Comparative genomics reveal processes implicated in host-specificity in species within the American clade of the *Fusarium fujikuroi* species complex

by

Claudette Dewing

Submitted in partial fulfilment of the requirements for the degree

MAGISTER SCIENTIAE

in the Faculty of Natural and Agricultural Sciences,

University of Pretoria,

Pretoria

April 2020

DECLARATION

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

C. Dewing

April 2020

ACKNOWLEDGEMENTS

I thank the DST/NRF SARChI Chair in Fungal Genetics, the Centre of Excellence in Tree Health Biotechnology (CTHB), the Tree Protection Cooperative Programme (TPCP) and the University of Pretoria for financial support.

I thank everyone who joined my journey and had a part in the completion of this dissertation.

To my supervisors:

I was very fortunate to have an all-woman supervisor committee. Profs. Emma Steenkamp and Brenda Wingfield and Drs. Lieschen De Vos and Magriet van der Nest, thank you for having me as your student.

Dr. Lieschen De Vos, thank you for taking me in as your student. Thank you for your interest in my project and constant reminder of the bigger picture. Thank you for the necessary genome assemblies and annotations and the hours you have put in to figure out most of the scripts used. Thank you for your patience, your open door and that entire Friday we spent tabulating the SynChro results. Thank you for always attending my presentations and supporting me during those stressful times.

Dr. Magriet van der Nest, thank you for walking the road with me since I was an honours student. Thank you for being my co-supervisor. Thank you for your enthusiasm and for always reminding me that I can complete this dissertation. Thank you for checking up on me when you have not heard from me in a while. Thank you for the help and input with the analyses. Thank you for your patience, draft after draft. Thank you for being such a good and funny person. Thank you for all the coffee and telling me about Snippie and Babatjie. Thank you for always being available to help me. Magriet, words will never be enough to show my appreciation.

Prof. Emma Steenkamp, thank you for walking the road with me since I was an honours student. Thank you for your patience and for always making sure I knew what the aims

of this dissertation were. Thank you for being such a strong and proud woman and for being an inspiration to me. Thank you for your enthusiasm and your advice on life. Thank you for reminding me of my self-worth and for believing that I can be a good scientist. Thank you for your patience and your open door, apart from your busy schedule. Thank you for always being kind and always smiling. Thank you for your input during Friday afternoon meetings. Thank you for helping me sort out the 119 phylogenetic trees. Thank you for being straightforward and honest. Being your student was a life-changing experience.

Prof. Brenda Wingfield, words can never be enough to explain the inspiration you are to me. Thank you for believing in me from the beginning. Thank you for your support and motivation during this journey. Thank you for Friday afternoon meetings where we could all relax and laugh about life. Thank you for your input and suggestions during the completion of this dissertation. Thank you for being so passionate about science and for your patience. Thank you for the opportunity to be one of your students and for always smiling. Thank you for the foundation you provided me to become the best scientist I can be. Brenda, thank you for being you.

Dr. Quentin Santana, thank you for your advice on analyses and for always being available to help with scripts. Thank you for taking such interest in all projects related to *Fusarium.* It was a privilege to work with you and to have you on my team.

To friends and family:

Ladies from culture collection (Dr. Seonju Marincowitz, Lydia and Valentina), thank you for always being friendly and for allowing me to complete my work-study experience hours with you. Thank you for the conversations we had and for the experience I have gained.

To Ricu, thank you for being my culture collection partner in crime, for the conversations we had, and the hard work we did, which made the time go by so quickly. Thank you for the best field trip I had with you and Juanita. To Benny, thank you for always being so supportive and funny. Thank you for your personal time and for the adventures we had, such as the Soweto trip. To Juanita, thank you for your friendliness and great company

in the car on the way to our field trip and to that one art exhibition. To Nam, thank you for being such a great lab partner. Thank you for your interest in my project and your support. Thank you for the laughter and treats.

To Ian, Elsa, Gomo and Phetogo, thank you for everything. Thank you for all the laughter and great conversations. Thank you for always checking up on me and making sure that I am smiling. Thank you for caring about me and the progress I made with my dissertation. Thank you for the ice cream, frozen yoghurt and peanut butter filled pretzels. Thank you for being my stress relievers, my feel-good people.

To Wian, thank you for playing such a significant part in my life at this moment. Thank you for your endless support and motivation. Thank you for your interest in my work, as well as your rationality. Thank you for all the laughter and coffee. Thank you for all the adventures, for being my friend and for loving me the way you do.

To my parents, thank you for being such good role models for me. Thank you for your undying love and support. Thank you for believing in me and my dreams. Thank you for being next to me in every step I take. Thank you for being the stability in my life and for celebrating every achievement. Thank you for being with me during happy, sad and anxious times. Thank you for the joy you bring to my life. Thank you for being my sunshine and my heroes. Thank you.

TABLE OF CONTENTS

PREFACE

Fusarium species have a set of genes conferring host-specificity and improving downstream adaptation and survival. The genomic regions in which these genes are located have shown to be vital determinants of host-specificity. Furthermore, the level of expression of these genes will likely differ at different time points of infection. This MSc took on two approaches, each identifying genes involved in host-specificity by implementing comparative genomics. The main focus of this study was based on contrasting *Fusarium* species associated with *Pinus patula* and members of Poaceae and identifying unique genes to each group. This study also included species considered as pathogens or endophytes on their respective hosts. The results obtained can be of great value for future comparative studies with regards to the role and functions of host-specific fungal genes.

The first chapter is written as a review article focussing on the diversity of fungal genomes and how different gene sets play a role in host-specificity. The focus was on fungi in general but primarily focussed on the diversity found within the genus *Fusarium*. Chapter 2 and Chapter 3 were independent studies on genes potentially involved in host-specificity. Chapter 2 focussed on the identification of genes unique to *Fusarium* species associated with different hosts which also have pivotal roles in conferring host-specificity. The aim of Chapter 2 was to characterise these genes in terms of the processes they encode for, their chromosomal location and phylogenetic origin. Chapter 3 focussed on the repertoire of genes encoding for Carbohydrate-Active enZymes (CAZymes) between *Fusarium* species associated with the two different hosts. The aim of Chapter 3 was to determine if this gene repertoire differed between these two groups and the role these genes play in host-specificity, taking the cell wall composition of the two different host plants into account.

The research conducted during this study was performed at the Forestry and Agricultural Biotechnology Institute (FABI), Department of Biochemistry, Genetics and Microbiology at the University of Pretoria. This study was performed under the supervision of Drs. Lieschen De Vos, and Magriet van der Nest, as well as Profs. Emma Steenkamp and Brenda Wingfield. Most of the genomes used in this study were

previously sequenced and are publicly available, except for two genomes which were kindly provided by Dr. Robert H. Proctor.

This dissertation consists of three independent units and, therefore, contain some duplicated information and references. Chapter 1 is written as a review article based on the available literature, whereas Chapter 2 and Chapter 3 are written in the format required for submission to scientific journals.

All the information from this dissertation can be retrieved from the following Google drive:

https://drive.google.com/drive/folders/1eK5hGcNRkECBVHwOno6XlOKbKUI43ZLg?usp=sharing

CHAPTER 1:

Literature review: Diversity in fungal genomes

INTRODUCTION

The genus *Fusarium* is diverse, harbouring species of great importance in industry, agriculture and medicine (Geiser *et al.*, 2013). *Fusarium* species have also shown to be significant members of microbial communities in the soil (Nelson *et al.*, 1994). Examples of species important in the industry are *F. venenatum* and *F. oxysporum* strain Fo47. The former is essential in the production of Quorn, which is a vegetarian protein food product (King *et al.*, 2018). The latter promotes plant growth by acting as a biocontrol agent to protect bananas and several other crops against Fusarium wilt (also known as Panama disease) (Alabouvette, 1986, Fuchs *et al.*, 1997, Forsyth *et al.*, 2006, Nel *et al.*, 2006b, Dita *et al.*, 2018). Species within the genus *Fusarium* are also well-known for the production of mycotoxins under favourable environmental conditions (Bertero *et al.*, 2018). These mycotoxins are secreted by *Fusarium* species leading to the contamination of agricultural products, food and animal feed, making it unsuitable for consumption (Jurjevic *et al.*, 2002, Logrieco *et al.*, 2003, Molinié *et al.*, 2005, Desjardins & Proctor, 2007, Ferrigo *et al.*, 2016). For example, fumonisin B_1 is mainly secreted by *F. verticillioides* and *F. proliferatum* (Voss *et al.*, 2007), deoxynivalenol is mainly produced by *F. graminearum* and *F. culmorum* (Logrieco *et al.*, 2002), and zearalenone is produced mainly by *F. graminearum* (Bertero *et al.*, 2018).

Most economically important *Fusarium* species are pathogens. These *Fusarium* species are associated with a wide range of plant hosts, which include *Pinus* species (Gordon *et al.*, 1996, Herron *et al.*, 2015), grasses (Desjardins *et al.*, 2000b, Zeller *et al.*, 2003) and cereal and vegetable crops (Parry *et al.*, 1995, Bottalico & Perrone, 2002, Armitage *et al.*, 2018). In plants, the typical symptoms of infection associated with this genus include vascular wilts, root and collar rot, cankers on branches, stems or roots, as well as branch and tip dieback (Bloomberg, 1981, Dwinell *et al.*, 1981, Kraft *et al.*, 1981, Wingfield *et al.*, 2008, Mitchell *et al.*, 2011). Also, species from this genus infect humans by causing keratitis, infections in the finger and toenails of immune-compromised patients and people with decreased white blood cell count (Chang *et al.*, 2006, Sutton & Brandt, 2011).

Not all *Fusarium* species are pathogenic, and some may engage in endophytic to partially endophytic lifestyles (Alabouvette, 1986, Sieber *et al.*, 1988, Boshoff *et al.*, 1996, Fuchs

et al., 1997, Sieber, 2002, Bacon & Yates, 2006, Forsyth *et al.*, 2006, Nel *et al.*, 2006b, Zakaria & Ning, 2013, Imazaki & Kadota, 2015, Dita *et al.*, 2018, Lofgren *et al.*, 2018, Nieva *et al.*, 2019). There are many examples in the scientific literature of fungi being pathogenic to one host but endophytic on others (Zabalgogeazcoa, 2008, Hardoim *et al.*, 2015). For example, *F. circinatum* (the pitch canker pathogen on pine trees) is known to be endophytic on grasses and dicots (such as maize and herbaceous plants) (Swett & Gordon, 2009, Swett & Gordon, 2012, Swett & Gordon, 2015, Hernandez-Escribano *et al.*, 2018, Carter & Gordon, 2020), as well as other conifers (Storer *et al.*, 1994, Martín-García *et al.*, 2018) and cause no symptoms of disease on these hosts. Different grass species growing in pitch canker infested *Pinus* plantations serve as a reservoir for the *F. circinatum* inoculum that contributes to the epidemiology of pitch canker (Carter & Gordon, 2020). Grass species, therefore, seem to act as a reproductive host for this pathogen to infect nearby susceptible hosts (Swett *et al.*, 2014). Furthermore, pine seedlings grown in *F. circinatum* infested soil have accelerated growth rates, increased host resistance to infections with pathogens and may even enhance the extensiveness of the pine-associated mycorrhizal community (Martin-Rodrigues *et al.*, 2013, Evira-Recuenco *et al.*, 2015, Swett *et al.*, 2016, Swett & Gordon, 2017). Fungal activity shows the ambiguous ecological adaptations of this fungus, which is not limited to its status as a primary pathogen.

Amongst the species complexes of *Fusarium*, the *Fusarium fujikuroi* species complex (FFSC) is well known for the plant pathogens it includes (Weerdt *et al.*, 2006, Herron *et al.*, 2015, Al-Hatmi *et al.*, 2016, Edwards *et al.*, 2016, Laurence *et al.*, 2016, Niehaus *et al.*, 2016, Moussa *et al.*, 2017). The FFSC broadly corresponds to those species contained within the section *Liseola* of *Fusarium* (Nelson *et al.*, 1994, O'Donnell *et al.*, 1998). Based on the phylogeographic origin of their hosts, species within this complex can be divided into one of three clades: the African, American and Asian clade (O'Donnell *et al.*, 1998) (Table 1, Figure 1). Species in the FFSC are phylogenetically distinct from one another, although their morphological characteristics can overlap (Kvas *et al.*, 2009, Summerell *et al.*, 2010).

Species of the FFSC display high levels of micro- and macro-synteny and this also extends to other *Fusarium* species complexes. Syntenic regions are defined as regions with similar gene content in which the gene order is also conserved. Sperschneider *et al.*

(2015) discovered syntenic overlaps between genomic regions of different species complexes, *Fusarium oxysporum* species complex (FOSC) and *Fusarium graminearum* species complex (FGSC), respectively. Syntenic regions have also been identified between *F. verticillioides* (FFSC) and *F. graminearum* (FGSC) (Zhao *et al.*, 2014), as well as between *F. fujikuroi* (FFSC) and *F. verticillioides* (Wiemann *et al.*, 2013) and between *F. verticillioides* and *F. oxysporum* f. sp. *lycopersici* (FOSC) (Ma *et al.*, 2010). Similarly, *F. circinatum* and *F. temperatum*, in comparison to *F. fujikuroi* and *F. verticillioides* (FFSC) (De Vos *et al.*, 2007) also show high levels of synteny.

Comparative genomics is often utilised to understand how evolution alters genomes. Moreover, these comparisons can identify the genes conserved amongst species and those unique to others. It is often hypothesised that the unique genes are involved in niche adaptation and the emergence of new phenotypes (Sperschneider *et al.*, 2015, Walkowiak *et al.*, 2016, Williams *et al.*, 2016, Niehaus *et al.*, 2017). The underlying mechanisms of phenotypes, such as host-specificity in *Fusarium* species, are poorly understood. This literature review aims to provide insight into the mechanisms involved in host-specificity, as well as genes implicated in various host-pathogen interactions. Studies on host-specificity can lead to downstream development of effective management strategies for the prevention of emerging fungal threats by providing vital insights into the interactions mentioned above. The subsequent chapters of this dissertation will focus primarily on interactions between species of the FFSC and their hosts. The review also provides a brief overview of what is known regarding these interactions, focussing on those *Fusarium* species associated with *Pinus patula* and members of Poaceae.

PINE-ASSOCIATED *FUSARIUM* WITHIN THE AMERICAN CLADE OF THE FFSC

The most economically important *Fusarium* species affecting pines is *F. circinatum* (Hepting & Roth, 1946, O'Donnell *et al.*, 1998). It first appeared in the USA where this pathogen colonised *P. virginiana* trees (Hepting & Roth, 1946) and caused pitch canker on this host. The symptoms of this disease include cankers soaked with resin on branches and trunks of trees (Dwinell *et al.*, 1985), dieback of shoots (Correll, 1991) and mortality

of female flowers and mature cones (Barrows-Broaddus, 1990). The pathogen then appeared in other parts of the world, such as Japan (Kobayashi & Muramoto, 1989, Muramoto & Dwinell, 1990), South Africa (Viljoen *et al.*, 1994), Spain (Landeras *et al.*, 2005), Mexico (Guerra-Santos, 1998), South Korea (Lee *et al.*, 2000), Chile (Wingfield *et al.*, 2002), France (OEPP, 2006), Italy (Carlucci *et al.*, 2007), Portugal (Bragança *et al.*, 2009), Uruguay (Alonso & Bettucci, 2009), Colombia (Steenkamp *et al.*, 2012) and Brazil (Pfenning *et al.*, 2014). These findings emphasise the fact that *F. circinatum* is spread across the world from where it was initially distributed and discovered. These findings also emphasise that this fungus is not limited to specific climates since it associates with Mediterranean, subtropical and temperate climates (Ganley *et al.*, 2009, EFSA, 2010).

The pitch canker disease causes significant losses of susceptible *Pinus* species all around the world, resulting in significant economic losses in plantations where *Pinus* species are planted (Hepting & Roth, 1946, Wingfield *et al.*, 2008, Mitchell *et al.*, 2011). A study based on *Pinus* species from Central America and Mexico indicated high levels of susceptibility of *F. circinatum* towards *P. radiata*, *P. patula*, *P. pseudostrobus*, *P. taeda* and to a lesser extent in *P. elliotti* and *P. greggii* (Hodge & Dvorak, 2000). Other species, such as *P. oocarpa*, *P. pringlei*, *P. jaliscana*, *P. tecunumanii*, *P. maximinoi* and *P. caribaea* are more tolerant towards the pitch canker fungus (Hodge & Dvorak, 2000). However, in South Africa, the most common *Pinus* species planted is the very susceptible *P. patula* in areas with summer rainfall and *P. radiata* in areas with winter rainfall (Mitchell *et al.*, 2011, Hongwane *et al.*, 2018). These two *Pinus* species have been replaced by more tolerant species, *P. elliotti* and *P. taeda*, to improve field survival. However, the two latter species have lower wood quality and growth, compared to the two former species and the only solution against susceptibility and quality of wood was to create and implement hybrid species (Kietzka, 1988, Malan, 2003, Mitchell *et al.*, 2011). Some hybrid species in South Africa, such as *P. elliottii* x *P. caribaea*, *P. patula* x *P. tecunumanii*, *P. patula* x *P. oocarpa* and *P. patula* x *P. tecunumanii* have shown significantly more tolerance against *F. circinatum* (Bayley & Blakeway, 2002, Nel *et al.*, 2006a, Mitchell *et al.*, 2011).

It was initially thought that *F. circinatum* is a necrotrophic pathogen (O'Donnell *et al.*, 1998). A study was performed on the response of pine hosts towards the necrotrophic

pathogen (Morse *et al.*, 2004) and found that the expression of genes, from both the pathogen and the pine host, were likely linked to a defence response and disease formation. Morse *et al.* (2004) studied the expression of genes from the pine host during the disease state and identified genes encoding chitinase, peroxidase, an antimicrobial peptide, a lipid transfer protein and a boiling-stable protein associated with drought responses that induce shoot desiccation during the disease state. However, *F. circinatum* is now considered to be a hemibiotrophic fungus (Swett *et al.*, 2016, Swett *et al.*, 2018) due to the asymptomatic fungal association with pine roots (Martin-Rodrigues *et al.*, 2013, Swett *et al.*, 2016, Swett & Gordon, 2017). When this pathogen reaches the collar of the pine host, infection occurs in both the root and shoot tissue. Infection is followed by extensive fungal growth in the roots, possibly due to nutrient availability from dead root tissue (Swett *et al.*, 2016). These authors suggest that damage to the root system is due to the location of the pathogen.

Of great concern are novel *Fusarium* species that were recently found on diseased *Pinus* species, some of which include *F. fracticaudum*, *F. pininemorale*, *F. sororula*, *F. marasasianum* and *F. parvisorum* (Herron *et al.*, 2015). Some strains of these species showed to be at least as pathogenic as a virulent strain of *F. circinatum.* Some variation occurred in the level of pathogenicity between isolates of the same species but is not uncommon and has been documented before for several other *Fusarium* species (Burgess, 1981, Gordon & Okamoto, 1992, Appel & Gordon, 1995, Miedaner *et al.*, 2001, Carter *et al.*, 2002). The variation in the pathogenicity is of particular concern as these emerging pathogens not only pose a significant economic risk to forestry activities in Colombia but also to all pine growing regions of the world.

Due to their economic importance, genomes of various *Fusarium* species associated with pines have been sequenced. These included three *F. circinatum* isolates, FSP34 (USA) (Wingfield *et al.*, 2018b), KS17 (South Africa) (Van Wyk *et al.*, 2018) and a laboratory strain (GL1327) (Van der Nest *et al.*, 2014). Also sequenced were *F. pininemorale* (Wingfield *et al.*, 2017) and *F. fracticaudum* (Wingfield *et al.*, 2018a). Sequencing of fungal genomes provides insight into genome evolution, metabolic pathways and fungal lifestyles (Aylward *et al.*, 2017). The availability of genome sequences allows for *in silico* predictions of secreted proteins involved in virulence (Kämper *et al.*, 2006, Liu *et al.*, 2015). Another essential aspect of sequenced genomes is to reveal chromosomes or genes

acquired through horizontal chromosome/gene transfer (Ma *et al.*, 2010). Genome sequencing enables gene prediction, identification of gene gains and losses and revealing the effect of genome plasticity on fungal adaptation and survival (Wöstemeyer & Kreibich, 2002, Keller *et al.*, 2005, Lee *et al.*, 2005). Fungal genomics is a component of comparative genomics and is quite crucial in genome characterisation, in terms of the biology and genetics of pathogens.

Fusarium circinatum resides within the American clade of the FFSC (O'Donnell *et al.*, 1998, Geiser *et al.*, 2013). A few members of this clade are known associates of Poaceae (Nelson *et al.*, 1994, Desjardins *et al.*, 2000b, Zeller *et al.*, 2003, Walsh *et al.*, 2010, Scauflaire *et al.*, 2011a). *Fusarium circinatum*, while being a well-known pine pathogen, is also reported to colonise members of the Poaceae family (grasses and maize) as an endophyte (Swett & Gordon, 2012, Swett & Gordon, 2015, Hernandez-Escribano *et al.*, 2018, Carter & Gordon, 2020). Pathogenesis in pines from *F. circinatum* is postulated to be secondary to the endophytic relationships on numerous Poaceae species as discussed in the next section (Swett & Gordon, 2009, Swett & Gordon, 2015, Swett *et al.*, 2016). The development of *F. circinatum* as a pathogen on pines is proposed to be a recent evolutionary adaptation, probably influenced by the overlapping geographic regions that pines and grasses share (Steenkamp *et al.*, 2002). The recent evolutionary adaptation is supported by the partial interfertility between *F. temperatum* (a maize pathogen) and *F. circinatum* (Desjardins *et al.*, 2000b, Steenkamp *et al.*, 2002).

GRASS-ASSOCIATED *FUSARIUM* WITHIN THE AMERICAN CLADE OF THE FFSC

The American clade contains numerous *Fusarium* species with vast host and geographic ranges (Table 1). *Fusarium* species, such as *F. temperatum* (Scauflaire *et al.*, 2011a), *F. subglutinans* (Desjardins *et al.*, 2006), *F. konzum* (Zeller *et al.*, 2003) and *F. werrikimbe* (Walsh *et al.*, 2010) in the American clade of the FFSC are phylogenetically closely related and all three associate with members of Poaceae. *Fusarium temperatum* infections are found on *Zea mays* in South Africa and Belgium (Scauflaire *et al.*, 2011b), Spain (Pintos *et al.*, 2013), Australia, Germany, Netherlands, Slovakia, Turkey (Susca *et al.*,

2013), China (Wang *et al.*, 2014), Poland (Czembor *et al.*, 2014), South Korea (Shin *et al.*, 2014), Argentina (Fumero *et al.*, 2015), Mexico (Robles-Barrios *et al.*, 2015), North America (Lanza *et al.*, 2016), Italy (Venturini *et al.*, 2016), France (Boutigny *et al.*, 2017) and Hungary (Molnár *et al.*, 2017). This pathogen causes malformation in maize seedlings and stem rot in maize (Varela *et al.*, 2013). *Fusarium temperatum* was initially classified as *F. subglutinans*, where it formed part of the cryptic species referred to as *F. subglutinans* group 1 (Steenkamp *et al.*, 2002). This pathogen was subsequently elevated to species level and described as *F. temperatum* (Scauflaire *et al.*, 2011a).

Fusarium subglutinans sensu stricto causes stalk and ear rot of maize and appears to be distributed all around the world, especially in South Africa (Desjardins *et al.*, 2000b), Spain (Pintos *et al.*, 2013), Germany, Slovakia, Portugal (Susca *et al.*, 2013), Poland (Czembor *et al.*, 2014), Argentina (Fumero *et al.*, 2015), Italy (Venturini *et al.*, 2016) and France (Boutigny *et al.*, 2017). Stalk rot reduces the growth patterns of maize but also results in rotted leaf sheaths and internal stalk tissue with characteristic brown streaks in the lower internodes. The internal stalk pith tissues of older plants changed colour from pink to salmon (Shaner & Scott, 1998) as a result of the colonisation of this fungus. The economic importance of *F. subglutinans* is as a consequence of causing premature death and stalk rot due to the obstruction in the translocation of water and nutrients to the leaves and developing ears of the plant.

Both *F. subglutinans* and *F. temperatum* are known to secrete toxic secondary metabolites or mycotoxins into the host plant during colonisation. These include beauvericin and moniliformin produced by *F. temperatum*, which is absent in *F. subglutinans* (Desjardins *et al.*, 2006, Moretti *et al.*, 2008, Scauflaire *et al.*, 2012). Fumonisin is produced by *F. temperatum* and *F. subglutinans* (Rheeder *et al.*, 2002, Stepien *et al.*, 2011), in contrast to what was previously thought (Nelson *et al.*, 1992, Proctor *et al.*, 2004). *Fusarium temperatum* seems to be the first *Fusarium* species in the American clade of the FFSC to produce enniatin homologues (Scauflaire *et al.*, 2012). These described mycotoxins accumulate in the plant and are a potential health danger for humans and animals consuming these plants (crops) if they occur in high enough concentrations (Desjardins *et al.*, 2006, Desjardins & Proctor, 2007). As a result of the infections on maize occurring from both *F. temperatum* and *F. subglutinans*, maize yields may decrease (Desjardins *et al.*, 2006, Desjardins & Proctor, 2007, Shin *et al.*, 2014).

Fusarium konzum was first identified in Kansas (USA) and isolated from native prairie grasses (Zeller *et al.*, 2003). Different strains of this fungus are known for their ability to synthesise mycotoxins, such as beauvericin, regardless of its non-pathogenic association with grass (Leslie *et al.*, 2004, Troncoso *et al.*, 2010). It is, therefore, considered that *F. konzum* might probably be an endophyte or, at the least, a potential pathogen of prairie grasses in Kansas (Zeller *et al.*, 2003).

The association of *Fusarium* species with their hosts and the genetic basis of the biological traits in these species remain largely unknown. Completed genome assemblies, therefore, provide many advantages in genomics and comparative genomics to understand the biology of host-specificity of species in the FFSC.

MECHANISMS OF GENOME CHANGE IN FUNGI

Fungal genomes are known for their diverse genomic architecture (Noble & Andrianopoulos, 2013). The plasticity of these genomes augments their dynamic nature, which is evidenced by numerous observable differences, such as dispensable chromosomes (Han *et al.*, 2001, Akagi *et al.*, 2009, Wittenberg *et al.*, 2009, Ma *et al.*, 2010), genome duplications (Feschotte, 2008, Kaessmann, 2010, Hua-Van *et al.*, 2011), as well as the partitioning of their genomes in genomic sub-compartments which differ in evolutionary rates (Van der Nest *et al.*, 2019). These genomic differences underpin some of the genetic factors underlying host-specificity, diversity and novel genetic features, most of which remain unknown.

Speciation is driven by chromosomal differentiation, as well as the accumulation of genomic mutations (Hu *et al.*, 2014). Within the genus *Fusarium*, chromosomal fusion events, as observed for *F. graminearum* (Cuomo *et al.*, 2007), and translocation events that occurred in *F. circinatum* and *F. temperatum* (De Vos *et al.*, 2014), have the potential to facilitate speciation. Some of the genomic features underlying this diversity, such as repetitive DNA and transposable elements, as well as the role of horizontal gene transfer and telomeres, will be discussed in the ensuing sections (additional information in Table 2).

Repetitive DNA

An important factor contributing to fungal diversity is the repetitive DNA content of genomes (Seidl & Thomma, 2014, Moller & Stukenbrock, 2017). Repetitive DNA is defined as segments of DNA that appear multiple times throughout the genome and can constitute more than half of the DNA content in the cell nucleus (Biscotti *et al.*, 2015). The distribution of these sequences is not equal within the genome but are rather conserved to specific genomic regions, such as the telomeric regions. The presence of repetitive sequences was shown to be involved in genome size expansion amongst diverse fungal lineages (Dean *et al.*, 2005, Spanu *et al.*, 2010, Raffaele & Kamoun, 2012). Repetitive DNA contributes towards both genome stability and plasticity through their involvement during gene duplication and recombination (Treangen *et al.*, 2009). These elements seem to be essential for genome function and stability since they contain coding and non-coding sequences that can be arranged in tandem repeats throughout the genome (Hall *et al.*, 2017). Their role during the recombination process is vital as the presence of these repeats results in increased rates of rearrangement, amplification and deletion of genetic material (Bzymek & Lovett, 2001, Biscotti *et al.*, 2015, Moller & Stukenbrock, 2017). These repeats can also alter genes during meiosis, create new combinations of pre-existing alleles, resulting in novel functions from pre-existing ones, and localise sequence diversification (Jacob, 1977). However, chromosomal rearrangements driven by repeats seem to lead to genome instability, resulting in plastic genomes (Bzymek & Lovett, 2001, Hall *et al.*, 2017).

Genome comparisons between several *Fusarium* species show that the repeat content can be as low as 6% and as high as 21%, as observed for *F. solani* and *F. oxysporum*, respectively (Coleman *et al.*, 2009, Stukenbrock & Croll, 2014). In other fungi this can also vary widely, *e.g.*, 4% in the necrotrophic *Botrytis cinerea* and 66% in the hemibiotrophic *Leptosphaeria maculans* (Gout *et al.*, 2007, Haas *et al.*, 2009, Amselem *et al.*, 2011). Pathogens exercising an obligate biotrophic or hemibiotrophic lifestyle seem to contain much higher DNA repeat contents compared to necrotrophic pathogens (Raffaele & Kamoun, 2012). Gene repertoire differences are possibly indicative of the putative role these repeats have in the biotrophic phase during host-pathogen interactions.

Transposable elements

Transposable elements are mobile segments of DNA that can move within and between genomes, the latter through horizontal gene transfer (Daniels *et al.*, 1990). Transposable elements are classified into two classes (class I and II) based on the structure and mode of transposition (Wicker *et al.*, 2007). The class I transposons are known as retrotransposons that contain either long terminal repeats or non-long terminal repeats. This mode of transposition is employed by reverse transcription and a copy-and-paste mechanism. Class I long terminal repeats seem to expand genome size by doubling the organism's copy number of the repeat during each transposition event (Elliott & Gregory, 2015). The second class, known as DNA transposons, is based on the cut-and-paste mechanism mediated by the transposase enzyme. Class II transposons also contain terminal inverted repeats.

The presence of transposable elements may result in intra- and inter-specific variability. Comparative genomics between *F. graminearum* and *F. oxysporum* f. sp. *lycopersici* revealed the effect of transposable elements on the structure and size of genomes (Stukenbrock & Croll, 2014). The genome expansion that occurred within *F. oxysporum* compared to *F. graminearum* was likely due to the activity of transposable elements (Ma *et al.*, 2010, King *et al.*, 2015). This type of expansion also occurred in the fungal species *Blumeria graminis* f. sp. *hordei* and *Magnaporthe grisea* (Dean *et al.*, 2005, Spanu *et al.*, 2010).

Transposable elements are crucial in mediating high rates of sequence mutations, chromosome rearrangements and ploidy changes (Stukenbrock & Croll, 2014). The insertion of transposable elements to new genomic locations may result in mutations and disrupt the regulatory sequence of a gene, resulting in deleterious effects (Raffaele & Kamoun, 2012). Such incidences occur when transposable elements are inserted in exons, which alters the reading frame and can directly affect the type of peptide encoded for, as well as inducing missense or nonsense mutations. These elements can also be inserted within introns that may induce alternative or novel splicing sites, the disruption of canonical splice sites and the introduction of polyadenylation signals (Deininger & Batzer, 1999, Batzer & Deininger, 2002, Chen *et al.*, 2005, Callinan & Batzer, 2006, Belancio *et al.*, 2008, Konkel & Batzer, 2010).

Alternatively, loss-of-function mutational changes can also be beneficial to a pathogen as it can promote pathogen fitness, resulting in selective advantage as reviewed by Raffaele & Kamoun (2012). An example of this is where an avirulent effector gene of a plant pathogen lost its function due to a mutational change within the gene, resulting in an opportunity for the pathogen to evade the recognition system of the specific plant immune receptor. Asexual species also seem to benefit from mutations of transposable elements by overcoming the lack of allele shuffling occurring during recombination (Stukenbrock & Croll, 2014). Chromosomal rearrangements in the asexual pathogen *Verticillium dahlia* and other pathogens seem to create genomic regions that are repetitive and lineage-specific but also contain genes involved in host-specificity and adaptation (Gout *et al.*, 2006, Fedorova *et al.*, 2008, De Jonge *et al.*, 2012, De Jonge *et al.*, 2013, Schardl *et al.*, 2013).

The effect of transposable elements on gene expression has also been investigated (Castanera *et al.*, 2016). Gene expression patterns differed in *F. graminearum* when the genes were under the influence of transposable elements. The expression of these genes was lower than those not under transposable element influence. The reason for this was due to the insertion of transposable elements within regions rich in introns which have shown to destabilise the mRNA and directly lead to a reduction in the levels of gene expression (Chen *et al.*, 2006).

Horizontal gene transfer

Horizontal gene transfer (HGT) allows for the stable integration of genetic material between species that are phylogenetically distant from one another and between cytoplasmic organelles and the nucleus in a manner where these genetic material are not inherited from parents to offspring (Doolittle *et al.*, 2003, Keeling & Palmer, 2008, Gao *et al.*, 2014). HGT, and by extension chromosome transfer, has the potential to affect host range and induce diversity amongst fungi (Friesen *et al.*, 2006, Zhang *et al.*, 2019). The availability of whole-genome sequences allows for genome comparisons to identify new DNA segments acquired and identify their possible origins (Friesen *et al.*, 2006).

Specific gene sequences are more abundant than others in the DNA that undergo HGT. It is thought that genes involved in intermediary and secondary metabolism undergo

higher frequencies of horizontal gene transfer compared to the genes directly involved in transcription, translation and replication (Jain *et al.*, 1999). A single event of gene transfer can also involve gene clusters that encode an entire metabolic pathway between distantly related fungi, as observed in *Aspergillus* and *Podospora* lineages (Slot & Rokas, 2011).

HGT generally occurs more frequently between organisms occupying the same niche, as observed from Slot & Rokas (2011). The newly acquired DNA sequences have proven to promote and improve the survival rate of fungi in novel niches but also improved their ability to confer pathogenicity on novel hosts. This latter occurrence was illustrated by Ma *et al.* (2010), where the chromosomal transfer from a pathogenic *F. oxysporum* strain to a non-pathogenic strain resulted in pathogenicity in the former strain. The horizontal transfer of a locus-specific gene in the genome of the wheat pathogen *Phaeosphaeria nodorum* to *Pyrenophora tritici-repentis* has been demonstrated (Ciuffetti *et al.,* 1997, Friesen *et al.,* 2006). *Pyrenophora tritici-repentis* causes tan spot disease on wheat, potentially due to the transfer of the *ToxA* gene, which encodes for a host-specific toxin, from *P. nodorum.* It is suspected that the fungal effector Ave1 in *Verticillium dahlia* has been transferred from plants to fungi since this gene is present in diverse plant pathogenic fungi and bacteria (De Jonge *et al.,* 2012).

HGT creates functional novelty in fungi, especially in *Fusarium* species from the FFSC where these genes promote fungal evolution, fitness or virulence through the emergence of species-specific traits (Stewart *et al.*, 2014, Chiara *et al.*, 2015, Glenn *et al.,* 2016, Van Wyk *et al.,* 2018). Gene clusters from *F. verticillioides* involved in the detoxification of antimicrobial compounds produced by maize were acquired from multiple external sources utilising the horizontal transfer of these genes (Stewart *et al.,* 2014, Glenn *et al.,* 2016). *Fusarium fujikuroi* isolates from different geographic regions have lineage-specific genes, with roles in host-pathogen interactions and adaptation to environmental changes, acquired through HGT (Chiara *et al.*, 2015). A similar trend was observed where five growth rate-determining genes in *F. circinatum* were acquired horizontally from various sources outside of the FFSC (Van Wyk *et al.,* 2018).

Telomeres

Telomeres are located at chromosomal ends and protect these ends against the loss of DNA during replication, as well as fusion with other chromosomes. The telomerase enzyme adds short repeated sequence motifs (TTAGGG/CCCTAA) to the chromosomal ends (Garcia-Pedrajas & Roncero, 1996). A subtelomere region exists between the telomere region and chromatin and contains multiple sequence segments at both chromosomal ends that share a high similarity (Tashiro *et al.*, 2017). The subtelomere region consists of two distinct domains, such as the distal and proximal domains (Flint *et al.*, 1997, Pryde *et al.*, 1997). The distal domain contains short, tandem repeat motifs at most or all chromosome ends, compared to the proximal domain containing longer duplicated segments of DNA that is less widely dispersed at these ends. These regions often contain clusters of related genes.

The subtelomeric gene clusters in microbial eukaryotes are often associated with niche adaptation. This association was observed in *Saccharomyces cerevisiae*, where genes involved in sugar utilisation were located within the subtelomeres (Denayrolles *et al.*, 1997). Human pathogens, such as *Plasmodium falciparum*, *Trypanosoma brucei* and *Pneumocystis carinii* also contain a variety of subtelomeric genes encoding for surface proteins that contribute to the host-pathogen interaction (Hernandez-Rivas *et al.*, 1997, Barry *et al.*, 2003, Keely *et al.*, 2005, Kutty *et al.*, 2013). These surface proteins also promote antigenic variation to facilitate the evasion of host immune responses. Subtelomeric regions in *Fusarium* tend to harbour genes enriched for secondary metabolite biosynthesis gene clusters contributing to fungal niche adaptation, host colonisation and accelerated growth rate (Wiemann *et al.*, 2013, Sieber *et al.*, 2014, Zhao *et al.*, 2014).

In conclusion, comparative genomics allows for observing various evolutionary trends amongst diverse fungal lineages (Raffaele & Kamoun, 2012). It is speculated that genome expansions directly influence the formation of unique genes and are considered as a potential driving force behind emerging fungal pathogens (Coleman *et al.*, 2009, Ma *et al.*, 2010, Croll & McDonald, 2012, Zhao *et al.*, 2014, Dong *et al.*, 2015). These are often located near telomeric regions which are highly variable and capable of influencing the biology of pathogens (Spanu *et al.*, 2010, Raffaele & Kamoun, 2012). Furthermore, the

repetitive and transposable elements can often be located in telomeric regions which allow for alterations in the sizes and shapes of chromosomes (Raskina *et al.*, 2008).

FUNGAL PATHOGENICITY FACTORS

Fungal survival depends on the host-pathogen interaction to survive and to thrive. Underlying molecular factors that promote host-pathogen interactions include cell wall degrading enzymes (CAZymes), secondary metabolites, peptidases and effector proteins (Tyler *et al.*, 2006, Haas *et al.*, 2009). All these molecular factors contribute to pathogen adaptation and survival by inferring fungal virulence (some information is also captured in Table 2).

Carbohydrate-Active enZymes (CAZymes)

Carbohydrate-Active enZymes (CAZymes) are involved in carbohydrate metabolism, which is responsible for the synthesis, degradation and recognition of carbohydrates (Lombard *et al.*, 2014). Carbohydrates from the plant host bind together and form polysaccharides (*e.g.*, cellulose, xylan, pectin) located within plant cell walls and can be degraded by CAZymes, specific for a polysaccharide, upon pathogen entry (Goubet *et al.*, 2002, Lombard *et al.*, 2014). These enzymes act to weaken the cell walls, not only to promote fungal penetration and infection but also to make nutrients available for the invading pathogen (An *et al.*, 2005, Cantarel *et al.*, 2009).

Several classes of CAZymes exist, *i.e.*, glycoside transferases (GTs) forming glycosidic bonds, glycoside hydrolases (GHs) for the hydrolysis of glycosidic bonds between carbohydrates (Cantarel *et al.*, 2009), polysaccharide lyases (PLs) for the degradation of glycosaminoglycans and pectin (Yip & Withers, 2006, Cantarel *et al.*, 2009), carbohydrate esterases (CEs) for the catalysis of esters or amides (Biely, 2012) and carbohydrate-binding modules (CBMs) for the degradation of insoluble polysaccharides (Boraston *et al.*, 2004). Together, these provide a repertoire of enzymes available to pathogens to assist in fungal growth and survival.

Genomic analyses revealed that the genome of *F. graminearum* encodes for 109 CAZymes (Brown *et al.*, 2012, Albarouki *et al.*, 2014, Heard *et al.*, 2015). Of these, CAZymes such as cellulases, xylanases and pectinases, are produced during the early stages of wheat spike infection. The initial transcript abundance during symptomless wheat spike infection was much lower compared to what was observed in symptomatic wheat spikes. This low CAZyme transcript abundance was potentially due to the limited availability of sugars. The increased CAZyme transcript abundance likely occurred when the pathogen was exposed to nutrient deprivation (Brown *et al.*, 2017). These enzymes alter the cell wall components in infected tissue, releasing nutrients, and thus emphasise that CAZymes are indeed required for successful host colonisation (Brown *et al.*, 2010, Brown *et al.*, 2017).

Fungal necrotrophs possess more genes encoding for CAZymes compared to the gene repertoire of saprophytic fungi (Liao *et al.*, 2013) and biotrophs (Zhao *et al.*, 2013). These genic differences might be because necrotrophs require CAZymes, not only for virulence on the host (Hoffmeister & Keller, 2007, Sperschneider *et al.*, 2016), but also for cell death (Sella *et al.*, 2013). Another possible example of this occurrence has been illustrated within *F. graminearum* during plant infection (Zhao *et al.*, 2013). This study specifically focussed on the expression of genes in the CEs, PLs and GHs CAZyme classes. The authors found that most genes were up-regulated during plant infection. They also found that some genes are only up-regulated during germination of spores with no difference in expression during plant infection, indicating that these genes are more critical for conidiation and germination. In contrast, some genes were down-regulated during germination and up-regulated during plant infection, indicating that these genes are not relevant during conidiation and germination.

Secondary metabolites

Most *Fusarium* species can synthesise secondary metabolites ranging in toxicity levels and typically have a low molecular weight (Vinnig, 1990). These metabolites are organic compounds that do not actively contribute to primary metabolism (*e.g.*, pathogen growth) but do, however, participate in the survival capability of a pathogen (Mylonakis *et al.*, 2007). The general function of secondary metabolites differs. They can promote the attractiveness of the host plant towards animals for pollination or seed dispersal. In some

cases, they promote defence mechanisms during plant-pathogen interactions by contributing to pathogen virulence. These metabolites have been utilised in industrial applications for the production of pigments, drugs and insecticides (Yim *et al.*, 2007, Fox & Howlett, 2008, Scharf *et al.*, 2014).

Enzymes known to be involved in secondary metabolite synthesis include the non-ribosomal peptide synthetases (NRPSs), polyketide synthases (PKSs), terpene synthases (TSs; terpenes and indole terpenes) and a class "other" given to those that are a mixture of the other classes whose identity can be confirmed by BLAST searches (Medema *et al.*, 2011, Weber *et al.*, 2015). Some examples of secondary metabolites within each class are known. For example, the non-ribosomal class is responsible for the synthesis of sirodesmin, peramine and siderophores, whereas the polyketide class contains the aflatoxins and fumonisins (Fox & Howlett, 2008). The T-2 toxin and deoxynivalenol are examples of secondary metabolites within the terpene class, while the indole terpene class contains paxilline and lolitrems (Fox & Howlett, 2008).

The genes encoding for the enzymes involved in the synthesis of a specific secondary metabolite are typically clustered in fungi (Maplestone *et al.*, 1992). The clustering of genes holds some selective advantage over those genes not being clustered. Clustering of genes promotes the co-regulation of a set of genes involved in a biosynthetic pathway due to their response to a common transcription factor and the ability of such a gene set to be co-inherited by either horizontal or vertical gene transfer (Osbourn, 2010). Similarly, clustering of secondary metabolite gene clusters has been documented for a variety of *Fusarium* species and other filamentous fungi (Sieber *et al.*, 2014, Hoogendoorn *et al.*, 2018).

Genomic comparisons have revealed a range of secondary metabolites produced by the genus *Fusarium* (Ma *et al.*, 2013). Secondary metabolites are an essential component of pathogenicity within this genus, as observed for *F. graminearum* as an example (Nicholson *et al.*, 2003, Bourdages *et al.*, 2006, Nogueira *et al.*, 2018). Mycotoxins form the largest group of secondary metabolites produced by these pathogenic fungi (Nelson *et al.*, 1994, Desjardins, 2003). The toxicity of these secondary metabolites has proven to impart a negative effect on humans and animals if consumed in significant amounts (Nelson *et al.*, 1994, Desjardins, 2003), while others can be used for pharmaceutical

purposes instead of only inducing pathogenesis (Porter *et al.*, 1990). Some examples of secondary metabolites previously discovered to be synthesised by *Fusarium* species include aflatoxins, fusaric acid, trichothecenes, fumonisins and zearalenones (Keller & Hohn, 1997, Desjardins *et al.*, 2000a, Yu & Keller, 2005, Summerell & Leslie, 2011).

Effectors

Effector proteins are known as small secretory proteins containing cysteine-rich regions (Rep *et al.*, 2004). These proteins selectively bind to proteins secreted by the host during pathogen attack. They can also regulate the biological activity of host proteins by increasing/decreasing enzyme activity, plant cell signalling or gene expression (Chisholm *et al.*, 2006, Jones & Dangl, 2006, Da Cunha *et al.*, 2007, He *et al.*, 2007, Tameling & Takken, 2008, Zhou & Chai, 2008). These proteins are capable of entering host cells and interfere with the host's basal immune responses, resulting in successful fungal colonisation (Jones & Dangl, 2006, Kamoun, 2006, Presti *et al.*, 2015). These proteins actively facilitate infection by the pathogen to derive nutrients from the host plant (Bielska *et al.*, 2014, Garnica *et al.*, 2014). Those effectors recognised by the host's resistance proteins are referred to as avirulence proteins (Avr) (Buttner & Bonas, 2002), from which some are expressed in vascular tissue or secreted in xylem (*SIX*), and target cellular processes (*e.g.*, signal transduction, transcription and vesical trafficking) (Da Cunha *et al.*, 2007, He *et al.*, 2007, Zhou & Chai, 2008).

Fusarium oxysporum f. sp. *lycopersici* is known for its release of effector/avirulence genes (*Avr*) during the induction of vascular wilt in tomatoes (Armstrong & Armstrong, 1981, Huang & Lindhout, 1997, Di Pietro *et al.*, 2003). This pathogen secretes proteins, such as Avr3 and Avr2, in the host xylem during colonisation which is known to promote virulence. The genes for these proteins are present in strains causing tomato wilt but absent in the strains that are non-pathogenic on tomatoes. These genes are located on chromosome 14 that is a known lineage-specific chromosome in *F. oxysporum*. Another study was performed on the secretion of effector proteins by *F. graminearum* on wheat, where the authors analysed the transcriptome of this fungus during symptomless and symptomatic stages of infection (Brown *et al.*, 2017). During investigations into the progression of symptomless infection to symptomatic infection, there was an increase in the number of effector proteins expressed. The increase in putative effector proteins may

be due to the need to induce host cell death, hence the visible symptoms of the disease. Thus, the arsenal of virulence factors encoded by plant pathogens is thought to be an essential factor for successful host colonisation.

Effector proteins have vital roles in pathogenicity of various fungi in the FFSC, enabling them to manipulate their host for successful colonisation (Brown *et al.*, 2014, Li *et al.*, 2019). In *F. proliferatum* these proteins were found to be potentially involved in the invasion of banana fruit, cell wall degradation, metabolic and biological processes, but also likely enabled the fungus to respond to stress induced by the banana (Li *et al.*, 2019). In *F. verticillioides*, these proteins allowed the transition of the fungus from being a saprophyte on maize stalks to grow as an endophyte or parasite within the plant (Brown *et al.*, 2014, Niu *et al.*, 2015). These proteins have variable roles during different infection stages.

Peptidases

Peptidases are essential enzymes that participate in peptide bond cleavage, resulting in smaller proteins/polypeptides or single amino acids as yield (Hamin Neto *et al.*, 2018). These enzymes are essential for regulating physiological processes, such as fertilisation, embryogenesis, cell signalling and immune responses, fungal morphogenesis and metabolism (Ng *et al.*, 2009, Yike, 2011). These hydrolytic enzymes have essential roles in the turnover of proteins, as well as in the degradation of damaged, misfolded and harmful proteins (Yike, 2011). Free amino acids are released upon the hydrolysis of peptide bonds, driving the synthesis of new proteins. These enzymes exert minor proteolysis at highly specific sites for the post-translational modification of proteins and also enable enzymes to mature and participate in protein assembly and subcellular targeting (Schaller, 2004).

In *Candida albicans*, peptidases seem to play a role in pathogenicity upon colonisation and disease progression (Naglik *et al.*, 2003, Schaller *et al.*, 2005, Pietrella *et al.*, 2010). The presence of these enzymes promotes the invasion of host tissue, releasing nutrients utilised for fungal growth and alters the host immune responses. Peptidases in *Conidiobolus coronatus* promotes morphogenesis, germination and the release of conidia for the dispersal of this asexual fungus (Phadatare *et al.*, 1992, Reichard *et al.*, 2000,

Papagianni & Moo-Young, 2002). Furthermore, the inactivation of fungal peptidases by plant hosts results in a deviation in the germination and growth patterns of fungi (Dunaevskii *et al.*, 2005), pointing to the crucial role of peptidases in the development, adaptation and survival of pathogens.

The synthesis of peptidases in *Fusarium* is not uncommon (Lanubile *et al.*, 2014, Lowe *et al.*, 2015, Sharma *et al.*, 2016). The transcriptome of *F. oxysporum* f. sp. *ciceris* contain a vast arsenal of genes encoding for several factors, including peptidases, which are up-regulated at a later stage of host infection to enhance pathogenicity (Sharma *et al.*, 2016). *Fusarium graminearum* secrete peptidases for the breakdown of the structural component of the plant cell wall and the acquisition of nutrients (Lowe *et al.*, 2015). Several peptidases were expressed in *F. verticillioides* during the colonisation of maize, which emphasises the role of these proteins during pathogen invasions (Lanubile *et al.*, 2014).

COMPARTMENTALISATION OF FUNGAL GENOMES

Genome comparisons are entirely dependent on accurate genome assemblies and annotations that improve the understanding of gene functions. In the post-genomics era, comparative genomics is a fundamental part of determining the factors influencing pathogenicity of a fungal species. The diverse underlying mechanisms of pathogenicity can be explained by the genes involved in each fungal species (Ma *et al.*, 2010), *i.e.*, the species-specific genes. The organisation of a genome can give valuable information in terms of how genes involved in fungal pathogenicity are related across several fungal species which will also enable downstream studies to determine how molecular diversity and genetic transmission are mediated (Rep & Kistler, 2010). Information regarding the differences in genomes can be improved as more genome sequence data become known.

The structure and organisation of a genome provide information such as the size of the genome, the number of genes present within the genome and also the density of these genes (Croll & McDonald, 2012, Ma *et al.*, 2013). It was previously thought that more complex pathogens require more genes and simple pathogens, fewer genes (Thomas, 1971, Petrov, 2001). This hypothesis was rejected since some simple pathogens have

more genes than complex pathogens (Thomas, 1971, Shahin *et al.*, 2012). Instead, the size of the genome is dependent on the developmental and ecological need of the pathogen. Genomes vary in size, not due to pathogen complexity, but rather due to the organism's need for adaptation to novel niches and natural selection (Petrov, 2001). Some pathogens will thus require more proteins and enzymes encoded by genes to promote a pathogenic lifestyle, where these genes may be silent/absent in those pathogens that do not live this sophisticated lifestyle. Employing comparative genomics is, thus, essential for understanding how host-pathogen interactions are mediated and also how different pathogen species can adapt to specific niches (Waalwijk *et al.*, 2017). Comparative genomics could portray the presence or absence of genes in pathogens and enhance our understanding of the different gene repertoires in different fungal species.

Traditional genome comparisons performed on *Fusarium* species revealed that these fungal genomes could be divided into two compartments, known as the core and accessory compartment (Croll & McDonald, 2012, Dong *et al.*, 2015). These two compartments differ from one another in terms of evolutionary rates, gene repertoire and the levels of gene expression (Ma *et al.*, 2013, Waalwijk *et al.*, 2017). These compartments consist of genes/regions that are common to all individuals of a species (core) and those that are dispensable (or lineage-specific) or present in only some individuals (accessory) (Stukenbrock & Croll, 2014). The core compartment consists of core chromosomes/chromosomal regions containing all the housekeeping genes involved in primary metabolism and which also evolve at slower rates due to the high gene density in this region (Croll & McDonald, 2012, Raffaele & Kamoun, 2012).

Fusarium accessory genomes contain regions that may span one or more entire chromosomes or may be restricted to specific regions of one or more chromosomes (Coleman *et al.*, 2009, Ma *et al.*, 2010). These regions are subjected to horizontal gene/chromosome transfer to other chromosomes or between individuals within a species and are also known to have higher evolutionary rates as opposed to the genetic material present in the core compartment (Croll & McDonald, 2012, Raffaele & Kamoun, 2012). Some regions contain genes that participate in ecological niche exploitation, thus contributing to host specialisation, pathogenicity and virulence. These regions may also be lineage-specific, as observed in *F. oxysporum*, where it contributes to the evolution of

these devastating pathogens (Hatta *et al.*, 2002, Ma *et al.*, 2010, Croll & McDonald, 2012, Galazka & Freitag, 2014).

