Genome-Based Selection and Characterization of Fusarium circinatum-Specific Sequences

Mkhululi N. Maphosa,* Emma T. Steenkamp,† and Brenda D. Wingfield*†

*Department of Genetics and †Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, Faculty of Natural and Agricultural Sciences, 0028 South Africa

ABSTRACT Fusarium circinatum is an important pathogen of pine trees and its management in the commercial forestry environment relies largely on early detection, particularly in seedling nurseries. The fact that the entire genome of this pathogen is available opens new avenues for the development of diagnostic tools for this fungus. In this study we identified open reading frames (ORFs) unique to F. circinatum and determined that they were specific to the pathogen. The ORF identification process involved bioinformatics-based screening of all the putative F. circinatum ORFs against public databases. This was followed by functional characterization of ORFs found to be unique to F. circinatum. We used PCR- and hybridization-based approaches to confirm the presence of selected unique genes in different strains of F. circinatum and their absence from other Fusarium species for which genome sequence data are not yet available. These included species that are closely related to F. circinatum as well as those that are commonly encountered in the forestry environment. Thirty-six ORFs were identified as potentially unique to F. circinatum. Nineteen of these encode proteins with known domains while the other 17 encode proteins of unknown function. The results of our PCR analyses and hybridization assays showed that three of the selected genes were present in all of the strains of F. circinatum tested and absent from the other Fusarium species screened. These data thus indicate that the selected genes are common and unique to F. circinatum. These genes thus could be good candidates for use in rapid, in-the-field diagnostic assays specific to F. circinatum. Our study further demonstrates how genome sequence information can be mined for the identification of new diagnostic markers for the detection of plant pathogens.

Fusarium circinatum is the causal agent of pitch canker, which is an economically important disease of pines (Wingfield et al. 2008). The first known incidence of an epidemic caused by this fungus occurred in the south-eastern United States in 1946 (Gordon et al. 2001), but the pathogen has since spread worldwide and is responsible for devastating forestry industry losses (Viljoen 2001), but the pathogen has since spread worldwide and is responsible for devastating forestry industry losses (Viljoen 2001; Wingfield et al. 2002, 2008; Landers et al. 2005; Carlucci et al. 2007; Alonso and Bettucci 2009; Bragança et al. 2009; Steenkamp et al. 2012; Pfenning et al. 2014). In regions where the pathogen occurs, it often also has a complicated life history and distribution that could not have been easily predicted. In South Africa, for example, the pathogen initially appeared to be confined to nurseries where it caused severe root disease on pine seedlings and cuttings (Viljoen et al. 1994; Wingfield et al. 2008), while it emerged as a plantation pathogen only recently (Coutinho et al. 2007; Steenkamp et al. 2014) by affecting established or mature pine trees in the Eastern Cape, Western Cape, and KwaZulu Natal Provinces (Britz et al. 2005; Coutinho et al. 2007; Steenkamp et al. 2014; Santana et al. 2015). The global spread of F. circinatum could be attributed to trade in seeds while the spread from nurseries to plantations is probably the consequence of practices that involve the planting of contaminated or infected seedlings (Wingfield et al. 2008). Therefore, a major challenge facing forestry industries has been the detection of the pathogen in plant growth media and in plant tissues especially during the early stages of infection. However, one of the most significant hurdles in terms of early detection has been the lack of rapid,
in-the-field pathogen detection tools. The currently available quantitative real-time PCR methodologies (Schweigkofler et al. 2004; Ioos et al. 2009; Dreaden et al. 2012) all utilize expensive and sophisticated equipment that are not practically and economically feasible for routine use in nurseries and field stations. Alternative tools such as the DNA-based loop-mediated isothermal amplification (LAMP) method (Tomita et al. 2008) and antigen-based enzyme-linked immunosorbent assay (ELISA) test kits (Gan et al. 1997) would be much more appropriate for in-the-field detection, but have not yet been developed for the pitch canker pathogen.

The development of diagnostic assays based on technologies such as LAMP and ELISA is dependent on the availability of pathogen-specific targets to allow unambiguous identification of F. circinatum. In the case of LAMP, the DNA target region should ideally span an area not exceeding 200 bp specific to the genome of F. circinatum (Notomi et al. 2000; Tomita et al. 2008), while the ELISA targets should represent antigenic proteins with epitopes specific to the pathogen (Gan et al. 1997). However, the available diagnostic tools for this fungus were mostly developed based on known taxonomic markers and accordingly rely on polymorphisms within highly conserved and/or noncoding DNA regions (Steenkamp et al. 1999; Schweigkofler et al. 2004; Ioos et al. 2009; Dreaden et al. 2012), which would not be suitable for LAMP purposes or for developing ELISA tools.

