Genomic Understanding of an Infectious Brain Disease from the Desert

Leandro F. Moreno,*†,† Abdalla A. O. Ahmed,§ Balázs Brankovics,*† Christina A. Cuomo,** Steph B. J. Menken,† Saad J. Taj-Aldeen,†† Hani Faidah,§ J. Benjamin Stielow,* Marcus de M. Teixeira,** Francesc X. Prenafeta-Boldú,‡‡ Vania A. Vicente,§ and Sybren de Hoog*†,**,††

*Westerdijk Fungal Biodiversity Institute, 3584 CT Utrecht, The Netherlands, †Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1090 GE, The Netherlands, ‡Department of Basic Pathology, Federal University of Paraná State, Curitiba, PR, Brazil, §Department of Microbiology, College of Medicine, Umm Al-Qura University, 24381 Makkah, Saudi Arabia, **Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02142, ††Microbiology Division, Department of Laboratory Medicine and Pathology, Hamad Medical Corporation, Box 3050, Doha, Qatar, ‡‡Division of Pathogen Genomics, Translational Genomics Research Institute, Flagstaff, Arizona 86005, ††GIRO Joint Research Unit IRTA-UPC, Torre Marimon, Caldes de Montbui, E-08140, Barcelona, Catalonia, Spain, and ***Center of Expertise in Mycology of Radboudumc/CWZ, Nijmegen, 6525 GA, The Netherlands

ORCID IDs: 0000-0002-2529-0855 (L.F.M.); 0000-0003-0536-7787 (B.B.); 0000-0002-5778-960X (C.A.C.); 0000-0001-9514-7029 (F.X.P.-B.)

ABSTRACT Rhinocladiella mackenziei accounts for the majority of fungal brain infections in the Middle East, and is restricted to the arid climate zone between Saudi Arabia and Pakistan. Neurotropic dissemination caused by this fungus has been reported in immunocompromised, but also immunocompetent individuals. If untreated, the infection is fatal. Outside of humans, the environmental niche of R. mackenziei is unknown, and the fungus has been only cultured from brain biopsies. In this paper, we describe the whole-genome resequencing of two R. mackenziei strains from patients in Saudi Arabia and Qatar. We assessed intraspecies variation and genetic signatures to uncover the genomic basis of the pathogenesis, and potential niche adaptations. We found that the duplicated genes (paralogs) are more susceptible to accumulating significant mutations. Comparative genomics with other filamentous ascomycetes revealed a diverse arsenal of genes likely engaged in pathogenicity, such as the degradation of aromatic compounds and iron acquisition. In addition, intracellular accumulation of trehalose and choline suggests possible adaptations to the conditions of an arid climate region. Specifically, protein family contractions were found, including short-chain dehydrogenase/reductase SDR, the cytochrome P450 (CYP) (E-class), and the G-protein B WD-40 repeat. Gene composition and metabolic potential indicate extremotolerance and hydrocarbon assimilation, suggesting a possible environmental habitat of oil-polluted desert soil.

Phylogenetic affinity to the so-called black yeasts. Rhinocladiella mackenziei is unique within the genus in that it causes primary central nervous system infection in humans, with the brain as yet being the only known habitat of the fungus. Other species such as R. aquaspersa and R. basitona may cause opportunistic skin infections; however, there is no report of extracranial dissemination. R. mackenziei is confined to the Middle East and adjacent desert zones, and is especially frequent in Saudi Arabia, Pakistan, and Kuwait (Figure 1 and Supplemental Material, Table S1). Cases reported outside the endemic area, e.g., in Europe or the U.S., have invariably concerned immigrants from countries in the Middle East (Cristini et al. 2010).

The pathogen probably invades the brain via the blood–brain barrier, presumably after inhalation, but the portal of entry has not been established with certainty (de Hoog et al. 2000). The infection mostly affects immunocompromised patients but occasionally immunocompetent...
individuals are affected, and despite antifungal therapy combined with surgical intervention, a case fatality rate of ≈70% is observed (Li and de Hoog 2009; Badali et al. 2010). Neurotropism, defined as the affinity of a pathogen for the central nervous system, is recurrent in the Herpotrichiellaceae, specifically in members of the bantiana and dermatitidis clades (Teixeira et al. 2017). Similar to the neurotropic species Cladophialophora bantiana, Fonseccaea monophora, and Exophiala dermatitidis, brain infection by R. mackenziei is primary, i.e., the first symptoms of the disease are of a neurological nature (Horré and de Hoog 1999). In this respect, the infection deviates from classical rhinocerebral infection by Mucorales and from orbital skull-base infection by Aspergillus, where the brain infection is secondary, having a portal of entry in the sinus and later enters the brain gray matter. Pathogenicity of members of Herpotrichiellaceae is partially explained by general virulence factors such as cell wall melanin (Paolo et al. 2006; Seyedmousavi et al. 2014) and thermotolerance, which are shared by numerous black fungi, while the assimilation of alkylbenzene hydrocarbons, which are structurally similar to neurotransmitters, has been suggested to be specific to this family (Prenafeta-Boldú et al. 2006). Although the exact mechanism by which melanin can enhance the virulence in pathogenic fungi remains unclear, recent literature argues that the pigmentation of the cell wall might play a protective role in fungal cells against free radicals, enzymatic or microbial lysis, and extreme temperatures (Jacobson 2000).

Since R. mackenziei thus far has not been isolated from any environmental source, its natural reservoir and route of infection remain unclear. Nevertheless, the presence of genes that could confer tolerance against environmental stressors has been reported in R. mackenziei (Teixeira et al. 2017), indicating that this fungus might reside under the conditions of heat and dryness of arid climates. Genome content determination and comparative genomics with closely related species may further our understanding of metabolic adaptations in R. mackenziei with clues to its potential natural habitat. Black yeasts in general display remarkably diverse lifestyles, with a predilection for extreme and toxic environments such as those rich in aromatic compounds or heavy metals, or with high temperatures, increased salinity, and scarcity of nutrients (Paolo et al. 2006; Prenafeta-Boldú et al. 2006; Zhao et al. 2010). A recent comparative genomic analysis of 23 black yeast species (Teixeira et al. 2017), including the type strain of R. mackenziei, CBS 650.93, revealed a wide diversity of mechanisms associated with nutrient acquisition. These olistrofotrophic fungi exhibit specific gene family expansions, including alcohol (ADH) and aldehyde dehydrogenases (ALDHs), membrane transport proteins, and a diverse ensemble of CYPs, which are believed to be essential for extremotolerance and survival of environmental stress conditions (Teixeira et al. 2017). On the other hand, gene family contractions have not been reported in black yeasts (Z. Chen et al. 2014). The R. mackenziei genome harbors a large number of gene clusters involved in secondary metabolite production, genes encoding proteolytic and carbohydrate-active enzymes, and a wide set of proteins involved in melanin production through distinctive pathways (Teixeira et al. 2017).