A comparative study between three species revealed that the genome of *F. oxysporum* f. sp. *lycopersici* is 44% bigger than the genome of *F. verticillioides* and 65% bigger than *F. graminearum* (Ma *et al.*, 2010). A total of 9000 conserved syntenic orthologs were present across all genomes that were enriched for transcription factors, hydrolytic enzymes and transmembrane transporters. These 9000 conserved orthologs were considered as the core genome of these three genomes. The remaining four out of 15 chromosomes of *F. oxysporum* that could not be mapped back to the genome of *F. verticillioides* were considered as chromosomes in the accessory genome. These chromosomes are often referred to as supernumerary chromosomes and are also well-known for the presence of lineage-specific regions and the different transposons it harbours (Huang *et al.*, 2016).

Comparative genomics, therefore, reveals that core genomes between species share high levels of synteny. The regions that are not conserved and unique to a species contain different gene content in terms of the number of genes and repetitive sequences. These non-conserved regions can also contain the lineage-specific regions as seen in *Fusarium* species, such as *F. oxysporum* and *F. solani.* The genomic regions that are not well-conserved are known to contain the genes participating in host-pathogen interactions (Coleman *et al.*, 2009, Ma *et al.*, 2010).

CONCLUSIONS

Comparative genomic studies in *Fusarium* are advancing as more whole-genome sequences become available and aids in the understanding of underlying mechanisms contributing to biological traits, such as pathogenicity and host-specificity (Ma *et al.*, 2010, Gardiner *et al.*, 2012). Currently, the available *Fusarium* genomes from the FFSC include *F. fujikuroi* (Wiemann *et al.*, 2013), *F. temperatum* (Wingfield *et al.*, 2015b), *F. verticillioides* (Ma *et al.*, 2010), *F. mangiferae* (Wiemann *et al.*, 2013), *F. circinatum* (Wingfield *et al.*, 2012), *F. nygamai* (Wingfield *et al.*, 2015a), *F. agapanthi* (Edwards *et*

al., 2016), *F. proliferatum* (Niehaus *et al.,* 2016), *F. pininemorale* (Wingfield *et al.,* 2017), *F. fracticaudum* (Wingfield *et al.*, 2018a) and *F. udum* (Srivastava *et al.*, 2018), with some species having numerous strains sequenced.

By using the genomes available, in-depth studies can be performed on the structure, function and regulation of the genes involved in host-specificity and pathogenicity in existing and new fungal genomes by performing comparative analyses. The effect of repetitive sequences on genome size and function can be investigated, along with the occurrence of pseudogenes, segmental duplications, tandem repeats and transposable elements. The factors contributing to the virulence of *Fusarium* species, such as the cell wall degrading enzymes (CAZymes), secondary metabolites, effector proteins and peptidases synthesized can be identified. The arsenal of these genes within a genome can be compared with genomes of pathogenic and non-pathogenic fungi.

The objective of the research in this dissertation is to identify genes and processes involved in host-specificity between two groups of *Fusarium* species within the American clade of the FFSC. The research aims to understand the contribution of these genes towards the biology of these two assemblages of species based on their respective hosts. The dissertation will also interrogate the ancestral origin and divergence of these genes, as well as their genomic location. Identification of these genes will infer an understanding of the molecular mechanisms of the processes involved in host-specificity within these selected *Fusarium* species, providing vital information on the biology of these economically important fungi. The provision of data retrieved from this study will be available for future use in comparative genomics.

REFERENCES

1. Akagi Y, Akamatsu H, Otani H & Kodama M (2009) Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell* **8**: 1732-1738.

2. Al-Hatmi AM, Mirabolfathy M, Hagen F, Normand AC, Stielow JB, Karami-Osbo R, *et al.* (2016) DNA barcoding, MALDI-TOF, and AFLP data support *Fusarium ficicrescens* as a distinct species within the *Fusarium fujikuroi* species complex. *Fungal Biology* **120**: 265-278.

3. Alabouvette C (1986) Fusarium-wilt suppressive soils from the Châteaurenard region: Review of 10-years study. *Agronomie* **6**: 273-284.

4. Albarouki E, Schafferer L, Ye F, Von Wirén N, Haas H & Deising HB (2014) Biotrophy-specific downregulation of siderophore biosynthesis in *Colletotrichum graminicola* is required for modulation of immune responses of maize. *Molecular Microbiology* **92**: 338-355.

5. Alonso R & Bettucci L (2009) First report of the pitch canker fungus *Fusarium circinatum* affecting *Pinus taeda* seedlings in Uruguay. *Australasian Plant Disease Notes* **4**: 91-92.

6. Amselem J, Cuomo CA, Van Kan JAL, Viaud M, Benito EP, Couloux A, *et al.* (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7**: 26-37.

7. An HJ, Lurie S, Greve LC, Rosenquist D, Kirmiz C, Labavitch JM, *et al.* (2005) Determination of pathogen-related enzyme action by mass spectrometry analysis of pectin breakdown products of plant cell walls. *Analytical Biochemistry* **338**: 71-82.

8. Appel DJ & Gordon TR (1995) Intraspecific variation within populations of *Fusarium oxysporum* based on RFLP analysis of the Intergenic Spacer Region of the rDNA. *Experimental Mycology* **19**: 120-128.

9. Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, *et al.* (2018) Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports* **8**: 13530.

10. Armstrong GM & Armstrong JK (1981) *Formae speciales* and races of *Fusarium oxysporum* causing wilt diseases. *Fusarium: Diseases, Biology and Taxonomy*, (Cook RJ, ed.), pp. 391-399. Pennsylvania State University Press, University Park, PA.

11. Aylward J, Steenkamp ET, Dreyer LL, Roets F, Wingfield BD & Wingfield MJ (2017) A plant pathology perspective of fungal genome sequencing. *IMA Fungus* **8**: 1-15.

12. Bacon CW & Yates IE (2006) Endophytic root colonization by *Fusarium* species: Histology, plant interactions, and toxicity. *Microbial Root Endophytes*, (Schulz BJE, Boyle CJC & Sieber TN, eds.), pp. 133-152. Springer Berlin Heidelberg, Berlin, Heidelberg.

13. Barrows-Broaddus J (1990) Colonization of cones and seed of loblolly pine following inoculation with *Fusarium subglutinans*. *Plant Disease* **74**: 1002-1005.

14. Barry JD, Ginger ML, Burton P & McCulloch R (2003) Why are parasite contingency genes often associated with telomeres? *International Journal of Parasitology* **33**: 29-45.

15. Bashyal BM, Rawat K, Sharma S, Kulshreshtha D, Krishnan SG, Singh AK, *et al.* (2017) Whole genome sequencing of *Fusarium fujikuroi* provides insight into the role of secretory proteins and cell wall degrading enzymes in causing bakanae disease of rice. *Frontiers in Plant Science* **8**: 2013.

16. Batzer MA & Deininger PL (2002) *Alu* repeats and human genomic diversity. *Nature Reviews in Genetics* **3**: 370-379.

17. Bayley AD & Blakeway F (2002) Deployment strategies to maximise value recovery from tree improvement: The experience of two South African companies. *Southern African Forestry Journal* **195**: 11-22.

18. Belancio VP, Hedges DJ & Deininger P (2008) Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Research* **18**: 343-358.

19. Bertero A, Moretti A, Spicer LJ & Caloni F (2018) *Fusarium* molds and mycotoxins: Potential species-specific effects. *Toxins* **10**: 244.

20. Bielska E, Higuchi Y, Schuster M, Steinberg N, Kilaru S, Talbot NJ, *et al.* (2014) Long-distance endosome trafficking drives fungal effector production during plant infection. *Nature Communications* **5**: 5097.

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

22. Biscotti MA, Olmo E & Heslop-Harrison JS (2015) Repetitive DNA in eukaryotic genomes. *Chromosome Research: Chromatin, Chromosomes and Genomes* **23**: 415-420.

23. Bloomberg CW (1981) Diseases caused by *Fusarium* in forest nurseries. *Fusarium: Diseases, Biology and Taxonomy*, (Nelson PE, Toussoun TA & Cook RJ, eds.), pp. 178-187. Pennsylvania State University Press, University Park, PA.

24. Boraston AB, Bolam DN, Gilbert HJ & Davies GJ (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal* **382**: 769-781.

25. Boshoff W, Pretorius Z, Swart W & Jacobs A (1996) A comparison of scab development in wheat infected by *Fusarium graminearum* and *Fusarium crockswellense*. *Phytopathology* **86**: 558-563.

26. Bottalico A & Perrone G (2002) Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. *European Journal of Plant Pathology* **108**: 611-624.

27. Bourdages JV, Marchand S, Belzile FJ & Rioux S (2006) Diversity and prevalence of *Fusarium* species from Quebec barley fields. *Canadian Journal of Plant Pathology* **28**: 419-425.

28. Boutigny A, Scauflaire J, Ballois N & Ioos R (2017) *Fusarium temperatum* isolated from maize in France. *European Journal of Plant Pathology* **148**: 997-1001.

29. Bragança H, Diogo E, Moniz F & Amaro P (2009) First report of pitch canker on pines caused by *Fusarium circinatum* in Portugal. *Plant Disease* **93**: 1079.

30. Brown DW, Busman M & Proctor RH (2014) *Fusarium verticillioides SGE1* is required for full virulence and regulates expression of protein effector and secondary metabolite biosynthetic genes. *Molecular Plant-Microbe Interactions* **27**: 809-823.

31. Brown NA, Antoniw J & Hammond-Kosack KE (2012) The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: A refined comparative analysis. *PLoS One* **7**: e33731.

32. Brown NA, Evan J, Mead A & Hammond-Kosack K (2017) A spatial temporal analysis of the *Fusarium graminearum* transcriptome during symptomless and symptomatic wheat infection. *Molecular Plant Pathology* **18**: 1295-1312.

33. Brown NA, Urban M, Van de Meene AML & Hammond-Kosack KE (2010) The infection biology of *Fusarium graminearum*: Defining the pathways of spikelet to spikelet colonisation in wheat ears. *Fungal Biology* **114**: 555-571.

34. Burgess TI (1981) General ecology. *Fusarium: Diseases, Biology and Taxonomy*, (Nelson PE, Toussoun TA & Cook RJ, eds.), pp. 225-235. Pennsylvania State University Press, University Park, PA.

35. Buttner D & Bonas U (2002) Port of entry- The type III secretion translocon. *Trends in Microbiology* **10**: 186-192.

36. Bzymek M & Lovett ST (2001) Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 8319-8325.

37. Callinan PA & Batzer MA (2006) Retrotransposable elements and human disease. *Genome Dynamics* **1**: 104-115.

38. Cantarel B, Coutinho P, Rancural C, Bernard T, Lombard V & Henrissat B (2009) The Carbohydrate-Active Enzymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research* **37**: D233-D238.

39. Carlucci A, Colatruglio L & Frisullo S (2007) First report of pitch canker caused by *Fusarium circinatum* on *Pinus halepensis* and *P. pinea* in Apulia (Southern Italy). *Plant Disease* **91**: 1683.

40. Carter JP, Rezanoor HN, Holden D, Desjardins AE, Plattner RD & Nicholson P (2002) Variation in pathogenicity associated with the genetic diversity of *Fusarium graminearum*. *European Journal of Plant Pathology* **108**: 573-583.

41. Carter JW & Gordon T (2020) Infection of the native California grass, *Bromus carinatus*, by *Fusarium circinatum*, the cause of pitch canker in pines. *Plant Disease* **104**: 194-197.

42. Castanera R, López-Varas L, Borgognone A, LaButti K, Lapidus A, Schmutz J, *et al.* (2016) Transposable elements versus the fungal genome: Impact on whole-genome architecture and transcriptional profiles. *PLoS Genetics* **12**: e1006108.

43. Chang DC, Grant GB, O'Donnell K, Wannemuehler KA, Noble-Wang J, Rao CY, *et al.* (2006) Multistate outbreak of *Fusarium* keratitis associated with use of a contact lens solution. *American Journal of Ophthalmology* **142**: 896-897.

44. Chen JM, Férec C & Cooper DN (2006) LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: Mutation detection bias and multiple mechanisms of target gene disruption. *Journal of Biomedicine and Biotechnology* **2006**: 56182.

45. Chen JM, Stenson PD, Cooper DN & Férec C (2005) A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Human Genetics* **117**: 411-427.

46. Chiara M, Fanelli F, Mule G, Logrieco AF, Pesole G, Leslie JF, *et al.* (2015) Genome sequencing of multiple isolates highlights subtelomeric genomic diversity within *Fusarium fujikuroi*. *Genome Biology and Evolution* **7**: 3062-3069.

47. Chisholm ST, Coaker G, Day B & Staskawicz BJ (2006) Host-microbe interactions: Shaping the evolution of the plant immune response. *Cell* **124**: 803-814.

48. Ciuffetti LM, Tuori RP & Gaventa JM (1997) A single gene encodes a selective toxin causal to the development of tan spot of wheat. *Plant Cell* **9**: 135-144.

49. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, *et al.* (2009) The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* **5**: e1000618.

50. Correll JC (1991) Pitch canker disease in California: Pathogenicity, distribution and canker development on Monterey pine (*Pinus radiata*). *Plant Disease* **75**: 676-682.

51. Croll D & McDonald BA (2012) The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens* **8**: e1002608.

52. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, *et al.* (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400-1402.

53. Czembor E, Stepień Ł & Waśkiewicz A (2014) *Fusarium temperatum* as a new species causing ear rot on maize in Poland. *Plant Disease* **98**: 1001.

54. Da Cunha L, Sreerekha MV & Mackey D (2007) Defense suppression by virulence effectors of bacterial phytopathogens. *Current Opinion in Plant Biology* **10**: 349-357.

55. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG & Chovnick A (1990) Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. *Genetics* **124**: 339-355.

56. De Jonge R, Bolton MD, Kombrink A, Van den Berg GCM, Yadeta KA & Thomma BPHJ (2013) Extensive chromosomal reshuffling drives evolution of virulence in an asexual pathogen. *Genome Research* **23**: 1271-1282.

57. De Jonge R, Van Esse PH, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, *et al.* (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 5110-5115.

58. De Vos L, Myburg AA, Wingfield MJ, Desjardins AE, Gordon TR & Wingfield BD (2007) Complete genetic linkage maps from an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Genetics and Biology* **44**: 701-714.

59. De Vos L, Steenkamp ET, Martin SH, Santana QC, Fourie G, Van der Merwe NA, *et al.* (2014) Genome-wide macrosynteny among *Fusarium* species in the *Gibberella fujikuroi* complex revealed by amplified fragment length polymorphisms. *PLoS One* **9**: e114682.

60. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, *et al.* (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980-986.

61. Deininger PL & Batzer MA (1999) *Alu* repeats and human disease. *Molecular Genetics and Metabolism* **67**: 183-193.

62. Denayrolles M, De Villechenon EP, Lonvaud-Funel A & Aigle M (1997) Incidence of SUC-RTM telomeric repeated genes in brewing and wild wine strains of *Saccharomyces*. *Current Genetics* **31**: 457-461.

63. Desjardins AE (2003) *Gibberella* from *A*(*venaceae*) to *Z*(*eae*). *Annual Review of Phytopathology* **41**: 177-198.

64. Desjardins AE, Bai G, Plattner RD & Proctor RH (2000a) Analysis of aberrant virulence of *Gibberella zeae* following transformation-mediated complementation of a trichothecene-deficient (*Tri5*) mutant. *Microbiology* **146**: 2059-2068.

65. Desjardins AE, Maragos CM & Proctor RH (2006) Maize ear rot and moniliformin contamination by cryptic species of *Fusarium subglutinans*. *Journal of Agricultural and Food Chemistry* **54**: 7383-7390.

66. Desjardins AE, Plattner RD & Gordon TR (2000b) *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. *Mycological Research* **104**: 865-872.

67. Desjardins AE & Proctor RH (2007) Molecular biology of *Fusarium* mycotoxins. *International Journal of Food Microbiology* **119**: 47-50.

68. Di Pietro A, Madrid MP, Caracuel Z, Delgado-Jarana J & Roncero MIG (2003) *Fusarium oxysporum:* Exploring the molecular arsenal of a vascular wilt fungus *Molecular Plant Pathology* **4**: 315-325.

69. Dita M, Barquero M, Heck D, Mizubuti ESG & Staver CP (2018) Fusarium wilt of banana: Current knowledge on epidemiology and research needs toward sustainable disease management. *Frontiers in Plant Science* **9**: 1468.

70. Dong S, Raffaele S & Kamoun S (2015) The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development* **35**: 57-65.

71. Doolittle WF, Boucher Y, Nesbø CL, Douady CJ, Andersson JO & Roger AJ (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philosophical Transactions of the Royal Society of London* **358**: 39-58.

72. Dunaevskii YE, Tsybina TA, Belyakova GA, Domash VI, Sharpio TP, Zabreiko SA, *et al.* (2005) Proteinase inhibitors as antistress proteins in higher plants. *Applied Biochemistry and Microbiology* **41**: 344-348.

73. Dwinell LD, Barrows-Broaddus J & Kuhlman EG (1985) Pitch canker: A disease complex of southern pines. *Plant Disease* **69**: 270-276.

74. Dwinell LD, Kuhlman EG & Blakeslee GM (1981) Pitch canker of southern pines. *Fusarium: Diseases, Biology and Taxonomy*, (Nelson PE, Toussoun TA & Cook RJ, eds.), pp. 188-194. Pennsylvania State University Press, University Park, PA.

75. Edwards J, Auer D, De Alwis SK, Summerell B, Aoki T, Proctor RH, *et al.* (2016) *Fusarium agapanthi* sp. nov., a novel bikaverin and fusarubin-producing leaf and stem spot pathogen of *Agapanthus praecox* (African lily) from Australia and Italy. *Mycologia* **108**: 981-992.

76. EFSA (2010) Risk assessment of *Gibberella circinata* for the EU territory and identification and evaluation of risk management options. *European Food Safety Authority Journal* **8**: 1620.

77. Elliott TA & Gregory TR (2015) Do larger genomes contain more diverse transposable elements? *BMC Evolutionary Biology* **15**: 69.

78. Evira-Recuenco M, Iturritxa E & Raposo R (2015) Impact of seed transmission on the infection and development of pitch canker disease in *Pinus radiata*. *Forests* **6**: 3353-3368.

79. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, *et al.* (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genetics* **4**: e1000046.

80. Ferrigo D, Raiola A & Causin R (2016) *Fusarium* toxins in cereals: Occurrence, legislation, factors promoting the appearance and their management. *Molecules* **21**: 627.

81. Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nature Reviews in Genetics* **9**: 397-405.

82. Flint J, Bates GP, Clark K, Dorman A, Willingham D, Roe BA, *et al.* (1997) Sequence comparison of human and yeast telomeres identifies structurally distinct subtelomeric domains. *Human Molecular Genetics* **6**: 1305-1314.

83. Forsyth LM, Smith LJ & Aitken EAB (2006) Identification and characterization of non-pathogenic *Fusarium oxysporum* capable of increasing and decreasing Fusarium wilt severity. *Mycological Research* **110**: 929-935.

84. Fox EM & Howlett BJ (2008) Secondary metabolism: Regulation and role in fungal biology. *Current Opinion in Microbiology* **11**: 481-487.

85. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, *et al.* (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics* **38**: 953-956.

86. Fuchs JG, Moënne-Loccoz Y & Défago G (1997) Nonpathogenic *Fusarium oxysporum* strain Fo47 induces resistance to Fusarium wilt in tomato. *Plant Disease* **81**: 492-496.

87. Fumero MV, Reynoso MM & Chulze S (2015) *Fusarium temperatum* and *Fusarium subglutinans* isolated from maize in Argentina. *International Journal of Food Microbiology* **199**: 86-92.

88. Galazka JM & Freitag M (2014) Variability of chromosome structure in pathogenic fungi - Of 'ends and odds'. *Current Opinion in Microbiology* **20**: 19-26.

89. Ganley RJ, Watt MS, Manning L & Iturritxa E (2009) A global climatic risk assessment of pitch canker disease. *Canadian Journal of Forest Research* **39**: 2246-2256.

90. Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, *et al.* (2014) Horizontal gene transfer in plants. *Functional and Integrative Genomics* **14**: 23-29.

91. Garcia-Pedrajas MD & Roncero MIG (1996) A homologous and self-replicating system for efficient transformation of *Fusarium oxysporum*. *Current Genetics* **29**: 191-198.

92. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, *et al.* (2012) Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. *PLoS Pathogens* **8**: e1002952.

93. Garnica DP, Nemri A, Upadhyaya NM, Rathjen JP & Dodds PN (2014) The ins and outs of rust haustoria. *PLoS Pathogens* **10**: e1004329.

94. Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, Brandt ME, *et al.* (2013) One fungus, one name: Defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400-408.

95. Glenn AE, Davis CB, Gao M, Gold SE, Mitchell TR, Proctor RH, *et al.* (2016) Two horizontally transferred xenobiotic resistance gene clusters associated with detoxification of benzoxazolinones by *Fusarium* species. *PLoS One* **11**: e0147486.

96. Gordon TR & Okamoto D (1992) Population structure and the relationship between pathogenic and nonpathogenic strains of *Fusarium oxysporum*. *Phytopathology* **82**: 73-77.

97. Gordon TR, Storer AJ & Okamoto D (1996) Population structure of the pitch canker pathogen, *Fusarium subglutinans* f. sp. *pini*, in California. *Mycological Research* **100**: 850-854.

98. Goubet F, Jackson P, Deery MJ & Dupree P (2002) Polysaccharide analysis using carbohydrate gel electrophoresis: A method to study plant cell wall polysaccharides and polysaccharide hydrolases. *Analytical Biochemistry* **300**: 53-68.

99. Gout L, Fudal I, Kuhn M, Blaise F, Eckert M, Cattolico L, *et al.* (2006) Lost in the middle of nowhere: The *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*. *Molecular Microbiology* **60**: 67-80.

100. Gout L, Kuhn ML, Vincenot L, Bernard-Samain S, Cattolico L, Barbetti M, *et al.* (2007) Genome structure impacts molecular evolution at the *AvrLm1* avirulence locus of the plant pathogen *Leptosphaeria maculans*. *Environmental Microbiology* **9**: 2978-2992.

101. Guerra-Santos J (1999) Pitch Canker on Monterey pine in Mexico. *Current and Potential Impacts of Pitch Canker in Radiata Pine*, (Devey ME, Matheson AC & Gordon TR, eds.), pp. 58–61. Proceedings of the IMPACT Monterey Workshop, Monterey, California, USA, 30 November to 3 December 1998, CSIRO, Australia.

102. Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, *et al.* (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* **461**: 393-398.

103. Hall AC, Ostrowski LA, Pietrobon V & Mekhail K (2017) Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. *Nucleus* **8**: 162-181.

104. Hamin Neto YAA, Da Rosa Garzon NG, Pedezzi R & Cabral H (2018) Specificity of peptidases secreted by filamentous fungi. *Bioengineered* **9**: 30-37.

105. Han Y, Liu X, Benny U, Kistler HC & Van Etten HD (2001) Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*. *Plant Journal* **25**: 305-314.

106. Hardoim PR, Van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, *et al.* (2015) The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews* **79**: 293-320.

107. Hatta R, Ito K, Hosaki Y, Tanaka T, Tanaka A, Yamamoto M, *et al.* (2002) A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics* **161**: 59-70.

108. He P, Shan L & Sheen J (2007) Elicitation and suppression of microbe-associated molecular pattern-triggered immunity in plant–microbe interactions. *Cellular Microbiology* **9**: 1385-1396.

109. Heard S, Brown NA & Hammond-Kosack K (2015) An interspecies comparative analysis of the predicted secretomes of the necrotrophic plant pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS One* **10**: e0130534.

110. Hepting GH & Roth ER (1946) Pitch canker, a new disease of some southern pines. *Journal of Forestry* **44**: 742-744.

111. Hernandez-Escribano L, Iturritxa E, Elvira-Recuenco M, Berbegal M, Campos JA, Renobales G, *et al.* (2018) Herbaceous plants in the understory of a pitch canker-affected *Pinus radiata* plantation are endophytically infected with *Fusarium circinatum*. *Fungal Ecology* **32**: 65-71.

112. Hernandez-Rivas R, Mattei D, Sterkers Y, Peterson DS, Wellems TE & Scherf A (1997) Expressed *var* genes are found in *Plasmodium falciparum* subtelomeric regions. *Molecular and Cellular Biology* **17**: 604-611.

113. Herron DA, Wingfield MJ, Wingfield BD, Rodas CA, Marincowitz S & Steenkamp ET (2015) Novel taxa in the *Fusarium fujikuroi* species complex from *Pinus* spp. *Studies in Mycology* **80**: 131-150.

114. Hodge GR & Dvorak WS (2000) Differential responses of Central American and Mexican pine species and *Pinus radiata* to infection by the pitch canker fungus. *New Forests* **19**: 241-258.

115. Hoffmeister D & Keller NP (2007) Natural products of filamentous fungi: Enzymes, genes and their regulation. *Natural Product Reports* **24**: 393-416.

116. Hongwane P, Mitchell G, Kanzler A, Verryn S, Lopez J & Chirwa P (2018) Alternative pine hybrids and species to *Pinus patula* and *P. radiata* in South Africa and Swaziland. *Southern Forests* **80**: 301-310.

117. Hoogendoorn K, Barra L, Waalwijk C, Dickschat JS, Van der Lee TAJ & Medema MH (2018) Evolution and diversity of biosynthetic gene clusters in *Fusarium*. *Frontiers in Microbiology* **9**: 1-12.

118. Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, *et al.* (2014) Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 10642-10647.

119. Hua-Van A, Le Rouzic A, Boutin TS, Filee J & Capy P (2011) The struggle for life of the genome's selfish architects. *Biology Direct* **6**: 19.

120. Huang CC & Lindhout P (1997) Screening for resistance in wild *Lycopersicon* species to *Fusarium oxysporum* f. sp. *lycopersici* race 1 and race 2. *Euphytica* **93**: 145-153.

121. Huang X, Das A, Sahu BB, Srivastava SK, Leandro LF, O'Donnell K, *et al.* (2016) Identification of highly variable supernumerary chromosome segments in an asexual pathogen. *PLoS One* **11**: e0158183.

122. Imazaki I & Kadota I (2015) Molecular phylogeny and diversity of *Fusarium* endophytes isolated from tomato stems. *FEMS Microbiology Ecology* **91**: fiv1098.

123. Jacob F (1977) Evolution and tinkering. *Science* **196**: 1161-1166.

124. Jain R, Rivera MC & Lake JA (1999) Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 3801-3806.

125. Jones JD & Dangl JL (2006) The plant immune system. *Nature* **444**: 323-329.

126. Jurjevic Z, Solfrizzo M, Cvjetkovic B, De Girolamo A & Visconti A (2002) Occurrence of beauvericin in corn from Croatia. *Food Technology and Biotechnology* **40**: 91-94.

127. Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Research* **20**: 1313-1326.

128. Kamoun S (2006) A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review of Phytopathology* **44**: 41-60.

129. Kämper J, Kahmann R, Bölker M, Ma L, Brefort T, Saville BJ, *et al.* (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**: 97-101.

130. Katoh K, Rozewicki J & Yamada KD (2017) MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**: 1160-1166.

131. Keeling PJ & Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Reviews in Genetics* **9**: 605-618.

132. Keely SP, Renauld H, Wakefield AE, Cushion MT, Smulian AG, Fosker N, *et al.* (2005) Gene arrays at *Pneumocystis carinii* telomeres. *Genetics* **170**: 1589-1600.

133. Keller NP & Hohn TM (1997) Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology* **21**: 17-29.

134. Keller NP, Turner G & Bennett JW (2005) Fungal secondary metabolism-From biochemistry to genomics. *Nature Reviews in Microbiology* **3**: 937-947.

135. Kietzka J (1988) *Pinus maximinoi*: A promising species in South Africa. *South African Forestry Journal* **145**: 33-38.

136. King R, Brown NA, Urban M & Hammond-Kosack KE (2018) Inter-genome comparison of the Quorn fungus *Fusarium venenatum* and the closely related plant infecting pathogen *Fusarium graminearum*. *BMC Genomics* **19**: 269.

137. King R, Urban M, Hammond-Kosack MC, Hassani-Pak K & Hammond-Kosack KE (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**: 544.

138. Kobayashi T & Muramoto M (1989) Pitch canker of *Pinus luchuensis*, a new disease of Japanese forests. *Forest Pests* **38**: 169-173.

139. Konkel MK & Batzer MA (2010) A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Seminars in Cancer Biology* **20**: 211-221.

140. Kraft JM, Burke DW & Haglund WA (1981) *Fusarium* diseases of beans, peas and lentils. *Fusarium: Diseases, Biology and Taxonomy*, (Nelson PE, Toussoun TA & Cook RJ, eds.), pp. 142-156. Pennsylvania State University Press, University Park, PA.

141. Kumar S, Stecher G & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870-1874.

142. Kutty G, Shroff R & Kovacs JA (2013) Characterization of pneumocystis major surface glycoprotein gene (*msg*) promoter activity in *Saccharomyces cerevisiae*. *Eukaryotic Cell* **12**: 1349-1355.

143. Kvas M, Marasas WFO, Wingfield BD, Wingfield MJ & Steenkamp ET (2009) Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. *Fungal Diversity* **34**: 1-21.

144. Landeras E, García P, Fernández Y, Braña M, Fernández-Alonso O, Mendez-Lodos S, *et al.* (2005) Outbreak of pitch canker caused by *Fusarium circinatum* on *Pinus* spp. in Northern Spain. *Plant Disease* **89**: 1015.

145. Lanubile A, Ferrarini A, Maschietto V, Delledonne M, Marocco A & Bellin D (2014) Functional genomic analysis of constitutive and inducible defense responses to *Fusarium verticillioides* infection in maize genotypes with contrasting ear rot resistance. *BMC Genomics* **15**: 710.

146. Lanza FE, Mayfield DA & Munkvold GP (2016) First report of *Fusarium temperatum* causing maize seedling blight and seed rot in North America. *Plant Disease* **100**: 1019.

147. Laurence M, Walsh J, Shuttleworth L, Robinson D, Johansen R, Petrovic T, *et al.* (2016) Six novel species of *Fusarium* from natural ecosystems in Australia. *Fungal Diversity* **77**: 349-366.

148. Lee B, Kroken S, Chou DY, Robbertse B, Yoder O & Turgeon BG (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.

149. Lee JK, Lee S, Yang S, II & Lee Y (2000) First report of pitch canker disease on *Pinus rigida* in Korea. *Plant Pathology Journal* **16**: 52-54.

150. Leslie JF, Zeller KA, Logrieco A, Mule G, Moretti A & Ritieni A (2004) Species diversity of and toxin production by *Gibberella fujikuroi* species complex strains isolated from native prairie grasses in Kansas. *Applied and Environmental Microbiology* **70**: 2254-2262.

151. Li T, Wu Y, Wang Y, Gao H, Gupta VK, Duan X, *et al.* (2019) Secretome profiling reveals virulence-associated proteins of *Fusarium proliferatum* during interaction with banana fruit. *Biomolecules* **9**: 246.

152. Liao X, Fang W, Lin L, Lu H & Leger RJS (2013) *Metarhizium robertsii* produces an extracellular invertase (mrinv) that plays a pivotal role in rhizospheric interactions and root colonization. *PLoS One* **8**: e78118.

153. Liu J, Yuan Y, Wu Z, Li N, Chen Y, Qin T, *et al.* (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.

154. Lofgren LA, LeBlanc NR, Certano AK, Nachtigall J, LaBine KM, Riddle J, *et al.* (2018) *Fusarium graminearum*: Pathogen or endophyte of North American grasses? *New Phytologist* **217**: 1203-1212.

155. Logrieco A, Bottalico A, Mulé G, Moretti A & Perrone G (2003) Epidemiology of toxigenic fungi and their associated mycotoxins for some Mediterranean crops. *European Journal of Plant Pathology* **109**: 645-667.

156. Logrieco A, Mule G, Moretti A & Bottalico A (2002) Toxigenic *Fusarium* species and mycotoxins associated with maize ear rot in Europe. *European Journal of Plant Pathology* **108**: 597-609.

157. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490-D495.

158. Lowe RGT, McCorkelle O, Bleackley M, Collins C, Faou P, Mathivanan S, *et al.* (2015) Extracellular peptidases of the cereal pathogen *Fusarium graminearum*. *Frontiers in Plant Science* **6**: 962.

159. Ma L, Geiser DM, Proctor RH, Rooney AP, O'Donnell K, Trail F, *et al.* (2013) *Fusarium* pathogenomics. *Annual Review of Microbiology* **67**: 399-416.

160. Ma L, Van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367-373.

161. Malan FS (2003) The wood properties of South African timber resource for high-value solid wood products and its role in sustainable forestry. *Southern African Forestry Journal* **198**: 53–62.

162. Maplestone RA, Stone MJ & Williams DH (1992) The evolutionary role of secondary metabolites - A review. *Gene* **115**: 151-157.

163. Martín-García J, Lukačevičová A, Flores-Pacheco JA, Diez JJ & Dvořák M (2018) Evaluation of the susceptibility of several Czech conifer provenances to *Fusarium circinatum*. *Forests* **9**: 72.

164. Martin-Rodrigues N, Espinel S, Sanchez-Zabala J, Ortiz A, Gonzalez-Murua C & Dunabeitia MK (2013) Spatial and temporal dynamics of the colonization of *Pinus radiata* by *Fusarium circinatum*, of conidiophora development in the pith and of traumatic resin duct formation. *New Phytologist* **198**: 1215-1227.

165. Medema MH, Blin K, Cimermancic P, De Jager V, Zakrzewski P, Fischbach MA, *et al.* (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.

166. Miedaner T, Schilling AG & Geiger HH (2001) Molecular genetic diversity and variation for aggressiveness in populations of *Fusarium graminearum* and *Fusarium culmorum* sampled from wheat fields in different countries. *Journal of Phytopathology* **149**: 641-648.

167. Mitchell RG, Steenkamp ET, Coutinho TA & Wingfield MJ (2011) The pitch canker fungus, *Fusarium circinatum*: Implications for South African forestry. *Southern Forests* **73**: 1-13.

168. Molinié A, Faucet V, Castegnaro M & Pfohl-Leszkowicz A (2005) Analysis of some breakfast cereals on the French market for their contents of ochratoxin A, citrinin and fumonisin B_1: Development of a method for simultaneous extraction of ochratoxin A and citrinin. *Food Chemistry* **92**: 391-400.

169. Moller M & Stukenbrock EH (2017) Evolution and genome architecture in fungal plant pathogens. *Nature Reviews in Microbiology* **15**: 756-771.

170. Molnár O, Marton LC & Szőke C (2017) First report of *Fusarium temperatum* infecting corn in Hungary. *Plant Disease* **101**: 1325.

171. Moretti A, Mule G, Ritieni A, Laday M, Stubnya V, Hornok L, *et al.* (2008) Cryptic subspecies and beauvericin production by *Fusarium subglutinans* from Europe. *Internation Journal of Food Microbiology* **127**: 312-315.

172. Morse AM, Nelson CD, Covert SF, Holliday AG, Smith KE & Davis JM (2004) Pine genes regulated by the necrotrophic pathogen *Fusarium circinatum*. *Theoretical and Applied Genetics* **109**: 922-932.

173. Moussa TAA, Al-Zahrani HS, Kadasa NMS, Ahmed SA, De Hoog GS & Al-Hatmi AMS (2017) Two new species of the *Fusarium fujikuroi* species complex isolated from the natural environment. *Antonie van Leeuwenhoek* **110**: 819-832.

174. Muramoto M & Dwinell L (1990) Pitch canker of *Pinus luchuensis* in Japan. *Plant Disease* **74**: 530.

175. Mylonakis E, Casadevall A & Ausubel FM (2007) Exploiting amoeboid and non-vertebrate animal model systems to study the virulence of human pathogenic fungi. *PLoS Pathogens* **3**: e101.

176. Naglik JR, Challacombe SJ & Hube B (2003) *Candida albicans* secreted aspartyl proteinases in virulence and pathogenesis. *Microbiology and Molecular Biology Reviews* **67**: 400-428.

177. Nel A, Kanzler A & Dvorak W (2006a) Development of a commercial breeding program for *Pinus tecunumanii* in South Africa. (Isik F, ed.), pp. 160-163. Antalya, Turkey.

178. Nel B, Steinberg C, Labuschagne N & Viljoen A (2006b) The potential of nonpathogenic *Fusarium oxysporum* and other biological control organisms for suppressing Fusarium wilt of banana. *Plant Pathology* **55**: 217-223.

179. Nelson PE, Dignani MC & Anaissie EJ (1994) Taxonomy, biology and clinical aspects of *Fusarium* species. *Clinical Microbiology* **7**: 479-504.

180. Nelson PE, Plattner RD, Shackelford DD & Desjardins AE (1992) Fumonisin B1 production by *Fusarium* species other than *F. moniliforme* in section *Liseola* and by some related species. *Applied and Environmental Microbiology* **58**: 984-989.

181. Ng NM, Pike RN & Boyd SE (2009) Subsite cooperativity in protease specificity. *Biological Chemistry* **390**: 401-407.

182. Nicholson P, Chandler E, Draeger RC, Gosman NE, Simpson DR, Thomsett M, *et al.* (2003) Molecular tools to study epidemiology and toxicology of Fusarium head blight of cereals. *European Journal of Plant Pathology* **109**: 691-703.

183. Niehaus E, Kim H, Münsterkötter M, Janevska S, Arndt B, Kalinina SA, *et al.* (2017) Comparative genomics of geographically distant *Fusarium fujikuroi* isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLoS Pathogens* **13**: e1006670.

184. Niehaus E, Münsterkötter M, Proctor RH, Brown DW, Sharon A, Idan Y, *et al.* (2016) Comparative "omics" of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. *Genome Biology and Evolution* **8**: 3574-3599.

185. Nieva AS, Vilas JM, Gárriz A, Maiale SJ, Menéndez AB, Erban A, *et al.* (2019) The fungal endophyte *Fusarium solani* provokes differential effects on the fitness of two *Lotus* species. *Plant Physiology and Biochemistry* **144**: 100-109.

186. Niu C, Payne GA & Woloshuk CP (2015) Transcriptome changes in *Fusarium verticillioides* caused by mutation in the transporter-like gene *FST1*. *BMC Microbiology* **15**: 90.

187. Noble LM & Andrianopoulos A (2013) Fungal genes in context: Genome architecture reflects regulatory complexity and function. *Genome Biology and Evolution* **5**: 1336-1352.

188. Nogueira MS, Decundo J, Martinez M, Dieguez SN, Moreyra F, Moreno MV, *et al.* (2018) Natural contamination with mycotoxins produced by *Fusarium graminearum* and *Fusarium poae* in malting barley in Argentina. *Toxins* **10**: 78.

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

190. OEPP (2006) Premier signalement de *Gibberella circinata* en France. 2006/104.

191. Osbourn A (2010) Secondary metabolite gene clusters: Evolutionary toolkits for chemical innovation. *Trends in Genetics* **26**: 449-457.

192. Papagianni M & Moo-Young M (2002) Protease secretion in glucoamylase producer *Aspergillus niger* cultures: Fungal morphology and inoculum effects. *Process Biochemistry* **37**: 1271-1278.

193. Parry D, Jenkinson P & McLeod L (1995) Fusarium ear blight (scab) in small grain cereals-A review. *Plant Pathology* **44**: 207-238.

194. Petrov D (2001) Evolution of genome size: New approaches to an old problem. *Trends in Genetics* **17**: 23-28.

195. Pfenning LH, Da Silva Costa S, Pereira de Melo M, Costa H, Ventura JA, Auer CG, *et al.* (2014) First report and characterization of *Fusarium circinatum*, the causal agent of pitch canker in Brazil. *Tropical Plant Pathology* **39**: 210-216.

196. Phadatare SU, Srinivasan MC & Deshpande VV (1992) Evidence for controlled autoproteolysis of alkaline protease: A mechanism for physiological regulation of conidial discharge in *Conidiobolus coronatus*. *European Journal of Biochemistry* **205**: 679-686.

197. Pietrella D, Rachini A, Pandey N, Schild L, Netea M, Bistoni F, *et al.* (2010) The inflammatory response induced by aspartic proteases of *Candida albicans* is independent of proteolytic activity. *Infection and Immunity* **78**: 4754-4762.

198. Pintos V, Aguín C, Chaves P, Ferreiroa-Martínez V, Sainz M, Scauflaire J, *et al.* (2013) First report of *Fusarium temperatum* causing seedling blight and stalk rot on maize in Spain. *Plant Disease* **97**: 1252.

199. Porter JK, Voss KA, Bacon CW & Norred WP (1990) Effects of *Fusarium moniliforme* and corn associated with equine leukoencephalomalacia on rat neurotransmitters and metabolites. *Proceedings of the Society for Experimental Biology and Medicine* **194**: 265-269.

200. Presti LL, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, *et al.* (2015) Fungal effectors and plant susceptibility. *Annual Review of Plant Biology* **66**: 513-546.

201. Proctor RH, Plattner RD, Brown DW, Seo JA & Lee YW (2004) Discontinuous distribution of fumonisin biosynthetic genes in the *Gibberella fujikuroi* species complex. *Mycological Research* **108**: 815-822.

202. Pryde FE, Gorham HC & Louis EJ (1997) Chromosome ends: All the same under their caps. *Current Opinion in Genetics and Development* **7**: 822-828.

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

204. Raskina O, Barber JC, Nevo E & Belyayev A (2008) Repetitive DNA and chromosomal rearrangements: Speciation-related events in plant genomes. *Cytogenetic and Genome Research* **120**: 351-357.

205. Reichard U, Cole GT, Hill TW, Rüchel R & Monod M (2000) Molecular characterization and influence on fungal development of ALP2, a novel serine proteinase from *Aspergillus fumigatus*. *International Journal of Medical Microbiology* **290**: 549-558.

206. Rep M & Kistler HC (2010) The genomic organization of plant pathogenicity in *Fusarium* species. *Current Opinion in Plant Biology* **13**: 420-426.

207. Rep M, Van der Does HC, Meijer M, Van Wijk R, Houterman PM, Dekker HL, *et al.* (2004) A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for *I-3*-mediated resistance in tomato. *Molecular Microbiology* **53**: 1373-1383.

208. Rheeder JP, Marasas WF & Vismer HF (2002) Production of fumonisin analogs by *Fusarium* species. *Applied and Environmental Microbiology* **68**: 2101-2105.

209. Robles-Barrios KF, Medina-Canales MG, Rodríguez-Tovar AV & Pérez NO (2015) Morphological and molecular characterization, enzyme production and pathogenesis of *Fusarium temperatum* on corn in Mexico. *Canadian Journal of Plant Pathology* **37**: 495-505.

210. Scauflaire J, Gourgue M, Callebaut A & Munaut F (2012) *Fusarium temperatum*, a mycotoxin-producing pathogen of maize. *European Journal of Plant Pathology* **133**: 911-922.

211. Scauflaire J, Gourgue M & Munaut F (2011a) *Fusarium temperatum* sp. nov. from maize, an emergent species closely related to *Fusarium subglutinans*. *Mycologia* **103**: 586-597.

212. Scauflaire J, Mahieu O, Louvieaux J, Foucart G, Renard F & Munaut F (2011b) Biodiversity of *Fusarium* species in ears and stalks of maize plants in Belgium. *European Journal of Plant Pathology* **131**: 59-66.

213. Schaller A (2004) A cut above the rest: The regulatory function of plant proteases. *Planta* **220**: 183-197.

214. Schaller M, Borelli C, Korting HC & Hube B (2005) Hydrolytic enzymes as virulence factors of *Candida albicans*. *Mycoses* **48**: 365-377.

215. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, *et al.* (2013) Plant-symbiotic fungi as chemical engineers: Multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genetics* **9**: e1003323.

216. Scharf DH, Heinekamp T & Brakhage AA (2014) Human and plant fungal pathogens: The role of secondary metabolites. *PLoS Pathogens* **10**: e1003859.

217. Seidl MF & Thomma BP (2014) Sex or no sex: Evolutionary adaptation occurs regardless. *Bioessays* **36**: 335-345.

218. Sella L, Gazzetti K, Faoro F, Odorizzi S, D'Ovidio R, Schafer W, *et al.* (2013) *Fusarium graminearum* xylanase expressed during wheat infection is a necrotizing factor but is not essential for virulence. *Plant Physiology and Biochemistry* **64**: 1-10.

219. Shahin A, Van Kaauwen M, Esselink D, Bargsten JW, Van Tuyl JM, Visser RG, *et al.* (2012) Generation and analysis of expressed sequence tags in the extreme large genomes *Lilium* and *Tulipa*. *BMC Genomics* **13**: 640.

220. Shaner G & Scott D (1998) Stalk rots of corn. Extension publication BP-59, Department of Botany and Plant Pathology, Purdue University West Lafayette. Available https://extension.purdue.edu/extmedia/BP/BP-59.html.

221. Sharma M, Sengupta A, Ghosh R, Agarwal G, Tarafdar A, Nagavardhini A, *et al.* (2016) Genome wide transcriptome profiling of *Fusarium oxysporum* f. sp. *ciceris* conidial germination reveals new insights into infection-related genes. *Scientific Reports* **6**: 37353.

222. Shin J, Han J, Lee JK & Kim KS (2014) Characterization of the maize stalk rot pathogens *Fusarium subglutinans* and *F. temperatum* and the effect of fungicides on their mycelial growth and colony formation. *The Plant Pathology Journal* **30**: 397-406.

223. Sieber CMK, Lee W, Wong P, Münsterkötter M, Mewes H, Schmeitzl C, *et al.* (2014) The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One* **9**: e110311.

224. Sieber T (2002) Fungal root endophytes. *Plant roots: The hidden half*, (Waisel Y, Eshel A & Kafkafi U, eds.), pp. 887-917. Marcel Dekker, New York.

225. Sieber T, Riesen T, Müller E & Fried P (1988) Endophytic fungi in four winter wheat cultivars (*Triticum aestivum L.*) differing in resistance against *Stagonospora nodorum* (Berk.) Cast & Germ = *Septoria nodorum* (Berk.) Berk. *Journal of Phytopathology* **122**: 289-306.

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

227. Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, Stuber K, *et al.* (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543-1546.

228. Sperschneider J, Gardiner DM, Dodds PN, Tini F, Covarelli L, Singh KB, *et al.* (2016) EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytologist* **210**: 743-761.

229. Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, *et al.* (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biology and Evolution* **7**: 1613-1627.

230. Srivastava AK, Kashyap PL, Chakdar H, Kumar M, Srivastava AK, Yadav J, *et al.* (2018) First *de novo* draft genome sequence of the pathogenic fungus *Fusarium udum* F02845, associated with pigeonpea (*Cajanus cajan* L. Millspaugh) wilt. *Microbiology Resource Announcements* **7**: e01001-e01018.

231. Steenkamp ET, Rodas CA, Kvas M & Wingfield MJ (2012) *Fusarium circinatum* and pitch canker of *Pinus* in Colombia. *Australasian Plant Pathology* **41**: 483-491.

232. Steenkamp ET, Wingfield BD, Desjardins AE, Marasas WF & Wingfield MJ (2002) Cryptic speciation in *Fusarium subglutinans*. *Mycologia* **94**: 1032-1043.

233. Stepien L, Koczyk G & Waskiewicz A (2011) FUM cluster divergence in fumonisins-producing *Fusarium* species. *Fungal Biology* **115**: 112-123.

234. Stewart J, Abdo Z & Glenn A (2014) Gene clusters *FDB1* and *FDB2* in *Fusarium verticillioides* were acquired through multiple horizontal gene transfer events. *Phytopathology* **104**: 114.

235. Storer AJ, Gordon TR, Dallara PL & Wood DL (1994) Pitch canker kills pines, spreads to new species and regions. *California Agriculture* **48**: 9-13.

236. Stukenbrock EH & Croll D (2014) The evolving fungal genome. *Fungal Biology Reviews* **28**: 1-12.

237. Summerell BA, Laurence MH, Liew ECY & Leslie JF (2010) Biogeography and phylogeography of *Fusarium*: A review. *Fungal Diversity* **44**: 3-13.

238. Summerell BA & Leslie JF (2011) Fifty years of *Fusarium*: How could nine species have ever been enough? *Fungal Diversity* **50**: 135-144.

239. Susca A, Villani G, Mulè GS, Logrieco AF & Moretti A (2013) Geographic distribution and multilocus analysis of *Fusarium subglutinans* and *F. temperatum* from maize worldwide. p. 170. Palais de la Bourse, Bordeaux, France. Available https://mafiadoc.com/12th-european-fusarium-seminar-mafiadoccom_5ca8fb34097c470c4b8b4576.html.

240. Sutton DA & Brandt ME (2011) *Fusarium and other opportunistic hyaline fungi*. ASM Press, Washington, DC.

241. Swett CL & Gordon TR (2009) Colonization of corn (*Zea mays*) by the pitch canker pathogen, *Fusarium circinatum*: Insights into the evolutionary history of a pine pathogen. *Phytopathology* **99**: S126-S127.

242. Swett CL & Gordon TR (2015) Endophytic association of the pine pathogen *Fusarium circinatum* with corn (*Zea mays*). *Fungal Ecology* **13**: 120-129.

243. Swett CL & Gordon TR (2017) Exposure to a pine pathogen enhances growth and disease resistance in *Pinus radiata* seedlings. *Forest Pathology* **47**: 1-10.

244. Swett CL & Gordon TR (2012) First report of grass species (Poaceae) as naturally occurring hosts of the pine pathogen *Gibberella circinata*. *Plant Disease* **96**: 908.

245. Swett CL, Kirkpatrick SC & Gordon TR (2016) Evidence for a hemibiotrophic association of the pitch canker pathogen *Fusarium circinatum* with *Pinus radiata*. *Plant Disease* **100**: 79-84.

246. Swett CL, Porter B, Fourie G, Steenkamp ET, Gordon TR & Wingfield MJ (2014) Association of the pitch canker pathogen *Fusarium circinatum* with grass hosts in commercial pine production areas of South Africa. *Southern Forests* **76**: 161-166.

247. Swett CL, Reynolds GJ & Gordon TR (2018) Infection without wounding and symptomless shoot colonization of *Pinus radiata* by *Fusarium circinatum*, the cause of pitch canker. *Forest Pathology* **48**: e12422.

248. Tameling WIL & Takken FLW (2008) Resistance proteins: Scouts of the plant innate immune system. *Sustainable disease management in an European context*, (Collinge DB, Munk L & Cooke BM, eds.), pp. 243-255. Springer Netherlands, Dordrecht.

249. 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.

250. Tashiro S, Nishihara Y, Kugou K, Ohta K & Kanoh J (2017) Subtelomeres constitute a safeguard for gene expression and chromosome homeostasis. *Nucleic Acids Research* **45**: 10333-10349.

251. Thomas C (1971) The genetic organization of chromosomes. *Annual Review of Genetics* **5**: 237-256.

252. Treangen TJ, Abraham A, Touchon M & Rocha EPC (2009) Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiology Reviews* **33**: 539-571.

253. Troncoso C, González X, Bömke C, Tudzynski B, Gong F, Hedden P, *et al.* (2010) Gibberellin biosynthesis and gibberellin oxidase activities in *Fusarium*

sacchari, *Fusarium konzum* and *Fusarium subglutinans* strains. *Phytochemistry* **71**: 1322-1331.

254. Tyler BM, Tripathy S, Zhang X, Dehal P, Jiang RHY, Aerts A, *et al.* (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **313**: 1261-1266.

255. Van der Nest MA, Beirn LA, Crouch JA, Demers JE, De Beer ZW, De Vos L, *et al.* (2014) Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntiella omanensis*, *Leptographium procerum*, *Rutstroemia sydowiana* and *Sclerotinia echinophila*. *IMA Fungus* **5**: 473-486.

256. Van der Nest MA, Steenkamp ET, Roodt D, Soal NC, Palmer M, Chan WY, *et al.* (2019) Genomic analysis of the aggressive tree pathogen *Ceratocystis albifundus*. *Fungal Biology* **123**: 351-363.

257. Van Wyk S, Wingfield BD, De Vos L, Santana QC, Van der Merwe NA & Steenkamp ET (2018) Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*. *IMA Fungus* **9**: 27-36.

258. Varela CP, Casal OA, Padin MC, Martinez VF, Oses MJS, Scauflaire J, *et al.* (2013) First report of *Fusarium temperatum* causing seedling blight and stalk rot on maize in Spain. *Plant Disease* **97**: 1252.

259. Venturini G, Toffolatti SL, Passera A, Pilu R, Quaglino F & Casati P (2016) First report of *Fusarium temperatum* causing ear rot on maize in Italy. *Journal of Plant Pathology* **98**: 686.

260. Viljoen A, Wingfield MJ & Marasas WFO (1994) First report of *Fusarium subglutinans* f. sp. *pini* on pine seedlings in South Africa. *Plant Disease* **78**: 309-312.

