



# Draft Genome Sequence of a Preterm Infant-Derived Isolate of *Candida parapsilosis*

Steve A. James,<sup>a</sup> Andrea Telatin,<sup>a</sup> David Baker,<sup>a</sup> Rhiannon Evans,<sup>a</sup> Paul Clarke,<sup>b,c</sup> Lindsay J. Hall,<sup>a,d</sup> Simon R. Carding<sup>a,c</sup>

<sup>a</sup>Gut Microbes and Health Research Programme, Quadram Institute Bioscience, Norwich, United Kingdom

<sup>b</sup>Neonatal Intensive Care Unit, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, United Kingdom

<sup>c</sup>Norwich Medical School, University of East Anglia, Norwich, United Kingdom

<sup>d</sup>Ziel-Institute for Food and Health, Technical University of Munich, Freising, Germany

**ABSTRACT** *Candida parapsilosis* is a human fungal pathogen of increasing incidence and causes invasive candidiasis, notably in preterm or low-birthweight neonates. Here, we present the genome sequence of *C. parapsilosis* NCYC 4289, a fecal isolate from a preterm male infant.

*Candida parapsilosis* is a dimorphic ascomycete yeast belonging to the *Lodderomyces* clade, a large monophyletic group of species, which includes a number of important human pathogens (e.g., *Candida albicans*) (1). Although often found in the human gut (2, 3), *C. parapsilosis* is primarily a skin commensal and is present in the hospital environment (4), especially the neonatal intensive care unit (NICU). With a capacity to form persistent biofilms, it can spread by horizontal transmission and is regarded as a significant neonatal pathogen, with low-birthweight preterm neonates at particular risk of infection (5). Here, we combined short- and long-read sequencing to obtain the genome sequence of *C. parapsilosis* NCYC 4289, a feces-derived isolate from a 31+5-week-old premature male infant delivered by caesarean section. A fecal homogenate was prepared in sterile phosphate-buffered saline (PBS) and cultured on Sabouraud dextrose (SD) agar plates containing penicillin (25 U/mL) and streptomycin (25 U/mL) at 30°C. Species identity, from single colonies, was determined by PCR amplification and Sanger sequencing of the ribosomal DNA internal transcribed spacer (ITS) region of the ribosomal DNA locus using primers ITS1F (6) and ITS4 (7). The ITS sequence of strain NCYC 4289 (GenBank accession number ERZ15609610) is 100% identical to that of the *C. parapsilosis* type strain CBS 604 (GenBank accession number AY391843).

For short- and long-read sequencing, total genomic DNA was extracted from a stationary-phase SD culture (MasterPure yeast DNA purification kit; Cambio), and cells were treated with Zymolyase (0.25 mg/mL) to aid cell wall disruption with an additional proteinase K treatment step included prior to DNA precipitation. Short-read Illumina sequencing was performed using a modified 20-fold dilution of DNA Prep (Flex) reagent and run on a NextSeq 500 sequencer, producing 7,418,311 paired-end 150-bp reads. Nanopore sequencing was obtained from two methods. First, a novel modified Illumina DNA Prep (Flex) approach used symmetrical 24-base barcoded primers and a long-range polymerase (8). Libraries were pooled and size selected on a SageELF 0.75% cassette, fractions from 4 kb and above were pooled, and long-read sequencing was performed using a MinION sequencer (Oxford Nanopore Technologies [ONT]), ligation sequencing kit SQK-LSK109 (ONT), and flow cell FLO-MIN106 R9.4.1 (ONT). This produced a total of 812,374 reads with an average read length of 4,591 bases. The second method followed the standard ligation protocol using the manufacturer's recommendations and loading on a second flow cell, producing 852,227 reads (average read length, 4,756 bases). Base calling was performed using Guppy (ONT; v.3.6.0) in high-accuracy mode (model dna\_r9.4.1\_450bps\_hac).

**Editor** Antonis Rokas, Vanderbilt University

**Copyright** © 2023 James et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Steve A. James, [steve.james@quadram.ac.uk](mailto:steve.james@quadram.ac.uk).

The authors declare no conflict of interest.

[This article was published on 27 February 2023 without the middle initial for Lindsay J. Hall. The byline was corrected in the current version, posted on 16 March 2023.]

