

Multiple independent origins for a subtelomeric locus associated with growth rate in *Fusarium circinatum*

Stephanie Van Wyk, Brenda D. Wingfield, Lieschen De Vos, Quentin C. Santana, Nicolaas A. Van der Merwe, and Emma T. Steenkamp

Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa; corresponding author email: emma.steenkamp@fabi.up.ac.za

Abstract: *Fusarium* is a diverse assemblage that includes a large number of species of considerable medical and agricultural importance. Not surprisingly, whole genome sequences for many *Fusarium* species have been published or are in the process of being determined, the availability of which is invaluable for deciphering the genetic basis of key phenotypic traits. Here we investigated the distribution, genic composition, and evolutionary history of a locus potentially determining growth rate in the pitch canker pathogen *F. circinatum*. We found that the genomic region underlying this locus is highly conserved amongst *F. circinatum* and its close relatives, except for the presence of a 12 000 base pair insertion in all of the examined isolates of *F. circinatum*. This insertion encodes for five genes and our phylogenetic analyses revealed that each was most likely acquired through horizontal gene transfer from polyphyletic origins. Our data further showed that this region is located in a region low in G+C content and enriched for repetitive sequences and transposable elements, which is situated near the telomere of Chromosome 3 of *F. circinatum*. As have been shown for other fungi, these findings thus suggest that the emergence of the unique 12 000 bp region in *F. circinatum* is linked to the dynamic evolutionary processes associated with subtelomeres that, in turn, have been implicated in the ecological adaptation of fungal pathogens.

Key words:

Fusarium temperatum
horizontal gene transfer
Pitch canker fungus
transposable elements

Article info: Submitted: 12 January 2018; Accepted: 19 February 2018; Published: 27 February 2018.

INTRODUCTION

Fusarium species are remarkably diverse (Leslie & Summerell 2006, O'Donnell *et al.* 2013). Despite the extensive genomic synteny characterizing this genus (Waalwijk *et al.* 2004, Ma *et al.* 2010, Lysøe *et al.* 2014), individual species are not only phenotypically complex but also display a wide range of species-specific traits (Wiemann *et al.* 2013, Herron *et al.* 2015, Sperschneider *et al.* 2015). Comparative studies are increasingly showing that this diversity also extends to their genomic architectures and genetic content (Waalwijk *et al.* 2004, De Vos *et al.* 2011, Chiara *et al.* 2015, Hansen *et al.* 2015). For example, the closely related species *F. circinatum* and *F. temperatum* are characterized by substantial levels of both macro- and micro-synteny (De Vos *et al.* 2014), but they are, respectively, pathogens of pine (Hepting & Roth 1946, Leslie *et al.* 2006) and maize (Scauflaire *et al.* 2011). They also differ dramatically in other phenotypic traits (Desjardins *et al.* 2000, De Vos *et al.* 2007, 2011), including growth rate for which a major Quantitative Trait Locus (QTL) has previously been identified (De Vos *et al.* 2011).

Certain parts of *Fusarium* genomes appear to be more variable than others (Cuomo *et al.* 2007, Coleman *et al.* 2009, Ma *et al.* 2010, Chiara *et al.* 2015, Sperschneider

et al. 2015). In addition to the telomeres and centromeres (Chiara *et al.* 2015, Sperschneider *et al.* 2015), areas of high sequence variability also occur in other chromosomal regions and may even extend across entire chromosomes such as the supernumerary or dispensable chromosomes (Ma *et al.* 2010, Van der Nest *et al.* 2014). Generally, these variable regions in diverse fungi are rich in repeats and transposable elements (TEs), have G+C contents that differ markedly from the rest of the genome (Goodwin *et al.* 2011), and often encode nonessential genes (Fedorova *et al.* 2008, Coleman *et al.* 2009, Sperschneider *et al.* 2015). Overall, such regions of variability are thought to accelerate genome evolution and plasticity and to promote adaptation (Fedorova *et al.* 2008, Coleman *et al.* 2009, Chiara *et al.* 2015).

The genomes of filamentous fungi are dynamic and capable of tolerating extensive gene gains and losses (Braun *et al.* 2000, Coleman *et al.* 2009, Spanu *et al.* 2010, Raffaele & Kamoun 2012). Gene gains may occur *via* internal genomic mutations (i.e. intra-genomic mutations) due to duplication, displacement and translocation events (Gac & Giraud 2008, Proctor *et al.* 2009, De Vos *et al.* 2014), or *via* gene introductions from external sources through horizontal gene transfer (HGT) (Ma *et al.* 2010, Chuma *et al.* 2011, Hansen *et al.* 2015). HGT refers to the exchange of

© 2018 International Mycological Association

You are free to share - to copy, distribute and transmit the work, under the following conditions:

Attribution: You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).

Non-commercial: You may not use this work for commercial purposes.

No derivative works: You may not alter, transform, or build upon this work.

For any reuse or distribution, you must make clear to others the license terms of this work, which can be found at <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>. Any of the above conditions can be waived if you get permission from the copyright holder. Nothing in this license impairs or restricts the author's moral rights.

genetic material between different strains or species, which would include those due to hybridization (Brown & Doolittle 1999). Nevertheless, such gains and differential losses have apparently given rise to species-specific regions in various fungi (Daboussi & Capy 2003, Coleman *et al.* 2009, Proctor *et al.* 2009, Ma *et al.* 2010, Spanu *et al.* 2010), e.g. lineages of *Magnaporthe*, *Aspergillus*, *Fusarium* and *Coccidioides* (Galagan *et al.* 2005, Behnsen *et al.* 2008, Skamnioti *et al.* 2008, Coleman *et al.* 2009, Proctor *et al.* 2009, Moran *et al.* 2011, Hansen *et al.* 2015). Recently it was also demonstrated that such gains and losses have been particularly important in driving the formation of species-specific regions within the telomeric regions of certain *Fusarium* species (Chiara *et al.* 2015).

The acquisition of genes *via* HGT is regarded as an important and ongoing source of functional novelty in fungi (Ma *et al.* 2010, Wisecaver *et al.* 2014, Jaramillo *et al.* 2015). Compared to other eukaryotes, and some prokaryotes (Nelson *et al.* 1999, Crisp *et al.* 2015), this form of gene gains is relatively high in fungi (Gardiner *et al.* 2013, Glenn *et al.* 2016). This is also true for *Fusarium* species, where HGTs are thought to have shaped their evolution and contributed to the emergence of species-specific traits (Ma *et al.* 2010, Alves *et al.* 2014, Sieber *et al.* 2014, Stewart *et al.* 2014, Wisecaver *et al.* 2014, Glenn *et al.* 2016). For example, *F. graminearum*, *F. verticillioides* and *F. oxysporum* f. sp. *lycopersici* have species-specific gene clusters that were likely acquired across species boundaries (Sieber *et al.* 2014, Glenn *et al.* 2016). In *F. verticillioides* it was also recently shown that certain gene clusters were acquired from multiple external sources as opposed to having been acquired through gene duplication and differential gene loss (Stewart *et al.* 2014, Glenn *et al.* 2016).