261. Vinnig LC (1990) Functions of secondary metabolites. *Annual Review of Microbiology* **44**: 395-427.

262. Voss K, Smith G & Haschek W (2007) Fumonisins: Toxicokinetics, mechanism of action and toxicity. *Animal Feed Science and Technology* **137**: 299-325.

263. Waalwijk C, Vanheule A, Audenaert K, Zhang H, Warris S, Van de Geest HC, *et al.* (2017) *Fusarium* in the age of genomics. *Tropical Plant Pathology* **42**: 184-189.

264. Walkowiak S, Rowland O, Rodrigue N & Subramaniam R (2016) Whole genome sequencing and comparative genomics of closely related Fusarium head blight fungi: *Fusarium graminearum*, *F. meridionale* and *F. asiaticum*. *BMC Genomics* **17**: 1014.

265. Walsh J, Laurence M, Liew ECY, Sangalang A, Burgess L, Summerell B, *et al.* (2010) *Fusarium*: Two endophytic novel species from tropical grasses of northern Australia. *Fungal Diversity* **44**: 149-159.

266. Wang J, Zhang J, Li H, Gong A, Xue S, Agboola RS, *et al.* (2014) Molecular identification, mycotoxin production and comparative pathogenicity of *Fusarium temperatum* isolated from maize in China. *Journal of Phytopathology* **162**: 147-157.

267. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, *et al.* (2015) antiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* **43**: W237-W243.

268. Weerdt M, Zijlstra C, Brouwershaven I, Van Leeuwen G, Gruyter J & Kox L (2006) Molecular detection of *Fusarium foetens* in Begonia. *Journal of Phytopathology* **154**: 694-700.

269. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**: 973-982.

270. Wiemann P, Sieber CMK, Von Bargen KW, Studt L, Niehaus E, Espino JJ, *et al.* (2013) Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathogens* **9**: e1003475.

271. Williams AH, Sharma M, Thatcher LF, Azam S, Hane JK, Sperschneider J, *et al.* (2016) Comparative genomics and prediction of conditionally dispensable sequences in legume–infecting *Fusarium oxysporum formae speciales* facilitates identification of candidate effectors. *BMC Genomics* **17**: 191.

272. Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W, Crouch JA, *et al.* (2015a) Draft genome sequences of *Chrysoporthe austroafricana*, *Diplodia scrobiculata*, *Fusarium nygamai*, *Leptographium lundbergii*, *Limonomyces culmigenus*, *Stagonosporopsis tanaceti* and *Thielaviopsis punctulata*. *IMA Fungus* **6**: 233-248.

273. Wingfield BD, Barnes I, De Beer ZW, De Vos L, Duong TA, Kanzi AM, *et al.* (2015b) Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysoporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum* and *Thielaviopsis musarum*. *IMA Fungus* **6**: 493-506.

274. Wingfield BD, Berger DK, Steenkamp ET, Lim H, Duong TA, Bluhm BH, *et al.* (2017) Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. *IMA Fungus* **8**: 385-396.

275. Wingfield BD, Bills GF, Dong Y, Huang W, Nel WJ, Swalarsk-Parry BS, *et al.* (2018a) Draft genome sequence of *Annulohypoxylon stygium*, *Aspergillus mulundensis*, *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora*, *Fusarium fracticaudum*, *Phialophora* cf. *hyalina* and *Morchella septimelata*. *IMA Fungus* **9**: 199-223.

276. Wingfield BD, Liu M, Nguyen HDT, Lane FA, Morgan SW, De Vos L, *et al.* (2018b) Nine draft genome sequences of *Claviceps purpurea* s. lat., including *C. arundinis*, *C. humidiphila*, and *C.* cf. *spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti*,

Grosmannia galeiformis, *Quambalaria eucalypti* and *Teratosphaeria destructans*. *IMA Fungus* **9**: 401-418.

277. Wingfield BD, Steenkamp ET, Santana QC, Coetzee MPA, Bam S, Barnes I, *et al.* (2012) First fungal genome sequence from Africa: A preliminary analysis. *South African Journal of Science* **108**: 1-9.

278. Wingfield MJ, Hammerbacher A, Ganley RJ, Steenkamp ET, Gordon TR, Wingfield BD, *et al.* (2008) Pitch canker caused by *Fusarium circinatum* - A growing threat to pine plantations and forests worldwide. *Australasian Plant Pathology* **37**: 319-334.

279. Wingfield MJ, Jacobs A, Coutinho TA, Ahumada R & Wingfield BD (2002) First report of the pitch canker fungus, *Fusarium circinatum*, on pines in Chile. *Plant Pathology* **51**: 397.

280. Wittenberg AHJ, Van der Lee TAJ, M'Barek SB, Ware SB, Goodwin SB, Kilian A, *et al.* (2009) Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. *PLoS One* **4**: e5863.

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

282. Yike I (2011). *Mycopathologia* **171**: 299-323.

283. Yim G, Wang HH & Davies J (2007) Antibiotics as signalling molecules. *Philosophical Transactions of the Royal Society B: Biological Science* **362**: 1195-1200.

284. Yip VL & Withers SG (2006) Breakdown of oligosaccharides by the process of elimination. *Current Opinion in Chemical Biology* **10**: 147-155.

285. Yu J & Keller NP (2005) Regulation of secondary metabolism in filamentous fungi. *Annual Review of Phytopathology* **43**: 437-458.

286. Zabalgogeazcoa I (2008) Fungal endophytes and their interaction with plant pathogens: A review. *Spanish Journal of Agricultural Research* **6**: 138.

287. Zakaria L & Ning CH (2013) Endophytic *Fusarium* spp. from roots of lawn grass (*Axonopus compressus*). *Tropical Life Sciences Research* **24**: 85-90.

288. Zeller KA, Summerell BA, Bullock S & Leslie JF (2003) *Gibberella konza* (*Fusarium konzum*) sp. nov. from prairie grasses, a new species in the *Gibberella fujikuroi* species complex. *Mycologia* **95**: 943-954.

289. Zhang Q, Chen X, Xu C, Zhao H, Zhang X, Zeng G, *et al.* (2019) Horizontal gene transfer allowed the emergence of broad host range entomopathogens. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 7982-7989.

290. Zhao C, Waalwijk C, De Wit PJ, Tang D & Van der Lee T (2014) Relocation of genes generates non-conserved chromosomal segments in *Fusarium graminearum* that show distinct and co-regulated gene expression patterns. *BMC Genomics* **15**: 191.

291. Zhao Z, Liu H, Wang C & Xu J (2013) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**: 274.

292. Zhou JM & Chai J (2008) Plant pathogenic bacteria type III effectors subdue host responses. *Current Opinion in Microbiology* **11**: 179-185.

TABLES

Table 1. The three different clades in the *Fusarium fujikuroi* species complex (FFSC) and the *Fusarium* species belonging to these clades (O'Donnell *et al.*, 1998, Kvas *et al.*, 2009, Herron *et al.*, 2015).

Table 2. Some genome features and statistics of *Fusarium* species discussed in this study.

Table 1. The three different clades in the *Fusarium fujikuroi* species complex (FFSC) and the *Fusarium* species belonging to these clades (O'Donnell *et al.*, 1998, Kvas *et al.*, 2009, Herron *et al.*, 2015).

African clade	American clade	Asian clade
Fusarium acutatum	*Fusarium agapanthi*	*Fusarium annulatum*
Fusarium brevicatenulatum	*Fusarium ananatum*	*Fusarium concentricum*
Fusarium denticulatum	*Fusarium anthophilum*	*Fusarium fractiflexum*
Fusarium dlamini	*Fusarium bactridioides*	*Fusarium fujikuroi*
Fusarium lactis	*Fusarium begoniae*	*Fusarium fractiflexum*
Fusarium napiforme	*Fusarium bulbicola*	*Fusarium globosum*
Fusarium nygamai	*Fusarium circinatum*	*Fusarium mangiferae*
Fusarium phyllophilum	*Fusarium fracticaudum*	*Fusarium neoceras*
Fusarium pseudoanthophilum	*Fusarium guttiforme*	*Fusarium proliferatum*
Fusarium pseudocircinatum	*Fusarium konzum*	*Fusarium sacchari*
Fusarium pseudonygamai	*Fusarium marasasianum*	
Fusarium ramigenum	*Fusarium mexicanum*	
Fusarium thapsinum	*Fusarium parvisorum*	
Fusarium udum	*Fusarium pininemorale*	
Fusarium verticillioides	*Fusarium sororula*	
Fusarium xylarioides	*Fusarium sterilihyphosum*	
	Fusarium subglutinans	
	Fusarium succisae	
	Fusarium temperatum	
	Fusarium tupiense	
	Fusarium werrikimbe	

Table 2. Some genome features and statistics of *Fusarium* species discussed in this study.

***Fusarium* species**	**Chromosome count**	**Genome size (Mb)**	**Repetitive element content, %**	**Transposable element content, %**	**Secondary metabolite biosynthetic genes, %**	**CAZymes[1], %**	**References**
F. circinatum	12	42.5	5.81	2.94	0.30*	4.40	Unpublished data
F. fracticaudum	12	46.3	3.39	2.51	0.37*	4.63	Unpublished data
F. fujikuroi	12	48.3	4.08	2.2	0.30	4.57	Wiemann *et al.* (2013), Bashyal *et al.* (2017)
F. graminearum	4	36.4	1.31	0.33	0.31	4.26	Ma *et al.* (2012), Wiemann *et al.* (2013)
F. mangiferae	11	45.6	1.35	0.54	0.30	4.71	Wiemann *et al.* (2013), Niehaus *et al.* (2016)
F. oxysporum	15	61.4	21.24	4.76	0.13	3.77	Ma *et al.* (2012), Wiemann *et al.* (2013)
F. pininemorale	12	47.8	8.80	3.16	0.31*	4.56	Unpublished data
F. solani (*Nectria haematococca*)	17	51.3	5.85	1.64	0.17	4.88	Coleman *et al.* (2009), Wiemann *et al.* (2013)
F. temperatum	12	45.5	8.45	2.01	0.31*	4.46	Unpublished data
F. verticillioides	11	41.8	1.28	0.47	0.20	3.96	Ma *et al.* (2012), Wiemann *et al.* (2013)

*Secondary metabolite biosynthetic genes identified with antiSMASH (https://fungismash.secondarymetabolites.org/).
[1]CAZymes were identified with dbCAN2 (http://bcb.unl.edu/dbCAN2/blast.php).

FIGURES

Figure 1. Phylogenetic relationship of closely related *Fusarium* species within the *Fusarium fujikuroi* species complex (FFSC).

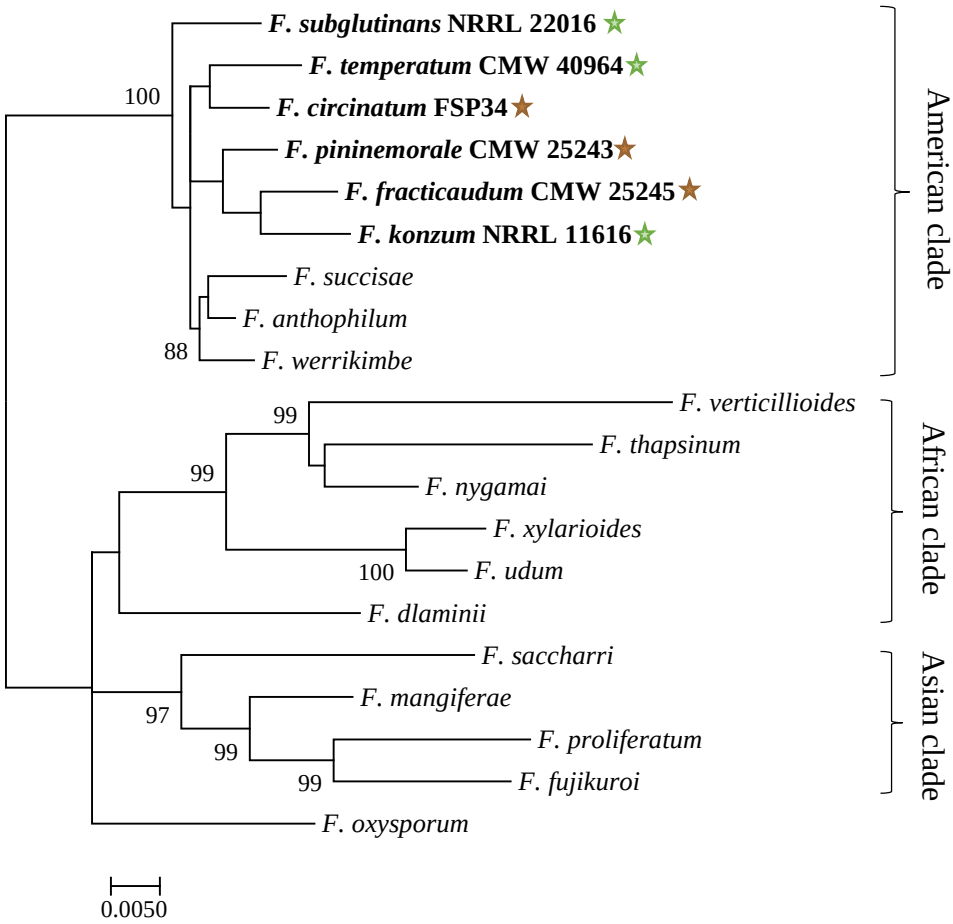


Figure 1. Phylogenetic relationship of closely related *Fusarium* species within the *Fusarium fujikuroi* species complex (FFSC). The Maximum Likelihood branch support was estimated using bootstrap analyses based on 100 pseudoreplicates. The brown stars represent the pine-associated *Fusarium* species, whereas the green stars represent the Poaceae-associates. Most genome sequences were obtained from Herron *et al.* (2015) and supplemented with the sequences of the six *Fusarium* genomes from Chapter 2 and 3. The elongation factor and β-tubulin regions of the respective genomes were aligned using MAFFT v. 7 (Katoh *et al.*, 2017). The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura & Nei, 1993). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA v. 7.0.26 (Kumar *et al.*, 2016).

CHAPTER 2:

Characterisation of host-specific genes from pine- and Poaceae-associated *Fusarium* species of the FFSC

ABSTRACT

Species in the *Fusarium fujikuroi* species complex (FFSC) comprise some of the most socio-economically important pathogens globally. Many of these fungi have genomic sequence data available, with some even assembled to chromosome level. In this study, we investigated the molecular basis of host-specificity by comparing the genomes of species associated with *Pinus* species (*i.e.*, *F. circinatum* FSP34, *F. pininemorale* CMW 25243 and *F. fracticaudum* CMW 25245) and members of Poaceae (*i.e.*, *F. temperatum* CMW 40964, *F. subglutinans* NRRL 22016 and *F. konzum* NRRL 11616). Genes common and unique to the two sets of species were identified and subjected to functional annotation. A total of 11 850 genes were shared amongst the six species, while 72 and 47 genes were unique to the respective sets of genomes. These two sets were enriched for genes implicated in carbohydrate and protein metabolism. Phylogenetic analysis of the unique genes suggested various evolutionary origins, indicating that they were acquired over time from numerous sources. This study found that the frequency distribution of unique genes differed significantly between telomeric and non-telomeric regions and that some tended to cluster together or were located close to another cluster. An instance of chromosome length polymorphism was found for chromosome 12. Chromosome 12 of *F. temperatum* CMW 40964 was larger than that of *F. circinatum* FSP34 and possessed more unique genes, which are potentially involved in niche-specificity amongst *Fusarium* species associated with Poaceae. These findings demonstrate the different molecular mechanisms employed by these *Fusarium* species to infect, inhabit and cause disease on their respective plant hosts.

INTRODUCTION

The genus *Fusarium* represents a diverse group of ascomycetes of significant importance in industry and agriculture (Geiser *et al.*, 2013). Species in this genus also represent essential pathogens of humans (O'Donnell *et al.*, 2004, Chang *et al.*, 2006, Sutton & Brandt, 2011), animals (Ortoneda *et al.*, 2004) and insects (Coleman *et al.*, 2011, Navarro-Velasco *et al.*, 2011). It is considered that the majority of these species are known as destructive plant pathogens (Parry *et al.*, 1995, Gordon *et al.*, 1996, Desjardins *et al.*, 2000, Bottalico & Perrone, 2002, Herron *et al.*, 2015, Armitage *et al.*, 2018). These pathogenic species affect the quality and quantity of agriculturally important crops, resulting in significant issues regarding food security (Fisher *et al.*, 2012).

Fusarium species associate with diverse plant hosts, ranging from gymnosperms through to angiosperms, including many dicots, as well as monocots such as those in the grass family Poaceae (O'Donnell *et al.*, 1998, Steenkamp *et al.*, 2002, Niehaus *et al.*, 2017). For example, *F. circinatum* is a pathogen of *Pinus* species and cause severe economic losses, whereas *F. fracticaudum* and *F. pininemorale* can also colonise the tissues of these gymnosperms, but without any apparent symptoms of a disease (Gordon *et al.*, 1996, Herron *et al.*, 2015). Amongst the grasses, *F. konzum* colonises prairie grass as an endophyte (Zeller *et al.*, 2003), whereas *F. subglutinans* and *F. temperatum* colonise and cause disease on maize (Desjardins *et al.*, 2000, Desjardins *et al.*, 2006). Furthermore, *F. circinatum* can also colonises maize (Swett & Gordon, 2012, Swett & Gordon, 2015), herbaceous plants (Hernandez-Escribano *et al.*, 2018), as well as other conifers (Martín-García *et al.*, 2018) as an endophyte.

The different underlying mechanisms of host-specificity and processes involved in the host-pathogen interaction vary amongst fungi (Nirmala *et al.*, 2007). Recent advances in genomics have enhanced the discovery of and provided new insights into the molecular and genomic basis of host-pathogen interactions (Abdullah *et al.*, 2017). These insights emerged from the study of the genomes of *F. graminearum*, *F. verticillioides*, *F. oxysporum* and *Nectria haematococca* (*F. solani*) (Coleman *et al.*, 2009, Ma *et al.*, 2010, Wiemann *et al.*, 2013). These insights can vary widely from plants in the group Gymnosperms.

Genomic analyses showed that fungal genomes, including those of *Fusarium*, are divided into a core and accessory compartment (Covert, 1998, Croll & McDonald, 2012). The core compartment of *Fusarium* genomes contributes to host-specificity and pathogenicity in a way which is different from the accessory compartment (Ma *et al.*, 2010). Genes within this compartment play a part in the primary metabolism, reproductive strategies and fungal development that also enable these fungi to overcome host defence responses. The synteny of the core compartment is conserved with regards to closely related species and is rich in genes (Armitage *et al.*, 2018). This genomic region is thought to drive the longevity of these fungi to enhance the survival rate.

The accessory compartment contains genes that are non-essential for fungal growth and development but encode for niche-associated traits, including pathogenicity and virulence factors (Croll & McDonald, 2012). The accessory compartment is gene-poor, and the synteny is non-conserved with regards to closely related species (Armitage *et al.*, 2018). Genomic comparisons have shown that the non-conserved regions tend to be more variable in terms of the type and distribution of genes (Cuomo *et al.*, 2007, Coleman *et al.*, 2009, Ma *et al.*, 2010, Sperschneider *et al.*, 2015, Van der Nest *et al.*, 2019). These genomic regions tend to harbour subtelomeric regions and supernumerary/accessory chromosomes (Raffaele & Kamoun, 2012). Subtelomeric regions are often associated with a break in synteny and have also been linked to genes promoting the adaptation and survival of pathogens (Zhao *et al.*, 2014). It is, therefore, speculated that these diverse genomic regions drive pathogen adaptation and niche utilisation.

Comparative genomics promotes the understanding of how genome evolution affects host-specificity and pathogenicity (Hardison, 2003, Wittenberg *et al.*, 2009, Goodwin *et al.*, 2011). For example, comparative genomic analyses identified several factors that have been implicated in evolutionary forces that shape the outcome of the host-pathogen interaction (Chain *et al.*, 2004, Akagi *et al.*, 2009, Ma *et al.*, 2010, Bäumler & Fang, 2013). Selection is an integral part of this process and occurs when some traits within a species are preferred above others. Genes that are not under selection become either mutated or deleted and can change the ways pathogens adapt to different host environments (Kisiela *et al.*, 2012).

An additional factor that can drive evolution in fungi is genomic gains or decay (Bäumler & Fang, 2013). Genomic gains can occur in numerous ways, such as horizontal gene transfer as explained by Akagi *et al.* (2009) and the movement of accessory chromosomes (Ma *et al.*, 2010). In contrast, genomic decay is due to gene deletions or gene inactivation through point mutations (Chain *et al.*, 2004). These gains or losses are also likely a mechanism of fungal adaptation from one host to another as a result of genome expansion. For example, the transfer of chromosomes between two genetically isolated members of *F. oxysporum*, a strain pathogenic to tomato and one non-pathogenic strain, demonstrated that chromosomal transfer from the pathogenic strain to the non-pathogenic strain results in pathogenicity in the latter strain (Ma *et al.*, 2010). The acquisition of genes, therefore, may directly impact the evolution of fungal pathogens.

Chromosomal rearrangements are another evolutionary force driving genome evolution. These rearrangements alter gene order and may also negatively affect growth rate, chromosome symmetry and may induce replication-transcription conflicts (Hill & Gray, 1988, Rebollo *et al.*, 1988, Campo *et al.*, 2004). However, chromosomal rearrangements can also have a positive effect on growth and survival, with the formation of novel genes (*e.g.*, secondary metabolite biosynthesis gene clusters) (Nierman *et al.*, 2005, Chiara *et al.*, 2015, Waalwijk *et al.*, 2017). When pathogens have to maximise growth for adaptation and survival, evolutionary forces rearrange specific genes on specific chromosomes to expose these genes to selection that may also be transferred to other strains to infer virulence (Lawley *et al.*, 2008, Van Dam *et al.*, 2017).

The genus *Fusarium* consists of a range of monophyletic groups that are referred to as species complexes (Geiser *et al.*, 2013). The complex investigated in this study is the *Fusarium fujikuroi* species complex (FFSC), which is characterised by high levels of synteny amongst its species (Wiemann *et al.*, 2013, De Vos *et al.*, 2014). Also, hybridisation is not unknown (Desjardins *et al.*, 2000, Leslie *et al.*, 2004, Scauflaire *et al.*, 2011), which is indicative of their shared ancestry. For example, two species that are known to hybridise, *F. circinatum* and *F. temperatum*, are economically important plant pathogens capable of inducing disease in their respective hosts, pine trees and maize, respectively (Hepting & Roth, 1946, Scauflaire *et al.*, 2011). This hybridisation led to the hypothesis that these two species shared a common ancestor, and co-evolved with their hosts (pine and maize), which have overlapping geographic ranges (Desjardins *et al.*,

2000, Steenkamp *et al.*, 2002). Differences in the genetic make-up between these two potentially point to genes and processes involved in host-specificity obtained through genomic gain, decay or chromosomal rearrangements.

Limited information is available on host-specificity within *Fusarium*. The economic importance of *Fusarium* species emphasises the need for research focussing on the mechanisms underlying their behaviour on different hosts. Fortunately, numerous genomes of species within the FFSC are available (Ma *et al.*, 2010, Wingfield *et al.*, 2012, Wiemann *et al.*, 2013, Wingfield *et al.*, 2015a, Wingfield *et al.*, 2015b, Edwards *et al.*, 2016, Niehaus *et al.*, 2016, Aylward *et al.*, 2017, Wingfield *et al.*, 2017, Srivastava *et al.*, 2018, Wingfield *et al.*, 2018a). The overall goal of this study was to identify a set of genes unique to *Fusarium* species associated with pine (*F. circinatum*, *F. fracticaudum* and *F. pininemorale*) versus species associated with Poaceae (*F. konzum*, *F. subglutinans* and *F. temperatum*). These genes were analysed and compared in terms of identity, gene ontology terms, predicted pathways and processes, chromosomal location and potential ancestral origin. Any information thus retrieved based on the genic structure of these pathogens will contribute to the current understanding of the molecular basis of the biology, diversity and evolution of these species.

MATERIALS AND METHODS

Genome sequences and annotation

The study included the genome sequence information for six *Fusarium* species, of which three were associated with *Pinus* species and three with members of the Poaceae family. The three associated with *Pinus* species included *F. circinatum* FSP34 isolated from *P. radiata* (Gordon *et al.*, 1996, De Vos *et al.*, 2020, unpublished), *F. fracticaudum* CMW 25245 isolated from *P. maximinoi* (Herron *et al.*, 2015, Wingfield *et al.*, 2018a) and *F. pininemorale* CMW 25243 isolated from *P. tecunumanii* (Herron *et al.*, 2015, Wingfield *et al.*, 2017). The three that were associated with members of the Poaceae family included *F. konzum* NRRL 11616 isolated from prairie grass (Zeller *et al.*, 2003), *F. subglutinans* NRRL 22016 isolated from maize (Desjardins *et al.*, 2006) and *F. temperatum* CMW

40964 isolated from teosinte (Desjardins *et al.*, 2000, Wingfield *et al.*, 2015b). The Illumina Mi-Seq was performed with one 500 bp paired-end library for the sequencing of the *F. subglutinans* NRRL 22016 and *F. konzum* NRRL 11616 genomes at Peoria (Illinois) (unpublished data, kindly provided by Robert H. Proctor).

The completeness of each genome assembly was evaluated through the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool v. 3.0.2. BUSCO considers the evolutionarily-informed expectations of gene content from the single-copy orthologs catalogue, OrthoDB v. 9, and using the Sordariomyceta database to identify all relevant orthologs from the input genes (Waterhouse *et al.*, 2017). Furthermore, using the significant synteny found within the FFSC (Wiemann *et al.*, 2013, De Vos *et al.*, 2014), scaffolds from the genomes of *F. pininemorale* (Wingfield *et al.*, 2017) and *F. fracticaudum* (Wingfield *et al.*, 2018a) were compared against the genomes of *F. circinatum* (Wingfield *et al.*, 2018b), *F. fujikuroi* (Wiemann *et al.*, 2013, Wingfield *et al.*, 2018a) and *F. temperatum* (Wingfield *et al.*, 2015b), as these have been assembled into chromosomes. The scaffolds were ordered and orientated into 12 contiguous pseudomolecules [representing the core chromosomes (1-11) and the accessory chromosome (12)], using the LASTZ plugin (Harris, 2007) of Geneious v. 7.0.4 (Kearse *et al.*, 2012). The genomes of *F. subglutinans* (733 contigs) and *F. konzum* (2 262 contigs) were considered too fragmented to be assembled into pseudomolecules.

A representative (reference) genome was selected that was associated with either pine trees or members of the Poaceae. The purpose of this was to achieve a detailed comparison with two well-assembled genomes between *Fusarium* species colonising different plant hosts. *Fusarium circinatum* FSP34 was chosen to represent the pine-associates and *F. temperatum* CMW 40964 represented Poaceae-associates. The abundance of the telomere-associated repeat sequence (TTAGGG/CCCTAA) (Garcia-Pedrajas & Roncero, 1996, Fulnecková *et al.*, 2013) was evaluated using a motif search performed in CLC Genomics Workbench v. 11 (CLC bio, Aarhus, Denmark), to further investigate the completeness of the two reference genomes (*F. circinatum* and *F. temperatum*). For the motif search, a window size of 10 000 bp with 5 000 bp increments was used. Only repeats with ≥ 80% similarity to the telomere repeat were considered in this analysis.

The conservation of chromosome size was investigated for four genomes assembled into pseudomolecules (*F. circinatum*, *F. fracticaudum*, *F. pininemorale* and *F. temperatum*). Firstly, each chromosome size was evaluated in all four genomes to identify the biggest chromosome per genome for all twelfth chromosomes. The biggest chromosome was divided by each of the other chromosomes, from the same chromosome number, in order to obtain the chromosome fold-difference between all twelfth chromosome from the four genomes. A factor of 1.00 was obtained when the biggest chromosome was divided by itself or if the factor obtained was more than 0.999 and rounded of to 1.00.

The MAKER pipeline v. 2.31.8 (Cantarel *et al.*, 2008) was utilised for functional annotations of all six genomes in order to identify protein-coding genes. Gene prediction was performed in MAKER using SNAP (Korf, 2004), GeneMark ES (Ter-Hovhannisyan *et al.*, 2008) and AUGUSTUS (Hoff *et al.*, 2019). As additional evidence, gene model data from *F. circinatum* (Wingfield *et al.*, 2018b), *F. fujikuroi* (Wiemann *et al.*, 2013), *F. verticillioides* and *F. graminearum* (Ma *et al.*, 2010), as well *as F. mangiferae* and *F. proliferatum* (Niehaus *et al.*, 2017), were included. These isolates were selected due to the availability of their genomic information on the NCBI public database (https://www.ncbi.nlm.nih.gov/).

Identification and functional annotation of genes unique to pine- and Poaceae-associated *Fusarium* species

The gene content of all six *Fusarium* genomes was evaluated to determine which genes are shared amongst all six *Fusarium* species and which ones are unique to the species associated with the two groups of plant hosts. For this purpose, OrthoFinder v. 2.3.1 was implemented (Emms & Kelly, 2015). OrthoFinder identified the orthologous genes (gene pairs descended from a single gene in the last common ancestor of two or more species) present in all six *Fusarium* genomes by firstly identifying the orthogroups (groups of genes descended from a single gene in the last common ancestor of a group of species) upon built-in BLASTp searches. The genes present in all six genomes were labelled as the "shared" genes, while those occurring only in the genomes of the pine-associates were labelled as "unique" genes. Those occurring only in the genomes of Poaceae-associates were also labelled as "unique".

Functional annotation was performed using the Blast2GO (Conesa & Gotz, 2008) plugin for CLC Genomics Workbench v. 11 (CLC bio, Aarhus, Denmark). A two-tailed Fisher exact test was implemented ($P < 0.05$) in Blast2GO (Conesa & Götz, 2008) to detect the gene ontology (GO) terms that were overrepresented in the unique genes set of both *Fusarium* representatives, using the whole genome of each as reference. Expressed Sequence Tag (EST) data for *F. circinatum* FSP34 from a previous study (Wingfield *et al.*, 2012) and RNA-seq data from a more recent study (Van Wyk *et al.*, 2019) were compared with the unique genes data set to identify which genes are expressed. All data were analysed in CLC Genomics Workbench v. 11.

Phylogenetic origins of the unique genes

The two unique gene sets were uploaded onto NCBI to perform BLASTp searches against the non-redundant database using the online position-specific iterative (psi) BLAST tool (Altschul *et al.*, 1997). All highly divergent protein sequences were excluded by considering only those sequences with at least 40% amino acid identity over 70% of the query sequence length and that had E-values $\leq 1 \times 10^{-5}$ (Van Wyk *et al.*, 2018). All the sequences were aligned using the constraint-based alignment tool (COBALT) (Papadopoulos & Agarwala, 2007) and the phylogenetic trees retrieved from COBALT were implemented to determine the ancestral origins of the unique genes. The unique genes with unclear ancestral origin were selected to manually construct phylogenetic trees. To infer phylogenies, all sequences were aligned with MAFFT (Multiple sequence Alignment based on Fast Fourier Transform) v. 7.0 with default settings (Katoh *et al.*, 2017). These alignments included the relevant *Fusarium* sequences, together with those from other Ascomycota. MEGA v. 7.0.26 (Kumar *et al.*, 2016) was used to draw initial tree(s) for the heuristic search by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances to estimate the best substitution model to use. The Maximum Likelihood branch support was estimated using bootstrap analyses based on 100 pseudoreplicates. The phylogenetic trees were viewed and edited in MEGA.

Genomic distribution of the unique genes

The location and distribution of the unique genes were plotted across the 12 chromosomes in each of the two reference genomes using KaryoploteR v. 3.9 (Gel & Serra, 2017).

Differences in the distribution of genes in the different genomic regions were evaluated with Chi-squared tests. These tests determine the frequency in which these genes differ significantly ($P < 0.05$) between telomeric and non-telomeric regions, assuming that the frequencies do not differ significantly as the null hypothesis. Telomeric regions were determined to be 100 000 bp (Van Wyk *et al.*, 2018) from each end of the assembled pseudomolecules representing the chromosomes.

The synteny and conservation of the location of the different unique genes were studied with SynChro which revealed the synteny breakpoints between the reference and query genomes (Drillon *et al.*, 2014). The input files for SynChro were the GFF files retrieved from MAKER, followed by the identification of the Reciprocal Best-Hits (RBH) between coding sequences to identify conserved and non-conserved syntenic blocks. The gene density per chromosome was determined based on the input GFF files.

RESULTS

Genome sequences and annotation

The corresponding BUSCO value for each genome portrayed the general completeness (97.3% for *F. circinatum* to 99.1% for *F. pininemorale*) of each of the six genomes examined (Supplemental Table 1). The genome sizes and gene density were in a range that was quite similar to one another (Table 1). The G+C content for these genomes varied from 46% for *F. pininemorale* to 49% for *F. konzum*. Comparison of the four genomes assembled into pseudomolecules (*F. circinatum, F. fracticaudum, F. pininemorale* and *F. temperatum*) showed that chromosome sizes seemed to be conserved throughout chromosomes 1-11 since they all did not differ by more than a factor of 1.00-1.10 (as evidenced in Supplemental Table 2). In comparison, chromosome 12 showed extreme chromosome length polymorphism (CLP) and differed by a factor of 2.08.

In the case of *F. pininemorale*, nine out of the 12 chromosomes were larger than similar chromosomes of the other species, and therefore this species has the largest genome size (Table 1). However, the genome of *F. pininemorale* has less protein-coding genes

compared to that of *F. fracticaudum* and *F. konzum*. The relationship between gene density and genome size was inversely proportional (Pearson product-moment correlation coefficient; r = -0.84) (Supplemental Table 3) since the smallest genome (*F. konzum*) had the largest gene density of 374.61 ORFs/Mb compared to the largest genome (*F. pininemorale*) with the smallest gene density of 332.35 ORFs/Mb.

For the reference genomes, the telomeric cap was present in most of the chromosomes (Table 2). *Fusarium circinatum* had seven chromosomes with telomeres at both chromosomal ends and five with a telomere at only one of the two ends. Chromosome 2 from *F. temperatum* lacked a telomere region at both chromosomal ends. In contrast, two chromosomes only had a telomere at one chromosomal end, and nine chromosomes had telomeres at both ends.

Identification and functional annotation of genes unique to pine- and Poaceae-associated *Fusarium* species

Data retrieved from OrthoFinder indicated that most genes (11 850 genes) were shared between all six *Fusarium* species, which ranged from 72.7% of all genes for *F. fracticaudum* to 76.4% of all genes for *F. circinatum* (Supplemental Table 4 and 5). Furthermore, 72 and 47 genes were uniquely identified in the *Fusarium* species associated with pine and members of Poaceae, respectively (Supplemental Table 6 and 7). Some of these unique genes were present in more than one copy. These were represented by one multiple-copy gene in *F. circinatum*, represented by FCIR_10_gene_25.58 and FCIR_10_gene_25.59. In *F. fracticaudum* there were two multiple-copy genes, with the first copy represented by FFRAC_10_gene_24.123 and FFRAC_6_gene_42.88 and the second copy by FFRAC_10_gene_17.127 and FFRAC_10_gene_17.128. *Fusarium pininemorale* had one multiple-copy gene represented by FPIN_12_gene_2.12 and FPIN_12_gene_2.24, while the multiple-copy genes in *F. konzum* were represented by FKON_2386_gene_0.6, FKON_2837_gene_0.2, and FKON_3361_gene_0.22. The two multicopy genes in *F. subglutinans* were represented by FSUB_25_gene_2.21 and FSUB_37_gene_0.24 and by FSUB_266_gene_0.11 and FSUB_309_gene_0.3. Interestingly, the multiple-copy genes were present on the same chromosome, except for FFRAC_10_gene_24.123, which was on chromosome 10, and the multiple-copy on chromosome 6. No inferences could be

made regarding the chromosomal location of the multiple-copy genes of *F. konzum* and *F. subglutinans* due to the fragmented condition of these assemblies.

Expression data for the unique genes of the pine pathogen, *F. circinatum*, were obtained from previous studies (Wingfield *et al.*, 2012, Van Wyk *et al.*, 2019) (Supplemental Table 8). This data specifically focussed on the expression of *F. circinatum* genes during carbon and nitrogen starvation versus expression on half-strength potato dextrose broth at room temperature. A gene was considered expressed if it had an RPKM (Reads Per Kilobase per Million mapped reads) value ≥ 0.2, and having at least three unique gene reads mapping to it (Wickramasinghe *et al.*, 2012). The combined results indicated that 62 of the 72 unique genes (86.11%) were transcribed under the described conditions (Wingfield *et al.*, 2012, Van Wyk *et al.*, 2019). Overall, the results indicated that a vast majority of the unique genes from the pine-associated *Fusarium* species are likely actively transcribed and encode for a protein.

Most of the unique genes had Blast2GO descriptions available (Supplemental Table 9 and 10). In the two reference genomes, *F. circinatum* (Supplemental Table 9A) and *F. temperatum* (Supplemental Table 10C), 11 (15.28%) and five genes (10.64%) respectively, had no Blast2GO descriptions. Of the 11 genes in *F. circinatum*, six had no BLAST hits in the non-redundant protein sequence database on NCBI, compared to three of the five equivalent genes in *F. temperatum*.

The Blast2GO results for the unique genes of both host-associated sets of *Fusarium* species were illustrated at level 4 GO terms for the two reference genomes (Supplemental Figure 1A-F). The biological processes of the 72 unique genes were mostly involved in the metabolism of macromolecules, organic cyclic compounds and cellular aromatic, nitrogen and heterocyclic compounds (Supplemental Figure 1A). The molecular functions associated with the 72 unique genes were predominantly involved in cation binding activity (Supplemental Figure 1B). In contrast, most of the cellular components of these genes were associated with the intrinsic component of the membrane (Supplemental Figure 1C). Apart from the 11 genes with no Blast2GO data, 26 genes (36.11%) were either hypothetical, predicted or uncharacterised proteins. The remaining 35 genes (48.61%) could be organised into seven groups each associated with different functions and characterised as either virulence promoters, genes acting as a cytochrome

P450, genes acting as transcriptional activators, genes being members of the major facilitator superfamily (MFS) transporters, genes involved in fungal development, genes involved in metabolic pathways and genes acting as hydrolases (Supplemental Table 9A).

The biological processes of the 47 unique genes associated with members of Poaceae were mostly involved in transport and the biosynthesis of organic substances (Supplemental Figure 1D). Anion binding, coenzyme binding and nucleoside phosphate-binding were some of the significant molecular functions associated with the 47 unique genes (Supplemental Figure 1E). In contrast, all the cellular components of these genes were associated with the intrinsic component of the membrane (Supplemental Figure 1F). Apart from the five genes with no Blast2GO data, 17 genes (36.17%) were either hypothetical, predicted or uncharacterised proteins. The remaining 25 genes (53.19%) could be organised into nine groups each associated with different functions and characterised as either virulence promoters, genes acting as a cytochrome P450, genes acting as transcriptional activators, genes being members of the major facilitator superfamily (MFS) transporters, genes involved in fungal development, genes involved in chemical reactions, genes involved in cell wall biosynthesis, genes involved in the transfer of electrons and genes containing a HET-domain (Supplemental Table 10C).

The Fischer exact test showed the enrichment of GO terms within the unique gene set, compared to the rest of each respective genome. In *F. circinatum*, five GO terms were significantly ($P < 0.05$) underrepresented in comparison to the whole genome (Supplemental Table 11A). In contrast, nine GO terms were enriched significantly ($P < 0.05$) in the *F. temperatum* genome (Supplemental Table 11B). Of genes associated with these enriched GO terms within the unique gene set, one gene had a biologically relevant role (FTEMP_12_gene_4.36), namely glutamate metabolism, and four genes (FTEMP_1_gene_58.115, FTEMP_7_gene_33.90, FTEMP_10_gene_4.59 and FTEMP_12_gene_4.36) had essential molecular functions, such as glutamate decarboxylase activity, steroid dehydrogenase activity and RNA-DNA hybrid ribonuclease activity. Furthermore, one gene (FTEMP_7_gene_33.90) was influential in cellular components, such as the microtubule and kinesin complex.

Phylogenetic origins of the unique genes

The phylogenetic origins of all the unique genes have been divided into nine groups (Table 3, Figure 1-8 for a representation of the first eight groups, Supplemental Figure 2A-I for the proposed phylogenetic history of each gene). The origins of the unique genes varied, but most belonged to groups 1, 4 and 9. Interestingly, group 8 was not represented in *F. circinatum* but was overrepresented in *F. temperatum* (4.30x more), in comparison to *F. circinatum*, and group 6 was overrepresented in *F. circinatum* (3.29x more), in comparison to *F. temperatum*.

Genomic distribution of the unique genes

The Chi-squared tests determined whether the unique genes differed significantly ($P < 0.05$) in their distribution frequency between telomeric and non-telomeric regions for both *F. circinatum* and *F. temperatum*, assuming that the frequencies do not differ significantly as the null hypothesis. Results indicated that the frequency of occurrence of the unique genes and their placement in telomeric regions differed significantly ($P < 0.05$, Supplemental Table 12). All the chromosomes of *F. circinatum* contained at least one unique gene, compared to chromosome 2 and 11 from *F. temperatum* that lacked any unique genes (Figure 9 and 10). The unique genes were located in areas that were neither gene-poor nor gene-rich (Figure 9 and 10). From the conservation of synteny (Figure 9 and 10) it was also apparent that for both *F. circinatum* and *F. temperatum*, in comparison to their respective genomes, inversions were frequently found in similar positions on the chromosomes and seemed to be closer to the telomeric regions. Chromosome 8 of both reference genomes had more inversions than the other chromosomes, and the CLP of chromosome 12 observed between *F. circinatum* and *F. temperatum* was portrayed in Figure 9 and 10.

The unique gene density per chromosome revealed some differences between the chromosomes for both *F. circinatum* and *F. temperatum* (Supplemental Table 13). Most chromosomes, except for chromosome 5, 10 and 12, had higher unique gene densities when the two assemblies were compared. Most chromosomes also had similar densities of unique genes, ranging from 1.0x (chromosome 4) – 1.97x (chromosome 7). Some differences appeared in the gene density (more than two-fold) of chromosome 9 and 12,

where *F. circinatum* contained more than *F. temperatum*, and vice versa, for these two chromosomes. Lastly, a major difference (more than seven-fold) occurred on chromosome 3 and 6, where *F. circinatum* contained more unique genes compared to the same chromosomes in *F. temperatum.* Chromosome 6 of *F. circinatum* did not have a telomeric cap at the "end" portion. It is also worth noting that chromosome 12 in *F. temperatum* contained much more unique genes compared to the same chromosome in *F. circinatum*.

Figure 11 and 12 illustrates that unique genes can cluster together, or be located in close proximity to one another, and either share the same ancestral origin or differ in their origins. This was the case for 44 of 72 (61.11%) unique genes from *F. circinatum*, and 17 of 47 (36.17%) unique genes from *F. temperatum*. Figure 11A shows that two unique genes (FCIR_6_gene_2.2 and FCIR_6_gene_2.142) are located next to each other, both having ancestral origins from group 3 (Table 3). Unique genes FCIR_6_gene_2.26 and FCIR_6_gene_2.131 were located next to each other close to the other two unique genes but had different ancestral origins, group 5 and 6, respectively. A similar trend was observed in Figure 11B and Figure 12B. Figure 12A is almost similar to 11A, except that FTEMP_10_gene_5.23 and FTEMP_10_gene_5.111 are located directly next to each other with origins from group 1, while FTEMP_10_gene_4.59 is nearby these two genes and also share an origin from group 1.

Investigations into the percentage of syntenous genes in the genomic location around each unique gene (this includes the unique gene, along with five genes upstream and downstream), shared by *F. circinatum* towards *F. temperatum*, and vice versa, indicated that these genomic locations in the pine- and Poaceae-associates, were normally distributed (Shapiro-Wilk's test for departure from normality; $P > 0.05$; W = 0.95 and $P > 0.05$; W = 0.86, respectively) (Figure 13, Supplemental Table 14 and 15). Here, the clusters displaying 5 out of 11 genes shared have been severely depleted.

DISCUSSION

Access to genomic data has enabled whole-genome comparisons, particularly focussing on *in silico* investigations into host-specificity (Chiara *et al.*, 2015, King *et al.*, 2015, Maphosa *et al.*, 2016, Bashyal *et al.*, 2017). With this study, we were able to examine host-specificity for two assemblages of *Fusarium* species within the FFSC, *i.e.*, those associated with pines and Poaceae, respectively. We identified two groups of genes unique to each of the two different groups of hosts. These genes differed between the two host groups in terms of identity and processes they are involved in. They also showed differences in their ancestral origins. Therefore, the processes that were involved in the adaptation of these two groups of *Fusarium* species to these two hosts (pine and Poaceae), likely involved different means of evolutionary acquisition. Genomic locations and distributions of these unique genes showed that they tended to cluster in telomeric regions, and they often also clustered together or close to each other. However, these two sets of unique genes could be grouped into functionally related genes, indicative of related processes that these genes are involved in.

The two sets of unique genes were identified and characterised, with 72 and 47 unique genes for the pine- and Poaceae-associated *Fusarium* species found. Investigations into the enrichment of GO terms indicated that glutamate decarboxylase activity, steroid dehydrogenase activity and RNA-DNA hybrid ribonuclease activity were representatives of these in *F. temperatum* indicative of the potential importance of these genes in the interaction between *Fusarium* and members of Poaceae. The presence of glutamate decarboxylase is known to play a role in fungal development during conidiation and germination and plays a role in oxidative stress tolerance in yeast (Christensen & Schmit, 1980, Coleman *et al.*, 2001). Steroid dehydrogenase is involved in the synthesis of active steroids, such as androgens and estrogens, where these steroids lead to increased virulence in bacteria (García-Gómez *et al.*, 2013). The enzyme RNA-DNA hybrid ribonuclease plays a role in bacteria by promoting cold growth, expression of virulence genes and acute infection (Cerritelli & Crouch, 1995, Lawal *et al.*, 2011). It has also been shown that *F. temperatum* CMW 40964 grows significantly faster at lower temperatures (10°C) than *F. circinatum* FSP34 (De Vos *et al.*, 2011). The presence of these enzymatic activities in other pathogens, such as the *Fusarium* species in this study, suggests that a

similar response could be expected in *Fusarium* species associated with members of Poaceae.

The identity of the unique genes itself could mostly be grouped as virulence factors, genes acting as a cytochrome P450 (CYP), genes acting as transcriptional activators, genes that are members of the major facilitator superfamily (MFS) transporters and genes involved in fungal development. Furthermore, *F. temperatum* had unique genes containing a heterokaryon incompatibility (HET)-domain, which was absent from *F. circinatum*. Virulence factors are molecules whose function is directly linked to damage and are produced by a range of pathogenic organisms, such as viruses, bacteria, parasites and fungi (Brunke *et al.*, 2016). These molecules enable pathogens to colonise a specific niche by promoting attachment to host cells, evade host immunity, produce toxic compounds for robustness, stress resistance and inflammation (Rhodes, 1988, Brunke *et al.*, 2016). These activities allow pathogen growth in the host upon nutrient acquisition and to perform pathogen morphological transition in the host to survive during different life-cycle phases. Additionally, fungal CYPs contribute to the developmental processes and pathogenesis of fungi (Fan *et al.*, 2013, Shin *et al.*, 2017). CYP genes are involved in the production of secondary metabolites to enable fungi to outcompete other pathogens localised in the same ecological niche. This competition between fungi, other pathogens and the respective host are mediated upon the secretion of toxins that promote virulence factors during the host-pathogen interaction (Fu & Viraraghavan, 2001, Casadevall, 2007, Evelin *et al.*, 2009, Gao *et al.*, 2010).

Transcriptional activators are transcription factors that increase the expression of genes. These are wide-ranging, including effector proteins (Heimel *et al.*, 2010, IpCho *et al.*, 2010, Wahl *et al.*, 2010, Zahiri *et al.*, 2010, Soyer *et al.*, 2015), other genes involved in plant-pathogen interactions, such as those genes encoding for cutinases, lipases and cellulases (Coradetti *et al.*, 2012, Van der Does *et al.*, 2016), and are even known to regulate a morphological switch between different lifestyles, which allows fungi to switch from being saprophytes to pathogens (Cain *et al.*, 2012). Furthermore, the major facilitator superfamily (MFS) is the largest of secondary transporters and includes members that function as uniporters, symporters or antiporters. Some MFS proteins in fungi are known for the role they play in resistance to natural toxic compounds and fungicides (Alexander *et al.*, 1999, Prasad & Kapoor, 2005, Roohparvar *et al.*, 2007) and

also to virulence (Pitkin *et al.*, 1996, Callahan *et al.*, 1999, Coleman & Mylonakis, 2009). Also, MFS proteins isolated from *Penicillium funiculosum* seem to be involved in acid resistance and intracellular pH homeostasis (Xu *et al.*, 2014).

Finally, genes with a HET-domain create a system for self/nonself recognition. This system allows for the prevention of heterokaryon formation between dissimilar individuals to restrict the horizontal transfer of cytoplasmic infectious elements (Saupe, 2000). These domains are also linked to programmed cell death in eukaryotes where it can be extremely beneficial to infected fungal cells, as a result of host defence responses, to prevent the spread of the infection to surrounding tissue (Gilchrist, 1998, Greenberg & Yao, 2004, Huckelhoven, 2007). The absence of HET-domains in pine-associated unique genes emphasises the different host defence responses between pines and members of Poaceae. It also emphasises the evolutionary changes Poaceae-associated *Fusarium* species had to undergo in order to acquire these genes to be successful colonisers of members of Poaceae. The presence of these groups of genes is indicative of the role in the survival and adaptation of these *Fusarium* species, likely providing a competitive advantage as well as providing a protective advantage against the host defence responses.

When examining the potential ancestral origins of the unique genes, most shared an evolutionary trajectory similar to the *Fusarium* species harbouring them. A few genes appear to have emerged from duplications of existing genes within the FFSC, or some facet of the *Fusarium* clade. Various genes had diverse and non-orthologous origins, such as other ascomycetous species, with others currently having no clear ancestral origin. It can be suggested that the evolution of these two unique gene sets involved multiple acquisitions of genes, possibly utilising horizontal gene transfer, from sources outside of the FFSC. Furthermore, it also appears as if some unique genes cluster together and have diverse ancestral origins, suggesting that certain genomic regions are more likely to harbour the acquisition of these unique genes. Some genomic regions are known for being "hotspots" for niche- or host-specific genes, such as subtelomeric regions. Virulence genes involved in host-specificity in *F. graminearum* (Galagan *et al.*, 2005, Perrin *et al.*, 2007, Zhao *et al.*, 2014) and *F. fujikuroi* (Wiemann *et al.*, 2013) have been detected to be located in subtelomeric regions. Furthermore, unique genes potentially involved in growth in *F. circinatum* (species-specific genes), were located in the subtelomeric region (Van Wyk *et al.*, 2019). In this study, some of the unique genes are significantly located

towards the chromosomal ends and are likely located within the subtelomeric regions. This suggests that these subtelomeric regions possessing unique genes involved areas where the frequent break in genic synteny is often indicative of "hotspots" for niche- and host-specific genes. These regions allowed the gain of new gene functions which could enable fungal colonisation on their respective hosts.

The variability of the two reference genomes was most evident when comparisons of inversions were investigated. More inversions were closer to the ends of chromosomes. This was known from other literature (Zhao *et al.*, 2014, Van Dam *et al.*, 2017, King *et al.*, 2018). Also, chromosome 8 was the most variable, in comparison to the other core chromosomes (1-11). Chromosome 8 is also known to possess a sizable reciprocal translocation with chromosome 11 within species of the American clade of the FFSC (De Vos *et al.*, 2014, Wingfield *et al.*, 2015b, Wingfield *et al.*, 2017, Wingfield *et al.*, 2018a, Wingfield *et al.*, 2018b). This suggests that this chromosome has a larger predisposition to chromosomal architectural changes.

The importance of chromosome 12 in possessing unique genes poses an interesting question. This chromosome is thought to be a dispensable chromosome within the FFSC (Xu *et al.*, 1995, Jurgenson *et al.*, 2002, Ma *et al.*, 2010, Van der Nest *et al.*, 2014) and can be strain-specific in members of the *F. fujikuroi* complex (Wiemann *et al.*, 2013, Van der Nest *et al.*, 2014). In *F. oxysporum* f. sp. *lycopersici* it is known that dispensable chromosomes impart pathogenicity to the strains possessing them, and being enriched in, amongst others, genes involved in virulence and transcription factors which contribute to niche-specificity in these species (Han *et al.*, 2001, Coleman *et al.*, 2009, Ma *et al.*, 2010). Chromosome 12 is not essential for pathogenicity (Wiemann *et al.*, 2013, Van der Nest *et al.*, 2014, Slinski *et al.*, 2016) and the role of this chromosome in niche-specificity is unknown. However, what is clear is that the larger chromosome 12 of *F. temperatum* possessed more unique genes potentially involved in niche-specificity amongst *Fusarium* species that are associated with Poaceae, which implies that the CLP between *F. circinatum* and *F. temperatum* accounts for the addition of unique genes. Furthermore, chromosome 12 harbours more unique genes than the core chromosomes. Taken together, this suggests that the variability of chromosome 12 contributes to genes involved in niche-specificity.

