AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Mycology | Announcement

Chromosome-level genome assembly of a kombucha isolate of *Zygotorulaspora florentina*

Steve A. James,[1,2] **Nerice Ryan,**[3] **Dave J. Baker,**[1] **Arjan Narbad**[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Zygotorulaspora florentina* is a fermentative yeast of growing interest to both the brewing and winemaking industries. Here, we report the draft genome sequence of *Z. florentina* NCYC 4520, a kombucha-derived isolate.

KEYWORDS fungi, fermented foods, kombucha, probiotics, *Zygotorulaspora florentina*, genome sequencing

Zygotorulaspora florentina, formerly *Zygosaccharomyces florentinus* (1), is an ascomycetous yeast that has been isolated from a variety of sources, including grape must, plant material, fruit flies, and fruit-based soft drinks (2). With its ability to ferment maltose and maltotriose, as well as enhance aroma complexity in co-fermentations with *Saccharomyces cerevisiae*, there is growing interest in its potential use in both brewing and winemaking (3–5). Here, we combined short- and long-read sequencing to assemble the genome sequence of *Z. florentina* NCYC 4520, isolated from a bottle of Norfolk-produced organic kombucha purchased in Norwich, UK. A 100 µL kombucha sample was plated onto Rose Bengal agar and incubated for 2 days at 25°C. Colonies were picked and purified by two additional rounds of re-streaking on Yeast Malt (YM) agar (10 g/L glucose, 3 g/L malt extract, 5 g/L peptone, and 3 g/L yeast extract) and incubated at 25°C. The species was identified by ITS sequencing using primers ITS1 and ITS4 (6, 7) (GenBank accession number PX480735). Sequence identity was 100% to the *Z. florentina* type strain (KY106082), based on a BLASTn query (8) against the NCBI rRNA_typestrains/ITS_RefSeq_Fungi database.

Genomic DNA was isolated using a MasterPure yeast DNA purification kit (Cambio), with additional zymolyase and proteinase K treatment steps, from cells harvested from a 2-day liquid YM culture grown at 25°C. A short-read library was prepared using a Nextera DNA Flex library prep kit (Illumina) with 20-fold diluted reagents and sequenced on an Illumina NextSeq 500, producing 11.6 million paired-end 150 bp reads. A Nanopore library was prepared from the same gDNA prep without fragmentation or size selection, using the SQK-LSK109 1D ligation kit (Oxford Nanopore Technologies, ONT), and sequenced on a MinION instrument (ONT) with a R9.4.1 flow cell (ONT). This produced 336,776 reads with an N_{50} of 15,734 bp. Default parameters were used for all software unless otherwise specified. Base calling was performed using Guppy (ONT; v3.6.0) in high-accuracy mode (model dna_r9.4.1_450bps_hac).

Illumina reads were filtered using fastp v.0.23.2 (9). Nanopore reads were filtered using NanoFilt v.2.3.0 (10), selecting reads with quality ≥10 and length ≥5,000 bp. Read quality was verified with FastQC v.0.12.1 (11). After filtration, 161,449 Nanopore reads were retained (mean length = 14,386 bp). *De novo* assembly of Nanopore reads was performed using Flye v2.9.1 (12), 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 (13) and correction with Pilon v.1.24 (14). The final assembly consists of 15 contigs, including 8 chromosome-sized contigs, 5 with a

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 26 February 2026
Accepted 30 March 2026
Published 23 April 2026

13 bp telomeric repeat sequence (5′-TGTAGGGGTGTGG-3′)$_n$ (15), identified using Tandem Repeats Finder v.4.09 (16), at one end. The total length is 11,137,782 bp, with an N_{50} of 1,404,365 bp, and a G + C content of 41.03%. Genome completeness was estimated at 99.6% using BUSCO v.5.5.0 (17) with the saccharomycete_odb10 data set. Augustus v.3.2.3 (18) predicted 5,065 protein-coding genes using the *Saccharomyces cerevisiae* training data set, and 194 tRNA genes were detected using tRNAscan-SE 2.0 (19).

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]School of Biological Sciences, University of East Anglia, 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 Nerice Ryan Dave J. Baker Arjan Narbad
Biotechnology and Biological Sciences Research Council	BBS/E/F/000PR13631	Steve A. James Nerice Ryan Dave J. Baker Arjan Narbad

AUTHOR CONTRIBUTIONS

Steve A. James, Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing – original draft | Nerice Ryan, Investigation | Dave J. Baker, Formal analysis | Arjan Narbad, Conceptualization, Funding acquisition, Project administration

DATA AVAILABILITY

This whole-genome shotgun project has been deposited in DDJB/ENA/GenBank (BioProject number PRJNA1333318, assembly accession number JBVFKW000000000). The version described in this paper is version 1. The raw reads are available in the NCBI Sequence Read Archive (SRA) under accession numbers SRX30640826 (ONT) and SRX30640825 (Illumina). The predicted protein and tRNA annotation data sets have been deposited in Zenodo (https://doi.org/10.5281/zenodo.18151368).

REFERENCES

1. Kurtzman CP. 2003. Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora. FEMS Yeast Res 4:233–245. https://doi.org/10.1016/S1567-1356(03)00175-2

2. Kurtzman CP. 2011. Zygotorulaspora Kurtzman, p 949–951. *In* Kurtzman CP, Fell JW, Boekhout T (ed), The yeasts; a taxonomic study, Fifth. Elsevier.
3. Lencioni L, Romani C, Gobbi M, Comitini F, Ciani M, Domizio P. 2016. Controlled mixed fermentation at winery scale using *Zygotorulaspora florentina* and *Saccharomyces cerevisiae*. Int J Food Microbiol 234:36–44. https://doi.org/10.1016/j.ijfoodmicro.2016.06.004
4. Canonico L, Galli E, Ciani E, Comitini F, Ciani M. 2019. Exploitation of three non-conventional yeast species in the brewing process. Microorganisms 7:11. https://doi.org/10.3390/microorganisms7010011
5. Nikulin J, Eerikäinen R, Hutzler M, Gibson B. 2020. Brewing characteristics of the maltotriose-positive yeast *Zygotorulaspora florentina* isolated from Oak. Beverages 6:58. https://doi.org/10.3390/beverages6040058
6. 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
7. 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.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
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. 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
11. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
12. 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
13. 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
14. 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
15. Krause DJ, Kominek J, Opulente DA, Shen X-X, Zhou X, Langdon QK, DeVirgilio J, Hulfachor AB, Kurtzman CP, Rokas A, Hittinger CT. 2018. Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts. Proc Natl Acad Sci USA 115:11030–11035. https://doi.org/10.1073/pnas.1806268115
16. 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
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