In this study, we examined the chromosomal location and evolutionary origins of the major QTL associated with growth rate variation in *F. circinatum*, that was previously identified in a genetic linkage map of an interspecific cross between *F. circinatum* and *F. temperatum* (De Vos *et al.* 2007, 2011). For this purpose, our study had four specific objectives. Firstly, we located the genetic marker linked to growth rate variation (i.e. marker AT/AC-625bh) in the genome of *F. circinatum* (Wingfield *et al.* 2012) and identified the genes encoded in the region underlying it by making use of various *in silico* approaches. Secondly, the identified region and the chromosomal areas surrounding it were examined in terms of the likely functions they encode, their G+C content, and the presence and distribution of repeats and TEs. Thirdly, the presence and distribution of the region identified was assessed in a broad collection of *Fusarium* species and in other isolates of *F. circinatum* by making use of PCR-based analyses and genome-based searches. For the latter, the two *F. circinatum* genomes already in the public domain (Wingfield *et al.* 2012, Van der Nest *et al.* 2014) were supplemented by sequencing the genome for a third isolate obtained from diseased pine seedling roots in South Africa (Steenkamp *et al.* 2014). Finally, the putative origin of the identified region was evaluated using various sequence alignments and phylogenetic analyses. These fine-scale synteny comparisons and phylogenetic information revealed genetic features that likely facilitated the emergence of a phenotype-

associated QTL and further broaden our understanding of genetic differentiation amongst related fungal lineages.

MATERIALS AND METHODS

Genome sequencing and assembly

Fusarium circinatum isolate KS17 (CMW 674; Culture collection of the Forestry and Agricultural Biotechnology Institute, FABI, University of Pretoria, South Africa) was obtained from the infected root tissue of a *Pinus radiata* seedling collected in a nursery in the Western Cape, South Africa in 2005 (Steenkamp *et al.* 2014). The isolate was grown in half strength potato dextrose broth (20 % w/v) and incubated at 25 °C in the dark on an orbital shaker at 120 rpm for 7 d, after which DNA was extracted (Möller *et al.* 1992). The DNA was used to prepare two mate-pair libraries (1000 base pair [bp] insert size) and a single-read library, which were then sequenced by SEQOMICS (Csongrád, Hungary) using the SOLiD™ V4 technology (Applied Biosystems, California, USA) producing reads containing *ca.* 50 bp. Sequence reads were quality filtered using CLC Genomics Workbench v.8.0 (CLCbio, Aarhus, Denmark), assembled into scaffolds using ABySS v.1.5.2 (Simpson *et al.* 2009), after which gapped regions within scaffolds were closed with GapFiller v1.11 (Boetzer & Pirovano 2012). Completeness of the genome assembly was evaluated with BUSCO v.2.0.1 using the *Sordariomycetes* gene set (Simão *et al.* 2015). WebAUGUSTUS (Hoff & Stanke 2013) to predict putative open reading frames (ORFs) based on the gene models for *F. graminearum* and mRNA data from *F. circinatum* (Wingfield *et al.* 2012).

Genomic localization of marker AT/AC-625bh, a major growth rate determining QTL in *F. circinatum*

The location of marker AT/AC-625bh (De Vos *et al.* 2011) within the genome sequence of isolate FSP34 of *F. circinatum* (Wingfield *et al.* 2012) was determined as described previously (De Vos *et al.* 2014). This was done with *in silico* Amplified Fragment Length Polymorphism (AFLP) analysis using AFLP*inSilico* v2 (Rombauts *et al.* 2003), which involved the use of simulated restriction enzyme digestion profiles for the entire genome of *F. circinatum*. The analysis used the restriction sites for *EcoRI* (GAATT↓C) and *MseI* (TTA↓A) with an adapter length of zero, as well as AC and AT selective nucleotides (De Vos *et al.* 2007). In order to account for initial variability in estimated restriction fragment sizes, all restriction fragments in the size range 595–635 bp were considered in the analysis. By making use of nucleotide BLAST (Basic Local Alignment Search Tool; Altschul *et al.* 1997) searches and alignments in CLC Main Workbench software (CLC Bio-Qiagen, Aarhus, Denmark, version 7.0.3), sequences of the *in silico* restriction fragments were then compared to those in the most recent version of the published assembly of *F. circinatum* (Wingfield *et al.* 2012). The latter was represented in the genome database of the National Centre for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>) by a draft pseudo-chromosome assembly (BioProject PRJNA41113) with accession AYJV00000000.2.

Sequence characterization of the genomic region containing marker AT/AC-625bh

The stretch of sequence containing marker AT/AC-625bh, as well as regions up- and downstream of it were characterized in terms of G+C content and the occurrence and distribution of repetitive elements. The G+C content was determined using CLC Genomics Workbench and a sliding window of 1 000 bp and step size of 500 bp. For identifying repeat elements, Repeat Masker (Tarailo-Graovac & Chen 2009) and Tandem Repeat Finder (Benson 1999) were used. Putative transposable elements (TE) were identified by using the CENSOR-EMBL fungal TEs database (Kohany *et al.* 2006, Li *et al.* 2015). Repeat and TE density were determined using a sliding window of 1 000 bp with 500 bp increments. In order to determine the abundance of the telomere-associated repeat sequence “TTAGGG/CCCTAA” (Garcia-Pedrajas & Roncero 1996, Fulnečková *et al.* 2013), a motif search was conducted in CLC Genomics Workbench using a sliding window of 1000 bp with 500 bp increments. All repeats showing 80 % similarity to the telomere-associated sequence were considered in this analysis.

The functions of genes encoded on the stretch of genome sequence containing marker AT/AC-625bh were also inferred. This was done using InterProScan (Zdobnov & Apweiler 2001) to determine Gene Ontologies (GO), protein family membership (PFAM) and protein functional domains. Putative secondary metabolism gene clusters were identified using Antibiotics and Secondary Metabolites Analysis Shell (antiSMASH) (Blin *et al.* 2013). Gene density was estimated using a window size was 10 000 bp and the step size 5000 bp.