CONCLUSIONS

This study has provided new insights into the host-specificity of agriculturally important *Fusarium* species. Comparative genomics was utilised with near-complete and high-quality genomes. This allowed for the understanding of the different relationships employed by a host and a pathogen during their interaction. From this, a group of genes were identified and characterised with regards to their potential roles in host-specificity. In this study, two different gene sets were identified, one unique to pine-associated *Fusarium* species, and another to Poaceae-associated *Fusarium* species. These genes are hypothesised to play a role in niche-specificity between these two different groups. These genes differed in function, products and processes as they are involved in fungal adaptation and survival. The ancestral origins of these genes showed to have polyphyletic origins with some possibly acquired through horizontal gene transfer from sources outside the FFSC.

The CLP of chromosome 12 between the four *Fusarium* species assembled to chromosomal level, in comparison to the core chromosomes, suggest that this chromosome is rapidly evolving. Chromosome 12 may or may not directly or indirectly contribute to pathogenicity. It is likely that the smallest chromosomes of *Fusarium* species from the FFSC forms part of the accessory genome and serve as an adaptive region of the genome where new genes can rapidly evolve to promote, amongst others, host-specificity. Since previous studies show that chromosome 12 is not involved in virulence, the presence of this chromosome likely contributes to niche-specificity in *Fusarium* species associated with members of Poaceae, specifically the part of the chromosome displaying the length polymorphism. In particular, chromosome 12 from *F. circinatum* is much smaller in comparison to *F. temperatum*, and the synteny corresponds to the internal portion of *F. temperatum* of chromosome 12. The synteny would imply that the extension in the distal and proximal ends of the chromosome in *F. temperatum* has driven the formation of novel genes, particularly the niche-specific genes of this study. The presence of unique genes and the length polymorphism on *F. temperatum* chromosome 12 might be a recent evolutionary event, in comparison to *F. circinatum*.

This study discovered that the frequency of genes potentially involved in host-specificity, niche-specificity or virulence differ significantly between telomeric and non-telomeric regions. Future work from this dissertation may include the studying of the functional relevance of the unique genes and their gene products associated with *Fusarium* species colonising pines and members of Poaceae. Furthermore, investigations into the expression of these genes in these two hosts and at different time points would allow for better quantification of how the pathogen interacts with its host, using the unique genes identified in this study.

REFERENCES

1. Abdullah AS, Moffat CS, Lopez-Ruiz FJ, Gibberd MR, Hamblin J & Zerihun A (2017) Host–multi-pathogen warfare: Pathogen interactions in co-infected plants. *Frontiers in Plant Science* **8**: 1806.

2. Akagi Y, Akamatsu H, Otani H & Kodama M (2009) Horizontal chromosome transfer, a mechanism for the evolution and differentiation of a plant-pathogenic fungus. *Eukaryotic Cell* **8**: 1732-1738.

3. Alexander NJ, McCormick SP & Hohn TM (1999) *TRI12*, a trichothecene efflux pump from *Fusarium sporotrichioides*: Gene isolation and expression in yeast. *Molecular and General Genetics* **261**: 977-984.

4. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

5. Armitage AD, Taylor A, Sobczyk MK, Baxter L, Greenfield BPJ, Bates HJ, *et al.* (2018) Characterisation of pathogen-specific regions and novel effector candidates in *Fusarium oxysporum* f. sp. *cepae*. *Scientific Reports* **8**: 13530.

6. Aylward J, Steenkamp ET, Dreyer LL, Roets F, Wingfield BD & Wingfield MJ (2017) A plant pathology perspective of fungal genome sequencing. *IMA Fungus* **8**: 1-15.

7. Bashyal BM, Rawat K, Sharma S, Kulshreshtha D, Gopala Krishnan S, Singh AK, *et al.* (2017) Whole genome sequencing of *Fusarium fujikuroi* provides insight into the role of secretory proteins and cell wall degrading enzymes in causing bakanae disease of rice. *Frontiers in Plant Science* **8**: 2013.

8. Bäumler A & Fang FC (2013) Host specificity of bacterial pathogens. *Cold Spring Harbor Perspectives in Medicine* **3**: a010041.

9. Bottalico A & Perrone G (2002) Toxigenic *Fusarium* species and mycotoxins associated with head blight in small-grain cereals in Europe. *European Journal of Plant Pathology* **108**: 611-624.

10. Brunke S, Mogavero S, Kasper L & Hube B (2016) Virulence factors in fungal pathogens of man. *Current Opinion in Microbiology* **32**: 89-95.

11. Cain CW, Lohse MB, Homann OR, Sil A & Johnson AD (2012) A conserved transcriptional regulator governs fungal morphology in widely diverged species. *Genetics* **190**: 511-521.

12. Callahan TM, Rose MS, Meade MJ, Ehrenshaft M & Upchurch RG (1999) CFP, the putative cercosporin transporter of *Cercospora kikuchii*, is required for wild type cercosporin production, resistance, and virulence on soybean. *Molecular Plant-Microbe Interactions* **12**: 901-910.

13. Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P & Le Bourgeois P (2004) Chromosomal constraints in gram-positive bacteria revealed by artificial inversions. *Molecular Microbiology* **51**: 511-522.

14. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, *et al.* (2008) MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188-196.

15. Casadevall A (2007) Determinants of virulence in the pathogenic fungi. *Fungal Biology Reviews* **21**: 130-132.

16. Cerritelli SM & Crouch RJ (1995) The non-RNase H domain of *Saccharomyces cerevisiae* RNase H1 binds double-stranded RNA: Magnesium modulates the switch between double-stranded RNA binding and RNase H activity. *RNA* **1**: 246-259.

17. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, *et al.* (2004) Insights into the evolution of *Yersinia pestis* through whole-genome

comparison with *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 13826-13831.

18. Chang DC, Grant GB, O'Donnell K, Wannemuehler KA, Noble-Wang J, Rao CY, *et al.* (2006) Multistate outbreak of *Fusarium* keratitis associated with use of a contact lens solution. *American Journal of Ophthalmology* **142**: 896-897.

19. Chiara M, Fanelli F, Mule G, Logrieco AF, Pesole G, Leslie JF, *et al.* (2015) Genome sequencing of multiple isolates highlights subtelomeric genomic diversity within *Fusarium fujikuroi*. *Genome Biology and Evolution* **7**: 3062-3069.

20. Christensen RL & Schmit JC (1980) Regulation and glutamic acid decarboxylase during *Neurospora crassa* conidial germination. *Journal of Bacteriology* **144**: 983-990.

21. Coleman JJ & Mylonakis E (2009) Efflux in fungi: La piece de resistance. *PLoS Pathogens* **5**: e1000486.

22. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, *et al.* (2009) The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* **5**: e1000618.

23. Coleman JJ, Wasmann CC, Usami T, White GJ, Temporini ED, McCluskey K, *et al.* (2011) Characterization of the gene encoding pisatin demethylase (*FoPDA*1) in *Fusarium oxysporum*. *Molecular Plant-Microbe Interactions* **24**: 1482-1491.

24. Coleman ST, Fang TK, Rovinsky SA, Turano FJ & Moye-Rowley WS (2001) Expression of a glutamate decarboxylase homologue is required for normal oxidative stress tolerance in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **276**: 244-250.

25. Conesa A & Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* **2008**: 619832.

26. Coradetti ST, Craig JP, Xiong Y, Shock T, Tian C & Glass NL (2012) Conserved and essential transcription factors for cellulase gene expression in ascomycete fungi. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 7397-7402.

27. Covert SF (1998) Supernumerary chromosomes in filamentous fungi. *Current Genetics* **33**: 311-319.

28. Croll D & McDonald BA (2012) The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens* **8**: e1002608.

29. Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, *et al.* (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400-1402.

30. De Vos L, Steenkamp ET, Martin SH, Santana QC, Fourie G, van der Merwe NA, *et al.* (2014) Genome-wide macrosynteny among *Fusarium* species in the *Gibberella fujikuroi* complex revealed by amplified fragment length polymorphisms. *PloS one* **9**: e114682.

31. De Vos L, Van der Nest MA, Santana QC, Wingfield BD & Steenkamp ET (2020, unpublished) Enhancement of *Fusarium circinatum* genomic assemblies. *BioTechniques*.

32. De Vos L, Van der Nest MA, Van der Merwe NA, Myburg AA, Wingfield MJ & Wingfield BD (2011) Genetic analysis of growth, morphology and pathogenicity in the F1 progeny of an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Biology* **115**: 902-908.

33. Desjardins AE, Maragos CM & Proctor RH (2006) Maize ear rot and moniliformin contamination by cryptic species of *Fusarium subglutinans*. *Journal of Agricultural and Food Chemistry* **54**: 7383.

34. Desjardins AE, Plattner RD & Gordon TR (2000) *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. *Mycological Research* **104**: 865-872.

35. Drillon G, Carbone A & Fischer G (2014) SynChro: A fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. *PLoS One* **9**: e92621.

36. Edwards J, Auer D, De Alwis S, Summerell B, Aoki T, Proctor RH, *et al.* (2016) *Fusarium agapanthi* sp. nov., a novel bikaverin and fusarubin-producing leaf and stem spot pathogen of *Agapanthus praecox* (African lily) from Australia and Italy. *Mycologia* **108**: 981-992.

37. Emms DM & Kelly S (2015) OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**: 157.

38. Evelin H, Kapoor R & Giri B (2009) Arbuscular mycorrhizal fungi in alleviation of salt stress: A review. *Annals of Botany* **104**: 1263-1280.

39. Fan J, Urban M, Parker JE, Brewer HC, Kelly SL, Hammond-Kosack KE, *et al.* (2013) Characterization of the sterol 14α-demethylases of *Fusarium graminearum* identifies a novel genus-specific CYP51 function. *New Phytologist* **198**: 821-835.

40. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, *et al.* (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**: 186-194.

41. Fu Y & Viraraghavan T (2001) Fungal decolorization of dye wastewaters: A review. *Bioresource Technology* **79**: 251-262.

42. Fulnecková J, Sevcíková T, Fajkus J, Lukesová A, Lukes M, Vlcek C, *et al.* (2013) A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution* **5**: 468-483.

43. Galagan JE, Calvo SE, Cuomo C, Ma L, Wortman JR, Batzoglou S, *et al.* (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105.

44. Gao D, Du L, Yang J, Wu W & Liang H (2010) A critical review of the application of white rot fungus to environmental pollution control. *Critical Reviews in Biotechnology* **30**: 70-77.

45. García-Gómez E, González-Pedrajo B & Camacho-Arroyo I (2013) Role of sex steroid hormones in bacterial-host interactions. *BioMed Research International* **2013**: 928290.

46. Garcia-Pedrajas MD & Roncero MI (1996) A homologous and self-replicating system for efficient transformation of *Fusarium oxysporum*. *Current Genetics* **29**: 191-198.

47. Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, Brandt ME, *et al.* (2013) One fungus, one name: Defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400-408.

48. Gel B & Serra E (2017) karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**: 3088-3090.

49. Gilchrist DG (1998) Programmed cell death in plant disease: The purpose and promise of cellular suicide. *Annual Reviews in Phytopathology* **36**: 393-414.

50. Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics* **7**: e1002070.

51. Gordon TR, Storer AJ & Okamoto D (1996) Population structure of the pitch canker pathogen, *Fusarium subglutinans* f. sp. *pini,* in California. *Mycological Research* **100**: 850-854.

52. Greenberg JT & Yao N (2004) The role and regulation of programmed cell death in plant-pathogen interactions. *Cellular Microbiology* **6**: 201-211.

53. Han Y, Liu X, Benny U, Kistler HC & Van Etten HD (2001) Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*. *Plant Journal* **25**: 305-314.

54. Hardison RC (2003) Comparative genomics. *PLoS Biology* **1**: e58.

55. Harris RS (2007) Improved pairwise alignment of genomic DNA. Thesis, The Pennsylvania State University.

56. Heimel K, Scherer M, Vranes M, Wahl R, Pothiratana C, Schuler D, *et al.* (2010) The transcription factor Rbf1 is the master regulator for *b*-mating type controlled pathogenic development in *Ustilago maydis*. *PLoS Pathogens* **6**: e1001035.

57. Hepting GH & Roth ER (1946) Pitch canker, a new disease of some southern pines. *Journal of Forestry* **44**: 742-744.

58. Hernandez-Escribano L, Iturritxa E, Elvira-Recuenco M, Berbegal M, Campos JA, Renobales G, *et al.* (2018) Herbaceous plants in the understory of a pitch canker-affected *Pinus radiata* plantation are endophytically infected with *Fusarium circinatum*. *Fungal Ecology* **32**: 65-71.

59. Herron DA, Wingfield MJ, Wingfield BD, Rodas CA, Marincowitz S & Steenkamp ET (2015) Novel taxa in the *Fusarium fujikuroi* species complex from *Pinus* spp. *Studies in Mycology* **80**: 131-150.

60. Hill CW & Gray JA (1988) Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. *Genetics* **119**: 771-778.

61. Hoff KJ, Lomsadze A, Borodovsky M & Stanke M (2019) Whole-genome annotation with BRAKER. *Methods in Molecular Biology* **1962**: 65-95.

62. Huckelhoven R (2007) Cell wall-associated mechanisms of disease resistance and susceptibility. *Annual Reviews of Phytopathology* **45**: 101-127.

63. IpCho SV, Tan KC, Koh G, Gummer J, Oliver RP, Trengove RD, *et al.* (2010) The transcription factor StuA regulates central carbon metabolism, mycotoxin production, and effector gene expression in the wheat pathogen *Stagonospora nodorum*. *Eukaryotic Cell* **9**: 1100-1108.

64. Jones DT, Taylor WR & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**: 275-282.

65. Jurgenson JE, Zeller KA & Leslie JF (2002) Expanded genetic map of *Gibberella moniliformis* (*Fusarium verticillioides*). *Applied and Environmental Microbiology* **68**: 1972-1979.

66. Katoh K, Rozewicki J & Yamada KD (2017) MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**: 1160-1166.

67. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

68. King R, Brown NA, Urban M & Hammond-Kosack KE (2018) Inter-genome comparison of the Quorn fungus *Fusarium venenatum* and the closely related plant infecting pathogen *Fusarium graminearum*. *BMC Genomics* **19**: 269.

69. King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K & Hammond-Kosack KE (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**: 544.

70. Kisiela DI, Chattopadhyay S, Libby SJ, Karlinsey JE, Fang FC, Tchesnokova V, *et al.* (2012) Evolution of *Salmonella enterica* virulence via point mutations in the fimbrial adhesin. *PLoS Pathogens* **8**: e1002733.

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

72. Kumar S, Stecher G & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870-1874.

73. Lawal A, Jejelowo O, Chopra AK & Rosenzweig JA (2011) Ribonucleases and bacterial virulence. *Microbial Biotechnology* **4**: 558-571.

74. Lawley TD, Bouley DM, Hoy YE, Gerke C, Relman DA & Monack DM (2008) Host transmission of *Salmonella enterica* serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. *Infection and Immunity* **76**: 403-416.

75. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**: 1307-1320.

76. Leslie JF, Zeller KA, Logrieco A, Mule G, Moretti A & Ritieni A (2004) Species diversity of and toxin production by *Gibberella fujikuroi* species complex strains isolated from native prairie grasses in Kansas. *Applied and Environmental Microbiology* **70**: 2254-2262.

77. Ma L, Van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367-373.

78. Maphosa MN, Steenkamp ET & Wingfield BD (2016) Genome-based selection and characterization of *Fusarium circinatum*-specific sequences. *G3* **6**: 631-639.

79. Martín-García J, Lukačevičová A, Flores-Pacheco JA, Diez JJ & Dvořák M (2018) Evaluation of the susceptibility of several Czech conifer provenances to *Fusarium circinatum*. *Forests* **9**: 72.

80. Navarro-Velasco GY, Prados-Rosales RC, Ortíz-Urquiza A, Quesada-Moraga E & Di Pietro A (2011) *Galleria mellonella* as model host for the trans-kingdom pathogen *Fusarium oxysporum*. *Fungal Genetics and Biology* **48**: 1124-1129.

81. Niehaus E, Kim H, Münsterkötter M, Janevska S, Arndt B, Kalinina SA, *et al.* (2017) Comparative genomics of geographically distant *Fusarium fujikuroi* isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLoS Pathogens* **13**: e1006670.

82. Niehaus E, Münsterkötter M, Proctor RH, Brown DW, Sharon A, Idan Y, *et al.* (2016) Comparative "omics" of the *Fusarium fujikuroi* species complex highlights differences in genetic potential and metabolite synthesis. *Genome Biology and Evolution* **8**: 3574-3599.

83. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, *et al.* (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**: 1151-1156.

84. Nirmala J, Dahl S, Steffenson BJ, Kannangara CG, Von Wettstein D, Chen X, *et al.* (2007) Proteolysis of the barley receptor-like protein kinase RPG1 by a proteasome pathway is correlated with *Rpg1*-mediated stem rust resistance. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 10276-10281.

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

86. O'Donnell K, Sutton DA, Rinaldi MG, Magnon KC, Cox PA, Revankar SG, *et al.* (2004) Genetic diversity of human pathogenic members of the *Fusarium oxysporum* complex inferred from multilocus DNA sequence data and amplified fragment length polymorphism analyses: Evidence for the recent dispersion of a geographically

widespread clonal lineage and nosocomial origin. *Journal of Clinical Microbiology* **42**: 5109-5120.

87. Ortoneda M, Guarro J, Madrid MP, Caracuel Z, Roncero MIG, Mayayo E, *et al.* (2004) *Fusarium oxysporum* as a multihost model for the genetic dissection of fungal virulence in plants and mammals. *Infection and Immunity* **72**: 1760-1766.

88. Papadopoulos JS & Agarwala R (2007) COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**: 1073-1079.

89. Parry D, Jenkinson P & McLeod L (1995) Fusarium ear blight (scab) in small grain cereals - A review. *Plant Pathology* **44**: 207-238.

90. Perrin RM, Fedorova ND, Bok JW, Cramer Jr RA, Wortman JR, Kim HS, *et al.* (2007) Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. *PLoS Pathogens* **3**: e50.

91. Pitkin JW, Panaccione DG & Walton JD (1996) A putative cyclic peptide efflux pump encoded by the *TOXA* gene of the plant-pathogenic fungus *Cochliobolus carbonum*. *Microbiology* **142**: 1557-1565.

92. Prasad R & Kapoor K (2005) Multidrug resistance in yeast *Candida*. *International Review of Cytology* **242**: 215-248.

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

94. Rebollo JE, Francois V & Louarn JM (1988) Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Proceedings of the National Academy of Sciences in the United States of America* **85**: 9391-9395.

95. Rhodes JC (1988) Virulence factors in fungal pathogens. *Microbiological Sciences* **5**: 252-254.

96. Roohparvar R, De Waard MA, Kema GHJ & Zwiers L (2007) MgMfs1, a major facilitator superfamily transporter from the fungal wheat pathogen *Mycosphaerella graminicola*, is a strong protectant against natural toxic compounds and fungicides. *Fungal Genetics and Biology* **44**: 378-388.

97. Saupe SJ (2000) Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews* **64**: 489-502.

98. Scauflaire J, Gourgue M & Munaut F (2011) *Fusarium temperatum* sp. nov. from maize, an emergent species closely related to *Fusarium subglutinans*. *Mycologia* **103**: 586-597.

99. Shin JY, Bui DC, Lee Y, Nam H, Jung S, Fang M, *et al.* (2017) Functional characterization of cytochrome P450 monooxygenases in the cereal head blight fungus *Fusarium graminearum*. *Environmental Microbiology* **19**: 2053-2067.

100. Slinski SL, Kirkpatrick SC & Gordon TR (2016) Inheritance of virulence in *Fusarium circinatum*, the cause of pitch canker in pines. *Plant Pathology* **65**: 1292-1296.

101. Soyer JL, Hamiot A, Ollivier B, Balesdent MH, Rouxel T & Fudal I (2015) The APSES transcription factor LmStuA is required for sporulation, pathogenic development and effector gene expression in *Leptosphaeria maculans*. *Molecular Plant Pathology* **16**: 1000-1005.

102. Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, Manners JM, *et al.* (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biology and Evolution* **7**: 1613-1627.

103. Srivastava AK, Kashyap PL, Chakdar H, Kumar M, Srivastava AK, Yadav J, *et al.* (2018) First *de novo* draft genome sequence of the pathogenic fungus *Fusarium udum* F02845, associated with pigeonpea (*Cajanus cajan* L. Millspaugh) wilt. *Microbiology Resource Announcements* **7**: e01001-e01018.

104. Steenkamp ET, Wingfield BD, Desjardins AE, Marasas WF & Wingfield MJ (2002) Cryptic speciation in *Fusarium subglutinans*. *Mycologia* **94**: 1032-1043.

105. Sutton DA & Brandt ME (2011) *Fusarium and other opportunistic hyaline fungi*. ASM Press, Washington, DC.

106. Swett CL & Gordon TR (2012) First report of grass species (Poaceae) as naturally occurring hosts of the pine pathogen *Gibberella circinata*. *Plant Disease* **96**: 908-908.

107. Swett CL & Gordon TR (2015) Endophytic association of the pine pathogen *Fusarium circinatum* with corn (*Zea mays*). *Fungal Ecology* **13**: 120-129.

108. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO & Borodovsky M (2008) Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Research* **18**: 1979-1990.

109. Van Dam P, Fokkens L, Ayukawa Y, Van der Gragt M, Horst A, Brankovics B, *et al.* (2017) A mobile pathogenicity chromosome in *Fusarium oxysporum* for infection of multiple cucurbit species. *Scientific Reports* **7**: 9042.

110. Van der Does HC, Fokkens L, Yang A, Schmidt SM, Langereis L, Lukasiewicz JM, *et al.* (2016) Transcription factors encoded on core and accessory chromosomes of *Fusarium oxysporum* induce expression of effector genes. *PLoS Genetics* **12**: e1006401.

111. Van der Nest MA, Beirn LA, Crouch JA, Demers JE, De Beer ZW, De Vos L, *et al.* (2014) Draft genomes of *Amanita jacksonii, Ceratocystis albifundus, Fusarium circinatum, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana,* and *Sclerotinia echinophila*. *IMA Fungus* **5**: 473-486.

112. Van der Nest MA, Steenkamp ET, Roodt D, Soal NC, Palmer M, Chan WY, *et al.* (2019) Genomic analysis of the aggressive tree pathogen *Ceratocystis albifundus*. *Fungal Biology* **123**: 351-363.

113. Van Wyk S, Wingfield BD, De Vos L, Santana QC, Van der Merwe NA & Steenkamp ET (2018) Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*. *IMA Fungus* **9**: 27-36.

114. Van Wyk S, Wingfield BD, De Vos L, Van der Merwe NA, Santana QC & Steenkamp ET (2019) Repeat-induced point mutations drive divergence between *Fusarium circinatum* and its close relatives. *Pathogens* **8**: E298.

115. Waalwijk C, Vanheule A, Audenaert K, Zhang H, Warris S, Van de Geest H, *et al.* (2017) *Fusarium* in the age of genomics. *Tropical Plant Pathology* **42**: 184-189.

116. Wahl R, Zahiri A & Kamper J (2010) The *Ustilago maydis b* mating type locus controls hyphal proliferation and expression of secreted virulence factors *in planta*. *Molecular Microbiology* **75**: 208-220.

117. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, *et al.* (2017) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35**: 543-548.

118. Whelan S & Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* **18**: 691-699.

119. Wickramasinghe S, Rincon G, Islas-Trejo A & Medrano JF (2012) Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* **13**: 45.

120. Wiemann P, Sieber CM, Von Bargen KW, Studt L, Niehaus EM, Espino JJ, *et al.* (2013) Deciphering the cryptic genome: Genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. *PLoS Pathogens* **9**: e1003475.

121. Wingfield BD, Ades PK, Al-Naemi FA, Beirn LA, Bihon W, Crouch JA, *et al.* (2015a) Draft genome sequences of *Chrysoporthe austroafricana, Diplodia scrobiculata, Fusarium nygamai, Leptographium lundbergii, Limonomyces culmigenus, Stagonosporopsis tanaceti,* and *Thielaviopsis punctulata*. *IMA Fungus* **6**: 233-248.

122. Wingfield BD, Barnes I, De Beer ZW, De Vos L, Duong TA, Kanzi AM, *et al.* (2015b) Draft genome sequences of *Ceratocystis eucalypticola, Chrysoporthe cubensis, C. deuterocubensis, Davidsoniella virescens, Fusarium temperatum, Graphilbum fragrans, Penicillium nordicum,* and *Thielaviopsis musarum*. *IMA Fungus* **6**: 493-506.

123. Wingfield BD, Berger DK, Steenkamp ET, Lim HJ, Duong TA, Bluhm BH, *et al.* (2017) Draft genome of *Cercospora zeina, Fusarium pininemorale, Hawksworthiomyces lignivorus, Huntiella decipiens* and *Ophiostoma ips*. *IMA Fungus* **8**: 385-396.

124. Wingfield BD, Bills GF, Dong Y, Huang W, Nel WJ, Swalarsk-Parry BS, *et al.* (2018a) Draft genome sequence of *Annulohypoxylon stygium, Aspergillus mulundensis, Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora, Fusarium fracticaudum, Phialophora* cf. *hyalina*, and *Morchella septimelata*. *IMA Fungus* **9**: 199-223.

125. Wingfield BD, Liu M, Nguyen HDT, Lane FA, Morgan SW, De Vos L, *et al.* (2018b) Nine draft genome sequences of *Claviceps purpurea* s. lat., including *C. arundinis, C. humidiphila,* and *C.* cf. *spartinae*, pseudomolecules for the pitch canker pathogen *Fusarium circinatum*, draft genome of *Davidsoniella eucalypti, Grosmannia galeiformis, Quambalaria eucalypti*, and *Teratosphaeria destructans*. *IMA Fungus* **9**: 401-418.

126. Wingfield BD, Steenkamp ET, Santana QC, Coetzee MPA, Bam S, Barnes I, *et al.* (2012) First fungal genome sequence from Africa: A preliminary analysis. *South African Journal of Science* **108**: 104-112.

127. Wittenberg AHJ, Van der Lee TAJ, M'Barek SB, Ware SB, Goodwin SB, Kilian A, *et al.* (2009) Meiosis drives extraordinary genome plasticity in the haploid fungal plant pathogen *Mycosphaerella graminicola*. *PLoS One* **4**: e5863.

128. Xu X, Chen J, Xu H & Li D (2014) Role of a major facilitator superfamily transporter in adaptation capacity of *Penicillium funiculosum* under extreme acidic stress. *Fungal Genetics and Biology* **69**: 75-83.

129. Xu J, Yan K, Dickman MB & Leslie F (1995) Electrophoretic karyotypes distinguish the biological species of *Gibberella fujikuroi* (*Fusarium* section *Liseola*). *Molecular Plant-Microbe Interactions* **8**: 7.

130. Zahiri A, Heimel K, Wahl R, Rath M & Kamper J (2010) The *Ustilago maydis* forkhead transcription factor Fox1 is involved in the regulation of genes required for the attenuation of plant defenses during pathogenic development. *Molecular Plant-Microbe Interactions* **23**: 1118-1129.

131. Zeller KA, Summerell BA, Bullock S & Leslie JF (2003) *Gibberella konza* (*Fusarium konzum*) sp. nov., a new biological species within the *Gibberella fujikuroi* species complex from prairie grasses. *Mycologia* **95**: 943-954.

132. Zhao C, Waalwijk C, De Wit PJGM, Tang D & Van der Lee T (2014) Relocation of genes generates non-conserved chromosomal segments in *Fusarium graminearum* that show distinct and co-regulated gene expression patterns. *BMC Genomics* **15**: 191.

TABLES

Table 1. Genome statistics for the *Fusarium* genomes examined in this study.

Table 2. Presence of telomeres at chromosomal ends for the two representative *Fusarium* species examined.

Table 3. The number of unique *Fusarium* genes associated with pine and members of the Poaceae, belonging to each proposed group based on the most probable phylogenetic origin of the respective genes.

Table 1. Genome statistics for the *Fusarium* genomes examined in this study.

***Fusarium* species[1,2]**	**Genome size (bp)**	**Number of scaffolds**	**Number of orientated chromosomes**	**Gene density (ORFs/Mb)**	**Genes with coding sequences[3]**	**GC content (%)**
F. circinatum FSP34 (CMW 51752)	45 018 643	49	12	344.50	15 509	47
F. fracticaudum CMW 25245	46 252 763	50	12	352.28	16 294	48
F. pininemorale CMW 25243	47 778 485	153	12	332.35	15 879	46
F. konzum NRRL 11616	43 487 959	2 262	N/A	374.49	16 286	49
F. subglutinans NRRL 22016	44 190 517	733	N/A	356.23	15 742	48
F. temperatum CMW 40964	45 458 781	43	12	341.67	15 532	47

[1]CMW: Culture collection of the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria.
[2]NRRL: Northern Regional Research Laboratory, United States Department of Agricultures (USDA), Illinois.
[3]As determined using MAKER v. 2.31.8 (Cantarel *et al.*, 2008).

Table 2. Presence of telomeres at chromosomal ends for the two representative *Fusarium* species examined.

Chromosome number	***F. circinatum* FSP34**		***F. temperatum* CMW 40964**	
	B[1]	E[2]	B[1]	E[2]
1	✓	✗	✓	✓
2	✗	✓	✗	✗
3	✓	✓	✓	✓
4	✓	✓	✓	✓
5	✗	✓	✓	✗
6	✓	✗	✓	✓
7	✓	✓	✓	✓
8	✓	✗	✓	✗
9	✓	✓	✓	✓
10	✓	✓	✓	✓
11	✓	✓	✓	✓
12	✓	✓	✓	✓

[1]Beginning of chromosome.
[2]End of chromosome.

Table 3. The number of unique *Fusarium* genes associated with pine and members of the Poaceae, belonging to each proposed group based on the most probable phylogenetic origin of the respective genes.

Group no.	Classification based on ancestral origin*	*F. circinatum* FSP34	*F. temperatum* CMW 40964
1	Emerged within the FFSC and were retained only in certain of its lineages	18 (25.0%)	13 (27.7%)
2	Emerged due to a duplication in the FFSC but retained/acquired only in a few FFSC species	3 (4.2%)	5 (10.6%)
3	Originated in the FOSC	5 (6.9%)	1 (2.1%)
4	Originated within the broader FFSC+FOSC clade but only retained in certain species	15 (20.8%)	14 (29.8%)
5	Originated in the *Fusarium* F2 clade but retained/acquired only in certain *Fusarium* species	5 (6.9%)	1 (2.1%)
6	Originated in *Fusarium* F3 clade that excludes *F. solani* (*Nectria haematococca*)	4 (5.6%)	3 (6.4%)
7	Ascomycetous origin outside the *Fusarium* F2 clade	7 (9.7%)	2 (4.3%)
8	Emerged due to a duplication in the broader *Fusarium* F2 clade but retained/acquired only in certain species complexes	0 (0.0%)	2 (4.3%)
9	No significant hits based on blast searches against the NCBI database	15 (20.8%)	6 (12.8%)
Total genes		**72**	**47**

*FFSC = *Fusarium fujikuroi* species complex.
FOSC = *Fusarium oxysporum* species complex.
F2 and F3 clade = Refer to article by Geiser *et al.* (2013).

FIGURES

Figure 9. Pine-associated unique genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FCIR = *Fusarium circinatum.*

Figure 10. Poaceae-associated unique genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FTEMP = *Fusarium temperatum*.

Figure 11. Integration of SynChro and phylogenetic results of the pine-associated genes clustering together or those in close proximity to each other.

Figure 12. Integration of SynChro and phylogenetic results of the Poaceae-associated genes clustering together or those in close proximity to each other.

Figure 13. Data retrieved from the Shapiro-Wilk's test for departure from normality for the unique genes of the pine- and Poaceae-associated *Fusarium* species, represented by *F. circinatum* FSP34 and *F. temperatum* CMW 40964, respectively.

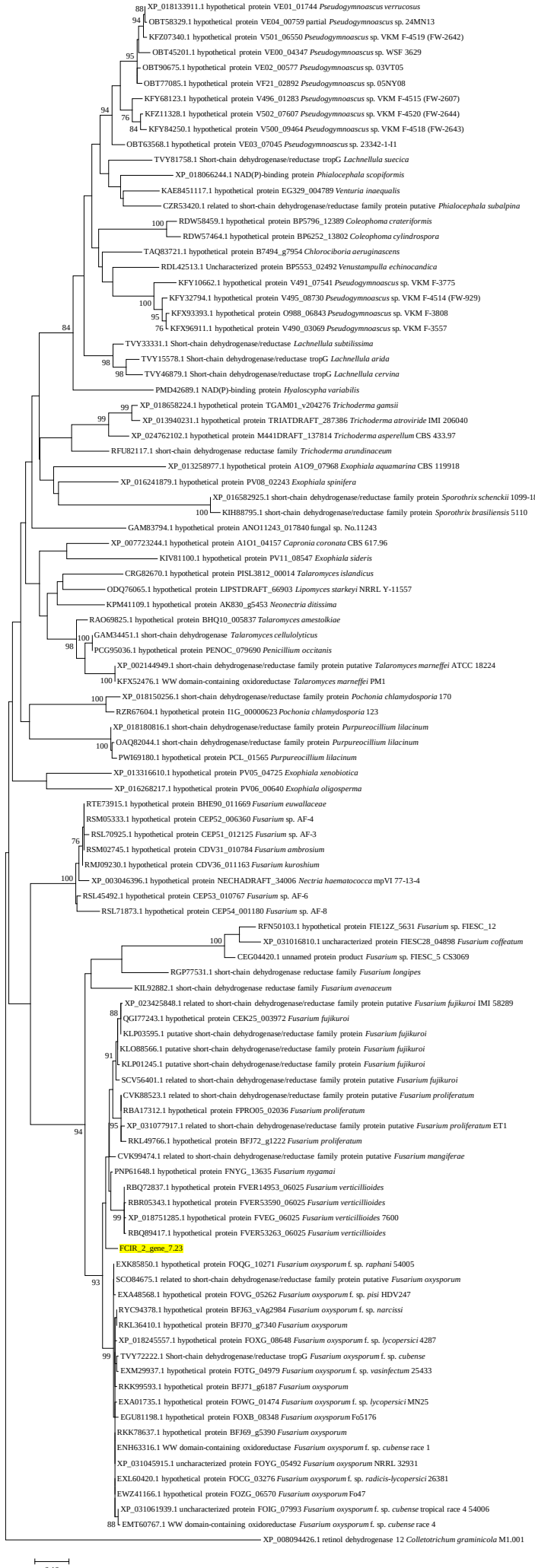

XP_018133911.1 hypothetical protein VE01_01744 Pseudogymnoascus verrucosus
OBT58329.1 hypothetical protein VE04_00759 partial Pseudogymnoascus sp. 24MN13
KFZ07340.1 hypothetical protein V501_06550 Pseudogymnoascus sp. VKM F-4519 (FW-2642)
OBT45201.1 hypothetical protein VE00_04347 Pseudogymnoascus sp. WSF 3629
OBT90675.1 hypothetical protein VE02_00577 Pseudogymnoascus sp. 03VT05
OBT77085.1 hypothetical protein VF21_02892 Pseudogymnoascus sp. 05NY08
KFY68123.1 hypothetical protein V496_01283 Pseudogymnoascus sp. VKM F-4515 (FW-2607)
KFZ11328.1 hypothetical protein V502_07607 Pseudogymnoascus sp. VKM F-4520 (FW-2644)
KFY84250.1 hypothetical protein V500_09464 Pseudogymnoascus sp. VKM F-4518 (FW-2643)
OBT63568.1 hypothetical protein VE03_07045 Pseudogymnoascus sp. 23342-1-I1
TVY81758.1 Short-chain dehydrogenase/reductase tropG Lachnellula suecica
XP_018066244.1 NAD(P)-binding protein Phialocephala scopiformis
KAE8451117.1 hypothetical protein EG329_004789 Venturia inaequalis
CZR53420.1 related to short-chain dehydrogenase/reductase family protein putative Phialocephala subalpina
RDW58459.1 hypothetical protein BP5796_12389 Coleophoma crateriformis
RDW57464.1 hypothetical protein BP6252_13802 Coleophoma cylindrospora
TAQ83721.1 hypothetical protein B7494_g7954 Chlorociboria aeruginascens
RDL42513.1 Uncharacterized protein BP5553_02492 Venustampulla echinocandica
KFY10662.1 hypothetical protein V491_07541 Pseudogymnoascus sp. VKM F-3775
KFY32794.1 hypothetical protein V495_08730 Pseudogymnoascus sp. VKM F-4514 (FW-929)
KFX93393.1 hypothetical protein O988_06843 Pseudogymnoascus sp. VKM F-3808
KFX96911.1 hypothetical protein V490_03069 Pseudogymnoascus sp. VKM F-3557
TVY33331.1 Short-chain dehydrogenase/reductase Lachnellula subtilissima
TVY15578.1 Short-chain dehydrogenase/reductase tropG Lachnellula arida
TVY46879.1 Short-chain dehydrogenase/reductase tropG Lachnellula cervina
PMD42689.1 NAD(P)-binding protein Hyaloscypha variabilis
XP_018658224.1 hypothetical protein TGAM01_v204276 Trichoderma gamsii
XP_013940231.1 hypothetical protein TRIATDRAFT_287386 Trichoderma atroviride IMI 206040
XP_024762102.1 hypothetical protein M441DRAFT_137814 Trichoderma asperellum CBS 433.97
RFU82117.1 short-chain dehydrogenase reductase family Trichoderma arundinaceum
XP_013258977.1 hypothetical protein A1O9_07968 Exophiala aquamarina CBS 119918
XP_016241879.1 hypothetical protein PV08_02243 Exophiala spinifera
XP_016582925.1 short-chain dehydrogenase/reductase family protein Sporothrix schenckii 1099-18
KIH88795.1 short-chain dehydrogenase/reductase family protein Sporothrix brasiliensis 5110
GAM83794.1 hypothetical protein ANO11243_017840 fungal sp. No.11243
XP_007723244.1 hypothetical protein A1O1_04157 Capronia coronata CBS 617.96
KIV81100.1 hypothetical protein PV11_08547 Exophiala sideris
CRG82670.1 hypothetical protein PISL3812_00014 Talaromyces islandicus
ODQ76065.1 hypothetical protein LIPSTDRAFT_66903 Lipomyces starkeyi NRRL Y-11557
KPM41109.1 hypothetical protein AK830_g5453 Neonectria ditissima
RAO69825.1 hypothetical protein BHQ10_005837 Talaromyces amestolkiae
GAM34451.1 short-chain dehydrogenase Talaromyces cellulolyticus
PCG95036.1 hypothetical protein PENOC_079690 Penicillium occitanis
XP_002144949.1 short-chain dehydrogenase/reductase family protein putative Talaromyces marneffei ATCC 18224
KFX52476.1 WW domain-containing oxidoreductase Talaromyces marneffei PM1
XP_018150256.1 short-chain dehydrogenase/reductase family protein Pochonia chlamydosporia 170
RZR67604.1 hypothetical protein I1G_00000623 Pochonia chlamydosporia 123
XP_018180816.1 short-chain dehydrogenase/reductase family protein Purpureocillium lilacinum
OAQ82044.1 short-chain dehydrogenase/reductase family protein Purpureocillium lilacinum
PWI69180.1 hypothetical protein PCL_01565 Purpureocillium lilacinum
XP_013316610.1 hypothetical protein PV05_04725 Exophiala xenobiotica
XP_016268217.1 hypothetical protein PV06_00640 Exophiala oligosperma
RTE73915.1 hypothetical protein BHE90_011669 Fusarium euwallaceae
RSM05333.1 hypothetical protein CEP52_006360 Fusarium sp. AF-4
RSL70925.1 hypothetical protein CEP51_012125 Fusarium sp. AF-3
RSM02745.1 hypothetical protein CDV31_010784 Fusarium ambrosium
RMJ09230.1 hypothetical protein CDV36_011163 Fusarium kuroshium
XP_003046396.1 hypothetical protein NECHADRAFT_34006 Nectria haematococca mpVI 77-13-4
RSL45492.1 hypothetical protein CEP53_010767 Fusarium sp. AF-6
RSL71873.1 hypothetical protein CEP54_001180 Fusarium sp. AF-8
RFN50103.1 hypothetical protein FIE12Z_5631 Fusarium sp. FIESC_12
XP_031016810.1 uncharacterized protein FIESC28_04898 Fusarium coffeatum
CEG04420.1 unnamed protein product Fusarium sp. FIESC_5 CS3069
RGP77531.1 short-chain dehydrogenase reductase family Fusarium longipes
KIL92882.1 short-chain dehydrogenase reductase family Fusarium avenaceum
XP_023425848.1 related to short-chain dehydrogenase/reductase family protein putative Fusarium fujikuroi IMI 58289
QGI77243.1 hypothetical protein CEK25_003972 Fusarium fujikuroi
KLP03595.1 putative short-chain dehydrogenase/reductase family protein Fusarium fujikuroi
KLO88566.1 putative short-chain dehydrogenase/reductase family protein Fusarium fujikuroi
KLP01245.1 putative short-chain dehydrogenase/reductase family protein Fusarium fujikuroi
SCV56401.1 related to short-chain dehydrogenase/reductase family protein putative Fusarium fujikuroi
CVK88523.1 related to short-chain dehydrogenase/reductase family protein putative Fusarium proliferatum
RBA17312.1 hypothetical protein FPRO05_02036 Fusarium proliferatum
XP_031077917.1 related to short-chain dehydrogenase/reductase family protein putative Fusarium proliferatum ET1
RKL49766.1 hypothetical protein BFJ72_g1222 Fusarium proliferatum
CVK99474.1 related to short-chain dehydrogenase/reductase family protein putative Fusarium mangiferae
PNP61648.1 hypothetical protein FNYG_13635 Fusarium nygamai
RBQ72837.1 hypothetical protein FVER14953_06025 Fusarium verticillioides
RBR05343.1 hypothetical protein FVER53590_06025 Fusarium verticillioides
XP_018751285.1 hypothetical protein FVEG_06025 Fusarium verticillioides 7600
RBQ89417.1 hypothetical protein FVER53263_06025 Fusarium verticillioides
FCIR_2_gene_7.23
EXK85850.1 hypothetical protein FOQG_10271 Fusarium oxysporum f. sp. raphani 54005
SCO84675.1 related to short-chain dehydrogenase/reductase family protein putative Fusarium oxysporum
EXA48568.1 hypothetical protein FOVG_05262 Fusarium oxysporum f. sp. pisi HDV247
RYC94378.1 hypothetical protein BFJ63_vAg2984 Fusarium oxysporum f. sp. narcissi
RKL36410.1 hypothetical protein BFJ70_g7340 Fusarium oxysporum
XP_018245557.1 hypothetical protein FOXG_08648 Fusarium oxysporum f. sp. lycopersici 4287
TVY72222.1 Short-chain dehydrogenase/reductase tropG Fusarium oxysporum f. sp. cubense
EXM29937.1 hypothetical protein FOTG_04979 Fusarium oxysporum f. sp. vasinfectum 25433
RKK99593.1 hypothetical protein BFJ71_g6187 Fusarium oxysporum
EXA01735.1 hypothetical protein FOWG_01474 Fusarium oxysporum f. sp. lycopersici MN25
EGU81198.1 hypothetical protein FOXB_08348 Fusarium oxysporum Fo5176
RKK78637.1 hypothetical protein BFJ69_g5390 Fusarium oxysporum
ENH63316.1 WW domain-containing oxidoreductase Fusarium oxysporum f. sp. cubense race 1
XP_031045915.1 uncharacterized protein FOYG_05492 Fusarium oxysporum NRRL 32931
EXL60420.1 hypothetical protein FOCG_03276 Fusarium oxysporum f. sp. radicis-lycopersici 26381
EWZ41166.1 hypothetical protein FOZG_06570 Fusarium oxysporum Fo47
XP_031061939.1 uncharacterized protein FOIG_07993 Fusarium oxysporum f. sp. cubense tropical race 4 54006
EMT60767.1 WW domain-containing oxidoreductase Fusarium oxysporum f. sp. cubense race 4
XP_008094426.1 retinol dehydrogenase 12 Colletotrichum graminicola M1.001
0.10

Figure 1. A representative phylogenetic tree from the group 1 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches ($\geq 75\%$). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).

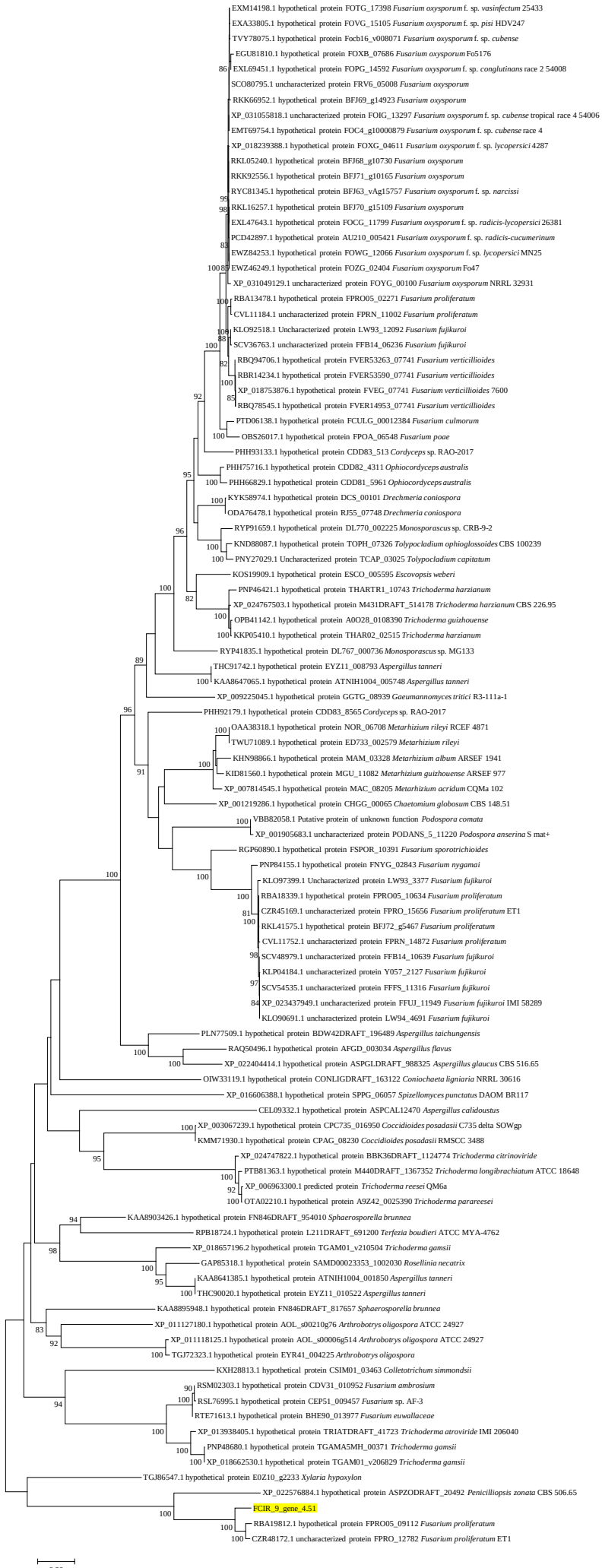

EXM14198.1 hypothetical protein FOTG_17398 Fusarium oxysporum f. sp. vasinfectum 25433
EXA33805.1 hypothetical protein FOVG_15105 Fusarium oxysporum f. sp. pisi HDV247
TVY78075.1 hypothetical protein Focb16_v008071 Fusarium oxysporum f. sp. cubense
EGU81810.1 hypothetical protein FOXB_07686 Fusarium oxysporum Fo5176
EXL69451.1 hypothetical protein FOPG_14592 Fusarium oxysporum f. sp. conglutinans race 2 54008
SCO80795.1 uncharacterized protein FRV6_05008 Fusarium oxysporum
RKK66952.1 hypothetical protein BFJ69_g14923 Fusarium oxysporum
XP_031055818.1 uncharacterized protein FOIG_13297 Fusarium oxysporum f. sp. cubense tropical race 4 54006
EMT69754.1 hypothetical protein FOC4_g10000879 Fusarium oxysporum f. sp. cubense race 4
XP_018239388.1 hypothetical protein FOXG_04611 Fusarium oxysporum f. sp. lycopersici 4287
RKL05240.1 hypothetical protein BFJ68_g10730 Fusarium oxysporum
RKK92556.1 hypothetical protein BFJ71_g10165 Fusarium oxysporum
RYC81345.1 hypothetical protein BFJ63_vAg15757 Fusarium oxysporum f. sp. narcissi
RKL16257.1 hypothetical protein BFJ70_g15109 Fusarium oxysporum
EXL47643.1 hypothetical protein FOCG_11799 Fusarium oxysporum f. sp. radicis-lycopersici 26381
PCD42897.1 hypothetical protein AU210_005421 Fusarium oxysporum f. sp. radicis-cucumerinum
EWZ84253.1 hypothetical protein FOWG_12066 Fusarium oxysporum f. sp. lycopersici MN25
EWZ46249.1 hypothetical protein FOZG_02404 Fusarium oxysporum Fo47
XP_031049129.1 uncharacterized protein FOYG_00100 Fusarium oxysporum NRRL 32931
RBA13478.1 hypothetical protein FPRO05_02271 Fusarium proliferatum
CVL11184.1 uncharacterized protein FPRN_11002 Fusarium proliferatum
KLO92518.1 Uncharacterized protein LW93_12092 Fusarium fujikuroi
SCV36763.1 uncharacterized protein FFB14_06236 Fusarium fujikuroi
RBQ94706.1 hypothetical protein FVER53263_07741 Fusarium verticillioides
RBR14234.1 hypothetical protein FVER53590_07741 Fusarium verticillioides
XP_018753876.1 hypothetical protein FVEG_07741 Fusarium verticillioides 7600
RBQ78545.1 hypothetical protein FVER14953_07741 Fusarium verticillioides
PTD06138.1 hypothetical protein FCULG_00012384 Fusarium culmorum
OBS26017.1 hypothetical protein FPOA_06548 Fusarium poae
PHH93133.1 hypothetical protein CDD83_513 Cordyceps sp. RAO-2017
PHH75716.1 hypothetical protein CDD82_4311 Ophiocordyceps australis
PHH66829.1 hypothetical protein CDD81_5961 Ophiocordyceps australis
KYK58974.1 hypothetical protein DCS_00101 Drechmeria coniospora
ODA76478.1 hypothetical protein RJ55_07748 Drechmeria coniospora
RYP91659.1 hypothetical protein DL770_002225 Monosporascus sp. CRB-9-2
KND88087.1 hypothetical protein TOPH_07326 Tolypocladium ophioglossoides CBS 100239
PNY27029.1 Uncharacterized protein TCAP_03025 Tolypocladium capitatum
KOS19909.1 hypothetical protein ESCO_005595 Escovopsis weberi
PNP46421.1 hypothetical protein THARTR1_10743 Trichoderma harzianum
XP_024767503.1 hypothetical protein M431DRAFT_514178 Trichoderma harzianum CBS 226.95
OPB41142.1 hypothetical protein A0O28_0108390 Trichoderma guizhouense
KKP05410.1 hypothetical protein THAR02_02515 Trichoderma harzianum
RYP41835.1 hypothetical protein DL767_000736 Monosporascus sp. MG133
THC91742.1 hypothetical protein EYZ11_008793 Aspergillus tanneri
KAA8647065.1 hypothetical protein ATNIH1004_005748 Aspergillus tanneri
XP_009225045.1 hypothetical protein GGTG_08939 Gaeumannomyces tritici R3-111a-1
PHH92179.1 hypothetical protein CDD83_8565 Cordyceps sp. RAO-2017
OAA38318.1 hypothetical protein NOR_06708 Metarhizium rileyi RCEF 4871
TWU71089.1 hypothetical protein ED733_002579 Metarhizium rileyi
KHN98866.1 hypothetical protein MAM_03328 Metarhizium album ARSEF 1941
KID81560.1 hypothetical protein MGU_11082 Metarhizium guizhouense ARSEF 977
XP_007814545.1 hypothetical protein MAC_08205 Metarhizium acridum CQMa 102
XP_001219286.1 hypothetical protein CHGG_00065 Chaetomium globosum CBS 148.51
VBB82058.1 Putative protein of unknown function Podospora comata
XP_001905683.1 uncharacterized protein PODANS_5_11220 Podospora anserina S mat+
RGP60890.1 hypothetical protein FSPOR_10391 Fusarium sporotrichioides
PNP84155.1 hypothetical protein FNYG_02843 Fusarium nygamai
KLO97399.1 Uncharacterized protein LW93_3377 Fusarium fujikuroi
RBA18339.1 hypothetical protein FPRO05_10634 Fusarium proliferatum
CZR45169.1 uncharacterized protein FPRO_15656 Fusarium proliferatum ET1
RKL41575.1 hypothetical protein BFJ72_g5467 Fusarium proliferatum
CVL11752.1 uncharacterized protein FPRN_14872 Fusarium proliferatum
SCV48979.1 uncharacterized protein FFB14_10639 Fusarium fujikuroi
KLP04184.1 uncharacterized protein Y057_2127 Fusarium fujikuroi
SCV54535.1 uncharacterized protein FFFS_11316 Fusarium fujikuroi
XP_023437949.1 uncharacterized protein FFUJ_11949 Fusarium fujikuroi IMI 58289
KLO90691.1 uncharacterized protein LW94_4691 Fusarium fujikuroi
PLN77509.1 hypothetical protein BDW42DRAFT_196489 Aspergillus taichungensis
RAQ50496.1 hypothetical protein AFGD_003034 Aspergillus flavus
XP_022404414.1 hypothetical protein ASPGLDRAFT_988325 Aspergillus glaucus CBS 516.65
OIW33119.1 hypothetical protein CONLIGDRAFT_163122 Coniochaeta ligniaria NRRL 30616
XP_016606388.1 hypothetical protein SPPG_06057 Spizellomyces punctatus DAOM BR117
CEL09332.1 hypothetical protein ASPCAL12470 Aspergillus calidoustus
XP_003067239.1 hypothetical protein CPC735_016950 Coccidioides posadasii C735 delta SOWgp
KMM71930.1 hypothetical protein CPAG_08230 Coccidioides posadasii RMSCC 3488
XP_024747822.1 hypothetical protein BBK36DRAFT_1124774 Trichoderma citrinoviride
PTB81363.1 hypothetical protein M440DRAFT_1367352 Trichoderma longibrachiatum ATCC 18648
XP_006963300.1 predicted protein Trichoderma reesei QM6a
OTA02210.1 hypothetical protein A9Z42_0025390 Trichoderma parareesei
KAA8903426.1 hypothetical protein FN846DRAFT_954010 Sphaerosporella brunnea
RPB18724.1 hypothetical protein L211DRAFT_691200 Terfezia boudieri ATCC MYA-4762
XP_018657196.2 hypothetical protein TGAM01_v210504 Trichoderma gamsii
GAP85318.1 hypothetical protein SAMD00023353_1002030 Rosellinia necatrix
KAA8641385.1 hypothetical protein ATNIH1004_001850 Aspergillus tanneri
THC90020.1 hypothetical protein EYZ11_010522 Aspergillus tanneri
KAA8895948.1 hypothetical protein FN846DRAFT_817657 Sphaerosporella brunnea
XP_011127180.1 hypothetical protein AOL_s00210g76 Arthrobotrys oligospora ATCC 24927
XP_011118125.1 hypothetical protein AOL_s00006g514 Arthrobotrys oligospora ATCC 24927
TGJ72323.1 hypothetical protein EYR41_004225 Arthrobotrys oligospora
KXH28813.1 hypothetical protein CSIM01_03463 Colletotrichum simmondsii
RSM02303.1 hypothetical protein CDV31_010952 Fusarium ambrosium
RSL76995.1 hypothetical protein CEP51_009457 Fusarium sp. AF-3
RTE71613.1 hypothetical protein BHE90_013977 Fusarium euwallaceae
XP_013938405.1 hypothetical protein TRIATDRAFT_41723 Trichoderma atroviride IMI 206040
PNP48680.1 hypothetical protein TGAMA5MH_00371 Trichoderma gamsii
XP_018662530.1 hypothetical protein TGAM01_v206829 Trichoderma gamsii
TGJ86547.1 hypothetical protein E0Z10_g2233 Xylaria hypoxylon
XP_022576884.1 hypothetical protein ASPZODRAFT_20492 Penicilliopsis zonata CBS 506.65
FCIR_9_gene_4.51
RBA19812.1 hypothetical protein FPRO05_09112 Fusarium proliferatum
CZR48172.1 uncharacterized protein FPRO_12782 Fusarium proliferatum ET1
0.20

Figure 2. A representative phylogenetic tree from the group 2 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The rate variation model allowed for some sites to be evolutionarily invariable. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches ($\geq$ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).