Increased access to whole genome sequence information for fungal pathogens has opened the possibility of mining these genomes for suitable targets to use in diagnostics. The genome sequences for various Fusarium species have been determined previously and are in the public domain; e.g., the Fusarium Comparative Sequencing Project (Broad Institute of Harvard and MIT; http://www.broadinstitute.org) and the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov). This is also true for the pitch canker fungus (Wingfield et al. 2012) and its close relatives F. verticillioides (Fusarium Comparative Sequencing Project) and F. fujikuroi (Wiemann et al. 2013). Although comparisons among these genomes have revealed high levels of synteny, various chromosomal regions in these fungi have been suggested to be strain- or species-specific (Wiemann et al. 2013; De Vos et al. 2014). The overall goal of this study was, therefore, to explore the possibility of using genome-based information to identify targets that would be suitable for future development of diagnostic methods based on technologies such as LAMP and ELISA. Our first aim was to analyze the protein-coding component of the F. circinatum genome against those of other Fusarium species in public databases to identify genes unique to the pitch canker fungus. We then characterized the identified sequences in terms of the proteins they encode, as well as the cellular localization and antigenicity of the inferred proteins. Finally, genes that were apparently specific to F. circinatum and that could potentially encode products unique to this fungus were then evaluated for their distribution among isolates of F. circinatum and their absence in other species of Fusarium, particularly those such as F. proliferatum (Stępień et al. 2011) and F. oxysporum (Fravel et al. 2003) which often occur in the same environment as the pitch canker fungus. This study will thus provide the foundation for future development of highly specific diagnostic assays for this important pathogen, both in terms of potential gene targets and the methodologies to identify suitable diagnostic targets.

### MATERIALS AND METHODS

#### Screening of the F. circinatum genome to identify species-specific genes

In this study, the genome sequence information for one strain (FSP34) of F. circinatum was used (Wingfield et al. 2012). Genome data and predicted protein sequences of F. oxysporum, F. graminearum, and F. verticillioides were obtained from the Broad Institute’s Fusarium Comparative Sequencing Project. The genomic data of F. fujikuroi that were generated by Wiemann et al. (2013) were obtained from the authors. A nucleotide database and a protein database of all these genomes were created on CLC Main Workbench 5.7 (CLC bio A/S). This platform was then used to search for homologs of the ca. 15,000 putative genes of F. circinatum (Wingfield et al. 2012) in the genomes of these other fungi by making use of BLASTn and a word size of 11. In a similar way, the protein sequences encoded by the screened genes were then analyzed on the protein database using BLASTp searches to identify potentially unique proteins in F. circinatum. All the identified genes were then screened against the nucleotide and protein sequences databases at the NCBI, using BLASTn and BLASTp searches. For the purposes of this study, unique open reading frames (ORFs) were defined as those showing less than 50% nucleotide sequence identity and encode proteins returning less than 30% positive amino acid identity from all screened databases.

Putative unique ORFs or ORFs that are potentially specific to F. circinatum were subjected to BLASTx and tBLASTn analyses using the search engines and databases of the Broad Institute and NCBI to characterize the potential protein products coded for by these putative genes. All putative genes that potentially coded for protein sequences similar to sequences available in either of these public databases were eliminated from our set of ORFs that are potentially unique to F. circinatum.

#### In silico characterization of possible F. circinatum-specific genes

To predict functions for the F. circinatum-specific candidate genes, their inferred amino acid sequences were analyzed on the following databases: Pfam (Punta et al. 2012) to determine which protein family they belong to; conserved domains (CDD) (Marchler-Bauer et al. 2011) to