**Received** 1 December 2022

**Accepted** 10 February 2023

**Published** 27 February 2023

Raw short- and long-read polishing, including the removal of adapters and low-quality bases, was performed using SeqFu 1.16 (9) and fastp 0.23 (10). In addition, long reads of <1 kb in length were discarded. The genome was assembled using Flye 2.9.1 (11), polished with four rounds of Pilon 1.24 (12), and refined using RagTag (13). The genome assembly comprised eight chromosome-sized contigs of >890 kb and a linear mitochondrial genome (29,583 bp). The total size of the genome was 13,082,726 bp, the  $N_{50}$  value was 2,085,264 bp, and the G+C content was 38.70%. The largest contig in the assembly was 3,026,395 bp. Genome completeness was estimated at 93.0% using BUSCO v5.4.4 (14). Dependencies and scripts are available at <https://github.com/quadram-institute-bioscience/ont-candida>.

**Data availability.** This whole-genome shotgun project has been deposited at DDJB/ENA/GenBank (BioProject number [PRJEB56866](https://www.ncbi.nlm.nih.gov/bioproject/PRJEB56866) and assembly accession number CAMXCU010000000.1). The version described in this paper is version 1. The raw reads were deposited at SRA (accession numbers [ERX9916594](https://www.ncbi.nlm.nih.gov/sra/ERX9916594) and [ERX9917623](https://www.ncbi.nlm.nih.gov/sra/ERX9917623)).

## ACKNOWLEDGMENTS

We gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC). This research was funded by BBSRC Core Capability Grant BB/CCG1860/1 and BBSRC Institute Strategic Programme Grant Gut Microbes and Health (BB/R012490/1) and constituent project BBS/E/F/000PR10353.

## REFERENCES

- Lachance M-A, Boekhout T, Scorzetti G, Fell JW, Kurtzman CP. 2011. *Candida* Berkhout (1923), p 987–1278. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts; a taxonomic study*, 5th ed, vol 2. Elsevier, Amsterdam, Netherlands.
- Strati F, Di Paola M, Stefanini I, Albanese D, Rizzetto L, Lionetti P, Calabro A, Jousson O, Donati C, Cavalieri D, De Filippo C. 2016. Age and gender affect the composition of fungal population of the human gastrointestinal tract. *Front Microbiol* 7:1227. <https://doi.org/10.3389/fmicb.2016.01227>.
- Schei K, Avershina E, Oien T, Rudi K, Follestad T, Salamati S, Odegard RA. 2017. Early gut mycobiota and mother-offspring transfer. *Microbiome* 5:107. <https://doi.org/10.1186/s40168-017-0319-x>.
- Sabino R, Sampaio P, Carneiro C, Rosado L, Pais C. 2011. Isolates from hospital environments are the most virulent of the *Candida parapsilosis* complex. *BMC Microbiol* 11:180. <https://doi.org/10.1186/1471-2180-11-180>.
- Pammi M, Holland L, Butler G, Gacser A, Bliss JM. 2013. *Candida parapsilosis* is a significant neonatal pathogen: a systematic review and meta-analysis. *Pediatr Infect Dis J* 32:E206–E216. <https://doi.org/10.1097/INF.0b013e3182863a1c>.
- Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol Ecol* 2:113–118. <https://doi.org/10.1111/j.1365-294x.1993.tb00005.x>.
- White TJ, Bruns TD, Lee SL, Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. In Innis MA, Gelfand DH, Sninsky JJ (ed), *PCR protocols: a guide to methods and applications*. Academic Press, San Diego, CA.
- Baker DJ, Aydin A, Le-Viet T, Kay GL, Rudder S, Martins LD, Tedim AP, Kolyva A, Diaz M, Alikhan NF, Meadows L, Bell A, Gutierrez AV, Trotter AJ, Thomson NM, Gilroy R, Griffith L, Adriaenssens EM, Stanley R, Charles IG, Elumogo N, Wain J, Prakash R, Meader E, Mather AE, Webber MA, Dervisevic S, Page AJ, O'Grady J. 2021. CoronaHiT: high-throughput sequencing of SARS-CoV-2 genomes. *Genome Med* 13:21. <https://doi.org/10.1186/s13073-021-00839-5>.
- Telatin A, Fariselli P, Birolo G. 2021. SeqFu: a suite of utilities for the robust and reproducible manipulation of sequence files. *Bioengineering* (Basel) 8:59. <https://doi.org/10.3390/bioengineering8050059>.
- Chen SF, Zhou YQ, Chen YR, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng QD, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Alonge M, Lebeigle L, Kirsche M, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2021. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *bioRxiv*. <https://doi.org/10.1101/2021.11.18.469135>.
- Manni M, Berkeley MR, Seppely M, Simao FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>.