In the present study, we performed the whole-genome resequencing of two additional R. mackenziei strains isolated from black biopsies of patients from Saudi Arabia and Qatar. Subsequent variant calling analysis provided a catalog of sequence variations and their putative biological consequences. In addition, we compared the R. mackenziei genomes to other closely related fungi in order to determine genetic signatures such as species-specific genes. Using Fisher’s exact test combined with q-value correction, we confirmed gene family expansions, but also noted gene family contractions shared by R. mackenziei and other neurotropic fungi. GO enrichment analyses were used to predict gene functions of species-specific genes, paralogs, and genes shared with other neurotropic black yeasts. A survey of potential pathogenicity-related genes present in R. mackenziei was carried out comparing its proteome against a curated database.

**MATERIALS AND METHODS**

**Strains and DNA extraction**

To extract genomic DNA, fungal mycelia of the R. mackenziei strains IHM 22877 and dH24460 were harvested from fresh cultures on Sabouraud’s Glucose Agar, washed using sterile Tris-EDTA buffer (TE), pH 8.0 in 2 ml screw-capped tubes, and then resuspended in 500 µl TE buffer. Fungal cell walls were disrupted using 0.5 mm glass beads in a BioSpec Mini-Beadbeater-16 (BioSpec) for 5 min and cooled on ice for an additional 5 min. DNA solutions were separated using two phenol/ chloroform (1:24, pH 8.0) extractions. DNA was then precipitated by isopropanol, washed with 70% ethanol, dried at room temperature, and resuspended in 35 µl TE buffer, pH 8.0. DNA quantity and quality were determined using Qubit (Invitrogen, Applied BioSystems), and an Agilent Bio Analyzer 2100 using a 1000 DNA Chip (Agilent).

**Genome sequencing and de novo assembly**

For genome sequencing, the libraries were constructed using an Illumina NexteraXT Library Preparation Kit and samples were barcoded using a NexteraXT Index Kit (Illumina). The libraries were validated and quantified without bead normalization using an Agilent Bio Analyzer 2100 1000 DNA Chip (Agilent). The dH24460 and IHM 22877 genomes were sequenced on the Illumina MiSeq platform using a paired-ends protocol and a version-3 600 cycles kit. Quality control was performed using FastQC v0.11.3 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) and low-quality sequences were removed by Trimmomatic (Leading: 3, Trailing: 3, Slidingwindow: 4:15) (Bolger et al. 2014). High-quality reads were assembled de novo using SPAdes v3.6.2 (Bankevich et al. 2012).

**Read mapping and SNP calling**

For SNP calling, high-quality reads were mapped against the reference genome of the type strain R. mackenziei CBS 650.93 (accession JYBU00000000) by using Burrows Wheeler Aligner (BWA) v0.7.5 mem (Li and Durbin 2009). The reference genome assembly consisted of 32.47 Mb with a GC content of 50.42%, and was organized in 130 contigs linked by paired-end reads into 17 scaffolds. Alignments were improved by GATK RealignerTargetCreator and IndelRealigner realigning reads around insertion/deletion (indels) in order to minimize bases mismatching the reference (GATK v. 3.4–46) (Van der Auwera et al. 2013). Variants were identified with GATK haplotypecaller v. 3.4–46. Hard filters were applied using the GATK VariantFiltration according to GATK best-practices v.3 guidelines [QD < 20, MQ < 40.0, FS > 60.0, HaploTypeScore > (average value + 2*SD), MQRankSum < −12.5, ReadPosRankSum < −8.0]. In addition, SNPs and indels were called with Pilon version 1.4 (Walker et al. 2014) using default settings. Predictions were combined using the tools SelectVariants and CombineVariants (https://software.broadinstitute.org/gatk/) to assess whether the SNP was found within an untranslated region, intron, or coding exon. Mutations were classified into synonymous (SYN), nonsynonymous (NSY), read-through (RTH), and nonsense (STP). We used the alignment-based method PROVEAN to predict the effect of NSY SNPs on protein function (Choi et al. 2012). For this purpose, proteins carrying single-amino acid substitutions were
compared to their homologs in the NCBI nonredundant protein database, and the substitution frequency and chemical properties of the changed amino acids were taken into account to estimate a score that was used to measure the effect of the variation (Choi et al. 2012).

**Structural variation (SV) identification**

To detect genomic SVs from high-throughput sequencing data, the *R. mackenziei* strains IHM 22877 and dH24460 were analyzed combining two different methods. First, we ran Breakdancer (Chen et al. 2009) to predict SV based on paired-end read mapping (default parameters). To increase the sensitivity and specificity of the call, we used the output of Breakdancer as input for Pindel (Ye et al. 2009), which employs split-read alignments to detect SV, setting the options –band–M10 as described elsewhere (Ghoneim et al. 2014).

**Protein annotation of strains IHM 22877 and dH24460**

To determine the protein sequences corresponding to the newly sequenced strains, 11,382 proteins of the type strain *R. mackenziei* CBS 650.93 (JYBU00000000) were aligned to the genome of the strains IHM 22877 and dH24460 using the homology-based predictor Exonerate v. 2.2.0 (Slater and Birney 2005) with the parameters–model protein2genome–percentage 90. Pathway and KOG prediction were assessed using KAAS (bidirectional best hit) (Moriya et al. 2007). Putative secreted proteins were identified using SignalP 4.1 (Petersen et al. 2011) and WoLF PSORT 0.2 (Horton et al. 2007). Secreted proteins were assessed for the presence of transmembrane helices by Phobius (Kall et al. 2007) and THMMH 2.0 (Sonnhammer et al. 1998). CYPs were predicted as described elsewhere (Teixeira et al. 2017). CYP proteins that could not be assigned to families or subfamilies based on the International P450 Nomenclature Committee were aligned and subjected to phylogenetic analyses, as described elsewhere (W. Chen et al. 2014). Pathogenicity-related genes were identified by means of BLASTP searches (identity > 50%) against the curated PHI-database version 4.0 (Winnenburg et al. 2006).