---

**Table 1. Primers used in this study indicating different annealing temperatures for each primer pair**

| F. circinatum Gene | Name         | Sequence 5’-3’  | Annealing Temperatures |
|--------------------|--------------|-----------------|------------------------|
| FCIRG_14470        | FCIRG_14470F | CTCCTTCCGCCCTCAACTA | 55                     |
| FCIRG_14470        | FCIRG_14470R | GAGCCGTTTAGGCGACCTG | 57                     |
| FCIRG_06550        | FCIRG_06550F | CTCCTCCAGCAACCCACCG | 54                     |
| FCIRG_06550        | FCIRG_06550R | CGACCAGTTCTCTGCTGACC | 57                     |
| FCIRG_06217        | FCIRG_06217F | AGAGGTCCAGTAGCACATG | 54                     |
| FCIRG_06217        | FCIRG_06217R | GCACCTTGCTTCTCTCGG  | 57                     |
| FCIRG_05181        | FCIRG_05181F | CGCAGACGCTGAAGAAA  | 57                     |
| FCIRG_05181        | FCIRG_05181R | TGGCAGTTGACAGTGAAAT | 58                     |
| FCIRG_10575        | FCIRG_10575F | TCTCGGAAATAGGTCTTGTATCAGC | 58                     |
| FCIRG_10575        | FCIRG_10575R | CCTGCGGAGGCCGACATTAGC | 58                     |
deduce any conserved domains they might encode; and Simple Modular Architecture Research Tool (SMART) (Letunic et al. 2012) to ascertain the arrangement of different domains (where applicable). To predict the cellular localization of the putative proteins, the following programs were used: SignalP (Dyrløv Bendtsen et al. 2004) to predict any signal peptides within the first 70 amino acids of the protein sequence; and WoLF PSORT (Horton et al. 2007) to predict subcellular localization. To evaluate if the proteins could be applicable in an immune assay such as ELISA, VaxiJen (Doytchinova and Flower 2007) was used to predict antigenicity. To determine if there could be paralogs within the F. circinatum genome we analyzed the ORF sequences against the F. circinatum genomic data using the BLASTn function on CLC Bio workbench. We further analyzed the unique candidate sequences against the available F. circinatum RNA sequence data (Wingfield et al. 2012) to ascertain the evidence of expression.

### Table 2 Fungal isolates and species used in this study as well as their hosts and geographic origins

| Isolates | Species | Host and Origin |
|----------|---------|----------------|
| CMWF530, CMWF1799, CMWF1800, CMWF1801, CMWF1802, CMWF1803 | F. circinatum | Pinus patula, Mexico, Hidalgo |
| CMWF550 | F. circinatum | Pinus leiophylla, Mexico, North-central Michoacan |
| CMWF567 | F. circinatum | Pinus douglasiana, Mexico, Jalisco |
| CMWF1804 | F. circinatum | Pinus greggii, Mexico, Laguna Atezca |
| CMWF39, CMWF30, CMWF45 | F. circinatum | Pinus patula, South Africa, Mpumalanga |
| CMWF56 | F. circinatum | Pinus greggii, South Africa, Mpumalanga |
| CMWF497 | F. circinatum | Pinus patula, South Africa, Mpumalanga |
| CMWF538, CMWF513, CMWF659, CMWF674 | F. circinatum | Pinus radiata, South Africa, Western Cape |
| CMWF350 | F. circinatum | Pinus radiata, USA, California |
| CMWF968, CMWF1002 | F. oxysporum | Syzygium cordatum, South Africa, Gauteng |
| CMWF915, CMWF927 | F. oxysporum | Syzygium cordatum, South Africa, KwaZulu Natal |
| CMWF940 | F. oxysporum | Syzygium cordatum, South Africa, Western Cape |
| CMWF985 | F. oxysporum | Syzygium cordatum, South Africa, Western Cape |
| CMWF978 | F. pallidoroseum | Syzygium cordatum, South Africa, Western Cape |
| CMWF948, CMWF898 | F. proliferatum | Syzygium cordatum, South Africa, Gauteng |
| CMWF1155, CMWF1161 | F. proliferatum | Syzygium cordatum, South Africa, KwaZulu Natal |
| CMWF1182 | F. proliferatum | Syzygium cordatum, South Africa, Western Cape |
| CMWF1005 | F. solani | Syzygium cordatum, South Africa, Western Cape |
| CMWF1147 | F. solani | Syzygium cordatum, South Africa, KwaZulu Natal |
| CMWF1474, CMWF1475 | F. subglutinans | Zea mays, USA, Illinois |

* CMWF refers to the Fusarium culture collection of the Forestry and Agriculture Biotechnology Institute, FABI, University of Pretoria, Pretoria, South Africa.

The isolates of F. circinatum were all reported from previous studies where those from Mexico and California were used by Wikler and Gordon (2000), while those from the Western Cape and Mpumalanga provinces of South Africa were respectively reported by Steenkamp et al. (2014) and Viljoen et al. (1997). The representatives of F. subglutinans came from the study of Steenkamp et al. (2001). All of the isolates from Syzygium cordatum originated from a previous survey of the diversity of Fusarium species associated with this host in South Africa (Kvas et al. 2008; E. Steenkamp, unpublished data).