Synteny analysis of the genomic region containing marker AT/AC-625bh

Synteny and collinearity across the region containing marker AT/AC-625bh were evaluated using nucleotide alignments of the relevant genomic sections in representative *Fusarium* isolates and species (Supplementary Table S1). Together with the genome data for *F. temperatum* and *F. circinatum* isolate FSP34, we also included those for two additional isolates of *F. circinatum* KS17 (this study) and GL1327 (Van der Nest *et al.* 2014) as well as additional taxa in the *F. fujikuroi* species complex (FFSC) (Geiser *et al.* 2013); i.e. *F. verticillioides* (Cuomo *et al.* 2007), *F. mangiferae* (Niehaus *et al.* 2016) and *F. fujikuroi* (Wiemann *et al.* 2013). For comparison we also included representatives of other well-known *Fusarium* complexes; i.e. *F. graminearum* (Cuomo *et al.* 2007), *F. oxysporum* (Ma *et al.* 2010) and *F. solani* (Coleman *et al.* 2009).

These genome-based synteny and collinearity analyses were complemented with PCRs and Sanger sequencing. This was done to confirm the assembly of the genomic region containing marker AT/AC-625bh in 22 diverse isolates of *F. circinatum* (Supplementary Table S1). The approach was also used to confirm breaks in synteny and collinearity in representative isolates of other *Fusarium* species. All primers (Supplementary Table S2) were designed using Primer3 (Untergasser *et al.* 2012). Genomic DNA was extracted from each isolate using a previously described protocol (Steenkamp *et al.* 1999). All amplification reactions

were performed using MyTaq™ DNA polymerase (Bioline Reagents Ltd., MA), according to the supplier's protocol. Purified PCR products were sequenced at the Department of Genetics at the University of Pretoria, using the ABI 3500xl Genetic Analyzer (Applied Biosystems, CA).

Putative origins of the genomic region containing marker AT/AC-625bh

For each of the genes encoded in the genomic region containing the QTL marker, a dataset of homologous protein sequences was assembled. The sequences included in these datasets were identified using BLAST searches against eight publically available *Fusarium* genomes (Supplementary Table S3), as well as the genome databases of MycoCosm (Grigoriev *et al.* 2013) (Joint Genome Institute [JGI], US Department of Energy) and the NCBI. For the latter, query sequences were searched against those in the non-redundant database using the online position-specific iterative (psi) BLAST tool (Altschul *et al.* 1997). In order to exclude highly divergent protein sequences, we only considered those BLAST sequences with at least 40 % amino acid identity over 70 % of the query sequence length and that had expect-values [E] < 1×10⁻⁵ and bit scores > 200. Also, predicted proteins classified as “partial proteins” were excluded, and only the fully predicted proteins were considered for further analyses.

Individual sequence datasets were aligned using CO-BALT (Constraint-based Multiple Protein Alignment Tool) (Papadopoulos & Agarwala 2007) with default settings (https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi). These alignments were then trimmed in BioEdit v.7.0.9.0 (Hall 1999) to ensure that all of the sequences spanned the same region. Each dataset was subjected to ProtTest 3.2 (Abascal *et al.* 2005) to determine the best-fit substitution model. These model parameters were then used to perform Maximum Likelihood phylogenetic analyses with MEGA 6.0 (Tamura *et al.* 2013). Branch support was evaluated using the same model parameters and 1000 bootstrap pseudo-replicates.

Relative to the overall phylogenetic relationships among the FFSC species and its *Fusarium* relatives, we also investigated the relationships between the *F. circinatum*-specific genes encoded in the AT/AC-625bh-containing region to those encoded elsewhere in the examined *Fusarium* genomes (Supplementary Table S3). Dataset construction and phylogenetic analyses were performed as described above, except that BLASTp was used to identify homologs and only full-length sequences were included. Another round of analyses was also conducted where we constructed overview trees of the top BLAST hits (irrespective of bit scores and query coverage) in the NCBI databases using a neighbor-joining approach in MEGA.

RESULTS

Fusarium circinatum isolate KS17 genome sequence

The draft genome assembly of *F. circinatum* isolate KS17 was 46 325 048 bp in size. It had an average coverage of 166x and G+C content of 44.69 %. The assembly consisted of 6033 contigs (>200bp) with an N50-value of 95 695 bp.

BUSCO suggests that the assembly was 76.2 % complete (i.e. complete BUSCOs = 76.2 %; complete and single-copy BUSCOs = 75.1 %; complete and duplicated BUSCOs = 1.1 %; fragmented BUSCOs = 17.0 %; missing BUSCOs = 6.8 %; number of BUSCOs searched = 3725) (Simão *et al.* 2015). WebAUGUSTUS predicted that it encodes 16 502 putative ORFs. The *F. circinatum* KS17 genome sequence data were deposited at DDBJ/EMBL/GenBank under the accession number LQBB00000000. The version described here is version LQBB01000000.

Genomic localization of marker AT/AC-625bh

In silico AFLP analysis and sequence comparisons (De Vos *et al.* 2011) revealed that marker AT/AC-625bh is 599 bp in size. It was located within the gene FCIRG_04559 of *F. circinatum* (FSP34). Marker AT/AC-625bh was positioned from nucleotides 39 762–40 361 on Chromosome 3 (NCBI accession CM004513.1). Note that this corresponds to position 9 351–9 950 on contig 02138 of the previous version of the assembly (Wingfield *et al.* 2012)

Sequence characterization of the genomic region containing marker AT/AC-625bh

The first 100 000 bp of Chromosome 3 of *F. circinatum* that contained marker AT/AC-625bh was characterized further. Based on InterProScan, this region encoded a diverse range of putative protein products (Supplementary Tables S4 and S5). However, it appeared to be enriched for those involved in transmembrane

substrate transportation (FCIRG_04551 and FCIRG_04555), transcriptional regulation (FCIRG_04552, FCIRG_04556 and FCIRG_04559), carbon metabolism (FCIRG_04550 and FCIRG_04553), and catalytic activities (FCIRG_04549, FCIRG_04553, FCIRG_04557, FCIRG_04558). The analysis with antiSMASH also predicted the presence of a biosynthetic gene cluster between 53 928–81 890bp (Supplementary Table S6; FCIRG_03382, FCIRG_03383, FCIRG_03384, FCIRG_03385 and FCIRG_03388) with similarity to the gene cluster involved in butirosin biosynthesis.

Large changes in G+C content, gene, TE and repeat density were found across the examined portion of Chromosome 3 (Fig. 1). Based on G+C content, the first 12 000 bp were markedly different from the remainder of the sequence. After averaging ca. 27 % in the first 12 000 bp, the G+C content increased to an average of ca. 48.5 %. In terms of gene density, this first section also encoded fewer genes compared to the rest of Chromosome 3. We observed a similar distribution pattern for the repeats (Supplementary Table S7–S8) and putative TEs (Supplementary Table S9 and Supplementary Fig. S1), which were notably more abundant in the first 14 000 bp compared to that of the remainder of the downstream regions. The same was also true for the telomere-specific “TTAGGG” repeat motif (Supplementary Fig. S2). Therefore, based on G+C, repeat, TE (Supplementary Fig. S1) and gene content, marker AT/AC-625bh is located in the subtelomeric region of Chromosome 3 of *F. circinatum*.

