae* often has a shorter time to onset of growth. If this this time period (lag phase) were taken as a proxy for fitness (or competitive ability), the shorter lag phase of *K. proteae* on several substrates suggests that it may have a higher fitness than *K. capensis* under certain conditions. The host-exclusivity studies of Roets *et al.* (2011a) showed that *K. proteae* grows better than *K. capensis* on complete medium supplemented with *P. repens* flowers, supporting its increased fitness on this host. The trade-off hypothesis, where generalist species are capable of using many resources while specialists use fewer resources with greater efficiency (Wilson & Yoshimura, 1994), could apply in this case. The specialist nature of *K. proteae* may, therefore, be a result of the generalist common ancestor of *Protea*-associated *Knoxdaviesia* increasing its efficiency on one host. Although sacrificing the ability to occupy other *Protea* hosts, the

specialised *K. proteae* could have gained superior fitness on *P. repens* that makes it dominant and able to outcompete *K. capensis* on this host (Aylward *et al.*, 2015).

Stable habitats lacking temporal variation favour specialisation (Wilson & Yoshimura, 1994) and it is possible that the stable *P. repens* infructescence micro-niche has driven specialisation in *K. proteae*. However, if this were true, it would be necessary to understand why the *Knoxdaviesia* ancestor specialised only on *P. repens* and not on each of its *Protea* hosts. One explanation may be that specialisation on *P. repens* does not limit host availability since it is the most widespread and common species in the African *Protea* clade (Rebelo, 2001; Protea Atlas Project, South African National Biodiversity Institute). In contrast, the numerous hosts of *K. capensis* have either a limited or patchy distribution throughout the CCR. The phylogenetic distance between host species could also influence this phenomenon. A study considering aphid parasitoids (Straub *et al.*, 2011) found that the fitness costs associated with being a generalists are more dependent on the taxonomic range of the hosts employed than the actual number of hosts. Therefore, a generalist may increase its fitness by specialising on a host that is phylogenetically distant from the others. Following the *Protea* phylogeny (Valente *et al.*, 2010), the hosts of *K. capensis* occur in three distinct groups. *Protea coronata* Lam., *P. longifolia* Andrews, *P. neriifolia* R. Br. and *P. obtusifolia* De Wild partly comprise a monophyletic taxon that is sister to the remaining five hosts. Further, *P. burchelli* Stapf., *P. laurifolia* Thunb., *P. lepidocarpodendron* L. and *P. magnifica* Link form part of a monophyly that is sister to the clade containing *P. repens*. Although the three host-groups are not necessarily phylogenetically close, members within the groups are closely related, leaving *P. repens* the "odd one out". Specialisation may have been the mechanism whereby the common ancestor of CCR *Knoxdaviesia* species increased its fitness on this lone host, whilst the generalist maintained its dominant presence in the remaining hosts.

Considering both the phenome and proteome results of this study, it is clear that genotype is not necessarily an accurate predictor of phenotype. Although *K. capensis* clearly utilised a wider range of substrates than *K. proteae*, the differences in this range could not be explained by the genes predicted to be involved in carbon metabolism. It is possible that silenced genes in *K. proteae* were still identified as functional due to recent divergence of these two *Knoxdaviesia* species (Wingfield *et al.*, 1999); enzymes in *K. proteae* may have evolved a more specific substrate affinity or these genes may not have been expressed during the phenome assays due to complex regulatory mechanisms. In the *K. proteae* specialist niche, the lack of selective pressure to degrade diverse substrates could have led to loss of function,

such as the inability of *K. proteae* to grow on glycogen. Although utilisation of a ubiquitous plant storage polysaccharide would be advantageous for a generalist species, more abundant or preferable substrates are likely available to *K. proteae*. Xylose may be such an example as it was one of the substrates on which *K. capensis* displayed the longest lag phase. In contrast, *K. proteae* readily used this sugar, probably because its *P. repens* host has the highest concentration (up to 5 %) of xylose in its nectar (Nicolson & Van Wyk, 1998).

Unexpected carbon resources for *C. albifundus*

Glucose and fructose are regarded as the simplest available sources of energy, but our phenome analysis showed that these carbon sources were not used by the pathogen *C. albifundu*s. Our results suggested that this might be due to a lack of proteins to transport these substrates across the plasma membrane. In *Saccharomyces cerevisiae* Meyen: E.C. Hansen, strains deficient in the major hexose transporters responsible for the uptake of glucose and fructose, cannot grow on these sugars (Batista *et al.*, 2005). Additionally, evidence from the bacterium *Aerobacter aerogenes* indicates that metabolism of externally provided D-fructose relies on the D-fructose-1P pathway (Kelker *et al.*, 1970; Sapico *et al.*, 1968), which also appears to be the case for *Knoxdaviesia* and *C. albifundus*. Even if this pathogen were, therefore, able to take up exogenous fructose, it would not be able to phosphorylate and metabolise it further due to the missing ketohexokinase enzyme (needed for the generation of fructose-1P from exogenous fructose). The inability of *C. albifundus* to use fructose and glucose was also clear from the fact that it did not grow on polysaccharides that released glucose or fructose monomers. The single exceptions for each were maltotriose (glucose trisaccharide) that induced little growth and inulin (fructose polymer) that induced copious growth in *C. albifundus*. The growth of *C. albifundus* on these polymers, coupled with its inability to utilise glucose and fructose, suggests that *C. albifundus* may transport short glucans and fructans and hydrolyse them intracellularly. Such a scenario would be congruent with the access that this pathogen has to phloem through which soluble oligosaccharides are transported, but in which reducing sugars such as glucose and fructose are relatively scarce or even absent (Taiz & Zeiger, 2006). The ability to take up and utilise oligosaccharides rather than monosaccharides could reflect an adaptation to nutrient availability in the environment occupied by *C. albifundus*.

The most surprising result of the phenome assays was the inability of *C. albifundus* to grow on sucrose, the most common carbohydrate in plant phloem (Pritchard, 2007). This was further confounded by the recent description of two invertases in *C. albifundus* that were predicted to facilitate sucrose uptake (Van der Nest *et al.*, 2015). Our results, however, suggest that the *C. albifundus* invertases may have affinity for storage carbohydrates other than sucrose. This pathogen grew well on stachyose, a tetrasaccharide that has been detected in the phloem of certain tree species (Pritchard, 2007). Since stachyose consists of two galactose, one glucose and one fructose residues, the growth of *C. albifundus* on this sugar can be explained either by extracellular hydrolysis of the two galactose residues or translocation of the tetrasaccharide into the cell. The affinity of *C. albifundus* for stachyose as well as for the glucose and fructose polymers mentioned above, suggests that oligosaccharides other than sucrose may be abundant in the phloem of its plant hosts.

The only three monosaccharides that induced growth in *C. albifundus* were galactose, fucose and ribose. The first two are components of plant cell walls and cell surface oligosaccharides (Staudacher *et al.*, 1999; Van den Brink & de Vries, 2011), whereas ribose is an essential precursor of nucleotide biosynthesis that may appear in its free form during nucleotide metabolism (Lager *et al.*, 2003). These monosaccharides could be used by *C. albifundus* when it does not have access to plant storage polysaccharides, for example during initial colonization of parenchyma cells in the pith (Lee *et al.*, 2016).

CONCLUSIONS

Knoxdaviesia species associated with *Protea* plants in the CCR have a lower than average number of carbohydrate hydrolysing enzymes in their proteome and this is likely due to their specialist habitat. These species grow on simple sugars abundant in *Protea* nectar and can also degrade complex cell wall polysaccharides once *Protea* inflorescences become senescent and nectar is depleted. We posit that the specialist species, *K. proteae*, has emerged from further specialisation on its single host, *P. repens* and this is likely due to a need for increased fitness. Although not fully reflected in the proteome analysis, phenome data showed that the specialisation of *K. proteae* has led to a loss in ability to use a large range of carbon sources. Saprotrophic fungi that consistently occupy *Protea* infructescences fungi are also known from the order Ophiostomatales (Roets *et al.*, 2013). Future characterisation of the carbohydrate metabolism genes in these species would, therefore, verify whether the limited

number of carbohydrate hydrolysing enzymes is due to the lack of diverse substrates in infructescences, as we hypothesise.

The tree pathogen *C. albifundus* displayed a carbon usage profile very different from the two saprotrophic *Knoxdaviesia* species. It uses few simple sugars and does not grow on either sucrose or its monomers, despite its direct access to fresh host vascular tissue. We speculate that stachyose may be the primary phloem carbohydrate used by *C. albifundus* and that this species translocates glucose and fructose oligosaccharides rather than monosaccharides into its cytoplasm. Unlike other plant pathogenic species, *C. albifundus* does not contain an array of cell wall-degrading enzymes and, in fact, has less than *Knoxdaviesia* species. Notably, however, several enzymes involved in pectin degradation were identified, suggesting that *C. albifundus* weakens the pectic fraction to penetrate the cell wall. *Ceratocystis albifundus*, therefore, seems to rely predominantly on soluble plant storage polysaccharides for its energy.

In order to evaluate our hypothesis that these species are unable to grow on certain substrates due to ineffective substrate translocation, genomic characterisation of transporters and experimental validation of substrate uptake is necessary. These experiments would specifically be important to determine the mechanism behind the inability of *C. albifundus* to grow on glucose, fructose and sucrose. Further, analysis of the phloem composition of *C. albifundus* hosts could show why *C. albifundus* uses stachyose as a carbon source rather than sucrose.

ACKNOWLEDGEMENTS

We thank the National Research Foundation (NRF) and the Department of Science and Technology (DST)-NRF Centre of Excellence in Tree Health Biotechnology (CTHB) for financial assistance. We are also grateful to Danielle Roodt for providing us with the annotated genome of *Ceratocystis albifundus* isolate CMW17620.

REFERENCES

Addou, S., Rentzsch, R., Lee, D. & Orengo, C. A. (2009). Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *Journal of Molecular Biology* **387**, 416-430.

Atanasova, L. & Druzhinina, I. (2010). Global nutrient profiling by Phenotype MicroArrays: a tool complementing genomic and proteomic studies in conidial fungi. *Journal of Zhejiang University SCIENCE B* **11**, 151-168.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015). *Knoxdaviesia proteae* is not the only *Knoxdaviesia*-symbiont of *Protea repens*. *IMA Fungus* **6**, 471-476.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D. & Wingfield, M. J. (2016). Genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa. *Standards in Genomic Sciences* **11**, 1-7.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Research* **28**, 304-305.

Batista, A. S., Miletti, L. C. & Stambuk, B. U. (2005). Sucrose fermentation by *Saccharomyces cerevisiae* lacking hexose transport. *Journal of Molecular Microbiology and Biotechnology* **8**, 26-33.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research* **41**, D36-D42.

Bergh, N. G., Verboom, G., Rouget, M. & Cowling, R. M. (2014). Vegetation types of the Greater Cape Floristic Region. In *Fynbos: Ecology, Evolution, and Conservation of a Megadiverse Region*, pp. 26-46. Edited by N. Allsopp, J. F. Colville & G. A. Verboom. New York: Oxford University Press

Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances* **30**, 1575-1588.

Brown, R. M. (2004). Cellulose structure and biosynthesis: what is in store for the 21st century? *Journal of Polymer Science Part A: Polymer Chemistry* **42**, 487-495.

Buléon, A., Colonna, P., Planchot, V. & Ball, S. (1998). Starch granules: structure and biosynthesis. *International Journal of Biological Macromolecules* **23**, 85-112.

Caffall, K. H. & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research* **344**, 1879-1900.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S. & Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**, 188-196.

Coutinho, P. M., Andersen, M. R., Kolenova, K., vanKuyk, P. A., Benoit, I., Gruben, B. S., Trejo-Aguilar, B., Visser, H., Solingen, P. v., Pakula, T., *et al.* (2009). Post-genomic insights into the plant polysaccharide degradation potential of *Aspergillus nidulans* and comparison to *Aspergillus niger* and *Aspergillus oryzae*. *Fungal Genetics and Biology* **46**, S161-S169.

Cowling, R. & Mitchell, D. (1981). Sugar composition, total nitrogen and accumulation of C-14 assimilates in floral nectaries of *Protea* species. *Journal of South African Botany* **47**, 743-750.

Crous, P. W., Denman, S., Taylor, J. E., Swart, L., Palm, M. E. & Samson, R. A. (2004). *Cultivation and diseases of Proteaceae: Leucadendron, Leucospermum and Protea*. Utrecht, The Netherlands: CBS Biodiversity Series 2.

Crous, P. W., Summerell, B. A., Shivas, R. G., Burgess, T. I., Decock, C. A., Dreyer, L. L., Granke, L. L., Guest, D. I., Hardy, G. E. S. T. J., Hausbeck, M. K., *et al.* (2012). Fungal Planet description sheets: 107-127. *Persoonia* **28**, 138-182.

De Beer, Z. W., Seifert, K. A. & Wingfield, M. J. (2013). A nomenclature for ophiostomatoid genera and species in the *Ophiostomatales* and *Microascales*. In *Ophiostomatoid fungi: expanding frontiers*, pp. 245-322. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield: CBS Biodiversity Series.

Di, C., Zhang, M., Xu, S., Cheng, T. & An, L. (2006). Role of poly-galacturonase inhibiting protein in plant defense. *Critical Reviews in Microbiology* **32**, 91-100.

Divakar, S. (2013). Glycosidases. In *Enzymatic Transformation*, pp. 5-21. India: Springer.

Engelbrecht, C. J., Harrington, T. C. & Alfenas, A. (2007). *Ceratocystis* wilt of cacao-a disease of increasing importance. *Phytopathology* **97**, 1648-1649.

Holt, C. & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491.

Howell, J. T. & Davis, M. (2005). Plant defense mechanisms against fungal pathogens: polygalacturonase inhibitor proteins. *Canadian Journal of Plant Pathology* **27**, 5-15.

Huang, Q. & Allen, C. (2000). Polygalacturonases are required for rapid colonization and full virulence of *Ralstonia solanacearum* on tomato plants. *Physiological and Molecular Plant Pathology* **57**, 77-83.

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

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32**, D277-D280.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. (2011). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, gkr988.

Kanehisa, M., Sato, Y. & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**, 726-731.

Kelker, N., Hanson, T. & Anderson, R. (1970). Alternate Pathways of D-Fructose Metabolism in *Aerobacter aerogenes*. *Journal of Biological Chemistry* **245**, 2060-2065.

Kolařík, M. & Hulcr, J. (2009). Mycobiota associated with the ambrosia beetle *Scolytodes unipunctatus* (Coleoptera: Curculionidae, Scolytinae). *Mycological Research* **113**, 44-60.