For these PCR-based analyses, we used 25-µl reaction mixtures consisting of 2.5 mM of each dNTP, 2.5 mM MgCl₂, 10 µM of each primer, 100 ng template DNA, 0.03U Taq DNA polymerase, and reaction buffer (Roche). The PCR cycling conditions were as follows: initial denaturation hold at 94°C for 5 min, 30 cycles of denaturation at 94°C for 30 sec, annealing for 30 sec (see Table 1 for specific temperatures), and elongation at 72°C for 30 sec, one hold for elongation at 72°C for 7 min, followed by a final hold at 4°C. The samples were analyzed using 2% agarose gel electrophoresis (Sambrook et al. 1989) using gel red as a DNA indicator and a 100 bp ladder (Promega) as a size marker.

All amplicons were purified using the Invitex PCR clean up kit and then sequenced in both directions using the original PCR primers. For this purpose the Big Dye kit (Applied Biosystems, Foster City, CA) and an ABI PRISM 3100 Autosequencer (Applied Biosystems) at the University of Pretoria’s sequencing facility were used. All sequence traces were analyzed and assembled into contigs using CLC Bio workbench, after which sequence alignments were conducted using ClustalW in Mega version 5 (Tamura et al. 2011). Sequences derived from F. circinatum isolates were analyzed against each other to check for variations and sequences from other Fusarium species were compared to the F. circinatum sequences to check for similarities.

We used dot blot hybridization assays to screen for the presence of the identified candidate genes in each of the isolates included in the study. These assays were also used to resolve instances where PCR resulted in no amplification and/or multiple amplicons that could not be sequenced. For these assays, we utilized Roche’s DIG (digoxigenin) High Prime DNA Labeling and Detection Kit (Roche, Manheim, Germany). Genomic DNA of the fungal isolates (Table 2) was blotted onto positively charged nylon membranes and hybridized at 42°C with the respective random primed...
DIG-labeled amplicons of *F. circinatum* isolate FSP34 (i.e., the labeled amplicon for each of the candidate genes was hybridized to the genomic DNA of each of the respective isolates). All hybridizations and detections were conducted according to the manufacturer’s instructions.

**Data availability**

All the genome sequences used in this study are available without restriction.

### RESULTS

**Screening of the *F. circinatum* genome to identify species-specific genes**

BLASTn analyses against the genomic database of *F. oxysporum*, *F. graminearum*, *F. verticillioides*, and *F. fujikuroi* returned 411 *F. circinatum* ORFs that were >50% similar to those of the other fungi. This set of ORFs also did not include smaller genes (<450 bp) that would encode proteins less than 140 amino acids long as their limited size might complicate detection assays based on ELISA technologies. BLASTp analyses using the 411 ORF sequences resulted in the identification of 214 predicted *F. circinatum* proteins that showed <30% amino acid sequence similarity to those in the other *Fusarium* genomes. Screening of these 214 ORFs against NCBI’s database identified three ORFs that were more than 50% similar at the nucleotide level to other genes in the database. After excluding these ORFs, screening of the predicted amino acid sequences for the remaining 211 ORFs against NCBI’s protein database returned 36 putative proteins that shared >30% amino acid similarity to other proteins in the database (Table 3). A final screening of these 36 ORFs against the NCBI and Broad Institute databases using BLASTx and tBLASTn confirmed that they all represented potentially unique sequences in the pitch canker fungus.

**In silico characterization of possible *F. circinatum*-specific genes**

Of the 36 putative genes potentially unique to *F. circinatum*, 19 encode proteins with known domains (Table 4) and 17 encode proteins of unknown function (Table 5). SignalP predicted that three of the putative proteins had signal peptides and were also predicted to be extracellular proteins by WoLF PSORT. Some putative proteins were predicted to represent mitochondrial proteins, but these were likely exported to this organelle as no significant hits were obtained when...
comparing the ORFs against the *F. circinatum* mitochondrial genome data (Fourie et al. 2013), thus confirming that all of the 36 ORFs are encoded on the nuclear genome. Twenty-four putative proteins were predicted to be potentially antigenic, suggesting that they are good candidates for an immune-based diagnostic assay. No paralogs of any of these ORFs were identified in the *F. circinatum* genomic data and we found evidence of expression in *F. circinatum* for 28 of the ORFs (Table 3).

