

Draft genome sequences of the pathogenic fungi *Scedosporium aurantiacum* and *Scedosporium apiospermum* from clinical isolates

Samuel T. Montgomery,^{1,2} Hamideh Rezaee,³ Nicolau Sbaraini,³ Joshua Iszatt,¹ Liza Mantjani,¹ Sarah E. Kidd,^{4,5} Yit-Heng Chooi,³ Luke W. Garratt^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Scedosporium* species are filamentous fungi with inherent broad antifungal resistance that pose opportunistic infection threats. We present draft genome assemblies of *S. aurantiacum* (11 contigs) and *S. apiospermum* (9 contigs), derived from Oxford Nanopore sequencing of one Australian clinical isolate each.

KEYWORDS *Scedosporium*, genomes, filamentous fungi

In the 2022 World Health Organization Fungal Priority Pathogen List (1), *Scedosporium* species ranked fourth for research needs. Genomic variation of *Scedosporium* is poorly described (2). Here, deidentified clinical isolates of *S. aurantiacum* (strain ID-16-90050374; otomycosis, Darwin, Australia) and *S. apiospermum* (strain ID-17-90028250; subcutaneous mycosis, Melbourne, Australia) were obtained from the National Mycology Reference Centre (Adelaide, Australia) for research, as permitted under the National Statement on Ethical Conduct in Human Research. Isolates were cultivated in Sabouraud Dextrose liquid media at 37°C under aerobic conditions with agitation for 4–5 days. The mycelial biomass was collected by filtration through Kimwipes, ground to a fine powder using mortar and pestle with liquid nitrogen, and high-molecular-weight DNA extracted using phenol/chloroform (3). Indexed paired-end libraries for short-read sequencing were prepared using Nextera XT DNA Library Prep (Illumina, USA) under manufacturer's instructions, quality assessed using a Bioanalyzer 2100 (Agilent Technologies, USA), and whole-genome sequencing performed on the Illumina HiSeq 2500 platform with 100 bp paired-end reads. For long-read sequencing, Oxford Nanopore libraries prepared using the Native Barcoding Kit (NBD-114-96), with no size selection or shearing, were sequenced on a PromethION flow cell on a P2 Solo. Data were basecalled using dorado (v0.5.3) using the dna_r10.4.1_e8.2_400bps_sup@v4.3.0 model.

Long-read sequencing yielded 3,469,118 and 1,864,815 reads for *S. aurantiacum* and *S. apiospermum*, respectively, with an N50 value of 12.50 kb and 11.91 kb and read lengths up to 133 kb (Table 1). Data were filtered using chopper (v0.10.0) (4) to remove reads <1 kb and <Q20, and sequencing adaptors were trimmed using porechop_abi (v0.5.0) (5). Filtered reads were assembled using flye (v2.9.6) (6) using default parameters specifying --scaffold, --asm-coverage 50, and --read-error 0.01. The resulting *de novo* *S. aurantiacum* and *S. apiospermum* assemblies are 39,272,335 bp and 43,356,729 bp long, comprising 11 and 9 contigs, respectively (Table 1). Short reads, filtered with AAFTE (v0.2.2) and default parameters (7), were used to polish contigs using both PyPOLCA (v0.3.1) (8, 9) with the --careful flag and polypolish (v0.6.0) (10). Reads were mapped back to the final assemblies using minimap2 (v2.2.29) (11) with long reads using the lr:hq preset and short reads using the sr preset and quality metrics obtained using qualimap (v2.3) (12). Telomeric repeats (AACCCT) were identified using tidk (v0.2.65) (13), with

Editor Jason E. Stajich, University of California
Riverside, Riverside, California, USA

Address correspondence to Yit-Heng Chooi,
yitheng.chooi@uwa.edu.au, or Luke W. Garratt,
Luke.Garratt@thekids.org.au.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 27 August 2025

Accepted 2 December 2025

Published 17 December 2025

Copyright © 2025 Montgomery et al. This is an
open-access article distributed under the terms of
the [Creative Commons Attribution 4.0 International
license](#).

TABLE 1 Summary statistics for the *Scedosporium aurantiacum* and *Scedosporium apiospermum* genome assembly and annotation

	<i>S. aurantiacum</i> (ID-16-90050374)	<i>S. apiospermum</i> (ID-17-90028250)
Nanopore reads (<i>n</i>)	3,469,118	1,864,815
Read N50 (bp)	12,498	11,905
Median read Q score	19.0	18.9
Illumina reads (<i>n</i>)	40,364,656	40,576,204
Assembly size (bp)	39,272,335	43,389,729
Contigs (<i>n</i>)	11	9
Contig N50 (bp)	4,772,094	5,346,190
GC %	49.82	50.43
BUSCO complete %	95.5	95.3
Repeat sequences (%)	6.33	4.74
Long-read mapping (%)	99.87	99.99
Long-read mean coverage	745.58	324.28
Short-read mapping (%)	83.14	99.25
Short-read mean coverage	127.89	139.04
Predicted genes (<i>n</i>)	10,523	12,291
Predicted CDS (<i>n</i>)	10,244	12,012
GenBank BioSample accession	SAMN48578032	SAMN48578033
Genbank Assembly accession	GCA_052249765.1	GCA_052249755.1

S. aurantiacum containing telomeric repeats on both ends of eight contigs and on a single end of two contigs, while the *S. apiospermum* assembly contains telomeric repeats on both ends of eight contigs and a single end of one contig. We used funannotate2 (v25.7.1) (14) and helixer (v0.3.5) (15) for repeat masking, gene prediction, and annotation using all default evidence-guided and *ab initio* predictions. Protein evidence for model training was derived using *Sordariomycetes* as the seed data for BUSCO searches.