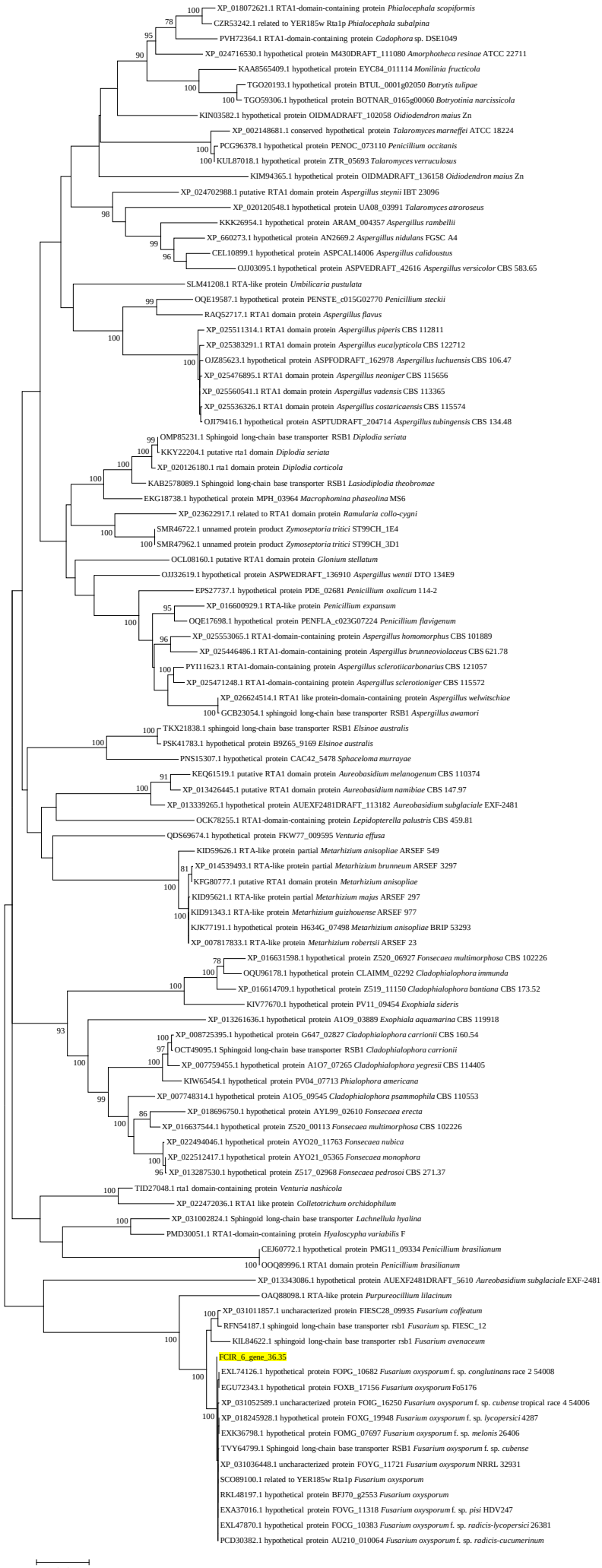

XP_018072621.1 RTA1-domain-containing protein Phialocephala scopiformis
CZR53242.1 related to YER185w Rta1p Phialocephala subalpina
PVH72364.1 RTA1-domain-containing protein Cadophora sp. DSE1049
XP_024716530.1 hypothetical protein M430DRAFT_111080 Amorphotheca resinae ATCC 22711
KAA8565409.1 hypothetical protein EYC84_011114 Monilinia fructicola
TGO20193.1 hypothetical protein BTUL_0001g02050 Botrytis tulipae
TGO59306.1 hypothetical protein BOTNAR_0165g00060 Botryotinia narcissicola
KIN03582.1 hypothetical protein OIDMADRAFT_102058 Oidiodendron maius Zn
XP_002148681.1 conserved hypothetical protein Talaromyces marneffei ATCC 18224
PCG96378.1 hypothetical protein PENOC_073110 Penicillium occitanis
KUL87018.1 hypothetical protein ZTR_05693 Talaromyces verruculosus
KIM94365.1 hypothetical protein OIDMADRAFT_136158 Oidiodendron maius Zn
XP_024702988.1 putative RTA1 domain protein Aspergillus steynii IBT 23096
XP_020120548.1 hypothetical protein UA08_03991 Talaromyces atroroseus
KKK26954.1 hypothetical protein ARAM_004357 Aspergillus rambellii
XP_660273.1 hypothetical protein AN2669.2 Aspergillus nidulans FGSC A4
CEL10899.1 hypothetical protein ASPCAL14006 Aspergillus calidoustus
OJJ03095.1 hypothetical protein ASPVEDRAFT_42616 Aspergillus versicolor CBS 583.65
SLM41208.1 RTA-like protein Umbilicaria pustulata
OQE19587.1 hypothetical protein PENSTE_c015G02770 Penicillium steckii
RAQ52717.1 RTA1 domain protein Aspergillus flavus
XP_025511314.1 RTA1 domain protein Aspergillus piperis CBS 112811
XP_025383291.1 RTA1 domain protein Aspergillus eucalypticola CBS 122712
OJZ85623.1 hypothetical protein ASPFODRAFT_162978 Aspergillus luchuensis CBS 106.47
XP_025476895.1 RTA1 domain protein Aspergillus neoniger CBS 115656
XP_025560541.1 RTA1 domain protein Aspergillus vadensis CBS 113365
XP_025536326.1 RTA1 domain protein Aspergillus costaricaensis CBS 115574
OJI79416.1 hypothetical protein ASPTUDRAFT_204714 Aspergillus tubingensis CBS 134.48
OMP85231.1 Sphingoid long-chain base transporter RSB1 Diplodia seriata
KKY22204.1 putative rta1 domain Diplodia seriata
XP_020126180.1 rta1 domain protein Diplodia corticola
KAB2578089.1 Sphingoid long-chain base transporter RSB1 Lasiodiplodia theobromae
EKG18738.1 hypothetical protein MPH_03964 Macrophomina phaseolina MS6
XP_023622917.1 related to RTA1 domain protein Ramularia collo-cygni
SMR46722.1 unnamed protein product Zymoseptoria tritici ST99CH_1E4
SMR47962.1 unnamed protein product Zymoseptoria tritici ST99CH_3D1
OCL08160.1 putative RTA1 domain protein Glonium stellatum
OJJ32619.1 hypothetical protein ASPWEDRAFT_136910 Aspergillus wentii DTO 134E9
EPS27737.1 hypothetical protein PDE_02681 Penicillium oxalicum 114-2
XP_016600929.1 RTA-like protein Penicillium expansum
OQE17698.1 hypothetical protein PENFLA_c023G07224 Penicillium flavigenum
XP_025553065.1 RTA1-domain-containing protein Aspergillus homomorphus CBS 101889
XP_025446486.1 RTA1-domain-containing protein Aspergillus brunneoviolaceus CBS 621.78
PYI11623.1 RTA1-domain-containing protein Aspergillus sclerotiicarbonarius CBS 121057
XP_025471248.1 RTA1-domain-containing protein Aspergillus sclerotioniger CBS 115572
XP_026624514.1 RTA1 like protein-domain-containing protein Aspergillus welwitschiae
GCB23054.1 sphingoid long-chain base transporter RSB1 Aspergillus awamori
TKX21838.1 sphingoid long-chain base transporter RSB1 Elsinoe australis
PSK41783.1 hypothetical protein B9Z65_9169 Elsinoe australis
PNS15307.1 hypothetical protein CAC42_5478 Sphaceloma murrayae
KEQ61519.1 putative RTA1 domain protein Aureobasidium melanogenum CBS 110374
XP_013426445.1 putative RTA1 domain protein Aureobasidium namibiae CBS 147.97
XP_013339265.1 hypothetical protein AUEXF2481DRAFT_113182 Aureobasidium subglaciale EXF-2481
OCK78255.1 RTA1-domain-containing protein Lepidopterella palustris CBS 459.81
QDS69674.1 hypothetical protein FKW77_009595 Venturia effusa
KID59626.1 RTA-like protein partial Metarhizium anisopliae ARSEF 549
XP_014539493.1 RTA-like protein partial Metarhizium brunneum ARSEF 3297
KFG80777.1 putative RTA1 domain protein Metarhizium anisopliae
KID95621.1 RTA-like protein partial Metarhizium majus ARSEF 297
KID91343.1 RTA-like protein Metarhizium guizhouense ARSEF 977
KJK77191.1 hypothetical protein H634G_07498 Metarhizium anisopliae BRIP 53293
XP_007817833.1 RTA-like protein Metarhizium robertsii ARSEF 23
XP_016631598.1 hypothetical protein Z520_06927 Fonsecaea multimorphosa CBS 102226
OQU96178.1 hypothetical protein CLAIMM_02292 Cladophialophora immunda
XP_016614709.1 hypothetical protein Z519_11150 Cladophialophora bantiana CBS 173.52
KIV77670.1 hypothetical protein PV11_09454 Exophiala sideris
XP_013261636.1 hypothetical protein A1O9_03889 Exophiala aquamarina CBS 119918
XP_008725395.1 hypothetical protein G647_02827 Cladophialophora carrionii CBS 160.54
OCT49095.1 Sphingoid long-chain base transporter RSB1 Cladophialophora carrionii
XP_007759455.1 hypothetical protein A1O7_07265 Cladophialophora yegresii CBS 114405
KIW65454.1 hypothetical protein PV04_07713 Phialophora americana
XP_007748314.1 hypothetical protein A1O5_09545 Cladophialophora psammophila CBS 110553
XP_018696750.1 hypothetical protein AYL99_02610 Fonsecaea erecta
XP_016637544.1 hypothetical protein Z520_00113 Fonsecaea multimorphosa CBS 102226
XP_022494046.1 hypothetical protein AYO20_11763 Fonsecaea nubica
XP_022512417.1 hypothetical protein AYO21_05365 Fonsecaea monophora
XP_013287530.1 hypothetical protein Z517_02968 Fonsecaea pedrosoi CBS 271.37
TID27048.1 rta1 domain-containing protein Venturia nashicola
XP_022472036.1 RTA1 like protein Colletotrichum orchidophilum
XP_031002824.1 Sphingoid long-chain base transporter Lachnellula hyalina
PMD30051.1 RTA1-domain-containing protein Hyaloscypha variabilis F
CEJ60772.1 hypothetical protein PMG11_09334 Penicillium brasilianum
OOQ89996.1 RTA1 domain protein Penicillium brasilianum
XP_013343086.1 hypothetical protein AUEXF2481DRAFT_5610 Aureobasidium subglaciale EXF-2481
OAQ88098.1 RTA-like protein Purpureocillium lilacinum
XP_031011857.1 uncharacterized protein FIESC28_09935 Fusarium coffeatum
RFN54187.1 sphingoid long-chain base transporter rsb1 Fusarium sp. FIESC_12
KIL84622.1 sphingoid long-chain base transporter rsb1 Fusarium avenaceum
FCIR_6_gene_36.35
EXL74126.1 hypothetical protein FOPG_10682 Fusarium oxysporum f. sp. conglutinans race 2 54008
EGU72343.1 hypothetical protein FOXB_17156 Fusarium oxysporum Fo5176
XP_031052589.1 uncharacterized protein FOIG_16250 Fusarium oxysporum f. sp. cubense tropical race 4 54006
XP_018245928.1 hypothetical protein FOXG_19948 Fusarium oxysporum f. sp. lycopersici 4287
EXK36798.1 hypothetical protein FOMG_07697 Fusarium oxysporum f. sp. melonis 26406
TVY64799.1 Sphingoid long-chain base transporter RSB1 Fusarium oxysporum f. sp. cubense
XP_031036448.1 uncharacterized protein FOYG_11721 Fusarium oxysporum NRRL 32931
SCO89100.1 related to YER185w Rta1p Fusarium oxysporum
RKL48197.1 hypothetical protein BFJ70_g2553 Fusarium oxysporum
EXA37016.1 hypothetical protein FOVG_11318 Fusarium oxysporum f. sp. pisi HDV247
EXL47870.1 hypothetical protein FOCG_10383 Fusarium oxysporum f. sp. radicis-lycopersici 26381
PCD30382.1 hypothetical protein AU210_010064 Fusarium oxysporum f. sp. radicis-cucumerinum
0.20

Figure 3. A representative phylogenetic tree from the group 3 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The rate variation model allowed for some sites to be evolutionarily invariable. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches ($\geq 75\%$). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).

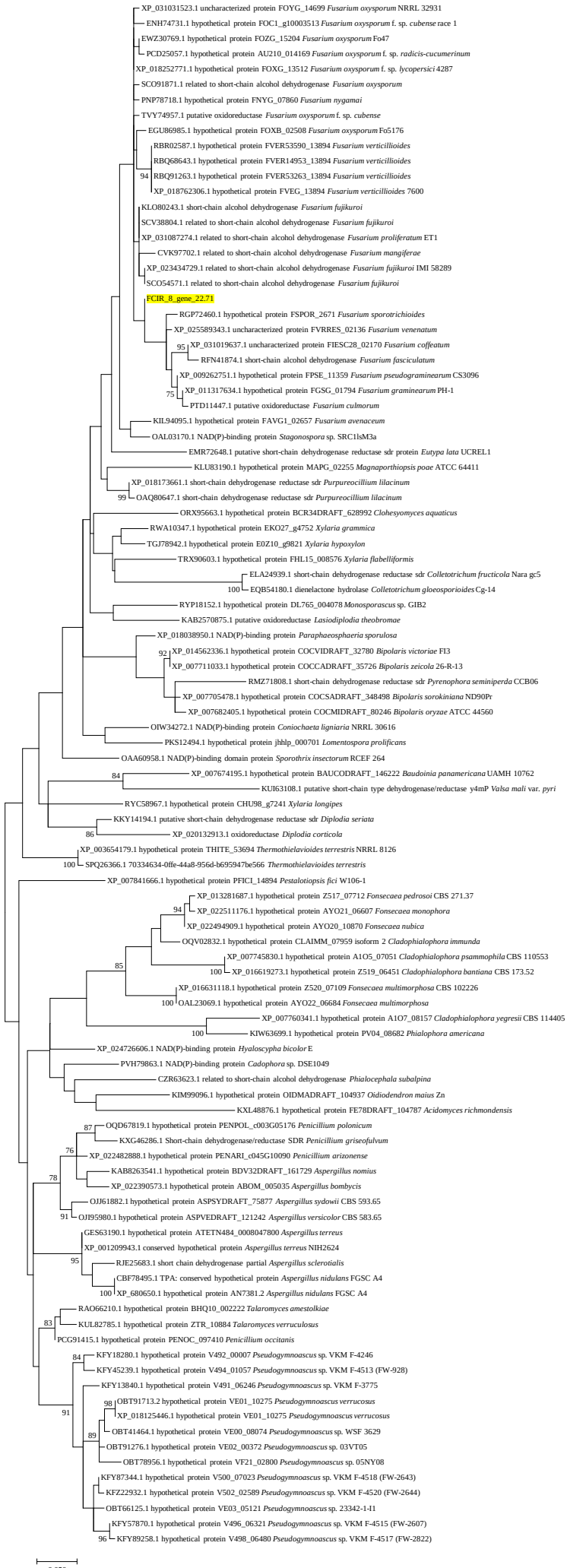

XP_031031523.1 uncharacterized protein FOYG_14699 Fusarium oxysporum NRRL 32931
ENH74731.1 hypothetical protein FOC1_g10003513 Fusarium oxysporum f. sp. cubense race 1
EWZ30769.1 hypothetical protein FOZG_15204 Fusarium oxysporum Fo47
PCD25057.1 hypothetical protein AU210_014169 Fusarium oxysporum f. sp. radicis-cucumerinum
XP_018252771.1 hypothetical protein FOXG_13512 Fusarium oxysporum f. sp. lycopersici 4287
SCO91871.1 related to short-chain alcohol dehydrogenase Fusarium oxysporum
PNP78718.1 hypothetical protein FNYG_07860 Fusarium nygamai
TVY74957.1 putative oxidoreductase Fusarium oxysporum f. sp. cubense
EGU86985.1 hypothetical protein FOXB_02508 Fusarium oxysporum Fo5176
RBR02587.1 hypothetical protein FVER53590_13894 Fusarium verticillioides
RBQ68643.1 hypothetical protein FVER14953_13894 Fusarium verticillioides
94
RBQ91263.1 hypothetical protein FVER53263_13894 Fusarium verticillioides
XP_018762306.1 hypothetical protein FVEG_13894 Fusarium verticillioides 7600
KLO80243.1 short-chain alcohol dehydrogenase Fusarium fujikuroi
SCV38804.1 related to short-chain alcohol dehydrogenase Fusarium fujikuroi
XP_031087274.1 related to short-chain alcohol dehydrogenase Fusarium proliferatum ET1
CVK97702.1 related to short-chain alcohol dehydrogenase Fusarium mangiferae
XP_023434729.1 related to short-chain alcohol dehydrogenase Fusarium fujikuroi IMI 58289
SCO54571.1 related to short-chain alcohol dehydrogenase Fusarium fujikuroi
FCIR_8_gene_22.71
RGP72460.1 hypothetical protein FSPOR_2671 Fusarium sporotrichioides
XP_025589343.1 uncharacterized protein FVRRES_02136 Fusarium venenatum
95
XP_031019637.1 uncharacterized protein FIESC28_02170 Fusarium coffeatum
RFN41874.1 short-chain alcohol dehydrogenase Fusarium fasciculatum
XP_009262751.1 hypothetical protein FPSE_11359 Fusarium pseudograminearum CS3096
75
XP_011317634.1 hypothetical protein FGSG_01794 Fusarium graminearum PH-1
PTD11447.1 putative oxidoreductase Fusarium culmorum
KIL94095.1 hypothetical protein FAVG1_02657 Fusarium avenaceum
OAL03170.1 NAD(P)-binding protein Stagonospora sp. SRC1lsM3a
EMR72648.1 putative short-chain dehydrogenase reductase sdr protein Eutypa lata UCREL1
KLU83190.1 hypothetical protein MAPG_02255 Magnaporthiopsis poae ATCC 64411
XP_018173661.1 short-chain dehydrogenase reductase sdr Purpureocillium lilacinum
99
OAQ80647.1 short-chain dehydrogenase reductase sdr Purpureocillium lilacinum
ORX95663.1 hypothetical protein BCR34DRAFT_628992 Clohesyomyces aquaticus
RWA10347.1 hypothetical protein EKO27_g4752 Xylaria grammica
TGJ78942.1 hypothetical protein E0Z10_g9821 Xylaria hypoxylon
TRX90603.1 hypothetical protein FHL15_008576 Xylaria flabelliformis
ELA24939.1 short-chain dehydrogenase reductase sdr Colletotrichum fructicola Nara gc5
100
EQB54180.1 dienelactone hydrolase Colletotrichum gloeosporioides Cg-14
RYP18152.1 hypothetical protein DL765_004078 Monosporascus sp. GIB2
KAB2570875.1 putative oxidoreductase Lasiodiplodia theobromae
XP_018038950.1 NAD(P)-binding protein Paraphaeosphaeria sporulosa
92
XP_014562336.1 hypothetical protein COCVIDRAFT_32780 Bipolaris victoriae FI3
XP_007711033.1 hypothetical protein COCCADRAFT_35726 Bipolaris zeicola 26-R-13
RMZ71808.1 short-chain dehydrogenase reductase sdr Pyrenophora seminiperda CCB06
XP_007705478.1 hypothetical protein COCSADRAFT_348498 Bipolaris sorokiniana ND90Pr
XP_007682405.1 hypothetical protein COCMIDRAFT_80246 Bipolaris oryzae ATCC 44560
OIW34272.1 NAD(P)-binding protein Coniochaeta ligniaria NRRL 30616
PKS12494.1 hypothetical protein jhhlp_000701 Lomentospora prolificans
OAA60958.1 NAD(P)-binding domain protein Sporothrix insectorum RCEF 264
84
XP_007674195.1 hypothetical protein BAUCODRAFT_146222 Baudoinia panamericana UAMH 10762
KUI63108.1 putative short-chain type dehydrogenase/reductase y4mP Valsa mali var. pyri
RYC58967.1 hypothetical protein CHU98_g7241 Xylaria longipes
KKY14194.1 putative short-chain dehydrogenase reductase sdr Diplodia seriata
86
XP_020132913.1 oxidoreductase Diplodia corticola
XP_003654179.1 hypothetical protein THITE_53694 Thermothielavioides terrestris NRRL 8126
100
SPQ26366.1 70334634-0ffe-44a8-956d-b695947be566 Thermothielavioides terrestris
XP_007841666.1 hypothetical protein PFICI_14894 Pestalotiopsis fici W106-1
XP_013281687.1 hypothetical protein Z517_07712 Fonsecaea pedrosoi CBS 271.37
94
XP_022511176.1 hypothetical protein AYO21_06607 Fonsecaea monophora
XP_022494909.1 hypothetical protein AYO20_10870 Fonsecaea nubica
OQV02832.1 hypothetical protein CLAIMM_07959 isoform 2 Cladophialophora immunda
XP_007745830.1 hypothetical protein A1O5_07051 Cladophialophora psammophila CBS 110553
85
100
XP_016619273.1 hypothetical protein Z519_06451 Cladophialophora bantiana CBS 173.52
XP_016631118.1 hypothetical protein Z520_07109 Fonsecaea multimorphosa CBS 102226
100
OAL23069.1 hypothetical protein AYO22_06684 Fonsecaea multimorphosa
XP_007760341.1 hypothetical protein A1O7_08157 Cladophialophora yegresii CBS 114405
100
KIW63699.1 hypothetical protein PV04_08682 Phialophora americana
XP_024726606.1 NAD(P)-binding protein Hyaloscypha bicolor E
PVH79863.1 NAD(P)-binding protein Cadophora sp. DSE1049
CZR63623.1 related to short-chain alcohol dehydrogenase Phialocephala subalpina
KIM99096.1 hypothetical protein OIDMADRAFT_104937 Oidiodendron maius Zn
KXL48876.1 hypothetical protein FE78DRAFT_104787 Acidomyces richmondensis
87
OQD67819.1 hypothetical protein PENPOL_c003G05176 Penicillium polonicum
KXG46286.1 Short-chain dehydrogenase/reductase SDR Penicillium griseofulvum
76
XP_022482888.1 hypothetical protein PENARI_c045G10090 Penicillium arizonense
78
KAB8263541.1 hypothetical protein BDV32DRAFT_161729 Aspergillus nomius
XP_022390573.1 hypothetical protein ABOM_005035 Aspergillus bombycis
OJJ61882.1 hypothetical protein ASPSYDRAFT_75877 Aspergillus sydowii CBS 593.65
91
OJI95980.1 hypothetical protein ASPVEDRAFT_121242 Aspergillus versicolor CBS 583.65
GES63190.1 hypothetical protein ATETN484_0008047800 Aspergillus terreus
XP_001209943.1 conserved hypothetical protein Aspergillus terreus NIH2624
95
RJE25683.1 short chain dehydrogenase partial Aspergillus sclerotialis
CBF78495.1 TPA: conserved hypothetical protein Aspergillus nidulans FGSC A4
100
XP_680650.1 hypothetical protein AN7381.2 Aspergillus nidulans FGSC A4
RAO66210.1 hypothetical protein BHQ10_002222 Talaromyces amestolkiae
83
KUL82785.1 hypothetical protein ZTR_10884 Talaromyces verruculosus
PCG91415.1 hypothetical protein PENOC_097410 Penicillium occitanis
84
KFY18280.1 hypothetical protein V492_00007 Pseudogymnoascus sp. VKM F-4246
KFY45239.1 hypothetical protein V494_01057 Pseudogymnoascus sp. VKM F-4513 (FW-928)
KFY13840.1 hypothetical protein V491_06246 Pseudogymnoascus sp. VKM F-3775
98
OBT91713.2 hypothetical protein VE01_10275 Pseudogymnoascus verrucosus
91
XP_018125446.1 hypothetical protein VE01_10275 Pseudogymnoascus verrucosus
OBT41464.1 hypothetical protein VE00_08074 Pseudogymnoascus sp. WSF 3629
89
OBT91276.1 hypothetical protein VE02_00372 Pseudogymnoascus sp. 03VT05
OBT78956.1 hypothetical protein VF21_02800 Pseudogymnoascus sp. 05NY08
KFY87344.1 hypothetical protein V500_07023 Pseudogymnoascus sp. VKM F-4518 (FW-2643)
KFZ22932.1 hypothetical protein V502_02589 Pseudogymnoascus sp. VKM F-4520 (FW-2644)
OBT66125.1 hypothetical protein VE03_05121 Pseudogymnoascus sp. 23342-1-I1
KFY57870.1 hypothetical protein V496_06321 Pseudogymnoascus sp. VKM F-4515 (FW-2607)
96
KFY89258.1 hypothetical protein V498_06480 Pseudogymnoascus sp. VKM F-4517 (FW-2822)
0.050

Figure 4. A representative phylogenetic tree from the group 4 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones *et al.*, 1992), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches ($\geq$ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).

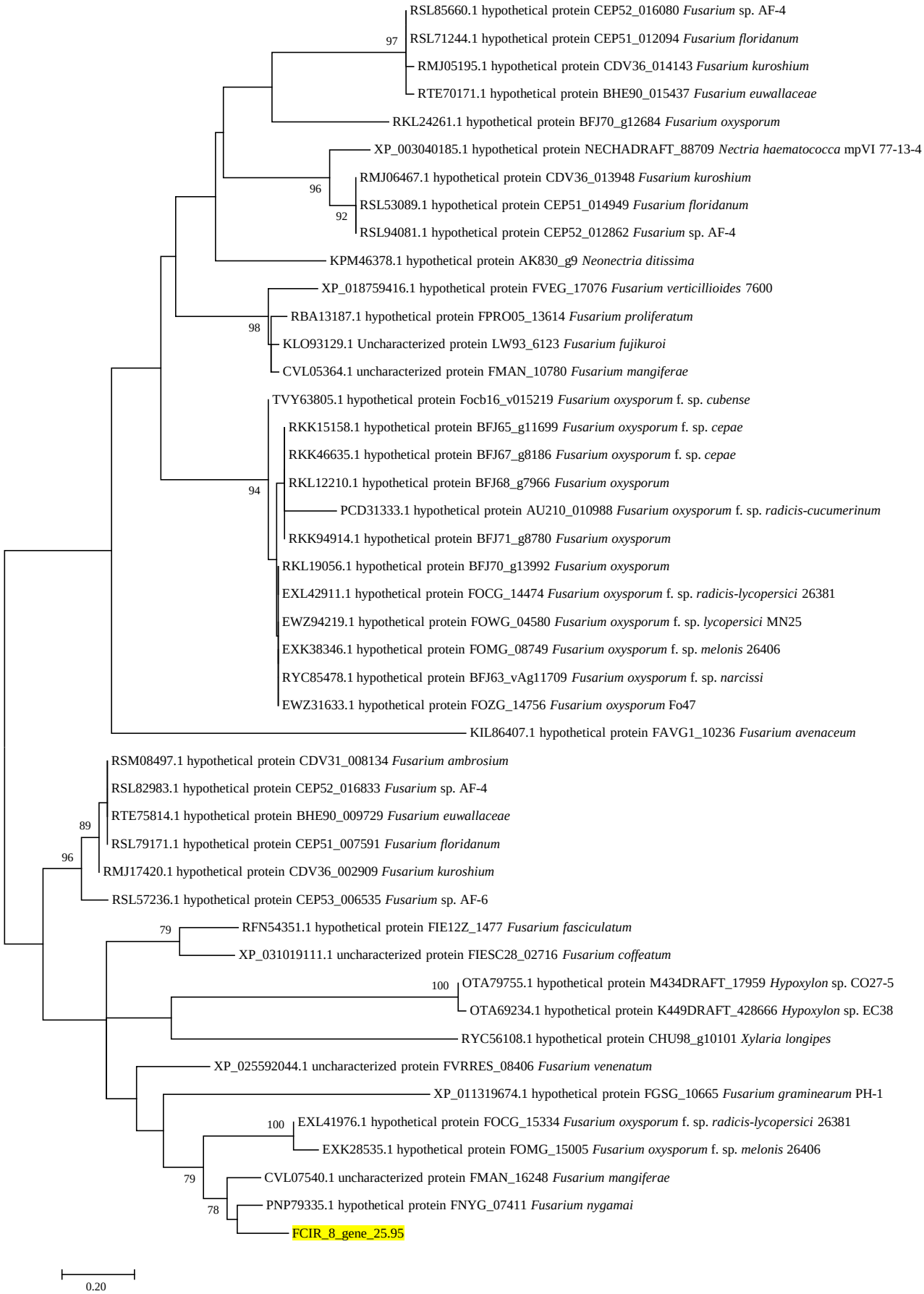

RSL85660.1 hypothetical protein CEP52_016080 Fusarium sp. AF-4
97
RSL71244.1 hypothetical protein CEP51_012094 Fusarium floridanum
RMJ05195.1 hypothetical protein CDV36_014143 Fusarium kuroshium
RTE70171.1 hypothetical protein BHE90_015437 Fusarium euwallaceae
RKL24261.1 hypothetical protein BFJ70_g12684 Fusarium oxysporum
XP_003040185.1 hypothetical protein NECHADRAFT_88709 Nectria haematococca mpVI 77-13-4
RMJ06467.1 hypothetical protein CDV36_013948 Fusarium kuroshium
96
RSL53089.1 hypothetical protein CEP51_014949 Fusarium floridanum
92
RSL94081.1 hypothetical protein CEP52_012862 Fusarium sp. AF-4
KPM46378.1 hypothetical protein AK830_g9 Neonectria ditissima
XP_018759416.1 hypothetical protein FVEG_17076 Fusarium verticillioides 7600
RBA13187.1 hypothetical protein FPRO05_13614 Fusarium proliferatum
98
KLO93129.1 Uncharacterized protein LW93_6123 Fusarium fujikuroi
CVL05364.1 uncharacterized protein FMAN_10780 Fusarium mangiferae
TVY63805.1 hypothetical protein Focb16_v015219 Fusarium oxysporum f. sp. cubense
RKK15158.1 hypothetical protein BFJ65_g11699 Fusarium oxysporum f. sp. cepae
RKK46635.1 hypothetical protein BFJ67_g8186 Fusarium oxysporum f. sp. cepae
RKL12210.1 hypothetical protein BFJ68_g7966 Fusarium oxysporum
94
PCD31333.1 hypothetical protein AU210_010988 Fusarium oxysporum f. sp. radicis-cucumerinum
RKK94914.1 hypothetical protein BFJ71_g8780 Fusarium oxysporum
RKL19056.1 hypothetical protein BFJ70_g13992 Fusarium oxysporum
EXL42911.1 hypothetical protein FOCG_14474 Fusarium oxysporum f. sp. radicis-lycopersici 26381
EWZ94219.1 hypothetical protein FOWG_04580 Fusarium oxysporum f. sp. lycopersici MN25
EXK38346.1 hypothetical protein FOMG_08749 Fusarium oxysporum f. sp. melonis 26406
RYC85478.1 hypothetical protein BFJ63_vAg11709 Fusarium oxysporum f. sp. narcissi
EWZ31633.1 hypothetical protein FOZG_14756 Fusarium oxysporum Fo47
KIL86407.1 hypothetical protein FAVG1_10236 Fusarium avenaceum
RSM08497.1 hypothetical protein CDV31_008134 Fusarium ambrosium
RSL82983.1 hypothetical protein CEP52_016833 Fusarium sp. AF-4
RTE75814.1 hypothetical protein BHE90_009729 Fusarium euwallaceae
89
RSL79171.1 hypothetical protein CEP51_007591 Fusarium floridanum
96
RMJ17420.1 hypothetical protein CDV36_002909 Fusarium kuroshium
RSL57236.1 hypothetical protein CEP53_006535 Fusarium sp. AF-6
79
RFN54351.1 hypothetical protein FIE12Z_1477 Fusarium fasciculatum
XP_031019111.1 uncharacterized protein FIESC28_02716 Fusarium coffeatum
100
OTA79755.1 hypothetical protein M434DRAFT_17959 Hypoxylon sp. CO27-5
OTA69234.1 hypothetical protein K449DRAFT_428666 Hypoxylon sp. EC38
RYC56108.1 hypothetical protein CHU98_g10101 Xylaria longipes
XP_025592044.1 uncharacterized protein FVRRES_08406 Fusarium venenatum
XP_011319674.1 hypothetical protein FGSG_10665 Fusarium graminearum PH-1
100
EXL41976.1 hypothetical protein FOCG_15334 Fusarium oxysporum f. sp. radicis-lycopersici 26381
EXK28535.1 hypothetical protein FOMG_15005 Fusarium oxysporum f. sp. melonis 26406
79
CVL07540.1 uncharacterized protein FMAN_16248 Fusarium mangiferae
78
PNP79335.1 hypothetical protein FNYG_07411 Fusarium nygamai
FCIR_8_gene_25.95
0.20

Figure 5. A representative phylogenetic tree from the group 5 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches ($\geq$ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).

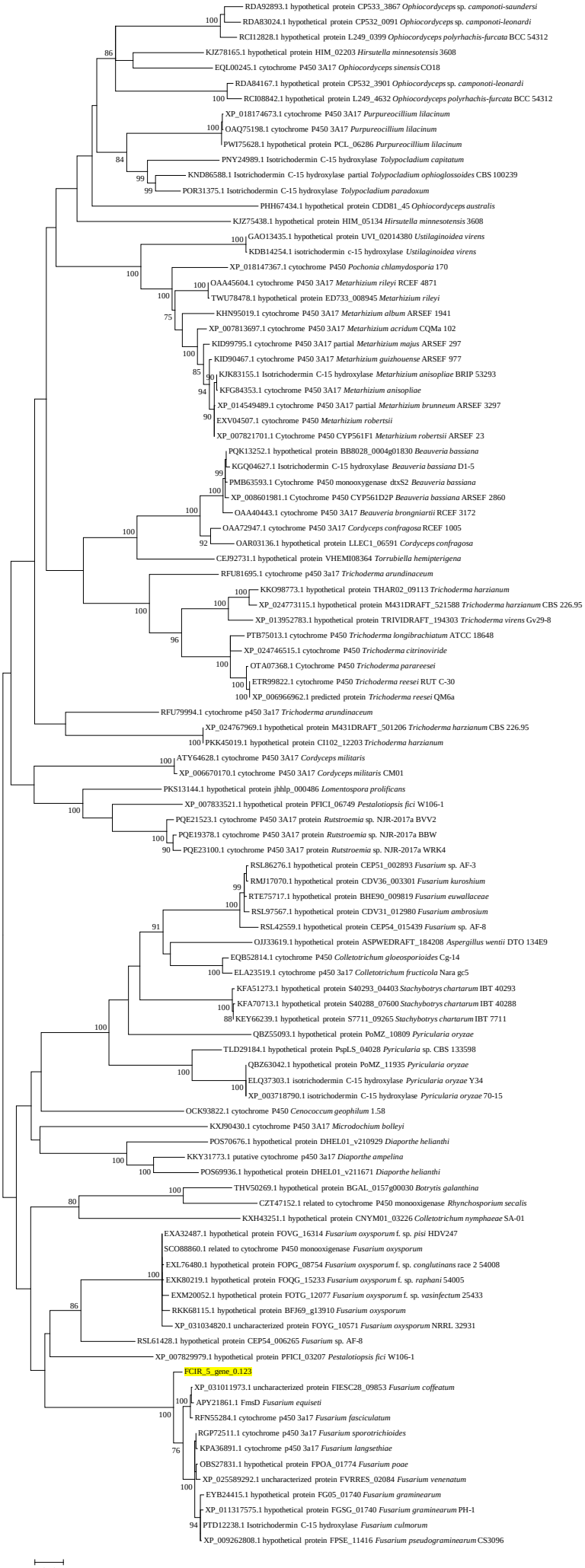

RDA92893.1 hypothetical protein CP533_3867 Ophiocordyceps sp. camponoti-saundersi
RDA83024.1 hypothetical protein CP532_0091 Ophiocordyceps sp. camponoti-leonardi
RCI12828.1 hypothetical protein L249_0399 Ophiocordyceps polyrhachis-furcata BCC 54312
KJZ78165.1 hypothetical protein HIM_02203 Hirsutella minnesotensis 3608
EQL00245.1 cytochrome P450 3A17 Ophiocordyceps sinensis CO18
RDA84167.1 hypothetical protein CP532_3901 Ophiocordyceps sp. camponoti-leonardi
RCI08842.1 hypothetical protein L249_4632 Ophiocordyceps polyrhachis-furcata BCC 54312
XP_018174673.1 cytochrome P450 3A17 Purpureocillium lilacinum
OAQ75198.1 cytochrome P450 3A17 Purpureocillium lilacinum
PWI75628.1 hypothetical protein PCL_06286 Purpureocillium lilacinum
PNY24989.1 Isotrichodermin C-15 hydroxylase Tolypocladium capitatum
KND86588.1 Isotrichodermin C-15 hydroxylase partial Tolypocladium ophioglossoides CBS 100239
POR31375.1 Isotrichodermin C-15 hydroxylase Tolypocladium paradoxum
PHH67434.1 hypothetical protein CDD81_45 Ophiocordyceps australis
KJZ75438.1 hypothetical protein HIM_05134 Hirsutella minnesotensis 3608
GAO13435.1 hypothetical protein UVI_02014380 Ustilaginoidea virens
KDB14254.1 isotrichodermin c-15 hydroxylase Ustilaginoidea virens
XP_018147367.1 cytochrome P450 Pochonia chlamydosporia 170
OAA45604.1 cytochrome P450 3A17 Metarhizium rileyi RCEF 4871
TWU78478.1 hypothetical protein ED733_008945 Metarhizium rileyi
KHN95019.1 cytochrome P450 3A17 Metarhizium album ARSEF 1941
XP_007813697.1 cytochrome P450 3A17 Metarhizium acridum CQMa 102
KID99795.1 cytochrome P450 3A17 partial Metarhizium majus ARSEF 297
KID90467.1 cytochrome P450 3A17 Metarhizium guizhouense ARSEF 977
KJK83155.1 Isotrichodermin C-15 hydroxylase Metarhizium anisopliae BRIP 53293
KFG84353.1 cytochrome P450 3A17 Metarhizium anisopliae
XP_014549489.1 cytochrome P450 3A17 partial Metarhizium brunneum ARSEF 3297
EXV04507.1 cytochrome P450 Metarhizium robertsii
XP_007821701.1 Cytochrome P450 CYP561F1 Metarhizium robertsii ARSEF 23
PQK13252.1 hypothetical protein BB8028_0004g01830 Beauveria bassiana
KGQ04627.1 Isotrichodermin C-15 hydroxylase Beauveria bassiana D1-5
PMB63593.1 Cytochrome P450 monooxygenase dtxS2 Beauveria bassiana
XP_008601981.1 Cytochrome P450 CYP561D2P Beauveria bassiana ARSEF 2860
OAA40443.1 cytochrome P450 3A17 Beauveria brongniartii RCEF 3172
OAA72947.1 cytochrome P450 3A17 Cordyceps confragosa RCEF 1005
OAR03136.1 hypothetical protein LLEC1_06591 Cordyceps confragosa
CEJ92731.1 hypothetical protein VHEMI08364 Torrubiella hemipterigena
RFU81695.1 cytochrome p450 3a17 Trichoderma arundinaceum
KKO98773.1 hypothetical protein THAR02_09113 Trichoderma harzianum
XP_024773115.1 hypothetical protein M431DRAFT_521588 Trichoderma harzianum CBS 226.95
XP_013952783.1 hypothetical protein TRIVIDRAFT_194303 Trichoderma virens Gv29-8
PTB75013.1 cytochrome P450 Trichoderma longibrachiatum ATCC 18648
XP_024746515.1 cytochrome P450 Trichoderma citrinoviride
OTA07368.1 Cytochrome P450 Trichoderma parareesei
ETR99822.1 cytochrome P450 Trichoderma reesei RUT C-30
XP_006966962.1 predicted protein Trichoderma reesei QM6a
RFU79994.1 cytochrome p450 3a17 Trichoderma arundinaceum
XP_024767969.1 hypothetical protein M431DRAFT_501206 Trichoderma harzianum CBS 226.95
PKK45019.1 hypothetical protein CI102_12203 Trichoderma harzianum
ATY64628.1 cytochrome P450 3A17 Cordyceps militaris
XP_006670170.1 cytochrome P450 3A17 Cordyceps militaris CM01
PKS13144.1 hypothetical protein jhhlp_000486 Lomentospora prolificans
XP_007833521.1 hypothetical protein PFICI_06749 Pestalotiopsis fici W106-1
PQE21523.1 cytochrome P450 3A17 protein Rutstroemia sp. NJR-2017a BVV2
PQE19378.1 cytochrome P450 3A17 protein Rutstroemia sp. NJR-2017a BBW
PQE23100.1 cytochrome P450 3A17 protein Rutstroemia sp. NJR-2017a WRK4
RSL86276.1 hypothetical protein CEP51_002893 Fusarium sp. AF-3
RMJ17070.1 hypothetical protein CDV36_003301 Fusarium kuroshium
RTE75717.1 hypothetical protein BHE90_009819 Fusarium euwallaceae
RSL97567.1 hypothetical protein CDV31_012980 Fusarium ambrosium
RSL42559.1 hypothetical protein CEP54_015439 Fusarium sp. AF-8
OJJ33619.1 hypothetical protein ASPWEDRAFT_184208 Aspergillus wentii DTO 134E9
EQB52814.1 cytochrome P450 Colletotrichum gloeosporioides Cg-14
ELA23519.1 cytochrome p450 3a17 Colletotrichum fructicola Nara gc5
KFA51273.1 hypothetical protein S40293_04403 Stachybotrys chartarum IBT 40293
KFA70713.1 hypothetical protein S40288_07600 Stachybotrys chartarum IBT 40288
KEY66239.1 hypothetical protein S7711_09265 Stachybotrys chartarum IBT 7711
QBZ55093.1 hypothetical protein PoMZ_10809 Pyricularia oryzae
TLD29184.1 hypothetical protein PspLS_04028 Pyricularia sp. CBS 133598
QBZ63042.1 hypothetical protein PoMZ_11935 Pyricularia oryzae
ELQ37303.1 isotrichodermin C-15 hydroxylase Pyricularia oryzae Y34
XP_003718790.1 isotrichodermin C-15 hydroxylase Pyricularia oryzae 70-15
OCK93822.1 cytochrome P450 Cenococcum geophilum 1.58
KXJ90430.1 cytochrome P450 3A17 Microdochium bolleyi
POS70676.1 hypothetical protein DHEL01_v210929 Diaporthe helianthi
KKY31773.1 putative cytochrome p450 3a17 Diaporthe ampelina
POS69936.1 hypothetical protein DHEL01_v211671 Diaporthe helianthi
THV50269.1 hypothetical protein BGAL_0157g00030 Botrytis galanthina
CZT47152.1 related to cytochrome P450 monooxigenase Rhynchosporium secalis
KXH43251.1 hypothetical protein CNYM01_03226 Colletotrichum nymphaeae SA-01
EXA32487.1 hypothetical protein FOVG_16314 Fusarium oxysporum f. sp. pisi HDV247
SCO88860.1 related to cytochrome P450 monooxigenase Fusarium oxysporum
EXL76480.1 hypothetical protein FOPG_08754 Fusarium oxysporum f. sp. conglutinans race 2 54008
EXK80219.1 hypothetical protein FOQG_15233 Fusarium oxysporum f. sp. raphani 54005
EXM20052.1 hypothetical protein FOTG_12077 Fusarium oxysporum f. sp. vasinfectum 25433
RKK68115.1 hypothetical protein BFJ69_g13910 Fusarium oxysporum
XP_031034820.1 uncharacterized protein FOYG_10571 Fusarium oxysporum NRRL 32931
RSL61428.1 hypothetical protein CEP54_006265 Fusarium sp. AF-8
XP_007829979.1 hypothetical protein PFICI_03207 Pestalotiopsis fici W106-1
FCIR_5_gene_0.123
XP_031011973.1 uncharacterized protein FIESC28_09853 Fusarium coffeatum
APY21861.1 FmsD Fusarium equiseti
RFN55284.1 cytochrome p450 3a17 Fusarium fasciculatum
RGP72511.1 cytochrome p450 3a17 Fusarium sporotrichioides
KPA36891.1 cytochrome p450 3a17 Fusarium langsethiae
OBS27831.1 hypothetical protein FPOA_01774 Fusarium poae
XP_025589292.1 uncharacterized protein FVRRES_02084 Fusarium venenatum
EYB24415.1 hypothetical protein FG05_01740 Fusarium graminearum
XP_011317575.1 hypothetical protein FGSG_01740 Fusarium graminearum PH-1
PTD12238.1 Isotrichodermin C-15 hydroxylase Fusarium culmorum
XP_009262808.1 hypothetical protein FPSE_11416 Fusarium pseudograminearum CS3096
0.10

Figure 6. A representative phylogenetic tree from the group 6 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The rate variation model allowed for some sites to be evolutionarily invariable. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches (≥ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).