### Evaluating the specificity of the identified ORFs to *F. circinatum*

The 17 genes that encode putative proteins without any known domains were regarded as good candidates for diagnostics. This is because their use might eliminate cross-reactivity associated with the use of proteins with conserved domains that can present the same epitopes. Among the 17 ORFs encoding proteins with no known domains, we selected five for which we found evidence for expression and that potentially encode antigenic proteins. Therefore, primers were designed to amplify the five *F. circinatum* genes FCIRG_14470, FCIRG_06550, FCIRG_06217, FCIRG_05181, and FCIRG_10575. Three primer sets designed for the genes FCIRG_14470, FCIRG_05181, and FCIRG_10575 resulted in amplicons of the expected size in all tested isolates of *F. circinatum*. Sequence analyses of the FCIRG_05181 amplicons revealed single nucleotide polymorphisms among different isolates of *F. circinatum*, while no differences were observed in FCIRG_10575 and FCIRG_14470. The primer set designed for FCIRG_06217 amplified different sized amplicons in the various *F. circinatum* strains. Sequence analyses of these amplicons revealed that the observed polymorphism is due to various indels (20–115 bp) in different *F. circinatum* isolates. The PCRs with the primers designed for FCIRG_06550 failed to generate amplicons in some *F. circinatum* species. Primers for FCIRG_10575 resulted in multiple-sized amplicons with most of the *Fusarium* species tested, and no sequence analysis was done on its amplicons. Sequencing of the amplicons obtained with the primers for FCIRG_06550 resulted in multiple-sized amplicons with most of the *Fusarium* species tested, and no sequence analysis was done on its amplicons. Sequencing of the amplicons obtained with the primers for FCIRG_06550, FCIRG_05181, FCIRG_10575, FCIRG_14470, and FCIRG_06217 from the non-*F. circinatum* isolates showed that they were all different from those of *F. circinatum*. Sequence comparison of the FCIRG_05181 amplicon obtained from

| Name of Gene in FSP34 | Pfam a | CDD b | SignalP c | WoLF PSORT d | Vaxijen e |
|-----------------------|--------|-------|-----------|---------------|----------|
| FCIRG_07223 Oxidored_FMN | OYE_like_FMN TIM_phosphate_binding superfamily NAD_binding_8 superfamily | N | cyto | Nonantigen |
| FCIRG_00789 Fungal_trans_2 RTA1 | Fungal_trans_2 superfamily RTA1 superfamily | N | cyto | Nonantigen |
| FCIRG_05207 RR_TM4-6 DUF4337 IFP_35_N | | N | cyto | Nonantigen |
| FCIRG_05759 DUF2935 | | N | cyto | Nonantigen |
| FCIRG_03368 DPBB_1 | PAT1 | Y | extr | Antigen |
| FCIRG_03489 TcdA_TcdB_pore Pfam-B-4370 Pfam-B-8938 | TcdA_TcdB_pore superfamily | N | mito | Antigen |
| FCIRG_14908 HET | HET superfamily | N | mito | Antigen |
| FCIRG_12843 Lysine_decarbox | Lysine_decarbox superfamily | N | cyto | Antigen |
| FCIRG_05207 RR_TM4-6 DUF4337 IFP_35_N | | N | cyto | Nonantigen |
| FCIRG_10765 MFS_1 HpaX | tail_TIGR02242 superfamily | N | plas | Nonantigen |
| FCIRG_02555 Aldedh | NAD_binding_8 superfamily | N | cyto | Antigen |
| FCIRG_09038 ADIP | | N | plas | Nonantigen |

Table 4 *F. circinatum* potentially unique candidate genes with known putative domains, indicating putative protein families and domains, the top predicted subcellular localization, and whether proteins are antigens or nonantigens.

Volume 6 March 2016 | *Fusarium circinatum*-Specific Genes | 635
characterize a set of genes that are potentially unique to F. circinatum. In this study, we utilized a genome-based approach to identify and characterize a set of genes that are potentially unique to F. circinatum. Although it is possible that we might have excluded suitable gene targets during the initial identification phase of the process, our use of >50% and >30% sequence similarity cut-off values, at the respective DNA and protein levels, ensured that the genes or ORFs identified in this fungus encode products that are quite distinct from other proteins. In other words, strongly conserved genes with homologous sequences in related fungi were excluded to limit the possibility of unwanted cross-reactivity in diagnostic assays. For example, a LAMP assay utilizes six primers targeting eight regions within a DNA fragment of between 130 bp and 200 bp; and for it to be unambiguous, all the primers have to be specific to the target organism (Notomi et al. 2000). Such cross-reactivity can also occur in an immune-based assay such as ELISA which utilizes the interactions between an antibody and epitopes on an antigen; and homologous proteins that potentially have similar folding patterns could present similar epitopes that would allow cross-reaction with antibodies. Our relatively conservative approach for identifying genes or ORFs unique to F. circinatum thus facilitated compilation of a list of putative gene targets that are sufficiently variable to ultimately allow for their potential use in the diagnostics of this pathogen.