Kötzler, M. P., Hancock, S. M. & Withers, S. G. (2014). Glycosidases: Functions, Families and Folds. In *eLS*. Chichester: John Wiley & Sons, Ltd.

Kubicek, C. P., Starr, T. L. & Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* **52**, 427-451.

Lager, I., Fehr, M., Frommer, W. B. & Lalonde, S. (2003). Development of a fluorescent nanosensor for ribose. *FEBS Letters* **553**, 85-89.

Lee, D.-H., Roux, J., Wingfield, B. D., Barnes, I., Mostert, L. & Wingfield, M. J. (2016). The genetic landscape of *Ceratocystis albifundus* populations in South Africa reveals a recent fungal introduction event. *Fungal Biology* **120**, 690-700.

Lee, S., Roets, F. & Crous, P. W. (2005). Biodiversity of saprobic microfungi associated with the infructescences of *Protea* species in South Africa. *Fungal Diversity* **19**, 69-78.

Lindqvist, R. (2006). Estimation of *Staphylococcus aureus* Growth Parameters from Turbidity Data: Characterization of Strain Variation and Comparison of Methods. *Applied and Environmental Microbiology* **72**, 4862-4870.

Lucas, W. J., Groover, A., Lichtenberger, R., Furuta, K., Yadav, S. R., Helariutta, Y., He, X. Q., Fukuda, H., Kang, J. & Brady, S. M. (2013). The plant vascular system: evolution, development and FunctionsF. *Journal of Integrative Plant Biology* **55**, 294-388.

Manners, D. J. (1991). Recent developments in our understanding of glycogen structure. *Carbohydrate Polymers* **16**, 37-82.

Marais, G. J. & Wingfield, M. J. (1994). Fungi associated with infructescences of *Protea* species in South Africa, including a new species of *Ophiostoma*. *Mycological Research* **98**, 369-374.

Martens-Uzunova, E. S. & Schaap, P. J. (2008). An evolutionary conserved d-galacturonic acid metabolic pathway operates across filamentous fungi capable of pectin degradation. *Fungal Genetics and Biology* **45**, 1449-1457.

Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., Chapman, J., Chertkov, O., Coutinho, P. M. & Cullen, D. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnology* **26**, 553-560.

Mchunu, N. P., Permaul, K., Alam, M. & Singh, S. (2013). Carbon utilization profile of a thermophilic fungus, *Thermomyces lanuginosus* using phenotypic microarray. *Advances in Bioscience and Biotechnology* **4**, 24.

Mittermeier, R. A., Myers, N., Thomsen, J. B., Da Fonseca, G. A. B. & Olivieri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conservation Biology* **12**, 516-520.

Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology* **11**, 266-277.

Moreira, L. & Filho, E. (2008). An overview of mannan structure and mannan-degrading enzyme systems. *Applied Microbiology and Biotechnology* **79**, 165-178.

Morgan-Jones, G. & Sinclair, R. C. (1980). Notes on hyphomycetes. XXXVI a new species of *Custingophora*. *Mycotaxon* **11**, 443-445.

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

Nelson, T. E., Scaletti, J. V., Smith, F. & Kirkwood, S. (1963). The Use of Enzymes in Structural Studies on Polysaccharides: I. The Mode of attack of a β-D-(l→3)-glucanase on Laminarin. *Canadian Journal of Chemistry* **41**, 1671-1678.

Nicolson, S. W. & Van Wyk, B.-E. (1998). Nectar Sugars in Proteaceae: Patterns and Processes. *Australian Journal of Botany* **46**, 489-504.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.

Pinnoi, A., McKenzie, E. H. C., Jones, E. B. G. & D, H. K. (2003). Palm fungi from Thailand: *Custingophora undulatistipes* sp. nov. and *Vanakripa minutiellipsoidea* sp. nov. *Nova Hedwigia* **77**, 213-219.

Pritchard, J. (2007). Solute transport in the Phloem. In *Plant Solute Transport*, pp. 235-274. Edited by A. R. Yeo & T. J. Flowers. Oxford, UK: Blackwell Publishing Ltd.

R Core Team (2014).R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Rebelo, T. (2001). *Proteas: A field guide to the Proteas of Southern Africa*. Vlaeberg, South Africa: Fernwood Press.

Roets, F., Dreyer, L. L. & Crous, P. W. (2005). Seasonal trends in colonisation of *Protea* infructescences by *Gondwanamyces* and *Ophiostoma* spp. *South African Journal of Botany* **71**, 307-311.

Roets, F., Crous, P. W., Wingfield, M. J. & Dreyer, L. L. (2009a). Mite-mediated hyperphoretic dispersal of *Ophiostoma* spp. from the Infructescences of South African *Protea* spp. *Environmental Entomology* **28**, 143-152.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2009b). Fungal radiation in the Cape Floristic Region: An analysis based on *Gondwanamyces* and *Ophiostoma*. *Molecular Phylogenetics and Evolution* **51**, 111-119.

Roets, F., Theron, N., Wingfield, M. J. & Dreyer, L. L. (2011a). Biotic and abiotic constraints that facilitate host exclusivity of *Gondwanamyces* and *Ophiostoma* on *Protea*. *Fungal Biology* **116**, 49-61.

Roets, F., Wingfield, M. J., Wingfield, B. D. & Dreyer, L. L. (2011b). Mites are the most common vectors of the fungus *Gondwanamyces proteae* in *Protea* infructescences. *Fungal Biology* **115**, 343-350.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2013). Taxonomy and ecology of ophiostomatoid fungi associated with *Protea* infructescences. In *Ophiostomatoid fungi: expanding frontiers*, pp. 177-187. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield. Utrecht, The Netherlands: CBS Biodiversity Series.

Roper, M. C., Greve, L. C., Warren, J. G., Labavitch, J. M. & Kirkpatrick, B. C. (2007). *Xylella fastidiosa* requires polygalacturonase for colonization and pathogenicity in *Vitis vinifera* grapevines. *Molecular Plant-Microbe Interactions* **20**, 411-419.

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering* **12**, 85-94.

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

Roux, J., Dunlop, R. & Wingfield, M. J. (1999b). Susceptibility of elite *Acacia mearnsii* families to Ceratocystis wilt in south Africa. *Journal of Forest Research* **4**, 187-190.

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

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

Roux, J. & Wingfield, M. J. (2009). *Ceratocystis* species: emerging pathogens of non-native plantation *Eucalyptus* and *Acacia* species. *Southern Forests* **71**, 115-120.

Roux, J. & Wingfield, M. (2013). *Ceratocystis* species on the African continent, with particular reference to *C. albifundus*, an African species in the *C. fimbriata sensu lato* species complex. In *The Ophiostomatoid Fungi: Expanding Frontiers*, pp. 131-138. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield: CBS Biodiversity Series.

Rowe, H. C. & Kliebenstein, D. J. (2007). Elevated genetic variation within virulence-associated *Botrytis cinerea* polygalacturonase loci. *Molecular Plant-Microbe Interactions* **20**, 1126-1137.

Saier, M. H. (2000). Families of transmembrane sugar transport proteins. *Molecular Microbiology* **35**, 699-710.

Sapico, V., Hanson, T., Walter, R. & Anderson, R. (1968). Metabolism of D-Fructose in *Aerobacter aerogenes*: analysis of mutants lacking D-Fructose 6-phosphate kinase and D-Fructose 1, 6-diphosphatase. *Journal of Bacteriology* **96**, 51-54.

Scheller, H. V. & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology* **61**, 263-289.

Schols, H. A., Bakx, E. J., Schipper, D. & Voragen, A. G. (1995). A xylogalacturonan subunit present in the modified hairy regions of apple pectin. *Carbohydrate Research* **279**, 265-279.

Sellick, C. A., Campbell, R. N. & Reece, R. J. (2008). Galactose metabolism in yeast—structure and regulation of the Leloir pathway enzymes and the genes encoding them. *International Review of Cell and Molecular Biology* **269**, 111-150.

Standaert, A. R., Francois, K., Devlieghere, F., Debevere, J., Van Impe, J. F. & Geeraerd, A. H. (2007). Modeling individual cell lag time distributions for *Listeria monocytogenes*. *Risk Analysis* **27**, 241-254.

Staudacher, E., Altmann, F., Wilson, I. B. H. & März, L. (1999). Fucose in N-glycans: from plant to man. *Biochimica et Biophysica Acta* **1473**, 216-236.

Straub, C. S., Ives, A. R. & Gratton, C. (2011). Evidence for a trade-off between host-range breadth and host-use efficiency in aphid parasitoids. *The American Naturalist* **177**, 389-395.

Sutton, D. & Lampen, J. (1962). Localization of sucrose and maltose fermenting systems in *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta* **56**, 303-312.

Taiz, L. & Zeiger, E. (2006). Translocation in the Phloem. In *Plant Physiology, 4th edn.*, pp. 221-254. Sunderland, Mass.: Sinauer Associates.

Tappy, L. & Lê, K.-A. (2010). Metabolic effects of fructose and the worldwide increase in obesity. *Physiological Reviews* **90**, 23-46.

Teeri, T. T. (1997). Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends in Biotechnology* **15**, 160-167.

Uzunovic, A., Yang, D.-Q., Gagné, P., Breuil, C., Bernier, L., Byrne, A., Gignac, M. & Kim, S. (1999). Fungi that cause sapstain in Canadian softwoods. *Canadian Journal of Microbiology* **45**, 914-922.

Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M. & Klenk, H.-P. (2012). Visualization and Curve-Parameter Estimation Strategies for Efficient Exploration of Phenotype Microarray Kinetics. *PLOS ONE* **7**, e34846.

Vaas, L. A. I., Sikorski, J., Hofner, B., Fiebig, A., Buddruhs, N., Klenk, H.-P. & Göker, M. (2013). opm: an R package for analysing OmniLog® phenotype microarray data. *Bioinformatics* **29**, 1823-1824.

Valente, L. M., Reeves, G., Schnitzler, J., Mason, I. P., Fay, M. F., Rebelo, T. G., Chase, M. W. & Barraclough, T. G. (2010). Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* **64**, 745-760.

Valentine, B. & Bainbridge, B. (1978). The relevance of a study of a temperature-sensitive ballooning mutant of *Aspergillus nidulans* defective in mannose metabolism to our understanding of mannose as a wall component and carbon/energy source. *Microbiology* **109**, 155-168.

Van den Brink, J. & de Vries, R. P. (2011). Fungal enzyme sets for plant polysaccharide degradation. *Applied Microbiology and Biotechnology* **91**, 1477-1492.

Van der Linde, J. A., Six, D. L., Wingfield, M. J. & Roux, J. (2012). New species of *Gondwanamyces* from dying *Euphorbia* trees in South Africa. *Mycologia* **104**, 574-584.

Van der Nest, M., Steenkamp, E., McTaggart, A., Trollip, C., Godlonton, T., Sauerman, E., Roodt, D., Naidoo, K., Coetzee, M. & Wilken, P. (2015). Saprophytic and pathogenic fungi in the Ceratocystidaceae differ in their ability to metabolize plant-derived sucrose. *BMC Evolutionary Biology* **15**, 1.

Van der Nest, M. A., Beirn, L. A., Crouch, J. A., Demers, J. E., de Beer, Z. W., De Vos, L., Gordon, T. R., Moncalvo, J. M., Naidoo, K., Sanchez-Ramirez, S., *et al.* (2014). IMA Genome-F

3: Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntiella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana*, and *Sclerotinia echinophila*. *IMA Fungus* **5**, 473-486.

Van Wilgen, B. W. (2009). The evolution of fire and invasive alien plant management practices in fynbos. *South African Journal of Science* **105**, 335-342.

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

Vehkala, M., Shubin, M., Connor, T. R., Thomson, N. R. & Corander, J. (2015). Novel R Pipeline for Analyzing Biolog Phenotypic Microarray Data. *PLOS ONE* **10**, e0118392.

Verduyn, C., Van Kleef, R., Frank, J., Schreuder, H., Van Dijken, J. & Scheffers, W. (1985). Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast *Pichia stipitis*. *Biochemical Journal* **226**, 669-677.

Wiens, D., Rourke, J. P., Casper, B. B., Rickart, E. A., LaPine, T. R., Peterson, C. J. & Channing, A. (1983). Nonflying Mammal Pollination of Southern African Proteas: A Non-Coevolved System. *Annals of the Missouri Botanical Garden* **70**, 1-31.

Wilson, D. S. & Yoshimura, J. (1994). On the Coexistence of Specialists and Generalists. *The American Naturalist* **144**, 692-707.

Wingfield, B. D., Viljoen, C. D. & Wingfield, M. J. (1999). Phylogenetic relationships of ophiostomatoid fungi associated with *Protea* infructescences in South Africa. *Mycological Research* **103**, 1616-1620.

Wingfield, M. J., Wyk, P. S. V. & Marasas, W. F. O. (1988). *Ceratocystiopsis proteae* sp. *nov*., with a new anamorph genus. *Mycologia* **80**, 23-30.

Wingfield, M. J. & Van Wyk, P. S. (1993). A new species of *Ophiostoma* from *Protea* infructescences in South Africa. *Mycological Research* **97**, 709-716.

Wingfield, M. J., De Beer, C., Visser, C. & Wingfield, B. D. (1996). A New *Ceratocystis* Species Defined Using Morphological and Ribosomal DNA Sequence Comparisons. *Systematic and Applied Microbiology* **19**, 191-202.

Wong, D. (2008). Enzymatic deconstruction of backbone structures of the ramified regions in pectins. *The Protein Journal* **27**, 30-42.

Zhao, Z., Liu, H., Wang, C. & Xu, J.-R. (2014). Erratum to: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **15**, 1-15.

SUPPLEMENTARY INFORMATION

The following supplementary information has been provided electronically:

Figure S5.1 Growth curves of *Knoxdaviesia capensis*, *K. proteae* and *C. albifundus* on all phenotype microarray substrates (E - F), active substrates (G - L) and shared substrates (M - P).

Table S5.1 Carbon sources contained in the BioLog PM1 and PM2 plates.

Table S5.2 Proportion of *Knoxdaviesia capensis*, *K. proteae* and *Ceratocystis albifundus* proteins predicted to belong to each KEGG category.

Table S5.3 Carbon-metabolising enzymes (excluding polysaccharide-degrading enzymes) identified only in *Knoxdaviesia* or *Ceratocystis albifundus*, respectively.