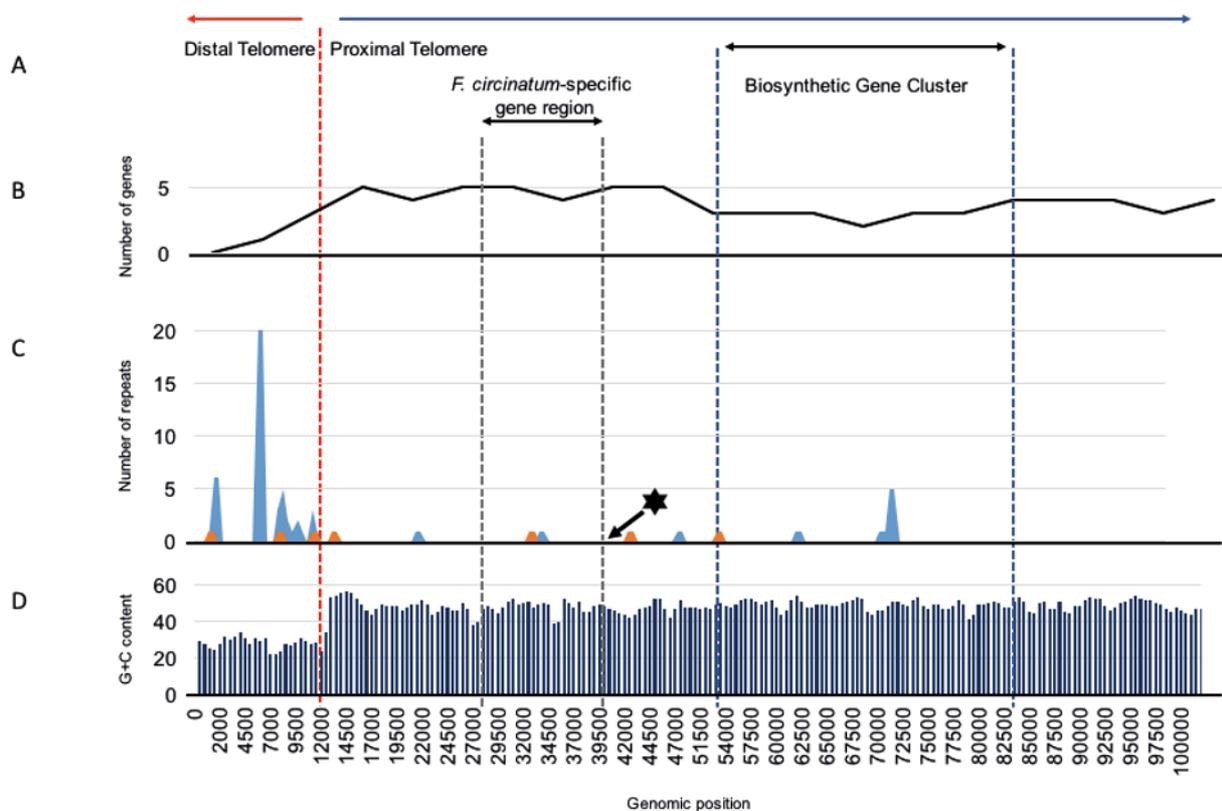


Fig. 1. Genomic features of the first 100 000 bp of Chromosome 3 of *Fusarium circinatum* (FSP34). (A) This region corresponds to the subtelomere of the chromosome. (B) Line graph illustrating the change in gene count determined through a 10 000 bp sliding window at 5 000 bp increments. (C) Chart showing the count of simple repeat and tandem repeat sequences in blue and the count of transposable element associated repeat sequences in orange; these were determined using a 10 bp sliding window at 500 bp increments, and the black star indicates the position of the QTL marker. (D) The data series represents G+C (%) content, which was determined with a 1000 bp sliding window at 500 bp increments.

Synteny analysis of the genomic region containing marker AT/AC-625bh

We first compared the gene content and orientation of the region containing the AFLP marker in the genome of *F. circinatum* FSP34 to those in the two other *F. circinatum* genomes (i.e. for isolate KS17 and GL1327). All 15 genes encoded in the region containing marker AT/AC-625bh were present in the same order and orientation in these three genomes. The intergenic PCR and Sanger sequencing analysis of this region, in 21 additional isolates of the fungus, further confirmed the genomic assembly of this region, as well as the order and orientation of genes (results not shown).

Subsequent interspecies comparisons revealed that the ca. 12 000 bp region containing marker AT/AC-625bh was absent from the corresponding genomic regions in other *Fusarium* species (Fig. 2). This 12 000 bp sequence encode five genes (FCIRG_04559, FCIRG_04558, FCIRG_04557, FCIRG_04556 and FCIRG_04555). This genome-based observation was confirmed with PCR and Sanger sequencing, where our primers were designed to span the synteny breakpoint (i.e. from the end of gene FCIRG_04560 to the start of gene FCIRG_04554). These analyses confirmed that the 12 000 bp region was indeed absent from the genomes of the other FFSC species examined (i.e. *F. temperatum*, *F. mangiferae*, *F. fujikuroi* and *F. verticillioides*). However, genome-based comparisons of the up- and downstream regions flanking the 12 000 bp insert in *F. circinatum*, revealed a high degree of conserved synteny amongst the FFSC species included. This homology also extended to the sequenced representatives of *F. oxysporum* (Supplementary Table S10), but not to the more distantly related *F. graminearum* and *F. solani* (Supplementary Table S11).

Putative origins of the genomic region containing marker AT/AC-625bh

To examine the potential origins of the AT/AC-625bh marker-containing region specific to *F. circinatum*, the five genes encoded on this 12 000 bp stretch of DNA were compared to those included in various local and public databases. This allowed for the identification of homologous proteins for all five of the genes encoded in this *F. circinatum*-specific region (Supplementary Tables S12–S15). However, none of the five genes co-occurred (i.e., located together on the same contig or chromosome) in any of the fungal genomes examined. Furthermore, the taxa with which the *F. circinatum*-specific sequences shared identity differed markedly among the five genes.

Phylogenetic analysis of datasets containing only *Fusarium* sequences revealed that none of FCIRG_04559, FCIRG_04558, FCIRG_04557, FCIRG_04556 and FCIRG_04555 grouped with other sequences from *F. circinatum* (Supplementary Fig. S3). The same pattern was observed in the overview trees inferred from the top BLAST hits for each gene in the NCBI database (Supplementary Fig. S4). This was also true even if the FSP34 genome contained a second homolog of the gene, as is the case for FCIRG_04559 and FCIRG_04557. In both instances, the gene encoded in the *F. circinatum*-specific region did not group with *F. circinatum* or other members of the FFSC. None of the five genes in the *F. circinatum* specific region was thus characterized by a phylogeny matching that expected for the FFSC.