Among the set of 36 ORFs potentially unique to the pitch canker fungus, 17 encode proteins with obscure features (POFs) (Armisén et al. 2008) that lack known and defined motifs or domains. Arguably, these ORFs would represent good candidates for diagnostics because of their apparent uniqueness and lack of domains common to other organisms. Although all 17 of these ORFs appear to be transcribed and 10 are predicted to be antigenic, more work is, however, needed to fully understand their expression and the types of proteins they encode, before utilizing them for immune-based procedures. The ideal candidates for an immune-based assay would be genes that are constitutively expressed in all the life stages of the pathogen, while their protein products are stable and easily accessible or extractable (Gan et al. 1997). The other 19 ORFs that are potentially unique to F. circinatum encode proteins involved in a range of different processes. These include cellular division (FCIRG_03368) (Wang et al. 1996), growth (FCIRG_12122) (Callebaut et al. 2001), and maintenance (FCIRG_10765) (Pao et al. 1998), as well as host colonization (FCIRG_10116, FCIRG_05800, and FCIRG_00789) (Soustre et al. 1996; Suárez et al. 1998), as well as host colonization (FCIRG_10116, FCIRG_05800, and FCIRG_00789) (Soustre et al. 1996; Suárez et al. 1998).
et al. 2007; Carr and Ollis 2009). Some of these ORFs also encode substrate-transforming proteins (FCIRG_03079, FCIRG_02555, FCIRG_12843, and FCIRG_07223) (Williams and Bruce 2002), while others encode products potentially involved in transcription (FCIRG_00789) (Shelest 2008) and nonself recognition (FCIRG_14908) (Espagne et al. 2002). One of the identified ORFs encoded the TcdA/TcdB pore motif (FCIRG_03489) of the Clostridium difficile toxin A and toxin B pore-forming region (Qa’Dan et al. 2000). Clostridial toxins A and B are a class of virulence factors that cause serious diseases in mammals (Qa’Dan et al. 2000) and their occurrence in fungi and effects on plants has not been reported.

All 36 ORFs were compared against the F. circinatum mitochondrial genome assembly data (Fouri et al. 2013) to check if any of them could represent mitochondrial genes. No significant hits were obtained indicating that these were all nuclear genes. Roughly 1% of mitochondrial proteins are typically encoded by the mitochondrial genome while the rest are encoded on the nuclear genome (Pfanner and Geissler 2001; Schmidt et al. 2010). As a result, the large majority of mitochondrial proteins are synthesized as precursor proteins in the cytoplasm and imported into the organelle (Schmidt et al. 2010). Our results thus suggest that at least four of the ORFs (FCIRG_03489, FCIRG_14908, FCIRG_06217, and FCIRG_10144) apparently unique to F. circinatum encode for proteins that are transported in a similar way into the mitochondrion. It would be interesting to understand exactly how they function in this cellular compartment and whether or not they potentially convey unique mitochondrial traits to the pathogen.

The available F. circinatum genome harbored no detectable paralogs of the 36 unique ORFs and all of them, therefore, appeared to represent single copy nuclear genes. Although multi-copy genes are usually regarded as good candidates for DNA-based diagnostics because of enhanced sensitivity compared to single copy genes (Ioos et al. 2009), there are limitations associated with their use in this context. Some of the notable limitations include intragenomic heterogeneity (Morandi et al. 2005) that could lead to misidentification of species (Graf 1999). Single copy genes, however, can often be quite useful as diagnostic markers (Alvarez et al. 2008) because they are less likely to be subject to complexities related to intragenomic polymorphisms (i.e., differences among the paralogs of a gene) (Simon and Weiß 2008).

By making use of a PCR-based approach and dot blot hybridization assays, we evaluated the ubiquitous presence of a subset of five unique ORFs in a diverse collection of F. circinatum isolates. These assays indicated that homologs for four of the five genes tested (i.e., FCIRG_14470, FCIRG_06217, FCIRG_05181, and FCIRG_10575) were present in all of the genetically and geographically diverse F. circinatum isolates evaluated, while only some isolates of this fungus appear to harbor a homolog of FCIRG_06550. Through sequence analysis, we also showed that the amplified products corresponded to the original FSP34 sequences, although we did observe various single nucleotide polymorphisms (FCIRG_05181) and indels (FCIRG_06217) among the isolates. Therefore, based on their ubiquitous presence in F. circinatum, at least four of the tested genes represent potential candidates for the development of rapid in-the-field diagnostic assays for this pathogen.