CHAPTER 6

Genome divergence explains differences in host range and competitive abilities of closely related fungi

ABSTRACT

Biodiversity hotspots are commonly defined by the numbers of macro-species that they accommodate. Ironically, the microbial biodiversity and symbioses are most often poorly studied or even ignored. In this regard, an intriguing plant-fungal symbiosis in the Core Cape Subregion biodiversity hotspot has received some focussed attention. *Knoxdaviesia capensis* and *K. proteae* are saprotrophic fungi found exclusively in the flower heads of *Protea* plants and they have complex relationships with insects and mites. The aim of this study was to determine whether the genomes of these fungi might reflect differences in their host ranges and competitive abilities. The genomes of *K. capensis* and *K. proteae* displayed macrosynteny (89.45 % identity) with approximately 12 % repeat content. The different composition of retrotransposons between the two species, however, suggests different evolutionary trajectories. Both *Knoxdaviesia* species had fewer secreted proteins than pathogens and other saprotrophs, reflecting their specialised habitat. The bulk of the predicted secretome coded for carbohydrate metabolism proteins, with the broad host-range *K. capensis* coding for more than double the number of unique carbohydrate-degrading proteins compared to species-specific *K. proteae*. Very few effector-like secreted proteins were identified, indicating that these species are likely unable to attack the living tissues of their hosts. These fungi had fewer secondary metabolite cluster genes than those known from other filamentous fungi, suggesting that competition with other microbes is minimal and may be controlled by antibiotic-producing actinomycete bacteria common in this niche. Genes for secreted proteins associated with detoxification processes and siderophore-synthesising clusters identified from their genomes most likely enable these *Knoxdaviesia* species to tolerate antifungal compounds and effectively compete for resources, thereby facilitating their unusual dominance in *Protea* flower heads.

INTRODUCTION

Microbial diversity and the role of microbes in ecosystems are, in many cases, poorly understood. In an ecosystem, the number of macro-species, specifically plants and vertebrates, typically form the basis for representing biodiversity (Myers *et al.*, 2000). The biodiversity hotspot concept, however, fails to consider microbial diversity (Cowan *et al.*, 2013; Marchese, 2015), an aspect that likely exceeds the diversity of macro-species by several orders of magnitude. For example, massive fungal diversity (more than 170 000 species) is estimated for South Africa, of which many species in biodiversity hotspots are likely to be endemic (Crous *et al.*, 2006). In order to understand the role of microbes in ecosystems, it is necessary to investigate this understudied "below surface" component of biodiversity.

One example of a well-studied microbial interaction with plants is the symbiotic relationship between ophiostomatoid fungi and *Protea* L. plants. This interaction occurs in the biodiversity hotspot known as the Core Cape Subregion (CCR) at the southern tip of Africa and has been the subject of numerous recent studies. Within the flower heads of serotinous *Protea* species, two genera of ophiostomatoid fungi dominate (Lee *et al.*, 2005; Marais & Wingfield, 1994; Seifert *et al.*, 2013; Wingfield & Van Wyk, 1993). These fungi are primarily dispersed by mites that are abundant in *Protea* flower heads (Roets *et al.*, 2007; 2011b). They are not associated with obvious disease in their hosts, but their status as harmless *Protea* symbionts remains untested. Personal observations suggest discolouration of the *Protea* style in regions of ophiostomatoid fungal growth on the young flowers. Many *Protea* species also yield a characteristically low proportion (< 30 %) of fertilized seeds (Coetzee & Giliomee, 1985; Rebelo & Rourke, 1985), for which there is no universal explanation (Mustart *et al.*, 1995). Many non-*Protea*-associated ophiostomatoid fungi are economically important pathogens that impact negatively on the forestry, agricultural and medical industries. Of the two *Protea*-associated lineages, members of *Sporothrix* Hektoen & C.F. Perkins *emend.* Z.W. de Beer, T.A. Duong & M.J. Wingf. are related to several plant and animal pathogens (Brasier, 1991; Christiansen & Solheim, 1990; De Beer *et al.*, 2016), whereas non-*Protea Knoxdaviesia* M.J. Wingf., P.S. van Wyk & Marasas species have been isolated from diseased plant tissue (Van der Linde *et al.*, 2012), although direct involvement with disease has not been established.

The *Protea* niche occupied by ophiostomatoid fungi is an inherently competitive one in which these fungi are apparently successful contenders (Lee *et al.*, 2005). These fungi are dispersed as ascospores that attach to their mite vectors (Roets *et al.*, 2007; 2011b) and are subsequently deposited in young *Protea* flower heads where they germinate and grow asexually. After fertilization, the flower heads mature and develop into seed storage organs known as infructescences. The closed infructescence is moist and protected from external factors such as wind and rain, providing an ideal micro-habitat for microorganisms, arthropods and nematodes (Coetzee & Giliomee, 1985; Human, 2013; Lee *et al.*, 2003; 2005; Roets *et al.*, 2005; 2006b; Theron *et al.*, 2012; Zwölfer, 1979). The arthropod and, potentially, the nematode colonizers in infructescences feed on ophiostomatoid fungi (Roets *et al.*, 2013; Ruess & Lussenhop, 2005). Fungi are, however, not only passively consumed, but may upregulate secondary metabolite biosynthesis and induce sexual reproduction to resist fungivory (Döll *et al.*, 2013). Additionally, at least 22 species of saprotrophic fungi other than ophiostomatoid symbionts occupy *Protea* infructescences (Lee *et al.*, 2005), generating further inter-organismal competition.

Genomics can provide useful opportunities to assess the pathogenicity and competitive ability of species. Most plant pathogenic fungi secrete an array of enzymes to degrade plant cell walls (Kubicek *et al.*, 2014) as well as numerous effector proteins that play a role in virulence and may suppress or elicit plant defence responses (Schneider & Collmer, 2010). Low molecular weight proteins, rich in the amino acid cysteine and often uncharacterised, have been particularly implicated in pathogen-host interactions (Lu & Edwards, 2016; Templeton *et al.*, 1994). Although functional prediction may be lacking, genomics has enabled secondary metabolite biosynthesis clusters to be identified from sequence data (Keller *et al.*, 2005) and potential cell defence proteins can be elucidated by considering the functional domains of predicted proteins. Genome predictions, therefore, have value for evaluating potential pathogenicity and tolerance to inter-organismal competition.

Two *Knoxdaviesia* species colonize the flower heads of *Protea* species in the CCR. *Knoxdaviesia proteae* M.J. Wingf., P.S. van Wyk & Marasas is thought to have evolved from a common *Protea*-associated ancestor that specialised on a single *Protea* host, consequently increasing its fitness (Chapter 5). In contrast, *K. capensis* M.J. Wingf. & P.S. van Wyk occurs in multiple *Protea* hosts (Roets *et al.*, 2009; Wingfield & Van Wyk, 1993) and, therefore, encounters a variety of host chemistries (Roets *et al.*, 2011a), reflected in its ability to grow on a wider range of substrates than *K. proteae* (Chapter 5). Both *K. capensis* and *K. proteae*

are, however, specialists in that they occupy a niche outside of which they have never been observed (Roets *et al.*, 2005; 2009). All available evidence shows that these putative saprotrophic species are obligate *Protea*-associates. Consequently, their biomass-degrading enzymes (Chapter 5), and likely other aspects of their lifestyle, have been reduced to the extent that only *Protea* species are suitable hosts. It is consequently plausible that their different host ranges and their specialisation on a single host genus will be reflected in their genomes.

In this study, we compare the publically available genomes of *K. capensis* and *K. proteae* (Aylward *et al.*, 2016a, Chapter 2) to investigate whether these reflect differences in their host ranges and their potential to contend with inter-organismal competition. *Knoxdaviesia capensis* and *K. proteae* are phylogenetically very closely related (Wingfield *et al.*, 1999) and previous genetic studies (Aylward *et al.*, 2016b, Chapters 4 & 5) have revealed few differences between them. Consequently, their genomes should display macrosynteny. In contrast, the distribution of transposable elements in these genomes may not be congruent, since repetitive sequences often play a role in divergence (Castanera *et al.*, 2016). Due to their confined niche, both species should have smaller secretomes than other generalist saprotrophs and fewer secreted protein effectors than plant pathogens. This reduction could, however, be more pronounced in *K. proteae*, since it has specialised on *P. repens* and presumably diverged from the broad-host range *K. capensis* due to restricted host association (Chapter 5). With its broad host-range, *K. capensis* must interact with multiple hosts and, therefore, may also encounter a greater range of competitive organisms within host infructescences. We, therefore, hypothesise that the secondary metabolite biosynthesis capability of these *Protea*-associated *Knoxdaviesia* species would be adapted to contending with microbial and animal competitors and that *K. capensis* requires a larger arsenal of secondary metabolites than *K. proteae*.

MATERIALS AND METHODS

Knoxdaviesia genomes and whole genome alignment

The genomes of *K. capensis* CBS139037 (LNGK00000000.1) and *K. proteae* CBS140089 (LNGL00000000.1), as well as their predicted proteins were utilised in this study. Both genomes were previously annotated (Aylward *et al.*, 2016a) with MAKER (Cantarel *et al.*, 2008; Holt & Yandell, 2011). The 29 scaffolds of the *K. capensis* genome and the 133

scaffolds of the *K. proteae* genome were masked with RepeatScout 1.0.5 (Price *et al.*, 2005) and aligned with the MUMmer 3.23 package (Kurtz *et al.*, 2004). Whole genome nucleotide alignments were performed with nucmer, using unique matches of at least 50 bases in both the reference (*K. capensis*) and query (*K. proteae*) sequences ("--mum" option). Alignments were filtered with delta-filter to remove repetitive matches, allowing for sequence rearrangements ("-1" option). A dotplot of the filtered matches was generated with mummerplot, using the "-R" and "-Q" options to manually reorder the scaffolds. Genome coordinates of the filtered matches were obtained with the show-coords script. Alignment statistics were computed with DNAdiff 1.3, distributed with MUMmer.

Repeats, transposable elements and repeat-induced point mutation (RIP)

To investigate whether transposable elements could have contributed to genome evolution and plasticity (Castanera *et al.*, 2016) in *Knoxdaviesia*, families of these elements were identified and classified with TEdenovo and annotated with TEannot that are both part of the REPET v2.2 package (Gilgado *et al.*, 2005). To accurately ascertain the total percentage of the genome taken up by transposable elements, the annotation files were filtered in Microsoft Office Excel 2010 (Microsoft Corp., Redmond, WA, USA) to remove duplicated regions where transposable element annotations overlap. Low complexity (tandem) repeats were identified with RepeatScout 1.0.5 (Price *et al.*, 2005) and mapped to the genome using RepeatMasker version open-4.0.5 (Smit *et al.*, 2013-2015). Telomere repeats were identified according to Fulnečková *et al.* (2013).

The genome sequences of the two *Knoxdaviesia* species were investigated for evidence of repeat-induced point mutation (RIP), a mechanism that limits the spread and diversification of transposable elements (Slotkin & Martienssen, 2007). Because RIP targets C-G nucleotides in repetitive regions and converts them to T-A (Cambareri *et al.*, 1989), RIP-affected regions will be depleted of C-G nucleotides and will have high (≥ 1.1) RIP product index (TpA / ApT) ratios (Margolin *et al.*, 1998; Selker *et al.*, 2003). Conversely, the RIP substrate index (CpA + TpG / ApC + GpT) ratios of these regions will be low (< 0.9; Margolin *et al.*, 1998; Selker *et al.*, 2003). A Perl script "RIP_index_calculation.pl" (available from https://github.com/hyphaltip/fungaltools/tree/master/scripts) was used to calculate the composite RIP index (substrate index subtracted from the product index) in 500

base windows, sliding the window 100 bases at a time. A positive composite RIP index indicates a RIP-affected region (Lewis *et al.*, 2009).

Prediction of the *Knoxdaviesia* secretome

The putative secreted proteins of *K. capensis* and *K. proteae* were predicted following the protocol of Brown *et al.* (2012). Proteins with a secretion signal were identified with SignalP 4.1 (Petersen *et al.*, 2011) and TargetP 1.1b (Emanuelsson *et al.*, 2000). These were filtered by discarding proteins with transmembrane domains identified by TMHMM 2.0 (Krogh *et al.*, 2001). All proteins with a secretion signal were interrogated for a glycosylphosphatidylinositol (GPI)-anchor site with PredGPI (http://gpcr2.biocomp.unibo.it/predgpi/, accessed 14 July 2016) (Pierleoni *et al.*, 2008). The secretome was subsequently refined by excluding sequences lacking an initial methionine residue and GPI-anchor proteins (proteins with a "highly probable" ($P \leq 0.001$) or "probable" ($P \leq 0.005$) GPI-anchor site). Protein location was predicted with ProtComp 9.0 (http://www.softberry.com), keeping proteins with extracellular and unknown localizations for further refinement. The localization of these remaining sequences was finally predicted with WoLFPSort 0.2 (https://github.com/fmaguire/WoLFPSort; Horton *et al.*, 2007). Sequences with an extracellular score above 17 were maintained as the final dataset of proteins with a high likelihood of being secreted ("high likelihood secretome dataset"; Brown *et al.*, 2012). Sequences that scored ≤ 17, but for which the extracellular environment was still the most probable localization, were maintained as proteins with a lower likelihood of secretion ("low likelihood secretome dataset").

The final dataset of secreted proteins was annotated by conducting BLASTp searches (Camacho *et al.*, 2009) against the NCBI (National Center for Biotechnology Information; https://www.ncbi.nlm.nih.gov) non-redundant protein database and the Pathogen Host Interactions (PHI) database version 4.1 (Urban *et al.*, 2015; Winnenburg *et al.*, 2006). Reversed Position Specific (RPS)-BLAST searches were also conducted against the Pfam (protein family) database (Finn *et al.*, 2016) to identify conserved protein domains. Blast2GO 3.3.5 (Götz *et al.*, 2008) was used to parse the results of the BLAST searches and add InterPro (Zdobnov & Apweiler, 2001), Gene Ontology (Gene Ontology Consortium, 2004) and enzyme code (Kanehisa *et al.*, 2004) identifiers to the annotations. An enrichment analysis was also performed in Blast2GO to investigate potential functional enrichment in

either *K. capensis* or *K. proteae*. Since many plant pathogen effectors have been described as cysteine-rich proteins (Lu & Edwards, 2016; Rep, 2005), the cysteine content of the secreted proteins was determined from the amino acid composition statistics of each protein as calculated by the EMBOSS PEPSTATS tool (Rice *et al.*, 2000).

OrthoMCL 2.0.9 (Li *et al.*, 2003) was used to determine which *K. capensis* and *K. proteae* secreted proteins are likely to be functionally similar and those that are unique. This program clusters proteins according to sequence similarity, grouping together orthologous proteins (derived from the same ancestor) and recent paralogs (generated by duplication, but likely retaining functional similarity; Li *et al.*, 2003). OrthoMCL was independently applied to the high and low likelihood secretome datasets. Proteins without identified orthologues were extracted and the clustering procedure was reapplied to compare these "unique" proteins across the two datasets. A secreted protein was only considered truly unique where it failed to have an ortholog in both the high and the low likelihood dataset. Ortholog groups and unique proteins were subsequently grouped into categories according to the Functional Catalogue (FunCat) annotation scheme (Ruepp *et al.*, 2004).

