



Genome Sequence of the Black Yeast-Like Strain *Aureobasidium pullulans* var. *aubasidani* CBS 100524

Gabriel A. Vignolle,^a Robert L. Mach,^a Astrid R. Mach-Aigner,^a  Christian Derntl^a

^aInstitute of Chemical, Environmental and Bioscience Engineering, TU Wien, Vienna, Austria

ABSTRACT In this work, we present the whole-genome sequence and the complete mitochondrial sequence of the black yeast-like strain *Aureobasidium pullulans* var. *aubasidani* CBS 100524, which produces the exopolysaccharide aubasidan and was previously isolated from *Betula* sp. slime flux from the Leningrad Region of Russia.

Aureobasidium pullulans is a yeast-like ascomycete with industrial relevance due to its extracellular polysaccharides (1). The main exopolysaccharide of *A. pullulans* var. *aubasidani* strain CBS 100524 is aubasidan rather than pullulan (2, 3). This strain was previously isolated from plant exudates of a *Betula* sp. from the Leningrad Region of Russia (2). Despite the difference in the secreted extracellular polysaccharides, *A. pullulans* var. *aubasidani* strain CBS 100524 is part of a main phylogenetic group (phylogenetic difference below 0.25 based on a multilocus alignment with a bootstrap value of 100) within the *A. pullulans* species complex. This group also includes the ex-neotype strain *A. pullulans* var. *pullulans* CBS 584.75 and the sequenced strain *A. pullulans* var. *pullulans* EXF-150 (3).

A. pullulans strain CBS 100524 was cultivated in malt extract medium (30 g/liter malt extract, 1 g/liter peptone) at 24°C and 220 rpm for 24 h. The biomass was filtered through Miracloth (EMD Millipore Corp., Burlington, MA, USA), lyophilized, and stored at –20°C. Genomic DNA was extracted as described in reference 4, sheared through sonication, purified using the GeneJET PCR purification kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA), and then size selected for 800-bp fragments using NEBNext Ultra sample purification beads (New England Biolabs, Ipswich, MA, USA). The library was prepared using the NEBNext Ultra II DNA library kit with purification beads and NEBNext multiplex oligos for Illumina (index primer set 2) (both New England Biolabs) and sequenced on a MiSeq instrument using a v3 reagent kit (600 cycles, 2 × 300-bp paired-end reads) (both Illumina, Inc., San Diego, CA, USA).

The sequencing yielded 2,892,731 read pairs. First, a crude *de novo* assembly was performed using SPAdes v3.13.1 (5) with default parameters. From this initial assembly, mitochondrial sequences were identified by a BLAST analysis against the nonredundant nucleotide database (6). Next, these sequences were used as seed input for NOVOplasty v3.7 (7) for a *de novo* assembly of the mitochondrial genome sequence (one circular contig; size, 37,556 bp; coverage, 358×). Using the mitochondrial genome sequence as index built with Bowtie v1.2.2 (8), the mitochondrial reads were extracted from the raw reads. The mitochondrion-free reads were then re-paired using Fastq-pair (9), quality checked and trimmed using Trimmomatic (10), leaving 2,543,186 read pairs, and then mapped against the reference genome *A. pullulans* strain EXF-150 (GenBank accession no. [GCA_000721785.1](https://www.ncbi.nlm.nih.gov/nuccore/GCA_000721785.1)) with BWA (11) and combined and sorted using SAMtools v1.7 (12) and Picard (13). A first genome representation was extracted using ANGSD v0.925 (Analysis of Next Generation Sequencing Data) (14). The genome assembly was iteratively improved using SSPACE-Standard v3.0 (15), GapFiller v1-10 (16), and Pilon v1.21 (17). tRNA genes were detected using tRNAscan-SE v1.3.1 (18). Genes

Citation Vignolle GA, Mach RL, Mach-Aigner AR, Derntl C. 2021. Genome sequence of the black yeast-like strain *Aureobasidium pullulans* var. *aubasidani* CBS 100524. *Microbiol Resour Announc* 10:e01293-20. <https://doi.org/10.1128/MRA.01293-20>.

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2021 Vignolle 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 Christian Derntl, christian.derntl@tuwien.ac.at.

Received 10 November 2020

Accepted 23 February 2021

Published 25 March 2021

were predicted with AUGUSTUS v3.3.2 (19), trained with the reference genome *A. pullulans* strain EXF-150 according to reference 20. The assembly was masked using RepeatMasker v4.0.9 (21), based on the Dfam_3.0 database to identify repetitive elements. We used QUAST v5.0.2 (22, 23), including the fungal (fungi_odb9) Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (24), for the final evaluation.

The assembly consists of 83 scaffolds (total sequence length, 30,265,078 bp; N_{50} , 1,201,293 bp; GC content, 50.50%; mean coverage, 28 \times), and 10,978 genes (99.31% complete BUSCO genes found) and 353 tRNAs were predicted.

Data availability. The raw reads were uploaded to the Sequence Read Archive (SRA) under the accession no. [SRR12830835](https://www.ncbi.nlm.nih.gov/sra/SRR12830835). The complete genome sequence was deposited at DDBJ/ENA/GenBank under the accession no. [JADGIM000000000](https://www.ncbi.nlm.nih.gov/genbank/JADGIM000000000). The version described in this paper is version [JADGIM000000000.1](https://www.ncbi.nlm.nih.gov/genbank/JADGIM000000000.1). The complete mitochondrial genome sequence was deposited under GenBank accession no. [MW148763](https://www.ncbi.nlm.nih.gov/genbank/MW148763).

ACKNOWLEDGMENTS

This study was supported by the Austrian Science Fund (FWF) (P29556) given to R.L.M. and by TU Wien (Ph.D. program TU Wien bioactive).

REFERENCES

1. Rekha MR, Sharma CP. 2007. Pullulan as a promising biomaterial for biomedical applications: a perspective. *Trends Biomater Artif Organs* 20:116–121.
2. Yurlova NA, de Hoog GS. 1997. A new variety of *Aureobasidium pullulans* characterized by exopolysaccharide structure, nutritional physiology and molecular features. *Antonie Van Leeuwenhoek* 72:141–147. <https://doi.org/10.1023/a:1000212003810>.
3. Zalar P, Gostiñar C, de Hoog GS, Uršič V, Sudhadham M, Gunde-Cimerman N. 2008. Redefinition of *Aureobasidium pullulans* and its varieties. *Stud Mycol* 61:21–38. <https://doi.org/10.3114/sim.2008.61.02>.
4. Paun O, Turner B, Trucchi E, Munzinger J, Chase MW, Samuel R. 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. *Syst Biol* 65:212–227. <https://doi.org/10.1093/sysbio/syv076>.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
6. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
7. Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18. <https://doi.org/10.1093/nar/gkw955>.
8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
9. Edwards JA, Edwards RA. 2019. Fastq-pair: efficient synchronization of paired-end fastq files. *bioRxiv* <https://doi.org/10.1101/552885>.
10. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
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. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
13. Broad Institute. 2019. Picard toolkit. <http://broadinstitute.github.io/picard/>.
14. Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356. <https://doi.org/10.1186/s12859-014-0356-4>.
15. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
16. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with Gap-Filler. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
17. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, 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>.
18. 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>.
19. Stanke M, Morgenstern B. 2005. AUGUSTUS: a Web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33:W465–W467. <https://doi.org/10.1093/nar/gki458>.
20. Hoff KJ, Stanke M. 2019. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics* 65:e57. <https://doi.org/10.1002/cpbi.57>.
21. Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. <http://repeatmasker.org>.
22. 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>.
23. Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>.
24. Simao 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>.