**Ortholog detection**

Orthology data set was generated clustering the *R. mackenziei* proteins with 20 previously sequenced black yeast species (*E. xenobiotica* CBS 118157, *E. aquamarina* CBS 119918, *E. mesophilica* CBS 402.95, *E. spinifera* CBS 899.68, *E. oligosperma* CBS 725.88, *E. sideris* CBS 121828, *E. dermatitidis* NIIH 8656, *Capronia epimyces* CBS 606.96, *C. coronata* CBS 617.96, *C. semiimmersa* CBS 121828, *C. carrionii* CBS 160.54, *C. immunda* CBS 834.96, *C. bantiana* CBS 173.52, *C. psammophila* CBS 110553, *C. yegresii* CBS 114405, *F. pedrosai* CBS 271.37, *F. multiformis* CBS 102226, *Phialophora europaea* CBS 101466, *P. attae* CBS 131958, *Coniosporium apollinis* CBS 100218, and *Verruconis gallopava* CBS 437.64). Orthologs and paralogs were determined by means of OrthoMCL (Li et al. 2003), with a Markov inflation of 1.5 and a maximum e-value of $10^{-5}$. Species-specific proteins were extracted using a custom Bash script.

**Mitochondrial genome assembly and annotation**

Mitochondrial genome sequences of the strains were assembled using GRaBB (Brankovics et al. 2016) (https://github.com/b-brankovics/grabb) from the raw sequencing data. The bait used for the GRaBB assembly was the mitochondrial genome of *E. dermatitidis* (NW_008751656). Initial mitochondrial genome annotations were done using MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) and manually adjusted. Annotation of tRNA genes was improved using tRNAscan-SE (Pavesi et al. 1994). Annotation of protein-coding genes was corrected by aligning intronless homologs to the genome.

**Data availability**

The sequencing data from this study have been submitted to the GenBank: accession numbers GCA_001723215 (strain IHM 22877) and GCA_001723235 (strain dH24460). Raw sequence data can be accessed through the NCBI Sequence Read Archive: accession number SRP128022. Table S1 summarizes the distribution of 27 cases of cerebral phaeohyphomycosis caused by *R. mackenziei* from 1983 to 2015.
cerebral phaeohyphomycosis caused by \textit{R. mackenziei}. Table S2 contains genes with significant mutations according to PROVEAN. Table S3 and Table S4 show the expanded GO categories between the neurotropic species. Table S5 contains the expanded GO categories in paralog genes of \textit{R. mackenziei}. Table S6 contains the singleton proteins found in \textit{R. mackenziei}. Table S7 contains the \textit{R. mackenziei} secretome. Table S8 contains the CYP in \textit{R. mackenziei}. Table S9 shows the gene family expansions and contractions observed in \textit{R. mackenziei}. Table S10 contains the pathogenicity related genes. Figure S1 shows the frequency and size of indels in coding and noncoding regions.

\textbf{RESULTS}

\textbf{Sequencing and mapping}

To characterize variants of two \textit{R. mackenziei} strains from Qatar and Saudi Arabia compared to the reference genome, each strain was deeply sequenced using Illumina MiSeq. After filtering, a total of 12,996,368
and 10,131,024 high-quality reads were generated for the strains dH24460 and IHM22877, respectively. Among these reads, 11,542,511 and 9,568,960 reads were mapped to the R. mackenziei CBS 650.93 reference genome (Figure 2).

Sequencing of both dH24460 and IHM22877 strains did not produce reads corresponding to the region of scaffold 15 in the reference genome. This scaffold was 4775 bp long and contained two protein-coding genes (Z518_11416 and Z518_11417) annotated as hypothetical proteins, holding PFAM functional domains associated with LTR retrotransposons (PF03732)/zinc finger domains (PF00098) and manipulation of chromatin (PF00385). In order to further investigate the genomic composition of scaffold 15, we used TransposonPSI (http://transposonpsi.sourceforge.net) to identify potential transposable elements. The scaffold harbored three putative LTR retroelements: one classified as Ty1/Copia and two classified as gypsy.

**Variant calling**

The total number of genomic variations relative to the reference genome was detected in the resequenced R. mackenziei strains using two methods. The intersection of the two SNP calling analyses is represented in Figure S3A and calls made by both programs were used for further analyses.

Genome-wide variation identification revealed that the strains IHM22877 and dH24460 contain 65,333 and 77,999 SNPs, respectively, compared to the reference. The highest concentration of SNPs was found in scaffold 11 with, on average, a SNP every 294 bp for IHM 22877 and one every 342 bp for strain dH24460. The lowest density of SNPs was found in scaffold 13 (one every 426 bp on average) for strain dH24460 and in scaffold 3 (one every 484 bp on average) for strain IHM 22877. The distribution of SNPs along the genomes is shown in Figure 2C and E. Regions of high SNP density were located near ends of the scaffolds 1, 4, 6, 7, 8, 9, 10, and 11, suggesting that these regions might correspond to subtelomeres, shown to be hypervariable regions in other studies, potentially undergoing higher rates of mutation (Cuomo et al. 2007).

In order to obtain a comprehensive variant annotation, VCFannotator (see Materials and Methods) was used to classify the mutations into four distinctive categories: SYN, NSY, RTH, and STP. Variant annotation showed that 78,041 SNPs (54.4%) were located in intergenic regions (dH24460: 42,249 and IHM 22877: 35,792), while 65,291 SNPs (45.6%) were located in intronic (dH24460: 3,757 and IHM 22877: 3080) and exonic (dH24460: 31,993 and IHM 22877: 26,461) regions. Of the total SNPs localized in coding regions, 31,409 were SYN (dH24460: 17,117 and IHM 22877: 14,292), 26,700 were NSY (dH24460: 14,679 and IHM 22877: 12,021), 86 were RTH (dH24460: 44; IHM 22877: 42), and 394 were STP (dH24460: 220; IHM 22877: 174) (Figure 3B).

Leveraging published genomes, we identified 2468 single-copy orthologs (SCOs) that were common to 22 black yeasts (see Materials and Methods). Among this gene set, 325 and 422 intronic SNPs, and 5287 (3258 SYN and 2024 NSY) and 6547 (2577 SYN and 3964 NSY) exonic SNPs, were identified in the strains IHM 22877 and dH24460, respectively. SNPs within SCO exons included only four STP (1%) and RTH mutations of stop codons (8%) (Table 1). STP mutation promotes early translational termination of messenger RNAs, resulting in truncated proteins and with possible deleterious effect on the protein function. As expected, the low frequency of mutations in SCO genes confirms the high level of conservation of this gene set in black yeasts.