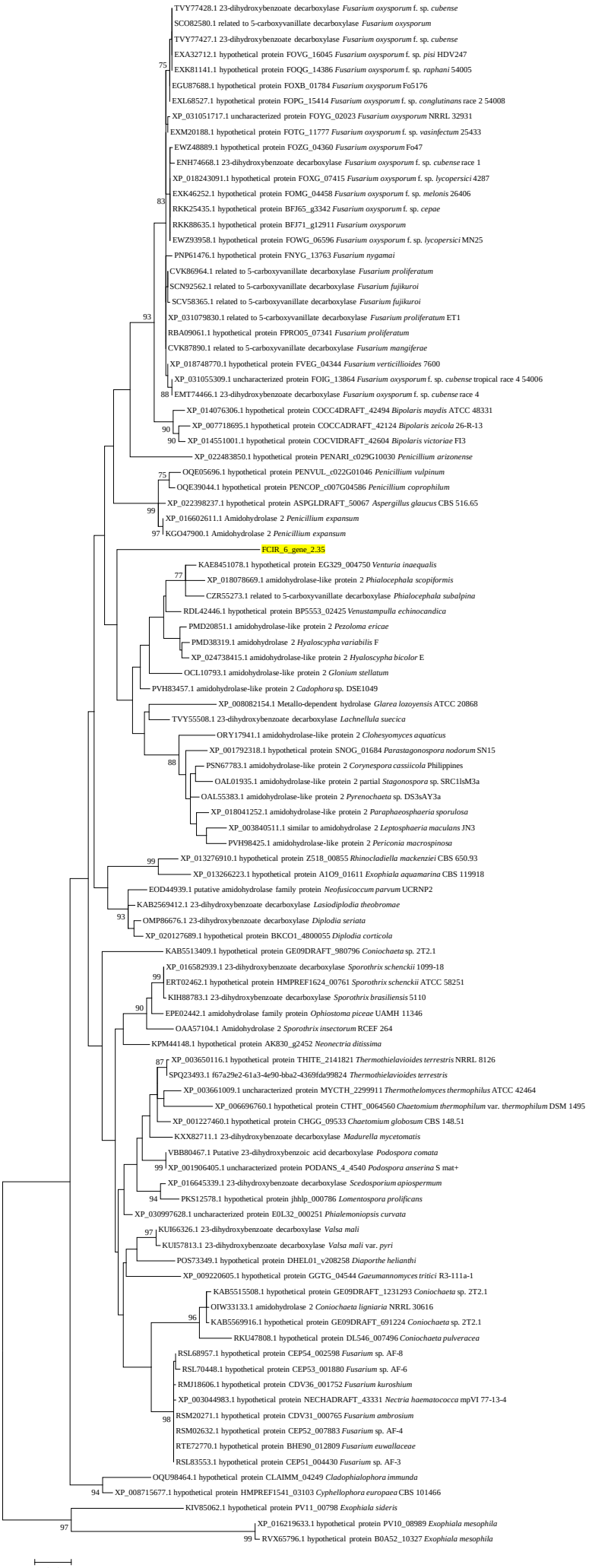

TVY77428.1 23-dihydroxybenzoate decarboxylase Fusarium oxysporum f. sp. cubense
SCO82580.1 related to 5-carboxyvanillate decarboxylase Fusarium oxysporum
TVY77427.1 23-dihydroxybenzoate decarboxylase Fusarium oxysporum f. sp. cubense
EXA32712.1 hypothetical protein FOVG_16045 Fusarium oxysporum f. sp. pisi HDV247
EXK81141.1 hypothetical protein FOQG_14386 Fusarium oxysporum f. sp. raphani 54005
EGU87688.1 hypothetical protein FOXB_01784 Fusarium oxysporum Fo5176
EXL68527.1 hypothetical protein FOPG_15414 Fusarium oxysporum f. sp. conglutinans race 2 54008
XP_031051717.1 uncharacterized protein FOYG_02023 Fusarium oxysporum NRRL 32931
EXM20188.1 hypothetical protein FOTG_11777 Fusarium oxysporum f. sp. vasinfectum 25433
EWZ48889.1 hypothetical protein FOZG_04360 Fusarium oxysporum Fo47
ENH74668.1 23-dihydroxybenzoate decarboxylase Fusarium oxysporum f. sp. cubense race 1
XP_018243091.1 hypothetical protein FOXG_07415 Fusarium oxysporum f. sp. lycopersici 4287
EXK46252.1 hypothetical protein FOMG_04458 Fusarium oxysporum f. sp. melonis 26406
RKK25435.1 hypothetical protein BFJ65_g3342 Fusarium oxysporum f. sp. cepae
RKK88635.1 hypothetical protein BFJ71_g12911 Fusarium oxysporum
EWZ93958.1 hypothetical protein FOWG_06596 Fusarium oxysporum f. sp. lycopersici MN25
PNP61476.1 hypothetical protein FNYG_13763 Fusarium nygamai
CVK86964.1 related to 5-carboxyvanillate decarboxylase Fusarium proliferatum
SCN92562.1 related to 5-carboxyvanillate decarboxylase Fusarium fujikuroi
SCV58365.1 related to 5-carboxyvanillate decarboxylase Fusarium fujikuroi
XP_031079830.1 related to 5-carboxyvanillate decarboxylase Fusarium proliferatum ET1
RBA09061.1 hypothetical protein FPRO05_07341 Fusarium proliferatum
CVK87890.1 related to 5-carboxyvanillate decarboxylase Fusarium mangiferae
XP_018748770.1 hypothetical protein FVEG_04344 Fusarium verticillioides 7600
XP_031055309.1 uncharacterized protein FOIG_13864 Fusarium oxysporum f. sp. cubense tropical race 4 54006
EMT74466.1 23-dihydroxybenzoate decarboxylase Fusarium oxysporum f. sp. cubense race 4
XP_014076306.1 hypothetical protein COCC4DRAFT_42494 Bipolaris maydis ATCC 48331
XP_007718695.1 hypothetical protein COCCADRAFT_42124 Bipolaris zeicola 26-R-13
XP_014551001.1 hypothetical protein COCVIDRAFT_42604 Bipolaris victoriae FI3
XP_022483850.1 hypothetical protein PENARI_c029G10030 Penicillium arizonense
OQE05696.1 hypothetical protein PENVUL_c022G01046 Penicillium vulpinum
OQE39044.1 hypothetical protein PENCOP_c007G04586 Penicillium coprophilum
XP_022398237.1 hypothetical protein ASPGLDRAFT_50067 Aspergillus glaucus CBS 516.65
XP_016602611.1 Amidohydrolase 2 Penicillium expansum
KGO47900.1 Amidohydrolase 2 Penicillium expansum
FCIR_6_gene_2.35
KAE8451078.1 hypothetical protein EG329_004750 Venturia inaequalis
XP_018078669.1 amidohydrolase-like protein 2 Phialocephala scopiformis
CZR55273.1 related to 5-carboxyvanillate decarboxylase Phialocephala subalpina
RDL42446.1 hypothetical protein BP5553_02425 Venustampulla echinocandica
PMD20851.1 amidohydrolase-like protein 2 Pezoloma ericae
PMD38319.1 amidohydrolase 2 Hyaloscypha variabilis F
XP_024738415.1 amidohydrolase-like protein 2 Hyaloscypha bicolor E
OCL10793.1 amidohydrolase-like protein 2 Glonium stellatum
PVH83457.1 amidohydrolase-like protein 2 Cadophora sp. DSE1049
XP_008082154.1 Metallo-dependent hydrolase Glarea lozoyensis ATCC 20868
TVY55508.1 23-dihydroxybenzoate decarboxylase Lachnellula suecica
ORY17941.1 amidohydrolase-like protein 2 Clohesyomyces aquaticus
XP_001792318.1 hypothetical protein SNOG_01684 Parastagonospora nodorum SN15
PSN67783.1 amidohydrolase-like protein 2 Corynespora cassiicola Philippines
OAL01935.1 amidohydrolase-like protein 2 partial Stagonospora sp. SRC1lsM3a
OAL55383.1 amidohydrolase-like protein 2 Pyrenochaeta sp. DS3sAY3a
XP_018041252.1 amidohydrolase-like protein 2 Paraphaeosphaeria sporulosa
XP_003840511.1 similar to amidohydrolase 2 Leptosphaeria maculans JN3
PVH98425.1 amidohydrolase-like protein 2 Periconia macrospinosa
XP_013276910.1 hypothetical protein Z518_00855 Rhinocladiella mackenziei CBS 650.93
XP_013266223.1 hypothetical protein A1O9_01611 Exophiala aquamarina CBS 119918
EOD44939.1 putative amidohydrolase family protein Neofusicoccum parvum UCRNP2
KAB2569412.1 23-dihydroxybenzoate decarboxylase Lasiodiplodia theobromae
OMP86676.1 23-dihydroxybenzoate decarboxylase Diplodia seriata
XP_020127689.1 hypothetical protein BKCO1_4800055 Diplodia corticola
KAB5513409.1 hypothetical protein GE09DRAFT_980796 Coniochaeta sp. 2T2.1
XP_016582939.1 23-dihydroxybenzoate decarboxylase Sporothrix schenckii 1099-18
ERT02462.1 hypothetical protein HMPREF1624_00761 Sporothrix schenckii ATCC 58251
KIH88783.1 23-dihydroxybenzoate decarboxylase Sporothrix brasiliensis 5110
EPE02442.1 amidohydrolase family protein Ophiostoma piceae UAMH 11346
OAA57104.1 Amidohydrolase 2 Sporothrix insectorum RCEF 264
KPM44148.1 hypothetical protein AK830_g2452 Neonectria ditissima
XP_003650116.1 hypothetical protein THITE_2141821 Thermothielavioides terrestris NRRL 8126
SPQ23493.1 f67a29e2-61a3-4e90-bba2-4369fda99824 Thermothielavioides terrestris
XP_003661009.1 uncharacterized protein MYCTH_2299911 Thermothelomyces thermophilus ATCC 42464
XP_006696760.1 hypothetical protein CTHT_0064560 Chaetomium thermophilum var. thermophilum DSM 1495
XP_001227460.1 hypothetical protein CHGG_09533 Chaetomium globosum CBS 148.51
KXX82711.1 23-dihydroxybenzoate decarboxylase Madurella mycetomatis
VBB80467.1 Putative 23-dihydroxybenzoic acid decarboxylase Podospora comata
XP_001906405.1 uncharacterized protein PODANS_4_4540 Podospora anserina S mat+
XP_016645339.1 23-dihydroxybenzoate decarboxylase Scedosporium apiospermum
PKS12578.1 hypothetical protein jhhlp_000786 Lomentospora prolificans
XP_030997628.1 uncharacterized protein E0L32_000251 Phialemoniopsis curvata
KUI66326.1 23-dihydroxybenzoate decarboxylase Valsa mali
KUI57813.1 23-dihydroxybenzoate decarboxylase Valsa mali var. pyri
POS73349.1 hypothetical protein DHEL01_v208258 Diaporthe helianthi
XP_009220605.1 hypothetical protein GGTG_04544 Gaeumannomyces tritici R3-111a-1
KAB5515508.1 hypothetical protein GE09DRAFT_1231293 Coniochaeta sp. 2T2.1
OIW33133.1 amidohydrolase 2 Coniochaeta ligniaria NRRL 30616
KAB5569916.1 hypothetical protein GE09DRAFT_691224 Coniochaeta sp. 2T2.1
RKU47808.1 hypothetical protein DL546_007496 Coniochaeta pulveracea
RSL68957.1 hypothetical protein CEP54_002598 Fusarium sp. AF-8
RSL70448.1 hypothetical protein CEP53_001880 Fusarium sp. AF-6
RMJ18606.1 hypothetical protein CDV36_001752 Fusarium kuroshium
XP_003044983.1 hypothetical protein NECHADRAFT_43331 Nectria haematococca mpVI 77-13-4
RSM20271.1 hypothetical protein CDV31_000765 Fusarium ambrosium
RSM02632.1 hypothetical protein CEP52_007883 Fusarium sp. AF-4
RTE72770.1 hypothetical protein BHE90_012809 Fusarium euwallaceae
RSL83553.1 hypothetical protein CEP51_004430 Fusarium sp. AF-3
OQU98464.1 hypothetical protein CLAIMM_04249 Cladophialophora immunda
XP_008715677.1 hypothetical protein HMPREF1541_03103 Cyphellophora europaea CBS 101466
KIV85062.1 hypothetical protein PV11_00798 Exophiala sideris
XP_016219633.1 hypothetical protein PV10_08989 Exophiala mesophila
RVX65796.1 hypothetical protein B0A52_10327 Exophiala mesophila
0.20

Figure 7. A representative phylogenetic tree from the group 7 ancestral origin classification system (unique gene investigated is highlighted in yellow; FCIR = *Fusarium circinatum*). Evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches ($\geq$ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (Kumar *et al.*, 2016).

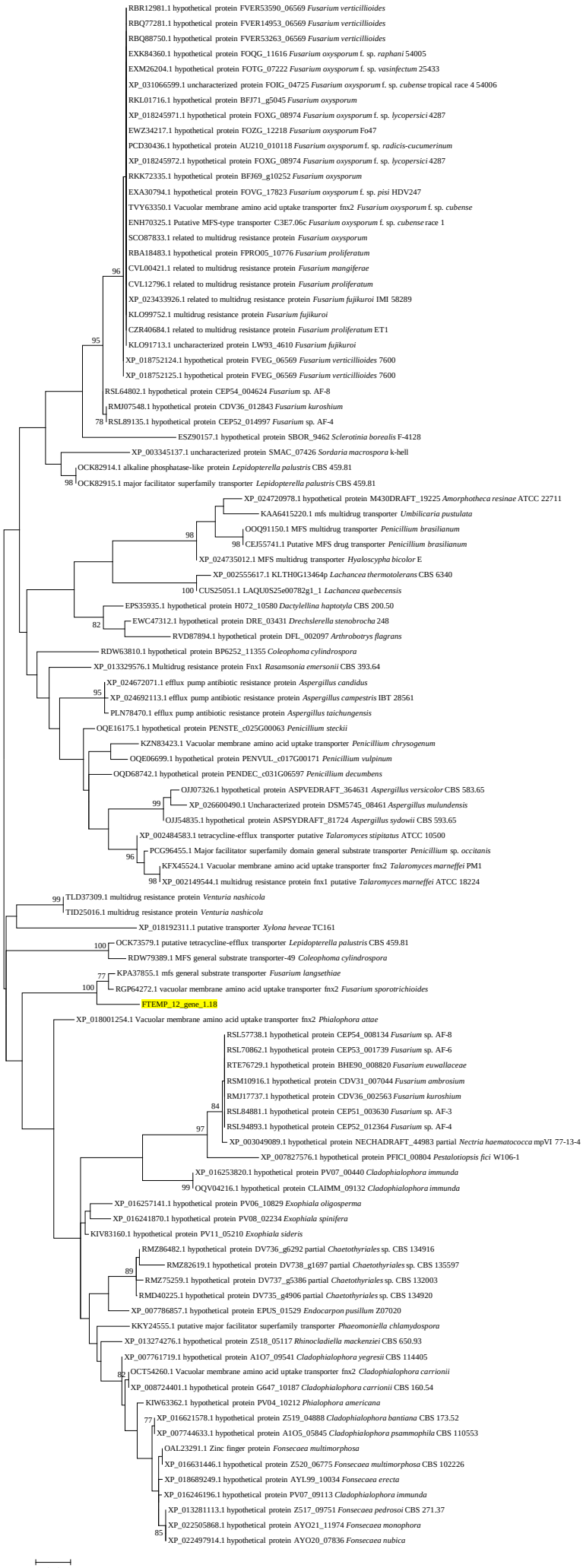
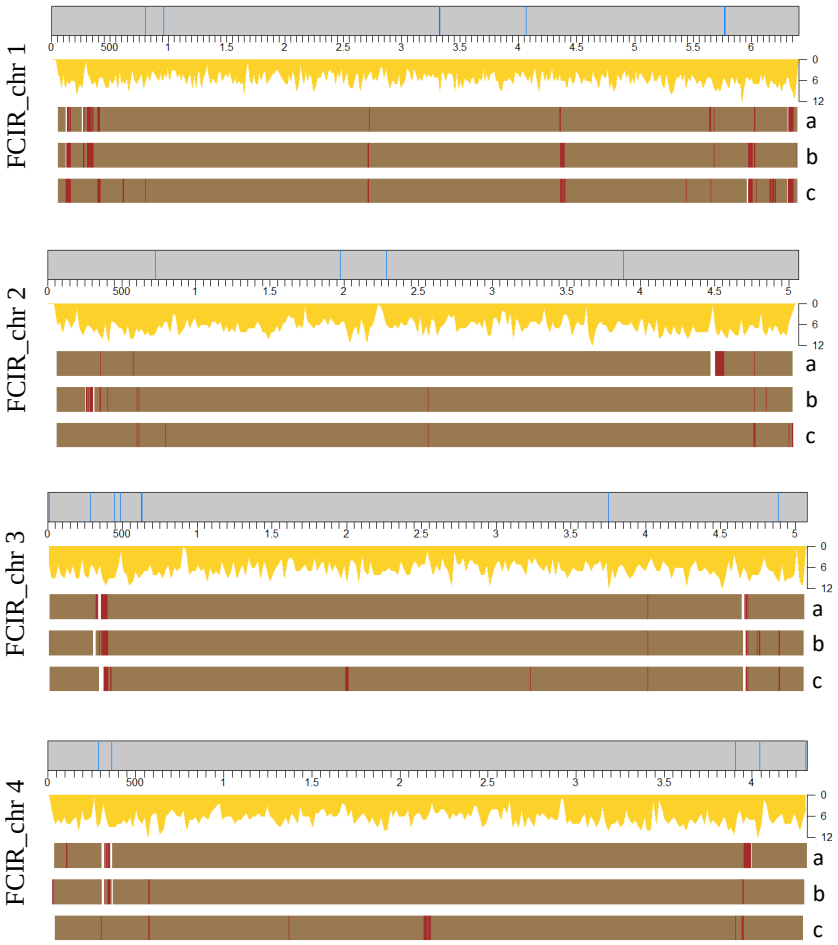

RBR12981.1 hypothetical protein FVER53590_06569 Fusarium verticillioides
RBQ77281.1 hypothetical protein FVER14953_06569 Fusarium verticillioides
RBQ88750.1 hypothetical protein FVER53263_06569 Fusarium verticillioides
EXK84360.1 hypothetical protein FOQG_11616 Fusarium oxysporum f. sp. raphani 54005
EXM26204.1 hypothetical protein FOTG_07222 Fusarium oxysporum f. sp. vasinfectum 25433
XP_031066599.1 uncharacterized protein FOIG_04725 Fusarium oxysporum f. sp. cubense tropical race 4 54006
RKL01716.1 hypothetical protein BFJ71_g5045 Fusarium oxysporum
XP_018245971.1 hypothetical protein FOXG_08974 Fusarium oxysporum f. sp. lycopersici 4287
EWZ34217.1 hypothetical protein FOZG_12218 Fusarium oxysporum Fo47
PCD30436.1 hypothetical protein AU210_010118 Fusarium oxysporum f. sp. radicis-cucumerinum
XP_018245972.1 hypothetical protein FOXG_08974 Fusarium oxysporum f. sp. lycopersici 4287
RKK72335.1 hypothetical protein BFJ69_g10252 Fusarium oxysporum
EXA30794.1 hypothetical protein FOVG_17823 Fusarium oxysporum f. sp. pisi HDV247
TVY63350.1 Vacuolar membrane amino acid uptake transporter fnx2 Fusarium oxysporum f. sp. cubense
ENH70325.1 Putative MFS-type transporter C3E7.06c Fusarium oxysporum f. sp. cubense race 1
SCO87833.1 related to multidrug resistance protein Fusarium oxysporum
RBA18483.1 hypothetical protein FPRO05_10776 Fusarium proliferatum
CVL00421.1 related to multidrug resistance protein Fusarium mangiferae
CVL12796.1 related to multidrug resistance protein Fusarium proliferatum
XP_023433926.1 related to multidrug resistance protein Fusarium fujikuroi IMI 58289
KLO99752.1 multidrug resistance protein Fusarium fujikuroi
CZR40684.1 related to multidrug resistance protein Fusarium proliferatum ET1
KLO91713.1 uncharacterized protein LW93_4610 Fusarium fujikuroi
XP_018752124.1 hypothetical protein FVEG_06569 Fusarium verticillioides 7600
XP_018752125.1 hypothetical protein FVEG_06569 Fusarium verticillioides 7600
RSL64802.1 hypothetical protein CEP54_004624 Fusarium sp. AF-8
RMJ07548.1 hypothetical protein CDV36_012843 Fusarium kuroshium
RSL89135.1 hypothetical protein CEP52_014997 Fusarium sp. AF-4
ESZ90157.1 hypothetical protein SBOR_9462 Sclerotinia borealis F-4128
XP_003345137.1 uncharacterized protein SMAC_07426 Sordaria macrospora k-hell
OCK82914.1 alkaline phosphatase-like protein Lepidopterella palustris CBS 459.81
OCK82915.1 major facilitator superfamily transporter Lepidopterella palustris CBS 459.81
XP_024720978.1 hypothetical protein M430DRAFT_19225 Amorphotheca resinae ATCC 22711
KAA6415220.1 mfs multidrug transporter Umbilicaria pustulata
OOQ91150.1 MFS multidrug transporter Penicillium brasilianum
CEJ55741.1 Putative MFS drug transporter Penicillium brasilianum
XP_024735012.1 MFS multidrug transporter Hyaloscypha bicolor E
XP_002555617.1 KLTH0G13464p Lachancea thermotolerans CBS 6340
CUS25051.1 LAQU0S25e00782g1_1 Lachancea quebecensis
EPS35935.1 hypothetical protein H072_10580 Dactylellina haptotyla CBS 200.50
EWC47312.1 hypothetical protein DRE_03431 Drechslerella stenobrocha 248
RVD87894.1 hypothetical protein DFL_002097 Arthrobotrys flagrans
RDW63810.1 hypothetical protein BP6252_11355 Coleophoma cylindrospora
XP_013329576.1 Multidrug resistance protein Fnx1 Rasamsonia emersonii CBS 393.64
XP_024672071.1 efflux pump antibiotic resistance protein Aspergillus candidus
XP_024692113.1 efflux pump antibiotic resistance protein Aspergillus campestris IBT 28561
PLN78470.1 efflux pump antibiotic resistance protein Aspergillus taichungensis
OQE16175.1 hypothetical protein PENSTE_c025G00063 Penicillium steckii
KZN83423.1 Vacuolar membrane amino acid uptake transporter Penicillium chrysogenum
OQE06699.1 hypothetical protein PENVUL_c017G00171 Penicillium vulpinum
OQD68742.1 hypothetical protein PENDEC_c031G06597 Penicillium decumbens
OJJ07326.1 hypothetical protein ASPVEDRAFT_364631 Aspergillus versicolor CBS 583.65
XP_026600490.1 Uncharacterized protein DSM5745_08461 Aspergillus mulundensis
OJJ54835.1 hypothetical protein ASPSYDRAFT_81724 Aspergillus sydowii CBS 593.65
XP_002484583.1 tetracycline-efflux transporter putative Talaromyces stipitatus ATCC 10500
PCG96455.1 Major facilitator superfamily domain general substrate transporter Penicillium sp. occitanis
KFX45524.1 Vacuolar membrane amino acid uptake transporter fnx2 Talaromyces marneffei PM1
XP_002149544.1 multidrug resistance protein fnx1 putative Talaromyces marneffei ATCC 18224
TLD37309.1 multidrug resistance protein Venturia nashicola
TID25016.1 multidrug resistance protein Venturia nashicola
XP_018192311.1 putative transporter Xylona heveae TC161
OCK73579.1 putative tetracycline-efflux transporter Lepidopterella palustris CBS 459.81
RDW79389.1 MFS general substrate transporter-49 Coleophoma cylindrospora
KPA37855.1 mfs general substrate transporter Fusarium langsethiae
RGP64272.1 vacuolar membrane amino acid uptake transporter fnx2 Fusarium sporotrichioides
FTEMP_12_gene_1.18
XP_018001254.1 Vacuolar membrane amino acid uptake transporter fnx2 Phialophora attae
RSL57738.1 hypothetical protein CEP54_008134 Fusarium sp. AF-8
RSL70862.1 hypothetical protein CEP53_001739 Fusarium sp. AF-6
RTE76729.1 hypothetical protein BHE90_008820 Fusarium euwallaceae
RSM10916.1 hypothetical protein CDV31_007044 Fusarium ambrosium
RMJ17737.1 hypothetical protein CDV36_002563 Fusarium kuroshium
RSL84881.1 hypothetical protein CEP51_003630 Fusarium sp. AF-3
RSL94893.1 hypothetical protein CEP52_012364 Fusarium sp. AF-4
XP_003049089.1 hypothetical protein NECHADRAFT_44983 partial Nectria haematococca mpVI 77-13-4
XP_007827576.1 hypothetical protein PFICI_00804 Pestalotiopsis fici W106-1
XP_016253820.1 hypothetical protein PV07_00440 Cladophialophora immunda
OQV04216.1 hypothetical protein CLAIMM_09132 Cladophialophora immunda
XP_016257141.1 hypothetical protein PV06_10829 Exophiala oligosperma
XP_016241870.1 hypothetical protein PV08_02234 Exophiala spinifera
KIV83160.1 hypothetical protein PV11_05210 Exophiala sideris
RMZ86482.1 hypothetical protein DV736_g6292 partial Chaetothyriales sp. CBS 134916
RMZ82619.1 hypothetical protein DV738_g1697 partial Chaetothyriales sp. CBS 135597
RMZ75259.1 hypothetical protein DV737_g5386 partial Chaetothyriales sp. CBS 132003
RMD40225.1 hypothetical protein DV735_g4906 partial Chaetothyriales sp. CBS 134920
XP_007786857.1 hypothetical protein EPUS_01529 Endocarpon pusillum Z07020
KKY24555.1 putative major facilitator superfamily transporter Phaeomoniella chlamydospora
XP_013274276.1 hypothetical protein Z518_05117 Rhinocladiella mackenziei CBS 650.93
XP_007761719.1 hypothetical protein A1O7_09541 Cladophialophora yegresii CBS 114405
OCT54260.1 Vacuolar membrane amino acid uptake transporter fnx2 Cladophialophora carrionii
XP_008724401.1 hypothetical protein G647_10187 Cladophialophora carrionii CBS 160.54
KIW63362.1 hypothetical protein PV04_10212 Phialophora americana
XP_016621578.1 hypothetical protein Z519_04888 Cladophialophora bantiana CBS 173.52
XP_007744633.1 hypothetical protein A1O5_05845 Cladophialophora psammophila CBS 110553
OAL23291.1 Zinc finger protein Fonsecaea multimorphosa
XP_016631446.1 hypothetical protein Z520_06775 Fonsecaea multimorphosa CBS 102226
XP_018689249.1 hypothetical protein AYL99_10034 Fonsecaea erecta
XP_016246196.1 hypothetical protein PV07_09113 Cladophialophora immunda
XP_013281113.1 hypothetical protein Z517_09751 Fonsecaea pedrosoi CBS 271.37
XP_022505868.1 hypothetical protein AYO21_11974 Fonsecaea monophora
XP_022497914.1 hypothetical protein AYO20_07836 Fonsecaea nubica
96
95
78
98
98
98
100
82
95
99
96
98
99
100
77
100
84
97
99
89
82
77
85
0.20

Figure 8. A representative phylogenetic tree from the group 8 ancestral origin classification system (unique gene investigated is highlighted in yellow; FTEMP = *Fusarium temperatum*). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences amongst sites. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches (≥ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016).


FCIR_chr 1
FCIR_chr 2
FCIR_chr 3
FCIR_chr 4
FCIR_chr 5
FCIR_chr 6
FCIR_chr 7
FCIR_chr 8
a
b
c

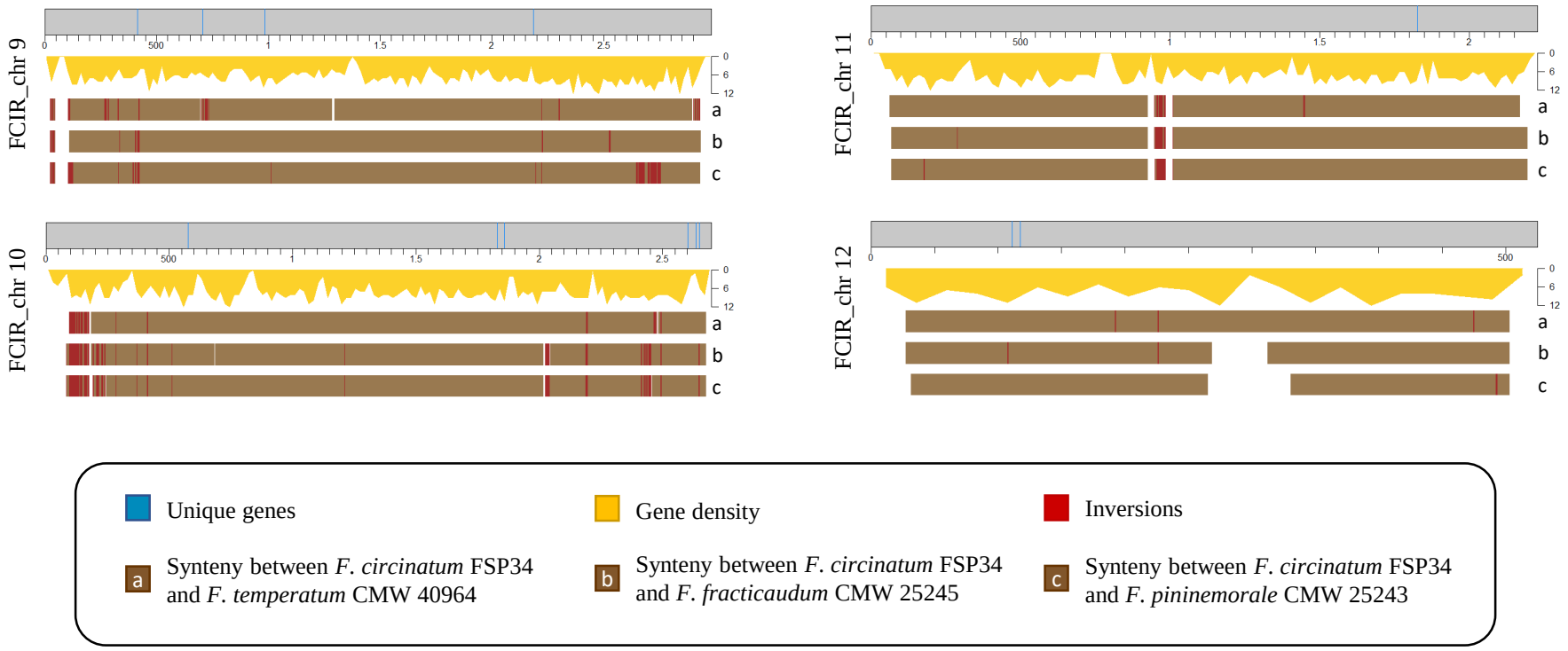
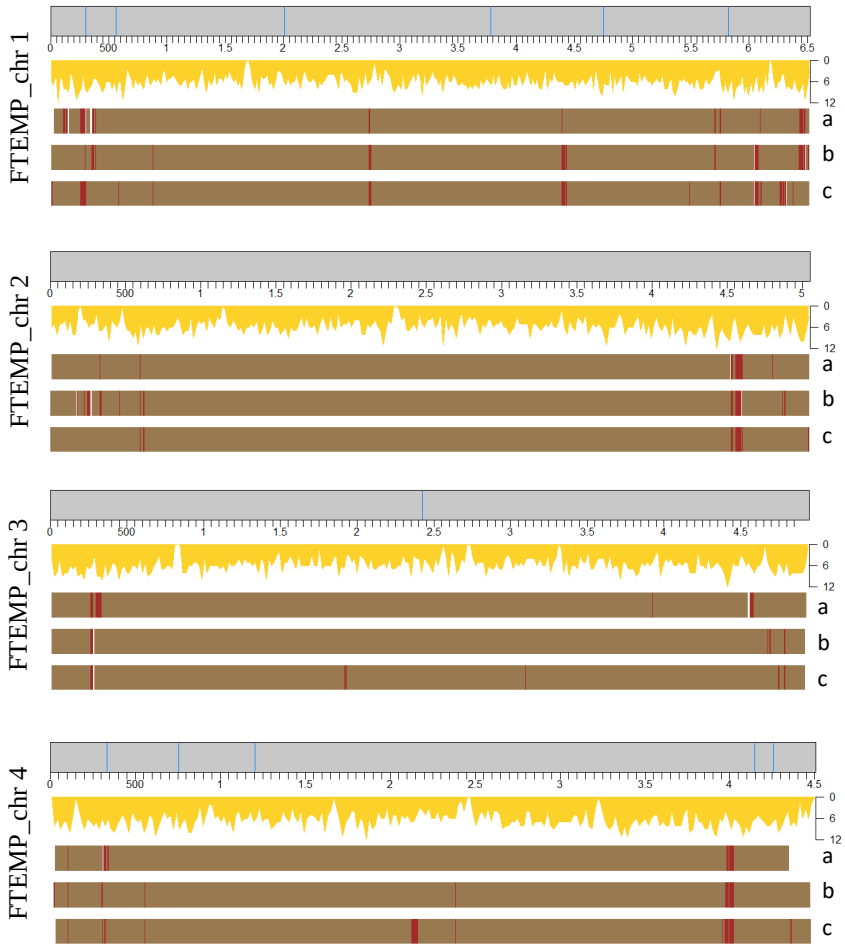


Figure 9. Pine-associated unique genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FCIR = *Fusarium circinatum*.


FTEMP_chr 1
FTEMP_chr 2
FTEMP_chr 3
FTEMP_chr 4
FTEMP_chr 5
FTEMP_chr 6
FTEMP_chr 7
FTEMP_chr 8
a
b
c

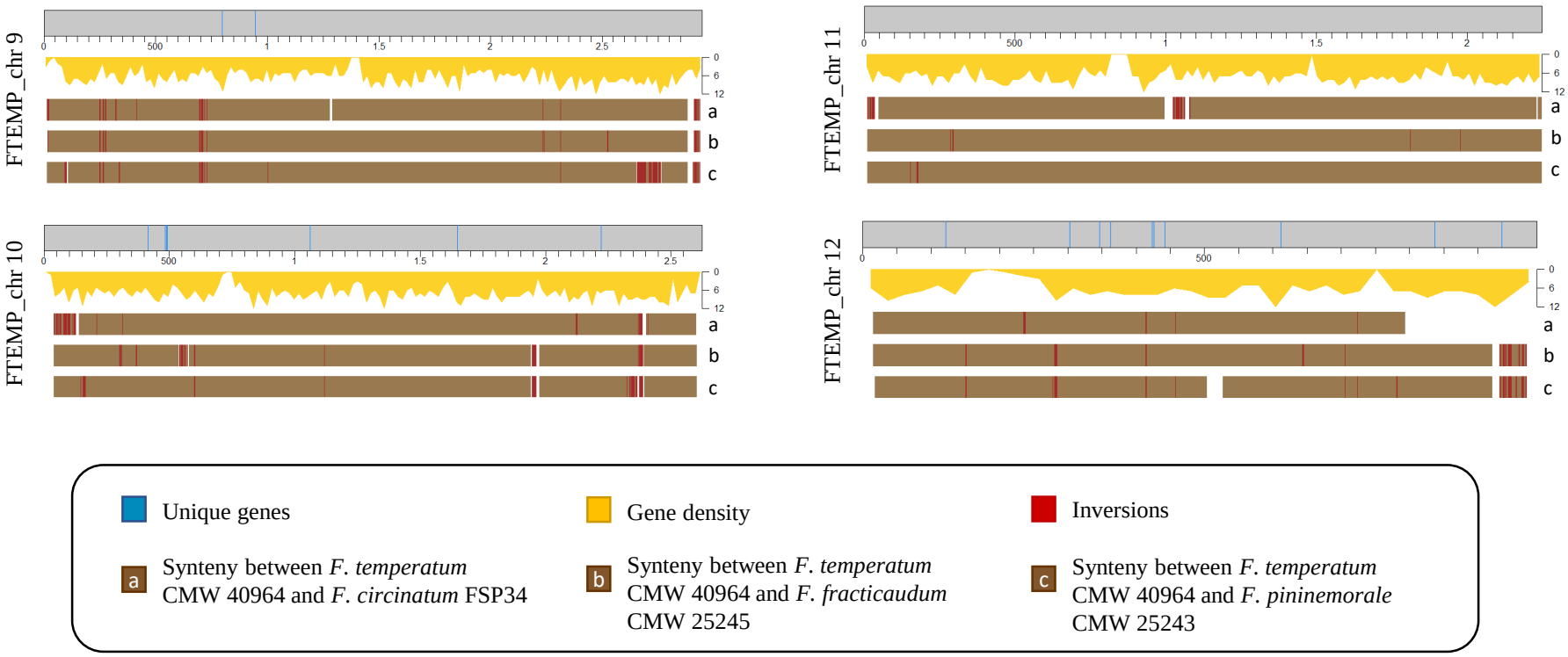


Figure 10. Poaceae-associated unique genes distribution across each of the chromosomes as indicated by the blue lines. The conservation of synteny and inversion between the relevant genomes are indicated in the brown blocks and red lines. FTEMP = *Fusarium temperatum*.

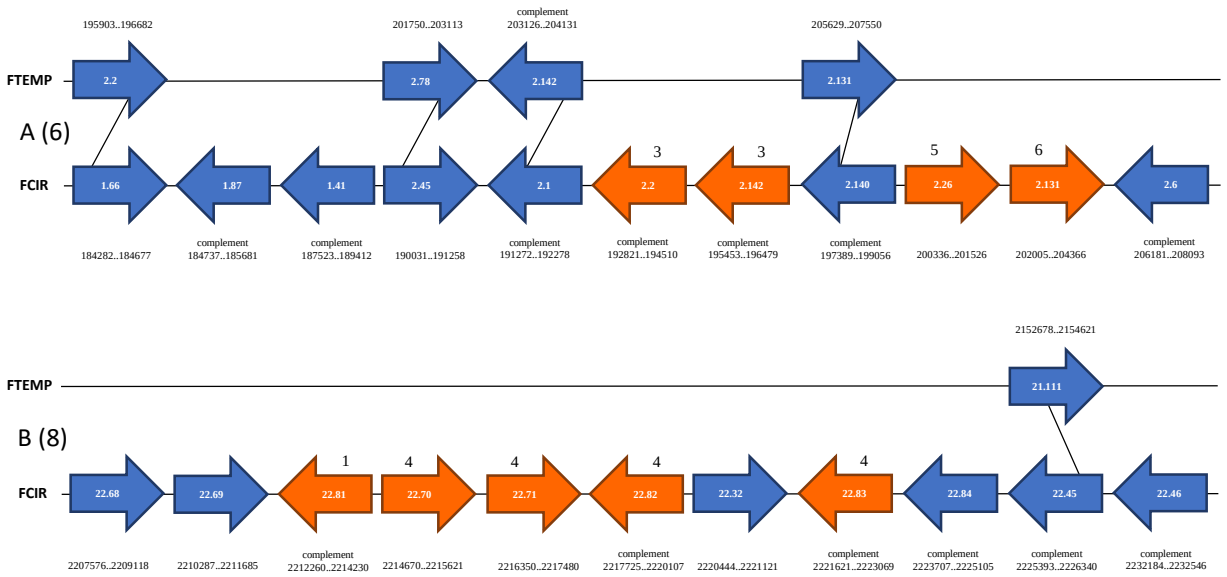


Figure 11. Integration of SynChro and phylogenetic results of the pine-associated genes clustering together or those in close proximity to each other. This is a representation of numerous *F. circinatum* FSP34 unique genes that are clustered. The *F. temperatum* CMW 40964 genes shared 36.4% synteny in (A) and 9.1% in (B), towards the *F. circinatum* FSP34 genes. Unique genes can be located next to each other or in close proximity to each other, and can have the same or different ancestral origins.

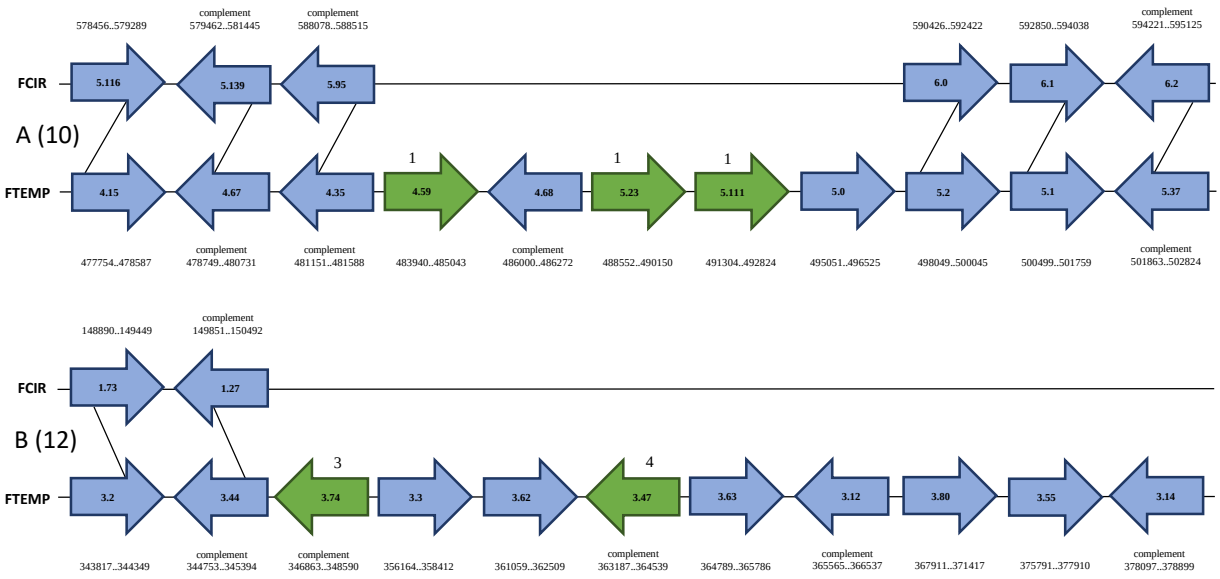


Figure 12. Integration of SynChro and phylogenetic results of the Poaceae-associated genes clustering together or those in close proximity to each other. This is a representation of numerous *F. temperatum* CMW 40964 unique genes that are clustered. The *F. circinatum* FSP34 genes shared 54.5% synteny in (A) and 18.2% in (B), towards the *F. temperatum* CMW 40964 genes. Unique genes can be located next to each other or in close proximity to each other, and can have the same or different ancestral origins.

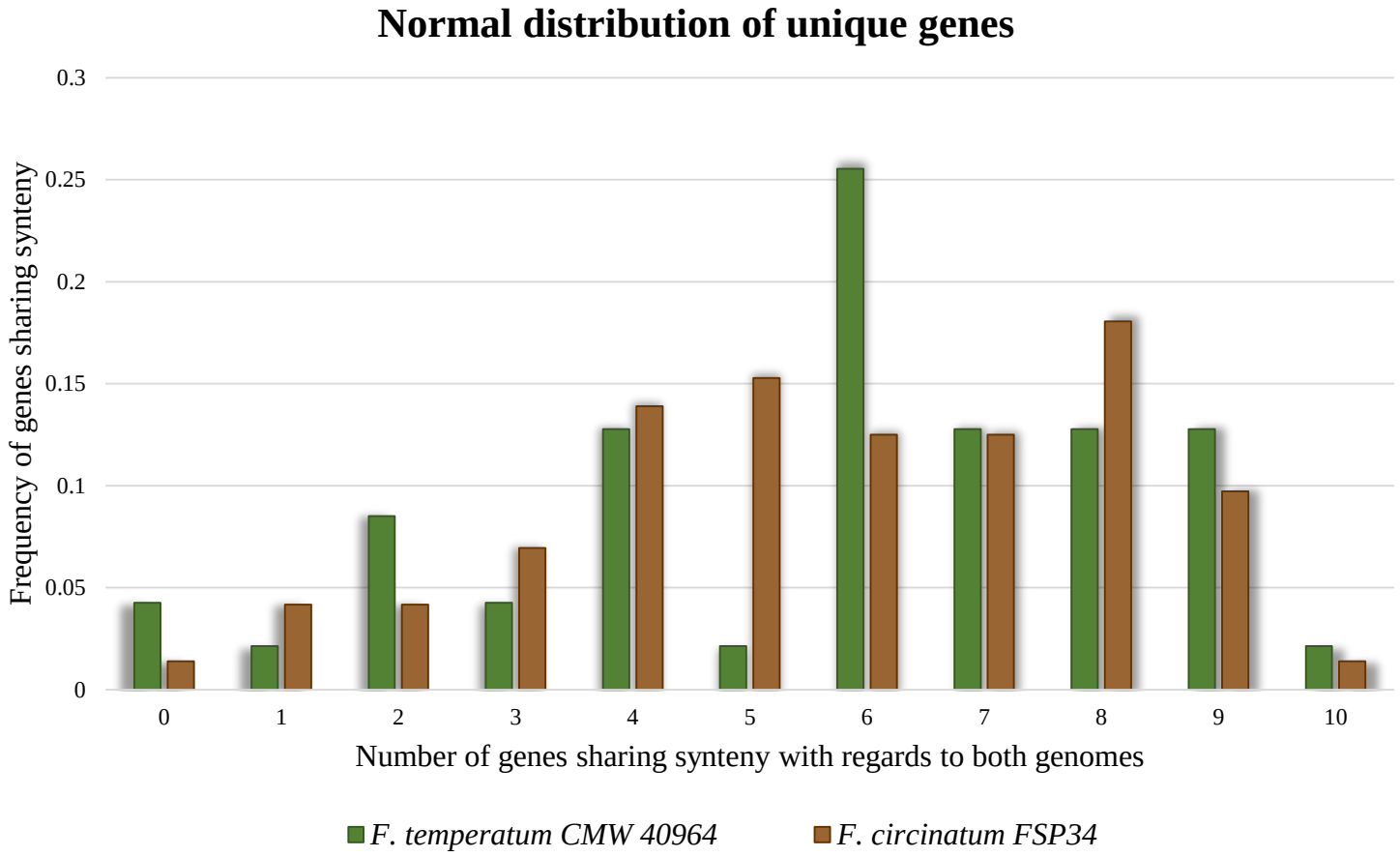


Figure 13. Data retrieved from the Shapiro-Wilk's test for departure from normality for the unique genes of the pine- and Poaceae-associated *Fusarium* species, represented by *F. circinatum* FSP34 and *F. temperatum* CMW 40964, respectively.

SUPPLEMENTAL TABLES

All the supplemental data tables can be retrieved from the following Google drive:
https://drive.google.com/drive/folders/1eK5hGcNRkECBVHwOno6XlOKbKUI43ZLg?usp=sharing

Supplemental Table 1. BUSCO results for the relevant *Fusarium* genomes.

Supplemental Table 2. The chromosome fold-difference between four of the six *Fusarium* species. Indicated in red is the chromosome which is the largest amongst the four investigated species.

Supplemental Table 3. Pearson product-moment correlation coefficient calculation.

Supplemental Table 4. The output retrieved from OrthoFinder, including all six *Fusarium* genomes. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12). In *F. konzum* and *F. subglutinans* it indicates the scaffold number.

Supplemental Table 5. The shared genes across all six *Fusarium* genomes, along with the amino acid (AA) sequence for each gene. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12). In *F. konzum* and *F. subglutinans* it indicates the scaffold number.

Supplemental Table 6. The 72 unique genes, along with their multiple copies, shared between all the pine-associated *Fusarium* species. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12).

Supplemental Table 7. The 47 unique genes, along with their multiple copies, shared between all Poaceae-associated *Fusarium* species. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12). In *F. konzum* and *F. subglutinans* it indicates the scaffold number.

Supplemental Table 8. The EST data for *F. circinatum*, obtained from Wingfield *et al*. (2012) and van Wyk *et al.* (2019).

Supplemental Table 9. The Blast2GO data for the 72 unique pine-associated genes, specifically for (A) *Fusarium circinatum*, (B) *F. fracticaudum* and (C) *F. pininemorale*.

Supplemental Table 10. The Blast2GO data for the 47 unique Poaceae-associated genes, specifically for (A) *Fusarium konzum*, (B) *F. subglutinans* and (C) *F. temperatum*.

Supplemental Table 11. The Fischer exact test data for (A) the 72 unique pine-associated genes and (B) the 47 unique Poaceae-associated genes.

Supplemental Table 12. The Chi-squared test calculations for the genomic locations of the unique genes from both *F. circinatum* and *F. temperatum*.

Supplemental Table 13. The unique gene density for both *F. circinatum* and *F. temperatum*.

Supplemental Table 14. The SynChro data for genes downstream and upstream of the unique genes of both the pine- and Poaceae-associated *Fusarium* species. FCIR = *Fusarium circinatum* FSP34 and FTEMP = *Fusarium temperatum* CMW 40964.

Supplemental Table 15. The Shapiro-Wilk's test for departure from normality.

SUPPLEMENTAL FIGURES

All the supplemental figures can be retrieved from the following Google drive:
https://drive.google.com/drive/folders/1eK5hGcNRkECBVHwOno6XlOKbKUI43ZLg?usp=sharing

Supplemental Figure 1A. A pie chart of the level 4 GO terms involved with the biological processes of the 72 unique genes from the pine-associated *Fusarium* species.

Supplemental Figure 1B. A pie chart of the level 4 GO terms involved with the molecular functions of the 72 unique genes from the pine-associated *Fusarium* species.

Supplemental Figure 1C. A pie chart of the level 4 GO terms involved with the cellular components of the 72 unique genes from the pine-associated *Fusarium* species.

Supplemental Figure 1D. A pie chart of the level 4 GO terms involved with the biological processes of the 47 unique genes from the Poaceae-associated *Fusarium* species.

Supplemental Figure 1E. A pie chart of the level 4 GO terms involved with the molecular functions of the 47 unique genes from the Poaceae-associated *Fusarium* species.

Supplemental Figure 1F. A pie chart of the level 4 GO terms involved with the cellular components of the 47 unique genes from the Poaceae-associated *Fusarium* species.

Supplemental Figure 2A. Unique genes with ancestral origins that emerged within the FFSC and were retained only in certain lineages.

Supplemental Figure 2B. Unique genes with ancestral origins that emerged due to a duplication in the FFSC but were retained/acquired only in a few FFSC species.

Supplemental Figure 2C. Unique genes with ancestral origins that originated in the FOSC.

Supplemental Figure 2D. Unique genes with ancestral origins within the broader FOSC+FFSC clade but were only retained in certain species.

Supplemental Figure 2E. Unique genes with ancestral origins that originated in the *Fusarium* F2 clade but were retained/acquired only in certain *Fusarium* species.

Supplemental Figure 2F. Unique genes with ancestral origins that originated in the *Fusarium* F3 clade, excluding *Fusarium solani* (*Nectria haematococca*).

Supplemental Figure 2G. Unique genes with ancestral origins that emerged from ascomycetes outside the *Fusarium* clade.

Supplemental Figure 2H. Unique genes with ancestral origins that emerged due to a duplication in the *Fusarium* clade but were retained/acquired only in certain species complexes.

Supplemental Figure 2I. Unique genes with ancestral origins that have no significant hits based on NCBI results and phylogenetic analyses.

CHAPTER 3:

CAZyme gene repertoire comparisons of selected *Fusarium* species from the American clade of the FFSC

ABSTRACT

The genus *Fusarium* is diverse and harbour species that can be organised into different species complexes. One well-known species complex is the *Fusarium fujikuroi* species complex (FFSC) that contains economically and agriculturally important pathogens. Fungi, in general, are known to occupy diverse niches and utilise many different strategies for survival and reproduction. The host-pathogen interaction between plants and fungi is dynamic, complex and co-dependent. The initial phase of host cell penetration initiates on the epidermal cells, followed by entry into intracellular spaces. The process of penetration is often facilitated by Carbohydrate-Active enZymes (CAZymes). In this study, we investigated the CAZyme repertoire of species associated with *Pinus* species (*F. circinatum* FSP34, *F. pininemorale* CMW 25243 and *F. fracticaudum* CMW 25245) and members of Poaceae (*F. temperatum* CMW 40964, *F. subglutinans* NRRL 22016 and *F. konzum* NRRL 11616). This study aims to identify CAZymes that could play a role in host defence mechanisms, but that also differ between the *Fusarium* species belonging to each host group. For this, we implemented comparative genomics to illustrate the difference in the CAZyme repertoire between the two groups of *Fusarium* species. We found that the glycoside hydrolases (GHs) class was the most prevalent in all six genomes, followed by the carbohydrate esterases (CEs), auxiliary activities (AAs) and glycosyltransferases (GTs) classes with an almost similar repertoire of genes, emphasising the carbohydrate degradation and utilisation potential of these fungi, irrespective of host. The similarities of the CAZyme families between the *Fusarium* species associated with monocots and gymnosperms were striking, emphasising the close relatedness of these fungi. Despite this correspondence, two CAZyme families were identified as different amongst the two groups of *Fusarium* species (GH43 and GH109). These genes differ in ancestral origins, where GH43 have originated within the *Fusarium oxysporum* species complex (FOSC), and GH109 originated within the FOSC+FFSC. We found that these two genes were possibly acquired through horizontal gene transfer.

INTRODUCTION

Some fungi employ cell-wall degrading enzymes, often referred to as Carbohydrate-Active enZymes (CAZymes), to penetrate host cells for nutrients (Jones & Dangl, 2006, Cantarel *et al.*, 2009, Borah *et al.*, 2018, Dracatos *et al.*, 2018). The secretion of these enzymes is especially prevalent in fungi lacking penetration structures. CAZymes permits the degradation of host macromolecules and play a part in the host-pathogen interaction. For example, these enzymes degrade host macromolecules to release nutrients for the pathogen (Brunner *et al.*, 2013). Host cell walls are enriched in carbon (in the form of carbohydrates) and nitrogen and provide the pathogen with a sustainable diet to develop to its full potential (Lowe *et al.*, 2015). Pathogens require carbon as a source of cellular energy and cell wall remodelling, and nitrogen for the synthesis of proteins and nucleic acids. Several CAZymes, such as pectinases (Babalola, 2010, Walter *et al.*, 2010), xylanases (Husaini *et al.*, 2018, Yang *et al.*, 2018, Yu *et al.*, 2018) and cellulases (Babalola, 2010, Quoc & Chau, 2017, Husaini *et al.*, 2018) are frequently involved in the host-pathogen interaction. Thus, the nutritional and survival needs of a pathogen are, therefore, dependant on a successful interaction between the host and a pathogen.

The diverse ecological functions associated with CAZymes allow fungi to adapt to a specific lifestyle (Zhao *et al.*, 2013b). Alternatively, their diverse nature allows a lifestyle, which can either be pathogenic or endophytic (Zhao *et al.*, 2013b, Bashyal *et al.*, 2017, Brown *et al.*, 2017, Sista Kameshwar & Qin, 2018, Yang *et al.*, 2019, Roy *et al.*, 2020). The contrasting lifestyles employed by fungi can be brought about by differing circumstances. During evolution, additional genes that were acquired from either external (*e.g.*, horizontal gene transfer) (Coleman *et al.*, 2009, Ma *et al.*, 2010, Stewart *et al.*, 2014, Glenn *et al.*, 2016, Van Wyk *et al.*, 2018) or internal (*e.g.*, mutations, duplications and translocations) (Coleman *et al.*, 2009, De Vos *et al.*, 2014, King *et al.*, 2015) processes possibly enabled various *Fusarium* species to contain a more extensive arsenal of genes. Differential expression of genes is also known to play a role in the ability of fungi to colonise different hosts, as observed for *F. graminearum* (Brown *et al.*, 2017).

The role of CAZymes has been investigated in different *Fusarium* species, such as *F. graminearum* (Zhao *et al.*, 2013b), *F. fujikuroi* (Bashyal *et al.*, 2017), *F. oxysporum*

(Chang *et al.*, 2016) and *Nectria haematococca* (*F. solani*) (Coleman *et al.*, 2009), amongst others. Identified CAZyme-encoding genes from these *Fusarium* species indicated that many of these genes are up-regulated during host-pathogen interactions, emphasising the importance of these enzymes during initial infection towards successful pathogen colonisation. Genomic analyses have shown that fungi are enriched for a variety of CAZyme families belonging to the glycoside hydrolases (GHs), carbohydrate esterases (CEs) and polysaccharide lyases (PLs) classes (Zhao *et al.*, 2013b, Kubicek *et al.*, 2014). For example, in *F. fujikuroi*, these enzymes promote the colonisation of root tissue and the degradation of plant biomass (Bashyal *et al.*, 2017). In *F. graminearum* these enzymes induce the degradation of plant cell walls and host cell death (Brown *et al.*, 2010, Brown *et al.*, 2012), whereas these enzymes promote successful host invasion in *F. oxysporum* (Roy *et al.*, 2020). These studies emphasise the role of CAZymes in overcoming the host's defence responses, occupying a niche, which will, in turn, promote the survival and reproduction of fungi.

