doi:10.5598/imafungus.2018.09.01.03 IMA FUNGUS · 9(1): 27–36 (2018)

ARTICLE

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

Stephanie Van Wyk, Brenda D. Wingfeld, 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 (Scaufaire *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-specifc 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-specifc 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-specifc 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-specifc 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 identifed in a genetic linkage map of an interspecifc cross between *F. circinatum* and *F. temperatum* (De Vos *et al.* 2007, 2011). For this purpose, our study had four specifc objectives. Firstly, we located the genetic marker linked to growth rate variation (i.e. marker AT/AC-625bh) in the genome of *F. circinatum* (Wingfeld *et al.* 2012) and identifed the genes encoded in the region underlying it by making use of various *in silico* approaches. Secondly, the identifed 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 identifed 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 (Wingfeld *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 identifed region was evaluated using various sequence alignments and phylogenetic analyses. These fne-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 fltered 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* (Wingfeld *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* (Wingfeld *et al.* 2012) was determined as described previously (De Vos *et al.* 2014). This was done with *in silico* Amplifed Fragment Length Polymorphism (AFLP) analysis using AFLPinSilico v2 (Rombauts *et al.* 2003), which involved the use of simulated restriction enzyme digestion profles for the entire genome of *F. circinatum*. The analysis used the restriction sites for *Eco*RI (GAATT↓C) and *Mse*I (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* (Wingfeld *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 COBALT (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 (Wingfeld *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*.

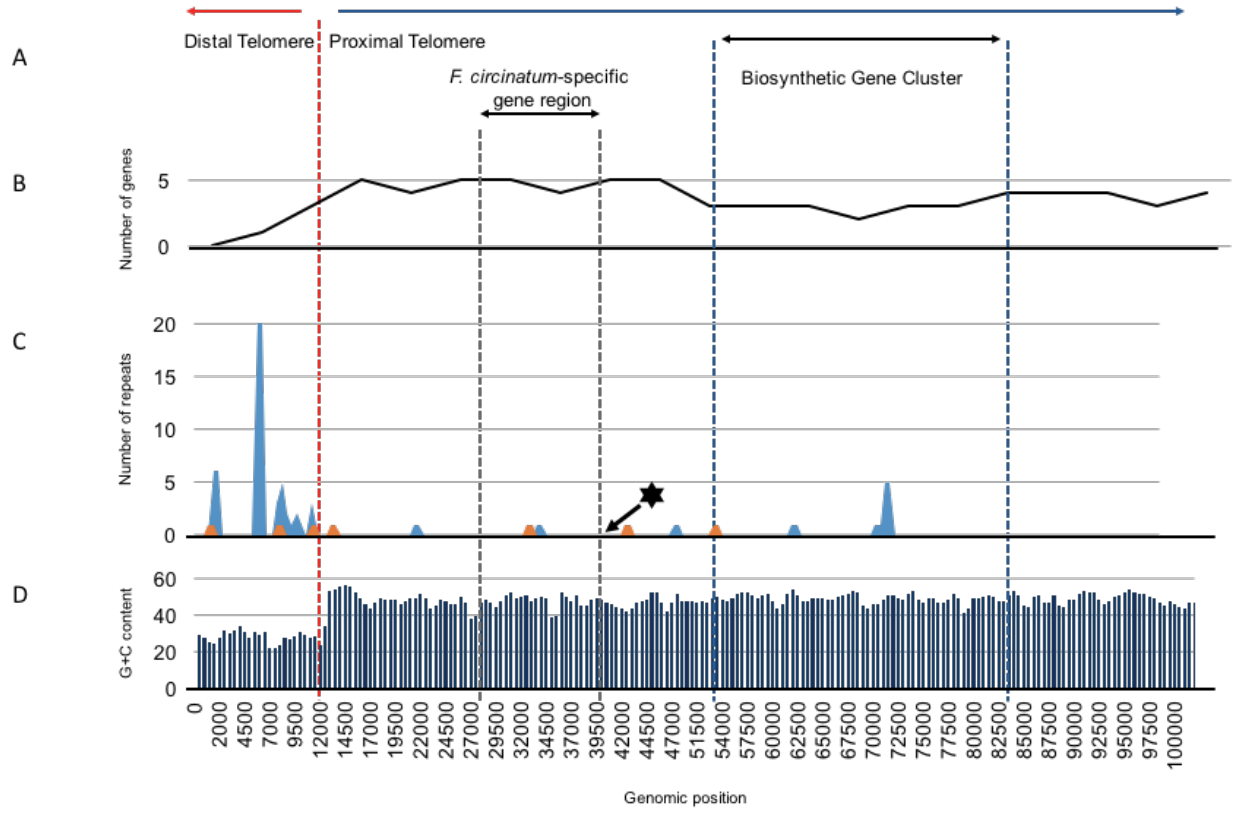

Fig. 1. Genomic features of the first 100 000 bp of Chromosome 3 of *Fusarium circinatum* (FSP34). (A) This region corresponds to the subtelomere of the chromosome. (B) Line graph illustrating the change in gene count determined through a 10 000 bp sliding window at 5 000 bp increments. (C) Chart showing 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

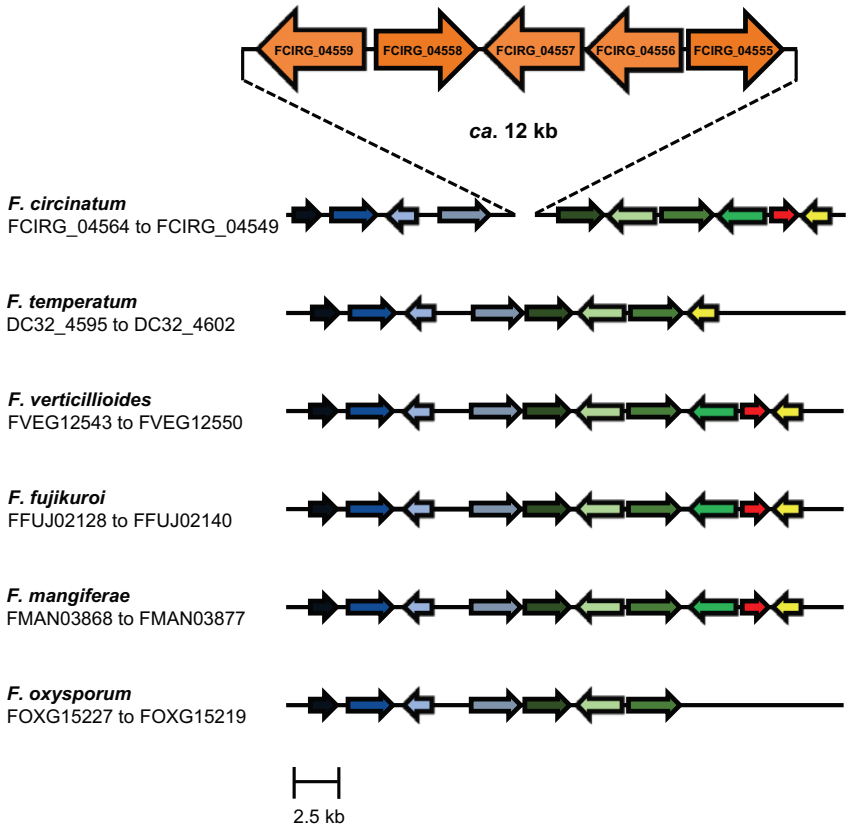

Fig. 2. Gene content and organization of the region containing the QTL marker associated with growth rate variation in *Fusarium circinatum*. Gene position and orientation are indicated with block arrows. Orange arrows illustrate genes only encoded in *F. circinatum*. Gene names are indicated below each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

