AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Food Microbiology | Announcement

Draft genome sequence of a kombucha isolate of *Pichia kudriavzevii*

Steve A. James,[1,2] Nicolle Som,[2,3] Adrian Turner,[2,4] David J. Baker,[1] Arjan Narbad[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Pichia kudriavzevii* (syn. *Candida krusei*) is a fermentative yeast of importance to both the food and biotechnology industries, as well as being of significant clinical concern. Here, we report the draft genome sequence of *P. kudriavzevii* NCYC 4488, a kombucha-derived isolate.

KEYWORDS fungi, yeasts, *Pichia kudriavzevii*, *Candida krusei*, fermented foods, probiotics, genome sequencing

Pichia kudriavzevii (syn. *Candida krusei*) is an ascomycetous yeast widely distributed in nature (1). Often found in spontaneous fermentations, it is used to produce a variety of traditional fermented drinks and beverages (1, 2), is a candidate probiotic (3), and has gained an increasing role in biotechnology (4, 5). Furthermore, in a clinical setting, it represents a notable opportunistic human pathogen (6, 7). Here, we combined short- and long-read sequencing to assemble the genome sequence of *P. kudriavzevii* NCYC 4488, isolated from a UK-produced organic kombucha. A 10 mL sample was 10-fold concentrated by centrifugation (10 mins and 4,000 rpm), plated onto Brain Heart Infusion (BHI) agar and incubated for 2 days at 30°C. Colonies were picked and purified by two further rounds of re-streaking on fresh BHI agar (incubated at 30°C). Colony isolate NCYC 4488 was identified as *P. kudriavzevii* by LSU D1/D2 sequencing using primers NL1 and NL4 (8). The D1/D2 sequence was deposited in GenBank (accession number PV526257.1).

Genomic DNA was isolated using a MasterPure yeast DNA purification kit (Cambio) with additional zymolyase and proteinase K treatment steps from a 2-day liquid Sabouraud dextrose culture grown at 30°C. A short-read library was prepared using a Nextera DNA Flex library prep kit (Illumina) with 20-fold diluted library reagents. Sequencing was performed on a NextSeq 500 sequencer, producing 7,997,848 paired-end 150 bp reads (~111× coverage). A Nanopore library was prepared from the same gDNA prep, without shearing or size selection, using the 1D ligation kit (catalog number SQK-LSK109, Oxford Nanopore Technologies, ONT), and loaded onto a FLO-MIN106 R9.4.1 flow cell (ONT) on a MinION instrument (ONT). This produced a total of 778,218 reads with an average read length of 4,687 bases (~337× coverage). Default parameters were used for all software unless otherwise specified. Base calling was performed using Guppy (ONT; v.3.6.0) in high-accuracy mode (model dna_r9.4.1_450bps_hac). Raw short- and long-read polishing was performed using fastp 0.23.2 (9).

De novo assembly with Nanopore reads was performed using Flye 2.9.1 (10), using long reads in "ONT regular reads, pre-Guppy5" mode. The assembly was polished with the Illumina data using three rounds of alignment with BWA-MEM v.0.7.17.1 (11) and Pilon v.1.24 (12) correction. Assembly statistics were generated using QUAST v.5.0.2 (13). The assembly consists of 10 contigs, including 4 chromosome-sized contigs, each terminating in either a 28 bp telomeric repeat sequence (5′-TTACAATATGAACTAGGAG CGAGGTGTG-3′)$_n$ (14), identified using Tandem Repeats Finder v.4.09 (15), or an rDNA array (detected by barrnap v.0.9) (16). Minimap v.2.28 (17) was used to align these to

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

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 6 May 2025
Accepted 30 July 2025
Published 11 August 2025

the reference assembly (CBS 573; GenBank assembly GCA_003054445.1) and determine they corresponded to chromosomes 1, 2, 3, and 5 (14). The total length is 10,881,549 bp, with an N_{50} value of 2,747,939 bp, and a G + C content of 38.40%. Genome completeness was estimated as 96.2% using BUSCO v.5.5.0 (18) with the saccharomycete_odb10 data set. Augustus v.3.2.3 (19) predicted 5,063 protein-coding genes using the *Saccharomyces cerevisiae* training data set, and 175 tRNA genes were detected using tRNAscan SE 2.0 (20).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Institute Strategic Programme Food Microbiome and Health BB/X011054/1 and its constituent project BBS/E/F/000PR13631.