**Functional effect of SNPs on proteins**

Among the NSY SNPs, 1266 (8.5%) and 1011 (8.3%) mutations in strains dH24460 and IHM 22877, respectively, were considered to have significant effects on protein function (Table S2). The proteins carrying these mutations, in both strains, were enriched for GO categories involved in oxidoreductase activity (Table 2), including several proteins annotated as CYPs (PF00067), alcohol dehydrogenases (ADHs) (PF08240), and ALDHs (PF00171). Furthermore, we found that paralogous genes, including several CYPs, are enriched for such mutations that can affect their properties (P < 2.2·10^{-16}), which may contribute to diversification of these functions. Among the 872 putative paralogous genes predicted in R. mackenziei, 186 carry at least one severe mutation.

**Indel variation**

We identified 26,754 potential indels (dH24460: 14,837 and IHM 22877: 11,917). Overall, 7.7% of these mutations were found in coding regions with a mean length of 7.3 and 8.4 bp for insertions, and 14.9 and 16.6 bp for deletions, in the strains dH24460 and IHM 22877, respectively. Indels that are multiples of 3 bp, and thus cannot cause frameshifts, account for 42% of the coding indels in dH24460 and 38% in IHM 22877 (Figure S1).
Intronic regions comprise 5.8% of the indels (dH24460: 845 and IHM 22877: 722). Rates of SNPs and indels were more frequent in *R. mackenziei* dH24460, revealing that this is more highly diverged from the reference isolate.

Overall, *R. mackenziei* strain dH24460 possessed 95 potential inversions with a mean length of 136,002 bp and strain IHM 22877 harbored 287 potential inversions with a mean length of 45,324 bp (Figure 2F). The three longest inversions were identified in scaffold 3 corresponding to 64.5 and 60.3% of the total length of inversions in strains IHM 22877 and dH24460, respectively. The highest density of inversions was found in scaffold 1 in both strains of *R. mackenziei* (Figure 2).

### Protein-coding gene annotation and general features

Protein-coding genes were predicted based on alignments of proteins of *R. mackenziei* CBS 650.93 against the genome of the strains IHM 22877 (GenBank accession: GCA_001723215) and dH24460 (GenBank accession: GCA_001723235). The reference protein set contained 11,382 proteins. In total, 11,148 genes were predicted in strain IHM 22877 and 11,191 genes in dH24460, covering 97.9 and 98.2% of the gene content of the type strain, respectively. Using the reference protein set, a BlastP analysis against the NCBI database revealed that many of these enzymes are often found in plant-associated fungi (Table S5), for example the putative glutathione S-transferase protein Z518_06694 (86% identity with *Eutypa lata*), nitrilase Z518_09980 (66% identity with *Botrytis cinerea*), CYP Z518_01777 (93% identity with *Oidiodendron maius*), the putative siderophore iron transporter Z518_06150 (identity 77% with *Phacemoniella chlamydospora*), and the myo-inositol transporter Z518_03615 (74% with *O. maius*). Other enzymes involved in the iron transport, such as the siderophore iron transporters Z518_10557 and Z518_10557, were similar to the saprofic fungus *Purpureocillium lilacinum*.

### R. mackenziei secretome

In *R. mackenziei*, the pool of secreted proteins (secretome) includes several virulence factors, such as proteases and chitinases. Analyses of secretion signal peptides, the N-terminal regions that control the entry of proteins to the secretory pathway, revealed the presence of 216 putative secretory proteins in this fungus. Among these enzymes, 20 peptides, 46 carbohydrate-active enzymes (CAZY), four laccases, and one CYP were identified (Table S7). The signal peptide was also found in three out of four tyrosinases reported in *R. mackenziei*. Interestingly, extracellular tyrosinases and laccases are involved in oxidation and scavenging of phenolic compounds, as well as in the melanin biosynthesis (van Gelder et al. 1997; Li et al. 2016).

### CYPs

CYPs are involved in primary, secondary, and xenobiotic metabolism, playing a diverse role in fungal pathogenicity and detoxification/degradation of toxic exogenous compounds. These enzymes may allow
the fungus to grow under different stressful conditions or adapt to polluted environments (Crešnar and Petrić 2011). Black yeasts possess one of the highest rates of CYP genes in Ascomycota reported to date (Teixeira et al. 2017). These include families of CYPs thought to be involved in the metabolism of phenolic compounds and aromatic hydrocarbons, such as CYP530, CYP682, CYP504, and CYP52 (Teixeira et al. 2017). In this study, we analyzed 50 CYPs previously reported in R. mackenziei in which their precise families or subfamilies could not be assigned based on the fungal P450 CYPs database (Nelson 1998) (http://blast.uthsc.edu) and following the International P450 Nomenclature Committee rule, which suggests that proteins with \( > 40\% \) identity and \( > 55\% \) identity may be clustered under the same family and subfamily, respectively. For that, we used phylogenetic analysis including 645 CYP protein sequences from the black yeasts C. bantiana, C. psammophila, E. dermatitidis, E. xenobiotica, and R. mackenziei (Teixeira et al. 2017). CYPs within the same family/subfamily clustered together in the tree, indicating that their annotation based on the BLASTP identity was correct. Among the 50 unannotated P450s, 33 were accommodated in 18 families with CYPs of other black yeasts, possibly with overlapping functions (Table S8). Three CYPs belonged to the CYP52 family, which is involved in alkane assimilation and fatty acid metabolism in yeasts (Seghezzi et al. 1992; Ohkuma et al. 1995). Consistent with other studies (Pedrini et al. 2013), on the phylogenetic tree, the CYP52 family is very close to the CYP584 family, indicating that these CYPs derived from a single ancestor. Seventeen CYPs were not assigned to any family and may consist of new P450s, specific to R. mackenziei, and their accurate nomenclature remains unclear. We identified seven clans (Figure 4), a higher level of P450 classification comprising evolutionarily related