Secondary metabolite clusters

Secondary metabolite (SM) clusters were predicted with the online tool antiSMASH 3.0.5 (Blin *et al.*, 2013; Medema *et al.*, 2011; Weber *et al.*, 2015) using the whole genome sequences as input. Genomic regions in which SM clusters had been predicted were investigated by extracting the protein predictions of the putative biosynthetic genes as well as approximately five upstream and five downstream flanking genes. These were subsequently annotated with BLASTp and Blast2GO as described above. The presence of the necessary catalytic domains in the backbone non-ribosomal peptide synthase (NRPS) and polyketide synthase (PKS) proteins (Keller *et al.*, 2005) were verified by searching for protein family (Pfam; Finn *et al.*, 2016) domains with HMMER 2.3.2 (Eddy, 2011).

RESULTS

Knoxdaviesia whole genome alignment

More than 85 % of the two *Knoxdaviesia* genomes were aligned to each other with an average nucleotide identity of 89.5 % (Table 6.1). Ninety-eight (73.7 %) of the *K. proteae* scaffolds were aligned to 24 (82.8 %) *K. capensis* scaffolds, with a mean alignment length of approximately 3702 bases (Fig. 6.1). Several large inversions were apparent in the alignments. The unaligned scaffolds were small (< 26 kb) and had a low percentage (< 40 %) of GC nucleotides. Uninterrupted diagonals of ≥ 20 kb in MUMmer genome dotplots have been proposed to constitute macrosynteny (Hane *et al.*, 2011). Considering that all scaffolds in Fig 6.1B are > 500 kb in length, these lengthy diagonals are indicative of macrosynteny between the two *Knoxdaviesia* genomes.

Table 6.1 Whole genome nucleotide alignment statistics for *Knoxdaviesia capensis* and *K. proteae*.[1]

	K. capensis	***K. proteae***
Total scaffolds	29	133
Aligned scaffolds[2]	24 (82.76%)	98 (73.68%)
Unaligned scaffolds	5 (17.24%)	35 (26.32%)
Total Bases	35537816	35489142
Aligned Bases[2]	30327269 (85.34%)	30341562 (85.50%)
Unaligned Bases	5210547 (14.66%)	5147580 (14.50%)
1-to-1 alignments	8198	8198
Total Length (bases)	30342580	30352057
Mean Length (bases)	3701.22	3702.37
Mean Identity	89.45%	89.45%

[1] Statistics were computed with DNAdiff (Kurtz *et al.*, 2004)

[2] Nucleotide alignments were computed with nucmer (Kurtz *et al.*, 2004)

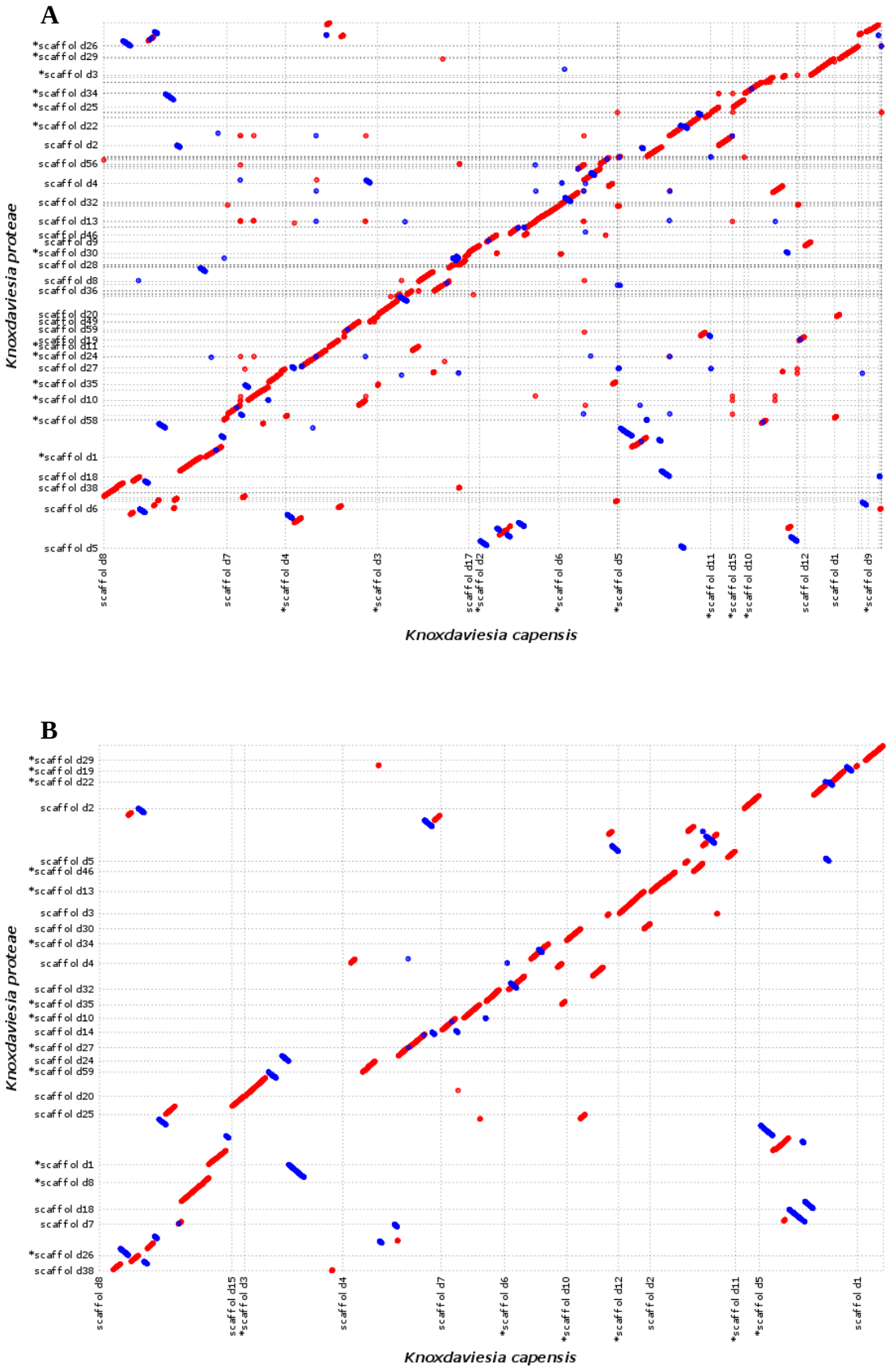

Figure 6.1 Dotplots representing nucleotide alignments of the *Knoxdaviesia capensis* and *K. proteae* genomes. The alignment of all scaffolds (A) as well as scaffolds > 500 kb (B) are

shown. Vertical grey bars indicate the start of *K. capensis* scaffolds and horizontal bars the start of *K. proteae* scaffolds. Only the names of scaffolds > 400 kb are annotated on the axes. Red diagonals represent alignments in the same direction, whereas blue diagonals indicate a reverse orientation in one of the species.

Repeats, transposable elements and repeat-induced point mutation (RIP)

REPET and RepeatScout identified 11.2 % of the *K. capensis* genome and 12.1 % of the *K. proteae* genome as repetitive (Table 6.2). Class I retrotransposons, specifically Long Terminal Repeat (LTR) transposons, constituted the bulk of the repetitive content (> 8 %) in both genomes. Low complexity (short tandem) repeats occupied approximately 3 % of each genome, whereas Class II DNA transposons were rare, comprising only up to 0.2 %. Since the transposable element content between different strains of a fungal species may vary (Castanera *et al.*, 2016), the slightly greater proportion of repeats in the *K. proteae* genome was not indicative of significant differences between *K. capensis* and *K. proteae*. However, *K. proteae* had a greater diversity of repeat families and 1803 additional individual transposable elements than *K. capensis*. Although the additional transposable element copies could be partly explained by the fragmented nature of the *K. proteae* genome assembly, increased numbers of the non-autonomous Class I elements, large retrotransposon derivatives (LARD) and terminal repeat retrotransposons in miniature (TRIM), were especially apparent in this species (Table 6.2). The increase in LARD and TRIM elements was reflected in that they occupy more than double the number of sequence bases in *K. proteae* compared to *K. capensis*. The Class I order DIRS that made up 0.23 % of the *K. capensis* genome were, however, not identified in *K. proteae*.

Table 6.2 Repetitive elements identified in the genomes of *Knoxdaviesia capensis* and *K. proteae*.

	K. capensis				***K. proteae***			
Repeats[1]	**Repeat families**[2]	**Copies in genome**[3]	**Length occupied (bases)**[3]	**% of genome**	**Repeat families**[2]	**Copies in genome**[3]	**Length occupied (bases)**[3]	**% of genome**
Class I elements								
DIRS	1	35	82141	0.23	0	-	-	-
LTR *(Gypsy & Copia)*	57	730	2490438	7.01	72	883	2602486	7.33
LTR *(LARD)**	3	86	162940	0.46	2	801	386806	1.09
LTR *(TRIM)**	5	72	40929	0.12	12	296	139610	0.39
Unknown	1	48	102632	0.29	1	1	1192	0.00
Class I Total	*67*	*971*	*2879080*	*8.10*	*87*	*1981*	*3130094*	*8.82*
Class II elements								
TIR (*MITE*)*	2	23	8417	0.02	1	12	4150	0.01
TIR (*Tc1-Mariner*)	3	70	53242	0.15	6	61	54498	0.15
TIR (*unknown*)	1	12	10506	0.03	4	15	6880	0.02
Class II Total	*6*	*105*	*72165*	*0.20*	*11*	*88*	*65528*	*0.18*
Low complexity repeats[4]	-	24894	1042985	2.93	-	25704	1089080	3.07
TOTAL REPEATS	*73*	*25970*	*3994230*	*11.23*	*98*	*27773*	*4284702*	*12.07*

[1] DIRS = *Dictyostelium* intermediate repeat sequence; LTR = long terminal repeat; LARD = large retrotransposon derivative; TRIM = terminal repeat retrotransposons in miniature; TIR = terminal inverted repeat; MITE = miniature inverted-repeat transposable element; asterisks (*) indicate non-autonomous elements

[2] Identified by TEdenovo package in REPET and clustered with NCBI Blastclust (Gilgado *et al.*, 2005)

[3] Identified by TEannot package in REPET (Gilgado *et al.*, 2005)

[4] Identified with RepeatScout (Price *et al.*, 2005) and annotated with RepeatMasker (Smit *et al.*, 2013-2015)

The proportion of each genome affected by RIP was summarised by considering the number of sliding windows with a positive RIP composite index and dividing them by the total number of sliding windows. Following this approach, 44.5 % of the entire *K. capensis* genome and 42.7 % of the entire *K. proteae* genome was affected by RIP. Many of the smaller sized scaffolds, however, had positive RIP scores over their entire length (Table S6.1), indicating that they were composed of repetitive sequences and this explained why they could not be integrated into larger scaffolds. All the scaffolds of *K. capensis* that could not be aligned to *K. proteae* were affected by RIP across their entire length, whereas 34 of the 35 unaligned *K. proteae* scaffolds had > 80 % RIP. Across the largest scaffolds (> 100 kb), considerably fewer regions were affected by RIP; 16.56 % in *K. capensis* and 13.91 % in *K. proteae*, and thus only slightly more than the repetitive proportion of each genome.

The *K. capensis* and *K. proteae* genomes did not contain any of the eukaryotic telomere repeats that have been described from animals, plants or fungi (Fulnečková *et al.*, 2013; Traut *et al.*, 2007). However, a minisatellite comprised of 10 nucleotides was identified at one terminal of seven *K. capensis* scaffolds and 10 of *K. proteae* (Table S6.2). This minisatellite had the sequence TTAGGGTTAC and is, therefore, a longer variant of the TTAGGG "vertebrate" telomere repeat (Fulnečková *et al.*, 2013; Traut *et al.*, 2007). In *K. capensis*, TTAGGGTTAC telomere repeats were also identified within a transposable element on scaffold 3 and approximately 3 kb upstream of the 3' terminal in scaffold 20 (Table S6.2). The latter may suggest that the 3' end of scaffold 20 was assembled incorrectly.

The *Knoxdaviesia* secretome

A maximum of 4.2 % and 4.0 % of the total putative proteins of *K. capensis* and *K. proteae*, respectively, were predicted to be secreted (Table 6.3). Initial identification of protein signal peptides by SignalP and TargetP identified more than 1 000 potentially secreted proteins in each of the *Knoxdaviesia* species. By excluding proteins that contain transmembrane domains, GPI-anchor sites and proteins that do not start with a methionine residue, this number was reduced to less than 600 for each species (Table 6.3). Analysis of protein localization indicated < 400 proteins from each species that were likely secreted to the outside. For *K. capensis*, WolfPSort predicted 248 proteins that have a high likelihood of being secreted (score > 17) and 108 proteins with lower likelihoods of being secreted. Slightly fewer secreted proteins were predicted in the *K. proteae* "high likelihood" secretome

dataset (229), whereas 125 *K. proteae* proteins have lower likelihoods of secretion. Annotation of these putative secreted proteins revealed classifications in both the high and low likelihood datasets that are not secreted, such as protein translation machinery and transmembrane transport proteins. Based on annotation, a further 25 proteins were excluded from the secretomes of both *K. capensis* and *K. proteae*.

Table 6.3 Number of secreted proteins predicted in *Knoxdaviesia capensis* and *K. proteae*.[1]

	K. capensis	***K. proteae***
Putative secretion signals	1163	1166
Excluded proteins		
GPI-anchor, TM region, no Met	586	583
Localization not extracellular	221	229
Annotation not extracellular	25	25
Refined secretome		
High likelihood[2]	234	215
Low likelihood[3]	97	114
Proportion of total predicted proteins	4.17%	4.02%

[1] Prediction methods are outlined in the text
[2] WolfPSort score > 17
[3] WolfPSort score ≤ 17

Functional annotation of secretome proteins

More than 90 % of the predicted secreted proteins in the high likelihood datasets could be annotated, whereas the lower likelihood datasets contained > 10 % unannotated proteins. No significant enrichment of functional annotations between the *K. capensis* and *K. proteae* datasets could be detected in Blast2GO at a critical value of 0.05. OrthoMCL identified 176 ortholog groups in the high likelihood dataset and 47 ortholog groups in the low likelihood dataset. An additional 26 ortholog groups were identified by comparing the remainder of the high and low likelihood datasets with each other. Considering the total secretome of both *K. capensis* and *K. proteae*, almost 75 % of the proteins are, therefore, present in both species, whereas the remaining 25 % occurs in only one of the two.