Rigorous phylogenetic analyses of the *F. circinatum*-specific region revealed that the genes in this locus have distinct evolutionary ancestries (Fig. 3). Based on these results, FCIRG_04556 and FCIRG_04559 were most closely related to proteins encoded by *F. solani*. In the phylogenetic trees containing homologs of FCIRG_04555

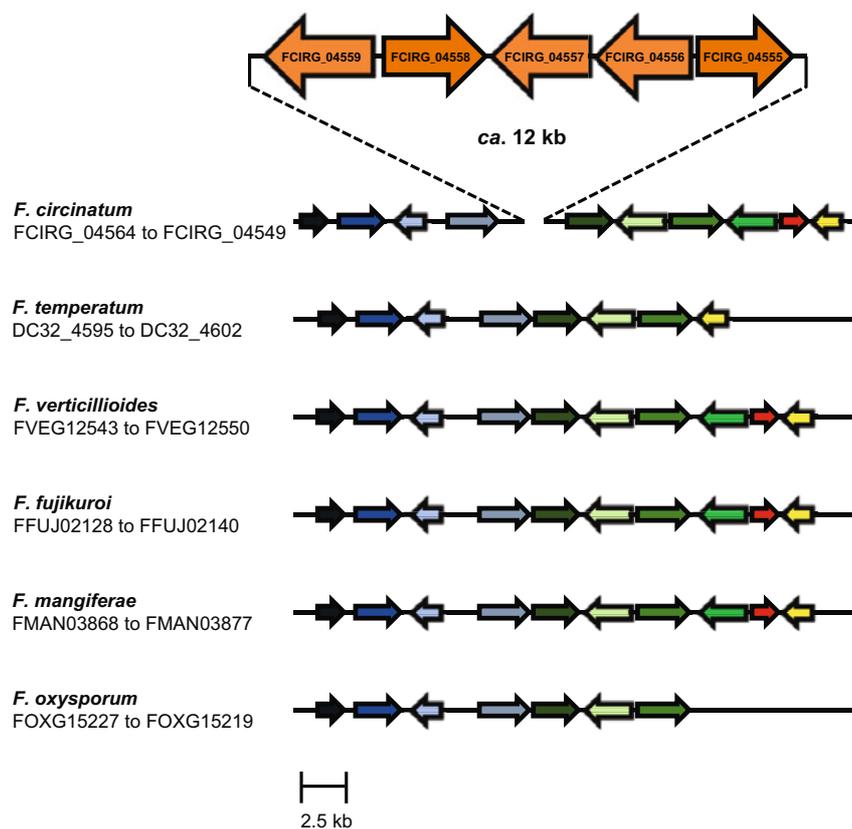


Fig. 2. Gene content and organization of the region containing the QTL marker associated with growth rate variation in *Fusarium circinatum*. Gene position and orientation are indicated with block arrows. Orange arrows illustrate genes only encoded in *F. circinatum*. Gene names are indicated below each species name. Similar colored genes illustrate shared collinearity and synteny. See Supplementary Table S4 for the predicted gene functions in *F. circinatum*.

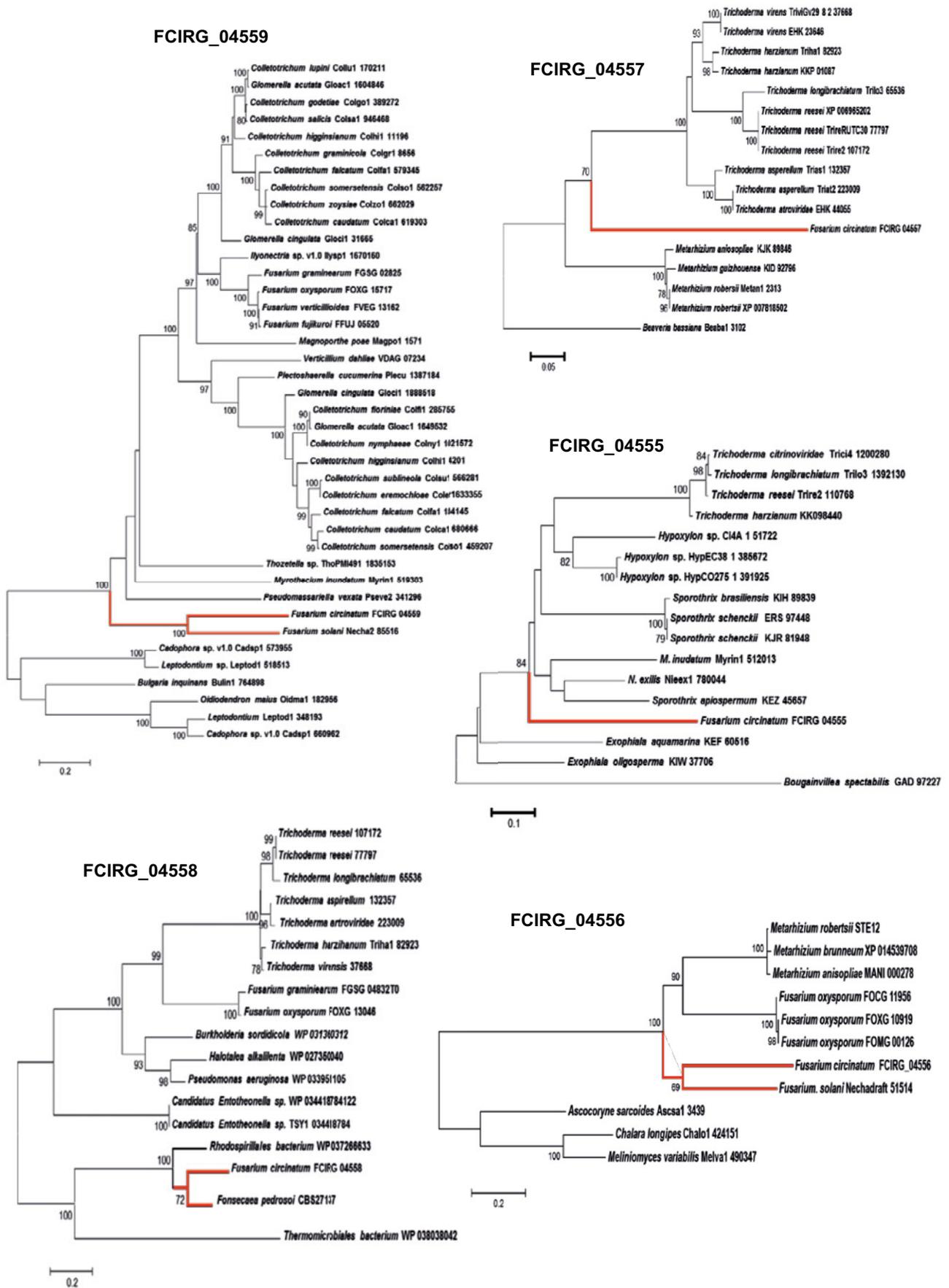


Fig. 3. Maximum likelihood trees constructed from the inferred *Fusarium circinatum* species-specific proteins FCIRG_04559, FCIRG_04558, FCIRG_04557, FCIRG_04556 and FCIRG_04555. Branches indicated in red show the position and closest relative or clade of *F. circinatum* in the five protein trees. Each alignment included only those protein sequences with >40 % amino acid similarity to that of the particular *F. circinatum* homologue. Bootstrap values (>70 %) are indicated at nodes, and the scale shows substitutions per site.

and FCIRG_04557, these *F. circinatum* genes grouped with diverse non-*Fusarium* fungi. The results showed that FCIRG_04558 was nested within a bacterial clade. These results thus pointed towards HGT-based origins for the *F. circinatum*-specific 12 000 bp region and its genes.