For diagnostic assays to be reliable, they should ideally produce unambiguous and conclusive diagnoses. In other words, if a specific marker region is used, it should be present in all individuals of the focal species to avoid recording false negatives; the results of our screenings with the diverse set of F. circinatum isolates allowed evaluation of this issue. However, the ideal diagnostic marker should also be absent from all nonfocal species to avoid recording false positives. This aspect was evaluated by screening a set of non-F. circinatum isolates for the presence/absence of the target genes. The PCR and dot blot hybridization assays showed that none of the evaluated isolates encodes a homolog of any of the five genes tested. The only exception was for FCIRG_10575, which appeared to be also present in F. subglutinans, which is closely related to F. circinatum (Kvas et al. 2009). Although F. subglutinans is unlikely to be encountered in the commercial forestry environment (Kvas et al. 2009; Leslie and Summerrell 2006), the fact that it apparently harbors a homolog of FCIRG_10575 points toward the potential presence of the gene in other species of the so-called “American Clade” of the Gibberella fujikuroi complex of which F. circinatum is also a member (Kvas et al. 2009). This considerably detracts from the potential value of gene FCIRG_10575 as a diagnostic marker because its use might lead to recording of false positives when non-F. circinatum members of the “American Clade” of the complex are encountered.

Taken together, these findings suggest that the four ORFs found in all of the F. circinatum isolates examined represent members of the so-called core genome of the fungus (Hsiang and Baillie 2005). However, our findings also indicated that only those core genome components not shared with those of other species would be useful for the development of robust diagnostic assays (i.e., the use of core genome regions that overlap with those of other species would lead to false negatives). The ORF that was absent from some F. circinatum isolates is potentially lineage-specific, forming part of its so-called accessory genome (Croll and McDonald 2012). Although the genes encoded on this component of the fungal genome is often associated with adaptive properties such as virulence and/or pathogenicity (Croll and McDonald 2012), their use in diagnostics is limited due to the high likelihood of recording false negatives.

Here we showed that comparative genomic studies allow for the identification of species-specific traits that can be used to identify a taxon. Species-specific traits might be genomic regions that are unique and fixed to a particular species or strongly modified compared to homologous loci in close relatives. In this study, genomic regions that are
unique to *F. circinatum* and are fixed in different strains of the pitch canker fungus were identified. Although care should be taken to avoid regions characterized by high levels of intraspecific polymorphism, these genomic regions appear to be good candidates for use as targets in a *F. circinatum* species-specific diagnostic assay. However, lack of functional annotation of these genes makes it very difficult to infer or speculate on their significance within the *F. circinatum* genome. Tracing the origins of these genes will also go a long way in validating any diagnostic assay that may be developed based on them. Nevertheless, the findings of this study thus represent a fundamental resource for the development of diagnostic tool(s) for the pitch canker pathogen as at least three of the gene targets identified could be used to develop rapid methods for in-the-field diagnosis of the pathogen. Our novel approach and the workflow employed can also easily be adapted for identifying species-specific diagnostic markers for other important taxa.

**ACKNOWLEDGMENTS**

The authors thank the University of Pretoria, members of the Tree Protection Cooperative Program (TPCP), the National Research Foundation (NRF)/Department of Science and Technology (DST), Centre of Excellence in Tree Health Biotechnology (CTHB) and the THRIP initiative of the Department of Trade and Industry (DTI) in South Africa for financial assistance. This work was based on the research supported in part by a number of grants from the National Research Foundation of South Africa (includes Grant specific unique reference number (UID) 83924). The Grant holders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author(s), and that the NRF accepts no liability whatsoever in this regard. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**LITERATURE CITED**

Abd-Elsalam, K. A., 2003 Bioinformatic tools and guideline for PCR primer design. Afr. J. Biotechnol. 2: 91–95.

Alonso, R., and L. Bettucci, 2009 First report of the pitch canker fungus *Fusarium circinatum* affecting *Pinus taeda* seedlings in Uruguay. Australas. Plant Dis. Notes 4: 91–92.

Álvarez, I., A. Costa, and G. N. Feliner, 2008 Selecting single-copy nuclear genes for plant phylogenetics: a preliminary analysis for the *Senecioneae* (Asteraceae). J. Mol. Evol. 66: 276–291.

Armisén, D., A. Lecharny, and S. Aubourg, 2008 Unique genes in plants: specificities and conserved features throughout evolution. BMC Evol. Biol. 8: 280.

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

Brito, H., T. Coutinho, B. Wingfield, W. Marasas, and M. Wingfield, 2005 Diversity and differentiation in two populations of *Gibberella circinata* in South Africa. Plant Pathol. 54: 46–52.

Callebaut, I., B. Goud, and J. P. Mornon, 2001 RUN domains: a new family of domains involved in Ras-like GTPase signaling. Trends Biochem. Sci. 26: 79–83.

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

Carr, P. D., and D. L. Ollis, 2009 α/β hydrolase fold: an update. Protein Pept. Lett. 16: 1137–1148.