The shared set of secreted proteins between *K. capensis* and *K. proteae* consisted of 249 orthologous groups. Based on their Blast2GO annotations and orthologous groups, the secreted proteins were divided into 10 of the "top level" FunCat hierarchical categories (Table 6.4 and S6.3), including unclassified proteins and those with an unclear function. Most

proteins in both the high (> 55 %) and low likelihood (> 36 %) datasets were associated with metabolism, including amino acid, carbohydrate and lipid metabolism. Of these, metabolism of carbon compounds was by far the largest category. The second largest category comprised proteolytic enzymes, indicating that > 60 % of all *K. capensis* and *K. proteae* secreted proteins are dedicated to degrading organic substrates, such as plant cell wall and plasma membrane components.

The largest category of secreted proteins unique to one of the *Knoxdaviesia* species was metabolism, specifically carbohydrate metabolism (Table S6.3). Twenty unique carbohydrate metabolism proteins could be identified from *K. capensis*, whereas *K. proteae* had had only 13. *Knoxdaviesia capensis* also had three unique enzymes with carbohydrate binding domains and double the number of unique proteolytic enzymes than those identified in *K. proteae* (six vs. three). In several cases, the annotations of the unique proteins were similar to proteins in the shared secretome (*e.g.* glycoside hydrolase family 3), but OrthoMCL did not classify them as orthologs or recent paralogs, suggesting functional divergence. Seventeen such cases were detected in *K. capensis* and *K. proteae*, likely representing paralogs that have evolved different substrate specificities. Many unique unclassified or poorly classified proteins were present in the low likelihood datasets, specifically in *K. proteae* in which 33 such proteins occurred.

Table 6.4 Functional categories present in the secretome datasets of *Knoxdaviesia capensis* and *K. proteae*.

FunCat Classification[1]	High Secretion Likelihood Dataset				Low Secretion Likelihood Dataset			
	K. capensis		*K. proteae*		*K. capensis*		*K. proteae*	
	Total (%)	Unique (%)	Total (%)	Unique (%)	Total (%)	Unique (%)	Total (%)	Unique (%)
01 METABOLISM	139 (59.4%)	28 (62.2%)	120 (55.8%)	12 (42.9%)	41 (42.3%)	13 (33.3%)	42 (36.8%)	18 (31.0%)
14 PROTEIN FATE	32 (13.7%)	5 (11.1%)	28 (13.0%)	2 (7.1%)	9 (9.3%)	3 (7.7%)	8 (7.0%)	1 (1.7%)
16 PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT	17 (7.3%)	5 (11.1%)	13 (6.1%)	0	6 (6.2%)	1 (2.6%)	5 (4.4%)	1 (1.7%)
18 REGULATION OF METABOLISM AND PROTEIN FUNCTION	1 (0.4%)	0	1 (0.5%)	0	0	0	0	0
30 CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM	5 (2.1%)	0	6 (2.8%)	2 (7.1%)	5 (5.2%)	1 (2.6%)	8 (7.0%)	2 (3.4%)
32 CELL RESCUE, DEFENSE AND VIRULENCE	22 (9.4%)	4 (8.9%)	21 (9.8%)	0	6 (6.2%)	3 (7.7%)	3 (2.6%)	2 (3.4%)
34 INTERACTION WITH THE ENVIRONMENT	4 (1.7%)	1 (2.2%)	4 (1.9%)	1 (3.6%)	4 (4.1%)	0	5 (4.4%)	0
40 CELL FATE	2 (0.9%)	0	2 (0.9%)	1 (3.6%)	2 (2.1%)	1 (2.6%)	3 (2.6%)	1 (1.7%)
98 CLASSIFICATION NOT YET CLEAR-CUT	11 (4.7%)	1 (2.2%)	15 (7.0%)	5 (17.9%)	11 (11.3%)	6 (15.4%)	13 (11.4%)	8 (13.8%)
99 UNCLASSIFIED PROTEINS	1 (0.4%)	1 (2.2%)	5 (2.3%)	5 (17.9%)	13 (13.4%)	11 (28.2%)	27 (23.7%)	25 (43.1%)
TOTAL SECRETOME	*234*	*45*	*215*	*28*	*97*	*39*	*114*	*58*

[1] FunCat = Functional Catalogue (Ruepp *et al.*, 2004)

Potential secreted virulence / defence proteins in Knoxdaviesia

In both *K. capensis* and *K. proteae*, 105 secreted proteins had significant hits to the Pathogen Host Interactions (PHI) database (Table S6.4). The majority of *K. capensis* and *K. proteae* orthologs had the same PHI-BLAST classification, excluding proteins orthologous between the high and low likelihood datasets. In the high likelihood dataset, two sets of orthologs had different classifications. One set consisted of putative necrosis inducing (NPP1) family proteins, where the best *K. capensis* hit was a protein that does not affect pathogenicity and the best *K. proteae* hit has a "reduced virulence" mutant phenotype. Such a difference in the best hits was also true for two ortholog groups (a protease and a glycosyltransferase) in the low likelihood datasets. In the high likelihood datasets, the majority of proteins were homologous to hits that do not affect pathogenicity. The second largest category was proteins with a "reduced virulence" phenotype. These were typically proteases, glycolyases or oxidative enzymes. Only three proteins in *K. capensis* and five in *K. proteae* were homologous to proteins required for pathogenicity ("loss of pathogenicity" mutant phenotype).

Less than 2 % of the secreted proteins of the *Knoxdaviesia* species were homologous to pathogen effector (plant avirulence determinant) proteins and approximately a third of the secreted proteins could be classified as "cysteine-rich". Five potential effector proteins were present in *K. capensis* and three in *K. proteae*. The unique effector identified in *K. capensis* was annotated as a lytic polysaccharide monooxygenase, enzymes that introduce internal breaks in polysaccharides, such as cellulose, by oxidation (Hemsworth *et al.*, 2015). For *K. capensis*, 83 (35.5 %) proteins in the high likelihood dataset and 24 (24.7 %) in the "low likelihood" secretome dataset had a cysteine content of ≥ 2 %. These numbers were, respectively, 77 (35.8 %) and 35 (30.7 %) for *K. proteae* (Table S6.4). In both the *K. capensis* and *K. proteae* high likelihood datasets, 29 of these proteins consisted of ≤ 200 amino acids and may, therefore, be classified as "small secreted cysteine-rich proteins" (SSCPs; Lu & Edwards, 2016). When considering the SSCPs in the low likelihood dataset, *K. proteae* had more than *K. capensis* (16 vs. 9).

Thirteen enzymes that potentially confront microbial competition in the environment were identified from the high and low secretome datasets (Table 6.5). These were glucose oxidase enzymes that produce hydrogen peroxide, nine different enzymes that detoxify antimicrobials or reactive compounds, two that potentially produce antimicrobial compounds, one involved in pigment production and a protein that binds to bacterial peptidoglycan. Of these defence-

related proteins, 20 were identified in *K. capensis* and 15 in *K. proteae*. The peptidoglycan-binding protein and the Aflatoxin B1 aldehyde reductase proteins were identified only in *K. capensis*, whereas the FAD-containing monooxygenase and Fumonisin B1 esterase were identified only in *K. proteae*. *Knoxdaviesia capensis* had double the number of Glucose oxidase, Tyrosinase and 2OG-Fe(II)oxygenase superfamily proteins than could be found in *K. proteae*.

Table 6.5 Secreted proteins potentially involved in cell defence.

Putative defence protein[2]	**Description**	***Kc***[1]	***Kp***[1]
Production of reactive oxygen species			
Glucose oxidase	Hydrogen peroxide production	4***	2
Secondary metabolite biosynthesis			
Snoal-like polyketide cyclase family	Nogalamycin biosynthesis	1	1
Tyrosinase	Pigment production	4**	2*
2OG-Fe(II)oxygenase superfamily	Antibiotic biosynthesis	2	1
Other			
Peptidoglycan-binding lysin domain		1	0
Detoxification			
Aflatoxin b1 aldehyde reductase member 2	Detoxifies ketones and aldehydes	1*	0
Carboxylesterase	Xenobiotic metabolism	1	1
Chlorocatechol 1,2-dioxygenase	Cleaves catechol (aromatic compound)	1	1
FAD-containing monooxygenase	Oxidises xenobiotics	0	1
Fumonisin B1 esterase	Degrades fumonisin B1	0	1*
Haloacid dehalogenase / Epoxide hydrolase family	Hydrolyses halogen-containing aromatic compounds	2**	2*
Intradiol ring-cleavage dioxygenase	Cleaves aromatic rings	1	1
Superoxide dismutase	Dismutates superoxide	2	2
TOTAL		**20**	**15**

[1] *Kc* = *Knoxdaviesia capensis*; *Kp* = *K. proteae*; asterisks (*) indicate how many proteins are from the low likelihood secretome dataset

[2] FAD = flavin adenine dinucleotide; OG-Fe(II) = oxoglutarate/iron-dependent

Secondary metabolite clusters

AntiSMASH identified 11 SM biosynthesis clusters in *K. capensis* and 10 in *K. proteae*. Three NRPS clusters (Fig. 6.2), two terpene clusters and one unclassified cluster (Fig. 6.3) were found in both species. In two *K. proteae* NRPS clusters, however, the adenylation (A) domain and / or peptidyl carrier protein (P) domain could not be detected (Fig 6.2), implying that these genes may not be functional (Keller *et al.*, 2005). Of the five identified PKS clusters (Fig. 6.4), the PKS gene in one *K. proteae* cluster was split into multiple ORFs of which only one had a detectable β-ketoacyl synthase, C-terminal domain (Fig. 6.4).

The homologous clusters between *K. capensis* and *K. proteae* were identified by considering the upstream and downstream flanking regions. In four of the clusters (NRPS-1, T1PKS-2, T1PKS-3 and T1PKS-4), retrotransposon likely affected the gene content of their flanking regions. In clusters T1PKS-2 and T1PKS-3 the synteny of genes beyond the transposon insertion appeared to be maintained. In cluster T1PKS-4, the synteny of the genes immediately upstream was minimally affected, however, synteny was lost downstream of the truncated *K. proteae* PKS gene. The genes further downstream of the *K. proteae* retrotransposons in cluster NRPS-1 did not coincide with the remainder of the *K. capensis* biosynthetic genes (results not shown). A tBLASTn search against the *K. proteae* genome revealed that these biosynthesis genes were located on scaffold 62. The retrotransposons had, therefore, interrupted this cluster and could have been the cause of the displacement or inversion of these genes.

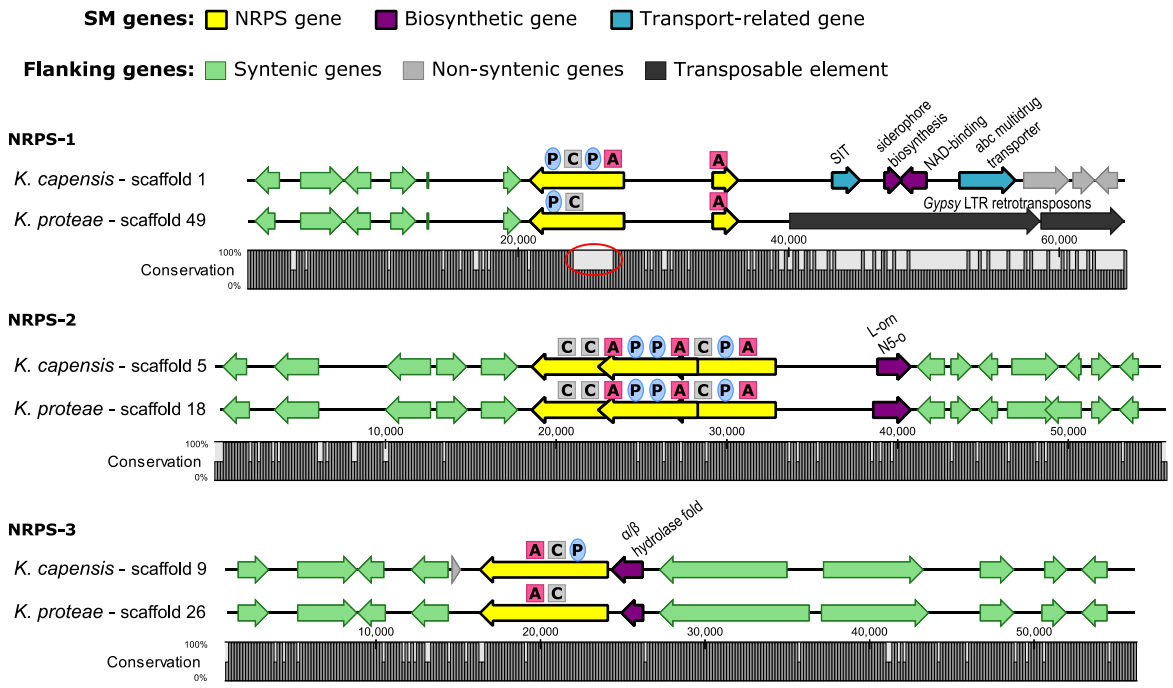

Figure 6.2 Alignment between the homologous non-ribosomal peptide synthase (NRPS) secondary metabolite (SM) biosynthesis clusters in *Knoxdaviesia capensis* and *K. proteae*. The identified NRPS domains are indicated above the genes: A = adenylation (PF00501), C = condensation (PF00668), P = peptidyl carrier protein (PF00550). Other putative SM genes and long terminal repeat (LTR) retrotransposons have been annotated: L-orn N5-o = L-ornithine-N5-oxygenase; SIT = siderophore iron transporter. The red circle indicates missing sequence data in *K. proteae*.