The non-vertical inheritance of the *F. circinatum*-specific region and its genes was also evident when we re-examined G+C content. It was characterized by an average G+C content of 51.2 %, which is significantly higher than the 47 % in the rest of the FSP34 genome (Supplementary Table S16). A similar pattern was also observed for some of the individual genes (i.e., FCIRG_0556 and FCIRG_0559) (Supplementary Table S17–S18), but particularly pronounced in FCIRG_0558 (Supplementary Table S19–S20). This gene and its xenolog in *F. pedrosoi* (KIW 84299) had G+C contents exceeding 53 % (Supplementary Table S19), which supported the bacterial ancestry of this gene is dramatically different from the rest of their genomes.

DISCUSSION

The results of this study showed that the QTL-marker AT/AC 625bh, which previously had been associated with growth rate (De Vos *et al.* 2011), is located on Chromosome 3 of *F. circinatum*. The genomic region underlying this marker is approximately 12 000 bp in size and is apparently unique to the species. It is absent from all of the examined genomes of other *Fusarium* species, including the closely related *F. temperatum*. It is, however, present in the genomes of all *F. circinatum* isolates we investigated, including the newly sequenced isolate KS17. The genomic regions directly adjacent to this unique region showed a high degree of synteny and collinearity across the FFSC and its sister taxa in the *F. oxysporum* species complex, but not in species more distantly related to the FFSC. This implies that the *F. circinatum*-specific gene region must have been introduced from elsewhere.

Detailed examination of the region up- and downstream of the *F. circinatum*-specific region suggested that it is located within Chromosome 3's subtelomere. This was evident from the high density of repeats and TEs that coincided with an AT-rich genomic environment. These genetic features are characteristic of distal subtelomeric regions (Flint *et al.* 1997, Cuomo *et al.* 2007, Wiemann *et al.* 2013, Chiara *et al.* 2015). In addition, the telomere-associated repeat motif, "TTAGGG", a known genetic feature of the distal parts of the telomeres (Garcia-Pedrajas & Roncero 1996, Fulnečková *et al.* 2013), was prominent in this region. Similarly, synteny often also breaks down within subtelomeric regions, and these regions previously have been implicated in the development of species-specific adaptations and niche specification (Galagan *et al.* 2005, Moran *et al.* 2011, Zhao *et al.* 2014). Thus, the *F. circinatum*-specific 12 000 bp located in a synteny break point is probably a consequence of the dynamic processes allowing genetic innovation in the telomeric regions of fungal chromosomes.

The genomic region in which the *F. circinatum* growth marker is located is predicted to be involved in producing proteins that have a diverse range of cellular, biological

and metabolic functions. Previous studies on growth rate variation in *F. circinatum* and *F. temperatum* showed that *F. circinatum* grows significantly faster than *F. temperatum* at 25 °C on solid media (De Vos *et al.* 2011). This QTL marker was also significantly correlated with growth rate variation amongst the F1 progeny of an interspecific cross between *F. circinatum* and *F. temperatum* (De Vos *et al.* 2011). In our study we showed that, comparable to the highly-variable telomeric regions in *F. fujikuroi* isolates (Chiara *et al.* 2015), this genomic region is particularly enriched for genes involved in carbohydrate metabolism, metabolite transportation and transcriptional regulation. This adaption may have been brought about through the combination of enhanced substrate transport and carbon metabolism that is further supported by tight, species-specific transcriptional regulation (Proctor *et al.* 2009). Moreover, the clustering and possible co-regulation of these genes may assist this fungus to grow faster at higher temperatures (De Vos *et al.* 2011), in a species-specific manner.

Examination of the genetic makeup of the subtelomere of *F. circinatum*'s Chromosome 3 allowed further insight regarding the evolution of such species-specific loci. Interspecific comparisons between homologous regions of *F. circinatum* and *F. temperatum* suggests that the differences in their TEs acquisition occurred in a species-specific manner. Transposable elements, specifically Retro- and DNA transposons, seemed to be confined to the supposed distal telomeric region of *F. temperatum*, whereas more TE integration in homologous *F. circinatum* regions continued into the adjacent telomere-proximal gene regions. Moreover, *F. circinatum*-specific TE acquisition also seemed to correlate with the location of the unique region. Previous studies established that *F. circinatum* and *F. temperatum* share a recent common ancestor (De Vos *et al.* 2014). Both the *F. circinatum*-specific TE acquisition and unique gene region were thus acquired after the divergence of these two species. It therefore stands to reason that the acquisition of the unique gene region probably involved a TE-mediated mechanism (see below). Future analysis of this region should seek to determine whether its acquisition coincided with (or potentially facilitated) the emergence of the pitch canker fungus as a separate species.

The introduction of a unique gene region into the *F. circinatum* genome may have been brought about by means of a number of possible mechanisms. It is generally thought that the repeat-rich nature of the distal and proximal telomeric regions of chromosomes frequently induce ectopic and non-homologous recombination allowing for species-specific gene gains (Davière *et al.* 2001, Chow *et al.* 2012, Starnes *et al.* 2012). However, variable genomic regions may be more susceptible to TE invasion through non-homologous recombination. The more extensive, species-specific repeat sequences and TE acquisition within the telomeric-proximal region of *F. circinatum* may have facilitated such events allowing the species-specific gene gains within this region.