### Table 2 GO categories enriched among genes with NSY SNPs and scored by PROVEAN as severe mutations

| Strain     | GO Category                               | Term                                                                 | P-Value | Top Five PFAM Domains                                      |
|------------|-------------------------------------------|-----------------------------------------------------------------------|---------|-----------------------------------------------------------|
| IHM 22877  | GO:0016491 Oxidoreductase activity         |                                                                       | 1.74E–08| PF00106, PF00107, PF08240, PF01266, PF00171               |
|            | GO:0055114 Oxidation-reduction process    |                                                                       | 3.00E–08| PF00067, PF08240, PF00107, PF01266, PF0732               |
|            | GO:0042623 ATPase activity, coupled        |                                                                       | 5.30E–08|                                                            |
|            | GO:0016887 ATPase activity                |                                                                       | 7.13E–08| PF00005                                                   |
|            | GO:0015399 Primary active transmembrane   |                                                                       | 7.90E–07|                                                            |
|            |                                           | transport activity                                                     |         |                                                            |
|            | GO:0015405 P-P-bond-hydrolysis-driven     |                                                                       | 7.90E–07|                                                            |
|            |                                           | transmembrane transporter activity                                     |         |                                                            |
|            | GO:0043492 ATPase activity, coupled to    |                                                                       | 1.37E–06|                                                            |
|            | movement of substances                     |                                                                       |         |                                                            |
|            | GO:0042626 ATPase activity, coupled to    |                                                                       | 2.08E–06| PF00664, PF06422                                         |
|            | transmembrane movement of substances      |                                                                       |         |                                                            |
|            | GO:0016462 Pyrophosphatase activity       |                                                                       | 2.43E–06|                                                            |
|            | GO:0016818 Hydrolase activity, acting on  |                                                                       | 4.52E–06|                                                            |
|            | acid anhydrides, in phosphorus-containing |                                                                       |         |                                                            |
|            | anhydrides                                |                                                                       |         |                                                            |
|            | GO:0016820 Hydrolase activity, acting on  |                                                                       | 4.57E–06|                                                            |
|            | acid anhydrides, catalyzing transmembrane |                                                                       |         |                                                            |
|            | movement of substances                     |                                                                       |         |                                                            |
|            | GO:0017111 Nucleoside-triphosphatase      |                                                                       | 5.58E–06|                                                            |
|            | activity                                  |                                                                       |         |                                                            |
|            | GO:0016817 Hydrolyase activity, acting on |                                                                       | 9.17E–06|                                                            |
|            | acid anhydrides                            |                                                                       |         |                                                            |
|            | GO:0003824 Catalytic activity             |                                                                       | 2.29E–05| PF00501, PF01048, PF02515, PF00128, PF01557            |
|            | GO:0022804 Active transmembrane transporter| activity                                                             | 5.85E–05|                                                            |
|            | activity                                  |                                                                       |         |                                                            |
|            | GO:0046906 Tetrapyrrole binding           |                                                                       | 1.30E–04|                                                            |
|            | GO:0020037 Heme binding                   |                                                                       | 1.30E–04| PF00067, PF01077, PF00199                               |
|            | GO:0016705 Oxidoreductase activity, acting|                                                                       | 1.59E–04| PF00067                                                   |
|            | on paired donors, with incorporation or   |                                                                       |         |                                                            |
|            | reduction of molecular oxygen             |                                                                       |         |                                                            |
|            | dH24460                                    |                                                                       |         |                                                            |
|            | GO:0016491 Oxidoreductase activity         |                                                                       | 3.11E–07| PF00106, PF00107, PF08240, PF02668, PF01565             |
|            | GO:0055114 Oxidation-reduction process    |                                                                       | 1.45E–06| PF00067, PF08240, PF00107, PF02668, PF00732             |
|            | GO:0016705 Oxidoreductase activity, acting|                                                                       | 9.97E–06| PF00067                                                   |
|            | on paired donors, with incorporation or   |                                                                       |         |                                                            |
|            | reduction of molecular oxygen             |                                                                       |         |                                                            |
|            | GO:0005506 Iron ion binding                |                                                                       | 4.32E–05| PF00067, PF10637, PF00848                               |

GO, gene ontology; PFAM, protein family database.
families that probably have common functions (Nelson 1998). Clan 1 was the most populated, although the majority of unannotated CYPs in *R. mackenziei* belonged to clan 2.

**Metabolism of fatty acids and two-carbon compounds**

*R. mackenziei* harbored genes for the glyoxylate cycle, which allows the assimilation of fatty acids and two-carbon compounds, such as acetate and ethanol, to produce glucose (Kornberg 1966). This pathway has been associated with fungal virulence, as it permits energy production in environments that are poor in complex carbon compounds (Lorenz and Fink 2001). Two enzymes are considered key in the glyoxylate cycle and were identified in *R. mackenziei*: isocitrate lyase (Z518_04722 and Z518_01820) and malate synthase (Z518_01715). Similarly to other black yeast species, *R. mackenziei* possessed a paralog copy for isocitrate lyase.

**Metabolism of aromatic compounds**

The genome of *R. mackenziei* CBS 650.93 was assessed in order to identify genes linked to the degradation of aromatic compounds, such as benzene, toluene, ethylbenzene, xylene, and styrene, important pollutants from petrochemical and chemical industries. Toluene is initially oxidized to benzyl alcohol by a membrane-bound CYP in toluene-growing cells of the closely related *C. saturnica* CBS 114326, previously confused with *Cladosporium sphaerospermum* (Luykx et al. 2003). Comparative analysis against the aromatic hydrocarbon-degrading fungus *C. immunda* revealed the presence of orthologs that resemble...
the published fungal toluene degradation pathway via protocatechuic acid (Parales et al. 2008; Blasi et al. 2017). *R. mackenziei* possesses five highly conserved homologs of CYP that were overexpressed when *C. immunda* was grown in the presence of toluene (Figure 5) (Blasi et al. 2017). Among them, the CYP (Z518_09427) belongs to the cytochrome family CYP53, a benzoate para-hydroxylase that is associated with the detoxification of xenobiotic compounds in other ascomycetes (Faber et al. 2001; Jawallapersand et al. 2014). Furthermore, CYP53 has been proposed as a novel alternative antifungal drug target (Podobnik et al. 2008). Benzy1 alcohol is converted to protocatechuic acid through a multistep reaction involving four enzymes (Figure 5: enzymes two–five). Contrary to *C. immunda*, where protocatechuic acid can be converted to catechol, in *R. mackenziei* the only route for the processing of this compound seems to be via the β-ketoacipate pathway (Figure 5). *R. mackenziei* also possesses enzymes of the styrene and the phenylalanine degradation pathways. Accordingly, styrene might be degraded via phenylacetic acid, which is converted to homogentisic acid, then oxidized to 4-maleylacetoacetate by the homogentisate 1,2-dioxygenase (Z518_05388, Z518_05995, and Z518_06895, and Z518_09726), which is converted to fumarylacetoacetate and to fumarate/acetocacetate by the enzymes of the fumurate pathway (Z518_00072) and fumarylacetocetase (Z518_00072), respectively. This data suggests that the catabolism of aromatic amino acids, such as phenylalanine, can occur via oxidation to phenylacetic acid. Previously studies suggested that a CYP504 family member is involved in the catabolism of both phenylacetate and 2-hydroxyphenylacetate acid, which is then oxidized to homogentisic acid and degraded to fumarate and acetocacetate (Rodriguez-Saiz et al. 2001). These pathways may be important to allow this fungus to thrive on environmental contaminants conferring both metabolic versatility and defense against toxic chemicals.