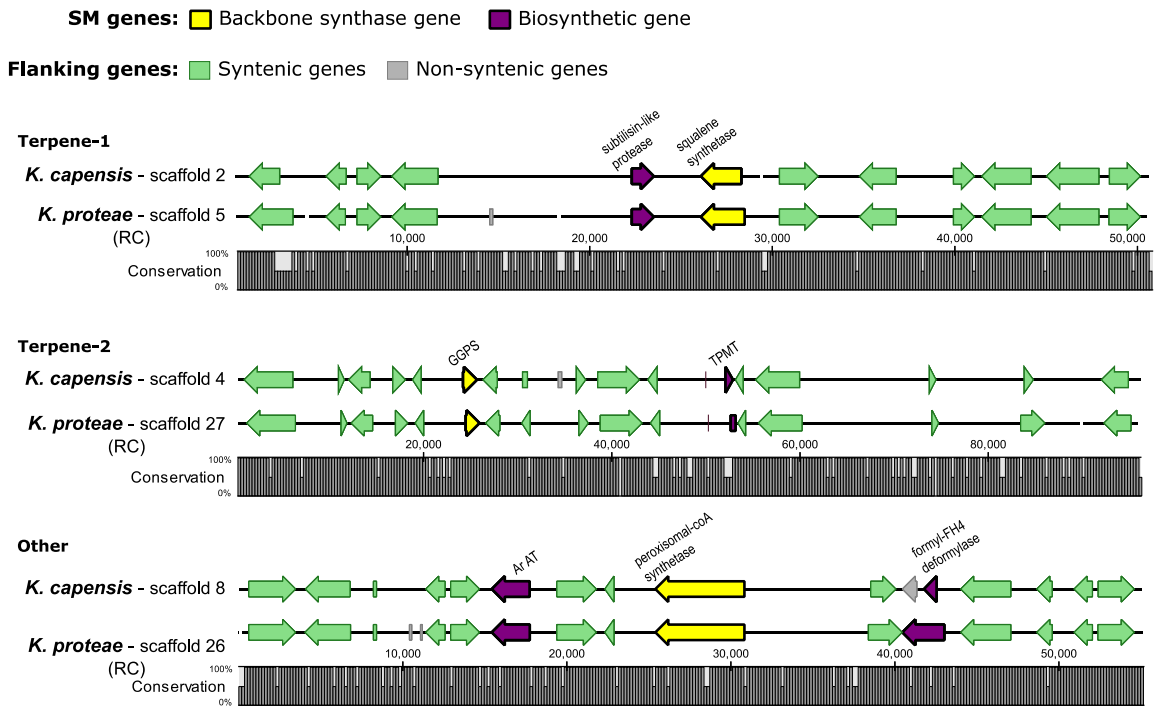

Figure 6.3 Alignment of the two homologous terpene biosynthesis clusters and one peroxisomal-coenzyme A synthetase cluster of *Knoxdaviesia capensis* and *K. proteae*. Other putative SM genes have been annotated: ArAT = aromatic amino acid aminotransferase; formyl-FH4 = formyltetrahydrofolate; GGPS = geranylgeranyl pyrophosphate synthetase; TPMT = thiopurine S-methyltransferase.

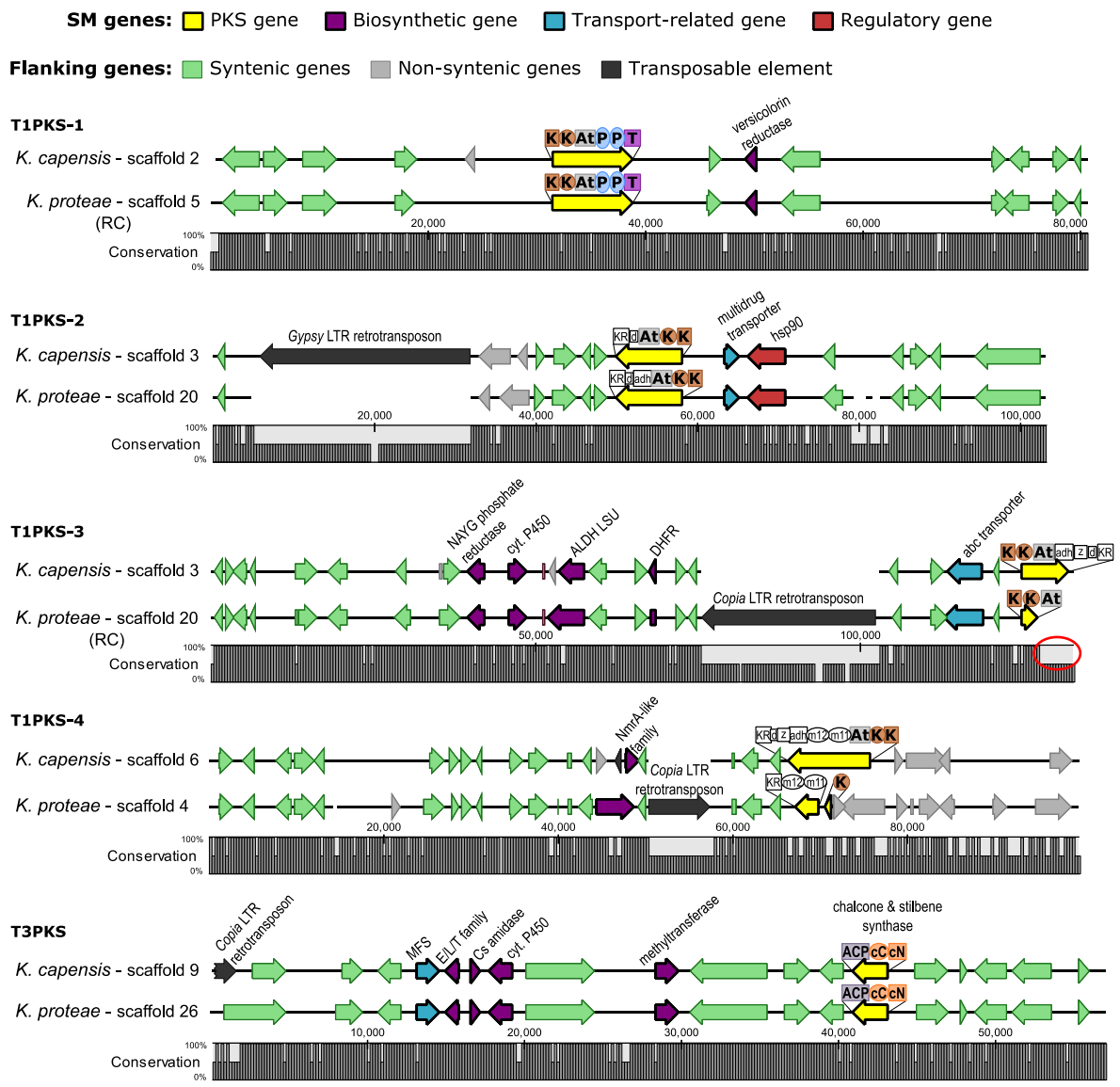

Figure 6.4 Alignment between the homologous polyketide synthase (PKS) secondary metabolite (SM) biosynthesis clusters of *Knoxdaviesia capensis* and *K. proteae*. Essential PKS domains are indicated in colour above the genes, whereas accessory domains are white: ACP = acyl carrier protein (PF08541); adh = alcohol dehydrogenase (PF08240); At = acyltransferase (PF00698); cC = chalcone and stilbene synthase, C-terminal (PF02797); cN = chalcone and stilbene synthase, N-terminal (PF00195); d = short-chain dehydrogenases/reductase (PF00106); K, boxed = β-ketoacyl synthase, N-terminal (PF00109); K, circled = β-ketoacyl synthase, C-terminal (PF02801); KR = β-ketoacyl-ACP reductase (PF08659); m11 = methyltransferase (PF08241); m12 = methyltransferase (PF08242); P = peptidyl carrier protein (PF00550); T = thioesterase (PF00975); z = zinc finger (PF00172). Other putative SM genes and long terminal repeat (LTR) retrotransposons

have been annotated: ALDH LSU = L-aminoadipate-semialdehyde dehydrogenase large subunit; Cyt. = cytochrome; DHFR = dihydrofolate reductase; E/L/T = esterase lipase thioesterase; Cs amindase = N-carbamoylsarcosine amidase; hsp90 = heat shock protein 90; MFS = major facilitator superfamily; NAYG = N-acetyl-gamma-glutamyl; NmrA = nitrogen metabolite repressor A. The T1PKS-3 cluster continues beyond the terminal of scaffold 20 in *K. proteae* (indicated by the red circle).

DISCUSSION

Knoxdaviesia proteae and *K. capensis* are saprotrophic, yet highly niche-specific fungi (Roets *et al.*, 2013). They have evolved in concert with the iconic *Protea* plants of the CCR biodiversity hotspot, together with arthropods, mites, microbes and most likely other organisms. The results of this study show that the overall genome structure of these two closely related species has remain conserved. However, genes that govern their interactions with their host and competitive organisms have diverged, likely in response to different selective pressures.

Macrosynteny and low repeat density

Knoxdaviesia capensis and *K. proteae* showed large-scale genome synteny and aligned to each other with an average of 89.5 % nucleotide identity. The few identified telomeres and the lack of telomeres at both scaffold ends in *K. capensis* and *K. proteae* implies that neither genome assembly contains a complete chromosome. Although several inversions and rearrangements were apparent in the genome dotplot, the lengthy aligned portions indicated macrosynteny between the two genomes (Hane *et al.*, 2011). Both genomes had similar sizes (35.54 vs. 35.49 Mb) and a comparable number of predicted genes (7940 vs. 8173) (Aylward *et al.*, 2016a) and, since these two species are phylogenetically very closely related (Wingfield *et al.*, 1999), large scale genome conservation was expected. The average nucleotide identity between them was less than that observed in comparisons of other closely related *Grosmannia* Goid. (98 %; Alamouti *et al.*, 2014) and dermatophyte species (94.8 %; Burmester *et al.*, 2011), but similar to the 85 - 91 % identity between three species of *Fusarium* Link (Ma *et al.*, 2010).

The repetitive proportion of the *Knoxdaviesia* genomes is congruent with the 0.8 – 28 % range that has been reported for other Sordariomycete fungi (Castanera *et al.*, 2016; Galagan *et al.*, 2003; King *et al.*, 2015; Ma *et al.*, 2010; Nowrousian *et al.*, 2010). Like other fungi (Castanera *et al.*, 2016), the *Knoxdaviesia* genomes contain a large number of Class I elements, specifically *Gypsy* and *Copia* LTR retrotransposons, that use an RNA intermediate to "copy and paste" themselves. Extended proliferation of repeat sequences in the *Knoxdaviesia* genomes, however, appears to be limited by RIP. The identified repetitive sequences occupied a smaller proportion of the genomes than has been affected by RIP, implying that certain repeats could have been mutated to the extent that they can no longer be detected.

Although the two *Knoxdaviesia* species have similar repetitive proportions, a difference in the repeat identity of retrotransposons was apparent and may reflect different evolutionary trajectories. Both *K. capensis* and *K. proteae* have massive gene and genotypic diversity (Aylward *et al.*, 2014; 2015; 2017), disperse over large geographic distances (Aylward *et al.*, 2015; 2017) and continually generate novel genetic combinations through an outcrossing reproductive strategy (Aylward *et al.*, 2016b, Chapter 4). Consequently, they have a high "evolutionary potential" or ability to adapt to changing environments (McDonald & Linde, 2002). Transposable elements are known to increase the plasticity of genomes through transposition to different locations and by providing sites for homologous recombination (Castanera *et al.*, 2016). The divergence of *K. capensis* and *K. proteae* may, therefore, have been partially facilitated by these repetitive elements.

Small secretomes and few effectors

The secreted proteins of fungal species mediate the interaction with their hosts and environments. *Knoxdaviesia capensis* and *K. proteae* have few secreted proteins in comparison to both plant pathogens and saprotrophs (Brown *et al.*, 2012). Previously, Aylward *et al*. (Chapter 5) noted that the lower number of polysaccharide-degrading enzymes in *Knoxdaviesia* compared to other saprotrophs can be attributed to their specialist niche. Carbohydrate degrading proteins were, however, the largest category identified in these species, congruent with *Protea* cell walls that provide their primary source of nutrition. The secreted peptidases and lipases in both species may indicate that they have some capacity to degrade plant components beyond the cell wall. The ecological success of *K. capensis* and *K.*

proteae, despite their small secretomes, most likely reflects their adaptation to a single host genus.

A small proportion of the *Knoxdaviesia* secretome was made up of proteins associated with plant pathogenesis, such as small secreted cysteine-rich proteins (SSCPs; Rep, 2005), proteins with homology to pathogen effector molecules and members of the necrosis and ethylene inducing-like protein (NLP) family. The presence of pathogenesis-related proteins is not uncommon in saprotrophic species (Seidl *et al.*, 2015), although they are usually distinguished from pathogens by their numbers. For example, 190 SSCPs were recently identified from the genome of the cereal head blight pathogen, *F. graminearum* (Lu & Edwards, 2016), in comparison with the < 50 potential SSCP candidates of *Knoxdaviesia*. The *Knoxdaviesia* species, therefore, have a very limited arsenal of potential effector proteins that can be targeted to a living host.

Successful competition

Several secreted defence enzymes were detected in *K. capensis* and *K. proteae*, allowing them to compete with arthropods, nematodes and microbes in infructescences. The glucose oxidase enzymes, which convert glucose to hydrogen peroxide (Bankar *et al.*, 2009), likely enable these species to establish their dominance early during the colonization of young flower heads. Glucose is abundant in *Protea* nectar (Nicolson & Van Wyk, 1998) and the production of hydrogen peroxide may prevent establishment of other microbes during the flowering season. Once nectar is depleted and the infructescence develops, the amount of free glucose would depend on the degradation of cellulose components. Numerous organisms colonize the enclosed infructescence, suggesting that the *Knoxdaviesia* species require additional measures to compete successfully during this stage of niche capture.

Besides the secreted defence proteins, a small arsenal of potential SM biosynthesis clusters was found in both *Knoxdaviesia* species. Whereas most Ascomycete fungi have upwards of 20 NRPS and PKS backbone genes (Kubicek *et al.*, 2011), *K. capensis* has only eight and *K. proteae* seven of these genes. These numbers are even lower than the 10 NRPS and PKS genes identified from *Neurospora crassa* Shear & B.O. Dodge, the model for RIP activity (Galagan *et al.*, 2003). Low repeat content and low gene duplication levels have been identified as the reason for the lack of SM cluster diversity in *N. crassa* (Galagan *et al.*,

2003). Similarly, the large sections of RIP-affected sequences in the *Knoxdaviesia* genomes could imply that RIP has played a role in keeping the evolution of SM clusters in check.

The highly specialised niches occupied by *Protea*-associated *Knoxdaviesia* species may have made large SM diversification unnecessary. The abundance of ophiostomatoid fungi in infructescences was previously thought to be due to their fitness and ability to resist colonization by other contaminating fungi (Lee *et al.*, 2005; Marais & Wingfield, 1994). However, three novel clades of antifungal-producing *Streptomyces* Waksman & Henrici bacteria have recently been isolated from *Protea* infructescences, to which *Knoxdaviesia* species have been shown to be at least partially resistant (Human, 2013). Although *Streptomyces* species may lessen inter-specific fungal competition to a certain extent, *Protea*-associated ophiostomatoid fungi would have to tolerate the antifungal compounds produced by these species.

