Microbiology Resource Announcements

| Mycology | Announcement

Draft genome sequence of a non-human primate-derived isolate of *Candida parapsilosis*

Steve A. James,[1] Aimee Parker,[1] Catherine Purse,[1] David J. Baker,[1] Simon G. P. Funnell,[1,2] Simon R. Carding[1,3]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Candida parapsilosis* is a common human commensal and opportunistic fungal pathogen that is also found in non-human primates (NHPs). Here, we report the first draft sequence of *C. parapsilosis* NCYC 4418, a fecal isolate from an adult cynomolgus macaque.

KEYWORDS *Candida parapsilosis*, fungal pathogen, gut mycobiome, non-human primate, cynomolgus macaque, genome sequence

Candida parapsilosis is a dimorphic ascomycete yeast belonging to the *Lodderomyces* clade (1). With the capacity to form persistent biofilms, on both biotic and abiotic surfaces, it has become a major global cause of invasive candidiasis, most notably in preterm and low-birthweight neonates (2). While considered a commensal member of the human gastrointestinal tract and skin (3–5), *C. parapsilosis* is also found in non-human primates (NHPs) (6, 7). Here, we combined short- and long-read sequencing to obtain the genome sequence of *C. parapsilosis* NCYC 4418, a fecal isolate from a healthy adult cynomolgus macaque housed at a managed breeding colony (UK). Fecal homogenate, prepared in sterile phospate-buffered saline (PBS), was spread onto solid Sabouraud dextrose (SD) medium containing penicillin (25 U/mL) and streptomycin (25 U/mL) and was incubated for 2 d at 37°C. Colonies were picked and purified by three rounds of re-streaking on fresh SD agar. Colony isolate NCYC 4418 was identified as *C. parapsilosis* by ITS1 sequencing using primers ITS1F and ITS2 (8, 9). The ITS1 sequence was deposited in GenBank (accession number OZ120394.1).

Genomic DNA was extracted from a 10-mL stationary phase SD culture using a MasterPure yeast DNA purification kit (Cambio) with additional zymolyase and proteinase K treatment steps. Illumina short-read sequencing used a modified 20-fold dilution of DNA Prep (Flex) reagent and run on a NextSeq 500 sequencer, producing 12,162,590 paired-end 150-bp reads (~134× coverage). A Nanopore library was prepared from non-sheared DNA (not size selected), and sequencing was performed using a MinION sequencer (Oxford Nanopore Technologies, ONT), ligation sequencing SQK-LSK109 kit (ONT), and flow cell FLO-MIN106 R9.4.1 (ONT). This produced a total of 180,907 reads with a N_{50} of 17,699 bp (~58× coverage). Default parameters were used for all software unless otherwise specified. Base calling was performed using Guppy (ONT; v3.6.0) in a high-accuracy mode (model dna_r9.4.1_450bps_hac). Raw short- and long-read polishing was performed using fastp 0.23.2 (10).

De novo assembly of the Nanopore reads was performed using Flye 2.9.1 (11), and the Illumina short reads were used for contig error correction (polishing) using BWA v0.7.17.1 (12) and Pilon v1.24 (13). Assembly statistics were generated using QUAST v5.0.2 (14). The resulting assembly comprised 14 contigs, including four chromosome-sized contigs with telomeric termini (15). The total genome size was 13,056,352 bp, with a N_{50} of 2,100,473 bp and a G+C content of 38.70%. The largest contig was 3,026,409 bp and corresponded to chromosome 2 (16). A mitochondrial genome sequence, representing

Editor Vincent Michael Bruno, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Steve A. James, steve.james@quadram.ac.uk.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 29 July 2024
Accepted 15 November 2024
Published 10 December 2024

94.7% of total length (16), was also recovered in three contigs. Genome completeness was estimated as 98.9% using BUSCO v5.5.0 (17) with the saccharomycete_odb10 data set. Augustus v3.2.3 (18) predicted 5,587 protein-coding genes using the *Candida albicans* training data set, and 105 tRNA genes (91 nuclear encoded) were detected using tRNAscan SE 2.0 (19). Sequencing coverage was assessed using SAMtools and SAMtools depth (20), with the primate isolate found to carry 10 copies of *ARR3*, a putative arsenate transporter-encoding gene frequently amplified in copy number in both human and environmental isolates (21).

ACKNOWLEDGMENTS

This research was funded in part by a Biotechnology and Biological Sciences Research Council (BBSRC) awarded Core Capability Grant BB/CCG1860/1, project grant BB/N000250/1, and Institute Strategic Programme Grant Gut Microbes and Health (BB/R012490/1) and its constituent projects BBS/E/F/000PR10353, BBS/E/F/000PR10355.

All procedures were conducted under the authority and in compliance with UK Home Office project license number PD28B8ED5, and were approved by UKHSA Porton Down Establishment Animal Welfare and Ethical Review Body.