ACKNOWLEDGMENTS

The authors acknowledge Ms. Emily Wong, Dr. Cameron Gilchrist, and Ms. Su Hui Yong for their assistance.

This research was funded by the Australian Research Council (FT160100233) and the National Health and Medical Research Council (2036521).

AUTHOR AFFILIATIONS

¹Wal-Yan Respiratory Research Centre, Kids Research Institute Australia, The University of Western Australia, Perth, Western Australia, Australia

²Centre for Child Health Research, The University of Western Australia, Perth, Western Australia, Australia

³School of Molecular Sciences, The University of Western Australia, Perth, Western Australia, Australia

⁴National Mycology Reference Centre, Microbiology & Infectious Diseases, SA Pathology, Adelaide, South Australia, Australia

⁵School of Biological Sciences, Faculty of Sciences, Engineering and Technology, University of Adelaide, Adelaide, South Australia, Australia

AUTHOR ORCIDs

Samuel T. Montgomery  <http://orcid.org/0000-0001-7132-852X>

Hamideh Rezaee  <http://orcid.org/0000-0002-9096-3104>

Nicolau Sbaraini  <http://orcid.org/0000-0002-7745-0590>

Joshua Iszatt  <http://orcid.org/0000-0002-6394-5058>

Liza Mantjani  <http://orcid.org/0009-0009-5560-6492>

Sarah E. Kidd  <http://orcid.org/0000-0002-5957-4178>

Yit-Heng Chooi  <http://orcid.org/0000-0001-7719-7524>

Luke W. Garratt  <http://orcid.org/0000-0001-9381-326X>

FUNDING

Funder	Grant(s)	Author(s)
Department of Education and Training Australian Research Council (ARC)	FT160100233	Yit-Heng Chooi
National Health and Medical Research Council	2035621	Joshua Iszatt Sarah E. Kidd Yit-Heng Chooi

AUTHOR CONTRIBUTIONS

Samuel T. Montgomery, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft | Hamideh Rezaee, Data curation, Investigation, Methodology, Project administration, Writing – review and editing | Nicolau Sbaraini, Investigation, Methodology, Resources, Writing – review and editing | Joshua Iszatt, Data curation, Formal analysis, Funding acquisition, Methodology, Writing – review and editing | Liza Mantjani, Investigation, Methodology, Project administration, Writing – review and editing | Sarah E. Kidd, Funding acquisition, Resources, Writing – review and editing | Yit-Heng Chooi, Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing | Luke W. Garratt, Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

All data are available under the National Center for Biotechnology Information BioProject number [PRJNA1249256](https://doi.org/10.5281/zenodo.17538694). Assembled genomes are also available at <https://doi.org/10.5281/zenodo.17538694>. Long-read sequencing data are available at [SRX30026839](https://doi.org/10.5281/zenodo.17538694) (*S. aurantiacum*) and [SRX30026840](https://doi.org/10.5281/zenodo.17538694) (*S. apiospermum*). Short-read sequencing data are available at (*S. aurantiacum*) [SRX30026841](https://doi.org/10.5281/zenodo.17538694) and [SRX30026842](https://doi.org/10.5281/zenodo.17538694) (*S. apiospermum*).

REFERENCES

- World Health Organization. 2022. WHO fungal priority pathogens list to guide research, development and public health action. Geneva, Switzerland
- Venice F, Spina F, Davolos D, Ghignone S, Varese GC. 2023. The genomes of *Scedosporium* from environmental challenges and opportunism. *IMA Fungus* 14:25. <https://doi.org/10.1186/s43008-023-00128-3>
- Sambrook J. 2001. Edited by D. W. Russell. Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
- Bonenfant Q, Noé L, Touzet H. 2022. Porechop_ABI: discovering unknown adapters in ONT sequencing reads for downstream trimming. *bioRxiv*. <https://doi.org/10.1101/2022.07.07.499093>
- Freire B, Ladra S, Parama JR. 2022. Memory-efficient assembly using flye. *IEEE/ACM Trans Comput Biol Bioinform* 19:3564–3577. <https://doi.org/10.1109/TCBB.2021.3108843>
- Palmer JM, Stajich JE. 2018. Automatic assembly for the fungi (AAFTF): genome assembly pipeline v0.2.2 Zenodo. Available from: <https://zenodo.org/records/6326242>
- Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. 2024. How low can you go? Short-read polishing of Oxford Nanopore bacterial genome assemblies. *Microb Genom* 10:001254. <https://doi.org/10.1099/mgen.0.001254>
- Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. *PLoS Comput Biol* 16:e1007981. <https://doi.org/10.1371/journal.pcbi.1007981>
- Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 18:e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294. <https://doi.org/10.1093/bioinformatics/btv566>
- Brown MR, Manuel Gonzalez de La Rosa P, Blaxter M. 2025. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics* 41. <https://doi.org/10.1093/bioinformatics/btaf049>
- Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation (v1.8.1). Available from: <https://zenodo.org/records/4054262>
- Stiehler F, Steinborn M, Scholz S, Dey D, Weber APM, Denton AK. 2021. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. *Bioinformatics* 36:5291–5298. <https://doi.org/10.1093/bioinformatics/btaa1044>