

Draft Genome of Australian Environmental Strain WM 09.24 of the Opportunistic Human Pathogen *Scedosporium aurantiacum*

Åsa Pérez-Bercoff,^{a,b} Alexie Papanicolaou,^{c,d} Marc Ramsperger,^a Jashanpreet Kaur,^e Hardip R. Patel,^b Azian Harun,^{a,f} Shu Yao Duan,^a Liam Elbourne,^d Jean-Philippe Bouchara,^{g,h} Ian T. Paulsen,^d Helena Nevalainen,^d Wieland Meyer,^a Gavin A. Huttley^b

Centre for Infectious Diseases and Microbiology, Molecular Mycology Research Laboratory, Marie Bashir Institute for Infectious Diseases and Biosecurity, Sydney Medical School-Westmead Hospital, University of Sydney, Westmead Millennium Institute, Sydney, NSW, Australia^a; John Curtin School of Medical Research, Australian National University, Arts and Culture Trust, Canberra, Australia^b; CSIRO Land and Water Flagship, Black Mountain Laboratories, CSIRO, Arts and Culture Trust, Canberra, Australia^c; Hawkesbury Institute for the Environment, University of Western Sydney, Arts and Culture Trust, Penrith, Australia^d; Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW, Australia^e; School of Medical Sciences, Universiti Sains Malaysia, Kota Bharu, Kelantan, Malaysia^f; L'UNAM Université, Université d'Angers, Groupe d'Etude des Interactions Hôte-Pathogène, Angers, France^g; Laboratoire de Parasitologie-Mycologie, Centre Hospitalier Universitaire, Angers, France^h

We report here the first genome assembly and annotation of the human-pathogenic fungus *Scedosporium aurantiacum*, with a predicted 10,525 genes, and 11,661 transcripts. The strain WM 09.24 was isolated from the environment at Circular Quay, Sydney, New South Wales, Australia.

Received 19 December 2014 Accepted 5 January 2015 Published 12 February 2015

Citation Pérez-Bercoff Å, Papanicolaou A, Ramsperger M, Kaur J, Patel HR, Harun A, Duan SY, Elbourne L, Bouchara J-P, Paulsen IT, Nevalainen H, Meyer W, Huttley GA. 2015. Draft genome of Australian environmental strain WM 09.24 of the opportunistic human pathogen *Scedosporium aurantiacum*. *Genome Announc* 3(1):e01526-14. doi:10.1128/genomeA.01526-14.

Copyright © 2015 Pérez-Bercoff et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Wieland Meyer, wieland.meyer@sydney.edu.au, or Gavin A. Huttley, gavin.huttley@anu.edu.au.

Scedosporium aurantiacum is a hyphomycetous filamentous fungus found in various environments such as soil, sewage and polluted waters (1). It is an emerging opportunistic pathogen capable of causing a range of infections that are especially abundant in Australia (2). To enable the genetic characterization of the virulence and high antifungal resistance potential of this species (3), the genome of the environmental, highly virulent *S. aurantiacum* strain WM 09.24, collected from Circular Quay, Sydney, NSW, Australia, in 2009, was sequenced (A. Harun and W. Meyer, unpublished data). The isolate selection was based on a global multilocus sequence typing (MLST) study (A. Harun and W. Meyer, unpublished data) and determination of its virulence using the *Galleria mellonella* model (S. Duan and W. Meyer, unpublished data).

Genomic DNA was sequenced (paired-end [PE] reads) on Illumina HiSeq 2000 at the Ramaciotti Centre for Genomics, University of New South Wales (UNSW), Sydney, NSW, Australia. Trimmomatic (version 0.27) (4) was used to clip off adapters and trim the reads prior to assembly with SPAdes (version 3.1.1) (5). After correction with REAPR (version 1.0.16) (6), and discarding of scaffolds <200 nucleotides (nt), the assembly consisted of 1,584 scaffolds (202 nt to 380,183 nt in length) and had a total genome size of 39,890,731 nt with 49.20% GC content. The assembly's N_{50} was 78,269 nt and N_{90} was 16,521 nt and was calculated to have an average coverage of 162× over all 1,584 contigs. The CEGMA pipeline (7) indicated an assembly completeness of 93.15%. RepeatMasker version open-4.0.3 (<http://www.repeatmasker.org>) was used to identify repeats in the genome assembly and 1.96% of the bases were masked.

To facilitate genome annotation, RNA-seq was obtained from *S. aurantiacum* strain WM 09.24 grown in different media at 25°C and 37°C, pertaining to environmental and human host growth conditions, respectively, as follows: (i) water medium for starva-

tion, (ii) potato dextrose (PD) to promote good general growth, and (iii) artificial sputum medium (ASM) to simulate growth conditions in an infected lung of a cystic fibrosis (CF) patient (8; M. Ramsperger and W. Meyer, unpublished data). Single-end (SE) reads of 100 bp were generated using the Illumina HiSeq 2500 at the Biomolecular Research Facility (BRF) at the John Curtin School for Medical Research (JCSMR), Australian National University (ANU), Canberra, ACT, Australia. Genome annotation was performed using the JAMg pipeline (9) and GMAP (version 2014-12-06) (10). We predicted 10,525 gene models and 11,661 transcripts for the genome assembly of *S. aurantiacum* strain WM 09.24. The number of inferred genes in *S. aurantiacum* is comparable to *Trichoderma virens*, which has 12,400 genes (11), and to the recently published genome of *Scedosporium apiospermum* strain IHEM 14462, which was found to have 10,919 coding sequences (CDSs) (12).