AUTHOR AFFILIATIONS

[1]Food, Microbiome and Health Programme, Quadram Institute Bioscience, Norwich, United Kingdom
[2]UK Health Security Agency, Porton Down, Salisbury, United Kingdom
[3]Norwich Medical School, University of East Anglia, Norwich, United Kingdom

AUTHOR ORCIDs

Steve A. James http://orcid.org/0000-0003-2292-563X
Simon R. Carding http://orcid.org/0000-0002-2383-9701

FUNDING

Funder	Grant(s)	Author(s)
UKRI | Biotechnology and Biological Sciences Research Council (BBSRC)	BB/CCG1860/1	Simon R. Carding
UKRI | Biotechnology and Biological Sciences Research Council (BBSRC)	BB/N000250/1	Simon R. Carding
UKRI | Biotechnology and Biological Sciences Research Council (BBSRC)	BB/R012490/1	Simon R. Carding
UKRI | Biotechnology and Biological Sciences Research Council (BBSRC)	BBS/E/F/000PR10353	Simon R. Carding
UKRI | Biotechnology and Biological Sciences Research Council (BBSRC)	BBS/E/F/000PR10355	Simon R. Carding

AUTHOR CONTRIBUTIONS

Steve A. James, Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft | Aimee Parker, Investigation, Methodology | Catherine Purse, Investigation | Simon G. P. Funnell, Resources.

DATA AVAILABILITY

This whole-genome shotgun project has been deposited in DDJB/ENA/GenBank (BioProject number PRJEB77687, assembly accession number CAXMZS000000000.1). The version described in this paper is version 1. The raw reads are available in the NCBI Sequence Read Archive (SRA) under accession numbers SRR29413522 (ONT) and SRR29413523 (Illumina).

REFERENCES

1. Lachance M-A, Boekhout T, Scorzetti G, Fell JW, Kurtzman CP. 2011. *Candida* berkhout,p 987–1278. In Kurtzman CP, Fell JW, Boekhout T (ed), The yeasts; a taxonomic study, 5th ed. Elsevier.
2. 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
3. Strati F, Di Paola M, Stefanini I, Albanese D, Rizzetto L, Lionetti P, Calabrò 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
4. Schei K, Avershina E, Øien T, Rudi K, Follestad T, Salamati S, Ødegård RA. 2017. Early gut mycobiota and mother-offspring transfer. Microbiome 5:107. https://doi.org/10.1186/s40168-017-0319-x
5. Trofa D, Gácser A, Nosanchuk JD. 2008. *Candida parapsilosis*, an emerging fungal pathogen. Clin Microbiol Rev 21:606–625. https://doi.org/10.1128/CMR.00013-08
6. James SA, Parker A, Purse C, Telatin A, Baker D, Holmes S, Durham J, Funnell SGP, Carding SR. 2022. The cynomolgus macaque intestinal mycobiome is dominated by the *Kazachstania* genus and *K. pintolopesii* species. J Fungi 8:1054. https://doi.org/10.3390/jof8101054
7. Hamad I, Keita MB, Peeters M, Delaporte E, Raoult D, Bittar F. 2014. Pathogenic eukaryotes in gut microbiota of western lowland gorillas as revealed by molecular survey. Sci Rep 4:6417. https://doi.org/10.1038/srep06417
8. 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
9. White TJ, Bruns TD, Lee SL, Taylor JW. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogeneticsp 315–322. In Innis MA, Gelfand DH, Sninsky JJ (ed), PCR protocols: a guide to methods and applications. Academic Press, San Diego.
10. 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
11. 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
12. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
13. 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
14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086
15. Tóth R, Nosek J, Mora-Montes HM, Gabaldon T, Bliss JM, Nosanchuk JD, Turner SA, Butler G, Vágvölgyi C, Gácser A. 2019. *Candida parapsilosis*: from genes to the bedside. Clin Microbiol Rev 32:38. https://doi.org/10.1128/CMR.00111-18
16. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. 2017. The *Candida* Genome Database (CGD): incorporation of assembly 22, systematic identifiers and visualization of high throughput sequencing data. Nucleic Acids Res 45:D592–D596. https://doi.org/10.1093/nar/gkw924
17. Manni M, Berkeley MR, Seppey M, Simão 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
18. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics 24:637–644. https://doi.org/10.1093/bioinformatics/btn013
19. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
21. Bergin SA, Zhao F, Ryan AP, Müller CA, Nieduszynski CA, Zhai B, Rolling T, Hohl TM, Morio F, Scully J, Wolfe KH, Butler G. 2022. Systematic analysis of copy number variations in the pathogenic yeast *Candida parapsilosis* identifies a gene amplification in *RTA3* that is associated with drug resistance. mBio 13:e0177722. https://doi.org/10.1128/mbio.01777-22