

# Genome sequence of the yeast *Candida solani* UCD1087, isolated from soil in Ireland

Matthieu Osborne,<sup>1</sup> Conor Hession,<sup>2</sup> Shareen Behnass,<sup>2</sup> Christina Biju,<sup>2</sup> John Dennison,<sup>2</sup> Adam Faulkner,<sup>2</sup> Anthony Flynn,<sup>2</sup> Joshua Haugh,<sup>2</sup> Sebastian Magner,<sup>2</sup> Sean O'Reilly,<sup>2</sup> Reema Saoud,<sup>2</sup> Tadhg Ó Cróinín,<sup>2</sup> Kenneth H. Wolfe,<sup>1</sup> Geraldine Butler,<sup>2</sup> Kevin P. Byrne<sup>1,2</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** *Candida solani* is a member of the Wickerhamomyces clade of budding yeasts. We present the genome sequence of *C. solani* strain UCD1087, which was isolated from soil on the University College Dublin (UCD) campus in Ireland. This genome is 12.85 Mb and was assembled into six chromosome-sized contigs plus a mitochondrial genome contig.

**KEYWORDS** yeasts, genome analysis

*Candida solani* is an asexual species that is a member of the Wickerhamomyces clade of budding yeasts (1). It was first isolated from a potato starch mill in the Netherlands (2). It has no known food or biotechnological applications. As it does not grow above 35°C, it is unlikely to have pathogenic capacity in humans. The only other genome assembly available for this species is from the type strain (NRRL Y-2224), which is fragmented into over 9,000 contigs (1).

*C. solani* UCD1087 was isolated from soil collected on the University College Dublin (UCD) campus, Ireland (GPS coordinates 53.3085254, -6.2220592) as part of an undergraduate research module (3).

Soil material was passaged twice at room temperature in 9 mL liquid yeast extract–peptone–dextrose (YPD) containing chloramphenicol (30 µg/mL) and ampicillin (100 µg/mL) and cultured on YPD agar plates. A single colony was isolated. The species was identified by PCR and Sanger sequencing of the ribosomal DNA internal transcribed spacer (ITS) and D1/D2 regions (accession numbers [PQ438597](#) and [PQ438670](#)), using primers ITS1 (TCCGTAGGTGAACCTGCGG) and ITS4 (TCCTCCGCTTATTGATATGC) (4), and for the D1/D2 region, NL1 (GCATATCAATAAGCGGAGGAA) and NL4 (GGTCCGTGTTCAAGACGG) (5). Sequence identity was 100% (554/554 bp) in the ITS and 98.94% (559/565 bp) in the D1/D2 region to the type strain of *C. solani* (accession numbers [KY102402](#) and [NG\\_055191](#)).

DNA was isolated by phenol/chloroform extraction from liquid YPD cultures grown at 20°C. Short-read library construction (Illumina DNA-Prep(M) Ref:20060060) and sequencing (UCD Conway Core Facility, Dublin, Ireland) used a NextSeq2000 and P1 flowcell, yielding 5.6 million read pairs (2 × 150 bp). Adapters and low-quality reads were removed (Skewer v0.2.2) (6). For long-read sequencing, after DNA extraction (Biosearch Technology Masterpure yeast DNA purification kit MPY80010), library preparation (Native Barcoding Kit SQK-NBD112-24), end-repair (NEB-M6630), and ligation (NEB-E6056), we selected for DNA >3 kb with long fragment buffer before Oxford Nanopore sequencing (MinION MK1C, flowcell FLO-MIN112 R10.4). Default fast basecalling was done on the instrument (MinKNOW v23.07.12; default demultiplexing; barcode trimming; reads ≥500 bp kept; no quality cut-off). NanoFilt (v2.3.0) (7) retained 98,802 reads with quality ≥10 and length ≥10,000 bp (reads N50 = 18,304 bp), which were then assembled using Canu (v2.2) (8), followed by five rounds of error correction using NextPolish (v1.4.1)

**Editor** Julie C. Dunning Hotopp, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Kevin P. Byrne, [kevin.byrne@ucd.ie](mailto:kevin.byrne@ucd.ie).

The authors declare no conflict of interest.

**Received** 23 October 2024

**Accepted** 17 March 2025

**Published** 28 March 2025

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

TABLE 1 Genome assembly characteristics of *Candida solani* UCD1087

Chromosome	Size	Left end	Right end
1	3,520,480	rDNA array	Telomere
2	2,242,305	rDNA array	rDNA array
3	2,132,486	Telomere	Telomere
4	1,914,830	Telomere	rDNA Telomere
5	1,507,218	rDNA array	Telomere
6	1,498,922	Telomere	Telomere
Mitochondrial	37,835	n/a (circular)	n/a (circular)

with the Illumina reads (9). Default parameters were used, except where otherwise noted.