The findings of this study suggest that the genes encoded on the *F. circinatum*-specific region of Chromosome 3 did not result from internal duplications, but rather from HGT. These five genes have polyphyletic origins as they are derived from more than one independent evolutionary

ancestor. Perhaps the most striking is gene FCIRG_04558 (encoding a class III aminotransferase) that share a recent common ancestor with bacteria. In fact, our data suggest that only two independent introductions of a FCIRG_04558 homolog have so far occurred in fungi (i.e., into unrelated lineages represented by *F. pedrosoi* [Eurotiomycetes] and *F. circinatum* [Sordariomycetes]). Also, the species-specific genes showed marked differences in G+C content compared to that of the surrounding gene regions, the rest of Chromosome 3, and the remainder of the *F. circinatum* genome. Interestingly, the lack of introns in the *F. pedrosoi* gene, together with the higher G+C content, would also fit the scenario of bacterial ancestry implied by the phylogeny. These findings are thus in line with the view that similarities in nucleotide composition of xenologs reflect features of both donor and recipient genomes involved in HGT (Lawrence & Ochman 1998).

This study has provided new insights into the origin and evolution of genes encoded within a locus implicated in growth rate regulation of the pitch canker fungus *F. circinatum* (De Vos et al. 2011). A main hypothesis emerging from our work is that the dynamic evolutionary processes associated with subtelomeric regions likely facilitated the emergence of the *F. circinatum*-specific sequence, which in turn enabled differentiation and adaptation of the fungus in a species-specific manner. Details regarding the precise evolutionary mechanisms involved in the origin of this *F. circinatum*-specific locus might become apparent when the genomes of *Fusarium* species with more recent common ancestry to that of *F. circinatum* are investigated. Additionally, establishing the functional relevance of each of the species-specific proteins identified in this study will be the focus of future studies.

ACKNOWLEDGMENTS

This work is based on research funded and supported in part by the South African National Department of Science and Technology (DST), National Research Foundation (NRF) and Technology and Human Resources of Industry Programme (THRIP), as well as the University of Pretoria and the Tree Protection Cooperative Programme (TPCP). The Grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by NRF supported research are that of the author(s) and that the NRF accepts no liability whatsoever in this regard.

REFERENCES

- Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104–2105.
- Altschul SF, Madden TL, Schäffer AA, Zhang JJ, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.
- Alves FL, Ribeiro MA, Hahn RC, DeMelo M, Teixeira M, De Camargo ZP, et al. (2014) Transposable elements and two other molecular markers as typing tools for the genus *Paracoccidioides*. *Medical Mycology*: myu074.
- Behnsen J, Hartmann A, Schmalzer J, Gehrke A, Brakhage AA, et al. (2008) The opportunistic human pathogenic fungus *Aspergillus fumigatus* evades the host complement system. *Infection and Immunity* **76**: 820–827.
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**: 573–580.
- Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, et al. (2013) antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*: gkt449.
- Boetzer M, Pirovano W (2012) Towards almost closed genomes with GapFiller. *Genome Biology* **13**: R56.
- Braun EL, Halpern AL, Nelson MA, Natvig DO (2000) Large-scale comparison of fungal sequence information: mechanisms of innovation in *Neurospora crassa* and gene loss in *Saccharomyces cerevisiae*. *Genome Research* **10**: 416–430.
- Brown JR, Doolittle WF (1999) Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *Journal of Molecular Evolution* **49**: 485–495.
- Chiara M, Fanelli F, Mulè G, Logrieco AF, Pesole G, et al. (2015) Genome sequencing of multiple isolates highlights subtelomeric genomic diversity within *Fusarium fujikuroi*. *Genome Biology and Evolution* **7**: 3062–3069.
- Chow EW, Morrow CA, Djordjevic JT, Wood IA, Fraser JA (2012) Microevolution of *Cryptococcus neoformans* driven by massive tandem gene amplification. *Molecular Biology and Evolution* **29**: 1987–2000.
- Chuma I, Isobe C, Hotta Y, Ibaragi K, Futamata N, et al. (2011) Multiple translocation of the AVR-Pita effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. *PLoS Pathogens* **7**: e1002147.
- Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, et al. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genetics* **5**: e1000618.
- Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G (2015) Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biology* **16**: 50.
- Cuomo CA, Güldener U, Xu JR, Trail F, Turgeon BG, et al. (2007) The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **317**: 1400–1402.
- Daboussi MJ, Capy P (2003) Transposable elements in filamentous fungi. *Annual Reviews in Microbiology* **57**: 275–299.
- Davière, JM, Langin T, Daboussi MJ (2001) Potential role of transposable elements in the rapid reorganization of the *Fusarium oxysporum* genome. *Fungal Genetics and Biology* **34**: 177–192.
- De Vos, L, Myburg AA, Wingfield MJ, Desjardins AE, Gordon TR, et al. (2007) Complete genetic linkage maps from an interspecific cross between *Fusarium circinatum* and *Fusarium subglutinans*. *Fungal Genetics and Biology* **44**: 701–714.
- De Vos L, Van der Nest MA, Van der Merwe NA, Myburg AA, Wingfield MJ, et al. (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.
- De Vos, L, Steenkamp ET, Martin SH, Santana, QC, Fourie G, et al. (2014) Genome-wide macrosynteny among *Fusarium* species in