In both *Knoxdaviesia* species, two of the SM clusters identified (NRPS-1 and 2) are involved in synthesising siderophores, whereas eight different types of secreted proteins potentially detoxify harmful compounds (Table 6.5). Siderophores sequestrate iron from the environment (Haas, 2014). As a scarce resource essential in many catalytic reactions, effective iron uptake plays an important role in determining whether a species will be a successful competitor (Loper & Buyer, 1991). The products synthesised by the remainder of the SM clusters are unknown, although chalcone and stilbene synthase (T3PKS) theoretically produces a flavonoid with a role in pigmentation or defence (Austin & Noel, 2003) and geranylgeranyl pyrophosphate synthetase (GGPS; Terpene-2) an antimicrobial diterpene or carotenoid (Keller *et al.*, 2005). SM clusters are thought to enable life in a specific niche by protection from predators and other competitors (Brakhage & Schroeckh, 2011). These SM clusters may, therefore, contribute toxic substances that enable *Knoxdaviesia* to compete in infructescences. In addition to iron sequestration and putative toxin production by SM clusters, the various secreted detoxification enzymes may enable the *Knoxdaviesia* species to tolerate some antifungal compounds produced by both the actinomycete bacteria and its plant host. Interestingly, a protein with a peptidoglycan binding domain was also detected in *K. capensis*, suggesting that it mediates an interaction with the actinomycete bacteria.

CONCLUSIONS

A high level of genetic similarity has been observed between *K. capensis* and *K. proteae* in previous studies (Aylward *et al.*, 2016b, Chapters 4 & 5). Despite this fact, this comparative genomics study has revealed several differences in their genome arrangements, secreted proteins and defence compounds. The macrosynteny between the *Knoxdaviesia* genomes, the similarities in their secreted proteins and their syntenic SM cluster localization, reflect the phylogenetic and ecological relatedness of these species. In contrast, their difference in host associations is illustrated by the unique component of the secretome that enables *K. capensis* to colonise multiple *Protea* hosts having different chemistries (Valente *et al.*, 2010). This is in contrast to the host-specific *K. proteae* that contained less secreted proteins.

The greater number of secreted enzymes associated with cell defence and the maintenance of the SM clusters in *K. capensis* likely reflects its wider host range. Due to the occurrence of this species in multiple hosts, it will likely also encounter greater numbers of competitors than the host-specific *K. proteae*. For example, two *Streptomyces* groups additional to those already known from *P. repens* have been isolated from a *K. capensis* host, *P. neriifolia* R. Br. (Human, 2013). Conversely, the fewer defence proteins and degraded SM gene cluster in *K. proteae* implies that a diversity of defence compounds is not required for this species to be successful on its *P. repens* host.

Knoxdaviesia capensis and *K. proteae* represent only one of the two lineages of ophiostomatoid fungi that occur in southern African *Protea* species (Marais & Wingfield, 2001). Several *Sporothrix* species occupy *Protea* hosts in the CCR and beyond the borders of this biodiversity hotspot (Roets *et al.*, 2013). *Protea*-associated species in *Sporothrix* are phylogenetically distant to *Knoxdaviesia*, but have adapted to the same niche (De Beer *et al.*, 2016; Wingfield *et al.*, 1999). It would, therefore, be interesting to investigate whether *Sporothrix* has followed the same path of adaptation as hypothesised for the *Knoxdaviesia* species. Specifically, the set of SM biosynthesis clusters and predicted secretomes of these species would indicate whether the *Protea* environment promotes a smaller genetic complement. Further, some *Sporothrix* species have multiple *Protea* hosts (Marais & Wingfield, 2001; Roets *et al.*, 2006a), while others are apparently host-specific (Roets *et al.*, 2008; 2010), enabling further investigation into how host-association has potentially resulted in speciation.

REFERENCES

Alamouti, S. M., Haridas, S., Feau, N., Robertson, G., Bohlmann, J. & Breuil, C. (2014). Comparative genomics of the pine pathogens and beetle symbionts in the genus *Grosmannia*. *Molecular Biology and Evolution* **31**, 1454-1474.

Austin, M. B. & Noel, J. P. (2003). The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports* **20**, 79-110.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2014). Panmixia defines the genetic diversity of a unique arthropod-dispersed fungus specific to *Protea* flowers. *Ecology and Evolution* **4**, 3444-3455.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015). Long-distance dispersal and recolonization of a fire-destroyed niche by a mite-associated fungus. *Fungal Biology* **119**, 245-256.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D. & Wingfield, M. J. (2016a). Genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa. *Standards in Genomic Sciences* **11**, 1-7.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, M. J. & Wingfield, B. D. (2016b). Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*. *Fungal Genetics and Biology* **96**, 47-57.

Aylward, J., Dreyer, L. L., Laas, T., Smit, L., & Roets, F. (2017). *Knoxdaviesia capensis*: dispersal ecology and population genetics of a flower-associated fungus. *Fungal Ecology* **26**, 28-36.

Bankar, S. B., Bule, M. V., Singhal, R. S. & Ananthanarayan, L. (2009). Glucose oxidase — An overview. *Biotechnology Advances* **27**, 489-501.

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E. & Weber, T. (2013). antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, gkt449.

Brakhage, A. A. & Schroeckh, V. (2011). Fungal secondary metabolites – Strategies to activate silent gene clusters. *Fungal Genetics and Biology* **48**, 15-22.

Brasier, C. M. (1991). *Ophiostoma novo-ulmi* sp. nov., causative agent of current Dutch elm disease pandemics. *Mycopathologia* **115**, 151-161.

Brown, N. A., Antoniw, J. & Hammond-Kosack, K. E. (2012). The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: a refined comparative analysis. *PLOS ONE* **7**, e33731.

Burmester, A., Shelest, E., Glöckner, G., Heddergott, C., Schindler, S., Staib, P., Heidel, A., Felder, M., Petzold, A. & Szafranski, K. (2011). Comparative and functional genomics provide insights into the pathogenicity of dermatophytic fungi. *Genome Biology* **12**, R7.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421-421.

Cambareri, E. B., Jensen, B. C., Schabtach, E. & Selker, E. U. (1989). Repeat-induced GC to AT mutations in *Neurospora*. *Science* **244**, 1571-1575.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S. & Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**, 188-196.

Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J., Grimwood, J., Pérez, G., Pisabarro, A. G., Grigoriev, I. V., *et al.* (2016). Transposable Elements

versus the Fungal Genome: Impact on Whole-Genome Architecture and Transcriptional Profiles. *PLOS Genetics* **12**, e1006108.

Christiansen, E. & Solheim, H. (1990). The bark beetle-associated blue-stain fungus *Ophiostoma polonicum* can kill various spruces and Douglas fir. *European Journal of Forest Pathology* **20**, 436-446.

Coetzee, J. H. & Giliomee, J. H. (1985). Insects in association with the inflorescence of *Protea repens* (L.) (Proteaceae) and their role in pollination. *Journal of the Entomological Society of South Africa* **48**, 303-314.

Cowan, D. A., Rybicki, E. P., Tuffin, M. I., Valverde, A. & Wingfield, M. J. (2013). Biodiversity: So much more than legs and leaves. *South African Journal of Science* **109**, 1-9.

Crous, P. W., Rong, I. H., Wood, A., Lee, S., Glen, H., Botha, W., Slippers, B., de Beer, W. Z., Wingfield, M. J. & Hawksworth, D. L. (2006). How many species of fungi are there at the tip of Africa? *Studies in Mycology* **55**, 13-33.

De Beer, Z. W., Duong, T. A. & Wingfield, M. J. (2016). The divorce of *Sporothrix* and *Ophiostoma*: solution to a problematic relationship. *Studies in Mycology* **83**, 165-191.

Döll, K., Chatterjee, S., Scheu, S., Karlovsky, P. & Rohlfs, M. (2013). Fungal metabolic plasticity and sexual development mediate induced resistance to arthropod fungivory. *Proceedings of the Royal Society of London B: Biological Sciences* **280**, 20131219.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLOS Computational Biology* **7**, e1002195.

Emanuelsson, O., Nielsen, H., Brunak, S. & Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005-1016.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., *et al.* (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**, D279-D285.

Fulnečková, J., Ševčíková, T., Fajkus, J., Lukešová, A., Lukeš, M., Vlček, Č., Lang, B. F., Kim, E., Eliáš, M. & Sýkorová, E. (2013). A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution* **5**, 468-483.

Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., FitzHugh, W., Ma, L.-J., Smirnov, S. & Purcell, S. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**, 859-868.

Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32**, D258-D261.

Gilgado, F., Cano, J., Gené, J. & Guarro, J. (2005). Molecular Phylogeny of the *Pseudallescheria boydii* Species Complex: Proposal of Two New Species. *Journal of Clinical Microbiology* **43**, 4930-4942.

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420-3435.

Haas, H. (2014). Fungal siderophore metabolism with a focus on *Aspergillus fumigatus*. *Natural Product Reports* **31**, 1266-1276.

Hane, J. K., Rouxel, T., Howlett, B. J., Kema, G. H., Goodwin, S. B. & Oliver, R. P. (2011). A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biology* **12**, 1.

Hemsworth, G. R., Johnston, E. M., Davies, G. J. & Walton, P. H. (2015). Lytic polysaccharide monooxygenases in biomass conversion. *Trends in Biotechnology* **33**, 747-761.

Holt, C. & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research* **35**, W585-W587.

Human, Z. R. (2013). The diversity and ecology of actinomycetes associated with environments dominated by ophiostomatoid fungi. MSc Thesis, University of Pretoria, Pretoria, South Africa.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32**, D277-D280.

Keller, N. P., Turner, G. & Bennett, J. W. (2005). Fungal secondary metabolism—from biochemistry to genomics. *Nature Reviews Microbiology* **3**, 937-947.

King, R., Urban, M., Hammond-Kosack, M. C., Hassani-Pak, K. & Hammond-Kosack, K. E. (2015). The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**, 1.

Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567-580.

Kubicek, C. P., Herrera-Estrella, A., Seidl-Seiboth, V., Martinez, D. A., Druzhinina, I. S., Thon, M., Zeilinger, S., Casas-Flores, S., Horwitz, B. A., Mukherjee, P. K., *et al.* (2011). Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. *Genome Biology* **12**, R40-R40.

Kubicek, C. P., Starr, T. L. & Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* **52**, 427-451.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology* **5**, R12.

Lee, S., Groenewald, J. Z., Taylor, J. E., Roets, F. & Crous, P. W. (2003). Rhynchostomatoid fungi occurring on Proteaceae. *Mycologia* **95**, 902-910.

Lee, S., Roets, F. & Crous, P. W. (2005). Biodiversity of saprobic microfungi associated with the infructescences of *Protea* species in South Africa. *Fungal Diversity* **19**, 69-78.

Lewis, Z. A., Honda, S., Khlafallah, T. K., Jeffress, J. K., Freitag, M., Mohn, F., Schübeler, D. & Selker, E. U. (2009). Relics of repeat-induced point mutation direct heterochromatin formation in *Neurospora crassa*. *Genome Research* **19**, 427-437.

Li, L., Stoeckert, C. J. & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **13**, 2178-2189.

Loper, J. E. & Buyer, J. S. (1991). Siderophores in microbial interactions on plant surfaces. *Molecular Plant-Microbe Interactions* **4**, 5-13.

Lu, S. & Edwards, M. C. (2016). Genome-Wide Analysis of Small Secreted Cysteine-Rich Proteins Identifies Candidate Effector Proteins Potentially Involved in *Fusarium graminearum*− Wheat Interactions. *Phytopathology* **106**, 166-176.

Ma, L.-J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M. & Henrissat, B. (2010). Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**, 367-373.

Marais, G. J. & Wingfield, M. J. (1994). Fungi associated with infructescences of *Protea* species in South Africa, including a new species of *Ophiostoma*. *Mycological Research* **98**, 369-374.

Marais, G. J. & Wingfield, M. J. (2001). *Ophiostoma africanum* sp. *nov*., and a key to ophiostomatoid species from *Protea* infructescences. *Mycological Research* **105**, 240-246.

Marchese, C. (2015). Biodiversity hotspots: A shortcut for a more complicated concept. *Global Ecology and Conservation* **3**, 297-309.

Margolin, B. S., Garrett-Engele, P. W., Stevens, J. N., Fritz, D. Y., Garrett-Engele, C., Metzenberg, R. L. & Selker, E. U. (1998). A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* **149**, 1787-1797.

McDonald, B. A. & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349-379.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E. & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research* **39**, W339-W346.

Mustart, P., Cowling, R. & Wright, M. (1995). Clustering of fertile seeds in infructescences of serotinous *Protea* species: an anti-predation mechanism? *African Journal of Ecology* **33**, 224-229.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

Nicolson, S. W. & Van Wyk, B.-E. (1998). Nectar Sugars in Proteaceae: Patterns and Processes. *Australian Journal of Botany* **46**, 489-504.

Nowrousian, M., Stajich, J. E., Chu, M., Engh, I., Espagne, E., Halliday, K., Kamerewerd, J., Kempken, F., Knab, B. & Kuo, H.-C. (2010). De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. *PLOS Genetics* **6**, e1000891.

Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* **8**, 785-786.

Pierleoni, A., Martelli, P. L. & Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9**, 1.

Price, A. L., Jones, N. C. & Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics* **21**, i351-i358.

Rebelo, A. & Rourke, J. (1985). Seed germination and seed set in southern African Proteaceae: ecological determinants and horticultural problems. *Acta Horticulturae* **185**, 75-88.

Rep, M. (2005). Small proteins of plant-pathogenic fungi secreted during host colonization. *Fems Microbiology Letters* **253**, 19-27.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276-277.

Roets, F., Dreyer, L. L. & Crous, P. W. (2005). Seasonal trends in colonisation of *Protea* infructescences by *Gondwanamyces* and *Ophiostoma* spp. *South African Journal of Botany* **71**, 307-311.

Roets, F., de Beer, Z. W., Dreyer, L. L., Zipfel, R., Crous, P. W. & Wingfield, M. J. (2006a). Multi-gene phylogeny for *Ophiostoma* spp. reveals two new species from *Protea* infructescences. *Studies in Mycology* **55**, 199-212.

Roets, F., Dreyer, L. L., Geertsema, H. & Crous, P. W. (2006b). Arthropod communities in Proteaceae infructescences: seasonal variation and the influence of infructescence phenology. *African Entomology* **14**, 257-265.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2007). Discovery of Fungus-Mite Mutualism in a Unique Niche. *Environmental Entomology* **36**, 1226-1237.

Roets, F., de Beer, Z. W., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2008). *Ophiostoma gemellus* and *Sporothrix variecibatus* from Mites Infesting *Protea* Infructescences in South Africa. *Mycologia* **100**, 496-510.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2009). Fungal radiation in the Cape Floristic Region: An analysis based on *Gondwanamyces* and *Ophiostoma*. *Molecular Phylogenetics and Evolution* **51**, 111-119.

Roets, F., Wingfield, B. D., de Beer, Z. W., Wingfield, M. & Dreyer, L. L. (2010). Two new *Ophiostoma* species from *Protea caffra* in Zambia. *Persoonia* **24**, 18-28.

