genomeA
Journals.ASM.org

Draft Genome Sequence of the Pathogenic Fungus *Scedosporium apiospermum*

Patrick Vandeputte,a,b Sarah Ghamrawi,a Mathias Rechenmann,c Agnès Iltis,c Sandrine Giraud,a Maxime Fleury,a Christopher Thornton,d Laurence Delhaès,e Wieland Meyer,f Nicolas Papon,g Jean-Philippe Boucharaa,b

L'UNAM Université, Université d'Angers, Groupe d'Etude des Interactions Hôte-Pathogène, Angers, Francea; Laboratoire de Parasitologie-Mycologie, Centre Hospitalier Universitaire, Angers, Franceb; Genostar, Montbonnot, Francec; Hybridoma, Exeter, United Kingdomd; Université de Lille II, Biologie et Diversité des Pathogènes Eucaryotes Emergents, Lille, Francee; Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Sydney Medical School, Westmead Hospital, Marie Bashir Institute for Infectious Diseases and Biosecurity, University of Sydney, Westmead Millenium Institute for Medical Research, Westmead, New South Wales, Australiaf; Université de Tours, Biomolécules et Biotechnologie Végétales, Faculté de Pharmacie, Tours, Franceg

The first genome of one species of the *Scedosporium apiospermum* complex, responsible for localized to severe disseminated infections according to the immune status of the host, will contribute to a better understanding of the pathogenicity of these fungi and also to the discovery of the mechanisms underlying their low susceptibility to current antifungals.

Address correspondence to Patrick Vandeputte, pvandepu@gmail.com.

*S*cedosporium apiospermum is a soil-borne opportunistic pathogen responsible for cutaneous or subcutaneous mycetomas following traumatic inoculation of fungal elements and for respiratory tract infections. Moreover, it becomes a redoubtable pathogen in immunocompromised patients, where it may cause disseminated infections in the skin, eye, bones, joints, and deep organs, such as the heart and central nervous system (1). Species of the *Scedosporium apiospermum* complex are the second most frequent molds, after *Aspergillus fumigatus*, and colonize the respiratory tract of cystic fibrosis patients, causing in this clinical context respiratory infections such as bronchitis and allergic bronchopulmonary mycoses, or disseminated life-threatening infections in cases of immunodeficiency such as corticosteroid-induced diabetes or after lung transplantation. Here we report the sequencing and annotation of the genome of one of the major species within this closely related species complex.

The genome sequence of *Scedosporium apiospermum* strain IHEM 14462, isolated in 1998 from a sputum sample from a cystic fibrosis patient in Tours, France, was resolved by two distinct high-throughput Illumina sequencing technologies on an HiSeq2000: a paired-end run, sequencing on average 50 bp at each extremity of approximately 120.3 million 250-bp inserts, and a mate-pairs run, sequencing on average 50 bp at each extremity of approximately 82.2 million 4-kb inserts. *De novo* assembly was achieved by an additional single-molecule real-time sequencing on a PacBio RSII instrument (Pacific Biosciences), thus fully sequencing 273,000 inserts with a mean size of 2.3 kb. After trimming bad-quality Illumina runs, the sequences were assembled by Genostar (Montbonnot, France) in 3,744 contigs with the CLC Genomics Workbench version 6.0.2 (http://www.clcbio.com/products/clc-genomics-workbench). PacBio RSII reads were then used to generate scaffolds from the contigs using the softwares BLASR (2) and SSPACE Premium scaffolder version 2.3 (3). Subsequently, gaps were closed with GapFiller version 1.10 (4).

Finally, 176 scaffolds were obtained with a mean size of 246,804 bp, representing a total length of 43.4 Mbp.

The genome annotation was performed by Genostar; prediction of coding DNA sequences (CDSs) was performed with Augustus version 2.5.5 (5), trained on an algorithm optimized for *Neurospora crassa*. The 10,919 putative CDSs identified were annotated by BLASTP (6) against the TrEmbl database (7). A function was inferred for a given CDS when the deduced protein sequence shared at least 80% similarity and 40% identity with a protein of known function in the TrEmbl database. A putative function was attributed to the remaining CDSs by functional domain searches through the Pfam database (8). Thus, a function was assigned to 8,818 out of the 10,919 CDSs (80.75%). For 813 CDSs, the functional annotation was refined by attribution of the enzyme classification numbers.

Genomic data will greatly improve our comprehension of the pathogenic mechanisms underlying scedosporiosis and will help our understanding of the low susceptibility of *Scedosporium apiospermum* to currently available antifungal drugs (9).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number JOWA00000000. The version described in this paper is the first version, JOWA01000000.

ACKNOWLEDGMENT

REFERENCES

1. **Guarro J, Kantarcioglu AS, Horré R, Rodriguez-Tudela JL, Cuenca Estrella M, Berenguer J, de Hoog GS.** 2006. *Scedosporium apiospermum*: changing clinical spectrum of a therapy-refractory opportunist. Med. Mycol. 44:295–327. http://dx.doi.org/10.1080/13693780600752507.
2. **Chaisson MJ, Tesler G.** 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): applica-

tion and theory. BMC Bioinformatics **13**:238. http://dx.doi.org/10.1186/1471-2105-13-238.

3. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**:578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.

4. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol. **13**:R56. http://dx.doi.org/10.1186/gb-2012-13-6-r56.

5. **Stanke M, Steinkamp R, Waack S, Morgenstern B.** 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res. **32**:W309–W312. http://dx.doi.org/10.1093/nar/gkh379.

6. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

7. **Magrane M, UniProt Consortium.** 2011. UniProt Knowledgebase: a hub of integrated protein data. Database 2011:bar009. http://dx.doi.org/10.1093/database/bar009.

8. **Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD.** 2012. The Pfam protein families database. Nucleic Acids Res. **40**:D290–D301. http://dx.doi.org/10.1093/nar/gkr1065.

9. **Lackner M, de Hoog GS, Verweij PE, Najafzadeh MJ, Curfs-Breuker I, Klaassen CH, Meis JF.** 2012. Species-specific antifungal susceptibility patterns of *Scedosporium* and *Pseudallescheria* species. Antimicrob. Agents Chemother. **56**:2635–2642. http://dx.doi.org/10.1128/AAC.05910-11.