**Iron acquisition strategies**

Different strategies for iron acquisition were also present in *R. mackenziei*, allowing the uptake and storage of iron from the host during infection. The enzymes that are required for the production of siderophores (iron-specific chelators) in *E. dermatitidis* were also found in *R. mackenziei*. In contrast to other neurotropic fungi (Z. Chen et al. 2014), the *SidD* and *SidF* genes (Z518_02441 and Z518_02442, respectively) are adjacent in *R. mackenziei*. The *SidA* and *SidC* genes (Z518_09385 and Z518_09384, respectively) are adjacent in both *R. mackenziei* and *E. dermatitidis*. The cluster containing the high-affinity iron permease FtrA and the ferroxidase FetC are duplicated in *R. mackenziei* (Z518_02600- Z518_06281 and Z518_02599-Z518_06280), as well as in the neurotropic black fungi *E. dermatitidis* (HMPREF1120_01590-HMPREF1120_04510 and HMPREF1120_01589-HMPREF1120_04509) and *C. bantiana* (Z519_09337- Z519_11106 and Z519_09338-Z519_1110).

**Mechanisms of osmotolerance**

*R. mackenziei* possesses distinctive mechanisms to enable its survival under extreme environmental conditions, such as the heat and the dryness of the desert where the environmental phase of the fungus is thought to reside. To avoid dehydration of cells, some fungi are able to accumulate neutral and low-molecular-weight compounds intracellularly,
Pathogenicity-related genes

In total, 520 pathogenicity-related genes were found in R. mackenziei. In general, the species shared many potential virulence-associated homologs with the neurotropic fungus Cryptococcus neoformans (Table S10). Sequence analyses revealed that gene Z518_10204, with equivalent copies found in all three R. mackenziei strains examined, corresponded to the DEAD-box RNA helicase VAD1 (Panepinto et al. 2005). Protein Z518_10204 in R. mackenziei possessed the IPRs related to VAD1 protein—IPR001650, IPR014001, IPR014014, IPR027417 and IPR011545—supporting the functional annotation. VAD1 protein is believed to be involved in transcriptional regulation as well as in RNA stability, and plays an essential role in stress response, salt tolerance, and regulation of the virulence factor laccase (Panepinto et al. 2005). This protein is present in Cr. neoformans (75% identity to Z518_10204) and is expressed during infection of human brain (Panepinto et al. 2005). Other pathogenic species belonging to the order Onygenales, such as Co. immitis (83% identity) and T. rubrum (83% identity), possessed homologs to Z518_10204. Additionally, gene Z518_03960, which encodes a phosphoenolpyruvate carboxykinase (PCK1) and is a virulence factor dependent on VAD1, was found conserved in R. mackenziei compared with Cr. neoformans (64% identity).

Urease is another interesting factor shared by R. mackenziei and C. neoformans. This protein has been postulated as an important pathogenicity-related gene facilitating fungal central nervous system invasion (Olszewski et al. 2004; Singh et al. 2013). In general, black yeasts carry a single gene copy coding for urease, except C. immunda, which possesses two urease copies. Despite the presence of conserved IPRs specific to the urease accessory proteins (UreF, UreD, UreG, and NIC1), their amino-acid sequences are poorly conserved compared to C. neoformans. In contrast, Z518_09873, the R. mackenziei homolog of URE1, shares 62% of BLASTP identity with URE1 in C. neoformans. Accessory proteins are responsible for activating the urease apoenzyme and are thought to be involved in human brain infection (Singh et al. 2013).

R. mackenziei also harbors candidate pathogenicity orthologs in species of Eurotiales. The glucosamine-fructose-6-phosphate aminotransferase (GFA1) in R. mackenziei shares 88% of BLASTP identity with that in A. fumigatus. Gfa1 protein catalyzes the first step in the chitin biosynthesis pathway, and is believed to be essential and a potential drug target enzyme in A. fumigatus (Hu et al. 2007).

Mitochondrial genomes

Contrary to nuclear DNA, the circular mitochondrial genome of R. mackenziei is AT-rich (~26% G+C). It contains 14 conserved protein coding genes, two rRNA genes, and 25 tRNA genes, which code for all 20 amino acids (Figure 6). All genes are located on the same strand and > 88% of the mitochondrial genome is coding sequence. Strain dH24460 contains three introns inside protein coding genes (cob, cox1, and cox3), while IHM 22877 contains no introns. The introns are group I and contain holding endonuclease genes; two of them belong to the LAGLIDADG family and one to the GIY-YIG family (cob). Besides the intron difference there are minor (0.3%) sequence differences between the two strains. In total, 66 positions are variable, 24 of which are in intergenic and 42 are in coding sequences. The tRNA genes contain no variation, the rRNA genes contain 14 variable positions, and the protein-coding genes 28, but only five of them result in amino acid substitutions.

The intron inside cob has its closest hit with an intron at the same position inside the cob gene of Fusarium oxysporum with 37% query

Figure 6 Mitochondrial genomes of R. mackenziei strains. IHM 22877 (inner circle) and dH24460 (outer circle). Blue arrows represent genes and yellow arrows are protein coding sequences. Red arrows indicate rRNA or tRNA coding sequences.
coverage, 8e—98 E-value, and 75% sequence identity. BLAST search of the NCBI database returned no results for the other two introns.