Roets, F., Theron, N., Wingfield, M. J. & Dreyer, L. L. (2011a). Biotic and abiotic constraints that facilitate host exclusivity of *Gondwanamyces* and *Ophiostoma* on *Protea*. *Fungal Biology* **116**, 49-61.

Roets, F., Wingfield, M. J., Wingfield, B. D. & Dreyer, L. L. (2011b). Mites are the most common vectors of the fungus *Gondwanamyces proteae* in *Protea* infructescences. *Fungal Biology* **115**, 343-350.

Roets, F., Wingfield, M. J., Crous, P. W. & Dreyer, L. L. (2013). Taxonomy and ecology of ophiostomatoid fungi associated with *Protea* infructescences. In *Ophiostomatoid fungi: expanding frontiers*, pp. 177-187. Edited by K. A. Seifert, Z. W. de Beer & M. J. Wingfield. Utrecht, The Netherlands: CBS Biodiversity Series.

Ruepp, A., Zollner, A., Maier, D., Albermann, K., Hani, J., Mokrejs, M., Tetko, I., Güldener, U., Mannhaupt, G. & Münsterkötter, M. (2004). The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Research* **32**, 5539-5545.

Ruess, L. & Lussenhop, J. (2005). Trophic interactions of fungi and animals. In *The fungal community—its organization and role in the ecosystem*, pp. 581–598. Edited by J. Dighton, J. K. White & P. Oudemans. Boca Raton, FL: CRC Press.

Schneider, D. J. & Collmer, A. (2010). Studying plant-pathogen interactions in the genomics era: beyond molecular Koch's postulates to systems biology. *Annual Review of Phytopathology* **48**, 457-479.

Seidl, M. F., Faino, L., Shi-Kunne, X., van den Berg, G. C., Bolton, M. D. & Thomma, B. P. (2015). The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its pathogenic relatives. *Molecular Plant-Microbe Interactions* **28**, 362-373.

Seifert, K. A., De Beer, Z. W. & Wingfield, M. J. (2013). The Ophiostomatoid Fungi: Expanding Frontiers. Utrecht, The Netherlands: CBS Biodiversity Series.

Selker, E. U., Tountas, N. A., Cross, S. H., Margolin, B. S., Murphy, J. G., Bird, A. P. & Freitag, M. (2003). The methylated component of the *Neurospora crassa genome*. *Nature* **422**, 893-897.

Slotkin, R. K. & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**, 272-285.

Smit, A. F. A., Hubley, R. & Green, P. (2013-2015). RepeatMasker Open-4.0.

Templeton, M. D., Rikkerink, E. H. & Beever, R. E. (1994). Small, cysteine-rich proteins and recognition in fungal-plant interactions. *Molecular Plant Microbe Interactions* **7**, 320-325.

Theron, N., Roets, F., Dreyer, L. L., Esler, K. J. & Ueckermann, E. A. (2012). A new genus and eight new species of Tydeoidea (Acari: Trombidiformes) from *Protea* species in South Africa. *International Journal of Acarology* **38**, 257-273.

Traut, W., Szczepanowski, M., Vítková, M., Opitz, C., Marec, F. & Zrzavý, J. (2007). The telomere repeat motif of basal Metazoa. *Chromosome Research* **15**, 371-382.

Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H. & Hammond-Kosack, K. E. (2015). The Pathogen-Host Interactions database (PHI-base): additions and future developments. *Nucleic Acids Research* **43**, D645-D655.

Valente, L. M., Reeves, G., Schnitzler, J., Mason, I. P., Fay, M. F., Rebelo, T. G., Chase, M. W. & Barraclough, T. G. (2010). Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution* **64**, 745-760.

Van der Linde, J. A., Six, D. L., Wingfield, M. J. & Roux, J. (2012). New species of *Gondwanamyces* from dying *Euphorbia* trees in South Africa. *Mycologia* **104**, 574-584.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., Lee, S. Y., Fischbach, M. A., Müller, R. & Wohlleben, W. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, gkv437.

Wingfield, B. D., Viljoen, C. D. & Wingfield, M. J. (1999). Phylogenetic relationships of ophiostomatoid fungi associated with *Protea* infructescences in South Africa. *Mycological Research* **103**, 1616-1620.

Wingfield, M. J. & Van Wyk, P. S. (1993). A new species of *Ophiostoma* from *Protea* infructescences in South Africa. *Mycological Research* **97**, 709-716.

Winnenburg, R., Baldwin, T. K., Urban, M., Rawlings, C., Köhler, J. & Hammond-Kosack, K. E. (2006). PHI-base: a new database for pathogen host interactions. *Nucleic Acids Research* **34**, D459-D464.

Zdobnov, E. M. & Apweiler, R. (2001). InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847-848.

Zwölfer, H. (1979). Strategies and counter strategies in insect population systems competing for space and food in flower heads and plant galls. *Fortschritte der Zoologie* **25**, 331-353.

SUPPLEMENTARY INFORMATION

Tables S6.2 and **S6.4** are below. **Tables S6.1** and **S6.3** have been provided electronically.

Table S6.1 Summary of repeat-induced-point mutation (RIP) per scaffold. * = unaligned scaffolds.

Table S6.3 Classification of secreted proteins into Functional Catalogue (FunCat; Ruepp et al., 2004) categories.

Table S6.2 Occurrence of the TTAGGGTTAC / GTAACCCTAA *Knoxdaviesia* telomere repeat in *K. capensis* and *K. proteae*.

scaffold	repeat	start	end	length
K. proteae				
scaffold16 (3')	(TTAGGGTTAC)8(TTA)	329642	329724	83
scaffold21 (5')	(GTAACCCTAA)8(GTAACCCTAT)2	1	100	100
scaffold36 (3')	(TTAGGGTTAC)8(TTAG)	400522	400605	84
scaffold45 (5')	(TAA)(GTAACCCTAA)8	1	83	83
scaffold50 (5')	(CTAA)(GTAACCCTAA)8	1	84	84
scaffold51 (5')	(CCTAA)(GTAACCCTAA)8	1	85	85
scaffold75 (5')	(GTAACCCTAA)7-(GTAA)(CTAA)-CC-(CTAA)(GTAA)(CTAA)-CC-(CTAA)(GTAA)-C-(CTAA)2-C-(CTAA)2-CC-[(CTAA)2-C]3(CTACCCTAA)(GTAACC)	1	164	164
scaffold86 (5')	(ACCCTAA)(GTAACCCTAA)8	1	87	87
scaffold88 (5')	(AACCCTAA)(GTAACCCTAA)8	1	88	88
scaffold107 (3')	(TTAGGGTTAC)8(TTAGG)	1987	2071	85
K. capensis				
scaffold3 (in a transposon)	(TTAGGGTTAC)10-T-(TTAGGGTTAC)5-TTAGGGTTNA-(GTAACCCTAA)3-(GAAACCCTAA)-(GTAACCCTAA)7	1165579	1165848	270
scaffold6 (3')	(TTAGGGTTAC)8-(TTAGGGTT)	2648838	2648925	88
scaffold7 (3')	(TTAGGGTTAC)11-(TTAGGGTT)	2670629	2670746	118
scaffold13 (3')	(TTAGGGTTAC)8-T-(TTAGGGTTAC)9-(TTAGGGTT)	212831	213010	180
scaffold17 (5')	(AA)-(GTAACCCTAA)8	1	82	82
scaffold18 (3')	(TTAGGGTTAC)11-(TTAGGGTT)	144117	144234	118
scaffold20 (3 kbp from 5')	(TAGGGTTAC)-(TTAGGGTTAC)13-(TTAGGGTTTC)-(TTAGGGTTAC)27-(TTA)	49916	50334	419
scaffold22 (5')	(CCCTAA)-(GTAACCCTAA)8	1	85	85
scaffold24 (5')	(TAA)-(GTAACCCTAA)8	1	83	83

Table S6.4 Number of cysteine-rich secreted proteins and proteins with hits to the Pathogen Host Interaction (PHI) database.

	High Likelihood Dataset				Low Likelihood Dataset			
	K. capensis		*K. proteae*		*K. capensis*		*K. proteae*	
	Total (%)	**Unique (%)**	**Total (%)**	**Unique (%)**	**Total (%)**	**Unique (%)**	**Total (%)**	**Unique (%)**
Cysteine-rich proteins								
≥ 2 % cysteine residues	83 (35.5%)	14 (31.1%)	77 (35.8%)	11 (39.3%)	24 (24.7%)	6 (15.4%)	35 (30.7%)	17 (29.3%)
SSCPs[1]	29 (12.4%)	5 (11.1%)	29 (13.5%)	4 (14.3%)	9 (9.3%)	3 (7.7%)	16 (14.0%)	10 (17.2%)
PHI-BLAST mutant phenotypes								
Effector (plant avirulence determinant)	5 (2.1%)	1 (2.2%)	3 (1.4%)	0	0	0	0	0
Increased virulence	4 (1.7%)	1 (2.2%)	3 (1.4%)	0	0	0	0	0
Loss of pathogenicity	1 (0.4%)	0	1 (0.5%)	0	1 (1.0%)	1 (2.6%)	2 (1.8%)	2 (3.4%)
Reduced virulence	31 (13.2%)	2 (4.4%)	32 (14.9%)	2 (7.1%)	18 (18.6%)	8 (20.5%)	10 (8.8%)	3 (5.2%)
Unaffected pathogenicity	33 (14.1%)	3 (6.7%)	40 (18.6%)	5 (17.9%)	8 (8.2%)	1 (2.6%)	9 (7.9%)	4 (6.9%)
No hits in PHI-database	160 (68.4%)	38 (84.4%)	136 (63.3%)	21 (75.0%)	80 (72.2%)	29 (74.4%)	93 (81.6%)	49 (84.5%)
TOTAL SECRETOME	*234*	*45*	*215*	*28*	*97*	*39*	*114*	*58*

[1] SSCPs = small secreted cysteine-rich proteins; these have a cysteine content of ≥ 2 % and a length of ≤ 200 residues

CHAPTER 7

Conclusions

This dissertation investigated the ecology of two species of *Knoxdaviesia* fungi associated with native *Protea* plants in the Western Cape Province of South Africa. The first major contribution from this work is the publically available genomes of *K. capensis* and *K. proteae* (Aylward *et al.*, 2016a). These genomes are the first to be determined for the fungal family Gondwanamycetaceae and, therefore, add to the goal of sampling fungal biodiversity (Spatafora, 2011).

The genome sequences of *K. capensis* and *K. proteae* were applied to investigate the genes that control mating recognition in these two species (Aylward *et al.*, 2016b). Results showed that *K. capensis* and *K. proteae* are heterothallic and have similar gene content to other characterised species in the order Microascales. The presence of a single mating type idiomorph per strain in both *K. capensis* and *K. proteae* means that these species are obligated to find a mate with the opposite idiomorph to induce sexual reproduction. This explains the massive gene and genotypic diversity observed in natural populations of both species (Aylward *et al.*, 2014; 2015; 2017). Sexual structures of *Knoxdaviesia* fungi are common in *Protea* infructescences and each of these produce offspring that are genetically novel.

The mating loci of the *Knoxdaviesia* species also shed light on the evolution of these loci in the Microascales. The pathogenic Ceratocystidaceae have rearrangements in the flanking regions of these loci (Wilken *et al.*, 2014; Wilson *et al.*, 2015), but *Knoxdaviesia* do not. Since the *Knoxdaviesia* species are sister to the Ceratocystidaceae, this indicates that the diversification of flanking regions did not take place in the Microascalean ancestor, but in the Ceratocystidaceae family. Further, similar features in the *Knoxdaviesia MAT1-1-1* and *MAT1-2-7* genes suggest an ancestral recombination event in the Microascalean ancestor – something rarely recorded in loci associated with reproduction.

The carbon utilisation study uncovered significant differences in the carbon usage profile of saprotrophic *Knoxdaviesia* species in comparison to a closely related plant pathogen. While the *Knoxdaviesia* species use simple sugars and dead plant cell walls, the pathogen,

Ceratocystis albifundus, likely exploits soluble polysaccharides in phloem and penetrates plant cells to access storage polysaccharides. Assumptions of simple sugars as universally preferred carbon sources were challenged, because *C. albifundus* did not grow on glucose, fructose or sucrose as a sole carbon source. Likely, this pathogen has lost its ability to translocate these sugars into the cell. These data did not only reveal aspects of saprotroph *versus* plant pathogen biology, but will inform future research on substrate transport and metabolism in *C. albifundus*.

Association with a single host genus appears to have decreased the substrate usage and competitive abilities of these *Knoxdaviesia* species, such that only *Protea* plants are suitable hosts. This pattern is even more pronounced in the specialist *K. proteae*. It displayed a smaller substrate range, less secreted proteins, less enzymes involved in cell defence and a lower number of potentially functional secondary metabolite biosynthesis clusters compared to its generalist counterpart, *K. capensis*. Host association appears to have been a potential driver of evolution in these *Knoxdaviesia* species.

REFERENCES

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2014). Panmixia defines the genetic diversity of a unique arthropod-dispersed fungus specific to *Protea* flowers. *Ecology and Evolution* **4**, 3444-3455.

Aylward, J., Dreyer, L. L., Steenkamp, E. T., Wingfield, M. J. & Roets, F. (2015). Long-distance dispersal and recolonization of a fire-destroyed niche by a mite-associated fungus. *Fungal Biology* **119**, 245-256.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, B. D. & Wingfield, M. J. (2016a). Genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa. *Standards in Genomic Sciences* **11**, 1-7.

Aylward, J., Steenkamp, E. T., Dreyer, L. L., Roets, F., Wingfield, M. J. & Wingfield, B. D. (2016b). Genetic basis for high population diversity in *Protea*-associated *Knoxdaviesia*. *Fungal Genetics and Biology* **96**, 47-57.

Aylward, J., Dreyer, L. L., Laas, T., Smit, L., & Roets, F. (2017). *Knoxdaviesia capensis*: dispersal ecology and population genetics of a flower-associated fungus. *Fungal Ecology* **26**, 28-36.

Spatafora, J. (2011). 1000 fungal genomes to be sequenced. *IMA Fungus* **2**, 41.

Wilken, P. M., Steenkamp, E. T., Wingfield, M. J., de Beer, Z. W. & Wingfield, B. D. (2014). DNA Loss at the *Ceratocystis fimbriata* Mating Locus Results in Self-Sterility. *PLOS ONE* **9**, e92180.

Wilson, A., Godlonton, T., Van der Nest, M. A., Wilken, P. M., Wingfield, M. J. & Wingfield, B. D. (2015). Unisexual reproduction in *Huntiella moniliformis*. *Fungal Genetics and Biology* **80**, 1-9.