Coutinho, T., E. Steenkamp, K. Mongwaketsi, M. Wilmot, and M. Wingfield, 2007 First outbreak of pitch canker in a South African pine plantation. Australas. Plant Pathol. 36: 256–261.

Croll, D., and B. A. McDonald, 2012 The accessory genome as a cradle for adaptive evolution in pathogens. PLoS Pathog. 8: e1002608.

De Vos, L., E. T. Steenkamp, S. H. Martin, Q. C. Santana, G. Fourie *et al.*, 2014 Genome-wide macro synteny among *Fusarium* species in the *Gibberella fujikuroi* complex revealed by Amplified Fragment Length Polymorphisms. PLoS One 9: e114682.

Doytchanova, I. A., and D. R. Flower, 2007 Vaxilen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8: 4.

Dreaden, T., J. Smith, E. Barnard, and G. Blakeslee, 2012 Development and evaluation of a real time PCR seed lot screening method for *Fusarium circinatum*, causal agent of pitch canker disease. For. Pathol. 42: 405–411.

Dyhrlov Bendsen, J., H. Nielsen, G. von Heijne, and S. Bruun, 2004 Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340: 783–795.

Espagne, E., P. Balladère, M. L. Penin, C. Barreau, and B. Turcq, 2002 HET-E and HET-D belong to a new subfamily of WD40 proteins involved in vegetative incompatibility specificity in the fungus *Podospora anserina*. Genetics 161: 71–81.

Fourie, G., N. A. Van der Merwe, B. D. Wingfield, M. Bogale, B. Tuzdynski *et al.*, 2013 Evidence for inter-specific recombination among the mitochondrial genomes of *Fusarium* species in the *Gibberella fujikuroi* complex. BMC Genomics 14: 605.

Fravel, D., C. Olivain, and C. Alabouvette, 2003 *Fusarium oxysporum* and its biocontrol. New Phytop. 157: 493–502.

Gan, Z. R., R. Marquardt, D. Abramson, and R. M. Clear, 1997 The characterization of chicken antibodies raised against *Fusarium* spp. by enzyme-linked immunosorbent assay and immunoblotting. Int. J. Food Microbiol. 38: 191–200.

Gordon, T., A. Storer, and D. Wood, 2001 The pitch canker epidemic in California. Plant Dis. 85: 1128–1139.

Graf, J., 1999 Diverse restriction fragment length polymorphism patterns of the PCR-amplified 16s rRNA genes in *Aeromonas veronii* strains and possible misidentification of *Aeromonas* species. J. Clin. Microbiol. 37: 3194–3197.

Horton, P., K. J. Park, T. Obayashi, N. Fujita, H. Harada *et al.*, 2007 WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35: W585–W587.

Hisang, T., and D. L. Bailie, 2005 Comparison of the yeast proteome to other fungal genomes to find core fungal genes. J. Mol. Evol. 60: 475–483.

Iooos, R., C. Fournier, G. Iancu, and T. R. Gordon, 2009 Sensitive detection of *Fusarium circinatum* in pine seed by combining an enrichment procedure with a real-time polymerase chain reaction using dual-labeled probe chemistry. Phytopathology 99: 582–590.

Kvas, M., E. T. Steenkamp, B. D. Wingfield, W. F. O. Marasas, and M. J. Wingfield, 2008 Diversity of *Fusarium* species associated with malformed ino florescences of *Syzygium cordatum*. S. Afr. J. Bot. 74: 370.

Kvas, M., W. F. O. Marasas, B. D. Wingfield, M. J. Wingfield, and E. T. Steenkamp, 2009 Diversity and evolution of *Fusarium* species in the *Gibberella fujikuroi* complex. Fungal Divers. 34: 1–21.

Landuras, E., P. García, Y. Fernández, M. Braña, O. Fernández-Alonso *et al.*, 2005 Outbreak of pitch canker caused by *Fusarium circinatum* on *Pinus spp.* in northern Spain. Plant Dis. 89: 1015.

Leslie, J. F., and B. A. Summerell, 2006 The *Fusarium* Laboratory Manual. Blackwell Professional, Ames, IA.

Letunic, I., T. Doerk, and P. Bork, 2012 SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40: D302–D305.

Marchler-Bauer, A., S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire *et al.*, 2011 CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39: D225–D229.

Morandi, A., O. Zhaxybayeva, J. P. Gogarten, and J. Graf, 2005 Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in *Aeromonas* strains. J. Bacteriol. 187: 6561–6564.

Notomi, T., H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe *et al.*, 2000 Loop-mediated isothermal amplification of DNA. Nucleic Acids Res. 28: e63–e63.