The assembly consists of six nuclear contigs (total 12.85 Mb; N50 = 2,132,486 bp) and the mitochondrial genome (37,835 bp circular unit; accession number [CAXWWC010000007](https://www.ncbi.nlm.nih.gov/nuccore/CAXWWC010000007)). All six nuclear contigs appear to be complete chromosomes because every end of every contig is either a probable telomere repeat sequence (GG GATGTA CTGTGTGGTGA)<sub>n</sub> (10) or an rDNA array (detected by BLASTN) (Table 1). All rDNA arrays are oriented with the 18S and 26S rRNA genes transcribed towards the nearby contig end.

The UCD1087 genome assembly has an average nucleotide identity (11) of 96.26% to the assembly of the *C. solani* type strain NRRL Y-2224 (accession no. [JAKVQS010000000](https://www.ncbi.nlm.nih.gov/nuccore/JAKVQS010000000)) (1). Using BUSCO v5.1.2 (12), genome completeness was estimated as 93.67% compared with the Ascomycota lineage data set. G + C content is 39.63%.

## ACKNOWLEDGMENTS

This work was supported by undergraduate teaching resources from University College Dublin and by Science Foundation Ireland (20/FFP-A/8795). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## AUTHOR AFFILIATIONS

<sup>1</sup>School of Medicine, Conway Institute, University College Dublin, Belfield, Ireland

<sup>2</sup>School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Ireland

## AUTHOR ORCIDs

Kenneth H. Wolfe  <http://orcid.org/0000-0003-4992-4979>

Geraldine Butler  <http://orcid.org/0000-0002-1770-5301>

Kevin P. Byrne  <http://orcid.org/0000-0002-6368-154X>

## DATA AVAILABILITY

This whole-genome shotgun project has been deposited as DDBJ/ENA/GenBank accession number [CAXWWC010000000](https://www.ncbi.nlm.nih.gov/nuccore/CAXWWC010000000) (BioProject no. [PRJEB80645](https://www.ncbi.nlm.nih.gov/bioproject/PRJEB80645)). The version described in this paper is version 1. The reads were deposited at ENA (accessions [ERR13731958](https://www.ncbi.nlm.nih.gov/nuccore/ERR13731958), [ERR13731962](https://www.ncbi.nlm.nih.gov/nuccore/ERR13731962)). The ITS sequence is [PQ438597](https://www.ncbi.nlm.nih.gov/nuccore/PQ438597). The D1/D2 sequence is [PQ438670](https://www.ncbi.nlm.nih.gov/nuccore/PQ438670). Isolate UCD1087 has been deposited in the CBS and PYCC culture collections as CBS 18653 and PYCC 10020.

## REFERENCES

- Opulente DA, LaBella AL, Harrison M-C, Wolters JF, Liu C, Li Y, Kominek J, Steenwyk JL, Stoneman HR, VanDenAvond J, et al. 2024. Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science* 384:eadj4503. <https://doi.org/10.1126/science.adj4503>
- Lodder J, Kreger-van Rij NJW. 1952. The Yeasts: a taxonomic study

3. Bergin SA, Allen S, Hession C, Ó Cinnéide E, Ryan A, Byrne KP, Ó Cróinín T, Wolfe KH, Butler G. 2022. Identification of European isolates of the lager yeast parent *Saccharomyces eubayanus*. *FEMS Yeast Res* 22:foac053. <https://doi.org/10.1093/femsyr/foac053>
4. 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>
5. Leaw SN, Chang HC, Sun HF, Barton R, Bouchara J-P, Chang TC. 2006. Identification of medically important yeast species by sequence analysis of the internal transcribed spacer regions. *J Clin Microbiol* 44:693–699. <https://doi.org/10.1128/JCM.44.3.693-699.2006>
6. Jiang H, Lei R, Ding SW, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182. <https://doi.org/10.1186/1471-2105-15-182>
7. 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>
8. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
9. Chen Z, Erickson DL, Meng J. 2021. Polishing the Oxford Nanopore long-read assemblies of bacterial pathogens with Illumina short reads to improve genomic analyses. *Genomics* 113:1366–1377. <https://doi.org/10.1016/j.ygeno.2021.03.018>
10. Cohn M, McEachern MJ, Blackburn EH. 1998. Telomeric sequence diversity within the genus *Saccharomyces*. *Curr Genet* 33:83–91. <https://doi.org/10.1007/s002940050312>
11. Rodriguez-R LM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species. *Microbe* 9:111–118. <https://doi.org/10.1128/microbe.9.111.1>
12. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>