**DISCUSSION**

Our study catalogs the mutations present in two strains of the neurotropic fungus *R. mackenziei*. Additionally, SVs were assessed, revealing an unexpected amount of large inversions. Hypervariable sections were seen in the ends of scaffolds (probably indicating telomeric regions) composed by repetitive nucleotides, which are subject to higher levels of mutations in other fungi (Cuomo et al. 2007) (Figure 2). Repetitive DNA prediction supported the high incidence of these elements at both ends of the scaffolds. As demonstrated by SNPs and indels, genomic variation was more frequent in *R. mackenziei* strain dH24460 than in strain IIM 22877 (Figure 3). This finding suggests that this strain is more genetically divergent from the reference strain. In contrast, SV from the reference, such as long inversions, was greater in strain IIM22877.

Expanded protein families in *R. mackenziei*, such as the CYPs, ADHs, and ALDHs, are more susceptible to accumulating significant mutations, which can be important for the emergence of evolutionary novelties or may have possible deleterious effects (Table S2). The notorious abundance of CYPs, ADHs, and ALDHs in the genomes of black yeasts is partially explained by gene duplication events, occurring in the common ancestor of these organisms, by which new genes with similar functions were copied (Teixeira et al. 2017). Redundantly, we showed that there are significant mutations found in the paralogous gene sets. Taking the results together, we propose two scenarios to explain the accumulation of significant mutations in genes originated by duplication in *R. mackenziei*. First, the redundant copy became a nonfunctional pseudogene due to deleterious mutations, and second, these enzymes could be under positive selection driving niche-specific adaptation via neo-functionalization, although no evidence of functional divergence via neo-functionalization has ever been experimentally detected in black yeasts. Further analyses are required to determine whether gene duplication is a major force in the evolution of black yeasts.

Sequence comparisons with the best-studied neurotropic fungus, *C. neoformans*, provided important insights into genes and possible mechanisms involved in brain infection. Although pathologies caused by *R. mackenziei* (brain abscess), *C. neoformans*, and *C. gattii* [meningoencephalitis and propensity of *C. gattii* to form cryptococcomas (Sorrell 2001)] differ significantly, genes conserved in both species might explain common strategies to invade the host. For instance, genes conserved in both species might explain common strategies to invade the host. For instance, genes conserved in both species might explain common strategies to invade the host.

### ADHs

As members of the major homolog family of *C. bantiana* and *C. neoformans*, most of these homologs are shared with recognized alkylbenzene-degrading fungi, such as *Phanerochaete chrysosporium* and *Trichoderma reesei*. In general, enzymes with this feature are more susceptible to accumulating significant mutations, which can be important for the emergence of evolutionary novelties or may have possible deleterious effects (Table S2). The notorious abundance of CYPs, ADHs, and ALDHs in the genomes of black yeasts is partially explained by gene duplication events, occurring in the common ancestor of these organisms, by which new genes with similar functions were copied (Teixeira et al. 2017). Redundantly, we showed that there are significant mutations found in the paralogous gene sets. Taking the results together, we propose two scenarios to explain the accumulation of significant mutations in genes originated by duplication in *R. mackenziei*. First, the redundant copy became a nonfunctional pseudogene due to deleterious mutations, and second, these enzymes could be under positive selection driving niche-specific adaptation via neo-functionalization, although no evidence of functional divergence via neo-functionalization has ever been experimentally detected in black yeasts. Further analyses are required to determine whether gene duplication is a major force in the evolution of black yeasts.

Sequence comparisons with the best-studied neurotropic fungus, *C. neoformans*, provided important insights into genes and possible mechanisms involved in brain infection. Although pathologies caused by *R. mackenziei* (brain abscess), *C. neoformans*, and *C. gattii* [meningoencephalitis and propensity of *C. gattii* to form cryptococcomas (Sorrell 2001)] differ significantly, genes conserved in both species might explain common strategies to invade the host. For instance, genes conserved in both species might explain common strategies to invade the host.
The amount of secretory laccases in *R. mackenziei* is consistent with what was recently described for black yeasts of the genus *Fonsecaea* (Moreno et al. 2017). Laccases may catalyze the oxidation of phenolic compounds and aromatic amines (Feng et al. 2015). The most abundant secreted peptidases belong to the S10 family, according to MEROPS (a proteolytic enzyme database), which contains a variable set of carboxypeptidases. Carboxypeptidase S10 was earlier described in the dermatophyte *T. rubrum* and was considered as a virulence factor (Zaugg et al. 2008). Moreover, peptidases of the families A01 (pepsin), S09 (serine-dependent peptidases), and S33 (exopeptidases) were found in multiple copies in the repertoire of proteins with secreted proteolytic activity. Thus, we speculate that these secreted proteases are important for the virulence of *R. mackenziei*, allowing this fungus to grow on the nitrogenous substrates during infection.

The evolution of true mammal pathogens, such as members of the order Onygenales (Sharpton et al. 2009), is often associated with gene family loss leading to pathways lacking enzymes. Conversely, *R. mackenziei* possesses a wide arsenal of genes encoding enzymes for nutrient acquisition and primary metabolism (e.g., nitrogen metabolism, carbohydrate metabolism, and fatty acid metabolism). Overall, the pathway collection in this fungus resembles that of black yeasts that are able to assimilate alkylbenzene hydrocarbons. These results confirm that, unlike other true pathogens, *R. mackenziei* is a versatile fungus, with several pathways to sequester carbon from a wide range of nutrient substrates from the environment, possibly suggesting an oligotrophic lifestyle. In this respect, the fungus differs from true pathogens that need to extract nutrients from their hosts. Opportunism may be explained by the occurrence of virulence factors that have other roles in the natural life cycle of the fungus, such as the presence of a glyoxylate cycle pathway, different strategies for acquiring iron, toxin neutralization, and secondary metabolism (Teixeira et al. 2017). Key enzymes of the glyoxylate cycle are highly expressed during infection, possibly in response to conditions of starvation, e.g., in the phagolysosome during phagocytosis (Lorenz and Fink 2001). However, recent publications suggest that the glyoxylate cycle might not be required for virulence in *Cr. neoformans* (Rude et al. 2002) and other fungi (Schobel et al. 2007), which opens up a debate on its clinical significance. Utilization of C2 compounds via the glyoxylate shunt might also be a nutritional strategy adopted by oligotrophic black yeasts to survive in nutrient-depleted environments, avoiding loss of carbons as CO2 due to the decarboxylation steps in the tricarboxylic acid cycle.