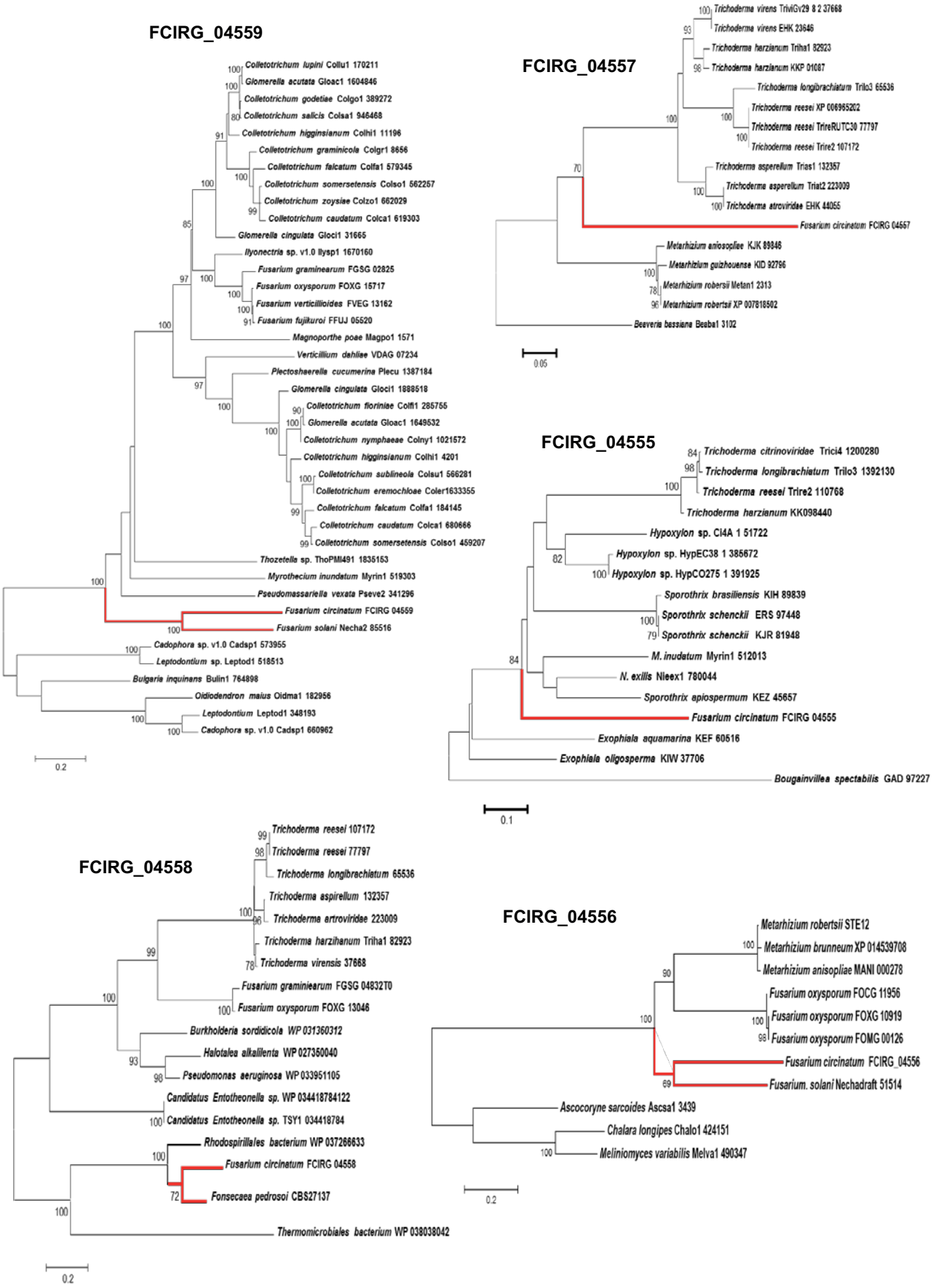

Fig. 3. Maximum likelihood trees constructed from the inferred *Fusarium circinatum* species-specifc proteins FCIRG_04559, FCIRG_04558, FCIRG_04557, FCIRG_04556 and FCIRG_04555. Branches indicated in red show the position and closest relative or clade of *F. circinatum* in the fve protein trees. Each 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*-specifc 12 000 bp region and its genes.

The non-vertical inheritance of the *F. circinatum*-specifc 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 signifcantly 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*-specifc gene region must have been introduced from elsewhere.

Detailed examination of the region up- and downstream of the *F. circinatum*-specifc 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-specifc adaptations and niche specifcation (Galagan *et al.* 2005, Moran *et al.* 2011, Zhao *et al.* 2014). Thus, the *F. circinatum*-specifc 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 signifcantly faster than *F. temperatum* at 25 °C on solid media (De Vos *et al.* 2011). This QTL marker was also signifcantly correlated with growth rate variation amongst the F1 progeny of an interspecifc 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-specifc 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-specifc manner.

Examination of the genetic makeup of the subtelomere of *F. circinatum*'s Chromosome 3 allowed further insight regarding the evolution of such species-specifc loci. Interspecifc comparisons between homologous regions of *F. circinatum* and *F. temperatum* suggests that the differences in their TEs acquisition occurred in a species-specifc manner. Transposable elements, specifcally Retro- and DNA transposons, seemed to be confned 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*-specifc 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*-specifc 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-specifc 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-specifc repeat sequences and TE acquisition within the telomeric-proximal region of *F. circinatum* may have facilitated such events allowing the species-specifc gene gains within this region.

The fndings of this study suggest that the genes encoded on the *F.* circinatum-specifc region of Chromosome 3 did not result from internal duplications, but rather from HGT. These fve 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-specifc 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 ft the scenario of bacterial ancestry implied by the phylogeny. These fndings are thus in line with the view that similarities in nucleotide composition of xenologs refect 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*-specifc sequence, which in turn enabled differentiation and adaptation of the fungus in a species-specifc manner. Details regarding the precise evolutionary mechanisms involved in the origin of this *F. circinatum*-specifc 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-specifc proteins identifed 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, fndings 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-ft 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, Schmaler 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 fnder: 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 amplifcation. *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 flamentous 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, Wingfeld MJ, Desjardins AE, Gordon TR, *et al.* (2007) Complete genetic linkage maps from an interspecifc 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, Wingfeld MJ, *et al.* (2011) Genetic analysis of growth, morphology and pathogenicity in the F1 progeny of an interspecifc 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**: 865–72.

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, Wingfeld MJ, Wingfeld 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) COBALT: 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.

Scaufaire 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 identifes 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, Wingfeld BD, Wingfeld MJ (2014) Evidence for a new introduction of the pitch canker fungus *Fusarium circinatum* in South Africa. *Plant Pathology* **63**: 530–538.

Steenkamp ET, Wingfeld BD, Coutinho TA, Wingfeld 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. *Phythopathology* **104**: 114

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

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, Huntiella omanensis, Leptographium procerum, Rutstroemia sydowiana,* 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.

Wingfeld 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.

