





Draft Genome Sequence of the Yeast *Blastobotrys aristata* Strain UCD613, Isolated from Soil in Ireland

Aida Don,^a Kevin P. Byrne,^b Fouz Alqaderi,^a Alexandra Mazhova,^a Ben O'Leary Chaney,^a Sarah Redmond,^a Stephen Allen,^a Jiaran Gao,^b Eoin Ó Cinnéide,^b Sean A. Bergin,^a Conor Hession,^a  Kenneth H. Wolfe,^b  Geraldine Butler^a

^aSchool of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Ireland

^bSchool of Medicine, Conway Institute, University College Dublin, Belfield, Ireland

ABSTRACT *Blastobotrys aristata* is a member of the Trichomonascaceae family in the order Saccharomycetales. Here, we present the genome sequence of *B. aristata* UCD613, which was isolated from soil in Dublin, Ireland. This genome is 13.3 Mb and was assembled into 4 chromosome-size scaffolds of >2.2 Mb in size plus a mitochondrial genome scaffold.

B *blastobotrys aristata* was first isolated from moldy plaster in the former Czechoslovakia in 1976 (Marvanova 1976, also known as *B. aristatus*) (1). We identified isolate *B. aristata* UCD613 from soil collected from the campus of University College Dublin (GPS coordinates 53.3034961, -6.2131910). Soil material was passaged twice in in 9 mL liquid yeast extract-peptone-dextrose (YPD) containing chloramphenicol (30 μ g/mL) and ampicillin (100 μ g/mL) and cultured on YPD plates at 30°C.

The species was identified from single colonies by PCR amplification and Sanger sequencing of the internal transcribed spacer (ITS) (OP221981) and D1/D2 (OP221771) regions of its ribosomal DNA (rDNA) locus. The D1/D2 region was 100% identical to that of the type strain of *B. aristata* (2) (DQ442686.1). No other ITS sequence is available.

For short-read sequencing, total genomic DNA was extracted from a YPD culture using phenol-chloroform-isoamyl alcohol and dissolved in 150 μ L water (3). Libraries were generated and sequenced by BGI Tech Solutions (Hong Kong). One microgram of DNA was fragmented using Covaris, size selected (200 to 400 bp) using magnetic beads, end repaired, and 3' adenylated, and primers were ligated. Fragments were amplified by PCR and heat denatured and circularized using the splint oligonucleotide sequence. The library was amplified with ϕ 29 DNA polymerase to make DNA nanoballs (DNBs). The DNBs were loaded on a patterned nanoarray, and 150 bases were sequenced from each end using combinatorial probe-anchor synthesis (cPAS) on a DNBSeg-G400, yielding ~6.1 million read pairs. Default parameters were used unless noted. Adapters and low-quality reads were removed first using SOAPnuke (4) and subsequently using Skewer v.0.2.2 (5). For long-read sequencing, genomic DNA was prepared using a Genomic Tip 100G kit (Qiagen). Two libraries were generated using the SQK-RBK004 kit from Oxford Nanopore Technologies (ONT) and cleaned with AMPure XP magnetic beads. Libraries were sequenced on primed R9.4.1 flow cells using MinKNOW v.4.1.22 on a MinION device. From run 1, raw data were base called using Guppy v.4.2.2 +effba8 (using the fast model [dna_r9.4.1_450bps_fast.cfg]) (ONT) and demultiplexed using qcat v.1.1.0 (ONT) with default settings. For the second run, Guppy v.4.2.2 +effba8 was used both for base calling and demultiplexing. Both sets of reads were concatenated together for downstream processing. NanoFilt v.2.3.0 (6) was used to select reads (minimum quality, ≥ 7 ; minimum length, $\geq 1,000$ bp) which retained 107,000 reads with an N_{50} of 6,639 bp.

The genome was assembled from the long reads using Canu v.2.2 (7), followed by five rounds of error correction with the DNBSeg short reads using NextPolish (8). Five

Editor Antonis Rokas, Vanderbilt University

Copyright © 2022 Don et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Geraldine Butler, gbutler@ucd.ie.

The authors declare no conflict of interest.

Received 9 September 2022

Accepted 24 September 2022

Published 12 October 2022

contigs of <45 kb (corresponding to rDNA and parts of the mitochondrial genome) were removed, leaving 4 chromosome-size contigs of >2.2 Mb in size and a circular mitochondrial genome (48,582 bp, manually edited; accession no. [OX291664.1](https://doi.org/10.1093/mra/00957-22)). The total size of the genome is 13.3 Mb, the N_{50} value is 3.5 Mb, the L_{50} value is 2 contigs, and the G+C content is 48%. The largest contig is 4.2 Mb. Using BUSCO v.5.1.2, genome completeness was estimated at 94.8% (compared to the Ascomycota lineage data set).

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank (BioProject no. [PRJEB55420](https://doi.org/10.1093/mra/00957-22)). The version described in this paper is version 1. The raw reads were deposited at SRA (accession no. [ERX9629577](https://doi.org/10.1093/mra/00957-22), [ERX9629578](https://doi.org/10.1093/mra/00957-22), and [ERX9629579](https://doi.org/10.1093/mra/00957-22)). The ITS sequence is at [OP221981](https://doi.org/10.1093/mra/00957-22) and the D1/D2 region sequence at [OP221771](https://doi.org/10.1093/mra/00957-22).

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.

REFERENCES

1. Marvanova L. 1976. Two new *Blastobotrys* species. *Trans Br Mycol Soc* 66: 217–222. [https://doi.org/10.1016/S0007-1536\(76\)80049-6](https://doi.org/10.1016/S0007-1536(76)80049-6).
2. Kurtzman CP, Robnett CJ. 1995. Molecular relationships among hyphal ascomycetous yeasts and yeastlike taxa. *Can J Bot* 75:S1.
3. Dymond JS. 2013. Preparation of genomic DNA from *Saccharomyces cerevisiae*. *Methods Enzymol* 529:153–160. <https://doi.org/10.1016/B978-0-12-418687-3.00012-4>.
4. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. 2018. SOAPnucle: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7:1–6. <https://doi.org/10.1093/gigascience/gix120>.
5. Jiang H, Lei R, Ding S-W, 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>.