- the *Gibberella fujikuroi* complex revealed by amplified fragment length polymorphisms. *PLoS One* **9**: e114682.
- Desjardins A, Plattner R, 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.
- Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, *et al.* (2008) Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genetics* **4**: e1000046.
- Flint J, Thomas K, Micklemz G, Raynhaml H, Clarkl K, *et al.* (1997) The relationship between chromosome structure and function at a human telomeric region. *Nature Genetics* **15**: 252–257.
- Fulnečková, J, Ševčíková T, Fajkus J, Lukešová A, Lukeš M, *et al.* (2013) A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution* **5**: 468–483.
- Gac M, Giraud T (2008) Existence of a pattern of reproductive character displacement in *Homobasidiomycota* but not in *Ascomycota*. *Journal of Evolutionary Biology* **21**: 761–772.
- Galagan JE, Calvo SE, Cuomo C, Ma L-J, Wortman JR, *et al.* (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105–1115.
- Garcia-Pedrajas MD, Roncero MIG (1996) A homologous and self-replicating system for efficient transformation of *Fusarium oxysporum*. *Current Genetics* **29**: 191–198.
- Gardiner DM, Kazan K, Manners JM (2013) Cross-kingdom gene transfer facilitates the evolution of virulence in fungal pathogens. *Plant Science* **210**: 151–158.
- Geiser DM, Aoki T, Bacon CW, Baker SE, Bhattacharyya MK, *et al.* (2013) One fungus, one name: defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology* **103**: 400–408.
- Glenn AE, Davis CB, Gao M, Gold SE, Mitchell TR, *et al.* (2016) Two horizontally transferred xenobiotic resistance gene clusters associated with detoxification of benzoxazolinones by *Fusarium* species. *PLoS One* **11**: e0147486.
- Goodwin SB, M'barek SB, Dhillon, B Wittenberg AH, Crane CF, *et al.* (2011) Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genetics* **7**: e1002070.
- Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, *et al.* (2013) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: D699–04.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98
- Hansen FT, Gardiner DM, Lysøe E, Fuertes PR, Tudzynski B, *et al.* (2015) An update to polyketide synthase and non-ribosomal synthetase genes and nomenclature in *Fusarium*. *Fungal Genetics and Biology* **75**: 20–29.
- Hepting GH, Roth ER (1946) Pitch canker, a new disease of some southern pines. *Journal of Forestry* **44**: 742–744.
- Herron DA, Wingfield MJ, Wingfield BD, Rodas C, Marincowitz S, *et al.* (2015) Novel taxa in the *Fusarium fujikuroi* species complex from *Pinus* spp. *Studies in Mycology* **80**: 131–150.
- Hoff KJ, Stanke M (2013) WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**: W123–128.
- Jaramillo VD, Sukno SA, Thon MR (2015) Identification of horizontally transferred genes in the genus *Colletotrichum* reveals a steady tempo of bacterial to fungal gene transfer. *BMC Genomics* **16**: 2.
- Kohany O, Gentles AJ, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7**: 474.
- Lawrence JG, Ochman H (1998) Molecular archaeology of the *Escherichia coli* genome. *Proceedings of the National Academy of Sciences, USA* **95**: 9413–9417.
- Leslie JF, Summerell BA (2006) *The Fusarium Laboratory Manual*. Ames, Iowa: Wiley Online Library.
- Li W, Cowley A, Uludag M, Gur T, McWilliam H, *et al.* (2015) The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*: gkv279.
- Lysøe E, Harris LJ, Walkowiak S, Subramaniam R, Divon HH, *et al.* (2014) The genome of the generalist plant pathogen *Fusarium avenaceum* is enriched with genes involved in redox, signaling and secondary metabolism. *PLoS One* **9**: e112703.
- Ma L-J, Van der Does HC, Borkovich KA, Coleman JJ, Daboussi M.-J, *et al.* (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* **464**: 367–373.
- Möller E, Bahnweg G, Sandermann H, Geiger H (1992) A simple and efficient protocol for isolation of high molecular weight DNA from filamentous fungi, fruit bodies, and infected plant tissues. *Nucleic Acids Research* **20**: 6115–6116.
- Moran GP, Coleman DC, Sullivan DJ (2011) Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. *Eukaryotic Cell* **10**: 34–42.
- Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323–329.
- Niehaus EM, Münsterkötter M, Proctor RH, Brown DW, Sharon A, *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.
- O'Donnell K, Rooney AP, Proctor RH, Brown DW, McCormick SP, *et al.* (2013) Phylogenetic analyses of *RPB1* and *RPB2* support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genetics and Biology* **52**: 20–31.
- Papadopoulos JS, Agarwala R (2007) COBAL: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**: 1073–1079.
- Proctor RH, McCormick SP, Alexander NJ, Desjardins AE (2009) Evidence that a secondary metabolic biosynthetic gene cluster has grown by gene relocation during evolution of the filamentous fungus *Fusarium*. *Molecular Microbiology* **74**: 1128–1142.
- Raffaele S, Kamoun S (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* **10**: 417–430.
- Rombauts S, Van de Peer Y, Rouzé P (2003) AFLPinSilico, simulating AFLP fingerprints. *Bioinformatics* **19**: 776–777.
- Scaufflaire J, Gourgue M, Munaut F (2011) *Fusarium temperatum* sp. nov. from maize, an emergent species closely related to *Fusarium subglutinans*. *Mycologia* **103**: 586–597.
- Sieber CM, Lee W, Wong P, Münsterkötter M, Mewes HW, *et al.* (2014) The *Fusarium graminearum* genome reveals more secondary metabolite gene clusters and hints of horizontal gene transfer. *PLoS One* **9**: e110311.

- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, et al. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Research* **19**: 1117–1123.
- Skamnioti P, Furlong RF, Gurr SJ (2008) Evolutionary history of the ancient cutinase family in five filamentous ascomycetes reveals differential gene duplications and losses and in *Magnaporthe grisea* shows evidence of sub-and neo-functionalization. *New Phytologist* **180**: 711–721.
- Spanu PD, Abbott JC, Amselem J, Burgis TA, Soanes DM, et al. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **330**: 1543–1546.
- Sperschneider J, Gardiner DM, Thatcher LF, Lyons R, Singh KB, et al. (2015) Genome-wide analysis in three *Fusarium* pathogens identifies rapidly evolving chromosomes and genes associated with pathogenicity. *Genome Biology and Evolution*: evv092.
- Starnes JH, Thornbury DW, Novikova OS, Rehmeyer CJ, Farman ML (2012) Telomere-targeted retrotransposons in the rice blast fungus *Magnaporthe oryzae*: agents of telomere instability. *Genetics* **191**: 389–406.
- Steenkamp ET, Makhari OM, Coutinho TA, Wingfield BD, Wingfield MJ (2014) Evidence for a new introduction of the pitch canker fungus *Fusarium circinatum* in South Africa. *Plant Pathology* **63**: 530–538.
- Steenkamp ET, Wingfield BD, Coutinho TA, Wingfield MJ, Marasas WFO (1999) Differentiation of *Fusarium subglutinans* f. sp. *pini* by histone gene sequence data. *Applied and Environmental Microbiology* **65**: 3401–3406.
- Stewart JE, Abdo Z, Glenn AE (2014) Gene clusters FDB1 and FDB2 in *Fusarium verticillioides* were acquired through multiple horizontal gene transfer events. *Phytopathology* **104**: 114
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.
- Tarailo-Graovac M, Chen N (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*: **4.10**: 11–14.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et al. (2012) Primer3 – new capabilities and interfaces. *Nucleic Acids Research* **40**: e115.
- Van der Nest MA, Beirn LA, Crouch JA, Demers JE, De Beer, ZW, et al. (2014) IMA Genome-F 3: Draft genomes of *Amanita jacksonii*, *Ceratocystis albifundus*, *Fusarium circinatum*, *Huntia omanensis*, *Leptographium procerum*, *Rutstroemia sydowniana*, and *Sclerotinia echinophila*. *IMA Fungus* **5**: 473–486.
- Waalwijk C, Van der Lee T, De Vries, I, Hesselink T, Arts J, et al. (2004) Synteny in toxigenic *Fusarium* species: the fumonisin gene cluster and the mating type region as examples. *European Journal of Plant Pathology* **110**: 533–544.
- Wiemann P, Sieber CM, Von Bargen KW, Studt L, Niehaus EM, 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.
- Wingfield BD, Steenkamp ET, Santana QC, Coetzee MPA, Bam S, et al. (2012) First fungal genome sequence from Africa: a preliminary analysis. *South African Journal of Science* **108**: 1–2.
- Wisecaver JH, Slot JC, Rokas A (2014) The evolution of fungal metabolic pathways. *PLoS Genetics* **10**: e1004816.
- Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**: 847–848.
- 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.