Supplementary Figure S3

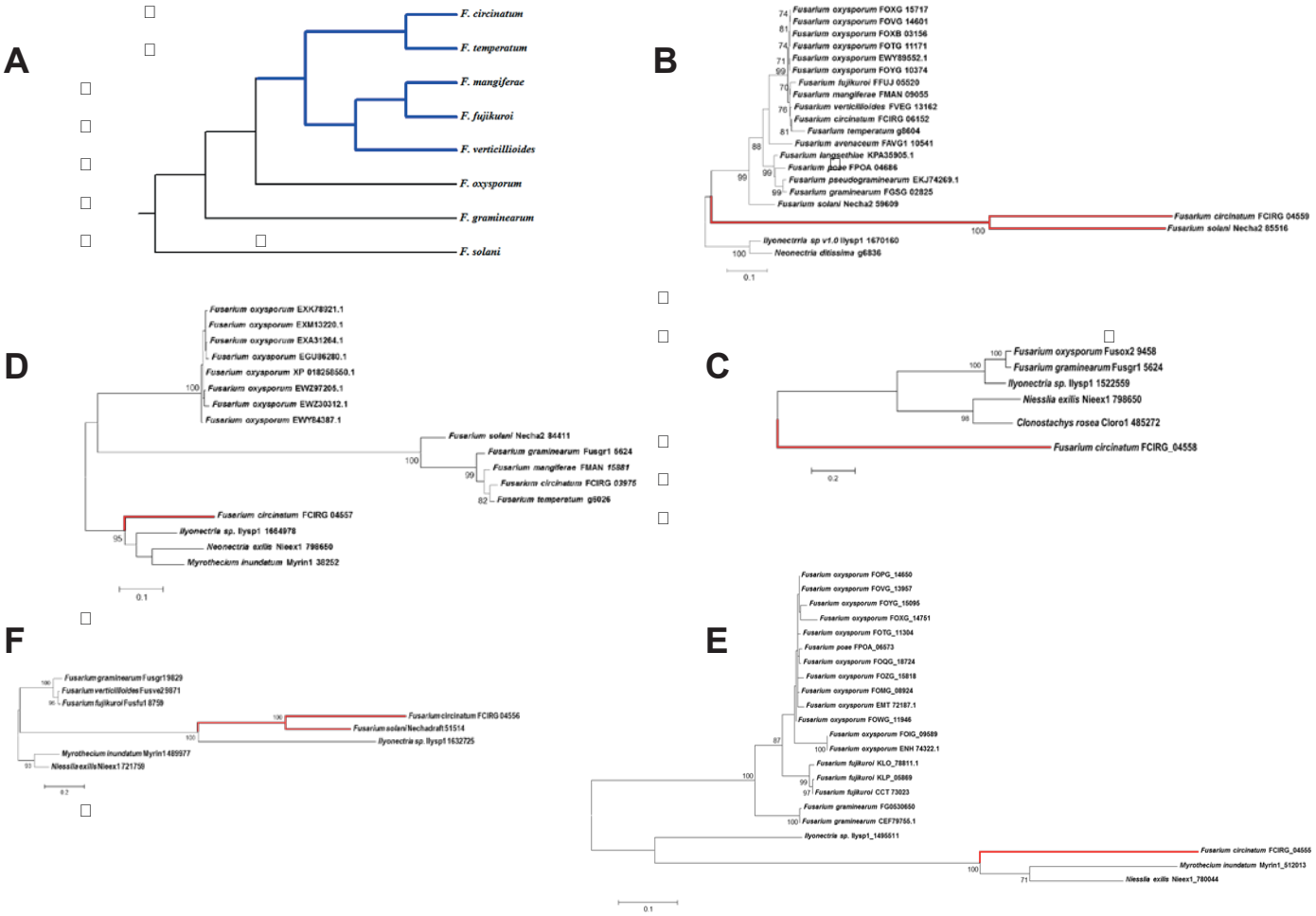

Fig. S3. [illegible]
[illegible] *F. circinatum* [illegible]
Branches indicated in red show the position and closest relative of the respective *F. circinatum* [illegible]
one or more of *Myrothecium inundatum* *Clonostachys rosea* *Niesslia exilis* *Neonectria ditissima* and *Ilyonectria* sp. were included for comparison or [illegible] *Fusarium* [illegible]
F. [illegible] *Fusarium* [illegible]
[illegible] *Fusarium* [illegible]
in the genomes of other *Fusarium* [illegible] *Fusarium*
[illegible] *Fusarium* species.

Supplementary Tables S1-S20

Table S1: List of *Fusarium* isolates used in this study.

Species	Isolate number	Origin
F. circinatum	CMWF 350 (FSP 34)	California, USA
F. circinatum	CMWF 530	Mexico
F. circinatum	CMWF 550	Mexico
F. circinatum	CMWF 560	Mexico
F. circinatum	CMWF 567	Mexico
F. circinatum	CMWF 1221	Mexico
F. circinatum	FCC 4881	Mexico
F. circinatum	FCC 4882	Mexico
F. circinatum	FCC 4883	Mexico
F. circinatum	FCC 4884	Mexico
F. circinatum	FCC 4885	Mexico
F. circinatum	UGIE 8.2.1	Eastern Cape, SA
F. circinatum	UGIE 10	Eastern Cape, SA
F. circinatum	UGIE 17.6	Eastern Cape, SA
F. circinatum	UGIE 27	Eastern Cape, SA
F. circinatum	CMWF 30	Mpumalanga, SA
F. circinatum	CMWF 39	KwaZulu Natal, SA
F. circinatum	CMWF 45	KwaZulu Natal, SA
F. circinatum	CMWF 487	Western Cape, SA
F. circinatum	CMWF 538	Western Cape, SA
F. circinatum	CMWF 513	Western Cape, SA
F. circinatum	CMWF 659	Western Cape, SA
F. temperatum	CMWF 1206	Texcoco, Mexico
F. mangiferae	CMWF1214	Ginosar, Israel
F. fujikuroi	CMWF1539	Taiwan, China
F. verticillioides	CMWF 1227	California, USA

CMWF, FCC, UGIE = Obtainable from the Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, South Africa

Table S2: Sequence, annealing temperature (Tm) and description of target genomic regions of primers used in this study.

Primer name	Primer sequence 5' → 3'	Target gene region
fwd60-59	TCCCGTCGCAGTTATGTCTT	Intergenic region from FCIRG_04560 to FCIRG_04559
rvs60-59	GGATCTTCTTTCGCAGCCTG	
fwd59-58	CAGAGCACCTAACCTTTCGC	Intergenic region from FCIRG_04559 to FCIRG_04558
rvs59-58	CTGGGGCAGGGTCTTATCAT	
fwd58-57	TCTAAGACCCCTGCTCCTCT	Intergenic region from FCIRG_04558 to FCIRG_04557
rvs58-57	TCGAGTGTGAAGGGTGTCAT	
fwd57-56	TCGAGTGTGAAGGGTGTCAT	Intergenic region from FCIRG_04557 to FCIRG_04556
rvs57-56	AGCTGTGTCTGATGCCTCAA	
fwd56-55	TCATCGCCGAGTGACTATCC	Intergenic region from FCIRG_04556 to FCIRG_04555
rvs56-55	CAGATGATGAGGGTGCTGGA	
fwd55-54	CATCATTGCGGGCTTGACTA	Intergenic region from FCIRG_04555 to FCIRG_04554
rvs55-54	TGCTCCGCCCATTACTAAGA	
Fusarium FWD	GTTGGTACGAAACAGCAGCA	Synteny break point corresponding gene region in *F. circinatum* intergenic region from FCIRG_04560 to FCIRG_04554
Fusarium RVS	ATTCGGGATTGGGGTTCAGT	

Table S3: List of database and sequenced fungal genomes investigated in this study.

Database source	Species name	Strain	Reference
NCBI	*F. circinatum*	FSP34	Wingfield *et al.* 2012
NCBI	*F. temperatum*	CMWF389	Wingfield *et al*. 2015
NCBI	*F. mangiferae*	MRC7560	Niehaus et al. 2017
JGI[a]	*F. fujikuroi*	IMI 58289	Weimann *et al*. 2013
NCBI	*F. verticillioides*	7600	Cuomo *et al*. 2007
JGI[a]	*F. oxysporum* f. sp. *lycopersici*	4287, race 2, VCG 0030	Ma *et al*. 2011
NCBI	*F. graminearum*	PH-1, NRRL 31084	Cuomo *et al*. 2007
JGI[a]	*F. solani*	Mating Population VI (MPVI) 77-13-4	Coleman *et al.* 2009

[a] JGI, Joint Genome Institute (http://genome.jgi.doe.gov/programs/fungi/index.jsf) (Grigoriev *et al.* 2013)

Table S4: InterPro terms and protein family membership (PFAM) of the proteins predicted within the genomic region underlying the genetic marker (AT/AC-625bh) of the *F. circinatum* genome assembly (39500-48000bp on chromosome 3).

Protein name	Description	InterPro term	Protein family
FCIRG_04548	Protein of unknown function	None predicted	None predicted
FCIRG_04549	Amidohydrolase	IPR006992	PF04909
FCIRG_04550	Rhamnose mutarotase	IPR008000	PF05336
	Dimeric alpha-beta barrel	IPR011008	SSF54909
FCIRG_04551	Major facilitator super family	IPR011701	PF07690
	Major facilitator super family domain	IPR020846	SSF103473
FCIRG_04552	Transcription factor, fungi	IPR007219	PF04082
FCIRG_04553	Mandelate racemase/ Muconate lactonizing enzyme methylaspartate amomonia lyase	IPR001354	None predicted
	Endolase N-terminal domain	IPR029017	None predicted
	Endolase C-terminal domain	IPR029065	SSF51604
	Mandelate racemase/ Muconate lactonizing enzyme methylaspartate amomonia lyase conserved site	IPR013342	PF01188
FCIRG_04554	Aldo/keto reductase	IPR001395	None predicted
	NADP-dependent oxireductase domain	IPR02310	PF00248
FCIRG_04555	Major facilitator super family	IPR011701	PF07690
	Major facilitator super family domain	IPR020846	SSF103473
FCIRG_04556	Transcription factor, fungi	IPR007219	PF04082
FCIRG_04557	Aminotransferase class III	IPR015424	PF00202
	Pyridoxal phosphate dependent transferase	IPR015424	SSF53383
	Pyridoxal phosphate dependent transferase major region subdomain 1	IPRO015421	None predicted
	Pyridoxal phosphate dependent transferase major region subdomain 2	IPR015422	None predicted
FCIRG_04558	Aminotransferase class III	IPR015424	PF00202
	Pyridoxal phosphate dependent transferase	IPR015424	SSF53383
	Pyridoxal phosphate dependent transferase major region subdomain 1	IPRO015421	None predicted
	Pyridoxal Phosphate dependent transferase major region subdomain 2	IPR015422	None predicted
FCIRG_04559	$Zn(II)_2$-C_6 fungal type DNA binding domain	IPR00138	PF00172
	Transcription factor, fungi	IPR007219	PF04082
FCIRG_04560	Peptidase M14, carboxypeptidase	IPR000834	PF00246
FCIRG_04562	RTA-like protein	IPR007568	PF04479
FCIRG_04563	Six bladed beta-propeller Tol-B-line	IPR011042	None predicted
FCIRG_04564	Protein of unknown function	IPR021369	PFR11204

Table S5: Gene Ontology (GO) terms of genes of interest predicted in the genomic region underlying the genetic-marker (AT/AC-625bh) for *Fusarium circinatum* (39500-48000 bp on chromosome 3).