CAZymes are considered CWDE (cell wall degrading enzymes), which contribute to the degradation of cellulose, hemicellulose and pectin. Fungi are also more likely to produce a bigger arsenal of CAZymes at later stages of host infection (Gibson *et al.*, 2011). The type and quantity of CAZymes produced are dependent on the host plant since the arsenal of these enzymes will differ for monocots and dicots (Cuomo *et al.*, 2007, King *et al.*, 2011). CWDEs can be divided into enzymes focussing on fungal cell wall degradation (FCWDEs) versus plant cell wall degradation (PCWDEs). The FCWDEs are more focussed on the fungal interaction between the environment and a host, whereas PCWDEs focuses on the degradation of plant cell wall components (Coutinho *et al.*, 2009, Battaglia *et al.*, 2011).

Gymnosperms and monocots have plant cell walls that differ significantly in their composition and structure and have evolved differently to protect them from pathogen attack (Malinovsky *et al.*, 2014, Houston *et al.*, 2016, Bacete *et al.*, 2018). Plant cell walls consist of a range of components, which include cellulose, non-cellulose and pectic polysaccharides, proteins, phenolic compounds and water (Rose *et al.*, 2004, Cosgrove, 2005, Houston *et al.*, 2016). Cell walls of grass monocots and gymnosperms differ in their amounts of non-cellulose polysaccharides and their associations/linkages (Carpita,

1996, Vogel, 2008). For example, primary cell walls in grasses are more abundant in hemicellulose, glucuronoarabinoxylan and glucan, and less abundant in xyloglucan, pectin and structural proteins (*e.g.*, arabinogalactan). In contrast, primary cell walls in gymnosperms are highly abundant in mannan and glucomannan as the predominant hemicelluloses, along with high abundances of pectin and structural proteins (Vogel, 2008). Secondary cell walls differ from primary cell walls in terms of structure and composition, typically containing more cellulose and less pectin, with the predominant non-cellulose components in secondary cell walls of grasses being xylan and lignin. The major non-cellulose components of secondary cell walls in gymnosperms are lignin and hemicellulose, with galactomannan and glucomannan as the more prominent hemicelluloses in these cell walls, in addition to xylan (Aspinall, 1980, Capek *et al.*, 2002, Pauly & Keegstra, 2008). Due to these differences in primary and secondary cell walls, the expectation is that differing CAZymes would be necessitated by fungal pathogens attacking these dissimilar plant hosts.

To complement the study into host-specificity, investigations into CAZymes were undertaken. This study aimed to implement comparative genomics to illustrate the difference in the CAZyme repertoire between six *Fusarium* species, all belonging to the American clade of the *Fusarium fujikuroi* species complex (FFSC). These *Fusarium* species associate either with pine trees or members of Poaceae. The results will provide insight into the carbohydrate degradation and utilisation potential of these two groups of fungi. The correlation between the diversity and specificity of CAZyme families between the *Fusarium* species associated with gymnosperms and monocots will be established. Genes that differ between two CAZyme families will be considered at a phylogenetic level. This is the first study focussing on CAZymes encoded by these *Fusarium* species. The results from this study will contribute to the current understanding of how these enzymes play a role in host colonisation between gymnosperms and monocots.

MATERIALS AND METHODS

CAZyme identification and distribution across the relevant *Fusarium* species

A total of six *Fusarium* genomes were used in this study. Three of these genomes were from *Fusarium* species associated with *Pinus* species [*F. circinatum* FSP34 isolated from *P. radiata* (Gordon *et al.*, 1996, De Vos *et al.*, 2020, unpublished), *F. fracticaudum* CMW 25245 isolated from *P. maximinoi* (Herron *et al.*, 2015, Wingfield *et al.*, 2018a) and *F. pininemorale* CMW 25243 isolated from *P. tecunumanii* (Herron *et al.*, 2015, Wingfield *et al.*, 2017)]. The remaining three genomes were from *Fusarium* species associated with members of the Poaceae family [*F. konzum* NRRL 11616 isolated from prairie grass (Zeller *et al.*, 2003), *F. subglutinans* NRRL 22016 isolated from maize (Desjardins *et al.*, 2006) and *F. temperatum* CMW 40964 isolated from teosinte (Desjardins *et al.*, 2000, Wingfield *et al.*, 2015b)]. The sequencing of the *F. subglutinans* NRRL 22016 and *F. konzum* NRRL 11616 genomes were performed by the Illumina Mi-Seq at Peoria (Illinois) with one 500 bp paired-end library (unpublished data, kindly provided by Robert H. Proctor).

The protein sequences predicted for the genes of each of the six *Fusarium* genomes were analysed for CAZymes identified by the DataBase for automated Carbohydrate-active enzyme Annotation (dbCAN2, http://bcb.unl.edu/dbCAN2/) (Zhang *et al.*, 2018). The putative CAZyme genes were grouped into different classes, based on their catalytic activities (Ferreira Filho *et al.*, 2017). The Carbohydrate-Active enZymes (CAZy) database currently covers five CAZyme classes, each containing multiple families (*i.e.*, GH43, GH109) that are classified based on amino acid sequences and structural similarities (Lombard *et al.*, 2014). These classes include glycoside hydrolases (GHs) that hydrolyse glycosidic bonds and degrade plant biomass (Berlemont & Martiny, 2016), glycosyltransferases (GTs) that form glycosidic bonds, resulting in the biosynthesis of oligosaccharides, polysaccharides and glycoconjugates (Breton *et al.*, 2005), polysaccharide lyases (PLs) that cleave uronic acid-containing polysaccharide chains into unsaturated polysaccharides (Lombard *et al.*, 2010), carbohydrate esterases (CEs) involved in the catalysation of de-*O* or de-*N*-acylation to remove esters of substituted saccharides (Cantarel *et al.*, 2009) and the auxiliary activities (AAs) that depolymerise

non-carbohydrate structural components (Rytioja *et al.*, 2014). Lastly, the carbohydrate-binding modules (CBMs) are not considered an official enzyme class but consist of polypeptides that adhere to carbohydrates (Boraston *et al.*, 2004).

The CAZyme-encoding genes will be compared with the unique genes identified. This comparison will allow for the identification of any unique genes encoding for CAZymes which aid in the host-specificity of these *Fusarium* species. Furthermore, the chromosomal location of the unique CAZyme-encoding genes will also be investigated.

Hierarchical clustering was performed using ClustVis to visualise the clustering of multi-variate data and to visualise groups of species with similar enzymatic profiles (Metsalu & Vilo, 2015). The Euclidean distance was used as the distance metric and average lineage clustering as the linkage method for the columns. In contrast, the rows were correlated, and complete linkage was applied. A heatmap was also created using ClustVis, where CAZyme families with the same number of genes in all species were automatically removed.

Availability of EST data

Expressed Sequence Tag (EST) data for *F. circinatum* FSP34 genes were implemented from previous studies. These include genes expressed during carbon and nitrogen starvation (Wingfield *et al.*, 2012) and genes expressed on half-strength potato dextrose broth at room temperature (Van Wyk *et al.*, 2019). These results were compared with the CAZyme genes in *F. circinatum* to identify which genes are transcribed. All data were analysed in CLC Genomics Workbench v. 11. Genes with RPKM (Reads Per Kilobase per Million mapped reads) values ≥ 0.2 and at least three unique gene reads mapping to it were considered transcribed (Wickramasinghe *et al.*, 2012).

CAZyme family-based phylogeny of selected genes

CAZyme families with constant (similar for all species in a group) and differing (different between the two groups) gene counts were selected and studied on a phylogenetic level. To infer phylogenies, all sequences were aligned with MAFFT (Multiple sequence

Alignment based on Fast Fourier Transform) v. 7.0 with default settings (Katoh *et al.*, 2017). MEGA v. 7.0.26 (Kumar *et al.*, 2016) was used to draw trees using the Maximum Likelihood branch support using bootstrap analyses based on 100 pseudoreplicates.

RESULTS

CAZyme identification and characterisation

Many putative CAZymes with similarity towards those previously characterised from the five main CAZyme classes, along with the carbohydrate-binding modules, have been identified in the six *Fusarium* genomes (Figure 1, Table 1, Supplemental Table 1-6). These included genes from 302-342 GHs, 107-115 GTs, 112-134 CEs, 113-121 AAs, 23-26 PLs families and 17-22 CBMs. The percentage of genes encoding for CAZymes in each genome was relatively similar, ranging from 4.40% of genes in *F. circinatum* compared to the 4.63% of genes in *F. fracticaudum* (Table 2).

Amongst all CAZyme classes, the GHs class was the most prevalent in all six genomes, followed by the classes CEs, AAs, GTs, PLs and CBMs (Figure 1). GH was also the class that had the most variation amongst the six *Fusarium* species (Table 2). All CAZyme families shared similar profiles amongst the six *Fusarium* species (Table 1). *Fusarium fracticaudum* was the most prevalent genome in the GHs class, having 19 GH-encoding genes more than the next prevalent genome, *F. konzum* (Figure 1). This increase extended over numerous GHs families, so it was not due to an increase in a specific family. *Fusarium fracticaudum* was also the most prevalent in the GTs and PLs classes, which accounts for the larger proportion of its genome encoding for CAZymes (Table 2), compared to the other *Fusarium* species. Conversely, the increase in GTs within *F. fracticaudum* was in one family (GT1) compared to the other *Fusarium* species. Similarly, the increase in PLs was due to the presence of one PL11_2 not found in any of the other *Fusarium* species.

CAZyme genes potentially involved in fungal cell wall degradation and plant cell wall degradation have also been identified (Zhao *et al.*, 2013b, Kubicek *et al.*, 2014). Potential

FCWDEs (between 63-70 total genes in each genome) involved in the degradation of β-glycan includes the GH16, GH17, GH55, GH71, GH72 and GH81 families. Families GH18, GH20, GH75 and GH76 involve the degradation of chitin. The genes in families GH16 and GH18 were much more abundant compared to the other families involved in FCWDE. The PCWDEs (between 166-184 total genes in each genome) included CAZyme families playing a role in pectin degradation (*e.g.*, rhamnogalacturonan exolyase, PL26; β-galactosidases, GH2; galacturonases, GH28; endo-arabinases, GH43; α-arabinofuranosidases, GH51; endo-β-1,4-galactanases, GH53; α-L-rhamnosidases, GH78; exo-α-L-1,5-arabinanases, GH93 and pectin methylesterase, CE8), cellulose degradation (*e.g.*, GH5, GH12 and GH45, β-1,4-endoglucanases; GH6 and GH7, endoglucanases/cellobiohydrolases; GH1 and GH3, β-glucosidases) and hemicellulose degradation (*e.g.*, GH1 and GH3, β-1,4-glucosidases; GH2, β-1,4-mannosidases; GH10 and GH11, β-1,4-endoxylanases; GH3 and GH43, β-1,4-xylosidases; GH5, β-1,4-endomannanases; CE3, CE4 and CE5, acetylxylan esterases; GH3, GH10, GH43 and GH51, α-L-arabinofuranosidases and GH74, β-1,4-endoglucanases). Genes from families AA3 (cellobiose dehydrogenase, glucose 1-oxidase, aryl alcohol oxidase, alcohol oxidase and pyranose oxidase) and AA7 (glucooligosaccharide oxidase) are involved in the degradation of lignin. Genes within families AA7 and GH3 were much more abundant compared to genes from families CE3, CE4, CE5, GH2 and GH78 which was less abundant but still more abundant compared to the rest of the PCWDEs.

CAZyme families that are present in all ascomycetes were also present in this study (Zhao *et al.*, 2013). These families include GH2, GH72, GH76, GH5, GH79, CE3 and CE5. The number of genes within these families displayed similar patterns to other CAZymes, with either the same number for each genome or slight variation amongst the genomes.

Distribution of CAZymes across the relevant *Fusarium* species

Putative CAZyme-encoding genes belonging to specific classes were found to be dispersed amongst chromosomes and were not necessarily located on the same chromosome (Supplemental Table 1-3, 6). Graphical distribution of genes in some CAZyme families is given in Figure 2A-F. The abundance of the different enzymes within a family was represented by a colour scale from blue to red. The darker the blue,

the more depleted the enzymes are in a specific family, while a darker red represented an enrichment of enzymes in a specific family. *Fusarium circinatum* was generally depleted in AAs and CBMs compared to the rest of the genomes (Figure 2A and 2B). There was an enrichment in AA1_2 and AA3 for *F. subglutinans* and *F. temperatum,* while *F. subglutinans* was enriched in AA1 and *F. konzum* enriched in AA1_3, AA3_2, AA4 and AA7, both species as outliers in comparison to the other Poaceae-associates. An enrichment for CBM24 was observed in *F. subglutinans* and *F. konzum,* while *F. pininemorale, F. subglutinans* and *F. temperatum* were enriched for CBM32, CBM1 and CBM67, respectively, as outliers in comparison to the remaining pine- and Poaceae-associates. *Fusarium temperatum* was generally depleted in CEs, except for CE1, which was enriched in the genomes of the Poaceae-associated *Fusarium* species (Figure 2C). The two pine-associates (*F. fracticaudum* and *F. pininemorale*) were generally more enriched in CE3, CE4, CE5 and CE10. *Fusarium konzum* and *F. subglutinans* were enriched in CE5 as outliers towards *F. temperatum.* The genome of *F. fracticaudum* was generally more enriched in PLs compared to the rest of the genomes (Figure 2D). Figure 2E indicated the severe depletion of GH5_22, GH10, GH12, GH16, GH28, GH29, GH43_29, GH43_36, GH54, GH67, GH78, GH114 and GH142, which clustered in *F. circinatum*. These genes are likely involved in PCWD (GH10, GH12, GH28, GH29, GH43_29, GH43_36, GH54, GH67, GH78, GH114 and GH142) and FCWD (GH5_22 and GH16). In *F. fracticaudum,* GH35, GH39, GH43_1, GH43_14, GH106, GH127 and GH145 was enriched and clustered, where these genes participate in PCWD. The Poaceae-associated *Fusarium* species were depleted in GH3, GH5, GH20, GH36, GH43_13, GH88 and GH106, compared to the remaining genomes. A clear difference in the CAZyme repertoire existed in GH43 and GH109 between the pine- and Poaceae-associated *Fusarium* species. Furthermore, compared to the rest of the genomes, a depletion of GT8, GT17 and GT64 was observed in *F. circinatum, F. temperatum* and *F. fracticaudum,* respectively (Figure 2F). In contrast, the latter genome was highly enriched in GT1 compared to the remaining genomes.

Availability of EST data

Expression data for the genes of the pine pathogen, *F. circinatum,* were obtained from previous studies (Wingfield *et al.*, 2012, Van Wyk *et al.*, 2019) (Supplemental Table 7).

The combined results indicated that 647 of the 665 putative CAZyme genes (97.29%) were transcribed under the described conditions. Overall, the results indicated that a vast majority the unique genes from the pine-associated *Fusarium* species are likely actively transcribed.

CAZyme family-based phylogeny of selected genes

The results from Table 1 and Figure 2E highlighted two CAZyme families (GH43 and GH109) that differ in genic repertoire between the two host groups. The putative CAZyme-encoding genes from this study correlated with three of the unique genes. Of the 72 unique genes from the pine-associated *Fusarium* species, two genes have been identified to encode for the CE10 and GH43 CAZyme families. The single gene (FCIR_3_gene_4.30, FFRAC_3_gene_4.125 and FPIN_3_gene_4.79) within the CE10 family were located on chromosome 3 in all three *Fusarium* species associated with pine. Similarly, the gene (FCIR_6_gene_2.142, FFRAC_6_gene_2.15 and FPIN_6_gene_2.27) within the GH43 family were located on chromosome 6. The only CAZyme family identified in the 47 unique genes from the Poaceae-associated *Fusarium* species was a CE3 which was located on chromosome 1 of *F. temperatum* (FTEMP_1_gene_3.65). The location of this gene in *F. konzum* and *F. subglutinans* could not be traced back to a specific chromosome due to the fragmented condition of these genomes.

The potential phylogenetic origin of these genes was demonstrated (Figure 3 and 4). A classification system for ancestral origins was constructed. The ancestral origin for the GH43 gene falls in group 3 of this system, meaning that this gene has likely originated from the *Fusarium oxysporum* species complex (FOSC). In contrast, the GH109 gene falls into group 4 of the classification system, indicating that this gene likely originated from the broader FFSC and FOSC, but was only retained in specific lineages (Figure 3 and 4).

DISCUSSION

This study comprehensively investigates the CAZyme repertoire of *Fusarium* species associated with *Pinus* species and members of Poaceae. CAZymes analyses found remarkable similarities between all the species investigated, with similar proportions of the genome encoded by CAZymes. CAZymes were found to be distributed across all chromosomes and were generally not localised to specific chromosomes. Furthermore, only two CAZyme families (GH43 and GH109) showed differences between the *Fusarium* species associated with the two different host groups. Phylogenetic analyses revealed that these genes were likely as a result of gene loss and/or gain.

The results of this study demonstrated that these species encode for more CAZymes compared to other fungi and *Fusarium* species. The six *Fusarium* species consist of 683-754 CAZyme-encoding genes, which differed from other plant pathogens, such as *Sclerotinia sclerotiorum* (415 genes) and *Botrytis cinerea* (441 genes), as well as from non-pathogenic fungi, such as *Neurospora crassa* (322 genes) and *Aspergillus niger* (439 genes) (Coutinho *et al.*, 2009, Amselem *et al.*, 2011). *Fusarium* species, such as *F. graminearum* (biotroph) and *F. fujikuroi* (necrotroph) contain 109 and 356 CAZyme-encoding genes, respectively (Bashyal *et al.*, 2017, Brown *et al.*, 2017). The number of CAZymes per class might differ between different pathogens depending on the host and pathogen lifestyle (Zhao *et al.*, 2013b, Kubicek *et al.*, 2014). *Fusarium circinatum* is considered a hemibiotroph (Martin-Rodrigues *et al.*, 2013, Swett *et al.*, 2016, Swett & Gordon, 2017, Swett *et al.*, 2018), *F. subglutinans* and *F. temperatum* are necrotrophs (Desjardins *et al.*, 2000, Scauflaire *et al.*, 2011) and *F. konzum* is an endophyte on prairie grass (Zeller *et al.*, 2003). The pathogenic status linked to *F. fracticaudum* and *F. pininemorale* is not yet known. Further research can improve current knowledge of how different lifestyles affect CAZyme gene expression, especially for those pathogens colonising one host as an endophyte and another as a pathogen.

CAZymes vary slightly in distribution and abundance amongst the *Fusarium* species of this study. For example, it was observed that *F. subglutinans* (AA1), *F. pininemorale* (CBM32 and GH43_28) and *F. fracticaudum* (PL11_2) contained a single gene in the specific families and was highly enriched in the respective genomes but these genes were

absent in all the other genomes from this study. The presence of laccase (p-diphenol oxygen oxidoreductase) (AA1) is expected since this enzyme catalyses the oxidative gelation of feruloylated arabinoxylans (hemicellulosic polysaccharides covalently linked to each other and lignin) upon the dimerization of their ferulic esters (Rosell & Marco, 2008). Arabinoxylans are present in the cell walls of monocots, hence the need for this specific enzyme (Hatfield *et al.*, 2017). CBM32 binds to galactose and lactose as a form of nutrient acquisition for the pathogen (*e.g.*, *F. pininemorale*) and can also bind to polygalacturonic acid for the breakdown of plant tissues into simpler molecules which may also serve as a source of nutrients to a pathogen (Newstead *et al.*, 2005, Abbott *et al.*, 2007). In addition, β-xylosidase (GH43_28) participates in the biodegradation of xylan, which is also the main hemicellulose component in the secondary cell walls of gymnosperms (*F. pininemorale*) (Aspinall, 1980, Capek *et al.*, 2002, Pauly & Keegstra, 2008). Rhamnogalacturonan endolyase (PL11_2) degrades the rhamnogalacturonan from plant cell walls to releases oligosaccharides, such as sucrose, lactose and maltose, which may act as an additional nutrient source for *F. fracticaudum* (Ochiai *et al.*, 2006). The absence of the genes mentioned in this paragraph possibly emphasises the diverse ecological needs and adaptation of different fungal species.

Genes from CBM38, GH43_29 and GH142 were absent in *F. circinatum* while being present in the remaining genomes. The CBM38 enzymes bind inulin that promotes the storage of carbohydrates (Lee *et al.*, 2004). Inulin reduces the energy density of food and is used to enrich food with dietary fibre or to replace sugar and fat (Lee *et al.*, 2004). The absence of this enzyme in *F. circinatum* suggests that this pathogen can survive in the absence of this enzyme and uses sugars other than inulin as sources of energy. The enzyme β-xylosidase, encoded by members of the GH43 family, are important in fungi and bacteria for providing thermostability, increased specificity and more tolerance to xylose (Zhao *et al.*, 2013a, Mustafa *et al.*, 2016). *Fusarium circinatum* does not necessarily require thermostability as much as increased specificity and tolerance to xylose. The absence of β-xylosidase (GH43_29) from *F. circinatum* suggests that this pathogen uses other β-xylosidases from different CAZyme families to degrade xylan and overcome the pine host's defence response (secretion of xylose). Furthermore, β-L-arabinofuranosidase (GH142) degrade arabinoxylans and arabinogalactans, which are present in the cell walls of monocots (Hatfield *et al.*, 2017). The absence of GH142

probably would not significantly affect pathogens associated with gymnosperms. The minor gains and losses of CBM38, GH43_29 and GH142 genes do not seem to be involved with host-specificity, as their gain/loss are not uniformly distributed.

Many putative GH families were identified from this study and are predicted to have activities in cell wall degradation (*i.e.*, PCWDE and FCWDE) (Van den Brink & De Vries, 2011, Kubicek *et al.*, 2014). This could potentially be due to the multifunctionality of these enzymes, as well as the variation in function and substrate specificity of enzymes in the GHs class (Zhao *et al.*, 2013b, Kubicek *et al.*, 2014). Several GH families from this study have been identified from previous literature to be involved in host cell wall modification, plant biomass degradation and fungal virulence (McDonough *et al.*, 2004, Kubicek *et al.*, 2014, Bashyal *et al.*, 2017, Brown *et al.*, 2017).

A large group of putative PCWDEs, potentially targeting pectin, have been identified in the six *Fusarium* genomes. Polygalacturonase (GH28) synthesised by fungi are considered important for pectin degradation and host infection. In this study, all six genomes contained multiple copies of these genes (8-10). No correlation could be made with regards to differences in these genes between the *Fusarium* species associated with the two host groups. Some CAZyme families involved in pectin degradation, such as PL1, PL3, PL4 and GH13, were absent from the genomes investigated. However, there are subfamilies in these categories and the enzymes in these subfamilies have activities which have the potential to replace these missing enzyme families (Zhao *et al.*, 2013b, Kubicek *et al.*, 2014).

Some putative PCWDEs have been associated with the degradation of celluloses and were mainly classified into three groups (*i.e.*, β-1,4-endoglucanases, exoglucanases/cellobiohydrolases and β-glucosidases). These enzymes are all important for the internal/external cleavage of a cellulose chain and to convert cellobiose to glucose monomers, respectively (Vlasenko *et al.*, 2010, Van den Brink & De Vries, 2011). The CAZyme families involved in cellulose degradation consist of 1-26 copies of these genes, depending on the family. Differences in this study, in comparison to other studies (Amselem *et al.*, 2011, Gibson *et al.*, 2011, Van den Brink & De Vries, 2011, Kubicek *et al.*, 2014), emphasised the differences amongst the species in the genes involved in the

enzymatic activity of cellulose degradation. This emphasises that pathogens adopt a lifestyle and synthesise proteins promoting this lifestyle.

The *Fusarium* species from this study had some CAZyme profiles potentially involved in the degradation of hemicellulose. After cellulose, hemicellulose is the second most abundant polysaccharide in plants. Hemicellulose consists of a variety of polysaccharides, including xylan, xyloglucan, glucogalactomannan and galactan (Amselem *et al.*, 2011, Kubicek *et al.*, 2014). The *Fusarium* species from this study had similar β-1,4-endoxylanases (GH10 and GH11) profiles compared to other pathogens (*e.g., Sclerotinia sclerotiorum, Verticillium dahlia, F. verticillioides* and *F. graminearum*) (Kubicek *et al.*, 2014), which suggest a role for these enzymes during host colonisation for a wide range of fungi.

The putative CAZymes involved in the degradation of fungal cell walls are predicted to participate in various metabolic pathways and processes to modify fungal cell walls (Zhao *et al.*, 2013b, Kubicek *et al.*, 2014). These enzymes have diverse functions, such as providing protection to cells during host infection and inferring pathogenicity and virulence. The β-glucanases from GH16 and chitinases from GH18 were the most abundant amongst the other identified FCWDEs. This correlates with literature since fungal cell walls are enriched in polysaccharides such as glucan and chitin (Karlsson & Stenlid, 2009, Mouyna *et al.*, 2013, Kang *et al.*, 2019). FCWDEs play a crucial part in host-pathogen interactions since components of the fungal cell wall trigger host defence responses upon recognition by host immune cells (Geoghegan *et al.*, 2017).

In the current study, investigations focussed into the two classes of CAZymes (GH43 and GH109) that varied between the two groups of investigated *Fusarium* genomes. The GH43 class is prevalent in most plant pathogens and plays a crucial role in biomass degradation (Kohler *et al.*, 2015). Compared to other pathogens, the *Fusarium* species from this study contain very few genes (1-2). For example, *N. haematococca, S. sclerotiorum, V. dahlia, F. graminearum* and *F. verticillioides* contain 32, 21, 22, 19 and 22 genes, respectively (Kubicek *et al.*, 2014). Pathogens associated with gymnosperms, such as *Dichomitus squalens* and *Heterobasidion annosum* contain less GH43 genes (7 and 5 genes, respectively), which is still more than what was found from this study

(Kubicek *et al.*, 2014). The GH43 family consists of different enzymatic activities, and some are known to play roles in host-pathogen interactions (Knoch *et al.*, 2014). The second CAZyme class, GH109, plays a role in the hydrolysis of the terminal α-*O*-glycoside-bonded residues of *N*-acetylgalactosamine (NAG) from the non-reducing ends of various complex carbohydrates and glycoconjugates (Bakunina *et al.*, 2013). This enzyme has some degree of specificity towards substrates, such as glycolipids, glycopeptides, and glycoproteins containing the structures with the *O*-glycoside core, and also oligosaccharides. These enzymes promote the integrity of cell structures, energy storage, pathogen defence, viral penetration, cellular signalling, fertilization, development of carcinomas, inflammatory events and lysosomal storage diseases, which all contribute to the physiological and pathological processes of organisms (Intra *et al.*, 2008).

The potential ancestral origins of the GH43 and GH109 genes identified in this study differed markedly. The evolution of these genes is likely as a result of gene gains and/or losses. The pine-associated *Fusarium* genes belonging to the GH43 family likely lost genes with ancestral origins in the FFSC or acquired genes with ancestral origins in the FOSC, over time. If these genes were lost, it could be suggested that these genes were not selected for anymore, however, if these genes were gained, then these genes were possibly selected for. If the GH43 genes were gained, then it emphasises the need of the pathogen to synthesise the enzymes encoded by these genes in order to adopt a lifestyle. In contrast, the GH109 genes still share origins in the FFSC and FOSC (Figure 3 and 4). It remains unsure whether these origins were as a result of gene gains. These two genes also probably co-evolved over time. Genes that co-evolved and selected for can generate biological diversity, as observed for the pine- and Poaceae-associated *Fusarium* species containing these GH43 and GH109 genes.

CONCLUSIONS

Comparative analysis of the six *Fusarium* species identified the genes encoding for gene products involved in plant cell wall alteration, penetration and degradation of and plant

biomass, which in turn, may significantly impact fungal virulence. CAZymes utilise plant material as a source of nutrition to a pathogen. Differences in the number of genes, as well as the presence/absence of particular CAZyme families, are possibly linked to the lifestyle to be maintained by a pathogen.

In this study, the *in silico* CAZyme analyses alone was not enough to determine the role these enzymes play in host-specificity. Downstream applications can be implemented in this study, such as the integration of RNA sequencing in order to understand the expression of genes at different stages of infection, but also to determine gene regulation and the prediction of biological processes encoded by the respective genes.

REFERENCES

1. Abbott DW, Hrynuik S & Boraston AB (2007) Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. *Journal of Molecular Biology* **367**: 1023-1033.

2. Amselem J, Cuomo CA, Van Kan JA, Viaud M, Benito EP, Couloux A, *et al.* (2011) Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genetics* **7**: e1002230.

3. Aspinall G (1980) Chemistry of cell wall polysaccharides. *The biochemistry of plants: A comprehensive treatise,* (Priess J, ed.) pp. 477-500. Academic Press, New York.

4. Babalola OO (2010) Pectinolytic and cellulolytic enzymes enhance *Fusarium compactum* virulence on tubercles infection of Egyptian broomrape. *International Journal of Microbiology* **2010**: 273264.

5. Bacete L, Mélida H, Miedes E & Molina A (2018) Plant cell wall-mediated immunity: Cell wall changes trigger disease resistance responses. *The Plant Journal* **93**: 614-636.
6. Bakunina I, Nedashkovskaya O, Balabanova L, Zvyagintseva T, Rasskasov V & Mikhailov V (2013) Comparative analysis of glycoside hydrolases activities from phylogenetically diverse marine bacteria of the genus Arenibacter. *Marine Drugs* **11**: 1977-1998.

7. Bashyal BM, Rawat K, Sharma S, Kulshreshtha D, Gopala Krishnan S, Singh AK, *et al.* (2017) Whole genome sequencing of *Fusarium fujikuroi* provides insight into the role of secretory proteins and cell wall degrading enzymes in causing bakanae disease of rice. *Frontiers in Plant Science* **8**: 2013.

8. Battaglia E, Benoit I, Van den Brink J, Wiebenga A, Coutinho PM, Henrissat B, *et al.* (2011) Carbohydrate-active enzymes from the zygomycete fungus *Rhizopus*

oryzae: A highly specialized approach to carbohydrate degradation depicted at genome level. *BMC Genomics* **12**: 38.

9. Berlemont R & Martiny AC (2016) Glycoside hydrolases across environmental microbial communities. *PLoS Computational Biology* **12**: e1005300.

10. Borah N, Albarouki E & Schirawski J (2018) Comparative methods for molecular determination of host-specificity factors in plant-pathogenic fungi. *International Journal of Molecular Science* **19**: E863.

11. Boraston AB, Bolam DN, Gilbert HJ & Davies GJ (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochemical Journal* **382**: 769-781.

12. Breton C, Šnajdrová L, Jeanneau C, Koča J & Imberty A (2005) Structures and mechanisms of glycosyltransferases. *Glycobiology* **16**: R29-R37.

13. Brown NA, Antoniw J & Hammond-Kosack KE (2012) The predicted secretome of the plant pathogenic fungus *Fusarium graminearum*: A refined comparative analysis. *PLoS One* **7**: e33731.

14. Brown NA, Evans J, Mead A & Hammond-Kosack KE (2017) A spatial temporal analysis of the *Fusarium graminearum* transcriptome during symptomless and symptomatic wheat infection. *Molecular Plant Pathology* **18**: 1295-1312.

15. Brown NA, Urban M, Van de Meene AM & Hammond-Kosack KE (2010) The infection biology of *Fusarium graminearum*: Defining the pathways of spikelet to spikelet colonisation in wheat ears. *Fungal Biology* **114**: 555-571.

16. Brunner PC, Torriani SFF, Croll D, Stukenbrock EH & McDonald BA (2013) Coevolution and life cycle specialization of plant cell wall degrading enzymes in a hemibiotrophic pathogen. *Molecular Biology and Evolution* **30**: 1337-1347.

17. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V & Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Research* **37**: D233-D238.

18. Capek P, Alfoldi J & Liskova D (2002) An acetylated galactoglucomannan from *Picea abies* L. Karst. *Carbohydrate Research* **337**: 1033-1037.

19. Carpita NC (1996) Structure and biogenesis of the cell walls of grasses. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 445-476.

20. Chang HX, Yendrek CR, Caetano-Anolles G & Hartman GL (2016) Genomic characterization of plant cell wall degrading enzymes and in silico analysis of xylanses and polygalacturonases of *Fusarium virguliforme*. *BMC Microbiology* **16**: 147.

21. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, Grimwood J, *et al.* (2009) The genome of *Nectria haematococca*: Contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* **5**: e1000618.

22. Cosgrove DJ (2005) Growth of the plant cell wall. *Nature Reviews Molecular Cell Biology* **6**: 850-861.

23. Coutinho PM, Andersen MR, Kolenova K, Van Kuyk PA, Benoit I, Gruben BS, *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.

24. Cuomo CA, Güldener U, Xu J, Trail F, Turgeon BG, Di Pietro A, *et al.* (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400-1402.

25. De Vos L, Steenkamp ET, Martin SH, Santana QC, Fourie G, Van der Merwe NA, *et al.* (2014) Genome-wide macrosynteny among *Fusarium* species in the *Gibberella*

fujikuroi complex revealed by amplified fragment length polymorphisms. *PLoS One* **9**: e114682.

26. De Vos L, Van der Nest MA, Santana QC, Wingfield BD & Steenkamp ET (2020, unpublished) Enhancement of *Fusarium circinatum* genomic assemblies. *BioTechniques*.

27. Desjardins AE, Maragos CM & Proctor RH (2006) Maize ear rot and moniliformin contamination by cryptic species of *Fusarium subglutinans*. *Journal of Agricultural and Food Chemistry* **54**: 7383.

28. Desjardins AE, Plattner RD & Gordon TR (2000) *Gibberella fujikuroi* mating population A and *Fusarium subglutinans* from teosinte species and maize from Mexico and Central America. *Mycological Research* **104**: 865-872.

29. Dracatos PM, Haghdoust R, Singh D & Park RF (2018) Exploring and exploiting the boundaries of host specificity using the cereal rust and mildew models. *New Phytologist* **218**: 453-462.

30. Ferreira Filho JA, Horta MAC, Beloti LL, Dos Santos CA & De Souza AP (2017) Carbohydrate-active enzymes in *Trichoderma harzianum*: A bioinformatic analysis bioprospecting for key enzymes for the biofuels industry. *BMC Genomics* **18**: 779.

31. Geoghegan I, Steinberg G & Gurr S (2017) The role of the fungal cell wall in the infection of plants. *Trends in Microbiology* **25**: 957-967.

32. Gibson DM, King BC, Hayes ML & Bergstrom GC (2011) Plant pathogens as a source of diverse enzymes for lignocellulose digestion. *Current Opinion in Microbiology* **14**: 264-270.

33. Glenn AE, Davis CB, Gao M, Gold SE, Mitchell TR, Proctor RH, *et al.* (2016) Two horizontally transferred xenobiotic resistance gene clusters associated with detoxification of benzoxazolinones by *Fusarium* species. *PLoS One* **11**: e0147486.

34. Gordon TR, Storer AJ & Okamoto D (1996) Population structure of the pitch canker pathogen, *Fusarium subglutinans* f. sp. *pini*, in California. *Mycological Research* **100**: 850-854.

35. Hatfield RD, Rancour DM & Marita JM (2017) Grass cell walls: A story of cross-linking. *Frontiers in Plant Science* **7**: 2056.

36. Herron D, Wingfield MJ, Wingfield B, Rodas CA, Marincowitz S & Steenkamp E (2015) Novel taxa in the *Fusarium fujikuroi* species complex from *Pinus spp*. *Studies in Mycology* **80**: 131-150.

37. Houston K, Tucker MR, Chowdhury J, Shirley N & Little A (2016) The plant cell wall: A complex and dynamic structure as revealed by the responses of genes under stress conditions. *Frontiers in Plant Science* **7**: 984.

38. Husaini AM, Sakina A & Cambay SR (2018) Host-Pathogen interaction in *Fusarium oxysporum* infections: Where do we stand? *Molecular Plant-Microbe Interactions* **31**: 889-898.

39. Intra J, Pavesi G & Horner DS (2008) Phylogenetic analyses suggest multiple changes of substrate specificity within the glycosyl hydrolase 20 family. *BMC Evolutionary Biology* **8**: 214.

40. Jones JD & Dangl JL (2006) The plant immune system. *Nature* **444**: 323-329.

41. Kang L, Zhu Y, Bai Y & Yuan S (2019) Characteristics, transcriptional patterns and possible physiological significance of glycoside hydrolase family 16 members in *Coprinopsis cinerea*. *FEMS Microbiology Letters* **366**: p.nfz083.

42. Karlsson M & Stenlid J (2009) Evolution of family 18 glycoside hydrolases: Diversity, domain structures and phylogenetic relationships. *Journal of Molecular Microbiology and Biotechnology* **16**: 208-223.

43. Katoh K, Rozewicki J & Yamada KD (2017) MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**: 1160-1166.

44. King BC, Waxman KD, Nenni NV, Walker LP, Bergstrom GC & Gibson DM (2011) Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. *Biotechnology for Biofuels* **4**: 4.

45. King R, Urban M, Hammond-Kosack MCU, Hassani-Pak K & Hammond-Kosack KE (2015) The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* **16**: 544.

46. Knoch E, Dilokpimol A & Geshi N (2014) Arabinogalactan proteins: Focus on carbohydrate active enzymes. *Frontiers in Plant Science* **5**: 198.

47. Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, *et al.* (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics* **47**: 410-415.

48. Kubicek CP, Starr TL & Glass NL (2014) Plant cell wall–degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* **52**: 427-451.

49. Kumar S, Stecher G & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**: 1870-1874.

50. Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**: 1307-1320.

51. Lee J, Kim K & Choi Y (2004) Identification and characterization of a novel inulin binding module (IBM) from the CFTase of *Bacillus macerans* CFC1. *FEMS Microbiology Letters* **234**: 105-110.

52. Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM & Henrissat B (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal* **432**: 437-444.

53. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research* **42**: D490-D495.

54. Lowe RGT, McCorkelle O, Bleackley M, Collins C, Faou P, Mathivanan S, *et al.* (2015) Extracellular peptidases of the cereal pathogen *Fusarium graminearum*. *Frontiers in Plant Science* **6**: 962.

55. Ma L, Van der Does HC, Borkovich KA, Coleman JJ, Daboussi MJ, Di Pietro A, *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367-373.

56. Malinovsky FG, Fangel JU & Willats WGT (2014) The role of the cell wall in plant immunity. *Frontiers in Plant Science* **5**: 178.

57. Martin-Rodrigues N, Espinel S, Sanchez-Zabala J, Ortiz A, Gonzalez-Murua C & Dunabeitia MK (2013) Spatial and temporal dynamics of the colonization of *Pinus radiata* by *Fusarium circinatum*, of conidiophora development in the pith and of traumatic resin duct formation. *New Phytologist* **198**: 1215-1227.

58. McDonough MA, Kadirvelraj R, Harris P, Poulsen J & Larsen S (2004) Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4. *FEBS Letters* **565**: 188-194.

59. Metsalu T & Vilo J (2015) ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research* **43**: W566-W570.

60. Mouyna I, Hartl L & Latgé J (2013) β-1,3-glucan modifying enzymes in *Aspergillus fumigatus*. *Frontiers in Microbiology* **4**: 81.

61. Mustafa G, Kousar S, Rajoka MI & Jamil A (2016) Molecular cloning and comparative sequence analysis of fungal β-Xylosidases. *AMB Express* **6**: 30.

62. Newstead SL, Watson JN, Bennet AJ & Taylor G (2005) Galactose recognition by the carbohydrate-binding module of a bacterial sialidase. *Acta Crystallographica, Section D, Biologucal Crystallogrophy* **61**: 1483-1491.

63. Ochiai A, Yamasaki M, Itoh T, Mikami B, Hashimoto W & Murata K (2006) Crystallization and preliminary X-ray analysis of the rhamnogalacturonan lyase YesW from *Bacillus subtilis* strain 168, a member of polysaccharide lyase family 11. *Acta Crystallographica, Section F, Structural Biology and Crystallization Communications* **62**: 438-440.

64. Pauly M & Keegstra K (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant Journal* **54**: 559-568.

65. Quoc NB & Chau NNB (2017) The role of cell wall degrading enzymes in pathogenesis of *Magnaporthe oryzae*. *Current Protein and Peptide Science* **18**: 1019-1034.

66. Rose JKC, Saladié M & Catalá C (2004) The plot thickens: New perspectives of primary cell wall modification. *Current Opinion in Plant Biology* **7**: 296-301.

67. Rosell CM & Marco C (2008) 4 - Rice. *Gluten-Free Cereal Products and Beverages*, Vol. II-III (Arendt EK & Dal Bello F, eds.), pp. 81-100. Academic Press, San Diego.

68. Roy A, Jayaprakash A, Rajeswary T R, Annamalai A & Lakshmi PTV (2020) Genome-wide annotation, comparison and functional genomics of carbohydrate-active enzymes in legumes infecting *Fusarium oxysporum formae speciales*. *Mycology* **11**:56-70.

69. Rytioja J, Hildén K, Yuzon J, Hatakka A, De Vries RP & Mäkelä MR (2014) Plant-polysaccharide-degrading enzymes from basidiomycetes. *Microbiology and Molecular Biology Reviews* **78**: 614-649.

70. Scauflaire J, Gourgue M & Munaut F (2011) *Fusarium temperatum* sp. nov. from maize, an emergent species closely related to *Fusarium subglutinans*. *Mycologia* **103**: 586-597.

71. Sista Kameshwar AK & Qin W (2018) Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. *Mycology* **9**: 93-105.

72. Stewart J, Abdo Z & Glenn A (2014) Gene clusters FDB1 and FDB2 in *Fusarium verticillioides* were acquired through multiple horizontal gene transfer events. *Phytopathology* **104**: 114.

73. Swett CL & Gordon TR (2017) Exposure to a pine pathogen enhances growth and disease resistance in *Pinus radiata* seedlings. *Forest Pathology* **47**: 1-10.

74. Swett CL, Kirkpatrick SC & Gordon TR (2016) Evidence for a hemibiotrophic association of the pitch canker pathogen *Fusarium circinatum* with *Pinus radiata*. *Plant Disease* **100**: 79-84.

75. Swett CL, Reynolds GJ & Gordon TR (2018) Infection without wounding and symptomless shoot colonization of *Pinus radiata* by *Fusarium circinatum*, the cause of pitch canker. *Forest Pathology* **48**: e12422.

76. Van den Brink J & De Vries RP (2011) Fungal enzyme sets for plant polysaccharide degradation. *Applied Microbiology and Biotechnology* **91**: 1477-1492.

77. Van Wyk S, Wingfield BD, De Vos L, Santana QC, Van der Merwe NA & Steenkamp ET (2018) Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*. *IMA Fungus* **9**: 27-36.

78. Van Wyk S, Wingfield BD, De Vos L, Van der Merwe NA, Santana QC & Steenkamp ET (2019) Repeat-induced point mutations drive divergence between *Fusarium circinatum* and its close relatives. *Pathogens* **8**: E298.

79. Vlasenko E, Schülein M, Cherry J & Xu F (2010) Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. *Bioresource Technology* **101**: 2405-2411.

80. Vogel J (2008) Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* **11**: 301-307.

81. Walter S, Nicholson P & Doohan FM (2010) Action and reaction of host and pathogen during *Fusarium* head blight disease. *New Phytologist* **185**: 54-66.

82. Wickramasinghe S, Rincon G, Islas-Trejo A & Medrano JF (2012) Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* **13**: 45.

83. Wingfield BD, Barnes I, De Beer ZW, De Vos L, Duong TA, Kanzi AM, *et al.* (2015b) Draft genome sequences of *Ceratocystis eucalypticola*, *Chrysoporthe cubensis*, *C. deuterocubensis*, *Davidsoniella virescens*, *Fusarium temperatum*, *Graphilbum fragrans*, *Penicillium nordicum*, and *Thielaviopsis musarum*. *IMA Fungus* **6**: 493-506.

84. Wingfield BD, Berger DK, Steenkamp ET, Lim HJ, Duong TA, Bluhm BH, *et al.* (2017) Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. *IMA Fungus* **8**: 385-396.

85. Wingfield BD, Bills GF, Dong Y, Huang W, Nel WJ, Swalarsk-Parry BS, *et al.* (2018a) Draft genome sequence of *Annulohypoxylon stygium*, *Aspergillus mulundensis*, *Berkeleyomyces basicola* (syn. *Thielaviopsis basicola*), *Ceratocystis smalleyi*, two *Cercospora beticola* strains, *Coleophoma cylindrospora*, *Fusarium fracticaudum*, *Phialophora* cf. *hyalina*, and *Morchella septimelata*. *IMA Fungus* **9**: 199-223.

86. Wingfield BD, Steenkamp ET, Santana Q, Coetzee M & Bam S (2012) First fungal genome sequence from Africa: A preliminary analysis. *South Africa Journal of Science* **108**: 104-112.

87. Yang Y, Liu X, Cai J, Chen Y, Li B, Guo Z, *et al.* (2019) Genomic characteristics and comparative genomics analysis of the endophytic fungus *Sarocladium brachiariae*. *BMC Genomics* **20**: 782.

88. Yang Y, Yang X, Dong Y & Qiu D (2018) The *Botrytis cinerea* xylanase BcXyl1 modulates plant immunity. *Frontiers in Microbiology* **9**: 02535.

89. Yu C, Li T, Shi X, Saleem M, Li B, Liang W, *et al.* (2018) Deletion of endo-β-1,4-xylanase *VmXyl1* impacts the virulence of valsa mali in apple tree. *Frontiers in Plant Science* **9**: 663-663.

90. Zeller KA, Summerell BA, Bullock S & Leslie JF (2003) *Gibberella konza* (*Fusarium konzum*) sp. nov., a new biological species within the *Gibberella fujikuroi* species complex from prairie grasses. *Mycologia* **95**: 943-954.

91. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, *et al.* (2018) dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **46**: W95-W101.

92. Zhao L, Xie J, Zhang X, Cao F & Pei J (2013a) Overexpression and characterization of a glucose-tolerant β-glucosidase from *Thermotoga thermarum* DSM 5069T with high catalytic efficiency of ginsenoside Rb1 to Rd. *Journal of Molecular Catalysis B: Enzymatic* **95**: 62-69.

93. Zhao Z, Liu H, Wang C & Xu J (2013b) Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* **14**: 274-274.

TABLES

Table 1. CAZyme identification and gene count.

Table 2. Statistics for the genes encoding for CAZymes.

Table 1. CAZyme gene identification and count. Genes highlighted in yellow represent FCWDES, green represent PCWDEs. Red blocks identify differing CAZyme repertoire between the species from each host group.

	FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]		FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]		FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]
AUXILIARY ACTIVITIES (AAs)																				
AA1	0	0	0	0	1	0	**AA2**	4	4	4	4	4	4	**AA5**	1	1	1	1	1	1
AA1_2	3	3	4	3	4	4	**AA3**	8	10	10	9	11	11	**AA5_2**	4	4	4	4	4	4
AA1_3	7	7	7	8	7	6	**AA3_1**	5	5	5	5	5	5	**AA6**	1	1	1	1	1	1
AA11	4	4	4	4	4	4	**AA3_2**	12	13	13	13	11	12	**AA7**	41	41	42	43	40	38
AA12	2	3	3	3	3	3	**AA3_3**	2	2	2	2	2	2	**AA9**	13	13	13	13	13	13
AA13	1	1	1	1	1	1	**AA3_4**	1	1	1	1	1	1							
AA14	1	1	1	1	1	1	**AA4**	3	5	4	5	4	4							
CARBOHYDRATE-BINDING MODULES (CBMs)																				
CBM1	1	0	1	1	2	1	**CBM32**	0	0	1	0	0	0	**CBM43**	1	1	1	1	1	1
CBM20	2	2	2	2	2	2	**CBM35**	1	1	1	1	1	1	**CBM6**	1	1	1	1	1	1
CBM21	1	1	1	1	1	1	**CBM38**	0	1	1	1	1	1	**CBM63**	2	2	2	2	2	2
CBM24	1	1	1	2	2	1	**CBM42**	2	3	3	3	3	3	**CBM67**	5	7	5	6	6	7
CARBOHYDRATE ESTERASES (CEs)																				
CE1	4	5	5	6	6	6	**CE2**	1	1	1	1	1	1	**CE8**	4	4	4	4	4	4
CE10	73	80	74	81	70	64	**CE3**	8	11	11	10	10	8	**CE9**	1	1	1	1	1	1
CE12	4	4	4	4	4	4	**CE4**	11	10	11	10	10	9							
CE16	6	6	6	6	6	6	**CE5**	9	10	11	11	11	9							
GLYCOSIDE HYDROLASES (GHs)																				
GH1	5	6	4	5	6	5	**GH28**	8	10	10	10	9	9	**GH5_16**	1	1	1	1	1	1
GH10	3	5	5	5	5	5	**GH29**	2	3	3	3	3	3	**GH5_22**	1	2	2	2	2	2
GH105	4	4	4	4	4	4	**GH3**	25	26	25	23	22	24	**GH5_23**	2	2	2	2	2	2
GH106	1	2	1	1	0	1	**GH30_7**	3	3	3	3	3	3	**GH5_24**	1	1	1	1	1	1
GH109	2	2	2	1	1	1	**GH31**	8	9	8	10	9	9	**GH5_27**	1	1	1	1	1	1
GH11	3	3	3	3	3	3	**GH32**	8	7	8	7	8	8	**GH5_31**	2	2	2	2	2	2

Table 1 (continued). CAZyme gene identification and count. Genes highlighted in yellow represent FCWDES, green represent PCWDEs. Red blocks identify differing CAZyme repertoire between the species from each host group.

	FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]		FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]		FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]
GLYCOSIDE HYDROLASES (GHs)																				
GH115	2	2	2	2	2	2	**GH35**	5	8	3	4	3	3	**GH5_7**	2	2	2	2	2	2
GH12	4	6	5	5	5	5	**GH36**	3	3	3	3	2	2	**GH5_9**	1	1	1	1	1	1
GH125	3	3	3	3	3	3	**GH37**	2	2	2	2	2	2	**GH51**	2	2	2	2	2	3
GH127	2	4	1	1	1	1	**GH38**	1	1	1	1	1	1	**GH53**	1	1	1	1	1	1
GH128	4	4	4	5	4	4	**GH39**	0	3	1	1	0	0	**GH54**	2	3	3	3	3	3
GH13_1	1	1	1	1	1	1	**GH43**	2	2	2	1	1	1	**GH55**	2	2	3	3	3	2
GH13_25	1	1	1	1	1	1	**GH43_1**	2	3	2	2	2	2	**GH6**	1	1	1	1	1	1
GH13_40	8	7	7	9	7	7	**GH43_11**	2	2	2	2	2	2	**GH62**	1	1	1	1	1	1
GH13_8	1	1	1	1	1	1	**GH43_13**	2	2	2	1	2	2	**GH63**	1	1	1	1	1	1
GH131	1	1	1	1	1	1	**GH43_14**	2	4	2	2	2	2	**GH64**	2	2	2	2	2	2
GH132	2	2	2	2	2	2	**GH43_21**	1	1	1	1	1	1	**GH65**	1	1	1	1	1	1
GH134	1	1	1	1	1	1	**GH43_22**	3	4	4	4	4	3	**GH67**	2	3	3	3	3	3
GH139	1	1	1	1	1	1	**GH43_24**	3	3	3	3	3	3	**GH7**	3	3	3	3	3	3
GH141	1	2	1	2	2	2	**GH43_26**	2	2	2	2	2	2	**GH71**	2	2	2	2	2	2
GH142	0	1	1	1	1	1	**GH43_28**	0	0	1	0	0	0	**GH72**	3	3	3	3	3	3
GH145	2	3	2	2	2	2	**GH43_29**	0	1	1	1	1	1	**GH74**	1	1	1	1	1	1
GH146	1	1	1	1	1	1	**GH43_30**	3	3	3	3	3	3	**GH75**	2	2	2	2	2	2
GH15	3	3	3	3	3	3	**GH43_33**	1	1	1	1	1	1	**GH76**	9	9	9	9	9	9
GH152	1	1	1	1	1	1	**GH43_36**	1	4	3	3	3	3	**GH78**	9	11	10	11	10	10
GH16	23	28	25	25	26	26	**GH43_5**	1	1	1	1	1	1	**GH79**	1	2	2	2	1	1
GH17	5	4	4	4	5	4	**GH45**	1	1	1	1	1	1	**GH81**	3	2	3	3	3	3
GH18	17	16	19	20	22	18	**GH47**	10	10	10	10	10	10	**GH88**	3	3	3	2	3	3
GH2	10	12	11	11	11	9	**GH49**	1	1	1	1	1	1	**GH93**	6	6	6	6	6	5
GH20	3	3	3	3	2	3	**GH5**	1	1	1	1	0	0	**GH95**	2	3	2	2	3	2

Table 1 (continued). CAZyme gene identification and count. Genes highlighted in yellow represent FCWDES, green represent PCWDEs. Red blocks identify differing CAZyme repertoire between the species from each host group.

	FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]		FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP		FCIR[1]	FFRAC[1]	FPIN[1]	FKON[1]	FSUB[1]	FTEMP[1]
GH24	1	1	1	1	1	1	**GH5_12**	2	2	2	2	2	2							
GH27	1	1	2	1	1	2	**GH5_15**	3	3	3	3	3	3							
GLYCOSYLTRANSFERASES (GTs)																				
GT1	11	18	13	14	13	13	**GT3**	1	1	1	1	1	1	**GT58**	1	1	1	1	1	1
GT15	5	5	5	5	5	5	**GT32**	5	5	5	5	5	5	**GT59**	1	1	1	1	1	1
GT17	2	2	2	2	2	1	**GT33**	1	1	1	1	1	1	**GT62**	3	3	3	3	3	3
GT2 Chitin synth_1	5	5	5	5	5	5	**GT34**	3	3	3	3	3	3	**GT64**	3	2	3	3	3	3
GT2 Chitin synth_2	13	12	13	13	13	12	**GT35**	1	1	1	1	1	1	**GT66**	1	1	1	1	1	1
GT2 Glyco tranf_2_3	8	9	8	9	9	8	**GT39**	3	3	3	3	3	3	**GT69**	3	3	3	3	3	3
GT2 Glyco transf_2	2	2	2	2	2	2	**GT4**	5	5	5	5	5	5	**GT71**	3	3	3	3	3	3
GT20	3	3	3	3	3	3	**GT48**	1	1	1	1	1	1	**GT76**	1	1	1	1	1	1
GT21	1	1	1	1	1	1	**GT50**	1	1	1	1	1	1	**GT8**	5	7	7	7	7	7
GT22	4	4	4	4	4	4	**GT54**	1	1	1	1	1	1	**GT90**	7	7	6	6	7	7
GT24	1	1	1	1	1	1	**GT57**	2	2	2	2	2	2							
POLYSACCHARIDE LYASES (PLs)																				
PL1_10	1	1	1	1	1	1	**PL1_9**	1	1	1	1	1	1	**PL4_1**	1	1	1	1	1	1
PL1_2	1	1	1	1	1	0	**PL11_2**	0	1	0	0	0	0	**PL4_3**	2	2	2	2	2	2
PL1_4	5	7	6	6	6	6	**PL26**	1	1	1	1	1	1	**PL9_3**	2	2	2	2	2	2
PL1_7	3	3	3	3	3	3	**PL3_2**	6	6	6	6	6	6							

[1]**FCIR:** *Fusarium circinatum*; **FFRAC:** *Fusarium fracticaudum*; **FPIN:** *Fusarium pininemorale*; **FKON:** *F. konzum*; **FSUB:** *Fusarium subglutinans*; **FTEMP:** *Fusarium temperatum*.

Table 2. Statistics for the genes encoding for CAZymes.

***Fusarium* species**	**AA[1], %**	**CBM[2], %**	**CE[3], %**	**GH[4], %**	**GT[5], %**	**PL[6], %**	**Total, %**
F. circinatum FSP34 (CMW 51752)	0.73	0.11	0.78	1.95	0.69	0.15	4.40
F. fracticaudum CMW 25245	0.73	0.12	0.81	2.10	0.71	0.16	4.63
F. pininemorale CMW 25243	0.76	0.13	0.81	2.03	0.69	0.15	4.56
F. konzum NRRL 11616	0.74	0.13	0.82	1.98	0.69	0.15	4.51
F. subglutinans NRRL 22016	0.75	0.14	0.78	2.02	0.71	0.15	4.55
F. temperatum CMW 40964	0.74	0.14	0.72	2.01	0.70	0.15	4.46

[1]AA: Auxiliary activities.
[2]CBM: Carbohydrate-binding modules.
[3]CE: Carbohydrate esterases.
[4]GH: Glycoside hydrolases.
[5]GT: Glycosyltransferases.
[6]PL: Polysaccharide lyases.

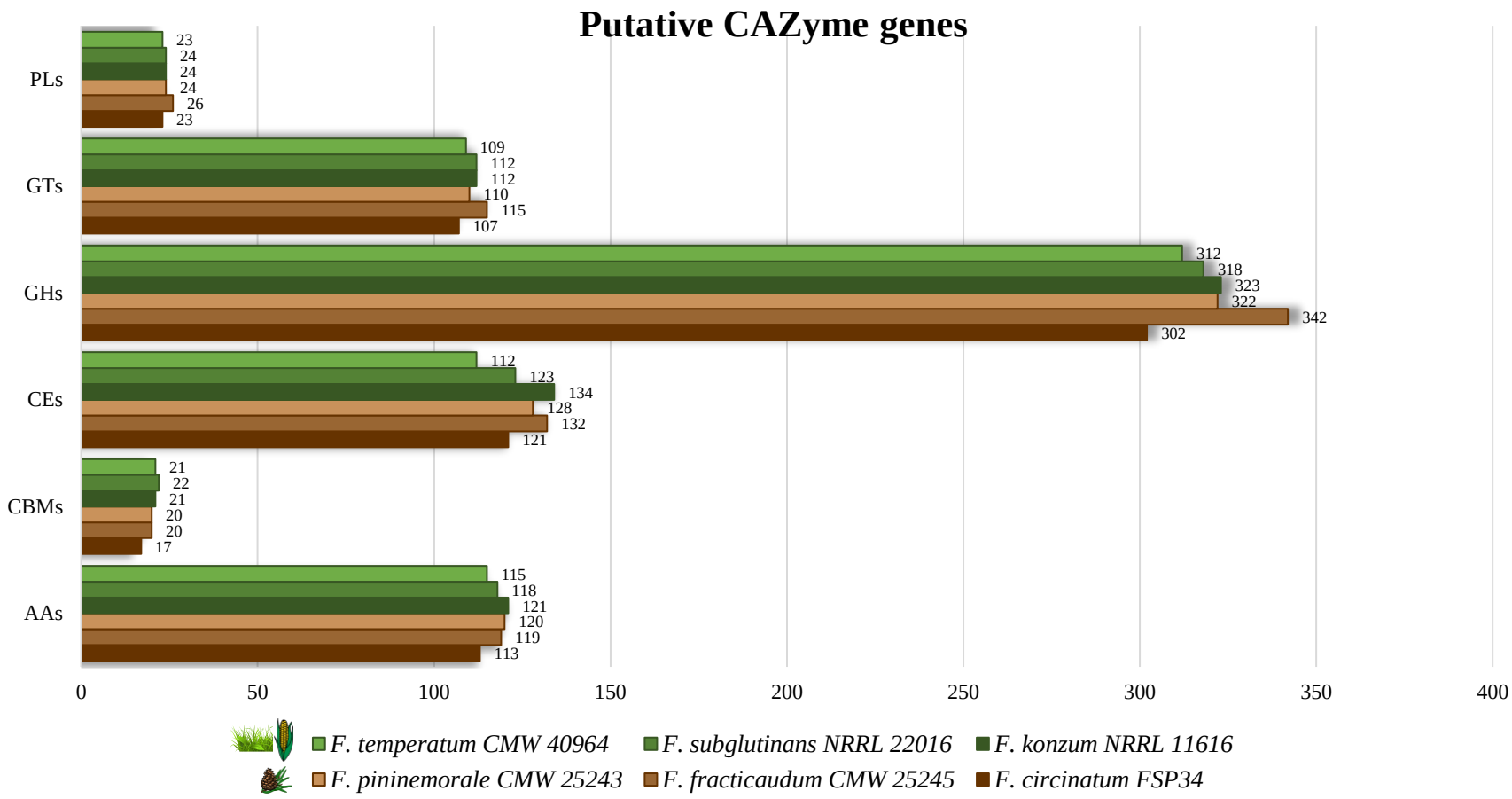
FIGURES

Figure 1. Identification of the number of putative CAZyme-encoding genes within each CAZyme class for all six *Fusarium* species.

Figure 2. Heatmaps generated by ClustVis to demonstrate gene distribution amongst the six *Fusarium* species.

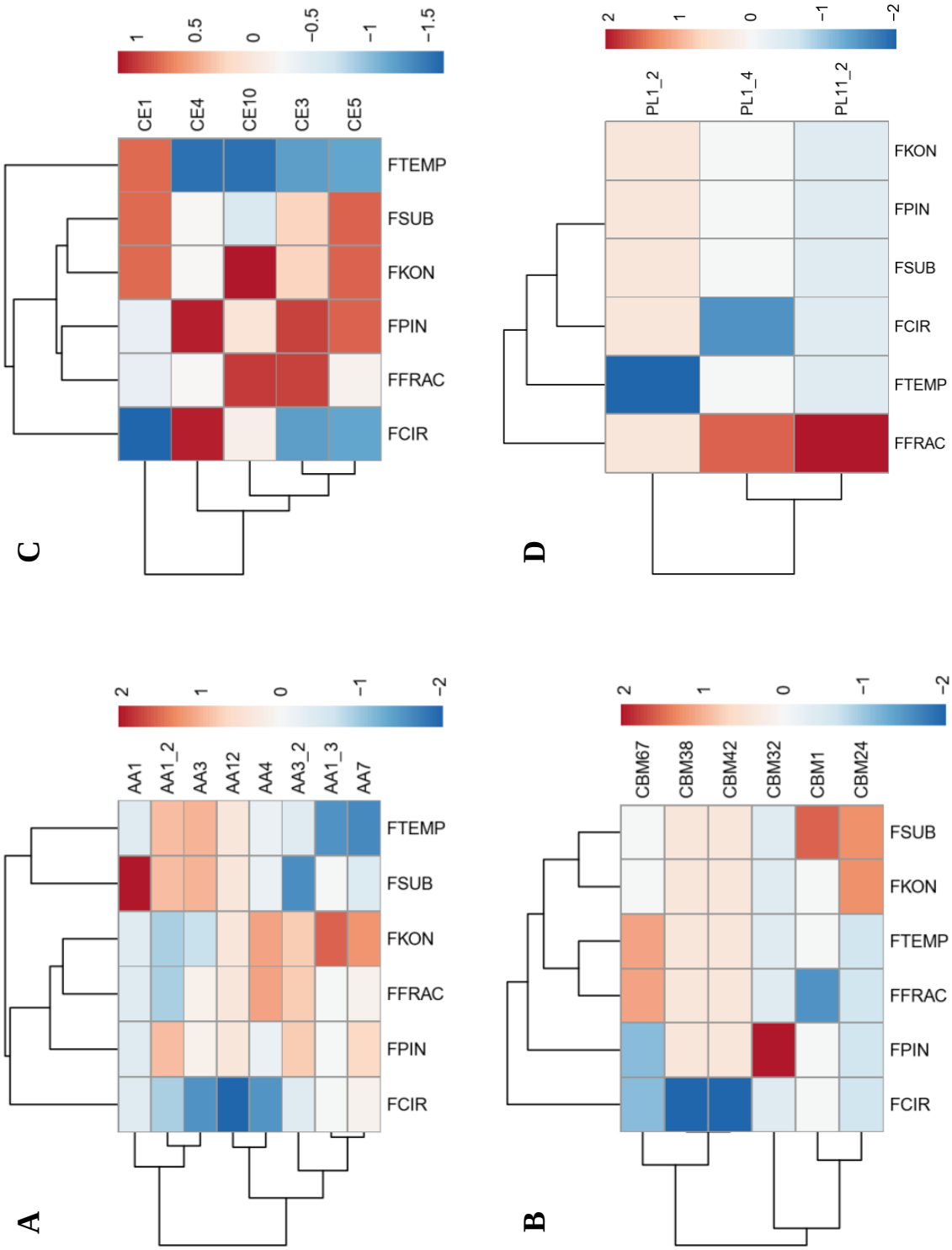
Figure 3. A representative phylogenetic tree for GH43.

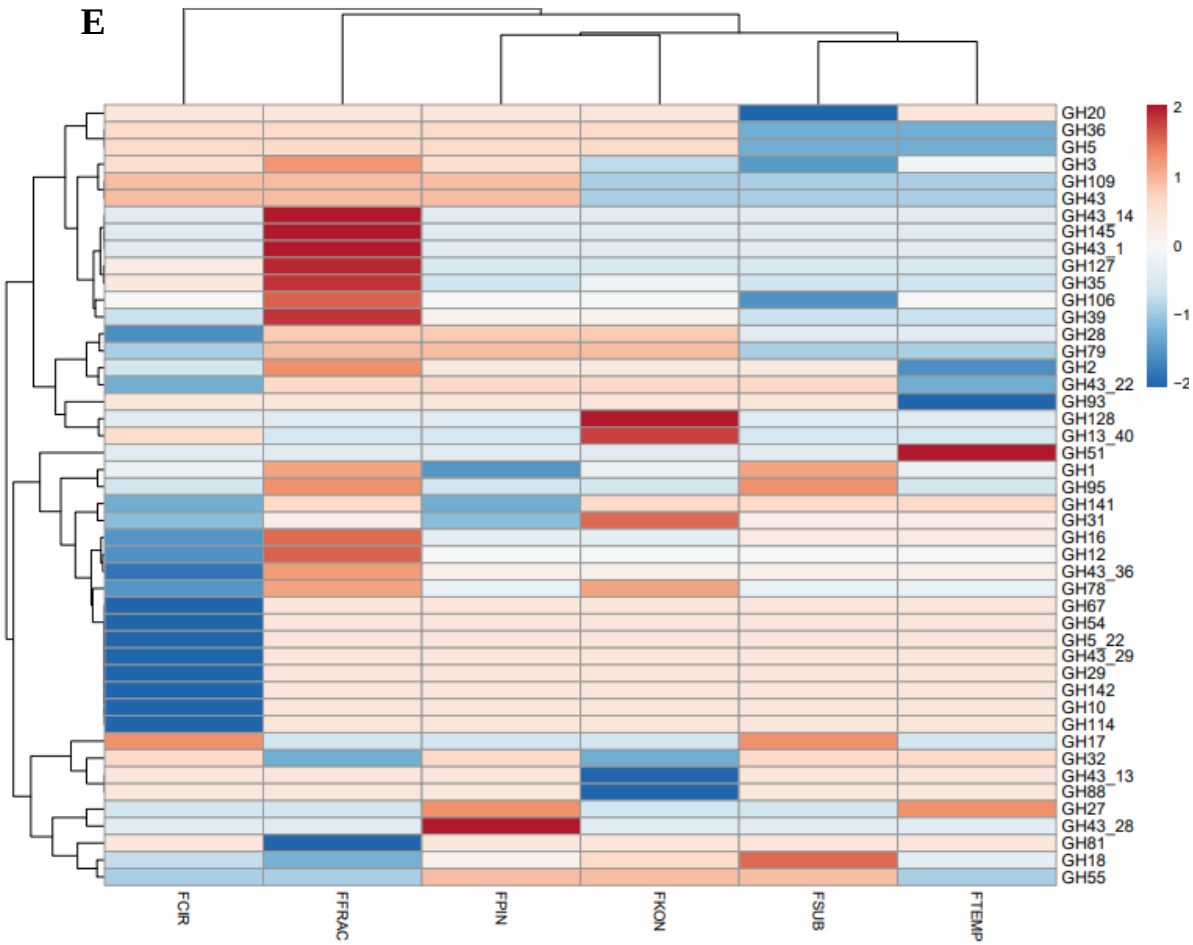
Figure 4. A representative phylogenetic tree GH109.



AAs: Auxiliary activities; CBMs: Carbohydrate-binding modules; CEs: Carbohydrate esterases; GHs: Glycoside hydrolases; GTs: Glycosyltransferases; PLs: Polysaccharide lyases

Figure 1. Identification of the number of putative CAZyme-encoding genes within each CAZyme class for all six *Fusarium* species.


A
2
1
0
−1
−2
AA1
AA1_2
AA3
AA12
AA4
AA3_2
AA1_3
AA7
FTEMP
FSUB
FKON
FFRAC
FPIN
FCIR
B
2
1
0
−1
−2
CBM67
CBM38
CBM42
CBM32
CBM1
CBM24
FSUB
FKON
FTEMP
FFRAC
FPIN
FCIR
C
1
0.5
0
−0.5
−1
−1.5
CE1
CE4
CE10
CE3
CE5
FTEMP
FSUB
FKON
FPIN
FFRAC
FCIR
D
2
1
0
−1
−2
PL1_2
PL1_4
PL11_2
FKON
FPIN
FSUB
FCIR
FTEMP
FFRAC


E
GH20
GH36
GH5
GH3
GH109
GH43
GH43_14
GH145
GH43_1
GH127
GH35
GH106
GH39
GH28
GH79
GH2
GH43_22
GH93
GH128
GH13_40
GH51
GH1
GH95
GH141
GH31
GH16
GH12
GH43_36
GH78
GH67
GH54
GH5_22
GH43_29
GH29
GH142
GH10
GH114
GH17
GH32
GH43_13
GH88
GH27
GH43_28
GH81
GH18
GH55
FCIR
FFRAC
FPIN
FKON
FSUB
FTEMP
2
1
0
−1
−2

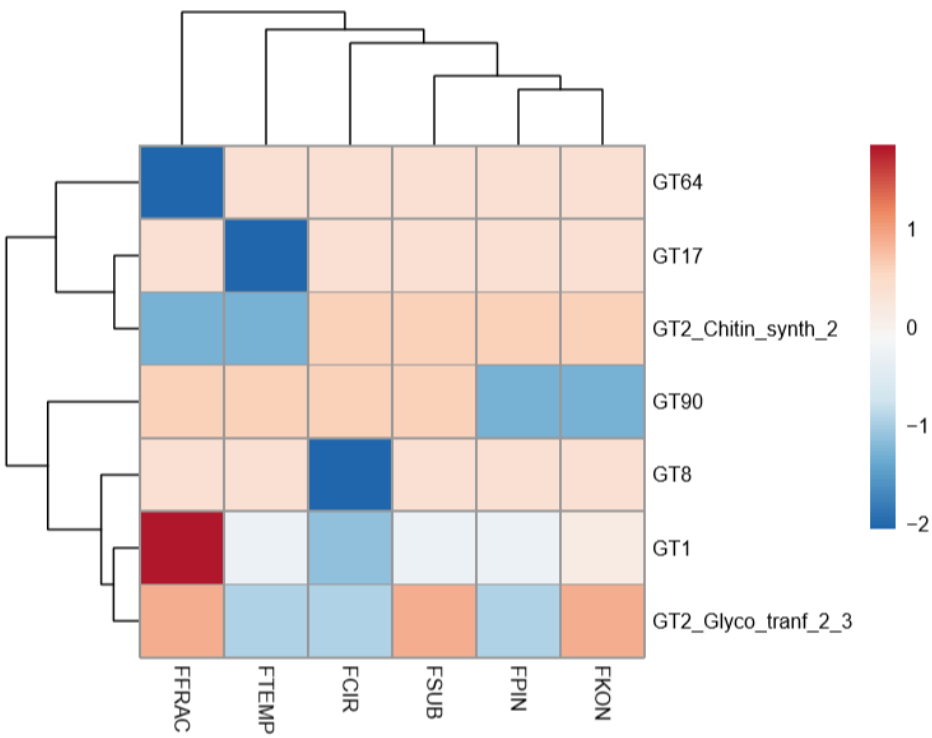

F
GT64
GT17
GT2_Chitin_synth_2
GT90
GT8
GT1
GT2_Glyco_tranf_2_3
FFRAC
FTEMP
FCIR
FSUB
FPIN
FKON
1
0
−1
−2

Figure 2. Heatmaps generated by ClustVis to demonstrate gene distribution among the six *Fusarium* species. The graphical representation follows for genes grouped in the A) Aas class, B) CBMs class, C) CEs class, D) PLs class, E) GHs class and F) GTs class, where red blocks indicate CAZymes that are highly enriched with genes compared to the depletion of families indicated in blue blocks. All heatmaps contain only the CAZyme families with genic variation between the six species. FCIR = *F. circinatum*; FPIN = *F. pininemorale*; FFRAC = *F. fracticaudum*; FKON = *F. konzum*; FSUB = *F. subglutinans*; FTEMP = *F. temperatum*.

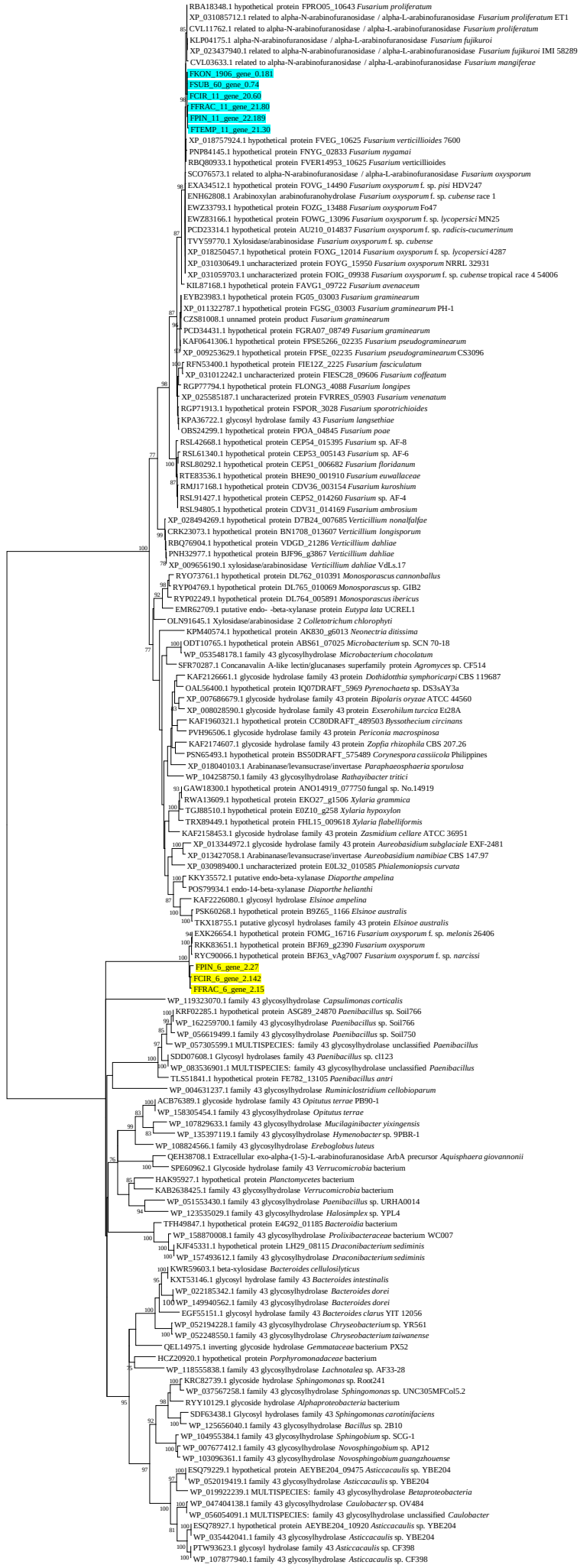

RBA18348.1 hypothetical protein FPRO05_10643 Fusarium proliferatum
XP_031085712.1 related to alpha-N-arabinofuranosidase / alpha-L-arabinofuranosidase Fusarium proliferatum ET1
CVL11762.1 related to alpha-N-arabinofuranosidase / alpha-L-arabinofuranosidase Fusarium proliferatum
KLP04175.1 alpha-N-arabinofuranosidase / alpha-L-arabinofuranosidase Fusarium fujikuroi
XP_023437940.1 related to alpha-N-arabinofuranosidase / alpha-L-arabinofuranosidase Fusarium fujikuroi IMI 58289
CVL03633.1 related to alpha-N-arabinofuranosidase / alpha-L-arabinofuranosidase Fusarium mangiferae
FKON_1906_gene_0.181
FSUB_60_gene_0.74
FCIR_11_gene_20.60
FFRAC_11_gene_21.80
FPIN_11_gene_22.189
FTEMP_11_gene_21.30
XP_018757924.1 hypothetical protein FVEG_10625 Fusarium verticillioides 7600
PNP84145.1 hypothetical protein FNYG_02833 Fusarium nygamai
RBQ80933.1 hypothetical protein FVER14953_10625 Fusarium verticillioides
SCO76573.1 related to alpha-N-arabinofuranosidase / alpha-L-arabinofuranosidase Fusarium oxysporum
EXA34512.1 hypothetical protein FOVG_14490 Fusarium oxysporum f. sp. pisi HDV247
ENH62808.1 Arabinoxylan arabinofuranohydrolase Fusarium oxysporum f. sp. cubense race 1
EWZ33793.1 hypothetical protein FOZG_13488 Fusarium oxysporum Fo47
EWZ83166.1 hypothetical protein FOWG_13096 Fusarium oxysporum f. sp. lycopersici MN25
PCD23314.1 hypothetical protein AU210_014837 Fusarium oxysporum f. sp. radicis-cucumerinum
TVY59770.1 Xylosidase/arabinosidase Fusarium oxysporum f. sp. cubense
XP_018250457.1 hypothetical protein FOXG_12014 Fusarium oxysporum f. sp. lycopersici 4287
XP_031030649.1 uncharacterized protein FOYG_15950 Fusarium oxysporum NRRL 32931
XP_031059703.1 uncharacterized protein FOIG_09938 Fusarium oxysporum f. sp. cubense tropical race 4 54006
KIL87168.1 hypothetical protein FAVG1_09722 Fusarium avenaceum
EYB23983.1 hypothetical protein FG05_03003 Fusarium graminearum
XP_011322787.1 hypothetical protein FGSG_03003 Fusarium graminearum PH-1
CZS81008.1 unnamed protein product Fusarium graminearum
PCD34431.1 hypothetical protein FGRA07_08749 Fusarium graminearum
KAF0641306.1 hypothetical protein FPSE5266_02235 Fusarium pseudograminearum
XP_009253629.1 hypothetical protein FPSE_02235 Fusarium pseudograminearum CS3096
RFN53400.1 hypothetical protein FIE12Z_2225 Fusarium fasciculatum
XP_031012242.1 uncharacterized protein FIESC28_09606 Fusarium coffeatum
RGP77794.1 hypothetical protein FLONG3_4088 Fusarium longipes
XP_025585187.1 uncharacterized protein FVRRES_05903 Fusarium venenatum
RGP71913.1 hypothetical protein FSPOR_3028 Fusarium sporotrichioides
KPA36722.1 glycosyl hydrolase family 43 Fusarium langsethiae
OBS24299.1 hypothetical protein FPOA_04845 Fusarium poae
RSL42668.1 hypothetical protein CEP54_015395 Fusarium sp. AF-8
RSL61340.1 hypothetical protein CEP53_005143 Fusarium sp. AF-6
RSL80292.1 hypothetical protein CEP51_006682 Fusarium floridanum
RTE83536.1 hypothetical protein BHE90_001910 Fusarium euwallaceae
RMJ17168.1 hypothetical protein CDV36_003154 Fusarium kuroshium
RSL91427.1 hypothetical protein CEP52_014260 Fusarium sp. AF-4
RSL94805.1 hypothetical protein CDV31_014169 Fusarium ambrosium
XP_028494269.1 hypothetical protein D7B24_007685 Verticillium nonalfalfae
CRK23073.1 hypothetical protein BN1708_013607 Verticillium longisporum
RBQ76904.1 hypothetical protein VDGD_21286 Verticillium dahliae
PNH32977.1 hypothetical protein BJF96_g3867 Verticillium dahliae
XP_009656190.1 xylosidase/arabinosidase Verticillium dahliae VdLs.17
RYO73761.1 hypothetical protein DL762_010391 Monosporascus cannonballus
RYP04769.1 hypothetical protein DL765_010069 Monosporascus sp. GIB2
RYP02249.1 hypothetical protein DL764_005891 Monosporascus ibericus
EMR62709.1 putative endo- -beta-xylanase protein Eutypa lata UCREL1
OLN91645.1 Xylosidase/arabinosidase 2 Colletotrichum chlorophyti
KPM40574.1 hypothetical protein AK830_g6013 Neonectria ditissima
ODT10765.1 hypothetical protein ABS61_07025 Microbacterium sp. SCN 70-18
WP_053548178.1 family 43 glycosylhydrolase Microbacterium chocolatum
SFR70287.1 Concanavalin A-like lectin/glucanases superfamily protein Agromyces sp. CF514
KAF2126661.1 glycoside hydrolase family 43 protein Dothidotthia symphoricarpi CBS 119687
OAL56400.1 hypothetical protein IQ07DRAFT_5969 Pyrenochaeta sp. DS3sAY3a
XP_007686679.1 glycoside hydrolase family 43 protein Bipolaris oryzae ATCC 44560
XP_008028590.1 glycoside hydrolase family 43 protein Exserohilum turcica Et28A
KAF1960321.1 hypothetical protein CC80DRAFT_489503 Byssothecium circinans
PVH96506.1 glycoside hydrolase family 43 protein Periconia macrospinosa
KAF2174607.1 glycoside hydrolase family 43 protein Zopfia rhizophila CBS 207.26
PSN65493.1 hypothetical protein BS50DRAFT_575489 Corynespora cassiicola Philippines
XP_018040103.1 Arabinanase/levansucrase/invertase Paraphaeosphaeria sporulosa
WP_104258750.1 family 43 glycosylhydrolase Rathayibacter tritici
GAW18300.1 hypothetical protein ANO14919_077750 fungal sp. No.14919
RWA13609.1 hypothetical protein EKO27_g1506 Xylaria grammica
TGJ88510.1 hypothetical protein E0Z10_g258 Xylaria hypoxylon
TRX89449.1 hypothetical protein FHL15_009618 Xylaria flabelliformis
KAF2158453.1 glycoside hydrolase family 43 protein Zasmidium cellare ATCC 36951
XP_013344972.1 glycoside hydrolase family 43 protein Aureobasidium subglaciale EXF-2481
XP_013427058.1 Arabinanase/levansucrase/invertase Aureobasidium namibiae CBS 147.97
XP_030989400.1 uncharacterized protein E0L32_010585 Phialemoniopsis curvata
KKY35572.1 putative endo-beta-xylanase Diaporthe ampelina
POS79934.1 endo-14-beta-xylanase Diaporthe helianthi
KAF2226080.1 glycosyl hydrolase Elsinoe ampelina
PSK60268.1 hypothetical protein B9Z65_1166 Elsinoe australis
TKX18755.1 putative glycosyl hydrolases family 43 protein Elsinoe australis
EXK26654.1 hypothetical protein FOMG_16716 Fusarium oxysporum f. sp. melonis 26406
RKK83651.1 hypothetical protein BFJ69_g2390 Fusarium oxysporum
RYC90066.1 hypothetical protein BFJ63_vAg7007 Fusarium oxysporum f. sp. narcissi
FPIN_6_gene_2.27
FCIR_6_gene_2.142
FFRAC_6_gene_2.15
WP_119323070.1 family 43 glycosylhydrolase Capsulimonas corticalis
KRF02285.1 hypothetical protein ASG89_24870 Paenibacillus sp. Soil766
WP_162259700.1 family 43 glycosylhydrolase Paenibacillus sp. Soil766
WP_056619499.1 family 43 glycosylhydrolase Paenibacillus sp. Soil750
WP_057305599.1 MULTISPECIES: family 43 glycosylhydrolase unclassified Paenibacillus
SDD07608.1 Glycosyl hydrolases family 43 Paenibacillus sp. cl123
WP_083536901.1 MULTISPECIES: family 43 glycosylhydrolase unclassified Paenibacillus
TLS51841.1 hypothetical protein FE782_13105 Paenibacillus antri
WP_004631237.1 family 43 glycosylhydrolase Ruminiclostridium cellobioparum
ACB76389.1 glycoside hydrolase family 43 Opitutus terrae PB90-1
WP_158305454.1 family 43 glycosylhydrolase Opitutus terrae
WP_107829633.1 family 43 glycosylhydrolase Mucilaginibacter yixingensis
WP_135397119.1 family 43 glycosylhydrolase Hymenobacter sp. 9PBR-1
WP_108824566.1 family 43 glycosylhydrolase Ereboglobus luteus
QEH38708.1 Extracellular exo-alpha-(1-5)-L-arabinofuranosidase ArbA precursor Aquisphaera giovannonii
SPE60962.1 Glycoside hydrolase family 43 Verrucomicrobia bacterium
HAK95927.1 hypothetical protein Planctomycetes bacterium
KAB2638425.1 family 43 glycosylhydrolase Verrucomicrobia bacterium
WP_051553430.1 family 43 glycosylhydrolase Paenibacillus sp. URHA0014
WP_123535029.1 family 43 glycosylhydrolase Halosimplex sp. YPL4
TFH49847.1 hypothetical protein E4G92_01185 Bacteroidia bacterium
WP_158870008.1 family 43 glycosylhydrolase Prolixibacteraceae bacterium WC007
KJF45331.1 hypothetical protein LH29_08115 Draconibacterium sediminis
WP_157493612.1 family 43 glycosylhydrolase Draconibacterium sediminis
KWR59603.1 beta-xylosidase Bacteroides cellulosilyticus
KXT53146.1 glycosyl hydrolase family 43 Bacteroides intestinalis
WP_022185342.1 family 43 glycosylhydrolase Bacteroides dorei
WP_149940562.1 family 43 glycosylhydrolase Bacteroides dorei
EGF55151.1 glycosyl hydrolase family 43 Bacteroides clarus YIT 12056
WP_052194228.1 family 43 glycosylhydrolase Chryseobacterium sp. YR561
WP_052248550.1 family 43 glycosylhydrolase Chryseobacterium taiwanense
QEL14975.1 inverting glycoside hydrolase Gemmataceae bacterium PX52
HCZ20920.1 hypothetical protein Porphyromonadaceae bacterium
WP_118555838.1 family 43 glycosylhydrolase Lachnotalea sp. AF33-28
KRC82739.1 glycoside hydrolase Sphingomonas sp. Root241
WP_037567258.1 family 43 glycosylhydrolase Sphingomonas sp. UNC305MFCol5.2
RYY10129.1 glycoside hydrolase Alphaproteobacteria bacterium
SDF63438.1 Glycosyl hydrolases family 43 Sphingomonas carotinifaciens
WP_125656040.1 family 43 glycosylhydrolase Bacillus sp. 2B10
WP_104955384.1 family 43 glycosylhydrolase Sphingobium sp. SCG-1
WP_007677412.1 family 43 glycosylhydrolase Novosphingobium sp. AP12
WP_103096361.1 family 43 glycosylhydrolase Novosphingobium guangzhouense
ESQ79229.1 hypothetical protein AEYBE204_09475 Asticcacaulis sp. YBE204
WP_052019419.1 family 43 glycosylhydrolase Asticcacaulis sp. YBE204
WP_019922239.1 MULTISPECIES: family 43 glycosylhydrolase Betaproteobacteria
WP_047404138.1 family 43 glycosylhydrolase Caulobacter sp. OV484
WP_056054091.1 MULTISPECIES: family 43 glycosylhydrolase unclassified Caulobacter
ESQ78927.1 hypothetical protein AEYBE204_10920 Asticcacaulis sp. YBE204
WP_035442041.1 family 43 glycosylhydrolase Asticcacaulis sp. YBE204
PTW93623.1 glycosyl hydrolase family 43 Asticcacaulis sp. CF398
WP_107877940.1 family 43 glycosylhydrolase Asticcacaulis sp. CF398
0.50

Figure 3. A representative phylogenetic tree for GH43. The genes investigated are highlighted in yellow and blue; yellow represent the genes unique to pine-associated *Fusarium* species, whereas blue represent the genes shared by pine- and Poaceae-associated *Fusarium* species. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches (≥ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016). FCIR = *F. circinatum*; FFRAC = *F. fracticaudum*; FPIN = *F. pininemorale*; FKON = *F. konzum*; FSUB = *F. subglutinans*; FTEMP = *F. temperatum.*

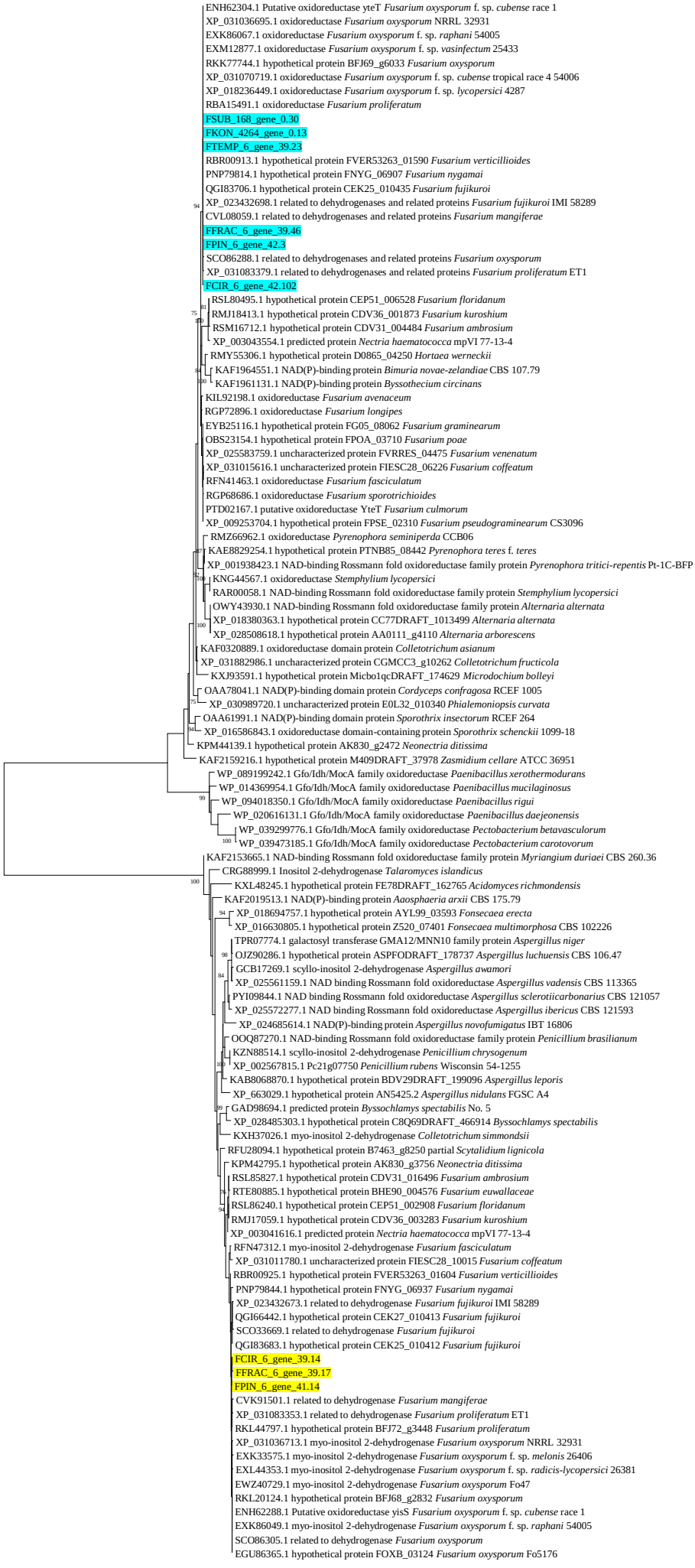

ENH62304.1 Putative oxidoreductase yteT *Fusarium oxysporum* f. sp. *cubense* race 1
XP_031036695.1 oxidoreductase *Fusarium oxysporum* NRRL 32931
EXK86067.1 oxidoreductase *Fusarium oxysporum* f. sp. *raphani* 54005
EXM12877.1 oxidoreductase *Fusarium oxysporum* f. sp. *vasinfectum* 25433
RKK77744.1 hypothetical protein BFJ69_g6033 *Fusarium oxysporum*
XP_031070719.1 oxidoreductase *Fusarium oxysporum* f. sp. *cubense* tropical race 4 54006
XP_018236449.1 oxidoreductase *Fusarium oxysporum* f. sp. *lycopersici* 4287
RBA15491.1 oxidoreductase *Fusarium proliferatum*
FSUB_168_gene_0.30
FKON_4264_gene_0.13
FTEMP_6_gene_39.23
RBR00913.1 hypothetical protein FVER53263_01590 *Fusarium verticillioides*
PNP79814.1 hypothetical protein FNYG_06907 *Fusarium nygamai*
QGI83706.1 hypothetical protein CEK25_010435 *Fusarium fujikuroi*
XP_023432698.1 related to dehydrogenases and related proteins *Fusarium fujikuroi* IMI 58289
CVL08059.1 related to dehydrogenases and related proteins *Fusarium mangiferae*
FFRAC_6_gene_39.46
FPIN_6_gene_42.3
SCO86288.1 related to dehydrogenases and related proteins *Fusarium oxysporum*
XP_031083379.1 related to dehydrogenases and related proteins *Fusarium proliferatum* ET1
FCIR_6_gene_42.102
RSL80495.1 hypothetical protein CEP51_006528 *Fusarium floridanum*
RMJ18413.1 hypothetical protein CDV36_001873 *Fusarium kuroshium*
RSM16712.1 hypothetical protein CDV31_004484 *Fusarium ambrosium*
XP_003043554.1 predicted protein *Nectria haematococca* mpVI 77-13-4
RMY55306.1 hypothetical protein D0865_04250 *Hortaea werneckii*
KAF1964551.1 NAD(P)-binding protein *Bimuria novae-zelandiae* CBS 107.79
KAF1961131.1 NAD(P)-binding protein *Byssothecium circinans*
KIL92198.1 oxidoreductase *Fusarium avenaceum*
RGP72896.1 oxidoreductase *Fusarium longipes*
EYB25116.1 hypothetical protein FG05_08062 *Fusarium graminearum*
OBS23154.1 hypothetical protein FPOA_03710 *Fusarium poae*
XP_025583759.1 uncharacterized protein FVRRES_04475 *Fusarium venenatum*
XP_031015616.1 uncharacterized protein FIESC28_06226 *Fusarium coffeatum*
RFN41463.1 oxidoreductase *Fusarium fasciculatum*
RGP68686.1 oxidoreductase *Fusarium sporotrichioides*
PTD02167.1 putative oxidoreductase YteT *Fusarium culmorum*
XP_009253704.1 hypothetical protein FPSE_02310 *Fusarium pseudograminearum* CS3096
RMZ66962.1 oxidoreductase *Pyrenophora seminiperda* CCB06
KAE8829254.1 hypothetical protein PTNB85_08442 *Pyrenophora teres* f. *teres*
XP_001938423.1 NAD-binding Rossmann fold oxidoreductase family protein *Pyrenophora tritici-repentis* Pt-1C-BFP
KNG44567.1 oxidoreductase *Stemphylium lycopersici*
RAR00058.1 NAD-binding Rossmann fold oxidoreductase family protein *Stemphylium lycopersici*
OWY43930.1 NAD-binding Rossmann fold oxidoreductase family protein *Alternaria alternata*
XP_018380363.1 hypothetical protein CC77DRAFT_1013499 *Alternaria alternata*
XP_028508618.1 hypothetical protein AA0111_g4110 *Alternaria arborescens*
KAF0320889.1 oxidoreductase domain protein *Colletotrichum asianum*
XP_031882986.1 uncharacterized protein CGMCC3_g10262 *Colletotrichum fructicola*
KXJ93591.1 hypothetical protein Micbo1qcDRAFT_174629 *Microdochium bolleyi*
OAA78041.1 NAD(P)-binding domain protein *Cordyceps confragosa* RCEF 1005
XP_030989720.1 uncharacterized protein E0L32_010340 *Phialemoniopsis curvata*
OAA61991.1 NAD(P)-binding domain protein *Sporothrix insectorum* RCEF 264
XP_016586843.1 oxidoreductase domain-containing protein *Sporothrix schenckii* 1099-18
KPM44139.1 hypothetical protein AK830_g2472 *Neonectria ditissima*
KAF2159216.1 hypothetical protein M409DRAFT_37978 *Zasmidium cellare* ATCC 36951
WP_089199242.1 Gfo/Idh/MocA family oxidoreductase *Paenibacillus xerothermodurans*
WP_014369954.1 Gfo/Idh/MocA family oxidoreductase *Paenibacillus mucilaginosus*
WP_094018350.1 Gfo/Idh/MocA family oxidoreductase *Paenibacillus rigui*
WP_020616131.1 Gfo/Idh/MocA family oxidoreductase *Paenibacillus daejeonensis*
WP_039299776.1 Gfo/Idh/MocA family oxidoreductase *Pectobacterium betavasculorum*
WP_039473185.1 Gfo/Idh/MocA family oxidoreductase *Pectobacterium carotovorum*
KAF2153665.1 NAD-binding Rossmann fold oxidoreductase family protein *Myriangium duriaei* CBS 260.36
CRG88999.1 Inositol 2-dehydrogenase *Talaromyces islandicus*
KXL48245.1 hypothetical protein FE78DRAFT_162765 *Acidomyces richmondensis*
KAF2019513.1 NAD(P)-binding protein *Aaosphaeria arxii* CBS 175.79
XP_018694757.1 hypothetical protein AYL99_03593 *Fonsecaea erecta*
XP_016630805.1 hypothetical protein Z520_07401 *Fonsecaea multimorphosa* CBS 102226
TPR07774.1 galactosyl transferase GMA12/MNN10 family protein *Aspergillus niger*
OJZ90286.1 hypothetical protein ASPFODRAFT_178737 *Aspergillus luchuensis* CBS 106.47
GCB17269.1 scyllo-inositol 2-dehydrogenase *Aspergillus awamori*
XP_025561159.1 NAD binding Rossmann fold oxidoreductase *Aspergillus vadensis* CBS 113365
PYI09844.1 NAD binding Rossmann fold oxidoreductase *Aspergillus sclerotiicarbonarius* CBS 121057
XP_025572277.1 NAD binding Rossmann fold oxidoreductase *Aspergillus ibericus* CBS 121593
XP_024685614.1 NAD(P)-binding protein *Aspergillus novofumigatus* IBT 16806
OOQ87270.1 NAD-binding Rossmann fold oxidoreductase family protein *Penicillium brasilianum*
KZN88514.1 scyllo-inositol 2-dehydrogenase *Penicillium chrysogenum*
XP_002567815.1 Pc21g07750 *Penicillium rubens* Wisconsin 54-1255
KAB8068870.1 hypothetical protein BDV29DRAFT_199096 *Aspergillus leporis*
XP_663029.1 hypothetical protein AN5425.2 *Aspergillus nidulans* FGSC A4
GAD98694.1 predicted protein *Byssochlamys spectabilis* No. 5
XP_028485303.1 hypothetical protein C8Q69DRAFT_466914 *Byssochlamys spectabilis*
KXH37026.1 myo-inositol 2-dehydrogenase *Colletotrichum simmondsii*
RFU28094.1 hypothetical protein B7463_g8250 partial *Scytalidium lignicola*
KPM42795.1 hypothetical protein AK830_g3756 *Neonectria ditissima*
RSL85827.1 hypothetical protein CDV31_016496 *Fusarium ambrosium*
RTE80885.1 hypothetical protein BHE90_004576 *Fusarium euwallaceae*
RSL86240.1 hypothetical protein CEP51_002908 *Fusarium floridanum*
RMJ17059.1 hypothetical protein CDV36_003283 *Fusarium kuroshium*
XP_003041616.1 predicted protein *Nectria haematococca* mpVI 77-13-4
RFN47312.1 myo-inositol 2-dehydrogenase *Fusarium fasciculatum*
XP_031011780.1 uncharacterized protein FIESC28_10015 *Fusarium coffeatum*
RBR00925.1 hypothetical protein FVER53263_01604 *Fusarium verticillioides*
PNP79844.1 hypothetical protein FNYG_06937 *Fusarium nygamai*
XP_023432673.1 related to dehydrogenase *Fusarium fujikuroi* IMI 58289
QGI66442.1 hypothetical protein CEK27_010413 *Fusarium fujikuroi*
SCO33669.1 related to dehydrogenase *Fusarium fujikuroi*
QGI83683.1 hypothetical protein CEK25_010412 *Fusarium fujikuroi*
FCIR_6_gene_39.14
FFRAC_6_gene_39.17
FPIN_6_gene_41.14
CVK91501.1 related to dehydrogenase *Fusarium mangiferae*
XP_031083353.1 related to dehydrogenase *Fusarium proliferatum* ET1
RKL44797.1 hypothetical protein BFJ72_g3448 *Fusarium proliferatum*
XP_031036713.1 myo-inositol 2-dehydrogenase *Fusarium oxysporum* NRRL 32931
EXK33575.1 myo-inositol 2-dehydrogenase *Fusarium oxysporum* f. sp. *melonis* 26406
EXL44353.1 myo-inositol 2-dehydrogenase *Fusarium oxysporum* f. sp. *radicis-lycopersici* 26381
EWZ40729.1 myo-inositol 2-dehydrogenase *Fusarium oxysporum* Fo47
RKL20124.1 hypothetical protein BFJ68_g2832 *Fusarium oxysporum*
ENH62288.1 Putative oxidoreductase yisS *Fusarium oxysporum* f. sp. *cubense* race 1
EXK86049.1 myo-inositol 2-dehydrogenase *Fusarium oxysporum* f. sp. *raphani* 54005
SCO86305.1 related to dehydrogenase *Fusarium oxysporum*
EGU86365.1 hypothetical protein FOXB_03124 *Fusarium oxysporum* Fo5176
0.50

Figure 4. A representative phylogenetic tree for GH109. The genes investigated are highlighted in yellow and blue; yellow represent the genes unique to pine-associated *Fusarium* species, whereas blue represent the genes shared by pine- and Poaceae-associated *Fusarium* species. The evolutionary history was inferred by using the Maximum Likelihood method based on the Le Gascuel model (Le & Gascuel, 2008), along with a discrete Gamma distribution to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) were shown next to the branches (≥ 75%). All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA 7 (Kumar *et al.*, 2016). FCIR = *F. circinatum*; FFRAC = *F. fracticaudum*; FPIN = *F. pininemorale*; FKON = *F. konzum*; FSUB = *F. subglutinans*; FTEMP = *F. temperatum.*

SUPPLEMENTAL TABLES

All the supplemental data tables can be retrieved from the following Google drive:
https://drive.google.com/drive/folders/1eK5hGcNRkECBVHwOno6XlOKbKUI43ZLg?usp=sharing

Supplemental Table 1. Putative CAZyme-encoding genes from *F. circinatum* as identified by dbCAN2. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12).

Supplemental Table 2. Putative CAZyme-encoding genes from *F. fracticaudum* as identified by dbCAN2. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12).

Supplemental Table 3. Putative CAZyme-encoding genes from *F. pininemorale* as identified by dbCAN2. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12).

Supplemental Table 4. Putative CAZyme-encoding genes from *F. konzum* as identified by dbCAN2. The number indicated after the species identifier is the scaffold number of this gene.

Supplemental Table 5. Putative CAZyme-encoding genes from *F. subglutinans* as identified by dbCAN2. The number indicated after the species identifier is the scaffold number of this gene.

Supplemental Table 6. Putative CAZyme-encoding genes from *F. temperatum* as identified by dbCAN2. The number indicated after the species identifier is the chromosomal location of this gene (chromosome 1-12).

Supplemental Table 7. The EST data for *F. circinatum*, obtained from Wingfield *et al*. (2012) and van Wyk *et al.* (2019).

SUMMARY

The *Fusarium fujikuroi* species complex (FFSC) contain a diversity of species associating with a variety of hosts. These species are economically important due to the disease caused on respective hosts. Genomic sequences make it possible to study these fungi using comparative genomics. Some genomic regions tend to be more variable in terms of genes and genetic structure. These regions are more subjected to the acquisition of novel genes through horizontal gene transfer, internal mutations, duplications and translocations. Genes within these regions may also promote host-specificity of these fungi by encoding for specific gene products and participate in processes and pathways. This study has provided evidence for two sets of unique genes promoting host-specificity in *Fusarium* species associated with pine trees and members of Poaceae, respectively. The location of these genes was more likely to be found in the telomeric regions and they were inclined to cluster together or were near another cluster. This corresponds to what is known about host specificity genes and their location in variable regions, of which the telomere is considered a part of. The variable nature of the accessory chromosome in the *Fusarium* species associated with Poaceae (chromosome 12) also plays a role in host-specificity, as it possessed more unique genes than the core chromosomes. These findings emphasise that *Fusarium* species employ different molecular mechanisms to successfully infect and colonise pine and members of Poaceae, respectively.

We furthermore provided evidence for the differences and similarities in the repertoire of CAZyme genes between the *Fusarium* species associated with different hosts. Some CAZyme classes were more prevalent than others and likely have vital roles in fungal colonisation, nutrient utilisation and plant biomass degradation. Fungi are known to also utilise different sets of CAZymes depending on the cell wall composition of the host plant. However, the CAZymes results from this study were similar between *Fusarium* species associated with gymnosperms (pine trees) and monocots (members of Poaceae) which would imply that CAZymes are not intrinsically involved in host-specificity. Notwithstanding, two glycoside hydrolases (GHs) were identified as different amongst the two groups of *Fusarium* species. The roles of each in host-specificity were not clear and further investigations into these are warranted.

Future studies investigating the functional characterisation of the host-specific unique genes identified in this study are merited. Expression analyses into the unique genes and CAZymes highlighted in this study at different points of infection and in different hosts would shed some light on their respective roles. Furthermore, the *Fusarium* population can be broadened by supplementing this study with additional and complete genomes. The findings from this study advance the current understanding of the diversity and evolution of species within the FFSC.