ACKNOWLEDGMENTS

L.F.M. is supported by a Science without Borders fellowship (grant number: 1331-13-2) from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (capes.gov.br). Sequencing of strains IHM 22877 and dH24460 was supported by grant number 12-BIO2319-10 from King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.

LITERATURE CITED

Arzanlou, M., J. Z. Groenewald, W. Gams, U. Braun, H. D. Shin et al., 2007 Phylogenetic and morphotaxonomic revision of *Ramichloridium* and allied genera. Stud. Mycol. 58: 57–93.

Badali, H., G. S. de Hoog, I. Curfs-Breuker, and J. F. Meis, 2010 *In vitro* activities of antifungal drugs against *Rhinochlaena mackenziei*, an agent of fatal brain infection. J. Antimicrob. Chemother. 65: 175–177.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin et al., 2012 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19: 455–477.

Blasi, B., H. Tafer, C. Kustor, C. Poyntner, K. Lopandic et al., 2017 Genomic and transcriptomic analysis of the toluene degrading black yeast *Cladosporium phialophora* immunda. Sci. Rep. 7: 11436.

Blomberg, A., and L. Adler, 1992 Physiology of osmotolerance in fungi. Adv. Microb. Physiol. 33: 145–212.10.1016/S0006-2911(08)60217-9

Bolger, A. M., L. Mohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Brankovics, B. Z., A. D. van Diepeningen, T. A. J. van der Lee, C. Waalwijk, and G. S. de Hoog, 2016 GRABr: selective assembly of genomic regions, a new niche for genomic research. PLoS Comput. Biol. 12: e1004753.

Chen, K., J. W. Wallis, M. D. McLellan, D. E. Larson, J. M. Kalicki et al., 2009 BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods 6: 677–681.

Chen, W., M. K. Lee, C. Jefcoate, S. C. Kim, F. Chen et al., 2014 Fungal cytochrome p450 monoxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. Genome Biol. Evol. 6: 1620–1634.

Chen, Z., D. A. Martinez, S. Guuja, S. M. Sykes, Q. Zeng et al., 2014 Comparative genomic and transcriptomic analysis of *Wangiella dermatitidis*, a major cause of phaeohyphomycosis and a model black yeast human pathogen. G3 (Bethesda) 4: 561–578.

Choi, Y., G. E. Simis, S. Murphy, J. R. Miller, and A. P. Chan, 2012 Predicting the functional effect of amino acid substitutions and indels. PLoS One 7: e66888.

Creissen, B., and S. Petric, 2011 Cytochrome P450 enzymes in the fungal kingdom. Biochim. Biophys. Acta 1814: 29–35.

Cristini, A., D. García-Hermoso, M. Celard, G. Albrand, and O. Lortholary, 2010 Cerebral phaeohyphomycosis caused by *Rhinochlaena mackenziei* in a woman native to Afghanistan. J. Clin. Microbiol. 48: 3451–3454.

Cuomo, C. A., U. Guldener, J. R. Xu, F. Trail, B. G. Turgeon et al., 2007 The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. Science 317: 1400–1402.

de Hoog, G. S., J. Guarro, J. Gené, and M. J. Figueras, 2000 *Fusarium* sp.: a review. Stud. Mycol. 43: 176.

de Hoog, G. S., J. S. Zeng, M. J. Harrak, and D. A. Sutton, 2006 *Exopolithia xenobiotica* sp. nov., an opportunistic black yeast inhabiting environments rich in hydrocarbons. Antonie van Leeuwenhoek 90: 257–268.

Faber, B. W., R. F. van Gorcom, and J. A. Duine, 2001 Purification and characterization of benzoate-para-hydroxylase, a cytochrome P450 (CYP53A1), from *Aspergillus niger*. Arch. Biochem. Biophys. 394: 245–254.

Frylinger, S. G., R. B. Gaines, and L. E. B. Jr, 1999 Quantitative determination of BTEX and total aromatic compounds in gasoline by comprehensive two-dimensional gas chromatography (GC×GC). J. High Resolut. Chromatogr. 22: 195–200.

Fuchs, G., M. Boll, and J. Heider, 2011 Microbial degradation of aromatic compounds - from one strategy to four. Nat. Rev. Microbiol. 9: 803–816.

Ghoneim, D. H., J. R. Myers, E. Tuttle, and A. R. Paciorkowski, 2014 Comparison of insertion/deletion calling algorithms on human next-generation sequencing data. BMC Res. Notes 7: 864.

Horrè, R., and G. de Hoog, 1999 Primary cerebral infections by melanized fungi: a review. Stud. Mycol. 43: 176–193.

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.

Hu, W., S. Sillaots, S. Lemieux, J. Davison, S. Kaufman et al., 2007 Essential gene identification and drug target prioritization in *Aspergillus fumigatus*. PLoS Pathog. 3: e24.

Huang, F. C., A. Peter, and W. Schwab, 2014 Expression and characterization of CYP52 genes involved in the biosynthesis of sophorolipid and alkane metabolism from *Starmerella bombicola*. Appl. Environ. Microbiol. 80: 766–776.

Isola, D., L. Selblmann, G. S. de Hoog, M. Fenice, S. Onofri et al., 2013 Isolation and screening of black fungi as degraders of volatile aromatic hydrocarbons. Mycopathologia 175: 369–379.
Van der Auwerda, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel et al., 2013 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43: 11.10.1–11.10.33.

van Gelder, C. W., W. H. Flurkey, and H. J. Wichers, 1997 Sequence and structural features of plant and fungal tyrosinases. Phytochemistry 45: 1309–1323.

Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel et al., 2014 Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9: e112963.

Winnenburg, R., T. K. Baldwin, M. Urban, C. Rawlings, J. Kohler et al., 2006 PHI-base: a new database for pathogen host interactions. Nucleic Acids Res. 34: D459–D464.

Ye, K., M. H. Schulz, Q. Long, R. Apweiler, and Z. Ning, 2009 Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25: 2865–2871.

Zaugg, C., O. Jousson, B. Lechenne, P. Staub, and M. Monod, 2008 Trichophyton rubrum secreted and membrane-associated carboxypeptidases. Int. J. Med. Microbiol. 298: 669–682.

Zhao, J., J. Zeng, G. S. de Hoog, D. Attili-Angelis, and F. X. Prenafeta-Boldú, 2010 Isolation and identification of black yeasts by enrichment on atmospheres of monoaromatic hydrocarbons. Microb. Ecol. 60: 149–156.

Communicating editor: A. Rokas