Gene name	Biological process	GO term	Molecular function	GO term	Cellular component	GO term
FCIRG_04548	None predicted	None predicted	None predicted	None predicted	None predicted	None predicted
FCIRG_04549	Metabolic process	8152	Catalytic activity	3824	None predicted	None predicted
FCIRG_04550	Rhamnose metabolic process	19299	Racemase and epimerase activity acting on carbohydrates and derivatives	16857	Cytoplasm	5737
FCIRG_04551	Transmembrane transport	55085	None predicted	None predicted	Integral component of membrane	16201
FCIRG_04552	Transcription, DNA-template	6351	DNA binding; Zinc ion binding	3677; 827	Nucleus	5634
FCIRG_04553	Metabolic process; Catabolic activity	8152; 9063	Catalytic activity	3824	None predicted	None predicted
FCIRG_04554	None predicted	None predicted	None predicted	None predicted	None predicted	None predicted
FCIRG_04555	Transmembrane transport	55085	None predicted	None predicted	Integral component of membrane	16201
FCIRG_04556	Transcription, DNA-template	6351	DNA binding; Zinc ion binding	3677; 827; 30170	Nucleus	5634
FCIRG_04557	None predicted	None predicted	Transaminase activity; Pyridoxal phosphate dependent binding	3824; 8483	None predicted	None predicted
FCIRG_04558	None predicted	None predicted	Catalytic activity; Transaminase activity; Pyridoxal phosphate dependent binding	3824; 8483; 30170	None predicted	None predicted
FCIRG_04559	Transcription DNA template; Regulation of transcription DNA-template	6351; 6355	Sequence-specific DNA binding RNA polymerase II transcription factor activity; Zinc ion binding	981; 8270	Nucleus	5634
FCIRG_04560	Proteolysis	4181	Metalocarboxy peptidase activity; Zinc ion binding	4181; 8270	None predicted	None predicted
FCIRG_04562	Response to stress	6950	None predicted	None predicted	Integral component of membrane	16201
FCIRG_04563	None predicted	None predicted	None predicted	None predicted	None predicted	None predicted
FCIRG_04564	None predicted	None predicted	None predicted	None predicted	None predicted	None predicted

Table S6: Terminology and description of genes located within the biosynthetic gene cluster of *Fusarium circinatum* investigated in this study (49 000-101 000 bp on Chromosome 3).

Gene name	Location (bp)	InterPro term	Description	GO term	Significance in cluster
FCIRG_03388	138-887	IPR013024	Butirosin biosynthesis	None predicted	Biosynthetic gene
		IPR009288	AIG2-like domain		
FCIRG_03387	1568-3212	None predicted	None predicted	None predicted	
FCIRG_03386	4360-5027	None predicted	None predicted	None predicted	
FCIRG_03385	5074-5293	IPR002085	Alcohol dehydrogenase, zinc type	5514	Biosynthetic gene
		IPR011032	Alcohol dehydrogenase	8270	
		IPR020843	GroES chaperone 10-like domain	16491	
		IPR013154	Polyketidesynthase enoul reductase		
		IPR013114	Alcoholdehydrogenase-C-terminal domain		
		IPR016040	NAD(P) binding domain		
FCIRG_03384	7430-8212	IPR002085	Alcohol dehydrogenase, zinc type	5514	Biosynthetic gene
		IPR011032	Alcohol dehydrogenase	8270	
		IPR020843	GroES chaperone 10-like domain	16491	
		IPR013154	Polyketidesynthase enoul reductase		
		IPR0131149	Alcoholdehydrogenase-C-terminal domain		
		IPR016040	NAD(P) binding domain		
FCIRG_03383	9495-11000	IPR000277	Cys/Met metabolism, pyridoxal phosphate-dependent enzyme	30170	Biosynthetic gene
		IPR015421	Pyridoxal phosphate-dependent transferase, major region, subdomain 1	3824	
		IPR015424	Pyridoxal phosphate-dependent transferase	30170	
		IPR006235	O-acetylhomoserine/O-acetylserine sulfhydrylase	6520; 16765	
FCIRG_03382	11657-15022	IPR000873	AMP-dependent synthetase/ligase	3824; 8152	Biosynthetic gene
		IPR009081	Acyl carrier protein-like		
		IPR013120	Male sterility, NAD-binding		
		IPR016040	NAD(P)-binding domain		
FCIRG_03381	15863-17316	IPR008259	FMN-dependent alpha-hydroxy acid dehydrogenase, active site	16491; 55114	
		IPR012133	Alpha-hydroxy acid dehydrogenase, FMN-dependent	10181; 16491; 55114	
		IPR013785	Aldolase-type TIM barrel	3824	
FCIRG_03380	19521-20821	IPR00138	Zn_2C_6 fungal type DNA binding domain		
		IPR007219	Transcription factor, fungi		
FCIRG_03379	25244-27122	IPR011701	Major facilitator superfamily	16021; 55085	Transport related gene
		IPR020846	Major facilitator superfamily domain		
FCIRG_03378	28330-29954	IPR007219	Transcription factor domain, fungi	3677; 5634; 6351; 8270	
FCIRG_03376	30034-32848	IPR000073	Alpha/Beta hydrolase fold-1		
		IPR029058	Alpha/Beta hydrolase fold		
		IPR013083	Zinc finger, RING/FYVE/PHD-type		
FCIRG_03375	34353-34841	IPR019791	Haem peroxidase, animal		
		IPR001128	Cytochrome P450	5506; 16705; 20037; 55114	
		IPR010255	Haem peroxidase	4601; 6979; 20037; 55114	

Table S7: List of simple repeat sequences identified in the first 100 000 bp of chromosome 3 of *Fusarium circinatum.*

Repeat motif	Orientation of repeat	Begin (bp)	End (bp)	Score[a]	Deletion (%)[b]	Insertion (%)[c]
(TTTCCT)n	+	1247	1278	16	0	3,2
(TTAA)n	+	917	1958	17	6,5	3,8
(TA)n	+	1959	1987	15	6,9	0
(TTAA)n	+	1988	1993	17	6,5	3,8
(ATA)n	+	2117	2166	19	0	8,7
(TAATA)n	+	6736	6798	15	3,2	4,8
(TTTCT)n	+	6843	6881	29	5,1	0
(TA)n	+	8231	8265	15	2,9	2,9
(TAATAC)n	+	8292	8374	16	7,2	2,3
(CTAT)n	+	9294	9356	18	3,2	3,2
(TA)n	+	10826	10851	17	3,9	0
(TACT)n	+	11819	11881	12	4,8	6,5
(TAGTAT)n	+	12097	12141	16	2,2	9,5
(CTG)n	+	22570	22594	12	8	3,9
(AATTA)n	+	35150	35191	13	0	5
(GCGAC)n	+	49266	49291	13	3,9	3,9
(CGCCCAT)n	+	63441	63483	13	0	4,9
(TGTCC)n	+	71676	71702	12	3,7	0
(GTC)n	+	72429	72461	14	0	0

[A] Alignment score of the repeated sequences based on the overall average in matches, mismatches, insertions and deletions between repeated sequences located within the index.
[b] Average percentage of deletions between repeat copies overall.
[c] Average percentage of insertions between repeat copies overall.

Table S8: Summary of tandem repeat sequences identified in the first 100 000 bp of chromosome 3 of *Fusarium circinatum.*

Indices[A]	Consensus repeat size	Copy number	Matches (%)[B]	Indels[C]	Score[D]
6812-6839	14	2	100	0	56
6843-6881	5	8	83	11	55
6842-6881	9	4	87	3	53
6843-6881	14	3	92	0	60
72425-72468	12	4	96	0	79

[A]Genomic region in which repeated sequences are located (bp).