AUTHOR AFFILIATIONS

[1]Food, Microbiome and Health Programme, Quadram Institute Bioscience, Norwich, United Kingdom
[2]Centre for Microbial Interactions, Quadram Institute Bioscience, Norwich, United Kingdom
[3]Microbes in the Food Chain, Quadram Institute Bioscience, Norwich, United Kingdom
[4]National Collection of Yeast Cultures, Quadram Institute Bioscience, Norwich, United Kingdom

AUTHOR ORCIDs

Steve A. James http://orcid.org/0000-0003-2292-563X

FUNDING

Funder	Grant(s)	Author(s)
Biotechnology and Biological Sciences Research Council	BB/X011054/1	Steve A. James Nicolle Som Adrian Turner David J. Baker Arjan Narbad
Biotechnology and Biological Sciences Research Council	BBS/E/F/000PR13631	Steve A. James Nicolle Som Adrian Turner David J. Baker Arjan Narbad

AUTHOR CONTRIBUTIONS

Steve A. James, Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Nicolle Som, Investigation, Methodology | Adrian Turner, Investigation, Methodology | David J. Baker, Investigation, Methodology | Arjan Narbad, Funding acquisition, Writing – review and editing

DATA AVAILABILITY

This whole-genome shotgun project has been deposited in DDJB/ENA/GenBank (BioProject number PRJNA1251030, assembly accession number JBNDHO000000000.1). The version described in this paper is version 1. The raw reads are available in the NCBI Sequence Read Archive (SRA) under accession numbers SRX28403477 (ONT) and

SRX28403476 (Illumina). The predicted protein and tRNA annotation data sets have been deposited in Zenodo (https://doi.org/10.5281/zenodo.16269560).

REFERENCES

1. Kurtzman CP. 2011. Pichia E. C. Hansen. In KurtzmanCP, FellJW, BoekhoutT (ed), The yeasts; a taxonomic study, Fifth ed. Vol. 2. Elsevier.
2. Chu Y, Li M, Jin J, Dong X, Xu K, Jin L, Qiao Y, Ji H. 2023. Advances in the application of the non-conventional yeast *Pichia kudriavzevii* in food and biotechnology industries. J Fungi (Basel) 9:170. https://doi.org/10.3390/jof9020170
3. Chelliah R, Ramakrishnan SR, Prabhu PR, Antony U. 2016. Evaluation of antimicrobial activity and probiotic properties of wild-strain *Pichia kudriavzevii* isolated from frozen idli batter. Yeast 33:385–401. https://doi.org/10.1002/yea.3181
4. Mukherjee V, Radecka D, Aerts G, Verstrepen KJ, Lievens B, Thevelein JM. 2017. Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. Biotechnol Biofuels 10:216. https://doi.org/10.1186/s13068-017-0899-5
5. Xiao H, Shao Z, Jiang Y, Dole S, Zhao H. 2014. Exploiting Issatchenkia orientalis SD108 for succinic acid production. Microb Cell Fact 13:121. https://doi.org/10.1186/s12934-014-0121-4
6. Pfaller MA, Diekema DJ, Gibbs DL, Newell VA, Ellis D, Tullio V, Rodloff A, Fu W, Ling TA, Global Antifungal Surveillance Group. 2010. Results from the ARTEMIS DISK Global Antifungal Surveillance Study, 1997 to 2007: a 10.5-year analysis of susceptibilities of *Candida* species to fluconazole and voriconazole as determined by CLSI standardized DISK diffusion. J Clin Microbiol 48:1366–1377. https://doi.org/10.1128/JCM.02117-09
7. Turner SA, Butler G. 2014. The *Candida* pathogenic species complex. Cold Spring Harb Perspect Med 4:a019778. https://doi.org/10.1101/cshperspect.a019778
8. Kurtzman CP, Robnett CJ. 1998. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. Antonie Van Leeuwenhoek 73:331–371. https://doi.org/10.1023/a:1001761008817
9. 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
10. 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
11. 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
12. 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
13. 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
14. Douglass AP, Offei B, Braun-Galleani S, Coughlan AY, Martos AAR, Ortiz-Merino RA, Byrne KP, Wolfe KH. 2018. Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names. PLoS Pathog 14:e1007138. https://doi.org/10.1371/journal.ppat.1007138
15. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580. https://doi.org/10.1093/nar/27.2.573
16. Seemann T. Barrnap 0.9: rapid ribosomal RNA prediction, v0.9. https://github.com/tseemann/barrnap.
17. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
18. 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
19. 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
20. 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