

Genome sequence of the yeast *Candida solani* UCD2211 isolated from soil in Ireland

Padraic G. Heneghan,¹ Róisín O'Neill,² Uliana Osinovskaia,² Adam P. Ryan,² Edvin Sebastian,² Fiona Silitrari-Chen,² Alexander Vial,² Isabella Voza,² Vincent Walsh Giguere,² Tadhg Ó. Cróinín,² Kenneth H. Wolfe,¹ Geraldine Butler,² Kevin P. Byrne^{1,2}

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 UCD2211 isolated from soil on the University College Dublin (UCD) campus in Ireland. This genome was 12.67 Mb and 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. Two other genome assemblies are available: from type strain NRRL Y-2224 fragmented in over 9,000 contigs (1) and a chromosome-level assembly from a related strain UCD1087 (3).

C. solani UCD2211 was isolated from soil collected on the University College Dublin (UCD) Campus in Dublin, Ireland (GPS coordinates 53.308583, -6.222999) as part of an undergraduate research module (4). 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. The species was identified by ribosomal DNA as previously described (3). Sequence identity was 96.98% (546/563 bp) in the ITS and 99.46% (556/559 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 room temperature. Short-read sequencing using a NovaSeq X Plus (Novogene [UK] Company, Ltd.) yielded 17.6 million read pairs (2 × 150 bp). Adapters and low-quality reads were removed (Skewer v0.2.2) (5). For long-read sequencing, after DNA extraction (MasterPure Yeast DNA Purification Kit MPY80010) and library preparation (Native Barcoding Kit SQK-NBD114-24), we selected for DNA >3 kb with long fragment buffer before Oxford Nanopore sequencing (MinION MK1B, flowcell FLO-MIN114), and super accurate basecalling (MinKNOW v24.06.16; default demultiplexing; barcode trimming; reads ≥ 500 bp kept; no quality cut-off). NanoFilt (v2.3.0) (6) retained 184,681 reads with quality ≥10 and length ≥1,000 bp (reads N50 = 6,932 bp), which were then assembled using Canu (v2.2) (7), followed by five rounds of error correction using NextPolish (v1.4.1) with the Illumina reads (8). Default parameters were used, except where otherwise noted.

The assembly consists of six nuclear contigs (total 12.63 Mb; N50 = 2,190,064 bp) and the mitochondrial genome (38,626 bp circular unit; accession number [CDRNKE010000007](#)). All six nuclear contigs appear to be complete chromosomes because every end of every contig is either a probable telomere repeat sequence (TG GTGTAGGGATGTACTGTG)_n (9) or an rDNA array (detected by BLASTN) (Table 1). All rDNA arrays are oriented with the 18S and 26S rRNA genes transcribed toward the nearby

Editor Jason E. Stajich, University of California
Riverside, Riverside, California, USA

Address correspondence to Kevin P. Byrne,
kevin.byrne@ucd.ie.

The authors declare no conflict of interest.

Received 3 December 2025

Accepted 27 January 2026

Published 9 February 2026

Copyright © 2026 Heneghan 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* UCD2211

Chromosome	Size (bp)	Left end	Right end	UCD1087 comparison
1	2,692,170	rDNA	rDNA	chr1 + part of chr4 (reciprocal translocation)
2	2,640,582	Telomere → rDNA	Telomere	chr4 + part of chr1 (reciprocal translocation)
3	2,190,064	rDNA	rDNA	chr2
4	2,155,114	rDNA	Telomere	chr3
5	1,523,230	Telomere	rDNA	chr6
6	1,426,753	Telomere	Telomere	chr5 (inversion on this chr)
mitochondrial	38,626	n/a (circular)	n/a (circular)	

contig end. The genome is collinear with that of UCD1087 (3), except for a reciprocal translocation and an inversion (Table 1).

The UCD2211 genome assembly has an average nucleotide identity (10) of 97.69% to the assembly of type strain NRRL Y-2224 ([JAKVQS010000000](https://doi.org/10.1093/mra/01168-24)) (1) and 93.2% to the assembly of UCD1087 ([CAXWWC010000000](https://doi.org/10.1093/mra/01168-24)), which we previously isolated (3), indicating the sequence diversity within *C. solani*. Using BUSCO v5.8.2 (11), genome completeness was estimated as 93.6% compared to the *Ascomycota* lineage data set. The guanine or cytosine (G + C) content was 39.86%.

ACKNOWLEDGMENTS

This work was supported by undergraduate teaching resources from University College Dublin and by Research Ireland (24/FFP-A/12509). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

AUTHOR AFFILIATIONS

¹School of Medicine, Conway Institute, University College Dublin, Dublin, Ireland
²School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin, 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 [CDRNKE010000000](https://doi.org/10.1093/mra/01168-24) (BioProject no. [PRJEB97773](https://doi.org/10.1093/mra/01168-24)). The version described in this paper is version 1. The reads were deposited at ENA (accessions [ERR15652871](https://doi.org/10.1093/mra/01168-24), [ERR15761027](https://doi.org/10.1093/mra/01168-24)). The ITS sequence is [PX481655](https://doi.org/10.1093/mra/01168-24). The D1/D2 sequence is [PX481656](https://doi.org/10.1093/mra/01168-24).

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*. North-Holland Publishing Company, Amsterdam.
- Osborne M, Hession C, Behnass S, Biju C, Dennison J, Faulkner A, Flynn A, Haugh J, Magner S, O'Reilly S, Saoud R, Ó Cróinín T, Wolfe KH, Butler G, Byrne KP. 2025. Genome sequence of the yeast *Candida solani* UCD1087, isolated from soil in Ireland. *Microbiol Resour Announc* 14. <https://doi.org/10.1128/mra.01168-24>
- 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. <https://doi.org/10.1093/femsyr/foac053>

5. 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>
6. 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>
7. 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>
8. 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>
9. 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>
10. Yoon SH, Ha SM, Lim JM, Kwon SJ, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>
11. 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>