[B] Average percentage of matches between copies repeated sequences within the index.

[C] Average percentage of copies of the repeated sequence containing a insertion or a deletion.

[D] Alignment score of the repeated sequences based on the overall average in matches, mismatches, insertions and deletions between repeated sequences located within the index.

Table S9: Predicted transposable element repeat sequences identified in the first 100 000 bp on chromosome 3 of *Fusarium circinatum* and on the corresponding homologous scaffold 3 of *F. temperatum* relative to the genetic marker AT/AC-625bh.

Species	Location (bp)	TE family	Distance (bp) from genetic marker
Fusarium circinatum (FSP34)	1 949-2 157	LTR-AO	37 605 downstream
	8 291-8 582	GYPSY	31 180 downstream
	12 294-12 388	Copia	27 374 downstream
	14 674-14 728	GYPSY	25 034 downstream
	34 340-34 439	GYPSY	5 323 downstream
	44 387-44 470	LTR-TCN4-I	4 026 upstream
	54 563-54 603	GYPSY PYGGY	14 292 upstream
Fusarium temperatum	37 376-37 418	GYPSY PYGGY	24 208 upstream
	9 734-9 991	DNA MARINER	3 219 downstream
	8 181-9 601	DNA MARINER	3 609 downstream
	5 962-7 818	GYPSY PYGGY	5 392 downstream
	4 732-5 830	DNA MARINER	7 380 downstream
	2 740-3 474	DNA MARINER	9 735 downstream

Table S10: Genomic regions in different *Fusarium oxysporum* strains homologous to that of the examined QTL (36455 bp) region of Chromosome 3 of *Fusarium circinatum*.

Species	Strain[a]	Homologous locus location (bp)	Genes encoded within locus	Genomic region size (bp)
F. oxysporum	fo47	Supercontig 3: 4 778 200-4 797 514	FOZG_06216 to FOZG_06224	19314
	GL57	Supercontig 1: 4 763 370-4 782 686	FOCG_01721 to FOCG_01729	19316
	NRRL	Supercontig 2: 5 004 332-5 023 803	FOYG_16234 to FOYG_03919	19471
	MN25	Supercontig 31: 142 755-162 232	FOWG_16234 to FOWG_16243	19477
	PHW815	Supercontig 33: 50 517-73 324	FOQG_10787 to FOQG_10777	22807
	II5	Supercontig 68: 46 532-66 016	FOIG_16204 to FOIG_16195	19484
	HDV247	Supercontig 2: 4 410 949-4 430 504	FOVG_03768 to FOVG_03777	19555
	Cotton	Supercontig 59: 155 962-175 536	FOTG_13448 to FOTG_13457	16574
	melonis	Supercontig 27: 65 068-84 555	FOMG_16516 to FOMG_16524	19487
	PHW808	Supercontig 108: 748 26-94 388	FOPG_11876 to FOPG_11885	19562

[a] Available from the NCBI

Table S11: Gene and protein features of *Fusarium graminearum* (A) and *Fusarium solani* (B) encoded genes homologous to that of the *Fusarium circinatum* (FSP34) genes investigated in this study.

A

F. circinatum	FCIRG_04562	FCIRG_04560	FCIRG_04554	FCIRG_04553	FCIRG_04552	FCIRG_04551	FCIRG_04550
F. graminearum	FGSG_02084	FGSG_07668	FGSG_12407	FGSG_09257	FGSG_01593	FGSG_08507	FGSG_10131
Location and position (bp)	Supercontig_3.1 6 819 541-6 820 559	Supercontig_3. 44 134 297-4 135 965	Supercontig_3. 23 163 166-3 164 215	Supercontig_3. 6 935 923-937 595	Supercontig_3. 15 249 420-5 250 965	Supercontig_3. 51 615 939-1 616 359	Supercontig_3 7 942 631-944 083
Strand	-	+	+	-	+	-	-
Number of exons	3	2	1	2	4	2	2

B

F. circinatum	FCIRG_04563	FCIRG_04562	FCIRG_04560	FCIRG_04554	FCIRG_04553	FCIRG_04552	FCIRG_04550	FCIRG_04549
F. solani	e_gw1.9.283.1	Scaffold_15	e_gw1.69.10.1	Scaffold 120 unmapped000001	e_gw1.120.8.1	Scaffold 8 chromosome 1_1_00358	e_gw1.2.1103.1	e_gw1.5.1409.1
Location and position (bp)	Scaffold 9 Chromosome 7_10_1 013 498-1 015387	Scaffold 15 Chromosome 12_5_77 554-78 520	Scaffold 69 Chromosome 11_6 12 203-13 867	Scaffold 120 4 241-5 449	Scaffold _120 5 874-7 416	scaffold_8 Chromosome 1_1_568 236-569 835	Scaffold_2_ Chromosome_3_3_1 181 703-1 182 113	Scaffold_5_ Chromosome_5_3_1 150 698-1 151 786
Strand	-	-	+	+	-	+	+	+
Number of exons	1	2	2	1	4	4	2	2

Table S12: The distribution of *Fusarium circinatum*-specific proteins, determined through tBLASTn analyses, amongst the *Fusarium* species analyzed in this study.

Query Gene	Species name	Subject ID	Location and position (bp)	Identity (%)	Alignment length (bp)	Query coverage (%)	E-value	Score
FCIRG_04559	*Fusarium circinatum*	FCIRG_06152	Chromosome 6: 419 433-421 570	41,54	2 137	89,3	1,08E-173	810
	Fusarium temperatum	DC32_8604	Scaffold 6: 3 614 255-3 616 201	41,07	1 947	97,2	1,29E-157	1314
	Fusarium mangiferae	FMAN_09055	Contig 1688: 42 837-44 974	41,07	1 947	97,2	2,68E-155	1296
	Fusarium fujikuroi	FFUJ_05520	Chromosome 6: 472 080-474 225	45,8	903	87,1	8,35E-129	1287
	Fusarium verticillioides	FVEG_13162	Supercontig 22: 252 692-253 840	47,8	646	92,6	1,35E-145	1227
	Fusarium oxysporum	FOXG_15717	Supercontig 2.29: 296 837-299 006	48,3	609	84,9	9,72E-145	1199
	Fusarium graminearum	FGSG_02825	Supercontig 3.2: 5 277 911-5 280 156	47,4	809	85,4	1,21E-114	436
	Fusarium solani	Necha2_85516	Chromosome 10: Scaffold 82: 321 452-323 813	45	857	87,8	0	1653
		Necha2_59609	Chromosome 3: Scaffold 31: 244 017-246 394	43,4	569	83,4	1,08E-125	413
FCIRG_04558	*Fusarium circinatum*	None identified	-	-	-	-	-	-
	Fusarium temperatum	None identified	-	-	-	-	-	-
	Fusarium mangiferae	None identified	-	-	-	-	-	-
	Fusarium fujikuroi	None identified	-	-	-	-	-	-
	Fusarium verticillioides	None identified	-	-	-	-	-	-
	Fusarium oxysporum	FOXG_10346	Supercontig 2.17: 1 278 078-1 279 412	40,2	552	80,2	4,40E-75	714
	Fusarium graminearum	FGSG_04832	Supercontig 3: 400 164-401 498	40,01	407	81,1	1,03E-73	700
	Fusarium solani	None identified	-	-	-	-	-	-
FCIRG_04557	*Fusarium circinatum*	FCIRG_03975	Chromosome 4: 105 978-107 309	44,85	1 272	88,9	1,81E-108	913
	Fusarium temperatum	DC32_6026	Scaffold 5: 101 205-102 506	44,93	1 302	91,3	7,02E-112	939
	Fusarium mangiferae	FMAN_15881	Contig_2896: 251 56-262 46	44,93	1 302	91,3	3,63E-110	925
	Fusarium fujikuroi	None identified	-	-	-	-	-	-
	Fusarium verticillioides	None identified	-	-	-	-	-	-
	Fusarium oxysporum	FOXG_17530	Supercontig 2.25: 93 747-95 379	64,9	391	80,4	1,37E-152	1252
		FOXG_13046	Supercontig 2.17: 1 278 079-1 279 412	51,5	371	83,6	7,64E-116	985
	Fusarium graminearum	FGSG_04832	Supercontig 3.3: 400 164-401 498	52,8	369	83,1	7.6 E-118	999
	Fusarium solani	Necha2_8441	Chromosome 9: Scaffold 27: 269 225-270 547	52	993	84,4	1,03E-122	993
FCIRG_04556	*Fusarium circinatum*	None identified	-	-	-	-	-	-
	Fusarium temperatum	None identified	-	-	-	-	-	-
	Fusarium mangiferae	None identified	-	-	-	-	-	-
	Fusarium fujikuroi	None identified	-	-	-	-	-	-
	Fusarium verticillioides	None identified	-	-	-	-	-	-
	Fusarium oxysporum	FOXG_10919	Supercontig 2.14: 288 749-291 052	54	552	72	0	1530
	Fusarium graminearum	None identified	-	-	-	-	-	-
	Fusarium solani	Necha2_51514	Chromosome 11:117 201-118 955	58	672	89,7	0	1676
FCIRG_04555	*Fusarium circinatum*	None identified	-	-	-	-	-	-
	Fusarium temperatum	DC32_11454	Scaffold 10: 97 130-98 567	44,26	1 335	93,2	4,29E-64	706
	Fusarium mangiferae	FMAN_12630	Contig 2325: 1 512-3 067	42,26	1 335	93,5	4,29E-64	792
	Fusarium fujikuroi	FFUJ_12922	Chromosome 8: 3 180 363-3 181 919	44,7	1334	67,8	9,77E-93	819