The herein-reported high-quality draft assembly of the environmental and highly virulent *S. aurantiacum* strain WM 09.24, together with the recently published genome of *S. apiospermum* strain IHEM 14462 (12) will provide fundamental genomics resources to study the genetic basis of virulence and antifungal resistance in this genus.

Nucleotide sequence accession numbers. This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number **JUDQ00000000**. The version described in this paper is the first version, **JUDQ01000000**.

ACKNOWLEDGMENTS

We are grateful to the staff at the Genome Discovery Unit (GDU) at the Biomolecular Research Facility (BRF) at the John Curtin School of Medical Research (JCSMR) at the Australian National University (ANU) for software installations and computer cluster provisions on which the anal-

yses were performed. We thank the staff at the BRF for carrying out the RNA sequencing. This work was supported by an ARC Super Science Fellowship FS110200026 (to I.T.P. and H.N.), and an NH&MRC grant APP1031943 (to W.M., G.A.H., and H.N.). This is a publication of the ISHAM working group on *Pseudallescheria/Scedosporium* Infections.

REFERENCES

- Kaltseis J, Rainer J, De Hoog GS. 2009. Ecology of *Pseudallescheria* and *Scedosporium* species in human-dominated and natural environments and their distribution in clinical samples. *Med Mycol* 47:398–405. <http://dx.doi.org/10.1080/13693780802585317>.
- Heath CH, Slavin MA, Sorrell TC, Handke R, Harun A, Phillips M, Nguyen Q, Delhaes L, Ellis D, Meyer W, Chen SC, Australian Scedosporium Study Group. 2009. Population-based surveillance for scedosporiosis in Australia: epidemiology, disease manifestations and emergence of *Scedosporium aurantiacum* infection. *Clin Microbiol Infect* 15: 689–693. <http://dx.doi.org/10.1111/j.1469-0691.2009.02802.x>.
- Lackner M, de Hoog GS, Yang L, Ferreira Moreno L, Ahmed SA, Andreas F, Kaltseis J, Nagl M, Lass-Flörl C, Risslegger B, Rambach G, Speth C, Robert V, Buzina W, Chen S, Bouchara J, Cano-Lira JF, Guarro J, Gené J, Fernández Silva F. 2014. Proposed nomenclature for *Pseudallescheria*, *Scedosporium* and related genera. *Fungal Divers* 67:1–10. <http://dx.doi.org/10.1007/s13225-014-0295-4>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
- Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biol* 14:R47. <http://dx.doi.org/10.1186/gb-2013-14-5-r47>.
- Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Res* 37:289–297. <http://dx.doi.org/10.1093/nar/gkn916>.
- Fung C, Naughton S, Turnbull L, Tingpej P, Rose B, Arthur J, Hu H, Harmer C, Fung C, Naughton S, Turnbull L, Tingpej P, Rose B, Arthur J, Hu H, Harmer C, Harbour C, Hassett DJ, Whitchurch CB, Manos J, Harbour C, Hassett DJ, Whitchurch CB, Manos J. 2010. Gene expression of *Pseudomonas aeruginosa* in a mucin-containing synthetic growth medium mimicking cystic fibrosis lung sputum. *J Med Microbiol* 59: 1089–1100. <http://dx.doi.org/10.1099/jmm.0.019984-0>.
- Papanicolaou A, Haas B. 2013. Just annotate my genome (JAMg). <http://jamg.sourceforge.net/>.
- Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875. <http://dx.doi.org/10.1093/bioinformatics/bti310>.
- Kersey PJ, Allen JE, Christensen M, Davis P, Falin LJ, Grabmueller C, Hughes DS, Humphrey J, Kerhornou A, Khobova J, Langridge N, McDowall MD, Maheswari U, Maslen G, Nuhn M, Ong CK, Paulini M, Pedro H, Toneva I, Tuli MA, Walts B, Williams G, Wilson D, Youens-Clark K, Monaco MK, Stein J, Wei X, Ware D, Bolser DM, Howe KL, Kulesha E, Lawson D, Staines DM. 2014. Ensembl genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res* 42:D546–D552. <http://dx.doi.org/10.1093/nar/gkt979>.
- Vandeputte P, Ghamrawi S, Rechenmann M, Iltis A, Giraud S, Fleury M, Thornton C, Delhaès L, Meyer W, Papon N, Bouchara JP. 2014. Draft genome sequence of the pathogenic fungus *Scedosporium apiospermum*. *Genome Announc* 2(5):e00988–14. <http://dx.doi.org/10.1128/genomeA.00988-14>.