Fusarium verticillioides	None identified	-	-	-	-	-	-
Fusarium oxysporum	None identified	-	-	-	-	-	-
Fusarium graminearum	None identified	-	-	-	-	-	-
Fusarium solani	Necha_85217	Chromosome 4: 321 425-323 813	43,9	775	78,7	4,67E-09	775
	Necha_5570	Chromosome 11: 243 660-244 996	42,5	702	79,1	6,32E-72	702

Table S13: The distribution of *Fusarium circinatum*-specific proteins, determined through BLASTp analyses, amongst the *Fusarium* species included in the local constructed database from sequences that were originally catalogued in the Broad Institute's database for the Fusarium Comparative Project.

Query Gene	Species name	Gene ID	Location and position (bp)	Score	E-value	Alignment length[a]	Identities[a]	Positives[a]
FCIRG_04559	*Fusarium oxysporum Cotton*	FOTG_11171.1	Supercontig 1.35: 314 743-317 112	520,78	0	703	295	407
	Fusarium oxysporum NRRL3293	FOYG_10374.1	Supercontig 1.6: 3 640 983-3 643 550	517,31	0	703	294	406
	Fusarium oxysporum PHW808	FOPG_14849.1	Supercontig 1.206: 4 190-6 559	517,31	0	703	294	406
	Fusarium oxysporum f. sp. melon	FOMG_15543.1	Supercontig 1.19: 322 322-324 889	517,31	0	703	294	406
	Fusarium oxysporum Fo5176	FOXB_03685.1	Contig01278: 284 092-286 753	517,31	0	703	294	406
	Fusarium oxysporum 4287 (FO2)	FOXG_15717.3	Supercontig 29: 296 531-299 098	517,31	0	703	294	406
	Fusarium oxysporum II5	FOIG_12535.1	Supercontig 27: 152 889-155 115	517,31	0	703	294	406
	Fusarium oxysporum PHW815	FOQG_12489.1	Supercontig 49: 178 362-180 731	517,31	0	703	294	406
	Fusarium oxysporum MN25	FOWG_12941.1	Supercontig 14: 1 031 420-1 033 987	517,31	0	703	294	406
	Fusarium oxysporum CL57	FOCG_08754.1	Supercontig 7: 475 582-478 149	517,31	0	703	294	406
	Fusarium oxysporum Fo47	FOZG_08162.1	Supercontig 5: 191 534-194 101	517,31	0	703	294	406
	Fusarium oxysporum HDV247	FOVG_14601.1	Supercontig 15: 52 302-54 671	515	0	703	293	405
	Fusarium verticillioides 7600	FVEG_13162.1	Supercontig 22: 252 692-253 840	575	0	573	256	356
	Fusarium graminearum PH-1	FGSG_02825.3	Supercontig 2: 5 277 911-5 280 156	436	0	578	246	342
FCIRG_04558	*Fusarium oxysporum 4287*	FOXG_13046.3	Supercontig 17: 1 278 078-1 279 412	271,52	0	407	144	231
	Fusarium graminearum PH-1	FGSG_04832.3	Supercontig 3: 400 164-401 498	261,92	0	407	144	233
FCIRG_04557	*Fusarium oxysporum MN25*	FOWG_04399.1	Supercontig 3: 3 041 609-3 042 979	518,85	0	445	258	316
	Fusarium oxysporum HDV247	FOVG_17336.1	Supercontig 51: 37 054-38 424	517,31	0	445	258	314
	Fusarium oxysporum f. sp. melon	FOMG_16055.1	Supercontig 22: 304 949-306 319	516,54	0	445	257	315
	Fusarium oxysporum 4287 (FO2)	FOXG_17530.3	Supercontig 52: 93 747-95 117	516,54	0	445	257	315
	Fusarium oxysporum NRRL3293	FOYG_11847.1	Supercontig 8: 118 573-119 943	515,77	0	445	257	315
	Fusarium oxysporum PHW808	FOPG_13302.1	Supercontig 144: 64 862-66 232	513,84	0	445	256	314
	Fusarium oxysporum Fo5176 (45	FOXB_03758.1	Supercontig 1 154: 40 325-41 695	513,84	0	445	256	314
	Fusarium oxysporum Fo47	FOZG_16470.1	Supercontig 15: 763 662-765 032	513,84	0	445	254	313
	Fusarium oxysporum Cotton	FOTG_18322.1	Supercontig 353: 7 128-8 498	513,46	0	445	256	314
	Fusarium graminearum PH-1 (FC	FGSG_04832.3	Supercontig 3: 400 164-401 498	362,07	0	424	194	256
	Fusarium oxysporum NRRL3293	FOYG_10585.1	Supercontig 7: 252 735-254 069	357,07	0	426	190	257
	Fusarium oxysporum f. sp. melon	FOMG_16286.1	Supercontig 24: 203 380-204 714	356,30	0	426	190	256
	Fusarium oxysporum 4287 (FO2)	FOXG_13046.3	Supercontig 17: 1 278 078-1 279 412	356,30	0	426	190	256
	Fusarium oxysporum MN25	FOWG_04454.1	Supercontig 4: 121 132-122 466	356,30	0	426	190	256
	Fusarium oxysporum CL57	FOCG_16632.1	Supercontig 21: 128 182-129 516	356,30	0	426	190	256
	Fusarium oxysporum Fo47	FOZG_14633.1	Supercontig 12: 199 432-200 766	356,30	0	426	190	256
FCIRG_04556	*Fusarium oxysporum GL57*	FOCG_119561	Supercontig 11: 126 761-129 569	590	0	565	52	67
	Fusarium oxysporum 4287 (F02)	FOXG_10919	Supercontig 14: 288 173-291 421	591	0	565	52	67
	Fusarium oxysporum f. sp. melon	FOMG_00126	Supercontig 1: 345 068-347 876	590	0	565	52	67
FCIRG_04555	None Identified							

[a]Given in amino acids.

Table S14: The distribution of *Fusarium circinatum*-specific proteins, determined through BLASTp analyses, amongst the Sordariomycetes genomes included in JGI's MycoCosm database.

Query Gene	Species	Protein Name	Location and Position (bp)	E-Value	Alignment Length[a]	Identity (%)[a]	Positives [a]
FCIRG_04559	*Nectria haematococca*	Necha2_85516	Scaffold 82: 321 452-323 813	0	631	53,41	337
	Colletotrichum graminicola	Colgr1_8656	Supercontig_50: 180 011-182 343	6,10E-140	514	51,95	267
	C. somersetensis	Colso1_562257	Scaffold_5: 5 546-7 862	4,87E-140	529	50,85	269
	C. zoysiae	Colzo1_662029	Scaffold 36: 116 255-118 577	2,43E-139	529	50,66	268
	Glomerella cingulata	Gloci1_06755	Scaffold_3: 3 858 958-3 861 286	6,66E-157	570	50,53	288
	C. falcatum	Colfa1_579345	Scaffold 159: 77 191-79 537	3,12E-145	542	50,37	273
	G. acutata	Gloac1_1604846	Scaffold_5: 349 856-352 862	8,50E-158	562	50,36	283
	G. cingulata	Gloci1_1888518	Scaffold_5: 2 203 672-2 206 213	1,52E-128	512	50,2	257
	C. higginsianum	Colhi1_4201	Supercontig_2 356: 957-3 348	1,62E-128	508	50,2	255
	G. acutata	Gloac1_1649532	Scaffold_50: 2 434-5 397	3,24E-122	506	50	253
	C. somersetensis	Colso1_359207	Scaffold_36: 98 497-99 622	3,15E-123	498	50	249
	C. eremochloae	Coler1_633355	Scaffold_183: 7 254-9 630	7,50E-129	512	49,8	255
	C. sublineola	Colsu1_566281	Scaffold_7: 181 246-183 629	8,36E-130	516	49,61	256
	Ilyonectria sp.	Ilysp1_1670160	Scaffold_11: 451 054-453 831	4,24E-145	560	49,11	275
	C. falcatum	Colfa1_184145	Scaffold_237: 61 036-63 721	5,44E-129	542	48,71	264
	Fusarium graminearum	Fusgr1_02825	Supercontig 3.2: 5 277 911-5 280 156	1,21E-114	532	47,37	252
	Verticillium dahliae	VDAG_07234	Supercontig_1.16: 153 242-155 530	1,50E-126	520	46,54	242
	F. verticillioides	FVEG_13162	Supercontig_2.4: 2 087 517-2 088 685	1,23E-128	550	46,36	255
	Trichoderma asperellum	Trias1 6221256	Scaffold_16: 192 146-194 524	9,42E-118	533	76	247
	F. oxysporum f. sp. lycopersici	FOXG_1517	Supercontig_2.29: 296 837-299 006	1,07E-133	533	46,34	247
	F. fujikuroi IMI 58289	Fusfu1_05520	Chromosome_06 : 472 080-474 225	8,35E-129	572	45,8	262
FCIRG_04558	*Trichoderma harzianum*	Triha1_82923	Scaffold_5: 1 567 355-1 568 758	5,62E-76	352	40,34	142
	T. virens Gv29-8	TriviGv29_8_2_3_7668	Scaffold_5: 1 353 185-1 354 507	1,00E-75	336	41,67	140
	T. longibrachiatum	Trilo3_65536	Scaffold_5:1 232 601-1 234 007	9,93E-72	336	41,18	136
	T. asperellum CBS 433.97	Trias1_132357	Scaffold_3: 1 656 338-1 657 741	1,77E-75	335	41,49	139
	T. reesei RUT C-30	TrireRUTC30_1_77797	Scaffold_7: 69 152-70 558	2,54E-71	336	40,18	135
	Fusarium oxysporum 4287	FOXG_13046.3	Supercontig 17: 1 278 078-1 279 412	0	407	40,2	231
	F. graminearum PH-1	FGSG_04832.3	Supercontig 3: 400 164-401 498	0	407	40,2	233
FCIRG_04557	*Phaeoacremonium aleophilum*	Phaal1_778	Scaffold_154: 162 296-163 690	0	389	77,12	300
	Trichoderma asperellum	Trias1_132357	Scaffold_3:1 656 338-1 657 741	0	399	75,69	302
	T. virens	TriviGv29_8_2_37668	Scaffold_5: 1 353 182-1 354 585	0	392	75,51	296
	Metarhizium robertsii	Metan1_2313	Scaffold_002: 2 663 647-2 665 053	0	394	74,62	294
	Niesslia exilis	Nieex1_798650	Scaffold_2: 522 055-523 455	0	412	74,51	307
	T. reesei	TrireRUTC30_1_77797	Scaffold_7: 69 152-70 558	0	401	74,31	298
	T. reesei	Trire2_107172	Scaffold_8: 1 336 705-1 338 111	0	401	74,31	298
	T. harzianum	Triha1_82923	Scaffold_5: 1 567 355-1 568 758	0	399	74,19	296
	T. longibrachiatum	Trilo3_65536	Scaffold_5: 1 232 601-1 234 007	0	401	72,82	292
	M. inundatum	Myrin1_38252	Scaffold_1: 4 180 495-4 182 177	0	407	74,2	302
	T. atroviride	Triat2_223009	Contig_25: 1 334 207-1 335 837	0	399	74,19	296
	Ilyonectria sp.	Ilysp1_1664978	Scaffold_10: 962 471-964 399	0	412	71,84	296
	Ophiostoma piceae	Ophpc1_6776	Scaffold_17: 104 917-106 287	0	378	71,43	270
	Eucalypta lata	Eutla1_1479	Scaffold_1294: 9719-11 119	0	412	68,2	281
	Beauveria bassiana	Beaba1_3102	Scaffold_00006: 1 079 862-1 081 274	1,52E-135	423	66,67	282
	F. oxysporum f. sp. lycopersici	Fusox1_14897	Supercontig_2.52: 93 747-95 379	1,37E-152	365	64,93	237
	Thozetella sp. PMI_491	ThoPMI491_1_631583	Scaffold_1: 1 449 042-1 450 367	7,11E-123	367	53,95	198
	F. graminearum	Fusgr1_5624	Supercontig_3.3: 400 164-401 498	7,65E-118	369	52,85	195
	Ilyonectria sp.	Ilysp1_1522559	Scaffold_5: 663 621-664 931	4,15E-121	366	52,73	193
	F. solani	Necha2_84411	Chromosome 9: Scaffold 27: 269 225-270 547	1,03E-122	373	52,01	194
	F. oxysporum f. sp. lycopersici	Fusox1_6035	Supercontig_2.17: 1 278 078-1 279 412	7,64E-116	371	51,48	191

FCIRG_04556	*Fusarium solani* v2.0	Necha2_51514	Chromosome 11: 117 201-118 955	0	524	58,78	308
	F. oxysporum v1.0	FOXG1_1091	Supercontig 2.14: 288 749-291 052	0	552	53,99	298
FCIRG_04555	*Hypoxylon* sp. EC38	HypEC38_1_385672	Scaffold_19: 346 386-348 063	4,94E-162	365	61,92	226
	Hypoxylon sp. CO27-5	HypCO275_1_391925	Scaffold_4: 285 371-286 943	9,50E-162	366	61,48	225
	Hypoxylon sp. CI-4A	HypCI4A_1_51722	Scaffold_3: 990 218-991 612	2,86E-156	364	60,44	220
	Trichoderma longibrachiatum	Trilo3_1392130	Scaffold_8: 280 933-282 444	1,47E-134	363	55,1	200
	T. harzianum	Triha1_490835	Scaffold_2: 2 158 760-2 160 237	1,15E-137	370	54,86	203
	Niesslia exilis	Nieex1_803415	Scaffold_18: 572 636-573 993	3,39E-75	331	42,9	142
	Metarhizium inundatum	Myrin1_512013	Scaffold_1: 3 666 365-3 667 773	5,38E-151	369	57,72	213

[a]Given in amino acids.

Table S15: The distribution of *Fusarium circinatum*-specific proteins , determined through psi-BLASTp analyses, amongst sequenced species in the NCBI database.

Query Gene	Species	Protein ID	Score	E-value	Alignment length[a]	Identity (%)[a]	Positives (%)[a]
FCIRG_04559	*Nectria haematococca* mpVI 77-13-4	gi_302881705_ref_XP_003039763.1; Necha2_85516	676	0	709	49,51	64,6
	N. haematococca mpVI 77-13-4	gi_256725063_gb_EEU38429.1; Necha2_59609	555	0	498	53,41	70,88
	Fusarium avenaceum	gi_751348423_gb_KIL86145.1	532	3,00E-176	702	42,31	58,4
	F. oxysporum Fo5176	gi_342886514; FOXB_01356	528	7,00E-175	701	42,37	58,49
	F. verticillioides 7600	gi_584145854; FVEG_13162	524	2,00E-173	701	41,94	58,2
	F. oxysporum f. sp. *pisi* HDV247	gi_587736448_gb_EXA34164.1; FOVG_14601	524	2,00E-173	701	42,23	58,35
	F. pseudograminearum CS3096	gi_685859853	524	2,00E-173	711	42,19	58,51
	Colletotrichum higginsianum	gi_380490613_emb_CCF35893.1	491	7,00E-160	713	40,39	56,66
	C. graminicola M1.001	gi_310799562_gb_EFQ34455.1; Colgr1_8656Colsu1_556281	486	1,00E-158	716	40,36	55,03
	C. sublineola	gi_640917358_gb_KDN62083.1	484	2,00E-157	661	42,36	58,4
	N. haematococca mpVI 77-13-4	gi_302888222; gi_302890517 XP_003044142.1	472	1,00E-153	582	44,16	61,34
	C. gloeosporioides Cg-14	gi_530473007_gb_EQB53507.1	470	1,00E-152	666	42,19	56,91
	F. fujikuroi IMI 58289	gi_517318728_emb_CCT69619.1	469	2,00E-152	681	40,23	56,09
	F. graminearum	gi_699038476_emb_CEF77502.1; FGSG_02825	466	5,00E-151	651	41,94	57,6
	C. gloeosporioides Nara gc5	gi_596680550; GLO06755	465	1,00E-150	661	42,06	56,73
	F. oxysporum Fo47	gi_587692313; FOZG_08162	459	4,00E-150	552	45,11	61,78
FCIRG_04558	*Fonsecaea pedrosoi*	KIW84299.1	698	0	435	75	86,44
	R. bacterium URHD0088	WP_037266633.1	633	0	414	72	82,37
	Candidatus Entotheonella sp. TSY1	WP_034418784.1	306	7,00E-96	408	41	58,33
	C. Entotheonella sp. TSY1	ETW98620.1	306	1,00E-95	408	40	58,33
	Thermomicrobiales bacterium	WP_038038042.1	301	1,00E-93	408	40	58,33
	Pseudomonas aeruginosa	WP_033951105	288	9,00E-89	403	40	58,22
	Halotalea alkalilenta	WP_027350040	278	1,00E-84	403	40	58,21
	Burkholderia sordidicola	WP_031360312.1	296	1,00E-91	405	40	58,52
FCIRG_04557	*Trichoderma virens*	gi_358386050_gb_EHK23646.1	582	0	446	65,47	73,77
	Metarhizium guizhouense	gi_743665647_gb_KID92796.1	588	0	448	65,18	74,78
	T. reesei	XP_006965202	579	0	467	63,81	72,38
	M. anisopliae	gi_589106359;	598	0	466	63,73	74,25
	T. atroviride	770404918_gb_KJK89846.1; EHK_044055	581	0	467	63,38	71,95
	T. harzianum	KKP_01087	588	0	399	74,19	74,19
	Ophiostoma piceae	gi_358394662_gb_EHK44055.1	556	0	432	62,96	74,77
	Nectria haematococca	gi_302888535	626	0	599	52,42	68,95
	Fusarium oxysporum f. sp. *melonis*	gi_590044546	590	0	565	52,74	67,79
	F. oxysporum f. sp. *radicis-lycopersici*	gi_591410826	590	0	565	52,74	67,79
	M. robertsii	gi_629736656 (MAA 10450)	595	0	600	51	67,83
FCIRG_04556	*Nectria haematococca*	gi_302888535	626	0	599	52,42	68,95
	Metarhizium robertsii	gi_629736656; MAA10450	595	0	600	51	67,83
	M. robertsii	XP_007826639; STE12	595	0	600	51	67,83
	M. anisopliae	gi_672383964	589	0	600	51	67,67
	Fusarium oxysporum f. sp. *melonis*	gi_590044546	590	0	565	52,74	67,79
	F. oxysporum f. sp. *radicis-lycopersici*	gi_591410826	590	0	565	52,74	67,79
	M. brunneum ARSEF 3297	gi_743630071	585	0	600	50	66,83
FCIRG_04555	*Scedosporium apiospermum*	gi_666869369; gb_KEZ45657.1	474	5,00E-161	446	56,05	69,06
	S. schenckii	gi_550805532; gb_ERS97448.1	460	2,00E-155	445	53,26	68,09
	Exophiala oligosperma	gi_759261176; gb_KIW37706.1	461	5,00E-156	433	52,66	70,21
	E. aquamarina	gi_656914937; gb_KEF60516.1	417	3,00E-138	430	51,16	68,14
	S. brasiliensis	gi_550805532; gb KIH 89839	460	2,00E-155	445	53,26	68,09
	S. schenckii	gi 780591189 gb KJR81948.1	458	2,00E-154	441	53,74	68,48
	Trichoderma harzianum	gi 818157569; gb KKO98440.1	447	8,00E-150	433	50,81	67,9
	T. reesei	gi 589112713 XP_006968379.1	437	5,00E-146	471	48,41	63,91
	Cladophialophora immunda	gi 759252803 gb KIW29467.1	378	3,00E-123	438	44,52	63,01
	Fonsecaea multimorphosa	gi 761332535 gb KIX94174.1	376	2,00E-122	437	44,62	62,24
	Bougainvillea spectabilis	gi 57724113 gb GAD97227.1	361	1,00E-116	434	43,09	63,13
	E. xenobiotica	gi 759280896 gb KIW57400.1	353	2,00E-113	409	44,99	64,06

[a]Given in amino acids.

Table S16: Statistical measures determining significance of the G+C content of the *Fusarium circinatum*-specific genes against different genomic regions of the *Fusarium circinatum* host genome using Students t-test (unpaired)[A].

Description	Mean G + C content (%)	Standard Deviation	Standard Error of Mean	*P*-value	t	Degrees of Freedom	Standard error of difference
Fusarium circinatum-specific genes	51,2	1,32	0,5932959				
Contig02138	48	2,05	0,6482669	0,008	3,1273	13	1,017
Chromosome 3	47,29	0,299	0,1337169	0,002	6,429	8	0.608
F. circinatum genome	47	0,499	0,1504542	0,0001	9,4382	14	0,445

[A] H_0 = G+C content (%) of *F. circinatum*-specific genes are similar to that of *F. circinatum* genomic region.

H_1= G + C content (%) of *F. circinatum*-specific genes is significantly different than other genomic regions of *F. circinatum.*

Table S17: Base composition and gene structure comparison between FCIRG_04559 and the xenologous protein Necha2_85516.

Genetic feature	*Fusarium circinatum* FCIRG_04559	*Fusarium solani* Necha2_85516
Adenine (A) %	24	25
Cytosine (C) %	27	27
Guanine (G) %	23	24
Thymine (T) %	26	24
G + C %	50	51
A + T %	50	49
Length (AA)[a]	702	718
Length (bp) [b]	2312	2362
Intron Count	2	4

[a]Amino Acid (AA).
[b]Base pair (bp).

Table S18: Base composition and gene structure comparison of FCIRG_04556 and phylogenetically inferred xenologous Necha2_51514 protein.

Genetic feature	*Fusarium circinatum* FCIRG_04556	*Fusarium solani* Necha2_51514
Adenine (A) %	24	22
Cytosine (C) %	26	29
Guanine (G) %	25	25
Thymine (T) %	25	24
G + C %	52	54
A + T %	48	46
Length(AA)[b]	655	584
Length (bp)[a]	1986	1755
Intron count	1	0

[a]Amino Acid (AA).
[b]Base pair (bp).

Table S19: Base composition and gene structure comparison of FCIRG_04558 and phylogenetically inferred xenologous KIW84299.1 protein.

Genetic feature	*Fusarium circinatum* FCIRG_4558	*Fonsecaea pedrosoi* CBS 271.37 KIW84299.1
Adenine (A) %	23	24
Cytosine (C) %	26	26
Guanine (G) %	27	29
Thymine (T) %	24	21
G + C %	53	55
A + T %	47	45
Length (AA)[a]	439	458
Length (bp)[b]	1362	1377
Intron Count	1	0

[a]Amino Acid (AA).
[b]Base pair (bp).

Table S20: Statistical measures determining significance of the G+C content of the FCIRG 04558 xenologous gene pair (KIW 84299) against different genomic regions of the *Fusarium circinatum* host genome using Students t-test (unpaired)[A].

Description	Mean G + C content (%)	Standard Deviation	Standard Error of Mean	P-value	t	Degrees of Freedom	Standard error of difference
FCIRG 04558	53,2						
KIW 84299	55						
Xenologous genes mean	54,1	0,9	0,6364				
Contig 02138 coding genes	49,7	2,05	0,5479	0,011	2,9249	14	1,504
F. circinatum chromosome 3	47,92	0,299	0,13372	0,0001	15,2854	5	0,404
F. circinatum genome	47,3	0,449	0,13538	0,001	17,4525	11	0,39

[A] H_0 = G+C content (%) of xenologous pairs are similar to that of *F. circinatum* genomic region.
H_1= G+C content (%) of xenologous gene pair (FCIRG_04558) is higher than other genomic regions of *F. circinatum*